**Universidade de São Paulo
Escola Superior de Agricultura Luiz de Queiroz**

**Computação bayesiana aproximada: aplicações em modelos de
dinâmica populacional**

Maria Cristina Martins

Tese apresentada para obtenção do título de Doutora em
Ciências. Área de concentração: Estatística e Experimen-
tação Agronômica

**Piracicaba
2017**

Maria Cristina Martins
Licenciada em Matemática

**Computação bayesiana aproximada: aplicações em modelos de
dinâmica populacional**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Profa. Dra. **ROSELI APARECIDA LEANDRO**

Tese apresentada para obtenção do título de Doutora em
Ciências. Área de concentração: Estatística e Experimen-
tação Agronômica

Piracicaba
2017

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Martins, Maria Cristina

Computação bayesiana aproximada: aplicações em modelos de dinâmica populacional / Maria Cristina Martins. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2017.

121 p.

Tese (Doutorado) – – USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Inferência livre de verossimilhança 2. Distribuição a posteriori 3. Estatísticas de resumo 4. Equações diferenciais estocásticas. I. Título.

DEDICATÓRIA

*Ao meu companheiro de todos
os momentos, Ricardo,
com amor dedico.*

AGRADECIMENTOS

Ao meu companheiro por sempre me apoiar e ajudar com paciência e amor. Por sua presença em minha vida tanto nos momentos bons quanto nos difíceis.

À minha família pelo apoio e confiança. Aos meus pais, que sempre acreditaram em mim, e irmãos que têm me acompanhado ao longo desses anos.

À minha orientadora de doutorado, pelas contribuições com o trabalho e pelos momentos compartilhados nos últimos anos. Ao orientador no exterior pelas colaborações e ideias para o trabalho.

Aos docentes e funcionários do LCE/ESALQ/USP, por sua dedicação, disponibilidade e conhecimentos compartilhados.

À todos os familiares e amigos que mesmo de longe, de forma especial, torcem por mim.

Aos amigos que conheci no LCE e no exterior com quem tive períodos sérios e de intenso estudo, mas também, momentos de descontração e lazer. Pessoas estas que muitas vezes me apoiaram como família tanto em momentos de alegria quanto de tristeza.

A cada pessoa, que ao longo de minha jornada, contribuiu direta ou indiretamente à minha formação.

À Deus por sempre estar presente em cada momento de minha vida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos no país e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudos no exterior.

O meu muito obrigada.

EPÍGRAFE

*“We have not succeeded in answering all our problems.
The answers we have found only serve to raise a whole set
of new questions. In some ways we feel we are as
confused as ever, but we believe we are confused on a
higher level and about more important things.”*

Citation from Bernt Øksendal’s “Stochastic Differential
Equations”, Springer-Verlag (2007)

SUMÁRIO

RESUMO	8
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	14
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Métodos de Estimação Dependentes de Verossimilhança	19
2.1.1 Método Clássico	19
2.1.2 Método bayesiano	20
2.2 Intratabilidade da Verossimilhança	22
2.3 Computação bayesiana Aproximada	24
2.3.1 Algoritmos ABC	26
2.3.1.1 Rejeição	26
2.3.1.2 Regressão Local	29
2.3.1.3 MCMC	31
2.3.1.4 Sequencial	32
2.4 Equações Diferenciais	34
2.4.1 Métodos para Resolução de Equações Diferenciais	37
2.4.2 Equações Diferenciais Estocásticas	39
2.4.3 Sistemas Dinâmicos	41
2.4.3.1 Modelos Epidemiológicos	41
2.4.3.2 Modelos de Interação Populacional	44
2.5 Inferência para Sistemas Dinâmicos	46
3 MATERIAIS E MÉTODOS	51
3.1 Estudos de Simulação	51
3.1.1 Regressão Linear	52
3.1.2 Regressão Poisson	54
3.2 Modelos de Dinâmica Populacional	55
3.2.1 Modelos Epidemiológicos	55
3.2.1.1 Casos de Influenza no Internato	56
3.2.1.2 Gripe Comum em Tristão da Cunha	57
3.2.2 Modelos de Interação Populacional	58
3.2.2.1 Estudo de Simulação para o Modelo Lotka-Volterra	59
3.2.2.2 Modelo Lotka-Volterra para Lebre (presa) e Lince (predador)	60
3.2.2.3 Estudo de Simulação para o Modelo de Competição entre	
Espécies	62
4 RESULTADOS E DISCUSSÃO	65

4.1	Estudos de Simulação	65
4.1.1	Regressão Linear	65
4.1.2	Regressão Poisson	80
4.2	Modelos de Dinâmica Populacional	82
4.2.1	Modelos Epidemiológicos	83
4.2.1.1	Casos de Influenza no Internato	83
4.2.1.2	Casos de Gripe Comum em Tristão da Cunha	85
4.2.2	Modelos de Interação Populacional	88
4.2.2.1	Estudo de Simulação para o Modelo Lotka-Volterra	88
4.2.2.2	Modelo Lotka-Volterra para Lebre (presa) e Lince (predador)	91
4.2.2.3	Estudo de Simulação para o Modelo de Competição entre Espécies	94
5	CONCLUSÕES	97
	REFERÊNCIAS	99
	APÊNDICES	107
	ANEXOS	121

RESUMO

Computação bayesiana aproximada: aplicações em modelos de dinâmica populacional

Processos estocásticos complexos são muitas vezes utilizados em modelagem, com o intuito de capturar uma maior proporção das principais características dos sistemas biológicos. A descrição do comportamento desses sistemas tem sido realizada por muitos amostradores baseados na distribuição a posteriori de Monte Carlo. Modelos probabilísticos que descrevem esses processos podem levar a funções de verossimilhança computacionalmente intratáveis, impossibilitando a utilização de métodos de inferência estatística clássicos e os baseados em amostragem por meio de MCMC. A Computação Bayesiana Aproximada (ABC) é considerada um novo método de inferência com base em estatísticas de resumo, ou seja, valores calculados a partir do conjunto de dados (média, moda, variância, etc.). Essa metodologia combina muitas das vantagens da eficiência computacional de processos baseados em estatísticas de resumo com inferência estatística bayesiana uma vez que, funciona bem para pequenas amostras e possibilita incorporar informações passadas em um parâmetro e formar uma priori para análise futura. Nesse trabalho foi realizada uma comparação entre os métodos de estimação, clássico, bayesiano e ABC, para estudos de simulação de modelos simples e para análise de dados de dinâmica populacional. Foram implementadas no *software* R as distâncias modular e do máximo como alternativas de função distância a serem utilizadas no ABC, além do algoritmo ABC de rejeição para equações diferenciais estocásticas. Foi proposto sua utilização para a resolução de problemas envolvendo modelos de interação populacional. Os estudos de simulação mostraram melhores resultados quando utilizadas as distâncias euclidianas e do máximo juntamente com distribuições a priori informativas. Para os sistemas dinâmicos, a estimação por meio do ABC apresentou resultados mais próximos dos verdadeiros bem como menores discrepâncias, podendo assim ser utilizado como um método alternativo de estimação.

Palavras-chave: Inferência livre de verossimilhança; Distribuição a posteriori; Estatísticas de resumo; Equações diferenciais estocásticas

ABSTRACT

Approximate Bayesian Computation: applications in population dynamics models

Complex stochastic processes are often used in modeling in order to capture a greater proportion of the main features of natural systems. The description of the behavior of these systems has been made by many Monte Carlo based samplers of the posterior distribution. Probabilistic models describing these processes can lead to computationally intractable likelihood functions, precluding the use of classical statistical inference methods and those based on sampling by MCMC. The Approximate Bayesian Computation (ABC) is considered a new method for inference based on summary statistics, that is, calculated values from the data set (mean, mode, variance, etc.). This methodology combines many of the advantages of computational efficiency of processes based on summary statistics with the Bayesian statistical inference since, it works well for small samples and it makes possible to incorporate past information in a parameter and form a prior distribution for future analysis. In this work a comparison between, classical, Bayesian and ABC, estimation methods was made for simulation studies considering simple models and for data analysis of population dynamics. It was implemented in the R software the modular and maximum as alternative distances function to be used in the ABC, besides the rejection ABC algorithm for stochastic differential equations. It was proposed to use it to solve problems involving models of population interaction. The simulation studies showed better results when using the Euclidean and maximum distances together with informative prior distributions. For the dynamic systems, the ABC estimation presented results closer to the real ones as well as smaller discrepancies and could thus be used as an alternative estimation method.

Keywords: Likelihood-free inference; Posterior distribution; Summary statistics; Stochastic differential equations

LISTA DE FIGURAS

2.1	Esquema para o algoritmo de rejeição exato	27
2.2	Esquema para o algoritmo de rejeição com tolerância	28
2.3	Esquema para o algoritmo de rejeição ABC	29
2.4	Esquema para o modelo compartimental SIR sem dinâmica vital	42
2.5	Esquema para o modelo compartimental SIR com dinâmica vital	43
3.1	Número de indivíduos infectados por influenza no internato durante 15 dias	56
3.2	Esquema para o modelo compartimental SIR sem dinâmica vital	56
3.3	Número de indivíduos infectados e recuperados por gripe comum na ilha de Tristão da Cunha durante 21 dias	58
3.4	Número de indivíduos gerados com e sem erro: (a) População de presas; (b) População de predadores	60
3.5	População de peles de presas (lebres) e predadores (lince)	60
3.6	Número de indivíduos gerados com e sem erro: (a) População da espécie 1; (b) População da espécie 2	62
4.1	Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 10^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 2)	66
4.2	Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 3)	67
4.3	Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 3 com variável centrada)	68
4.4	Estudo do parâmetro β_1 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 3)	69
4.5	Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(7, 8; 6, 9^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 5)	70
4.6	Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_1 = 50$ e distribuição a priori $N(7, 8; 6, 9^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 5 com variável centrada)	71
4.7	Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 10^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 7)	72

4.8	Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 10^2)$ para β_0 e β_1 , $IGamma(0.5, 0.5)$ para σ^2 (cenário 8)	73
4.9	Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 9)	74
4.10	Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 9 com variável centrada) . .	75
4.11	Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 9)	76
4.12	Estudo do parâmetro σ^2 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 9 com variável centrada) . .	77
4.13	Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(7, 8; 7, 5^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 11)	78
4.14	Estudo do parâmetro σ_2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(7, 8; 7, 5^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 12)	79
4.15	Estudo do parâmetro β_1 para o modelo de regressão Poisson com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 (cenário 1) .	81
4.16	Estudo do parâmetro β_0 para o modelo de regressão Poisson com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 (cenário 2) .	82
4.17	Modelo SIR para dados de influenza: (a) Gráfico de contorno para a densidade dos parâmetros β e ν e o ponto estimado; (b) Valores preditos e observados para o número de garotos infectados durante o período de 15 dias	83
4.18	Modelo SIR pra dados de influenza estimado por meio do ABC com distância euclidiana: (a) Histograma das médias a posteriori para cada parâmetro estimado $(\beta, \nu, I(0))$; (b) Valores preditos pelo modelo para dados de influenza no internato	84
4.19	Modelo SIR pra dados de influenza estimado por meio do ABC com distância de Mahalanobis: (a) Histograma das médias a posteriori para cada parâmetro estimado $(\beta, \nu, I(0))$; (b) Modelo ajustado para dados de influenza no internato	85

4.20	Modelo SIR para os dados de Tristão da Cunha: (a) Gráfico de contorno para a densidade dos parâmetros β e ν e o ponto estimado; (b) Valores preditos e observados para o número de infectados e recuperados durante o período de 21 dias	86
4.21	Resíduos padronizados para o modelo SIR considerando a máxima verossimilhança para os dados de Tristão da Cunha: (a) Número de infectados; (b) Número de recuperados	86
4.22	Modelo SIR pra dados de Tristão da Cunha estimado por meio do ABC com distância de Mahalanobis: (a) Histograma das médias a posteriori para cada parâmetro estimado ($\beta, \nu, I(0)$); (b) Modelo ajustado para dados de gripe comum	87
4.23	Resíduos padronizados para o modelo SIR considerando o ABC para os dados de Tristão da Cunha: (a) Número de infectados; (b) Número de recuperados	87
4.24	Modelo Lotka-Volterra para dados simulados considerando: estimação por máxima verossimilhança - (a) Valores observados e preditos para presa e predador; (b) Gráfico de diagnóstico bivariado; estimação via MCMC - (c) Valores observados e preditos para presa e predador; (d) Gráfico de diagnóstico bivariado; estimação via ABC - (e) Valores observados e preditos para presa e predador; (f) Gráfico de diagnóstico bivariado	89
4.25	Histograma dos valores a posteriori estimados pelo MCMC para dados simulados, com sua média a posteriori e o verdadeiro valor do parâmetro	90
4.26	Histograma dos valores a posteriori estimados pelo ABC para dados simulados, com sua média a posteriori e o verdadeiro valor do parâmetro	90
4.27	Resíduos para o modelo Lotka-Volterra considerando a máxima verossimilhança para os dados reais de lebre (presa): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade	91
4.28	Resíduos para o modelo Lotka-Volterra considerando a máxima verossimilhança para os dados reais de lince (predador): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade	92
4.29	Resíduos para o modelo Lotka-Volterra considerando o ABC para lebre (presa): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade	92
4.30	Resíduos para o modelo Lotka-Volterra considerando o ABC para lince (predador): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade	92
4.31	Ajuste do modelo Lotka-Volterra para os dados de lebre e lince por meio da metodologia ABC	93

4.32	Estimação para dados simulados do modelo de competição utilizando o método MCMC: (a) Valores observados e preditos para espécie 1 e espécie 2; (b) Gráfico de diagnóstico bivariado	94
4.33	Estimação para dados simulados do modelo de competição utilizando o método ABC: (a) Valores observados e preditos para espécie 1 e espécie 2; (b) Gráfico de diagnóstico bivariado	94
A.1	Soluções da equação diferencial de forma explícita, y_a , e por meio do método de Euler, y_e	108

LISTA DE TABELAS

3.1	Conjuntos de dados simulados para a regressão linear com tamanho amostral $N_1 = 50$	52
3.2	Conjuntos de dados simulados para a regressão linear com tamanho amostral $N_2 = 10$	53
3.3	Tabela resumo para os casos de estudo da regressão linear	53
3.4	Conjuntos de dados simulados para a regressão Poisson com tamanhos amostrais $N_1 = 50$ e $N_2 = 10$	54
3.5	Tabela resumo para os casos de estudo da regressão Poisson	55
3.6	Número de infectados (I) por influenza em um internato inglês em 1978 . .	56
3.7	Dados de gripe comum em Tristão da Cunha coletados em outubro de 1967	58
3.8	Dados simulados para o modelo Lotka-Volterra e valores com erro aleatório $N(0; 0, 5^2)$	59
3.9	Quantidade, em milhares, de presas (lebre) e predadores (lince) ao longo do tempo	61
3.10	Dados simulados para o modelo de competição e valores com erro aleatório $N(0; 0, 5^2)$	63
4.1	Medidas de discrepância e tempo computacional do modelo SIR para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de gripe comum em Tristão da Cunha	87
4.2	Medidas de discrepância e tempo computacional do modelo Lotka-Volterra para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de predação simulados	91
4.3	Medidas de discrepância e tempo computacional do modelo Lotka-Volterra para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de predação de lebre e lince	93
4.4	Medidas de discrepância e tempo computacional do modelo de competição para os métodos MCMC e ABC considerando dados simulados	95
A.1	Resultados das soluções explícita (y_a) e por meio do método de Euler (y_e)	108
A.2	Medidas de discrepância para os parâmetros nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_1 = 50$, com $d_{\beta_0} = 3,5197$, $d_{\beta_1} = 0,1340$ e $d_{\sigma^2} = 2,1676$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança	109
A.3	Medidas de discrepância para os parâmetros nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_2 = 10$, com $d_{\beta_0} = 294,0437$, $d_{\beta_1} = 11,7774$ e $d_{\sigma^2} = 6,2438$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança	110

- A.4 Medidas de discrepância para os parâmetros da regressão Poisson nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_1 = 50$, com $d_{\beta_0} = 113,6138$ e $d_{\beta_1} = 112,6345$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança . . 111
- A.5 Medidas de discrepância para os parâmetros da regressão Poisson nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_2 = 10$, com $d_{\beta_0} = 115,3696$ e $d_{\beta_1} = 112,5526$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança . . 111

1 INTRODUÇÃO

Processos estocásticos complexos são frequentemente utilizados em modelagem a fim de capturar o máximo possível das características dos sistemas naturais. Compreender as variações dentro e entre populações tem sido, há décadas, um dos objetivos da comunidade científica (CSILLÉRY *et al.*, 2010).

Com o propósito de explicar os processos dinâmicos envolvidos nas populações, uma série de técnicas estatísticas e computacionais têm sido desenvolvidas. A identificação do comportamento desses sistemas geralmente é realizada por amostradores baseados em conceitos probabilísticos a fim de determinar uma função, denominada função de verossimilhança, para um parâmetro a partir da informação fornecida pelos dados. Atualmente, os métodos que utilizam a função de verossimilhança para estimação são os mais amplamente empregados, sendo que esta pode ser realizada de forma clássica ou por uma variedade de algoritmos, como métodos baseados na construção de cadeias de Markov.

Algumas das limitações desses métodos são a dificuldade de verificação da convergência da cadeia de Markov (CORNUET *et al.*, 2008) e obtenção da função de verossimilhança, restringindo seu uso a cenários mais simples como no estudo de uma única população (BEAUMONT, 1999; GRIFFITHS e TAVARÉ, 1994; STEPHENS e DONNELLY, 2000).

Para dados complexos ou de alta dimensão a função de verossimilhança é, frequentemente, intratável ou difícil de ser computada. No caso de sistemas dinâmicos, nem sempre a obtenção da verossimilhança é uma tarefa fácil, podendo apresentar diversos máximos locais, instabilidade numérica ou pontos de singularidade nos quais não existe solução.

Uma das mais recentes e satisfatórias aproximações para problemas probabilísticos envolvendo soluções complexas é a Computação bayesiana Aproximada (ABC) proposta por BEAUMONT *et al.* (2002) que surgiu, primeiro na área de genética e depois em um espectro mais amplo de aplicações. Essa metodologia contorna as dificuldades encontradas por métodos dependentes de verossimilhança, bem como seu cálculo por meio da utilização de resumos estatísticos e simulações (CSILLÉRY *et al.*, 2010). Resumos estatísticos são valores calculados a partir dos dados que representam o máximo de informação da forma mais simples possível, por exemplo, a média e a variância, entre outras.

BEAUMONT *et al.* (2002) mencionam que um fato importante da aproximação bayesiana é a resolução de problemas teóricos, com a presença de muitos parâmetros de perturbação que, podem ser incorporados no passo de simulação do modelo.

Este método para ajuste de modelos apresenta como vantagens a construção de um modelo gerador que produz o mesmo tipo de dados que se tenta modelar, assume-se probabilidade a priori sobre os parâmetros a serem estimados e essas distribuições representam as informações do modelo sobre os parâmetros antes da observação dos dados,

sendo intuitivo e exigindo apenas as especificações do modelo gerador.

O ABC ainda pode ser utilizado para obter valores independentes da distribuição a posteriori por meio da simulação de conjuntos de dados. Apresenta duas características adicionais que podem ser de interesse. A primeira é a possibilidade de avaliar o viés e a precisão das estimativas para os conjuntos de dados simulados, produzidos com valores de parâmetros conhecidos (ESTOUP *et al.*, 2004; EXCOFFIER *et al.*, 2005; CORNUET *et al.*, 2008). A segunda é a maneira simples pela qual a probabilidade a posteriori de diferentes cenários aplicados ao mesmo conjunto de dados pode ser estimada (MILLER *et al.*, 2005; PASCUAL *et al.*, 2007).

Aplicações desse método podem ser encontradas em estatística genética (TAVARÉ *et al.*, 1997; STEPHENS e DONNELLY, 2000) e modelagem de transmissão de doenças (BLUM e TRAN, 2010; LU *et al.*, 2013; KYPRAIOS *et al.*, 2017). Esse procedimento pode ser aplicado a vários sistemas biológicos conhecidos, e assim, os parâmetros e seus intervalos de credibilidade podem ser inferidos.

De maneira geral, os trabalhos apresentados na literatura, estão focados na aplicação da metodologia ABC, bem como em melhorias do algoritmo, não sendo observadas comparações com os métodos tradicionais de estimação. Nesse sentido, o presente trabalho visa preencher essa lacuna com a comparação dos métodos, clássico, bayesiano e ABC por meio de estudos de simulação, para modelos simples (regressão linear e Poisson) e análises de dados e simulações para modelos de dinâmica populacional.

Além disso, por meio de estudos de simulação, os resultados foram avaliados ao empregar diferentes, tolerâncias e distribuições a priori além da utilização das distâncias euclidiana, modular e do máximo. Foram implementadas no *software* R (R CORE TEAM, 2015) as distâncias modular e do máximo como alternativas de função distância a serem utilizadas, além do algoritmo ABC de rejeição para equações diferenciais estocásticas. Foi proposto sua utilização para a resolução de problemas envolvendo modelos de interação populacional.

O presente trabalho foi organizado da seguinte forma: na seção 2, apresentou-se uma revisão de literatura tanto para os métodos ABC quanto para equações diferenciais estocásticas bem como a inferência para tais modelos, na seção 3 a metodologia utilizada para o estudo de simulação é apresentada, bem como os conjuntos de dados e modelos utilizados para os sistemas dinâmicos e na seção 4, podem ser observados os resultados alcançados por meio do estudo de simulação além das comparações entre as metodologias utilizadas na inferência para sistemas dinâmicos. Por fim um apêndice com detalhes da implementação para essa metodologia no *software* R (R CORE TEAM, 2015) é apresentado.

2 REVISÃO BIBLIOGRÁFICA

2.1 Métodos de Estimação Dependentes de Verossimilhança

A análise e resumo de informações contidas em um conjunto de dados consiste na suposição de que existe um mecanismo estocástico gerador do processo. Este procedimento, geralmente, é representado por modelos probabilísticos que apresentam um ou mais parâmetros, conforme características do fenômeno observado. Dessa forma, a inferência estatística visa encontrar valores plausíveis para esses parâmetros, bem como, informar intervalos de variação e opinar sobre a estimativa em relação a alguma suposição especificada.

Partindo destes conceitos, um fenômeno aleatório ou estocástico pode ser descrito por uma distribuição de probabilidade, que por sua vez é indexada por seus parâmetros e respectivos campos de variação (espaço paramétrico), além do campo de variação da própria variável aleatória, que deve ser compatível com o suporte da distribuição. Com essas suposições, é possível fazer o processo de inferência sobre os parâmetros a partir dos dados de uma amostra por meio da função de verossimilhança, tanto para o método clássico quanto para o bayesiano (GLICKMAN e VAN DYK, 2007).

2.1.1 Método Clássico

A estimação clássica de parâmetros pode ser realizada por diversos procedimentos como o método dos momentos, mínimos quadrados e, provavelmente pelo mais popular, método da máxima verossimilhança (WACKERLY *et al.*, 2007; CASELLA e BERGER, 2002). Sejam $Y_1, \dots, Y_n$ variáveis aleatórias independentes e identicamente distribuídas (iid) de uma população com função de probabilidade ou densidade de probabilidade $f(y|\theta_1, \dots, \theta_k)$, no qual a função de verossimilhança pode ser definida como

$$L(\boldsymbol{\theta}|\mathbf{y}) = L(\theta_1, \dots, \theta_k|y_1, \dots, y_n) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta}).$$

Definição 1: Para cada ponto amostrado y , seja $\hat{\theta}(y)$ o valor do parâmetro que $L(\theta|y)$ atinge o máximo da função para θ com y fixo. Um estimador de máxima verossimilhança (EMV) do parâmetro θ baseado na amostra Y é $\hat{\theta}(y)$.

Note que, pela sua construção, a amplitude do EMV coincide com a amplitude do parâmetro. Existem casos em que o EMV pode não existir ou não ser único. No entanto, se existirem, não apresentam viés assintoticamente. Esta propriedade é muito importante, fundamentando grande parte da estatística inferencial (MOOD *et al.*, 1974). Apesar do EMV ser uma escolha razoável por ser uma medida de compatibilidade da amostra observada com um particular vetor de parâmetros, existem duas limitações inerentes associadas

a este método. A primeira consiste na determinação do máximo global e na sua verificação. Segundo CASELLA e BERGER (2002) em muitos casos este problema é facilmente resolvido, no entanto, mesmo para densidades comuns dificuldades podem surgir. A segunda esta relacionada com a sensibilidade numérica, isto é, quão sensível é a estimativa a pequenas variações nos dados. Neste sentido, a abordagem por verossimilhança não pode ser considerada a solução de todo e qualquer problema, sendo necessária a utilização de métodos alternativos que estendam, aproximem ou substituam a verossimilhança quando necessário.

2.1.2 Método bayesiano

A fim de incorporar informações a priori sobre parâmetros de interesse, BAYES e PRICE (1763) apresentaram uma nova metodologia denominada Inferência Estatística bayesiana, sendo o principal resultado desse trabalho o “Teorema de Bayes”. Para introduzir esse teorema, são necessários os axiomas de Kolmogorov (KOLMOGOROV, 1950) que definem o espaço de probabilidade (Ω, F, P) , em que:

- i. Ω é um conjunto não vazio de elementos $\omega, \omega \in \Omega$, e subconjuntos $A, A \subseteq \Omega$, chamado espaço amostral.
- ii. F é uma σ -álgebra de subconjuntos de Ω .

Definição 2: Uma classe F de eventos é uma σ -álgebra se:

- a. $\Omega \in F$
- b. $A \in F \Rightarrow A^c \in F$
- c. $A_1, A_2, \dots \in F \Rightarrow \bigcup_{i=1}^{\infty} A_i \in F$.

- iii. P é uma probabilidade para os eventos $A \subseteq F$, em que $P(A)$ é a probabilidade de A .

Definição 3: Uma função $P(A)$ é uma probabilidade se:

- a. $P[A] \geq 0, \forall A \in F$
- b. $P[\Omega] = 1$
- c. Se $A_i \cap A_j = \emptyset, \forall i \neq j, j = 1, 2, \dots$ e se $A_1 \cup A_2 \cup \dots = \bigcup_{i=1}^{\infty} A_i \in F$, então

$$P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i].$$

Considere uma partição finita ou infinita de Ω , $A_1, A_2, \dots, A_m$, com $P(A_i) > 0, A_i \cap A_j = \emptyset, i \neq j$ e $\cup_i(A_i) = \Omega$. Dado um evento $B \subset \Omega$ com $P(B) > 0$ e sua decomposição em conjuntos disjuntos $B = \cup_i(A_i \cap B)$. Pela definição de probabilidade condicional, tem-se, $P(B) = \sum_i P(A_i \cap B)$. Logo,

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B),$$

e portanto,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)},$$

é denominado Teorema de Bayes.

Este teorema é importante nas relações algébricas de probabilidades condicionais, pois relaciona a probabilidade de ocorrência no presente com a probabilidade a priori (DEGROOT e SCHERVISH, 2012).

Seja θ um parâmetro desconhecido e sua probabilidade de ocorrência $\pi(\theta)$, que pode ser atualizada por meio de uma variável aleatória Y de interesse, cuja distribuição de probabilidade está relacionada com θ . Essa relação pode ser descrita por meio da distribuição amostral dada a seguir:

$$\pi(\theta|y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(y|\theta)\pi(\theta)}{\int \pi(\theta, y)d\theta} = \frac{L(\theta|y)\pi(\theta)}{\int \pi(\theta, y)d\theta}, \quad (2.1)$$

sendo $1/\pi(y)$ a constante normalizadora de $\pi(\theta|y)$.

Em inferência bayesiana as probabilidades envolvidas no teorema de Bayes podem fornecer diferentes interpretações, uma vez que são expressas com um grau subjetivo de crença que pode mudar racionalmente as evidências.

De acordo com GLICKMAN e VAN DYK (2007) os ingredientes chave para a análise bayesiana são a função de verossimilhança e a distribuição a priori. A primeira reflete a informação sobre os parâmetros contidos nos dados para cada um dos possíveis valores de θ , com $L(\theta|y) = \pi(y|\theta)$. Já a distribuição a priori de θ , $\pi(\theta)$, quantifica o que é conhecido a respeito dos parâmetros antes de observar os dados (BOX e TIAO, 1992), ou seja, representa a informação sobre a incerteza a respeito do vetor de parâmetros (GELMAN *et al.*, 2004).

A distribuição a priori pode ser informativa (subjetiva) ou não informativa (objetiva). A distribuição a priori não informativa é utilizada quando, não existem ou as informações prévias disponíveis a respeito dos parâmetros são pouco relevantes.

Distribuição a priori e verossimilhança podem ser combinadas para resultar na distribuição a posteriori, a qual representa conhecimento total sobre os parâmetros depois dos dados serem observados. Sob essas condições BOX e TIAO (1992) citam que a forma usual do Teorema de Bayes, é dada por,

$$\pi(\theta|y) \propto L(\theta|y)\pi(\theta).$$

Essa maneira simplificada do Teorema de Bayes pode ser utilizada para isolar quantidades de interesse e finalmente para escrever conclusões substantivas (GLICKMAN e VAN DYK, 2007).

A constante normalizadora pode ser recuperada pois, $\pi(\theta|y) = k\pi(y|\theta)\pi(\theta)$ em que $k^{-1} = \pi(y)$ é chamada distribuição preditiva a priori. A distribuição preditiva de x dado y é dada por:

$$\pi(x|y) = \int \pi(x, \theta|y) d\theta = \int \pi(x|\theta, y) \pi(\theta|y) d\theta.$$

Quando há independência condicional entre X e Y dado θ , tem-se que $\pi(x|y) = \int \pi(x|\theta) \pi(\theta|y) d\theta$.

Para grandes amostras os resultados por meio da inferência clássica e bayesiana são similares. Entretanto, existem algumas vantagens na utilização da análise bayesiana. É possível incorporar informações passadas em um parâmetro e formar uma priori para futuras análises. As novas observações da distribuição a posteriori podem ser usadas como priori. Esse procedimento fornece inferências mais precisas, e da mesma forma que para grandes, funciona para pequenas amostras.

Os métodos de Monte Carlo via cadeias de Markov (MCMC) são algoritmos de simulação iterativos utilizados para a obtenção de uma amostra da distribuição a posteriori. Valores não independentes, gerados por meio de um passeio aleatório no espaço amostral, convergem para a distribuição de interesse do problema. Para isso a escolha da distribuição a priori é de grande importância, uma vez que esta pode interferir nos resultados da posteriori e assim, para muitos casos uma análise de sensibilidade da priori se faz necessária.

Uma análise de sensibilidade, também referida como análise de simulação, é uma técnica usada para determinar como diferentes valores de uma variável independente influenciam uma determinada variável dependente dado um conjunto de premissas (SALTELLI *et al.*, 2008). De acordo com GELMAN *et al.* (2004) a inferência a posteriori baseada em grandes amostras não é particularmente sensível a distribuição a priori. Mesmo assim, uma boa análise bayesiana, deve incluir pelo menos alguma verificação de plausibilidade do modelo. Isso às vezes é discutido como um problema de sensibilidade para a distribuição a priori (BERGER, 1990).

Apesar de sua precisão e eficiência para pequenas amostras, as estimativas baseadas na simulação das quantidades a posteriori podem variar. Além disso, alguns problemas envolvendo verossimilhanças intratáveis podem ser encontrados. Mais detalhes serão discutidos a seção 2.2.

2.2 Intratabilidade da Verossimilhança

A realização da inferência é, na maioria dos contextos estatísticos, baseada em métodos que envolvem a função de verossimilhança. Entretanto, para problemas complexos, a avaliação da verossimilhança pode se tornar árdua ou até mesmo impossível, dependendo do caso de estudo. Com a revolução da estatística computacional, algumas

metodologias para a inferência baseadas em verossimilhança, foram desenvolvidas, entre elas, os métodos Monte Carlo via cadeias de Markov.

Embora tais métodos sejam amplamente empregados para a realização da inferência, existem limitações envolvendo grande dimensionalidade de parâmetros ou até mesmo conjuntos de dados extensos, que geralmente, não podem ser abordados usando métodos já existentes baseados em verossimilhança.

No entanto, para problemas envolvendo verossimilhanças intratáveis, métodos alternativos livres de verossimilhança como o ABC tem sido estabelecidos. Esses avanços, aliados aos desenvolvimentos computacionais, têm enorme potencial para enfrentar os desafios mais difíceis das questões científicas modernas.

MURRAY *et al.* (2006) além de WILKINSON (2013) definiram algumas categorias de problemas em que a função de verossimilhança é intratável. Um dos casos é quando parte da função de verossimilhança é uma constante desconhecida e sua expressão é dada da seguinte forma:

$$L(\theta|y) = \pi(y|\theta)/C,$$

em que π é a parte “conhecida” da função densidade avaliada para os dados observados Y , ou seja, dado um valor de θ é possível calcular $\pi(y|\theta)$, e C é uma constante difícil de ser calculada (ou intratável), que independe do parâmetro de interesse θ .

O amostrador MCMC da distribuição a posteriori $\pi(\theta|y)$, precisa que a verossimilhança seja conhecida a menos de uma constante de proporcionalidade. Uma vez que,

$$L(\theta|y) \propto \pi(y|\theta),$$

em que o fato de a constante C ser desconhecida, não é um real problema para a inferência bayesiana.

Um tipo de intratabilidade mais complexa surge quando a parte desconhecida da verossimilhança, C , depende do parâmetro θ de interesse. Ou seja, a verossimilhança é dada por

$$L(\theta|y) = \pi(y|\theta)/C(\theta).$$

Neste caso, para avaliar a verossimilhança de um dado parâmetro, é preciso determinar $C(\theta)$. A constante geralmente depende do modelo e envolve integrais com alta dimensão e difíceis de calcular. Alguns métodos numéricos podem ser empregados. A principal desvantagem é que uma pequena alteração no modelo pode resultar em um novo $C(\theta)$ que não seja computável com o método numérico previamente planejado.

O método MCMC não pode ser utilizado para amostrar a distribuição a posteriori dos parâmetros sem modificação adicional, quando a constante é da forma $C(\theta)$. MURRAY *et al.* (2006) propuseram uma aproximação adaptada para este tipo de intratabilidade. Para superar esta dificuldade, eles sugerem a utilização do ABC.

A intratabilidade que motiva os métodos ABC foi primeiramente estudada por DIGGLE e GRATTON (1984). Os autores consideram que, sob as restrições dos parâmetros,

os dados podem ser gerados a partir do modelo, mas essa é toda a informação disponível sobre a verossimilhança. Nesse caso, nenhuma parte da verossimilhança $L(\theta|y)$ é explícita apesar de ser possível gerar observações independentes dos dados que são denotados por y^* a partir do modelo $\pi(y|\theta)$.

Os métodos ABC realizam inferência aproximada para a distribuição a posteriori dos parâmetros de tais modelos. Seu sucesso depende da simulação de conjuntos de pseudo-dados sob um modelo estocástico de interesse.

De acordo com RAYNER e MACGILLIVRAY (2002) a distribuição *g-and-k* é um excelente exemplo em que a função quantil é simples de ser expressa enquanto sua verossimilhança não está disponível:

$$Q(u; A, B, g, k) = A + B \left[1 + c \frac{1 - \exp\{-g\Phi(u)\}}{1 + \exp\{-g\Phi(u)\}} \right] (1 + \Phi(u)^2)^k \Phi(u),$$

em que A e $B > 0$ são parâmetros de locação e escala, g mensura assimetria, $k > -\frac{1}{2}$ a curtose da distribuição, $\Phi(u)$ é o u -ésimo quantil da normal padrão, e c é uma constante escolhida para ajudar a produzir distribuições adequadas. A verossimilhança da distribuição *g-and-k* é dada por:

$$L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f_x(x_i|\theta) = \prod_{i=1}^n \frac{\partial}{\partial x_i} Q_x^{-1}(x_i|\theta) = \left(\prod_{i=1}^n Q'_x(Q_x^{-1}(x_i|\theta)|\theta) \right).$$

A maximização da função de verossimilhança para os parâmetros não pode ser expressa analiticamente porque não existe forma fechada para a expressão $Q_x^{-1}(x_i|\theta)$. A verossimilhança pode ser avaliada numericamente (RAYNER e MACGILLIVRAY, 2002), entretanto a simulação a partir do modelo *g-and-k* é mais direta. Mais especificamente, um desenho da distribuição *g-and-k* envolve a simulação de uma variação aleatória normal padrão com $\Phi(u) = z(p)$ (DROVANDI *et al.*, 2011).

2.3 Computação bayesiana Aproximada

Os métodos ABC não necessitam da avaliação da função de verossimilhança. Dessa forma, com a realização de aproximações, mais modelos podem ser considerados para a realização da inferência. Isso faz com que o ABC ganhe cada vez mais popularidade para a análise de problemas complexos que surgem em genética (BEAUMONT *et al.*, 2002), epidemiologia (TANAKA *et al.*, 2006), genômica (CSILLÉRY *et al.*, 2010), e ecologia (TONI *et al.*, 2009; BEAUMONT, 2010).

As primeiras ideias sobre ABC surgiram com RUBIN (1984), que descreveu um mecanismo de amostragem que produz uma amostra independente da distribuição a posteriori. Além disso, RUBIN (1984) argumenta que na estatística bayesiana, os estatísticos

aplicados não deveriam dispor apenas de modelos analiticamente tratáveis, mas também considerar métodos computacionais que permitem estimar a distribuição a posteriori de interesse. Esses argumentos são de grande relevância no contexto ABC.

DIGGLE e GRATTON (1984) sugeriram, para casos em que a forma analítica é intratável, aproximar a função de verossimilhança por meio de simulações que utilizam valores amostrados pertencentes ao espaço paramétrico. Tal aproximação pode ser melhorada aplicando técnicas de suavização aos resultados da simulação.

Embora a aproximação de DIGGLE e GRATTON (1984) tenha aberto uma nova fronteira, seu método ainda não é exatamente idêntico ao que é conhecido por ABC, pois buscava aproximar a verossimilhança e não a distribuição a posteriori.

TAVARÉ *et al.* (1997) foram os primeiros a propor um algoritmo de rejeição exato para inferência a posteriori e podem ser considerados os precursores dos métodos ABC. Em seu trabalho, foi considerada a inferência para a genealogia de dados de uma sequência de DNA, e em particular o problema de decidir a distribuição a posteriori do tempo para o ancestral comum mais recente dos indivíduos amostrados. Essa inferência é analiticamente intratável para muitos modelos demográficos, mas os autores apresentaram maneiras de simular dados desses modelos. Uma amostra da posteriori dos parâmetros do modelo foi obtida aceitando/rejeitando amostras entre dados sintéticos e reais.

Posteriormente, PRITCHARD *et al.* (1999) foram os primeiros a utilizar verossimilhanças aproximadas com o algoritmo de TAVARÉ *et al.* (1997) em uma aplicação na modelagem da variação do cromossomo humano. Finalmente, o termo Computação bayesiana Aproximada foi apresentado por BEAUMONT *et al.* (2002) que estenderam ainda mais a metodologia para problemas envolvendo genética populacional. Eles propuseram a utilização do algoritmo de rejeição previamente desenvolvido por TAVARÉ *et al.* (1997) e implementaram a ideia de inferência por meio de resumos estatísticos, S_y . Desde então, o ABC vem sendo usado em aplicações fora da genética populacional, como em biologia, epidemiologia e fitogeografia.

O ABC tem suas raízes no algoritmo de rejeição o qual consiste na técnica de simular um grande número de amostras em um cenário hipotético da distribuição de probabilidade (TAVARÉ *et al.*, 1997; PRITCHARD *et al.*, 1999). A partir de uma priori para os parâmetros de interesse em estudo, o ABC realiza simulações do modelo, que são utilizadas para aproximar a verossimilhança sem explicitamente avaliá-la. Tais métodos baseiam-se no algoritmo de rejeição primeiramente tratados por DIGGLE e GRATTON (1984) e RUBIN (1984).

A inferência por meio do ABC tem sido realizada por resumos estatísticos dos dados ao invés do conjunto de dados completo, os quais são aceitos ou rejeitados, com base em uma distância métrica entre o resumo obtido pela predição e o observado (BEAUMONT *et al.*, 2002; CSILLÉRY *et al.*, 2010). A menos que essa estatística seja suficiente, geralmente há perda de informação e assim, as probabilidades são aproximadas.

Geralmente, há duas aproximações envolvidas nos procedimentos ABC. A primeira é que à medida que o esforço computacional aumenta, o erro diminui. Fato este devido a correção pós-amostragem dos valores independentes da posteriori no contexto do amostrador de rejeição (BEAUMONT *et al.*, 2002; BLUM e FRANCOIS, 2010). A segunda aproximação envolve a substituição do conjunto de dados completo por resumos estatísticos, que frequentemente não são estatísticas suficientes.

Os métodos ABC têm sido utilizados em muitos problemas práticos de inferência, especialmente em cenários que requerem a modelagem estocástica complexa com dados e/ou espaço paramétrico de alta dimensão.

Alguns exemplos envolvendo dados genéticos são inferências sobre o histórico demográfico das populações (PRITCHARD *et al.*, 1999; BLUM e JAKOBSSON, 2011) e espécies (ESTOUP *et al.*, 2004; EXCOFFIER *et al.*, 2005) como o tamanho efetivo da população, a detecção e a duração dos últimos eventos demográficos e a taxa de propagação de patógenos. Em ecologia, o ABC também tem sido utilizado para a realização de inferências, como em BEAUMONT (2010) ou por exemplo, para identificar a biodiversidade de florestas tropicais (JABOT e CHAVE, 2009).

2.3.1 Algoritmos ABC

Fundamentalmente, os algoritmos ABC baseiam-se na simulação de observações da distribuição a posteriori sem a função de verossimilhança analítica. O modelo de probabilidade $\pi(y|\theta)$ é utilizado para gerar dados, em que θ é um parâmetro de interesse. A distribuição a posteriori é dada por $\pi(\theta|y)$. Se o modelo $\pi(y|\theta)$ é acessível, a verossimilhança dos dados observados $L(\theta|y)$ pode ser avaliada, e a inferência bayesiana sobre $\pi(\theta|y) \propto L(\theta|y)\pi(\theta)$, em que $\pi(\theta)$ é a distribuição a priori, pode prosseguir por meio de métodos computacionais padrão (GELMAN *et al.*, 2004).

Quando esses métodos padrão empregam a verossimilhança exata $L(\theta|y)$, a inferência é exata, a menos de um erro Monte Carlo, no sentido em que o objetivo é a distribuição a posteriori $\pi(\theta|y)$. Os métodos computacionais projetados para amostrar a distribuição a posteriori sem avaliar a função de verossimilhança são precursores dos métodos ABC.

Diversos algoritmos ABC têm sido propostos e melhorados ao longo dos anos. Dentre eles destacam-se quatro algoritmos base que são rejeição (TAVARÉ *et al.*, 1997), regressão (BEAUMONT *et al.*, 2002), MCMC (MARJORAM *et al.*, 2003) e Sequencial Monte Carlo, SMC, (SISSON *et al.*, 2007; TONI *et al.*, 2009).

2.3.1.1 Rejeição

A ideia básica do algoritmo de rejeição foi introduzida por RUBIN (1984) e posteriormente apresentada por MARIN *et al.* (2011). Com base na priori $\pi(\theta)$ e nos dados

observados $\mathbf{Y}_{(n \times 1)}$, é amostrado um valor de parâmetro candidato θ^* e novos dados y^* são gerados a partir do modelo $\pi(\cdot|\cdot)$.

2.3.1.1.1 Algoritmo de rejeição exato

Não requer uma forma explícita da verossimilhança, mas é necessária a simulação de dados a partir do modelo proposto. As etapas são dadas a seguir:

-
1. Gerar θ^* a partir de $\pi(\theta)$.
  2. Gerar $\mathbf{Y}^*$ a partir de $\pi(\cdot|\theta^*)$.
  3. Aceitar θ^* se $\mathbf{Y}^* = \mathbf{Y}$; retornar ao passo 1.
-

Figura 2.1: Esquema para o algoritmo de rejeição exato

Na Figura 2.1 é apresentado o algoritmo de rejeição exato em que, θ^* são valores amostrados da priori $\pi(\theta)$, $\mathbf{Y}^*$ são novos valores simulados a partir de $\pi(\cdot|\theta^*)$, $\mathbf{Y}^*$ são valores aceitos da distribuição conjunta $\pi(\theta^*, \mathbf{Y}^*)$ e θ^* são valores aceitos da distribuição a posteriori, isto é, a distribuição marginal $\pi(\theta|\mathbf{Y})$. Infelizmente, amostrar da posteriori por métodos exatos pode ser computacionalmente difícil para modelos complexos. O problema com esse algoritmo é que muitos valores precisam ser gerados para a aceitação de poucos. Por causa desse problema, melhorias no algoritmo têm sido realizadas, como em TAVARÉ *et al.* (1997) e PRITCHARD *et al.* (1999).

2.3.1.1.2 Algoritmo de rejeição com tolerância

Parâmetros são aceitos somente se a distância entre $\mathbf{Y}$ e $\mathbf{Y}^*$ é menor que um determinado limite de tolerância, δ . As etapas são as seguintes:

-
1. Gerar θ^* a partir $\pi(\theta)$.
  2. Gerar $\mathbf{Y}^*$ a partir $\pi(\cdot|\theta^*)$.
  3. Aceitar θ^* se $d(\mathbf{Y}^*, \mathbf{Y}) < \delta$; retornar ao passo 1.
-

Figura 2.2: Esquema para o algoritmo de rejeição com tolerância

Observa-se na Figura 2.2 um esquema para o algoritmo de rejeição com tolerância em que, θ^* são valores amostrados da priori $\pi(\theta)$, $\mathbf{Y}^*$ são novos valores simulados a partir de $\pi(\cdot|\theta^*)$, valores $\mathbf{Y}^*$ da distribuição conjunta $\pi(\theta^*, \mathbf{Y}^*)$ são aceitos com uma tolerância δ e valores θ^* da distribuição a posteriori, ou seja, da distribuição marginal $\pi(\theta|d(\mathbf{Y}^*, \mathbf{Y}) < \delta)$ são aceitos com $d(\cdot, \cdot)$ uma função distância. Nesse caso, mais valores a posteriori são aceitos. Esse procedimento foi melhorado ainda mais para criar o algoritmo ABC.

A substituição dos dados por resumos estatísticos S tornou-se uma prática comum. Em particular, $\mathbf{Y}$ e $\mathbf{Y}^*$ são substituídos por S_y e S^* calculados a partir dos respectivos conjuntos de dados. Embora não haja perda de informação se as estatísticas S forem suficientes, os modelos complexos que admitem tais estatísticas são exceções na prática. A perda de informação é o primeiro nível de aproximação no ABC.

A caracterização da relação entre $\pi(\theta|S_y)$ e $\pi(\theta|y)$ continua sendo uma área de pesquisa ativa. Quando a posteriori é direcionada $\pi(\theta|S_y)$, satisfazer exatamente $S^* = S_y$ pode não ser possível. Este problema pode ser superado no ABC, relaxando a igualdade rigorosa na condição de aceitação $S^* = S_y$ com um intervalo de tolerância δ (BEAUMONT *et al.*, 2002).

2.3.1.1.3 Algoritmo de rejeição ABC

Os dados $\mathbf{Y}$ são substituídos por resumos estatísticos S_y .

-
1. Gerar θ^* a partir de $\pi(\theta)$.
  2. Gerar $\mathbf{Y}^*$ a partir de $\pi(\cdot|\theta^*)$.
  3. Computar resumos estatísticos S^* a partir de $\mathbf{Y}^*$.
  4. Aceitar θ^* se $d(S^*, S_y) < \delta$; retornar ao passo 1.
-

Figura 2.3: Esquema para o algoritmo de rejeição ABC

Observa-se na Figura 2.3 o algoritmo de rejeição ABC em que, θ^* são valores amostrados da priori $\pi(\theta)$, $\mathbf{Y}^*$ são novos valores simulados a partir de $\pi(\cdot|\theta^*)$ e suas respectivas estatísticas resumo S^* são calculadas, valores S^* da distribuição conjunta $\pi(\theta^*, S^*)$ são aceitos com uma tolerância δ e valores θ^* da distribuição a posteriori $\pi(\theta|d(S^*, S_y) < \delta)$ são aceitos com $d(\cdot, \cdot)$ uma função distância. Grandes valores de tolerância, δ , definida pelo usuário levam a maiores taxas de aceitação, mas também diminuem a precisão (TAVARÉ *et al.*, 1997).

2.3.1.2 Regressão Local

Para a estruturação do algoritmo de regressão ABC primeiramente é necessária a definição de regressão local. Trata-se de um método não paramétrico que utiliza suavização para ajustar curvas e superfícies aos dados. Sejam y_i , $i = 1, \dots, n$, os valores observados. É possível escrever

$$y_i = s(x_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2),$$

em que x_i é a variável independente e ε_i é o termo de erro aleatório.

A regressão local estima a função $s(\cdot)$ na vizinhança de cada ponto de interesse $x = x_0$. Uma forma de estimar a função localmente é considerar uma média ponderada de observações que estão na vizinhança de um ponto de interesse x_0 .

Dado o tamanho da vizinhança δ , o ponto $x = x_0$ e a função K_δ , que pondera o conjunto de pontos vizinhos para x_0 , chamado *kernel*, a equação para a média local ponderada por K_δ é dada por:

$$\hat{s}(x_0) = \frac{\sum_{i=1}^n K_\delta(x_i - x_0) y_i}{\sum_{i=1}^n K_\delta(x_i - x_0)}.$$

Entretanto, alguns problemas de viés nas regiões de fronteira (perto de y_1 ou y_n) podem surgir. Uma forma de resolver este problema é por meio da regressão linear local ponderada. Se uma reta é localmente estimada, este problema é eliminado.

Contudo, é necessário fazer três escolhas: função *kernel*, parâmetro de suavização e grau polinomial. Para mais detalhes consulte CLEVELAND *et al.* (1988).

De acordo com BEAUMONT *et al.* (2002) a densidade condicional pode ser estimada por meio do modelo de regressão a seguir que corrige a discrepância entre S_y e S_i^* .

$$\theta_i = \alpha + (S_i^* - S_y)' \beta + \varepsilon_i,$$

em que $\varepsilon_i \sim N(0, \sigma^2)$ e para cada ponto aceito tem-se: valor do parâmetro θ_i , resumo estatístico S_i^* , e medida de distância $d(S_i^*, S_y)$. Dessa forma, BEAUMONT *et al.* (2002) estabeleceram

$$\hat{\alpha} = \frac{\sum_{i=1}^n \{\theta_i - (S_i^* - S_y)' \hat{\beta}\} K_\delta(\|S_i^* - S_y\|)}{\sum_{i=1}^n K_\delta(\|S_i^* - S_y\|)},$$

que de acordo com NADARAYA (1964) e WATSON (1964), é o estimador de regressão da esperança condicional $E[\theta|S]$. Quando $\beta \equiv 0$,

$$\hat{\alpha} = \frac{\sum_{i=1}^n \theta_i K_\delta(\|S_i^* - S_y\|)}{\sum_{i=1}^n K_\delta(\|S_i^* - S_y\|)}.$$

DOKSUM e LO (1990) afirmaram a posteriori parcial como

$$\pi(\theta|S_y) = \frac{\pi(S_y|\theta)\pi(\theta)}{\pi(S_y)},$$

e disseram que essa posteriori coincide com a completa se os resumos estatísticos forem suficientes. Essa técnica pode ser estendida para ajustar múltiplos parâmetros simultaneamente utilizando regressão multivariada, no caso em que β é uma matriz e α e θ são vetores (BEAUMONT *et al.*, 2002).

A densidade a posteriori para o valor candidato de θ_0 para θ pode ser aproximada usando a estimativa da densidade *kernel* indicadora aplicando um peso na amostra

$$\pi(\hat{\theta}_0|S_y) = \frac{\sum_{i=1}^n K_\Delta(\theta_i - \theta_0) I_\delta(\|S_i^* - S_y\|)}{\sum_{i=1}^n I_\delta(\|S_i^* - S_y\|)},$$

em que Δ é a largura de banda de estimativa de densidade.

Uma das vantagens é que o algoritmo ABC fornece amostras independentes de

$$\pi(\theta|d(S^*, S_y) \leq \delta).$$

Se S_y é uma estatística suficiente, como $\delta \rightarrow 0$, e o número de simulações $\rightarrow \infty$,

$$\pi(\theta|d(S^*, S_y)) = \pi(\theta|\mathbf{Y}),$$

mais detalhes podem ser investigados em WILKINSON (2013).

Dessa forma, BEAUMONT *et al.* (2002) propuseram uma extensão do algoritmo (2.3.1.1.3), no qual uma densidade de *kernel* $K_\delta(S^*, S_y)$ é empregada, com tolerância δ . A cada conjunto de dados simulado que está dentro de δ é atribuído um peso pelo

kernel escolhido $K_\delta(S^*, S_y)$ com relação a alguma função de distância (WILKINSON, 2013; ROBERT *et al.*, 2011).

A primeira aplicação dessa ideia é devido a PRITCHARD *et al.* (1999) que usaram um *kernel* com suporte $(S_y - \delta, S_y + \delta)$ para obter um algoritmo de rejeição ABC como uma versão aproximada do método de TAVARÉ *et al.* (1997).

2.3.1.2.1 Algoritmo de regressão ABC

É realizada uma suavização dos parâmetros por meio da regressão local.

-
1. Gerar θ^* a partir de $\pi(\theta)$.
  2. Gerar $\mathbf{Y}^*$ a partir de $\pi(\cdot|\theta^*)$.
  3. Computar resumos estatísticos S^* a partir de $\mathbf{Y}^*$.
  4. Calcular $d(S^*, S_y)$.
  5. Escolher a tolerância δ de forma que a proporção de pontos aceitos seja $P_\delta = N/M$.
  6. Ponderar os valores simulados S_i^* utilizando $K_\delta(d(S_i^*, S_y))$, em que K_δ é, usualmente, a função *kernel* de Epanechnikov.
  7. Aplicar a regressão linear ponderada aos N pontos que possuem peso diferente de zero para obter a estimativa de $E(\theta|S_y)$.
  8. Ajustar $\theta_i^* = \theta_i - \hat{E}(\theta|S_i^*) + \hat{E}(\theta|S_y)$.
  9. Os valores θ_i^* , com pesos $K_\delta(d(S_i^*, S_y))$, são considerados aleatórios da aproximação da distribuição a posteriori $\pi(\theta|y)$.
-

2.3.1.3 MCMC

A fim de superar problemas causados pela baixa taxa de aceitação quando as distribuições a priori e a posteriori são muito diferentes entre si, MARJORAM *et al.* (2003) desenvolveram um algoritmo que incorpora os passos de simulação em um algoritmo Metropolis-Hastings (M-H) Monte Carlo via Cadeia de Markov (MCMC) padrão. Dessa forma o algoritmo ABC com métodos MCMC (ABC-MCMC) é dado na subseção seguinte.

2.3.1.3.1 Algoritmo ABC-MCMC

São necessários os dados observados $\mathbf{Y}$, a função distância $d(\cdot, \cdot)$, a expressão do resumo estatístico S , a função $\pi(\cdot|\theta^*)$, parâmetros da priori $\pi(\theta)$ e estado inicial θ^0 .

-
1. Seja $\theta^c = \theta^0$.
  2. Amostrar θ^* da distribuição proposta $\pi(\cdot|\theta^c)$.
  3. Simular K conjuntos de dados $\mathbf{Y}_1^*, \dots, \mathbf{Y}_K^*$ do modelo utilizando o parâmetro θ^* e calcular

$$r(Y, \theta^*) = \frac{1}{K} \sum_{k=1}^K \mathcal{I}(d(S_y, S_k^*)) \leq \delta,$$

em que $\mathcal{I}(E) = 1$ se E é verdadeiro, e 0 caso contrário.

4. Aceitar θ^* com probabilidade

$$\min \left(1, \frac{r(Y, \theta^*)}{r(Y, \theta^c)} \frac{\pi(\theta^*)}{\pi(\theta^c)} \frac{\pi(\theta^c|\theta^*)}{\pi(\theta^*|\theta^c)} \right).$$

5. Se θ^* é aceito, então tome $\theta^c = \theta^*$.
  6. Registre o estado atual.
  7. Retorne ao passo 2 e repita até que as amostras posteriori requerida sejam obtidas.
-

O algoritmo ABC-MCMC é muito similar ao algoritmo M-H padrão. A única, mas crucial diferença é na razão da probabilidade de aceitação. No algoritmo M-H padrão a razão das verossimilhanças é avaliada no valor candidato atual da cadeia de Markov, enquanto que no ABC-MCMC cada termo da verossimilhança é aproximado pela proporção de K simulações que “condizem com os dados” (KYPRAIOS *et al.*, 2017). MCKINLEY *et al.* (2009) demonstraram que a repetição da simulação dos dados dado θ não parece contribuir muito na melhora da aproximação da posteriori pela amostragem ABC.

Uma revisão sobre os algoritmos ABC-MCMC pode ser encontrada em Sisson e FAN (2011). Uma dificuldade que muitas vezes se enfrenta ao utilizar o ABC-MCMC é que a seleção requer um número expansivo de simulações piloto, além de gerar amostras correlacionadas (BEAUMONT, 2003).

2.3.1.4 Sequencial

Sisson *et al.* (2007) propuseram a inclusão do Sequencial Monte Carlo (SMC) dentro das etapas de simulação do algoritmo ABC (2.3.1.1.3) e depois deles, vários autores desenvolveram algoritmos com essa proposta. O mais utilizado e difundido que será apresentado nessa seção é o proposto por TONI *et al.* (2009). A ideia é amostrar de uma sequência de distribuições a posteriori ABC aproximadas sob tolerâncias de aceitação sucessivamente menores. A amostra de uma iteração, conhecida como partículas, é usada para ajudar a escolher quais valores de θ devem ser os próximos a serem simulados.

Os métodos SMC para o ABC têm uma série de potenciais vantagens em relação ao MCMC (2.3.1.3.1) e ao ABC rejeição (2.3.1.1.3). Diferentemente deste último, eles se concentram na simulação de conjuntos de dados de regiões de θ com probabilidades de aceitação relativamente altas, evitando desperdiçar recursos computacionais. Além disso ele pode adaptar a precisão do ajuste, como tolerâncias de aceitação, durante sua operação e resulta em distribuições a posteriori independentes, ao contrário do algoritmo MCMC. Este algoritmo é referido por algoritmo ABC-SMC (2.3.1.4.1).

Há também uma família mais ampla de algoritmos ABC-SMC relacionados, incluindo SISSON *et al.* (2007) e DEL MORAL *et al.* (2006), que atualizam suas partículas de maneiras mais complexas com base nos movimentos MCMC, conforme descrito em DEL MORAL *et al.* (2012).

2.3.1.4.1 Algoritmo ABC-SMC

São utilizados valores dos dados observados $\mathbf{Y}$, tolerâncias $1, \dots, M$, função distância $d(\cdot, \cdot)$, resumo estatístico S , número de partículas N , função *kernel* $K(\cdot|\cdot)$ e parâmetros da priori $\pi(\theta)$.

-
1. Seja $t = 1$.
  2. Repetir os passos a seguir até obter N aceitações.
    - (a) Amostrar θ^* da densidade $q_t(\theta)$ dado na equação (2.2) a seguir.
    - (b) Se $\pi(\theta^*) = 0$ rejeitar e voltar para (a).
    - (c) Simular o conjunto de dados Y^* a partir do modelo, utilizando os parâmetros θ^* .
    - (d) Aceitar θ^* se $d(S_y, S_y^*) \leq \varepsilon_t$.

Denotar por $\theta_1^t, \theta_2^t, \dots, \theta_N^t$ os valores de parâmetros aceitos.

3. Seja $w_i^t = \pi(\theta_i)/q_t(\theta_i)$ para $i = 1, 2, \dots, N$.
4. Incrementar $t = t + 1$.
5. Repetir os passos 2 e 3 até $t = M$.
6. Retornar a amostra final aceita e os pesos.

A densidade no passo 2a) é dada por

$$q_t(\theta) = \begin{cases} \pi(\theta) & \text{se } t = 1, \\ \sum_{i=1}^N w_i^{t-1} K_t(\theta|\theta_i^{t-1}) / \sum_{i=1}^N w_i^{t-1} & \text{caso contrário.} \end{cases} \quad (2.2)$$

Em outras palavras, na primeira iteração $q_t(\theta)$ é a priori e para as iterações subsequentes, as amostras ponderadas da população são descritas utilizando o *kernel* K_t .

A escolha do *kernel* é arbitrária mas BEAUMONT *et al.* (2009) ilustraram que uma boa opção é

$$K(\theta^t|\theta^{t-1}) = \phi(\theta^{t-1}, 2\Sigma^{t-1}),$$

em que $\phi(\cdot, \cdot)$ é a densidade de uma distribuição Normal (multivariada) e Σ^{t-1} é a matriz de covariância empírica dos elementos da população no tempo $t-1$, $\{\theta_i^{t-1}\}_{1 \leq i \leq N}$, calculado utilizando pesos $\{w_i^{t-1}\}_{1 \leq i \leq N}$.

Uma variação comum do algoritmo 2.3.1.4.1 (DROVANDI *et al.*, 2011), é determinar a sequência de tolerâncias adaptativas durante o algoritmo. Entretanto, KYPRAIOS *et al.* (2017) encontraram que as tolerâncias adaptativas as vezes funcionam mal para estatísticas resumo discretas. O problema ocorre quando a maioria das distâncias das simulações aceitas é exatamente igual à tolerância. Na prática, isso corresponde tipicamente a epidemias simuladas em que não ocorrem infecções.

Devido a flexibilidade dessa metodologia, algumas aplicações e expansões têm sido realizadas em modelos mais complexos com grande quantidade de parâmetros, ou onde a função de verossimilhança é intratável ou apresenta muitos máximos locais, como é o caso de modelos que envolvem Equações Diferenciais Estocásticas, EDE, (CAMPBELL *et al.*, 2012; KYPRAIOS *et al.*, 2017; BLUM e TRAN, 2010).

2.4 Equações Diferenciais

Uma equação diferencial (ED) é uma expressão matemática que relaciona uma função com suas derivadas, sejam ordinárias ou parciais. Nesse caso, a função que dá origem à equação é desconhecida, e tanto a função quanto suas derivadas devem aparecer na equação (STEWART, 2006).

As equações diferenciais estão presentes em diversas áreas como engenharia, física, economia e biologia e normalmente representam taxas. São essenciais para a descrição matemática da natureza, além de estarem presentes em muitas teorias físicas. Por exemplo, as equações de Newton e Lagrange são essenciais para a mecânica, as de Maxwell para o eletromagnetismo clássico, a de Schrödinger para a mecânica quântica, e a de Einstein para a teoria gravitacional.

Muitos estudos são realizados na matemática pura a fim de encontrar as soluções dessas equações. Apesar de somente equações diferenciais mais simples possuírem fórmulas explícitas, algumas propriedades de soluções de uma equação diferencial podem ser determinadas sem encontrar sua forma exata. Caso a equação diferencial não possua forma fechada, sua solução pode ser aproximada numericamente.

As equações diferenciais tiveram seu início com o desenvolvimento do cálculo por meio de Newton e Leibniz. No capítulo 2 do seu trabalho em 1671 “Methodus fluxionum et Serierum Infinitorum”, Newton listou três tipos de equações diferenciais, as quais ele resolveu utilizando series infinitas, além de discutir a não unicidade das soluções.

De acordo com HAIRER *et al.* (1993), em 1965 Bernoulli propôs as equações diferenciais de Bernoulli que são equações diferenciais ordinárias da forma

$$y' + P(x)y = Q(x)y^n,$$

a qual, no ano seguinte, Leibniz obteve soluções simplificando-a.

Vários autores propuseram diferentes métodos para a solução de diversas equações diferenciais, dentre eles, Fourier em 1822 publicou seu trabalho “Théorie analytique de la chaleur” o qual é baseado na lei de Newton sobre o resfriamento, ou seja, no fato de que o fluxo de calor entre duas moléculas adjacentes é proporcional a diferença de suas temperaturas. Este livro contém a proposta de Fourier com sua equação para a difusão de calor que é ensinada aos alunos de física e matemática.

São também exemplos clássicos de equações diferenciais abordadas nos cursos de graduação os modelos de crescimento populacional, que baseiam-se na premissa de que uma população cresce a uma taxa proporcional ao tamanho da populacional STEWART (2006). Nesse caso as variáveis do modelo são o tempo t e o número de indivíduos na população P . A taxa de crescimento populacional é dada por dP/dt e descrita como na equação

$$\frac{dP}{dt} = kP, \quad (2.3)$$

em que k é uma constante de proporcionalidade.

Desconsiderando a população nula, se $P(t) > 0, \forall t$ então, se $k > 0 \Rightarrow P'(t) > 0, \forall t$. Dessa forma, a população estará sempre aumentando, ou seja, a taxa de crescimento aumenta quando a população cresce.

A solução da equação (2.3) é dada por uma função cuja derivada é um múltiplo constante dela mesma. A função exponencial possui essa propriedade.

$$P'(t) = c(ke^{kt}) = k(ce^{kt}) = k(P(t)),$$

dessa forma, $P(t) = ce^{kt}$ é solução da equação diferencial.

Muitas populações começam a crescer exponencialmente, mas se estabilizam, por exemplo, devido a limitação de recursos. O nível da população se estabiliza quando ela se aproxima de sua capacidade suporte k . Considera-se as seguintes premissas:

$$\begin{aligned} \frac{dP}{dt} &\approx kP, & \text{se } P \text{ for pequeno,} \\ \frac{dP}{dt} &< 0, & \text{se } P > k. \end{aligned}$$

Uma expressão que incorpora ambas as premissas é a equação logística STEWART (2006),

$$\frac{dP}{dt} = kP \left(1 - \frac{P}{k} \right),$$

são soluções de equilíbrio da equação $P(t) = 0$ ou $P(t) = k$, pois em ambos os casos a população tende a permanecer dessa maneira. Se a população inicial for $0 < P(0) < k$ então $dP/dt > 0$ e a população aumenta. Se $P > k$ então $1 - P/k < 0$, logo $dP/dt < 0$ e a população diminui.

Outros exemplos de equações diferenciais são dados a seguir:

- (a) **Lei de Newton:** Força é igual a massa vezes aceleração, $f = ma$, em que m é massa, $a = d^2x/dt^2$, é a aceleração e f a força atuando na partícula. Dessa forma, a lei de Newton é a equação diferencial

$$m \frac{d^2x}{dt^2}(t) = f \left(t, x(t), \frac{dx}{dt}(t) \right),$$

em que $x(t)$ é a parte desconhecida da posição da partícula no espaço no tempo t . A força pode depender do tempo, da posição das partículas no espaço e da velocidade da partícula. Essa é uma equação diferencial ordinária (EDO) de segunda ordem.

- (b) **Decaimento radioativo:** A quantidade u de material radioativo muda com o tempo como descrito a seguir

$$\frac{du}{dt}(t) = -ku(t), \quad k > 0,$$

em que k é uma constante positiva representando propriedades radioativas do material. Essa é uma EDO de primeira ordem.

- (c) **A equação de calor:** A temperatura T em um material sólido muda com o tempo e em três dimensões espaciais, denotada por $w = (x, y, z)$, de acordo com a equação

$$\frac{dT}{dt}(t, w) = k \left(\frac{\partial^2 T}{\partial x^2}(t, w) + \frac{\partial^2 T}{\partial y^2}(t, w) + \frac{\partial^2 T}{\partial z^2}(t, w) \right), \quad K > 0,$$

em que k é uma constante positiva que representa propriedades termais do material. Essa é uma equação diferencial parcial (EDP) de primeira ordem no tempo e segunda ordem no espaço.

- (d) **Equação de ondas:** Uma perturbação de onda u propagando no tempo t e em três dimensões do espaço, denotada por $w = (x, y, z)$, com velocidade de onda $\nu > 0$ é uma EDP no espaço e tempo

$$\frac{du}{dt}(t, w) = \nu^2 \left(\frac{\partial^2 u}{\partial x^2}(t, w) + \frac{\partial^2 u}{\partial y^2}(t, w) + \frac{\partial^2 u}{\partial z^2}(t, w) \right).$$

Equação Diferencial Ordinária: Uma EDO é uma equação que contém uma função de uma variável independente e suas derivadas (STEWART, 2006). Exemplos de EDOs podem ser encontrados em (a) e (b), visto que a função desconhecida depende de uma única variável independente t . Frequentemente sistemas dinâmicos unidimensionais são modelados por EDOs.

Normalmente as EDOs lineares possuem soluções fechadas exatas enquanto EDOs não lineares raramente podem ser representadas em uma forma fechada, o que comumente conduz a soluções por meio de métodos numéricos que podem ser aplicados computacionalmente para aproximar as soluções. Alguns exemplos são o método de Euler e o algoritmo LSODA que utiliza os métodos de Adams e fórmulas de diferenciação regressivas (*Backward Differentiation Formulas* - BDF).

Equação Diferencial Parcial: Uma equação que contém funções multivariadas e suas derivadas parciais são chamadas de equações diferenciais parciais (EDP). São utilizadas para formular problemas envolvendo funções com muitas variáveis e podem ser resolvidas analiticamente ou por métodos iterativos. As equações (c) e (d) são exemplos de EDPs uma vez que a função desconhecida depende de duas ou mais variáveis independentes t, x, y e z e suas derivadas parciais aparecem nas equações.

Além de modelar sistemas multidimensionais, as EDPs são amplamente empregadas para a descrição de uma variedade de fenômenos como som, calor, eletrostática, eletrodinâmica, fluxo de fluídos, elasticidade e mecânica quântica.

Ordem da Equação Diferencial: A ordem de uma ED é a maior ordem de derivada que aparece na equação. A equação de Newton no exemplo (a) é de segunda ordem, a equação decaimento do tempo no exemplo (b) é de primeira ordem, a equação de calor no exemplo (c) é de primeira ordem no tempo e de segunda ordem no espaço, e a equação de ondas no exemplo (d) é de segunda ordem para as variáveis tempo e espaço.

Equação Diferencial Linear: Uma ED é linear (EDL) se a função desconhecida e suas derivadas possuem grau 1, caso contrário ela é chamada equação diferencial não linear. Sua definição é dada da seguinte forma

Definição 4: Uma EDO de primeira ordem do termo desconhecido y é

$$y'(t) = f(t, y(t))$$

em que f é dado e $y' = \frac{dy}{dt}$. A equação é linear se a função f é linear no seu segundo argumento,

$$y' = a(t)y + b(t)$$

A equação linear possui coeficientes constantes se ambos a e b são constantes. Caso contrário a equação possui coeficientes variáveis.

2.4.1 Métodos para Resolução de Equações Diferenciais

A resolução de equações diferenciais envolve diretamente o cálculo de integrais. Para modelos simples, formas explícitas podem ser obtidas por meio, por exemplo, de equações separáveis e equações lineares (STEWART, 2006; BOYCE e DIPRIMA, 2012). Uma equação separável consiste em uma diferencial de primeira ordem, na qual a expressão

dy/dx pode ser fatorada como uma função de x multiplicada por uma função de y . Em outras palavras, pode ser escrita como

$$\frac{dy}{dx} = k(x)l(y), \text{ ou ainda, } \frac{dy}{dx} = \frac{k(x)}{q(y)}, \text{ em que } q(y) = \frac{1}{l(y)}.$$

Para resolver essa equação, basta reescrevê-la na forma diferencial, desmembrando os termos dependentes de y e x em lados opostos da igualdade, permitindo assim, a integração da equação como se segue

$$\int q(y)dy = \int k(x)dx.$$

Uma equação diferencial linear de primeira ordem é aquela que pode ser escrita na forma

$$\frac{dy}{dx} + P(x)y = Q(x),$$

em que P e Q são funções contínuas em um dado intervalo. Um exemplo de uma equação linear de primeira ordem é $xy' + y = 2x$ porque, para $x \neq 0$, esta pode ser escrita na forma

$$y' + \frac{1}{x}y = 2.$$

Note que essa equação diferencial não é separável mas pode ser resolvida utilizando-se a Regra do Produto,

$$xy' + y = (xy)',$$

e assim reescrever a equação como

$$(xy)' = 2x.$$

Integrando dos dois lados da equação tem-se

$$y = x + C/x,$$

sendo C uma constante de integração qualquer.

De acordo com STEWART (2006) toda a equação diferencial linear de primeira ordem pode ser resolvida de maneira similar pela multiplicação em ambos os lados da equação por uma função adequada de $I(x) = \exp\{\int P(x)dx\}$ chamada fator integrante.

Infelizmente na maioria das vezes não é possível obter fórmulas explícitas para a solução das equações diferenciais, sendo necessários métodos numéricos. Um dos métodos mais antigos e amplamente empregados é o método de Euler (BOYCE e DiPRIMA, 2012; KEELING e ROHANI, 2008), em que os dois primeiros termos da expansão em série de Taylor da equação são utilizados para a obtenção de uma solução aproximada

$$y(x+h) \approx y(x) + \Delta y'(x, y),$$

no qual Δ é o passo ou deslocamento estabelecido para a aproximação. Para um melhor entendimento do método um exemplo é apresentado no Apêndice A.1.

Segundo CHAPRA (2012) a solução numérica de equações diferenciais envolvem basicamente dois tipos de erros: os causados pelo número limitado de dígitos significativos que podem ser mantidos por um computador e os erros causados pela natureza da técnica empregada para aproximar valores de y . O primeiro caso é solucionado com a evolução constante dos computadores, já no segundo caso, esforços têm sido realizados com o intuito de se trabalhar com mais elementos da série de Taylor, ou como no caso do método Runge-Kutta, em que derivadas de ordem superior são evitadas devido sua complexidade e uma aproximação é realizada apenas com combinações da função original. Além disso, algumas funções podem ser consideradas rígidas (*stiff equations*), ou seja, são numericamente instáveis, a menos que pequenos passos sejam tomados para sua aproximação (CHAPRA, 2012). Dessa forma, métodos que avaliam o comportamento da curva a cada passo tem sido amplamente utilizados, como o caso do LSODA desenvolvido por (PETZOLD, 1983; HINDMARSH, 1983). Esta abordagem alterna entre o método de Adam, que apresenta passos maiores e inicia o processo iterativo, com a fórmula de diferença reversa (*backward differentiation formula*), que apresenta passos menores, a fim de melhorar tanto a aproximação quanto o custo computacional envolvido. Na biblioteca `deSolve` do *software* R podem ser encontrados os algoritmos abordados nessa seção, sendo o LSODA a opção padrão de resolução.

2.4.2 Equações Diferenciais Estocásticas

Modelos de dinâmica populacional possuem tanto componentes determinísticos quanto estocásticos que operam simultaneamente. Uma equação diferencial estocástica (EDE) é aquela em que um ou mais termos são processos estocásticos, ou seja, é usualmente definida como a coleção de variáveis aleatórias que resulta em uma solução na qual também é um processo estocástico. Normalmente as EDEs possuem variáveis que representam um ruído aleatório calculado como a derivada de um movimento Browniano. Entretanto, outros tipos de comportamentos são possíveis.

A seguir pode ser observada uma equação diferencial ordinária estocástica (EDOE) definida por meio de um movimento Browniano.

$$\frac{dx(t)}{dt} = f(t, x), \quad dx(t) = f(t, x)dt,$$

com condições iniciais $x(0) = x_0$ que podem ser escritas na forma de integral

$$x(t) = x_0 + \int_0^t f(s, x(s))ds, \quad (2.4)$$

em que $x(t, x_0, t_0)$ é a solução com condições iniciais $x(t_0) = x_0$. Um exemplo é dado como

$$\frac{dx(t)}{dt} = a(t)x(t), \quad x(0) = x_0. \quad (2.5)$$

Ao tomar a EDO (2.5) e assumir que $a(t)$ não é um parâmetro determinístico mas sim estocástico, se tem uma EDE. O parâmetro estocástico $a(t) = dX^2(t)/dt$ é dado por

$$a(t) = f(t) + h(t)\xi(t),$$

em que, $\xi(t)$ denota um ruído.

Portanto, obtém-se

$$\frac{dX(t)}{dt} = f(t)X(t) + h(t)X(t)\xi(t). \quad (2.6)$$

Escrevendo (2.6) na forma diferencial e utilizando $dW(t) = \xi(t)dt$, em que $dW(t)$ denota a forma diferencial de um movimento Browniano, obtém-se

$$dX(t) = f(t)X(t)dt + h(t)X(t)dW(t).$$

E geram uma EDE dada como

$$dX(t, w) = f(t, X(t, w))dt + g(t, X(t, w))dW(t, w).$$

em que w denota que $X = X(t, w)$ é uma variável aleatória e possui a condição inicial $X(0, W) = X_0$ com probabilidade um. No caso do exemplo

$$dY(t, w) = \mu(t)dt + \sigma(t)dW(t, w). \quad (2.7)$$

Além disso, $f(t, X(t, w)) \in \mathbb{R}$, $g(t, X(t, w)) \in \mathbb{R}$ e $W(t, w) \in \mathbb{R}$. Da mesma forma que em (2.4), (2.7) pode ser escrita na forma de integrais

$$X(t, w) = X_0 + \int_0^t f(s, X(s, w))ds + \int_0^t g(s, X(s, w))dW(s, w).$$

Uma das formas de se definir uma EDE é como mencionado anteriormente por um movimento Browniano, entretanto LANDE *et al.* (2003) e TURCHIN (2003) citam que existem pelo menos outras três formas diferentes de incluir aleatoriedade ou estocasticidade.

1. A aproximação mais direta, quando os dados são observados de forma regular e discreta no tempo $t = 0, 1, 2, \dots$, é adicionando ruído a taxa de mudança populacional, isto é:

$$r_{t+1} = f(\mathbf{Z}_t) + \varepsilon_t,$$

em que $r_{t+1} = \log(N_{t+1}/N_t)$, é o vetor de estado das variáveis e ε_t é uma variável aleatória com alguma distribuição de probabilidade.

2. Uma segunda aproximação para incluir estocasticidade no modelo é adicionar um componente aleatório diretamente em N_t . Um possível mecanismo que suporta essa aproximação é a inclusão de eventos de imigração.

3. A terceira aproximação é variar aleatoriamente os parâmetros do modelo. Essa aproximação é útil quando pesquisadores sabem qual parte da população processa os efeitos ambientais mais importantes.

Analisar dados de dinâmica populacional é um problema não padrão, e muitas aproximações têm sido desenvolvidas. Em geral, a inferência para estes problemas é complicada pelos fatos que (i) os dados são inerentemente dependentes e (ii) os dados são usualmente incompletos no sentido de que o processo real de infecção não é observado (O'NEILL, 2002).

A vantagem desses modelos é que dependendo da aplicação, eles podem incorporar períodos latente, a variável infectividade, a redução da susceptibilidade após a recuperação, etc. Similarmente, aspectos como heterogeneidade da população, estrutura etária e susceptibilidade podem ser propriamente incluídos.

A inferência por máxima verossimilhança, bem como a bayesiana requer um modelo de amostragem que produz a verossimilhança e a distribuição condicional dos dados dado o modelo dos parâmetros. Dessa forma, a inferência estatística para esses modelos pode ser um desafio uma vez que as verossimilhanças associadas podem ser computacionalmente custosas ou mesmo não ser acessível de forma tratável. Recentemente o ABC surgiu na tentativa de resolver esses problemas.

2.4.3 Sistemas Dinâmicos

A modelagem de “sistemas dinâmicos” utilizando espécies de plantas e animais herbívoros surgiram como um exemplo em LOTKA (1920) e em 1925 ele utilizou equações diferenciais para analisar a interação entre presa e predador em seu livro “Elementos da biologia física”. VOLTERRA (1926) também desenvolveu alguns conceitos para um modelo que descreve a evolução de um predador e uma presa em populações de peixes no mar Adriático. Depois disso, esses modelos passaram a ser chamados de modelos Lotka-Volterra como é conhecido até os dias atuais.

KERMACK e MCKENDRICK (1927) estenderam esses modelos para a área epidêmica e desde então, as atenções têm se voltado para a modelagem de epidemias (CLÉMENÇON *et al.*, 2008).

2.4.3.1 Modelos Epidemiológicos

Predizer o curso futuro de uma epidemia, compreender as estruturas nas quais doenças se disseminam e como controlá-las, bem como comparar e avaliar políticas públicas de saúde são os principais objetivos de técnicas matemáticas e modelos utilizados no estudo de epidemias.

Quando soluções numéricas são possíveis, modelos determinísticos são amplamente utilizados, entretanto precisam ser simples. Eles podem ser um guia para descrever

epidemias, entretanto por causa da necessidade de quantificar a precisão das estimativas e a variação nos dados, os modelos estocásticos são o caminho mais natural para prever dados de doenças infecciosas (BECKER e BRITTON, 1999).

Essencialmente modelos em epidemiologia são chamados compartimentais e expressos como sistemas de equações matemáticas os quais descrevem como a população é estruturada em diferentes classes (BLUM e TRAN, 2010). Nesse sentido é possível estudar como um sistema muda de um compartimento para o próximo além das relações entre as variáveis definidas nas equações (HENS *et al.*, 2012).

O número de indivíduos em cada compartimento do modelo pode ser calculado por meio da resolução do sistema de equações. O principal problema é que normalmente alguns parâmetros são desconhecidos. ROSS (1916) propôs uma solução para este problema, em que se assume um conhecimento de causas para construir as equações diferenciais, e testar os resultados calculados por meio da comparação com as estatísticas observadas.

Modelo SIR

Os modelos compartimentais são ferramentas atrativas que ajudam na compreensão da dinâmica de doenças (CORREIA-GOMES *et al.*, 2014). Para os indivíduos Suscetíveis, Infectados, e Recuperados (ou Removidos) (SIR) é formulado o modelo de mudanças de indivíduos na população entre diferentes estados da doença em termos de um sistema de equações diferenciais ordinárias (EDO), conhecido como modelo de Kermack e McKendrick (DIEKMANN e HEESTERBEEK, 2000).

Se essa modelagem for útil no controle de doenças infecciosas, é fundamental ter as melhores estimativas possíveis das taxas de transição de um compartimento para outro e também para modelar a propagação da infecção.

O tamanho populacional é N , em que $N = S + I + R$. Na Figura 2.4 é possível observar um esquema para um modelo SIR básico sem dinâmica vital, ou seja, sem taxas de migração, mortalidade ou nascimento durante o período considerado. A EDO que representa este modelo é mostrada nas equações (2.8). Os parâmetros ν e $\beta(t)$ são taxas de recuperação e transmissão por tempo, respectivamente. Para os exemplos tratados nessa tese, o período de infecção é de somente 15 e 21 dias, então a taxa de transmissão $\beta(t)$ é considerada constante β .

Figura 2.4: Esquema para o modelo compartimental SIR sem dinâmica vital

$$\begin{cases} \frac{dS(t)}{dt} = -\beta(t)S(t)I(t), \\ \frac{dI(t)}{dt} = \beta(t)S(t)I(t) - \nu I(t), \\ \frac{dR(t)}{dt} = \nu I(t). \end{cases} \quad (2.8)$$

O modelo SIR sem dinâmica vital pode ser aplicado somente quando a doença estudada acontece durante um curto período de tempo. Para doenças em que a duração é maior, meses ou anos, o mais adequado é a utilização do modelo SIR com dinâmica vital, porque nesse caso há necessidade de inclusão do efeito demográfico.

Uma taxa importante a ser considerada em epidemiologia é a taxa de reprodução básica $\mathcal{R}_0$, que é definida como o número esperado de casos secundários por caso primário durante seu período infeccioso (DIEKMANN e HEESTERBEEK, 2000).

Para o modelo SIR, podem ser observadas quais são as condições para uma epidemia. Uma epidemia ocorre se o número de indivíduos infectados aumenta, isto é, $\frac{di}{dt} > 0$.

$$\beta SI - \nu I > 0 \Rightarrow \frac{\beta SI}{\nu} > I.$$

No início de uma epidemia, quase todos são suscetíveis. Então é possível dizer que $S \approx 1$. Substituindo $S = 1$, obtém-se a inequação a seguir

$$\frac{\beta}{\nu} = \mathcal{R}_0 > 1,$$

$\mathcal{R}_0 > 1$ indica uma epidemia persistente, enquanto que, quando $\mathcal{R}_0 < 1$ a epidemia irá desaparecer.

Modelo SIR com dinâmica vital

Considerando que existem novos suscetíveis entrando e saindo da população, estes devem ser incluídos no modelo. Nesse modelo cujo fluxograma é apresentado na Figura 2.5 inclui-se taxas de nascimento e mortalidade. Os recém-nascidos estão dentro da classe de suscetíveis. Supõe-se que as taxas tanto de natalidade quanto mortalidade são μ , o que mantém o tamanho populacional constante.

Figura 2.5: Esquema para o modelo compartimental SIR com dinâmica vital

Sendo $N = S + I + R$ o tamanho total da população, as equações diferenciais dadas a seguir representam esse modelo

$$\begin{cases} \frac{dS}{dt} = \mu N - \beta(t)SI - \mu S, \\ \frac{dI}{dt} = \beta(t)SI - \nu I - \mu I, \\ \frac{dR}{dt} = \nu I - \mu R. \end{cases}$$

2.4.3.2 Modelos de Interação Populacional

Normalmente as espécies não estão sozinhas e, ao contrário, normalmente existem várias espécies interagindo no mesmo habitat. Dessa forma, a dinâmica populacional de cada espécie afeta a relação entre elas. Foram evidenciadas duas das três principais formas de interação definidas por MURRAY (2002):

1. Se a densidade de uma população diminui enquanto a da outra aumenta, as populações estão em uma situação *presa-predador*.
2. Se as taxas de crescimento de ambas as populações decrescem simultaneamente, então existe *competição*.
3. Se as taxas de crescimento de ambas as populações aumentam simultaneamente então é chamado *simbiose* ou *mutualismo*.

Modelo Presa-Predador

LOTKA (1925) e VOLTERRA (1926) foram os primeiros a introduzir um modelo que acessa a relação entre presas e predadores por meio de suposições simples. Para isso, assume-se que exceto pela presença de predadores, presas são capazes de sobreviver satisfatoriamente no ambiente sem fontes de limitações e predadores são limitados somente pela disponibilidade de presas. O modelo preliminar foi construído de forma a ser auto sustentável infinitamente, gerando ciclos periódicos. Entretanto, para produzir um modelo mais realístico, mais suposições foram feitas com a inclusão de taxas que descrevem a dinâmica do sistema biológico. Assim, o modelo é dado pela equação diferencial como

$$\begin{cases} \frac{dx(t)}{dt} = ax - cxy, \\ \frac{dy(t)}{dt} = bxy - dy, \end{cases} \quad (2.9)$$

em que dy/dt e dx/dt representam as taxas de crescimento das duas populações ao longo do tempo, x é o número de presas, y é o número de predadores, t é o evento tempo, ax é o crescimento da presa, cxy é a taxa com que predadores e presas se encontram, bxy

representa o crescimento dos predadores e dy representa a taxa de perda dos predadores devido à morte natural ou à emigração.

Para esse modelo, o equilíbrio populacional ocorre quando nenhum dos níveis da população está mudando. Isso pode acontecer de duas maneiras: com a extinção de ambas as espécies

$$\begin{aligned}x(a - cy) &= 0, \\ -y(d - bx) &= 0,\end{aligned}$$

ou quando ambas as populações sustentam seus números atuais diferentes de zero com valores específicos dos parâmetros a, b, c e d

$$y = \frac{a}{c} \text{ e } x = \frac{d}{b}.$$

Modelo de Competição

Para construir esse modelo são necessárias duas suposições:

1. Na ausência de outra, a taxa de mudança populacional utiliza um modelo logístico;
2. A taxa de mudança em cada população decresce a uma taxa proporcional ao número de interações entre espécies.

Sejam $x(t)$ e $y(t)$ duas espécies. Então essas premissas produzem o sistema de equações diferenciais, em que as constantes são assumidas não negativas.

$$\begin{cases} \frac{dx(t)}{dt} = x(a_1 - b_1x - c_1y), \\ \frac{dy(t)}{dt} = y(a_2 - b_2y - c_2x). \end{cases} \quad (2.10)$$

Alguns equilíbrios podem ser encontrados. A origem $(0, 0)$ bem como os pontos de intersecção das retas são equilíbrios. Se $x = 0$ (ou $y = 0$), então

$$x = 0 \Rightarrow y' = y(a_2 - b_2y) \quad \text{ou} \quad y = 0 \Rightarrow x' = x(a_1 - b_1x).$$

Modelo de Mutualismo ou Simbiose

Existem muitos exemplos em que a interação de duas ou mais espécies é vantajosa para todas. Mutualismo ou simbiose muitas vezes desempenha o papel crucial na promoção e manutenção dessas espécies: a dispersão de plantas e sementes é um exemplo. Mesmo que a sobrevivência não esteja em jogo, a vantagem mútua do mutualismo ou da simbiose pode ser muito importante. Como um tema de ecologia teórica, mesmo para

duas espécies, essa área não foi tão amplamente estudada quanto as outras, embora sua importância seja comparável à das interações predador-presa e competição.

$$\begin{cases} \frac{dx(t)}{dt} = a_1x + b_1xy, \\ \frac{dy(t)}{dt} = a_2y + b_2xy, \end{cases}$$

em que a_1, a_2, b_1 e b_2 são constantes positivas. Como $dx/dt > 0$ e $dy/dt > 0$, x e y simplesmente crescem sem limites.

2.5 Inferência para Sistemas Dinâmicos

O problema de estimação dos parâmetros para sistemas dinâmicos aparece em muitas áreas como finanças, física, química, engenharia, biologia, neurociência entre outras. Frequentemente o modelo é um sistema de equações diferenciais estocástico que pode derivar de algum conhecimento sobre o processo sob investigação, mas os parâmetros devem ser inferidos a partir de observações dos dados. O conhecimento desses parâmetros é fundamental para o estudo do processo em questão e, portanto, sua estimativa com base nos dados observacionais é de grande importância.

A modelagem de doenças infecciosas é realizada a fim de estudar os mecanismos nos quais as doenças se espalham. Dessa forma, com a predição do curso futuro de um surto, podem ser criadas estratégias de controle da epidemia bem como políticas públicas de intervenção (DALEY e GANI, 2001). Além de serem utilizados para compreender a disseminação de parasitas, vírus e doenças (SINGH e UYENOYAMA, 2004), sistemas dinâmicos também são considerados quando há interação entre espécies, e como consequência observa-se como a população de cada espécie é afetada. Em geral, existe uma rede inteira de espécies que interagem, o que faz comunidades estruturalmente complexas.

A modelagem das interações existentes na natureza permitem uma melhor compreensão de como as relações funcionam, podendo auxiliar no entendimento de como a quantidade de indivíduos de uma espécie muda ao longo do tempo em relação a outra (WORSTER, 1994). Esses modelos são usados para entender a dinâmica das invasões biológicas e para a conservação ambiental.

Os métodos usualmente utilizados para a inferência desses sistemas são a estimação por máxima verossimilhança e via MCMC.

Considerando sistemas dinâmicos com vetor de parâmetros θ , o modelo matemático é dado da seguinte forma:

$$\frac{d\mathbf{x}}{dt}(t) = \mathbf{g}(t, \mathbf{x}(t), \boldsymbol{\theta}), \quad (2.11)$$

com o processo observado

$$\mathbf{y}(t) = C\mathbf{x}(t; \boldsymbol{\theta}).$$

O modelo matemático é um sistema determinístico, que para essa tese são equações diferenciais ordinárias. Supõe-se que o sistema possui soluções únicas que dependem dos parâmetros e dos dados iniciais. Seguindo a convenção usual (que corresponde à forma de dados geralmente disponível a partir de experiências), é assumida uma forma discreta das observações as quais possuem n observações longitudinais correspondentes a

$$\mathbf{y}(t_j) = C\mathbf{x}(t_j; \boldsymbol{\theta}),$$

em que C é um operador de observação descrito a seguir. Em geral, as observações ou dados correspondentes y_j não serão exatamente iguais para os diferentes tempos $y(t_j)$. Devido à natureza dos fenômenos que levam a essa discrepância, essa incerteza relativa às observações é tratada com um modelo estatístico para o processo de observação dado por

$$\mathbf{Y}_j = \mathbf{f}(t_j, \boldsymbol{\theta}_0) + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, n,$$

em que $\mathbf{f}(t_j, \boldsymbol{\theta}) = C\mathbf{x}(t_j, \boldsymbol{\theta})$, $j = 1, \dots, n$, corresponde a solução do modelo matemático (2.11) no j -ésimo tempo de observação para um vetor de parâmetros $\boldsymbol{\theta} \in \mathbb{R}^p$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f} \in \mathbb{R}^m$ e C é uma matriz $m \times n$. O termo $\boldsymbol{\theta}_0$ representa o verdadeiro vetor de parâmetros vindos das observações $\{\mathbf{Y}_j\}_{j=1}^n$. Os termos $\boldsymbol{\varepsilon}_j$ são variáveis aleatórias que podem representar o erro mensurado. Uma vez que esses erros são desconhecidos, assume-se que $\boldsymbol{\varepsilon}_j$ é gerado a partir de uma distribuição de probabilidade, usualmente $N(0, \sigma^2)$.

Para estimar o parâmetro $\hat{\boldsymbol{\theta}}$ para $\boldsymbol{\theta}_0$ técnicas utilizando máxima verossimilhança podem ser utilizadas. Supondo $m = 1$ com erro gerado por uma distribuição normal $\varepsilon_j \sim N(0, \sigma_0^2)$ e assim, $Y_j \sim N(f(t_j, \boldsymbol{\theta}_0), \sigma_0^2)$. Dessa forma a função de verossimilhança com $\varepsilon_j = Y_j - f(t_j, \boldsymbol{\theta}_j)$ é definida por CHOWELL *et al.* (2009)

$$L(\boldsymbol{\theta}; \sigma^2 | \mathbf{Y}) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} [Y_j - f(t_j, \boldsymbol{\theta})]^2 \right\}. \quad (2.12)$$

O máximo da equação (2.12) culminam nos estimadores de máxima verossimilhança θ_{MV} e σ_{MV}^2 para $\boldsymbol{\theta}_0$ e σ^2 . As soluções correspondentes as estimativas por máxima verossimilhança são obtidas por meio da maximização de 2.12 com $\mathbf{Y} = (Y_1, \dots, Y_n)^T$.

Maximizar (2.12) é equivalente a maximizar o logaritmo da função de verossimilhança

$$\log L(\boldsymbol{\theta}, \sigma^2 | \mathbf{Y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{j=1}^n [Y_j - f(t_j, \boldsymbol{\theta})]^2. \quad (2.13)$$

O máximo da equação (2.13) pode ser determinado por meio da diferenciação com respeito a $\boldsymbol{\theta}$ (com σ^2 fixado) e com respeito a σ^2 (com $\boldsymbol{\theta}$ fixado) fazendo o resultado das equações igual a zero e resolvendo para $\boldsymbol{\theta}$ e σ^2 . Dessa forma,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}, \sigma^2 | \mathbf{Y}) = \sum_{j=1}^n [Y_j - f(t_j, \boldsymbol{\theta})] \frac{\partial}{\partial \boldsymbol{\theta}} f(t_j, \boldsymbol{\theta}) = 0 \quad (2.14)$$

Resolver a equação (2.14) que é o mesmo que a otimização por mínimos quadrados

$$\theta_{MV}(\mathbf{Y}) = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{j=1}^n [Y_j - f(t_j, \boldsymbol{\theta})]^2.$$

Da mesma forma, fixando $\boldsymbol{\theta}$ e fazendo

$$\frac{\partial}{\partial \sigma^2} \log L(\boldsymbol{\theta}, \sigma^2 | \mathbf{Y}) = 0,$$

obtém-se,

$$\sigma_{MV}^2(\mathbf{Y}) = \frac{1}{n} \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{j=1}^n [Y_j - f(t_j, \boldsymbol{\theta})]^2.$$

Os métodos bayesianos também podem ser empregados para esses modelos. Neste caso a verossimilhança, $L(\boldsymbol{\theta}, \sigma^2 | \mathbf{Y})$, é complementada com a informação a priori, $\pi(\boldsymbol{\theta})$, disponível dos parâmetros para gerar uma distribuição a posteriori, $\pi(\boldsymbol{\theta} | \mathbf{Y})$, como descrito na equação (2.1).

Em casos em que a verossimilhança, é complexa nem sempre a solução analítica desses sistemas é viável. Entretanto, com os recentes avanços computacionais uma forma de contornar esse problema é utilizar métodos baseados em amostragem, dentre eles está o método MCMC. Essa abordagem permite obter amostras da distribuição a posteriori indiretamente, que possibilita o cálculo de estimativas de interesse dessa distribuição.

Para m iterações e p parâmetros desconhecidos simulam-se amostras $\{\theta^1, \dots, \theta^m\}$ em que $\theta^i = (\theta_1^i, \dots, \theta_p^i)$, $i = 1, \dots, m$ cuja distribuição de equilíbrio seja a posteriori. Assim, a estimativa de θ_j , $j = 1, \dots, p$ é dada por

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \theta_j^i = \hat{\theta}_j.$$

A cadeia de Markov é um processo no qual o próximo estado da cadeia, θ^{t+1} , depende somente do atual θ^t , dos dados e do estado inicial da cadeia θ^0 . Assim, poderá haver convergência para uma distribuição de equilíbrio. Os principais métodos de construção da cadeia consistem no amostrador de Gibbs (GEMAN e GEMAN, 1984) e o algoritmo Metropolis-Hasting (METROPOLIS *et al.*, 1953).

Algumas vezes a confiabilidade das estimativas para modelos de equações diferenciais estocásticas pode ser reduzida quando impactadas por mudanças nas condições

iniciais. Esses problemas tornam-se piores a medida que a dimensão do vetor de parâmetros desconhecido θ é aumentado gerando verossimilhanças complexas, tais como ondulações, cumes, seções planas e/ou com múltiplos máximos locais (CAMPBELL *et al.*, 2012; ESPOSITO e FLOUDAS, 2000).

Dadas as limitações da verossimilhança para esses modelos, métodos alternativos tem sido estudados, como é o caso do ABC. Essa forma de inferência para equações estocásticas já foi realizada por alguns autores com um sucesso considerável. BLUM e TRAN (2010) fizeram um estudo para a comparação de diferentes resumos estatísticos para um modelo Suscetível-Exposto-Infectado-Recuperado (SEIR) considerando dados de HIV-AIDS em Cuba, no qual foi avaliada a eficiência de predição da evolução da doença. LU *et al.* (2013) fizeram uma aplicação para dados de *Salmonella* Kentucky para um rebanho leiteiro adulto, no qual utilizam o modelo Suscetível-Infectado-Recuperado-Suscetível (SIRS) e calcularam a taxa de transmissão da doença, o período de infecção e a taxa básica de reprodução. KYPRAIOS *et al.* (2017) consideraram algumas aplicações, entre elas uma para um surto de varíola Abakaliki, e outra para gastroenterite.

No entanto, para problemas de interação populacional poucos trabalhos tem sido encontrados na literatura como é o caso de TONI *et al.* (2009) que fizeram uma análise utilizando o método ABC para dados simulados do modelo Lotka-Volterra com dois parâmetros desconhecidos. Nessa tese foi investigado o comportamento desses modelos por meio das três abordagens descritas (máxima verossimilhança, MCMC e ABC) quando o modelo possui mais parâmetros desconhecidos como apresentado na seção 4.2.

3 MATERIAIS E MÉTODOS

Nesse trabalho foram realizadas duas abordagens. A primeira, descrita na seção 3.1 consiste em um estudo de sensibilidade utilizando o método de regressão ABC para os modelos de regressão linear e Poisson no qual diferentes escolhas de priori, tolerâncias, distâncias e tamanhos amostrais foram avaliados e comparados com as estimativas obtidas por meio do método de máxima verossimilhança. Na seção 3.2, cinco estudos de caso envolvendo modelos epidemiológicos e de interação populacional, com dados reais e simulados, foram utilizados na verificação da performance dos métodos de máxima verossimilhança, MCMC e ABC com algoritmo de rejeição.

Para todos os casos estudados, a comparação entre os métodos de estimação por máxima verossimilhança, MCMC e ABC foi realizada pela medida de discrepância dada por

$$d_p = \sum_{j=1}^A (\hat{p}_j - p_v)^2, \quad (3.1)$$

em que $\hat{p}_j$ é a quantidade de interesse estimada na amostra j , $j = 1, \dots, A$, e p_v é o verdadeiro valor ou o valor observado da quantidade de interesse.

3.1 Estudos de Simulação

Para os estudos de simulação realizados nas subseções a seguir a inferência foi realizada por meio dos métodos da máxima verossimilhança e ABC com regressão local (2.3.1.2.1). Para o método de regressão ABC foi realizada uma análise de sensibilidade para verificar como os parâmetros do modelo, respondem a diferentes distribuições a priori. Foram simulados $A = 50$ conjuntos de dados para cada tamanho amostral ($N_1 = 50$ e $N_2 = 10$). Empregou-se as tolerâncias (0,01; 0,1 e 0,5) que representam a porcentagem de valores aceitos, e as diferentes distâncias métricas dadas a seguir (a definição de espaço euclidiano pode ser encontrado no Anexo A.1):

i) Distância euclidiana

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{\frac{1}{2}};$$

ii) Distância modular

$$d'(x, y) = |x_1 - y_1| + \dots + |x_n - y_n| = \sum_{i=1}^n |x_i - y_i|;$$

iii) Distância do máximo

$$d''(x, y) = \max\{|x_1 - y_1|, \dots, |x_n - y_n|\} = \max_{1 \leq i \leq n} |x_i - y_i|.$$

A comparação das estimativas por máxima verossimilhança com o ABC em cada cenário foi realizada por meio de uma medida de discrepância para cada parâmetro (Equação 3.1). Para a realização da inferência por meio do algoritmo regressão ABC, os resumos estatísticos utilizados para os parâmetros foram as estimativas por máxima verossimilhança.

Foram estudados os modelos de regressão linear simples e regressão Poisson com função de ligação canônica. Os principais aspectos das simulações são discutidos nas subseções seguintes. Todas as análises foram realizadas utilizando o *software* livre R (R CORE TEAM, 2015).

3.1.1 Regressão Linear

Primeiramente, utilizou-se um modelo de regressão linear em que a variável dependente y_i , mensurada na observação i , $i = 1, \dots, N$, apresenta uma relação linear com a variável regressora x_i . Esta relação é modelada por meio de um termo de perturbação ou erro ε_i , ou seja, uma variável aleatória não observada que adiciona um ruído à relação linear entre a variável dependente e regressora. Assim, o modelo assume a forma

$$y_i = \beta_0 + \beta_1 x_i \quad \varepsilon_i \sim N(0, \sigma^2), \quad (3.2)$$

em que β_0 e β_1 são parâmetros desconhecidos do modelo e σ^2 é a variabilidade do erro experimental ε_i . Com base na Equação 3.2, para N_1 e N_2 , 50 conjuntos de dados foram gerados (Tabelas 3.1 e 3.2 respectivamente) considerando arbitrariamente os parâmetros $\beta_0 = -3,7$, $\beta_1 = 2,3$, $\sigma^2 = 1$.

Tabela 3.1: Conjuntos de dados simulados para a regressão linear com tamanho amostral $N_1 = 50$

Observação	X	Y_1	Y_2	$\dots$	Y_{50}
1	0,0000	-3,5932	-3,4702	$\dots$	-4,1825
2	0,2041	-4,6633	-1,3680	$\dots$	-4,0571
3	0,4082	-2,2869	-3,6068	$\dots$	-3,7300
4	0,6122	-2,7803	-3,6673	$\dots$	-3,4097
5	0,8163	-0,0967	-1,3994	$\dots$	-1,1669
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
46	9,1837	17,0294	18,3598	$\dots$	16,8745
47	9,3877	18,8471	17,7638	$\dots$	17,8612
48	9,5918	18,5783	17,1598	$\dots$	17,4828
49	9,7959	18,5467	18,7235	$\dots$	18,5804
50	10,0000	19,1202	18,0173	$\dots$	18,5882

Tabela 3.2: Conjuntos de dados simulados para a regressão linear com tamanho amostral $N_2 = 10$

Observação	X	Y_1	Y_2	$\dots$	Y_{50}
1	0,0000	-4,3472	-2,0881	$\dots$	-5,0398
2	1,1111	-2,6077	-1,4459	$\dots$	-2,0154
3	2,2222	1,4437	0,8489	$\dots$	1,6342
4	3,3333	3,3793	1,7349	$\dots$	3,3256
5	4,4444	6,7879	6,1866	$\dots$	6,2578
6	5,5556	10,6965	7,9934	$\dots$	8,7938
7	6,6667	10,3151	10,3537	$\dots$	12,6823
8	7,7778	13,7540	15,3313	$\dots$	15,9338
9	8,8889	17,2384	17,5630	$\dots$	17,8970
10	10,0000	21,2717	19,4149	$\dots$	20,7343

Tabela 3.3: Tabela resumo para os casos de estudo da regressão linear

Cenários	Tamanho amostral	Priori para β_0 e β_1	Priori para σ^2
1	$N_1 = 50$	$N(0; 10^2)$ (vaga)	IGamma(10; 10) (informativa)
2		$N(0; 10^2)$ (vaga)	IGamma(0, 5; 0, 5) (vaga)
3		$N(0; 1)$ (informativa)	IGamma(10; 10) (informativa)
4		$N(0; 1)$ (informativa)	IGamma(0, 5; 0, 5) (vaga)
5		$N(7, 8; 6, 9^2)$ (informativa)	IGamma(10; 10) (informativa)
6		$N(7, 8; 6, 9^2)$ (informativa)	IGamma(0, 5; 0, 5) (vaga)
7	$N_2 = 10$	$N(0; 10^2)$ (vaga)	IGamma(10; 10) (informativa)
8		$N(0; 10^2)$ (vaga)	IGamma(0, 5; 0, 5) (vaga)
9		$N(0; 1)$ (informativa)	IGamma(10; 10) (informativa)
10		$N(0; 1)$ (informativa)	IGamma(0, 5; 0, 5) (vaga)
11		$N(7, 8; 7, 5^2)$ (informativa)	IGamma(10; 10) (informativa)
12		$N(7, 8; 7, 5^2)$ (informativa)	IGamma(0, 5; 0, 5) (vaga)

No método ABC, para cada um dos 50 conjunto de dados gerado, foram realizadas 50.000 simulações para cada combinação de distribuição a priori, tolerância e distância métrica. Utilizou-se para β_0 e β_1 uma distribuição a priori vaga proveniente da

distribuição normal $N(0; 10^2)$ e distribuições a priori informativas $N(0; 1)$, $N(7, 8; 6, 9^2)$ e $N(7, 8; 7, 5^2)$. Os valores 7, 8 e 6, 9² foram obtidos com base na média e variância dos 50 conjuntos gerados de tamanho N_1 e 7, 8 e 7, 5 com base na média e variância das 50 amostras de tamanho N_2 . Já para variância, foi utilizada uma distribuição a priori vaga proveniente da distribuição gama inversa $\text{IGamma}(0, 5; 0, 5)$ e distribuições a priori informativas $\text{IGamma}(10; 10)$. Os cenários avaliados para este caso podem ser consultados na Tabela 3.3.

3.1.2 Regressão Poisson

Em estatística, a regressão Poisson é a forma linear generalizada do modelo de análise de regressão, que é a escolha natural para análise de dados de contagem e tabelas de contingência. Seja Y_i uma variável aleatória com distribuição Poisson de média λ_i , denotada por $Y_i \sim \text{Poi}(\lambda_i)$. A função de probabilidade é escrita como

$$f(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0, 1, 2, \dots, \quad (3.3)$$

com função de ligação canônica e preditor linear dado por $\eta_i = \log(\lambda_i) = \beta_0 + \beta_1 x_i$. De forma análoga ao estudo de regressão linear, 50 conjuntos de dados para dois tamanhos amostrais, N_1 e N_2 , foram gerados (Tabela 3.4) para o modelo na Equação 3.3 considerando arbitrariamente os parâmetros $\beta_0 = 0, 5$ e $\beta_1 = 2$.

Tabela 3.4: Conjuntos de dados simulados para a regressão Poisson com tamanhos amostrais $N_1 = 50$ e $N_2 = 10$

Obs.	N_1	X	Y_1	Y_2	$\dots$	Y_{50}	Obs.	N_2	X	Y_1	Y_2	$\dots$	Y_{50}
1	0,0000	9	6	$\dots$	7	1	0,0000	9	11	$\dots$	4		
2	0,1020	5	9	$\dots$	10	2	0,5556	11	13	$\dots$	8		
3	0,2041	10	7	$\dots$	13	3	1,1111	9	13	$\dots$	10		
4	0,3061	14	10	$\dots$	14	4	1,6667	15	20	$\dots$	22		
5	0,4082	7	11	$\dots$	8	5	2,2222	23	19	$\dots$	16		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	6	2,7778	29	21	$\dots$	38		
47	4,6939	78	73	$\dots$	63	7	3,3333	38	45	$\dots$	38		
48	4,7959	77	91	$\dots$	93	8	3,8889	58	44	$\dots$	57		
49	4,8980	75	69	$\dots$	81	9	4,4444	66	71	$\dots$	67		
50	5,0000	95	79	$\dots$	93	10	5,0000	86	93	$\dots$	65		

Neste caso, foram consideradas distribuições a priori vagas com distribuições uniformes $U(0; 4)$ e $U(-1, 5; 2, 5)$ para β_0 e β_1 respectivamente e informativa $N(0; 1)$ para ambos. Não foi possível utilizar como distribuição a priori vaga a distribuição $N(0; 10^2)$ devido aos elevados valores obtidos, impossibilitando assim, os cálculos pelo *software* R (R CORE TEAM, 2015). O desempenho do ABC foi comparado com o da máxima verossimilhança por meio da medida de discrepância (Equação 3.1). Os cenários avaliados para este caso podem ser consultados na Tabela 3.5.

Tabela 3.5: Tabela resumo para os casos de estudo da regressão Poisson

Cenário	Tamanho amostral	Priori para β_0 e β_1
1	$N_1 = 50$	$N(0; 1)$ (informativa)
2		$U(0; 4)$, $U(-1, 5; 2, 5)$ (vaga)
3	$N_2 = 10$	$N(0; 1)$ (informativa)
4		$U(0; 4)$, $U(-1, 5; 2, 5)$ (vaga)

3.2 Modelos de Dinâmica Populacional

Para todos os modelos de dinâmica populacional estudados, para a realização da estimação dos parâmetros e predição das taxas como o número de infectados e recuperados para modelos epidemiológicos e número de presas e predadores para modelos populacionais, foram consideradas as metodologias da máxima verossimilhança, MCMC e ABC com o algoritmo (2.3.1.1.2), ou seja, devido a dificuldade de obtenção de resumos estatísticos para os parâmetros de EDOE (seção 2.4.2), o ABC é realizado por meio da comparação do conjunto de dados completo. Além disso, para a solução do sistema de equação diferencial ordinária, foi utilizado o algoritmo LSODA de HINDMARSH (1983) e PETZOLD (1983). A medida de discrepância (Equação 3.1) entre valores ajustados e observados foi obtida a fim de realizar um diagnóstico do modelo ajustado.

3.2.1 Modelos Epidemiológicos

Para os dois casos de estudo para modelos epidemiológicos tratados nesta seção, foi considerado o modelo SIR sem dinâmica vital (2.8) devido ao fato de o tempo observacional ser curto (15 dias para o primeiro caso e 21 para o segundo) e dessa forma não haver variações populacionais consideráveis neste período. O modelo SIR pode ser utilizado para estimar as taxas de infecção, recuperação e reprodução da doença (β , ν e $\mathcal{R}_0$) além do período de infecção, $1/\nu$.

O número básico de reprodução, $R_0 = \beta/\nu$, é definido como o número esperado de novos casos por pessoa infectada durante o período infeccioso (DIEKMANN e HEES-

TERBEEK, 2000). Quando $R_0 < 1$ a epidemia irá desaparecer, enquanto que, $R_0 > 1$ indica uma epidemia persistente.

3.2.1.1 Casos de Influenza no Internato

Em 1978 o jornal médico britânico, *The Lancet*, reportou uma epidemia em que 763 garotos eram residentes em um internato e 512 ficaram confinados na cama durante a epidemia que durou 15 dias. A quantidade de infectados para cada dia de observação é dada a seguir, pressupondo-se que um garoto infectado iniciou a epidemia (Tabela 3.6).

Tabela 3.6: Número de infectados (I) por influenza em um internato inglês em 1978

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$I(t)$	1	3	7	25	72	222	282	256	233	189	123	70	25	11	4

Na Figura 3.1 é possível observar que a infecção cresceu até o sétimo dia e quase 300 garotos ficaram doentes, depois disso, o número de casos decresceu.

Figura 3.1: Número de indivíduos infectados por influenza no internato durante 15 dias

Foi adotado o modelo SIR sem dinâmica vital (Equações 3.4) denotando o atual estado do indivíduo, que pode ser dividido em três compartimentos: Suscetível (S), Infectado (I) e Recuperado ou Removido da população (R) como pode ser observado na Figura 3.2.

Figura 3.2: Esquema para o modelo compartimental SIR sem dinâmica vital

$$\begin{cases} \frac{dS(t)}{dt} = -\beta(t)S(t)I(t), \\ \frac{dI(t)}{dt} = \beta(t)S(t)I(t) - \nu I(t), \\ \frac{dR(t)}{dt} = \nu I(t). \end{cases} \quad (3.4)$$

em que $\beta(t) = \beta$ é a taxa de transmissão e ν é a taxa de recuperação.

A inferência foi realizada por meio do método da máxima verossimilhança, MCMC e rejeição ABC. Para o estudo ABC foram consideradas as distâncias métricas euclidiana e de Mahalanobis a qual difere da euclidiana por considerar a correlação entre os conjuntos de dados. A expressão para essa distância entre dois vetores da mesma distribuição que tem matriz de covariância Σ é dada por

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})}.$$

Tanto para o ABC, as amostras simuladas, quanto para o MCMC, as iterações, foram 10.000. As distribuições a priori para o ABC e para o MCMC foram, para β , $10^{U(-4;-2)}$ conforme MURRAY (2002), para ν uma $U(0; 1)$ e para $I(0)$ (número inicial de infectados considerado desconhecido) foi $\text{Pois}(1)+1$ e a medida de tolerância tomada igual a 200. A distância para comparar a amostra verdadeira com as amostradas foi a euclidiana. Para análise bayesiana, foram utilizadas 3 cadeias com *thin*= 10 e *burn-in* = 1.000.

Além do cálculo da discrepância, para diagnóstico do modelo foram obtidos os resíduos ordinários e observadas as tendências por meio de gráficos.

3.2.1.2 Gripe Comum em Tristão da Cunha

Tristão da Cunha é uma ilha isolada no oceano Atlântico com aproximadamente 300 habitantes, e foi observado que uma doença viral, como gripe comum, apareceu na ilha após a chegada de navios vindos da Cidade do Cabo. Foram tomados registros diários do desenvolvimento da infecção das vias respiratórias superiores para os indivíduos envolvidos (HAMMOND e TYRRELL, 1971). O conjunto de dados possui 21 dias de observação como mostrado na Tabela 3.7 e foi obtido por SHIBLI *et al.* (1971) e HAMMOND e TYRRELL (1971) em outubro de 1967. Os dados fornecem o número de indivíduos infectados e recuperados, $I(t)$ e $R(t)$, enquanto que a quantidade inicial de suscetíveis, $S(0)$, não é conhecida. Dessa forma, $S(0)$ é um parâmetro desconhecido extra a ser estimado. É possível observar a evolução da doença na Figura 3.3.

Assim como para o caso de influenza, foi considerado o modelo SIR sem dinâmica vital devido ao fato de o tempo de observação ser de apenas 21 dias. Além disso a inferência também foi realizada por meio das três metodologias (máxima verossimilhança,

MCMC e rejeição ABC). O diagnóstico também foi realizado por meio da medida de discrepância e gráficos de tendência.

O estudo considerando os métodos MCMC e ABC foi realizado com 10.000 simulações e iterações respectivamente, e as mesmas distribuições a priori propostas por TONI *et al.* (2009) em que $S(0) \sim U(37; 100)$, $\beta \sim U(0; 1)$ e $\nu \sim U(0; 1)$. Na análise bayesiana, para 3 cadeias foram realizadas 10.000 interações, com $thin = 10$ e $burn-in = 1.000$. Assim como no estudo anterior as medidas de discrepância e tendências na dispersão dos resíduos foram observadas.

Tabela 3.7: Dados de gripe comum em Tristão da Cunha coletados em outubro de 1967

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$I(t)$	1	1	3	7	6	10	13	13	14	14	17	10	6	6	4	3	1	1	1	1	0
$R(t)$	0	0	0	0	5	7	8	13	13	16	16	24	30	31	33	34	36	36	36	36	37

Figura 3.3: Número de indivíduos infectados e recuperados por gripe comum na ilha de Tristão da Cunha durante 21 dias

3.2.2 Modelos de Interação Populacional

Para os modelos de interação populacional foram estudados três casos. O primeiro é um estudo de simulação para o modelo Lotka-Volterra (Equações 2.9) enquanto que o segundo trata da aplicação do mesmo modelo a um conjunto de dados reais. O último caso apresentado trata-se de um estudo de simulação para o modelo de competição (Equações 2.10). Para todos os casos estudados nessa seção foram realizadas 100.000 simulações para o ABC e mesmo número de iterações para o MCMC com $thin = 10$ e $burn-in = 10.000$ e utilizadas 3 cadeias.

O diagnóstico foi realizado por meio da avaliação da medida de discrepância e para os estudos de simulação também foi avaliado o gráfico bivariado para os resíduos conforme apresentado por MORAL e DEMÉTRIO (2016). A ideia geral desse gráfico consiste gerar valores aleatórios da distribuição empírica, obtendo assim, uma nuvem de pontos. Para esta nuvem de pontos é considerado um polígono convexo que contenha todos os

pontos gerados. Após isso a área desse polígono é reduzida a 95% da sua área original. Dessa forma, para cada tempo estudado é gerado um polígono convexo e observado se o ponto inferido ao tempo correspondente está no seu interior ou exterior. Quanto maior a quantidade de pontos dentro do polígono, maior o índice de um bom ajuste. Para o caso dos gráficos apresentados, pontos vermelhos indicam que os valores preditos estão fora do polígono e pontos pretos indicam que os valores preditos estão dentro.

3.2.2.1 Estudo de Simulação para o Modelo Lotka-Volterra

TONI *et al.* (2009) fizeram um estudo de simulação no qual, a partir do modelo dado pelas Equações 2.9 geraram um conjunto de dados (x, y) , com $(a, d) = (1, 1)$ e os parâmetros b e c fixados como $(1, 1)$. Um erro aleatório $N(0; 0, 5^2)$ foi computado para ser adicionado ao conjunto de dados simulados gerando um novo conjunto (xe, ye) . A fim de encontrar o valor de tolerância (δ) a ser utilizado no ABC, uma distância foi calculada da seguinte forma:

$$\delta = d((x, y), (xe, ye)) = \sum_i ((x_i - xe_i)^2 + (y_i - ye_i)^2) \quad (3.5)$$

Analogamente a TONI *et al.* (2009), para este trabalho, foram simulados os conjuntos de dados (x, y) e (xe, ye) e calculada a tolerância δ para os dois estudos de simulação dessa seção. O conjunto de dados (x, y) foi gerado a partir do modelo dado pelas Equações 3.6. O par ordenado (xe, ye) foi obtido adicionando-se um erro com distribuição $N(0; 0, 5^2)$ a (x, y) . O valor de tolerância foi calculado por meio da Equação 3.5 no qual $\delta = 4,3764$. Os valores (x, y) e (xe, ye) podem ser observados na Tabela 3.8.

$$\begin{cases} \frac{dx(t)}{dt} = ax - cxy, \\ \frac{dy(t)}{dt} = bxy - dy, \end{cases} \quad (3.6)$$

Tabela 3.8: Dados simulados para o modelo Lotka-Volterra e valores com erro aleatório $N(0; 0, 5^2)$

Tempo	1	2	3	4	5	6	7	8
Presas (x)	1,0000	1,5684	1,6455	0,8738	0,5240	0,5393	0,7875	1,2845
Predadores (y)	0,5000	0,6614	1,3335	1,7361	1,2296	0,7540	0,5279	0,5367
Presas (xe)	1,0900	1,7698	0,9497	1,3917	0,1422	0,7102	0,6589	1,5763
Predadores (ye)	0,3429	1,8631	0,9744	0,8558	0,6670	0,3942	1,1830	0,7626

Na Figura 3.4(a) é possível observar a quantidade de presas geradas (x) e a quantidade de presas com erro (xe) ao longo do tempo. Já na Figura 3.4(b) se nota a quantidade de predadores gerados (y) e a quantidade de predadores com erro (ye).

Figura 3.4: Número de indivíduos gerados com e sem erro: (a) População de presas; (b) População de predadores

Diferentemente de TONI *et al.* (2009), para este caso, a inferência foi realizada considerando todos os parâmetros desconhecidos $(a, b, c, d) = (0, 5; 0, 5; 0, 5; 0, 5)$. Foram utilizadas as abordagens da máxima verossimilhança, MCMC e rejeição ABC. As distribuições a priori utilizadas para a inferência tanto para o ABC quanto para o MCMC para os parâmetros a, b, c e d foram $U(0; 1)$. Neste caso gráfico bivariado para os resíduos foi obtido de acordo com MORAL e DEMÉTRIO (2016) e foram verificadas as porcentagens de pontos dentro dos polígonos de simulação.

3.2.2.2 Modelo Lotka-Volterra para Lebre (presa) e Lince (predador)

Desde os anos de 1920 que ecologistas discutem oscilações cíclicas na população de lebres. Um exemplo clássico de oscilação cíclica é entre a lebre (presa) e o lince (predador) (BEGON *et al.*, 1986). A empresa canadense Hudson Bay Company mantém registros das peles comercializadas de animais capturados desde o ano de 1800, aproximadamente.

Figura 3.5: População de peles de presas (lebres) e predadores (lince)

Embora esta seja uma medida indireta de predação, a suposição é que existe uma relação direta entre o número de peles coletadas e o número de lebre e lince no habitat natural. A quantidade peles, em milhares, de presas e predadores ao longo do tempo pode ser observada na Tabela 3.9 e na Figura 3.5.

Tabela 3.9: Quantidade, em milhares, de presas (lebre) e predadores (lince) ao longo do tempo

Tempo	Presa	Predador	Tempo	Presa	Predador	Tempo	Presa	Predador
1845	19,58	30,09	1876	97,12	46,34	1906	28,68	54,70
1846	19,60	45,15	1877	86,51	45,77	1907	23,37	6,30
1847	19,61	49,15	1878	72,17	44,15	1908	21,54	3,41
1848	11,99	39,52	1879	38,32	36,33	1909	26,34	5,44
1849	28,04	21,23	1880	10,11	12,03	1910	53,10	11,65
1850	58,00	8,42	1881	7,74	12,60	1911	68,48	20,35
1851	74,60	5,56	1882	9,67	18,34	1912	75,58	32,88
1852	75,09	5,08	1883	43,12	35,14	1913	57,92	39,55
1853	88,48	10,17	1884	52,21	43,77	1914	40,97	43,36
1854	61,28	19,60	1885	134,85	65,69	1915	24,95	40,83
1855	74,67	32,91	1886	134,86	79,35	1916	12,59	30,36
1856	88,06	34,38	1887	103,79	51,65	1917	4,97	17,18
1857	68,51	29,59	1888	46,10	32,59	1918	4,50	6,82
1858	32,19	21,30	1889	15,03	22,45	1919	11,21	3,19
1859	12,64	13,69	1890	24,20	16,16	1920	56,60	3,52
1860	21,49	7,65	1891	41,65	14,12	1921	69,63	9,94
1861	30,35	4,08	1892	52,34	20,38	1922	77,74	20,30
1862	2,18	4,09	1893	53,78	33,33	1923	80,53	31,99
1863	152,65	14,33	1894	70,40	46,00	1924	73,38	42,36
1864	148,36	38,22	1895	85,81	51,41	1925	36,93	49,08
1865	85,81	60,78	1896	56,69	46,43	1926	4,64	53,99
1866	41,41	70,77	1897	16,59	33,68	1927	2,54	52,25
1867	14,75	72,77	1898	6,16	18,01	1928	1,80	37,70
1868	2,28	42,68	1899	2,30	8,86	1929	2,39	19,14
1869	5,91	16,39	1900	12,82	7,13	1930	4,23	6,98
1870	9,95	9,83	1901	4,72	9,47	1931	19,52	8,31
1871	10,44	5,80	1902	4,73	14,86	1932	82,11	16,01
1872	70,64	5,26	1903	37,22	31,47	1933	89,76	24,82
1873	50,12	18,91	1904	69,72	60,57	1934	81,66	29,70
1874	50,13	30,95	1905	57,78	63,51	1935	15,76	35,40
1875	101,25	31,18						

Para este estudo, assim como para o estudo de simulação, foi considerado o modelo 3.6 e comparadas as mesmas três formas de inferência descritas no estudo de simulação. Para a realização da inferência por meio do ABC as distribuições a priori utilizadas para os parâmetros a e d foram vagas $U(0; 1)$ enquanto que para os parâmetros b e c foram informativas $U(0, 01; 0, 02)$ e $U(0, 02; 0, 03)$ respectivamente. Para o método MCMC as distribuições a priori para a e d foram informativas $U(0, 6; 0, 7)$ e $U(0, 5; 0, 65)$ devido ao fato de não ter havido convergência quando vagas. O diagnóstico foi realizado por meio da medida de discrepância e gráficos de tendência como o gráfico meio normal.

3.2.2.3 Estudo de Simulação para o Modelo de Competição entre Espécies

A seguir um exemplo que modela a população de duas espécies que competem por mesmos recursos é apresentado. Da mesma forma que TONI *et al.* (2009), foram gerados os conjuntos de dados (x, y) e (xe, ye) a partir do modelo 3.7 com parâmetros $\theta = c(a_1, b_1, c_1, a_2, b_2, c_2) = (1; 1; 1; 0, 75; 1; 0, 5)$, e $\delta = 0, 1319$ foi obtido. A quantidade de indivíduos da espécie 1 e espécie 2 pode ser observada na Tabela 3.10 e na Figura 3.6.

$$\begin{cases} \frac{dx(t)}{dt} = x(a_1 - b_1x - c_1y), \\ \frac{dy(t)}{dt} = y(a_2 - b_2y - c_2x). \end{cases} \quad (3.7)$$

Figura 3.6: Número de indivíduos gerados com e sem erro: (a) População da espécie 1; (b) População da espécie 2

Para este estudo, em que há seis parâmetros para serem estimados, a inferência foi realizada somente por meio das abordagens MCMC e rejeição ABC, devido a dificuldade de obtenção de valores iniciais e convergência do modelo por meio do método da máxima verossimilhança. As distribuições a priori utilizadas para a inferência tanto para o ABC quanto para o MCMC para os parâmetros $a_1, b_1, c_1, a_2, b_2, c_2$ e d foram $U(0, 2)$. Também

foram verificadas a medida de discrepância e o gráfico bivariado para os resíduos bem como suas porcentagens fora do polígono de simulação.

Tabela 3.10: Dados simulados para o modelo de competição e valores com erro aleatório $N(0; 0, 5^2)$

Tempo	1	2	3	4	5	6	7	8
Espécie 1 (x)	2,0000	0,7847	0,6716	0,6254	0,5945	0,5712	0,5536,	0,5402
Espécie 2 (y)	0,5000	0,3523	0,3744	0,4036	0,4279	0,4464	0,4603	0,4706
Espécie 1 (xe)	2,0180	0,8250	0,5325	0,7290	0,5181	0,6054	0,5279	0,5985
Espécie 2 (ye)	0,5235	0,3192	0,3432	0,1734	0,4108	0,4605	0,3106	0,3696

4 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados obtidos neste trabalho. Na subseção 4.1, foram realizados estudos de simulação usando o método de regressão ABC para os modelos de regressão linear e Poisson. Na seção 4.2 são trabalhados, cinco estudos de caso envolvendo modelos epidemiológicos e de interação populacional.

4.1 Estudos de Simulação

Para as 50 amostras simuladas em cada tamanho amostral e cenário, os parâmetros dos modelos de regressão linear e Poisson foram estimados. Os valores das medidas de discrepância para os parâmetros β_0, β_1 e σ^2 do modelo de regressão linear podem ser observados no Apêndice A.2 em que, estão apresentadas as Tabelas A.2 e A.3. Já para o modelo de regressão Poisson, os valores de discrepância para β_0 e β_1 foram apresentados nas Tabelas A.4 e A.5.

4.1.1 Regressão Linear

Nas Figuras 4.1 a 4.16 foram apresentados os níveis de tolerância (0,01; 0,1 e 0,5) e distâncias métricas (euclidiana, modular e do máximo) para alguns dos cenários e parâmetros estudados. Além disso, a linha tracejada preta representa o verdadeiro valor do parâmetro, a área cinza forma o intervalo de credibilidade de cada uma das 50 amostras a posteriori, os pontos pretos são as estimativas por máxima verossimilhança e os pontos brancos as estimativas das médias a posteriori para cada amostra por meio do método de regressão ABC.

Todos os cenários estudados nesta subseção podem ser observados na Tabela 3.3. Para os cenários 1 e 2, as estimativas dos parâmetros β_0 e β_1 foram similares para os métodos de estimação por meio da máxima verossimilhança e regressão ABC. No entanto, maiores valores de discrepância foram observados para as estimativas da variância por meio do método ABC quando distribuições a priori não informativas foram utilizadas. Esses resultados corroboram com o trabalho de MCKINLEY *et al.* (2009), no qual se tem uma influência negativa de distribuições a priori vagas nas estimativas dos parâmetros, causando posterioris inconsistentes com os dados. Na Figura 4.1 observa-se que a distância modular forneceu maiores intervalos de credibilidade, sendo que a tolerância não influenciou sua amplitude. Neste sentido, visando menores intervalos, recomenda-se para este caso a utilização de tolerâncias menores e as distâncias euclidiana e do máximo.

Figura 4.1: Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0;10^2)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 2)

Nos cenários 3 e 4 as distribuições a priori dos parâmetros não foram sensíveis a variações de tolerâncias e distâncias, apresentando intervalos de credibilidade semelhantes para todas as situações estudadas. As distribuições a priori informativas de β_0 e β_1 utilizadas nesse contexto possuem informação equivocada, dessa forma, o ABC não foi capaz de obter boas estimativas como pode ser observado nas Figuras 4.2 e 4.4. Observa-se também que o método superestima e subestima os valores dos parâmetros β_0 e β_1 respectivamente. Isso ocorre devido ao fato de o algoritmo aceitar valores vindos da

distribuição a priori, que neste caso, é informativa $N(0; 1)$, mas está distante do verdadeiro valor do parâmetro.

Figura 4.2: Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0;1)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 3)

No cenário 3, em que a distribuição a priori para σ^2 era informativa e baseada nos dados, d_{σ^2} para todos os casos estudados no ABC foi menor que para a máxima verossimilhança. Já no cenário 4, em que a distribuição a priori para σ^2 foi vaga, observou-se os piores desempenhos do método ABC.

Figura 4.3: Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_1 = 50$ e distribuição a priori $N(0;1)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 3 com variável centrada)

A Figura 4.3 representa o estudo para o parâmetro β_0 no cenário 3 considerando a variável tempo centrada. Ao comparar as Figuras 4.2 e 4.3 é possível observar que apesar de as estimativas para β_0 estarem mais próximas do verdadeiro valor, o método ABC não eliminou totalmente o viés para este parâmetro, evidenciando a importância da escolha de uma distribuição a priori adequada.

Figura 4.4: Estudo do parâmetro β_1 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0;1)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 3)

Os fatos apresentados nos cenários 3 e 4 mostram que para o ABC, assim como no MCMC, distribuições a priori equivocadas podem conduzir a resultados enganosos já que estes dependem da escolha da distribuição a priori (BEAUMONT *et al.*, 2002; BEAUMONT, 2003).

Para os cenários 5 e 6 os resultados foram similares entre si para os parâmetros β_0 e β_1 , gerando estimativas próximas dos valores simulados ao considerar todas as combinações de distância e tolerância como pode ser observado, veja a Figura 4.5. Nestes casos,

as distribuições a priori fornecidas corresponderam a distribuições normais com valores de média e variância provenientes das 50 amostras simuladas de tamanho N_1 . Para σ^2 , melhores desempenhos foram observados para o cenário 5, em que uma distribuição a priori informativa foi utilizada. Logo, é possível notar que o método ABC é sensível a escolha da distribuição a priori.

Figura 4.5: Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_1 = 50$ e distribuição a priori $N(7, 8; 6, 9^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 5)

Figura 4.6: Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_1 = 50$ e distribuição a priori $N(7, 8; 6, 9^2)$ para β_0 e β_1 , $\text{IGamma}(10, 10)$ para σ^2 (cenário 5 com variável centrada)

Além das distribuições a priori informativas baseadas nos dados consideradas para os cenários 5 e 6, considerou-se distribuições a priori informativas baseadas nas estimativas por máxima verossimilhança dos parâmetros com a variável tempo centrada. Na Figura 4.6 é possível observar o estudo de simulação para o parâmetro β_0 , em que, as estimativas tanto por máxima verossimilhança quanto por meio do algoritmo ABC regressão são semelhantes quando comparadas com o cenário 5 (Figura 4.5).

Figura 4.7: Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 10^2)$ para β_0 e β_1 , $\text{IGamma}(10,10)$ para σ^2 (cenário 7)

Em geral para os cenários estudados com tamanho amostral $N_1 = 50$, as estimativas de β_0 e β_1 para o ABC independem da priori, entretanto para σ^2 , existem diferenças pois as maiores discrepâncias ocorreram nos casos em que foram fornecidas informações vagas, gerando assim, distribuições a posteriori inconsistentes com os dados. Desse forma, é necessário uma maior atenção para o parâmetro de variância. Era esperado também que as estimativas vindas de distribuições a priori vagas para β_0 e β_1 não fossem satisfatórias,

uma vez que o ABC é baseado no algoritmo de rejeição, isto é, aceitará ou rejeitará valores da respectiva priori.

Figura 4.8: Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 10^2)$ para β_0 e β_1 , $\text{IGamma}(0.5, 0.5)$ para σ^2 (cenário 8)

Uma possível explicação para este resultado é o fato de que uma priori adequada não deve afetar os resultados do estudo (GELMAN *et al.*, 2004), entretanto, quando a priori é vaga, melhores resultados são dependentes de menores tolerâncias, como observado no

cenário 2, enquanto que para distribuições a priori informativas baseadas nos dados tanto tolerância quanto distância não interferem no resultado.

Figura 4.9: Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 9)

No cenário 7, o método da máxima verossimilhança superestimou o parâmetro β_0 e subestimou β_1 , isso acontece devido ao fato de os parâmetros serem correlacionados negativamente. Entretanto, apesar de utilizar as estimativas de máxima verossimilhança como resumo estatístico, o ABC apresentou bons resultados, como pode ser verificado

na Figura 4.7, e medidas de discrepância expressivamente baixas. Para esse caso, o comportamento para diferentes tolerâncias é parecido e o parâmetro σ^2 é similar para todos os casos apesar de tolerâncias intermediárias (0, 1) fornecerem menores discrepâncias para todas as distâncias.

Figura 4.10: Estudo do parâmetro β_0 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $\text{IGamma}(10,10)$ para σ^2 (cenário 9 com variável centrada)

Figura 4.11: Estudo do parâmetro σ^2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $\text{IGamma}(10,10)$ para σ^2 (cenário 9)

Para o cenário 8, o comportamento de β_0 e β_1 foi similar com o cenário 7, mas diferente para σ^2 (Figura 4.8). A medida de discrepância para este parâmetro foi maior para o ABC na maioria dos casos devido a priori vaga utilizada para este parâmetro. A distância modular produziu os maiores intervalos de credibilidade, enquanto que para tolerâncias pequenas as distâncias euclidiana e do máximo forneceram intervalos de credibilidade com menores amplitudes. Este fato sugere o uso de níveis de tolerância menores

quando distribuições a priori vagas são utilizadas. Para esse cenário as distribuições a priori estudadas responderam de forma satisfatória e forneceram boas estimativas.

Figura 4.12: Estudo do parâmetro σ^2 para o modelo de regressão linear com a variável tempo centrada, tamanho amostral $N_2 = 10$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 , $IGamma(10,10)$ para σ^2 (cenário 9 com variável centrada)

No cenário 9, para todos os parâmetros, tanto as estimativas por máxima verossimilhança quanto para o ABC foram viesadas (Figura 4.9). No caso da máxima verossimilhança o estimador é viesado para pequenas amostras (MOOD *et al.*, 1974), já

no caso do ABC, a distribuição a priori informativa estava equivocada, de forma semelhante aos cenários 3 e 4 para amostras de tamanho 50.

Figura 4.13: Estudo do parâmetro β_0 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(7, 8; 7, 5^2)$ para β_0 e β_1 , $\text{IGamma}(10, 10)$ para σ^2 (cenário 11)

Mesmo com distribuição a priori informativa equivocada, o ABC apresentou menores discrepâncias e consequentemente estimativas mais próximas das reais de β_0 e β_1

para todos os casos estudados (Figura 4.9). Já para σ^2 , o método da máxima verossimilhança apresentou melhores resultados como observado na Figura 4.11.

Figura 4.14: Estudo do parâmetro σ_2 para o modelo de regressão linear com tamanho amostral $N_2 = 10$ e distribuição a priori $N(7, 8; 7, 5^2)$ para β_0 e β_1 , $IGamma(10, 10)$ para σ^2 (cenário 12)

Ao considerar as variáveis regressoras centradas observa-se melhora nas estimativas tanto por meio do método da máxima verossimilhança quanto por meio do método ABC (Figuras 4.10 e 4.12), entretanto para este último ainda continua viesado devido a

distribuição a priori equivocada. Tal fato sugere que, para modelos de regressão linear, mesmo centrando a variável regressora para eliminar a correlação entre os parâmetros, existe a necessidade de escolher a distribuição a priori adequada.

De forma análoga ao cenário 8, menores discrepâncias para o método ABC foram obtidas quando as distâncias euclidiana e do máximo com tolerância de 0,01 foram empregadas. Os resultados foram semelhantes para o cenário 10, entretanto d_{σ^2} foram bem maiores. Esses fatos evidenciam que distribuições a priori vagas ou equivocadas influenciam negativamente o valor da variância que precisará ser melhor calibrada pelas tolerâncias (WILKINSON, 2013).

Assim como para grandes amostras, o método ABC para pequenas amostras forneceu boas estimativas tanto centrando quanto não centrando a variável regressora. Para os cenários 11 e 12 os resultados foram similares. As Figuras 4.13 e 4.14 mostram o estudo de simulação para os parâmetros β_0 e σ^2 respectivamente. É possível verificar que mesmo para pequenas amostras, as estimativas foram próximas do verdadeiro valor para todos os parâmetros. No cenário 11 em que todas as distribuições a priori foram informativas, os valores de discrepância para todos os parâmetros foram menores que para a função verossimilhança. Para o cenário 12, que apresenta distribuição a priori não informativa para σ^2 , a discrepância para este parâmetro foi maior.

4.1.2 Regressão Poisson

Todos os cenários estudados para o modelo de regressão Poisson foram apresentados na Tabela 3.5. Na Figura 4.15 pode ser observado o estudo para β_1 no cenário 1 da regressão Poisson.

Nota-se que níveis de tolerância pequenos fornecem menores intervalos e que para este cenário a distância modular é melhor para tolerâncias maiores. Para este caso a distância modular foi melhor devido ao fato de os parâmetros utilizados serem positivos. Para o cenário 1, o comportamento do parâmetro β_0 é o mesmo que β_1 . Dessa forma, os modelos ajustados pelo ABC para cada conjunto de dados são muito próximos do verdadeiro para todos os casos deste cenário.

As Figuras referentes aos cenários 3 e 4 foram omitidas devido sua similaridade com os cenários 1 e 2 respectivamente. Isso implica que tanto para pequenas quanto para grandes amostras, o ABC fornece boas estimativas, ou seja, a eficiência do ABC independe do tamanho amostral. Distribuições a priori vagas também foram utilizadas nos cenários 2 e 4 sendo recuperados os verdadeiros valores dos parâmetros pelo método ABC.

Finalmente é apresentado o estudo para β_0 no cenário 2 como apresentado na Figura 4.16. Para todas as combinações de distância e tolerância, os valores estimados são próximos do verdadeiro valor tanto para β_0 quanto para β_1 . Isso implica que o ABC

além de ser eficiente para pequenas amostras também o é para distribuições a priori vagas no caso da regressão Poisson. As medidas de discrepância para a grande maioria dos cenários estudados neste modelo foram menores para o ABC.

Figura 4.15: Estudo do parâmetro β_1 para o modelo de regressão Poisson com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 (cenário 1)

Figura 4.16: Estudo do parâmetro β_0 para o modelo de regressão Poisson com tamanho amostral $N_1 = 50$ e distribuição a priori $N(0; 1)$ para β_0 e β_1 (cenário 2)

4.2 Modelos de Dinâmica Populacional

São apresentados 5 estudos de caso para dinâmica populacional, dentre eles dois exemplos de modelos epidemiológicos para dados reais, um estudo de simulação e uma análise para dados reais considerando o modelo Lotka-Volterra, e um estudo de simulação para o modelo de competição entre espécies. Suas análises foram realizadas por meio dos métodos da máxima verossimilhança, MCMC e rejeição ABC (algoritmo 2.3.1.1.2).

4.2.1 Modelos Epidemiológicos

Os itens 4.2.1.1 e 4.2.1.2 dessa seção exploram o modelo SIR (equações 2.8) para casos de influenza em um internato na Inglaterra e gripe comum em Tristão da Cunha. Para o primeiro caso são conhecidos apenas o número de infectados ao longo do tempo, sendo necessária a estimação de dois parâmetros $\theta = (\beta, \nu)$. Para o segundo caso são conhecidos o número de infectados e recuperados ao longo do tempo, entretanto o número inicial de suscetíveis $S(0)$ é desconhecido, sendo este, um parâmetro extra a ser estimado.

4.2.1.1 Casos de Influenza no Internato

Ao registrar os casos de influenza no internato supõe-se que os erros possuem distribuição normal, e dessa forma a função de verossimilhança para o modelo e seus parâmetros pode ser calculada. Na Figura 4.17(a) é apresentado o gráfico de contorno para o logaritmo da função verossimilhança que mostra a região que minimiza o erro do modelo. É possível observar que a função de verossimilhança não é simétrica, o que torna mais difícil encontrar um bom valor inicial para o método iterativo, uma vez que o algoritmo pode se deparar com um máximo local e não global ou mesmo pode não convergir. Os parâmetros inferidos $\beta = 0,0022$ e $\nu = 0,5000$ estão cobertos por esta área. Eles apresentaram boas estimativas quando comparados com os previamente obtidos na literatura, $\beta = 0,0022$ e $\nu = 0,4410$.

Figura 4.17: Modelo SIR para dados de influenza: (a) Gráfico de contorno para a densidade dos parâmetros β e ν e o ponto estimado; (b) Valores preditos e observados para o número de garotos infectados durante o período de 15 dias

O modelo faz um bom trabalho para capturar o padrão dos dados (Figura 4.17(b)). Os valores preditos e observados para os indivíduos infectados são próximos entre si. O período de infecção foi de $1/\nu = 1/0,5 = 2$ dias e a taxa de reprodução básica $\mathcal{R}_0 = \beta/\nu = 0,5/0,0022 = 227,2727 > 1$ que indica uma epidemia persistente.

A análise desse conjunto de dados foi também realizada por meio do MCMC e do ABC que não apresentaram muitas diferenças de estimativa entre si. Para ambos os métodos, tanto quantidades de valores simulados quanto distribuições a priori utilizadas foram as mesmas. Dessa forma, é apresentado a seguir os resultados pelo método ABC considerando-se as distâncias euclidiana e de Mahalanobis.

Figura 4.18: Modelo SIR pra dados de influenza estimado por meio do ABC com distância euclidiana: (a) Histograma das médias a posteriori para cada parâmetro estimado ($\beta, \nu, I(0)$); (b) Valores preditos pelo modelo para dados de influenza no internato

Para o ABC foram aceitas 70 amostras por meio da utilização da distância euclidiana. A média a posteriori foi $\beta = 0,0024, \nu = 0,5544, I(0) = 2,7714$. Na Figura 4.18(a) é possível observar que a média a posteriori foi próxima da previamente obtida na literatura ($\beta = 0,0022, \nu = 0,4410$ e $I(0) = 1$). O tempo infeccioso inferido foi de $1/\nu = 1/0,55 \approx 1,8$ dias. As previsões para o modelo podem ser observadas na Figura 4.18(b).

Para que as covariâncias entre os vetores comparados sejam consideradas, outro cenário utilizando a distância de Mahalanobis foi estudado, no qual 32 amostras foram aceitas. As médias a posteriori foram $\beta = 0,0032, \nu = 0,4972, I(0) = 1$, novamente muito próximas das previamente obtidas na literatura. O número inicial de indivíduos infectados foi melhor para este caso. O tempo inferido de infecção foi de $1/\nu = 1/0,4972 \approx 2$ dias.

A Figura 4.19(a) mostra que para o parâmetro ν e o valor inicial de infectados $I(0)$ as estimativas foram bem próximas das verdadeiras, enquanto que para o parâmetro β_0 a melhor estimativa foi utilizando a distância euclidiana. As previsões para o número de infectados utilizando a distância de Mahalanobis é mostrada na Figura 4.19(b) a qual apresentou piores resultados quando comparada com as outras.

Nenhuma das figuras apresenta comportamento incomum, indicando adequação dos modelos. Para ambas as distâncias estudadas os valores de parâmetro estimados foram próximos do verdadeiro, entretanto, para este caso de estudo que é mais simples,

o método da máxima verossimilhança apresentou estimativas mais próximas da real e menores medidas de discrepância.

Figura 4.19: Modelo SIR pra dados de influenza estimado por meio do ABC com distância de Mahalanobis: (a) Histograma das médias a posteriori para cada parâmetro estimado ($\beta, \nu, I(0)$); (b) Modelo ajustado para dados de influenza no internato

4.2.1.2 Casos de Gripe Comum em Tristão da Cunha

Para os estudo da ocorrência de gripe comum em Tristão da Cunha, observa-se o gráfico de contorno para a função logaritmo da função de verossimilhança apresentado na Figura 4.20(a). Assim como para o caso de influenza, é possível observar que a função de verossimilhança não é simétrica, o que também torna mais difícil encontrar bons valores iniciais para o método iterativo. Os parâmetros inferidos $\beta = 0,0200$ e $\nu = 0,2527$ estão cobertos por esta área. Entretanto, ao observar a Figura 4.20(b) é possível verificar que os valores ajustados se afastam um pouco dos observados, além de as medidas de discrepância para os parâmetros serem maiores como apresentado na Tabela 4.1. Isso ocorre devido ao fato de a função verossimilhança ser irregular e a quantidade de parâmetros a ser estimado ser maior. Outro fato a ser observado é o gráfico de resíduos padronizados que possuem tendência tanto para o número de infectados (Figura 4.21(a)) quanto para o número de recuperados (Figura 4.21(b)). Dessa forma, a medida que a complexidade da função de verossimilhança aumenta, são necessários métodos alternativos de estimação.

É possível observar na Figura 4.22(a) o histograma contendo os valores a posteriori para β , ν e $S(0)$ quando estimados pelo método da rejeição ABC. As médias a posteriori foram $\beta = 0,0231$, $\nu = 0,2889$ e $S(0) = 39,6810$ enquanto os valores inferidos considerando o método MCMC foram $\beta = 0,0201$, $\nu = 0,2717$, $S(0) = 41,7600$.

Figura 4.20: Modelo SIR para os dados de Tristão da Cunha: (a) Gráfico de contorno para a densidade dos parâmetros β e ν e o ponto estimado; (b) Valores preditos e observados para o número de infectados e recuperados durante o período de 21 dias

Figura 4.21: Resíduos padronizados para o modelo SIR considerando a máxima verossimilhança para os dados de Tristão da Cunha: (a) Número de infectados; (b) Número de recuperados

A Figura 4.22(b) apresenta a curva predita para os compartimentos infectado e recuperado considerando o ABC. É possível observar um melhor ajuste quando comparado com o método da máxima verossimilhança (Figura 4.20(b)). Neste exemplo, o período de infecção foi de $1/\nu = 1/0,2889 = 3,46$ dias, e a taxa básica de reprodução da doença foi de $\mathcal{R}_0 = \beta/\nu = 0,02306/0,2889 = 0,79 < 1$ o que indica que a doença irá desaparecer.

Assim como o método da máxima verossimilhança, o método MCMC apresentou resíduos com tendência similar e valores de discrepância maiores que o ABC como apresentado na Tabela 4.1. Além disso, os resíduos padronizados do método ABC apresentaram menor tendência que os demais tanto para o número de infectados (Figura 4.23(a)) quanto para o número de recuperados (Figura 4.23(b)). Dessa forma, para este caso de estudo o ABC apresentou as estimativas com menores discrepâncias e melhor padrão para os

resíduos. Esses fatos mostram que a medida que a dimensão dos dados e parâmetros aumentam, a estimação por máxima verossimilhança deixa de ser interessante (CAMPBELL *et al.*, 2012).

Figura 4.22: Modelo SIR pra dados de Tristão da Cunha estimado por meio do ABC com distância de Mahalanobis: (a) Histograma das médias a posteriori para cada parâmetro estimado ($\beta, \nu, I(0)$); (b) Modelo ajustado para dados de gripe comum

Figura 4.23: Resíduos padronizados para o modelo SIR considerando o ABC para os dados de Tristão da Cunha: (a) Número de infectados; (b) Número de recuperados

Tabela 4.1: Medidas de discrepância e tempo computacional do modelo SIR para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de gripe comum em Tristão da Cunha

Método	Discrepância		Tempo Computacional (min)
	Número de Infectados	Número de Recuperados	
Verossimilhança	158,05	236,50	0,05
MCMC	111,81	92,39	39,73
ABC	47,41	79,73	24,71

4.2.2 Modelos de Interação Populacional

Para os itens 4.2.2.1 e 4.2.2.2 dessa seção, foi estudado o modelo Lotka-Volterra (equações 2.9), que descreve a interação entre espécies de presa, x , e predador, y , com vetor de parâmetros $\theta = (a, b, c, d)$. A primeira aplicação trata de um estudo de simulação enquanto que a segunda investiga dados reais de populações de lince e lebre. O item 4.2.2.3 examina um conjunto de dados simulados para o modelo de competição entre duas espécies (equações 2.10) x e y com parâmetros $\theta = (a_1, b_1, c_1, a_2, b_2, c_2)$.

4.2.2.1 Estudo de Simulação para o Modelo Lotka-Volterra

A inferência foi realizada considerando-se os 3 métodos de estimação descritos anteriormente. Por meio da máxima verossimilhança, apesar de o tempo computacional ter sido o menor (vide Tabela 4.2), as medidas de discrepância foram maiores. Além disso, o ajuste é o mais distante dos dados (Figura 4.24(a)) e o gráfico bivariado dos resíduos mostra que a maioria dos pontos está fora dos polígonos de confiança simulados já que estão em vermelho (Figura 4.24(b)). Devido ao fato da função de verossimilhança ser irregular e, a medida que a quantidade de parâmetros e consequentemente a complexidade do modelo aumentam devido a interação da população, isso introduz instabilidade ao sistema (MURRAY, 2002).

Ao considerar o MCMC os valores da média a posteriori para os parâmetros foram $(a, b, c, d) = (0,9958; 0,9956; 0,9946; 0,9941)$, ou seja, distantes do verdadeiro valor como pode ser observado na Figura 4.25. Esse fato provavelmente se dá devido a instabilidade na função verossimilhança que forneceu como estimativa máximos locais (CAMPBELL *et al.*, 2012).

Os valores preditos e observados tanto para presa quanto para predador estão distantes entre si (Figura 4.24(c)), apesar das medidas de discrepância serem um pouco melhores quando comparadas com as da máxima verossimilhança como apresentado na Tabela 4.2. O gráfico bivariado para os resíduos apresenta 5 de 8 pontos fora dos polígonos de simulação, indicando que o ajuste não foi satisfatório (4.24(d)).

A abordagem ABC foi realizada com $\delta = 4,38$ e é possível observar que os valores de média a posteriori estimados foram próximos dos verdadeiros além de os valores a posteriori estarem distribuídos de forma mais simétrica (Figura 4.26). Apesar do maior tempo computacional, a medida de discrepância calculada foi a menor entre os métodos apresentados (Tabela 4.2) e o gráfico bivariado para os resíduos apresenta apenas um ponto fora do polígono de simulação (Figura 4.24(f)) indicando um bom ajuste tanto para o número de presas quanto para o número de predadores como pode ser observado na Figura 4.24(e). Dessa forma, mesmo com a utilização de distribuições a priori pouco informativas para os parâmetros, o ABC apresentou bons resultados e superou os demais.

(a)

(b)

(c)

(d)

(e)

(f)

Figura 4.24: Modelo Lotka-Volterra para dados simulados considerando: estimação por máxima verossimilhança - (a) Valores observados e preditos para presa e predador; (b) Gráfico de diagnóstico bivariado; estimação via MCMC - (c) Valores observados e preditos para presa e predador; (d) Gráfico de diagnóstico bivariado; estimação via ABC - (e) Valores observados e preditos para presa e predador; (f) Gráfico de diagnóstico bivariado

Figura 4.25: Histograma dos valores a posteriori estimados pelo MCMC para dados simulados, com sua média a posteriori e o verdadeiro valor do parâmetro

Figura 4.26: Histograma dos valores a posteriori estimados pelo ABC para dados simulados, com sua média a posteriori e o verdadeiro valor do parâmetro

Tabela 4.2: Medidas de discrepância e tempo computacional do modelo Lotka-Volterra para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de predação simulados

Método	Discrepância		Tempo Computacional (min)
	Número de Presas	Número de Predadores	
Verossimilhança	4,22	3,85	0,07
MCMC	2,87	2,66	1,03
ABC	0,36	0,63	6,77

4.2.2.2 Modelo Lotka-Volterra para Lebre (presa) e Lince (predador)

Outro estudo para o caso Lotka-Volterra é o exemplo apresentado em MURRAY (2002) sobre variações populacionais entre lebres (presa) e linces (predador) durante 91 anos no Canadá.

A estimação por máxima verossimilhança demonstrou baixo desempenho para o conjunto de dados, uma vez que, apesar do reduzido tempo computacional, apresentou elevadas medidas de discrepância como pode ser observado na Tabela 4.3. Os resíduos do ajuste para o número de presas (Figura 4.27(a)) assim como para o número de predadores (Figura 4.28(a)) não apresentam padrão aleatório além de haver evidências de não normalidade devido ao fato de mais de 30% dos resíduos apresentarem-se fora do envelope de simulação para a distribuição normal (Figuras 4.27(b) e 4.28(b)).

Figura 4.27: Resíduos para o modelo Lotka-Volterra considerando a máxima verossimilhança para os dados reais de lebre (presa): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade

Figura 4.28: Resíduos para o modelo Lotka-Volterra considerando a máxima verossimilhança para os dados reais de linçe (predador): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade

Figura 4.29: Resíduos para o modelo Lotka-Volterra considerando o ABC para lebre (presa): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade

Figura 4.30: Resíduos para o modelo Lotka-Volterra considerando o ABC para linçe (predador): (a) Resíduos padronizados; (b) Envelope simulado e gráfico de normal probabilidade

Os valores da média a posteriori para os parâmetros foram $(a, b, c, d) = (0,6400; 0,0181; 0,0232; 0,5944)$, proporcionando resíduos mais próximos da distribuição normal assim como os obtidos pelo ABC para as quantidades de presas e predadores (Figuras 4.29(a) e 4.29(b)). Apesar de o ABC utilizar distribuições a priori vagas, os valores de discrepância para presa e tempo computacional foram menores que o MCMC (vide Tabela 4.3).

Tabela 4.3: Medidas de discrepância e tempo computacional do modelo Lotka-Volterra para os métodos de máxima verossimilhança, MCMC e ABC considerando os dados de predação de lebre e lince

Método	Discrepância		Tempo Computacional (min)
	Número de Presas	Número de Predadores	
Verossimilhança	220930,60	56323,40	1,14
MCMC	139601,10	25658,40	62,45
ABC	87853,02	34920,38	31,95

Figura 4.31: Ajuste do modelo Lotka-Volterra para os dados de lebre e lince por meio da metodologia ABC

Os valores preditos pelo modelo podem ser observados na Figura 4.31 que descreve razoavelmente o comportamento dos animais, entretanto os elevados índices de discrepância para esse estudo (Tabela 4.3) devem-se ao fato da grande variabilidade observada nos dados ($\sigma_{presa} = 36, \sigma_{predador} = 28$). O modelo Lotka-Volterra, apesar de recomendado para o estudo em questão, é pouco flexível quando ajustado para dados reais o que sugere a utilização de modelos com mais parâmetros que captem de forma mais eficiente a variabilidade.

4.2.2.3 Estudo de Simulação para o Modelo de Competição entre Espécies

Para este caso serão apresentados somente os resultados obtidos por meio do MCMC e do ABC devido ao fato da não convergência do modelo para o método da máxima verossimilhança, uma vez que se trata do ajuste de um modelo de equação diferencial estocástica com 6 parâmetros.

Figura 4.32: Estimação para dados simulados do modelo de competição utilizando o método MCMC: (a) Valores observados e preditos para espécie 1 e espécie 2; (b) Gráfico de diagnóstico bivariado

Figura 4.33: Estimação para dados simulados do modelo de competição utilizando o método ABC: (a) Valores observados e preditos para espécie 1 e espécie 2; (b) Gráfico de diagnóstico bivariado

Observa-se na Tabela 4.4 que os valores de discrepância para os dois métodos de estimação foram muito próximos, com uma ligeira vantagem para o método ABC. O mesmo pode ser observado em relação as Figuras 4.32(a) e 4.33(a) do modelo ajustado

bem como nas Figuras 4.32(b) e 4.33(b) dos resíduos bivariados no qual o método MCMC apresentou mais pontos fora do polígono de simulação.

Tabela 4.4: Medidas de discrepância e tempo computacional do modelo de competição para os métodos MCMC e ABC considerando dados simulados

Método	Discrepância		Tempo Computacional (min)
	Quantidade da Espécie 1	Quantidade da Espécie 2	
MCMC	0,0085	0,0961	1,03
ABC	0,0095	0,0008	8,48

Dessa forma, o ABC pode ser utilizado como método alternativo de estimação devido ao seu bom desempenho e ligeira vantagem em relação ao MCMC.

5 CONCLUSÕES

Os algoritmos ABC fornecem amostras independentes da posteriori, ao invés das correlacionadas obtidas por meio do MCMC, é fácil simular conjuntos de dados a partir do modelo dados os parâmetros, e estatísticas encontradas na literatura podem ser utilizadas sem a necessidade de ter o conjunto de dados original.

Os estudos de simulação sugerem que, em geral, o ABC fornece boas estimativas para os parâmetros tanto para pequenas quanto para grandes amostras e quando a priori é vaga é aconselhável a utilização de menores tolerâncias, o que implica em uma menor taxa de aceitação e, conseqüentemente, mais simulações são necessárias.

A distância modular gerou os maiores intervalos de credibilidade, e conseqüentemente outras alternativas como as distâncias euclidiana e do máximo podem ser empregadas. Mesmo utilizando um resumo estatístico viesado para pequenas amostras, com distribuições a priori apropriadas, o ABC foi capaz de obter boas estimativas para os parâmetros. Os cenários estudados permitem concluir que é preferível fornecer uma priori vaga a uma priori informativa equivocada. O algoritmo ABC é sensível a escolha da priori, o que implica que deve haver cuidado com essa escolha uma vez que o algoritmo de rejeição aceita valores vindos da priori. No caso da regressão linear, para o tamanho amostral 50, para os parâmetros β_0 e β_1 não foi observada sensibilidade do método a escolha de distribuições a priori, entretanto para σ^2 existe diferença, pois as maiores discrepâncias estão nos casos em que as distribuições a priori são vagas.

Já para os modelos de dinâmica populacional, a estimação por máxima verossimilhança mostrou-se insatisfatória devido ao fato da função de verossimilhança para EDOs ser irregular e possuir muitos máximos locais, dificultando a maximização da função. Dessa forma, na medida que a complexidade do modelo aumenta, é aconselhável a utilização dos métodos MCMC e ABC. Devido ao fato do método MCMC também precisar da função de verossimilhança para a realização da inferência, observou-se por meio do estudo de simulação para o modelo Lotka-Volterra, que o mesmo pode convergir para máximos locais.

Portanto, embora o tempo computacional por meio do ABC tenha sido maior para alguns casos, devido ao algoritmo LSODA que aproxima as soluções da EDO, as estimativas por meio do ABC foram mais próximas das verdadeiras, bem como foi o método que proporcionou menores medidas de discrepância e melhores resíduos. Dessa forma, o ABC pode ser utilizado como um método alternativo para a inferência de modelos de equações diferenciais.

Para os próximos passos do estudo, pretende-se implementar os algoritmos ABC-MCMC e ABC-Sequencial para modelos de dinâmica populacional, bem como compará-los com o algoritmo de rejeição ABC. Além disso, modelos envolvendo equações diferenciais estocásticas com mais parâmetros serão avaliados.

REFERÊNCIAS

- BAYES, M. e M. PRICE, 1763 An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. *Philosophical Transactions of the Royal Society of London* **53**: 370–418.
- BEAUMONT, M. A., 1999 Detecting population expansion and decline using microsatellites. *Genetics* **153**: 2013–2029.
- BEAUMONT, M. A., 2003 Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**: 1139–1160.
- BEAUMONT, M. A., 2010 Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**: 379–406.
- BEAUMONT, M. A., J.-M. CORNUET, J.-M. MARIN, e C. P. ROBERT, 2009 Adaptive approximate Bayesian computation. *Biometrika* **96**: 983–990.
- BEAUMONT, M. A., W. ZHANG, e D. J. BALDING, 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.
- BECKER, N. G. e T. BRITTON, 1999 Statistical studies of infectious disease incidence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**: 287–307.
- BEGON, M., J. L. HARPER, e C. R. TOWNSEND, 1986 *Ecology. Individuals, populations and communities*. Blackwell scientific publications.
- BERGER, J. O., 1990 Robust bayesian analysis: sensitivity to the prior*. *Journal of statistical planning and inference* **25**: 303–328.
- BLUM, M. e V. C. TRAN, 2010 HIV with contact-tracing: a case study in Approximate Bayesian Computation. *Biostatistics* **11**: 644–660.
- BLUM, M. G. B. e O. FRANCOIS, 2010 Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing* **20**: 63–73.
- BLUM, M. G. B. e M. JAKOBSSON, 2011 Deep divergences of human gene trees and models of human origins. *Molecular Biology and Evolution* **28**: 889–898.
- BOX, G. E. P. e G. C. TIAO, 1992 *Bayesian Inference in statistical analysis*. New York.
- BOYCE, W. E. e R. C. DiPRIMA, 2012 *Elementary Differential Equations*. WileyPLUS, 10th edition.

- CAMPBELL, D. A., G. HOOKER, e K. B. MCAULEY, 2012 Parameter estimation in differential equation models with constrained states. *Journal of Chemometrics* **26**: 322–332.
- CASELLA, G. e R. BERGER, 2002 *Statistical Inference*. L.Pacific Grove, CA: Duxbury.
- CHAPRA, S. C., 2012 *"Applied numerical methods."With MATLAB for Engineers and Scientists*. McGRAW-HILL INTERNATIONAL EDITION, fourth edition.
- CHOWELL, G., J. M. HYMAN, L. . M. A. BETTENCOURT, e C. CASTILLO-CHAVEZ, 2009 *Mathematical and Statistical Estimation Approaches in Epidemiology*. Dordrecht: Springer.
- CLÉMENÇON, S., V. CHI TRAN, e H. DE ARAZOZA, 2008 A stochastic SIR model with contact-tracing: large population limits and statistical inference. *Journal of Biological Dynamics* **2**: 392–414.
- CLEVELAND, W. S., S. J. DELVIN, e E. GROSSE, 1988 Regression by local fitting: methods, properties, and computational algorithms. *Journal of econometrics. Journal of econometrics* **37**: 87–114.
- CORNUET, J. M., F. SANTOS, M. A. BEAUMONT, C. P. ROBERT, J. M. MARIN, D. J. BALDING, T. GUILLEMAUD, e A. ESTOUP, 2008 Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713–2719.
- CORREIA-GOMES, C., T. ECONOMOU, T. BAILEY, P. BRAZDIL, L. ALBAN, e J. NIZA-RIBEIRO, 2014 Transmission parameters estimated for *Salmonella typhimurium* in swine using susceptible-infectious-resistant models and a Bayesian approach. *BMC veterinary research* **10**: 1–11.
- CSILLÉRY, K., M. G. B. BLUM, O. E. GAGGIOTTI, e O. FRANÇOIS, 2010 Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution* **25**: 410–418.
- DALEY, J. D. e J. GANI, 2001 *Epidemic modelling: an introduction*. Cambridge University Press.
- DEGROOT, M. H. e M. J. SCHERVISH, 2012 *Probability and Statistics*. Boston, fourth edition.
- DEL MORAL, P., A. DOUCET, e A. JASRA, 2006 Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B* **68**: 411–436.

- DEL MORAL, P., A. DOUCET, e A. JASRA, 2012 An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing* **22**: 1009–1020.
- DIEKMANN, O. e J. A. P. HEESTERBEEK, 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. John Wiley & Sons.
- DIGGLE, P. J. e R. J. GRATTON, 1984 Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society. Series B (Methodological)* **26**: 193–227.
- DOKSUM, K. A. e A. Y. LO, 1990 Consistent and robust bayes procedures for location based on partial information. *The Annals of Statistics* **18**: 443–453.
- DROVANDI, C. C., A. N. PETTITT, e M. J. FADDY, 2011 Approximate Bayesian Computation using indirect inference. *Journal of the Royal Statistical Society Series C-Applied Statistics* **60**: 317–337.
- ESPOSITO, W. R. e C. A. FLOUDAS, 2000 Deterministic global optimization in nonlinear optimal control problems. *Journal of Global Optimization* **17**: 97–126.
- ESTOUP, A., M. BEAUMONT, F. SENNETOT, C. MORITZ, e J.-M. CORNUET, 2004 Genetic Analysis of Complex Demographic Scenarios: Spatially Expanding Populations of the Cane Toad, *Bufo Marinus*. *Evolution* **58**: 2021.
- EXCOFFIER, L., A. ESTOUP, e J.-M. CORNUET, 2005 Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727–1738.
- GELMAN, A., J. B. CARLIN, H. S. STERN, e D. B. RUBIN, 2004 *Bayesian data analysis*. Chapman & Hall/CRC, Washington, second edition.
- GEMAN, S. e D. GEMAN, 1984 Estochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEE Transactions on pattern analysis and machine intelligence* **6**: 721–741.
- GLICKMAN, M. E. e D. A. VAN DYK, 2007 Basic Bayesian Methods **404**: 319–338.
- GRIFFITHS, R. C. e S. TAVARÉ, 1994 Simulating Probability Distributions in the Coalescent. *Theoretical Population Biology* **46**: 131–159.
- HAIRER, E., S. P. NØRSETT, e G. WANNER, 1993 *Solving ordinary differential equations I: Nonstiff problems*. Springer-Verlag.
- HAMMOND, B. J. e D. A. J. TYRRELL, 1971 A mathematical model of common-cold epidemics on Tristan da Cunha. *Epidemiology & Infection* **69**: 423–433.

- HENS, N., Z. SHKEDY, M. AERTS, C. FAES, P. VAN DAMME, e P. BEUTELS, 2012 *Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective*. Springer Science & Business Media.
- HINDMARSH, A. C., 1983 ODEPACK, A Systematized Collection of ODE Solvers. "In Scientific Computing". IMACS Transactions on Scientific Computation **1**: 55–64.
- HOFFMAN, K., 2013 *Analysis in Euclidean space*. Courier Corporation.
- JABOT, F. e J. CHAVE, 2009 Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. Ecology Letters **12**: 239–248.
- KEELING, M. J. e P. ROHANI, 2008 *Modeling infectious diseases in humans and animals*. Princeton University Press.
- KERMACK, W. O. e A. G. MCKENDRICK, 1927 A contribution to the mathematical theory of epidemics. In: Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society pp. 700–721.
- KOLMOGOROV, A. N., 1950 *Foundations of the Theory of Probability*.
- KYPRAIOS, T., P. NEAL, e D. PRANGLE, 2017 A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. Mathematical Biosciences **287**: 42–53.
- LANDE, R., S. ENGEN, e B.-E. SAETHER, 2003 *Stochastic populated dynamics in ecology and conservation*. Oxford University Press on Demand.
- LOTKA, A. J., 1920 Analytical note on certain rhythmic relations in organic systems. Proceedings of the National Academy of Sciences **6**: 410–415.
- LOTKA, A. J., 1925 Principles of physical biology. Baltimore: Waverly .
- LU, Z., R. R. M. MITCHELL, R. L. R. SMITH, J. S. J. KARNS, J. J. A. S. VAN KESSEL, D. D. R. WOLFGANG, Y. H. Y. SCHUKKEN, e Y. T. Y. GROHN, 2013 Invasion and transmission of Salmonella Kentucky in an adult dairy herd using approximate Bayesian computation. BMC veterinary research **9**: 245.
- MARIN, J.-M., P. PUDLO, C. P. ROBERT, e R. RYDER, 2011 Approximate Bayesian Computational methods pp. 1–28.
- MARJORAM, P., J. MOLITOR, V. PLAGNOL, e S. TAVARE, 2003 Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America **100**: 15324–8.

- McKINLEY, T., A. R. COOK, e R. DEARDON, 2009 Inference in Epidemic Models without Likelihoods. *The International Journal of Biostatistics* **5**.
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, e E. TELLER, 1953 Equation of state calculations by fast computing machines. *Journal Chemical Physics* **21**: 1087–1092.
- MILLER, N., A. ESTOUP, S. TOEPFER, D. BOURGUET, e L. LAPCHIN, 2005 Multiple Transatlantic Introductions of the Western Corn Rootworm. *Science* **310**: 992–992.
- MOOD, A. M., F. A. GRAYBILL, e D. C. BOES, 1974 *Introduction to the Theory of Statistics*. McGRAW-HILL INTERNATIONAL EDITION.
- MORAL, R. D. A. e C. G. B. DEMÉTRIO, 2016 Package bivrp' .
- MURRAY, I., Z. GHAHRAMANI, e D. MACKAY, 2006 MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, Arlington, Virginia.
- MURRAY, J. D., 2002 *Mathematical biology: I. An introduction*. Springer Science & Business Media, third edition.
- NADARAYA, E. A., 1964 On estimating regression. *Theory of Probability & Its Applications* **9**: 141–142.
- O'NEILL, P. D., 2002 A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Biosciences, Mathematical* **180**: 103–114.
- PASCUAL, M., M. P. CHAPUIS, F. MESTRES, J. BALANYÀ, R. B. HUEY, G. W. GILCHRIST, L. SERRA, e A. ESTOUP, 2007 Introduction history of *Drosophila subobscura* in the New World: A microsatellite-based survey using ABC methods. *Molecular Ecology* **16**: 3069–3083.
- PETZOLD, L., 1983 Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations*. *SIAM journal on scientific and statistical computing* **4**: 136–148.
- PRITCHARD, J. K., M. T. SEIELSTAD, A. PEREZ-LEZANN, e M. W. FELDMAN, 1999 Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**: 1791–1798.
- R CORE TEAM, 2015 R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria.

- RAYNER, G. D. e H. L. MACGILLIVRAY, 2002 Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing* **12**: 57–75.
- ROBERT, C. P., J.-M. CORNUET, J.-M. MARIN, e N. S. PILLAI, 2011 Lack of confidence in Approximate Bayesian Computation model choice. *Proceedings of the National Academy of Sciences* **108**: 15112–15117.
- ROSS, R., 1916 An application of the theory of probabilities to the study of a priori pathometry. Part I. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* **92**: 204–230.
- RUBIN, D. B., 1984 Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics* **12**: 1151–1172.
- SALTELLI, A., M. RATTO, T. ANDRES, F. CAMPOLONGO, J. CARIBONI, D. GATELLI, M. SAISANA, e S. TARANTOLA, 2008 *Global Sensitivity Analysis. The Primer*. John Wiley & Sons.
- SHIBLI, M., S. GOOCH, H. LEWIS, e D. TYRRELL, 1971 Common Colds on Tristan da Cunha. *The Journal of Hygiene* **69**: 255–262.
- SINGH, R. S. e M. K. UYENOYAMA, 2004 *The evolution of population biology*. Cambridge University Press.
- SISSON, S. A. e Y. FAN, 2011 Likelihood-free Markov chain Monte Carlo, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC pp. 313–335.
- SISSON, S. A., Y. FAN, e M. M. TANAKA, 2007 Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **106**: 1760–1765.
- STEPHENS, M. e P. DONNELLY, 2000 Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**: 605–635.
- STEWART, J., 2006 *Calculus*. Cengage Learning, fifth edition.
- TANAKA, M. M., A. R. FRANCIS, F. LUCIANI, e S. A. SISSON, 2006 Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics* **173**: 1511–1520.
- TAVARÉ, S., D. J. BALDING, R. C. GRIFFITHS, e P. DONNELLY, 1997 Inferring coalescence times from DNA sequence data. *Genetics* **145**: 505–518.
- TONI, T., D. WELCH, N. STRELKOWA, A. IPSEN, e M. P. H. STUMPF, 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* **6**: 187–202.

- TURCHIN, P., 2003 *Complex Population Dynamics: A Theoretical/ Empirical Synthesis*. Princeton University Press.
- VOLTERRA, V., 1926 Fluctuations in the abundance of a species considered mathematically. *Nature* **118**: 558–560.
- WACKERLY, D. D., W. MENDENHALL, e R. L. SCHEAFFER, 2007 *Mathematical statistics with applications*. Thomson Higher Education.
- WATSON, G. S., 1964 Smooth Regression Analysis. *Sankhyā: The Indian Journal of Statistics, Series A* **26**: 359–372.
- WILKINSON, R. D., 2013 Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error. *Statistical Applications in Genetics and Molecular Biology* **12**: 129–141.
- WORSTER, D., 1994 *Nature's economy: a history of ecological ideas*. Cambridge University Press.

APÊNDICES

Apêndice A

A.1 Ilustração do método de Euler

Considere a equação diferencial

$$dy/dx = \frac{y \ln(y)}{x},$$

tomando 10 iterações com $\Delta = 0,1$, em que os valores iniciais das variáveis x e y são $y(2) = e$, ou seja, $(x_0, y_0) = (2, e)$. O cálculo da derivada no ponto inicial é dado por

$$y' = \frac{dy}{dx} = \frac{y(2) \ln(y(2))}{2} = \frac{e \ln(e)}{2} = \frac{e}{2} \approx 1,359141,$$

o que significa dizer que o coeficiente angular da reta entre $x_0 = 2$ e $x_1 = x_0 + \Delta = 2 + 0,1 = 2,1$ é aproximadamente 1,359141. Para as próximas iterações, mantém-se fixo o valor de Δ , atualizando os demais valores da expansão em série de Taylor, como apresentado a seguir:

$$y(x + h) \approx y(x) + \Delta y'(x, y),$$

Assim,

$$y(2, 0 + 0,1) = y(2, 1) \approx y(2) + 0,1 \frac{e \ln(e)}{2} = e + 0,1 \frac{e}{2} = 2,854196,$$

$$y(2, 1 + 0,1) = y(2, 2) \approx y(2, 1) + 0,1 \frac{y(2, 1) \ln(y(2, 1))}{2}$$

$$= y(2, 1) + 0,1 \frac{2,854196 \ln(2,854196)}{2}$$

$$= 2,854196 + 0,1 \cdot 1,425454 = 2,996741,$$

⋮

⋮

$$y(2, 9 + 0,1) = y(3, 0) \approx 4,209439 + 0,1(2,086328) = 4,418072.$$

Para a equação diferencial em questão, a solução explícita é facilmente obtida por $y = \exp\{x/2\}$. Por meio da Tabela (A.1), verifica-se que o método de Euler gera valores muito próximos aos valores obtidos de forma explícita, apresentando um erro relativo baixo. Verifica-se também que para aumentar a precisão do método, pode-se reduzir o valor h . Contudo, para um número grande de pequenos passos a quantidade de cálculos envolvidos aumenta consideravelmente.

Figura A.1: Soluções da equação diferencial de forma explícita, y_a , e por meio do método de Euler, y_e

Tabela A.1: Resultados das soluções explícita (y_a) e por meio do método de Euler (y_e)

x	y_e	$\frac{dy}{dx} = \frac{y \ln(y)}{x}$	y_a	% erro relativo $\left(\frac{y_a - y_e}{y_a} \right)$
2,0	2,71828	1,35914	2,71828	0,00000
2,1	2,85420	1,42545	2,85765	0,12000
2,2	2,99674	1,49500	3,00417	0,25000
2,3	3,14624	1,56793	3,15819	0,38000
2,4	3,30304	1,64442	3,32012	0,51000
2,5	3,46748	1,72462	3,49034	0,66000
2,6	3,63994	1,80872	3,66930	0,80000
2,7	3,82081	1,89691	3,85743	0,95000
2,8	4,01050	1,98938	4,05520	1,10000
2,9	4,20944	2,08633	4,26312	1,26000
3,0	4,41807	2,18798	4,48169	1,42000

A.2 Medidas de discrepância para o estudo de simulação

Tabela A.2: Medidas de discrepância para os parâmetros nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_1 = 50$, com $d_{\beta_0} = 3,5197$, $d_{\beta_1} = 0,1340$ e $d_{\sigma^2} = 2,1676$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança

Parâmetro	Tolerância	Cenário	Dist. Euclidiana	Dist. Modular	Dist. do Máximo	Cenário	Dist. Euclidiana	Dist. Modular	Dist. do Máximo
β_0			3,51823	3,54682	3,49304		3,60687	3,51584	3,54478
β_1	0,01		0,13166	0,13407	0,13092		0,13546	0,13398	0,13787
σ^2			1,39388	1,34044	1,33370		2,72248	2,63851	2,73300
β_0			3,52687	3,51255	3,50642		3,50987	3,52054	3,52088
β_1	0,1	1	0,13404	0,13368	0,13373	2	0,13367	0,13342	0,13298
σ^2			1,35249	1,33988	1,35169		2,68018	2,70442	2,68739
β_0			3,51189	3,50517	3,51304		3,51668	3,51018	3,50427
β_1	0,5		0,13381	0,13399	0,13370		0,13388	0,13359	0,13354
σ^2			1,35249	1,35079	1,35212		2,64631	2,64861	2,65069
β_0			8,31864	7,17540	7,92335		9,32555	9,57613	10,51317
β_1	0,01		0,24984	0,21935	0,23969		0,25665	0,28619	0,30074
σ^2			1,45101	1,46627	1,55064		4,05998	2,88096	3,61284
β_0			7,67057	7,25961	7,81065		9,77767	9,63592	11,89546
β_1	0,1	3	0,23207	0,21712	0,23319	4	0,27738	0,27692	0,32666
σ^2			1,39857	1,34830	1,50955		3,47168	2,92642	3,62410
β_0			7,71698	7,17182	7,72175		11,96557	10,90587	15,13783
β_1	0,5		0,22867	0,21745	0,22806		0,32754	0,30444	0,39881
σ^2			1,41201	1,34452	1,43876		3,11063	2,84924	3,38274
β_0			3,51024	3,39925	3,49243		3,63614	3,42032	3,55129
β_1	0,01		0,13351	0,13333	0,13372		0,13556	0,13219	0,13646
σ^2			1,35395	1,41230	1,34429		2,59068	2,65239	2,66709
β_0			3,56158	3,56115	3,55970		3,53677	3,56357	3,54324
β_1	0,1	5	0,13448	0,13512	0,13427	6	0,13454	0,13427	0,13406
σ^2			1,33397	1,34802	1,33664		2,70389	2,63509	2,62773
β_0			3,52588	3,51423	3,47275		3,53008	3,51150	3,52080
β_1	0,5		0,13435	0,13370	0,13276		0,13459	0,13380	0,13326
σ^2			1,35228	1,34732	1,35174		2,65163	2,68754	2,64989

Tabela A.3: Medidas de discrepância para os parâmetros nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_2 = 10$, com $d_{\beta_0} = 294,0437$, $d_{\beta_1} = 11,7774$ e $d_{\sigma^2} = 6,2438$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança

Parâmetro	Tolerância	Cenário	Dist. Euclidiana	Dist. Modular	Dist. do Máximo	Cenário	Dist. Euclidiana	Dist. Modular	Dist. do Máximo
β_0			18,3670	18,4992	18,5883		18,5991	18,6841	19,1270
β_1	0,01		0,5980	0,5939	0,6055		0,6109	0,5921	0,6272
σ^2			1,1671	0,4872	1,1190		14,6722	21,3553	14,3221
β_0			18,6747	18,4897	18,8150		18,8657	18,5385	18,5752
β_1	0,1	7	0,6055	0,5998	0,6047	8	0,6089	0,5998	0,6060
σ^2			0,2472	0,4349	0,2659		31,1009	5,6967	54,4637
β_0			18,6085	18,6258	18,7068		18,2867	18,1575	18,0615
β_1	0,5		0,6015	0,6035	0,6046		0,5914	0,5890	0,5879
σ^2			0,3246	0,4411	0,2923		14,9171	3,9601	35,3026
β_0			73,6783	73,1243	72,1648		146,0009	141,5660	158,3189
β_1	0,01		1,7138	1,6714	1,6685		3,3609	3,0337	3,5813
σ^2			7,8058	77,8189	6,2908		284,4792	680,6575	247,3107
β_0			77,2986	74,4779	77,6709		161,2005	144,9891	161,0228
β_1	0,1	9	1,7827	1,7265	1,7711	10	3,7512	3,1322	3,6910
σ^2			10,6092	73,0139	13,3129		347,3191	647,0477	292,2341
β_0			76,6504	69,4482	75,0931		157,3587	143,5996	149,5330
β_1	0,5		1,7483	1,6058	1,7081		3,6076	3,1027	3,3823
σ^2			17,8647	52,5776	23,8457		408,0832	629,6562	369,5487
β_0			18,0384	18,0338	18,2946		18,8666	17,5311	19,0296
β_1	0,01		0,5863	0,5862	0,5872		0,6065	0,5786	0,6039
σ^2			0,9720	0,4071	0,9125		13,5753	16,3955	15,9599
β_0			18,1972	17,7958	17,7460		18,1696	16,4778	18,5165
β_1	0,1	11	0,5852	0,5784	0,5719	12	0,5939	0,5482	0,5893
σ^2			0,0554	0,3894	0,0396		54,1450	9,4442	121,5541
β_0			17,9870	18,1065	18,1436		16,9619	16,1300	17,4183
β_1	0,5		0,5830	0,5869	0,5833		0,5391	0,5235	0,5468
σ^2			0,5430	0,5371	0,5027		149,4758	195,5900	37,5013

Tabela A.4: Medidas de discrepância para os parâmetros da regressão Poisson nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_1 = 50$, com $d_{\beta_0} = 113,6138$ e $d_{\beta_1} = 112,6345$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança

Cenário	Tolerância	Parâmetro	Dist. Euclidiana	Dist. Modular	Dist. do Máximo
1	0,01	β_0	111,7338	112,2007	111,7216
		β_1	112,1832	112,2957	112,1775
	0,1	β_0	111,4870	111,8547	111,5614
		β_1	112,1244	112,1979	112,1463
	0,5	β_0	110,7333	110,2931	111,2218
		β_1	111,7954	111,8245	111,6473
2	0,01	β_0	113,2082	113,3439	113,4582
		β_1	112,5600	112,6639	112,6332
	0,1	β_0	113,2703	113,3020	113,2745
		β_1	112,6066	112,6218	112,6072
	0,5	β_0	113,2517	113,2576	113,2676
		β_1	112,9421	112,6066	113,0228

Tabela A.5: Medidas de discrepância para os parâmetros da regressão Poisson nos diferentes cenários (priori, tolerância e distância), considerando o tamanho amostral $N_2 = 10$, com $d_{\beta_0} = 115,3696$ e $d_{\beta_1} = 112,5526$ as discrepâncias para os parâmetros do modelo considerando a máxima verossimilhança

Cenário	Tolerância	Parâmetro	Dist. Euclidiana	Dist. Modular	Dist. do Máximo
3	0,01	β_0	107,4727	109,0379	107,5361
		β_1	110,7314	111,0774	110,7324
	0,1	β_0	106,7591	108,3828	106,8131
		β_1	110,4895	110,8839	110,5320
	0,5	β_0	103,1874	102,0377	104,3928
		β_1	108,5566	109,2086	108,6202
4	0,01	β_0	114,1670	114,0660	114,1991
		β_1	112,3525	112,3814	112,3850
	0,1	β_0	114,1325	114,2155	114,2048
		β_1	112,4257	112,3734	112,4631
	0,5	β_0	114,1888	114,0086	114,2834
		β_1	113,4136	112,1326	113,6197

Apêndice B

Códigos realizados no R:

B.1 Estudo dos casos de influenza de garotos em um internato por meio da máxima verossimilhança

```
#-----
# Dados de incidência de influenza em garotos de um internato
#-----
dat<-c(1, 3, 7, 25, 72, 222, 282, 256, 233, 189, 123, 70, 25, 11, 4)

#-----
# Função para gerar dados do modelo SIR utilizando o pacote deSolve
#-----
require(deSolve)
closed.sir.model <- function (t, x, params) {
  inc <- params["beta"]*x[1]*x[2]          # incidence
  list(c(-inc,inc-params["gamma"]*x[2]))
}

prediction <- function (params, times) {
  out <- ode(
    func=closed.sir.model,
    y=params[c("X.0","Y.0")],
    times=c(0,times),
    parms=params
  )
  out[-1,3]
}

#-----
# Função para avaliar o logaritmo da função de verossimilhança
#-----
loglik <- function (params, data) {
  times <- (0:15)
  pred <- prediction(params,times)
  sum(dnorm(x=dat,mean=pred,sd=params["sigma"],
            log=TRUE))
}
```

```

params <- c(X.0=762,Y.0=1,gamma=NA,beta=NA,sigma=1)

# Função objetiva
f <- function (beta, gamma) {
  params[c("beta","gamma")] <- c(beta,gamma)
  -loglik(params,dat)
}

require(bbmle)
guess <- list(beta=0.003,gamma=0.5)
fit0 <- mle2(f,start=guess)

#-----
# Gráfico de contorno para a função de verossimilhança
#-----
grid <- expand.grid(beta=seq(0.0008,to=0.005,length=50),
                    nu=seq(from=0.14,to=1,length=50))
grid$loglik <- with(grid,mapapply(f,beta,nu))

require(lattice)
contourplot(loglik~beta+nu,data=grid,cuts=40,
            ylab=list(expression(nu),cex=2),
            xlab=list(expression(beta),cex=2),
            scales=list(cex=1.6))
trellis.focus("panel", 1, 1, highlight=FALSE)
lpoints(x=coef(fit0)[1],y=coef(fit0)[2],col='black',pch=16)

B.2 Análise dos dados de gripe comum na ilha de Tristão da Cunha, por meio da rejeição ABC

#-----
# Análise para os dados de gripe comum em Tristão da Cunha utilizando a
# distância euclidiana
#-----
It = c(1, 1, 3, 7, 6, 10, 13, 13, 14, 14, 17, 10, 6, 6, 4, 3, 1, 1, 1, 1, 0)
Rt = c(0, 0, 0, 0, 5, 7, 8, 13, 13, 16, 16, 24, 30, 31, 33, 34, 36, 36, 36,
       36, 37)

```

```

#-----
# Gerando um erro aleatório
#-----

it <- It+rnorm(21,0,0.5)
rt <- Rt+rnorm(21,0,0.5)
(refdist <- sum((It-it)^2+(Rt-rt)^2))

#-----
# Função para gerar dados do modelo SIR sem a utilização do pacote deSolve
#-----
SIR<-function(S0, I0=1, beta, nu,t){
  S<-rep(0,t)
  I<-rep(0,t)
  R<-rep(0,t)
  S[1]<-S0-I0
  I[1]<-I0
  for(step in 1:(t-1)){
    ds= -beta*S[step]*I[step]
    di= beta*S[step]*I[step] - nu*I[step]
    dr= nu*I[step]
    S[step+1]=S[step]+ds
    I[step+1]=I[step]+di
    R[step+1]=R[step]+dr
  }
  datasir<-cbind(S,I,R)
}

#-----
# Ajuste do modelo utilizando o método de rejeição ABC
#-----

number_samples = 100000

beta_prior =  runif(number_samples, 0, 1)
nu_prior   =  runif(number_samples, 0, 1)
S0_prior   =  runif(number_samples,37,100)

param_t = 21  # The data cover 21 days

```

```

beta_posterior = NULL
nu_posterior = NULL
S0_posterior = NULL
fit_posterior = NULL

for(run in 1:number_samples){
  sim_data = SIR(
    S0=S0_prior[run],
    beta=beta_prior[run],
    nu=nu_prior[run],
    t=param_t)
  distance = (sum((It-sim_data[,2])^2+(Rt-sim_data[,3])^2))
  a<-is.nan(distance)
  if((a == FALSE)&(distance<= reldist)){
    beta_posterior <- cbind(beta_posterior, beta_prior[run])
    nu_posterior <- cbind(nu_posterior, nu_prior[run])
    S0_posterior <- cbind(S0_posterior, S0_prior[run])
    fit_posterior <- cbind(fit_posterior, distance)
  }
}

posterior <- cbind(as.vector(beta_posterior),
                  as.vector(nu_posterior),
                  as.vector(S0_posterior))

fittedSIR<-SIR(S0=mean(posterior[,3]), beta=mean(posterior[,1]),
              nu=mean(posterior[,2]),t=21)

```

B.3 Simulação de dados e ajuste do modelo de Lotka-Volterra por meio da abordagem MCMC

```

#-----
# Função para gerar dados do modelo Lotka-Volterra (presa-predador)
#-----

library(deSolve)
LotVmod <- function (Time, State, Pars) {
  with(as.list(c(State, Pars)), {
    dx = x*(a - c*y)
    dy = -y*(d - b*x)
    return(list(c(dx, dy))))}

```



```

Pars <- c(a = .5, b = .5, c=.5, d=.5)
State <- c(x = 1, y = 0.5)
Time <- seq(0, 15, by = 2)

dat <- as.data.frame(ode(func = LotVmod, y = State,
                        parms = Pars, times = Time))

x<-dat$x
y<-dat$y

#-----
# Ajuste do modelo utilizando a abordagem MCMC
#-----

library(R2OpenBUGS)

model <- function(){

solution[1:ngrid, 1:ndim] <- ode(init[1:ndim], tgrid[1:ngrid],
                                D(C[1:ndim], t), origin, tol)

  D(C[X], t) <- a*C[X] - c*C[X]*C[Y]
  D(C[Y], t) <- b*C[X]*C[Y] - d*C[Y]

  init[1] <- 1
  init[2] <- 0.5

  for (i in 1:ngrid){
    Xt[i] ~ dnorm(solution[i, 1], tau[1])
    Yt[i] ~ dnorm(solution[i, 2], tau[2])
  }

  a ~ dunif(0,1)
  b ~ dunif(0,1)
  c ~ dunif(0,1)
  d ~ dunif(0,1)

  for(k in 1:2){
    tau[k] ~ dgamma(0.001,0.001)
    stdev[k] <- pow(tau[k], -0.5)}}

```

```

data<- list(X=1, Y=2,
            ndim = 2,
            origin = 1,
            tol = 1.0E-3,
            ngrid = 8,
            tgrid = c(1,2,3,4,5,6,7,8),
            Xt = c(1.0000000, 1.5684173, 1.6454801, 0.8737835,
                   0.5240209, 0.5392857, 0.7874535, 1.2845173),
            Yt = c(0.5000000, 0.6614022, 1.3335418, 1.7360608,
                   1.2295908, 0.7540122, 0.5278987, 0.5366954)
)

inits <- list(list(a = 0.5, b = 0.5, c = 0.5, d = 0.5, tau=c(1,1)),
              list(a = 0.55, b = 0.4, c = 0.6, d = 0.65, tau=c(1,1)),
              list(a = 0.4, b = 0.7, c = 0.5, d = 0.5, tau=c(1,1)))

out <- bugs(data = data, inits = inits,
            parameters.to.save = c("a", "b","c","d","tau"),
            model.file = model, n.chains = 3, n.iter = 100000,
            n.burnin = 10000, debug=TRUE)

pars<-as.numeric(out$summary[-7,1])

```

B.4 Ajuste do modelo de Lotka-Volterra para lebre e lince por meio da abordagem rejeição ABC

```

#-----
# Função para gerar dados do modelo Lotka-Volterra (presa-predador)
#-----

library(deSolve)
LotVmod <- function (Time, State, Pars) {
  with(as.list(c(State, Pars)), {
    dx = x*(a - c*y)
    dy = -y*(d - b*x)
    return(list(c(dx, dy)))
  })
}

```

```

State <- c(x = 19.58, y = 30.09)
Time <- c(1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855,
         1856, 1857, 1858, 1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866,
         1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877,
         1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888,
         1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899,
         1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910,
         1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921,
         1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932,
         1933, 1934, 1935)

# Dados reais de Lebre e Lince
x<- c(19.58, 19.6, 19.61, 11.99, 28.04, 58, 74.6, 75.09, 88.48, 61.28,
      74.67, 88.06, 68.51, 32.19, 12.64, 21.49, 30.35, 2.18, 152.65,
      148.36, 85.81, 41.41, 14.75, 2.28, 5.91, 9.95, 10.44, 70.64,
      50.12, 50.13, 101.25, 97.12, 86.51, 72.17, 38.32, 10.11, 7.74,
      9.67, 43.12, 52.21, 134.85, 134.86, 103.79, 46.1, 15.03, 24.2,
      41.65, 52.34, 53.78, 70.4, 85.81, 56.69, 16.59, 6.16, 2.3, 12.82,
      4.72, 4.73, 37.22, 69.72, 57.78, 28.68, 23.37, 21.54, 26.34, 53.1,
      68.48, 75.58, 57.92, 40.97, 24.95, 12.59, 4.97, 4.5, 11.21, 56.6,
      69.63, 77.74, 80.53, 73.38, 36.93, 4.64, 2.54, 1.8, 2.39, 4.23,
      19.52, 82.11, 89.76, 81.66, 15.76)

y<- c(30.09, 45.15, 49.15, 39.52, 21.23, 8.42, 5.56, 5.08, 10.17, 19.6,
      32.91, 34.38, 29.59, 21.3, 13.69, 7.65, 4.08, 4.09, 14.33, 38.22,
      60.78, 70.77, 72.77, 42.68, 16.39, 9.83, 5.8, 5.26, 18.91, 30.95,
      31.18, 46.34, 45.77, 44.15, 36.33, 12.03, 12.6, 18.34, 35.14, 43.77,
      65.69, 79.35, 51.65, 32.59, 22.45, 16.16, 14.12, 20.38, 33.33, 46,
      51.41, 46.43, 33.68, 18.01, 8.86, 7.13, 9.47, 14.86, 31.47, 60.57,
      63.51, 54.7, 6.3, 3.41, 5.44, 11.65, 20.35, 32.88, 39.55, 43.36,
      40.83, 30.36, 17.18, 6.82, 3.19, 3.52, 9.94, 20.3, 31.99, 42.36,
      49.08, 53.99, 52.25, 37.7, 19.14, 6.98, 8.31, 16.01, 24.82, 29.7,
      35.4)

set.seed(138)
xx = x + rnorm(length(x), mean = 0, sd =30)
set.seed(112)
yy = y + rnorm(length(y), mean = 0, sd =sd(y))
refdist<-sum((x-xx)^2+(y-yy)^2)

```

```

#-----
# Ajuste do modelo por meio do método de rejeição ABC
#-----

number_samples = 100000

a_prior = runif(number_samples, 0, 1)
b_prior = runif(number_samples, 0.01, 0.02)
c_prior = runif(number_samples, 0.02, 0.03)
d_prior = runif(number_samples, 0, 1)

a_posterior = NULL
b_posterior = NULL
c_posterior = NULL
d_posterior = NULL

system.time(
for(run in 1:number_samples){
  Pars = c(a=a_prior[run], b=b_prior[run], c=c_prior[run], d=d_prior[run])
  sim_out <- as.data.frame(ode(func = LotVmod, y = State,
                             parms = Pars, times = Time))
  if(sum((x-sim_out[,2])^2+(y-sim_out[,3])^2) <= reldist){
    a_posterior <- cbind(a_posterior, a_prior[run])
    b_posterior <- cbind(b_posterior, b_prior[run])
    c_posterior <- cbind(c_posterior, c_prior[run])
    d_posterior <- cbind(d_posterior, d_prior[run])
  }
}
)

posterior <- cbind(as.vector(a_posterior),
                  as.vector(b_posterior),
                  as.vector(c_posterior),
                  as.vector(d_posterior))

```

B.5 Simulação de dados para o modelo de competição entre duas espécies

```
#-----
# Função para gerar dados do modelo de competição entre espécies
#-----

library(deSolve)

LotVmod <- function (Time, State, Pars) {
  with(as.list(c(State, Pars)), {
    dx = x*(a1 - b1*x - c1*y)
    dy = y*(a2 - b2*y - c2*x)
    return(list(c(dx, dy)))
  })
}

Pars <- c(a1 = 1, b1 = 1, c1 = 1, a2 = 0.75, b2 = 1, c2 = 0.5)
State <- c(x = 2, y = .5)
Time <- seq(0, 15, by = 2)

out <- as.data.frame(ode(func = LotVmod, y = State,
                        parms = Pars, times = Time))

x<-out$x
y<-out$y
```

ANEXOS

Anexo A

A.1 Espaço Euclidiano

De acordo com HOFFMAN (2013) uma métrica em um conjunto M é uma função $d : M \times M \rightarrow \mathbb{R}$, associando-se a cada par de elementos ordenados $x, y \in M$ um número real $d(x, y)$, chamado distância de x a y , de modo que as seguintes condições sejam satisfeitas para quaisquer $x, y, z \in M$:

- d1. $d(x, x) = 0$;
- d2. Se $x \neq y$ então $d(x, y) > 0$;
- d3. $d(x, y) = d(y, x)$;
- d4. $d(x, z) \leq d(x, y) + d(y, z)$.

Um espaço métrico é um par (M, d) , em que M é um conjunto e d é uma métrica em M . Para o espaço euclidiano, os pontos de $\mathbb{R}^n$ são vetores $x = (x_1, \dots, x_n)$ em que cada uma das n coordenadas x_i é um número real. Dados $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ existem algumas formas naturais de definir a distância entre dois pontos no $\mathbb{R}^n$, por exemplo, distâncias euclidiana, modular, do máximo e de Mahalanobis entre outras.