

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS
POSTGRADUATE PROGRAM

Bayesian inference for some deterministic epidemic models

By
Magda D. Drougka

A THESIS
Submitted to the Department of Statistics of the Athens University of Economics and Business in partial fulfilment of the requirements for the degree of Master of Science in Statistics (full-time)

Athens, Greece
September 2019




ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ



ΣΧΟΛΗ ΤΩΝ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Μπεϋζιανή συμπερασματολογία για ορισμένα ντετερμινιστικά επιδημιολογικά μοντέλα

Μάγδα Δ. Δρούγκα

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο τμήμα στατιστικής του Οικονομικού
Πανεπιστημίου Αθηνών ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Στατιστική (Πλήρους Φοίτησης)

Αθήνα, Ελλάδα
Σεπτέμβριος 2019




ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Acknowledgements

First of all, I must thank my supervisor Nikolaos Demiris, without whom this thesis would not be what it is. He gave me absolute freedom in writing my thesis, encouraged me to become independent, and was always available to discuss and to provide valuable insights. I am also indebted to all my fellow students from the graduate who taught me what teamwork means and helped me overcome many hindrance. Especially Eleni, Angeliki, Marianthi, Despina, Dimitra, Sofia our boys Vasilis, Thanasis and Ion. Also all the people in the lab who provided feedback and advice in various stages of my thesis, especially Anastasia Chatzilena and Marios.

I also have to thank my parents Dimitris and Stauroula who put enormous efforts into bringing me up and supported me in every decision I ever made. Overall my family Vaso, Fotis, Aspa, Costas, Magda, Efi, Fotini, Fotis, Maria, their encouragement to learn and to develop my mind and their unconditional love have shaped me. In particular my uncle Antonis who is no longer with us but is in our minds and hearts.

Especially my friends Georgia, Konstantina, Georgia, Afroditi who has tolerated me for so long and gave me advice. For all these years they have shown me what true friendship means. Also many dear friends made sure I never lacked good company. I have to thank Giannis Constantinou for his help, without him my thesis would have been markedly inferior. Popi, Thanasis and Mimi for the unlimited love they showed me from the first moment.

Finally, Giannis for everything. He has contributed more than anyone to my way of thinking and has helped me realise so many things I would never have thought by myself. I omit many others but they already know how much I am indebted to them.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Vita

I was born on January 27rd 1993 in Goura Korinthias, where I got my diploma from high school. In 2011, I entered the Department of Mathematics at the University of Crete. At the last academic semester of my undergraduate studies I have been accepted as intern at the Foundation for Research and Technology Greece for three months in the city of Heraklion, Crete. At the beginning of the next academic year after my graduation from University of Crete, I started the full time master program of Athens University of Economics and Business in Statistics, which ends with this thesis.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Abstract

Markov Chain Monte Carlo (MCMC) is a family of stochastic algorithms which are used to draw random samples from arbitrary probability distributions. This task is necessary to solve a variety of problems in Bayesian modelling, e.g. prediction and model comparison, making MCMC a fundamental tool in modern statistics. As model complexity increases, the performance of classical MCMC algorithms deteriorates due to their potentially limited exploration of the target distribution. The latest developments in statistics and machine learning suggest that Hamiltonian Monte Carlo (HMC) method can offer statistical and computational efficiency com-pared to MCMC. Hamiltonian Monte Carlo (HMC) is a state-of-the-art sampling algorithm for Bayesian computation. Popular probabilistic programming language Stan rely on HMC's generality and efficiency to provide automatic Bayesian inference platforms for epidemic models and more.

Our main purpose in this thesis is to provide a brief description of the most important features of Stan's implementation of HMC in the context of epidemic models. We present the HMC algorithm as well as the advantages of the Stan platform and we introduce information criteria like AIC, WAIC and DIC, the most common methods for selecting statistical models. By applying the above techniques to simulated and real data, it is shown that Stan is a (perhaps the only) general purpose software which provides a built-in mechanism for specifying and solving systems of ordinary differential equations, thus facilitating inference in epidemic models.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Περίληψη

Οι μέθοδοι Μόντε Κάρλο είναι μια μεγάλη κατηγορία υπολογιστικών αλγορίθμων που στηρίζονται σε τυχαία δειγματοληψία με σκοπό την εξαγωγή αριθμητικών αποτελεσμάτων. Η χρήση του MCMC είναι απαραίτητη για την επίλυση ποικίλων προβλημάτων στα Μπεϋζιανά μοντέλα, π.χ. πρόβλεψη και σύγκριση μοντέλων, καθιστώντας το MCMC θεμελιώδες εργαλείο στις σύγχρονες στατιστικές. Καθώς η πολυπλοκότητα των μοντέλων αυξάνεται, οι επιδόσεις των κλασικών αλγορίθμων MCMC εξασθενούν λόγω της ενδεχομένως περιορισμένης εξερεύνησης της κατανομής στόχου.

Οι μεγαλύτερες εξελίξεις στη στατιστική και στη μηχανική μάθηση υποδηλώνουν ότι η μέθοδος Hamiltonian Monte Carlo (HMC) μπορεί να προσφέρει στατιστική και υπολογιστική αποτελεσματικότητα σε σύγκριση με το MCMC. Το Hamiltonian Monte Carlo (HMC) είναι ένας υπερσύγχρονος αλγόριθμος δειγματοληψίας για την Μπεϋζιανή. Η δημοφιλής πιθανολογική γλώσσα προγραμματισμού Stan βασίζεται στη γενίκευση και την αποτελεσματικότητα του HMC για να παρέχει αυτόματες πλατφόρμες Μπεϋζιανών συμπερασμάτων για επιδημιολογικά μοντέλα και πολλά άλλα.

Ο κύριος στόχος μας σε αυτή τη διπλωματική εργασία είναι να παρέχουμε μια σύντομη περιγραφή των πιο σημαντικών χαρακτηριστικών της εφαρμογής του HMC από το Stan στο πλαίσιο των επιδημικών μοντέλων. Παρουσιάζουμε τον αλγόριθμο HMC καθώς και τα πλεονεκτήματα της πλατφόρμας Stan και εισάγουμε κριτήρια πληροφόρησης όπως τα AIC, WAIC και DIC, τις πιο κοινές μεθόδους για την επιλογή στατιστικών μοντέλων. Με την εφαρμογή των παραπάνω τεχνικών σε προσομοιωμένα και πραγματικά δεδομένα, δείχνει ότι το Stan είναι ένα (ίσως το μόνο) λογισμικό γενικού σκοπού που παρέχει έναν ενσωματωμένο μηχανισμό για τον προσδιορισμό και την επίλυση συστημάτων συνήθων

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

διαφορικών εξισώσεων, διευκολύνοντας έτσι τη συμπερίληψη σε επιδημικά μοντέλα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Contents

Acknowledgements **i**

Abstact **v**

Περίληψη **vii**

List of Tables **x**

List of Figures **xiii**

1 Introduction **1**

2 Bayesian Inference **3**

2.1 Bayesian Computation 7

2.2 Bayesian Model Selection 11

3 Epidemic models **13**

3.1 Types of epidemic models 14

3.2 The SIR model 15

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.3 Basic reproduction number 16

3.4 Single strain SIR models 17

4 Applications **19**

4.1 Simulated data 19

4.2 Real data 27

5 Discussion **35**

Bibliography **36**

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

List of Tables

4.1 HMC results for Binomial and Poisson models with simulated data . 21

4.2 HMC results for Neg.Binomial and Beta Binomial models with simulated data 21

4.3 Information criteria for models with simulated data 22

4.4 Execution time (minutes) for simulated data 23

4.5 HMC results for Binomial and Poisson models with real data 28

4.6 HMC results for Neg.Binomial and Beta Binomial models with real data 28

4.7 Information criteria for models with real data 29

4.8 Execution time (minutes) for real data 30

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

List of Figures

3.1 Different stages of SIR model 14

3.2 The change in the epidemic compartments over time, for $\beta = 0.6$ and $\gamma = 0.1$ 15

4.1 A better way to do experiments 20

4.2 Models fit to simulated data 22

4.3 Markov Chains for Binomial model with simulated data 23

4.4 Markov Chains for Poisson model with simulated data 24

4.5 Markov Chains for Neg. Binomial model with simulated data 24

4.6 Markov Chains for Beta Binomial model with simulated data 25

4.7 Trace plots for Binomial model with simulated data 25

4.8 Trace plots for Neg. Binomial model with simulated data 26

4.9 Trace plots for Beta Binomial model with simulated data 26

4.10 Boarding school 27

4.11 Models fit to real data 29

4.12 Markov Chains for Binomial model with real data 30

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

4.13 Markov Chains for Poisson model with real data 31

4.14 Markov Chains for Neg. Binomial model with real data 31

4.15 Markov Chains for Beta Binomial model with real data 32

4.16 Trace plots for Binomial with real data 32

4.17 Trace plots for Neg. Binomial with real data 33

4.18 Trace plots for Beta Binomial with real data 33

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chapter 1

Introduction

In epidemiology one of the most basic steps is the estimation of model parameters describing the mechanisms of disease spread. One or more of the model variables may be unobserved (latent), for this reason inference for epidemic models is complicated, to this end employing a Bayesian approach has the advantage of including uncertainties within the model. To practical applications of statistics, the Bayesian approach has evolved since the introduction of Markov Chain Monte Carlo (MCMC) methods with the use of MCMC in practical applications being greatly facilitated by the early development of the BUGS software (Lunn et al. 2000). Most of the recent studies in infectious disease epidemiology use MCMC for learning the model parameters from data (McKinley et al. 2014, O'Neill & Roberts 1999).

As model complexity increases, the performance of classical MCMC algorithms deteriorates due to their potentially limited exploration of the target distribution. The latest developments in statistics and machine learning suggest that Hamiltonian Monte Carlo (HMC) method can offer statistical and computational efficiency compared to MCMC (Betancourt 2017). While there have been a number of successors to the BUGS software, especially over the last decade, including JAGS (Plummer 2017) and Nimble (de Valpine et al. 2017), a relatively new software package called Stan enables the implementation of HMC within a probabilistic modeling language. In addition, it appears that Stan is the first such software which has built-in solvers

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

for systems of ordinary differential equations (ODEs) (Carpenter et al. 2017).

Our main purpose in this thesis is to provide a brief description of the most important features of Stan's implementation of HMC in the context of epidemic models. Since the epidemic models which we use are based in the Bayesian statistical inference, we give the basic theory and tools of Bayesian Inference in Chapter 2. In this chapter we will introduce the Bayes' formula and the components of Bayesian inference such as likelihood, the prior distributions and the posterior distributions. Moreover, we present the algorithm HMC as well as the advantages of the platform Stan and at the end of this chapter we introduce the information criteria like AIC, WAIC and DIC most common methods of assessing the goodness of fit of anestimated statistical models.

In chapter 3 we make an introduction to the epidemic models and the types of them, specifically in this thesis we will deal with deterministic epidemiological models such as SIR. The major assumptions that we make in these models is the population size in a subgroup is differentiable with respect to time and that the epidemic process is deterministic. In other words, the changes in population of a subgroups can be calculated using only the history that was used to develop the model. This assumption is tested in order to see whether it affects the spread of epidemic and its transmission rate, which is seen through the basic reproduction number, which analyze later in the chapter. Afterward we present the four models we use and their assumptions

In the next chapter we test extensively the aforementioned models, we use a way to extract the likelihood and calculate the basic reproduction number. We applied simulated data and data gathered from epidemic at British boarding school, in order to provide the results calculated in Rstan.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chapter 2

Bayesian Inference

Bayesian inference is a method of statistical inference in which Bayes's theorem is used to update the probability for a hypothesis as more information becomes available, in other words Bayesian inference techniques specify how one should update one's beliefs upon observing data. Let us see how Bayes's theorem is defined for two contingencies A and B. Bayes' theorem shows the relationship between two conditional probabilities that are the reverse of each other, indeed the conditional probability, or 'posterior probability', of an event A after B is observed in terms of the 'prior probability' of A, prior probability of B, and the conditional probability of B given A. The Bayes theorem has the following form:

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} \tag{2.1}$$

The rationale behind Bayes' inference is similar to that of classical statistics, ie we have a parameter θ of the population we want to estimate, and the probability $f(x|\theta)$ that determines the probability of observing different x, under different values of the parameter θ. The main difference is that θ is used as a random quantity. This difference leads us to a completely different approach to the interpretation of classical statistics.

In essence, our inference will be based on $f(\theta|x)$ and not on $f(x|\theta)$, ie the probability of parameter distribution given x (data) rather than x given the parameter. In

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

order to be able to do this, we need to determine the prior distribution $f(\theta)$ (prior probability distribution), which represents our beliefs about the distribution of θ before we get any information about the data. The idea of prior distribution of the parameter θ is the core of Bayes theory.

At this point we will formulate Bayes' theorem in a form suitable for random variables, where f denote the densities of the random variables. In equation (2.1) we replace B with observations x, A with parameter set Θ, and probabilities Pr with densities f, resulting in the following:

$$f(\Theta|x) = \frac{f(\Theta)f(x|\Theta)}{f(x)} \tag{2.2}$$

Where $f(\Theta)$ is the set of prior distributions of parameter set Θ before x is observed, $f(x|\Theta)$ is the likelihood of x under a model, and $f(\Theta|x)$ is the joint posterior distribution, sometimes called the full posterior distribution, of parameter set Θ that expresses uncertainty about parameter set Θ after taking both the prior and data into account. Since there are usually multiple parameters, Θ represents a set of j parameters, and may be considered here as

$$\Theta = \theta_1, ..., \theta_j \tag{2.3}$$

The denominator $f(x)$ is written as

$$f(x) = \int f(\Theta)f(x|\Theta)d\Theta \tag{2.4}$$

defining the "marginal likelihood" of x and may be set to an unknown constant c. The marginal likelihood function indicates what x should look like, given the model, before x has been observed. Only the set of prior probabilities and the model's likelihood function are used for the marginal likelihood of x. From the moment we have completed, with respect to Θ, the denominator in the Bayes theorem, it is a function leg as x. Therefore for given x observations, the denominator is constant and is called the normalization constant. Based on this an alternative way of presenting the theorem (2.2) is as follows:

$$f(\Theta|x) \propto f(\Theta)f(x|\Theta) \tag{2.5}$$

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

We see that the posterior distribution is proportional to the prior distribution multiplied by the likelihood function $f(x|\Theta)$.

If we make the equation a little more explicit, Θ is what we are interested in representing the set of parameters. For example, if we are trying to estimate the values of the Gaussian distribution parameters, then Θ represents both the mean (μ) and the standard deviation (σ), while the data represents the total of observations we have. So now the Bayes theorem in a model, according to the above, is written as:

$$\begin{aligned} P(\theta|data) &= \frac{P(\theta)P(data|\theta)}{P(data)} \\ f(\theta|data) &\propto f(\theta)f(data|\theta) \end{aligned} \tag{2.6}$$

So the components of Bayesian inference are:

1. $f(x|\Theta)$ is the likelihood or likelihood function, in which all variables are related in a full probability model.
2. $f(\Theta)$ is the set of prior distributions for parameter set Θ, and uses probability as a means of quantifying uncertainty about Θ before taking the data into account.
3. $f(\Theta|x)$ is the joint posterior distribution that expresses uncertainty about parameter set Θ after taking both the prior and the data into account.

Likelihood

In order to complete the definition of a Bayesian model, both the prior distributions and the likelihood must be approximated or fully specified (Fisher 1921). In the presence of independent observations, the likelihood factorises as

$$f(x|\Theta) = \prod_{i=1}^{n} f(x_i|\Theta) \tag{2.7}$$

The data x affects the posterior distribution $f(\Theta|x)$ only through the likelihood function $f(x|\Theta)$. In this way, Bayesian inference obeys the likelihood principle, which states that for a given sample of data, any two probability models $p(x|\Theta)$ that have the same likelihood function yield the same inference for Θ.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Prior Distributions

In Bayesian inference, a prior probability distribution of an uncertain parameter Θ is a probability distribution that expresses uncertainty about Θ before the data are taken into account (Berger et al. 2009). The parameters of a prior distribution are called hyperparameters, to distinguish them from the parameters (Θ) of the model. When applying Bayes' theorem, the prior is multiplied by the likelihood function and then normalized to estimate the posterior probability distribution, which is the conditional distribution of Θ given the data, moreover the prior distribution affects the posterior distribution. Prior probability distributions traditionally belonged to one of two categories: **informative priors** and **uninformative priors**.

- **Informative priors**: When prior information is available about Θ, it should be included in the prior distribution of Θ, that is definite information about a variable.

- **Uninformative priors**: An uninformative prior expresses vague or general information about a variable, can express "objective" information such as "the variable is positive" or "the variable is less than some limit", uninformative priors are a way of making minimal assumptions about the model.

Posterior Distribution

The best way to fully describe the unknown parameter Θ can be done only through the Posterior distribution, in other words Posterior $p(\Theta|x)$ is our inference, which has as many dimensions as the parameter Θ. The posterior distribution can be written as:

$$f(\Theta|x) = \frac{f(x,\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{f(x)} \propto f(x|\theta)f(\theta) \tag{2.8}$$

where $f(x|\theta)$ is the model likelihood. Inference is conducted via posterior summary statistics from the marginal posterior distributions of parameter vector θ such as the posterior mean, median and credible intervals. When the posterior distribution is not available analytically, we may obtain the posterior summary statistics of interest by generating values from the posterior distribution. Thus, Markov Chain Monte Carlo

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

(MCMC) algorithms can be used to generate values from this distribution. In our case we use adaptive HMC, more details we see following.

2.1 Bayesian Computation

For simple models, an analytical formula for the likelihood function can typically be derived. However for more complex models, an analytical formula is not available raising a computational issue. The likelihood may be unavailable for mathematical reasons, it is not available in closed form as a function of θ, or computational reasons, it is too expensive too calculate.

Usually, the true posterior must be approximated with numerical methods. To see why, let's return to the definition of the posterior distribution:

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\int f(\theta)f(x|\theta)d\theta} \tag{2.9}$$

This integral usually does not have a closed-form solution, so we need an approximation. Since the 90s a method of approximating our posterior is by using Markov Chain Monte Carlo (MCMC), which generates samples in a way that mimics the unknown distribution. The idea of the algorithm is as follows: We begin at a particular value, and "propose" another value as a sample according to a stochastic process. We may reject the sample if the proposed value seems unlikely and propose another. If we accept the proposal, we move to the new value and propose another. The statistical software for Bayesian analysis using Markov chain Monte Carlo (MCMC) methods is WinBugs, which it is based on Bayesian inference Using Gibbs Sampling (BUGS), project started in 1989.

Towards that direction in recent years, with many improvements in the literature and starting with the very notion of MCMC, Hamiltonian Monte Carlo (HMC) is one method of approximating our posterior, which stems from Physics (Duane et al. 1987). After a rise into the statistical community, it is now central in statistical software like Stan.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Stan

Stan is an open-source general purpose inference software, which implements full Bayesian inference with gradient based sampling techniques, a method to approximate posteriors with variational inference as well as penalized maximum likelihood estimation via optimization. The software performs the required computations using advanced techniques such as automatic differentiation (Hoffman & Gelman 2011). Moreover, Stan is one of the few general purpose software which provides a built-in mechanism for specifying and solving systems of ordinary differential equations which makes it suitable for inference in epidemic models (Chatzilena et al. 2019).

The user can write a Bayesian model in a BUGS-like language which is similar to standard statistical notation. The model is then compiled to C++, making the inference faster. A difference between Stan code and other automated platforms such as BUGS and JAGS, is the necessary variable type declarations and statements. Variables are declared by their type, in blocks according to their use, and constraints upon them need to be defined carefully. The first blocks of Stan's model statement consist of data, transformed data, parameters, transformed parameters and generated quantities. Within the model block, sampling notation is very similar to BUGS. Log probability variables can also be accessed directly, or user-defined probability functions can be employed. The Stan model file should have the extension .stan and is portable across the different interfaces.

There are interfaces for R, Python, MATLAB, Julia, Stata, Mathematica, Scala and the command line. According to the interface used, the user needs to call a different function for the different inference methods offered. All these functions include an argument which defines the location and name of the Stan model file. Moreover, extensive diagnostics for the inference are provided. In the following we highlight the basic concepts of HMC focusing on the way they are implemented in Stan.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a Markov Chain Monte Carlo method that uses the derivatives of the density function being sampled to generate efficient transitions spanning the posterior (Betancourt et al. 2014). It uses an approximate Hamiltonian dynamics simulation based on numerical integration which is then corrected by performing a Metropolis acceptance step. It is important to mention that this section translates the presentation of Hamiltonian Monte Carlo by Betancourt and Girolami into the notation of Gelman et al (2013). (Betancourt & Girolami 2015)

The algorithm introduces auxiliary momentum variables ρ to the parameters of the unknown posterior density, θ, with the joint density

$$P(\rho, \theta) = P(\rho, \theta)P(\theta) \tag{2.10}$$

The density of the auxiliary variables is deemed to be the multivariate normal (d dimensions) in most of the cases, as well as in the automatic procedure the probabilistic language Stan offers

$$\rho \sim N_d(0, \Sigma) \tag{2.11}$$

where the covariance matrix Σ works as a Euclidean metric to rotate and scale the unknown-target density (Betancourt & Stein 2011)

In Stan - in which the HMC algorithm is being implemented - the covariance matrix is usually replaced from the identity matrix or estimated from warmups samples and optionally restricted to a diagonal matrix. After specifying the conditional density of the momentum variables, the joint density defines a Hamiltonian,

$$\begin{aligned} H(\rho, \theta) &= -logP(\rho, \theta) \\ &= -logP(\rho|\theta) - logP(\theta) \\ &= T(\rho|\theta) + V(\theta) \end{aligned} \tag{2.12}$$

with the "kinetic energy"

$$T(\rho|\theta) = -logP(\rho|\theta) \tag{2.13}$$

and the "potential energy"

$$V(\theta) = -logP(\theta) \tag{2.14}$$

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

This Hamiltonian function generates a transition by first sampling the auxiliary momentum variables $\rho \sim P(\rho, \theta)$, where $P(\rho, \theta)$ is a distribution independent from the parameters θ when a $N_d(0, \Sigma)$ is assumed. The joint system is evolving via Hamilton's equations

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\partial H}{\partial \rho} = \frac{\partial T}{\partial \rho} \\ \frac{d\rho}{dt} &= -\frac{\partial H}{\partial \rho} = -\frac{\partial T}{\partial \theta} - \frac{\partial V}{\partial \theta} \end{aligned} \tag{2.15}$$

The assumption that the momentum variables are independent from the parameters θ, instantly makes the derivative $\partial T/\partial \theta$ equal to zero, leading to the following

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\partial T}{\partial \rho} \\ \frac{d\rho}{dt} &= -\frac{\partial V}{\partial \theta} \end{aligned} \tag{2.16}$$

This two-state differential equation is being solved by using the leapfrog integrator, as Stan suggests. The leapfrog integrator is a numerical integration algorithm that is specifically adapted to provide stable results for Hamiltonian systems of equations. The leapfrog integrator takes discrete steps of some small time integral ϵ. It begins by drawing a momentum value from the ρ density and then alternates half-step updates of the momentum and full-step updates of the position.

$$\begin{aligned} \rho &\Leftarrow \rho - \frac{\epsilon}{2}\frac{\partial V}{\partial \theta} \\ \theta &\Leftarrow \theta - \epsilon \Sigma \rho \\ \rho &\Leftarrow -\frac{\epsilon}{2}\frac{\partial V}{\partial \theta} \end{aligned} \tag{2.17}$$

L leapfrog steps are applied and a total of Lϵ time is simulated. After the L repetitions of the above three steps, the derived result is two vectors; one for the momentum variable ρ and the other for the parameters θ. A detailed analysis of the numerical integration of the Hamiltonian systems provide in (Leimkuhler & Reich 2004).

After the leapfrog implementation, in order to account for numerical errors, a Metropolis Hastings acceptance step is applied and a decision is made whether to update to the new state $(\pi^{new}, \theta^{new})$ or keep the existing state.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

2.2 Bayesian Model Selection

In Bayesian inference, the most common method of assessing the goodness of fit of an estimated statistical model is a generalization of the frequentist Akaike Information Criterion (AIC). The Bayesian method, like AIC, is not a test of the model in the sense of hypothesis testing. Instead Bayesian inference provides a model fit statistic that is to be used as a tool to refine the current model or select the better-fitting model of different methodologies. To begin with, model fit can be summarized with deviance, which is defined as -2 times the log-likelihood (Gelman et al. 2004), such as

$$D(x|\Theta) = -2log[p(x|\Theta)] \tag{2.18}$$

Due to the development of BUGS software deviance is defined differently in Bayesian inference than frequentist inference. In frequentist inference, deviance is -2 times the log-likelihood ratio of a reduced model compared to a full model, whereas in Bayesian inference, deviance is simply -2 times the log-likelihood. In Bayesian inference, the lowest expected deviance has the highest posterior probability. It is possible to have a negative deviance. Deviance is derived from the likelihood, which is derived from probability density functions (PDF). Evaluated at a certain point in parameter space, a PDF can have a density larger than 1 due to a small standard deviation or lack of variation. Likelihoods greater than 1 lead to negative deviance, and are appropriate.

On its own, the deviance is an insufficient model fit statistic, because it does not take model complexity into account. The effect of model fitting pD is used as the 'effective number of parameters' of a Bayesian model. The sum of the differences between the posterior mean of the model-level deviance and the deviance at each draw i of θ_i is the pD. A related way to measure model complexity is as half the posterior variance of the model-level deviance, known as pD (Gelman et al. 2004)

$$pD = var(D)/2 \tag{2.19}$$

The effect of model fitting, pD can be thought of as the number of 'unconstrained' parameters in the model, where a parameter counts as: 1 if it is estimated with no constraints or prior information, 0 if it is fully constrained or if all the information about the parameter comes from the prior distribution or an intermediate value if both

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

the data and the prior are informative (Gelman et al. 2004). Therefore, by including prior information, Bayesian inference is more efficient in terms of the effective number of parameters than frequentist inference.

Model complexity, pD should be positive. Although pD must be positive since it is related to variance, it is possible for pD to be negative, which indicates one or more problems: log-likelihood is non-concave, a conflict between the prior and the data, or that the posterior mean is a poor estimator. The sum of both the mean model-level deviance and the model complexity is the Deviance Information Criterion (DIC), a model fit statistic that is also an estimate of the expected loss, with deviance as a loss function (Spiegelhalter et al. 2002). DIC is

$$DIC = \bar{D} + pD \tag{2.20}$$

DIC may be compared across different models and even different methods, as long as the dependent variable does not change between models, making DIC the most flexible model fit statistic. DIC is a hierarchical modeling generalization of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Like AIC and BIC, it is an asymptotic approximation as the sample size becomes large. DIC is valid only when the joint posterior distribution is approximately multivariate normal, models should be preferred with smaller DIC. Since DIC increases with model complexity (pD), simpler models are preferred.

The Widely Applicable Information Criterion (WAIC) is an information criterion that is more fully Bayesian than DIC. WAIC is more difficult to calculate because the record-level log-likelihood is required over numerous samples. However, when available, the result more closely resembles leave-one-out cross-validation (LOO-CV).

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chapter 3

Epidemic models

Mathematical models can analyze how infectious diseases evolve to show the likely outcome of an epidemic and help inform public health interventions. Models use some basic assumptions and mathematics to determine parameters for various infectious diseases and use these parameters to measure the impact of various interventions such as bulk vaccination programs. One of the earliest studies of the non-linearity of an epidemic model was contained in a paper by (Hamer 1906). Hamer postulated that the probability of a new infection in the next discrete time-step was proportional to the product of the number of Susceptibles and the number of infectives. Modeling helps us decide which intervention we could avoid and which one to test. An epidemic model is a simplified way of explaining the transmission of contagious disease through human species. (Daley & Gani 2005)

The investigation of the disease event is called epidemiology. An epidemic is an unusually large, short-term outbreak of a disease. A disease is called endemic if it persists in a population. The spread of infectious disease raises not only the factors associated with the disease, such as infectious agents, transmission mode, latent period, infectious period, sensitivity and resistance, but also social, cultural, demographic, economic and geographical factors. The model examined here is the simplest prototype of three different equations of epidemiological models, it is important to understand its behavior.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.1 Types of epidemic models

There are two types of epidemiological models stochastic and deterministic

Stochastic

A stochastic model is a tool for estimating probability distributions of potential outcomes by allowing for random variation in one or more inputs over time. "Stochastic" means being or having a random variable, therefore the stochastic models depend on the chance variations in risk of exposure, disease and other illness dynamics.

Deterministic

When dealing with large populations deterministic mathematical models are often used. In a deterministic model, individuals in the population are assigned to different subgroups, each representing a specific stage of the epidemic and the letters such as S, I, and R are often used to represent different stages. The transition rates from one class to another are mathematically expressed as derivatives, hence the model is formulated using differential equations. While building such models, it must be

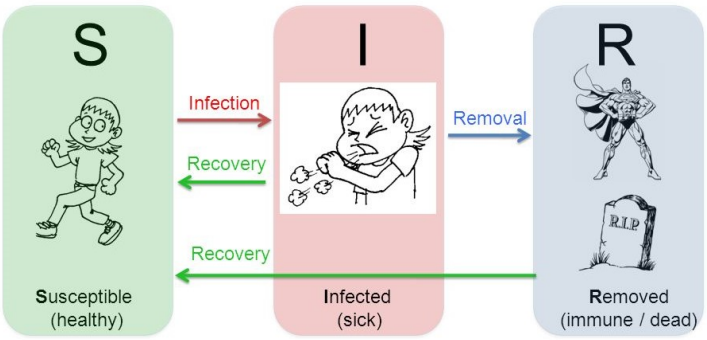


Figure 3.1: Different stages of SIR model

assumed that the population size in a subgroup is differently able with respect to time and that the epidemic process is deterministic. In other words, the changes in population of a subgroups can be calculated using only the history that was used to develop the model. Such a model is described below.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.2 The SIR model

Let S(t), I(t) and R(t) respectively denote the number of susceptibles, infectives and removed (= recovered or dead) individuals. The population is considered to be of fixed size N, where $S(t) + I(t) + R(t) = N$ for all t. The model is then defined by the following set of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta S_t \frac{I_t}{N} \\ \frac{dI}{dt} &= \beta S_t \frac{I_t}{N} - \gamma I_t \\ \frac{dR}{dt} &= \gamma I_t \end{aligned} \tag{3.1}$$

where β is the transmission rate and γ is the recovery rate and with initial state $(S(0), I(0), R(0)) = (s_0, i_0, 0)$. In Figure we show how (S, I, R) varies over time for $\beta = 0.6$ and $\gamma = 0.1$

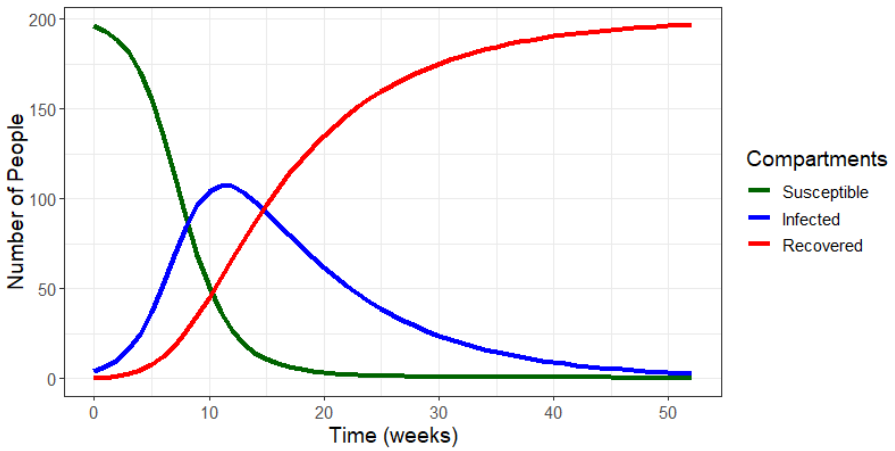


Figure 3.2: The change in the epidemic compartments over time, for $\beta = 0.6$ and $\gamma = 0.1$

Inference for the parameters linking compartments is complicated by the fact that one or several of the model variables may be unobserved (latent) and the presence of a non-linear system of ODEs makes the estimation a difficult task. Along with these parameters, we usually want to obtain estimates of the initial number of susceptible individuals and the basic reproduction number which is discussed below.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Stan has two built-in ODE solvers which provide a user-friendly solution to this problem appearing in epidemic analysis. In particular, the first one which is suitable for non-stiff systems, uses fourth and fifth order Runge-Kutta method being relatively fast while the second one is designed to deal with stiff systems and even though it is slower, it is more robust (Luo & Jiao 2018). In what follows, we provide a setting where the ODE solver role is highlighted in the context of an SIR model and finally we implement HMC.

3.3 Basic reproduction number

The basic reproduction number, denoted by R_0, is a measure of how transferable a disease is. It is the average number of people that a single infectious person will infect over the course of their infection, this quantity determines whether the infection will spread exponentially, die out, or remain constant. The basic reproduction number can be computed as a ratio of known rates over time: if an infectious individual contacts β other people per unit time, if all of those people are assumed to contract the disease, and if the disease has a mean infectious period of $1/\gamma$, then the basic reproduction number is just $R_0 = \beta/\gamma$ where:

- if $R_0 > 1$, then each person on average infects more than one other person so the disease will spread.
- if $R_0 < 1$, then each person infects less than one person on average so the disease will die out.
- if $R_0 = 1$, then each person will infect exactly one other person, so the disease will become endemic, it will move throughout the population but not increase or decrease.

In this model it has a natural interpretation as the expected number of infections generated by one infectious individual in a large, completely susceptible population. It is a key quantity to transmission dynamics as it determines the critical proportion

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

of individuals that need to be vaccinated in order to eliminate the disease (Dietz 1993).

3.4 Single strain SIR models

In the context of a deterministic SIR, in order to estimate the transmission and the recovery rate, we use four different specification. First, we assume that the total number of infected each week Y_t, follows a Binomial distribution:

Model 1

$$Y_t \sim Binomial(N, p_t) \tag{3.2}$$

$$p_t = \int_0^t (\beta i_s s_s - \gamma i_s) ds \tag{3.3}$$

Where s_s is the fraction of susceptible, i_s is the fraction of infected and N is the total population size. In order to perform HMC, we use a log-normal prior for the transmission rate, $\beta \sim lognormal(0, 1)$ and assuming that the mean infectious period is 5 days we employ a Gamma prior for the recovery rate with $\gamma \sim Gamma(0.004, 0.002)$. Finally, we use Jeffrey's prior for the initial proportion of susceptible individuals, $s(0) \sim Beta(0.5, 0.5)$.

Using the same prior distributions, we also estimate a Poisson model, formulated as follows:

Model 2

$$Y_t \sim Poisson(\mu_t) \tag{3.4}$$

$$\mu_t = \int_0^t (\beta \frac{I_s}{N} S_s - \gamma I_s) ds \tag{3.5}$$

Where S_s represents the number of susceptible, I_s the number of infected and N is the total population size. In this case we assume that the mean is equal to the variance.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Otherwise, a different parameterization of the total number of infected each week, follows a Negative Binomial distribution:

Model 3

$$Y_t \sim NegBinomial(\mu_t, \delta) \tag{3.6}$$

$$\mu_t = \int_0^t (\beta \frac{I_s}{N} S_s - \gamma I_s) ds \tag{3.7}$$

Recall that Poisson(μ_t) has variance μ, so $\mu^2/\delta > 0$ is the additional variance of the negative binomial above that of the Poisson with mean μ. So the inverse of parameter δ controls the over-dispersion, scaled by the square of the mean μ^2. We use a Gamma prior for inverse of parameter δ where $1/\delta \sim Gamma(10, 0.1)$, for the initial proportion of susceptible individuals we use Beta prior where $s(0) \sim Beta(2, 2)$ and for the other parameters using the same prior distributions like before.

We also estimate a Beta Binomial model, formulated as follows:

Model 4

$$Y_t \sim BetaBinomial(N, p_t, \delta) \tag{3.8}$$

$$p_t = \int_0^t (\beta i_s s_s - \gamma i_s) ds \tag{3.9}$$

In this case we assume a different parameterization of the Beta Binomial with α and β were $\alpha = \frac{\delta}{1-p}$ and $\beta = \frac{\delta}{p}$. The parameter δ controls the over-dispersion, where $\delta \sim Gamma(10, 0.01)$, for the initial proportion of susceptible individuals we use Beta prior where $s(0) \sim Beta(2, 2)$ also in this case for the other parameters we will using the same prior distributions like before. Below we will see how the deterministic approach it will gives us an insight into the dynamics of the disease for each corresponding instance of data sets.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chapter 4

Applications

4.1 Simulated data

First, we will generate data by integrating an SIR model with known β and γ parameters. This will simulate an epidemic window, which represents a time period over which data can be collected. For instance, we can go to the field and measure the proportion of hosts infected at various time points to capture the rise and fall of infection. This data we will used to fit the model.

More detail the parameters represent a particularly virulent pathogen, where by the end of 52 weeks, most of the population has been infected and has recovered or died. This gives us the "true" epidemic dynamics of the system, from which we can simulate data that an ecologist might collect. For instance, an ecologist could go to the field and sample a given number of individuals from the population and figure out how many are infected. We assign transmission and pathogen-induced death rates, $\beta = 0.6$ and $\gamma = 0.10$ respectively. Sampled could be repeated a number of times throughout the epidemic period, ideally this would be daily, but we all know that is difficult. For this reason we choose the sampling every week, were taken 52 weeks sampled throughout the epidemic and afterwards we choose 200 individuals sampled per week sampled.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

So we have the "true" fraction of the population that is infected in each sampling week, the number of weeks sampled, and the number of individuals sampled per week. To simulate the data we can use Binomial Distribution with parameters n, p where n the number of host individuals sampled per week, and p the fraction of the population that is infected on each of the sampled weeks.

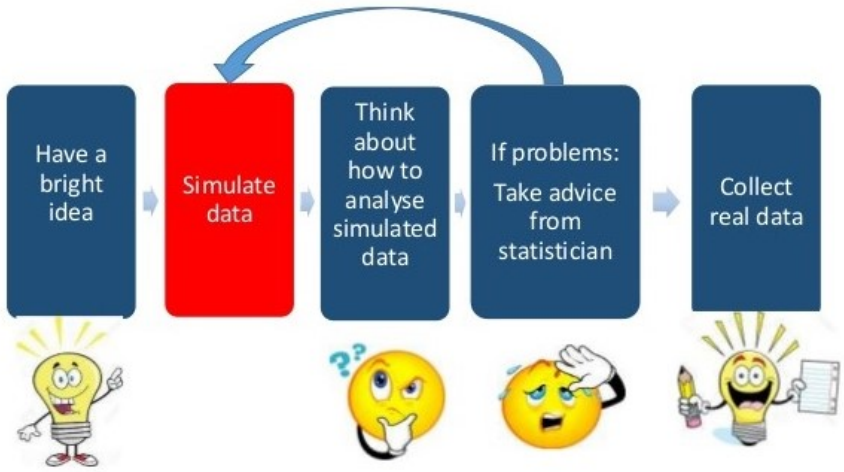


Figure 4.1: A better way to do experiments

Now that we have some simulation data, let's fit the mechanistic models using the Stan MCMC sampling software. In the models, we'll estimate the parameters of interest, as well as the initial conditions which were likely unknown in the field.

All four models are fitted using Stan's NUTS algorithm using 5 chains, each with 100500 iterations of which the first 2500 are warm-up to automatically tune the sampler, and then a sample is saved every forty-nine samples, leading to a total of 10000 posterior samples. We examine the convergence of the parameters by inspecting the trace plots of all chains indicating that there is no lack of convergence for all models and by checking the $\hat{R}$ convergence statistic reported by Stan.

Therefore, if the chains have not yet converged to a common distribution the $\hat{R}$ statistic will be greater than one. However, if it is equal to 1, it does not necessarily indicate convergence. As all convergence diagnostics, $\hat{R}$ can only detect failure to convergence but it cannot guarantee convergence. In our example, all models show good mixing according to the effective sample size, $\hat{R}$ and the trace plots. All models

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

were sensitive to initial values so we initialize our parameters using values drawn uniformly from the credible intervals we obtain from NUTS.

For the deterministic setting, posterior medians and 95% credible intervals of the parameters are summarized in Table 4.1 and 4.2. In addition to the parameters characterizing the transmission dynamics of the disease, we also report posterior estimates for over-dispersion parameter δ, of the Negative Binomial and Beta Binomial models.

Hamiltonian Monte Carlo						
	Binomial			Poisson		
	mean	95% CI	ESS	mean	95% CI	ESS
β	0.60	0.55-0.64	10378	0.61	0.55-0.67	9539
γ	0.09	0.09-0.10	9727	0.09	0.09-0.10	9731
s(0)	0.97	0.97-0.98	10343	0.97	0.97-0.98	9737
R_0	6.05	5.56-6.57	10223	6.23	5.56-6.69	9478

Table 4.1: HMC results for Binomial and Poisson models with simulated data

Hamiltonian Monte Carlo						
	Negative Binomial			Beta Binomial		
	mean	95% CI	ESS	mean	95% CI	ESS
β	0.62	0.55-0.71	9268	0.62	0.56-0.66	9867
γ	0.09	0.09-0.11	9879	0.09	0.09-0.10	9987
δ	175.92	49.61-433.25	10181	106.49	27.46-303.17	10224
s(0)	0.98	0.97-0.98	9610	0.97	0.97-0.98	9813
R_0	6.40	5.54-7.43	9473	6.16	5.55-6.85	9866

Table 4.2: HMC results for Neg.Binomial and Beta Binomial models with simulated data

We observe that the estimates for parameters β and γ are very close to the values we set to simulate the data, in general the posterior estimates for R_0 are in line with the estimated R_0. In Binomial model results in narrower credible intervals for β compared to the other models, suggesting the greatest uncertainty of estimation especially for models with over-dispersion parameter. To selection the better-fitting

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

model of different methodologies, we calculated information criteria for models like DIC and WAIC, the models to be preferred are those with the smaller DIC and WAIC. As expected, the smaller DIC and WAIC has the Binomial model, since we know that data generated from Binomial distribution.

	Binomial	Poisson	Neg Binomial	Beta Binomial
DIC	309.97	313.02	320.27	313.60
WAIC	310.54	313.01	317.18	312.08

Table 4.3: Information criteria for models with simulated data

Summing up, the results of the deterministic setting bring us to the preliminary conclusion that if we are interested in real-time inference the method are feasible and efficient. As seen from Figure 4.2, has a reasonable fit to the data but underestimates the overall uncertainty thus resulting in overly precise estimates which fail to capture the simulate data appropriately.

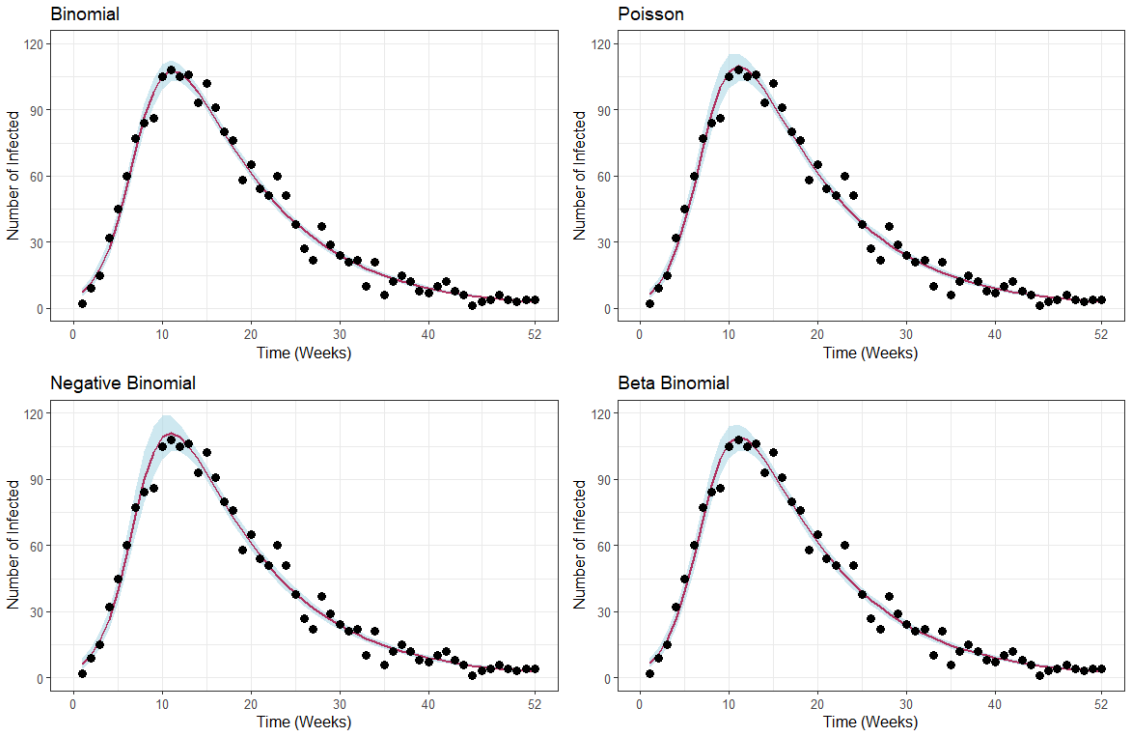


Figure 4.2: Models fit to simulated data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

In terms of computational time that every model needed is presented in the table below 4.4, HMC is extremely efficient in all four models. To summarize, we observe that the estimates of the models for R_0 are very close to the value of the simulated data, which confirms their good performance.

	Binomial	Poisson	Neg Binomial	Beta Binomial
HMC	10.04	12.57	17.13	13.95

Table 4.4: Execution time (minutes) for simulated data

We also note that the closest estimation and the narrower confidence interval are given by the binomial model, which we expected because the data were generated by a binomial distribution. This is also confirmed by the values of information criteria DIC and WAIC they get the lowest value for the binomial model.

In the following diagrams Figure 4.3, 4.4, 4.5, 4.6 we focuses on plotting parameter estimates from MCMC diagnostic plots. The "bayesplot" package provides various plotting functions for visualizing Markov Chain Monte Carlo (MCMC) draws from the posterior distribution of the parameters of a Bayesian model. Trace plots are time series plots of Markov chains in Figure 4.7, 4.8, 4.9 we show the trace plots confirming the convergence of the series. With all these diagrams we consider the convergence of parameters and series.

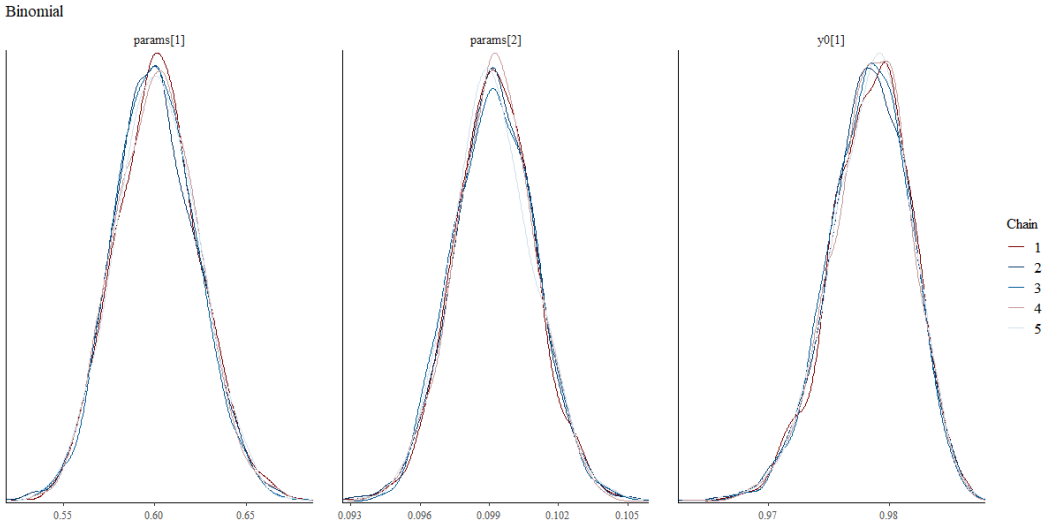


Figure 4.3: Markov Chains for Binomial model with simulated data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

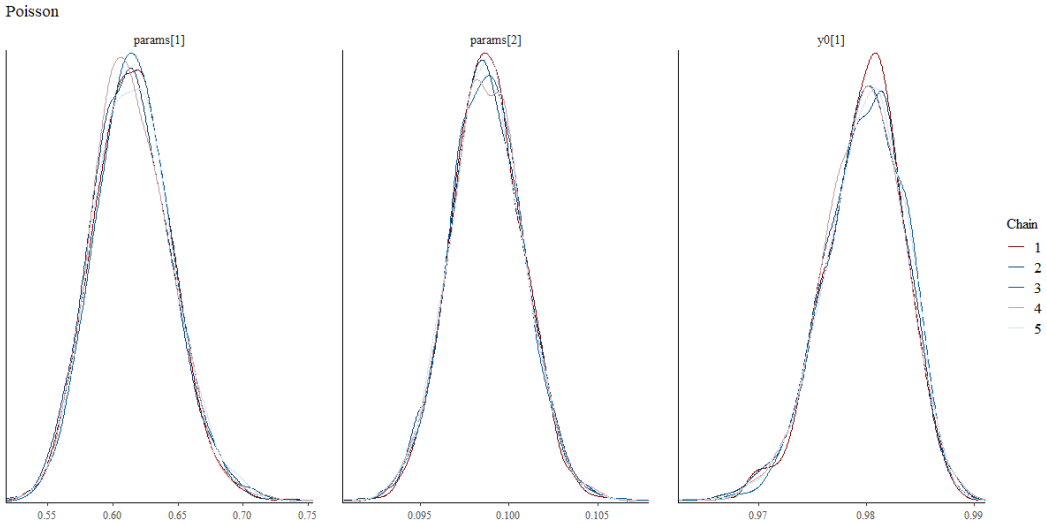


Figure 4.4: Markov Chains for Poisson model with simulated data

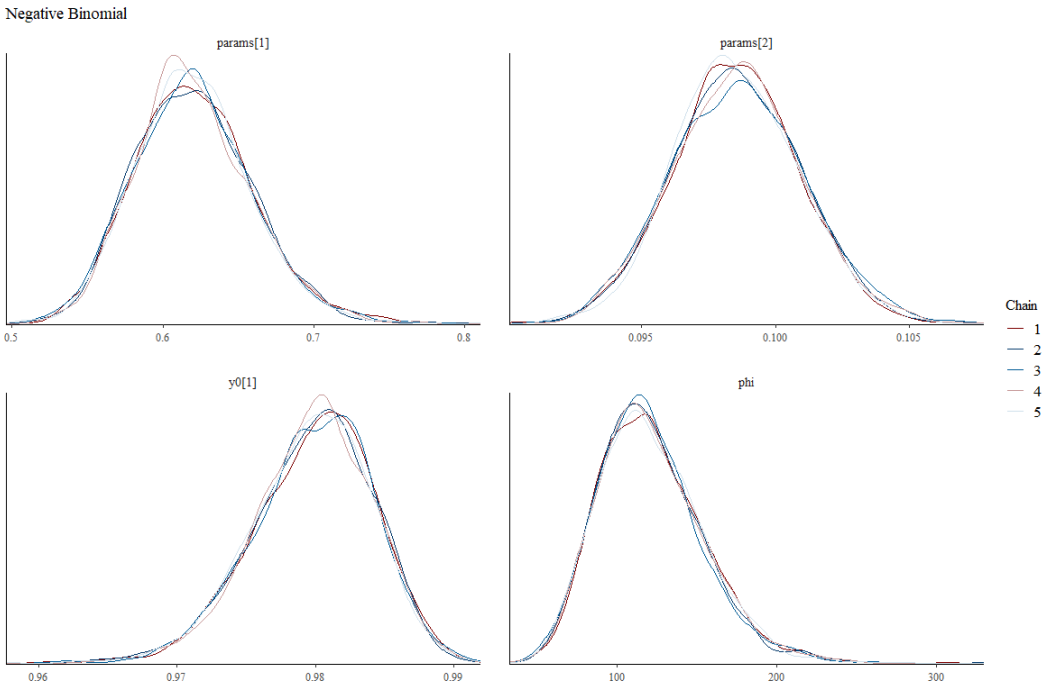


Figure 4.5: Markov Chains for Neg. Binomial model with simulated data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

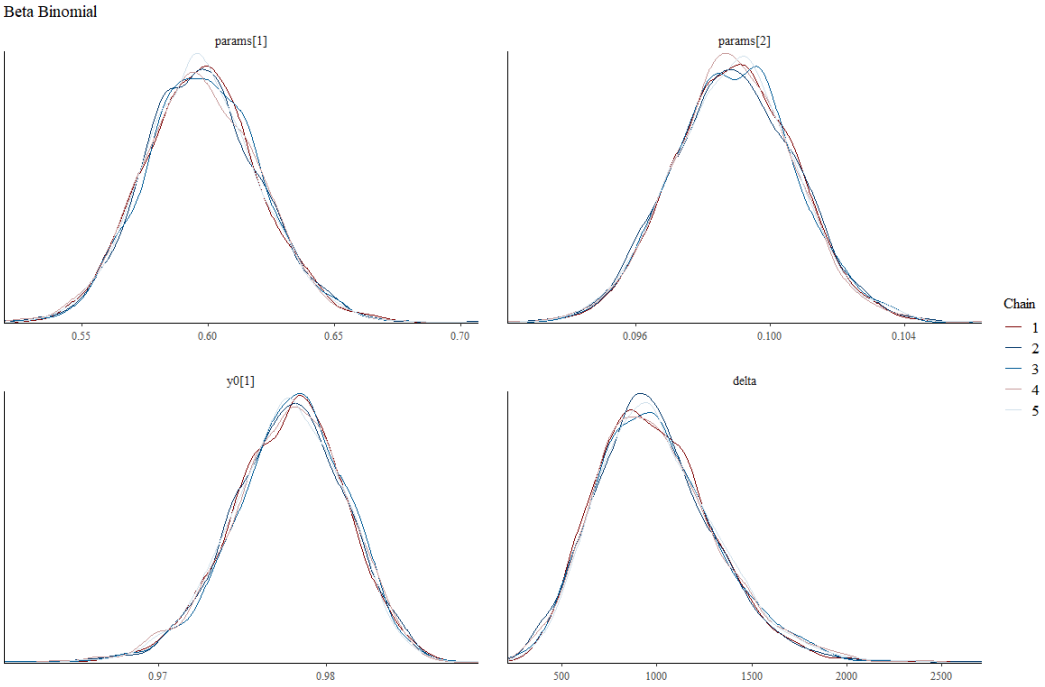


Figure 4.6: Markov Chains for Beta Binomial model with simulated data

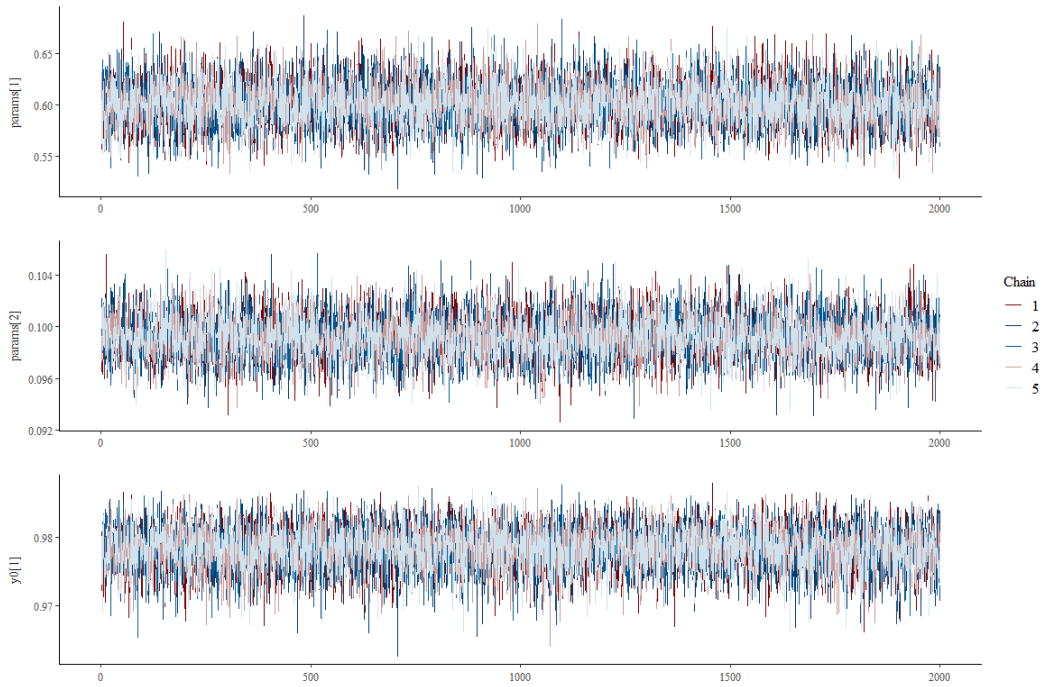


Figure 4.7: Trace plots for Binomial model with simulated data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

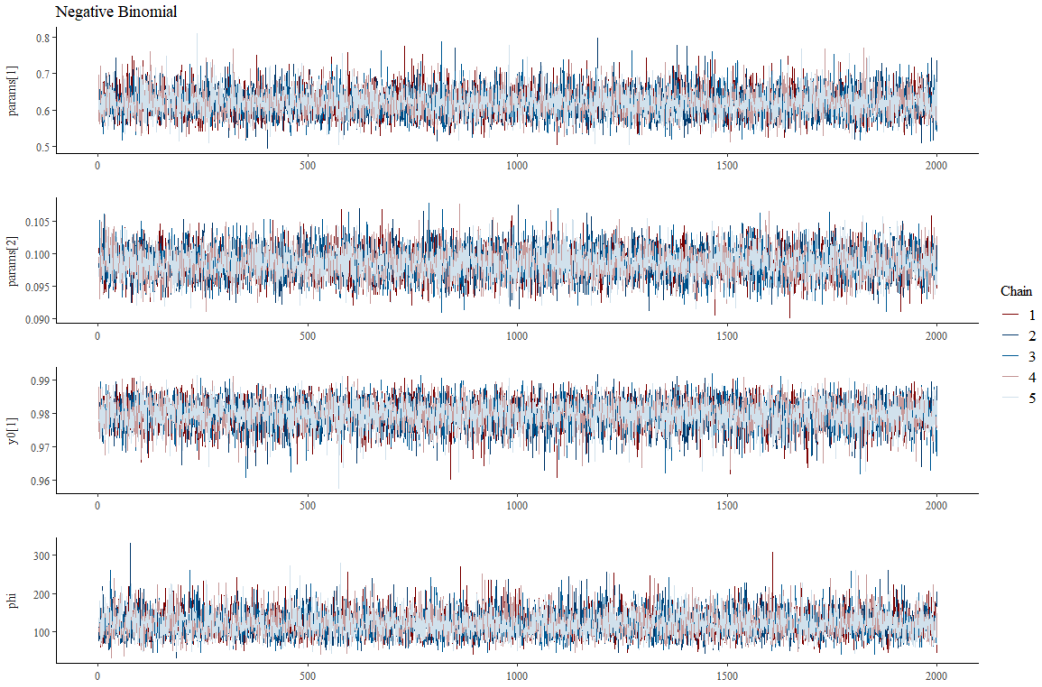


Figure 4.8: Trace plots for Neg. Binomial model with simulated data

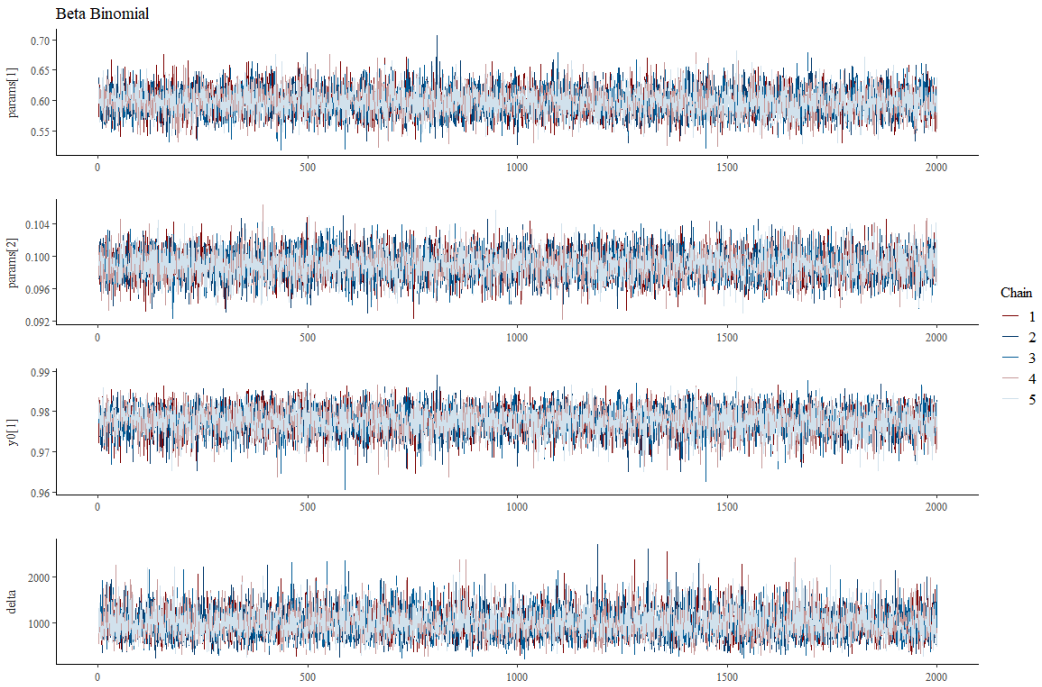


Figure 4.9: Trace plots for Beta Binomial model with simulated data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

4.2 Real data

In 1978, there was a report to the British Medical Journal for an influenza outbreak in a boarding school in the north of England. There were 763 male students which were mostly full boarders and 512 of them became ill. The outbreak lasted from the 22nd of January to the 4th of February and it is reported that one infected boy started the epidemic and then it spread rapidly in the relatively closed community of the boarding school.



Figure 4.10: Boarding school

We use the data from Chapter 9 of de Vries (Gerda) & Hillen [2006] which are freely available in the R package outbreaks, maintained as part of the R Epidemics Consortium (RECON; http://www.repidemicsconsortium.org). Data consist of the number of bedridden students rather than the number of infected students, so we assume that the number of students who are confined to bed each day are those who are newly infected and all these who are still in bed after being infected on a prior day, consisting all together the total number of infected students each day.

All four models are fitted using Stan's NUTS algorithm and in this case using 5 chains, each with 100500 iterations of which the first 500 are warm-up to automatically tune the sampler, and then a sample is saved every forty-nine samples, leading to a

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

total of 10000 posterior samples. We examine the convergence of the parameters by inspecting the trace plots of all chains indicating that there is no lack of convergence for all models and by checking the $\hat{R}$ convergence statistic reported by Stan.

For the deterministic setting, posterior medians and 95% credible intervals of the parameters are summarized in Table 4.5 and 4.6. In addition to the parameters characterizing the transmission dynamics of the disease, we also report posterior estimates for over-dispersion parameter δ, of the Negative Binomial and Beta Binomial models.

Hamiltonian Monte Carlo						
	Binomial			Poisson		
	mean	95% CI	ESS	mean	95% CI	ESS
β	1.87	1.78-1.98	9988	1.88	1.78-1.99	9427
γ	0.47	0.45-0.49	10032	0.48	0.45-0.50	9874
s(0)	0.99	0.99-0.99	10099	0.99	0.99-0.99	9375
R_0	3.93	3.71-4.17	10231	3.92	3.67-4.21	9956

Table 4.5: HMC results for Binomial and Poisson models with real data

Hamiltonian Monte Carlo						
	Negative Binomial			Beta Binomial		
	mean	95% CI	ESS	mean	95% CI	ESS
β	1.85	1.70-2.00	9422	1.86	1.76-1.96	9897
γ	0.50	0.46-0.54	15308	0.47	0.45-0.50	9874
δ	69.16	30.16-127.81	16081	836.84	323.67-1533.66	10018
s(0)	0.99	0.99-0.99	9361	0.99	0.99-0.99	10113
R_0	3.67	3.24-4.13	9983	3.89	3.65-4.15	9964

Table 4.6: HMC results for Neg.Binomial and Beta Binomial models with real data

We observe that the estimates for parameters β and γ are very close for all models. Especially we notice the estimates for the Binomial and Poisson models is narrower credible intervals for β and R_0 compared to Negative Binomial and Beta Binomial models (models with over-dispersion parameter), suggesting that the first two models may be underestimating the posterior uncertainty. In general, the posterior estimates

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

for R_0 are in line with the estimated R_0 obtained by Wearing (Wearing et al. 2005). To selection the better-fitting model of different methodologies, we calculated information criteria for models like DIC and WAIC, the models to be preferred are those with the the smallest value DIC and WAIC.

	Binomial	Poisson	Neg Binomial	Beta Binomial
DIC	145.52	141.16	130.46	143.68
WAIC	152.93	146.61	131.26	149.07

Table 4.7: Information criteria for models with real data

As seen from Figure 4.5, has a reasonable fit to the real data but underestimates the overall uncertainty thus resulting in overly precise estimates which fail to capture the simulate data appropriately, as has been observed in the simulated data.

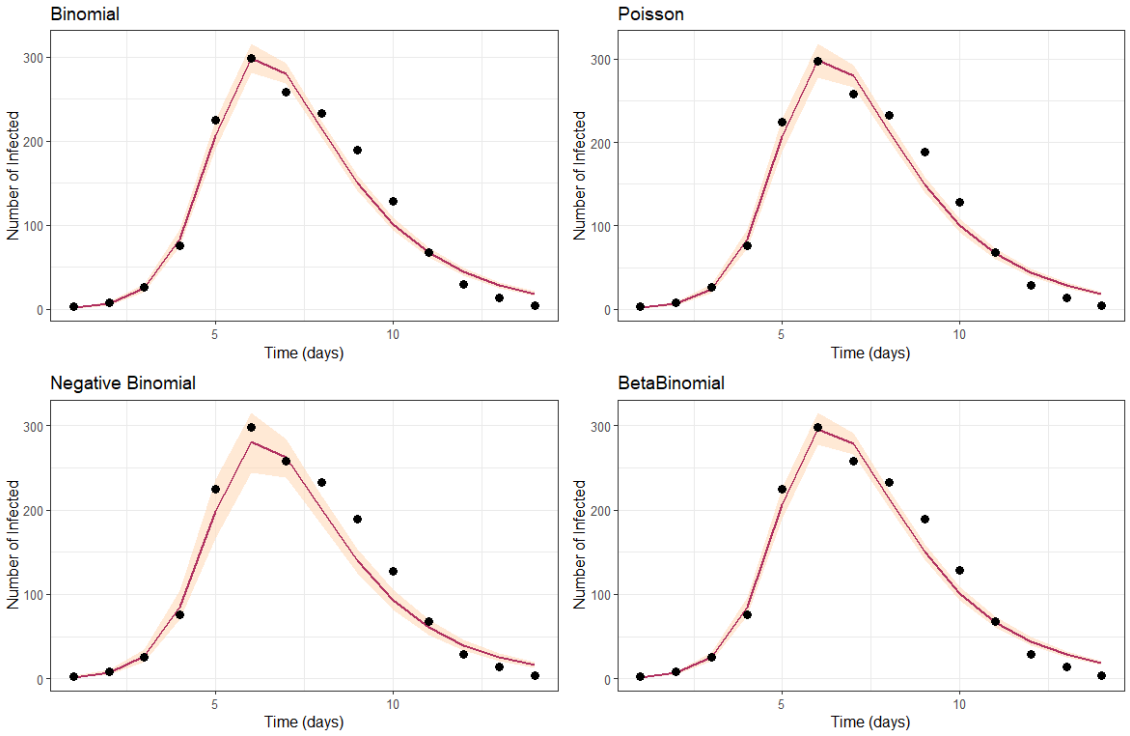


Figure 4.11: Models fit to real data

In terms of computational time that every model needed is presented in the table below 4.8, HMC is extremely efficient in all four models. To summarize, we observe

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

that the estimates of the models for R_0 are in line with the estimated R_0 obtained by Wearing (Wearing et al. 2005). We also note the values of information criteria DIC and WAIC they get the lowest value for the Negative Binomial model.

	Binomial	Poisson	Neg Binomial	Beta Binomial
HMC	16.88	15.49	18.55	18.90

Table 4.8: Execution time (minutes) for real data

In the following diagrams Figure 4.12, 4.13, 4.14, 4.15 we focuses on plotting parameter estimates from MCMC diagnostic plots. The "bayesplot" package provides various plotting functions for visualizing Markov Chain Monte Carlo (MCMC) draws from the posterior distribution of the parameters of a Bayesian model. Trace plots are timeseries plots of Markov chains in Figure 4.16, 4.17, 4.18 we show the trace plots confirming the convergence of the series. With all these diagrams we consider the convergence of parameters and series.

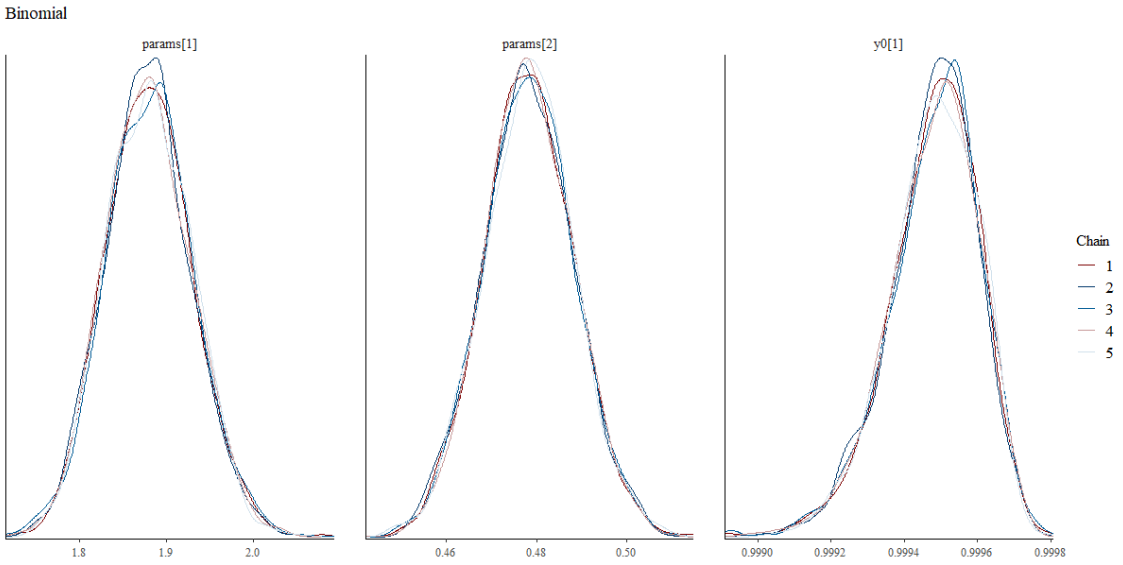


Figure 4.12: Markov Chains for Binomial model with real data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

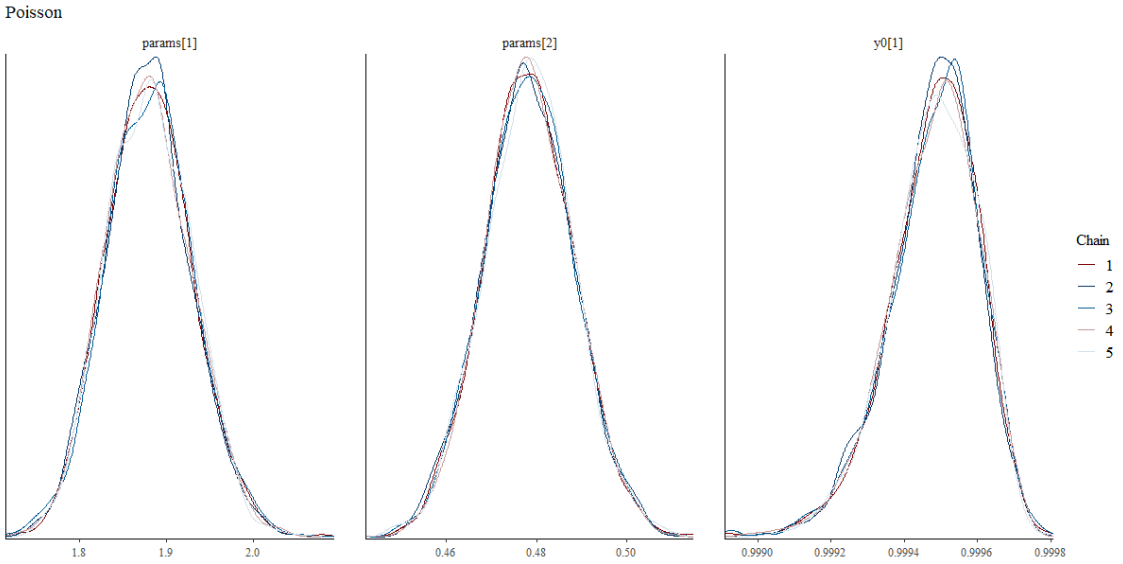


Figure 4.13: Markov Chains for Poisson model with real data

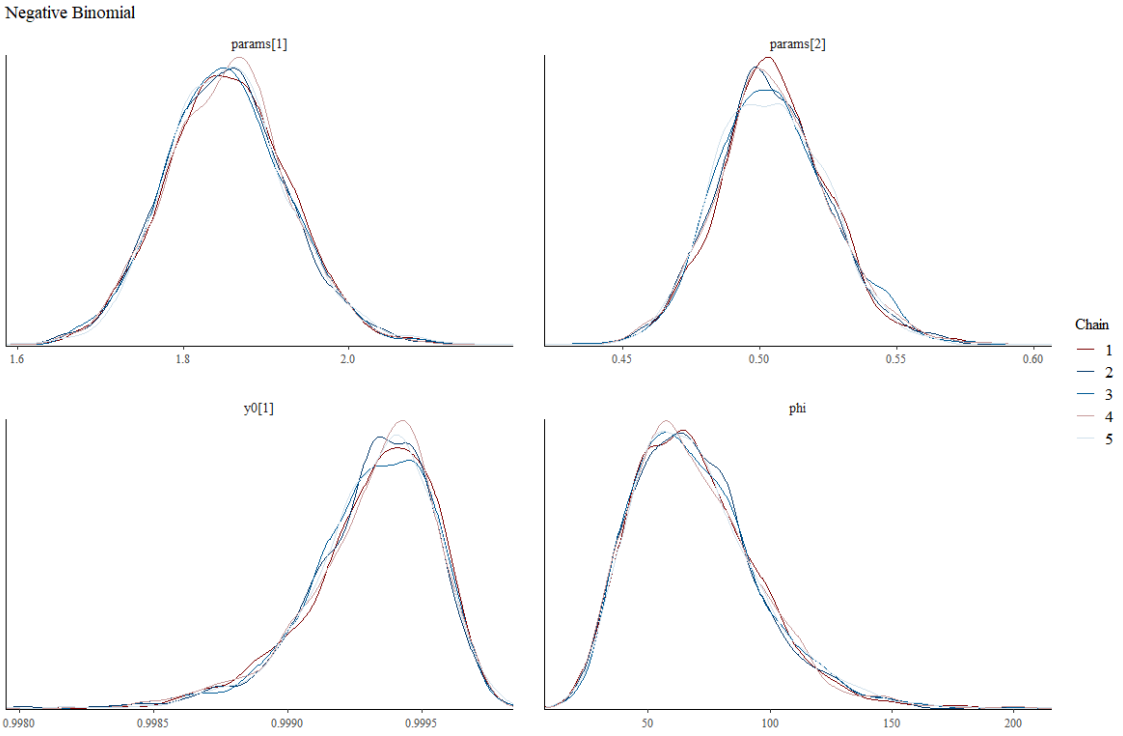


Figure 4.14: Markov Chains for Neg. Binomial model with real data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

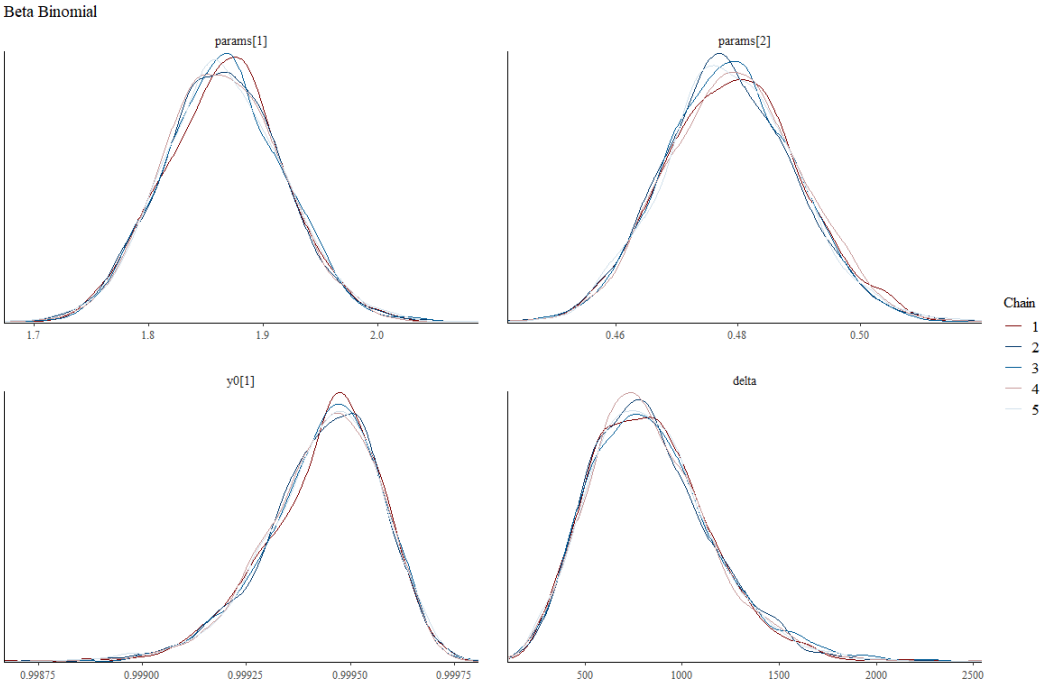


Figure 4.15: Markov Chains for Beta Binomial model with real data

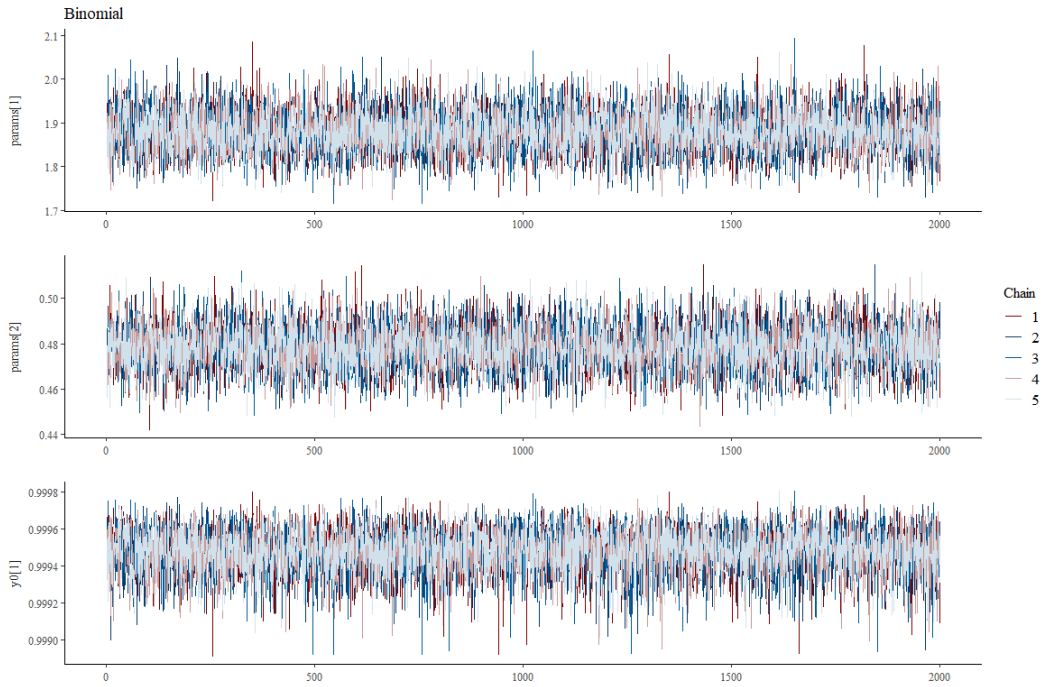


Figure 4.16: Trace plots for Binomial with real data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

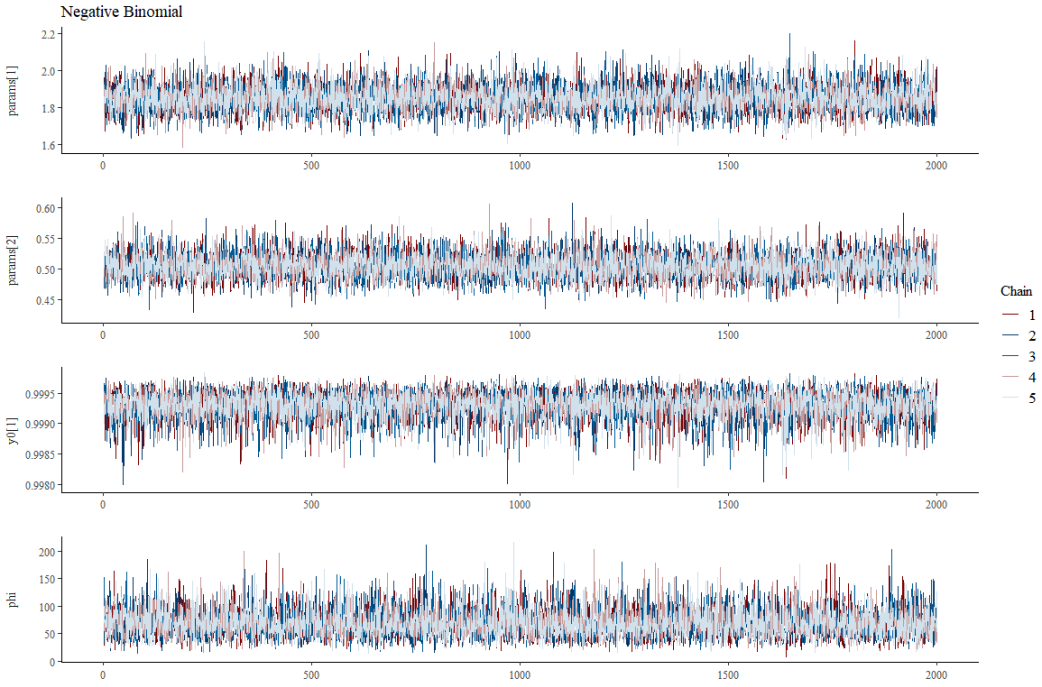


Figure 4.17: Trace plots for Neg. Binomial with real data

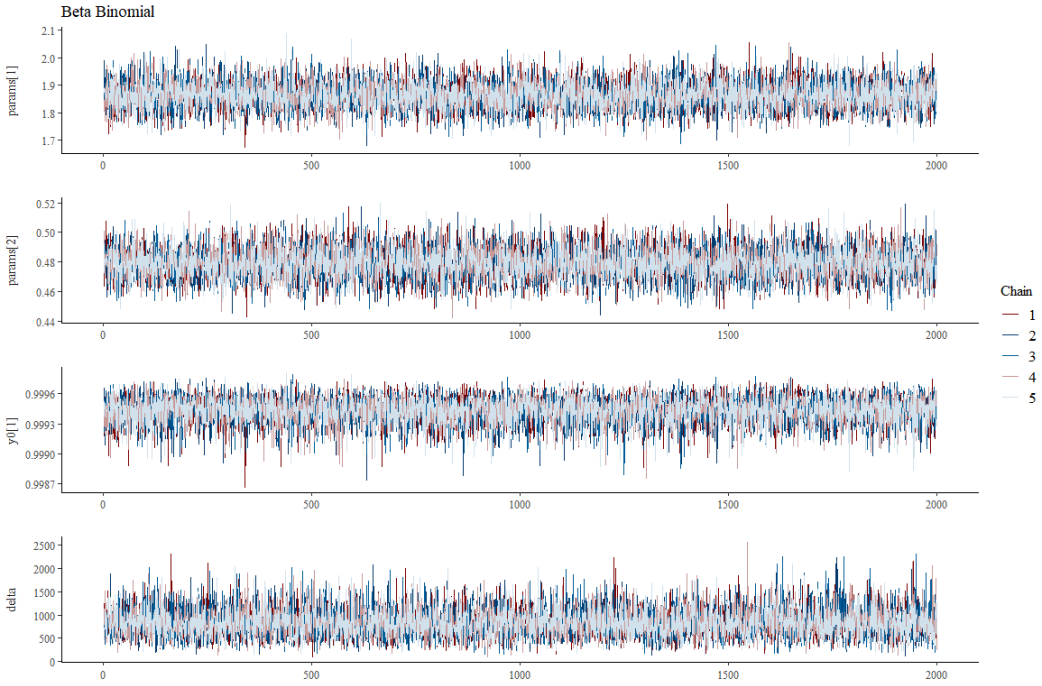


Figure 4.18: Trace plots for Beta Binomial with real data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chapter 5

Discussion

The scope of the present thesis is summarize the basic concepts required to perform HMC using Stan, in the framework of infectious disease modelling. In practice we need a statistical software that will allow the fitting of ODE-based models, Stan is the first general statistical software for this purpose. The idea comes from the fact that in ordinary differential equations, the respective likelihood function may have ridged regions resulting in a failure of standard regularity conditions and classical likelihood or inference based on MCMC. So in these cases, we apply the HMC algorithm that can produce more accurate results and is readily available to epidemiologists in the form of Stan.

The advantages has Stan are many, in the sense that to estimate different models either by changing the distributional assumptions or adding more components, we only need to change a few lines of code Thus, as a generic and flexible software package along with the fact that it may perform inference fast, Stan makes real-time inference feasible. Also, HMC seems to perform better than classical techniques in epidemiological models, where the posterior distributions may be characterised by highly correlated parameter spaces.

We also saw that there are many factors that can affect the performance of the different models we have implemented, the chosen parameterization, priors, starting values and tuning parameters, are only a few of these factors. To the present, HMC

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

in Stan, does not allow for discrete parameters, but they can simply be marginalized out.

Summing up, the deterministic approach it gives us an sight into the dynamics of the disease and behavior of the infection spread for each corresponding instance of data sets. The information that is being extracted from this method for R_0 we believe can be a possible answer on whether an epidemic will die out or not. This possible answer is quite interesting since is can be very useful for applications like the design and assessment of different vaccination strategies.

Finally, the results of the deterministic setting with real and simulated data, bring us to the conclusion that if we are interested in real-time inference the method are feasible and efficient. As seen from the results of models which applied has a reasonable fit to the data either simulated or real, but underestimates the overall uncertainty thus resulting in overly precise estimates which fail to capture the data appropriately in both cases. Our believe that a good idea for future research is to apply stochastic epidemic models to improve overly precise estimates.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Bibliography

Berger, J. O., Bernardo, J. M., Sun, D., et al. (2009). The formal definition of reference priors. *The Annals of Statistics*, *37*(2), 905–938.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*.

Betancourt, M., Byrne, S., & Girolami, M. (2014). Optimizing the integrator step size for hamiltonian monte carlo. *arXiv preprint arXiv:1411.6669*.

Betancourt, M., & Girolami, M. (2015). Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, *79*, 30.

Betancourt, M., & Stein, L. C. (2011). The geometry of hamiltonian monte carlo. *arXiv preprint arXiv:1112.4118*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Chatzilena, A., van Leeuwen, E., Ratmann, O., Baguelin, M., & Demiris, N. (2019). Contemporary statistical inference for infectious disease models using stan. *arXiv preprint arXiv:1903.00423*.

Daley, D., & Gani, J. (2005). *Epidemic modeling: An introduction. sl.* Cambridge University Press.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: writing statistical algorithms for

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

general model structures with nimble. *Journal of Computational and Graphical Statistics*, *26*(2), 403–413.

de Vries (Gerda), & Hillen, T. (2006). *A course in mathematical biology: quantitative modeling with mathematical and computational methods*. SIAM.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, *2*(1), 23–41.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, *195*(2), 216–222.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 3–32.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. (2004). Bayesian data analysis 2nd edn chapman & hall. *CRC, Boca Raton FL*.

Hamer, W. H. (1906). *Epidemic disease in england: the evidence of variability and of persistency of type*. Bedford Press.

Hoffman, M. D., & Gelman, A. (2011). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. arxiv preprint arxiv: 1111.4246.

Leimkuhler, B., & Reich, S. (2004). *Simulating hamiltonian dynamics* (Vol. 14). Cambridge university press.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, *10*(4), 325–337.

Luo, Y., & Jiao, H. (2018). Using the stan program for bayesian item response theory. *Educational and psychological measurement*, *78*(3), 384–408.

McKinley, T. J., Ross, J. V., Deardon, R., & Cook, A. R. (2014). Simulation-based bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, *71*, 434–447.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

O'Neill, P. D., & Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(1), 121–129.

Plummer, M. (2017). Jags version 4.3. 0 user manual [computer software manual]. *Retrieved from sourceforge. net/projects/mcmc-jags/files/Manuals/4. x*, *2*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, *64*(4), 583–639.

Wearing, H. J., Rohani, P., & Keeling, M. J. (2005). Appropriate models for the management of infectious diseases. *PLoS medicine*, *2*(7), e174.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ