POLITECNICO DI TORINO

Master's degree course
in Mathematical Engineering
Statistics and optimization on data and networks

Master's degree Thesis

Deterministic and stochastic SIR for the analysis of COVID-19 pandemic in Piemonte.

Supervisor
Prof. Enrico Bibbona

Candidate
Francesca Collini

Academic Year 2021-2022

Summary

The recent COVID-19 pandemic has shown the importance of modeling the spread of infectious diseases. Indeed, these phenomena are able to influence not only people's everyday lives, but they can cause even greater and permanent consequences.
Epidemic models can be brought into play to get a deeper understanding of the disease transmission. At the same time, they may be useful to generate simulated scenarios caused by the adoption of different control decisions. Having reliable prediction may help to guaranty an efficient health care system and to setup the best measures for the limitation of contagions. The first step that allows to make these models predictive is the *estimation* of their parameters. For this purpose, computational Bayesian techniques can be adopted.
Two different approaches can be followed in order to define an epidemic model. A *deterministic* model is useful to get an initial overview of the phenomenon, while a *stochastic* model sounds like the most obvious choice. Indeed, with a deterministic model, the output will always be the same, once the parameters and the initial conditions are defined. On the other hand, for a stochastic model with the same inputs, a certain output will be reached with some probability. Intuitively, it seems quite natural describe the evolution of an epidemic, providing the probabilities of a contagion or a recovery, rather than affirming these events take place for sure. Nevertheless this is not so obvious and both models should be considered useful for the analysis. A particular kind of epidemic model is the SIR. The main hypothesis of this approach is that the population can be divided into: Susceptible, Infectious and Removed with respect to a specific disease. The model is used in order to understand how a specific individual may move from one group to another.
An empirical analysis of the data from the first COVID-19 wave in Piemonte (an Italian region) is the goal of this thesis. All data considered are extracted from the official repository of the Italian Protezione Civile. The analysis is based both on a deterministic and a stochastic SIR model. In each situation, the parameters involved into the models are estimated via a particular kind of MCMC (Markov Chain Monte Carlo) algorithm.

Sommario

La recente pandemia di COVID-19 ha dimostrato quanto sia importante conoscere e modellare, nel modo corretto, le malattie infettive. Queste hanno iniziato a ricevere, con il tempo, un'attenzione sempre maggiore. Infatti, sono in grado di influenzare non solo le abitudini di vita di milioni di persone nel mondo, ma possono causare anche impatti più a lungo termine come ad esempio nella vita economica o climatica di uno Stato.
I modelli epidemici sono lo strumento principale da utilizzare innanzitutto per avere una descrizione migliore del processo di diffusione della malattia, ma anche per simulare dei possibili scenari futuri e fare previsioni. In questo contesto infatti, sarebbe importante poter capire l'impatto di una specifica misura preventiva sulla limitazione del contagio. Il primo passo per ottenere un modello predittivo è quello di *stimare* il valore dei parametri che entrano in gioco. Per farlo, è possibile utilizzare la statistica Bayesiana.
Per applicare i modelli epidemici è possibile percorrere due strade alternative. La prima è quella di un modello epidemico *deterministico*, dove una volta che i parametri di ingresso vengono scelti, il risultato rimane fissato. L'altro tipo di modello è quello *stocastico*, dove invece, anche se vengono fissati i parametri e le condizioni iniziali, il risultato finale rimane comunque legato ad una probabilità. Intuitivamente, si potrebbe pensare che il secondo approccio sia il migliore perché definire la probabilità che avvenga un contagio o una guarigione sembrerebbe il modo più naturale per descrivere questo tipo di dinamiche. Tuttavia, il confronto non è così ovvio ed entrambe le alternative sono degli ottimi strumenti per ottenere una descrizione del sistema. Un particolare tipo di modello epidemico è il SIR, secondo cui la popolazione totale può essere divisa in soggetti Suscettibili, Infetti e Rimossi. Successivamente si vanno a studiare le dinamiche che permettono ad un individuo di spostarsi da uno stato all'altro.
L'argomento principale di questo lavoro di tesi consiste in un'analisi empirica dei dati, raccolti dalla Protezione Civile Italiana, relativi alla prima ondata di COVID-19 in Piemonte. L'analisi si è basata sia su un SIR Deterministico che su uno Stocastico. In entrambi i casi, per la stima dei diversi parametri, si è scelto di utilizzare un algoritmo di simulazione MCMC.

Contents

List of Figures

Ho fatto passi avanti, non dico giganti,
non so neanche quanti, non li conto mai.
Poi mi ha detto un saggio:
spesso conta il viaggio più di dove vai.

[Willie Peyote, Aglio e olio]

Chapter 1

Introduction

Throughout all human history, infectious diseases have always been serious threats. They affect not only the global health, but also other aspects of human life such as economy and culture. In fact based on past events, we learn how, when there is the diffusion of a pandemic, the lifestyle of people all over the world is radically modified. In addiction, there may be also some economic, climatic or cultural impacts connected to this phenomenon. Some of the more extreme changes indeed occurred during the most severe pandemics such as the Black Death in 1300, the cholera during the Nineteenth Century or the Spanish flu which started to spread in 1918.
Since the main role played by infectious disease throughout past history and especially in the last period, it is really important to understand how the diffusion of a particular infection works. Indeed, we can expect this to continue to be a very important topic also in the next future. To this aim, we are interested in studying mathematical models which are able to represent (and simplify) the spread of a disease in the real world and also some statistical methods to analyze it. In other words, we would like work with models which are simple enough to be analyzed but, at the same time, sufficiently complex to accurately represent the reality.
In particular, in this thesis we focus on *deterministic* and *stochastic SIR epidemic models*. This is why, we propose a very detailed description of this model and we use it in the simulations. We collect the main theoretical results reached in this field. These are applied to recent COVID-19 data, published by Protezione Civile in Italy.
The final goal of this thesis is the estimation of parameters involved in different models with different techniques. In this way, we will be able to better understand how the process of the spread of the disease works. The estimation of these parameters can be really useful not only in order to model events happened in the past, but also to make some reliable forecasting for future days. Moreover, with the previous results, we will able to understand how a particular prevention measure can influence the contagion in order to prevent the outbreak of a pandemic or even what is the number of infective in the next days for the purpose of providing them the best assistance. Prediction, however, is a really important and complex topic which is not discussed in this thesis.

Part I

Background material

Chapter 2

Mathematical preliminaries

2.1 Statistical methods

Mathematical models used to represent the spread of a disease in a community can be used for two different and equally important purposes. First of all, they are really useful to get a deeper understanding of the infections, to get more information about their characteristics and properties. However, another important goal of modeling the spread of disease is to provide a guidance of health authorities. This is the reason why, we are also interested in drawing inferences about model parameters. In this way, for example we can realize how a particular preventive measure influences the future spread or what is the best action in order to avoid future epidemics.
In this particular situation, available data are based on *partial observations* because unfortunately, in real life, very detailed information aren't available. In fact hardly ever during an epidemic, all the infectious periods are known: sometimes we just know at what time an infected shows symptoms. In addiction, in particular for large communities, there may be several problems for data collection and moreover when we register a new infected, this doesn't correspond to the date of infection, but only to the date when we have the test results.
Anyway, statistical models provide a method to help you understand why and how infections spread and how they might be prevented or restricted. For instance, when a new infectious disease emerges or there is an outbreak of a known infectious disease, epidemiologists collect and analyze data to provide indications for halting further diffusion. In this section, we briefly introduce some fundamental concepts about Bayesian inference and statistical analysis using Markov Chain Monte Carlo algorithms. They find several applications in very different fields and in this particular context, they are very helpful tools to deal with epidemic models.

2.1.1 Bayesian inference

Before introducing the specific MCMC algorithm applied in this thesis, we give a brief introduction of some essential statistical tools. The main reference in this part is [1].
Classical frequentist inference is based on events frequencies, where model parameters

and hypotheses are fixed. In contrast to this theory, the Bayesian definition of probability also takes into account our knowledge of events and here uncertainty is quantified using probability distribution. Through Bayesian statistics, we understand how to combine our beliefs and some observed data. In fact, supposing we consider a specific hypothesis, with this technique, we are able to modify its probability, as more evidence (observations) becomes available. In this setting, a leading role is played by **Bayes' theorem**, indeed we want to learn some probability distribution for parameter(s) of a model, further denoted by θ, when both prior beliefs (about those parameters) and some data are given. The two building blocks in the theorem are the prior distribution and the likelihood distribution. The first one is denoted with $\pi(\theta)$ and it represents the *subjective* beliefs about θ. For example, it can be based on the results of previous studies, on the experience of an expert in a specific subject, or it can simply be a nearly flat non-informative distribution, when we have no expectation about the behavior of the parameter(s). On the other hand, the likelihood, $\mathscr{L}(X|\theta)$, indicates the compatibility of the evidence with the given hypothesis. Thus, it is a function of the hypothesis and it represents the distribution of the observed data, conditional on a particular realization of θ. On the contrary, the posterior distribution $\pi(\theta|X)$, which is the final goal in Bayesian inference, is the distribution of the parameter(s) after taking into account the observed data. It is given by, following Bayes' rule:

$$\pi(\theta|X) = \frac{\mathscr{L}(X|\theta)\pi(\theta)}{\int \mathscr{L}(X|y)\pi(y)dy}.$$

In general, it could be really difficult to evaluate the integral at the denominator (the evidence), especially when we are in an high dimensional space. This term ensures that the posterior distribution is indeed a valid probability density and this is why, it is called "normalizing" constant. However, if we use the Metropolis-Hastings algorithm or other Markov Chain Monte Carlo methods in order to sample from the target distribution, this problem will be overcome, because of the simplification of normalization factors in a fraction. This is the reason why, very often the theorem is enunciated as follows:

$$\pi(\theta|X) \propto \mathscr{L}(X|\theta)\pi(\theta),$$

or also said by words:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

The posterior distribution is a much richer object than just a point estimator or a confidence interval, so this technique is very popular and it can be used in different applications. A particular one we further analyze is in the epidemic field. In this setting, we want to estimate some parameters involved in the illness diffusion, when we have some available data for the estimation.

2.1.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a statistical tool used to sample from complicated target distributions. In fact, in most real word applications, we consider complex probability distribution in high-dimensional space from which we are unable to sample directly.

MCMC methods are an extension of simple Monte Carlo technique and they find application in very different fields. In particular, MCMC is widely used in inference to sample from an intractable posterior distribution. This method was described for the first time by Nicholas C. Metropolis and his team in 1953. The aim is to sample from a distribution $\pi(x), x \in E \subseteq \mathbb{R}^n$ which is sufficiently complex, so that Monte Carlo method cannot be used directly. A possible strategy is to build an *aperiodic* and *irreducible* Markov chain, whose state space is E and its stationary distribution is given by $\pi(x)$. In this way, once the stationarity is reached by the chain, it will be possible to extract a sample from the target distribution. The idea behind is that we build a sequence of correlated samples by incremental movements in the region of high probability. Even if the samples aren't independent, under some general conditions, we are able to find an approximation for the expectation we are interested in. There are lots of advantages using MCMC, for example this method can be applied when we deal with very complicated distribution in high-dimensional space and it is fairly easy to implement. Anyway, on the other hand it can be very difficult to assess accuracy and evaluate convergence and it is a slower method with respect to simple Monte Carlo or importance sampling because it requires more samples in order to reach the same level of accuracy.

Markov chain

A Markov process is a stochastic process where the future is conditionally independent of the past, given the present. Indeed this property is referred as *Markov property*. A Markov process with a discrete state space (it will take on values in a finite or countable set) is a Markov chain. This is a stochastic process mainly characterized by the *memory-less property*. This means that the current state of the chain is able to determine the next state, no matter all the previous history of the process. We can distinguish between Continuous and Discrete time Markov Chain. The previous differentiation is based on the time when we decide to observe the process. A Discrete Time Markov Chain (DTMC) is a system associated with a clock. It is able to quantify how much time goes by between events. On the other hand, in a CTMC (Continuous Time Markov Chain) events may occur at any time. On the following, we focus on DTMC and we are going to discuss some important properties. In a more formal way, a Markov chain $\mathbf{X}$, for any choice of $k+2$ times $t_0, t_1, ..., t_{k+1}$, where k is an arbitrary integer parameter, and possible states $x_0, x_1, ..., x_{k+1}$, is a sequence of random variables $X(t_0), X(t_1), ..., X(t_{k+1})$ which satisfies *conditional independence*:

$$\begin{aligned}\mathbb{P}(X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, ..., X(t_0) = x_0) \\ = \mathbb{P}(X(t_{k+1}) = x_{k+1} | X(t_k) = x_k).\end{aligned}$$

In fact, we don't need memory of the past (past information is negligible) or memory of the age of the state (the time spent by the process in the current state is negligible), to predict the future state. In other words, knowledge of the previous state is all that is necessary to determine the probability distribution of the current state.
A Markov chain is called **irreducible** if the state space E is irreducible. A closed subset of state space is a collection of state for which we have a zero probability to get out of

this subset. Once the chain enters in this set, it will be "trapped" in it. A closed set is also irreducible when it doesn't contain another closed set as own subset. A chain is **aperiodic** if all its states are aperiodic. A state is aperiodic if it is possible to come back in it with a number of steps whose largest common denominator is equal to one. A state is known as ergodic if it is aperiodic and positive recurrent (it is expected to return in this state within a finite number of steps). A Markov chain is ergodic if all its states are. For this special kind of Markov chains, we have an unique stationary distribution. A stationary distribution of a Markov chain is a probability distribution that remains unchanged in the Markov chain as time progresses. Typically, we denote it with a column vector π whose entries are the probabilities, one for each state, describing the behavior of the chain in the long term. In other words, in this vector, we collect the probability that we will find the chain in a specific state after an initial transitory period. We can also give the definition of the transition matrix $\mathbf{P}$, as a $n \times n$ matrix, where n is the number of different states in the chain, which in position (i, j) contains the probability for the chain to leave the state j and reach the state i. This is a particular matrix because in every column the sum of all the elements is equal to 1 and it is called "(left) stochastic matrix". The stationarity distribution $\pi(x)$, at this point can be defined also as the only vector which satisfies $\pi = \mathbf{P}\pi$ or in other words the only eigenvector corresponding to eigenvalue 1 for the transition matrix $\mathbf{P}$. Moreover, a Markov chain is said to to satisfy *detailed balance equation* or be reversible with respect to π if, for each state i and j:

$$\pi(i)\mathbb{P}(j|i) = \pi(j)\mathbb{P}(i|j) \tag{2.1}$$

where, by definition, $\mathbb{P}(i|j) = \mathbb{P}(X_{\nu+1} = i|X_\nu = j)$ is the Markov transition probability from state j to state i at any time ν and $\pi(i)$ is the equilibrium probabilities of being in state i. We can prove the following

Theorem 2.1 *A reversible Markov Chain is always stationary.*

Proof. Starting from (2.1), when we works with a DTMC we can sum over j and we obtain:

$$\pi(i) = \sum_j \mathbb{P}(i|j)\pi(j).$$

In fact, $\sum_j \mathbb{P}(j|i) = 1$, because you always have to transit to somewhere.Otherwise, if the chain is continuous, we simply have to integrate over j. Now if we write the vector form of the previous equation, we obtain exactly the definition of the stationarity distribution. If the chain is also ergodic, we can prove that π is unique. This implies that no matter the initial conditions, the chain will always converges to the asymptotic distribution.

□

Monte Carlo

Under Monte Carlo methods, we group a collection of different computational techniques used in order to find the (usually approximated) solution of mathematical problems involving random variables. With a Monte Carlo simulation, we can introduce and analyze

the impact of risk and uncertainty in prediction models. In general, we use a Monte Carlo simulation, when we are unable to use any other alternative. The main idea behind these approaches is to use repeated random sampling to solve very difficult problems. In fact, we assume that an estimation of a random variable can be found by assigning different values and then averaging the results. In conclusion, with a Monte Carlo simulation, several values are randomly sampled from the input probability distributions. After a very large number of repetitions, at the end, the result is again a probability distribution of possible outcomes. In this way, we obtain not only an information about what could happen, but also how likely it will be to happen.
All Monte Carlo simulators have a general structure which follows these steps:

- (pseudo)-random number generation,
- transformation to random-variates,
- domain-specific sample path generation,
- output analysis.

To this aim, it is necessary first of all to analyze the input and characterize the uncertainty. Then, we need to generate uniform random variables and transform them to specific random variables related to the problem. Finally, to get the estimation, we have to simulate the model and examine the output.

2.1.3 Metropolis-Hastings

In some situations, the target distribution is quite complex and we can't apply directly the previous method. For this reason, we may choose to apply ones of MCMC technique. In the following, we decide to consider a DTMC because of simplicity in the demonstrations, but all results can be extended to the continuous case. First of all, it is necessary to build an irreducible and aperiodic Markov chain with state space E and stationary distribution π. Suppose to deal with a Markov chain $X_\nu = X(t_\nu), \nu = 0, 1, 2, ...$, which starts in a suitable point in E. We know that $X_\nu = x$ and we want to determine the next state $X_{\nu+1}$. To this aim, we introduce the so called *proposal distribution* $q(\cdot|x)$ and we suggest a candidate point y, sampling from it. At this point, we decide to accept this proposal with probability $\alpha(x, y)$ where:

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right). \tag{2.2}$$

If we accept the candidate, then $X_{\nu+1} = y$, otherwise $X_{\nu+1} = x$.
Summarizing, in order to start with the algorithm, we have select some arbitrary transition distribution $q(x, y)$, which we are able to sample from, but which doesn't necessary have $\pi(x)$ as stationary density. We also need to initialize the iteration counter $j = 1$ and set the starting point $x^{(0)}$, the maximum number of iterations K and possibly the length B of the burn-in. Then we have to generate a proposed value y sampling from the proposal distribution $q(x^{(j-1)}, y)$ and accept the proposal with probability $\alpha(x^{(j-1)}, y)$, where $x^{(j-1)}$

is the value at the previous step. Operationally, we can generate a uniform random variable $U \sim U(0,1)$ and accept $x^{(j)} = y$ if $u \leq \alpha(x^{(j-1)}, y)$. Lastly, we can update j to $j+1$ and if $j < K$ repeat the previous steps, otherwise the process is terminated.
From this algorithm, we derive that, when a suitable proposal distribution is available for the specific problem, the Metropolis-Hastings algorithm is easy. Moreover, we can observe that normalization factors, if any, disappear in the fraction defining α. This is the reason why, this method can be used even if the distribution of interest is only known up to a scaling constant or when they are tricky to determine. At this point, we have to prove that the distribution of X_ν converges to the target distribution π, when $\nu \to \infty$ and irrespective of the choice of proposal distribution $q(\cdot|x)$. To this aim, it is sufficient to show that the chain is detailed balanced, because of Theorem 2.1.

Proof. From the definition of the algorithm, we know that the transition probabilities are:

$$\mathbb{P}(y|x) = q(y|x)\alpha(x,y), \qquad y \neq x, \tag{2.3}$$

$$\mathbb{P}(x|x) = 1 - \sum q(y|x)\alpha(x,y), \tag{2.4}$$

Now, first we consider the case $y \neq x$ in two separated steps: the case where

- $\dfrac{\pi(y)q(x|y)}{\pi(x)q(y|x)} < 1$: $\Rightarrow \alpha(x,y) = \dfrac{\pi(y)q(x|y)}{\pi(x)q(y|x)}$ and $\alpha(y,x) = 1$.

  Then, we obtain that

  $$\mathbb{P}(x|y) = q(x|y), \tag{2.5}$$

  $$\mathbb{P}(y|x) = \frac{\pi(y)q(x|y)}{\pi(x)}. \tag{2.6}$$

  Finally, if we substitute (2.5) in (2.6), we simply obtain the equation (2.1), proving that the Markov Chain is detailed balanced.

- $\dfrac{\pi(y)q(x|y)}{\pi(x)q(y|x)} > 1$:

  Here we can conduct a mirror-like reasoning, where:

  $$\mathbb{P}(x|y) = \frac{\pi(x)q(y|x)}{\pi(y)},$$

  $$\mathbb{P}(y|x) = q(y|x).$$

To conclude the proof, we consider the case $y = x$ and we would like to conclude that $\mathbb{P}(x|x)$ is symmetric. In this situation, $\alpha(x,x) = 1$, then:

$$\begin{aligned}
\mathbb{P}(x|x) &= \alpha(x,x) - \sum q(y|x) \min\left(1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right) \\
&= \alpha(x,x) - \sum \min\left(q(y|x), \frac{\pi(y)q(x|y)}{\pi(x)}\right) \\
&= \alpha(x,x) - \sum \min\left(q(y|x)\pi(x), \pi(y)q(x|y)\right).
\end{aligned}$$

This prove that the integrand is clearly symmetric in x and y and conclude the proof.
□

Moreover, if we assume that the proposal distribution has some property of weak regularity, it can be proved that that $\mathbf{X}$ converges in distribution, no matter the initial value. The same theory can be extended with some adaptation to the case when the state space is continuous [2].
The performance of Metropolis-Hastings algorithm strongly depends on the choice of the proposal distribution $q(x, y)$. Of course we want that it will be simple enough to be sampled and evaluated, but we'd like to make another request to speed up the convergence. In fact, we want that the acceptance probability α falls into a reasonable range $(15-45\%)$. It comes from the fact that we would like α to be large enough to reach the convergence as soon as possible, but at the same time we'd like to be small enough to explore the support of π in detail, without risking not to converge. In fact, when the proposal has a huge variance, lots of proposed points will be rejected, but we will explore the state space efficiently. On the other hand, if the proposal is more concentrated than the posterior, most proposed values will be accepted, but since we will always move to a close point, it will take a lot to explore the parameter space and reach convergence. In addiction, we can observe that when the proposal distribution is symmetric $q(x|y) = q(y|x)$, the acceptance ratio α simply becomes:

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right), \tag{2.7}$$

and what really matters in the ratio is the stationary distribution.
Regarding the proposal of new values for y, a typical (but not necessary) choice is to define $y_j = x_{j-1} + w_j$, where w_j is a sequence of independent and identically distributed random variables, independent of the state where the chain is. Quite often, we decide to pick the w_j normally distributed with mean zero and a fixed variance. In this setting, the variance has an impact on the acceptance rate, and thereby on the performances. So, the variance should be chosen appropriately to obtain the desired acceptance ration values.
The MH algorithm can be use for Bayesian inference in two distinct context: we can use MH both alone, when the parameter is one-dimensional or use MH withing Gibbs, when we perform a MH step to simulate from the full-conditional inside the *Gibbs algorithm*.

Gibbs sampling

Gibbs sampling is a special case of the Metropolis–Hastings algorithm. This method is particularly useful when we are dealing with a multivariate distribution from which we want to sample. The core argument of this technique is that the simulation is based only on the capability of sampling from a **conditional distribution**. In more detail, consider E the subset of a possibly high-dimensional space $\mathbb{R}^n$ where the chain is defined. Indeed suppose again that the density we want to find is $\pi(x)$, where $x = (x_1, ..., x_n)^T$. We consider sequential indices i, with $1 \leq i \leq n$ and for each one of them, we choose the

proposal distribution $q(y|x)$ as:

$$\begin{aligned} q(y|x) &= \pi(y_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) && \text{if } y_j = x_j, \forall j \neq i, \\ q(y|x) &= 0 && \text{otherwise.} \end{aligned}$$

With this algorithm, the candidate point y is simply obtained by changing the i-th coordinate of x according to the **full conditional** distribution. This is simply the conditional distribution of a variable, given all the others. A very important observation in this context is that $\alpha(x, y) = 1$ for all $x, y \in E$. This is the reason why, the Gibbs sampling is considered a special case of the Metropolis-Hastings algorithm. Indeed now the candidate is always accepted, without considering a specific definition for the acceptance ratio. In general, the performance of an MCMC method is called *mixing rate*. This parameter represents how quickly the sample averages converge to the expected value we want to find. Of course, we prefer an algorithm which has a good mixing, or equivalently which quickly converges to the solution.
To better understand the algorithm, suppose $\pi(x_1, x_2, ..., x_n)$ be the joint distribution we want to sample from. At the starting point, we initialize the iteration counter to $j = 1$, the starting point to $\mathbf{x}^{(0)} = \left(x_1^{(0)}, ..., x_n^{(0)}\right)$, the maximal number of iterations K, and if necessary the length of the burn-in period B. In an ideal world the starting point of the algorithm should be chosen in a region of high probability. Anyway in several applications it's also difficult to find it, so we would prefer not to take into account the initial observations to avoid mistakes. In fact we discard the first B samples and we start compute sample averages only when the *burn-in period* is over. After the initialization, we can obtain the next values of the variables as below and repeat the following steps. Suppose we are at step j, we want to find $\left(x_1^{(j)}, ..., x_n^{(j)}\right)$. This vector can be obtained sequentially simulating the variables:

$$\begin{aligned} x_1^{(j)} &\sim \pi\left(x_1 | x_2^{(j-1)}, ..., x_n^{(j-1)}\right); \\ x_2^{(j)} &\sim \pi\left(x_2 | x_1^{(j)}, x_3^{(j-1)}, ..., x_n^{(j-1)}\right); \\ &\vdots \qquad\qquad\qquad \vdots \\ x_i^{(j)} &\sim \pi\left(x_i | x_1^{(j)}, ..., x_{i-1}^{(j)}, x_{i+1}^{(j-1)} ..., x_n^{(j-1)}\right); \\ &\vdots \qquad\qquad\qquad \vdots \\ x_n^{(j)} &\sim \pi\left(x_n | x_1^{(j)}, ..., x_n^{(j)}\right). \end{aligned}$$

At the end, we can update j to $j + 1$ and, if $j \leq K$, repeat the previous steps, otherwise, stop. The values obtained will be approximately distributed according to a stationary vector valued process whose stationary density is $\pi(x_1, ..., x_n)$. A very important observation can be made. We always use the most recent value of variables, even if we are in the middle of an iteration. Indeed if in iteration j we are considering at x_i, we use the updated values for the previous variables from 1 to $i - 1$.

2.2 Stochastic Differential Equations

In this Section, we introduce what a stochastic process is and we give the definition of a Stochastic Differential Equations, by referring to [3]. To this aim, first of all, we recall that a collection of sets $\mathcal{A}$ is a σ-algebra if and only if:

- $\emptyset \in \mathcal{A}$,
- $A \in \mathcal{A} \Rightarrow , \bar{A} \in \mathcal{A}$, where by $\bar{A}$ we denote the complement of A.
- if $A_1, A_2, ... \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$;

and that a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is the triplet made up of: Ω that is the sample space of possible outcomes of the experiment, $\mathcal{A}$, the σ-algebra, and $\mathbb{P}$ which represents a probability measure. In other words, $\mathcal{A}$ is the collection of events for which a probability can be assigned. Then a **stochastic process** is a family of random variables $\{\mathbf{X}_\gamma, \gamma \in \Gamma \subseteq \mathbb{R}\}$ assuming real values, so we have:

$$\mathbf{X}(\gamma, \omega) : \Gamma \times \Omega \to \mathbb{R}.$$

We can denote $\mathbf{X}$ by:

$$\mathbf{X} = \{\mathbf{X}_t;\, t \geq 0\} \qquad \text{or similarly} \qquad \mathbf{X} = \{\mathbf{X}(t);\, t \geq 0\}.$$

When we fix a specific value for ω, $\mathbf{X} = \{\mathbf{X}(t, \bar{\omega});\, t \geq 0\}$ is a *path* or *trajectory* of the process, this means that it depicts the evolution of the process over time, a possible realization. On the other hand, if we choose a fixed value for t, $\mathbf{X} = \{\mathbf{X}(\bar{t}, \omega);\, t \geq 0\}$ is the set of possible states the process is visiting, at time $\bar{t}$.
If we consider again a space of probability $(\Omega, \mathcal{A}, \mathbb{P})$, a *filtration* $\{\mathcal{F}_t;\, t \geq 0\}$ is an increasing family of sub-σ-algebras of $\mathcal{A}$ indexed by $t \geq 0$. What this definition really means is that, for each t, s fixed such that $s < t$, we get $\mathcal{F}_s \subset \mathcal{F}_t$ where $\mathcal{F}_0 = \{\Omega, \emptyset\}$. In the same probability space, a random variable $X : \Omega \to \mathbb{R}$ is called *measurable* to denote that it is always possible to evaluate any probability related to X. In order to give a more accurate definition of what *measurable* indicates, we introduce $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra on $\mathbb{R}$ and X^{-1} the inverse function of X. Then X is *measurable* if

$$\forall A \in \mathcal{B}(\mathbb{R}), \qquad \exists B \in \mathcal{A} : X^{-1}(A) = B,$$

so, we say that it is always possible to measure the set of values taken by X using the probability measure $\mathbb{P}$ on the original space Ω:

$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(\{\omega \in \Omega : \omega \in X^{-1}(A)\}) = \mathbb{P}(B),$$

where $A \in \mathcal{B}(\mathbb{R})$ and $B \in \mathcal{A}$.
Coming back to definitions related to filtration, given a process $\{\mathbf{X}(t);\, t \geq 0\}$, for each t, we can associate a σ-algebra denoted by $\mathcal{F}_t = \sigma(\mathbf{X}(s);\, 0 \leq s \leq t)$. This is the σ-algebra generated by the process $\mathbf{X}$ up to time t, namely the smallest σ-algebra of $\mathcal{A}$ such that $\mathbf{X}(s, \omega)$ is measurable for every $0 \leq s \leq t$. Now, if we have a stochastic process $\{\mathbf{X}_t;\, t \geq 0\}$ and a filtration $\{\mathcal{F}_t;\, t \geq 0\}$, then the process $\mathbf{X}$ is said to be *adapted* to the filtration if $\forall t \geq 0, \mathbf{X}(t)$ is $\mathcal{F}_t$-measurable.

Definition 2.1 *Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a filtration* $\mathbb{F} = \{\mathcal{F}_t; t \geq 0\}$ *on* $\mathcal{F}$*, a* ***martingale*** *is a stochastic process* $\{\boldsymbol{X}(t);\, t \geq 0\}$ *such that:*

- $\mathbb{E}[\mathbf{X}_t] < \infty, \qquad \forall t \geq 0;$
- *it is adapted to the filtration* $\mathbb{F}$;
- $\mathbb{E}[\mathbf{X}_t|\mathcal{F}_s] = \mathbf{X}_s \qquad 0 \leq s \leq t < \infty.$
  In other words $\mathbf{X}_s$ *is the best predictor of* $\mathbf{X}_t$ *given* $\mathcal{F}_s$*. From this property, we can also characterize martingales as stochastic processes with a constant mean* $\forall t \geq 0$.

2.2.1 Brownian motion

With the aim of describing the evolution of a Markov process with **continuous states**, we can introduce a particular kind of martingale: the *Brownian motion*. It is also called *Wiener process* and in the following it will denoted by $\mathbf{W} = \{W(t), t \geq 0\}$. There are several options to define this concept, let's see some of them:

Definition 2.2a *A* ***Brownian motion*** *is a Gaussian process with:*

1a) *continuous paths;*

2a) *independent increments;*

3a) $W(0) = 0$*, with probability* 1.

Alternatively, we can give an equivalent definition of the same mathematical object, recalling the notion of measurability.

Definition 2.2b *The process* $\mathbf{W} = \{W(t), t \geq 0\}$ *is a* ***Brownian motion*** *if it is a adapted to* $\mathbb{F}$ *and:*

1b) $W(0) = 0$ *almost surely;*

2b) $W(t) - W(s) \sim N(0, t-s), \qquad \forall s < t;$

3b) *it has independent and stationary increments:* $W(t) - W(s)$ *is independent of* $W(t') - W(s'),\ \forall t > s \geq t' > s'$.

In order to simulate a trajectory of the Wiener process on a specific time interval $[0, T]$, when the time increment $\Delta t > 0$ is fixed, we use some properties of the Gaussian process. In fact, if we consider a specific time t, because of property 2b), it is true that:

$$W(t + \Delta t) - W(t) \sim N(0, \Delta t) \sim \sqrt{\Delta t}\, N(0, 1).$$

At this point, the simulation is really easy to perform. Indeed, after we built a grid on the interval $[0, T]$: $0 = t_1 < t_2 < ... < t_{N+1} < t_N = T$ such that $t_{i+1} - t_i = \Delta t$, we can repeat the following steps until we reach the number N of different points on the grid:

1. It's necessary to sample z from the Gaussian distribution $N(0, 1)$ and to increment the index i,

2. Now we can evaluate $W(t_i) = W(t_{i-1}) + z\sqrt{\Delta t}$.

Of course, we start from $i = 1$ and with $W(0) = W(t_1) = 0$. Once we get the values of the process on the points of the grid, we can approximate the trajectory in between two consecutive points with different techniques, even if the most used is the linear interpolation.

2.2.2 Stochastic integrals

When we consider a Brownian motion, we observe that it has continuous paths which are nowhere differentiable. This is a direct consequence of the independence of the increments and also of their distribution. In particular we notice that:

$$\lim_{\Delta t \to 0} \frac{|W(t+\Delta t) - W(t)|}{\Delta t} \simeq \lim_{\Delta t \to 0} \frac{\sqrt{\Delta t}}{\Delta t} = +\infty.$$

We define the variation of a process $\Delta W = W(t+\Delta t) - W(t)$, and this simple observation is just an insight into why the variations of a Wiener process $\mathrm{d}W(t) = \lim_{\Delta t \to 0} \Delta W$ are not finite. As a consequence we have some problems when we'd like to describe the evolution of a process with a differential equation.
For example, suppose we consider a certain quantity $S(t)$ depending on some parameter, say the time. We assume that its evolution is due to a deterministic contribution and a stochastic one, related to a standard Wiener process. At this point, we can consider the variation in a small time interval $[t, t + \Delta t]$. A possible process description is given by:

$$\Delta S = \mu S(t)\Delta t + \sigma S(t)\Delta W.$$

If we consider the previous difference equation for an infinitesimal time interval, we obtain a stochastic differential equation which can be written as:

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W(t). \tag{2.8}$$

Unfortunately, this equation has no meaning from a mathematical point of view because of previous considerations. In order to obtain a meaningful equation, we convert (2.8) into an integral form and we introduce the concept of *stochastic integral* with respect to the Brownian motion. Without going into details, given a generic integrand $f : [0, T] \times \Omega \to \mathbb{R}$, the stochastic integral,

$$I(f) = \int_0^T f(u)\mathrm{d}W(u),$$

is defined as the unique limit in quadratic mean for $n \to \infty$ of the sequence of the integrals $I(f^n)$. Here f^n are simple processes which are defined as

$$f^n(t, \omega) = f(t_j, \omega), \qquad t_j \le t < t_{j+1},$$

over the partition $\Pi_n([0, T])$. So those integrals are simply given by a summation:

$$I(f^{(n)}) = \sum_{j=0}^{n-1} f(t_j)\{W(t_{j+1}) - W(t_j)\}.$$

A particular kind of stochastic process used in this thesis is the **diffusion process**. The main stochastic differential equation is:

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{2.9}$$

with some initial condition $X(0)$ which can be random or not. Equation (2.9) can be also written into the equivalent integral form:

$$X(t) = X(0) + \int_0^T b(t, X(t))dt + \int_0^T \sigma(t, X(t))dW(t). \tag{2.10}$$

Here b and σ are two deterministic functions and they are called respectively the *drift* and *diffusion* coefficients. A very important result when dealing with an equation of the same type as (2.9) is the existence and uniqueness of the solution. To state this outcome we need two previous assumption.

Assumption 2.1 ***(Global Lipschitz)*** *For all $x, y \in \mathbb{R}$ and $t \in [0, T]$, there is a constant $K < +\infty$:*

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| < K|x - y|.$$

Assumption 2.2 ***(Linear growth)*** *For all $x \in \mathbb{R}$ and $t \in [0, T]$, there is a constant $C < +\infty$:*

$$|b(t, x)| + |\sigma(t, x)| < C(1 + |x|).$$

Theorem 2.2 *If Assumption 2.1 and Assumption 2.2 are true, the stochastic differential equation (2.9) has a unique, continuous and adapted strong solution such that*

$$\mathbb{E}\left\{\int_0^T |X_t|^2 dt\right\} < \infty.$$

This solution is called *diffusion process*. The Theorem 2.2 states the solution X is oof strong type and this simply means that it is path-wise unique.

2.3 Euler approximation

When a particular kind of process is described by a SDE, in most cases we are interested in the simulation of its trajectory. This is done in order to have an idea of both the shape of the whole trajectory and the expected value of some quantity functional of the process. In this section we refer to and further information can be found in [3] and [4]. The general reasoning used by simulation techniques is the discretization of continuous solution of the SDE. Based on the different properties of approximation methods, we can distinguish between them. In particular, we can state two different criteria of optimality for approximation method which are widely used in literature: the strong and the weak

orders of convergence.

Strong and weak order of convergence

Given a continuous-time process Y and δ the maximum time increment if the discretization, a time-discretized approximation Y_δ is of general *strong* order of convergence γ to Y if, for any fixed time horizon T,it is true that:

$$\mathbb{E}[|Y_\delta(T) - Y(T)|] \leq C\delta^\gamma, \qquad \forall \delta < \delta_0,$$

where $\delta_0 > 0$ and C is a constant not depending on δ.

Y_δ is said to converge *weakly* of order ω to Y if for any fixed time horizon T and any $2(\omega+1)$- continuous differentiable function g of polynomial growth, it holds true that:

$$|\mathbb{E}_g[Y(T)] - \mathbb{E}_g[Y_\delta(T)]| \leq C\delta^\omega, \qquad \forall \delta < \delta_0,$$

where again $\delta_0 > 0$ and C is a constant not depending on δ.

Weak order of convergence corresponds to requesting something less rather than strong convergence. This is the reason why, in general, schemes of approximation that strongly converge usually have an higher order of weak convergence. For example the Euler scheme is strongly convergent of order $\gamma = \frac{1}{2}$ and weakly convergent of order $\omega = 1$. **Euler scheme** is one of the most used approximation method. Let's see in detail how this technique works. Suppose that we are in 1-dimensional space and we know that $\{X(t), 0 \leq t \leq T\}$ is the solution of the stochastic equation:

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t),$$

with an initial deterministic condition. In an attempt to render the entire notation more manageable in the following we use an equivalent form:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t.$$

At this point, we'd like to evaluate the so called *Euler-Maruyama approximation*. At the beginning, we discretize the time interval $[0, T]$ into $\Pi_n = \{0 = t_0 < t_1 < ... < t_n = T\}$. At this point, the approximation is a continuous stochastic process Y which satisfies the following scheme:

$$Y_{i+1} = Y_i + b(t_i, Y_i)(t_{i+1} - t_i) + \sigma(t_i, Y_i)(W_{i+1} - W_i), \qquad i = 0,1,...N-1;$$

with $Y_0 = X_0$. For sake of simplicity, but it is not mandatory, we can choose time points equally spaced. In such a way, $\Delta t = t_{i+1} - t_i = \Delta$ and $t_i = i\Delta$ where $\Delta = \frac{T}{N}$. In any case, in between any two time points, the process can be determined through linear interpolation, and so, for any t between two different points in the grid t_i and t_{i+1}, we have:

$$Y(t) = Y_i + \frac{t - t_i}{t_{i+1} - t_i}(Y_{i+1} - Y_i), \qquad t \in [t_i, t_{i+1}).$$

Once we know this recursive scheme, we just need to generate random increments of the Wiener process. This can be done using a really easy method. The steps to be followed are described in the last part of Section 2.2.1.

2.3.1 Multidimensional SDE

On the other hand, if we are in a d-dimensional space and we are dealing with the following differential equation:

$$d\boldsymbol{X}_t = \boldsymbol{b}(t, \boldsymbol{X}_t)dt + \sum_{j=1}^{m} \boldsymbol{s}^j(t, \boldsymbol{X}_t)dW_t^j, \quad t \in [0, T], X_0 \in \mathbb{R}^d.$$

The iterative process we obtain for the k-th component of the Euler approximation is:

$$Y_{i+1}^k = Y_i^k + b^k(t_{i+1} - t_i) + \sum_{j=1}^{m} s^{k,j}\Delta W, \tag{2.11}$$

where $i = 0,1,...N-1$, $Y_0 = X_0$ and $k = 1, ..., d$. Into formula (2.11) $\Delta W = (W_{i+1}^j - W_i^j)$ is the independent Gaussian distributed increment of the j-th component o the m-dimensional Wiener process W on $[t_n, t_{n+1}]$. Moreover ΔW^{j_1} and ΔW^{j_2} are independent for $j_1 \neq j_2$. Finally the diffusion coefficient $\mathbf{s} = [s^{k,j}]_{k,j=1}^{d,m}$ in this context is a $d \times m$-matrix. The Euler scheme is often used when the drift and diffusion coefficients are nearly constant and the time step is sufficiently small. In those situations this method is able to provide a good approximation for the solution. However, if possible, it's recommended to use some more sophisticated techniques which are able to get an higher order of convergence. One of possible alternative method is the Milstein scheme.

2.4 Density dependent process

The main topic of this thesis consists in understanding, modeling and interpreting biochemical reaction systems. They are used in very different applications, but we focus manly on chemical kinetics and epidemic diffusion. First of all, we give a general definition of this kind of systems and then we study a method in order to find a suitable model to describe the process we are interested in. The main reference in this Section are [5], [6] and [7]. There, further insight can be found.
To introduce this topic, it is really important to understand that **biochemical reaction systems** are made up of two different elements. The first one is a *reaction network*, while the other is a *choice of dynamics*. A reaction network is the triple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ where:

- $\mathcal{S} = \{\mathcal{S}_1, ..., \mathcal{S}_d\}$ is the set of species, namely the chemical components whose evolution we would like to describe into the model;
- $\mathcal{C}$ is the set of complexes, which are non-negative linear combination of the species;
- $\mathcal{R}$ is a finite set of ordered couples of complexes which tells how to convert one complex to another.

The last element is connected to the dynamic of the system and it is governed by chemical rules. In particular, a chemical reaction is defined as an event that modify the state of

the network according to stoichiometric equations. *Stoichiometric equations* are the set of rules controlling a reaction network. The k-equation can be written as:

$$\sum_{i=1}^{d} c_{ki} S_i \rightarrow \sum_{i=1}^{d} c_{ki}^{'} S_i, \qquad k = 1, ...K. \tag{2.12}$$

The equation (2.12) describes how some complexes are consumed in order to produce other ones. In the formula c_{ki} and $c_{ki}^{'}$ are non-negative integers and they are associated with the source and product of a specific complex. Using an abuse notation, we can write directly $\mathcal{R} = \{c_k \rightarrow c_k^{'} : c_k, c_k^{'} \in \mathcal{C} \text{ and } c_k \neq c_k^{'}\}$. This allows for a more immediate intuition of the reason why a reaction network can be modeled by an unique directed graph. There the set of nodes coincides with the set of complexes.
In order to model the dynamical behavior of a reaction network, we have different alternatives. One of the most diffused method, we describe in Section 2.4.1, is a Markov chain. More in detail, Continuous Time Markov Chain (CTMC) are widely used in order to describe stochastic models of reaction networks. Anyway, a very critical point using this approach is that it's not feasible finding an analytic solution in real applications. In this context, in fact the number of reactions or interactions is so high that the only possible strategy is using an approximation. Otherwise we can use a deterministic approach. For several years, these two different paths were followed in parallel. The *deterministic modes* were used in order to investigate more theoretical aspects. Whereas *stochastic models* were used to study the process from a computational point. At the beginning, these models were independent. Their relationship was explained for the first time by T. Kurtz's works. It was proved that, when the dimension of the system is large, stochastic models converge to the deterministic ones. This result is really important. Indeed, it was clarified that the deterministic model was not an alternative to the stochastic model. On the contrary, it can be considered as an approximation of the stochastic one. Actually, in many practical applications the number of states is so large that even numerical simulations becomes computationally infeasible. In this chapter, we first introduce the definition of density process that we would like to approximate. Then we analyze how to model it. This is the reason why, we introduce the family of density processes. Thereafter, because of previous justifications, we analyze more in detail the two different approximations developed by Kurtz. In Section 2.4.2 we explore more in detail in which situations we can use a deterministic approximation. Otherwise we can use another kind of approximation developed into the framework of diffusion theory, as explained in Section 2.4.3.

2.4.1 Markov Chain

The dynamic of a reaction network can be described for example through a stochastic model. One of the most used method is a Markov chain. In the following, we consider a particular kind of Markov process used in a very large set of applications. In particular it is an important instruments used to model reaction networks.

Definition 2.2 *A continuous-time Markov process defined on the d-dimensional lattice $\mathbb{Z}^d$, where we can explicitly indicate the dependency on a parameter n:*

$$Z_n = \{Z_n(t); t \geq 0\}, \qquad \textit{for } n \geq 1,$$

is called ***density dependent*** *if it's true the following. Considering any state* $z, \ell \in \mathbb{R}^d$*, its transition rates can be written as:*

$$q^{(n)}_{z,z+\ell} = n\beta_\ell\left(\frac{z}{n}\right).$$

Here the transition rates $\beta_\ell(x)$ are continuous non-negative functions. In simple terms, two conditions are required because the process is considered "density dependent":

a) there is a linear relation between transition rates and the parameter n,

b) there is a dependence on the density of the state rather than on the state value.

Before going ahead, we look for an interpretation of the parameter n. If we are dealing with a system in the context of reaction networks, it can interpreted as the volume of the container where molecules are located. On the other hand, if we are considering population dynamics, it represents the number of individuals in a particular population. Let's see more in detail how a density dependent Markov process works. First of all, we recall that in a Poisson process, there is a dependence on the length of the interval in the rates. Then, if we consider the Definition in 2.2, we can derive that:

$$\begin{aligned}\mathbb{P}(Z_n(t+h) = z+\ell | Z_n(t) = z) &= hn\beta_\ell\left(\frac{z}{n}\right) + o(h), \quad \ell \neq \mathbf{0},\\ \mathbb{P}(Z_n(t+h) = z | Z_n(t) = z) &= 1 - hn\sum_\ell \beta_\ell\left(\frac{z}{n}\right) + o(h).\end{aligned} \tag{2.13}$$

Observing these probabilities, it is clear the reason for the name of this kind of process. Indeed in the jump rates there is a normalization by n and the fraction exactly represents the density. Before proceeding, we also suppose that the number of possible transitions ℓ, for which $\sup_x \beta_\ell(x) > 0$, is finite. This is a quite normal assumption in the context of chemical kinetics. It prescribe that there is a finite number of chemical reactions and all of them have a finite speed. Moreover the starting point $Z_n(0)$ is known. In the definition of jump intensities, we observe the multiplication by n, so when this parameter becomes larger and larger, the approximation with a *diffusion process* is more and more accurate. We will go into more detail this concept in Section 2.4.3.
At this point, we observe that $Z_n(t)$ can be written into two equivalent forms in the sense of probability law. The first one is by means of the following stochastic differential equation:

$$\mathrm{d}Z_n(t) = \sum_\ell \ell \mathrm{d}M_\ell(t), \tag{2.14}$$

where the state dependent rate associated with $M_\ell(t)$ is $q = n\beta_\ell(n^{-1}Z_n(t))$. Or equivalently equation (2.14) can be written as:

$$Z_n(t) = Z_n(0) + \sum_\ell \ell Y_\ell\left(n\int_0^t \beta_\ell(n^{-1}Z_n(s))ds\right). \tag{2.15}$$

Here for each possible transition ℓ, we introduced an independent standard Poisson process $Y_\ell = \{Y_\ell(t); t \geq 0\}$. So, $Y_\ell(t)$ counts the occurrences of events whose effect is to increase $Z_n(t)$ by ℓ. Before moving on, we have to prove that this process satisfies the probabilities given by definition in (2.13). To this aim, we recall that for a Poisson process, when the time step is short, the probability of a jump is proportional to the length of the interval. So, under the assumption that $Z_n(t) = z$, the probability to have a jump in $Y_\ell\left(n\int_0^t \beta_\ell(n^{-1}Z_n(s))ds\right)$ during the time interval $(t, t+h)$ is $hn\beta_\ell(n^{-1}z)+o(h)$, because the integrand function is constant until the first jump after t. This observation simply verifies that equations (2.15) complies with the requirements in Definition 2.2.
On the other hand, if our final aim is to characterize this system through independent unit-rate Poisson processes, we can substitute the counting process $Y_\ell(t)$ rewriting the precious expression. Before doing so, in many applications, it is common to rescale the process because of mathematical simplification. For this reason, we can introduce a new definition.

Definition 2.3 *For every n and any density dependent family $Z_n = \{Z_n(t); t \geq 0\}$, we can define the family of **density process** $\bar{Z}_n = \{\bar{Z}_n(t); t \geq 0\}$ as:*

$$\bar{Z}_n(t) = n^{-1}Z_n(t).$$

It is also called the normalized process.

We can also define $\hat{Y}_\ell(t) = Y_\ell(t) - t$ the *centered Poisson process* and $F(x) = \sum_\ell \ell\beta_\ell(x)$, where $x \in \mathbb{Z}^d$, the *drift function*. Then equation (2.15) can be rewritten in the following equivalent form:

$$\bar{Z}_n(t) = \bar{Z}_n(0) + \frac{1}{n}\sum_\ell \ell\hat{Y}_\ell\left(n\int_0^t \beta_\ell(\bar{Z}_n(s))ds\right) + \int_0^t F(\bar{Z}_n(s))ds. \tag{2.16}$$

In equation (2.16), $\hat{Y}_\ell(t)$ is an independent unit-rate Poisson process that counts the occurrences of the events which are able to increase the density process $\bar{Z}(t)$ by $\frac{\ell}{n}$. There are several techniques used to characterize both the initial transient period and the long run behavior of a CTMC. Anyway, in many real applications the state space of a the chain is so large that an analytical treatment would be infeasible. This is the reason why, we can use approximations. The main idea is to find a simpler process which is able to obtain good results in terms of approximating the real evolution of the phenomenon.

2.4.2 Fluid Limit

The main goal of this section is to find a technique in order to approximate a density dependent Markov process when the population size, denoted by the parameter n, increases. We emphasize that the approximation doesn't concern only the last period of the process, but we are interested into approximating its whole evolution over time. The first observation we can make is that when the number of states of the chain becomes more and more higher, the jumps of the stochastic process become more frequent. Another consequence is

that the jump magnitude becomes smaller and smaller. The previous observation justifies the approximation of a trajectory by a continuous function. This is the so called *fluid limit* or *fluid approximation* and it is the first approximation technique we analyze.
When the model describes the interactions of large groups, it can be proved that a set of ordinary differential equations (ODE) can be used for the approximation. Therefore we define the deterministic vector function $z(t)$, which is the solution to the integral equation:

$$z(t) = z_0 + \int_0^t F(z(s))ds. \tag{2.17}$$

At this point we can make the following:

Assumption 2.3 *For each compact $K \in \mathbb{R}^d$ (or in the state space) exists a constant $M_K > 0 : |F(x) - F(y)| \leq M_K|x - y|, \forall x, y \in K$. This equivalent to the requirement that the function F is Lipshitz continuous in K.*

Now, all required elements have been introduced in order to enunciate one of the main result in terms of approximation. We consider again the process $\bar{Z}(t)$ defined in (2.16).

Theorem 2.3 *Suppose that $\lim_{n\to\infty} \bar{Z}_n(0) = z_0$ and that Assumption 2.3 is verified. Then, $\lim_{n\to\infty} \sup_{s\leq t} |\bar{Z}_n(s) - z(s)| = 0$ a.s., where $z(s)$ is the unique solution of (2.17).*

Proof. First of all, we enunciate the Gronwall's inequality which will be used in the following demonstration.

Lemma 2.1 *(**Gronwall's inequality**) Assume f is a real function satisfying $0 \leq f(t) \leq a + b\int_0^t f(s)ds, \forall t \geq 0$, with a and b positive constants. Then $f(t) \leq ae^{bt}, t \geq 0$.*

Proof. We use the first inequality iteratively:

$$\begin{aligned}
f(t) &\leq a + b\int_0^t f(s_1)ds_1 \leq a + b\int_0^t \left(a + b\int_0^{s_1} f(s_2)ds_2\right) \\
&= a + abt + b^2 \int_0^t ds_1 \int_0^{s_1} (f(s_2)ds_2) \\
&\leq a + abt + b^2 \int_0^t ds_1 \int_0^{s_1} ds_2 \left(a + b\int_0^{s_2} f(s_3)ds_3\right) \\
&= a + abt + b^2 \left(a\frac{t^2}{2} + b\int_0^t ds_1 \int_0^{s1} ds_2 \int_0^{s_2} f(s_3)ds_3\right) \\
&\leq a\left(1 + bt + \frac{(bt)^2}{2} + \frac{(bt)^3}{6}\right) + b^4 \int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 \int_0^{s_3} f(s_4)ds_4 \\
&\leq \ldots \leq a\sum_{k=0}^{\infty} \frac{(bt)^k}{k!} = ae^{bt}.
\end{aligned}$$

□

Now, we are ready to prove the theorem. In the following, we use one of important properties related to a Poisson process $Y = \{Y(t); t \geq 0\}$. It can be proved that:

$$\lim_{n\to\infty} \sup_{s\leq t} |n^{-1}Y(ns) - s| = \lim_{n\to\infty} \sup_{s\leq t} n^{-1}|\hat{Y}(s)| = 0 \text{ almost surely, for any } t \geq 0. \quad (2.18)$$

More general results and other extension can be found in [8]. This is why, the second addendum in (2.16) gets smaller as n increases. From this observation, we can conclude that, under this condition, $\bar{Z}$ is really similar to z. In order to get a more compact notation, we introduce: $\bar{\beta}_\ell = \sup_{x\in K} \beta_\ell(x)$ which is finite because K is compact and β_ℓ is a continuous function, so we can use Weierstrass theorem. We recall definition in (2.16) and (2.17) so that we can write, also using Gronwall's inequality (Lemma 2.1):

$$\begin{aligned}
|\bar{Z}_n - z(s)| &= |\bar{Z}_n(0) - z(0) + \frac{1}{n}\sum_\ell \ell \hat{Y}_\ell \left(n\int_0^s \beta_\ell(\bar{Z}_n(u))du\right) \\
&\quad + \int_0^s \Big(F(\bar{Z}_n(u)) - F(z(u))\Big)\, du| \\
&\leq |\bar{Z}_n(0) - z(0)| + \frac{1}{n}\sum_\ell |\ell| \sup_{u\leq s} |\hat{Y}_\ell(n\bar{\beta}_\ell u)| + \int_0^s M_K |\bar{Z}_n(u) - z(u)|du \\
&\leq \left(|\bar{Z}_n(0) - z(0)| + \sum_\ell |\ell| \sup_{u\leq s} n^{-1}|\hat{Y}_\ell(n\bar{\beta}_\ell u)|\right) e^{M_K s}.
\end{aligned}$$

If we consider the supremum, to prove the theorem:

$$\sup_{s\leq t} |\bar{Z}_n(s) - z(s)| \leq \left(|\bar{Z}_n(0) - z(0)| + \sum_\ell |\ell| \sup_{u\leq t} n^{-1}|\hat{Y}_\ell(n\bar{\beta}_\ell u)|\right) e^{M_K t}.$$

By assumption, the first term on right side converges to 0 and so does the second one because of (2.18), while the exponential doesn't depend on n. In this way, the desired result is obtained. □

Intuitively, we can justify the deterministic approximation recalling that the process $\bar{Z}_n$ starts at $Z_n(0)$ and the average drift of $\bar{Z}_n(s)$ at s is $F(\bar{Z}_n(s))ds = \sum_\ell \ell\beta_\ell(\bar{Z}_n(s))$, thus the process should be approximated by $Z_n(0) + \int_0^t F(\bar{Z}_n(s))ds$.

Once we obtained this convergence result, we can study how much the two processes $\bar{Z}(t)$ and z differ. The conclusion, we are looking for, is given by the *Central Limit Theorem*. Its final result is that deviations are of order $\sqrt{n}$. Let's see more in detail the reason behind this outcome. To this aim, we introduce two related scaled processes in order to give a simpler formal derivation. The first one is:

$$W_\ell^{(n)}(t) = \sqrt{n}(n^{-1}Y_\ell(nt) - t) = n^{-\frac{1}{2}}\hat{Y}(nt).$$

It can be demonstrated that this process converges weakly to the standard Brownian motion denoted by W_ℓ. We can introduce also the process $V^{(n)}(t)$ which is centered and

scaled by $\sqrt{n}$:

$$V_n(t) = \sqrt{n}(\bar{Z}_n(t) - z(t)) =$$

$$= \sqrt{n}\left(\bar{Z}_n(0) + \frac{1}{n}\sum_\ell \ell \hat{Y}_\ell\left(n\int_0^t \beta_\ell(\bar{Z}_n(s))ds\right) + \int_0^t F(\bar{Z}_n(s))ds - z_0 - \int_0^t F(z(s))ds\right) =$$

$$= \sqrt{n}(\bar{Z}_n(0) - z_0) + \sum_\ell \ell n^{-\frac{1}{2}}\hat{Y}_\ell\left(n\int_0^t \beta_\ell(\bar{Z}_n(s))ds\right) + \int_0^t \sqrt{n}(F(\bar{Z}_n(s)) - F(z(s)))ds$$

$$= v_n(0) + \sum_\ell \ell W_\ell^{(n)}\left(\int_0^t \beta_\ell(\bar{Z}_n(s))ds\right) + \int_0^t \sqrt{n}(F(\bar{Z}_n(s)) - F(z(s)))ds.$$

This expression can be further simplified. We use Taylor's expansion for the last integrand so that we obtain the following expression:

$$\sqrt{n}(F(\bar{Z}_n(s)) - F(z(s))) = \sqrt{n}\partial F(z(s))(\bar{Z}_n(s) - z(s)) + O(\sqrt{n}|\bar{Z}_n(s) - z(s)|^2)$$

$$= \partial F(z(s))V_n(s) + V_n(s) + O(|\bar{Z}_n(s) - z(s)|),$$

where $\partial F = (\partial_j F_i)$ is the matrix of partial derivatives. Recalling the previous convergence results in Theorem 2.3, we can define the following process in order to find a possible convergence result for V_n:

$$V(t) = v_0 + \sum_\ell \ell W_\ell\left(\int_0^t \beta_\ell(z(s))ds\right) + \int_0^t \partial F(z(s))V(s)ds, \tag{2.19}$$

and state the following theorem in which we use the notation $L(x) = \sum_\ell \ell\ell^T\beta_\ell(x)$,

Theorem 2.4 ***(Central Limit Theorem)*** *Suppose ∂F is continuous and $\lim_{n\to\infty} v_n(0) = v_0$ (constant). Then $V_n \Rightarrow V$, defined in equation (2.19). Moreover, the process V is a Gaussian vector process with covariance matrix*

$$Cov(V(t), V(r)) = \int_0^{r\wedge t} \Phi(t,s)L(z(s))(\Phi(r,s))^T ds,$$

with Φ is a matrix function defined as the solution of:

$$\Phi_2'(t,s) = -\Phi(t,s)\partial F(z(s)) \qquad \Phi(s,s) = I,$$

where $\Phi_2'(t,s)$ denotes the partial derivative with respect to s.

Theorem 2.4 is really important because it states that a stochastic density dependent process converges to a Gaussian process with the specified covariance matrix. This simply means that the first process can be approximated by a Gaussian process. It's important to underline that this approximation may be used only when n is large enough. For example, suppose we want to use the previous model to describe the diffusion of an infectious disease. The result of Central Limit Theorem is valid only when there are many infectious individuals. This implies that we have to exclude the initial and final phases of the epidemic.

2.4.3 Diffusion Approximation

In the previous section, we have considered a density dependent Markov process and its approximation when the parameter n gets bigger. This is related to the number of states in the chain. Suppose for example that we are modeling an epidemic process, saying that n is increasing means to assume that the initial population which is susceptible to the virus is increasing as well. We already mentioned that the approximation doesn't hold when the epidemic is starting or ending because the parameter n is not large enough. On the contrary, for large n, Theorems 2.3 and 2.4 are applicable. So, every trajectory remains bounded in a small interval around the deterministic solution. However with the previous technique, every stochastic effect is lost. In fact, in that context only the mean of the process is relevant, while its stochastic nature is ignored. In many application, this may be a problem. In fact, stochastic effects may play an important role and the deterministic approximation simply is not justified. In this context effects like for example variance and skewness should be included into the model. This is the reason why, we need another method to approximate the solution. This aim can be reached in terms of the *diffusion process.*
First of all we recall the definition of $\bar{Z}(t)$ given by (2.16) and that one of $z(t)$ in (2.17). Then we define the diffusion process $G_n(t)$ which solves:

$$G_n(t) = G(0) + \sum_{\ell} \frac{\ell}{n} \left[n \int_0^t \beta_\ell(G_n(s))ds + W_\ell(n \int_0^t \beta_\ell(G_n(s)ds)) \right]. \tag{2.20}$$

Here W_ℓ are independent standard Wiener processes. In addiction $G_n(0) = \bar{Z}(0)$. Now we are ready to state the other important approximation result.

Theorem 2.5 *Suppose to consider two already defined processes: 2.17 and 2.20. Moreover assume that:*

- $\lim_{n\to\infty} \bar{Z}(0) = z_0$,
- $E \in \mathbb{R}^d$ *is the smallest hyper-rectangle containing the discrete state space of* $\bar{Z}(t)$,
- U *is any open connected subset in* E *that contains* $z(t)$ *for every* $0 \le t \le T$.
- $\exists M > 0$ *such that :*

$$|\beta_\ell(x) - \beta_\ell(y)| \le M|x - y|$$
$$|F(x) - F(y)| \le M|x - y|,$$

  $\forall x, y \in U$.

Define $\tau_n = \inf\{t : \bar{Z}_n(t) \notin U \text{ or } G_n(t) \notin U\}$. *We can observe that* $\mathbb{P}(\tau_n > T) \to 1$ *when* $n \to \infty$. *At this point, we can conclude that:*

$$\sup_{0\le t\le \tau_n \wedge T} |\bar{Z}_n(t) - G_n(t)| = O\left(\frac{\log n}{n}\right), \tag{2.21}$$

for any fixed time horizon T.

Again, for the proof and better estimation of the distance, we refer to [8]. The previous approximation is quite complex, anyway we can reformulate it drawing main conclusion. In easy words Theorem 2.5 states that trajectories of process $\bar{Z}(t)$ and process $G_n(t)$ can be constructed on the same probability space. In addiction the maximum distance between them decreases when $n \to \infty$ at a rate equal to $\frac{\log n}{n}$. Because of theory of time changed, the equation (2.20) defining the process $G_n(t)$ can be rewritten as:

$$G_n(t) = G_n(0) + \sum_{\ell} \frac{\ell}{\sqrt{n}} \left[\sqrt{n} \int_0^t \beta_\ell(G_n(s))ds + \int_0^t \sqrt{\beta_\ell(G_n(s))} \right]. \tag{2.22}$$

This formula is widely used in chemistry to model reactions and is known as *Langevin equation*. Thanks to this rewriting, we are able to compare more easy this solution with $z(t)$ in (2.17). In fact, we stress the fact that both $z(t)$ and $G_n(t)$ provide the strong approximation of $\bar{Z}_n(t)$. We notice that the first terms in the equations are really similar. The news in (2.22) is the second addendum. It adds noise and simply represents the stochastic nature of the process. This is why with this kind of approximation we are able to obtain a lower error into the approximation. As result, the diffusion approximation can be applied in many circumstances even when the deterministic one fails.

Chapter 3

Deterministic Epidemic Models

Ordinary differential equations (ODEs) can be used to describe dynamic systems in various fields, including ecology, economics, biology and also epidemics. In this Chapter we focus on the last kind of applications and our main reference is [7]. As well as for the next one where we explore an alternative model used also in epidemics. Some of the earliest and most common epidemic models are, for example SIS and SIR. They are both, as the overwhelming majority of the epidemic models, examples of *compartment models*. As the name says, they are based on the compartmentalization of individuals according to their status with respect to a particular disease. In this way, they try to capture patterns of dynamic changes in the sizes of the compartments over time. In the simpler compartmental model, the SI model, the population is divided into two different groups: Susceptible and Infectious and the size of each group changes following the defining differential equations. Considering more realistic conditions, such as reinfection, recovery, immunity and so on, the SI model has been extended to SIS (susceptible, infectious, susceptible), SIR (susceptible, infectious, removed), SIRD (susceptible, infectious, recovered, deceased), SIRV (susceptible, infectious, recovered, vaccinated), SEIR (susceptible, exposed, infectious, recovered), and many others.
Epidemic models are useful tools which study the transmission mechanisms of a disease and which try to predict the future development of an outbreak. In 1662, John Graunt, was the first one who systematically analyzes the causes of death in London's population. His studies are considered the earliest application of the "theory of competing risks", which nowadays is still one of the fundamental principles in epidemiology. In 1760, the first attempt to model the spread of a disease was proposed by Daniel Bernoulli in order to defend the practice of inoculating against smallpox. In the early 20th century, William Hamer (1906) and Ronald Ross (1908) formulate the earliest studies of the non-linearity of an epidemic model. Hamer first guessed that the probability that there will be a new infection in the next discrete time-step was proportional to the product of actual susceptibles and infectives. While, Ross translated this intuition of "mass action principle" to the

continuous-time setting. Anyway, we have to wait until 1927 for the first complete mathematical model for the spread of an infectious disease by Kermack–McKendrick. This model describes the deterministic relationship between susceptible, infective and removed individuals in a population (SIR model). At that time, also the Reed–Frost model (1928), gained the attention of many researchers even if it was never published, but it was included in a series of lectures. This model has been used as an introduction to stochastic epidemic theory and it is described in Section 4.1.
In this Chapter, first we introduce the Kermack–McKendrick model. Then we provide some general observations that will be really useful even in other models. Finally, we briefly discuss and compare deterministic and stochastic model. In fact, the final objective of Section 3.3 is to show how the two models may be considered as both useful tools and not as two independent alternatives to describe real world situations.

3.1 The Kermack-McKendrick model

In this Section, we describe more in detail the first mathematical model for the spread of an infectious disease. In 1927, W. O. Kermack and A. G. McKendrick proposed a theory which predicts the number of infective in a fixed population over time. This will be one of the building block of the subsequent SIR models and their relatives.
The model is known as **deterministic general epidemic** even if, after its publication, there were been several generalization. In fact this model is based on some fundamental (and also restrictive) assumptions. First of all, we clarify what kind of disease we are going to describe. In this context, an infectious disease, which is able to spread from person to person when there is a contact between an infective and a susceptible, is considered. In fact, as in any SIR model, we assume that population can be divided into Susceptible, Infectious and Removed. At the beginning, people are susceptible to the disease, which means they may contract the illness. In general, when we consider a new disease for which there has not already been the development of an epidemic, the entire population could be susceptible, unless someone is already immune due to previous vaccination or genetic characteristics. Then, each infected person remains infectious for some time and finally he is removed. A removed individual is one which can no longer be considered in the number of those who are sick, because of recovery or death, but in any case he doesn't play any further role in the spread of the disease. This model in fact, doesn't consider the possibility of reinfection. This assumption remains true for many infectious disease.
Moreover, we assume that the time interval in which the spread of the disease occurs, is shorter compared to the life of an individual. This is true in the vast majority of cases and so we can consider the population constant during the analysis. This is why, if we consider a specific time instant t and we denote with $x(t)$ the number of susceptible, $y(t)$ the number of infective and $z(t)$ the number of removed individuals in that moment; then $x(t) + y(t) + z(t) = n$. Where n is a constant used to denote the total size of the population we are considering. To summarize, even if we are ignoring any process of birth or emigration, we expect the approximation will be still accurate.

The process is described by the following set of differential equations:

$$\begin{cases} x'(t) = -\dfrac{\lambda}{n} x(t) y(t) \\ y'(t) = \dfrac{\lambda}{n} x(t) y(t) - \gamma y(t) \\ z'(t) = \gamma y(t), \end{cases} \tag{3.1}$$

where $\lambda, \gamma > 0$. In order to allow the spread of the disease, it is necessary to introduce a certain number of infective in the population, while the number of removed could be also zero when time begins to flow. So, the initial state is given and it may be equal to $(x(0), y(0), z(0)) = (x_0, y_0, 0)$.
In order get a clearer idea of the model, we underline what is the interpretation we can give to parameters. In (3.1), λ plays a double role. On the one hand it measure the ability of spreading by the virus. At the same time it is also related to the behavior of the individuals and on how many of contacts they have with each other. The other parameter is γ. It measures the rate at which infective people are no longer contagious. Before going on, we analyze more in detail what happens at the beginning of the epidemic in order to better understand the meaning of previous expressions. When the disease starts to spread, we suppose infective individuals are only a few. Moreover, we assume that nobody is already removed, so in principle everyone can be infected. This means that during the first period $x(t) \approx n$. So, the second equation can be written as:

$$y'(t) = (\lambda - \gamma)\, y(t). \tag{3.2}$$

Equation (3.2) is a linear ODE with constant coefficients. We are able to solve it and its solutions is an exponential. In fact:

$$y(t) \approx y(0) e^{-(\lambda-\gamma)t} = y(0) e^{\gamma(R_0-1)t},$$

where:

$$R_0 = \frac{\lambda}{\gamma}. \tag{3.3}$$

From definition in (3.3), we can derive important conclusions. In particular, we can observe the behavior of the exponential solution. In fact, if $R_0 < 1$ the disease will die off. On the contrary, if $R_0 > 1$, the number of infective people will increase exponentially, and there will be a large diffusion. The only thing that will slow down and eventually stop this growth will be herd immunity, when most of the people is not susceptible anymore. From model in (3.1), we can observe that the non-linear term $\lambda x(t) y(t)$ indicates that the spread of the disease is faster when, for the same λ, there are both more susceptibles and also more infectives. This seems quite intuitive. Roughly, if we start considering the ratio between the first and the last equations, we can obtain the following solutions.

$$\frac{dx}{dz} = -\frac{\lambda}{\gamma} x(t) = -R_0 x(t) \implies x(t) = x_0 \exp(-R_0 z(t)).$$

Now, considering a constant population, we can evaluate the number of infective as: $y(t) = n - z(t) - x(t) = n - z(t) - x_0 \exp(-R_0 z(t))$.
From this solution, we can derive two really important observations:

- $y(t)$ is a decreasing function unless $x_0 > \frac{1}{R_0}$. In fact, from the second equation:

$$y'(t) = (\lambda x(t) - \gamma) y(t) > 0 \iff \lambda x(t) - \gamma > 0 \iff x(t) > \frac{\gamma}{\lambda} = \frac{1}{R_0}.$$

  From the first equation, we can observe that the number of susceptible will be always smaller with respect to the initial condition, because of the sign of its derivative. In fact, $x(t), y(t), z(t) \geq 0$ because they represent the number (or the proportion) of individuals. The intuitive explanation is that there will be fewer and fewer susceptibles as the infection spreads and more people become infected. This is why, $y(t)$ can increase only if $x_0 > \frac{1}{R_0}$ and until $x(t)$ is larger than this threshold value. In this case we can conclude that there will be a growing epidemic. The result which formalizes the behavior of the system according to the initial condition on the number of susceptibles is referred as threshold theorem.

- From the first observation, we can derive that in any case the number of infectives tends to zero. So the number of removed $z(t)$, when $t \to \infty$ tends to z_∞, which is smaller than n. This means that in a certain future time the epidemic will stop, without that everyone is became infected, indeed z_∞ is the solution of $z = n - x_0 \exp(-R_0 z)$, where the second term is something positive.

These observations refer to the previous simple model, but they remains true also in other models as we will see in the following. In particular, the interpretation of parameters continues to be valid in other SIR models. This is the reason why, in real applications, we are so interested in finding the value of parameters.

3.2 Deterministic SIR model

As we already mentioned, a deterministic model is not very accurate to describe an epidemic diffusion process which occurs in reality. Anyway, it can be used as first approach to the phenomenon and as an useful benchmark. In fact, Theorem 2.3 ensures that, under some conditions, the stochastic solution converges to the deterministic one. Unfortunately, in real applications, we are unable to specify parameters in System (3.1), so we can't apply it directly. One of possible strategies is to use Bayesian statistics. The main goal of this section is to use some observed or simulated data in order to estimate all the parameters in the SIR model.
In this context, some simplistic assumptions need to be taken into account. First of all, we suppose that the population is closed, homogeneous and homogeneously mixed. *Closed* means that no one can leave or be added to the initial population, so we are ignoring births, deaths (not caused by the disease). The simplification is justified by the fact that the time period in which the infection takes place usually is shorter than the life time of an individual. Moreover, we are ignoring immigration and emigration processes. Even

if it is clear their impact on the dynamic, it's impossible to keep track of every single journey made by the whole population. As well as other characteristic for which we have no available data . Anyway this assumption implies that the size of the population remains **constant** during the analysis. So, even if we divide the population in Susceptible, Infective and Removed individuals and their single values change over the time, their sum is always constant. Assuming that the population is *homogeneous* indicates that every individual is identical to the others: there are not differences between them and they all act in the same way. The last simplification is that the population is *homogeneously mixed* so a contact between any two individuals in the population occurs randomly always with the same probability. This last assumption may be the most problematic one because, in this setting, an infective individual has exactly the same probability of infecting one of his familiar or an individual who lives completely in another region. Anyway, we can assume that this statement is approximately true when we consider a suitable large population. Last but not least, for very simplistic model, we decide to ignore any effect of latent periods, time varying infectivity or immunity.
First of all, since all theoretical results refers to normalized processes, in the following we denote with $x(t), y(t)$ and $z(t)$ the *proportion* of susceptibles, infectives and removed respectively in the same population. Then we suppose to know their values in different moments. We simply observe that the sum of the previous variables is always equal to one. For this reason, we decide arbitrarily to consider only the observations in the last two variables, while the first one can be found as difference. Now our final goal is to approximate the value of parameters. In particular, we are interested in estimating:

- $\boldsymbol{\lambda}$: it represents the strength of the disease, the chance of falling ill. Since, this parameter is connected to the possibility of contracting the infection when a susceptible has a contact with an infective, it should be modeled as a function of time. For this reason, a time grid can be fixed over the observed period, and then the aim is the estimation of λ_i in these different points over the grid. We can also assume for simplicity that two following points are connected with a line (linear interpolation). With this method, we are able to find the value $\lambda(t)$ for any t.

- $\boldsymbol{\gamma}$: it represents the rate of recovery. It is the reciprocal of the time needed for an Infective to move in Removed group. When we consider the Markovian case, we are assuming that the infectious period has the lack of memory property. In this situation, an infective can heal following an exponentially distribution with γ as intensity parameter.

- $\boldsymbol{\tau}$: it is an additional parameter used to introduce the noise. In fact if we try to describe a process in itself stochastic with a deterministic model, it's necessary to introduce a term that can represent the error. In the following, we explain how to choose the value of this parameter in the representative equations.

At this point, we can recall equations in (3.1) and write down the system of differential equations (ODEs) which are able to describe the process:

$$\begin{cases} x'(t) = -\lambda(t)x(t)y(t) \\ y'(t) = \lambda(t)x(t)y(t) - \gamma y(t) \\ z'(t) = \gamma y(t), \end{cases} \tag{3.4}$$

3.3 Deterministic vs stochastic models

When we have to model the spread of an infectious disease, we can follow two different strategies. The first one is the *deterministic* model as we saw in Section 3.2. Even if there are several motivations to prefer a *stochastic* approach, a deterministic model can be easier to analyze and it can be used as first instrument in order to study new phenomena. Anyway, of course, a natural strategy to define the spread of a disease is stochastic. In fact, we can state the probability that a transmission occurs, rather than declare for sure whether or not an infection will happen. In addiction, a deterministic model is based on the law of large numbers and in particular, on the assumption of mass action. However, there are processes stochastic by their-self, which don't satisfy the law of large numbers and for this reason their analysis is possible only in a stochastic setting. For instance, when we study the spread of an infectious disease in a large population, we can observe a minor outbreak, where only few individuals will ever get infected, or a major outbreak, where a deterministic positive proportion of individuals become sick by the end of the process. In this particular example, we can carry out a correct analysis only trough a stochastic model, by assigning the corresponding probabilities of the two events. Lastly, a very important advantage of using a stochastic model is about the estimation. The knowledge of uncertainty in the approximation, indeed requires a stochastic approach and an estimation is useful only when we know something about its uncertainty. On the other hand, however, a stochastic epidemic model needs to be quite simple in order to be mathematically manageable and so it won't be truly realistic. On the contrary, we can define a more complicate deterministic model and still analyze it.
In order to summarize, we underline that in a deterministic model the output is fully determined by the parameters and initial conditions, while for a stochastic one the outcome is not uniquely determined by the given input but it takes a range of possible values. In conclusion, we can state that, when its analysis is possible, a stochastic model should be preferred to a deterministic one. Anyway the two different approaches should not be seen in opposition, but as both useful instruments in order to better analyze the mechanisms of disease spread.

Chapter 4

Stochastic Epidemic Models

An alternative family of models used to study and to predict the spread of a disease is that one of Stochastic Epidemic Models. As described before, using a stochastic approach means that there is one or more random elements such that for a specific input to the model, the outcome takes a range of possible values. In this Chapter, first of all we explore more in detail a very basic example of Stochastic model used to describe a very simple epidemic process. Although this model contains several simplifications, it is a very intuitive example of stochastic models. Then we introduce some fundamental concepts for the stochastic SIR, different from the principles presented in Section 3.2. In addiction, we analyze an alternative construction for the stochastic SIR. This one is really useful in Section 4.2.3, where we present some fundamental theoretical results. Finally, in the last section, the model used for the simulation part, again with a MCMC algorithm, is described.

4.1 The Reed-Frost model

Lowell Reed and Wade Hampton Frost considered in their model a *discrete time dynamic*. As a first approach and to simplify the dynamics, the length of the infectious period is considered deterministic. In this context, the infectious period is shorter than and preceded by a latent period. According to this hypothesis, we assume that a contact between two individuals is independent from all the others contacts. This is the reason why, we can suppose that new infections will occur in generations and these generations are separated by the latent period as the discrete time unit. In this model, when an individual is susceptible to the infection, then he may become infected and he is firstly infective (when he is able to infect other people) and then removed (when he recovers and becomes immune). We denote with the term "infected" an individual who is infectious or removed, so when he comes into contact with the disease. The main idea introduced by Reed and Frost is that the process, in a stochastic setting, can be studied using a Markov chain. This means that the probabilities for the future in a specific time (or generation) only depend on the state of the epidemic in the previous time and not on all the previous history. Let's try to understand more in detail this idea. We assume to consider a chain

whose states are the numbers of susceptible denoted by X_j, and infective individuals Y_j, in any time j. Of course, the assumption of the Markov chain produces a really important simplification, indeed using only these data, we are able to evaluate the probabilities for infective and susceptibles in the following step, at time $j+1$:

$$\mathbb{P}(Y_{j+1} = y_{j+1} | X_0 = x_0, Y_0 = y_0, ..., X_j = x_j, Y_j = y_j)$$
$$= \mathbb{P}(Y_{j+1} = y_{j+1} | X_j = x_j, Y_j = y_j) = \binom{x_j}{y_{j+1}} (1 - q^{y_j})^{y_{j+1}} (q^{y_j})^{x_j - y_{j+1}}.$$

and at the same time, the number of susceptibles is simply given by $X_{j+1} = X_j - Y_{j+1}$. Here we assumed that the probability of avoiding the infection from a particular infective is *independent* from all the others and it is equal to q. The previous formula states that, given the number of susceptible and infective at time j, say x_j and y_j respectively, we can evaluate the probability of having y_{j+1} infective in the next time. In particular, this event is equal to the probability of getting y_{j+1} individuals who become ill, while the remaining part of the population is still susceptible. The first event is given by the binomial coefficient associated to the different ways in which y_{j+1} individuals can be chosen between x_j susceptible, multiplied by the probability of **not** avoiding the disease from each infective: $(1 - q^{y_j})$ for each individual who becomes infective. The second probability instead, is the chance not to get in contact with any infective for the remaining part of susceptible.
From this evaluation, we can observe that new infections can occur only if there is some individual that is infective in the population. In fact, $Y_j = 0$ implies that at time $j+1$, the number of infective is necessary equal to zero. This is due to the fact that, when $y_j = 0$, for any y_{j+1} the previous probability is zero, except when also y_{j+1} is null. So, a really important result is that the number of possible chains is *finite* because the length of the Markov chain cannot be longer than the total number of infected.
Now, we discuss the probability of the *complete chain*, which means observing the chain until the epidemic stops, that is until there aren't more infectives. Given this defining equation and the initial state ($x_0 = n$ and $y_0 = m$), thanks to the Markov property, it can be evaluated by conditioning sequentially. Here we consider $y_1, y_2, ..., y_k, y_{k+1} = 0$:

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, ..., Y_k = y_k, Y_{k+1} = 0 | X_0 = n, Y_0 = m)$$
$$= \mathbb{P}(Y_1 = y_1 | X_0 = n, Y_0 = m) \mathbb{P}(Y_2 = y_2 | X_1 = x_1, Y_0 = y_1) ...$$
$$... \ \mathbb{P}(Y_{k+1} = 0 | X_k = x_k, Y_k = y_k)$$
$$= \binom{n}{y_1} (1 - q^m)^{y_1} (q^m)^{n - y_1} \binom{x_1}{y_2} (1 - q^{y_1})^{y_2} (q^{y_1})^{x_1 - y_2} ... \binom{x_k}{0} (1 - q^{y_k})^0 (q^{y_k})^{x_k}.$$

The same formula can be used also to evaluate the final size of the epidemic, which means the total number of individuals who became infected during the spread of the disease, so excluding the initial infected introduced in the population: $Z = \sum_{j \geq 1} Y_j$. To evaluate the probability of having the final size equal to z, we simply sum all the previous probabilities where $|y| = \sum_{j \geq 1}^{k} y_j = z$.

$$\mathbb{P}(Z = z|X_0 = n, Y_0 = m) = \sum_{y:|y|=z} \mathbb{P}(Y_1 = y_1, ..., Y_k = y_k, Y_{k+1} = 0|X_0 = n, Y_0 = m).$$

Of course, the previous formula becomes very complicated, even for moderate sized groups. For this reason, we need to introduce some approximation techniques, when we are dealing with large communities. The main important result in this field, is the threshold limit theorem where a leading role is played by the *basic reproduction number* R_0. This is a function of the model and, in this simple situation, it represents the average number of infective caused by a typical infective during the early stages of the epidemic. The theorem says that a *major outbreak*, which is equivalent to require that a deterministic proportion of the population will be infected by the end of the epidemic, is possible if and only if $R_0 > 1$. In general, from a statistical point of view, only this case is considered since minor outbreaks aren't observed so much.

4.2 Stochastic SIR model

In this Section, we analyze more in detail the stochastic SIR model used to represent the spread of an infectious disease. As in the deterministic case, we suppose that the population is *closed, homogeneous* and *homogeneously mixed*. Moreover, to study the spread of this infection, we also need to know the initial conditions, so we fix m, the number of infectious individuals (who just became infected) and n, the number of susceptible individuals. At this point, we have to define the dynamics of the process, that means in which way the sizes of different classes (Susceptibles, Infective and Removed) can change. The size of the susceptible group can only decrease, in fact, in this model, there is no way for an individual to return susceptible to the disease once he left the class. There is a single way in which an individual leaves the susceptible class: becoming infected. An infection happens with a certain given probability when there is a contact between an infective and a susceptible. Suppose that, during his infectious period, an infective individual comes into contact with a given person, following the time points of a time homogeneous Poisson process with intensity $\frac{\lambda}{n}$, where λ is a positive parameter. The choice of the distribution's parameter is not random, but it is chosen in order to analyze the process without considering the population size. Indeed at the beginning, the susceptible individuals are n, so the rate of getting in touch with a given person is simple equal to λ multiplied by the number of infective. All Poisson processes describing the contacts are independent of each other and also of the infectious periods. If the individual who has just been contacted is still susceptible, then he become infected and he is instantaneously able to infect in turn others susceptibles. Once an individual became infected, his infectious period is independent and identically distributed with respect to all the other periods, following the given distribution of a random variable I, and we denote with ι and σ^2 its mean and variance. A very common choice for the distribution of variable I is the exponential one because of its really important mathematical properties such as the memory-less one, which is necessary if our aim is to build a Markov chain. Lastly, when the infectious period is over, the individual is considered removed and he plays no further role into the spread process. In this model, we don't consider the distinction between recovered and death because of the infection, these two kind of individuals are both considered removed. The epidemic

ends when there are no longer infective individuals into the population. In fact, when there are no infectives, the probability of infection for a susceptible person is null and the disease is no longer able to spread.
The model which considers all the previous assumptions is called **standard SIR epidemic model** and it is denoted by $E_{n,m}(\lambda, I)$. To better understand the behavior of the epidemic, we can introduce two quantities, which are widespread in epidemiological studies. The *final size of the epidemic* Z corresponds to the total number of initial susceptibles that become infected during the spread of the disease. It is a finite random variable which can assume values into the range $[0, n]$. It can be proved that we are able to obtain a linear system to find its correct distribution. Another quantity which is really important for the analysis is the *basic reproduction number* R_0. In the situations where the population is homogeneous, the variable is simply defined as the expected number of infections caused by one infectious individual, when there is a large susceptible population. In this simple case, $R_0 = \lambda\iota$, because ι is the average length of infectious period and at the beginning a susceptible has a probability equals to λ to enter in contact with the infection. The basic reproduction number plays a very critical role because of the threshold theorem and it indicates whether a large outbreak may occur or not.
Another possible construction for the model and a special case of SIR model will be presented below. Finally, the Section ends with some important results about the final size of the epidemic when a large population is considered.

4.2.1 The Sellke construction

The standard SIR epidemic model can be built also following a different and more elegant strategy. The core idea, introduced by Sellke in 1983, is that each individual is characterized by a certain critical level of *exposure to infection*. Only if and when this level is exceeded, the individual become infected. This subjective level is denoted as the threshold of the individual, while our analysis is focused on the total infection pressure generated by the infectious individuals. This construction simply reflects some mathematical ideas and not a characteristic in the spread of the disease in the real-life, anyway it is really useful to better understand the mechanism of the model. In fact, in the following, we prove that with this construction, one can model exactly the same dynamics as described so far. To follow this idea and build the model, we introduce:

- $Q_1, Q_2, ..., Q_n$ is a set of independent and identically distributed exponential random variables, with mean equals to 1. The subscripts $1, 2, ..., n$ indicate the initial susceptibles and the variables represent individual thresholds, so when the total pressure is greater than this level, the individual denoted by i turns infective.

- $I_{-(m-1)}, I_{-(m-2)}, ..., I_0, I_1, ..., I_n$ is a set of independent and identically distributed random variables, each one following the distribution of I, where $-(m-1), -(m-2)...0$ are the initial infectives (at the beginning, we introduce m infected individuals). The infective labeled with i remains infectious for a time I_i and then he is removed.

- $A(t)$ is the total infection pressure on a given susceptible up to time t. If we denote with $Y(t)$ the number of infectives at time t, the total pressure can be evaluated as:

$$A(t) = \frac{\lambda}{n}\int_0^t Y(u)du$$

From this definition, we note that the slope of the total pressure is proportional to to the number of infectives $Y(t)$, this means that it's proportional to the number of infectious periods which covers the time t.

Given these objects, actually we can describe the spread of the disease. A susceptible denoted with i become infected when $A(t)$ reaches Q_i. Meanwhile, the j-th infective individual (we denote with $Q_{(j)}$ the order statistic for his infectious period) remains infected for a time I_j and finally he is removed. The infection continues until there are no longer infectious individual in the population.
The model built using this construction is equivalent to the standard SIR model, we have previously defined. In order to prove this equivalence, given a certain time t, we suppose that $Y(t) = y$ and that the individual labeled with i is still susceptible. We want to prove that he become infected in the interval $(t, t+\Delta t)$ with probability $\frac{\lambda}{n}y\Delta t + o(\Delta t)$, because of the superposition of y independent Poisson processes with parameter $\frac{\lambda}{n}$, following the previous construction. We start from the lack-of-memory property of Q_i and we evaluate the complementary event, so the probability that the i-th susceptible won't become infected. In the following evaluation, we use the fact that we are considering an exponential variable and in a small enough increment, no infection will occur, so $Y(u)$ inside the integral is constant and it is equal to y:

$$\mathbb{P}(Q_i > A(t+\Delta t)|Q_i > A(t)) = \mathbb{P}(Q_i > A(t+\Delta t)-A(t)) = 1-\mathbb{P}(Q_i \leq A(t+\Delta t)-A(t)) =$$
$$= e^{-[A(t+\Delta t)-A(t)]} = \exp\left(-\frac{\lambda}{n}y\Delta t + o(\Delta t)\right) = 1 - \frac{\lambda}{n}y\Delta t + o(\Delta t).$$

This conclude our demonstration because the probability of the complementary event is given exactly by the same formula.
With this kind of construction, it's easy to derive a linear (triangular) system of equations for $P^n = (P_0^n, P_1^n, ..., P_n^n)$, where P_k^n is the probability that k of the initial n susceptibles get infected during the epidemic. In fact, in this context, we are able to find the exact results. First of all, we express

$$Z = min\left\{i : \mathcal{Q}_{(i+1)} > \frac{\lambda}{n}\sum_{j=-(m-1)}^{i} I_j\right\}, \qquad A := A(\infty) = \frac{\lambda}{n}\int_0^\infty Y(u)du = \frac{\lambda}{n}\sum_{j=-(m-1)}^{Z} I_j,$$

where the first equivalence is because $\mathcal{Q}_{(i)}$ denotes the order statistics, so the epidemic will stop when the total pressure, exercised by all the infective individuals, is insufficient to make another susceptible sick. While, for the total pressure, the last equivalence arises from the fact that the function we are interested in integrating is piecewise linear, so the integral is simply the sum of the infectious period, until the epidemic stops. At this point, we can introduce the following

Theorem 4.1 *In the standard SIR epidemic $E_{n,m}(\lambda, I)$, the probability that the final size of the epidemic is equal to k, denoted by P_k^n, where $0 \leq k \leq n$, is given by:*

$$\sum_{k=0}^{\ell} \binom{n-k}{\ell-k} \frac{P_k^n}{[\phi(\lambda(n-\ell)/n)]^{k+m}} = \binom{n}{\ell} \qquad \text{where } 0 \leq \ell \leq n,$$

and $\phi(\theta) = E(exp(-\theta I))$ is the Laplace transform of I.

This theorem provide a triangular system of equations, so the final-size probabilities can be solved recursively. Another important observation we can derive from the proof of Theorem 4.1, is that the probabilities depend only on the infection pressure. So, even if we introduce latent period or time-dependent infectious rates, we will get the same results. Even though this result is really important because it provides a closed formula for the final size of the epidemic, in many situation we are keen to find an approximation to the whole process and not to understand the behavior only for its final part. This is why in the following, we will introduce other methods and important results about approximation.

4.2.2 Markovian case

A fundamental and very common SIR model is the *general stochastic epidemic* one introduced for the first time by Bartlett in 1949. This model considers the standard SIR model $E_{n,m}(\lambda, I)$, where the distribution of I is exponential with parameter γ. Actually, this assumption is not justified by an epidemiological point of view, although the analysis of the process will be simpler from a mathematical prospective. In fact, under this assumption, the infectious periods have the lack-of-memory property, because of exponential characteristics. This is why, if we indicate with $X(t)$ the number of susceptibles at time t and with $Y(t)$ the number of infectives at the same time, the process $(X, Y) = \{(X(t), Y(t)); t \geq 0\}$ is a **Markov process**. In this situation, we can easily evaluate the transition probabilities of the chain. If we start from state (i, j), which means that in the population we are considering in a certain time, there are i individuals susceptibles and j infective individuals (the remaining part is made up of those individual that are removed who don't play any further role in the spread of the epidemic), only two different states can be reached. The first one can be visited by the chain when an infection occurs: a susceptible becomes infected with probability $\lambda ij/n$. The other possible situation is when an infective is removed (because of healing or death), this can happen with probability γj, since it is a superposition of Poisson processes. To summarize, the rules determining the evolution of the process are contained in the scheme in Figure 4.1. This special case of stochastic epidemic model is modeled in the following part.

4.2.3 The threshold limit theorem

Before describing the specific model we used in the application, in this Section we describe again some fundamental proprieties of stochastic models. In fact, usually we are not only interested in describing the way the disease behaves at the end of the epidemic, but rather in modeling its entire spread. In order to provide a more detailed process description, we

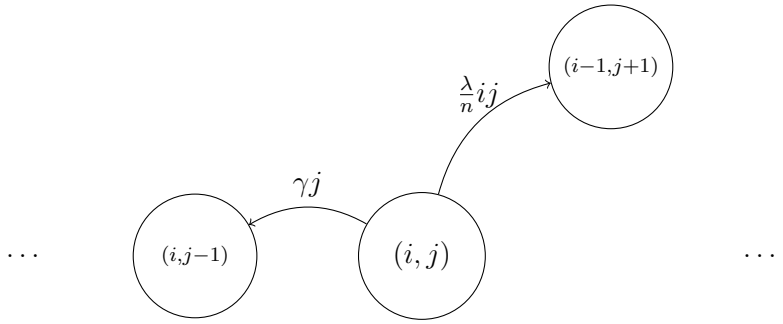


Figure 4.1: Sketchy representation of possible developments of Markov chain for the general stochastic epidemic in state (i, j)

consider again the standard SIR epidemic model $E_{n,m}(\lambda, I)$ described in the previous section and we determine the limit of the final size distribution, for large population. From different experiments, we can conclude that the epidemic has a different behavior according with the value that variable R_0 assumes. With the view to achieving this result in a formal way, we start again from Sellke construction and we need to introduce two new processes, where Q_j denote the individual thresholds and I_j are the infectious periods:

– the infection pressure process:

$$\mathcal{I}(t) = \frac{\lambda}{n} \sum_{j=-(m-1)}^{[t]-m} I_j, \qquad 0 \leq t \leq n+m$$

– the threshold process:

$$\mathcal{Q}(t) = \sum_{j=1}^{n} \mathbf{1}_{\{Q_j \leq t\}}, \qquad t \geq 0$$

Now we can express the final size of the epidemic in function of these processes. Recalling the definition, we know that $Z = i$ if and only if

$$\begin{aligned} \mathcal{Q}_{(1)} &\leq \mathcal{I}(0+m), \\ \mathcal{Q}_{(2)} &\leq \mathcal{I}(1+m) \\ &\vdots \\ \mathcal{Q}_{(i)} &\leq \mathcal{I}(i-1+m), \\ \mathcal{Q}_{(i+1)} &> \mathcal{I}(i+m). \end{aligned}$$

Here we denoted $\mathcal{Q}_{(i)}$ the ordered individual thresholds. This means that the total infectious pressure exercised by the number of infectives (i caused by the diffusion and the

m initial infectives) is not sufficient to infect one more susceptible individual. Expressing the same concept from another point of view:

$$\begin{aligned}
\mathcal{Q}(\mathcal{I}(0+m)) &> 0,\\
\mathcal{Q}(\mathcal{I}(1+m)) &> 1,\\
&\vdots\\
\mathcal{Q}(\mathcal{I}(i-1+m)) &> i-1,\\
\mathcal{Q}(\mathcal{I}(i+m)) &\leq i.
\end{aligned}$$

This means that the pressure $\mathcal{I}(0+m)$ allows the infection of at least one individual, but the pressure $\mathcal{I}(i+m)$ is not enough to infect i individuals. At this point, we are able to write the final size of the epidemic as:

$$Z = min\{t \geq 0 : \mathcal{Q}(\mathcal{I}(t+m)) = t\}. \tag{4.1}$$

To study the *convergence* of this variable as the total population gets larger, we consider a sequence of epidemic processes $E_{n,m}(\lambda, I), n \geq 1$. In order to get a more compact notation we introduce new variables and previously results which can be proved, where convergence is in probability and uniformly on compact sets:

$$\begin{aligned}
\bar{\mathcal{I}}_n(t) &= \mathcal{I}_n(nt) \quad \text{and} \quad \bar{\mathcal{I}}_n(t) \to \lambda\iota t,\\
\bar{\mathcal{Q}}_n(t) &= \frac{\mathcal{Q}_n(t)}{n} \quad \text{and} \quad \bar{\mathcal{Q}}_n(t) \to 1 - e^{-t},\\
\bar{\mathcal{C}}_n(t) &= \bar{\mathcal{Q}}_n(\bar{\mathcal{I}}_n(t)) \quad \text{and} \quad \bar{\mathcal{C}}_n(t) \to 1 - e^{-\lambda\iota t},
\end{aligned}$$

At this point, we can consider a special case of initial conditions. We suppose that $m_n/n \to \mu > 0$ as $n \to \infty$. Manipulating definition 4.1, we can consider:

$$\frac{Z_n}{n} = \frac{1}{n} \min\left\{t \geq 0 : \bar{\mathcal{C}}_n\left(\frac{t}{n} + \frac{m_n}{n}\right) = \frac{t}{n}\right\}.$$

Now we can work with $Z_n^{'} = Z_n + m_n$. This parameter includes in the count also the initial infectives. It can be proved that $\bar{Z}_n^{'} = Z_n^{'}/n$ converges to τ, which is the unique solution to the equation:

$$1 - e^{-\lambda\iota\tau} = \tau - \mu.$$

This equation can be rewritten as:

$$1 - \tau + \mu = e^{-\lambda\iota\tau}, \tag{4.2}$$

where the two sides of the equation both represent the probability of escaping of the infection. In fact this probability is given by the proportion of initial susceptibles how remain uninfected during the diffusion (left side):

$$\frac{1}{n}(m_n + n - y) = \frac{m_n}{n} + 1 - \frac{y}{n} = 1 + \frac{m_n}{n} - \bar{Z}_n \xrightarrow[n\to+\infty]{} 1 + \mu - \tau.$$

On the contrary, if we consider the right side, we can observe that the probability of avoid to become infected is:

$$\left(\mathbb{E}(\exp^{-\frac{\lambda\iota}{n}})\right)^{Z_n^{'}} \approx \left(1-\frac{\lambda\iota}{n}\right)^{n\tau} \approx \exp^{-\lambda\iota\tau}.$$

The following theorem gives a more detailed result.

Theorem 4.2 *Consider a sequence of standard SIR epidemic $E_{n,m_n}(\lambda, I)$, where $m_n/n \to \mu > 0$ as $n \to \infty, \tau$ is the solution of 4.2, $Z_n^{'}$ is the final epidemic proportion as before and $\rho = 1-\tau+\mu = e^{-\lambda\iota\tau}$.*
Then the sequence $\sqrt{n}(Z_n^{'}/n-\tau)$ converges to a normally distributed random variable with mean 0 and variance

$$\frac{\rho(1-\rho)+\lambda^2\sigma^2\tau\rho^2}{(1-\lambda\iota\rho)^2}.$$

This is the final result we expected to find. In fact when R_0 exceed 1, it is reasonable to think that the final epidemic size will satisfy a law of large number.

4.3 Density dependent Stochastic SIR

The main goal of this section is applying the results exposed in Section 2.4 to a special case of epidemic model. We focus of the Markov Stochastic SIR denoted by $E_{n,\mu n}(\lambda, I)$, where I is exponentially distributed with intensity γ. We highlight the fact that with the approximations we found in that Chapter, we are able to approximate the whole epidemic process and not only the final size of the epidemic. Nevertheless we are going to apply the Theorem 2.4. This implies that the approximation is valid only when the number of infectious individuals is quite large. So, we have to not consider the initial or last phase of the epidemic. In this context, we can use different methods such that coupling methods. In the following, we denote with $X_n(t)$ and $Y_n(t)$ the number of susceptibles and infectives at time$t$. While $Z_n = (X_n, Y_n)$ is the two-dimensional process we are going to approximate. Finally, we assume that the initial values are $X_n(0) = n$ and $Y_n(0) = \mu n$, where μ is assumed positive, but usually very small.
In this particular kind of process, there are only two different kind of jumps allowed:

- $\ell_1 = (-1, 1) \Rightarrow$ a susceptible becomes infective:

$$q^{(n)}_{z,z+\ell_1} = n\beta_{\ell_1}\left(\frac{z}{n}\right) = \frac{\lambda}{n}xy \qquad \Rightarrow \qquad \beta_{\ell_1}(\bar{x},\bar{y}) = \lambda\bar{x}\bar{y};$$

- $\ell_2 = (0, -1) \Rightarrow$ an infected is removed.

$$q^{(n)}_{z,z+\ell_2} = n\beta_{\ell_2}\left(\frac{z}{n}\right) = \gamma y \qquad \Rightarrow \qquad \beta_{\ell_2}(\bar{x},\bar{y}) = \gamma\bar{y}.$$

In the previous formulas, we denoted by $\bar{x} = \dfrac{X_n(t)}{n}$ and $\bar{y} = \dfrac{Y_n(t)}{n}$.
So, in order to sum up, we stated that the process is characterized by the following jump

intensities:

$$\mathbb{P}((X_n(t+h), Y_n(t+h)) = (x-1, y+1) | (X_n(t), Y_n(t)) = (x,y)) = hn\beta_{\ell_1}(\bar{x}, \bar{y}) + o(h),$$
$$\mathbb{P}((X_n(t+h), Y_n(t+h)) = (x, y-1) | (X_n(t), Y_n(t)) = (x,y)) = hn\beta_{\ell_2}(\bar{x}, \bar{y}) + o(h).$$

Moreover, the drift function in this specific case is defined as:

$$F(x,y) = \sum_{\ell} \ell\beta_\ell(x,y) = (-\lambda xy, \lambda xy - \gamma y).$$

At this point, we are able to evaluate the deterministic solution $z = (x, y)$ to the integral equation defined in (2.17). This last one is equivalent to the following pair of differential equations:

$$\begin{aligned} x'(t) &= -\lambda x(t)y(t), & x(0) &= 1, \\ y'(t) &= \lambda x(t)y(t) - \gamma y(t), & y(0) &= \mu. \end{aligned}$$

This is exactly the same system proposed by Kermack-McKendrick. In Section 3.1, we already evaluated a parametric solution to the previous system:

$$\begin{aligned} x(t) &= x(0)\exp(-R_0 z(t)) = e^{-R_0 z(t)} \\ y(t) &= 1 + y(0) - z(t) - x(t) = 1 + \mu - z(t) - e^{-R_0 z(t)}, \end{aligned}$$

where $z(t)$ is the solution of the implicit differential equation $z'(t) = \gamma y(t)$ with initial value $z(0) = 0$.
Now all the elements are available in order to apply the Theorem 2.3. Our aim is to consider the process $\bar{Z}_n = \left(\frac{X_n}{n}, \frac{Y_n}{n}\right)$ and we would like to prove that it converges to the deterministic solution just evaluated. With this aim in mind, we have to prove the hypothesis.

1. **Convergence of initial conditions:**

   $\bar{X}(0) = \dfrac{X_n(0)}{n} = \dfrac{n}{n} = 1 = x(0)$,

   $\bar{Y}(0) = \dfrac{Y_n(0)}{n} = \dfrac{\mu n}{n} = \mu = y(0)$. So easily the first assumption is valid (this is a punctual equality and not just a limit equation).

2. **Lipshitz function:**

   To prove Assumption 2.3, we consider any two sates: (x_1, y_1) and (x_2, y_2). Since the previous definitions, we know that $0 \le x_1, x_2, y_1, y_2 \le 1 + \mu$. In order to find a suitable bound, we recall the elementary inequality:

   $$(a^2 + b^2) \ge 0 \Rightarrow ab \le \frac{1}{2}(a^2 + b^2) \qquad \forall a, b \in \mathbb{R}.$$

   Then it can be proved that:

   $$|F(x_1, y_1) - F(x_2, y_2)| \le (2\lambda(1+\mu) + \gamma)|(x_1, y_1) - (x_2, y_2)|. \tag{4.3}$$

Even if the estimation in (4.3) is quite rough, all the assumptions remain proved. Thus it states that the process $(\bar{X}_n(t), \bar{Y}_n(t))$ converges almost surely to $(x(t), y(t))$, uniformly on bounded intervals.
Last but not least, we would like to apply the Central Limit Theorem in order to prove that there are asymptotically Gaussian fluctuations around the deterministic solution $(x(t), y(t))$. To this aim, firstly we define:

$$\tilde{X}_n(t) = \sqrt{n}(\bar{X}_n(t) - x(t)) \qquad \text{and} \qquad \tilde{Y}_n(t) = \sqrt{n}(\bar{Y}_n(t) - y(t)),$$

and we consider $V_n = (\tilde{X}_n, \tilde{Y}_n)$. Then we evaluate the matrix of partial derivatives and matrix L involved in the theorem:

$$\partial F(x, y) = \begin{pmatrix} -\lambda y & -\lambda x \\ \lambda y & \lambda x - \gamma \end{pmatrix},$$

$$L(x, y) = \begin{pmatrix} \lambda xy & -\lambda xy \\ -\lambda xy & \lambda xy + \gamma y \end{pmatrix}.$$

Since we observe that ∂F is continuous and that $V_n(0) = (\tilde{X}_n(0), \tilde{Y}_n(0)) = (0, 0), \forall n$, the Theorem 2.4 can be applied. Thus it remains proved that the two-dimensional process V_n converges to a Gaussian process V. The covariance function Φ is defined by the following independent pairs of differential equation, where $\Phi_{11}(s, s) = \Phi_{22}(s, s) = 1, \Phi_{12}(s, s) = \Phi_{21}(s, s) = 1$ and the derivatives are defined with respect to s:

$$\begin{pmatrix} \Phi'_{11}(t, s) & \Phi'_{12}(t, s) \\ \Phi'_{21}(t, s) & \Phi'_{22}(t, s) \end{pmatrix} = - \begin{pmatrix} \Phi_{11}(t, s) & \Phi_{12}(t, s) \\ \Phi_{21}(t, s) & \Phi_{22}(t, s) \end{pmatrix} \begin{pmatrix} \lambda y(s) & -\lambda x(s) \\ -\lambda y(s) & \lambda x(s) + \gamma \end{pmatrix}.$$

These solutions haven't an explicit formulation, but they can be derived for the previous equations and they are used in the evaluation of covariance for the process.

Part II

Results

The aim of this second part is applying the theoretical knowledge exposed so far in this thesis to a real world application. Our case-study concerns the COVID-19 epidemic and in particular a short time window during the first wave, in a specific region in Italy, Piemonte. Before everything else, in the following Section, we explore what the term "infectious disease" really means and we also provide a briefly description of the main characteristics of COVID-19. Then we comment and discuss in detail which strategies were implemented in the simulations.
Prior to presenting all the achieved results, it is useful to make a few general remarks. In all the following situations, we use a MCMC algorithm with a Gibbs step. Our final goal is the estimation of all the parameters involved in different models. These results will be helpful in order to understand how the evolution of this particular disease works and also to make some predictions. Moreover all the following simulations were carried out by programming in the **R language** and using its corresponding Software.
In order to apply the different models proposed in the previous part and more importantly to test them, our first approach was focused on simulated data. This was really useful because knowing the real value of parameters allows to test the model and understand how reliable our predictions are. In order to use simulated data, we need to define the parameters we are interested in and then randomly generate some values. They are used instead of real observations. The data generation can be done, following two different paths. The first one is via a system of deterministic equations. The other one is based on stochastic differential equations. In the next Chapter, we analyze both in detail. Instead, in Chapter 7 we don't need to generate random observations. In fact we can consider directly real data published by Italian Protezione Civile during the emergency period. In the following, the structure of data is presented.
Once the matrix containing all data is ready, we have to estimate different characteristic parameters of the model. We underline that the two different alternatives presented above are both useful to better understand the evolution of the system. A the triplet of ODEs can be defined if we are using a *deterministic model*. An alternative method is to use stochastic differential equations and a stochastic SIR. Both methods will be explored in the next Chapters.
The final part draws conclusions, compares the results obtained and also leaves some space for possible modifications and improvements.

Chapter 5

Infectious disease

Infectious diseases should be caused by a wide range of pathogens, in particular bacteria or viruses. Many of them live in and on humans' bodies and they are normally harmless or even helpful. Anyway, under certain conditions, bacteria and viruses may give rise to problems and cause diseases. Among infectious diseases, we can distinguish a particular group which is characterized by a person-to-person transmission mechanism. This kind of disease is said to be contagious. In this thesis, we are interested only in contagious diseases and some practical examples are chickenpox, measles, mumps, sexually transmitted diseases, influenza, common cold and also the most recent COVID-19. Contagious diseases may spread through the population with two different processes. The first one is through direct physical contact, like touching or kissing an infective individual. Alternatively, an infection occurs when an infectious microbe travels through the air after someone nearby sneezes, coughs or simply breath.
Before understanding how to model this kind of illness, we have to clarify the distinction between different terms: *endemic, epidemic* and *pandemic*, [9]. The first one can be used to indicate a disease whose outbreak is consistently present, but limited to a particular area and for which the spread and rates are predictable. Nowadays for instance, in certain countries, malaria is considered endemic. On the other hand, an epidemic has an unexpected increase in the number of infective in a specific geographical area. Polio and smallpox are two examples of epidemic. Finally, when a disease's growth is exponential, the World Health Organization (WHO) declares a pandemic. This means that the spread of a virus takes place in several countries and populations. Thus, the main difference between an epidemic and a pandemic doesn't lie in the severity of the illness, but in the degree to which it spreads. In particular, during a pandemic, the disease is widespread across international boundaries, unlike regional epidemics. A very important and present-day example, is COVID-19. In fact, firstly the epidemic of Coronavirus was declared in China, but on 11th March 2020, the WHO defined COVID-19 a pandemic, because of its diffusion all over the world.
When we are interested into describing a contagious disease, we can state that all models have two main factors in common:

- *Strong dependency among different individuals in the population.* The event of getting the infection for a specific person is strongly related to the status of people in his proximity. In fact, when a large number of individuals in his neighborhood is sick, the probability to get infected for him, of course, is higher. This is the reason why, the model becomes really complex, even if we are considering a small group of people.

- *Not completely observed process.* Actually, in general, we are not able to observe all the steps in a contagious epidemic process. For example, it's rare that we know the exact time when an individual becomes infected or who transmitted the disease to whom.

For this reason, when we decide to describe a contagious disease, regardless of the model used, we have to keep these issues in mind and find a way to manage them.

5.1 COVID-19

A particular kind of contagious infection, we subsequently consider in this thesis, is the Coronavirus disease 2019 (COVID-19). Our references in the following are [10], [11], and [12]. Coronaviruses are a type of virus including different kinds of infections. One of them, first identified in 2019, is SARS-CoV-2, which caused the so called COVID-19 pandemic. SARS stands for Severe Acute Respiratory Syndrome, which is one of the main consequences when you contract the virus. The reason for its name is that the coronavirus caused COVID-19 pandemic is very related to that one responsible of the first SARS outbreak in 2003, when several people were affected in different countries by severe respiratory problems.
Symptoms of COVID-19 depend on the individual, but the most common are fever or chills, cough, headache, breathing difficulties, loss of smell and taste. Some people infected may have no symptoms at all (although they are however capable of transmitting the virus), but in other situations, COVID-19 can lead to respiratory or kidney failure, lasting lung and heart muscle damage, nervous system problems or death. Usually the symptoms appear 2 to 14 days after the exposure and an infective is contagious to others for up to two days before symptoms appear, remaining contagious for 10 to 20 days [13]. We have to underline that these estimations are true for the first wave. Indeed during different waves and variants there were be several changes in the duration of infection, symptoms and also among those who were most likely to become infected [14].
A case of COVID-19 was identified for the first time in Wuhan, China, in December 2019 and since then, the disease has spread all over the world, giving rise to an ongoing pandemic. SARS-CoV-2 is thought to spread from person to person through *droplets* (secretions emitted by breath) released when an infective person coughs, sneezes, talks or just breath. The diffusion may also take place by touching a surface with the virus on it and then touching one's mouth, nose, or eyes, but this is less common.
COVID-19 is diagnosed through a test, in particular, there are two different kinds of tests: *viral test* and *antibody test*. The first one is able to tell if you are currently infected, while the latter one is a blood test that can show if you were previously exposed to the virus

that causes COVID-19, and if your body has created antibodies in an attempt to defend itself. This means that the person has some level of immune protection either from prior infection or from vaccination, but it is not a guarantee that there is not a re-infection with the virus. In fact, antibody levels decrease over time, and nowadays it is not clear how long antibody protection lasts. However, for our future purposes, we will only consider viral tests to understand the number of infectives day by day. Specifically, molecular tests on nasopharyngeal and oropharyngeal respiratory samples are the mainly used during the first wave and they are still the international *gold standard* for diagnosing COVID-19 in terms of sensitivity and specificity. This kind of swabs uses the Real-time RT-PCR (Reverse Transcription-Polymerase Chain Reaction) technique which enables the presence of the viral genome. With this method it is possible to detect the viral genome not only in symptomatic individuals, but also in those whether pre-symptomatic or asymptomatic. Two different COVID-19 vaccines, based on mRNA, have been developed and approved. They play a very important role in preventing serious disease, hospitalization and death from COVID-19. There are also many others preventive measures to reduce the chances of infection. It is very recommended to stay at home, wear a mask in public, avoid crowded places, keep social distancing, ventilate indoor spaces and so on.

Chapter 6

Simulated data

In this Chapter, we explore a preliminary step. Before everything else, now the main challenge is testing our methodology. In fact, we would like to understand if and how much we can trust the results. So rather than applying the models presented in the previous part directly to real data, we use simulated data. This is useful because in this context, we know the true value of each parameter. In fact, we choose arbitrarily the value of parameters involved in each model and we use them to generate the data. Nevertheless, we pretend not to know their value and we try to estimate them. This is done with the purpose of finding out whether the method is able to identify the correct values. In the following, two different strategies are described to estimate parameters: the deterministic and the stochastic one. They are analyzed and discussed in Sections 6.2 and 6.3. Before that, however, we clarify how we generate the simulated data in Section 6.1. Two different paths can be followed. The first method consists in solving the system of ODEs in (3.4), while the other is based on the resolution of SDEs by the Euler-Maruyama scheme.

6.1 Data Generation

As first step, we *generate* some synthetic data in order to evaluate the performance of the algorithm. We denote by $\boldsymbol{X}$ the matrix containing the **proportion** of individuals in states *Infective* and *Removed* at each time t over the observed period. They are evaluated as the ratio between the total number of individuals in that state over the total fixed population. Moreover, since the sum given by Susceptibles, Infective and Removed needs to be always equal to 1, we keep track only of proportions of Infective, in the first row of the matrix, and Removed, in the second row. The proportion of Susceptibles can be obtained by difference. The number of columns of the matrix depends on the length of the time window, we decide to consider. So first of all, we have to fix the number of *days*. Since in the following we would like to analyze the first wave of Covid-19 pandemic in Italy, we decide to set this parameter at 126 in order to have comparable sizes when we use real data from late February 2020 to the end of June.
Then we define the number of parameters. In fact, for simplicity we assume that the rate of recovery γ and the error terms τ and ν, remain constant during the observed period.

On the contrary, we need to model the infection rate as a function of time. As already discussed, it is infeasible to estimate the parameter $\lambda(t)$ each day. Then we build a grid on the time window, we decide to estimate the parameters on these points, and then linearly interpolate them. The parameter used to build the grid is *step*. It represents how often (after how many days) we are going to estimate the infection rate. For reasons we will explain when we consider real data, in Section 7.2, we fix this value at 14. So at the end, the total number of parameters we have to estimate is 11, of which 9 are the $\lambda(t)$, then we have γ, and only one between τ and ν, in the deterministic and stochastic case respectively.
Matrix $\boldsymbol{X}$ can be filled using two different strategies. The first method is based on the resolution of a system of ordinary differential equations to which some measurement white-noise is added. Thus we are considering a *deterministic* approach. The system to be solved is described in (3.4). An important parameter needed in order to find the unique solution of the system is the *initial conditions*. It contains the initial proportions of individuals in each of the three categories. An insight into how these parameters are chosen can be found in Section 6.1.1.
On the other hand we can assume that the process is described by a multidimensional stochastic differential equation. If we use this second approach, we need to use the Euler-Maruyama scheme in order to simulate a trajectory describing the evolution of the process. As described in Section 2.3, also in this case, we need a vector of *initial conditions* to start with the algorithm. In addiction, we have to define the partition of the time interval. In this context, our parameter *delta* will be always fixed to 1. This is done in order to have a correspondence with real data that are collected once per day.

6.1.1 Initialization

In order to apply all the previous models, it is necessary to fix some parameters to allow the algorithm initialization. First of all, we have to decide a strategy to fix the initial conditions of the specific model we are considering (deterministic or stochastic). In fact, if we decide to generate data, we have to set also the value of the initial proportions as well as the value of the parameters. If we decide to use a deterministic approach, we have to solve a system of ODEs. For this reason, as described in Section 3.2, we have to fix the proportion of Susceptible, Infective and Removed individuals at $t = 0$. This is the time when the analysis starts. Similarly, we have to specify some initial values, when we decide to apply a stochastic model. Indeed the *ode_sol* and *euler_approx* functions (in A.1 and A.5) need a starting point in order to apply the recursive formula. The first function needs three different values: one for the initial proportion of Infective, one for the initial proportion of Removed, and the latter for the initial proportion of Susceptibles (this value can be obtained by difference to one). Instead, the second function only requires the first two values. For sake of simplicity, we decide to define only once the initial conditions and then keep them fixed in every iteration. We can set them in an arbitrary manner. A rather reasonable assumption seems to assume that the number of Removed individuals is equal to zero or very low anyway. Moreover, we know from the theory that at least one individual should be Infective so that the disease is capable of spreading. However, we are analyzing the beginning of the epidemic, so even this number should be quite small.

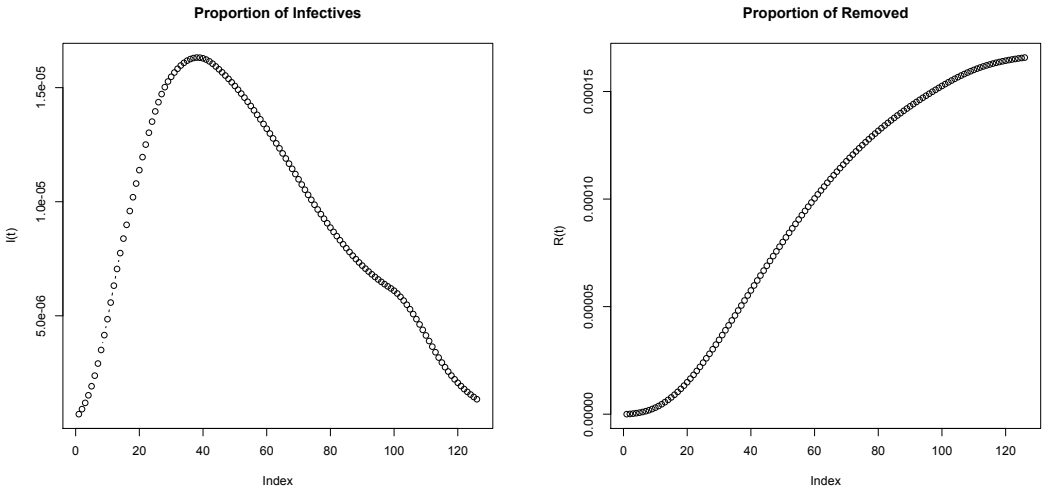
In the simulations we decide that there is no noise at the beginning. So we assume that the initial conditions can be directly extracted from the first column of available data.
Another important aspect, is the initialization and definition of parameters required by the MCMC algorithm. The Metropolis-Hastings technique requires an arbitrary initial value for each parameter we are going to estimate. This represents the starting pint of the chain and it may have an effect on the time required to achieve convergence. Moreover, in order to make a proposal for the new value of each parameter, it is necessary to choose a standard deviation. Then, this proposal can be accepted or refused according with the value of α, as described into Section 2.1.3. Just for sake of simplicity, we set the values of standard deviations (one for each parameter to be estimated) and we don't treat them as additional parameters to be estimated. Anyway, we manually adjust their values according with the *acceptance ratio*. This parameter is easily given by the ratio between how many times we accept the proposal and the total number of the iterations. According with the literature, the acceptance ratio should assume a value between 15% and 45%. The idea behind this heuristics is that our proposals should not be accepted too many or too few times. Indeed, when the standard deviation is too low, the chain's movements are too small. As a consequence, probably we will not be able to explore the whole space. On the other hand, if the standard deviation is too high, the chain moves too fast and we will not able to reach the convergence. This is the reason why we would like to reach a compromise between these two complementary requirements.
Once all parameters are initialized, we can start the estimation with the chosen method.

6.2 MCMC for deterministic SIR

In this thesis, we decided to apply the Gibbs algorithm with a Metropolis-Hastings step. At this point, we analyze more in detail the technique, when we decide to follow a deterministic approach. This means that we are going to make several proposals for the parameters in order to find their best value. It should be able to explain the data contained in $\boldsymbol{X}$. Then for each proposal, we are going to accept it according to a the corresponding posterior distribution which involves the resolution of a system of ODEs. The definition of these functions can be found in A.1.
Before starting, we define the vector of parameters we would like to estimate:

$$\boldsymbol{\theta} = (\boldsymbol{\lambda}, \gamma, \tau),$$

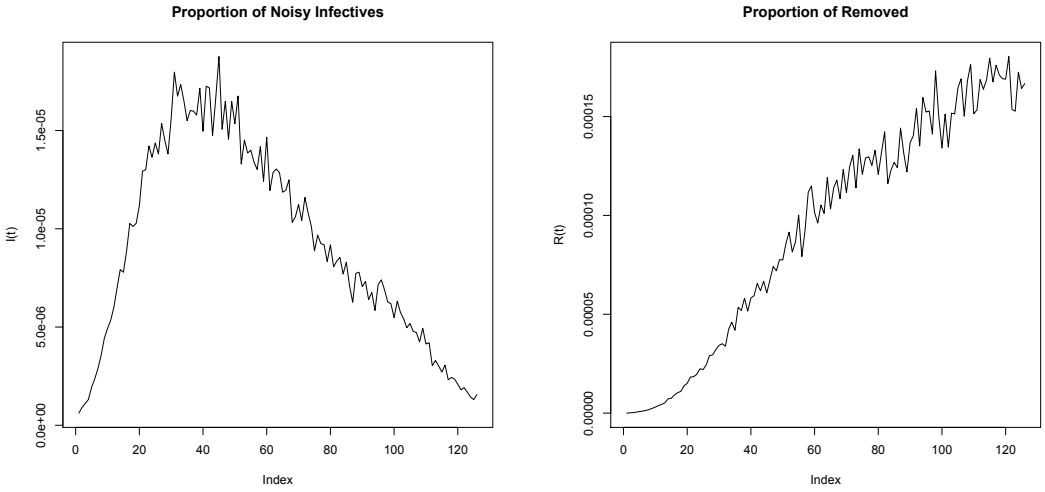
with their interpretations explained in Section 3.2. We start considering a period of time where we know for each day, the proportion of Infective and Removed individuals. We collect the daily data in the matrix $\boldsymbol{X}$, whose dimensions are $2 * d$, where d is the number of days during the period observed. In the first row, there are the proportions of infective day per day, while in the second row there are the proportions of removed. This is a key element in Bayes' theorem: the observations. Matrix $\boldsymbol{X}$ in this context is filled by deterministic data generation, as explained in A.2. In Figures 6.1a and 6.1b, the evolution in the proportion of Infective and Removed individuals for the particular choice of parameters are plotted. Then in Figures 6.2a and 6.2b some noise (with parameter τ) is added to the solution of ODEs in order to introduce some variability.



(a) Proportion of infective.

(b) Proportion of removed.

Figure 6.1: Data generation solving a system of ODEs with parameters chosen in A.2.



(a) Proportion of noisy infective.

(b) Proportion of noisy removed.

Figure 6.2: Data generation solving the same system in Figure 6.1 with noise $\tau = 200$.

Once matrix $\boldsymbol{X}$ is filled, we are ready to introduce all the elements we will need in order to apply the previous described statistical methods. We recall Section 2.1 and first of all we define the likelihood function, then the prior distributions, and finally the posterior and the full-conditional distributions to perform Gibbs sampling.

Likelihood
First of all, we define the *likelihood* function $\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta})$ in order to link the data with what we expect from the theoretic model. Since $\boldsymbol{X}$ contains proportions of individuals for several days, we decide to work on the logarithm of $\boldsymbol{X}$. In this way, no matter what, the matrix is always filled with positive values. In the following we denote by $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_d)$ the matrix of observations, one observation per day. Here, $\boldsymbol{X}_i = (I_i, R_i)$. At this point, assuming we know the values of the parameters, we can write the likelihood function. In particular, we suppose that the logarithm of $\boldsymbol{X}$ follows a normal distribution centered on logarithm of the solution of ODEs and variance $1/\tau$. In formulas, we can write:

$$\log(\boldsymbol{X}_t)|\lambda_t, \gamma, \tau \sim N\left(\log(\bar{\boldsymbol{X}}_t), \frac{1}{\tau}\right),$$
$$\mathscr{L}\left(\log(\boldsymbol{X}_t)|\lambda_t, \gamma, \tau\right) \propto \exp\left(-\frac{\tau}{2}(\log(\boldsymbol{X}_t) - \log(\bar{\boldsymbol{X}}_t))\right),$$
$$\mathscr{L}\left(\log(\boldsymbol{X})|\boldsymbol{\lambda}, \gamma, \tau\right) \propto \exp\left(-\frac{\tau}{2}\sum_{t=1}^{d}(\log(\boldsymbol{X}_t) - \log(\bar{\boldsymbol{X}}_t))\right).$$

Here, $\bar{\boldsymbol{X}}$ denotes the ODE solution given by System in (3.4). Of course to solve the previous equation, we need to propose at each iteration a specific value for the parameters. Then we substitute their values in the equations and finally we decide if we should accept or not the initial proposal. These steps are repeated several times in order to reach stationarity in the Markov chain.

Priors
At this point, we have to set the prior distributions in order to apply the Bayes' theorem. For the kind of parameters defined in the model, we want to ensure that they are positive. Unfortunately if we make a random proposal, there is a risk of considering a negative value for them. For this reason, also in this case, a possible solution could be working with the logarithm. We decide to assume that these parameters follow a **normal prior distribution**. In this way, the logarithms can assume any real value according to the normal distribution which we fix *a priori*, but $\lambda(t)$ and γ are necessary positive values. For the parameter τ instead the reasoning is different, in fact since it is a sort of precision, we can assume it follows a Gamma distribution. In conclusion, to summarize we get:

- Infection rate:

  if we consider any point over the grid we built on the time period, we assume that the parameter λ_i follows

  $$\log(\lambda_i)|m_i, s_i \sim N\left(m_i, \frac{1}{s_i}\right) \quad \Longrightarrow \quad \lambda_i|m_i, s_i \sim \log N\left(m_i, \frac{1}{s_i}\right).$$

  For sake of simplicity, we assume that $m_i = m \ \forall i \in [1, n]$ and $s_i = s \ \forall i \in [1, n]$, where n is the total number of points over the grid. So we simply assume that the infection rates follow the same distribution in any time. We can define $\boldsymbol{\lambda} = (\lambda(1), ..., \lambda(n)) = (\lambda_1, ..., \lambda_n)$. There the notion with subscripts is only for simplicity.

By the properties of a normal distribution, we can obtain that:
$\log(\boldsymbol{\lambda})|m, s \sim N(\boldsymbol{m}, T)$ where $\boldsymbol{m} = m(1, 1, ..., 1)$ and $T = diag\left(\frac{1}{s}\right)$. Hence,

$$\pi(\log(\lambda_i)|m, s) \propto \exp\left(-\frac{s}{2}(\lambda_i - m)^2\right),$$

$$\pi(\log(\boldsymbol{\lambda})|m, s) \propto \prod_{i=1}^{n} \exp\left(-\frac{s}{2}(\lambda_i - m)^2\right) = \exp\left(-\frac{s}{2}\sum_{i=1}^{n}(\lambda_i - m)^2\right).$$

We have to recall that in order to find the value of λ for each day, the previous obtained values have to be linearly interpolated. Indeed, the vector with one value per day is required by the previous log-likelihood function.

- Recovery rate:
  we can repeat the same reasoning as before and obtain the following distribution.

$$\log(\gamma)|n, r \sim N\left(n, \frac{1}{r}\right) \quad \Longrightarrow \quad \gamma|n, r \sim \log N\left(n, \frac{1}{r}\right),$$

$$\pi(\log(\gamma)|n, r) \propto \exp\left(-\frac{r}{2}(\gamma - n)^2\right).$$

- Error rate:

$$\tau|a, b \sim \Gamma(a, b),$$

$$\pi(\tau|a, b) \propto \tau^{a-1} \exp(-\tau b).$$

At the end, putting together all the components and assuming that the hyper-parameters (m, s, n, r, a, b) are given, we can obtain the prior distribution, where we replaced

$$\boldsymbol{\theta} = (\log(\boldsymbol{\lambda}), \log(\gamma), \tau)$$

.

$$\pi(\boldsymbol{\theta}) = \pi(\log(\boldsymbol{\lambda}), \log(\gamma), \tau) = \pi(\log(\boldsymbol{\lambda}))\pi(\log(\gamma))\pi(\tau) \propto$$
$$\propto \tau^{a-1} \exp\left\{-\frac{1}{2}\left[s\sum_{i=1}^{n}(\lambda_i - m)^2 + r(\gamma - n)^2 + 2\tau b\right]\right\}.$$

Finally, if we decide to work with logarithm of this function, it is not necessary to calculate the exponential. Furthermore, it is sufficient to consider the sum of the values we obtain individually with the previously defined distributions:

$$\log(\pi(\boldsymbol{\theta})) = \log(\pi(\log(\boldsymbol{\lambda}))) + \log(\pi(\log(\gamma))) + \log(\pi(\tau)).$$

Posterior
Finally, by applying Bayes' theorem, we can find an expression for the posterior distribution. This will be the distribution which we would like to sample from. The overall result

is:

$$\pi(\boldsymbol{\theta}|\log(\boldsymbol{X})) = \pi(\log(\boldsymbol{\lambda}), \log(\gamma), \tau|\log(\boldsymbol{X})) \propto \mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto$$
$$\tau^{a-1}\exp\left\{-\frac{1}{2}\left[\tau\sum_{t=1}^{d}(\log(\boldsymbol{X}_t) - \log(\bar{\boldsymbol{X}}_t)) + s\sum_{i=1}^{n}(\lambda_i - m)^2 + r(\gamma - n)^2 + 2\tau b\right]\right\},$$

$$\log(\pi(\boldsymbol{\theta}|\log(\boldsymbol{X}))) \propto \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})\pi(\boldsymbol{\theta})) = \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})) + \log(\pi(\boldsymbol{\theta})) =$$
$$= \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})) + \log(\pi(\log(\boldsymbol{\lambda}))) + \log(\pi(\log(\gamma))) + \log(\pi(\tau)). \tag{6.1}$$

Full Conditionals

The ability to sample from the posterior distribution is essential to the Bayesian statistics because it allows Monte Carlo estimation of all the parameters we are looking at. Anyway in this case (6.1) has a quite difficult form. A possible strategy consists of applying the Gibbs sample algorithm. With this technique, we define the full-conditional distributions and we sample from them. In order to obtain the full-conditionals for all the parameters, we look at the kernel of the posterior distribution and we simply ignore all the portions that not contain the variable we are interested in. We define a full-conditional for each parameter and in each step of the algorithm, we sample subsequently from these.

- Infection rate:

$$\log(\pi(\log(\boldsymbol{\lambda})|\log(\boldsymbol{X}), \log(\gamma), \tau)) = \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})) + \log(\pi(\log(\boldsymbol{\lambda}))).$$

- Recovery rate:

$$\log(\pi(\log(\gamma)|\log(\boldsymbol{X}), \log(\boldsymbol{\lambda}), \tau)) = \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})) + \log(\pi(\log(\gamma))).$$

- Error rate:

$$\log(\pi(\tau|\log(\boldsymbol{X}), \log(\boldsymbol{\lambda}), \log(\gamma))) = \log(\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta})) + \log(\pi(\tau)).$$

  It is still necessary to pay particular attention at the parameter τ. In fact we suppose that it follows a Gamma prior distribution, which isn't a symmetric one. On the other hand, we'd like to use a symmetric distribution because of simplicity in the acceptance ratio, as exposed in (2.7). For this reason, we can propose a random-walk on $\log(\tau)$ and, in order to evaluate α, we have to consider the right (symmetric) distribution. We know, $\tau \sim \Gamma(a, b) \implies \pi(\tau|a, b) \propto \tau^{a-1}\exp(-\tau b)$.
  We consider $w = \log(\tau)$ and we obtain:

$$\pi(e^w|a, b) \propto (e^w)^{a-1}\exp(-be^w)e^w = e^{aw}\exp(-be^w) = \exp(aw - be^w)$$

$$\implies \pi(w|a, b) = aw - b\exp(w). \tag{6.2}$$

All the assumptions and definitions contained in this Section are implemented in A.3.

6.2.1 Deterministic SIR results

Now, we are ready to implement this model in R. In this Section, we comment and analyze what we obtained applying this technique. The code implementation can be found in A.4. To begin with, we decide to follow a quite natural path. We use a deterministic approach to generate the data and then the resolution of a system of Ordinary Differential Equations in the MCMC algorithm.
The general idea is that at each step of the Metropolis-Hastings algorithm, we make a new proposal for each parameter. Then we use them to evaluate the posterior/full conditional distributions. Lastly, if we accept the proposal, we store in *chain* the values of proposed parameters. On the contrary, we store in *chain* their old values. The final goal is that the chain for each parameter has to reach the stationarity. In fact, we make some proposals which should gradually be concentrated on the value used in the generation. We experimentally set the total number of Metropolis-Hastings iterations. Moreover, in an attempt to eliminate some noise, correlation or computational errors, we can define an other parameter: *thin*. Thinning is a process which allows to store in our matrix $\boldsymbol{X}$ only some values. We keep only a value every *thin* iterations. This of course is useful to save memory, but also to reduce some correlation effects in the iterations. They contribute to slowing down the process of achieving stationarity. For this reason, it could be really helpful to introduce this parameter that in the simulations is set at 20. According with what explained in Section 6.1.1, we also adjust the values for each standard deviation of parameter proposals.
In Figure 6.3 all the chains of parameters are shown. Each sub-figure represents only the final part of the chain when stationarity is reached. We observe that the acceptance ratio is in the prefixed range and all the chains more or less reach the area were the value used for the simulations is. An other measurement of our performance can be Figure 6.4, where the parameter R_t is evaluated for each day. We notice that its value is really similar to that one evaluated in order to generate data. Finally, always with the same setting of parameters, we can compare the trajectories. We consider separately the first and the second row of matrix $\boldsymbol{X}$ and we evaluate the mean and two quantiles. The first with probability 0.25 and the other with probability 0.75. They are also called first and third quartiles, respectively. From Figure 6.5, we can observe that simulated data remains in the bandwidth most of the time and we get a good approximation of what we simulated. We implemented an R code which is able to evaluate the trajectories and plot all the previous results. It can be found at A.9. All the following plots are obtained by giving the results to this function.
In Figures 6.6, 6.7 and 6.8, we can observe the trend of different chains, the effective R_t and the trajectories respectively. In this case, all the parameters used for data generation maintain the same values, except for τ, which represents the noise we add to the deterministic solution. Now, its value is fixed at 80. Since it appears in a fraction, decreasing its value is equivalent to increase the added noise. As a consequence, we start from more noisy data as shown in Figure 6.8 where there are higher peaks than before. Here also the band obtained with quartiles is wider. Another approach for our model is to generate data using a stochastic approach. With this idea our goal is to estimate the common parameters between two models. In particularly, now we decide to generate our data

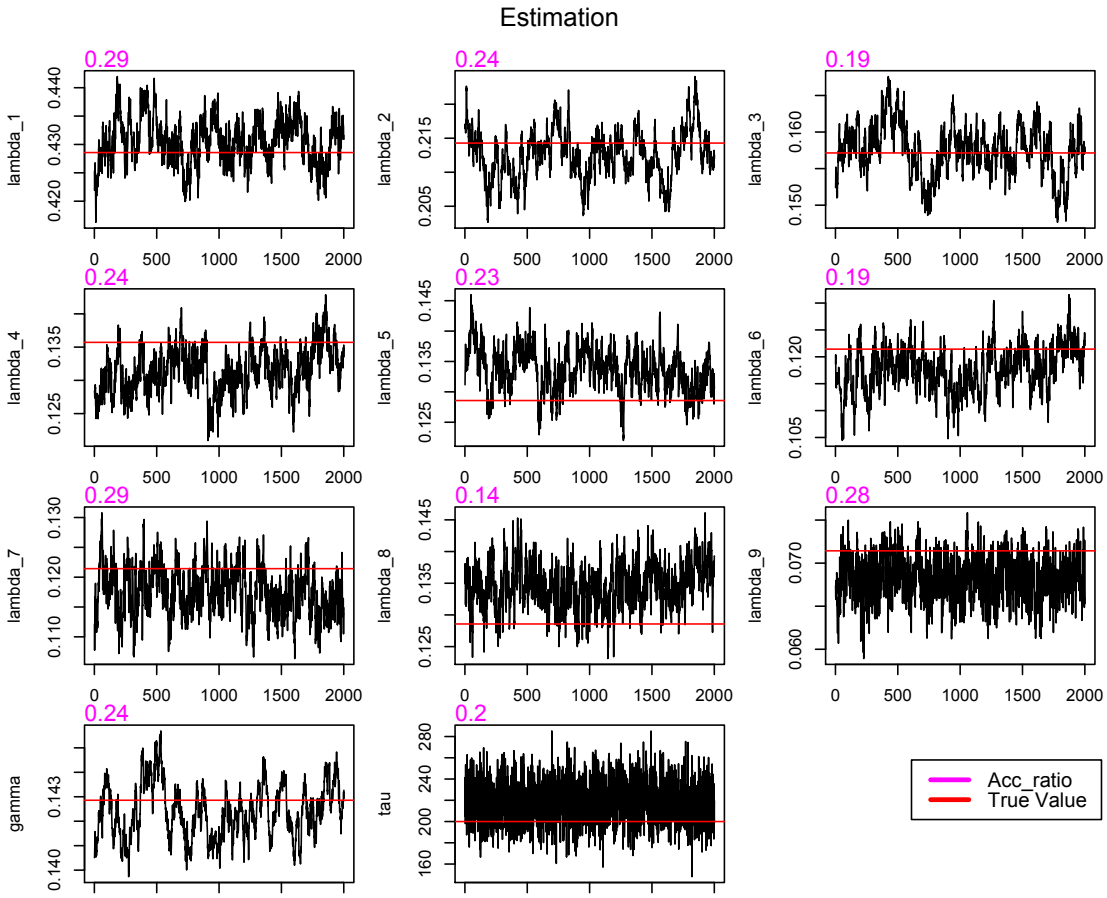


Figure 6.3: The evolution of different chains when stationarity is reached for a deterministic SIR with $\tau = 200$.

solving a SDEs system, and than use the same deterministic approach as before to obtain the estimation. The method to generate data with a stochastic approach is described in Section 6.3. In Figure 6.9, we can observe that almost all chains find out some difficulties to reach the stationarity, even if for a larger number of Metropolis-Hastings steps. For this reason, we decide not to continue on this path.

6.3 MCMC for stochastic SIR

As already discussed, an alternative approach in order to model an epidemic process is using a stochastic technique. Mirroring the deterministic part, we apply again the Gibbs algorithm with a Metropolis-Hastings step. We recall Section 4.2 and exactly as before, the parameters $\boldsymbol{\lambda}$ and γ represent the rate of infection (depending on time) and the rate of recovery, respectively. In addiction, we may introduce a new parameter ν. Indeed,

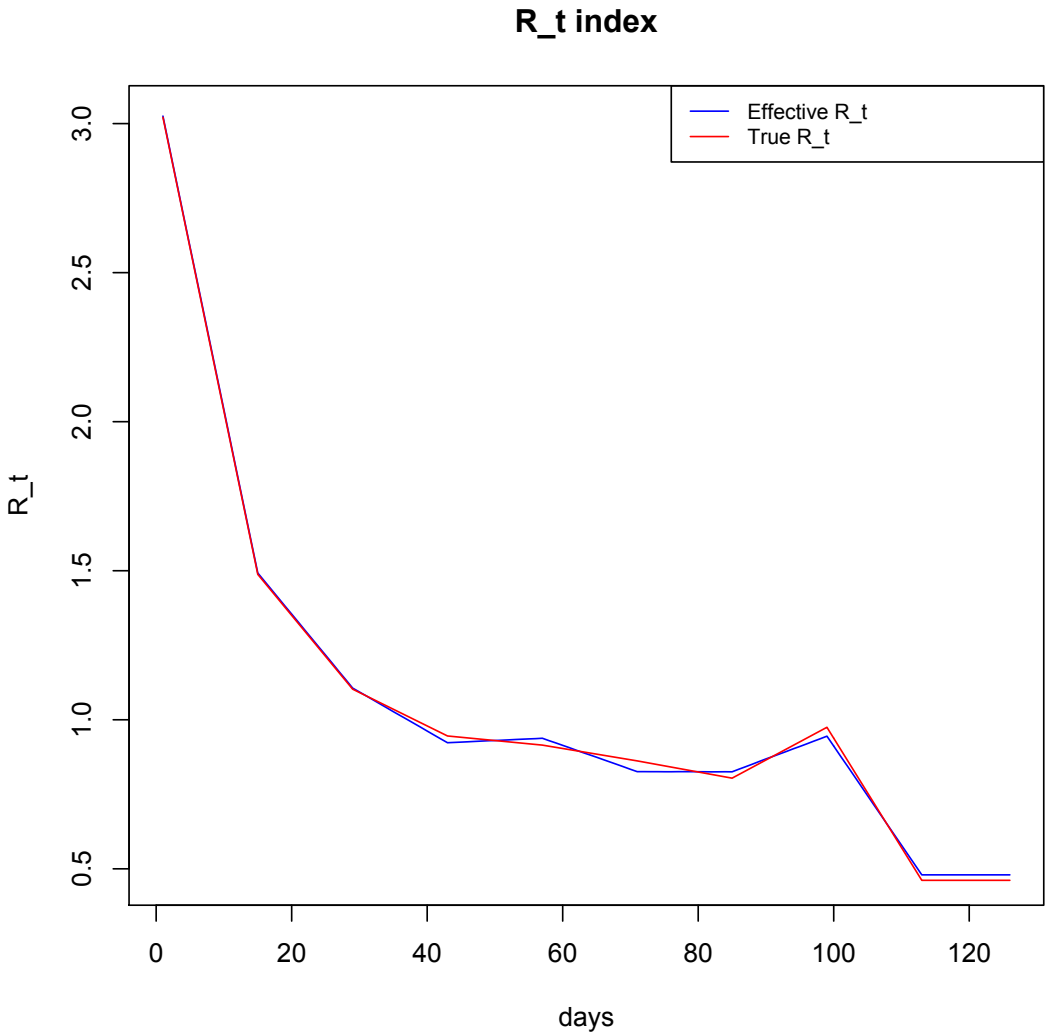


Figure 6.4: The estimation of R_t each day for a deterministic SIR with $\tau = 200$ compared with the R_t used for the data generation.

when we deal with a stochastic model, we have two different sources of noise. The first one is inherent in the system, while the other is due to the fact that we don't have enough information about the system. In particular, the diffusion of an illness is a process that is itself stochastic, even if we ideally had clean data. On the contrary, the second kind of error is the proper "measurement error". If we suppose to describe the process via a Markov chain, we would expect ν equal to 1. In fact, all the intrinsic variability should be explained by the model. Anyway, we come up against data collected in real word, which are noisy. So we can consider this additional parameter. It is really important underline

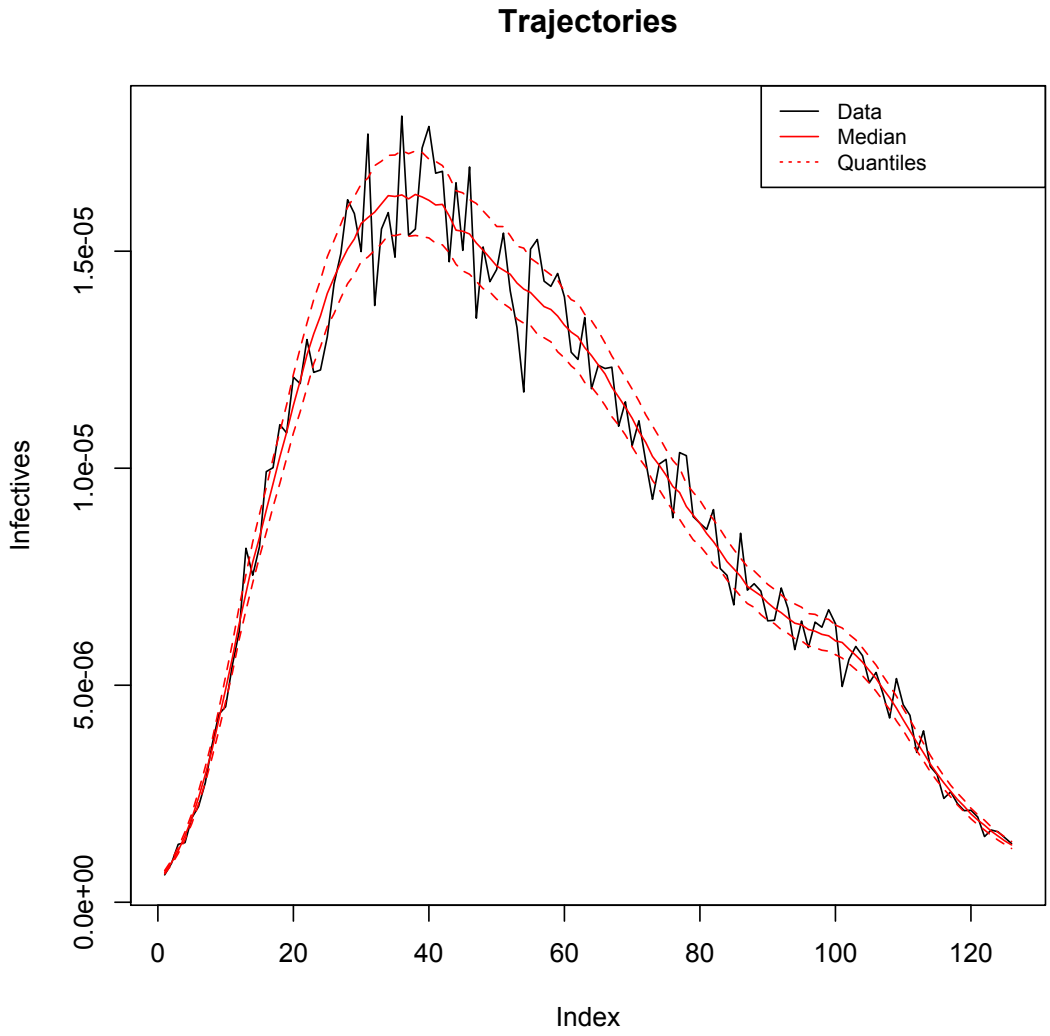


Figure 6.5: Changes in the proportion of infective individuals and the trajectory obtained with the median and two quantiles with probability 0.25 and 0.75 evaluated on the results of the deterministic SIR with $\tau = 200$.

that this new parameter has a completely different meaning with respect to the parameter τ introduced in the deterministic model. To summarize, the vector of parameters we have to estimate in a stochastic SIR is given by: $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \gamma, \nu)$.

Likelihood

The first step is the definition of the likelihood function $\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta})$. Also in this context, the matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_d)$ contains our observations. In the following, it will be helpful to consider the logarithm of the function to simplify calculations, just as before.

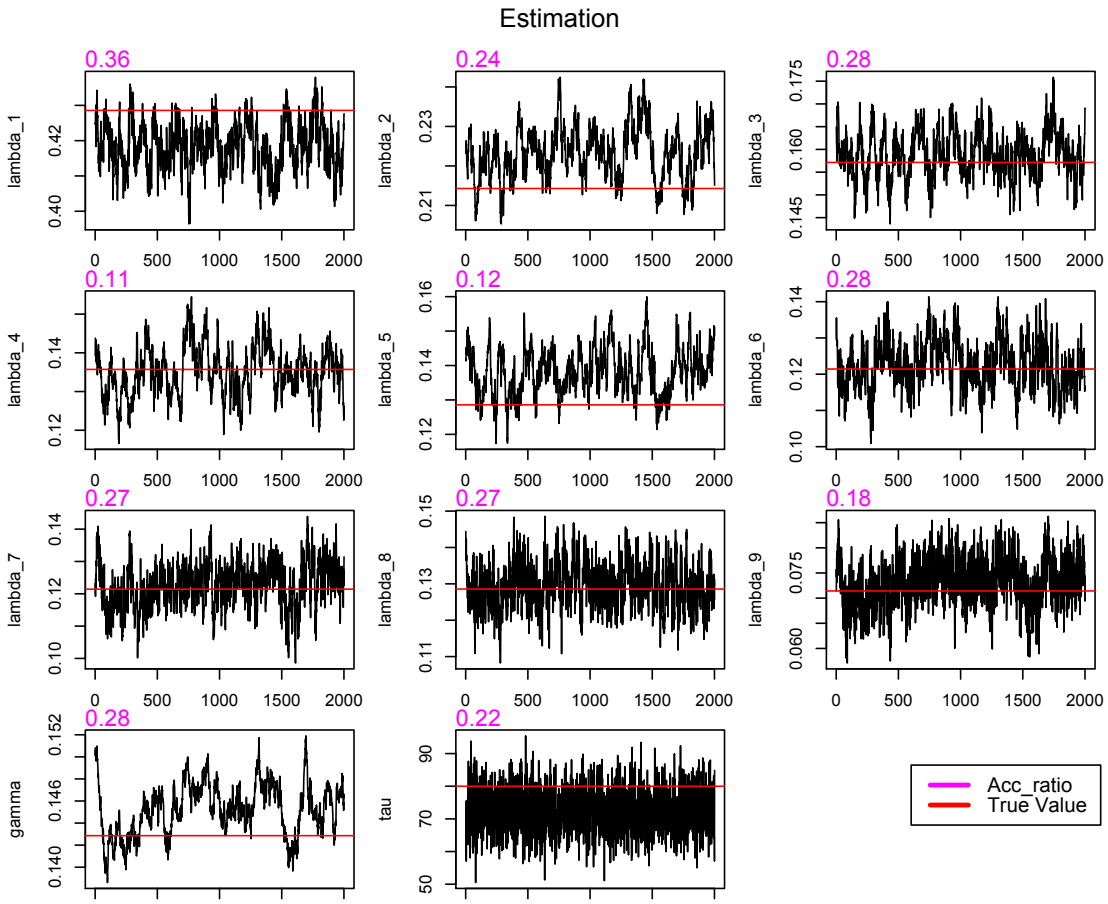


Figure 6.6: The evolution of different chains when stationarity is reached for a deterministic SIR with $\tau = 80$.

We underline that in the stochastic case, we are dealing with a **Markov process**, so there are several simplifications in evaluating the likelihood. First of all, we recall the most relevant characteristic: the Markov property. Because of it, we use the following trick in order to get some advantages:

$$\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta}) = \mathscr{L}(\boldsymbol{X}_1 = x_1, \boldsymbol{X}_2 = x_2, ..., \boldsymbol{X}_d = x_d|\boldsymbol{\theta}) =$$

$$\mathscr{L}(\boldsymbol{X}_d|\boldsymbol{X}_{d-1}, \boldsymbol{\theta})\mathscr{L}(\boldsymbol{X}_{d-1}|\boldsymbol{X}_{d-2}, \boldsymbol{\theta})...\mathscr{L}(\boldsymbol{X}_2|\boldsymbol{X}_1, \boldsymbol{\theta})\mathscr{L}(\boldsymbol{X}_1|\boldsymbol{X}_0, \boldsymbol{\theta}) =$$

$$\prod_{t=1}^{d} \mathscr{L}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}, \boldsymbol{\theta}) = \prod_{t=1}^{d} \mathscr{L}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}, \boldsymbol{\lambda}, \gamma, \nu).$$

At this point, the problem is thrown back on finding the likelihood function at each instant of time during the observed period. We also recall that we are using a *density process* to model the phenomenon. Starting from the reactions, we end with the stochastic

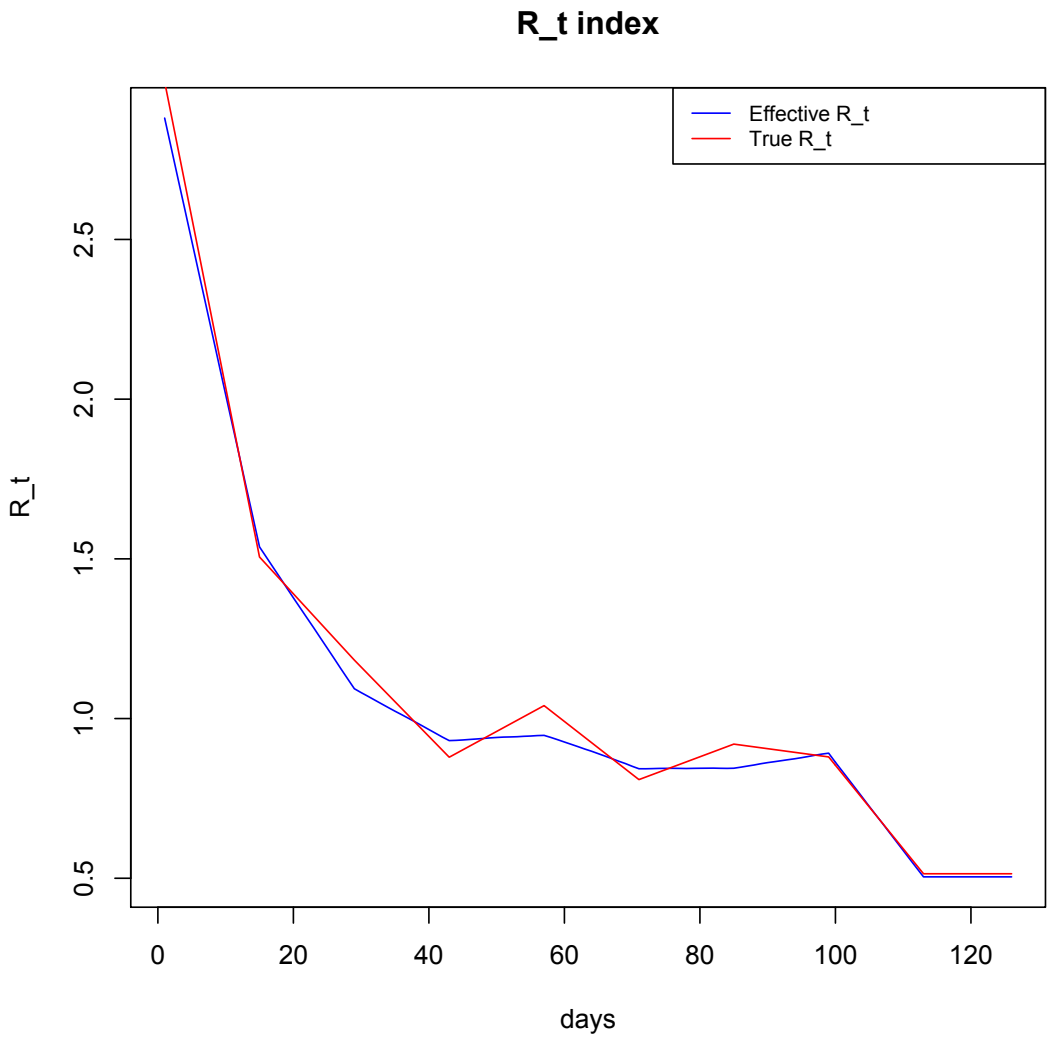


Figure 6.7: The estimation of R_t each day for a deterministic SIR with $\tau = 80$ compared with the R_t used for the data generation.

differential equations which are able to describe the state of the system. There we use $I(t)$ and $R(t)$ to denote the proportion of Infective and Removed at time t over a fixed population, respectively. The two-dimensional SDE describing the system is:

$$d\boldsymbol{X_t} = \boldsymbol{a_t}dt + \boldsymbol{B}_t d\boldsymbol{W}_t, \tag{6.3}$$

where:

$$\boldsymbol{X}_t = \begin{pmatrix} I(t) \\ R(t) \end{pmatrix}$$

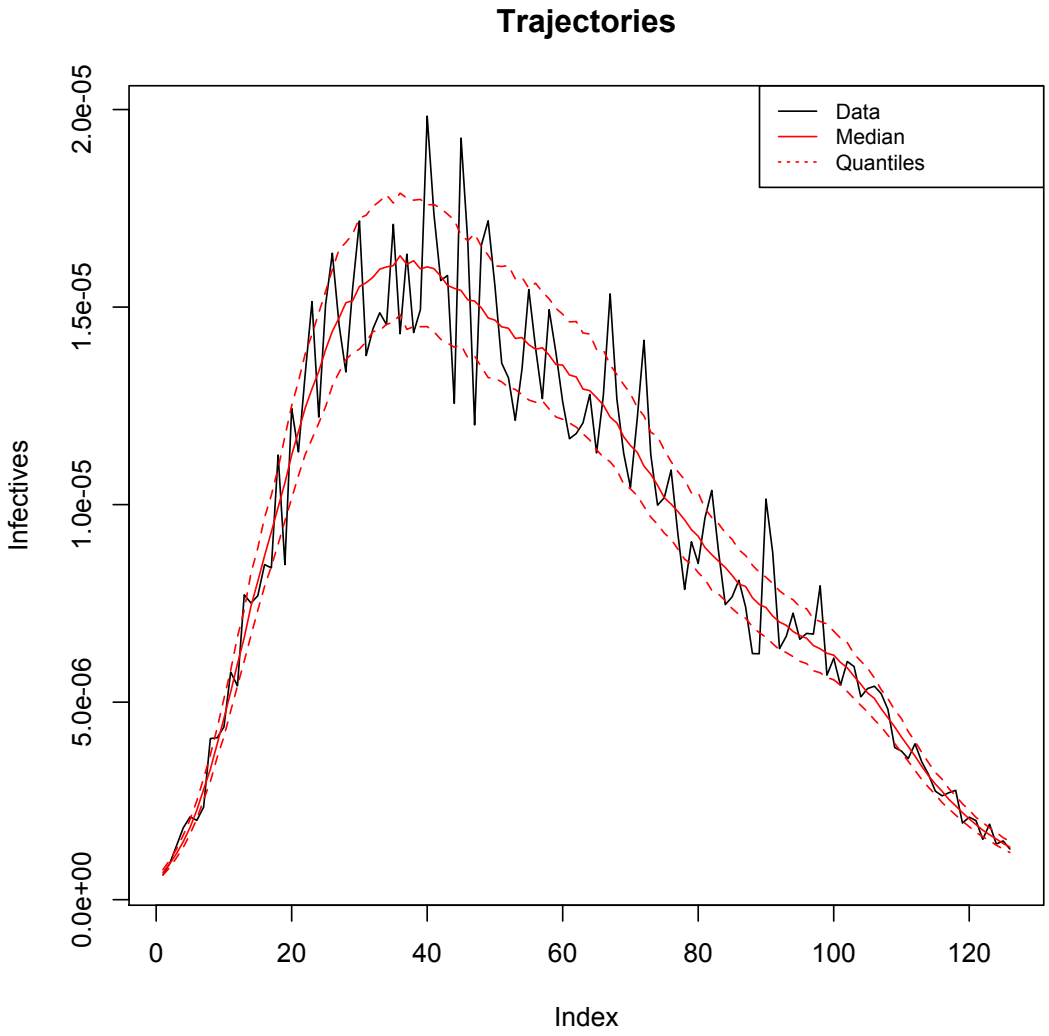


Figure 6.8: Changes in the proportion of infective individuals and the trajectory obtained with the median and two quantiles with probability 0.25 and 0.75 evaluated on the results of the deterministic SIR with $\tau = 80$.

$$\boldsymbol{a_t} = \begin{pmatrix} \boldsymbol{\lambda}[t]\boldsymbol{X}[1,t-1](1-\boldsymbol{X}[1,t-1]-\boldsymbol{X}[2,t-1]) - \gamma\boldsymbol{X}[1,t-1] \\ \\ \gamma\boldsymbol{X}[1,t-1] \end{pmatrix}$$

$$\boldsymbol{B_t} = \frac{dt}{\sqrt{\nu N}} \begin{pmatrix} \sqrt{\boldsymbol{\lambda}[t]\boldsymbol{X}[1,t-1](1-\boldsymbol{X}[1,t-1]-\boldsymbol{X}[2,t-1])} & -\sqrt{\gamma\boldsymbol{X}[1,t-1]} \\ 0 & \sqrt{\gamma\boldsymbol{X}[1,t-1]} \end{pmatrix}$$

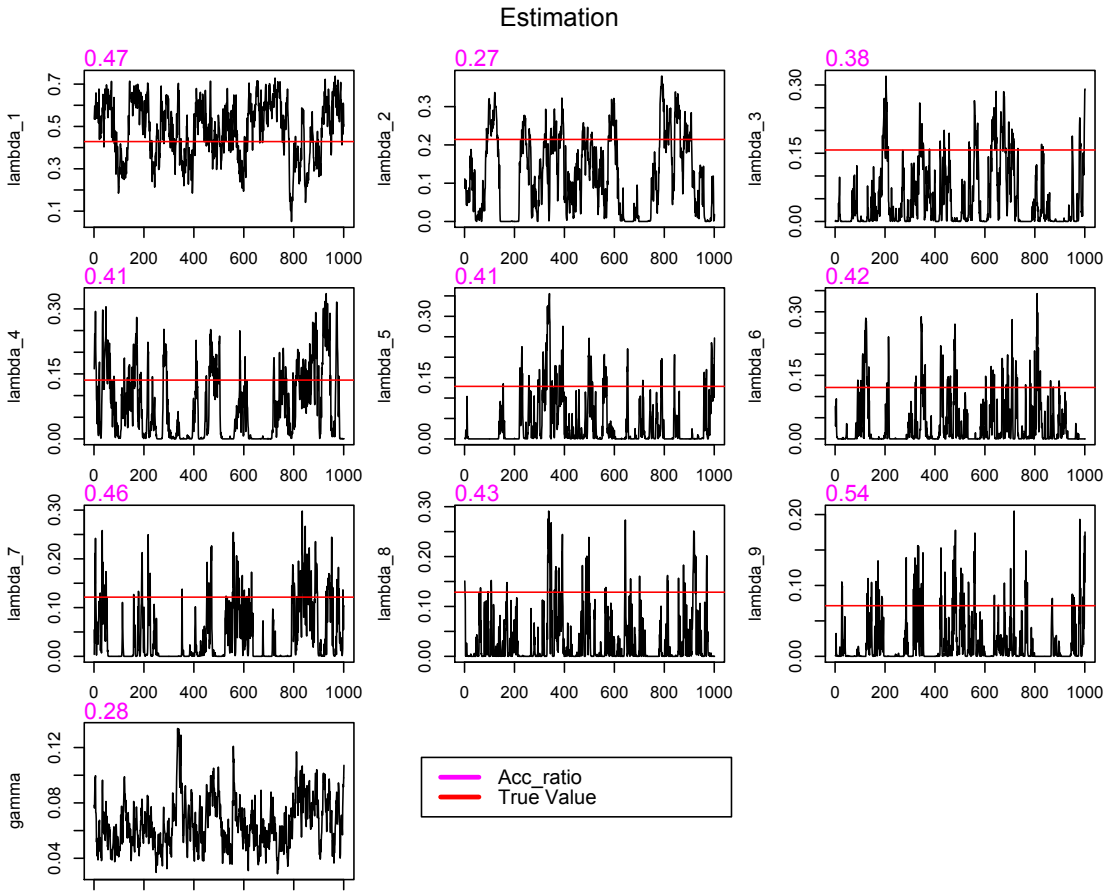


Figure 6.9: The evolution of different chains when stationarity is reached for SIR parameters generated using SDEs and estimated via a deterministic model.

dt is the limit approximation of the time grid increment Δ_t, while $W = \{W_t, t \in [0, T]\}$ is the standard Wiener process, so we know that its increments are independent random variables following a normal distribution with zero mean and variance dt.
Putting all the previous pieces together, we can conclude that:

$$(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}, \boldsymbol{\theta}) \sim \boldsymbol{N}_2 (\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t).$$

in which:

$$\boldsymbol{\mu_t} = \boldsymbol{X_{t-1}} + \boldsymbol{a_t} dt,$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{B}_t \boldsymbol{B}_t^T.$$

From these derivations, coming back to the likelihood distribution, we get:

$$\begin{aligned}\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta}) = \mathscr{L}(\boldsymbol{X}|\boldsymbol{\lambda},\gamma,\nu) &= \prod_{t=2}^{d}\mathscr{L}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\boldsymbol{\lambda},\gamma,\nu)\\ &\propto \prod_{t=1}^{d}\exp\left(-\frac{1}{2}(\boldsymbol{X_t}-\boldsymbol{\mu_t})^T{\boldsymbol{\Sigma_t}}^{-1}(\boldsymbol{X_t}-\boldsymbol{\mu_t})\right)\\ &= \exp\left\{-\frac{1}{2}\sum_{t=1}^{d}(\boldsymbol{X_t}-\boldsymbol{\mu_t})^T{\boldsymbol{\Sigma_t}}^{-1}(\boldsymbol{X_t}-\boldsymbol{\mu_t})\right\}.\end{aligned}$$

Now if we consider the logarithm of this function, we can obtain some simplifications.

$$\log(\mathscr{L}(\boldsymbol{X}|\boldsymbol{\lambda},\gamma,\nu)) = log\left(\prod_{t=1}^{d}\mathscr{L}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\boldsymbol{\lambda},\gamma,\nu)\right) = \sum_{t=1}^{d} log\left(\mathscr{L}(\boldsymbol{X_t}|\boldsymbol{X_{t-1}},\boldsymbol{\lambda},\gamma,\nu)\right)$$

Priors

As before, the next step consist into defining the prior distributions. We can repeat exactly the same reasoning. So, we can assume that the logarithm of the infection rate $\boldsymbol{\lambda}$ is distributed according to a multidimensional normal distribution and the logarithm of the recovery rate γ follows a normal distribution. Finally, as in the previous context, we do the same as τ for the new parameter. For $\log(\nu)$ we propose a particular kind of distribution such that the parameter follow a Gamma distribution. Its distribution is described in equation (6.2). In fact, in the MCMC algorithm, because of the previous simplifications, we'd like to make a proposal on a symmetric distribution. This is why we work with the logarithm of the parameters.

Posterior

In conclusion, by applying Bayes' theorem, we are able to get the final result.

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{X}) = \pi(\log(\boldsymbol{\lambda}),\log(\gamma),\nu|\boldsymbol{X}) \propto \mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto\\ \nu^{a-1}\exp\left\{-\frac{1}{2}\left[\nu\sum_{t=1}^{d}(\log(\boldsymbol{X}_i)-\log(\bar{\boldsymbol{X}}_i)) + s\sum_{i=1}^{n}(\lambda_i-m)^2 + r(\gamma-n)^2 + 2\nu b\right]\right\},\end{aligned}$$

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\log(\boldsymbol{X}))) \propto \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})\pi(\boldsymbol{\theta})) = \log(\mathscr{L}(\boldsymbol{X}|\boldsymbol{\theta})) + \log(\pi(\boldsymbol{\theta})) =\\ \log(\mathscr{L}(\log(\boldsymbol{X})|\boldsymbol{\theta})) + \log(\pi(\log(\boldsymbol{\lambda}))) + \log(\pi(\log(\gamma))) + \log(\pi(\nu)).\end{aligned}$$

Full Conditionals

Once again we have to repeat the previous considerations. Indeed, the posterior distribution has a very uncomfortable form and it could be really difficult sample from it. This is the reason why we can use the Gibbs sample and consider the full-conditional distributions. In practice, we analyze the kernel of the posterior distribution and easily we consider only the portions in which the variable we are interested in appears. The code implementation of these functions just defined is in A.7.

6.3.1 Stochastic SIR results

At this point we are ready to implement also this alternative method. The code developed in order to obtain the following results is implemented in A.8.
The same general considerations discussed in Section 6.2.1 for the MCMC algorithm remains true. We decide the number of total iterations, the *thin* value, and the standard deviation for each parameter by trial and error. What is really different now is of course the definition of the *likelihood* and all the other distributions that depend on it. The other main difference is the definition of Euler-Maruyama function which we use in order to build the trajectories of the stochastic process. Its implementation in R language can be found in A.5. Once the Euler function is implemented, we can generate simulated data by a stochastic approach like in A.6.
The first attempt consisting in using a deterministic approach to generate data and a stochastic one in the Metropolis-Hastings algorithm. As we can simply observe from the trend of different chains in Figure 6.10, this does not seem to be a promising path. Indeed we are introducing some kind of intrinsic noise in a deterministic model and there are several difficulties in reaching the stationarity near the true value used for data generation. Then we decide not to mixing the two previous methods, but only use the stochastic technique in order to generate data for a stochastic estimation.
Thus, we use the Euler-Maruyama scheme and we solve an SDE equation for the estimation. In Figures 6.11 and 6.12, we can observe a better performance of the algorithm rather than before, when we use a stochastic approach to generate data. In these figures we decide to use $\nu = 2$ to generate data. This value is quite similar to what we expect from theory if our data were without noise. We can observe that our estimation process works quite well, the trajectories for infective built in Figure 6.13 are not so accurate. In fact, even if the median is quite similar to the true realization, there is a very large variability which doesn't seem a good signal.
However a distinct improvement can be observed in Figures 6.14, 6.15, and 6.16. Here we decide to generate data with $\nu = 80$. Of course as a consequence the variability is significantly reduced both in observation and in the estimation. In this case the median is not as overlapping as before with "real" data, but on the other hand the band is tighter. In addiction, except for a few points, the black line is always contained in the band. This was built with the same values of probability as before: 0.25 and 0.75.

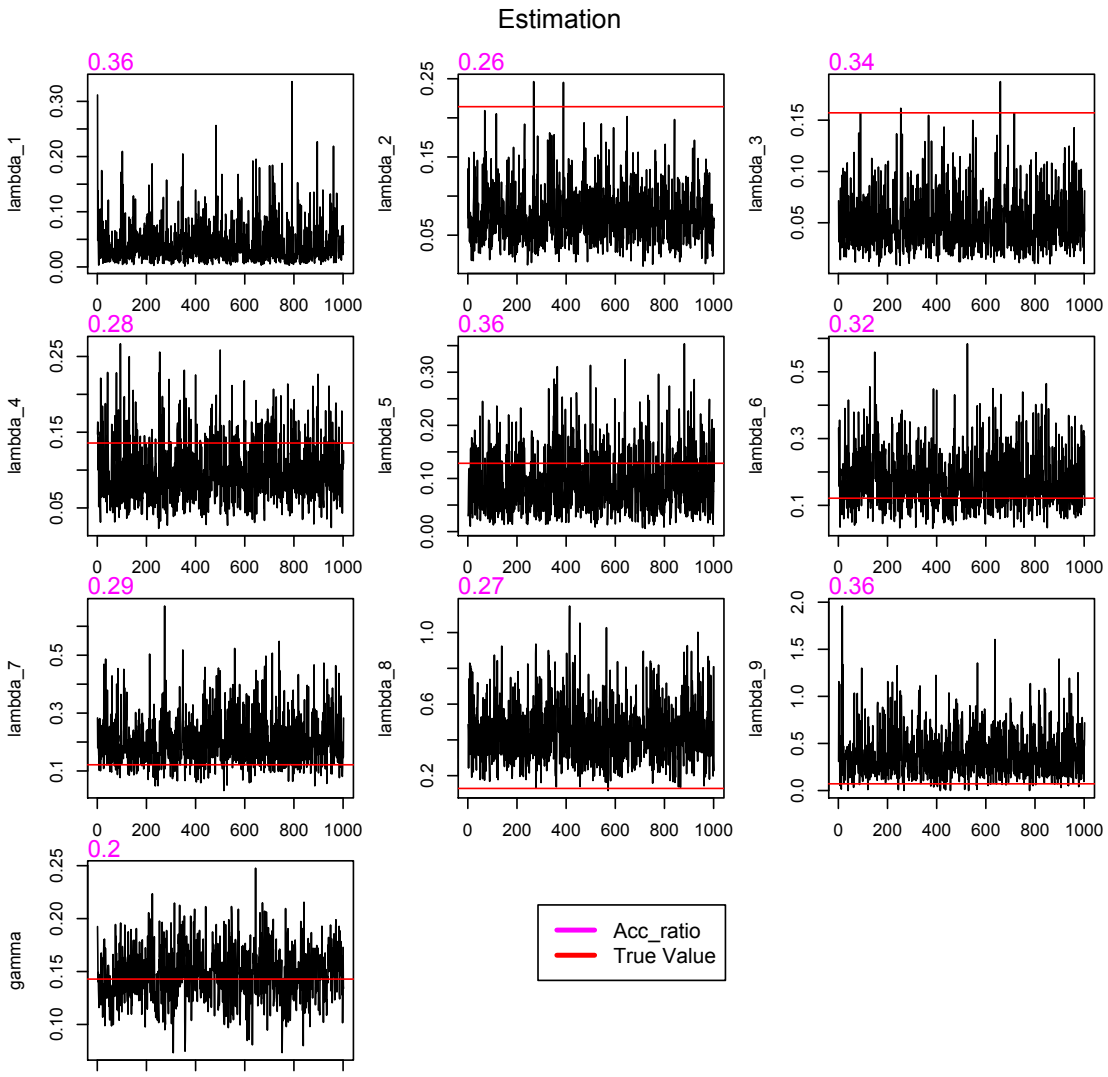


Figure 6.10: The evolution of different chains when stationarity is reached for SIR parameters generated using ODEs and estimated via a stochastic model.

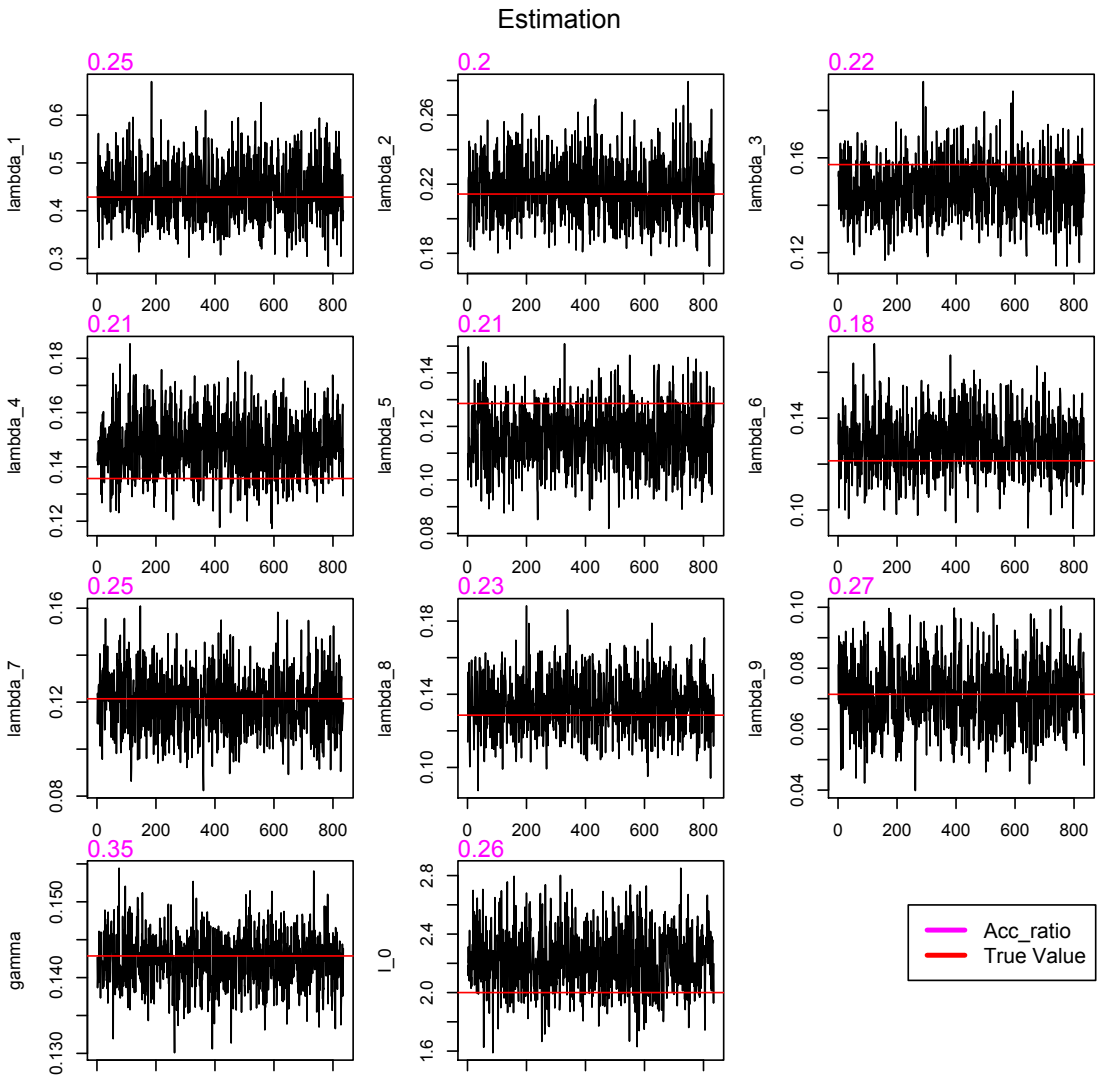


Figure 6.11: The evolution of different chains when stationarity is reached for a stochastic SIR with $\nu = 2$.

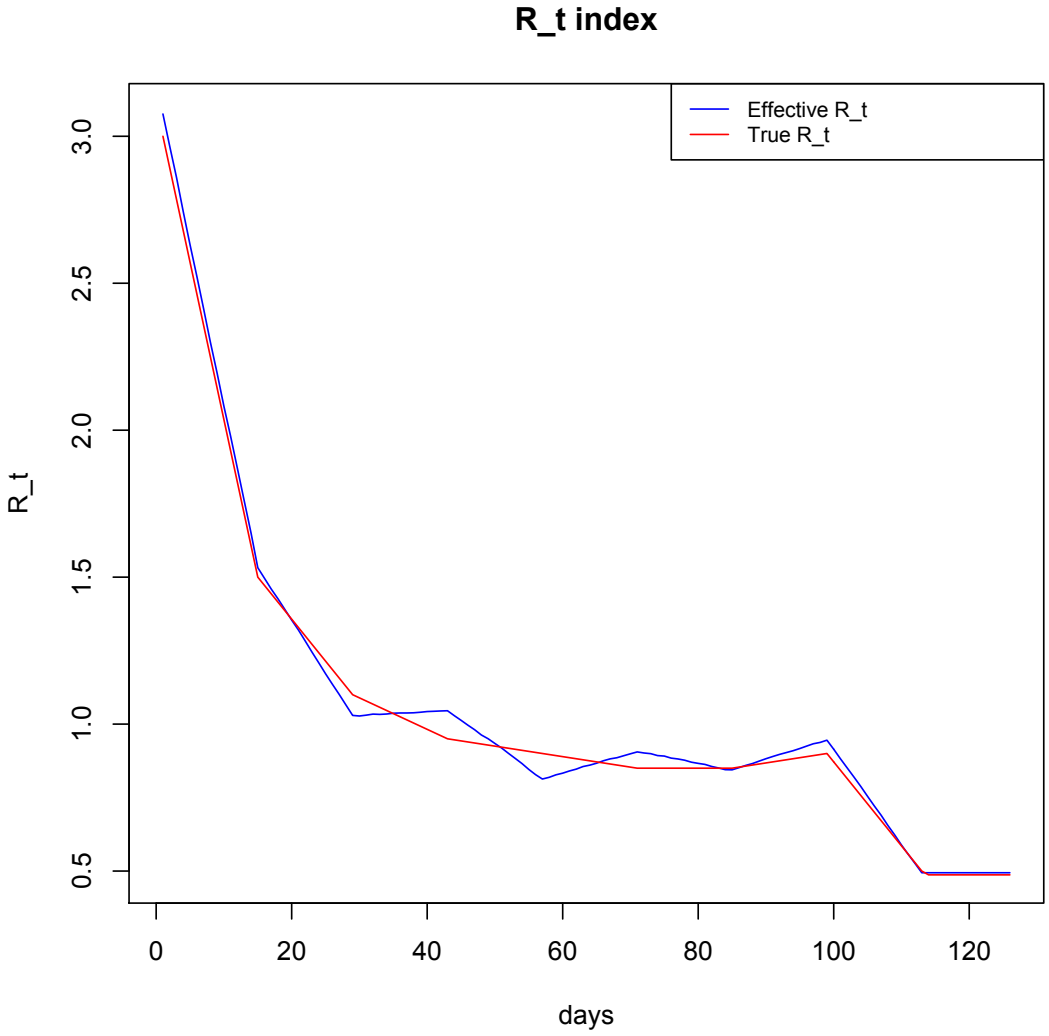


Figure 6.12: The estimation of R_t each day for a stochastic SIR with $\nu = 2$ compared with the R_t used for the data generation.

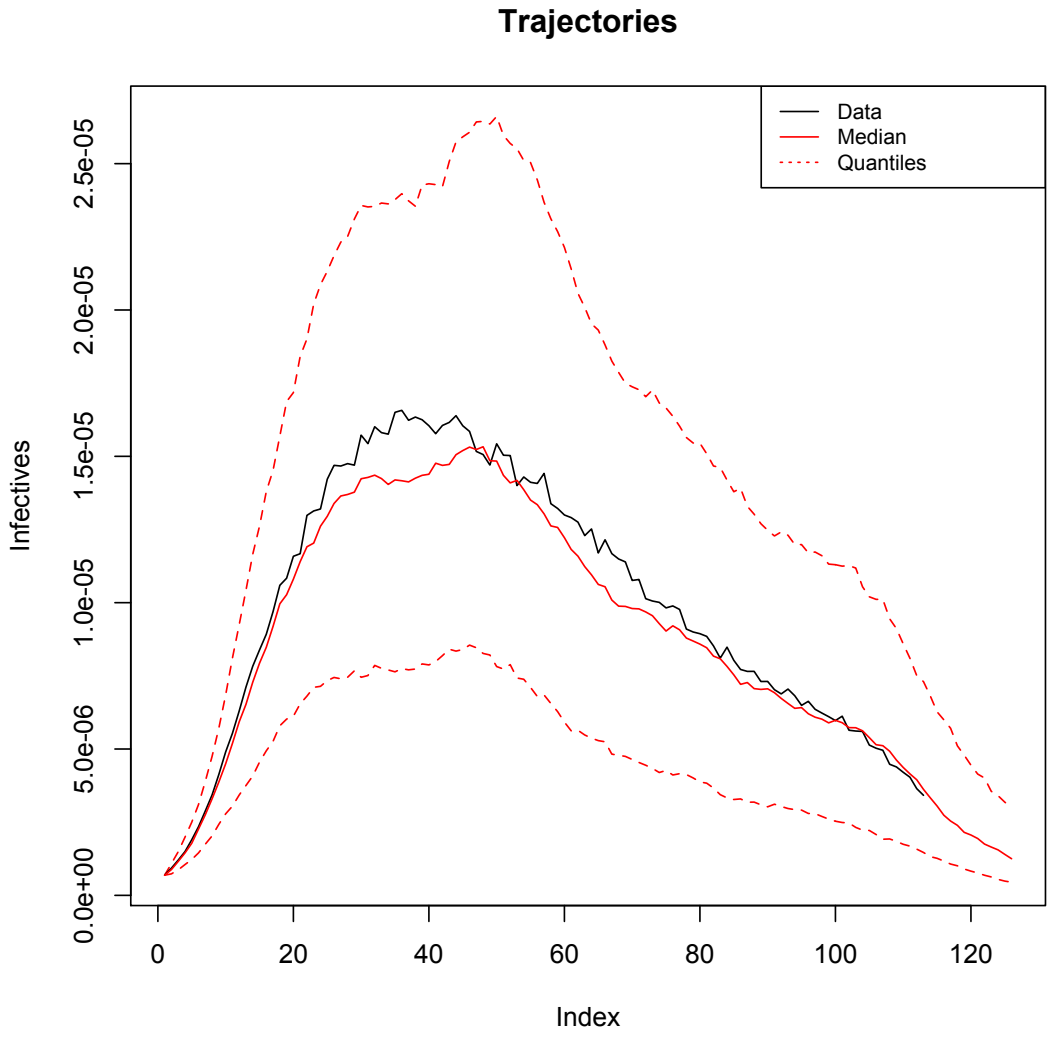


Figure 6.13: Changes in the proportion of infective individuals and the trajectory obtained with the median and two quantiles: the first and the third quartiles. They are evaluated on the results of the stochastic SIR for $\nu = 2$.

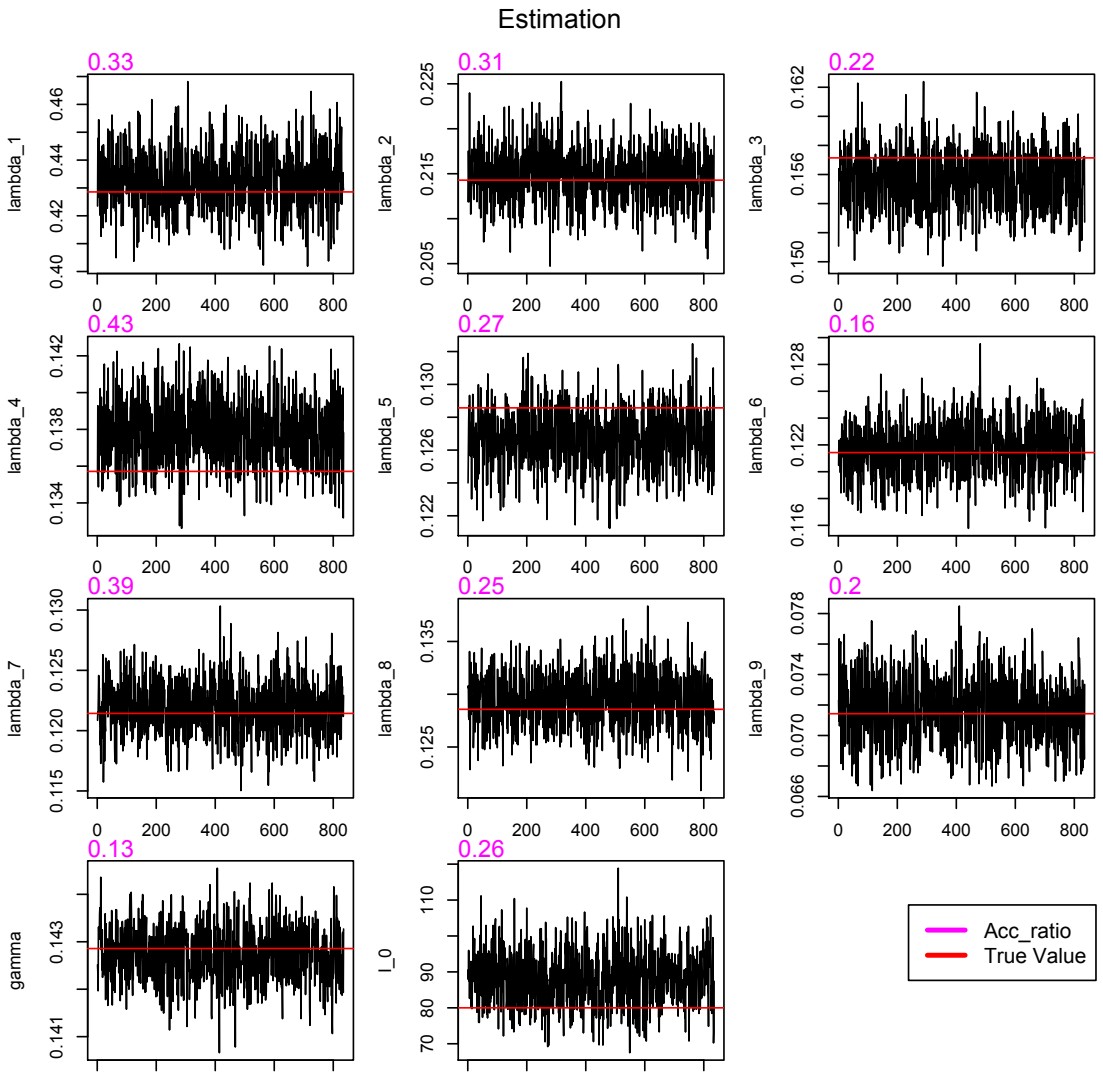


Figure 6.14: The evolution of different chains when stationarity is reached for a stochastic SIR with $\nu = 80$.

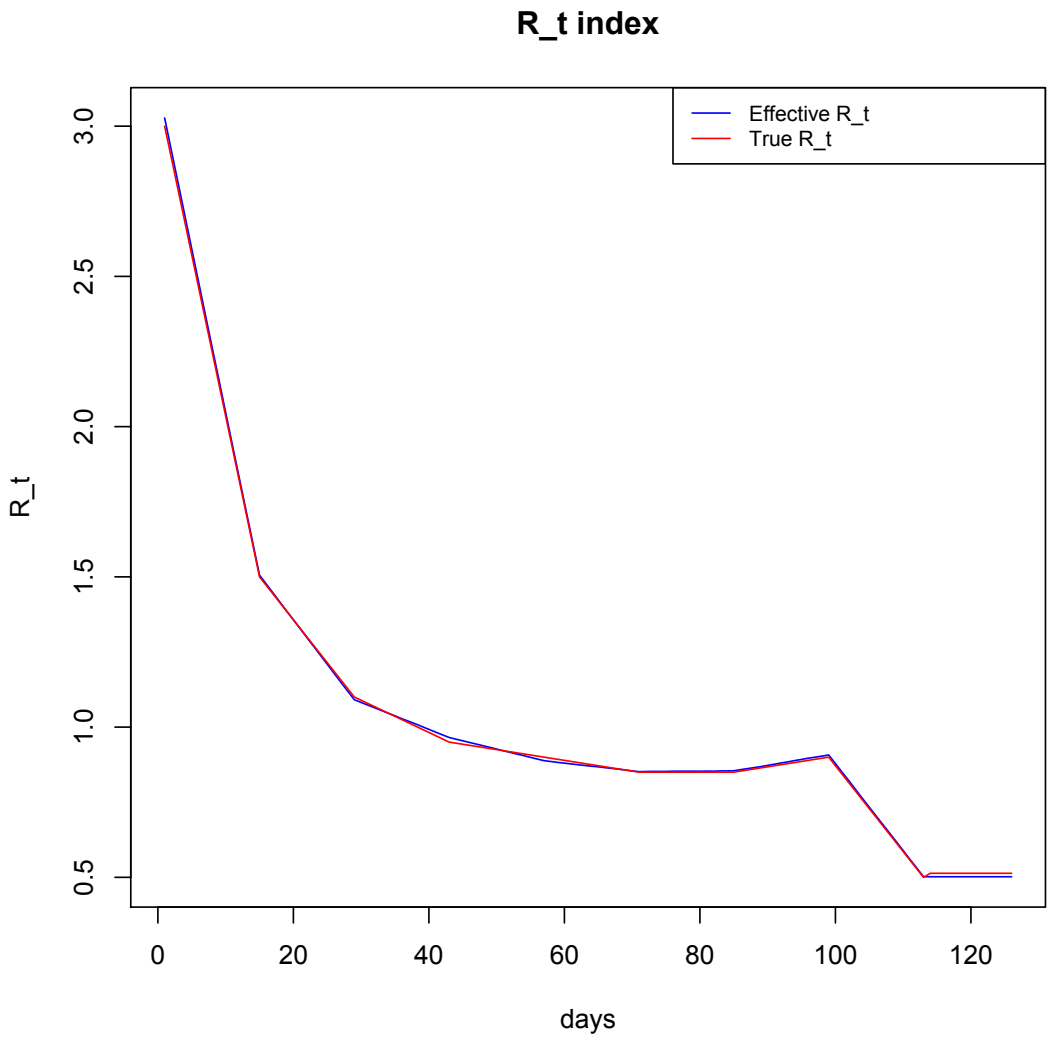


Figure 6.15: The estimation of R_t each day for a stochastic SIR with $\nu = 80$ compared with the R_t used for the data generation.

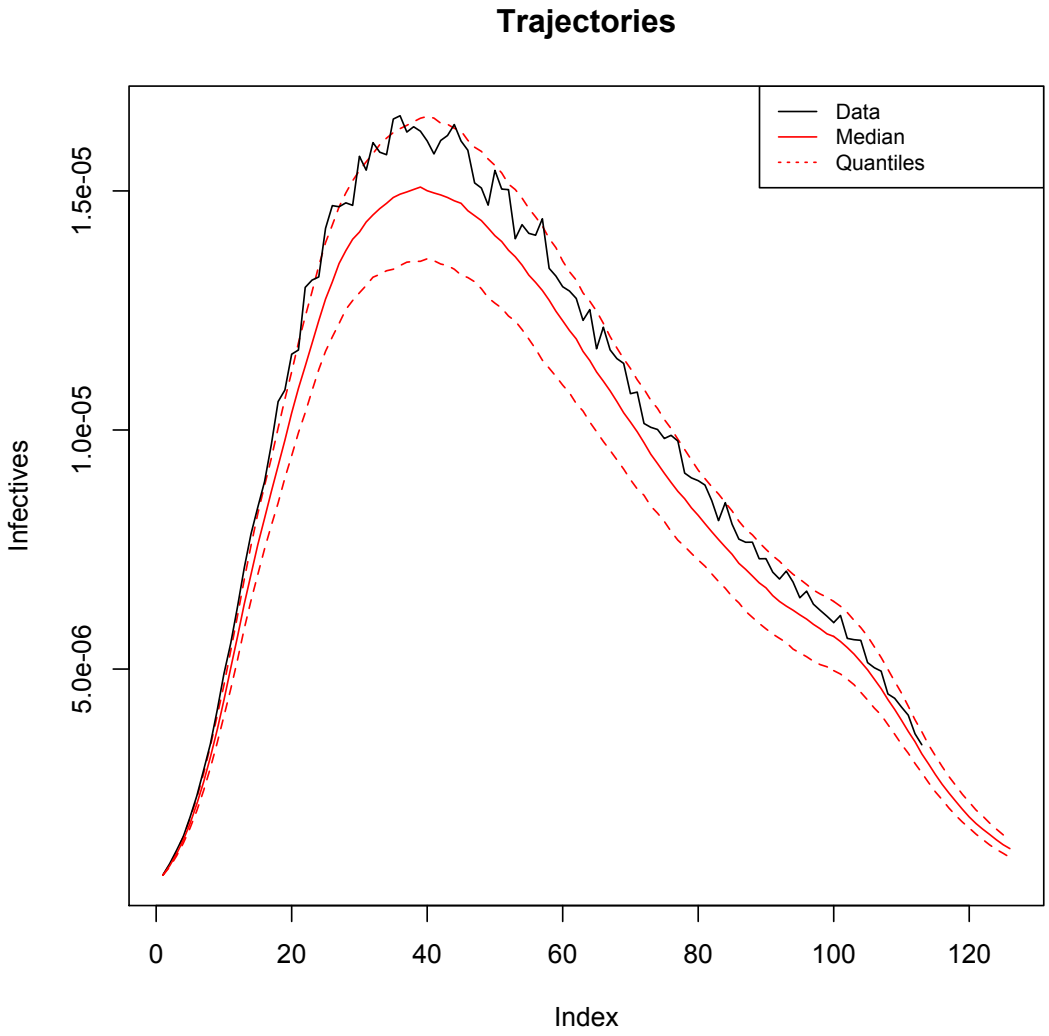


Figure 6.16: Changes in the proportion of infective individuals and the trajectory obtained with the median and two quantiles with probability 0.25 and 0.75 evaluated on the results of the stochastic SIR for $\nu = 80$.

Chapter 7

Real data

In this Chapter, we apply exactly the same algorithms presented earlier, but without first having to generate data. In fact, our final goal is not testing the previous models, but use them with real data. In this way, we are able to model what really happened during the epidemic outbreak. Indeed since these data are available, via Bayesian statistics, we are able to mix our prior knowledge with the realization in the real world.
First of all, we select only the data we are interested in. We visit the site [15], where in real time, it is possible to consult important information about the COVID-19 diffusion everywhere in Italy. In this thesis, we will further focus on data from Piemonte. All the data and relevant graphics are published and collected by Italian Protezione Civile. In particular, at [16], we can access to all the historical data. In Figure 7.1 the whole curve of contagion is plotted, only to get a general introduction on what kind of data we are handling. In the following Section, the data structure is described more in detail. In Section 7.2, some useful considerations are made in order to threat properly these data. In addiction, Section 7.3 contains an interesting observation about the initial conditions which we have to provide both the epidemic models. Lastly, in the following two Sections, the deterministic and stochastic method respectively are presented.

7.1 Data structure

Before applying our models, we examine more in detail the data structure. In each row, in the first column, there is the timestamp when the registration was made. The following columns contain a series of information more or less relevant for our purposes concerning epidemiological trends. Among other variables there are:

- **data** : the exact moment when that specific data was registered in the format $YYYY-MM-DD\ hh:mm:ss$.
- **stato** : the country to which the data refers. It isn't particularly useful because in our case it always describes contagion in Italy.
- **codice_regione** and **denominazione_regione** : respectively the code and the name associated to a specific region in Italy to which the data refer. In this thesis,

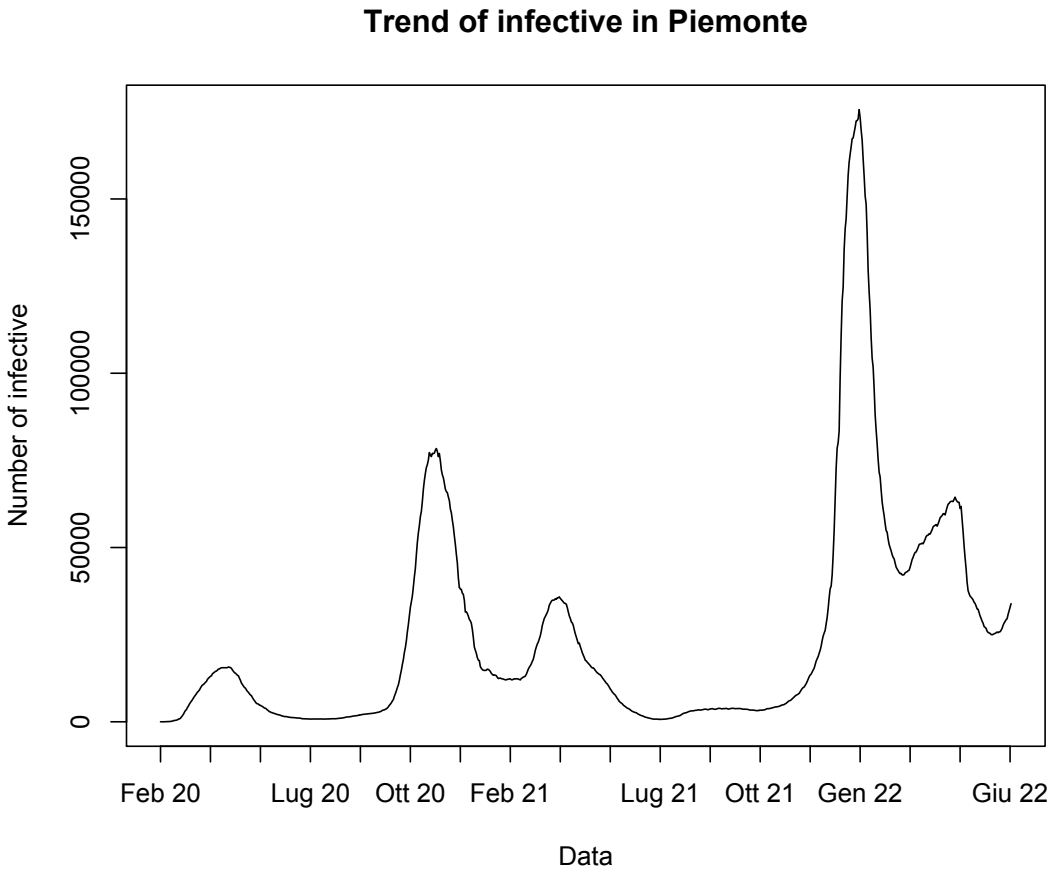


Figure 7.1: The overall evolution of infective individuals in Piemonte from the beginning of the pandemic to the current situation.

we focus on data related to Piemonte, a specific region in north-east of the country. Its code is "01" and we use it to make a *selection* of all data available.

- **totale_ospedalizzati** : it is given by the sum of other two variables. **ricoverati_con_sintomi** and **terapia_intensiva**. It represents the effective number of infective people who are hospitalized and who are in intensive care unit, on a daily basis. The distinction into these more detailed information could be helpful because in Italy, especially during the first wave of the epidemic there were problems in the hospitals. In fact, there weren't intensive care beds for all those who needed it.

- **totale_positivi** : the total number of Infective people. It is given by the total of **totale_ospedalizzati** and **isolamento_domiciliale**. In this last variable is

contained the number of those who are infected but not sick enough to need hospitalization. For this reason, they are are forced to stay at home and respect quarantine. This variable, normalized by the population, is contained in the first row of matrix $\boldsymbol{X}$.

- **nuovi_positivi** : the number of swabs with a positive result. It is worth specifying this variable doesn't (necessary) coincide with **variazione_totale_positivi**. This calculates the difference between **totale_positivi** in the current day and **totale_positivi** in the day before. The variation isn't given only by the addiction of the new infective, but we must subtract those who are no longer positive or those who have contracted the infection and whose test isn't negative yet.

- **dimessi_guariti** and **deceduti** : the number of recovered and death individuals. The sum of this two variables, normalized by the population, is collected in the second row of matrix $\boldsymbol{X}$. In the simpler model, we do not distinguish between these two classes and consider them all as representatives of Removed.

Once the organization of variables is clear, we can decide our time horizon. Pooling past experience with experimental data, we can narrow down our analysis to the *first wave*. In Italy, it began in late February, when the first cases of infection were confirmed [17]. Nevertheless, later it was discovered that the disease was already circulating for weeks, even if it wasn't officially registered. During the first phase of the epidemic there is a clear increasing in the total number of infective individuals, which then begins to decrease. This behavior is well described in Figure 7.2. On 21 of February 2020, was confirmed the first COVID case of an Italian citizen. In the next few days, there is an huge diffusion of the disease especially in the North of Italy. Then on 11/03/2020 the Italian government states the beginning of a lock down, where every unnecessary shift is forbidden and markets, productive activities and premises are closed. The immediate consequence was an evident decreasing in the number of Infectives. In early June, the containment measures are relaxed. At that time in fact, in many regions the epidemic was about to be eradicated. However, in late Summer/early Autumn, movements between Italian regions and even outside of Italy caused what is called the *second wave*, that goes beyond our analysis. Figures 7.3 and 7.4 show the trend of Removed and Susceptibles respectively, during the first wave in Piemonte. The time window considered is again from the beginning of March to the end of June in 2020. In particular, as shown in the last Figure, one of the main hypothesis concern the susceptible population. In fact, all the individuals may become eventually infected, so at the beginning we consider the whole registered population in Piemonte.
Another key point worth discussing concerns a simplification we are implicitly assuming. In fact, we say that the process can be described by a SIR model. So, once an individual become infected, then after some time he recovers and he plays no further role in the diffusion. From experimental data, it's been discovered that for those who became infected, there is a quite high probability contracting again the infection. Nowadays it isn't not so clear how this process works or how long does the immunity last. However, this hypothesis is quite reasonable because we are considering a very short time period and the probability of being again infective after such a short time is sufficiently low.

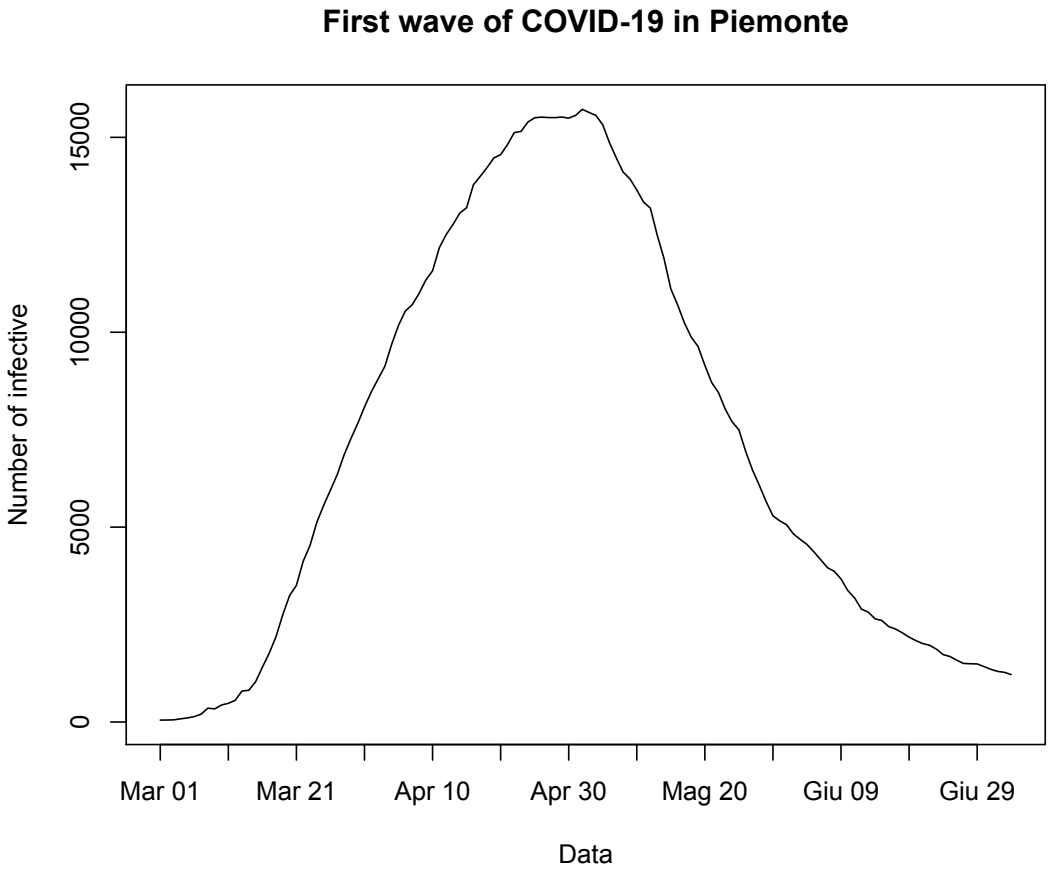


Figure 7.2: Evolution of the number of infective individuals on the fixed time window.

7.2 Data analysis

Data introduced in the previous section are important because they allow us to keep track of the *real* evolution of the epidemic process. However, there are several problems considering this kind of data because of their quite huge connected **noise**. First of all, we have to consider an *intrinsic error*. This is the variability which tries to capture the natural variations of data. In particular, we have to consider some other sources of noise. Of course, we can imagine several problems caused by the registration. Trivially, when the values of the variables involved take such high numbers, it is very likely that a mistake can been made. At the same time, there are bigger problems. First of all the **delay** for data collection. A new infective or removed individual is registered when the result of a swab is available. So, for example, the day in which a new Infective is inserted in the database is the day when the swab results positive. This of course does not correspond to the day when the individual becomes infected neither the day when the swab was made.

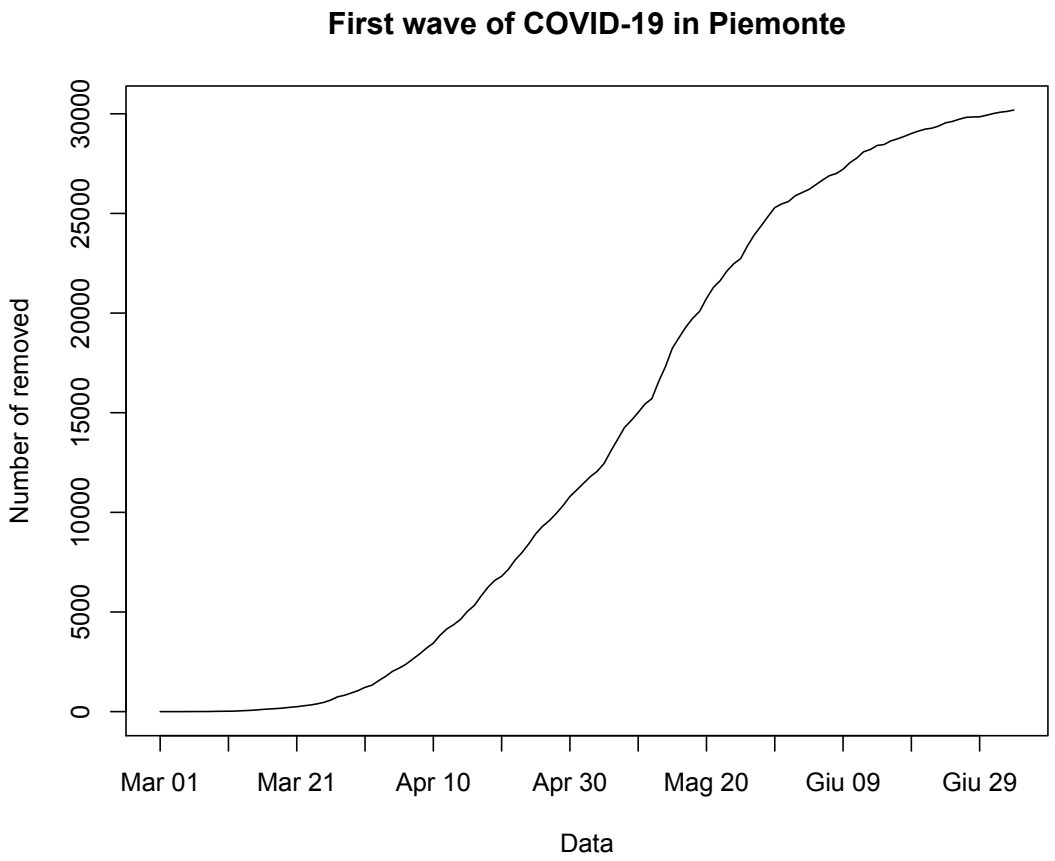


Figure 7.3: The evolution of the number of removed individuals on the fixed time window.

This delay was a big problem especially during the first wave, that is the particular period we are considering. Because of the lack of swabs, the number of tests is minimized. As a consequence, data available are not reflecting the real situation because many people were infective, but they were not tested and we can't keep track of them. A different source of delay is the analysis of swabs. Indeed at the beginning of the epidemic, labs were not ready to accommodate such a large amount of work. Then the result may arrive even three days after swabbing.
Another interesting problem, we have to deal with, is the **seasonal trend**. In fact, if we observe Figure 7.5 which represents the total amount of *new* infective people, we can observe that there is an alternation of peaks and valleys. In particular, we can analyze more in detail our data in order to understand if we can give a simple explanation to this behavior. If we observe Figure 7.6, we notice that there is a fixed pattern that repeats itself from week to week. We observe that in general on Mondays or Sundays, there is a sharp decline in Infective. This pattern is not reflecting a real decrease, but it is due to the fact that on weekends there is a decrease in the number of performed and analyzed swabs.

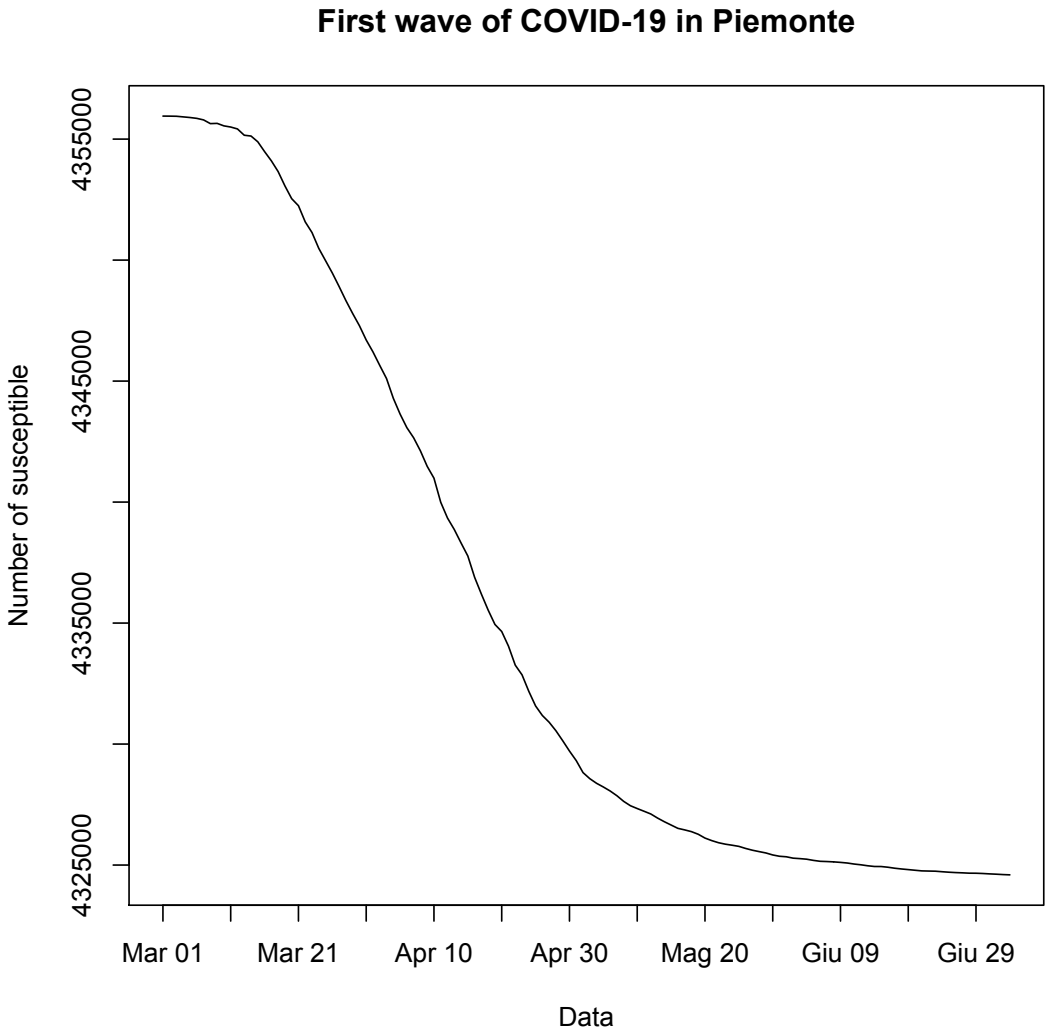


Figure 7.4: The evolution of the number of infective individuals on the fixed time window.

A connected consequence is that in general, we can also observe an increase in the number of Infective on Wednesdays. Indeed, we can notice that very often there is a peak. This of course isn't a reliable representation on what happens in real word. There is no reason for the epidemic to act in this way, but it is just a measurement error. In our analysis, we wouldn't like to introduce also this kind of error. A possible strategy to mitigate this effect is to set a suitable value for the partition interval in order to evaluate $\boldsymbol{\lambda}$. In fact, if we decide to adopt an uniform partition, we can assume that the step remains fixed to 14 days. In this way, we hope to reduce the seasonal trend.
Last but not least, a very important problem is the **lack of information**. Indeed even

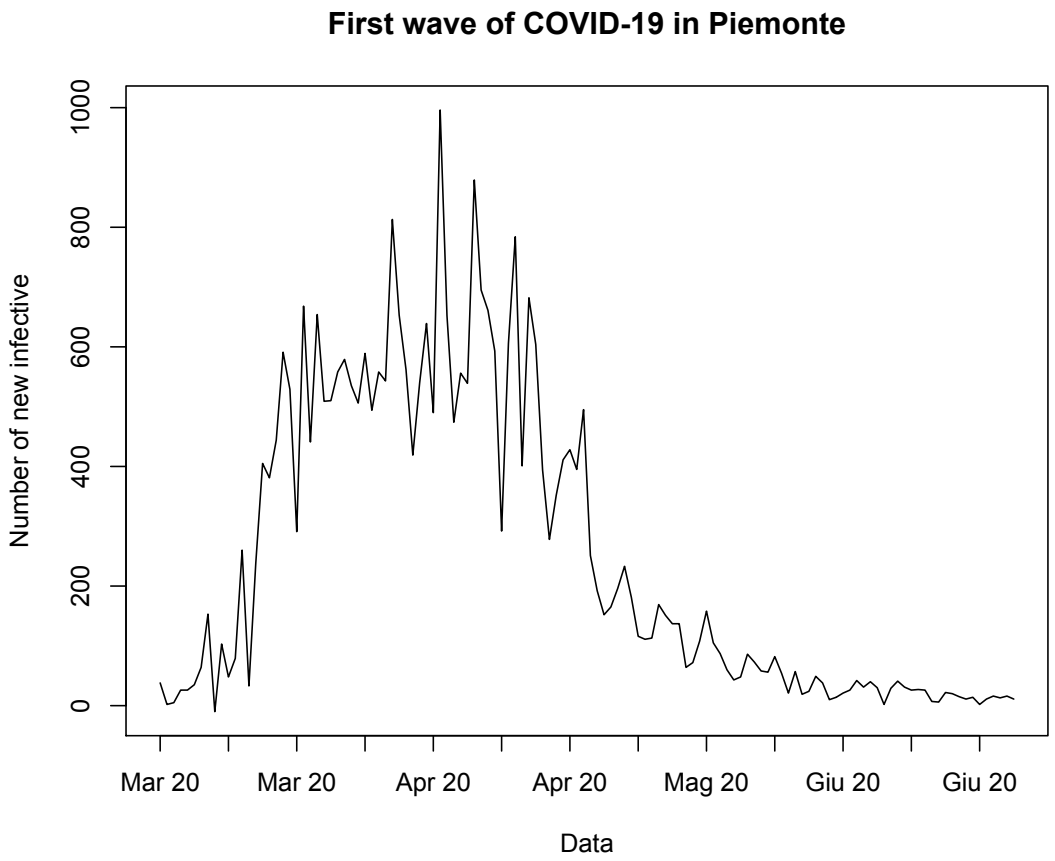


Figure 7.5: Seasonal pattern in the evolution of new infective individuals.

if the database seems quite complete, there are many missing elements. For example, to get a complete picture, it will be useful to get some data about the movement of every individual. If we could record all the shifts and meets, it would be easier to understand how the disease has been transmitted. Moreover, in this dataset, many asymptomatic individuals are not considered. Unfortunately, however, they are very important in the dynamics and they contribute to the diffusion of the epidemic. Again, among those who took a swab, there were who had contracted the disease, but they had not yet recovered. They are registered as new Infective, even though they had already been counted.

7.3 Initial Conditions

When we deal with real data, all the previous considerations in Section 6.1.1 remain true. In particular, we treat parameters involved in the MCMC algorithm in the same way. However, in this context, a new problem arises. In fact, we notice that in particular

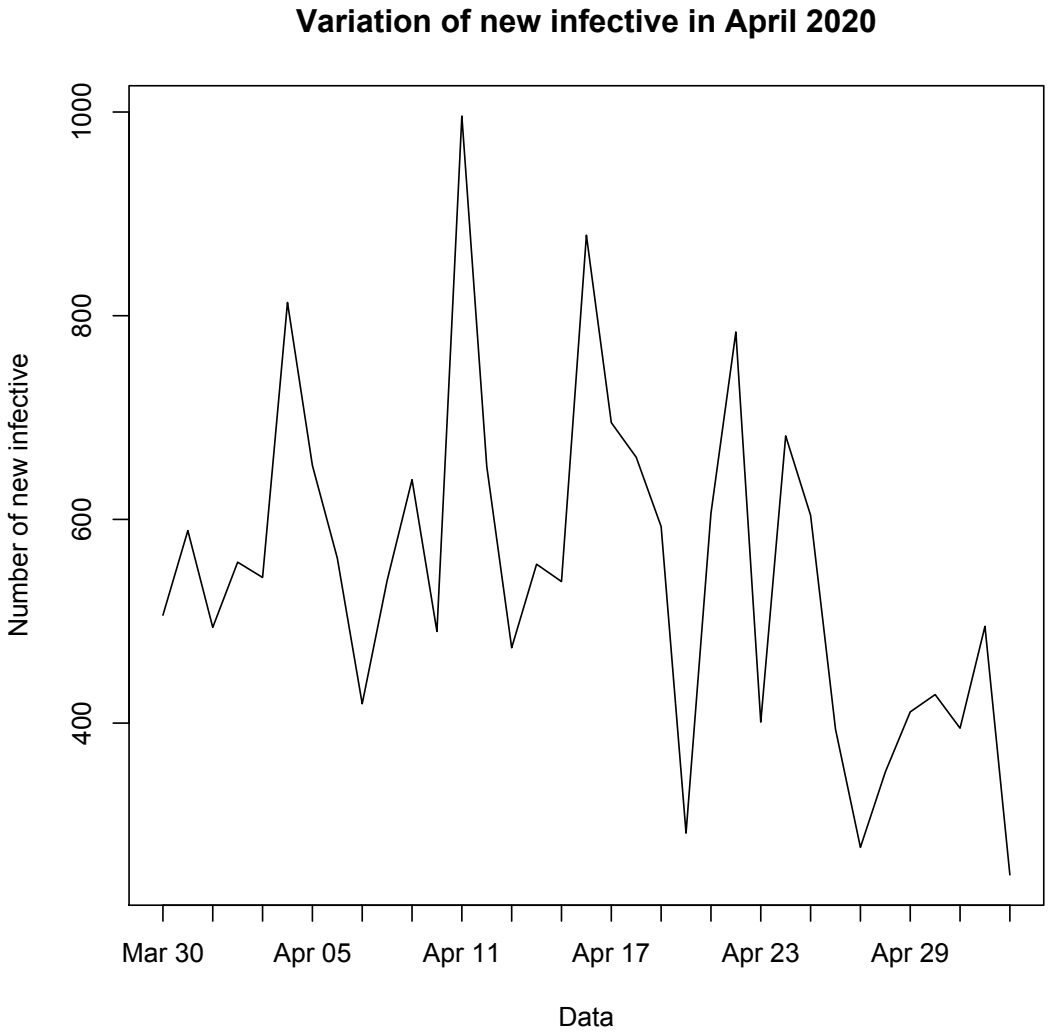


Figure 7.6: Alternating peaks and valleys for the number of new Infective over a shorter period of time.

the deterministic model encounters difficulties in the estimation process because of initial conditions. Actually, we suppose that at each iteration the initial conditions are so far from real values that it's so difficult for the algorithm to find a promising path. This is the reason why, in this second part, we decide to consider the initial proportions of Infective and Removed as **additional parameters**. All that remains is to decide what prior distribution these two new parameters should follow. We know that these parameters are strong correlated, so it doesn't seem a particularly smart decision to choose two independent distributions. On the contrary, we can use a two-dimensional **logistic normal**

distribution. This is a generalization of the logit-normal distribution to D-dimensional probability vectors. In this context, we have to consider a multivariate normal distribution and then take a logistic transformation of them. In particular, we are dealing with:

$$x = (I_0, R_0, S_0),$$

so the vector $\boldsymbol{x}$ has 3 components whose sum is equal to 1. Now, we can consider the following transformation:

$$y = \left(\log\left(\frac{I_0}{S_0}\right), \log\left(\frac{R_0}{S_0}\right)\right), \qquad \boldsymbol{y} \in \mathbb{R}^{D-1}.$$

We know that $\boldsymbol{y} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the probability density function of $\boldsymbol{x}$ is:

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp\left(-\frac{1}{2}\left[\log\left(\frac{\boldsymbol{x}_{-D}}{x_D}\right) - \boldsymbol{\mu}\right]' \boldsymbol{\Sigma}^{-1} \left[\log\left(\frac{\boldsymbol{x}_{-D}}{x_D}\right) - \boldsymbol{\mu}\right]\right), \qquad \boldsymbol{x} \in \mathbf{S}^D.$$

Here $\boldsymbol{x}_{-D}$ denotes the vector of all $\boldsymbol{x}$ components except the last one and $\mathbf{S}^D$ is the simplex of D-dimensional probability vectors. In this specific setting $D = 3$.
Now, the last step is choosing the hyper parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$. The idea behind this choice is to relate the logistic normal distribution with the Dirichlet distribution. It can be proved that if we use some moment properties of the Dirichlet distribution, these parameters can assume a form related with the parameter $\boldsymbol{\alpha}$ of the Dirichlet distribution. More information can be found at [18], [19], and [20]. In particular, this theoretical information are implemented in R in B.1.

7.4 Deterministic SIR

To begin with, we decide to model our data with a *deterministic SIR*. We already discussed all the criticisms of using this kind of approach when we model an epidemic process. In fact, we are ignoring all inherent stochastic components. Anyway we try to compensate, this deficiency by adding some noise with the parameter τ, which was already introduced. In the implementation, we have to follow exactly the same steps as in Chapter 6. The only difference is in the filling of matrix $\boldsymbol{X}$. When we deal with simulated data there is a preliminary task which consists into data generation. Now we already have on hands the data and the only requirement is import and prepare these data in such a way they will be feasible for the algorithm implementation. All necessary procedures are implemented in B.2. To summarize what we have already discussed, we recall that now also the initial conditions are parameters to be estimated. So the total number of parameters now is 13. Even if the model was already described in Section 6.2, we need to add all the transformations and changes connected with initial conditions. These adjustments are contained in B.3.
In Figures 7.7 and 7.8, we can observe the achievement of stationary values by the chains. The first image shows only the infection rates. One infection rate for each point over the grid. We also plotted and displayed the median value in order to get a better idea of what happens in this situation. From this analysis, it is clear that $\lambda(t)$ starts from an high value

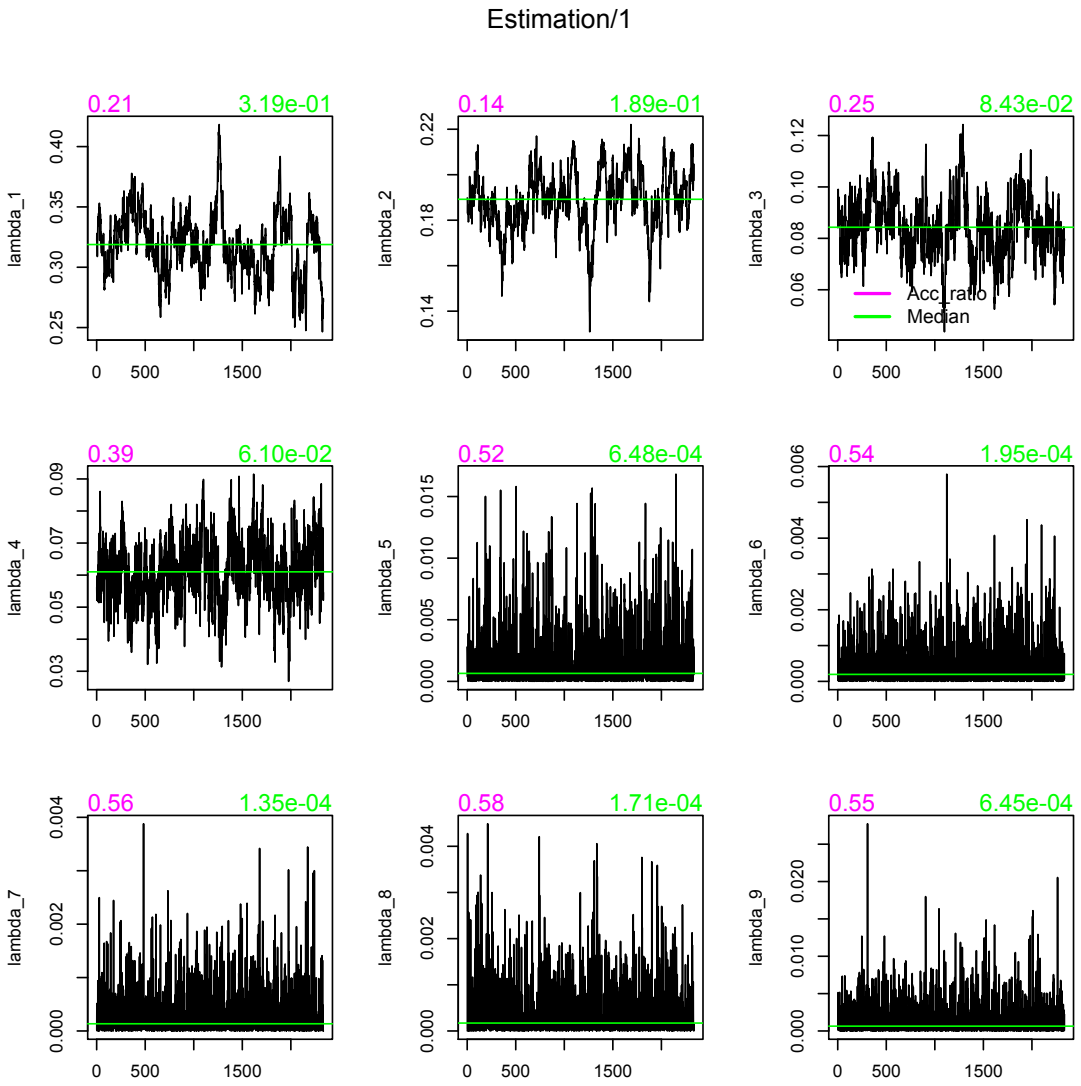


Figure 7.7: The evolution of different infection rates over the grid when stationarity is reached for a deterministic SIR for real data.

and it is always decreasing in this time period. The only exception is the last point over the grid. There's a minimal growth there. It corresponds to the second half of June 2020. Thus the infection rates show a quite reasonable trend. Indeed, at posterior we expect that at the beginning of the pandemic the infection rate assumes an high value. Then because of social distancing, masks, quarantine, lock down, and all the others preventive measures, we believe that this value has to decrease. Lastly, we know that at the end of the summer, the conditions were in place for the outbreak of the second wave. So also the last value seems plausible. Anyway, to better understand if they are reasonable values,

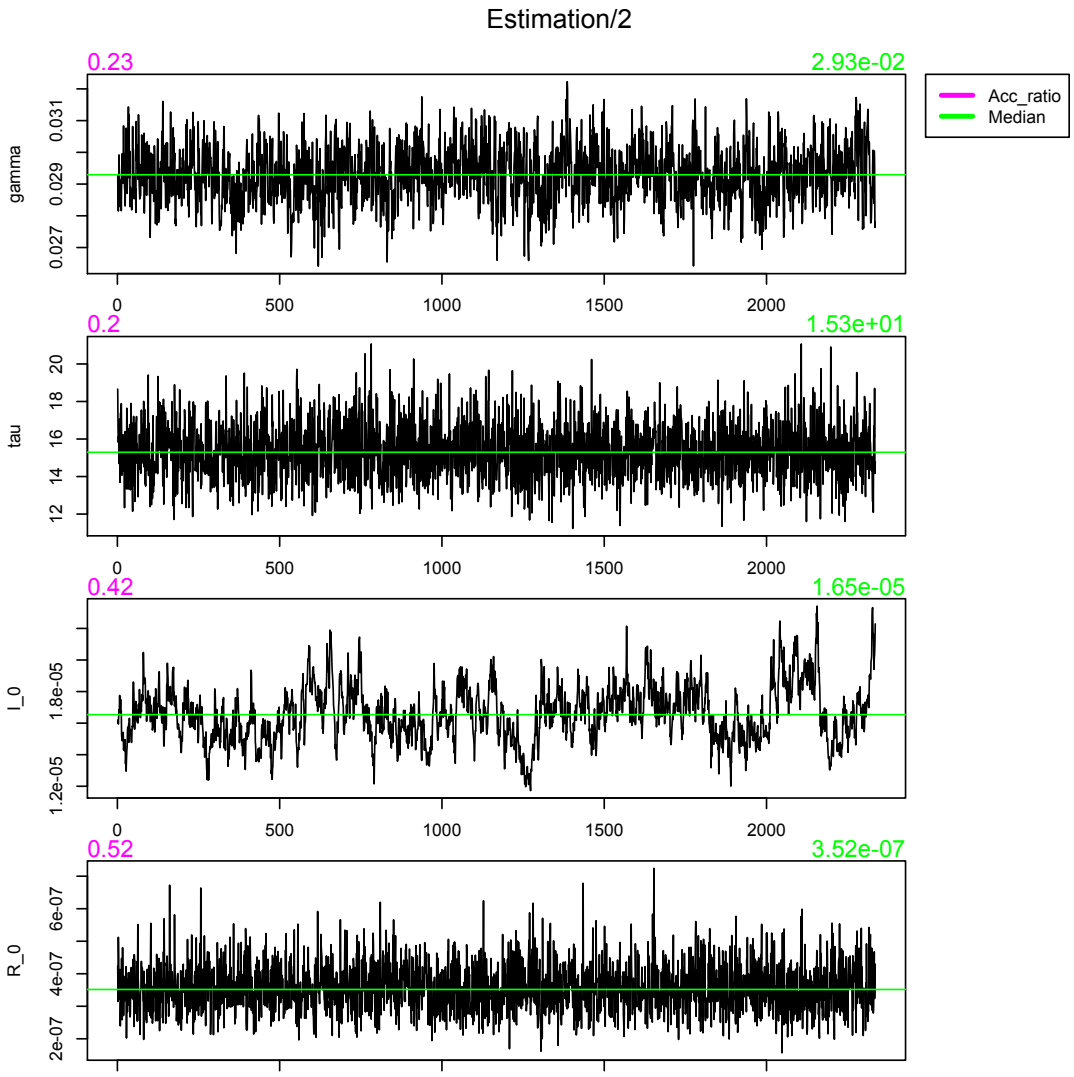


Figure 7.8: The evolution of different parameters when stationarity is reached for a deterministic SIR for real data.

we can observe Figure 7.9. Here the effective R_t index is shown. In order to obtain this figure firstly we linearly interpolate the previous $\lambda(t)$ obtained from the estimation. Then for each day, we divide this value for γ which is the recovery rate. We can notice in Figure 7.8 that its median is almost 0.03. If we consider the reciprocal of this value, we can conclude that the time it takes for an infective person to heal is something more than 30 days. It seems a quite high rate compared with what we would expect. Anyway, we can assume that this is due to different factors such as measurement errors. In fact, if we observe the second plot always in the same image, τ assumes a quite small values and we

expect a thick band built with two quartiles. Finally the variations of the new chains for the initial conditions can be analyzed. We can observe some correlations in that one of I_0 that may arise some problems. In Figure 7.10 again the comparison between real data and our estimation is shown. We can see that the median is often overlapping the real curve, even if we obtained a quite reasonable band with the two quartiles. Just for sake of completeness in Figure 7.11 is shown the behavior of the removed individuals. We can draw the same conclusions from this image as from the previous one.

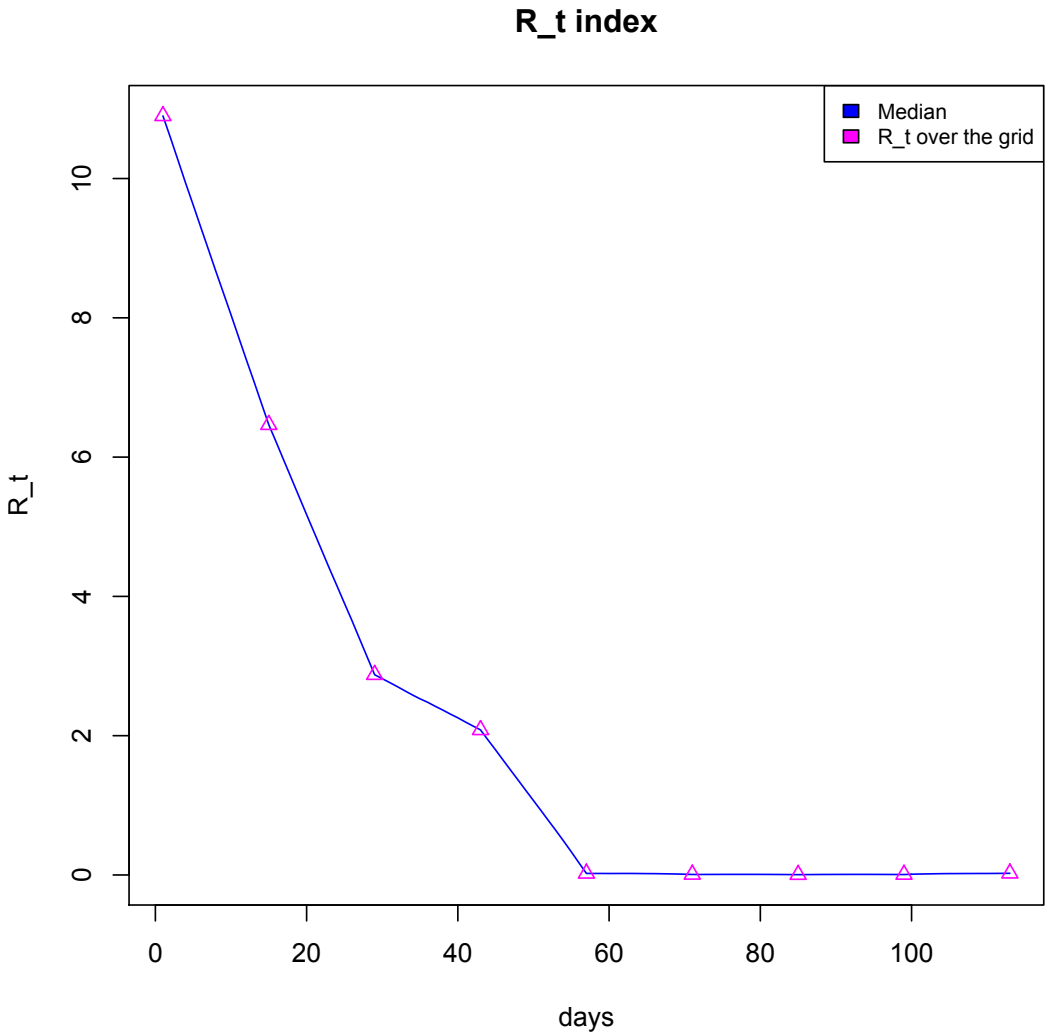


Figure 7.9: Evolution of R_t index over time estimated using a deterministic model.

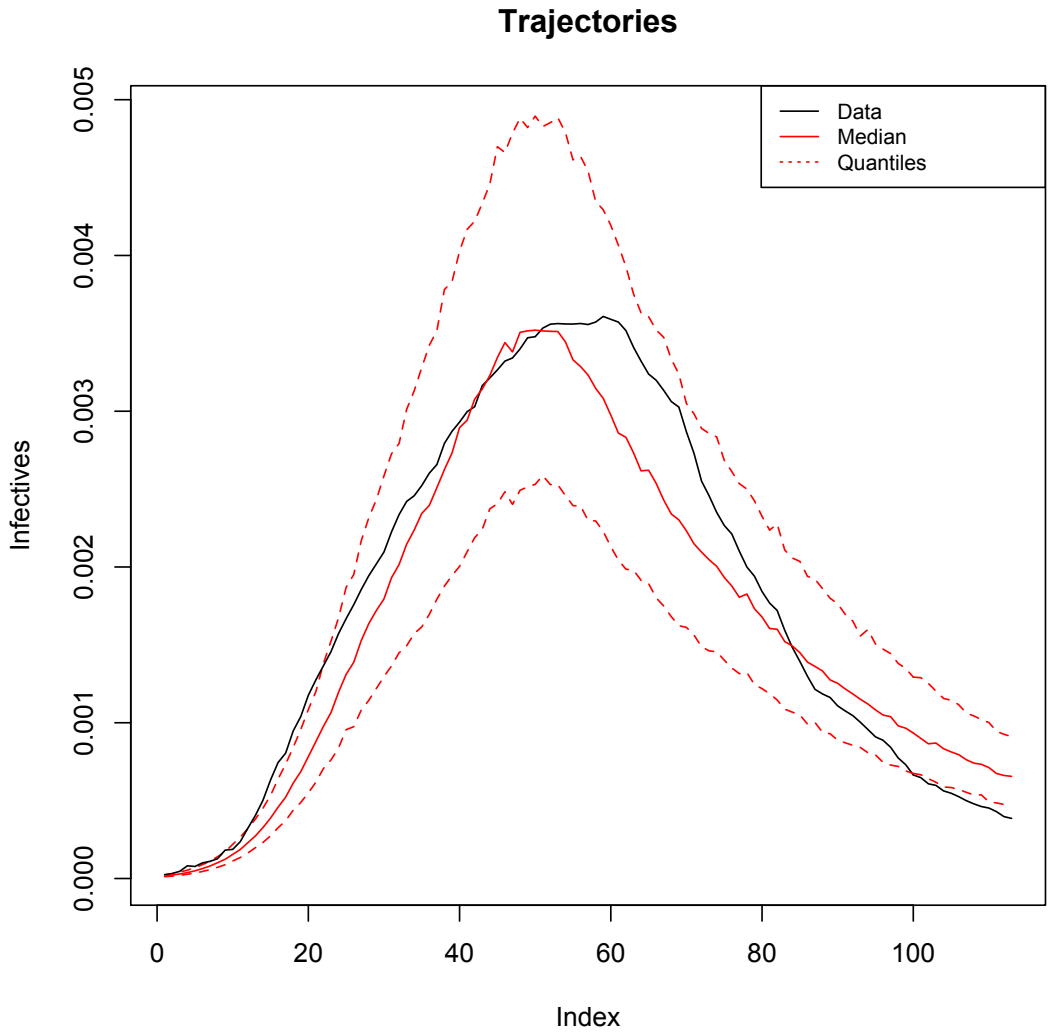


Figure 7.10: Changes in the proportion of infective individuals and the trajectory obtained with the median and two quantiles (the first and the third quartiles) evaluated on the results of the deterministic SIR when real data are considered.

7.5 Stochastic SIR

As already discussed, a different strategy consists in modeling our data via a *stochastic SIR*. To this aim we have to solve a group of stochastic differential equations and then use the Euler-Maruyama approximation to simulate the trajectories. Also in this case, the implementation is already described in Chapter 6. Anyway, in order to use real data, we have to import and prepare them to fill the matrix $\boldsymbol{X}$ rather than generate simulated data, just like in Section 7.4.

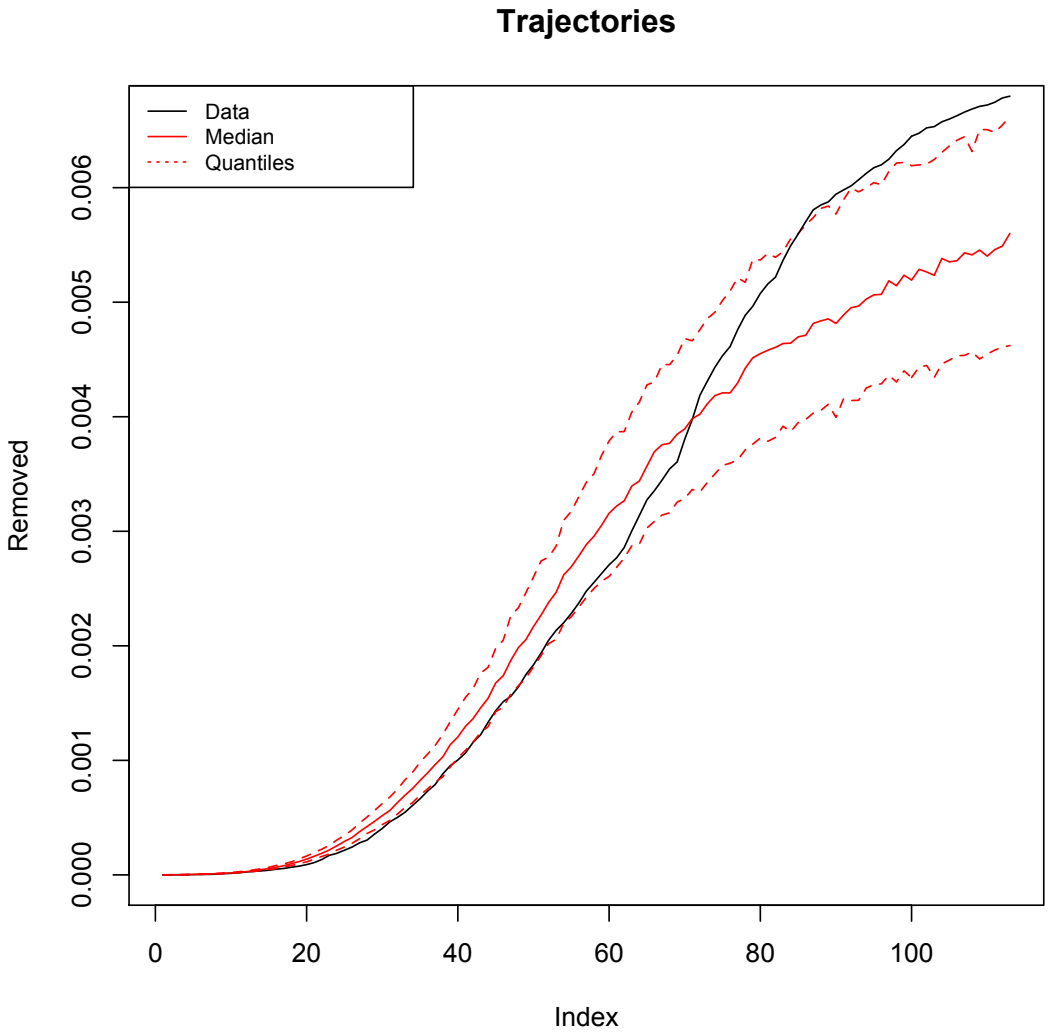


Figure 7.11: Changes in the proportion of removed individuals and the trajectory obtained with the median and two quantiles (the first and the third quartiles) evaluated on the results of the deterministic SIR when real data are considered.

If we decide to follow the stochastic approach, we can choose between two alternatives. The first one consists into fixing the value of parameter $\nu = 1$. Indeed we can assume that the evolution of a Markov chain is able to explain all the variability in the data. This might sound as a quite radical approach. For this reason, the alternative technique could be considering ν as a parameter. In this way, we create a more elastic model which should be able to detect added noise. In both situations, we decide to consider the two initial conditions as additional parameters. As in the previous case, we assume that their

appropriate transformations follow a two-dimensional normal distribution such that they are described by a logistic-normal distribution. Again the implementation code is exactly the same as the simulated case. Anyway, some modifications were necessary in order to consider initial conditions as additional parameters. The new version of functions involved in the stochastic approach can be found at B.4 and B.5.
In our first attempt, we decide to apply the model which comes directly from the theoretical analysis. For this reason, we decide not to consider ν as a parameter and we fix it at 1. As a consequence we have to estimate 12 parameters. The 9 infection rates in the same points over the grid already built (in Figure 7.12), the recovery rate γ and the two initial conditions I_0 and R_0 (in Figure 7.13). By a quick comparison, we notice that the values of infection rates are higher than the parameters obtained with a deterministic model. On the other side, the value of γ is quite similar to the deterministic one. Therefore the index value will be higher than before, at least in the initial period. Its behavior can be observed in Figure 7.14. Another observations we can make is about the chains of initial conditions. In Figure 7.13, we can notice that the median value of I_0 is again similar to the deterministic one even if now the problems of auto-correlation seem to be solved. On the contrary the median of R_0 now is smaller than an order of magnitude. In Figure 7.15 there is the comparison between our stochastic estimation and real data about the infective individuals. While in Figure 7.16 there are the results for the removed individuals. In both cases, the median of our trajectories is near to the real curve, even if the model seems rather rigid. Indeed it is not able to detect all the quick variations. This is why, we decide to implement the same model, but now ν is an other parameter.
In Figures 7.17 and 7.18 we can see the estimation of all parameters. The parameters $\lambda(t)$ and γ assumes quite the same values as before. Now the initial conditions are quite similar to the respective values obtained with the deterministic model. However the median of ν is very far from 1, which is the value we expected. In particular, the median of ν is about 0.02, which is a very small value. Indeed from Figure 7.19 and 7.20 we can conclude that now the model is more elastic than the previous case. However there is so much freedom and variability that this approach isn't so good for our purposes. Lastly, Figure 7.21 shows the trend of R_t in this last situation. Its values don't seem so far from those obtained before. In order to compare the different effective R_t evaluated using all three models, we plotted Figure 7.22.

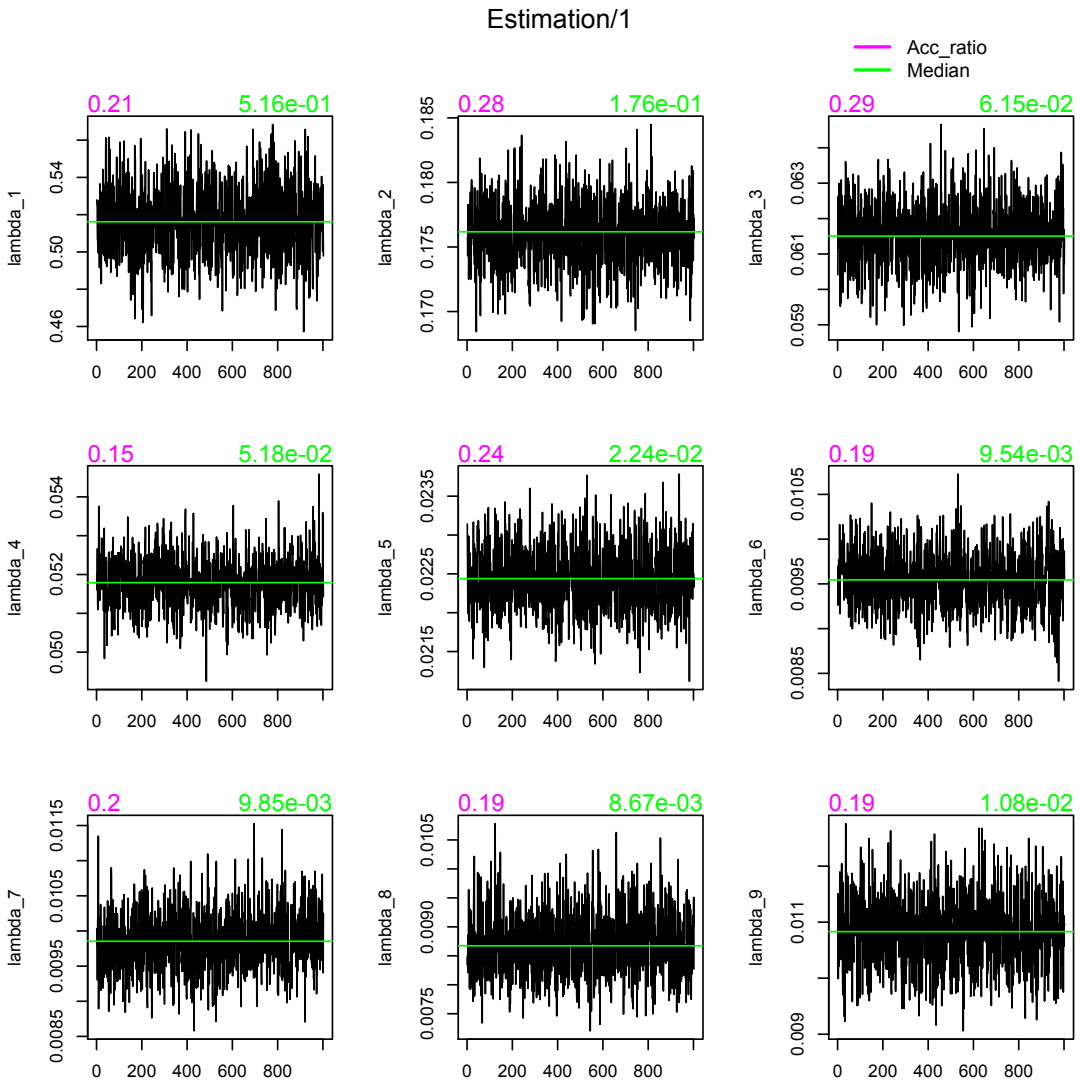


Figure 7.12: The evolution of different infection rates over the grid when stationarity is reached for a stochastic SIR (with $\nu = 1$) for real data.

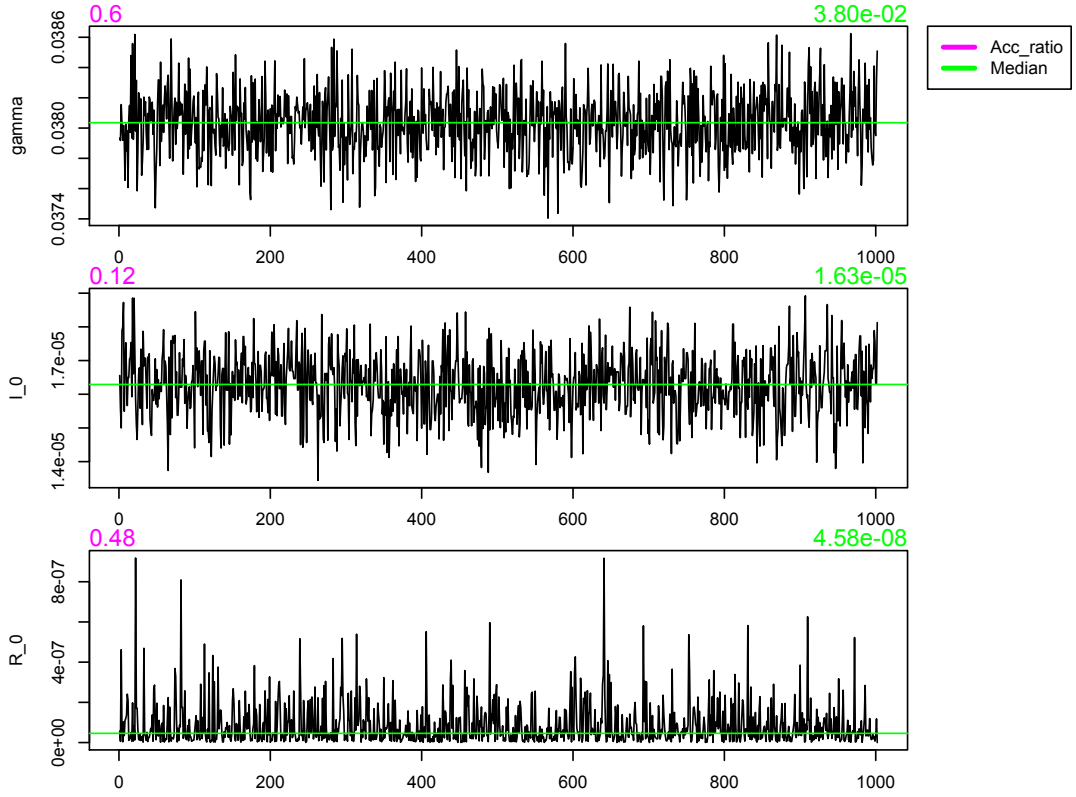


Figure 7.13: The evolution of different parameters when stationarity is reached for a stochastic SIR (with $\nu = 1$) for real data.

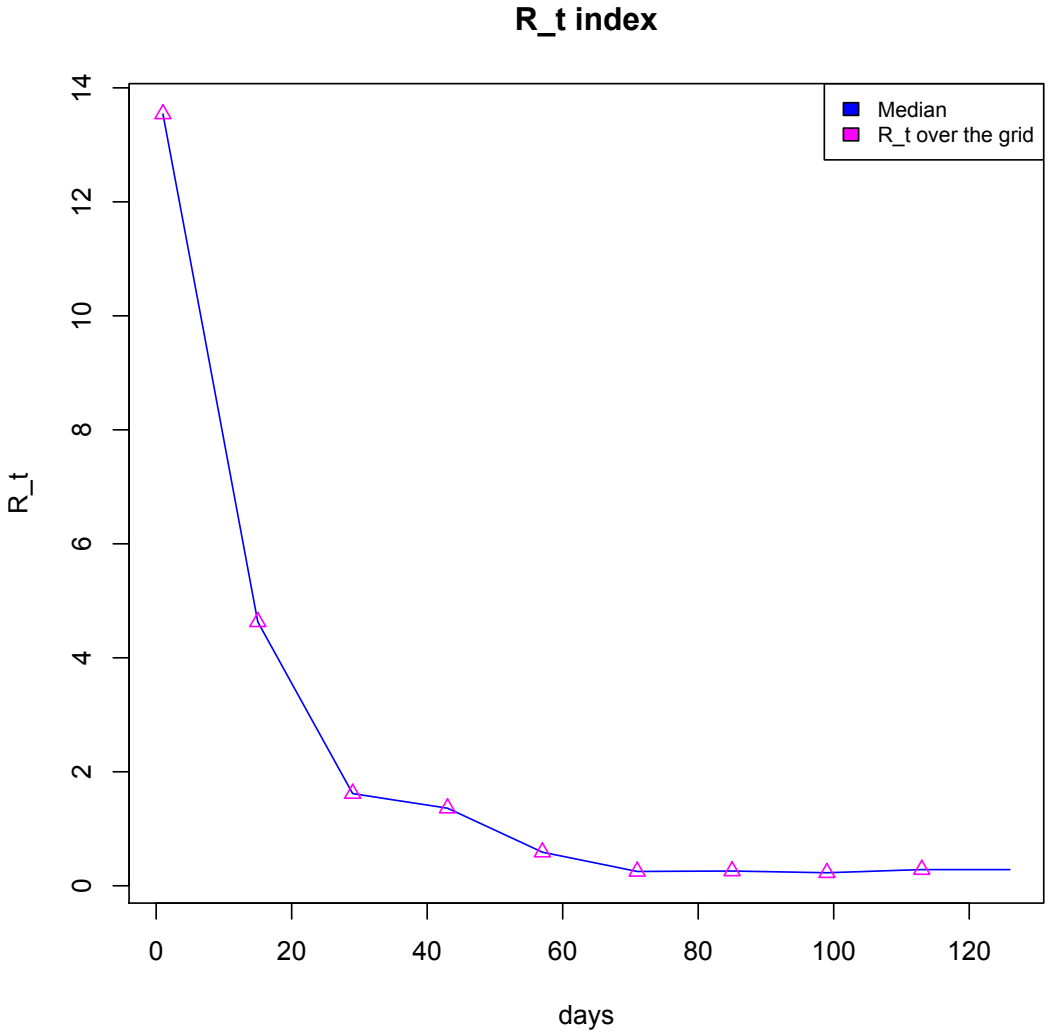


Figure 7.14: Evolution of R_t index over time estimated using a stochastic model with $\nu = 1$.

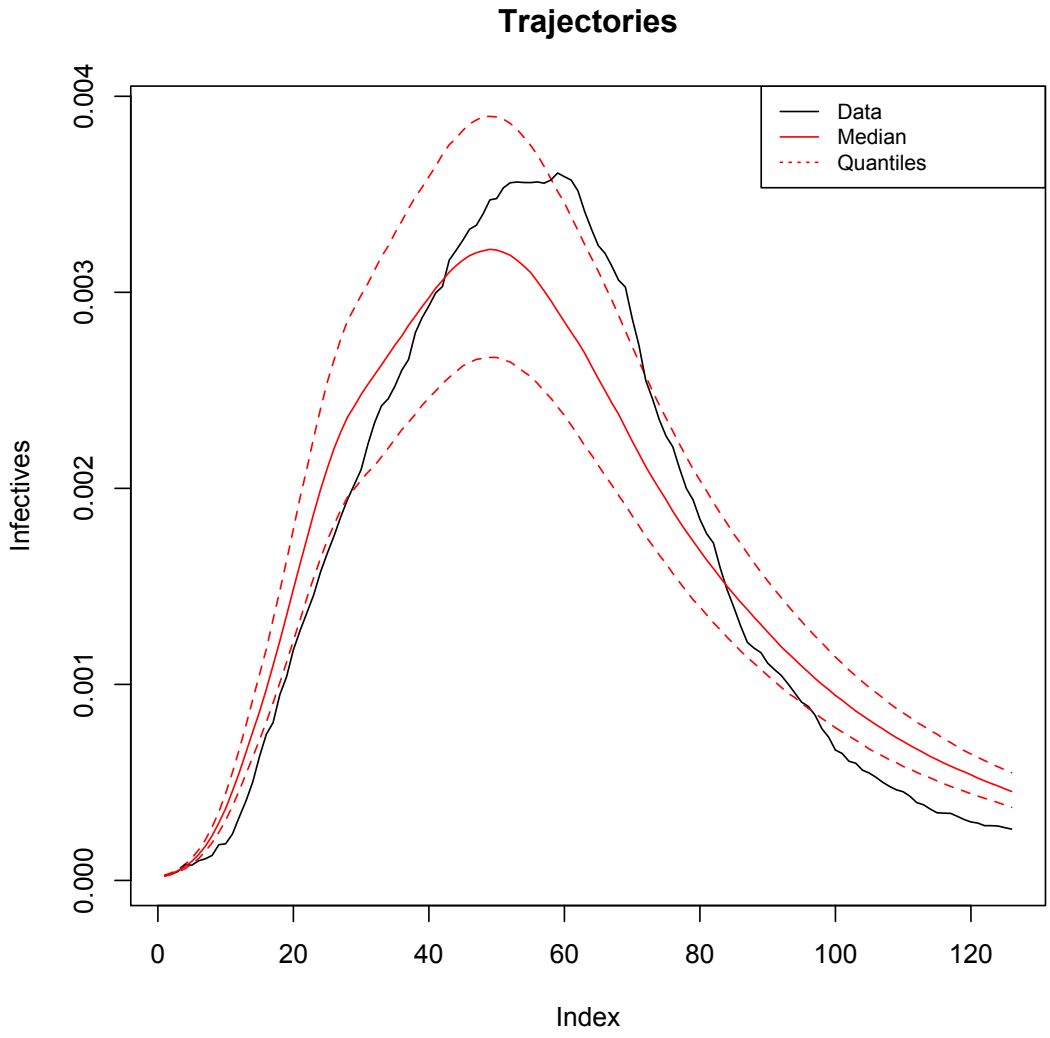


Figure 7.15: Changes in the proportion of Infective individuals and the trajectory obtained with the median and the first and the third quartiles evaluated on the results of the stochastic SIR and $\nu = 1$ when real data are considered.

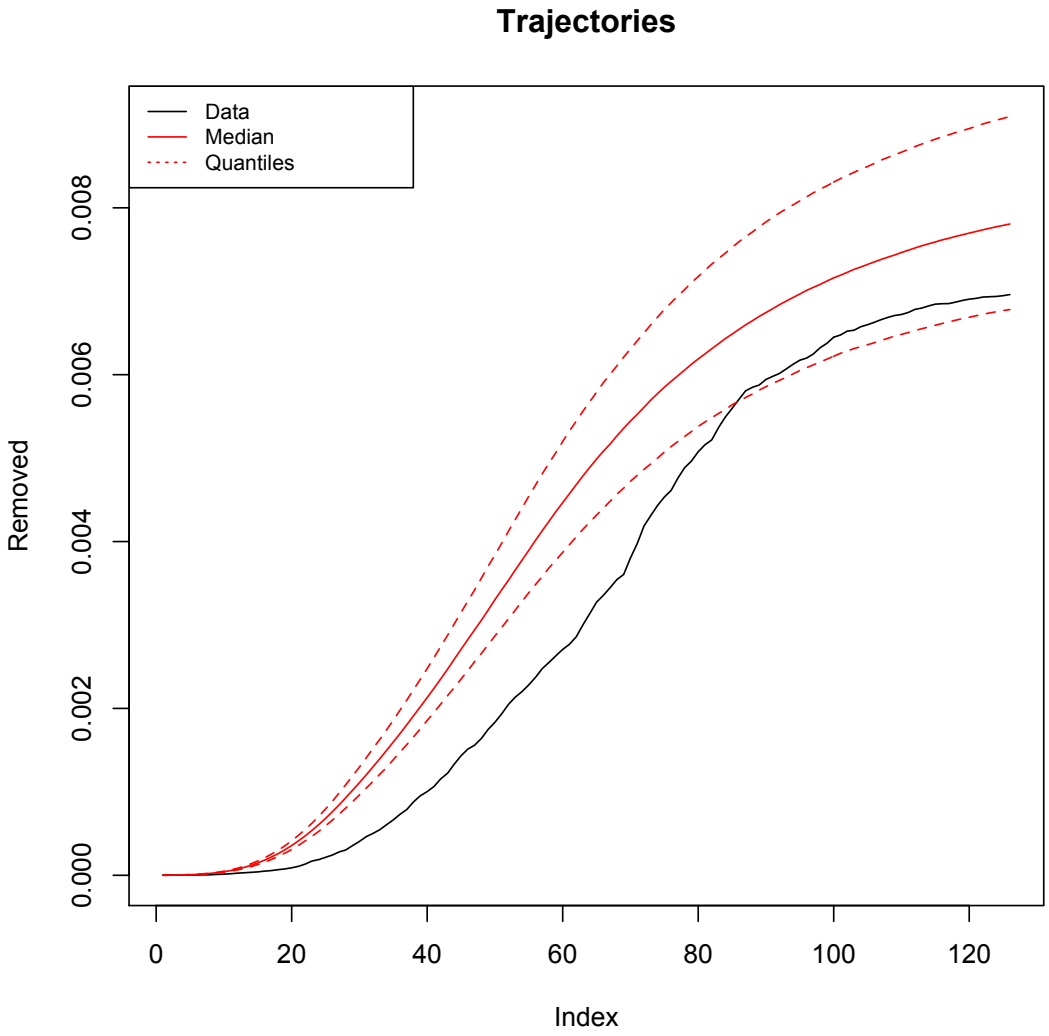


Figure 7.16: Changes in the proportion of Removed individuals and the trajectory obtained with the median and two quantiles (the first and the third quartiles) evaluated on the results of the stochastic SIR and $\nu = 1$ when real data are considered.

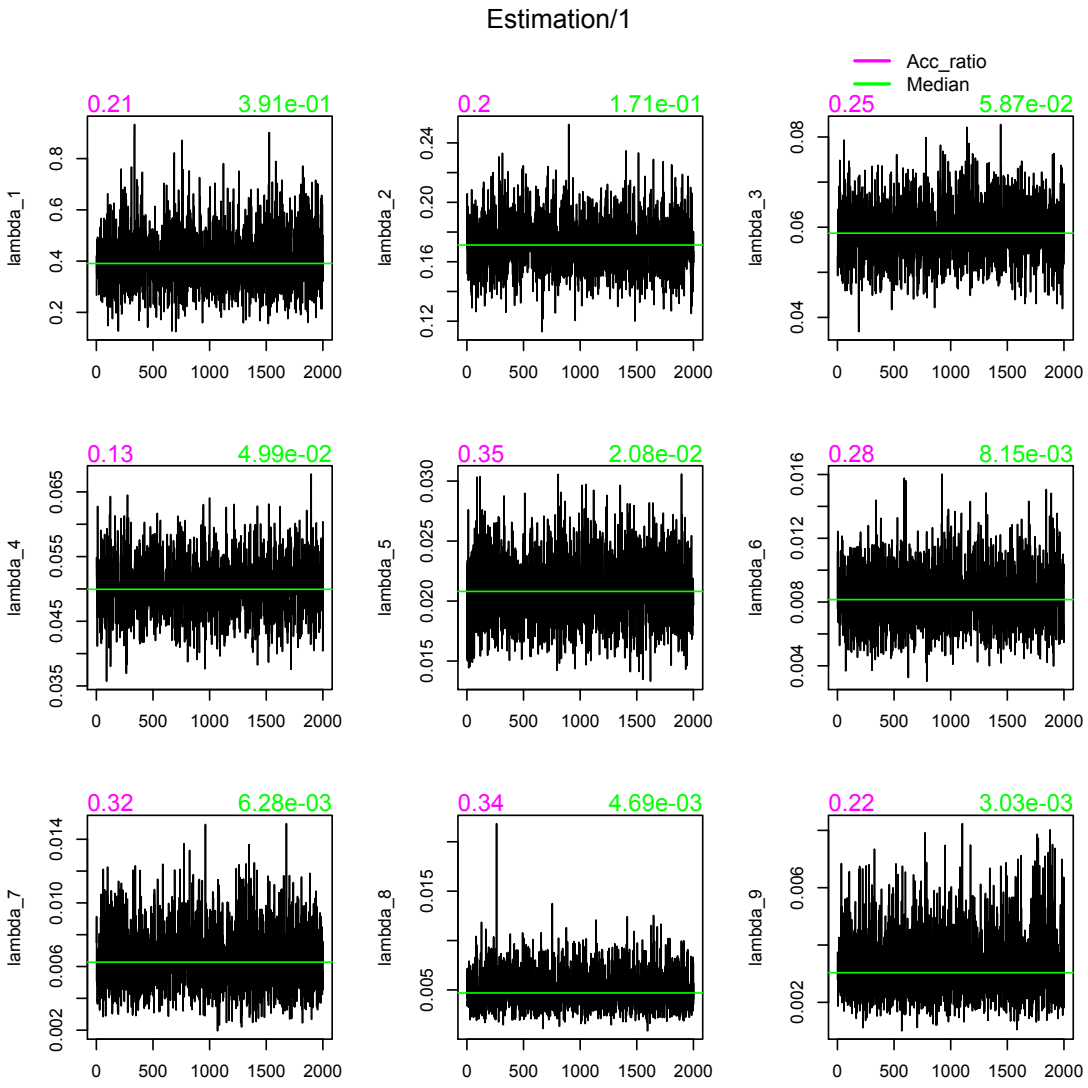


Figure 7.17: The evolution of different infection rates over the grid when stationarity is reached for a stochastic SIR for real data.

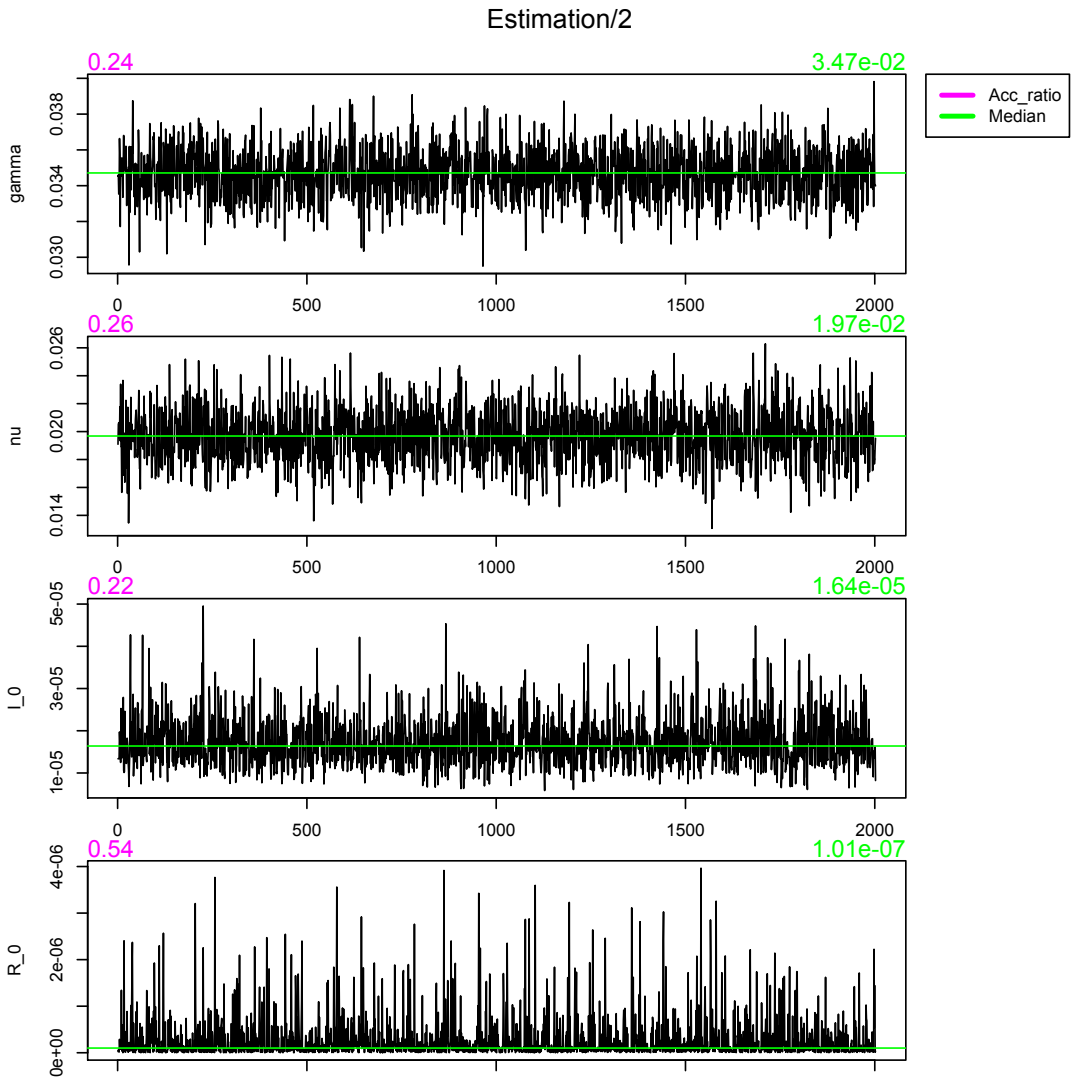


Figure 7.18: The evolution of different parameters when stationarity is reached for a stochastic SIR for real data.

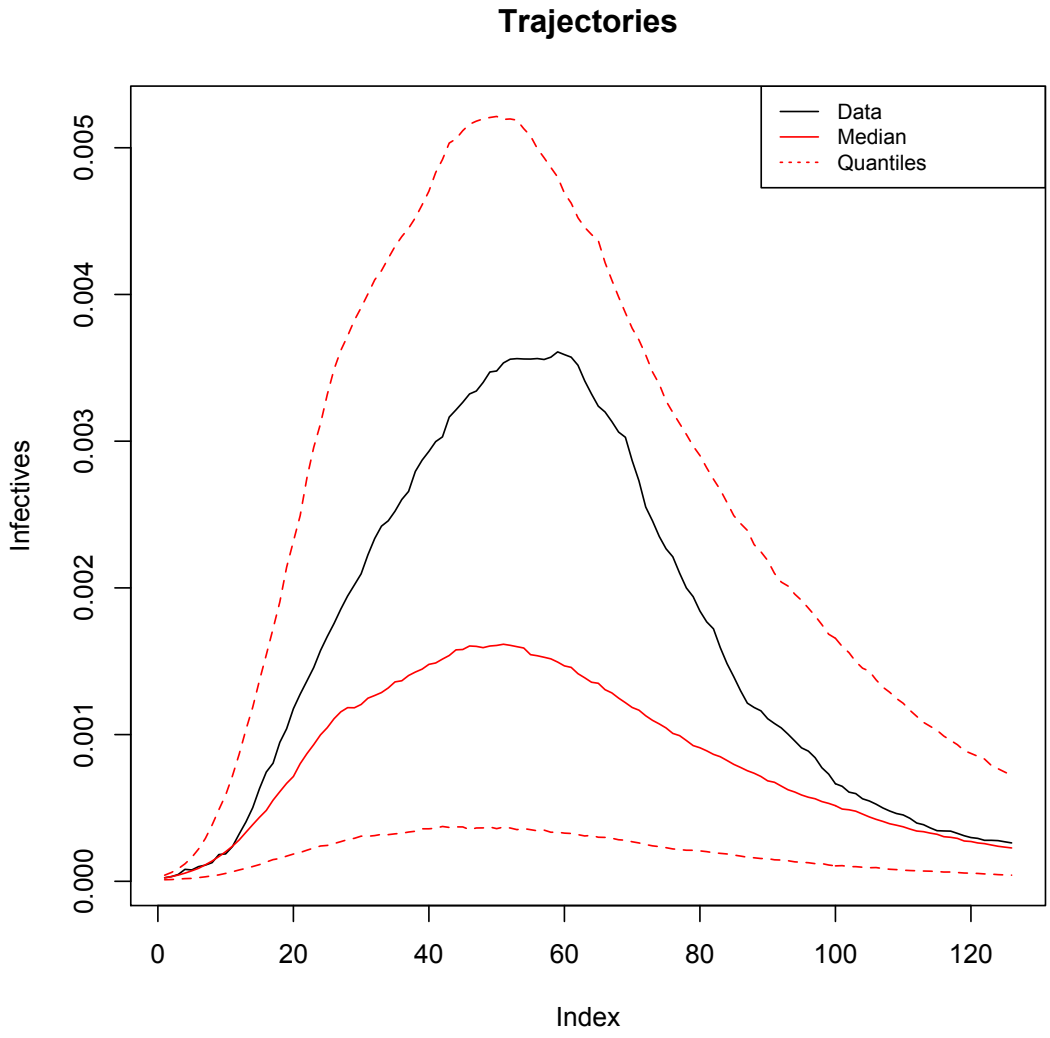


Figure 7.19: Changes in the proportion of Infectives and the trajectory obtained with the median and two quantiles with probability 0.25 and 0.75 evaluated on the results of the stochastic SIR when real data are considered.

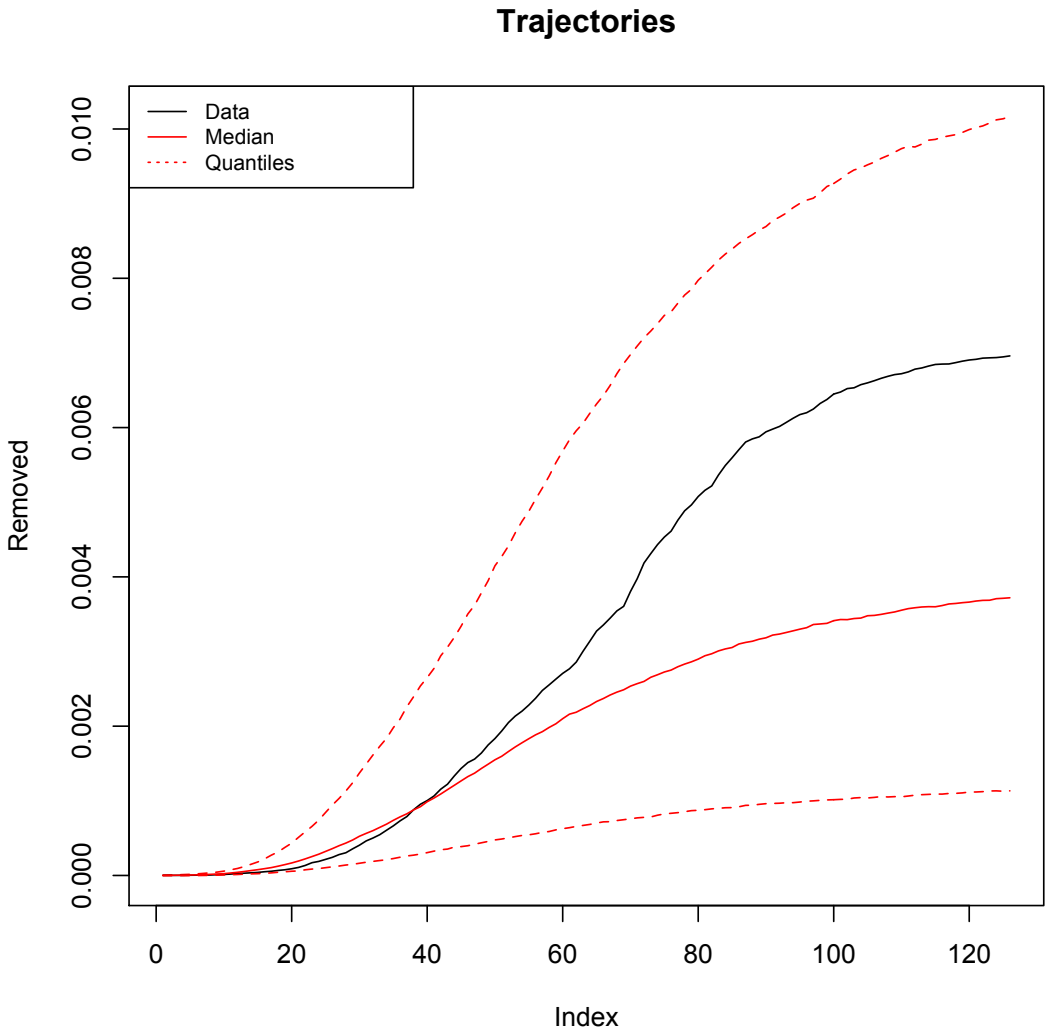


Figure 7.20: Changes in the proportion of removed individuals and the trajectory obtained with the median and two quantiles with probability 0.25 and 0.75 evaluated on the results of the stochastic SIR when real data are considered.

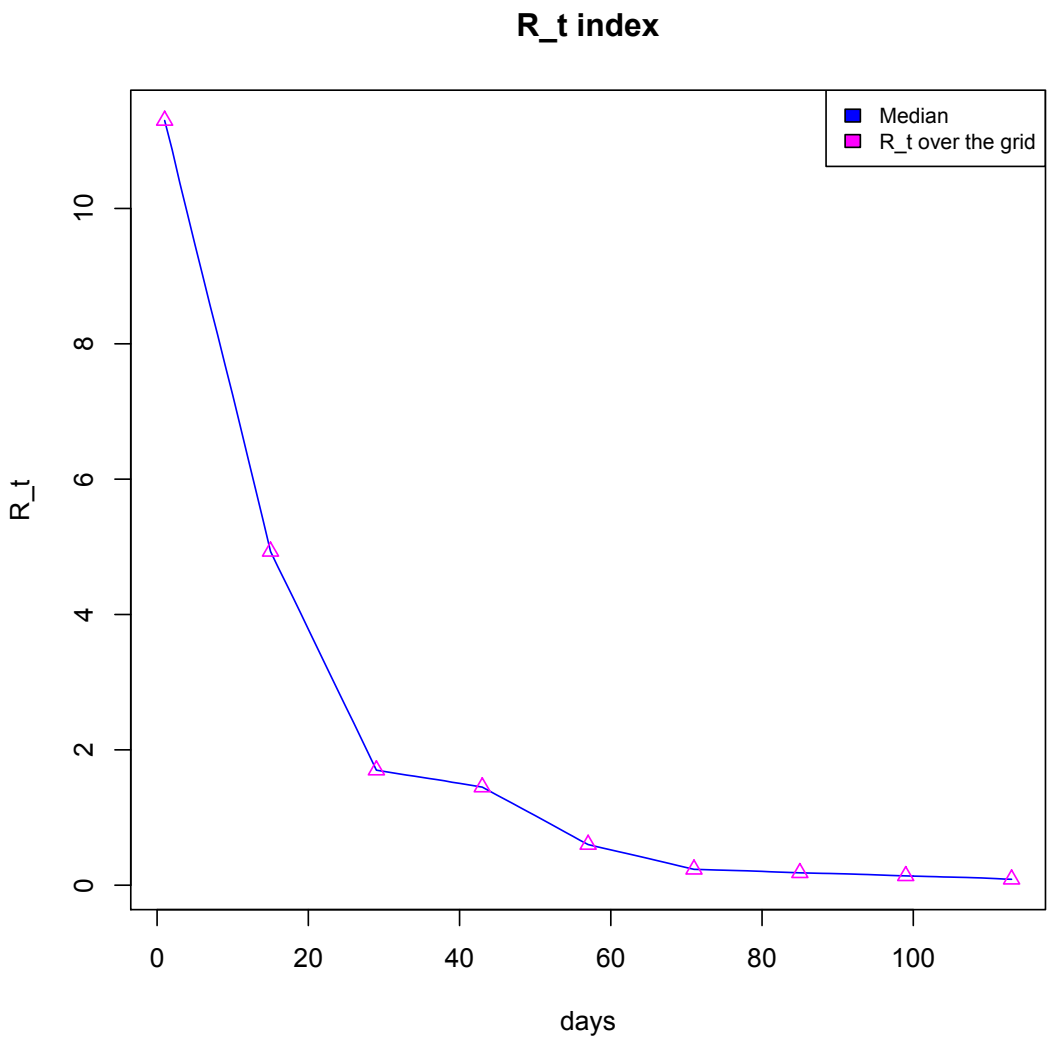


Figure 7.21: Evolution of R_t index over time estimated using a stochastic model.

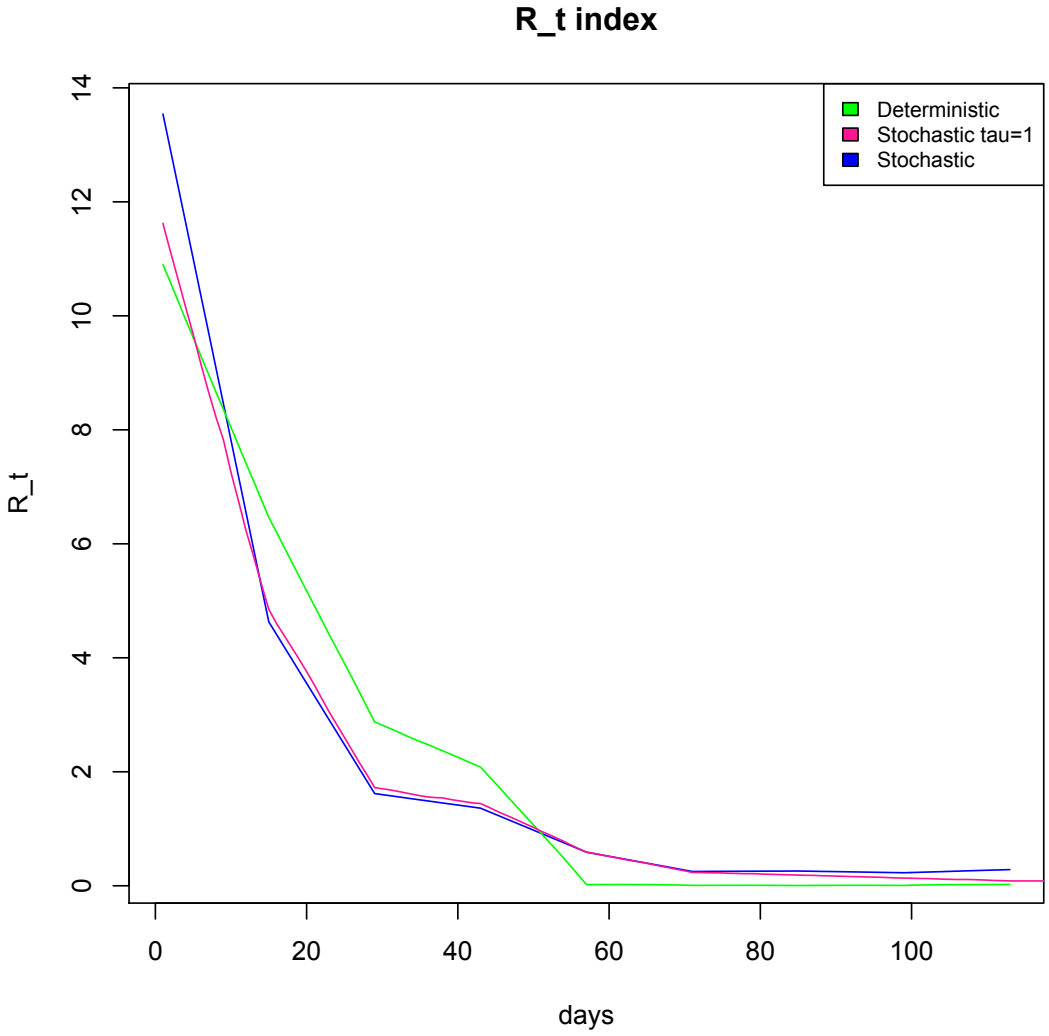


Figure 7.22: Different evolution of R_t using real data in Piemonte and estimated with a deterministic model, a stochastic one, and a stochastic one with $\nu = 1$.

Chapter 8

Conclusions

Nowadays, infectious diseases play a very important role in everyday life. The outbreak of the COVID-19 pandemic stressed this point. Thus, it is really important to find a mathematical technique which is able to model properly this kind of phenomena.
Of all available epidemic models, we decided to describe the COVID-19 pandemic with a SIR (Susceptibles - Infectives - Removed). Despite all the required simplifications and limitations of the model, we were expecting to obtain quite reasonable results. Indeed one of the main simplification is considering a closed population which can be divided only in three groups. This is why, at the beginning we assume that all the population (except the initial Infective) is Susceptible. Anyway, the previous assumption doesn't feel too demanding. Indeed, especially at the beginning, vaccines were not widespread. In addiction, there is still no certain evidence about some genetic characteristics which render immune an individual. Another problem about the SIR model is that an individual doesn't play any further role once the infection ends. First of all, we are not searching for a distinction between *recovered* individuals and *deaths*. In this model, no matter what they are considered as Removed and they are no longer able to contract the disease. From the experience, we know that this isn't a right assumption. It is not clear how long immunity lasts after contracting the disease. The number of individuals who is reinfected in fact, is always increasing. Anyway, also in this case the hypothesis almost seems correct. We are considering only the first wave of pandemic, so we can assume that in this setting the assumption of immunity remains true. Another simplification, is that there is no distinction between those who needed hospital care, those who had not serious symptoms, and asymptomatic individuals.
In this thesis, we analyzed in detail two different techniques in order to apply a SIR model. Indeed, the main difference lies in the way we define the equations. They determine how an individual can pass from one group to the other. The first option is thought Ordinary Differential Equations. It may sounds like a strange choice because of the inherit stochasticity in the process. Anyway this model is simpler (even if not necessarily simple) than a stochastic epidemic model. This latter option needs to be quite easy in order to be mathematically manageable. As a consequence the stochastic model may not be exactly realistic. On the other hand, there are many advantages when we decide to use a stochastic approach. For example, in modeling processes that are genuinely stochastic or

in the estimations. Anyway, we can decide to add some noise to the deterministic model in order to somehow "dirty" the results and add variability. The data we have at our disposal are indeed noisy. Moreover, we proved that under particular assumptions, when the population size is large, the stochastic model converges to a deterministic one. Thus, we can conclude that they are both useful tools in order to analyze a process. Indeed, both of them are able to describe the phenomenon from a new prospective providing important information.
With regard to data considered in the implementation, we have to recall all the possible sources of error such as measurement error, transcription error, delay and seasonal trend due to analysis process and so on. In addiction, we are ignoring a lot of determining factors because of lack of information. All in all, this is the best available dataset and we rely on it. In particular, we considered the situation in Piemonte. In fact, it can be observed that the general trend is fairly similar among all Italian regions. Anyway, each one of them has some specific parameters and speeds with which the results are obtained.
At the end, we can analyze the results achieved by different methods. The deterministic method produced what appears to be the best results. Here we considered an additional parameter τ as noise. We can affirm that despite all simplifications, we obtained quite accurate results which are able to reproduce what happens in reality. The main criticisms can be met during the first and the very last period of our time window. There, the most rapid changes happen. On the remaining period the real curve of infected is contained in the bandwidth and also the median value is really near to it.
Meanwhile, with the stochastic approach we found some difficulties in mapping the real behavior of the epidemic. Indeed there seem to be problems in finding and reaching the real value of parameters. In the first stochastic model, we don't consider ν. It seems quite a rigid model. Indeed, even if the median is nearly always *close* to the observed data, we can notice that for a considerable time period they are out of the band. In addiction, not even the shape of the curve seems to be recognized by the model. This means that the noisy contained in the Markov chain is not compatible with what we observed. On the other hand, if we consider ν as a variable, the model is too free. There is a lot of variability in our observations and we are not able to reach an overlapping curve. We obtained that the real curve is always contained in the band built with the first and third quartiles. However the band is too large and they are not reliable results.
One of possible future improvements would be using a stochastic model with added noise. Indeed, when we consider a stochastic model, there are two different kind of noises. On the one hand, there is the inherit noise which is modeled by a Markov chain. It tries to capture the true variability in our data. Ideally it should be equal to one, even if we notice that there may be an amplification of it when the model isn't able to explain real data. On the other hand, just like the deterministic case, there is an "measurement" error. This latter kind of noise is due to the fact that we have not many information about the process. So a possible strategy could be consider an additional parameter of noise. Hopefully, the improved model should be able to distinguish these two different sources of noise and it should provides better results. Anyway, at the same time, we are adding a new degree of freedom to the system, so we are not sure about really improvements.
With all the previous simplifications and (limiting) assumptions of the specific situation, we can conclude that the deterministic model should be applied in order to obtain the

more reliable results. They can be used not only from a mathematical point of view. Indeed based on these results, also some prediction about possible future realizations can be made. The future number of infected individuals could be predicted with the aim of being prepared to provide the best possible healthcare. At the same, they could be used to understand the effectiveness of a preventive measure in order to slow down or avoid the outbreak of the epidemic process.

Appendix A

R scripts for simulated data

A.1 Resolution of ODEs

```r
#ODESol

#Function to solve a deterministic SIR

library(deSolve)

closed.sir.model <- function (t, x, params, time_grid_beta) {
S <- x[1]
I <- x[2]
R <- x[3]
llltimes<-length(time_grid_beta)
if(length(params)!=llltimes+1) stop("The length of time_grid_beta should
    be one less than that of params")
beta_p <- params[1:llltimes]
beta_f<-approxfun(x=time_grid_beta,y=beta_p,yright=beta_p[llltimes]
   ,yleft=beta_p[1])
mu <- params[llltimes+1]
## now code the model equations
dSdt <- -beta_f(t)*S*I
dIdt<- beta_f(t)*S*I-mu*I
dRdt <- mu*I
dxdt <- c(dSdt,dIdt,dRdt)
list(dxdt)
}

odesolution <- function(f, parameters, initial_cond, days,
   time_grid_beta){
sol <- ode(func=f,y= initial_cond, times= 1:days, parms=exp(parameters
   [1:10]), time_grid_beta=time_grid_beta)
t(sol[,c("I","R")])
}
```

Script : Function to solve a system of ODEs for deterministic SIR.

A.2 Deterministic data generation

```
#Det_DataGen
source("ODESol.R")

#DATA GENERATION
days <- 126
time_grid_data <- 1:days
step <- 14
time_grid_beta <- seq(1, days , step)
initial_cond <- c(S=0.9999, I=6.89e-7, R=1e-20)
data <- matrix(0,nrow=2,ncol=days)

tau <- 200
gamma <- 1/7
beta_grid <- gamma*c(3,1.5,1.1,0.95,0.9,0.85,0.85,0.9, 0.5)
true_par <- c(beta_grid, gamma, tau)
parameters  <- log(true_par)

data <- odesolution(closed.sir.model, parameters, initial_cond, days,
    time_grid_beta)

for (i in 1:2){
  for (j in 1: days){
      data[i, j] <- exp(rnorm(1, mean=log(data[i, j]), sd=sqrt(1/tau)))
  }
}
```

Script : Deterministic data generation.

A.3 MCMC functions for deterministic SIR

```
#DetMCMC
source("ODESol.R")
loglikelihood <- function(f, parameters, initial_cond, days,
    time_grid_beta){
odesol <- odesolution(f, parameters, initial_cond, days, time_grid_beta)
if(sum(odesol<=0)==0) sum(dnorm(lYmat, mean=log(odesol),sd= sqrt(1/exp(
    parameters[par.num])),log=TRUE)) else -Inf}

logpriorllambda <- function(lbetas) sum(dnorm(lbetas,mean=0,sd=sqrt(100)
    ,log=TRUE))
logpriorlgamma <- function(lmu) dnorm(lmu,mean=0,sd=sqrt(100),log=TRUE)
logpriorltau <- function(w) w*0.01-0.01*exp(w)

fc_llambda <-function(f, parameters, initial_cond, days, time_grid_beta)
    {
loglikelihood(f, parameters, initial_cond, days, time_grid_beta)+
    logpriorllambda(parameters[1:(length(parameters)-2)])}
fc_lgamma <- function(f, parameters, initial_cond, days, time_grid_beta)
    {
loglikelihood(f, parameters, initial_cond, days, time_grid_beta)+
    logpriorlgamma(parameters[length(parameters)-1])}
```

```r
fc_ltau <- function(f, parameters, initial_cond, days, time_grid_beta){
loglikelihood(f, parameters, initial_cond, days, time_grid_beta)+
    logpriorltau(parameters[length(parameters)])}
```

Script: Definition of MCMC functions for deterministic SIR.

A.4 Deterministic SIR with simulated data

```r
rm(list=ls())
source("Det_DataGen.R")
source("DetMCMC.R")

lYmat <- log(data)
post.samp.size <- 6000 #number of stored values
thin <- 20
par.num <- 11
llambda <- par.num-2
sd.proposal <- c(0.02, 0.008, 0.01, 0.003, 0.03, 0.01, 0.02, 0.04, 0.09,
     0.004, 0.5)
chain <- matrix(rep(0,par.num*post.samp.size), byrow= TRUE, nrow=
    post.samp.size)

old <- new <- chain[1, ] <- rep(log(2), par.num) #arbitrarily choosen
    initial values
acceptances <- rep(0,par.num)

for (iter in 2:post.samp.size){
  if (iter%%100==0) print(iter)
    for (inner in 1:thin){
      old <- new
      for (i in 1:par.num){
        updated.pars <- new
        candidate <- old[i]+rnorm(1,mean=0,sd=sd.proposal[i])
        updated.pars[i] <- candidate
        if (i<= llambda){
          log.acc.ratio <- fc_llambda(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_llambda(closed.sir.model,
    new, initial_cond, days, time_grid_beta)
        }else if (i==llambda+1){
          log.acc.ratio <- fc_lgamma(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_lgamma(closed.sir.model,
    new, initial_cond, days, time_grid_beta)
        }else{
          log.acc.ratio <- fc_ltau(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_ltau(closed.sir.model, new,
     initial_cond, days, time_grid_beta)
        }
        if (log(runif(1))<log.acc.ratio){new[i] <- candidate
          acceptances[i] <- acceptances[i]+1
        } else new[i] <- old[i]
    }
  chain[iter,] <- new
  }
```

```
}
acc_ratio <- acceptances/(post.samp.size*thin)
save.image(file="Det_Sim.Rdata")
```

Script : Deterministic SIR for simulated data generated by a deterministic approach.

A.5 Euler approximation

```
#Euler-Maruyama scheme

library(mvtnorm)
euler_approx <- function(delta, parameters, initial_conditions, days, N)
    {
lgrid=length(parameters)-2
beta_grid <- parameters[1:lgrid]
beta_f <- approxfun(x=seq(1, days, 14), y=beta_grid,yright=beta_grid[
    ltime],yleft=beta_grid[1])
betas <- beta_f(seq(1, days))

gamma <- parameters[lgrid+1]
tau <- parameters[lgrid+2]

x <- matrix(0, nrow=2, ncol=days+1)
x[, 1] <- initial_conditions

delta_W <- matrix(rnorm(days*2),ncol=days,nrow=2)*sqrt(delta)

for (i in 1:days) {
a <- x[, i]+c(betas[i]*x[1, i]*(1-x[1, i]-x[2, i])-gamma*x[1, i], gamma*
    x[1, i])*delta
B <- matrix(c(sqrt(betas[i]*x[1,i]*(1-x[1, i]-x[2, i])), 0, -sqrt(gamma*
    x[1, i]),sqrt(gamma*x[1, i])), nrow = 2, ncol=2)
sigma1 <- delta/sqrt(tau* N)*B
x[, i+1] <- a+sigma1%*%delta_W[,i]
}
return(x)
}
```

Script : Euler approximation for simulated data.

A.6 Stochastic data generation

```
source("euler_approx.R")

#DATA GENERATION
days <- 126
step <- 14
time_grid_beta <- seq(1, days, step)
ltime <- length(time_grid_beta)
time_grid_data <- 1:(step*ltime)

tau <- 2
gamma <- 1/7
```

```
beta_grid <- gamma*c(3,1.5,1.1,0.95,0.9,0.85,0.85,0.9, 0.5)
true_par <- c(beta_grid, gamma, tau)
initial_conditions <- c(6.89e-7, 1e-20)
population <- 4356000

dt <- 1
set.seed(123)
x <- euler_approx(dt, true_par, initial_conditions, days, population)
```

Script : Stochastic data generation.

A.7 MCMC functions for stochastic SIR

```
#StoMCMC
library(mvtnorm)

evaluate_sigma <- function(A, i){
B <- matrix(c(A[3, i], 0, -A[4, i], A[4, i]), nrow=2, ncol=2)
return(B%*%t(B))
}

loglikelihood <- function(x, parameters, dt, N){
lgrid <- length(parameters)-2 #9
ldata <- dim(x)[2] #127
beta_grid <- parameters[1:lgrid]
beta_f <- approxfun(x=seq(1, ldata-1, 14),y=beta_grid,yright=beta_grid[
    ltime],yleft=beta_grid[1])
betas <- beta_f(seq(1, ldata-1))
days <- length(betas) #126

gamma <- parameters[lgrid+1]
tau <- parameters[lgrid+2]

A <- matrix(0, ncol=days, nrow=4)
x_r <- x[, 1:days]
A[1,] <- x_r[1,]+betas*x_r[1,]*(rep(1, days)-x_r[1, ]-x_r[2,])-gamma*x_r
    [1,]
A[2, ] <- x_r[2,]+gamma*x_r[1, ]
A[3,] <- sqrt(betas*x_r[1, ]*(rep(1, days)-x_r[1, ]-x_r[2,]))
A[4, ] <- sqrt(gamma*x_r[1,])
x <- x[, 2:ldata]
return(sum(sapply(1:ncol(x), function(i) dmvnorm(x[, i], mean=A[1:2, i],
     sigma=dt/(tau*N)*evaluate_sigma(A, i),log=TRUE))))
}


logpriorlgamma <- function(lgamma) dnorm(lgamma,mean=0,sd=sqrt(100), log
    =TRUE)
fc_lgamma <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N)+
    logpriorlgamma(par[length(par)-1])}


logpriorltau <-  function(ltau) {(ltau*0.01)-(0.01*exp(ltau))}
```

```
fc_ltau <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N)+
    logpriorltau(par[length(par)])}


logpriorlbetas <- function(lbetas) sum(dnorm(lbetas,mean=0,sd=sqrt(100),
    log=TRUE))
fc_lbetas <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N)+
    logpriorlbetas(par[1:(length(par)-2)])}
```

Script: Definition of MCMC functions for stochastic SIR for simulated data.

A.8 Stochastic SIR with simulated data

```
rm(list=ls())
source("SDE_DataGen")
source("StoMCMC.R")

post.sample.size <- 30000
thin <- 20
npar <- 11

chain <- matrix(0, nrow=npar,ncol=(K/thin+1))
old_par <- chain[,1] <- rep(2, npar)

count=rep(0,npar)
sd.proposal <- c(0.05, 0.02, 0.05, 0.009, 0.04, 0.095, 0.03, 0.02, 0.01,
    0.008, 0.4)

for (iter in 2:post.samp.size){
  if (iter%%100==0) print(iter)
    for (inner in 1:thin){
      old <- new
      for (i in 1:par.num){
        updated.pars <- new
        candidate <- old[i]+rnorm(1,mean=0,sd=sd.proposal[i])
        updated.pars[i] <- candidate
        if (i<= llambda){
          log.acc.ratio <- fc_llambda(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_llambda(closed.sir.model,
    new, initial_cond, days, time_grid_beta)
        }else if (i==llambda+1){
          log.acc.ratio <- fc_lgamma(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_lgamma(closed.sir.model,
    new, initial_cond, days, time_grid_beta)
        }else{
          log.acc.ratio <- fc_ltau(closed.sir.model, updated.pars,
    initial_cond, days, time_grid_beta) - fc_ltau(closed.sir.model, new,
     initial_cond, days, time_grid_beta)
        }
        if (log(runif(1))<log.acc.ratio){new[i] <- candidate
          acceptances[i] <- acceptances[i]+1
        } else new[i] <- old[i]
    }
```

```
  chain[iter,] <- new
  }
}

acc_ratio <- count/K
exp_chain <- exp(chain)

save.image(file = "Sto_Sim.RData")
```

Script : Stochastic SIR for simulated data generated by a stochastic approach.

A.9 Analysis of results

```
rm(list=ls())

Grafici <- function(filename, simulated, det, all, ic){

load(file=filename)
end <- max(dim(chain)[1], dim(chain)[2])
start <- end-ceiling(end/3)
if(dim(chain)[1]==end){flag <- 1}else{flag <- 2}
if (flag==2){chain <- t(chain)
if(all==TRUE){p <- npar}else{p <- npar-1}
yname <- c("gamma", "nu", "I_0", "R_0")
}else{
yname <- c("gamma", "tau", "I_0", "R_0")
acc_ratio <- acceptances/(post.samp.size*thin)
if(all==TRUE){p <- par.num}else{p <- par.num-1}
}

if (ic==1){
lgrid <- length(beta_grid)
dev.new()
par(mfrow = c(4,3),oma=c(0,0,3,0), mar = c(1, 4, 2, 1))

for (i in 1:p){
if (i<=lgrid){
plot(exp(chain[start:end, i]),type="l",ylab=paste("lambda_",i, sep=""))
}else{plot(exp(chain[start:end, i]),type="l",ylab=yname[i-lgrid])
}
mtext(round(acc_ratio[i], 2), line=0,adj=0, cex=0.9, col="magenta",
    outer=FALSE)
if (simulated==TRUE) {abline(h=true_par[i], col="red")}
}

mtext("Estimation", outer = TRUE, cex = 1, side=3)
legend(x="bottomright",inset=c(-1.4,0.4),c("Acc_ratio","True Value"),lty
    =1,col=c("magenta", "red"),lwd = 3,cex=1.2, xpd=NA)

}else{
lgrid <- p-4
dev.new()
par(mfrow = c(3,3),oma=c(0,0,3,0), mar = c(2, 4,4,2 ))
```

```
for (i in 1:lgrid){
plot(exp(chain[start:end, i]),type="l",ylab=paste("lambda_",i, sep=""))
abline(h=median(exp(chain[start:end, i])), col="green")
mtext(round(acc_ratio[i], 2), line=0,adj=0, cex=0.9, col="magenta",
    outer=FALSE)
numb <- median(exp(chain[start:end, i]))
r <- formatC(numb, format = "e", digits = 2)
mtext(r, line=0,adj=1, cex=0.9, col="green", outer=FALSE)
if (i==3){legend(x=c(0.102, 0.0675), c("Acc_ratio","Median"),lty=1,col=c
    ("magenta", "green"),lwd = 2,cex=1.1, bty="n",  xpd=TRUE)}
}

mtext("Estimation/1", outer = TRUE, cex = 1, side=3)
dev.new()
par(mfrow = c(4,1),oma=c(0,0,3,0), mar = c(1, 4, 2, 10))


for (i in (lgrid+1):p){

plot(exp(chain[start:end, i]),type="l",ylab=yname[i-lgrid])
abline(h=median(exp(chain[start:end, i])), col="green")
numb <- median(exp(chain[start:end, i]))
r <- formatC(numb, format = "e", digits = 2)
mtext(r, line=0,adj=1, cex=0.9, col="green", outer=FALSE)
mtext(round(acc_ratio[i], 2), line=0,adj=0, cex=0.9, col="magenta",
    outer=FALSE)
if (i==(lgrid+1)){legend(x="topright",inset=c(-0.2, 0) ,c("Acc_ratio","
    Median"),lty=1,col=c("magenta", "green"),lwd = 3,cex=1, xpd=NA)
}
}

mtext("Estimation/2", outer = TRUE, cex = 1, side=3)

}
iterations <- seq(start,end,2)
llii <- length(iterations)
inf <- rec <- matrix(0,nrow=llii,ncol=days)
r_t <- matrix(0, ncol=days, nrow=llii)

if (det==TRUE){
library(deSolve)
for (i in 1:llii){

if(ic==2){initial_cond <- c(S=1-exp(chain[iterations[i], 12])-exp(chain[
    iterations[i], 13]), I=exp(chain[iterations[i], 12]), R=exp(chain[
    iterations[i], 13]))}
inf[i,] <- odesolution(closed.sir.model, initial_cond, days, chain[
    iterations[i], ], time_grid_beta)[1,]
inf[i,] <- exp(rnorm(days,mean=log(inf[i,]), sd=sqrt(1/exp(chain[
    iterations[i], lgrid+2]))))
rec[i,] <- odesolution(closed.sir.model, initial_cond, days, chain[
    iterations[i], ], time_grid_beta)[1,]
```

```
rec[i,] <- exp(rnorm(days,mean=log(rec[i,]), sd=sqrt(1/exp(chain[
    iterations[i], lgrid+2]))))

beta_grid <- exp(chain[iterations[i],1: lgrid])
beta_f <- approxfun(x=seq(1, days, 14),y=beta_grid,yright=beta_grid[
    lgrid],yleft=beta_grid[1])
betas <- beta_f(seq(1, days))
r_t[i, ] <- betas/exp(chain[iterations[i], lgrid+1])
}
}else{
count <- 0
if(ic==2){
for (i in 1:llii){
par <- exp(chain[iterations[i],])
inf[i,] <- euler_approx(1, par,population)$trajectories[1,2:(days+1)]
rec[i,] <- euler_approx(1, par,population)$trajectories[2,2:(days+1)]
r_t[i, ] <- euler_approx(1, par,population)$r
}
}else{

for (i in 1:llii){
beta_grid <- exp(chain[iterations[i], ])[1:ltime]
beta_f <- approxfun(x=seq(1, days, 14),y=beta_grid,yright=beta_grid[
    ltime],yleft=beta_grid[1])
betas <- beta_f(seq(1, days))
r_t[i, ] <- betas/exp(chain[iterations[i], (ltime+1)])
par <- c(betas, exp(chain[iterations[i], (ltime+1):(ltime+2)]))
suppressWarnings(inf[i,] <- euler_approx(1, par, x[,1], population)[1,1
    :(days)])
suppressWarnings(rec[i,] <- euler_approx(1, par, x[,1], population)[2,1
    :(days)])
}
}
y <- rowSums(is.na(inf))
cl <- c()
for (i in 1:nrow(inf)){
if(y[i]!=0){
cl <- c(cl, i)
}
}
if (is.null(cl)==FALSE){inf=inf[-cl, ]}
}
firstquant_i <- apply(inf,2,quantile,0.15)
medians_i <- apply(inf,2,median)
thirdquant_i <- apply(inf,2,quantile,0.85)

dev.new()
plot(thirdquant_r,type="l",col="red",lty=2, main="Trajectories", ylab="
    Infective")
lines(exp(lYmat[1,]),type="l")
lines(medians_i,type="l",col="red")
lines(firstquant_i,type="l",col="red",lty=2)
```

```
128 legend(x="topleft",c("Data", "Median", "Quantiles"), lty=c(1, 1, 3) ,col
    =c("black", "red", "red"),lwd = 1,cex=0.8)
129
130 dev.new()
131 R_t <- rep(0,days)
132 R_t <- apply(r_t, 2, median)
133 plot(R_t, col="blue", type="l", main="R_t index", xlab="days", ylab="R_t
    ")
134 points(time_grid_beta, apply(r_t[,time_grid_beta],2, median), col="
    magenta", pch=2)
135 legend(x="topright",c("Median", "R_t over the grid"), fill=c( "blue", "
    magenta"),cex=0.8)
136 }
```

Script : Function used to evaluate the model perfromance.

Appendix B

R scripts for real data

B.1 Prior distribution for initial conditions

```
#IC_prior
population <- 4356000

#Distribuzione di Dirichlet che vorrei approssimare

#means
alphatilda <- c(100/population, 2/population, (population-102)/
    population)
alpha0 <- 1e6
alpha <- alpha0*alphatilda #parametri della dirichlet

#var
varianze <- alphatilda*(1-alphatilda)/(alpha0+1)

#Logistic-normal
mu1 <- digamma(alpha[1])-digamma(alpha[3])
mu2 <- digamma(alpha[2])-digamma(alpha[3])
muD <- c(mu1, mu2)
s1 <- trigamma(alpha[1])+trigamma(alpha[3])
s4 <- trigamma(alpha[2])+trigamma(alpha[3])
s2 <- trigamma(alpha[3])
S <- matrix(c(s1, s2, s2, s4), nrow=2, ncol=2)
S <- S*8
S[1, 1] <- S[1, 1]*4
S[2, 2] <- S[2, 2]/12

prioric <- function(x) {dmvnorm(x, mean=muD, sigma=S, log=TRUE)}
```

Script : Logistic normal distribution for initial conditions.

B.2 Data importing

```
#Initialization and import data
```

```
step <- 14
days <- 126
time_grid_beta <- seq(1,days,step)
ltime <- length(time_grid_beta)
time_grid_data <- 1:126
population <- 4356000
delta <- 1
urlfilereg <- "https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv"
mydata <- read_csv(url(urlfilereg), show_col_types = FALSE)
mydata <- mydata[mydata$codice_regione=="01", ]

Idata <- mydata$totale_positivi
Rdata <- mydata$dimessi_guariti+mydata$deceduti
x <- rbind(Idata[1:days]/population,Rdata[1:days]/population)
lYmat <- log(x)
```

Script : Data import.

B.3 MCMC function for deterministic SIR/2

```
library(deSolve)
source("IC_prior.R")
odesolution <- function(f, initial_cond, days, parameters, time_grid_beta){
sol <- ode(func=f,y= initial_cond, times= 1:days, parms=exp(parameters[1:10]), time_grid_beta=time_grid_beta)
t(sol[,c("I","R")])
}

loglikelihood <- function(f, days, parameters, time_grid_beta){
sum <- 1+exp(parameters[12])+exp(parameters[13])
infected <- exp(parameters[12])/sum
removed <- exp(parameters[13])/sum
initial_cond <- c(S=1/sum, I=infected, R=removed)
odesol <- odesolution(f, initial_cond, days, parameters, time_grid_beta)
if(sum(odesol<=0)==0) sum(dnorm(lYmat, mean=log(odesol),sd= sqrt(1/exp(parameters[11])),log=TRUE)) else -Inf}

logpriorlbetas  <- function(lbetas) sum(dnorm(lbetas,mean=0,sd=sqrt(50),log=TRUE))
logpriorlmu <- function(lmu) dnorm(lmu,mean=0,sd=sqrt(50),log=TRUE)
logpriorlogtau <- function(w) w*0.01-0.01*exp(w)



logposterior <- function(f, days, parameters, time_grid_beta){
loglikelihood(f, days, parameters, time_grid_beta)+logpriorlbetas(parameters[1:9]) +logpriorlmu(parameters[10])+logpriorlogtau(parameters[11])+prioric(parameters[12:13])
}
```

B.4 Euler approximation/2

```
euler_approx <- function(delta, parameters, days, N, fixed_tau){

library(mvtnorm)

#Parameters definition

if (fixed_tau==1){ lgrid <- length(parameters)-3
  tau <- 1
}else{ lgrid <- length(parameters)-4
  tau <- parameters[lgrid+2]
}

beta_grid <- parameters[1:lgrid]
beta_f <- approxfun(x=seq(1, days, 14),y=beta_grid,yright=beta_grid[
  lgrid],yleft=beta_grid[1])
betas <- beta_f(seq(1, days))

gamma <- parameters[lgrid+1]
r_t <- betas/gamma

#Initial Conditions transformation
sum <- 1+parameters[lgrid+2]+parameters[lgrid+3]
infected <- parameters[lgrid+2]/sum
removed <- parameters[lgrid+3]/sum

x <- matrix(0, nrow=2, ncol=days+1)
x[, 1] <- c(infected, removed)

delta_W <- matrix(rnorm(days*2),ncol=days,nrow=2)*sqrt(delta)

for (i in 1:days) {

a <- x[, i]+c(betas[i]*x[1, i]*(1-x[1, i]-x[2, i])-gamma*x[1, i],
  gamma*x[1, i])*delta

B <- matrix(c(sqrt(betas[i]*x[1,i]*(1-x[1, i]-x[2, i])), 0, -sqrt(
  gamma*x[1, i]),sqrt(gamma*x[1, i])), nrow = 2, ncol=2)

sigma1 <- (delta/sqrt(tau*N))*B

x[, i+1] <- a+sigma1%*%delta_W[,i]
}

nlist <- list("trajectories"=x, "r"=r_t)
return(nlist)
}

```

Script : Euler approximation when initial conditions are addictional parameters.

B.5 MCMC functions for stochastic SIR/2

```
evaluate_sigma=function(A, i){
B <- matrix(c(A[3, i], 0, -A[4, i], A[4, i]), nrow=2, ncol=2)
return(B%*%t(B))
}

loglikelihood <- function(x, parameters, dt, N, fixed_tau){
lgrid <- length(parameters)-3
days <- dim(x)[2]

if (fixed_tau==1){
  sum <- 1+parameters[lgrid+2]+parameters[lgrid+3]
  infected <- parameters[lgrid+2]/sum
  removed <- parameters[lgrid+3]/sum
  ic <- c(infected, removed)
  x <- cbind(ic, x)
  tau <- 1
  }else{
    tau <- parameters[lgrid+2]
  }

beta_grid <- parameters[1:lgrid]
beta_f <- approxfun(x=seq(1, days, 14),y=beta_grid,yright=beta_grid[
  lgrid],yleft=beta_grid[1])
betas <- beta_f(seq(1, days))

gamma <- parameters[lgrid+1]


A <- matrix(0, ncol=days, nrow=4)
x_r <- x[, 1:days]
A[1,] <- x_r[1,]+betas*x_r[1,]*(1-x_r[1, ]-x_r[2,])-gamma*x_r[1,]
A[2, ] <- x_r[2,]+gamma*x_r[1, ]
A[3,] <- sqrt(betas*x_r[1, ]*(1-x_r[1, ]-x_r[2,]))
A[4, ] <- sqrt(gamma*x_r[1,])
y<- x[, 2:(days+1)]
return(sum(sapply(1:ncol(y), function(i) dmvnorm(y[, i], mean=A[1:2, i
  ], sigma=(dt/(tau*N))*evaluate_sigma(A, i),log=TRUE))))
}


logpriorlgamma <- function(lgamma) dnorm(lgamma,mean=0,sd=sqrt(100),
  log=TRUE)
fc_lgamma <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N,
   fixed_tau)+logpriorlgamma(par[length(par)-3])}


logpriorltau <-  function(ltau) {(ltau*0.01)-(0.01*exp(ltau))}
fc_ltau <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N,
  fixed_tau)+logpriorltau(par[length(par)-2])}


```

```
47  logpriorlbetas <- function(lbetas) sum(dnorm(lbetas,mean=0,sd=100, log
      =TRUE))
48  fc_lbetas <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N,
       fixed_tau)+logpriorlbetas(par[1:(length(par)-4)])}
49
50  prioric <- function(x) {dmvnorm(x, mean=muD, sigma=S, log=TRUE)}
51  fc_ic <- function(x, par, dt, N){loglikelihood(x, exp(par), dt, N,
      fixed_tau)+prioric(par[(length(par)-2: length(par))])}
52
53
```

Script: Definition of MCMC functions for stochastic SIR when iniditial conditions are considered as additional parameters.

Bibliography

[1] D. J. Wilkinson, *Stochastic Modelling for Systems Biology. 2nd edition.* CRC Press, 2011.

[2] S. M. Ross, *Stochastic Processes. 2nd edition.* John Wiley and Sons, 1196.

[3] S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations.* Springer, 2008.

[4] E. Platen and N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance.* Springer, 2010.

[5] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems.* Springer, 2015.

[6] P. Mozgunov, M. Beccuti, A. Horvath, T. Jaki, R. Sirovich, and E. Bibbona, "A review of the deterministic and diffusion approximations for stochastic chemical reaction networks," *Springer*, 2018.

[7] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis.* Springer, 2000.

[8] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence.* Wiley, 1986.

[9] *Epidemic, Endemic, Pandemic: What are the Differences?* [Online]. Available: https://www.publichealth.columbia.edu/public-health-now/news/epidemic-endemic-pandemic-what-are-differences

[10] *Cosa sono SARS-CoV-2 e Covid-19.* [Online]. Available: https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?lingua=italiano&id=257

[11] *Symptoms of COVID-19.* [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[12] *Coronavirus Diagnosis: What Should I Expect?* [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/diagnosed-with-covid-19-what-to-expect

[13] C. Mamo, M. Dalmasso, O. Pasqualini, D. Quarta, D. Catozzi, F. Cigliano, E. Pompili, A. Gallone, G. Greco, E. Procopio, and A. Castella, "Durata dell'isolamento dei casi covid-19 nella prima ondata epidemica in una asl della città metropolitana di torino," *Boll Epidemiol Naz*, 2021.

[14] *Tutte le notevoli differenze tra la prima e la seconda ondata di contagi* . [Online]. Available: https://www.agi.it/cronaca/news/2021-01-12/contagi-coronavirus-differenze-prima-seconda-ondata-10989684/

[15] *Coronavirus. La situazione*. [Online]. Available: https://mappe.protezionecivile.gov.it/it/mappe-emergenze/mappe-coronavirus/situazione-desktop

[16] *GitHub dati regioni*. [Online]. Available: https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv

[17] *Covid-19: la pandemia in 10 date da ricordare*. [Online]. Available: https://www.fondazioneveronesi.it/magazine/articoli/da-non-perdere/covid-19-la-pandemia-in-10-date-da-ricordare

[18] *Logit-normal distribution*. [Online]. Available: https://en.wikipedia.org/wiki/Logit-normal_distribution

[19] *Dirichlet distribution*. [Online]. Available: https://en.wikipedia.org/wiki/Dirichlet_distribution

[20] J. Atchison and S. Shen, "Logistic-normal distributions: Some properties and uses." *Biometrika*, 1980.