Stochastic SIS Epidemic Models and Corresponding Statistical Inference

Jiafeng Pan

Department of Mathematics and Statistics
University of Strathclyde

Degree of Doctor of Philosophy, 2013

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I am grateful for my supervisors Professor Xuerong Mao and Dr David Greenhalgh for their constant support and insights during the course of writing this thesis.

Special thanks should also go to Dr Alison Gray for her patience, kindness and additional advice on numerous occasions.

I also wish to mention the support that my friends Wenwen, Wei and Hongrui have given me in various areas of research.

In addition, I thank SIAM Journal of Applied Mathematics and Journal of Mathematical Analysis and Applications for their invaluable suggestions and for accepting my papers.

Last but not least, I would also like to thank my parents, sister and husband for their love and support during this period.

Notation

a.s. :	Almost surely, or with probability one.
$A := B$:	A is defined by B or A is denoted by B.
$A(x) \equiv B(x)$:	$A(x)$ and $B(x)$ are identically equal, i.e. $A(x) = B(x)$ for all x.
$\emptyset$:	The empty set.
I_A :	The indicator function of a set A, that is $I_A(x) = 1$ if $x \in A$ or otherwise 0.
$\sigma(\mathcal{C})$:	The σ-algebra generated by $\mathcal{C}$.
$a \vee b$:	The maximum of a and b.
$a \wedge b$:	The minimum of a and b.
$f : A \to B$:	The mapping f from A to B.
$\mathbb{R}_+$:	The set of all nonnegative real numbers, that is $R_+ = [0, \infty)$.
$\mathbb{R}^n$:	The n-dimensional Euclidean space.
$\mathcal{B}^n$:	The Borel σ-algebra on $\mathbb{R}^n$.
$\mathbb{R}^{n\times m}$:	The space of real $n \times m$-matrices.
$\lvert x\rvert$:	The Euclidean norm of a vector x.
$\dot{x}(t)$:	$\dot{x}(t) = \frac{dx(t)}{dt}$.
A^T :	The transpose of a vector or matrix A.
V_x :	$V_x = \nabla V = (V_{x_1}, \cdots, V_{x_n}) = \left(\frac{\partial V}{\partial x_1}, \cdots, \frac{\partial V}{\partial x_n}\right)$.
V_{xx} :	$V_{xx} = (V_{x_i x_j})_{n\times n} = \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_{n\times n}$.
$C(D;\mathbb{R}^n)$:	The family of continuous $\mathbb{R}^n$-valued functions defined on D.
$C^m(D;\mathbb{R}^n)$:	The family of continuously m-times differentiable $\mathbb{R}^n$-valued functions defined on D.
$C^{2,1}(D \times \mathbb{R}_+;\mathbb{R})$:	The family of all real-valued functions $V(x,t)$ defined on $D \times \mathbb{R}_+$ which are continuously twice differentiable in $x \in D$ and once differentiable in $t \in \mathbb{R}_+$.
$L^p(\Omega;\mathbb{R}^n)$:	The family of $\mathbb{R}^n$-valued random variables X with $\mathbb{E}\lvert X\rvert^p < \infty$.
$L^p([a,b];\mathbb{R}^n)$:	The family of Borel measurable functions $h : [a,b] \to \mathbb{R}^n$ such that $\int_a^b \lvert h(t)\rvert^p dt < \infty$.
$\mathcal{L}^p([a,b];\mathbb{R}^n)$:	The family of $\mathbb{R}^n$-valued $\mathcal{F}_t$-adapted processes $f(t)_{a\le t\le b}$ such that $\int_a^b \lvert f(t)\rvert^p dt < \infty$ a.s..
$\mathcal{M}^p([a,b];\mathbb{R}^n)$:	The family of processes $f(t)_{a\le t\le b}$ in $\mathcal{L}^p([a,b];\mathbb{R}^n)$ such that $\mathbb{E}\int_a^b \lvert f(t)\rvert^p dt < \infty$.

Other notation will be explained where it first appears.

Abstract

This thesis considers the deterministic SIS epidemic model, which has applications to transmission of real-life diseases, such as pneumococcus, gonorrhea and tuberculosis. Environmental noise can affect the deterministic system significantly. There are various types of noise which can be incorporated into the deterministic dynamics according to different situations. The effect of three types of noise on the deterministic SIS epidemic model have been examined in this thesis, which has not been discussed in previous literature.

Firstly, assuming that there exists environmental noise in the disease transmission coefficient, we extend the classical SIS epidemic model from a deterministic framework to a stochastic one by incorporating white noise using the parameter perturbation technique, and formulate it as a stochastic differential equation (SDE) for the number of infectious individuals $I(t)$. For the model to make sense, we then prove that this SDE has a unique global positive solution $I(t)$ and establish conditions for extinction and persistence of $I(t)$ and compare these with the corresponding conditions for the deterministic SIS epidemic model. We also discuss perturbation by stochastic noise. In the case of persistence we show the existence of a stationary distribution and derive expressions for its mean and variance. Secondly, assuming that the parameters in the SIS epidemic model experience an abrupt change around the point of threshold value, we incorporate telegraph noise in the deterministic model. We then establish the explicit solution of the stochastic SIS epidemic model, which is useful in performing computer simulations. We also obtain the conditions for extinction and persistence for this model. Afterwards, we take a further step of incorporating both types of the aforementioned noise in the SIS epidemic model. We not only show the existence of a unique global positive solution but also examine the asymptotic properties, including extinction and persistence. The results are illustrated by computer simulations, including examples based on real life diseases for the first and second stochastic models. Computer simulations based on the explicit solution and the Euler–Maruyama scheme are compared for the SIS model with telegraph noise.

Furthermore, statistical inference is always essential in disease analysis. That is the motivation for us to conduct parameter estimation for the SDE SIS model with white noise introduced. Three estimation methods, least squares estimation, the pseudo-Maximum Likelihood Estimation (pseudo-MLE) method and the Bayesian approach are applied to the SDE SIS model. Our main contribution in least squares estimation and pseudo-MLE is variance estimation. We obtain not only the point estimators but also the interval estimators and the joint confidence regions for both estimation techniques. Additionally we investigate the factors which influence variance in estimation. As for the Bayesian approach, although strong results have been obtained by using the MCMC technique, we use a different method where analytic results are obtained without the need to deal with

the significant computational cost. Computer simulations are performed to illustrate our theory. The three estimation methods are compared both analytically and in the simulation examples.

Chapter 1

Introduction and Literature Review

1.1 Epidemic Modelling

Mathematical modelling of infectious diseases is a tool to investigate the mechanisms for the spread of diseases and to make predictions in order to control an epidemic. The first such model for an infectious disease was that for smallpox by the well-known mathematician, Daniel Bernoulli, in 1760 [46]. His calculations from this model showed that the universal inoculation against smallpox could increase the expectation of life. Later Sir Ronald Ross studied the spread of malaria and developed the important idea that one did not need to eradicate all mosquitoes to eliminate the disease. Modern epidemiology aims to model the spread of a disease and show that if certain conditions are met, then a disease will become extinct.

Epidemics are commonly modelled by using deterministic compartmental models where the population amongst whom the disease is spreading is divided into several compartments. The possible compartments are the susceptible compartment S, the exposed compartment E (in which individuals are exposed but not yet infectious), the infectious compartment I and the removed compartment R (in which individuals have permanent or temporary immunity). The classical Kermack-McKendrick model [51] is sometimes used for modelling common childhood diseases where a typical individual starts off susceptible, at some stage catches the disease and after a short infectious period becomes permanently immune. This is sometimes called the SIR (susceptible-infected-removed) model. However some diseases, in particular some sexually transmitted and bacterial diseases, do not have permanent immunity. For these diseases individuals start off susceptible, at some stage catch the disease and after a short infectious period become susceptible again. There is no protective immunity. For these diseases SIS (susceptible-infected-susceptible) models are appropriate [47]. Based on the ideas of these two fundamental compartmental models, more models have been developed to give a better description for certain diseases. For example, for some diseases such as scarlet fever [6] and measles [91], individuals experience a latent stage for a period of time between being exposed to the disease and becoming infectious. The latent period is neither short nor negligible compared with the infectious period. These diseases can be modelled by the SEIR (Susceptible-Exposed-Infective-Removed) or the SEIS (Susceptible-Exposed-Infective-Susceptible) models. For some special cases, where the diseases have no

recovery, for example HIV [15], SI (Susceptible-Infective) or SEI (Susceptible-Exposed-Infective) models can be proposed. SIR and SEIR models can be generalised to SIRS (Susceptible-Infective-Removed-Susceptible) and SEIRS (Susceptible-Exposed-Infective-Removed-Susceptible) type respectively by including the possibility that immunity is only temporary and allowing removed individuals to become susceptible again.

During the initial stage of research on these compartmental models, one common assumption being made to simplify the problem is that the diseases spread in a population with constant size. This assumption is reasonable if the disease spreads through the population in a short duration or if the disease spreads over many years with limited effects on mortality so births are being approximately balanced by the deaths. However the assumption fails to hold when there are many disease-related deaths. Examples of such diseases which substantially reduce the population sizes can be found in [6]. Recently studies of epidemic models that incorporate disease-caused death and varying total population size have become one of the important areas in epidemic modelling and a lot of research has been conducted on these topics (see e.g. [16, 22, 41, 42, 43, 88]).

1.2 SIS Epidemic model

The SIS epidemic model is one of the simplest epidemic compartmental models. It can be applied to many real life diseases which we will explain in detail in this section. In this thesis, we will concentrate on this simple epidemic model and develop various types of stochastic version from this deterministic model and then conduct statistical inference on the SDE SIS model. First we give a definition for the deterministic SIS model [47]. If $S(t)$ denotes the number of susceptibles and $I(t)$ the number of infecteds at time t, then the SIS model can be described by the following two differential equations:

$$\begin{cases} \frac{dS(t)}{dt} = \mu N - \beta S(t)I(t) + \gamma I(t) - \mu S(t), \\ \frac{dI(t)}{dt} = \beta S(t)I(t) - (\mu + \gamma)I(t), \end{cases} \tag{1.2.1}$$

with initial values $S_0 + I_0 = N$. N is the total size of the population amongst whom the disease is spreading. Here μ is the per capita death rate, and γ is the rate at which infected individuals become cured, so $1/\gamma$ is the average infectious period. The parameter β is the disease transmission coefficient, so that $\beta = \lambda/N$ where λ is the per capita disease contact rate. The parameter λ is the average number of adequate contacts of an infective per day. An adequate contact is one which is sufficient for the transmission of an infection if it is between a susceptible and an infected individual. This is one of the simplest possible epidemic models and because it is so simple it and its variants are commonly studied. For example, SIS models are discussed by Brauer et al. [13]. Ianelli, Milner and Pugliese [49] study age-structured epidemic models, as do Feng, Huang and Castillo-Chavez [28]. Li, Ma and Zhu [59] analyse backward bifurcation in an SIS epidemic model with vaccination and Van den Driessche and Watmough [92] study backward bifurcation in an SIS epidemic model with hysteresis. There are many other examples of SIS epidemic models in the literature.

In their excellent monograph Hethcote and Yorke [47] outline several mathematical

models for gonorrhea with increasing levels of complexity. The simplest of these is the above model (1.2.1) with μ set to zero (i.e. no demography). As $S + I = N$, the total population is constant, the two models with and without demographics are equivalent; just replace γ in the model with no demography by $\mu + \gamma$ to get the model with demography.

This is a very simple model for gonorrhea. It assumes that the population is homogeneous (so is more suitable for a homosexual than a heterosexual population) and mixing is homogeneous, whereas in practice sexual mixing is extremely heterogeneous. It also ignores the small but non-zero disease incubation period and assumes that contact rates remain constant and do not vary seasonally. Lajmanovich and Yorke [56] and Nold [76] discuss heterogeneously mixing SIS epidemic models for the spread of gonorrhea. The model (1.2.1) also ignores screening. Hethcote and Yorke define the contact number to be $\overline{\sigma} = \lambda/\gamma$, so in our model $\overline{\sigma} = \beta N/(\mu + \gamma)$.

If $\tilde{I} = I/N$ is the fraction of the population infected at time t then they show that in our notation the solution is

$$I(t) = \begin{cases} \left[\frac{\beta}{\beta N-\mu-\gamma}(1 - e^{-(\beta N-\mu-\gamma)t}) + \frac{1}{I_0}e^{-(\beta N-\mu-\gamma)t}\right]^{-1}, & \text{if } \frac{\beta N}{\mu+\gamma} \neq 1, \\ \left[\beta t + \frac{1}{I_0}\right]^{-1}, & \text{if } \frac{\beta N}{\mu+\gamma} = 1. \end{cases} \tag{1.2.2}$$

It is straightforward to show that $\overline{\sigma}$ has the usual interpretation as the basic reproduction number R_0. This is the expected number of secondary cases produced by a single newly infected individual entering a disease-free population at equilibrium. In such a situation each newly infected individual remains infectious for time $1/(\mu + \gamma)$ and during this period infects βN of the N susceptibles present. Hence

$$R_0 = \overline{\sigma} = \frac{\beta N}{\mu + \gamma}. \tag{1.2.3}$$

From now on we denote this R_0 by R_0^D to emphasise that it is R_0 for the deterministic model. It is a straightforward consequence of equations (1.2.2) that [47]:

- If $R_0^D \leq 1$, $\lim_{t\to\infty} I(t) = 0$.
- If $R_0^D > 1$, $\lim_{t\to\infty} I(t) = N\left(1 - \frac{1}{R_0^D}\right)$.

Another disease for which it is possible to use an SIS model is pneumococcal carriage. Streptococcus pneumoniae (*S.pneumoniae*) is a bacterium commonly found in the throat of young children. When an individual carries pneumococcus the infectious carriage normally lasts around seven weeks [93] and at the end of this carriage period the individual is susceptible again. R_0 for pneumococcal carriage and transmission is 1.8-2.2 [96]. Lipsitch [61] discusses mathematical models for the transmission of *S.pneumoniae* with multiple serotypes and vaccination. Lamb, Greenhalgh and Robertson [57] discuss a mathematical model for the transmission of a single serotype of *S.pneumoniae* with vaccination. If we have only a single serotype and no vaccination then the disease can be modelled by equations (1.2.1). Other bacterial diseases, for example tuberculosis, can also be modelled by SIS models [28].

1.3 Environmental Noise in Epidemic Modelling

We have briefly introduced some deterministic epidemic models and their applications to certain diseases in the previous two sections. Indeed the foundations of epidemic modelling are largely based upon deterministic equations for the dynamics of populations. However, deterministic models are only suitable when the numbers of susceptibles and infectives are both large [7]. Usually disease outbreaks start from only a few cases, and therefore incorporating stochasticity into the deterministic compartmental models is necessary. There are various types of noise which can be incorporated into the deterministic dynamics according to different situations. Four types of noise will be discussed in detail in this section.

1.3.1 Demographic Noise

In recent years event-based (demographic) stochasticity has been used increasingly by applied researchers. Much research suggests that demographic stochastic effects, due to the random nature of population events, can lead to significant deviations of disease spread from the description of the deterministic model (see e.g. [8, 30, 84]).

One common way to model the stochastic population with demographic noise is performed by using stochastic integer-based event-driven simulations. It is widely favored primely because it describes the supposed behaviour of the real system, where the population is composed with an integer number of susceptible or infected individuals. However, large numbers of replicate simulations are required to establish confidence in results since the dynamics through only one simulation are not necessarily representative of average behaviour. The computational cost can be significant. A number of approximation methods, moment closure techniques [73, 74] and diffusion approximations [54, 55], have overcome the problems from the requirement of performing large amounts of simulations. However, such methods are generally more accurate when the population size is large, where introducing the stochasticity is not so necessary. Due to this disadvantage of a stochastic event-driven simulation approach, Keeling and Ross [50] presented a Markov process model for the SIS and SIR models by using the Kolmogorov forward equation to simultaneously consider the probability of each possible state occurring. This equation is linear and has a natural matrix formulation and by solving this equation a complete description of all possible behaviours of the stochastic system can be obtained.

Alternatively, Allen [1] presented a stochastic differential equation (SDE) SIS epidemic model with demographic stochasticity, which is an approximation to its corresponding continuous time Markov chain model. This is done using the following procedure: The discrete stochastic SIS model is constructed by determining the possible changes with their corresponding transition probabilities for a small time interval. Then the expected change and the covariance matrix for the change are determined. This information leads to the SDE for the system. Also, McCormack and Allen [71] construct a similar SDE approximation to an SIS multihost epidemic model and explore the stochastic and deterministic models numerically. The latter stochastic model is obtained by assuming that events occurring at a constant rate in the deterministic model occur according to a Poisson process with the same rate.

1.3.2 White Noise

White noise is another type of environmental noise which will affect the population system significantly (see e.g. [37, 66, 67, 68]). Parameter perturbation is a routine method to incorporate white noise into a deterministic system. Dalal, Greenhalgh and Mao [21] have previously used the technique of parameter perturbation to examine the effect of environmental stochasticity in a model of AIDS and condom use. They found that the introduction of stochastic noise changes the basic reproduction number of the disease and can stabilise an otherwise unstable system. Ding, Xu and Hu [23] apply a similar technique to a simpler model of HIV/AIDS transmission. Other previous work on parameter perturbation in epidemic models seems to have concentrated on the SIR model. Tornatore, Buccellato and Vetro [90] discuss an SDE SIR system with and without delay with a similar parameter perturbation as we shall discuss here. The system for the SDE SIR model with no delay is

$$\begin{cases} d\tilde{S}(t) = [\mu - \tilde{\beta}\tilde{S}(t)\tilde{I}(t) - \mu\tilde{S}(t)]dt - \tilde{\sigma}\tilde{S}(t)\tilde{I}(t)dB(t), \\ d\tilde{I}(t) = [\tilde{\beta}\tilde{S}(t)\tilde{I}(t) - (\mu+\gamma)\tilde{I}(t)]dt + \tilde{\sigma}\tilde{S}(t)\tilde{I}(t)dB(t), \\ d\tilde{R}(t) = [\gamma\tilde{I}(t) - \mu\tilde{R}(t)]dt, \end{cases} \tag{1.3.1}$$

where $B(t)$ is a Brownian motion. Here $\tilde{S}, \tilde{I}$ and $\tilde{R}$ denote respectively the susceptible, infected and removed fractions of the population, rather than absolute numbers, so that $\tilde{\beta}$ in this model corresponds to βN in (1.2.1). They study the stability of the disease-free equilibrium (DFE). They find that

$$0 < \tilde{\beta} < \min\left\{\gamma + \mu - \frac{\tilde{\sigma}^2}{2}, 2\mu\right\}$$

is a sufficient condition for the asymptotic stability of the DFE. Their computer simulations for the SDE SIR model agree well with the analytical results and show that the introduction of noise into the system raises the threshold to $\mu + \gamma + (\tilde{\sigma}^2/2)$, so if

$$\min\left\{\mu + \gamma - \frac{\tilde{\sigma}^2}{2}, 2\mu\right\} < \tilde{\beta} < \gamma + \mu + \frac{\tilde{\sigma}^2}{2}$$

then the DFE $E_0 = (S(0), I(0), R(0)) = (1, 0, 0)$ is stable and the disease does not occur, whereas if $\tilde{\beta} > \gamma + \mu + (\tilde{\sigma}^2/2)$ then the DFE is unstable. These results are similar to those of [21]. Chen and Li [19] study an SDE version of the SIR model both with and without delay, but introduce stochastic noise in a different way than Tornatore, Buccellato and Vetro [90] do. Lu [63] studies an SIRS model and extends their results by including the possibility that immunity is only temporary and improving the analytical bound on the sufficient condition for the stability of the DFE to $\beta < \mu + \gamma - (\tilde{\sigma}^2/2)$.

For an epidemic model, it is important to include the effect of environmental variation in the disease transmission coefficient. Whilst several papers study the effect of stochastic parameter perturbation on SIR and SIRS epidemic models, we are not aware of any literature addressing this issue in SIS epidemic models. This is the motivation for us to examine the effect of white noise in the SIS epidemic model in Chapter 3.

1.3.3 Telegraph Noise

Some parameters in the deterministic population dynamic system may experience an abrupt change around the point of a threshold value. This abrupt change can be incorporated into the system in terms of telegraph noise. Telegraph noise can affect the population system significantly (see e.g. [25, 83]). It is therefore critical to discover whether the presence of such noise does affect population systems significantly.

For example, consider a predator-prey Lotka-Volterra model

$$\begin{cases} \dot{x}_1(t) = x_1(t)(a_1 - b_1 x_2(t)), \\ \dot{x}_2(t) = x_2(t)(-c_1 + d_1 x_1(t)), \end{cases} \tag{1.3.2}$$

where a_1, b_1, c_1 and d_1 are positive numbers. It is well known that the population develops periodically if there is no influence of environmental noise (see e.g. [35, 86]). However, if the factor of telegraph noise is taken into account, the system will change significantly. Telegraph noise can be illustrated as a switching between two or more regimes of environment, which differ by factors such as nutrition or rainfall (see e.g. [25, 83]). The switching is memoryless and the waiting time for the next switch has an exponential distribution. We can hence model the regime switching by a finite-state Markov chain. To make it simple, assume that there are only two regimes and the system obeys equation (1.3.2) when it is in regime 1, while it obeys another predator-prey Lotka-Volterra model

$$\begin{cases} \dot{x}_1(t) = x_1(t)(a_2 - b_2 x_2(t)), \\ \dot{x}_2(t) = x_2(t)(-c_2 + d_2 x_1(t)) \end{cases} \tag{1.3.3}$$

in regime 2. The switching between these two regimes is governed by a Markov chain $r(t)$ on the state space $\mathbb{S} = \{1, 2\}$. The population system under regime switching can therefore be described by the stochastic model

$$\begin{cases} \dot{x}_1(t) = x_1(t)(a_{r(t)} - b_{r(t)} x_2(t)), \\ \dot{x}_2(t) = x_2(t)(-c_{r(t)} + d_{r(t)} x_1(t)). \end{cases} \tag{1.3.4}$$

This system is operated as follows: If $r(0) = 1$, the system obeys equation (1.3.2) till time τ_1 when the Markov chain jumps to state 2 from state 1; the system will then obey equation (1.3.3) from time τ_1 till time τ_2 when the Markov chain jumps to state 1 from state 2. The system will continue to switch as long as the Markov chain jumps. If $r(0) = 2$, the system will switch similarly. In other words, equation (1.3.4) can be regarded as equations (1.3.2) and (1.3.3) combined, switching from one to the other according to the law of the Markov chain. Equations (1.3.2) and (1.3.3) are hence called the subsystems of equation (1.3.4).

Clearly, equations (1.3.2) and (1.3.3) have their unique positive equilibrium states as $(p_1, q_1) = (c_1/d_1, a_1/b_1)$ and $(p_2, q_2) = (c_2/d_2, a_2/b_2)$, respectively. Recently, Takeuchi et al. [87] revealed a very interesting and surprising result: If the two equilibrium states of the subsystems are different, then all positive trajectories of equation (1.3.4) always exit from any compact set of $\mathbb{R}^2_+$ with probability one; on the other hand, if the two equilibrium states coincide, then the trajectory either leaves from any compact set of $\mathbb{R}^2_+$ or converges to the equilibrium state. In practice, the two equilibrium states are usually different, whence Takeuchi et al. [87] show that equation (1.3.4) is neither permanent nor

dissipative. This is an important result as it reveals the significant effect of environmental noise on the population system: both subsystems (1.3.2) and (1.3.3) develop periodically, but switching between them makes them become neither permanent nor dissipative.

Markovian environments are also very popular in many other fields of biology. As examples Padilla and Adolph [79] present a mathematical model for predicting the expected fitness of phenotypically plastic organisms experiencing a variable environment and discuss the importance of time delays in this model, and Anderson [2] discusses optimal exploitation strategies for an animal population in a Markovian environment. Additionally Peccoud and Ycart [80] propose a Markovian model for the gene induction process, and Caswell and Cohen [18] discuss the effects of the spectra of the environmental variation in the coexistence of metapopulations.

Motivated by Takeuchi et al. [87], we will examine the effect of telegraph noise on the SIS epidemic model in Chapter 4. To the best knowledge of the author, there is no existing literature regarding incorporating this type of noise into compartmental epidemic models.

1.3.4 Combination of White and Telegraph Noise

Some recent research has examined the effect of taking both white and telegraph noise simultaneously into modelling. For example, Basak, Bisi and Ghosh [9] discussed the stability of a general linear SDE with Markovian switching and later Mao and Yuan [69] discussed the general non-linear case. Luo and Mao [64, 65] examined the combined effect of both types of noise in a more applied ecological model, the predator-prey Lotka-Volterra model, where the types of white noise introduced are different in the two papers. The white noise they introduced in [64] is dependent on the population sizes while the white noise in [65] is not. They showed in [64] that incorporating such two types of noise in the deterministic system will suppress the explosion of the population which is a desirable property. They showed in [65] that large white noise will force the population to become extinct while the population is bounded when the white noise is relatively small. Li et al. [60] considered the stochastic logistic population model with regime switching and obtained the sufficient and necessary conditions for stochastic permanence and extinction.

From the literature review, we see that including both types of noise into the deterministic system affects the system significantly. However there is no existing research regarding incorporating these two types of noise into deterministic epidemic models. Our research in Chapter 5 is aimed at developing a stochastic SIS model with regime switching and also to examine the new conditions for the permanence and extinction of the system.

1.4 Statistical Inference in Stochastic Compartmental Epidemic Models

Statistical inference is always essential in disease analysis since the parameter estimates can be used to characterise the infection process and also provide information on key indicators of disease spread, such as the basic reproduction number, which is often regarded

as a threshold indicating whether an epidemic is likely to persist or die out. Therefore providing the model used is sufficiently realistic, estimation of model parameters can help to inform the disease control policies. In general parameter estimation in stochastic differential equations (SDEs) is a non-trivial problem [10, 48]. Many SDEs are non-linear, making simpler approaches to estimation impossible to implement. Recently, many authors have discussed parameter estimation techniques for stochastic models. For example, Young [95] reviews parameter estimation methods for continuous time models. Nielsen, Madsen and Young [75] updates this to include newer methods for discretely observed SDEs. Timmer [89] discusses the relation between Maximum Likelihood Estimators (MLEs) and quasi-MLEs and compares the quasi-MLE approach with the $\Delta t = \delta t$ approach in simulations. Kristensen, Madsen and Young [53] considers the stochastic 'grey box' model and presents the approximate MLE approach based on the normal approximation and use of the extended Kalman filter and a software package CTSM. Bishwal [10] discusses the asymptotic properties of MLEs and Bayes estimators of real valued drift parameters in SDEs.

In Chapter 4, we will apply pseudo-MLE and least squares estimation to the SDE SIS epidemic model which we will derive in Chapter 3. This SDE is non-linear so applying either estimation techniques is not a trivial problem. Although the application of these two estimation techniques to the SDEs have been widely discussed in previous research, as far as the author knows, variance estimation has not been touched in most of the papers. In Chapter 4, we will obtain not only the point estimators but also the interval estimators and also the joint confidence regions for both estimation techniques. Afterwards we will investigate the factors which influence variance in estimation.

The Bayesian approach has recently been a more popular way to apply statistical inference to stochastic models since it can include previous information in the estimation procedure [48]. An advanced computational technique, i.e. the Markov Chain Monte Carlo (MCMC) method (see e.g.[32]) has been developed, which enables Bayesian inference to be applied for a broad spectrum of stochastic models. Although the MCMC method has been well established in the recent literature, it still needs further development before it can be routinely applied to the different models, and also the computational cost for the analysis can be significant. In this paper we apply the Bayesian approach to the stochastic SIS model, where an analytical expression is obtained for the posterior distribution as well as the Bayesian estimators, so that intensive computation is not needed.

Bayesian inference for stochastic compartmental models using the MCMC method has been discussed extensively in previous papers. Different MCMC algorithms have been developed to tackle stochastic models where the noise is modelled in different ways. One common way to develop the stochastic compartmental model is to represent the transitions between the classes of the model as stochastic processes, which we have mentioned in the last section. The choice of these stochastic processes can make the MCMC method very difficult to apply. Many efforts have been made to design a MCMC algorithm to apply Bayesian inference to more realistic but more complicated stochastic models. For example Streftaris and Gibson [85] extend a Markovian SIR model which allows the infectious period of an individual to follow a Weibull distribution and design an efficient independence-type Metropolis-Hastings algorithm to obtain parameter estimates within a Bayesian framework. Furthermore, Boys and Giles [12] extend the Markovian

SEIR model to let the rate parameter of the exponential distribution for describing the infectious period be a step function in time and then develop a reversible jump MCMC methodology to apply Bayesian inference. There is also much more research concerning how a MCMC algorithm can be developed to enable Bayesian inference for different stochastic compartmental models (see eg. [45, 77, 78]).

We mentioned in the last section that parameter perturbation is another technique to introduce stochastic noise. In Chapter 3, we will extend the classical SIS epidemic model to a stochastic version to include the effect of environmental variation in the disease transmission coefficient. There is no previous literature regarding Bayesian estimation using the MCMC method to this kind of stochastic compartmental model derived by parameter perturbation. Note that there is existing research dealing with Bayesian inference for diffusion processes using the MCMC method, among which strong results have been obtained. Roberts and Stramer [82] develop efficient MCMC algorithms to deal with the Bayesian inference for non-linear univariate diffusion processes. Later Golightly and Wilkinson [36] extended the method to tackle multivariate problems. It seems that the MCMC method is well established for diffusion problems. However the problem we deal with in this paper is different. In the two papers [36, 82], it was assumed that the time gap between the two closest observations was too large to be used as a time step for the Euler method. Therefore m latent time steps were introduced between every pair of adjacent observations, which made the analytical method very complicated to apply. For our stochastic epidemic SIS model, we assume that sufficient close observations are available for the Euler method to be applied to discretise the path of the process, so that the discretised form of the process has a likelihood that is useable. Therefore we can proceed to use analytical Bayesian inference in this thesis for the stochastic SIS epidemic model. Our main contribution in Chapter 5 is to apply Bayesian inference analytically for a certain type of stochastic compartmental model instead of using a data imputation method, and this has not been discussed previously.

In summary, in this chapter we have introduced a few compartmental models which are widely used to model epidemics. The deterministic SIS epidemic model is one of the simplest possible epidemic models, which has applications to the transmission of real-life diseases, such as pneumococcus, gonorrhea and tuberculosis. It is important to include the effect of environmental noise to the SIS epidemic model. We have reviewed four commonly used types of noise and the effect of these on the deterministic model in this chapter. We are not aware of any literature incorporating white noise, telegraph noise and the combination of these two types of noise in the SIS epidemic models. Therefore we are going to examine the effect of including these three types of noise into the SIS epidemic model in Chapters 3-5 respectively. Furthermore, statistical inference is always essential in disease analysis. That is the motivation for us to conduct parameter estimation for our SDE SIS model with white noise introduced. Estimation methods have been reviewed in this chapter. Three estimation methods, least squares estimation, the pseudo-MLE method and the Bayesian approach will be applied to our SDE SIS model in Chapters 6 and 7. These three estimation methods have been widely discussed in the existing literature. However, our main contribution in least squares estimation and pseudo-MLE is variance estimation. For the Bayesian approach, although strong results have been obtained for general SDE problems by using MCMC techniques, we use a different method

where analytic results will be obtained without the need to deal with the significant computational cost. In the next chapter we will give some background information about stochastic calculus which will be very useful for the work we are going to carry out in the subsequent chapters.

Chapter 2

Stochastic Calculus

The purpose of this chapter is to give an introduction to the theory of Itô SDEs and SDEs with Markovian switching. The main topics discussed in this chapter include random variables, stochastic processes, Brownian motion, stochastic integration, SDEs, SDEs with Markovian switching and the stability theory. There are many books available on both theory and application of SDEs [5, 31, 33, 68, 69]. The contents of this chapter are mainly based on [68] and [69].

2.1 Random Variables

A random variable is a real-valued function defined on the set of outcomes of a random experiment. Random variables are important in understanding the stochastic theory because a stochastic integral is a random variable and the solution of a stochastic differential equation at any fixed time is a random variable as well. In this section, probability space, random variables and expectation are introduced.

A process that has random outcomes is called a *random experiment*. The set of all possible outcomes of a random experiment is called the *sample space* and is denoted as Ω. A combination of outcomes, a subset of Ω, is called an *event*. In general not every subset of the sample space is an observable or interesting event. Therefore we only group these observable or interesting events as a family $\mathcal{F}$ of subsets of Ω. For the purpose of probability theory, such a family $\mathcal{F}$ should have the following properties:

1. $\emptyset \in \mathcal{F}$, where $\emptyset$ denotes the empty set;
2. $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$, where $A^C = \Omega - A$ is the complement of A in Ω;
3. $\{A_i\}_{i\geq 1} \subset \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A family $\mathcal{F}$ with these three properties is called a *σ-algebra*. The pair $(\Omega, \mathcal{F})$ is called a *measurable space*, and the elements of $\mathcal{F}$ are called *$\mathcal{F}$-measurable sets*. If $\mathcal{C}$ is a family of subsets of Ω, then there exists a smallest σ-algebra $\sigma(\mathcal{C})$ on Ω which contains $\mathcal{C}$. This $\sigma(\mathcal{C})$ is called the *σ-algebra generated by $\mathcal{C}$*. If $\Omega = \mathbb{R}^n$ and $\mathcal{C}$ is the family of all open sets in $\mathbb{R}^n$, then $\mathcal{B}^n = \sigma(\mathcal{C})$ is called the *Borel σ-algebra* and the elements of $\mathcal{B}^n$ are called the *Borel sets*.

A *random function* X is a real-valued function that assigns the value $X(\omega) \in \mathbb{R}$ to each outcome $\omega \in \Omega$, that is $X : \Omega \to \mathbb{R}$. The random function X is said to be $\mathcal{F}$-*measurable* if

$$\{\omega : X(\omega) \leq a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}.$$

Any $\mathcal{F}$-measurable random function $X : \Omega \to \mathbb{R}$ is called a random variable on $(\Omega, \mathcal{F})$. An $\mathbb{R}^n$-valued function $X(\omega) = (X_1(\omega), \cdots, X_n(\omega))^T$ is said to be $\mathcal{F}$-*measurable* if all the elements X_i are $\mathcal{F}$-measurable. Similarly, an $n \times m$ matrix-valued function $X(\omega) = (X_{ij}(\omega))_{n\times m}$ is said to be $\mathcal{F}$-*measurable* if all the elements X_{ij} are $\mathcal{F}$-measurable.

The *indicator function* I_A of a set $A \subset \Omega$ is defined by

$$I_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A, \\ 0 & \text{for } \omega \notin A. \end{cases}$$

A *probability measure* $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

1. $\mathbb{P}(\Omega) = 1$;
2. for any disjoint sequence $\{A_i\}_{i\geq 1} \subset \mathcal{F}$ (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$)

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.*

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we set

$$\bar{\mathcal{F}} = \{A \subset \Omega : \exists\ B, C \in \mathcal{F} \text{ such that } B \subset A \subset C,\ \mathbb{P}(B) = \mathbb{P}(C)\}.$$

Then $\bar{\mathcal{F}}$ is a σ-algebra and is called the *completion* of $\mathcal{F}$. If $\mathcal{F} = \bar{\mathcal{F}}$, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *complete.*

Next we indroduce the concept of expectation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If X is a real-valued random variable and is *integrable* with respect to the probability measure $\mathbb{P}$, then the number

$$\mathbb{E}X = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

is called the *expectation* of X (with respect to $\mathbb{P}$).

For $p \in (0, \infty)$, let $L^p = L^p(\Omega; \mathbb{R}^n)$ be the family of $\mathbb{R}^n$-valued random variables X with $\mathbb{E}|X|^p < \infty$.

We now give the definition of conditional probability. Let $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. The *conditional probability of A under condition B* is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

We also introduce a more general concept of *conditional expectation.* Let $X \in L^1(\Omega; \mathbb{R})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$ so $(\Omega, \mathcal{G})$ is a measurable space. In general X is not

$\mathcal{G}$-measurable. We now seek an integrable $\mathcal{G}$-measurable random variable Y such that it has the same values as X on average in the sense that

$$\mathbb{E}(I_G Y) = \mathbb{E}(I_G X) \quad \text{i.e.} \quad \int_G Y(\omega) d\mathbb{P}(\omega) = \int_G X(\omega) d\mathbb{P}(\omega) \quad \forall\, G \in \mathcal{G}.$$

By the Radon-Nikodym theorem, there exists one such Y, almost surely unique. It is called the *conditional expectation of X under the condition $\mathcal{G}$*, and we write

$$Y = \mathbb{E}(X|\mathcal{G}).$$

2.2 Stochastic Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a family $\{\mathcal{F}_t\}_{t\geq 0}$ of increasing sub-σ-algebras of $\mathcal{F}$ (i.e. $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for all $0 \leq t < s < \infty$). The filtration is said to be *right continuous* if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \geq 0$. When the probability space is complete, the filtration is said to satisfy the *usual conditions* if it is right continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.

From now on, unless otherwise specified, we let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions.

A *stochastic process* is a family of $\mathbb{R}^n$-valued random variables $\{X_t\}_{t\in I}$ with *parameter set (or index set) I and set space $\mathbb{R}^n$*. The parameter set I is usually the halfline $\mathbb{R}_+ = [0, \infty)$, but it may also be an interval $[a, b]$, the nonnegative integers or even subsets of $\mathbb{R}^n$. Note that for each fixed $t \in I$ we have a random variable

$$\Omega \ni \omega \to X_t(\omega) \in \mathbb{R}^n.$$

On the other hand, for each fixed $\omega \in \Omega$ we have a function

$$I \ni t \to X_t(\omega) \in \mathbb{R}^n,$$

which is called a *sample path* of the process, and we shall write $X.(\omega)$ for the path. Sometimes we will write $X(t, \omega)$ instead of $X_t(\omega)$, and the stochastic process may be regarded as a function of two variables (t, ω) from $I \times \Omega$ to $\mathbb{R}^n$. Similarly, one can define matrix-valued stochastic processes etc. We often write a stochastic process $\{X_t\}_{t\geq 0}$ as $\{X_t\}$, X_t or $X(t)$.

An $\mathbb{R}^n$-valued stochastic process $\{X_t\}_{t\geq 0}$ is said to be *continuous* (resp. *right continuous, left continuous*) if for almost all $\omega \in \Omega$ function $X_t(\omega)$ is continuous (resp. *right continuous, left continuous*) on $t \geq 0$. It is said to be *integrable* if for every $t \geq 0$, X_t is an integrable random variable. It is said to be $\{\mathcal{F}_t\}-$*adapted* (or simply, *adapted*) if for every t, X_t is $\mathcal{F}_t$-measurable. A real-valued stochastic process $\{A_t\}_{t\geq 0}$ is called an *increasing process* if for almost all $\omega \in \Omega$, $A_t(\omega)$ is non-negative increasing right continuous on $t \geq 0$.

A random variable $\tau : \Omega \to [0, \infty]$ (it may take the value ∞) is called an $\{\mathcal{F}_t\}$-*stopping time* (or simply, *stopping time*) if $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for any $t \geq 0$.

We now give the useful definition of a martingale. An $\mathbb{R}^n$-valued $\{\mathcal{F}_t\}$-adapted integrable process $\{M_t\}_{t\geq 0}$ is called a *martingale with respect to* $\{\mathcal{F}_t\}$ (or simply, *martingale)* if

$$\mathbb{E}(M_t|\mathcal{F}_s) = M_s \quad \text{a.s. for all } 0 \leq s < t < \infty.$$

A stochastic process $X = \{X_t\}_{t\geq 0}$ is called *square-integrable* if $\mathbb{E}|X_t|^2 < \infty$ for every $t \geq 0$. If $M = \{M_t\}_{t\geq 0}$ is a real-valued square-integrable continuous martingale, then there exists a unique continuous integrable adapted increasing process denoted by $\{\langle M, M\rangle_t\}$ such that $\{M_t^2 - \langle M, M\rangle_t\}$ is a continuous martingale vanishing at $t = 0$. The process $\{\langle M, M\rangle_t\}$ is called the *quadratic variation* of M.

A right continuous adapted process $M = \{M_t\}_{t\geq 0}$ is called a *local martingale* if there exists a nondecreasing sequence $\{\tau_k\}_{k\geq 1}$ of stopping times with $\tau_k \uparrow \infty$ a.s. such that $\{M_{\tau_k \wedge t} - M_0\}_{t\geq 0}$ is a martingale. While every martingale is a local martingale, the opposite is not true.

Now we state the strong law of large numbers.

Theorem 2.1 *Let $M = \{M_t\}_{t\geq 0}$ be a real-valued continuous local martingale vanishing at $t = 0$. Then*

$$\limsup_{t\to\infty} \frac{\langle M, M\rangle_t}{t} < \infty \quad a.s. \quad \Longrightarrow \quad \lim_{t\to\infty} \frac{M_t}{t} = 0 \ a.s.$$

2.3 Brownian Motion and the Stochastic Integral

In 1827 the biologist Robert Brown made an observation on pollen grains in water through a microscope and he noted that the grains moved through the water but the mechanisms that caused the motion could not be determined. This transport phenomenon is therefore called Brownian Motion.

In mathematics, Brownian motion is described by the Wiener process which is a continuous-time stochastic process. The probability distribution of the position of the particle at time $t + dt$, given that its position at time t is p, follows a normal distribution with mean $p + \mu dt$ and variance $\sigma^2 dt$, where the parameter μ is the drift velocity and the parameter σ is the power of the noise. Hence we can see clearly that the process is Markovian (see section 2.4 below).

Definition 2.2 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A (standard) one-dimensional Brownian motion is a real-valued continuous $\{\mathcal{F}_t\}$-adapted process $\{B_t\}_{t\geq 0}$ with the following properties:*

1. *$B_0 = 0$ a.s.;*
2. *for $0 \leq s < t < \infty$, the increment $B_t - B_s$ is normally distributed with mean zero and variance $t - s$;*
3. *for $0 \leq s < t < \infty$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$.*

Some important properties of Brownian motion are summarised below:

1. $\{-B_t\}$ is a Brownian motion with respect to the same filtration $\{\mathcal{F}_t\}$.

2. Let $c > 0$. Define
$$X_t = \frac{B_{ct}}{\sqrt{c}} \quad \text{for } t \geq 0.$$
Then $\{X_t\}$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_{ct}\}$.

3. $\{B_t\}$ is a continuous square-integrable martingale and its quadratic variation $\langle B, B\rangle_t = t$ for all $t \geq 0$.

4. The strong law of large numbers states that
$$\lim_{t\to\infty} \frac{B_t}{t} = 0 \quad \text{a.s..}$$

5. For almost every $\omega \in \Omega$, the Brownian sample path $B.(\omega)$ is nowhere differentiable.

Definition 2.3 *An m-dimensional process $\{B_t = (B_t^1, \cdots, B_t^m)\}_{t\geq 0}$ is called an m-dimensional Brownian motion if every $\{B_t^i\}$ is a one-dimensional Brownian motion, and $\{B_t^1\}$,$\cdots$,$\{B_t^m\}$ are independent.*

We now define the stochastic integral
$$\int_0^t f(t)dB_s$$
with respect to an m-dimensional Brownian motion $\{B_t\}$ for a class of $n \times m$-matrix-valued stochastic processes $\{f(t)\}$. Since for almost all $\omega \in \Omega$, the Brownian sample path $B.(\omega)$ is nowhere differentiable, the integral cannot be defined in an ordinary way. However, we can define the integral for a large class of stochastic processes by making use of the stochastic nature of Brownian motion.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. Let $B = \{B_t\}_{t\geq 0}$ be a one-dimensional Brownian motion defined on the probability space adapted to the filtration.

Definition 2.4 *A real-valued stochastic process $g = \{g(t)\}_{a\leq t\leq b}$ is called a simple (or step) process if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ of $[a, b]$, and bounded random variables ξ_i, $0 \leq i \leq k-1$ such that ξ_i is $\mathcal{F}_{t_i}$-measurable and*
$$g(t) = \xi_0 I_{[t_0,t_1]}(t) + \sum_{t=1}^{k-1} \xi_i I_{(t_i,t_{i+1}]}(t). \tag{2.3.1}$$
We denote the family of all such processes by $\mathcal{M}_0([a, b]; \mathbb{R})$.

Definition 2.5 *For a simple process g with the form of (2.3.1) in $\mathcal{M}_0([a, b]; \mathbb{R})$, define*
$$\int_a^b g(t)dB_t = \sum_{i=0}^{k-1} \xi_i (B_{t_{i+1}} - B_{t_i})$$
and call it the stochastic integral of g with respect to the Brownian motion $\{B_t\}$ or the Itô integral.

We now extend the integral definition from simple processes to processes in $\mathcal{M}^2([a,b];\mathbb{R})$.

Definition 2.6 *Let $f \in \mathcal{M}^2([a,b];\mathbb{R})$. The Itô integral of f with respect to $\{B_t\}$ is defined by*

$$\int_a^b f(t)dB_t = \lim_{k\to\infty} \int_a^b g_k(t)dB_t \quad \text{in } L^2(\Omega,\mathbb{R}),$$

where $\{g_k\}$ is a sequence of simple processes such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t)-g_k(t)|^2\,dt = 0.$$

Some properties of the stochastic integral are summarised below:

Theorem 2.7 *Let $f,\ g \in \mathcal{M}^2([a,b];\mathbb{R})$, and $\alpha,\ \beta$ be two real numbers. Then*

1. *$\int_a^b f(t)dB_t$ is $\mathcal{F}_b$-measurable;*
2. *$\mathbb{E}\int_a^b f(t)dB_t = 0$;*
3. *$\mathbb{E}\left|\int_a^b f(t)dB_t\right|^2 = \mathbb{E}\int_a^b |f(t)|^2\,dt$;*
4. *$\int_a^b |\alpha f(t)+\beta g(t)|dB_t = \alpha\int_a^b f(t)dB_t + \beta\int_a^b g(t)dB_t$.*

We now define the Itô formula, which is not only useful in evaluating the Itô integrals but also plays a key role in stochastic analysis. Let $B(t) = (B_1(t),...,B_m(t))^T$, $t \geq 0$ be an m-dimensional Brownian motion defined on the complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

Definition 2.8 *An n-dimensional Itô process is an $\mathbb{R}^n$-valued continuous adapted process $x(t) = (x_1(t),\cdots,x_n(t))^T$ on $t \geq 0$ of the form*

$$x(t) = x(0) + \int_0^t f(s)ds + \int_0^t g(s)dB(s),$$

where $f = (f_1,\cdots,f_n)^T \in \mathcal{L}^1(\mathbb{R}_+;\mathbb{R}^n)$ and $g = (g_{ij})_{n\times m} \in \mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$. Then $x(t)$ has a stochastic differential $dx(t)$ on $t \geq 0$ given by

$$dx(t) = f(t)dt + g(t)dB(t).$$

We now state the multi-dimensional Itô formula.

Theorem 2.9 *Let $x(t)$ be an n-dimensional Itô process on $t \geq 0$ with the stochastic differential*

$$dx(t) = f(t)dt + g(t)dB(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+;\mathbb{R}^n)$ and $g \in \mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$. Let $V \in C^{2,1}(\mathbb{R}^n\times\mathbb{R}_+;\mathbb{R})$. Then $V(x(t),t)$ is a real-valued Itô process with its stochastic differential given by

$$\begin{aligned} dV(x(t),t) = &\left[V_t(x(t),t) + V_x(x(t),t)f(t) + \frac{1}{2}trace\left(g^T(t)V_{xx}(x(t),t)g(t)\right)\right]dt \\ &+ V_x(x(t),t)g(t)dB(t) \quad a.s. \end{aligned}$$

We now state the Gronwall inequality which will be useful in our research.

Theorem 2.10 *Let $T > 0$ and $c \geq 0$. Let $u(\cdot)$ be a Borel measurable bounded nonnegative function on [0,T], and let $v(\cdot)$ be a nonnegative integrable function on [0,T]. If*

$$u(t) \leq c + \int_0^t v(s)u(s)ds \quad \textit{for all} \quad 0 \leq t \leq T,$$

then

$$u(t) \leq c \; exp\left(\int_0^t v(s)ds\right) \quad \textit{for all} \quad 0 \leq t \leq T.$$

2.4 Markov Processes and the Generalised Itô Formula

In this section we introduce some basic concepts regarding the Markov process. An n-dimensional $\mathcal{F}_t$-adapted process $X = \{X_t\}_{t\geq 0}$ is called a *Markov process* if the following *Markov property* is satisfied: for all $0 \leq s \leq t < \infty$ and $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(X(t) \in A|\mathcal{F}_s) = \mathbb{P}(X(t) \in A|X(s)).$$

Equivalently, for any bounded Borel measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $0 \leq s \leq t < \infty$,

$$\mathbb{E}(\varphi(X(t))|\mathcal{F}_s) = \mathbb{E}(\varphi(X(t))|X(s)).$$

An n-dimensional process $\{X_t\}_{t\geq 0}$ is called a *strong Markov process* if the following *strong Markov property* is satisfied: for any bounded Borel measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$, any finite $\{\mathcal{F}_t\}$-stopping time τ and $t \geq 0$,

$$\mathbb{E}(\varphi(X(t+\tau))|\mathcal{F}_\tau) = \mathbb{E}(\varphi(X(t+\tau))|X(\tau)).$$

Especially, in the homogeneous case, this becomes

$$\mathbb{E}(\phi(X(t+\tau))|\mathcal{F}_\tau) = \mathbb{E}_{X(\tau)}\phi(X(t)).$$

A stochastic process $\{X_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in a countable set Ξ (to be called the *state space* of the process), is called a *continuous-time Markov chain* if for any finite set $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$ of 'times', and corresponding set $i_1, i_2, \cdots, i_{n-1}, i, j$ of states in Ξ such that $\mathbb{P}\{X(t_n) = i, X(t_{n-1}) = i_{n-1}, \cdots, X(t_1) = i_1\} > 0$, we have

$$\mathbb{P}\{X(t_{n+1}) = j|X(t_n) = i, X(t_{n-1}) = i_{n-1}, \cdots, X(t_1) = i_1\} = \mathbb{P}\{X(t_{n+1} = j|X(t_n) = i\}.$$

If for all $s,\ t$ such that $0 \leq s \leq t < \infty$ and all $i, j \in \Xi$ the conditional probability $\mathbb{P}\{X(t) = j|X(s) = i\}$ depends only on $t - s$, we say that the process $X = \{X(t)\}_{t\geq 0}$ is *homogeneous*. In this case, then, $\mathbb{P}\{X(t) = j|X(s) = i\} = \mathbb{P}\{X(t-s) = j|X(0) = i\}$, and the function

$$P_{i,j}(t) =: \mathbb{P}\{X(t) = j|X(0) = i\}, \quad i, j \in \Xi, t \geq 0,$$

is called the *transition function* or *transition probability* of the process. The function $P_{i,j}(t)$ is called *standard* if $\lim_{t\to 0} P_{ii}(t) = 1$ for all $i \in \Xi$.

Theorem 2.11 *Let $P_{i,j}(t)$ be a standard transition function, then $\gamma_i := \lim_{t\to 0}[1-P_{ii}(t)]/t$ exists (but may be ∞) for all $i \in \Xi$.*

A state $i \in \Xi$ is said to be *stable* if $\gamma_i < \infty$.

Theorem 2.12 *Let $P_{i,j}(t)$ be a standard transition function, and let j be a stable state. Then $\gamma_{ij} = P'_{ij}(0)$ exists and is finite for all $i \in \Xi$.*

Let $\gamma_{ii} = -\gamma_i$ and $\Gamma = (\gamma_{ij})_{i,j\in\Xi}$. Γ is called the *generator* of the Markov chain. If the state space is *finite* which we can take to be $\mathbb{S} = \{1, 2, ..., N\}$, then the process is called a continuous-time *finite* Markov chain. From now on, unless otherwise specified, we assume that all Markov chains are finite and all states are stable. For such a Markov chain, almost every sample path is a right continuous step function.

Theorem 2.13 *Let $P(t) = (P_{ij}(t))_{N\times N}$ be the transition probability matrix and $\Gamma = (\gamma_{ij})_{N\times N}$ be the generator of a finite Markov chain. Then*

$$P(t) = e^{t\Gamma}.$$

A continuous -time Markov chain $X(t)$ with generator $\Gamma = (\gamma_{ij})_{N\times N}$ can be represented as a stochastic integral with respect to a Poisson random measure. Let $\Delta_{i,j}$ be consecutive, left closed, right open intervals of the real line each having length γ_{ij} such that

$$\begin{aligned}
\Delta_{12} &= [0, \gamma_{12}),\\
\Delta_{13} &= [\gamma_{12}, \gamma_{12} + \gamma_{13}),\\
&\vdots\\
\Delta_{1N} &= \left[\sum_{j=2}^{N-1}\gamma_{1j}, \sum_{j=2}^{N}\gamma_{1j}\right),\\
\Delta_{21} &= \left[\sum_{j=2}^{N}\gamma_{1j}, \sum_{j=2}^{N}\gamma_{1j} + \gamma_{21}\right),\\
\Delta_{22} &= \left[\sum_{j=2}^{N}\gamma_{1j} + \gamma_{21}, \sum_{j=2}^{N}\gamma_{1j} + \gamma_{21} + \gamma_{23}\right),\\
&\vdots\\
\Delta_{2N} &= \left[\sum_{j=2}^{N}\gamma_{1j} + \sum_{j=1,j\neq 2}^{N-1}\gamma_{2j}, \sum_{j=2}^{N}\gamma_{1j} + \sum_{j=1,j\neq 2}^{N}\gamma_{2j}\right)
\end{aligned}$$

and so on. Define a function

$$h : \mathbb{S}\times\mathbb{R} \to \mathbb{R}$$

by

$$h(i, y) = \begin{cases} j - i & \text{if } y \in \Delta_{ji},\\ 0 & \text{otherwise.}\end{cases} \tag{2.4.1}$$

Then

$$dX(t) = \int_{\mathbb{R}} h(X(t-), y)\nu(dt, dy), \tag{2.4.2}$$

with initial condition $X(0) = i_0$, where $\nu(dt, dy)$ is a Poisson random measure with intensity $dt \times \mu(dy)$, in which μ is the Lebesgue measure on $\mathbb{R}$.

Let $B(t) = (B_1(t), ..., B_m(t))^T$, $t \geq 0$ be an m-dimensional Brownian motion defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let $r(t)$, $t \geq t_0$, be a right-continuous Markov chain on the probability space taking values in a finite state space $\mathbb{S} = 1, 2, \cdots, N$ with generator $\Gamma = (\gamma_{ij})_{N\times N}$ given by

$$\mathbb{P}(r(t+\delta) = j | r(t) = i) = \begin{cases} \gamma_{ij}\delta + o(\delta) & \text{if } i \neq j, \\ 1 + \gamma_{ii}\delta + o(\delta) & \text{if } i = j, \end{cases}$$

where $\delta > 0$. Here $\gamma_{ij} \geq 0$ is the transition rate from i to j if $i \neq j$ while

$$\gamma_{ii} = -\sum_{j\neq i} \gamma_{ij}.$$

We assume that the Markov chain $r(\cdot)$ is $\mathcal{F}_t$-adapted but independent of the Brownian motion $B(\cdot)$.

Let $x(t)$ be an n-dimensional Itô process on $t \geq 0$ with the stochastic differential

$$dx(t) = f(t)dt + g(t)dB(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ and $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times m})$. The Itô formula established in the last section shows that a $C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+; \mathbb{R})$-function V maps the Itô process $x(t)$ onto another Itô process $V(x(t), t)$. However, here we will consider the paired process $(x(t), r(t))$ so we define a new function $V : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}$, which will map $(x(t), r(t))$ onto another process $V(x(t), t, r(t))$. Let $C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$ denote the family of all real-valued functions $V(x, t, i)$ on $\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}$ which are continuously twice differentiable in x and once in t. If $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$, define an operator LV from $\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}$ to $\mathbb{R}$ by

$$LV(x, t, i) = V_t(x, t, i) + V_x(x, t, i)f(t) + \frac{1}{2}\text{trace}[g^T(t)V_{xx}(x, t, i)g(t)] + \sum_{j=1}^{N} \gamma_{ij}V(x, t, j),$$

where

$$V_t(x, t, i) = \frac{\partial V(x, t, i)}{\partial t}, \quad V_x(x, t, i) = \left(\frac{\partial V(x, t, i)}{\partial x_1}, \cdots, \frac{\partial V(x, t, i)}{\partial x_n}\right)$$

and

$$V_{xx}(x, t, i) = \left(\frac{\partial^2 V(x, t, i)}{\partial x_i \partial x_j}\right)_{n\times n}.$$

We now state the generalised Itô formula, which reveals how V maps the paired process $(x(t), r(t))$ onto a new process $V(x(t), t, r(t))$.

Theorem 2.14 *If $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$, then for any $t \geq 0$*

$$\begin{aligned} V(x(t), t, r(t)) =& V(x(0), 0, r(0)) + \int_0^t LV(x(s), s, r(s))ds \\ &+ \int_0^t V_x(x(s), s, r(s))g(x(s), s, r(s))dB(s) \\ &+ \int_0^t \int_R (V(x(s), s, i_0 + h(r(s), l)) - V(x(s), s, r(s)))\mu(ds, dl), \end{aligned}$$

where the function h is defined by (2.4.1) and $\mu(ds, dl) = \nu(ds, dl) - \mu(dl)ds$ is a martingale measure while ν and μ have been defined in (2.4.2).

2.5 Stochastic Differential Equations

In this section we present some concepts regarding SDEs which are essential to our research. We define a general version of SDE and specify the existence and uniqueness criteria for the solution.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. Let $B(t) = (B_1(t), ..., B_m(t))^T$, $t \geq 0$ be an m-dimensional Brownian motion defined on the space. Let $0 \leq t_0 < T < \infty$. Let $x_0 \in L^2_{\mathcal{F}_{t_0}}(\Omega, \mathbb{R}^n)$, i.e. an $\mathcal{F}_{t_0}$-measurable $\mathbb{R}^n$-valued random variable such that $\mathbb{E}|x_0|^2 < \infty$. Let $f : \mathbb{R}^n \times [t_0, T] \to \mathbb{R}^n$ and $g : \mathbb{R}^n \times [t_0, T] \to \mathbb{R}^{n\times m}$ be both Borel measurable. Consider the n-dimensional stochastic differential equation of Itô type

$$dx(t) = f(x(t), t)dt + g(x(t), t)dB(t), \quad t_0 \leq t \leq T \tag{2.5.1}$$

with initial value $x(t_0) = x_0$. By the definition of the stochastic differential, this equation is equivalent to the following stochastic integral equation

$$x(t) = x_0 + \int_{t_0}^t f(x(s), s)ds + \int_{t_0}^t g(x(s), s)dB(s) \quad \forall t \in [t_0, T]. \tag{2.5.2}$$

We first give the definition of the solution.

Definition 2.15 *An $\mathbb{R}^n$-valued stochastic process $\{x(t)\}_{t_0\leq t\leq T}$ is called a* solution *of equation (2.5.1) if it has the following properties:*

1. *$\{x(t)\}$ is continuous and $\mathcal{F}_t$-adapted;*
2. *$\{f(x(t), t)\} \in \mathcal{L}^1([t_0, T]; \mathbb{R}^n)$ and $\{g(x(t), t)\} \in \mathcal{L}^2([t_0, T]; \mathbb{R}^{n\times m})$;*
3. *equation (2.5.2) holds for every $t \in [t_0, T]$ with probability one.*

A solution $\{x(t)\}$ is said to be unique *if any other solution $\{\bar{x}(t)\}$ is indistinguishable from $\{x(t)\}$, that is*

$$\mathbb{P}\{x(t) = \bar{x}(t) \text{ for all } t_0 \leq t \leq T\} = 1.$$

We now state the conditions that guarantee the existence and uniqueness of the solution to equation (2.5.1).

Theorem 2.16 *Assume that there exist two positive constants $\bar{K}$ and K such that (Lipschitz condition) for all $x, y \in \mathbb{R}^n$ and $t \in [t_0, T]$*

$$|f(x,t) - f(y,t)|^2 \vee |g(x,t) - g(y,t)|^2 \leq \bar{K}|x-y|^2;$$

(Linear growth condition) for all $(x,t) \in \mathbb{R}^n \times [t_0, T]$

$$|f(x,t)|^2 \vee |g(x,t)|^2 \leq K(1+|x|^2).$$

Then there exists a unique solution $x(t)$ to equation (2.5.1) and the solution belongs to $\mathcal{M}^2([t_0,T];\mathbb{R}^n)$.

Explicit solutions are usually not obtainable for most of SDEs. It is therefore important to construct approximate solutions. There are various ways to do this [52]. We use the Euler-Maruyama (EM) approximate solutions to equation (2.5.1), which are defined as follows: Let $\{x_k\}_{k=0}^n$ be the observations from (2.5.1). Given a step size Δt and defining $x(t_0) = x_0$, and then for small time intervals $k\Delta t \leq t \leq (k+1)\Delta t$

$$x_j = x_{j-1} + \Delta t f(x_{j-1}, t_{j-1}) + g(x_{j-1}, t_{j-1})\Delta W_k,$$

where $\Delta W_k = W_{k+1} - W_k$ is an increment of a Wiener process and follows $N(0, \sqrt{\Delta t})$.

Now we construct the explicit solution to the general n-dimensional linear stochastic differential equation

$$dx(t) = (F(x)x(t) + f(t))dt + \sum_{k=1}^{m}(G_k(t)x(t) + g_k(t))dB_k(t) \tag{2.5.3}$$

on $[t_0, T]$, where $F(\cdot)$, $G(\cdot)$ are $n \times n$-valued functions, $f(\cdot)$, $g(\cdot)$ are $\mathbb{R}^n$-valued functions. We shall assume that $F,\ f,\ G_k,\ g_k$ are all Borel-measurable and bounded on $[t_0, T]$. Therefore, by the existence-and-uniqueness Theorem 2.16, the linear equation (2.5.3) has a unique continuous solution in $\mathcal{M}^2([t_0,T];\mathbb{R}^n)$ for every initial value $x(t_0) = x_0$, which is $\mathcal{F}_{t_0}$-measurable and belongs to $L^2(\Omega;\mathbb{R}^n)$.

Consider the linear stochastic differential equation

$$dx(t) = F(t)x(t)dt + \sum_{k=1}^{m} G_k(t)x(t)dB_k(t) \tag{2.5.4}$$

on $[t_0, T]$. As assumed,

$$F(t) = (F_{ij}(t))_{n\times n} \quad G_k(t) = (G_{ij}^k(t))_{n\times n}$$

are all Borel-measurable and bounded. For every $j = 1, ..., n$, let e_j be the unit column-vector in the x_j-direction, i.e.

$$e_j = (\underbrace{0, \cdots, 0, 1}_{j}, 0, \cdots, 0)^T.$$

Let $\Phi_j(t) = (\Phi_{1j}(t) \cdots, \Phi_{nj}(t))^T$ be the solution of equation (2.5.4) with initial value $x(t_0) = e_j$. Define the $n \times n$ matrix

$$\Phi(t) = (\Phi_1(t) \cdots, \Phi_n(t)) = (\Phi_{ij})_{n\times n}.$$

We call $\Phi(t)$ the fundamental matrix of equation (2.5.4). The following theorem states that any solution of equation (2.5.4) can be expressed in terms of $\Phi(t)$ and that is why $\Phi(t)$ is called the fundamental matrix.

Theorem 2.17 *Given the initial value $x(t_0) = x_0$, the unique solution of equation (2.5.4) is*

$$x(t) = \Phi(t)x_0.$$

Let us turn to the general n-dimensional linear stochastic differential equation (2.5.3). We now state the variation of constants formula.

Theorem 2.18 *The unique solution of equation (2.5.3) can be expressed as*

$$x(t) = \Phi(t)\left(x_0 + \int_{t_0}^{t} \Phi^{-1}(s)\left(f(s) - \sum_{k=1}^{m} G_k(s)g_k(s)\right)ds + \sum_{k=1}^{m}\int_{t_0}^{t} \Phi^{-1}(s)g_k(s)dB_k(s)\right),$$

where $\Phi(t)$ is the fundamental matrix of the corresponding homogeneous equation (2.5.4).

2.6 Stochastic Differential Equations with Markovian Switching

We now discuss SDEs with Markovian switching which will be very useful when we conduct our research in Chapter 5.

Let $B(t) = (B_1(t), ..., B_m(t))^T$, $t \geq 0$ be an m-dimensional Brownian motion defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let $r(t)$, $t \geq t_0$, be a right-continuous Markov chain which has the same definition as in section 2.4. We assume that the Markov chain $r(\cdot)$ is $\mathcal{F}_t$-adapted but independent of the Brownian motion $B(\cdot)$. Consider an SDE with Markovian switching of the form

$$dx(t) = f(x(t), t, r(t))dt + g(x(t), t, r(t))dB(t), \quad t_0 \geq t \geq T \tag{2.6.1}$$

with initial data $x(t_0) = x_0 \in L^2_{\mathcal{F}_{t_0}}(\Omega; \mathbb{R}^n)$ and $r(t_0) = r_0$, where r_0 is an $\mathbb{S}$-valued $\mathcal{F}_{t_0}$-measurable random variable and

$$f : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S} \to \mathbb{R}^n \quad \text{and} \quad f : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S} \to \mathbb{R}^{n\times m}.$$

Definition 2.19 *An $\mathbb{R}^n$-valued stochastic process $\{x(t)\}_{t_0\leq t\leq T}$ is called a* solution *of equation (2.6.1) if it has the following properties:*

1. *$\{x(t)\}_{t_0\leq t\leq T}$ is continuous and $\mathcal{F}_t$-adapted;*

2. $\{f(x(t),t,r(t))\}_{t_0\le t\le T}\in\mathcal{L}^1([t_0,T];\mathbb{R}^n)$ *while* $\{g(x(t),t,r(t))\}_{t_0\le t\le T}\in\mathcal{L}^2([t_0,T];\mathbb{R}^{n\times m})$;

3. *for any* $t\in[t_0,T]$, *the equation*

$$x(t)=x(t_0)+\int_{t_0}^{t}f(x(s),s,r(s))ds+\int_{t_0}^{t}g(x(s),s,r(s))dB(s)$$

holds with probability 1.

Theorem 2.20 *Assume that there exist two positive constants* $\bar{K}$ *and* K *such that*

(Lipschitz condition) for all $x,y\in\mathbb{R}^n$ *and* $t\in[t_0,T]$ *and* $i\in\mathbb{S}$

$$|f(x,t,i)-f(y,t,i)|^2\vee|g(x,t,i)-g(y,t,i)|^2\le\bar{K}|x-y|^2;$$

(Linear growth condition) for all $(x,t,i)\in\mathbb{R}^n\times[t_0,T]\times\mathbb{S}$

$$|f(x,t,i)|^2\vee|g(x,t,i)|^2\le K(1+|x|^2).$$

Then there exists a unique solution $x(t)$ *to equation (2.6.1) and the solution belongs to* $\mathcal{M}^2([t_0,T];\mathbb{R}^n)$.

2.7 Stochastic Stability

Stability of a system means the insensitivity of the system to small changes in the initial value of the system. Stability is a desired property for a system to have. The main technique to show the stability of the solutions to SDEs or SDEs with Markovian switching is the method of Lyapunov functions which is discussed in detail in the books of Mao [68] and [69]. In this section we are going to introduce the definitions of only a few types of stability. We will then examine these types of stability for our stochastic version of SIS epidemic models in the subsequent chapters.

We consider the n-dimensional SDE which is defined in equation (2.5.1). We shall assume that the assumptions of the existence-and-uniqueness Theorem 2.16 are fulfilled. Hence, for any given initial value $x(t_0)=x_0\in\mathbb{R}^n$, equation (2.5.1) has a unique global solution which is denoted by $x(t;t_0,x_0)$. We know that the solution has continuous sample paths and each of its moments is finite. Assume furthermore that

$$f(0,t)=0\quad\text{and}\quad g(0,t)=0\quad\text{for all } t\ge t_0.$$

So equation (2.5.1) has the solution $x(t)\equiv0$ corresponding to the initial value $x(t_0)=0$. This solution is called the *trivial solution* or *equilibrium position*.

We now define three different types of stability.

Definition 2.21 *1. The trivial solution of equation (2.5.1) is said to be stochastically stable or stable in probability if for every pair of* $\varepsilon\in(0,1)$ *and* $r>0$, *there exists a* $\delta=\delta(\varepsilon,r,t_0)>0$ *such that*

$$P\{|x(t;t_0,x_0)|<r\quad\textit{for all}\quad t\ge t_0\}\ge1-\varepsilon$$

whenever $|x_0|\le\delta$.

2. *The trivial solution is said to be stochastically asymptotically stable if it is stochastically stable and, moreover, for every $\varepsilon \in (0,1)$, there exists a $\delta_0 = \delta_0(\varepsilon, t_0) > 0$ such that*

$$P\{\lim_{t\to\infty} x(t;t_0,x_0) = 0\} \geq 1-\varepsilon$$

whenever $|x_0| \leq \delta_0$.

3. *The trivial solution is said to be stochastically asymptotically stable in the large if it is stochastically stable and, moreover, for all $x_0 \in \mathbb{R}^n$*

$$P\{\lim_{t\to\infty} x(t;t_0,x_0) = 0\} = 1.$$

Definition 2.22 *The trivial solution of equation (2.5.1) is said to be almost surely exponentially stable if*

$$\limsup_{t\to\infty} \frac{1}{t}\log|x(t;t_0,x_0)| \leq 0 \quad a.s.$$

for all $x_0 \in \mathbb{R}^n$.

Definition 2.23 *The trivial solution of equation (2.5.1) is said to be pth moment exponentially stable if there is a pair of positive constants λ and C such that*

$$\mathbb{E}|x(t;t_0,x_0)|^p \leq C|x_0|^p e^{-\lambda(t-t_0)} \quad \text{on } t \geq t_0$$

for all $x_0 \in \mathbb{R}^n$. When $p = 2$, it is usually said to be exponentially stable in mean square.

Now we consider the non-linear SDE with Markovian switching which is defined in equation (2.6.1). We assume that conditions in Theorem 2.20 are fulfilled so that equation (2.6.1) has a unique global solution, which is denoted by $x(t;t_0,x_0,r_0)$. We shall let $f(0,t,i) \equiv 0$ and $g(0,t,i) \equiv 0$ so equation (2.6.1) has the trivial solution $x(t;t_0,0,i) = 0$ or the equilibrium state zero. We now discuss the exponential stability.

Definition 2.24 *For $p > 0$, the trivial solution of equation (2.6.1) or, simply, equation (2.6.1), is said to be pth moment exponentially stable if the pth moment Lyapunov exponent*

$$\limsup_{t\to\infty} \frac{1}{t}(\mathbb{E}|x(t;t_0,x_0,r_0)|^p) < 0$$

for all $(t_0,x_0,r_0) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{S}$. When $p = 2$, it is said to be exponentially stable in mean square. Moreover, it is said to be almost surely exponentially stable if the sample Lyapunov exponent

$$\limsup_{t\to\infty} \frac{1}{t}(\mathbb{E}|x(t;t_0,x_0,r_0)|) < 0 \quad a.s.$$

for all $(t_0,x_0,r_0) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{S}$.

Theorem 2.25 *If for each $i \in \mathbb{S}$, there are constant triples α_i, ρ_i and θ_i, such that*

$$\begin{aligned} &x^T f(x,t,i) \leq \alpha_i|x|^2, \\ &|g(x,t,i)| \leq \rho_i|x|, \\ &|x^T g(x,t,i)| \geq \theta_i|x|^2, \end{aligned}$$

for all $(x,t) \in \mathbb{R}^n \times \mathbb{R}_+$, *the solution of equation (2.6.1) satisfies*

$$\limsup_{t\to\infty} \frac{1}{t}\log(|x(t;t_0)|) \le \sum_{j=1}^{N} \pi_j \left(\alpha_j + \frac{1}{2}\rho_j^2 - \theta_j^2\right) \quad a.s.$$

for all $x_0 \in \mathbb{R}^n$. *In particular, the non-linear SDE with Markovian switching (2.6.1) is almost surely exponentially stable, if*

$$\sum_{j=1}^{N} \pi_j \left(\alpha_j + \frac{1}{2}\rho_j^2 - \theta_j^2\right) < 0.$$

Next we give the definition for moment and almost sure asymptotic stability.

Definition 2.26 *For* $p > 0$, *the trivial solution of equation (2.6.1) or, simply, equation (2.6.1), is said to be asymptotically stable in pth moment if*

$$\limsup_{t\to\infty}(\mathbb{E}|x(t;t_0,x_0,r_0)|^p) = 0$$

for all $(t_0,x_0,r_0) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{S}$. *When* $p = 2$, *it is said to be asymptotically stable in mean square. Moreover, it is said to be almost surely asymptotically stable or asymptotically stable with probability one if*

$$\limsup_{t\to\infty} x(t;t_0,x_0,r_0) = 0 \quad a.s.$$

for all $(t_0,x_0,r_0) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{S}$.

Now we state the definition for stability in probability and we shall also use the notation $\mathbb{S}_\delta = \{x \in \mathbb{R}^n : |x| < \delta\}$.

Definition 2.27 *1. The trivial solution of equation (2.6.1) is said to be stochastically stable or stable in probability if for every triple of* $\varepsilon \in (0,1)$, $\rho > 0$ *and* $t_0 \ge 0$, *there exists a* $\delta = \delta(\varepsilon,\rho,t_0) > 0$ *such that*

$$\mathbb{P}\{|x(t;t_0,x_0,i)| < \rho \text{ for all } t \ge t_0\} \ge 1-\varepsilon$$

for any $(x_0,i) \in \mathbb{S}_\delta \times \mathbb{S}$.

2. The trivial solution is said to be stochastically asymptotically stable or asymptotically stable in probability if it is stochastically stable and, moreover, for every pair of $\varepsilon \in (0,1)$ *and* $t_0 \ge 0$, *there exists a* $\delta_0 = \delta_0(\varepsilon,t_0) > 0$ *such that*

$$\mathbb{P}\{\lim_{t\to\infty} x(t;t_0,x_0) = 0\} \ge 1-\varepsilon$$

whenever $(x_0,i) \in \mathbb{S}_{\delta_0} \times \mathbb{S}$.

3. The trivial solution is said to be stochastically asymptotically stable in the large *if it is stochastically stable and, moreover,*

$$\mathbb{P}\{\lim_{t\to\infty} x(t;t_0,x_0,i) = 0\} = 1 \quad \forall (t_0,x_0,i) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{S}.$$

In this chapter, we have reviewed some background knowledge about the stochastic theory which will be very useful for our following research work. In the next chapter we extend the classical SIS epidemic model model from a deterministic framework to a stochastic one, with white noise being incorporated by the parameter perturbation technique.

Chapter 3

A Stochastic Differential Equation SIS Epidemic Model

3.1 Introduction

The deterministic SIS epidemic model is one of the simplest possible epidemic models, which has applications to transmission of real life diseases, such as pneumococcus, gonorrhea and tuberculosis. Recall that in section 1.2 we discussed the SIS epidemic model in detail, where the model is defined in equation (1.2.1) and its corresponding basic reproduction number R_0^D is defined in equation (1.2.3). Due to the random nature of the population system, environmental noise such as white noise can affect the population system significantly. Recall that in section 1.3.2, we reviewed the existing literature regarding introducing white noise into deterministic compartmental models and its influence on the deterministic system. We found that the common technique for introducing white noise is parameter perturbation and the introduction of white noise changes the conditions for the system to become extinct. The stochastic version of the SIR and SIRS epidemic models obtained by parameter perturbation have been discussed in the previous research but we are not aware of any literature addressing this issue in SIS epidemic models. This chapter is an attempt to fill this gap.

This chapter will be organised as follows: in section 3.2, we formulate the SIS epidemic model as a stochastic differential equation (SDE) for the number of infectious individuals $I(t)$. In section 3.3, we prove that this SDE has a unique global positive solution $I(t)$. In sections 3.4 and 3.5, we establish conditions for extinction and persistence of $I(t)$ and then we discuss perturbation by stochastic noise. In section 3.6, we show the existence of a stationary distribution and derive expressions for its mean and variance for the case of persistence. The results are illustrated by computer simulations, including two examples based on real life diseases, which are discussed in section 3.7. We draw conclusions and outline the future work required in section 3.8.

3.2 Stochastic Differential Equation SIS Model

Throughout this chapter, we let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions and we let $B(t)$ be a scalar Brownian motion defined on the probability space. Let us now consider the second equation of (1.2.1). To establish the stochastic differential equation (SDE) model, we naturally rewrite this equation in the differential form

$$dI(t) = [\beta S(t)I(t) - (\mu + \gamma)I(t)]dt. \tag{3.2.1}$$

Here $[t, t+dt)$ is a small time interval and we use the notation $d\cdot$ for the small change in any quantity over this time interval when we intend to consider it as an infinitesimal change, for example $dI(t) = I(t+dt) - I(t)$ and the change $dI(t)$ is described by (3.2.1). Consider the disease transmission coefficient β in the deterministic model. This can be thought of as the rate at which each infectious individual makes potentially infectious contacts with each other individual, where a potentially infectious contact will transmit the disease if the contact is made by an infectious individual with a susceptible individual. Thus the total number of new infections in the small time interval $[t, t+dt)$ is

$$\beta S(t)I(t)dt$$

and a single infected individual makes

$$\beta dt$$

potentially infectious contacts with each other individual in the small time interval $[t, t+dt)$.

Now suppose that some stochastic environmental factor acts simultaneously on each individual in the population. In this case β changes to a random variable $\tilde{\beta}$. Each infected individual makes precisely

$$\tilde{\beta}dt = \beta dt + \sigma dB(t)$$

potentially infectious contacts with each other individual in $[t, t+dt)$. Here $dB(t) = B(t+dt) - B(t)$ is the increment of a standard Brownian motion. Thus the number of potentially infectious contacts that a single infected individual makes with another individual in $[t, t+dt)$ is normally distributed with mean βdt and variance $\sigma^2 dt$. Hence $E(\tilde{\beta}dt) = \beta dt$ and $\text{var}(\tilde{\beta}dt) = \sigma^2 dt$. As $\text{var}(\tilde{\beta}dt) \to 0$ as $dt \to 0$ this is a biologically reasonable model. Indeed this is a well-established way of introducing stochastic environmental noise into biologically realistic population dynamic models. See [26, 29, 33, 38, 58, 63, 90] and many other references.

To motivate our assumption we argue as follows: Suppose that the number of potentially infectious contacts between an infectious individual and another individual in successive time intervals $[t, t+T), [t+T, t+2T), \ldots, [t+(n-1)T, t+nT)$ are independent, identically distributed random variables and n is very large. Then by the Central Limit Theorem the total number of potentially infectious contacts made in $[t, t+nT)$

has approximately a normal distribution with mean $n\mu_0$ and variance $n\sigma_0^2$, where μ_0 and σ_0^2 are respectively the mean and variance of the underlying distribution in each of the separate time intervals of length T. Thus it is reasonable to assume that the total number of potentially infectious contacts has a normal distribution whose mean and variance scale as the total length of the time interval as in our assumptions.

Therefore we replace βdt in equation (3.2.1) by $\tilde{\beta}dt = \beta dt + \sigma dB(t)$ to get

$$dI(t) = S(t)I(t)(\beta dt + \sigma dB(t)) - (\mu + \gamma)I(t)dt. \tag{3.2.2}$$

Note that βdt now denotes the mean of the stochastic number of potentially infectious contacts that an infected individual makes with another individual in the infinitesimally small time interval $[t, t+dt)$. Similarly, the first equation of (1.2.1) becomes another SDE. That is, the deterministic SIS model (1.2.1) becomes the Itô SDE

$$\begin{cases} dS(t) = [\mu N - \beta S(t)I(t) + \gamma I(t) - \mu S(t)]dt - \sigma S(t)I(t)dB(t), \\ dI(t) \ = [\beta S(t)I(t) - (\mu + \gamma)I(t)]dt + \sigma S(t)I(t)dB(t). \end{cases} \tag{3.2.3}$$

This SDE is called an SDE SIS model.

Given that $S(t) + I(t) = N$, it is sufficient to study the SDE for $I(t)$

$$dI(t) = I(t)\Big([\beta N - \mu - \gamma - \beta I(t)]dt + \sigma(N - I(t))dB(t)\Big) \tag{3.2.4}$$

with initial value $I(0) = I_0 \in (0, N)$. In the following sections we will concentrate on this SDE only.

3.3 Existence of Unique Positive Solution

The SDE SIS model (3.2.4) is a special SDE. In order for the model to make sense, we need to show at least that this SDE SIS model does not only have a unique global solution but also the solution will remain within $(0, N)$ whenever it starts from there. The existing general existence-and-uniqueness theorem on SDEs, Theorem 2.16, can not be applied to this special SDE in order to guarantee these properties. It is therefore necessary to establish such a new theory.

Theorem 3.1 *For any given initial value $I(0) = I_0 \in (0, N)$, the SDE (3.2.4) has a unique global positive solution $I(t) \in (0, N)$ for all $t \geq 0$ with probability one, namely*

$$\mathbb{P}\{I(t) \in (0, N) \text{ for all } t \geq 0\} = 1.$$

Proof. Regarding equation (3.2.4) as an SDE on $\mathbb{R}$, we see that its coefficients are locally Lipschitz continuous. It is known (see e.g. [66, 67, 68]) that for any given initial value $S_0 \in (0, N)$ there is a unique maximal local solution $I(t)$ on $t \in [0, \tau_e)$, where τ_e is the explosion time. Let $k_0 > 0$ be sufficiently large for $1/k_0 < I_0 < N - (1/k_0)$. For each integer $k \geq k_0$, define the stopping time

$$\tau_k = \inf\{t \in [0, \tau_e) : I(t) \notin (1/k,\ N - (1/k))\},$$

where throughout this thesis we set $\inf \emptyset = \infty$. Clearly, τ_k is increasing as $k \to \infty$. Set $\tau_\infty = \lim_{k\to\infty} \tau_k$, whence $\tau_\infty \leq \tau_e$ a.s. If we can show that $\tau_\infty = \infty$ a.s., then $\tau_e = \infty$ a.s. and $I(t) \in (0, N)$ a.s. for all $t \geq 0$. In other words, to complete the proof all we need to show is that $\tau_\infty = \infty$ a.s. If this statement is false, then there is a pair of constants $T > 0$ and $\varepsilon \in (0, 1)$ such that

$$\mathbb{P}\{\tau_\infty \leq T\} > \varepsilon.$$

Hence there is an integer $k_1 \geq k_0$ such that

$$\mathbb{P}\{\tau_k \leq T\} \geq \varepsilon \quad \text{for all } k \geq k_1. \tag{3.3.1}$$

Define a function $V : (0, N) \to \mathbb{R}_+$ by

$$V(x) = \frac{1}{x} + \frac{1}{N - x}.$$

By the Itô formula which is stated in Theorem 2.9, we have, for any $t \in [0, T]$ and $k \geq k_1$,

$$\mathbb{E}V(I(t \wedge \tau_k)) = V(I_0) + \mathbb{E}\int_0^{t\wedge\tau_k} LV(I(s))ds, \tag{3.3.2}$$

where $LV : (0, N) \to \mathbb{R}$ is defined by

$$\begin{aligned} LV(x) &= x\bigg(-\frac{1}{x^2} + \frac{1}{(N-x)^2}\bigg)[\beta N - \mu - \gamma - \beta x] \\ &+ \sigma^2 x^2 (N-x)^2 \bigg(\frac{1}{x^3} + \frac{1}{(N-x)^3}\bigg). \end{aligned} \tag{3.3.3}$$

It is easy to show that

$$LV(x) \leq \frac{\mu + \gamma}{x} + \frac{\beta N}{N - x} + \sigma^2 N^2 \bigg(\frac{1}{x} + \frac{1}{N - x}\bigg) \leq CV(x), \tag{3.3.4}$$

where $C = (\mu + \gamma) \vee (\beta N) + \sigma^2 N^2$. Substituting this into (3.3.2) we get

$$\mathbb{E}V(I(t \wedge \tau_k)) \leq V(I_0) + \mathbb{E}\int_0^{t\wedge\tau_k} CV(I(s))ds \leq V(I_0) + C\int_0^t \mathbb{E}V(I(s \wedge \tau_k))ds.$$

The Gronwall inequality (Theorem 2.10) yields that

$$\mathbb{E}V(I(T \wedge \tau_k)) \leq V(I_0)e^{CT}. \tag{3.3.5}$$

Set $\Omega_k = \{\tau_k \leq T\}$ for $k \geq k_1$ and, by (3.3.1), $\mathbb{P}(\Omega_k) \geq \varepsilon$. Note that for every $\omega \in \Omega_k$, $I(\tau_k, \omega)$ equals either $1/k$ or $N - (1/k)$, and hence

$$V(I(\tau_k, \omega)) \geq k.$$

It then follows from (3.3.5) that

$$V(I_0)e^{CT} \geq \mathbb{E}\Big[I_{\Omega_k}(\omega)V(I(\tau_k, \omega))\Big] \geq k\,\mathbb{P}(\Omega_k) \geq \varepsilon k.$$

Letting $k \to \infty$ leads to the contradiction

$$\infty > V(I_0)e^{CT} = \infty,$$

so we must therefore have $\tau_\infty = \infty$ a.s., whence the proof is complete.

3.4 Extinction

In the study of population systems, extinction and persistence are two of the most important issues. We will discuss the extinction of the SDE SIS model (3.2.4) in this section but leave its persistence to the next section.

Theorem 3.2 *If*

$$R_0^S := R_0^D - \frac{\sigma^2 N^2}{2(\mu+\gamma)} = \frac{\beta N}{\mu+\gamma} - \frac{\sigma^2 N^2}{2(\mu+\gamma)} < 1 \quad \textit{and} \quad \sigma^2 \le \frac{\beta}{N}, \tag{3.4.1}$$

then for any given initial value $I(0) = I_0 \in (0, N)$, the solution of the SDE (3.2.4) obeys

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \le \beta N - \mu - \gamma - 0.5\sigma^2 N^2 < 0 \quad a.s., \tag{3.4.2}$$

namely, $I(t)$ tends to zero exponentially almost surely. In other words, the disease dies out with probability one.

Proof. By the Itô formula in Theorem 2.9, we have

$$\log(I(t)) = \log(I_0) + \int_0^t f(I(s))ds + \int_0^t \sigma(N - I(s))dB(s), \tag{3.4.3}$$

where $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \beta N - \mu - \gamma - \beta x - 0.5\sigma^2 (N - x)^2. \tag{3.4.4}$$

However, under condition (3.4.1), we have

$$\begin{aligned} f(I(s)) &= \beta N - \mu - \gamma - 0.5\sigma^2 N^2 - (\beta - \sigma^2 N) I(s) - 0.5\sigma^2 I^2(s), \\ &\le \beta N - \mu - \gamma - 0.5\sigma^2 N^2, \end{aligned}$$

for $I(s) \in (0, N)$. It then follows from (3.4.3) that

$$\log(I(t)) \le \log(I_0) + (\beta N - \mu - \gamma - 0.5\sigma^2 N^2)t + \int_0^t \sigma(N - I(s))dB(s). \tag{3.4.5}$$

This implies that

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \le \beta N - \mu - \gamma - 0.5\sigma^2 N^2 + \limsup_{t\to\infty} \frac{1}{t} \int_0^t \sigma(N - I(s))dB(s) \quad a.s. \tag{3.4.6}$$

But by the large number theorem for martingales (see Theorem 2.1), we have

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t \sigma(N - I(s))dB(s) = 0 \quad a.s.$$

We therefore obtain the desired assertion (3.4.2) from (3.4.6).

It is useful to observe that in the classical deterministic SIS model (1.2.1), $I(t)$ tends to 0 if and only if $R_0^D \le 1$; while in the SDE SIS model (3.2.3), $I(t)$ tends to 0 if $R_0^S = R_0^D - 0.5\sigma^2 N^2/(\mu+\gamma) < 1$ and $\sigma^2 \le \beta/N$. In other words, the conditions for $I(t)$ to become extinct in the SDE SIS model are weaker than in the classical deterministic SIS model. The following example illustrates this result more explicitly:

Example 3.3 Throughout the thesis we shall assume that the unit of time is one day and the population sizes are measured in units of one million, unless otherwise stated. With these units assume that the system parameters are given by

$$\beta = 0.5,\ N = 100,\ \mu = 20,\ \gamma = 25,\ \sigma = 0.035.$$

So the SDE SIS model (3.2.4) becomes

$$dI(t) = I(t)\Big([5 - 0.5I(t)]dt + 0.035(100 - I(t))dB(t)\Big). \tag{3.4.7}$$

Noting that

$$\begin{aligned} R_0^S &= \frac{\beta N}{\mu+\gamma} - \frac{\sigma^2 N^2}{2(\mu+\gamma)} = \frac{50}{45} - \frac{12.25}{90} = 1.111 - 0.136 < 1, \\ \text{and} \quad \sigma^2 &= 0.001225 \leq \frac{\beta}{N} = 0.005, \end{aligned}$$

we can therefore conclude, by Theorem 3.2, that for any initial value $I(0) = I_0 \in (0, 100)$, the solution of (3.4.7) obeys

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \leq -1.125 \quad a.s.$$

That is, $I(t)$ will tend to zero exponentially with probability one.

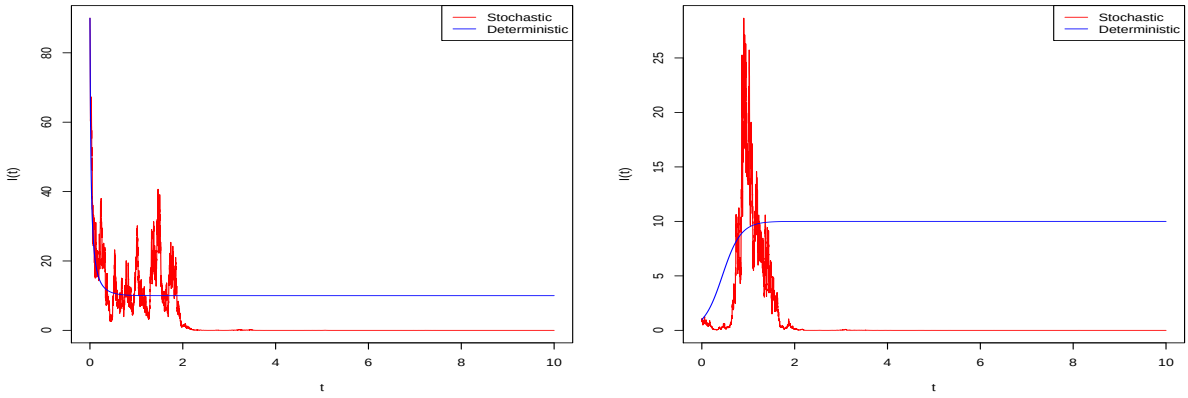
On the other hand, the corresponding deterministic SIS model (1.2.1) becomes

$$\frac{dI(t)}{dt} = I(t)(5 - 0.5I(t)). \tag{3.4.8}$$

For $R_0^D > 1$, it is known that, for any initial value $I(0) = I_0 \in (0, 100)$, this solution has the property

$$\lim_{t\to\infty} I(t) = N\left(1 - \frac{1}{R_0^D}\right) = 10 \quad \text{(section 1.2)}.$$

The computer simulations in Figure 3.1, using the EM method, support these results clearly, illustrating extinction of the disease.



(a) (b)

Figure 3.1: Computer simulation of the path $I(t)$ for the SDE SIS model (3.4.7) and its corresponding deterministic SIS model (3.4.8), using the EM method with step size $\Delta = 0.001$, using initial values (a) $I(0) = 90$ and (b) $I(0) = 1$.

In Theorem 3.2 we require the noise intensity $\sigma^2 \leq \beta/N$. The following theorem covers the case when $\sigma^2 > \beta/N$:

Theorem 3.4 *If*

$$\sigma^2 > \frac{\beta}{N} \vee \frac{\beta^2}{2(\mu+\gamma)}, \tag{3.4.9}$$

then for any given initial value $I(0) = I_0 \in (0, N)$, the solution of the SDE SIS model (3.2.4) obeys

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \leq -\mu - \gamma + \frac{\beta^2}{2\sigma^2} < 0 \quad a.s., \tag{3.4.10}$$

namely, $I(t)$ tends to zero exponentially almost surely. In other words, the disease dies out with probability one.

Proof. We use the same notation as in the proof of Theorem 3.2. It is easy to see that the quadratic function $f : \mathbb{R} \to \mathbb{R}$ defined by (3.4.4) takes its maximum value $f(\hat{x})$ at

$$x = \hat{x} := \frac{\sigma^2 N - \beta}{\sigma^2}.$$

By condition (3.4.9), it is easy to see that $\hat{x} \in (0, N)$. Compute

$$f(\hat{x}) = \beta N - \mu - \gamma - 0.5\sigma^2 N^2 + \frac{(\sigma^2 N - \beta)^2}{2\sigma^2} = -\mu - \gamma + \frac{\beta^2}{2\sigma^2},$$

which is negative by condition (3.4.9). It therefore follows from (3.4.3) that

$$\log(I(t)) \leq \log(I_0) + f(\hat{x})t + \int_0^t \sigma(N - I(s))dB(s).$$

This implies, in the same way as in the proof of Theorem 3.2, that

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le f(\hat{x}) \quad a.s.,$$

as required. The proof is hence complete.

Note that condition (3.4.9) implies that $R_0^S \le 1$.

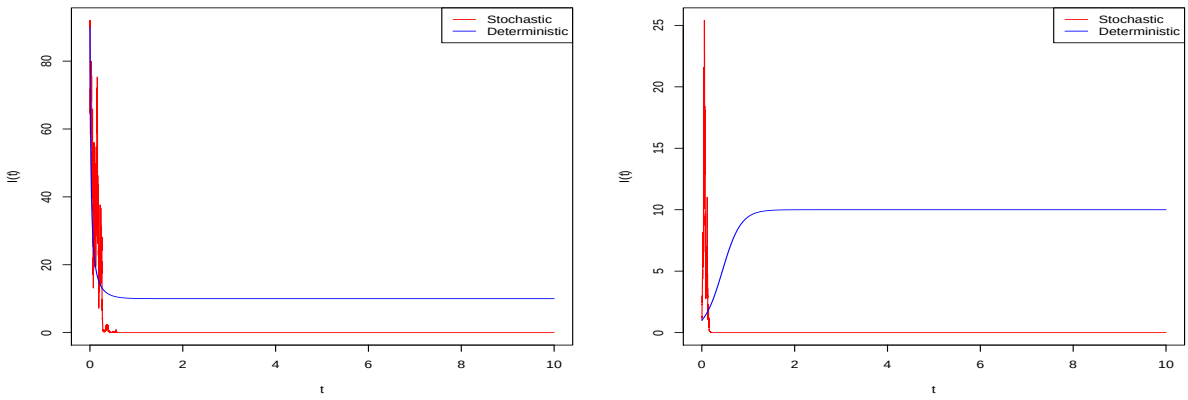
Example 3.5 We keep the system parameters the same as in Example 3.3 but let $\sigma = 0.08$, so the SDE SIS model (3.2.4) becomes

$$dI(t) = I(t)\Big([5 - 0.5I(t)]dt + 0.08(100 - I(t))dB(t)\Big). \tag{3.4.11}$$

It is easy to verify that the system parameters obey condition (3.4.9). We can therefore conclude, by Theorem 3.4, that for any initial value $I(0) = I_0 \in (0, 100)$, the solution of (3.4.11) obeys

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le -45 + \frac{0.5^2}{2\times 0.08^2} = -25.4688 \quad a.s.$$

That is, $I(t)$ will tend to zero exponentially with probability one. The computer simulations shown in Figure 3.2 support these results clearly.



(a) (b)

Figure 3.2: Computer simulation of the path $I(t)$ for the SDE SIS model (3.4.11) and its corresponding deterministic SIS model (3.4.8), using the EM method with step size $\Delta = 0.001$, with initial values (a) $I(0) = 90$ and (b) $I(0) = 1$.

3.5 Persistence

Theorem 3.6 *If*

$$R_0^S := \frac{\beta N}{\mu+\gamma} - \frac{\sigma^2 N^2}{2(\mu+\gamma)} > 1 \tag{3.5.1}$$

then for any given initial value $I(0) = I_0 \in (0, N)$, the solution of the SDE SIS model (3.2.4) obeys

$$\limsup_{t\to\infty} I(t) \geq \xi \quad a.s. \tag{3.5.2}$$

and

$$\liminf_{t\to\infty} I(t) \leq \xi \quad a.s. \tag{3.5.3}$$

where

$$\xi = \frac{1}{\sigma^2}\Big(\sqrt{\beta^2 - 2\sigma^2(\mu+\gamma)} - (\beta - \sigma^2 N)\Big) \tag{3.5.4}$$

which is the unique root in $(0, N)$ of

$$\beta N - \mu - \gamma - \beta\xi - 0.5\sigma^2(N-\xi)^2 = 0. \tag{3.5.5}$$

That is, $I(t)$ will rise to or above the level ξ infinitely often with probability one.

Proof. Recall the definition (3.4.4) of function $f : \mathbb{R} \to \mathbb{R}$. By condition (3.5.1), it is easy to see that equation $f(x) = 0$ has a positive root and a negative root. The positive one is

$$\begin{aligned}&\frac{1}{\sigma^2}\Big(\sqrt{(\beta - \sigma^2 N)^2 + 2\sigma^2(\beta N - \mu - \gamma - 0.5\sigma^2 N^2)} - (\beta - \sigma^2 N)\Big)\\ &\quad = \frac{1}{\sigma^2}\Big(\sqrt{\beta^2 - 2\sigma^2(\mu+\gamma)} - (\beta - \sigma^2 N)\Big) = \xi.\end{aligned}$$

Noting that

$$f(0) = \beta N - \mu - \gamma - 0.5\sigma^2 N^2 > 0 \quad \text{and} \quad f(N) = -\mu - \gamma < 0,$$

we see that $\xi \in (0, N)$ and

$$f(x) > 0 \quad \text{is strictly increasing on } x \in (0, 0 \vee \hat{x}), \tag{3.5.6}$$

$$f(x) > 0 \quad \text{is strictly decreasing on } x \in (0 \vee \hat{x}, \xi), \tag{3.5.7}$$

while

$$f(x) < 0 \quad \text{is strictly decreasing on } x \in (\xi, N). \tag{3.5.8}$$

We now begin to prove assertion (3.5.2). If it is not true, then there is a sufficiently small $\varepsilon \in (0, 1)$ such that

$$\mathbb{P}(\Omega_1) > \varepsilon, \tag{3.5.9}$$

where $\Omega_1 = \{\limsup_{t\to\infty} I(t) \leq \xi - 2\varepsilon\}$. Hence, for every $\omega \in \Omega_1$, there is a $T = T(\omega) > 0$ such that

$$I(t, \omega) \leq \xi - \varepsilon \quad \text{whenever } t \geq T(\omega). \tag{3.5.10}$$

Clearly we may choose ε so small (if necessary reduce it) that $f(0) > f(\xi - \varepsilon)$. It therefore follows from (3.5.6), (3.5.7) and (3.5.10) that

$$f(I(t, \omega)) \geq f(\xi - \varepsilon) \quad \text{whenever } t \geq T(\omega). \tag{3.5.11}$$

Moreover, by the large number theorem for martingales (Theorem 2.1), there is a $\Omega_2 \subset \Omega$ with $\mathbb{P}(\Omega_2) = 1$ such that for every $\omega \in \Omega_2$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \sigma(N - I(s,\omega))dB(s,\omega) = 0. \tag{3.5.12}$$

Now, fix any $\omega \in \Omega_1 \cap \Omega_2$. It then follows from (3.4.3) and (3.5.11) that, for $t \geq T(\omega)$,

$$\begin{aligned} \log(I(t,\omega)) &\geq \log(I_0) + \int_0^{T(\omega)} f(I(s,\omega))ds + f(\xi - \varepsilon)(t - T(\omega)) \\ &+ \int_0^t \sigma(N - I(s,\omega))dB(s,\omega). \end{aligned} \tag{3.5.13}$$

This yields

$$\liminf_{t\to\infty} \frac{1}{t} \log(I(t,\omega)) \geq f(\xi - \varepsilon) > 0,$$

whence

$$\lim_{t\to\infty} I(t,\omega) = \infty.$$

But this contradicts (3.5.10). We therefore must have the desired assertion (3.5.2).

Let us now prove assertion (3.5.3). If it were not true, then there is a sufficiently small $\delta \in (0,1)$ such that

$$\mathbb{P}(\Omega_3) > \delta, \tag{3.5.14}$$

where $\Omega_3 = \{\liminf_{t\to\infty} I(t) \geq \xi + 2\delta\}$. Hence, for every $\omega \in \Omega_3$, there is a $\tau = \tau(\omega) > 0$ such that

$$I(t,\omega) \geq \xi + \delta \quad \text{whenever } t \geq \tau(\omega). \tag{3.5.15}$$

Now, fix any $\omega \in \Omega_3 \cap \Omega_2$. It then follows from (3.4.3) and (3.5.8) that, for $t \geq \tau(\omega)$,

$$\begin{aligned} \log(I(t,\omega)) &\leq \log(I_0) + \int_0^{\tau(\omega)} f(I(s,\omega))ds + f(\xi + \delta)(t - \tau(\omega)) \\ &+ \int_0^t \sigma(N - I(s,\omega))dB(s,\omega). \end{aligned} \tag{3.5.16}$$

This, together with (3.5.12), yields

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t,\omega)) \leq f(\xi + \delta) < 0,$$

whence

$$\lim_{t\to\infty} I(t,\omega) = 0.$$

But this contradicts (3.5.15). We therefore must have the desired assertion (3.5.3). The proof is therefore complete.

Example 3.7 Assume that the system parameters are given by

$$\beta = 0.5,\ N = 100,\ \mu = 20,\ \gamma = 25,\ \sigma = 0.03.$$

That is, we keep all the system parameters the same as in Example 3.3 except that σ is reduced to 0.03 from 0.035. So the SDE SIS model (3.2.4) becomes

$$dI(t) = I(t)\Big([5 - 0.5I(t)]dt + 0.03(100 - I(t))dB(t)\Big). \tag{3.5.17}$$

Noting that

$$R_0^S = \frac{\beta N}{\mu + \gamma} - \frac{\sigma^2 N^2}{2(\mu + \gamma)} = \frac{50}{45} - 0.1 > 1,$$

we compute

$$\xi = \frac{1}{\sigma^2}\Big(\sqrt{\beta^2 - 2\sigma^2(\mu + \gamma)} - (\beta - \sigma^2 N)\Big) = 1.2179.$$

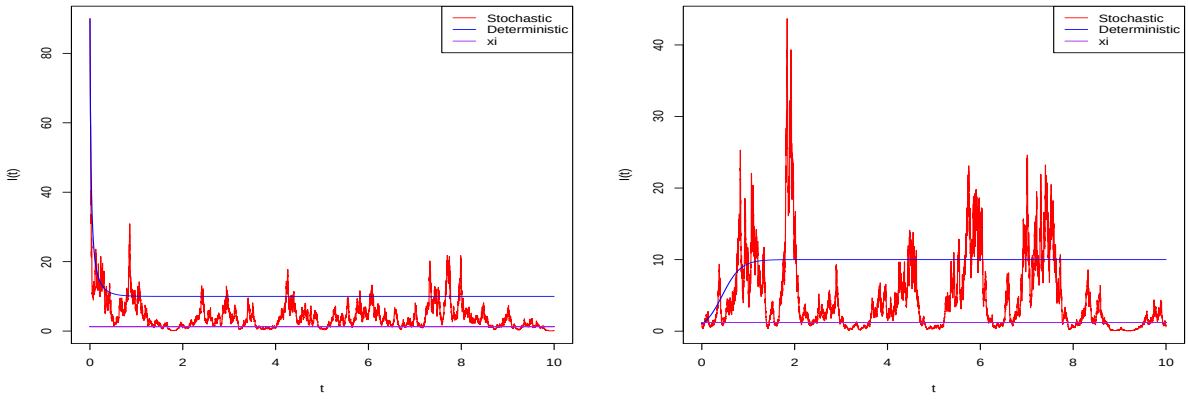
We can therefore conclude, by Theorem 3.6, that for any initial value $I(0) = I_0 \in (0, 100)$, the solution of (3.5.17) obeys

$$\liminf_{t\to\infty} I(t) \le 1.2179 \le \limsup_{t\to\infty} I(t) \quad a.s.$$

In comparison, we recall that the solution of the corresponding deterministic SIS model (1.2.2) has the property

$$\lim_{t\to\infty} I(t) = N\left(1 - \frac{1}{R_0^D}\right) = 10.$$

The computer simulations in Figure 3.3 support these results clearly, showing fluctuation around the level 1.2179.



(a) (b)

Figure 3.3: Computer simulation of the path $I(t)$ for the SDE SIS model (3.5.17) and its corresponding deterministic SIS model (3.4.8), using the EM method with step size $\Delta = 0.001$ and initial values (a) $I(0) = 90$ and (b) $I(0) = 1$.

Example 3.8 To further illustrate the effect of the noise intensity σ on the SDE SIS model, we keep all the parameters in Example 3.7 unchanged but reduce σ to $\sigma = 0.01$, namely we have

$$\beta = 0.5,\ N = 100,\ \mu = 20,\ \gamma = 25,\ \sigma = 0.01.$$

So the SDE SIS model (3.2.4) now becomes

$$dI(t) = I(t)\Big([5 - 0.5I(t)]dt + 0.01(100 - I(t))dB(t)\Big). \tag{3.5.18}$$

Noting that

$$R_0^S = \frac{\beta N}{\mu + \gamma} - \frac{\sigma^2 N^2}{2(\mu + \gamma)} = \frac{50}{45} - 0.011 > 1,$$

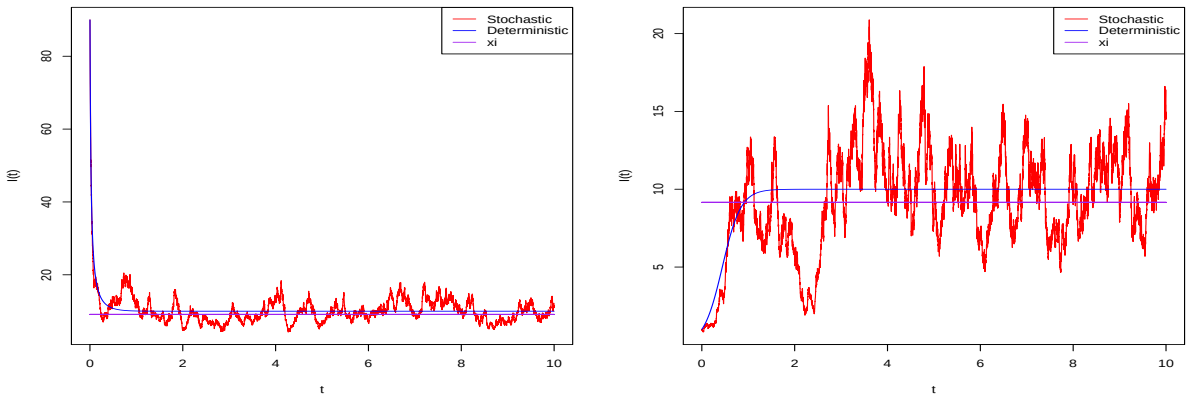
we compute

$$\xi = \frac{1}{\sigma^2}\Big(\sqrt{\beta^2 - 2\sigma^2(\mu + \gamma)} - (\beta - \sigma^2 N)\Big) = 9.1751.$$

We can therefore conclude, by Theorem 3.6, that for any initial value $I(0) = I_0 \in (0, 100)$, the solution of (3.5.18) obeys

$$\liminf_{t\to\infty} I(t) \le 9.1751 \le \limsup_{t\to\infty} I(t) \quad a.s.$$

The computer simulations in Figure 3.4 support these results clearly, illustrating persistence and that the effect of reducing the standard deviation σ is to increase the level ξ, which becomes closer to the limiting value of the corresponding deterministic SIS model.



(a) (b)

Figure 3.4: Computer simulation of the path $I(t)$ for the SDE SIS model (3.5.18) and its corresponding deterministic SIS model (3.4.8), using the EM method with step size $\Delta = 0.001$ and initial values (a) $I(0) = 90$ and (b) $I(0) = 1$.

These computer simulations indicate strongly that ξ will increase to $N(1 - (1/R_0^D))$, which is the equilibrium state of the deterministic SIS model (1.2.1), as the noise intensity σ decreases to zero. This is of course not surprising. The following proposition describes this situation rigorously:

Proposition 3.9 *Assume that $R_0^S > 1$ and regard ξ defined by (3.5.4) as a function of σ for*

$$0 < \sigma < \frac{\sqrt{2(\beta N - \mu - \gamma)}}{N} := \hat{\sigma}.$$

Then ξ is strictly decreasing and

$$\lim_{\sigma\to 0}\xi = N\Big(1-\frac{1}{R_0^D}\Big) \quad \textit{and} \quad \lim_{\sigma\to\hat{\sigma}}\xi = \begin{cases} 0, & \textit{if } \ 1\le R_0^D\le 2,\\ N\Big(\frac{R_0^D-2}{R_0^D-1}\Big), & \textit{if } \ R_0^D>2. \end{cases}$$

Proof. Compute

$$\frac{d\xi}{d\sigma} = \frac{1}{2}\Big(\frac{\beta^2}{\sigma^4}-\frac{2(\mu+\gamma)}{\sigma^2}\Big)^{-\frac{1}{2}}\Big(-\frac{4\beta^2}{\sigma^5}+\frac{4(\mu+\gamma)}{\sigma^3}\Big)+\frac{2\beta}{\sigma^3},$$

$$= \frac{-2\beta^2+2\sigma^2(\mu+\gamma)+2\beta\sqrt{\beta^2-2\sigma^2(\mu+\gamma)}}{\sigma^3\sqrt{\beta^2-2\sigma^2(\mu+\gamma)}},$$

$$= \frac{-(\sqrt{\beta^2-2\sigma^2(\mu+\gamma)}-\beta)^2}{\sigma^3\sqrt{\beta^2-2\sigma^2(\mu+\gamma)}}.$$

Since $\sigma>0$ we have $\sqrt{\beta^2-2\sigma^2(\mu+\gamma)}-\beta\neq 0$. We therefore have that $\frac{d\xi}{d\sigma}<0$ which implies that ξ is strictly decreasing as σ increases. Moreover, by the well-known L'Hopital's rule,

$$\lim_{\sigma\to 0}\xi = \lim_{\sigma\to 0}[-(\beta^2-2\sigma^2(\mu+\gamma))^{-\frac{1}{2}}(\mu+\gamma)+N] = -\frac{\mu+\gamma}{\beta}+N = N\Big(1-\frac{1}{R_0^D}\Big)$$

as desired. Furthermore, it is obvious that

$$\lim_{\sigma\to\hat{\sigma}}\xi = \frac{\sqrt{\beta^2-2\hat{\sigma}^2(\mu+\gamma)}-\beta+\hat{\sigma}^2N}{\hat{\sigma}^2}.$$

The numerator equals

$$\frac{|\beta N-2(\mu+\gamma)|}{N}+\frac{\beta N-2(\mu+\gamma)}{N},$$

so if $1\le R_0^D\le 2$ we have $\lim_{\sigma\to\hat{\sigma}}\xi=0$, but if $R_0^D>2$ we have $\lim_{\sigma\to\hat{\sigma}}\xi = N\Big(\frac{R_0^D-2}{R_0^D-1}\Big)$. The proof is complete.

Note that Proposition 3.9 implies that for $R_0^S>1$, ξ lies between the deterministic equilibrium value (and limiting value)

$$N\Big(1-\frac{1}{R_0^D}\Big)$$

for $I(t)$ and

$$\max\Big(0, N\Big(1-\frac{1}{R_0^D-1}\Big)\Big).$$

If R_0^D is large then ξ will be close to but beneath the deterministic equilibrium value for $I(t)$.

Example 3.10 The computer simulations of the solution to the SDE SIS model in the persistent case also suggest for higher σ that the distribution of the solution is skewed, as there are larger oscillations above ξ than below ξ, while for lower σ the oscillations about ξ appear to be more symmetrically distributed. This is confirmed by the histograms in Figure 3.5, showing the distribution of $I(t)$ in the case of $\beta = 0.5, N = 100, \mu = 20, \gamma = 25$, and σ= 0.03, 0.02, 0.01, 0.005 and 0.001, respectively. The simulations were run for 100,000 iterations with step size $\Delta = 0.001$, i.e. for 100 time steps, and the first 90,000 iterations were discarded to allow for $I(t)$ to reach its recurrent level. The distribution is positively skewed for $\sigma = 0.03$ and 0.02, but as σ reduces it becomes more symmetric about ξ, so that the distribution appears closer to a normal distribution. The corresponding sample skewness coefficients are $4.8774, 0.9319, 0.1692, 0.1798$, and -0.2106.

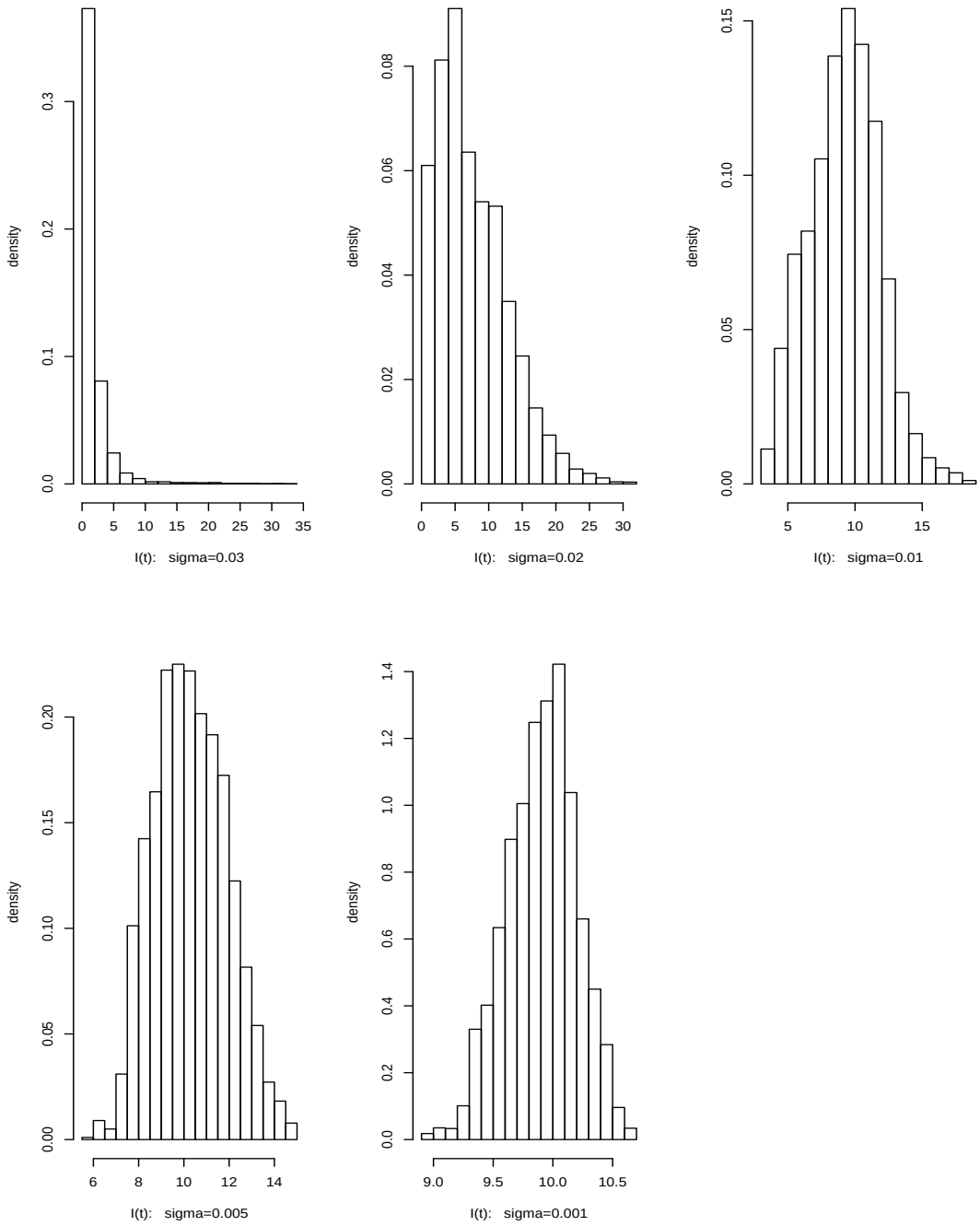


Figure 3.5: Histograms of the values of the path $I(t)$ for the recurrent SDE SIS model (3.2.4), for parameter values $\beta = 0.5, N = 100, \mu = 20, \gamma = 25$, $I(0) = 90$, and differing values of σ= 0.03, 0.02, 0.01, 0.005 and 0.001. The values are for the last 10,000 iterations of 100,000 iterations using step size $\Delta = 0.001$.

Testing these data for normality, all tests used were highly significant, conclusively

rejecting normality in all cases. This is not surprising in view of the very large sample sizes (10,000), as even moderate deviations from the tested distribution will be significant, however the normal QQ plots in Figure 3.6 suggest that these data are not far from being normally distributed for smaller values of σ.

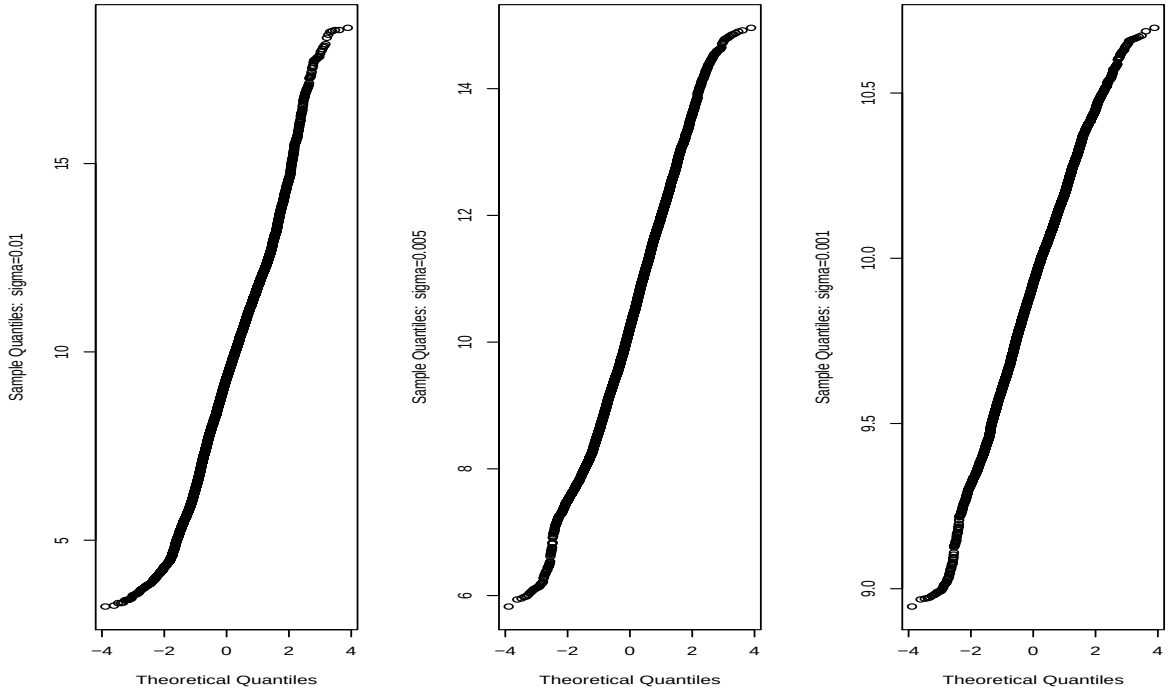


Figure 3.6: Normal quantile-quantile plots of the values of the path $I(t)$ for the recurrent SDE SIS model (3.2.4), for parameter values $\beta = 0.5, N = 100, \mu = 20, \gamma = 25$, and differing values of σ=0.01, 0.005 and 0.001, corresponding to the last three histograms in Figure 3.5.

3.6 Stationary Distribution

In the previous section we showed that $I(t)$ will fluctuate around the level $\xi \in (0, N)$ with probability 1 when $R_0^S > 1$. The computer simulations also strongly indicate that the SDE SIS model (3.2.4) has a stationary distribution. To be more precise, let $P_{I_0,t}(\cdot)$ denote the probability measure induced by $I(t)$ with initial value $I(0) = I_0$, that is

$$P_{I_0,t}(A) = \mathbb{P}(I(t) \in A), \quad A \in \mathcal{B}(0, N),$$

where $\mathcal{B}(0, N)$ is the σ-algebra of all the Borel sets $A \subset (0, N)$. If there is a probability measure $P_\infty(\cdot)$ on the measurable space $((0, N), \mathcal{B}(0, N))$ such that

$$P_{I_0,t}(\cdot) \to P_\infty(\cdot) \text{ in distribution for any } I_0 \in (0, N) \text{ as } t \to \infty,$$

we then say that the SDE (3.2.4) has a stationary distribution $P_\infty(\cdot)$ (see e.g. [44, 69]). To show the existence of a stationary distribution, let us first cite a known result from Has'minskii [44, pp.118–123] as a lemma.

Lemma 3.11 *The SDE SIS model (3.2.4) has a unique stationary distribution if there is a strictly proper sub-interval (a, b) of $(0, N)$ such that $\mathbb{E}(\tau) < \infty$ for all $I_0 \in (0, a] \cup [b, N)$, where*

$$\tau = \inf\{t \geq 0 : I(t) \in (a, b)\},$$

and, moreover,

$$\sup_{I_0 \in [\bar{a}, \bar{b}]} \mathbb{E}(\tau) < \infty \quad \textit{for every interval } [\bar{a}, \bar{b}] \subset (0, N).$$

It should be pointed out that in the original Has'minskii theorem, there is one more condition which states that the square of the diffusion coefficient of the SDE (3.2.4), namely $\sigma^2 I^2 (N - I)^2$, is bounded away from zero for $I \in (a, b)$. But this is obvious for the SDE, hence there is no point in stating this condition.

Theorem 3.12 *If $R_0^S > 1$, then the SDE SIS model (3.2.4) has a unique stationary distribution.*

Proof. We will use the same notation as used in the proofs of Theorems 3.2 and 3.6. Fix any $0 < a < \xi < b < N$. We observe from (3.5.6)–(3.5.8) that

$$f(x) \geq f(0) \wedge f(a) > 0 \text{ if } 0 < x \leq a \quad \text{and} \quad f(x) \leq f(b) < 0 \text{ if } b \leq x < N. \tag{3.6.1}$$

Define τ as in Lemma 3.11. For any $I_0 \in (0, a)$, it then follows from (3.4.3) and (3.6.1) that

$$\log(a) \geq \mathbb{E}(\log(I(\tau \wedge t)) \geq \log(I_0) + (f(0) \wedge f(a))\mathbb{E}(\tau \wedge t), \quad \forall t \geq 0.$$

Letting $t \to \infty$ yields

$$\mathbb{E}(\tau) \leq \frac{\log(a/I_0)}{f(0) \wedge f(a)}, \quad \forall I_0 \in (0, a). \tag{3.6.2}$$

Similarly, for any $I_0 \in (b, N)$,

$$\log(b) \leq \mathbb{E}(\log(I(\tau \wedge t)) \leq \log(I_0) - |f(b)|\mathbb{E}(\tau \wedge t), \quad \forall t \geq 0.$$

Letting $t \to \infty$ yields

$$\mathbb{E}(\tau) \leq \frac{\log(N/b)}{|f(b)|}, \quad \forall I_0 \in (b, N). \tag{3.6.3}$$

The conditions in Lemma 3.11 follow clearly from (3.6.2) and (3.6.3). Hence the SDE SIS model (3.2.4) has a unique stationary distribution. The proof is complete.

The following theorem gives the mean and variance of the stationary distribution. Such explicit formulae are particularly useful in the test of computer simulations.

Theorem 3.13 *Assume that $R_0^S > 1$. Let m and v denote the mean and variance of the stationary distribution of the SDE SIS model (3.2.4). Then*

$$m = \frac{2\beta(R_0^S - 1)(\mu + \gamma)}{2\beta(\beta - \sigma^2 N) + \sigma^2(\beta N - \mu - \gamma)} \tag{3.6.4}$$

and

$$v = \frac{m(\beta N - \mu - \gamma)}{\beta} - m^2. \tag{3.6.5}$$

Proof. Fix any $I_0 \in (0, N)$. It follows from (3.2.4) that

$$I(t) = I_0 + \int_0^t I(s)[\beta N - \mu - \gamma - \beta I(s)]ds + \int_0^t \sigma(N - I(s))dB(s).$$

Dividing both sides by t, letting $t \to \infty$ and applying the ergodic property of the stationary distribution (see e.g. [44, 62]) and the large number theorem for martingales (Theorem 2.1), we obtain

$$0 = (\beta N - \mu - \gamma)m - \beta m_2, \tag{3.6.6}$$

where m_2 denotes the second moment of the stationary distribution. Similarly, dividing both sides of (3.4.3) by t and letting $t \to \infty$ we get

$$\lim_{t\to\infty} \frac{\log(I(t))}{t} = \beta N - \mu - \gamma - 0.5\sigma^2 N^2 - (\beta - \sigma^2 N)m - 0.5\sigma^2 m_2 \quad a.s. \tag{3.6.7}$$

This, together with Theorem 3.6, implies that

$$\lim_{t\to\infty} \frac{\log(I(t))}{t} = 0 \quad a.s.$$

Writing $\beta N - \mu - \gamma - 0.5\sigma^2 N^2 = (R_0^S - 1)(\mu + \gamma)$, we then have

$$0 = (R_0^S - 1)(\mu + \gamma) - (\beta - \sigma^2 N)m - 0.5\sigma^2 m_2. \tag{3.6.8}$$

Substituting (3.6.6) into (3.6.8) yields

$$0 = (R_0^S - 1)(\mu + \gamma) - (\beta - \sigma^2 N)m - \frac{\sigma^2(\beta N - \mu - \gamma)m}{2\beta}.$$

This implies assertion (3.6.4). Moreover, it follows from (3.6.6) that

$$m_2 = \frac{m(\beta N - \mu - \gamma)}{\beta}.$$

Hence

$$v = m_2 - m^2 = \frac{m(\beta N - \mu - \gamma)}{\beta} - m^2,$$

which is the other assertion (3.6.5). The proof is therefore complete.

Example 3.14 We now use the same parameter values $\beta = 0.5, N = 100, \mu = 20$ and $\gamma = 25$ for both $\sigma = 0.02$ and 0.001 and show the results of running 1,000 simulations of the path $I(t)$ for the recurrent SDE SIS model, for a longer run of 200,000 iterations with step size $\Delta = 0.001$, but storing only the last of these $I(t)$ values in each case. Figure 6.1 shows the histogram of the last 10,000 samples from a single run of 200,000 iterations, beside the histogram of the last $I(t)$ values from each of the 1,000 simulations, and also the corresponding two empirical cumulative distribution functions (ecdfs), for comparison, for both values of σ. In each of the two cases, the corresponding histograms are similar and the ecdfs are close to each other.

The similarity of these distributions in each case of σ may be taken as an illustration of the existence of the stationary distribution of $I(t)$, and that in the simulations the probability distribution of $I(t)$ has more or less reached this stationary distribution. From (3.6.4) and (3.6.5), for these parameter values the mean and variance of the stationary distribution are 6.493506 and 22.76944 respectively when $\sigma = 0.02$, compared to the two sample means of 7.306251 and 6.557697 and the corresponding unbiased sample variances of 19.50794 and 23.18525, for the first and second histograms respectively. For $\sigma = 0.001$, the mean and variance of the stationary distribution are 9.991898 and 0.08094976, compared to the two sample means of 10.01330 and 9.98392 and unbiased sample variances of 0.06149254 and 0.07802619 respectively for the lower two histograms in Figure 3.7.

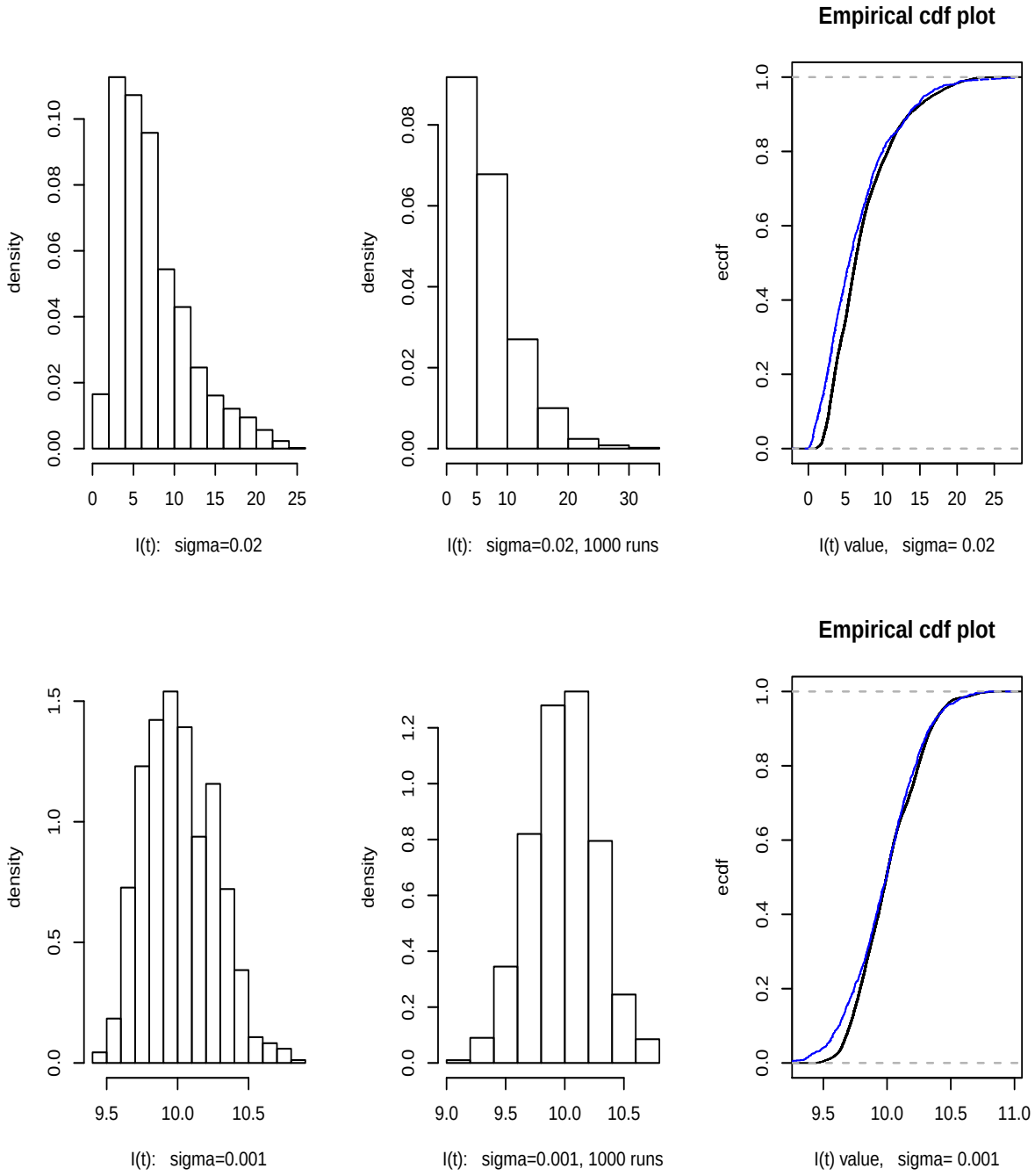


Figure 3.7: Histograms of the values of the path $I(t)$ for the recurrent SDE SIS model (3.2.4) for the last 10,000 samples of a single run of 200,000 iterations (left plot in each row) and also for the last iteration from each of 1,000 such runs (middle plot in each row), and the empirical cumulative distribution plot of each of these (right plot in each row; the black line corresponds to the first histogram and the blue line to the second one), for parameter values $\beta = 0.5, N = 100, \mu = 20, \gamma = 25$, and $\sigma = 0.02$ (top row) and $\sigma = 0.001$ (bottom row).

3.7 Two More Realistic Examples

As slightly more realistic examples to illustrate our theory, we suggest two SIS epidemic models with parameters estimated from actual disease situations. In this section the unit of time is still one day, but the population values are not scaled as previously:

Model A Gonorrhea amongst homosexuals [47].

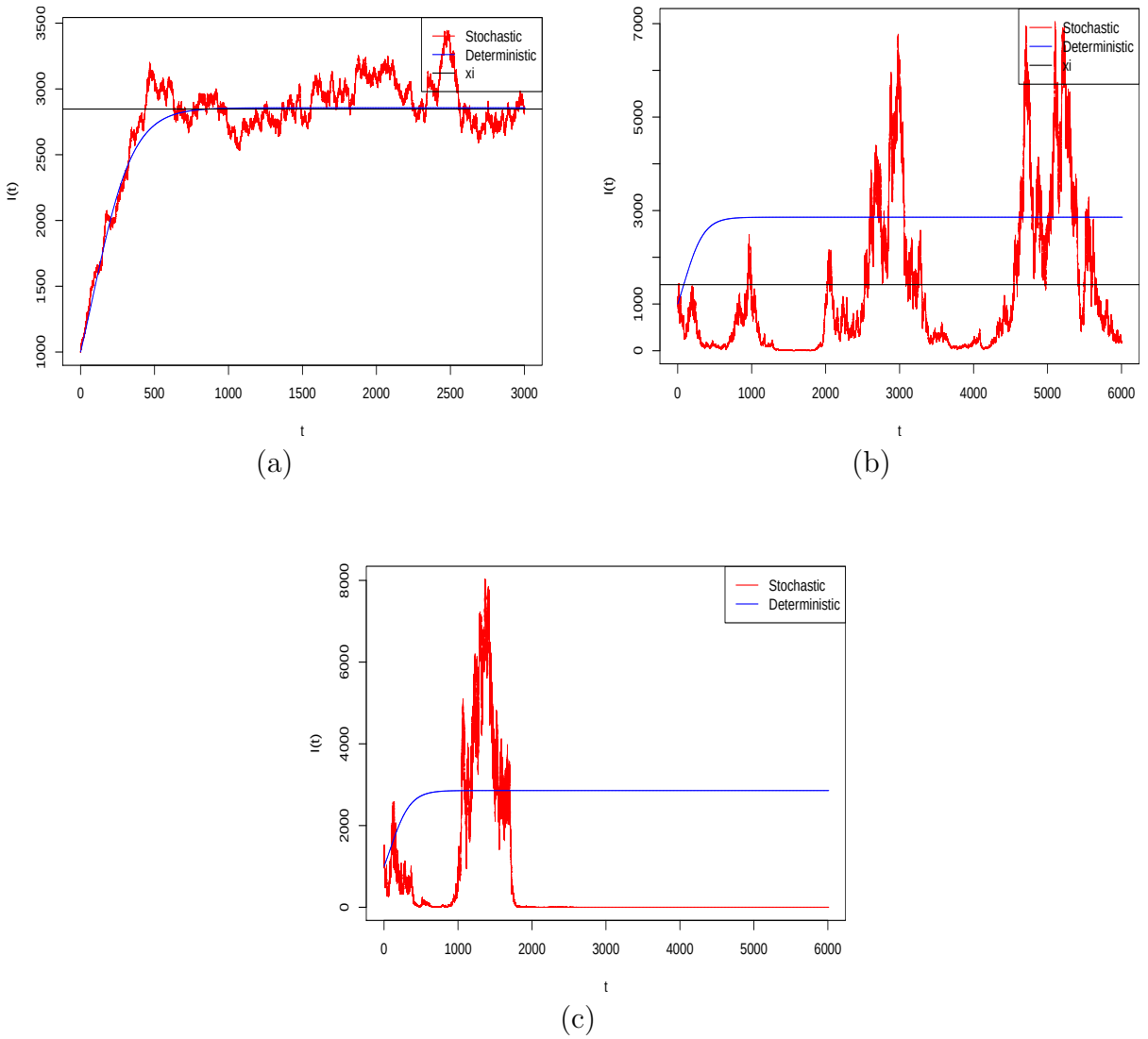


Figure 3.8: Computer simulation of the path $I(t)$ for Model A using the EM method with step size $\Delta = 0.001$ and initial value $I(0) = 1,000$. The deterministic case means $\sigma = 0$, while for the stochastic case we use (a) $\sigma = 10^{-6}$, (b) $\sigma = 10^{-5}$, (c) $\sigma = 1.5 \times 10^{-5}$.

In this model, the parameters are given by $N = 10,000, R_0^D = 1.4, \mu = (1/(40 \times 365.25))/\text{day} = 6.84463 \times 10^{-5}/\text{day}$ (average sexually active lifetime), $\gamma = (1/55)/\text{day}$

= 0.018182/day (based on Yorke, Hethcote and Nold [94]). Benenson [11] says that the infectious period is several months, but is not more precise, while Hethcote and Yorke ([47], Table 5.1) take $1/\gamma$ = 20-40 days for men and 80-160 days for women), and $\beta = (\mu + \gamma)R_0^D/N = 2.55504 \times 10^{-6}$/day.

Note that $R_0^D = 1.4$. Hence for the corresponding deterministic SIS model (1.2.1), we have

$$\lim_{t\to\infty} I(t) = 2,857$$

for any initial value $I_0 \in (0, 10,000)$. It is also easy to compute, for the SDE SIS model (3.2.4),

$$R_0^S = 1.4 - 2.739659 \times 10^9 \sigma^2.$$

To see the effect of the noise intensity, we consider three different values of σ: 10^{-6}, 10^{-5} and 1.5×10^{-5}. The corresponding values of R_0^S are 1.397, 1.126 and 0.784, respectively. By Theorem 3.6 we see that the SDE SIS model is persistent in the first two cases. However, in the last case, verifying $\sigma^2 = 2.25 \times 10^{-10} < \beta/N = 2.55504 \times 10^{-10}$ we conclude by Theorem 3.2 that the SDE SIS model is extinctive. The computer simulations shown in Figure 3.8 support these results clearly. Figure 3.9 shows the level ξ and the value of the mean m in (3.6.4) as a function of σ in the range given by Proposition 3.9.

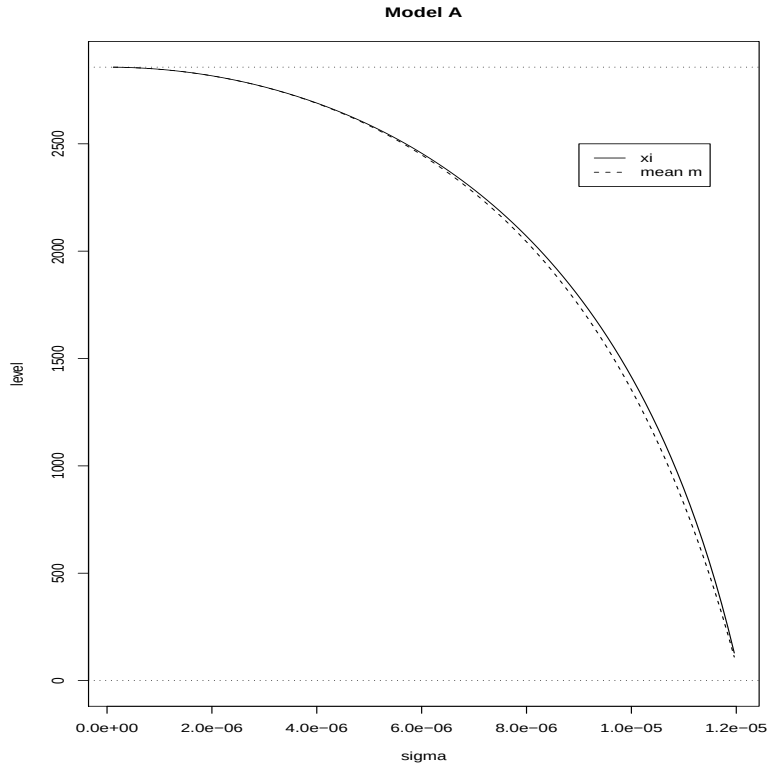


Figure 3.9: Plot of level of ξ (solid curve) and the mean m in (3.6.4) (dotted curve) against the value of σ in the range given in Proposition 3.9, for Model A. The horizontal dotted lines show the levels 0 and $N\left(1 - \frac{1}{R_0^D}\right)$ as limiting values for ξ.

Figure 3.10 shows histograms of the approximate stationary distribution of the two persistent cases (a) and (b) in Figure 3.8, resulting from the last two million iterations (last 2,000 days) for case (a) and the last four million iterations (4,000 days) for case (b). Both appear skewed to the right. The sample mean and unbiased sample variance are 2,901.894 and 31,045.18 respectively for case (a), compared to the theoretical mean and variance of the stationary distribution of 2,847.062 and 28,522.01 from (3.6.4) and (3.6.5). For case (b) the sample mean and unbiased sample variance are 1,051.77 and 2,125,489, compared to 1,354.592 and 2,035,258 from (3.6.4) and (3.6.5).

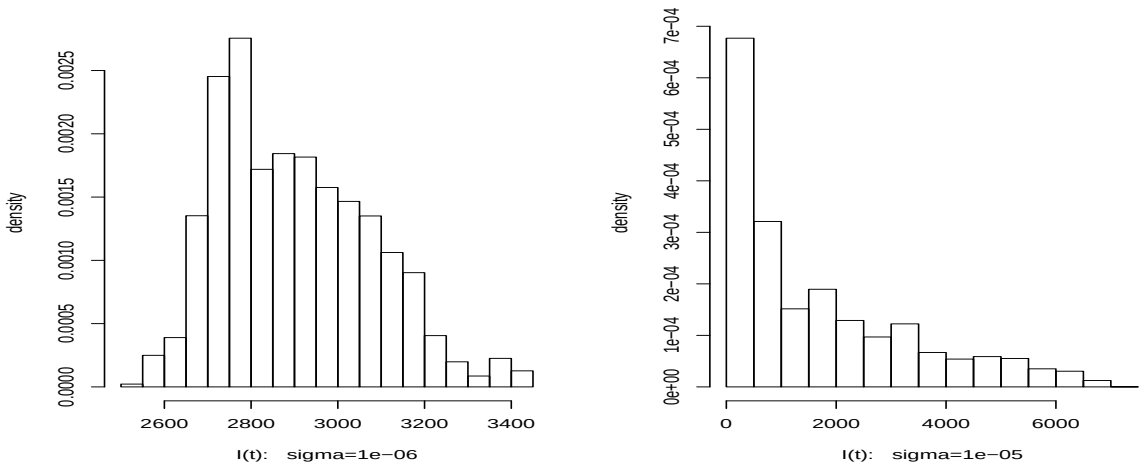


Figure 3.10: Histograms of the values of the path $I(t)$ for Model A for Figure 3.8(a) and 3.8(b), using the last two million iterations (2,000 days) for case (a) and the last four million iterations (4,000 days) for case (b).

Model B Pneumococcus amongst children under 2 years in Scotland (Lamb, Greenhalgh and Robertson [57]).

In this model, the parameters are given by $N = 150{,}000$, $\gamma = 1/(7.1 \text{ wk}) = 0.02011/\text{day}$ (Weir [93]), $\mu = 1/(104 \text{ wk}) = 1.3736 \times 10^{-3}/\text{day}$, and $\beta = 2.0055 \times 10^{-6}/\text{wk} = 2.8650 \times 10^{-7}/\text{day}$ (Zhang et al. [96]). (Farrington [27] has $R_0^D = 1.5$, which gives $\beta = 2.1486 \times 10^{-7}/\text{day}$.)

It is easy to compute $R_0^D = 2$. Hence for the corresponding deterministic SIS model (1.2.1), we have

$$\lim_{t\to\infty} I(t) = 75,000$$

for any initial value $I_0 \in (0, 150, 000)$. It is also easy to compute, for the SDE SIS model (3.2.4),

$$R_0^S = 2 - 5.23655 \times 10^{11}\sigma^2.$$

To see the effect of the noise intensity, we consider three different values of σ: 10^{-6}, 1.3×10^{-6} and 1.5×10^{-6}. The corresponding values of R_0^S are 1.476, 1.115 and 0.822, respectively. By Theorem 3.6 we see that the SDE SIS model is persistent in the first two

cases. However, in the last case, verifying $\sigma^2 = 2.25 \times 10^{-12} > (\beta/N) \vee (\beta^2/2(\mu+\gamma)) = 1.910347 \times 10^{-12}$ we conclude by Theorem 3.4 that the SDE SIS model is extinctive. The computer simulations shown in Figure 3.11 support these results clearly. In Figure 3.11(c) the deterministic simulation goes off the scale but is the same as in the other two simulations. Figure 3.12 shows the level ξ and the value of the mean m in (3.6.4) as a function of σ in the range given by Proposition 3.9.

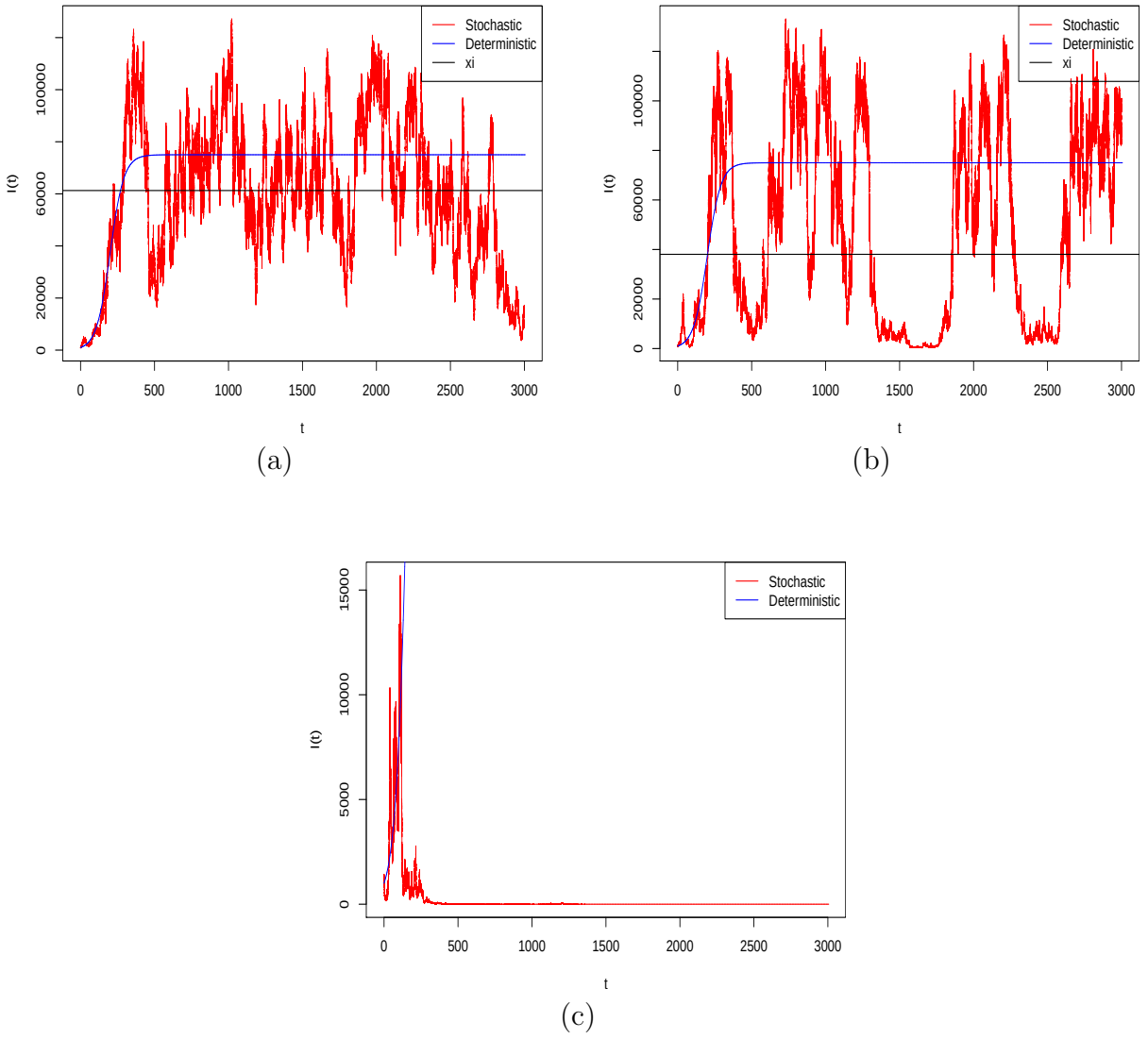


Figure 3.11: Computer simulation of the path $I(t)$ for Model B using the EM method with step size $\Delta = 0.001$ and initial value $I(0) = 50,000$. The deterministic case means $\sigma = 0$, while for the stochastic case we use (a) $\sigma = 10^{-6}$, (b) $\sigma = 1.3 \times 10^{-6}$, (c) $\sigma = 1.5 \times 10^{-6}$.

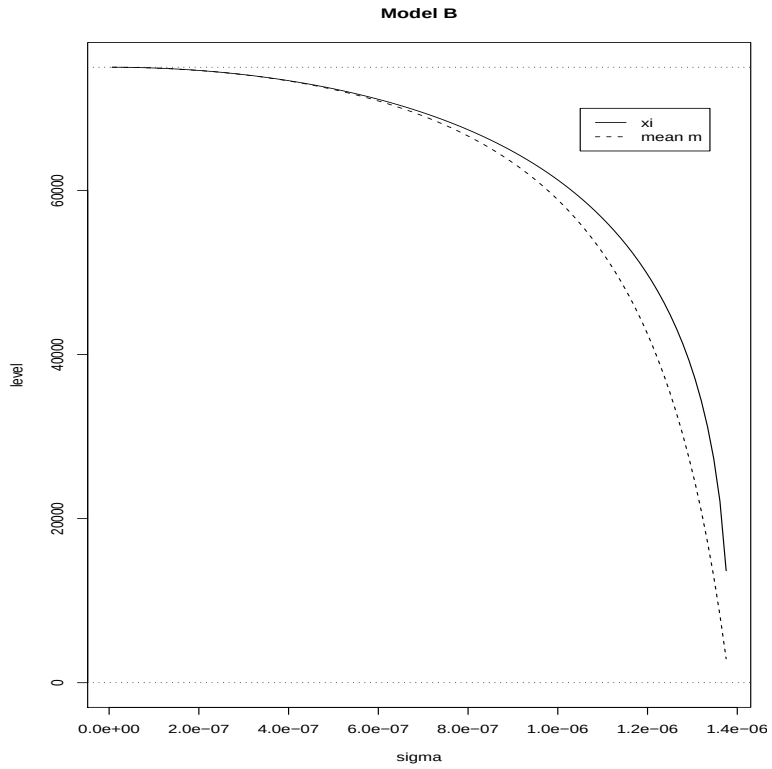


Figure 3.12: Plot of level of ξ (solid curve) and the mean m in (3.6.4) (dotted curve) against the value of σ in the range given in Proposition 3.9, for Model B. The horizontal dotted lines show the levels 0 and $N\left(1-\frac{1}{R_0^D}\right)$ as limiting values for ξ.

Figure 3.13 shows histograms of the approximate stationary distribution of the two persistent cases (a) and (b) in Figure 3.11, resulting from the last two million iterations (the last 2,000 days). The first appears symmetric, the second positively skewed. The sample mean and unbiased sample variance are 61,448.49 and 647,526,916 respectively for case (a), compared to the theoretical mean and variance of the stationary distribution of 58,856.32 and 950,958,848 from (3.6.4) and (3.6.5). For case (b) the sample mean and unbiased sample variance are 45,633.17 and 1,494,311,814, compared to 25,718.33 and 1,267,792,373 from (3.6.4) and (3.6.5).

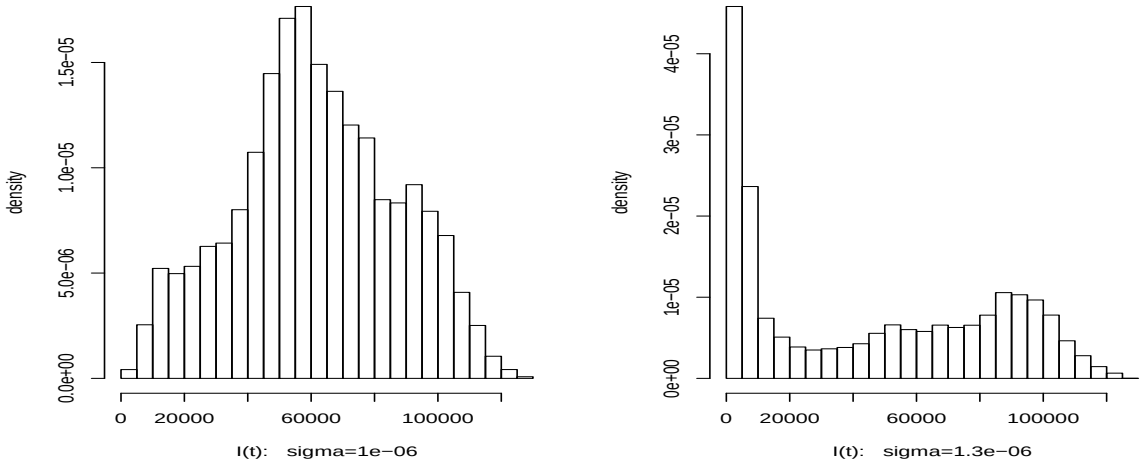


Figure 3.13: Histograms of the values of the path $I(t)$ for Model B for Figure 3.11(a) and 3.11(b), using the last two million iterations (2,000 days) in each case.

3.8 Discussion

Consider the stochastic SIS epidemic model in a neighbourhood of the DFE ($I = 0$). Then equation (3.2.4) becomes approximately

$$dI = [\beta N - (\mu + \gamma)]Idt + \sigma NIdB(t),$$

with solution

$$I(t) = I_0 \exp\left[\left(\beta N - (\mu + \gamma) - \frac{1}{2}\sigma^2 N^2\right)t + \sigma N B(t)\right].$$

Hence as $\lim_{t\to\infty} |B(t)|/t = 0$ [68] we expect that if

$$R_0^S = \frac{\beta N}{\mu + \gamma} - \frac{\sigma^2 N^2}{2(\mu + \gamma)} < 1,$$

then the approximate solution will die out, but if $R_0^S > 1$ then the approximate solution will diverge from the DFE. Thus in this sense R_0^S is the natural interpretation of R_0 in the SDE SIS model (3.2.4), although it is negative unless $\sigma^2 < 2\beta/N$.

This is almost what we have shown. Theorems 3.2 and 3.4 show that if either

$$(i)\ R_0^S < 1 \quad \text{and} \quad \sigma^2 \leq \frac{\beta}{N} \qquad \text{or} \qquad (ii)\ \sigma^2 > \frac{\beta}{N} \vee \frac{\beta^2}{2(\mu + \gamma)},$$

the disease will die out, whereas Theorem 3.6 shows that if $R_0^S > 1$ then the disease will persist. It is natural to make the following conjecture:

Conjecture 3.15 *If*

$$R_0^S < 1 \quad \textit{and} \quad \frac{\beta^2}{2(\mu+\gamma)} \geq \sigma^2 > \frac{\beta}{N}, \tag{3.8.1}$$

then the disease will die out with probability one.

While we have not so far been able to prove this, Example 3.16 provides an illustration of it.

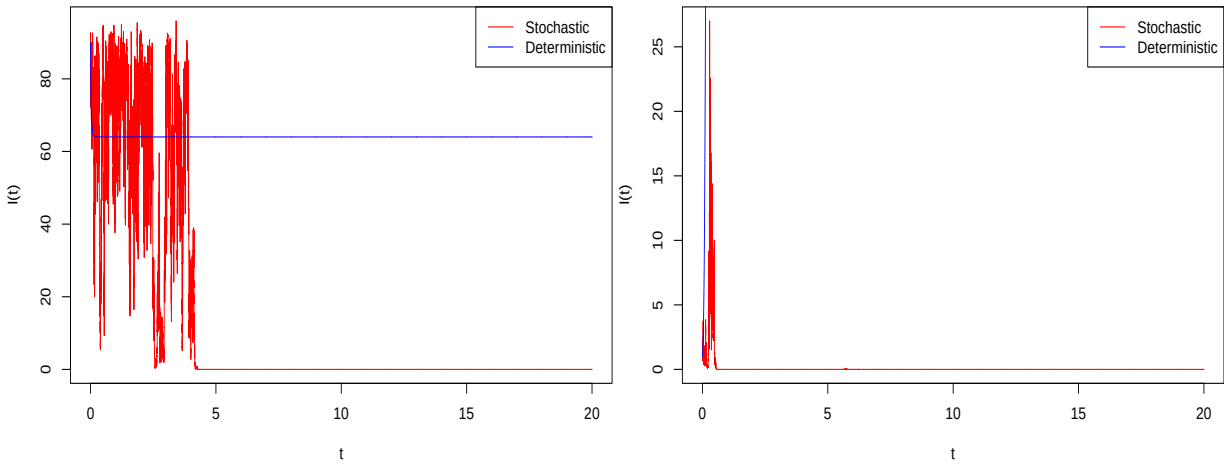
Example 3.16 We now use the system parameters

$$\beta = 0.5, \ N = 100, \ \mu = 10, \ \gamma = 8,$$

and now let $\sigma = 0.0825$, so that condition (3.8.1) is satisfied, and so the SDE SIS model (3.2.4) becomes

$$dI(t) = I(t)\Big([32 - 0.5I(t)]dt + 0.0825(100 - I(t))dB(t)\Big). \tag{3.8.2}$$

Figure 3.14 shows two simulations of the path $I(t)$, both becoming extinctive quickly. In Figure 3.14(b) the deterministic trajectory is as in Figure 3.14(a) although off the scale.



(a) (b)

Figure 3.14: Computer simulation of the path $I(t)$ for the SDE SIS model (3.8.2) and its corresponding deterministic SIS model (3.4.8), using the EM method with step size $\Delta = 0.001$, using initial values (a) $I(0) = 90$ and (b) $I(0) = 1$.

An alternative approach to including environmental stochasticity outlined by Allen [1] is to model the per capita disease transmission coefficient as a time dependent stochastic process $\overline{\beta}(t)$. The problem is discretised with a small timestep Δt so that $\overline{\beta}(n\Delta t)$ follows a random walk with state-dependent transition probabilities. These transition probabilities include both a diffusion term which causes the random walk to diverge and a mean-reverting term which drives the process back to a given mean value, say β_0. Under these assumptions the limiting process as $\Delta t \to 0$, $\overline{\beta}(t)$, follows an Ornstein-Uhlenbeck SDE

$$d\overline{\beta} = \gamma(\beta_0 - \overline{\beta})dt + \sigma dB(t).$$

This differential equation can be solved exactly and for large times the disease transmission coefficient is approximately normally distributed with mean β_0 and variance $\sigma^2/(2\gamma)$ [1].

In our method of including environmental stochasticity the total number of potentially infectious contacts between an infected individual and another individual in the infinitesimally small time interval $[t, t + dt)$ is given by

$$\tilde{\beta}dt = \beta_0 dt + \sigma dB(t)$$

(where $\tilde{\beta}$ is as in section 3.2 and β_0 is the given value as above) which implies that

$$\int_0^t \tilde{\beta}dt = \beta_0 t + \sigma B(t)$$

i.e. the total number of potentially infectious contacts between them in $[0, t)$ has a normal distribution with mean $\beta_0 t$ and variance $\sigma^2 t$. This is a well-established method [26, 29, 33, 38, 58, 63, 90], although both methods are biologically reasonable.

One way to compare the two approaches for including environmental stochasticity in the disease transmission coefficient $\tilde{\beta}$ is to consider the average value of the parameter $\tilde{\beta}$ over a small time interval Δt, i.e., let $\beta_a = \frac{1}{\Delta t}\int_t^{t+\Delta t} \tilde{\beta}dt$ for large time t. The well-established method used in this chapter gives mean $\mathbb{E}(\beta_a) = \beta_0$ with variance $V(\beta_a) = \sigma^2/\Delta t$ that blows up as Δt goes to zero. However, the method using an Ornstein-Uhlenbeck SDE gives the same mean $\mathbb{E}(\beta_a) = \beta_0$ but with variance $V(\beta_a) = \sigma^2/(2\gamma) + O(\Delta t)$ that approaches the constant value $\sigma^2/(2\gamma)$ as Δt goes to zero.

In this chapter we have looked at an SDE version of the classical SIS epidemic model, with noise introduced in the disease transmission term. We showed that the SDE had a unique positive global solution and established conditions for extinction and persistence of disease. A key parameter was the basic reproduction number R_0^S, which was less than the corresponding deterministic version R_0^D. Theorems 3.2 and 3.4 show that if $R_0^S \leq 1$, under mild extra conditions the disease would die out. Theorem 3.6 shows that if $R_0^S > 1$ then the disease will persist. We also showed (Theorem 3.12) that if $R_0^S > 1$ then the model has a unique stationary distribution and derived expressions for its mean and variance (Theorem 3.13). We made a conjecture about the disease behaviour if $R_0^S \leq 1$ and the conditions of Theorems 3.2 and 3.4 are not satisfied. Throughout the chapter we have illustrated our theoretical results with computer simulations, including two sets with realistic parameter values for gonorrhea amongst homosexuals and pneumococcus amongst young children.

Most of the contents of this chapter have been published in [39]. In the next chapter we examine the effect of telegraph noise on the SIS epidemic model.

Chapter 4

The SIS Epidemic Model with Markovian Switching

4.1 Introduction

We still consider the SIS epidemic model (1.2.1) in this chapter. All the details about the SIS epidemic model can be found in section 1.2 including the solutions of the system (1.2.2) and the basic reproduction number R_0^D (1.2.3). In this chapter we will discuss the effect of telegraph noise on the SIS epidemic model. Recall that telegraph noise can be illustrated as a switching between two or more regimes of environment. The switching is memoryless and the waiting time for the next switch has an exponential distribution, which can be modelled by a finite state Markov chain. In section 1.3.3 we used an example of a predator-prey Lotka-Volterra (LV) model to explain the significant effect of telegraph noise on the biological model, and also, we briefly reviewed the existing literature regarding the Markovian environments in the fields of biology. Motivated by these, we examine the effect of telegraph noise in the SIS epidemic model.

It is easy to see that $I(t)$ in (1.2.1) obeys the scalar Lotka–Volterra model

$$\frac{dI(t)}{dt} = I(t)[\beta N - \mu - \gamma - \beta I(t)], \tag{4.1.1}$$

which has the explicit solution as equation (1.2.2).

And we can conclude (see e.g. [86]):

- If $R_0^D \leq 1$, $\lim_{t\to\infty} I(t) = 0$.
- If $R_0^D > 1$, $\lim_{t\to\infty} I(t) = \frac{\beta N-\mu-\gamma}{\beta}$. In this case, $I(t)$ will monotonically decrease or increase to $\frac{\beta N-\mu-\gamma}{\beta}$ if $I(0) > \frac{\beta N-\mu-\gamma}{\beta}$ or $< \frac{\beta N-\mu-\gamma}{\beta}$, respectively, while $I(t) \equiv \frac{\beta N-\mu-\gamma}{\beta}$ if $I(0) = \frac{\beta N-\mu-\gamma}{\beta}$.

Taking into account the environmental noise, the system parameters μ, β and γ may experience abrupt changes. In the same fashion as in Takeuchi et al. [87], we may model

these abrupt changes by a Markov chain. As a result, the classical SIS model (1.2.1) evolves to a stochastic SIS model with Markovian switching of the form

$$\begin{cases} \frac{dS(t)}{dt} = \mu_{r(t)}N - \beta_{r(t)}S(t)I(t) + \gamma_{r(t)}I(t) - \mu_{r(t)}S(t), \\ \frac{dI(t)}{dt} = \beta_{r(t)}S(t)I(t) - (\mu_{r(t)} + \gamma_{r(t)})I(t), \end{cases} \tag{4.1.2}$$

where $r(t)$ is a Markov chain with a finite state space. The main aim of this chapter is to discuss the effect of the noise in the form of Markov switching. This chapter is organised as follows: To make the theory more understandable, we begin with the special case where the Markov chain has only two states as in Takeuchi et al. [87]. In section 4.2, we establish the explicit solution of the stochastic SIS epidemic model, which is useful in performing computer simulations. In sections 4.3, 4.4 and 4.5 we establish the conditions for extinction and persistence for the stochastic SIS epidemic model and compare these with the corresponding conditions for the deterministic SIS epidemic model. In section 4.6, we perform computer simulations based on the explicit solution and the Euler–Maruyama scheme to illustrate our theory. We then generalise our theory to the general case where the Markov chain has a finite number of states, M, in section 4.7. We then include a more realistic example using appropriate parameter values for the spread of *S.Pneumoniae* in children in section 4.8. We draw a conclusion for this chapter in section 4.9.

4.2 SIS Model with Markovian Switching

Throughout this chapter, we let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. We first define the Markovian switching for our model (4.1.2). Recall that we give the background knowledge about Markov chains in section 2.4, which will be very useful here. Let $r(t)$, $t \geq 0$, be a right-continuous Markov chain on the probability space taking values in the state space $\mathbb{S} = \{1, 2\}$ with the generator

$$\Gamma = \begin{pmatrix} -\nu_{12} & \nu_{12} \\ \nu_{21} & -\nu_{21} \end{pmatrix}.$$

Here $\nu_{12} > 0$ is the transition rate from state 1 to 2, while $\nu_{21} > 0$ is the transition rate from state 2 to 1, that is

$$\mathbb{P}\{r(t+\delta) = 2 | r(t) = 1\} = \nu_{12}\delta + o(\delta) \quad \text{and} \quad \mathbb{P}\{r(t+\delta) = 1 | r(t) = 2\} = \nu_{21}\delta + o(\delta),$$

where $\delta > 0$. We know from the review in section 2.4 that almost every sample path of $r(\cdot)$ is a right continuous step function with a finite number of sample jumps in any finite subinterval of $\mathbb{R}_+$. More precisely, there is a sequence $\{\tau_k\}_{k\geq 0}$ of finite-valued $\mathcal{F}_t$-stopping times such that $0 = \tau_0 < \tau_1 < \cdots < \tau_k \to \infty$ almost surely and

$$r(t) = \sum_{k=0}^{\infty} r(\tau_k) I_{[\tau_k, \tau_{k+1})}(t). \tag{4.2.1}$$

Moreover, given that $r(\tau_k) = 1$, the random variable $\tau_{k+1} - \tau_k$ follows the exponential distribution with parameter ν_{12}, namely

$$\mathbb{P}(\tau_{k+1} - \tau_k \geq T | r(\tau_k) = 1) = e^{-\nu_{12}T}, \quad \forall T \geq 0,$$

while given that $r(\tau_k) = 2$, $\tau_{k+1} - \tau_k$ follows the exponential distribution with parameter ν_{21}, namely

$$\mathbb{P}(\tau_{k+1} - \tau_k \geq T | r(\tau_k) = 2) = e^{-\nu_{21} T}, \quad \forall T \geq 0.$$

The sample paths of the Markov chain can therefore be simulated easily using these exponential distributions (we will illustrate this in section 4.6 below). Furthermore, this Markov chain has a unique stationary distribution $\Pi = (\pi_1, \pi_2)$ given by

$$\pi_1 = \frac{\nu_{21}}{\nu_{12} + \nu_{21}}, \quad \pi_2 = \frac{\nu_{12}}{\nu_{12} + \nu_{21}}. \tag{4.2.2}$$

Now we return to the stochastic SIS epidemic model (4.1.2). We assume that the system parameters β_i, μ_i, γ_i ($i \in \mathbb{S}$) are all positive numbers. Given that $I(t) + S(t) = N$, we see that $I(t)$, the number of infectious individuals, obeys the stochastic Lotka–Volterra model with Markovian switching given by

$$\frac{dI(t)}{dt} = I(t)[\alpha_{r(t)} - \beta_{r(t)} I(t)], \tag{4.2.3}$$

where

$$\alpha_i := \beta_i N - \mu_i - \gamma_i, \quad i \in \mathbb{S}. \tag{4.2.4}$$

It is sufficient to study equation (4.2.3) in order to understand the full dynamics of the stochastic SIS epidemic model (4.1.2), hence we will concentrate on this equation only in the remainder of this chapter. The following theorem shows that this equation has an explicit solution for any given initial value in $(0, N)$.

Theorem 4.1 *For any given initial value $I(0) = I_0 \in (0, N)$, there is a unique solution $I(t)$ on $t \in \mathbb{R}_+$ to equation (4.2.3) such that*

$$\mathbb{P}(I(t) \in (0, N) \text{ for all } t \geq 0) = 1.$$

Moreover, the solution has the explicit form

$$I(t) = \frac{\exp\left(\int_0^t \alpha_{r(s)} ds\right)}{\frac{1}{I_0} + \int_0^t \exp\left(\int_0^s \alpha_{r(u)} du\right) \beta_{r(s)} ds}. \tag{4.2.5}$$

Proof. Fix any sample path of the Markov chain. Without loss of generality we may assume that this sample path has its initial value $r(0) = 1$, as the proof is the same if $r(0) = 2$. We first observe from (4.2.1) that $r(t) = 1$ for $t \in [\tau_0, \tau_1)$. Hence equation (4.2.3) becomes

$$\frac{dI(t)}{dt} = I(t)[\alpha_1 - \beta_1 I(t)]$$

on $t \in [\tau_0, \tau_1)$. But this equation has a unique solution on the entire set of $t \in \mathbb{R}_+$ and the solution will remain within $(0, N)$. Hence the solution of equation (4.2.3), $I(t)$, is uniquely determined on $t \in [\tau_0, \tau_1)$ and, by continuity, for $t = \tau_1$ as well. Obtaining $I(\tau_1) \in (0, N)$, we further consider equation (4.2.3) for $t \in [\tau_1, \tau_2)$, which has the form

$$\frac{dI(t)}{dt} = I(t)[\alpha_2 - \beta_2 I(t)].$$

This equation has a unique solution on $t \geq \tau_1$ and the solution will remain within $(0, N)$. Hence the solution of equation (4.2.3), $I(t)$, is uniquely determined on $t \in [\tau_1, \tau_2)$ and, by continuity, for $t = \tau_2$ as well. Repeating this procedure, we see that equation (4.2.3) has a unique solution $I(t)$ on $t \in \mathbb{R}_+$ and the solution remains within $(0, N)$ with probability one.

After showing $I(t) \in (0, N)$, we may define

$$y(t) = \frac{1}{I(t)}, \quad t \geq 0,$$

in order to obtain the explicit solution. Compute

$$\begin{aligned}
\frac{dy(t)}{dt} &= -\frac{1}{I(t)^2}\frac{dI(t)}{dt} \\
&= -\frac{1}{I(t)^2} I(t)(\alpha_{r(t)} - \beta_{r(t)} I(t)) \\
&= \beta_{r(t)} - \frac{\alpha_{r(t)}}{I(t)} \\
&= \beta_{r(t)} - \alpha_{r(t)} y(t).
\end{aligned}$$

By the well-known variation-of-constants formula (Theorem 2.18), we have

$$y(t) = \Phi(t)\Big(y(0) + \int_0^t \Phi^{-1}(s)\beta_{r(s)}\, ds\Big),$$

where $\Phi(t) = e^{-\int_0^t \alpha_{r(s)}\, ds}$. This yields the desired explicit solution (4.2.5) immediately.

4.3 The Basic Reproduction Number

Naturally we wish to examine the behaviour of the stochastic SIS epidemic model (4.2.3) and we may ask what is the corresponding basic reproduction number R_0^S. Recall in section 1.2 that the basic reproduction number is the expected number of secondary cases caused by a single newly-infected case entering the disease-free population at equilibrium.

In our case the disease-free equilibrium (DFE) is $S = N,\ I = 0$. The individuals can be divided into two types, those who arrive when $r(t) = 1$, and those that arrive when $r(t) = 2$. Suppose that a newly infected individual enters the DFE when the Markov chain is in state 1. Then the next events that can happen are that the individual dies at rate μ_1, recovers at rate γ_1 or the Markov chain switches at rate ν_{12}. Hence the expected number of individuals infected before the first switch is

$$\frac{\beta_1 N}{\mu_1 + \gamma_1 + \nu_{12}}.$$

The expected number of individuals infected between the first and second switches is

$$\frac{\nu_{12}}{\mu_1 + \gamma_1 + \nu_{12}} \frac{\beta_2 N}{\mu_2 + \gamma_2 + \nu_{21}}$$

and between the second and third switches

$$\frac{\nu_{21}}{\mu_2+\gamma_2+\nu_{21}}\,\frac{\nu_{12}}{\mu_1+\gamma_1+\nu_{12}}\,\frac{\beta_1 N}{\mu_1+\gamma_1+\nu_{12}} = p\,\frac{\beta_1 N}{\mu_1+\gamma_1+\nu_{12}}$$

where $p = \dfrac{\nu_{12}\nu_{21}}{(\mu_1+\gamma_1+\nu_{12})(\mu_2+\gamma_2+\nu_{21})}$.

Hence this individual infects in total

$$m_{11} = \frac{\beta_1 N}{\mu_1+\gamma_1+\nu_{12}}\,(1+p+p^2+\dots) = \frac{\beta_1 N}{\mu_1+\gamma_1+\nu_{12}}\,\frac{1}{1-p}$$

individuals while the Markov chain is in state 1 and

$$m_{12} = \frac{\nu_{12}}{\mu_1+\gamma_1+\nu_{12}}\,\frac{\beta_2 N}{\mu_2+\gamma_2+\nu_{21}}\,\frac{1}{1-p}$$

individuals while the Markov chain is in state 2.

Similarly we can derive the expected number of individuals infected by a single newly infected individual entering the DFE when the Markov chain is in state 2. We deduce that the next generation matrix giving the expected number of secondary cases caused by a single newly infected individual entering the DFE is

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \frac{1}{1-p}\begin{pmatrix} a_1 & p_1 a_2 \\ p_2 a_1 & a_2 \end{pmatrix},$$

where $a_1 = \dfrac{\beta_1 N}{\mu_1+\gamma_1+\nu_{12}}$, $a_2 = \dfrac{\beta_2 N}{\mu_2+\gamma_2+\nu_{21}}$, $p_1 = \dfrac{\nu_{12}}{\mu_1+\gamma_1+\nu_{12}}$ and

$p_2 = \dfrac{\nu_{21}}{\mu_2+\gamma_2+\nu_{21}}$.

The basic reproduction number for the stochastic epidemic model is the largest eigenvalue of this matrix

$$\tilde{R}_0^S = \frac{a_1+a_2+\sqrt{(a_1+a_2)^2-4a_1a_2(1-p)}}{2(1-p)}. \tag{4.3.1}$$

However we do not pursue this further here.

4.4 Extinction

In this section we will establish extinction conditions. Recall that for the deterministic SIS epidemic model (4.1.1), the basic reproduction number R_0^D was also the threshold between disease extinction and persistence, with extinction for $R_0^D \leq 1$ and persistence for $R_0^D > 1$. Recall in section 2.7 for the stochastic model, there are different types of extinction and persistence, for example almost sure extinction, extinction in mean square and extinction in probability. In the rest of this chapter we examine a threshold

$$T_0^S = \frac{\pi_1\beta_1 N + \pi_2\beta_2 N}{\pi_1(\mu_1+\gamma_1)+\pi_2(\mu_2+\gamma_2)} \tag{4.4.1}$$

for almost sure extinction or persistence of our stochastic epidemic model. However this threshold is different to $\tilde{R}_0^S$ (4.3.1) which might be more relevant to other types of extinction or persistence with different conditions.

We will see later that the stochastic SIS model (4.2.3) will become extinct with probability one if $T_0^S < 1$. Before we state this result, let us state a proposition which gives an equivalent condition for $T_0^S < 1$ in terms of the system parameters α_i and the stationary distribution of the Markov chain.

Proposition 4.2 *We have the following alternative condition on the value of T_0^S:*

- $T_0^S < 1$ *if and only if* $\pi_1\alpha_1 + \pi_2\alpha_2 < 0$;
- $T_0^S = 1$ *if and only if* $\pi_1\alpha_1 + \pi_2\alpha_2 = 0$;
- $T_0^S > 1$ *if and only if* $\pi_1\alpha_1 + \pi_2\alpha_2 > 0$.

The proof of this proposition is straightforward, so is omitted. We can now state our theory on extinction.

Theorem 4.3 *If $T_0^S < 1$, then, for any given initial value $I_0 \in (0, N)$, the solution of the stochastic SIS epidemic model (4.2.3) obeys*

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le \alpha_1\pi_1 + \alpha_2\pi_2 \quad a.s. \tag{4.4.2}$$

By Proposition 4.2, we hence conclude that $I(t)$ tends to zero exponentially almost surely. In other words, the disease dies out with probability one.

Proof. It is easy to see that

$$\frac{d\log(I(t))}{dt} = \alpha_{r(t)} - \beta_{r(t)}I(t). \tag{4.4.3}$$

This implies that, for any $t > 0$,

$$\frac{\log(I(t))}{t} \le \frac{\log(I(0))}{t} + \frac{1}{t}\int_0^t \alpha_{r(s)}ds,$$

since $\beta_{r(t)} > 0$ and $I(t) \in (0, N)$. Letting $t \to \infty$ we hence obtain

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le \limsup_{t\to\infty} \frac{1}{t}\int_0^t \alpha_{r(s)}ds.$$

However, by the ergodic theory of the Markov chain (see e.g. [3]) we have

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \alpha_{r(s)}ds = \alpha_1\pi_1 + \alpha_2\pi_2 \quad a.s.$$

We therefore must have

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le \alpha_1\pi_1 + \alpha_2\pi_2 \quad a.s.,$$

as required.

Let us now make a few comments. First of all, let us recall that the stochastic SIS model (4.2.3) can be regarded as the result of the following two subsystems:

$$\frac{dI(t)}{dt} = I(t)[\alpha_1 - \beta_1 I(t)] \tag{4.4.4}$$

and

$$\frac{dI(t)}{dt} = I(t)[\alpha_2 - \beta_2 I(t)], \tag{4.4.5}$$

switching from one to the other according to the law of the Markov chain. If both $\alpha_1 < 0$ and $\alpha_2 < 0$, then the corresponding R_0^D values for both subsystems (4.4.4) and (4.4.5) are less than 1, whence both subsystems become extinct. In this case, T_0^S for the stochastic SIS model (4.2.3) is less than one, hence it will become extinct, and of course this is not surprising. However, if only one of α_1 and α_2 is negative, say $\alpha_1 < 0$ and $\alpha_2 > 0$, for example, one subsystem (4.4.4) becomes extinct but the other (4.4.5) is persistent. However, if the rate of the Markov chain switching from state 2 to 1 is relatively faster than that from 1 to 2, so that $\alpha_1\pi_1 + \alpha_2\pi_2 < 0$, then the overall system (4.2.3) will become extinct. This reveals the important role of the Markov chain in the extinction.

We next recall that in the deterministic SIS model (4.1.1) the disease will always go extinct even if $R_0^D = 1$. The reader may ask what happens to the stochastic SIS model (4.2.3) if the corresponding $T_0^S = 1$. Although we have a strong feeling that the disease will always become extinct, we have not been able to prove it so far. In Section 4.6.3 we show some simulations to illustrate this case.

4.5 Persistence

Let us now turn to the case when $T_0^S > 1$. The following theorem shows that the disease will be persistent in this case.

Theorem 4.4 *If $T_0^S > 1$, then, for any given initial value $I_0 \in (0, N)$, the solution of the stochastic SIS model (4.2.3) has the properties that*

$$\liminf_{t\to\infty} I(t) \le \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} \quad a.s. \tag{4.5.1}$$

and

$$\limsup_{t\to\infty} I(t) \ge \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} \quad a.s. \tag{4.5.2}$$

In other words, the disease will reach the neighbourhood of the level $\frac{\pi_1\alpha_1+\pi_2\alpha_2}{\pi_1\beta_1+\pi_2\beta_2}$ infinitely many times with probability one.

Proof. Let us first prove assertion (4.5.1). If this were not true, then we can find an $\varepsilon > 0$ sufficiently small for $\mathbb{P}(\Omega_1) > 0$ where

$$\Omega_1 = \left\{ \omega \in \Omega : \liminf_{t\to\infty} I(t) > \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon \right\}. \tag{4.5.3}$$

On the other hand, by the ergodic theory of the Markov chain, we have that $\mathbb{P}(\Omega_2) = 1$, where for any $\omega \in \Omega_2$,

$$\begin{aligned} \lim_{t\to\infty} \frac{1}{t} \int_0^t \Big(\alpha_{r(s)} - \beta_{r(s)} \Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon\Big]\Big) ds & \\ = \pi_1\Big(\alpha_1 - \beta_1\Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon\Big]\Big) + \pi_2\Big(\alpha_2 - \beta_2\Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon\Big]\Big) & \\ = -(\pi_1\beta_1 + \pi_2\beta_2)\varepsilon. & \end{aligned} \tag{4.5.4}$$

Now consider any $\omega \in \Omega_1 \cap \Omega_2$. Then there is a positive number $T = T(\omega)$ such that

$$I(t) \geq \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon \quad \forall t \geq T.$$

It then follows from (4.4.3) that

$$\log(I(t)) \leq \log(I_0) + \int_0^T (\alpha_{r(s)} - \beta_{r(s)} I(s)) ds + \int_T^t \Big(\alpha_{r(s)} - \beta_{r(s)} \Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} + \varepsilon\Big]\Big) ds$$

for all $t \geq T$. Dividing both sides by t and then letting $t \to \infty$, we obtain that

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \leq -(\pi_1\beta_1 + \pi_2\beta_2)\varepsilon,$$

where (4.5.4) has been used. This implies that

$$\lim_{t\to\infty} I(t) = 0.$$

But this contradicts (5.2.3). The required assertion (4.5.1) must therefore hold.

The procedure to prove assertion (4.5.2) is very similar. In fact if (4.5.2) were not true, we can then find an $\varepsilon > 0$ sufficiently small for $\mathbb{P}(\Omega_3) > 0$, where

$$\Omega_3 = \left\{ \omega \in \Omega : \limsup_{t\to\infty} I(t) < \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon \right\}. \tag{4.5.5}$$

By the ergodic theory we also have that $\mathbb{P}(\Omega_4) = 1$, where for any $\omega \in \Omega_4$,

$$\begin{aligned} \lim_{t\to\infty} \frac{1}{t} \int_0^t \Big(\alpha_{r(s)} - \beta_{r(s)} \Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon\Big]\Big) ds & \\ = \pi_1\Big(\alpha_1 - \beta_1\Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon\Big]\Big) + \pi_2\Big(\alpha_2 - \beta_2\Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon\Big]\Big) & \\ = (\pi_1\beta_1 + \pi_2\beta_2)\varepsilon. & \end{aligned} \tag{4.5.6}$$

If we consider any $\omega \in \Omega_3 \cap \Omega_4$, there is a positive number $T = T(\omega)$ such that

$$I(t) \leq \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon \quad \forall t \geq T.$$

From (4.4.3) we have that

$$\log(I(t)) \geq \log(I_0) + \int_0^T (\alpha_{r(s)} - \beta_{r(s)}I(s))ds + \int_T^t \Big(\alpha_{r(s)} - \beta_{r(s)}\Big[\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} - \varepsilon\Big]\Big)ds$$

for all $t \geq T$. Dividing both sides by t and then letting $t \to \infty$ while using (4.5.6) as well, we obtain that

$$\liminf_{t\to\infty} \frac{1}{t} \log(I(t)) \geq (\pi_1\beta_1 + \pi_2\beta_2)\varepsilon.$$

This implies that

$$\lim_{t\to\infty} I(t) \to \infty,$$

which contradicts (4.5.5). Therefore assertion (4.5.2) must hold.

To reveal more properties of the stochastic SIS model, we observe from Proposition 4.2 that $T_0^S > 1$ is equivalent to the condition that $\pi_1\alpha_1 + \pi_2\alpha_2 > 0$. This may be divided into two cases: (a) both α_1 and α_2 are positive; and (b) only one of α_1 and α_2 is positive. Without loss of generality, we may assume that $0 < \alpha_1/\beta_1 = \alpha_2/\beta_2$ or $0 < \alpha_1/\beta_1 < \alpha_2/\beta_2$ in Case (a), while $\alpha_1/\beta_1 \leq 0 < \alpha_2/\beta_2$ in Case (b). So there are three different cases to be considered under condition $T_0^S > 1$. Let us present a lemma in order to show another new result.

Lemma 4.5 *The following statements hold with probability one:*

(i) If $0 < \alpha_1/\beta_1 = \alpha_2/\beta_2$, then $I(t) = \alpha_1/\beta_1$ for all $t > 0$ when $I_0 = \alpha_1/\beta_1$.

(ii) If $0 < \alpha_1/\beta_1 < \alpha_2/\beta_2$, then $I(t) \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$ for all $t > 0$ whenever $I_0 \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$.

(iii) If $\alpha_1/\beta_1 \leq 0 < \alpha_2/\beta_2$, then $I(t) \in (0, \alpha_2/\beta_2)$ for all $t > 0$ whenever $I_0 \in (0, \alpha_2/\beta_2)$.

Proof. Case (i) is obvious. To prove Case (ii), we may assume, without loss of generality, that $r(0) = 1$. Recalling (4.2.1) and the properties of the deterministic SIS model (4.1.1) which we stated in section 4.1, we see that $I(t)$ will monotonically decrease during the time interval $[\tau_0, \tau_1]$ but never reach α_1/β_1, whence $I(t) \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$. At time τ_1, the Markov chain switches to state 2 and will not jump to state 1 until time τ_2. During this time interval $[\tau_1, \tau_2]$, $I(t)$ will monotonically increase but never reach α_2/β_2, whence $I(t) \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$ again. Repeating this argument, we see that $I(t)$ will remain within $(\alpha_1/\beta_1, \alpha_2/\beta_2)$ forever. Similarly, we can show Case (iii).

In the following study we will use the Markov property of the solutions (see section 2.4). For this purpose, let us denote by $\mathbb{P}_{I_0,r_0}$ the conditional probability measure generated by the pair of processes $(I(t), r(t))$ given the initial condition $(I(0), r(0)) = (I_0, r_0) \in (0, N) \times \mathbb{S}$.

Theorem 4.6 *Assume that $T_0^S > 1$ and let $I_0 \in (0, N)$ be arbitrary. The following statements hold with probability one:*

(i) If $0 < \alpha_1/\beta_1 = \alpha_2/\beta_2$, then $\lim_{t\to\infty} I(t) = \alpha_1/\beta_1$.

(ii) If $0 < \alpha_1/\beta_1 < \alpha_2/\beta_2$, then

$$\frac{\alpha_1}{\beta_1} \le \liminf_{t\to\infty} I(t) \le \limsup_{t\to\infty} I(t) \le \frac{\alpha_2}{\beta_2}.$$

(iii) If $\alpha_1/\beta_1 \le 0 < \alpha_2/\beta_2$, then

$$0 \le \liminf_{t\to\infty} I(t) \le \limsup_{t\to\infty} I(t) \le \frac{\alpha_2}{\beta_2}.$$

Proof. Case (i). If $I_0 = \alpha_1/\beta_1$, then $I(t) = \alpha_1/\beta_1$ for all $t \ge 0$, whence the assertion holds. If $I_0 < \alpha_1/\beta_1$, it is easy to see that $I(t)$ increases monotonically on $t \ge 0$, hence $\lim_{t\to\infty} I(t)$ exists. By Theorem 4.4, we therefore have

$$\lim_{t\to\infty} I(t) = \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} \quad a.s.$$

But, given $\alpha_1/\beta_1 = \alpha_2/\beta_2$, we compute

$$\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} = \frac{\pi_1\alpha_1 + \pi_2\alpha_1\beta_2/\beta_1}{\pi_1\beta_1 + \pi_2\beta_2} = \frac{\alpha_1}{\beta_1}.$$

We therefore have $\lim_{t\to\infty} I(t) = \alpha_1/\beta_1$ a.s. Similarly, we can show this for $I_0 > \alpha_1/\beta_1$.

Case (ii). If $I_0 \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$, then the assertion follows from Lemma 4.5 directly. Let us now assume that $I_0 \ge \alpha_2/\beta_2$. Given $0 < \alpha_1/\beta_1 < \alpha_2/\beta_2$, it is easy to show that

$$\frac{\alpha_1}{\beta_1} < \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} < \frac{\alpha_2}{\beta_2}.$$

Consider a number

$$\kappa \in \Big(\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2},\ \frac{\alpha_2}{\beta_2}\Big),$$

and define the stopping time

$$\rho_\kappa = \inf\{t \ge 0 : I(t) \le \kappa\}.$$

By Theorem 4.4 we have

$$\mathbb{P}(\rho_k < \infty) = 1,$$

while by the continuity of $I(t)$ we have $I(\rho_\kappa) = \kappa$. Set

$$\bar\Omega = \Big\{\alpha_1/\beta_1 \le \liminf_{t\to\infty} I(t) \le \limsup_{t\to\infty} I(t) \le \alpha_2/\beta_2\Big\}$$

and denote its indicator function by $\mathbb{I}_{\bar\Omega}$. By the strong Markov property, we compute

$$\mathbb{P}(\bar\Omega) = \mathbb{E}(\mathbb{I}_{\bar\Omega}) = \mathbb{E}(\mathbb{E}(\mathbb{I}_{\bar\Omega}|\mathcal{F}_{\rho_\kappa})) = \\ \mathbb{E}(\mathbb{E}(\mathbb{I}_{\bar\Omega}|I(\rho_\kappa), r(\rho_\kappa))) = \mathbb{E}(\mathbb{P}_{I(\rho_\kappa),r(\rho_\kappa)}(\bar\Omega)) = \mathbb{E}(\mathbb{P}_{\kappa,r(\rho_\kappa)}(\bar\Omega)).$$

But, by Lemma 4.5, $\mathbb{P}_{\kappa,r(\rho_\kappa)}(\bar{\Omega}) = 1$ and hence we have $\mathbb{P}(\bar{\Omega}) = 1$ as required. Similarly, we can show that $\mathbb{P}(\bar{\Omega}) = 1$ for $I_0 \le \alpha_1/\beta_1$.

Case (iii). It is obvious that $0 \le \liminf_{t\to 0} I(t)$, while the assertion that $\liminf_{t\to 0} I(t) \le \alpha_2/\beta_2$ can be proved in the same way as Case (ii) was proved. The proof is therefore complete.

Under the condition $T_0^S > 1$, the theorem above shows precisely that $I(t)$ will tend to α_1/β_1 with probability one if $\alpha_1/\beta_1 = \alpha_2/\beta_2$. However, it is quite rare to have $\alpha_1/\beta_1 = \alpha_2/\beta_2$ in practice. It is therefore more useful to study the case when, say, $\alpha_1/\beta_1 < \alpha_2/\beta_2$ in a bit more detail. In the proof above, we have in fact shown a slightly stronger result than Theorem 4.6 states, namely we have shown that

$$\mathbb{P}(I(t) \in (0 \vee (\alpha_1/\beta_1),\ \alpha_2/\beta_2) \text{ for all } t \ge \rho_\kappa) = 1. \tag{4.5.7}$$

It would be interesting to find out how $I(t)$ will vary within the interval $(0\vee(\alpha_1/\beta_1),\ \alpha_2/\beta_2)$ in the long term. The following theorem shows that $I(t)$ can take any value up to the boundaries of the interval infinitely many times (though never reach them) with positive probability.

Theorem 4.7 *Assume that $T_0^S > 1$ and $0 < \frac{\alpha_1}{\beta_1} < \frac{\alpha_2}{\beta_2}$, and let $I_0 \in (0, N)$ be arbitrary. Then for any $\varepsilon > 0$, sufficiently small for*

$$\frac{\alpha_1}{\beta_1} + \varepsilon < \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} < \frac{\alpha_2}{\beta_2} - \varepsilon,$$

the solution of the stochastic SIS epidemic model (4.2.3) has the properties that

$$\mathbb{P}\Big(\liminf_{t\to\infty} I(t) < \frac{\alpha_1}{\beta_1} + \varepsilon\Big) \ge e^{-\nu_{12}T_1(\varepsilon)}, \tag{4.5.8}$$

and

$$\mathbb{P}\Big(\limsup_{t\to\infty} I(t) > \frac{\alpha_2}{\beta_2} - \varepsilon\Big) \ge e^{-\nu_{21}T_2(\varepsilon)}, \tag{4.5.9}$$

where $T_1(\varepsilon) > 0$ and $T_2(\varepsilon) > 0$ are defined by

$$T_1(\varepsilon) = \frac{1}{\alpha_1}\Big(\log\Big(\frac{\beta_1}{\alpha_1} - \frac{\beta_2}{\alpha_2}\Big) + \log\Big(\frac{\alpha_1}{\beta_1} + \varepsilon\Big) - \log\Big(\frac{\varepsilon\beta_1}{\alpha_1}\Big)\Big) \tag{4.5.10}$$

and

$$T_2(\varepsilon) = \frac{1}{\alpha_2}\Big(\log\Big(\frac{\beta_1}{\alpha_1} - \frac{\beta_2}{\alpha_2}\Big) + \log\Big(\frac{\alpha_2}{\beta_2} - \varepsilon\Big) - \log\Big(\frac{\varepsilon\beta_2}{\alpha_2}\Big)\Big). \tag{4.5.11}$$

Proof. Let $T > 0$ be arbitrary. Define the stopping time

$$\sigma_1 = \inf\{t \ge T : I(t) \in (\alpha_1/\beta_1 + \varepsilon, \alpha_2/\beta_2 - \varepsilon)\}.$$

By Theorem 4.4, we have $\mathbb{P}(\sigma_1 < \infty) = 1$, while we see from the proof of Theorem 4.6 that

$$\mathbb{P}(I(t) \in (\alpha_1/\beta_1, \alpha_2/\beta_2) \text{ for all } t \ge \sigma_1) = 1. \tag{4.5.12}$$

To prove assertion (4.5.8), we define another stopping time

$$\sigma_2 = \inf\{t \geq \sigma_1 : r(t) = 1\}.$$

Clearly, $\mathbb{P}(\sigma_2 < \infty) = 1$ and by the right-continuity of the Markov chain, $r(\sigma_2) = 1$. By the memoryless property of an exponential distribution, the probability that the Markov chain will not jump to state 2 before $\sigma_2 + T_1(\varepsilon)$ is

$$\mathbb{P}(\Omega_1) = e^{-\nu_{12}T_1(\varepsilon)}, \tag{4.5.13}$$

where $\Omega_1 = \{r(\sigma_2 + t) = 1 \text{ for all } t \in [0, T_1(\varepsilon)]\}$. Now, consider any $\omega \in \Omega_1$ and consider $I(t)$ on $t \in [\sigma_2, \sigma_2 + T_1(\varepsilon)]$. Note that it obeys the differential equation

$$\frac{dI(t)}{dt} = I(t)(\alpha_1 - \beta_1 I(t)),$$

with initial value $I(\sigma_2) \in (\alpha_1/\beta_1, \alpha_2/\beta_2)$. By the explicit solution of this equation (1.2.2), we have

$$I(\sigma_2 + T_1(\varepsilon)) = \Big[e^{-\alpha_1 T_1(\varepsilon)}\Big(\frac{1}{I(\sigma_2)} - \frac{\beta_1}{\alpha_1}\Big) + \frac{\beta_1}{\alpha_1}\Big]^{-1}.$$

On the other hand, by (4.5.10), we have

$$\Big[e^{-\alpha_1 T_1(\varepsilon)}\Big(\frac{\beta_2}{\alpha_2} - \frac{\beta_1}{\alpha_1}\Big) + \frac{\beta_1}{\alpha_1}\Big]^{-1} = \frac{\alpha_1}{\beta_1} + \varepsilon.$$

Since $I(\sigma_2) < \alpha_2/\beta_2$, we must therefore have

$$I(\sigma_2 + T_1(\varepsilon)) < \frac{\alpha_1}{\beta_1} + \varepsilon.$$

Consequently

$$\mathbb{P}\Big(\inf_{T \leq t < \infty} I(t) < \frac{\alpha_1}{\beta_1} + \varepsilon\Big) \geq \mathbb{P}(\Omega_1) = e^{-\nu_{12}T_1(\varepsilon)}. \tag{4.5.14}$$

Noting that

$$\Big(\liminf_{t\to\infty} I(t) < \frac{\alpha_1}{\beta_1} + \varepsilon\Big) = \bigcup_{0<T<\infty} \Big(\inf_{T \leq t < \infty} I(t) < \frac{\alpha_1}{\beta_1} + \varepsilon\Big),$$

we can let $T \to \infty$ in (4.5.14) to obtain assertion (4.5.8). Similarly, we can prove the other assertion (4.5.9).

Theorem 4.8 *Assume that $T_0^S > 1$ (namely $\pi_1\alpha_1 + \pi_2\alpha_2 > 0$) and $\frac{\alpha_1}{\beta_1} \leq 0 < \frac{\alpha_2}{\beta_2}$. Let $I_0 \in (0, N)$ be arbitrary. Then for any $\varepsilon > 0$, sufficiently small for*

$$\varepsilon < \frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2} < \frac{\alpha_2}{\beta_2} - \varepsilon,$$

the solution of the stochastic SIS model (4.2.3) has the properties that

$$\mathbb{P}\Big(\liminf_{t\to\infty} I(t) < \varepsilon\Big) \geq e^{-\nu_{12}T_3(\varepsilon)}, \tag{4.5.15}$$

and

$$\mathbb{P}\Big(\limsup_{t\to\infty} I(t) > \frac{\alpha_2}{\beta_2} - \varepsilon\Big) \geq e^{-\nu_{21}T_4(\varepsilon)}, \tag{4.5.16}$$

where $T_3(\varepsilon) > 0$ *and* $T_4(\varepsilon) > 0$ *are defined by*

$$T_3(\varepsilon) = \frac{1}{\alpha_1}\Big(\log\Big(\frac{\beta_2}{\alpha_2} - \frac{\beta_1}{\alpha_1}\Big) - \log\Big(\frac{1}{\varepsilon} - \frac{\beta_1}{\alpha_1}\Big)\Big) \tag{4.5.17}$$

and

$$T_4(\varepsilon) = \frac{1}{\alpha_2}\Big(\log\Big(\frac{2}{\varepsilon} - \frac{\beta_2}{\alpha_2}\Big) + \log\Big(\frac{\alpha_2}{\beta_2} - \varepsilon\Big) - \log\Big(\varepsilon\frac{\beta_2}{\alpha_2}\Big)\Big). \tag{4.5.18}$$

Proof. Let $T > 0$ be arbitrary. Define the stopping time

$$\sigma_3 = \inf\{t \geq T : I(t) \in (\varepsilon, \alpha_2/\beta_2 - \varepsilon)\}.$$

By Theorem 4.4, we have $\mathbb{P}(\sigma_3 < \infty) = 1$, while we see from the proof of Theorem 4.6 that

$$\mathbb{P}(I(t) \in (0, \alpha_2/\beta_2) \text{ for all } t \geq \sigma_3) = 1. \tag{4.5.19}$$

To prove assertion (4.5.15), we define another stopping time

$$\sigma_4 = \inf\{t \geq \sigma_3 : r(t) = 1\}.$$

Clearly, $\mathbb{P}(\sigma_4 < \infty) = 1$ and by the right-continuity of the Markov chain, $r(\sigma_4) = 1$. By the memoryless property of an exponential distribution, the probability that the Markov chain will not jump to state 2 before $\sigma_4 + T_3(\varepsilon)$ is

$$\mathbb{P}(\Omega_2) = e^{-\nu_{12}T_3(\varepsilon)}, \tag{4.5.20}$$

where $\Omega_2 = \{r(\sigma_4 + t) = 1 \text{ for all } t \in [0, T_3(\varepsilon)]\}$. Now, consider any $\omega \in \Omega_2$ and consider $I(t)$ on $t \in [\sigma_4, \sigma_4 + T_3(\varepsilon)]$. Note that it obeys the differential equation

$$\frac{dI(t)}{dt} = I(t)(\alpha_1 - \beta_1 I(t)),$$

with initial value $I(\sigma_4) \in (0, \alpha_2/\beta_2)$. By the explicit solution of this equation (1.2.2), we have

$$I(\sigma_4 + T_3(\varepsilon)) = \Big[e^{-\alpha_1 T_3(\varepsilon)}\Big(\frac{1}{I(\sigma_4)} - \frac{\beta_1}{\alpha_1}\Big) + \frac{\beta_1}{\alpha_1}\Big]^{-1}.$$

On the other hand, by (4.5.17), we have

$$\Big[e^{-\alpha_1 T_3(\varepsilon)}\Big(\frac{\beta_2}{\alpha_2} - \frac{\beta_1}{\alpha_1}\Big) + \frac{\beta_1}{\alpha_1}\Big]^{-1} = \varepsilon.$$

Since $I(\sigma_4) < \alpha_2/\beta_2$, we must therefore have

$$I(\sigma_4 + T_1(\varepsilon)) < \varepsilon.$$

Consequently

$$\mathbb{P}\Big(\inf_{T\leq t<\infty} I(t) < \varepsilon\Big) \geq \mathbb{P}(\Omega_2) = e^{-\nu_{12}T_3(\varepsilon)}. \tag{4.5.21}$$

Noting that

$$\Big(\liminf_{t\to\infty} I(t) < \varepsilon\Big) = \bigcup_{0<T<\infty} \Big(\inf_{T\le t<\infty} I(t) < \varepsilon\Big),$$

we can let $T \to \infty$ in (4.5.21) to obtain assertion (4.5.15).

To prove the other assertion (4.5.16) we define the stopping time

$$\sigma_5 = \inf\{t \ge T : r(t) = 2\},$$

where $T > 0$ is arbitrary. Clearly $\mathbb{P}(\sigma_5 < \infty) = 1$. We define another stopping time

$$\sigma_6 = \inf\Big\{t \ge \sigma_5 : r(t) = 2, I(t) \ge \frac{1}{2}\Big(\frac{\pi_1\alpha_1 + \pi_2\alpha_2}{\pi_1\beta_1 + \pi_2\beta_2}\Big)\Big\}.$$

Suppose that $I(t) < \frac{1}{2}\Big(\frac{\pi_1\alpha_1+\pi_2\alpha_2}{\pi_1\beta_1+\pi_2\beta_2}\Big)$ when $t = \sigma_5$, $I(t)$ will eventually increase across this level by Theorem 4.4. Note that $I(t)$ increases monotonically when $r(t) = 2$ whilst it decreases monotonically when $r(t) = 1$. If $r(t)$ switches back to state 1 before $I(t)$ increases over this level and starts decreasing, since the $\limsup_{t\to\infty} I(t) \ge \frac{\pi_1\alpha_1+\pi_2\alpha_2}{\pi_1\beta_1+\pi_2\beta_2}$, $I(t)$ will increase across this level later on i.e. $r(t) = 2$ when $I(t)=\frac{1}{2}\Big(\frac{\pi_1\alpha_1+\pi_2\alpha_2}{\pi_1\beta_1+\pi_2\beta_2}\Big)$. Therefore we have $\mathbb{P}(\sigma_6 < \infty) = 1$. And by the right-continuity of the Markov chain, $r(\sigma_6) = 2$. By the memoryless property of an exponential distribution, the probability that the Markov chain will not jump to state 1 before $\sigma_6 + T_4(\varepsilon)$ is

$$\mathbb{P}(\Omega_3) = e^{-\nu_{21}T_4(\varepsilon)}, \tag{4.5.22}$$

where $\Omega_3 = \{r(\sigma_6 + t) = 2 \text{ for all } t \in [0, T_4(\varepsilon)]\}$. Now, consider any $\omega \in \Omega_3$ and consider $I(t)$ on $t \in [\sigma_6, \sigma_6 + T_4(\varepsilon)]$. Note that it obeys the differential equation

$$\frac{dI(t)}{dt} = I(t)(\alpha_2 - \beta_2 I(t)),$$

with initial value $I(\sigma_6) \ge \frac{1}{2}\Big(\frac{\pi_1\alpha_1+\pi_2\alpha_2}{\pi_1\beta_1+\pi_2\beta_2}\Big) > \frac{\varepsilon}{2}$. By the explicit solution of this equation (1.2.2), we have

$$I(\sigma_6 + T_4(\varepsilon)) = \Big[e^{-\alpha_2 T_4(\varepsilon)}\Big(\frac{1}{I(\sigma_6)} - \frac{\beta_2}{\alpha_2}\Big) + \frac{\beta_2}{\alpha_2}\Big]^{-1}.$$

On the other hand, by (4.5.18), we have

$$\Big[e^{-\alpha_2 T_4(\varepsilon)}\Big(\frac{2}{\varepsilon} - \frac{\beta_2}{\alpha_2}\Big) + \frac{\beta_2}{\alpha_2}\Big]^{-1} = \frac{\alpha_2}{\beta_2} - \varepsilon.$$

Since $I(\sigma_6) > \frac{\varepsilon}{2}$, we must therefore have

$$I(\sigma_6 + T_4(\varepsilon)) > \frac{\alpha_2}{\beta_2} - \varepsilon.$$

Consequently

$$\mathbb{P}\Big(\sup_{T\le t<\infty} I(t) > \frac{\alpha_2}{\beta_2} - \varepsilon\Big) \ge \mathbb{P}(\Omega_3) = e^{-\nu_{21}T_4(\varepsilon)}. \tag{4.5.23}$$

Noting that

$$\Big(\limsup_{t\to\infty} I(t) > \frac{\alpha_2}{\beta_2} - \varepsilon\Big) = \bigcap_{0<T<\infty} \Big(\sup_{T\le t<\infty} I(t) > \frac{\alpha_2}{\beta_2} - \varepsilon\Big),$$

we can let $T \to \infty$ in (4.5.23) to obtain assertion (4.5.16).

Define

$$R_{01}^D = \frac{\beta_1 N}{\mu_1 + \gamma_1} \text{ and } R_{02}^D = \frac{\beta_2 N}{\mu_2 + \gamma_2}.$$

Note that if $\alpha_j > 0$ then $R_{0j}^D > 1$ for $j = 1, 2$ and

$$\frac{\alpha_j}{\beta_j} = N\left(1 - \frac{1}{R_{0j}^D}\right)$$

is the endemic level of disease after a long time in the SIS model (1.2.1) with $\beta = \beta_j$, $\mu = \mu_j$ and $\gamma = \gamma_j$. If $\alpha_1 \le 0$ then $R_{01}^D \le 1$ and disease eventually dies out in the corresponding SIS model. So in general in the first model the disease prevalence eventually approaches $0 \vee (\alpha_1/\beta_1)$ and in the second model the disease prevalence eventually approaches α_2/β_2. These are the two levels between which the disease oscillates in the Markov chain switching model.

4.6 Simulations

In this section we shall assume that all parameters are given in appropriate units as in Chapter 3.

4.6.1 Extinction case

Example 4.9 *Assume that the system parameters are given by*

$$\mu_1 = 0.45, \mu_2 = 0.05, \gamma_1 = 0.35, \gamma_2 = 0.15, \beta_1 = 0.001, \beta_2 = 0.004, N = 100,$$

$$\nu_{12} = 0.6, \text{ and } \nu_{21} = 0.9.$$

So $\alpha_1 = -0.7$, $\alpha_2 = 0.2$, $\pi_1 = 0.6$, *and* $\pi_2 = 0.4$ *(see section 4.2 for definitions).*

Noting that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = -0.34,$$

we can therefore conclude, by Theorem 4.3, that for any given initial value $I(0) = I_0 \in (0, N)$, *the solution of (4.2.3) obeys*

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le -0.34 < 0 \quad a.s.$$

That is, $I(t)$ *will tend to zero exponentially with probability one.*

The computer simulation in Figure 4.1(a) supports this result clearly, illustrating extinction of the disease. Furthermore, $\alpha_1 < 0$ while $\alpha_2 > 0$ in this case, which means that one subsystem dies out while the other subsystem is persistent. Figure 4.1(a) shows some decreasing then increasing behaviour early on, but the general trend tends to zero, illustrating extinction for the system as a whole. The EM method with step size $\Delta = 0.001$ is also applied to approximate the solution $I(t)$. The two lines are very close to each other, showing that the EM method gives a very good approximation to the true solution in this case.

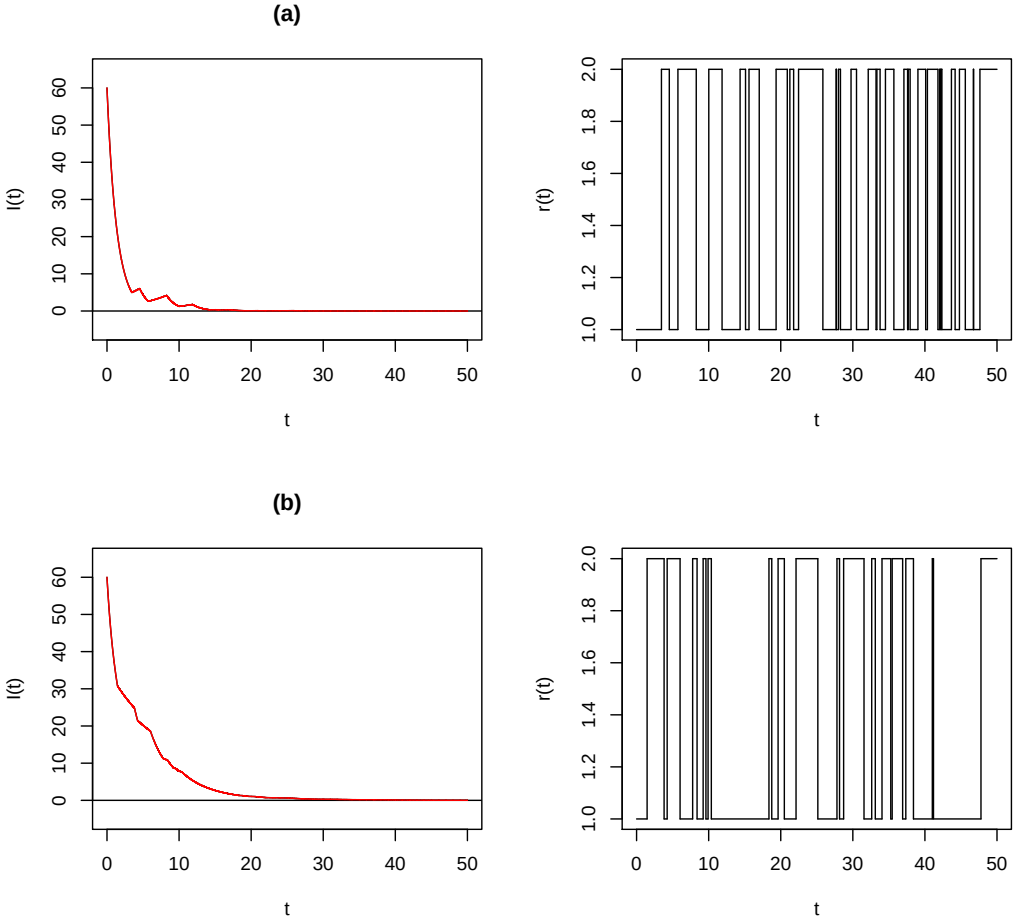


Figure 4.1: Computer simulation of $I(t)$ and its corresponding Markov chain $r(t)$, using the parameter values in Example 4.9 for (a) and in Example 4.10 for (b), $I(0) = 60$ for both cases, and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$. The black line is for $I(t)$ using formula (4.2.5) and the red line is for the EM method with step size $\Delta = 0.001$. (The two lines are very close to each other, so we hardly see the black line in the plot.)

Example 4.10 *Assume that the system parameters are given by*

$$\mu_1 = 0.45, \mu_2 = 0.05, \gamma_1 = 0.35, \gamma_2 = 0.15, \beta_1 = 0.006, \beta_2 = 0.0015, N = 100,$$

$$\nu_{12} = 0.6, \text{ and } \nu_{21} = 0.9.$$

So $\alpha_1 = -0.2$, $\alpha_2 = -0.05$, $\pi_1 = 0.6$, *and* $\pi_2 = 0.4$ *(see section 4.2 for definitions).*

Noting that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = -0.14,$$

we can therefore conclude, by Theorem 4.3, that for any given initial value $I(0) = I_0 \in (0, N)$, *the solution of (4.2.3) obeys*

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \leq -0.14 < 0 \quad a.s.$$

That is, $I(t)$ *will tend to zero exponentially with probability one. The computer simulation in Figure 1(b) supports this result clearly, illustrating extinction of the disease. Both* α_1 *and* α_2 *are less than zero in this case, which means that both subsystems die out. Figure 4.1(b) shows a trend of decreasing all the time but at different speeds, which reveals that property. As before, the EM method with step size* $\Delta = 0.001$ *gives a good approximation in this case as well.*

4.6.2 Persistence case

Example 4.11 *Assume that the system parameters are given by*

$$\mu_1 = 0.45, \mu_2 = 0.05, \gamma_1 = 0.35, \gamma_2 = 0.15, \beta_1 = 0.01, \beta_2 = 0.012, N = 100,$$

$$\nu_{12} = 0.6, \text{ and } \nu_{21} = 0.9.$$

So $\alpha_1 = 0.2$, $\alpha_2 = 1$, $\pi_1 = 0.6$, *and* $\pi_2 = 0.4$.

Noting that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = 0.52,$$

we can therefore conclude, by Theorem 4.6, that for any given initial value $I(0) = I_0 \in (0, N)$, *the solution of (4.2.3) obeys*

$$\frac{\alpha_1}{\beta_1} = 20 \leq \liminf_{t\to\infty} I(t) \leq \limsup_{t\to\infty} I(t) \leq 83.33 = \frac{\alpha_2}{\beta_2}.$$

That is, $I(t)$ *will eventually enter the region (20, 83.33) if* $I(0)$ *is not in this region, and will be attracted in this region once it has entered. Also, by Theorem 4.7,* $I(t)$ *can take any value up to the boundaries of (20, 83.33) but never reach them.*

The computer simulations in Figure 4.2(a), (b) and (c), using different initial values $I(0)$, *support these results clearly. As before, the EM method with step size* $\Delta = 0.001$ *gives a good approximation of the true solution.*

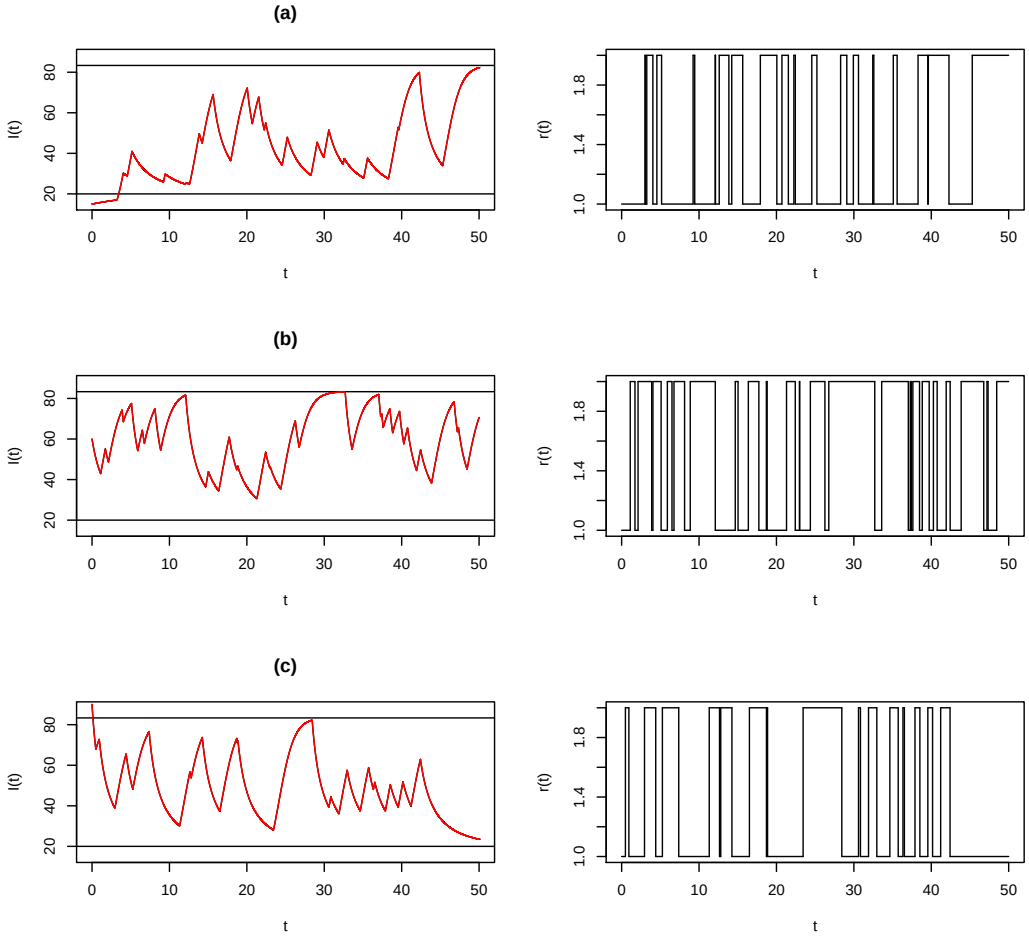


Figure 4.2: Computer simulation of $I(t)$ and its corresponding Markov chain $r(t)$, using the parameter values in Example 4.11, with $I(0) = 15$ for (a), $I(0) = 60$ for (b) and $I(0) = 90$ for (c), and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$. The black line is for $I(t)$ using formula (4.2.5) and the red line for the EM method with step size $\Delta = 0.001$. (The two lines are very close to each other, so we hardly see the black line in the plot.) The horizontal lines in the plot of $I(t)$ indicate levels $\frac{\alpha_1}{\beta_1}$ and $\frac{\alpha_2}{\beta_2}$.

Example 4.12 *Assume that the system parameters are given by*

$$\mu_1 = 0.45,\ \ \mu_2 = 0.05, \gamma_1 = 0.35, \gamma_2 = 0.15, \beta_1 = 0.004, \beta_2 = 0.012, N = 100,$$

$$\nu_{12} = 0.6,\ \ \textit{and}\ \nu_{21} = 0.9.$$

So $\alpha_1 = -0.4$, $\alpha_2 = 1$, $\pi_1 = 0.6$, *and* $\pi_2 = 0.4$.

Noting that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = 0.16,$$

we can therefore conclude, by Theorem 4.6, that for any given initial value $I(0) = I_0 \in (0, N)$*, the solution of (4.2.3) obeys*

$$0 \leq \liminf_{t\to\infty} I(t) \leq \limsup_{t\to\infty} I(t) \leq 83.33 = \frac{\alpha_2}{\beta_2}.$$

That is, $I(t)$ *will eventually enter the region (0, 83.33) if* $I(0)$ *is not in this region, and will be attracted in this region once it has entered. Also, by Theorem 4.8,* $I(t)$ *can take any value up to the boundaries of (0, 83.33) but never reach them.*

The computer simulations in Figure 4.3 support this result clearly.

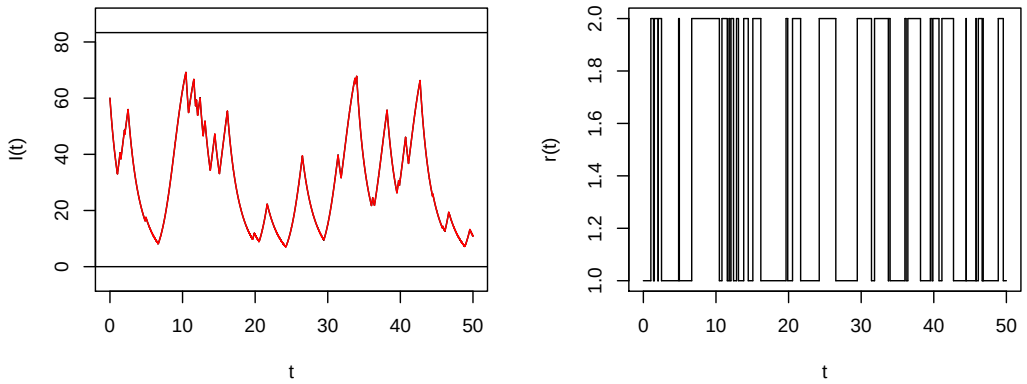


Figure 4.3: Computer simulation of $I(t)$ using the parameter values in Example 4.12 and its corresponding Markov chain $r(t)$, using formula (4.2.5) (black line) and the EM method with step size $\Delta = 0.001$ (red line) for $I(t)$, with $I(0) = 60$, and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$. (The two lines are very close to each other, so we hardly see the black line in the plot.) The horizontal lines in the plot of $I(t)$ indicate the levels 0 and $\frac{\alpha_2}{\beta_2}$.

4.6.3 T_0^S=1 Case

Example 4.13 *Assume that the system parameters are given by*

$$\mu_1 = 0.45, \mu_2 = 0.05, \gamma_1 = 0.35, \gamma_2 = 0.15, \beta_1 = 0.006, \beta_2 = 0.005, N = 100,$$

$$\nu_{12} = 0.6, \text{ and } \nu_{21} = 0.9.$$

So $\alpha_1 = -0.2$*,* $\alpha_2 = 0.3$*,* $\pi_1 = 0.6$*, and* $\pi_2 = 0.4$*.*

Note that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = 0$$

in this case, which is equivalent to $T_0^S = 1$*. As mentioned in section 4.4, we have not been able to prove the behaviour of* $I(t)$ *in this case. However, the simulation results in Figure 4.4 confirm our suspicion that the disease will always become extinct.*

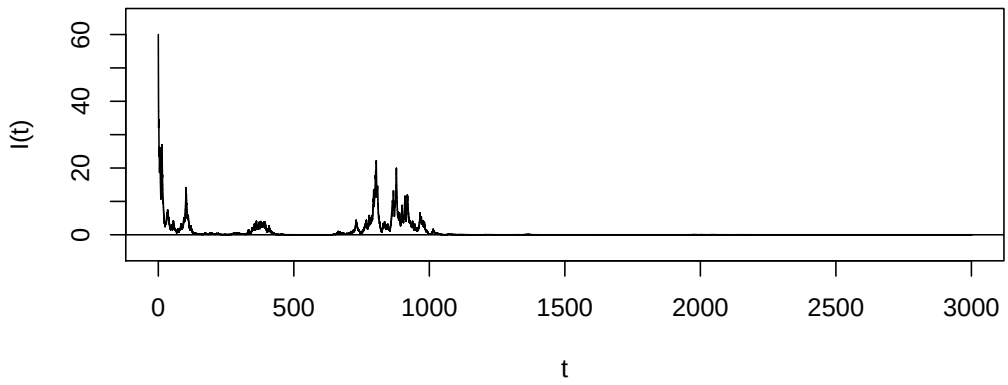


Figure 4.4: Computer simulation of $I(t)$ using the parameter values in Example 4.13, using formula (4.2.5) for $I(t)$, with $I(0) = 60$, and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

4.7 Generalisation

We have discussed the simplest case where the Markov chain has only two states, in the previous sections. Now we are going to generalise the results to the case where the Markov chain $r(t)$ has finite state space $\mathbb{S} = \{1, 2, ..., M\}$. The generator for $r(t)$ is defined as

$$\Gamma = (\nu_{ij})_{M\times M},$$

where $\nu_{ii} = -\sum_{1\le j\le M, j\neq i} \nu_{ij}$, and $\nu_{ij} > 0$ $(i \neq j)$ is the transition rate from state i to j, that is

$$\mathbb{P}\{r(t+\delta) = j | r(t) = i\} = \nu_{ij}\delta + o(\delta),$$

where $\delta > 0$. As before, there is a sequence $\{\tau_k\}_{k\ge 0}$ of finite-valued $\mathcal{F}_t$-stopping times such that $0 = \tau_0 < \tau_1 < \cdots < \tau_k \to \infty$ almost surely and

$$r(t) = \sum_{k=0}^{\infty} r(\tau_k) I_{[\tau_k, \tau_{k+1})}(t).$$

Moreover, given that $r(\tau_k) = i$, the random variable $\tau_{k+1} - \tau_k$ follows the exponential distribution with parameter $-\nu_{ii}$, namely

$$\mathbb{P}(\tau_{k+1} = j | \tau_k = i) = \frac{\nu_{ij}}{-\nu_{ii}}, \quad j \neq i, \quad \mathbb{P}(\tau_{k+1} - \tau_k \ge T | r(\tau_k) = i) = e^{\nu_{ii}T}, \quad \forall T \ge 0.$$

Furthermore, the unique stationary distribution of this Markov chain $\Pi = (\pi_1, \pi_2, ..., \pi_M)$ satisfies

$$\begin{cases} \Pi\Gamma = 0 \\ \sum_{i=1}^{M} \pi_i = 1. \end{cases}$$

Following a similar procedure we still can show that for any given initial value $I(0) = I_0 \in (0, N)$, there is a unique solution $I(t)$ on $t \in \mathbb{R}_+$ to equation (4.2.3) such that

$$\mathbb{P}(I(t) \in (0, N) \text{ for all } t \ge 0) = 1,$$

and the solution still has the form (4.2.5).

In the general finite state space Markov chain case it is possible to derive an explicit expression for the basic reproduction number R_0^S in the stochastic Markov switching model analogous to (4.3.1) expressed as the largest eigenvalue of a positive matrix. We define T_0^S for the general case as

$$T_0^S = \frac{\sum_{k=1}^M \pi_k \beta_k N}{\sum_{k=1}^M \pi_k (\mu_k + \gamma_k)}.$$

Similarly to Proposition 4.2, we have the following alternative conditions on the value of T_0^S:

Proposition 4.14 *We have the following alternative condition on the value of T_0^S:*

- $T_0^S < 1$ *if and only if* $\sum_{k=1}^M \pi_k \alpha_k < 0$;
- $T_0^S = 1$ *if and only if* $\sum_{k=1}^M \pi_k \alpha_k = 0$;
- $T_0^S > 1$ *if and only if* $\sum_{k=1}^M \pi_k \alpha_k > 0$.

If $T_0^S < 1$, similarly to Theorem 4.3, we can show:

Theorem 4.15 *If $T_0^S < 1$, for any given initial value $I_0 \in (0, N)$, the solution of the stochastic SIS model (4.2.3) obeys*

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \leq \sum_{k=1}^M \pi_k \alpha_k \quad a.s.$$

By the more general condition stated above, we hence conclude that $I(t)$ tends to zero exponentially almost surely. This means that the disease dies out with probability one.

For the case that $T_0^S > 1$, Theorem 4.4 can be generalised as follows:

Theorem 4.16 *If $T_0^S > 1$, for any given initial value $I_0 \in (0, N)$, the solution of the stochastic SIS model (4.2.3) has the properties that*

$$\liminf_{t\to\infty} I(t) \leq \frac{\sum_{k=1}^M \pi_k \alpha_k}{\sum_{k=1}^M \pi_k \beta_k} \quad a.s.$$

and

$$\limsup_{t\to\infty} I(t) \geq \frac{\sum_{k=1}^M \pi_k \alpha_k}{\sum_{k=1}^M \pi_k \beta_k} \quad a.s.,$$

which means the disease will reach the neighbourhood of the level $\frac{\sum_{k=1}^M \pi_k \alpha_k}{\sum_{k=1}^M \pi_k \beta_k}$ infinitely many times with probability one. This shows that the disease will be persistent in this case.

Lemma 4.5 can be generalised as follows:

Lemma 4.17 *Without loss of generality we assume that* $\alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$ *and the following statements hold with probability one:*

(i) If $0 < \alpha_1/\beta_1 = \alpha_2/\beta_2 = ... = \alpha_M/\beta_M$, *then* $I(t) = \alpha_1/\beta_1$ *for all* $t > 0$ *when* $I_0 = \alpha_1/\beta_1$.

(ii) If $0 < \alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$, *then* $I(t) \in (\alpha_1/\beta_1, \alpha_M/\beta_M)$ *for all* $t > 0$ *whenever* $I_0 \in (\alpha_1/\beta_1, \alpha_M/\beta_M)$.

(iii) If $\alpha_j/\beta_j \leq 0$ *(for some* $j \in (1, M-1)$*) and* $\alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$ *then* $I(t) \in (0, \alpha_M/\beta_M)$ *for all* $t > 0$ *whenever* $I_0 \in (0, \alpha_M/\beta_M)$.

Theorem 4.6 can be generalised as follows:

Theorem 4.18 *Assume that* $T_0^S > 1$ *and let* $I_0 \in (0, N)$ *be arbitrary. The following statements hold with probability one:*

(i) If $0 < \alpha_1/\beta_1 = \alpha_2/\beta_2 = ... = \alpha_M/\beta_M$, *then* $\lim_{t\to\infty} I(t) = \alpha_1/\beta_1$.

(ii) If $0 < \alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$, *then*

$$\frac{\alpha_1}{\beta_1} \leq \liminf_{t\to\infty} I(t) \leq \limsup_{t\to\infty} I(t) \leq \frac{\alpha_M}{\beta_M}.$$

(iii) If $\alpha_j/\beta_j \leq 0$ *(for some* $j \in (1, M-1)$*) and* $\alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$, *then*

$$0 \leq \liminf_{t\to\infty} I(t) \leq \limsup_{t\to\infty} I(t) \leq \frac{\alpha_M}{\beta_M}.$$

These stronger results indicate that $I(t)$ *will enter the region* $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ *in finite time and with probability one will stay in this region once it is entered.*

Theorem 4.7 can be generalised as follows:

Theorem 4.19 *Assume that* $T_0^S > 1$ *and* $0 < \alpha_1/\beta_1 \leq \alpha_2/\beta_2 \leq ... \leq \alpha_M/\beta_M$, *and let* $I_0 \in (0, N)$ *be arbitrary. Then for any* $\varepsilon > 0$, *sufficiently small for*

$$\frac{\alpha_1}{\beta_1} + \varepsilon < \frac{\sum_{k=1}^M \pi_k \alpha_k}{\sum_{k=1}^M \pi_k \beta_k} < \frac{\alpha_M}{\beta_M} - \varepsilon,$$

the solution of the stochastic SIS model (4.2.3) has the properties that

$$\mathbb{P}\Big(\liminf_{t\to\infty} I(t) < \frac{\alpha_1}{\beta_1} + \varepsilon \Big) \geq e^{\nu_{11} T_1(\varepsilon)},$$

and

$$\mathbb{P}\Big(\limsup_{t\to\infty} I(t) > \frac{\alpha_M}{\beta_M} - \varepsilon \Big) \geq e^{\nu_{MM} T_2(\varepsilon)},$$

where $T_1(\varepsilon) > 0$ *and* $T_2(\varepsilon) > 0$ *are defined by*

$$T_1(\varepsilon) = \frac{1}{\alpha_1}\Big(\log\Big(\frac{\beta_1}{\alpha_1} - \frac{\beta_M}{\alpha_M}\Big) + \log\Big(\frac{\alpha_1}{\beta_1} + \varepsilon\Big) - \log\Big(\varepsilon\frac{\beta_1}{\alpha_1}\Big)\Big) \tag{4.7.1}$$

and

$$T_2(\varepsilon) = \frac{1}{\alpha_M}\Big(\log\Big(\frac{\beta_1}{\alpha_1} - \frac{\beta_M}{\alpha_M}\Big) + \log\Big(\frac{\alpha_M}{\beta_M} - \varepsilon\Big) - \log\Big(\varepsilon\frac{\beta_M}{\alpha_M}\Big)\Big). \tag{4.7.2}$$

Also, Theorem 4.8 can be generalised as follows:

Theorem 4.20 *Assume that* $T_0^S > 1$*, that is* $\sum_{k=1}^M \pi_k\alpha_k > 0$*, and* $\alpha_j/\beta_j \le 0$ *(for some* $j \in (1, M-1)$*). Let* $I_0 \in (0, N)$ *be arbitrary. Then for any* $\varepsilon > 0$*, sufficiently small for*

$$\varepsilon < \frac{\sum_{k=1}^M \pi_k\alpha_k}{\sum_{k=1}^M \pi_k\beta_k} < \frac{\alpha_M}{\beta_M} - \varepsilon,$$

the solution of the stochastic SIS model (4.2.3) has the properties that

$$\mathbb{P}\Big(\liminf_{t\to\infty} I(t) < \varepsilon\Big) \ge e^{\nu_{11}T_3(\varepsilon)},$$

and

$$\mathbb{P}\Big(\limsup_{t\to\infty} I(t) > \frac{\alpha_M}{\beta_M} - \varepsilon\Big) \ge e^{\nu_{MM}T_4(\varepsilon)}.$$

Here $T_3(\varepsilon) > 0$ *and* $T_4(\varepsilon) > 0$ *are defined by*

$$T_3(\varepsilon) = \frac{1}{\alpha_1}\Big(\log\Big(\frac{\beta_M}{\alpha_M} - \frac{\beta_1}{\alpha_1}\Big) + \log\Big(\varepsilon\frac{\alpha_1}{\beta_1}\Big) - \log\Big(\frac{\alpha_1}{\beta_1} - \varepsilon\Big)\Big) \tag{4.7.3}$$

and

$$T_4(\varepsilon) = \frac{1}{\alpha_M}\Big(\log\Big(\frac{2}{\varepsilon} - \frac{\beta_M}{\alpha_M}\Big) + \log\Big(\frac{\alpha_M}{\beta_M} - \varepsilon\Big) - \log\Big(\varepsilon\frac{\beta_M}{\alpha_M}\Big)\Big). \tag{4.7.4}$$

Theorem 4.19 and Theorem 4.20 show that $I(t)$ will take any value arbitrarily close to the boundaries $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ but never reach them.

The proofs are all very similar to the simple case, so they are omitted here.

To prove (4.7.4) analogously to the simple case we define the stopping times

$$\sigma_5 = \inf\{t \ge T : r(t) = M\}$$

where $T > 0$ is arbitrary and

$$\sigma_6 = \inf\left\{ t \ge \sigma_5 : r(t) = M,\ I(t) \ge \frac{1}{2}\left(\frac{\sum_{k=1}^M \pi_k\alpha_k}{\sum_{k=1}^M \pi_k\beta_k}\right)\right\}.$$

By Theorem 4.16 if $I(t)$ ever goes beneath $\frac{1}{2}\frac{\sum_{k=1}^M \pi_k\alpha_k}{\sum_{k=1}^M \pi_k\beta_k}$ it will eventually increase above this level. Hence $I(t)$ is above this level when the Markov chain switches state infinitely

often. Each time that this happens it is either initially in state M, or switches to state M with probability at least

$$q = \min_{n\in[1,2,\ldots M-1]} \frac{\nu_{nM}}{-\nu_{nn}} > 0.$$

Therefore each time after σ_5 that $I(t)$ reaches the level $\frac{1}{2}\frac{\sum_{k=1}^{M}\pi_k\alpha_k}{\sum_{k=1}^{M}\pi_k\beta_k}$ we will have a value of $t \geq \sigma_5$ with $r(t) = M$ and $I(t)$ above the level $\frac{1}{2}\frac{\sum_{k=1}^{M}\pi_k\alpha_k}{\sum_{k=1}^{M}\pi_k\beta_k}$ with probability at least q. So considering the first X times after σ_5 that $I(t)$ reaches this level

$$P(\sigma_6 < \infty) \geq 1 - (1-q)^X.$$

Letting $X \to \infty$ we deduce that $P(\sigma_6 < \infty) = 1$. The proof proceeds as in the simple case.

4.8 A Slightly More Realistic Example

As a slightly more realistic example to illustrate the two state case, we consider *S. pneumoniae* amongst children under two years in Scotland. This may display a phenomenon called capsular switching, such that when an individual is co-infected with two strains (or serotypes) of pneumococcus, the outer polysaccharide capsule that surrounds the genetic pneumococcal material may switch, thus giving serotypes with possibly different infectivities and infectious periods [14, 20]. In reality the situation is very complicated, with many pneumococcal serotypes and sequence types (sequence types are ways of coding the genetic material). This is thought to be due to genetic transfer of material between the two serotypes.

Example 4.21 *We illustrate our model by applying it with suitable parameter values to two strains of pneumococcus with switching between them, although the real situation is much more complicated than the model allows. The parameter values used are taken from Lamb, Greenhalgh and Robertson [57] as:*

N = 150,000, $\gamma_1 = \gamma_2$ = 1/(7.1 wk) = 0.1408/wk = 0.02011/day [93],
μ = 1/(104 wk) = 9.615 $\times 10^{-3}$/wk = 1.3736 $\times 10^{-3}$/day,
β_1 = 1.5041 $\times 10^{-6}$/wk = 2.1486 $\times 10^{-7}$/day corresponding to R_{01}^D = 1.5 [27],
β_2 = 2.0055 $\times 10^{-6}$/wk = 2.8650 $\times 10^{-7}$/day corresponding to $R_{02}^D = 2$ [96].
So $\alpha_1 = 0.0107454$/day and $\alpha_2 = 0.0214914$/day. We set

$$\nu_{12} = 0.06/day \text{ and } \nu_{21} = 0.09/day.$$

So $\pi_1 = 0.6$, and $\pi_2 = 0.4$.

From these values, T_0^S is about 1.7 in this case. Noting that

$$\alpha_1\pi_1 + \alpha_2\pi_2 = 0.0150438 > 0,$$

we can therefore conclude, by Theorem 4.6, that for any given initial value $I(0) = I_0 \in (0, N)$*, the solution of (4.2.3) obeys*

$$\frac{\alpha_1}{\beta_1} = 50011.17 \leq \liminf_{t\to\infty} I(t) \leq \limsup_{t\to\infty} I(t) \leq 75013.61 = \frac{\alpha_2}{\beta_2}.$$

That is, $I(t)$ *will eventually enter the region (50,011.17, 75,013.61) if* $I(0)$ *is not in this region, and will be attracted in this region once it has entered. The computer simulations in Figure 4.5 support this result clearly.*

We vary the values for the transition rates ν_{12} *and* ν_{21}*. Figure 4.6 shows how the different values of the transition rates affect the behaviour of* $I(t)$*. We notice that it takes longer to switch between the two states when the transition rates are small, so* $I(t)$ *is more likely to approach the boundaries.*

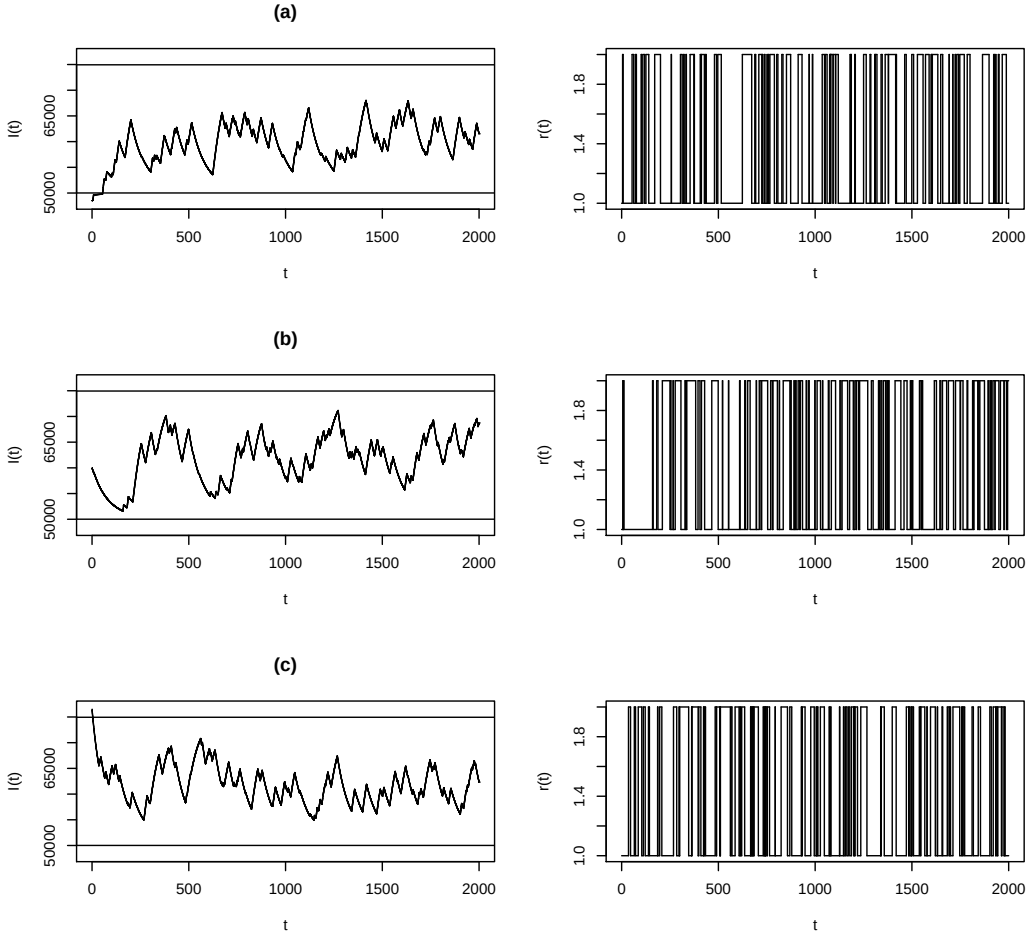


Figure 4.5: Computer simulation of $I(t)$ using the parameter values in Example 4.21 and its corresponding Markov chain $r(t)$, using formula (4.2.5) for $I(t)$, with $I(0) = 48,500$ for (a), $I(0) = 60,000$ for (b) and $I(0) = 76,500$ for (c), and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$. The horizontal lines in the plot of $I(t)$ indicate the levels $\frac{\alpha_1}{\beta_1}$ and $\frac{\alpha_2}{\beta_2}$.

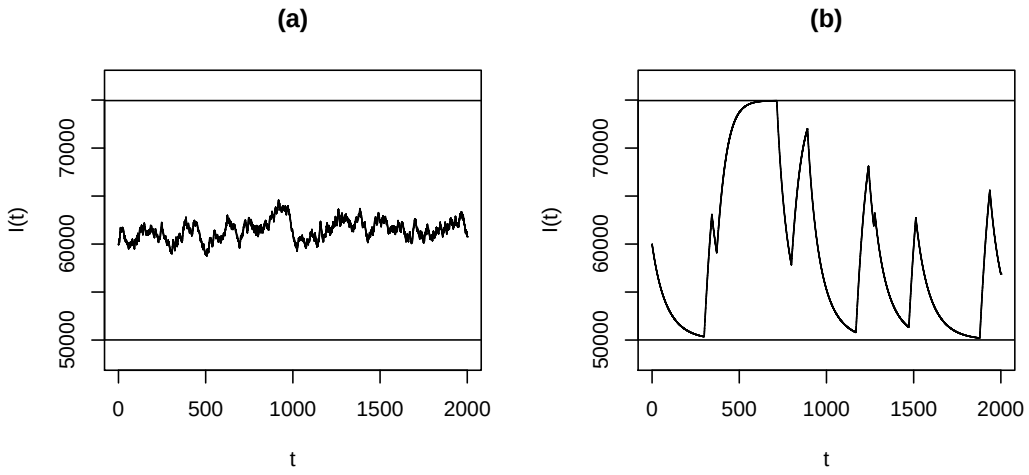


Figure 4.6: Computer simulation of $I(t)$ using the parameter values in Example 4.21, with $\nu_{12} = 0.6$/day and $\nu_{21} = 0.9$/day for (a), and $\nu_{12} = 0.006$/day and $\nu_{21} = 0.009$/day for (b), using formula (4.2.5) for $I(t)$ with $I(0) = 60,000$, and the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$. The horizontal lines in the plot of $I(t)$ indicate the levels $\frac{\alpha_1}{\beta_1}$ and $\frac{\alpha_2}{\beta_2}$ (which the values of $I(t)$ never quite reach).

4.9 Summary

In this chapter, we have introduced telegraph noise to the classical SIS epidemic model and set up the corresponding stochastic SIS model. We have established the explicit solution for the stochastic SIS model and also established conditions for extinction and persistence of the disease. For the stochastic Markov switching model a threshold value T_0^S was defined for almost sure persistence or extinction. We started with the special case in which the Markov chain has only two states and then generalised our theory to the general case where the Markov chain has M states. Theorem 4.15 shows that if $T_0^S < 1$, the disease will die out. Theorem 4.16 shows that if $T_0^S > 1$, then the disease will persist. We also showed Theorem 4.18 that if $T_0^S > 1$ the number of infectious individuals will enter $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ in finite time, and with probability one will stay in the interval once entered, and moreover the number of infectious individuals can take any value up to the boundaries of $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ but never reach them (Theorems 4.19 and 4.20).

For $j = 1, 2, \ldots M$, define $R_{0j}^D = \frac{\beta_j N}{\mu_j + \gamma_j}$. Note that if $\alpha_j > 0$ then $R_{0j}^D > 1$ and

$$\frac{\alpha_j}{\beta_j} = N\left(1 - \frac{1}{R_{0j}^D}\right)$$

is the long-term endemic level of disease in the SIS model (1.2.1) with $\beta = \beta_j,\ \mu = \mu_j$ and $\gamma = \gamma_j$. If $\alpha_j \leq 0$ then $R_{0j}^D \leq 1$ and disease eventually dies out in the same SIS model. Hence $0 \vee (\alpha_1/\beta_1)$ is the smallest and α_M/β_M is the largest long-term endemic level of disease in each of the M separate SIS models between which the Markov chain switches.

We have not been able to prove extinction for the case when $T_0^S = 1$, but the computer simulation shows that the disease would die out after a long period of time, as we suspect. We have illustrated our theoretical results with computer simulations, including an example with realistic parameter values for *S.Pneumoniae* amongst children.

Most of the contents of this chapter have been published in [40]. Next, we are going to examine the combined effect of incorporating both white noise and the noise in the form of Markovian switching in the SIS epidemic model.

Chapter 5

A Stochastic Differential Equation SIS Epidemic Model with Markovian Switching

5.1 Introduction

In this chapter, we still consider the SIS epidemic model (1.2.1). The introduction of the SIS epidemic model (1.2.1) can be found in section 1.2. The solutions of the system is shown in (1.2.2) and the basic reproduction number R_0^D is defined in (1.2.3). In Chapters 3 and 4, we examined the effect of white noise and telegraph noise respectively in the SIS epidemic model. We found that the introduction of stochastic noise changes the basic reproduction number of the original deterministic SIS model (1.2.1) in both cases. We take a further step in this chapter to examine the effect of taking both white and telegraph noise into account simultaneously. In section 1.3.4, we reviewed some recent research which considered combining the influence of both types of environmental noise into modelling and shows the significant effect of combined noise on the deterministic system.

We now briefly outline how we obtain the stochastic SIS model incorporating both types of noise as follows: We first use the parameter perturbation technique to introduce the white noise to (1.2.1) as in Chapter 3, so the deterministic SIS model (1.2.1) becomes the Itô SDE (3.2.3). Next we consider the telegraph noise. Recall that in Chapter 4 we model the regime switching by a finite-state Markov chain. Now we take into account the telegraph noise in the SDE SIS model (3.2.3). Assuming that there are M different environmental regimes operating within the population system, the system parameters μ, β, γ and σ may experience abrupt changes between regimes. As a result, the SDE SIS model (3.2.3) becomes a SDE SIS model with Markovian switching of the form

$$\begin{cases} dS(t) = [\mu_{r(t)}N - \beta_{r(t)}S(t)I(t) + \gamma_{r(t)}I(t) - \mu_{r(t)}S(t)]dt - \sigma_{r(t)}S(t)I(t)dB(t), \\ dI(t) = [\beta_{r(t)}S(t)I(t) - (\mu_{r(t)} + \gamma_{r(t)})I(t)]dt + \sigma_{r(t)}S(t)I(t)dB(t), \end{cases} \tag{5.1.1}$$

where $r(t)$ is a Markov chain on the state space $\mathbb{S} = \{1, 2, ..., M\}$ and has the same definition as in section 4.7.

The main aim of this chapter is to discuss the effect of the two noise types, namely white noise and the noise in the form of Markov switching, adding to the system simultaneously as in model (5.1.1). We will not only show the existence of a unique global positive solution (section 5.2) but also examine the asymptotic properties, including extinction (section 5.3) and persistence (section 5.4). Computer simulations will be performed to illustrate the theory and the conclusions will be drawn in section 5.5.

5.2 Stochastic Differential Equation SIS Model with Markovian Switching

We assume that the system parameters β_i, μ_i, γ_i and σ_i ($i \in \mathbb{S}$) are all positive numbers. Given that $I(t) + S(t) = N$, we see that $I(t)$, the number of infectious individuals with initial value $I(0) = I_0 \in (0, N)$, obeys

$$dI(t) = I(t)\Big([\alpha_{r(t)} - \beta_{r(t)}I(t)]dt + \sigma_{r(t)}(N - I(t))dB(t)\Big) \tag{5.2.1}$$

where $\alpha_i := \beta_i N - \mu_i - \gamma_i$. It is sufficient to study equation (5.2.1) for understanding the behaviour of the stochastic SIS epidemic model with Markovian switching in (5.1.1). So in the following sections we will concentrate on this equation only.

As $I(t)$ in (5.2.1) is the number of infected people at time t, for the model to make sense, we need to show that this SDE SIS model with Markovian switching not only has a unique global solution but also that the solution will remain within $(0, N)$ if it starts from there. The existing general existence-and-uniqueness theorem on SDEs with Markovian switching (Theorem 2.20) is not applicable to our model (5.2.1). Therefore we use the following theorem to show the required property for our model (5.1.1).

Theorem 5.1 *For any given initial value $I(0) = I_0 \in (0, N)$, there is a unique global positive solution $I(t) \in (0, N)$ for all $t \geq 0$ to equation (5.2.1) such that*

$$\mathbb{P}\{I(t) \in (0, N) \text{ for all } t \geq 0\} = 1.$$

Proof. Since the coefficients of the equation (5.2.1) are locally Lipschitz piecewise continuous, it is known (see e.g. [64]) that for any given initial value $I_0 \in (0, N)$ there is a unique maximal local solution $I(t)$ on $t \in [0, \tau_e)$, where τ_e is the explosion time. Let $k_0 > 0$ be sufficiently large so that I_0 lies in the interval $[1/k_0, N - 1/k_0]$. For each integer $k \geq k_0$, we define the stopping time

$$\tau_k = \inf\{t \in [0, \tau_e) : I(t) \notin (1/k,\ N - (1/k))\}.$$

Clearly, τ_k increases as $k \to \infty$. We set $\tau_\infty = \lim_{k\to\infty} \tau_k$, whence $\tau_\infty \leq \tau_e$ a.s. If we can show that $\tau_\infty = \infty$ a.s., then $\tau_e = \infty$ a.s. and $I(t) \in (0, N)$ a.s. for all $t \geq 0$. In other words, to complete the proof all we need to show is that $\tau_\infty = \infty$ a.s. If this statement is false, then there is a pair of constants $T > 0$ and $\varepsilon \in (0, 1)$ such that

$$\mathbb{P}\{\tau_\infty \leq T\} > \varepsilon.$$

Hence there is an integer $k_1 \geq k_0$ such that

$$\mathbb{P}\{\tau_k \leq T\} \geq \varepsilon \quad \text{for all } k \geq k_1. \tag{5.2.2}$$

We define a function $V : (0, N) \times [0, \infty) \to [0, \infty)$ by

$$V(x,t) = \frac{1}{x} + \frac{1}{N-x}.$$

By the Itô formula (Theorem 2.9), we have, for any $t \in [0,T]$ and $k \geq k_1$,

$$\mathbb{E}V(I(t \wedge \tau_k), t) = V(I_0, t_0) + \mathbb{E}\int_0^{t\wedge\tau_k} LV(I(s), s)ds, \tag{5.2.3}$$

where $LV : (0, N) \times [0, \infty) \to \mathbb{R}$ is defined by

$$\begin{aligned} LV(x,t) &= x\bigg(-\frac{1}{x^2} + \frac{1}{(N-x)^2}\bigg)[\beta_{r(t)}N - \mu_{r(t)} - \gamma_{r(t)} - \beta_{r(t)}x] \\ &+ \sigma^2_{r(t)}x^2(N-x)^2\bigg(\frac{1}{x^3} + \frac{1}{(N-x)^3}\bigg). \end{aligned} \tag{5.2.4}$$

We shall use $\hat{f}$ to denote $\min_{i\in\mathbb{S}} f_i$ and $\check{f}$ to denote $\max_{i\in\mathbb{S}} f_i$, where $\{f_i\}_{i\in\mathbb{S}}$ is a constant vector. It is easy to show that

$$LV(x,t) \leq \frac{\check{\mu} + \check{\gamma}}{x} + \frac{\check{\beta}N}{N-x} + \check{\sigma}^2N^2\bigg(\frac{1}{x} + \frac{1}{N-x}\bigg) \leq K_1V(x,t), \tag{5.2.5}$$

where $K_1 = (\check{\mu} + \check{\gamma}) \vee (\check{\beta}N) + \check{\sigma}^2N^2$. From here we can show that $\tau_\infty = \infty$ a.s. in the same way as in the proof of Theorem 3.1. So the proof is complete.

5.3 Extinction

We will discuss extinction conditions for our model (5.2.1) in this section. It was shown in Chapter 3 that for the SDE SIS epidemic model (3.2.3), if $R_0^S = \frac{\beta N}{\mu+\gamma} - \frac{\sigma^2N^2}{2(\mu+\gamma)} < 1$, under mild extra conditions, the disease will die out, and if $R_0^S > 1$ then the disease will persist. Naturally we wish to establish a similar theory on extinction and persistence for the SDE SIS epidemic model with Markovian switching (5.2.1) in terms of a corresponding threshold parameter. For the SIS epidemic model with Markovian switching in Chapter 4, we defined the threshold

$$T_0^S = \frac{\sum_{i=1}^M \pi_i\beta_iN}{\sum_{i=1}^M \pi_i(\mu_i + \gamma_i)}.$$

We showed that the disease would die out for $T_0^S < 1$ and persist if $T_0^S > 1$. Recalling the definition of R_0^S above for the SDE SIS epidemic model (3.2.3), we would attempt to define the corresponding T_0^{MC} by

$$T_0^{MC} = \frac{\sum_{i=1}^M \pi_i(\beta_iN - 0.5\sigma_i^2N^2)}{\sum_{i=1}^M \pi_i(\mu_i + \gamma_i)}. \tag{5.3.1}$$

In the rest of the chapter we examine this threshold (5.3.1) for almost sure extinction and stochastic permanence of our model (5.2.1). Before we show the conditions for extinction, we state a proposition which gives an equivalent condition on the value of T_0^{MC}.

Proposition 5.2 *We have the following alternative conditions on the value of* T_0^{MC}*:*

- $T_0^{MC} < 1$ *if and only if* $\sum_{i=1}^{M} \pi_i \left(\alpha_i - \frac{1}{2}\sigma_i^2 N^2\right) < 0$*;*
- $T_0^{MC} = 1$ *if and only if* $\sum_{i=1}^{M} \pi_i \left(\alpha_i - \frac{1}{2}\sigma_i^2 N^2\right) = 0$*;*
- $T_0^{MC} > 1$ *if and only if* $\sum_{i=1}^{M} \pi_i \left(\alpha_i - \frac{1}{2}\sigma_i^2 N^2\right) > 0$.

The proof of this proposition is straightforward, so we omit it here. We now state our theorems on extinction.

Theorem 5.3 *If*

$$T_0^{MC} < 1 \quad \textit{and} \quad \sigma_i^2 \leq \frac{\beta_i}{N},\ i \in \mathbb{S}, \tag{5.3.2}$$

then, for any given initial value $I(0) = I_0 \in (0, N)$*, the solution of the SDE (5.2.1) obeys*

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \leq \sum_{i=1}^{M} \pi_i \left(\alpha_i - \frac{1}{2}\sigma_i^2 N^2\right) < 0 \quad a.s., \tag{5.3.3}$$

namely, $I(t)$ *tends to zero exponentially almost surely. In other words, the disease dies out with probability one.*

Proof. By the Itô formula, we have

$$\log(I(t)) = \log(I_0) + \int_0^t f(I(s), s, r(s))ds + \int_0^t \sigma_{r(s)}(N - I(s))dB(s), \tag{5.3.4}$$

where $f : (0, N) \times [0, \infty) \times \mathbb{S} \to \mathbb{R}$ is defined by

$$f(x, t, i) = \alpha_i - \beta_i x - \frac{1}{2}\sigma_i^2 (N - x)^2. \tag{5.3.5}$$

However, under condition (5.3.2), we have

$$\begin{aligned} f(I(s), s, r(s)) &= \alpha_{r(s)} - \frac{1}{2}\sigma_{r(s)}^2 N^2 - (\beta_{r(s)} - \sigma_{r(s)}^2 N)I(s) - \frac{1}{2}\sigma_{r(s)}^2 I^2(s), \\ &\leq \alpha_{r(s)} - \frac{1}{2}\sigma_{r(s)}^2 N^2, \end{aligned}$$

for $I(s) \in (0, N)$. It then follows from (5.3.4) that

$$\log(I(t)) \leq \log(I_0) + \int_0^t \left(\alpha_{r(s)} - \frac{1}{2}\sigma_{r(s)}^2 N^2\right) ds + \int_0^t \sigma_{r(s)}(N - I(s))dB(s). \tag{5.3.6}$$

By the ergodic theory we have that

$$\limsup_{t\to\infty}\frac{1}{t}\int_0^t\left(\alpha_{r(s)}-\frac{1}{2}\sigma_{r(s)}^2N^2\right)ds=\sum_{1=1}^{M}\pi_i\left(\alpha_i-\frac{1}{2}\sigma_i^2N^2\right)\quad a.s.$$

Therefore from (5.3.6)

$$\limsup_{t\to\infty}\frac{1}{t}\log(I(t))\le\sum_{1=1}^{M}\pi_i\left(\alpha_i-\frac{1}{2}\sigma_i^2N^2\right)+\limsup_{t\to\infty}\frac{1}{t}\int_0^t\sigma_{r(s)}(N-I(s))dB(s)\quad a.s.\tag{5.3.7}$$

But note that

$$\left|\int_0^t\sigma_{r(s)}(N-I(s))dB(s)\right|^2\le\int_0^t\check{\sigma}^2N^2ds=\check{\sigma}^2N^2t.$$

So

$$\left|\frac{1}{t}\int_0^t\sigma_{r(s)}(N-I(s))dB(s)\right|^2\le\frac{\check{\sigma}^2N^2}{t}\to0\quad\text{as}\quad t\to\infty.$$

Hence we have

$$\limsup_{t\to\infty}\frac{1}{t}\int_0^t\sigma_{r(s)}(N-I(s))dB(s)=0.$$

By Proposition 5.2, we therefore obtain the desired assertion (5.3.3) from (5.3.7).

Example 5.4 *We only consider the case where the Markov chain has $M=2$ states in our simulation examples. With the same units we assumed in Chapter 3, the system parameters are given by*

$$\mu_1=0.95,\mu_2=0.55,\gamma_1=0.95,\gamma_2=0.55,\beta_1=0.02,\beta_2=0.01,\sigma_1=0.01,$$

$$\sigma_2=0.005,N=100,\nu_{12}=0.8\ \textit{and}\ \nu_{21}=0.5.$$

So $\alpha_1-\frac{1}{2}\sigma_1^2N^2=-0.40$, $\alpha_2-\frac{1}{2}\sigma_2^2N^2=-0.225$, $\pi_1=0.3846$, and $\pi_2=0.6154$ (see section 4.2 and 4.7 for definitions).

Noting that

$$\left(\alpha_1-\frac{1}{2}\sigma_1^2N^2\right)\pi_1+\left(\alpha_2-\frac{1}{2}\sigma_2^2N^2\right)\pi_2=-0.2923,$$

and

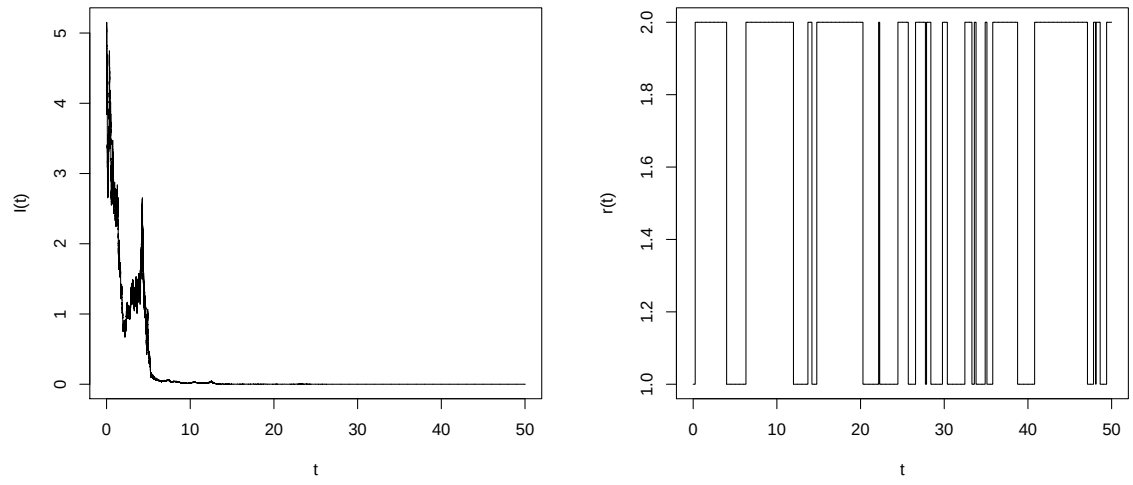
$$\sigma_1^2=1\times10^{-4}<\frac{\beta_1}{N}=2\times10^{-4},\ \sigma_2^2=2.5\times10^{-5}<\frac{\beta_2}{N}=1\times10^{-4},$$

we can therefore conclude by Theorem 5.3 that for any given initial value $I(0)=I_0\in(0,N)$, the solution of (5.2.1) obeys

$$\limsup_{t\to\infty}\frac{1}{t}\log(I(t))\le-0.2923<0\ \ a.s.$$

That is, $I(t)$ will tend to zero exponentially with probability one.

The computer simulation in Figure 5.1(a), using the EM method supports this result clearly, illustrating extinction of the disease. We also note that $\alpha_1-\frac{1}{2}\sigma_1^2N^2<0$ with



(a) (b)

Figure 5.1: (a) Computer simulation of the path $I(t)$ for the SDE SIS model with Markovian switching (5.2.1) using the EM method with step size $\Delta t = 0.001$, using parameter values from Example 5.4 and initial value $I(0) = 5$; (b) the corresponding Markov chain $r(t)$, using the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

$\sigma_1^2 < \frac{\beta_1}{N}$ while $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2 < 0$ with $\sigma_2^2 < \frac{\beta_2}{N}$ in this case, which means that both subsystems die out (see Theorem 3.2). The overall system is of course extinct in this case. In the next example we will examine the situation where one subsystem becomes extinct and another one persists.

Example 5.5 *Assume that the system parameters are given by*

$$\mu_1 = 0.95, \mu_2 = 0.55, \gamma_1 = 0.95, \gamma_2 = 0.55, \beta_1 = 0.02, \beta_2 = 0.016, \sigma_1 = 0.01,$$

$$\sigma_2 = 0.005, N = 100, \nu_{12} = 0.5 \text{ and } \nu_{21} = 0.8.$$

So $\alpha_1 - \frac{1}{2}\sigma_1^2 N^2 = -0.40$, $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2 = 0.375$, $\pi_1 = 0.6154$, and $\pi_2 = 0.3846$.

Noting that

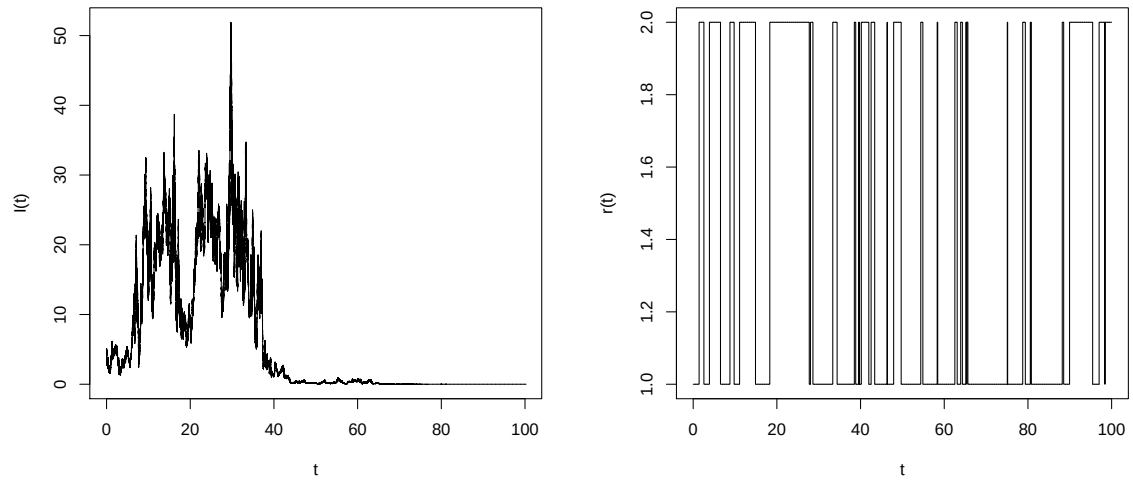
$$\left(\alpha_1 - \frac{1}{2}\sigma_1^2 N^2\right)\pi_1 + \left(\alpha_2 - \frac{1}{2}\sigma_2^2 N^2\right)\pi_2 = -0.1019,$$

and

$$\sigma_1^2 = 1 \times 10^{-4} < \frac{\beta_1}{N} = 2 \times 10^{-4}, \ \sigma_2^2 = 2.5 \times 10^{-5} < \frac{\beta_2}{N} = 1.6 \times 10^{-4},$$

we can therefore conclude, by Theorem 5.3, that for any given initial value $I(0) = I_0 \in (0, N)$, the solution of (5.2.1) obeys

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le -0.1019 < 0 \quad a.s.$$



(a) (b)

Figure 5.2: (a) Computer simulation of the path $I(t)$ for the SDE SIS model with Markovian switching (5.2.1) using the EM method with step size $\Delta t = 0.001$, using parameter values from Example 5.5 and initial value $I(0) = 5$; (b) the corresponding Markov chain $r(t)$, using the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

That is, $I(t)$ will tend to zero exponentially with probability one.

The computer simulation in Figure 5.2(a) supports this result clearly, illustrating extinction of the disease. In this case, $\alpha_1 - \frac{1}{2}\sigma_1^2 N^2 < 0$ with $\sigma_1^2 < \frac{\beta_1}{N}$ while $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2 > 0$ and this means that the first subsystem dies out while the second one persists (see Theorems 3.2 and 3.6). We see that $\nu_{21} > \nu_{12}$ which means that the transition rate from subsystem two (persistent) to subsystem one (extinct) is relatively faster than from subsystem one to subsystem two and therefore the overall system becomes extinct. We see clearly from Figure 5.2(b) that at time periods 18 to 26, the system stays in state 2 for a longer period and if we check the same time period in Figure 5.2(a) the system is drawn back to a high occurrence of disease in that period. However after that most time is spent in state 1 and therefore the disease dies out.

In Theorem 5.3 we require the noise intensity to be such that $\sigma_i^2 \leq \beta_i/N$, $i \in \mathbb{S}$. The following theorem covers the case when $\sigma_i^2 > \beta_i/N$:

Theorem 5.6 *If*

$$\sigma_i^2 > \frac{\beta_i}{N} \quad \forall\ i \in \mathbb{S}, \quad \sum_{i=1}^{M} \pi_i \Big(-\mu_i - \gamma_i + \frac{\beta_i^2}{2\sigma_i^2} \Big) < 0, \tag{5.3.8}$$

then for any given initial value $I(0) = I_0 \in (0, N)$, the solution of the SDE SIS model with Markovian switching (5.2.1) obeys

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \leq \sum_{i=1}^{M} \pi_i \Big(-\mu_i - \gamma_i + \frac{\beta_i^2}{2\sigma_i^2} \Big) < 0 \quad a.s., \tag{5.3.9}$$

namely, $I(t)$ tends to zero exponentially almost surely. In other words, the disease dies out with probability one.

Proof. We use the same notation as in the proof of Theorem 5.3. It is easy to see that the quadratic function $f : (0, N) \times [0, \infty) \times \mathbb{S} \to \mathbb{R}$ defined by (5.3.5) takes its maximum value for each $(t, r(t)) \in [0, \infty) \times \mathbb{S}$ at

$$x(t) = \tilde{x}(t) := \frac{\sigma_{r(t)}^2 N - \beta_{r(t)}}{\sigma_{r(t)}^2}.$$

By condition (5.3.8), it is easy to see that $\tilde{x} \in (0, N)$. Compute

$$\begin{aligned} f(\tilde{x}, t, r(t)) &= \beta_{r(t)} N - \mu_{r(t)} - \gamma_{r(t)} - \frac{1}{2}\sigma_{r(t)}^2 N^2 + \frac{(\sigma_{r(t)}^2 N - \beta_{r(t)})^2}{2\sigma_{r(t)}^2} \\ &= -\mu_{r(t)} - \gamma_{r(t)} + \frac{\beta_{r(t)}^2}{2\sigma_{r(t)}^2}. \end{aligned} \tag{5.3.10}$$

It therefore follows from (5.3.4) that

$$\log(I(t)) \le \log(I_0) + \int_0^t f(\tilde{x}, s, r(s))ds + \int_0^t \sigma_{r(s)}(N - I(s))dB(s).$$

This implies, in the same way as in the proof of Theorem 5.3, that

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le \limsup_{t\to\infty} \frac{1}{t}\int_0^t f(\tilde{x}, s, r(s))ds \quad a.s., \tag{5.3.11}$$

which tends to $\sum_{i=1}^M \pi_i f_i(\tilde{x})$ *a.s.* by the ergodic theory of Markov chains. Here

$$f_i(x) = \beta_i N - \mu_i - \gamma_i - \beta_i x - \frac{1}{2}\sigma_i^2 (N - x)^2.$$

Therefore from (5.3.10) and (5.3.11)

$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \le -\sum_{i=1}^M \pi_i \Big(\mu_i + \gamma_i - \frac{\beta_i^2}{2\sigma_i^2}\Big),$$

as required. The proof is hence complete.

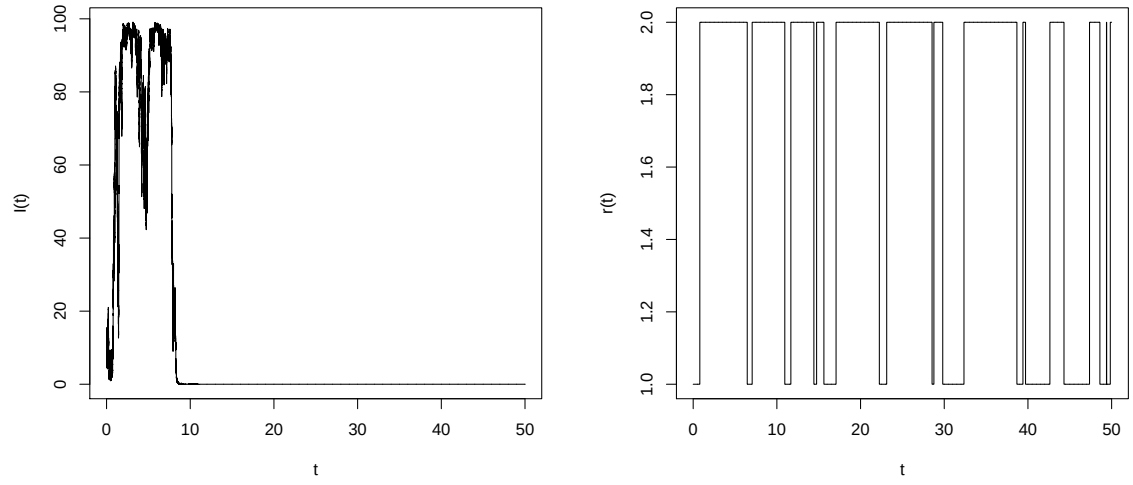
Noting that condition (5.3.8) implies that $T_0^{MC} \le 1$.

Example 5.7 *Assume that the system parameters are given by*

$$\mu_1 = 0.5, \mu_2 = 0.1, \gamma_1 = 0.4, \gamma_2 = 0.2, \beta_1 = 0.05, \beta_2 = 0.02, \sigma_1 = 0.04,$$

$$\sigma_2 = 0.03, N = 100, \nu_{12} = 0.8 \text{ and } \nu_{21} = 0.5.$$

So $\alpha_1 - \frac{1}{2}\sigma_1^2 N^2 = -3.9$, $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2 = -2.8$, $\pi_1 = 0.3846$, *and* $\pi_2 = 0.6154$.



(a) (b)

Figure 5.3: (a) Computer simulation of the path $I(t)$ for the SDE SIS model with Markovian switching (5.2.1) using the EM method with step size $\Delta t = 0.001$, using parameter values from Example 5.7 and initial value $I(0) = 5$; (b) the corresponding Markov chain $r(t)$, using the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

It is easy to verify that the system parameters obey condition (5.3.8). We can therefore conclude, by Theorem 5.6, that for any given initial value $I(0) = I_0 \in (0, N)$, the solution of (5.2.1) obeys

$$\limsup_{t\to\infty} \frac{1}{t} \log(I(t)) \leq -0.0935 < 0 \quad a.s.$$

That is, $I(t)$ will tend to zero exponentially with probability one.

The computer simulation in Figure 5.3(a) supports this result clearly, illustrating extinction of the disease.

We state another extinction condition in the following theorem.

Theorem 5.8 *If*

$$\sum_{i=1}^{M} \pi_i \left(\alpha_i + \frac{1}{2}\sigma_i^2 N^2 \right) < 0 \tag{5.3.12}$$

then for any given initial value $I(0) = I_0 \in (0, N)$, the solution of the SDE (5.2.1) obeys

$$\limsup_{t\to\infty} \frac{1}{t} log(I(t)) \leq \sum_{i=1}^{M} \pi_i \left(\alpha_i + \frac{1}{2}\sigma_i^2 N^2 \right) < 0 \ a.s., \tag{5.3.13}$$

namely, $I(t)$ tends to zero exponentially almost surely. In other words, the disease dies out with probability one.

Proof. Equation (5.2.1) can be written in the following form:

$$dI(t) = f(I(t), t, i)dt + g(x(t), t, i)dB(t), \tag{5.3.14}$$

where $f(I(t), t, i) = I(t)\big(\beta_i N - \mu_i - \gamma_i - \beta_i I(t)\big)$ and $g(x(t), t, i) = \sigma_i I(t)(N - I(t))$.

Using the fact that $I(t) \in (0, N)$ and β_i, μ_i and γ_i are all nonnegative, we have that

$$I(t)\cdot f(I(t), t, i) = \beta_i N I(t)^2 - \mu_i I(t)^2 - \gamma_i I(t)^2 - \beta_i I(t)^3 \leq (\beta_i N - \mu_i - \gamma_i)|I(t)|^2 = \alpha_i |I(t)|^2,$$

$$|g(x(t), t, i)| = |\sigma_i I(t)(N - I(t))| \leq \sigma_i N |I(t)|,$$

$$|I(t) \cdot g(x(t), t, i)| = \sigma_i I(t)^2 (N - I(t)) \geq 0 \cdot |I(t)|^2.$$

By Theorem 2.25 we can immediately obtain the extinction condition as (5.3.12).

By Proposition 5.2 we see clearly that the extinction condition (5.3.12) is stronger than the condition $T_0^{MC} < 1$.

Example 5.9 *Assume that the system parameters are given by*

$$\mu_1 = 0.9, \mu_2 = 0.8, \gamma_1 = 0.9, \gamma_2 = 0.8, \beta_1 = 0.02, \beta_2 = 0.01, \sigma_1 = 0.002,$$

$$\sigma_2 = 0.001, N = 100, \nu_{12} = 0.8 \text{ and } \nu_{21} = 0.5.$$

So $\pi_1 = 0.3846$, *and* $\pi_2 = 0.6154$.

Noting that

$$\left(\alpha_1 + \frac{1}{2}\sigma_1^2 N^2\right)\pi_1 + \left(\alpha_2 + \frac{1}{2}\sigma_2^2 N^2\right)\pi_2 = -0.2816,$$

we can therefore conclude, by Theorem 5.8, that for any given initial value $I(0) = I_0 \in (0, N)$, *the solution of (5.2.1) obeys*

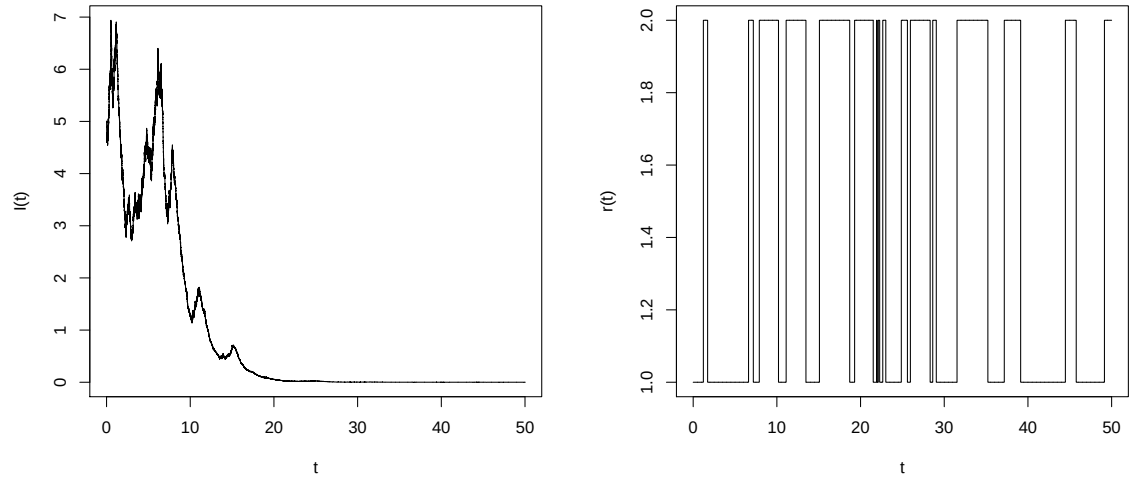
$$\limsup_{t\to\infty} \frac{1}{t}\log(I(t)) \leq -0.2816 < 0 \quad a.s.$$

That is, $I(t)$ *will tend to zero exponentially with probability one.*

The computer simulation in Figure 5.4(a) supports this result clearly, illustrating extinction of the disease.

5.4 Persistence

To show persistence of the system, unlike the technique we used for the stochastic SIS models in the previous two chapters, the M-matrix technique is applied here since in this way we can show persistence of the system with fewer conditions. Before stating our theory on persistence, we first give the definition of stochastic permanence [60].



(a) (b)

Figure 5.4: (a) Computer simulation of the path $I(t)$ for the SDE SIS model with Markovian switching (5.2.1) using the EM method with step size $\Delta t = 0.001$, using parameter values from Example 5.9 and initial value $I(0) = 5$; (b) the corresponding Markov chain $r(t)$, using the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

Definition 5.10 *SDE (5.2.1) is said to be stochastically permanent if for any $\epsilon \in (0, 1)$, there exist positive constants $C_1 = C_1(\epsilon)$ and $C_2 = C_2(\epsilon)$ such that*

$$\liminf_{t\to\infty} P\{I(t) \le C_1\} \ge 1 - \epsilon$$

and

$$\liminf_{t\to\infty} P\{I(t) \ge C_2\} \ge 1 - \epsilon,$$

where $I(t)$ is the solution of the SDE (5.2.1) with $I(0) \in (0, N)$.

Lemma 5.11 *If $T_0^{MC} > 1$, i.e. by Proposition 5.2*

$$\sum_{i=1}^{M} \pi_i \left(\alpha_i - \frac{1}{2} N^2 \sigma_i^2 \right) > 0,$$

then there exists a constant $\theta > 0$ such that the matrix

$$A(\theta) = diag(\xi_1(\theta), \xi_2(\theta), ..., \xi_M(\theta)) - \Gamma \tag{5.4.1}$$

is a non-singular M-matrix, where

$$\xi_i(\theta) = \theta b_i - \theta^2 \frac{1}{2} N^2 \sigma_i^2,$$

for $\forall i \in \mathbb{S}$ and here for convenience we write $b_i = \alpha_i - \frac{1}{2} N^2 \sigma_i^2$.

Proof. The method of proof is very similar to that of Lemma 3.4 in [60]. So we omit the proof here and just use the results.

Lemma 5.12 *If there exists a constant $\theta > 0$ such that $A(\theta)$ (5.4.1) is a nonsingular M-matrix, the solution $I(t)$ of the SDE (5.2.1) has the property that*

$$\limsup_{t\to\infty} E\left(\frac{1}{I^\theta(t)}\right) \leq K_3, \tag{5.4.2}$$

where K_3 will be defined later in (5.4.12)

Proof. We define $U(t)$, $t \geq 0$, that

$$U(t) = \frac{1}{I(t)}. \tag{5.4.3}$$

By the Itô formula (Theorem 2.9), we have

$$\begin{aligned} dU(t) = &\left(-U^2(t)I(t)(\alpha_{r(t)} - \beta_{r(t)}I(t)) + U^3(t)I^2(t)\sigma^2_{r(t)}\left(N - I(t)\right)^2\right)dt \\ &- U^2(t)I(t)\sigma_{r(t)}(N - I(t))dB(t), \end{aligned} \tag{5.4.4}$$

which can be simplified as

$$\begin{aligned} dU(t) = &\left(-\alpha_{r(t)}U(t) + \beta_{r(t)} + \sigma^2_{r(t)}\frac{1}{U(t)} - 2\sigma^2_{r(t)}N + N^2\sigma^2_{r(t)}U(t)\right)dt \\ &- U(t)\sigma_{r(t)}\left(N - \frac{1}{U(t)}\right)dB(t). \end{aligned} \tag{5.4.5}$$

For a vector $x \in \mathbb{R}^M$ we write $x >> 0$ to mean that $x_i \geq 0$ for $i = 1, 2, \cdots, M$. We assume that $A(\theta)$ is a nonsingular M-matrix, so for given θ, there is a vector $\vec{q} = (q_1, q_2, ..., q_M)^T >> 0$ such that

$$\vec{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_M)^T = A(\theta)\vec{q} >> 0,$$

which is

$$q_i\left(\theta b_i - \theta^2\frac{\sigma_i^2N^2}{2}\right) - \sum_{j=1}^{M}\nu_{ij}q_j > 0, \tag{5.4.6}$$

for all $1 \leq i \leq M$.

We define the function V as

$$V(U(t), i) = q_i\left(1 + U(t)\right)^\theta. \tag{5.4.7}$$

By the generalized Itô formula (Theorem 2.14), we have

$$EV(U(t), r(t)) = V(U(0), r(0)) + E\int_0^t LV(U(s), r(s))ds,$$

where

$$\begin{aligned} LV(U,i) = q_i\theta(1+U)^{\theta-1}\left(-\alpha_i U+\beta_i+\sigma_i^2\frac{1}{U}-2\sigma_i^2N+N^2\sigma_i^2U\right) \\ +\frac{1}{2}q_i\theta(\theta-1)\left(1+U\right)^{\theta-2}\left(U\sigma_iN-\sigma_i\right)^2+\sum_{j=1}^{M}\nu_{ij}q_j\left(1+U\right)^{\theta}, \end{aligned} \tag{5.4.8}$$

which is equal to

$$\begin{aligned} &(1+U)^{\theta-2}\left(q_i\theta(1+U)\left(-\alpha_iU+\beta_i+\sigma_i^2\frac{1}{U}-2\sigma_i^2N+N^2\sigma_i^2U\right)\right. \\ &\left.+\frac{1}{2}q_i\theta(\theta-1)\left(U\sigma_iN-\sigma_i\right)^2+\sum_{j=1}^{M}\nu_{ij}q_j\left(1+U\right)^2\right). \end{aligned} \tag{5.4.9}$$

This can be expanded as

$$\begin{aligned} &\frac{1}{U}\left(1+U\right)^{\theta-2}\left(U^3\left(-\alpha_iq_i\theta+N^2\sigma_i^2q_i\theta+\frac{1}{2}N^2\sigma_i^2q_i\theta^2-\frac{1}{2}N^2\sigma_i^2q_i\theta+\sum_{j=1}^{M}\nu_{ij}q_j\right)\right. \\ &+U^2\left(-\alpha_iq_i\theta+N^2\sigma_i^2q_i\theta+\beta_iq_i\theta-2N\sigma_i^2q_i\theta-N\sigma_i^2q_i\theta^2+N\sigma_i^2q_i\theta+2\sum_{j=1}^{M}\nu_{ij}q_j\right) \\ &\left.+U\left(\beta_iq_i\theta-2N\sigma_i^2q_i\theta+\sigma_i^2q_i\theta+\frac{1}{2}\sigma_i^2q_i\theta^2-\frac{1}{2}\sigma_i^2q_i\theta+\sum_{j=1}^{M}\nu_{ij}q_j\right)+q_i\sigma_i^2\theta\right), \end{aligned}$$

which is less than or equal to

$$\begin{aligned} &\frac{1}{U}\left(1+U\right)^{\theta-2}\left(-U^3\left(q_i\left(b_i\theta-\frac{1}{2}N^2\sigma_i^2\theta^2\right)-\sum_{j=1}^{M}\nu_{ij}q_j\right)\right. \\ &\left.+U^2\left(q_i\theta\left(N^2\sigma_i^2+\beta_i\right)+2\sum_{j=1}^{M}\nu_{ij}q_j\right)+U\left(q_i\theta\left(\beta_i+\frac{1}{2}\sigma_i^2+\frac{1}{2}\sigma_i^2\theta\right)+\sum_{j=1}^{M}\nu_{ij}q_j\right)+q_i\sigma_i^2\theta\right). \end{aligned} \tag{5.4.10}$$

Now we choose a constant K_2 which is sufficiently small such that for $1 \leq i \leq M$

$$\vec{\lambda}-K_2\vec{q} >> 0,$$

which is

$$q_i\left(\theta b_i-\theta^2\frac{\sigma_i^2N^2}{2}\right)-\sum_{j=1}^{M}\nu_{ij}q_j-K_2q_i>0. \tag{5.4.11}$$

Again by the generalised Itô formula (Theorem 2.14), we have

$$E\left(e^{K_2t}V(U(t),r(t))\right)=V(U(0),r(0))+E\int_0^t L\left(e^{K_2s}V(U(s),r(s))\right)ds,$$

where

$$L\left(e^{K_2 t}V(U,i)\right) = K_2 e^{K_2 t}V(U,i) + e^{K_2 t}LV(U,i).$$

Substituting expression (5.4.10) into this gives that $L\left(e^{K_2 t}V(U,i)\right)$ is less than or equal to

$$e^{K_2 t}\frac{1}{U}(1+U)^{\theta-2}\left(K_2 q_i (1+U)^2 U - U^3\left(q_i\left(b_i\theta - \frac{1}{2}N^2\sigma_i^2\theta^2\right) - \sum_{j=1}^{M}\nu_{ij}q_j\right)\right.$$
$$\left. + U^2\left(q_i\theta\left(N^2\sigma_i^2 + \beta_i\right) + 2\sum_{j=1}^{M}\nu_{ij}q_j\right) + U\left(q_i\theta\left(\beta_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sigma_i^2\theta\right) + \sum_{j=1}^{M}\nu_{ij}q_j\right) + q_i\sigma_i^2\theta\right),$$

which is equal to

$$e^{K_2 t}\frac{1}{U}(1+U)^{\theta-2}\left(-U^3\left(q_i\left(b_i\theta - \frac{1}{2}N^2\sigma_i^2\theta^2\right) - \sum_{j=1}^{M}\nu_{ij}q_j - K_2 q_i\right)\right.$$
$$+ U^2\left(q_i\theta\left(N^2\sigma_i^2 + \beta_i\right) + 2\sum_{j=1}^{M}\nu_{ij}q_j + 2K_2 q_i\right)$$
$$\left. + U\left(q_i\theta\left(\beta_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sigma_i^2\theta\right) + \sum_{j=1}^{M}\nu_{ij}q_j + K_2 q_i\right) + q_i\sigma_i^2\theta\right).$$

We can write this as

$$L\left(e^{K_2 t}V(U,i)\right) \leq \hat{q}K_2 K_3 e^{K_2 t},$$

where

$$K_3 = \frac{1}{\hat{q}K_2}max_{1\leq i\leq M}\left(sup_{\frac{1}{N}\leq x\leq +\infty}\left(\frac{1}{x}(1+x)^{\theta-2}\left(-x^3\left(q_i\left(b_i\theta - \frac{1}{2}N^2\sigma_i^2\theta^2\right)\right.\right.\right.\right.$$
$$\left. - \sum_{j=1}^{M}\nu_{ij}q_j - K_2 q_i\right) + x^2\left(q_i\theta\left(N^2\sigma_i^2 + \beta_i\right) + 2\sum_{j=1}^{M}\nu_{ij}q_j + 2K_2 q_i\right) \qquad (5.4.12)$$
$$\left. + x\left(q_i\theta\left(\beta_i + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sigma_i^2\theta\right) + \sum_{j=1}^{M}\nu_{ij}q_j + K_2 q_i\right) + q_i\sigma_i^2\theta\right)\right), 1\right).$$

From here we can obtain the required assertion (5.4.2) in the same way as in the proof of Lemma 3.6 in [60]. So the proof is complete.

Theorem 5.13 *If* $T_0^{MC} > 1$, *i.e.*

$$\sum_{i=1}^{M}\pi_i\left(\alpha_i - \frac{1}{2}N^2\sigma_i^2\right) > 0,$$

then the SDE (5.2.1) is stochastically permanent.

Proof. By Lemma 5.11 and Lemma 5.12 and by the same method as in the proof of Theorem 3.2 in [60] if we set $C_2 = \left(\frac{\epsilon}{K_3}\right)^{\frac{1}{\theta}}$, we have that

$$\liminf_{t\to\infty} P\{I(t) \geq C_2\} \geq 1 - \epsilon.$$

The other part of Definition 5.10 required for Theorem 5.13 follows from the fact that $I(t) \in (0, N)$ for $t \geq 0$.

Example 5.14 *We keep the system parameters the same as in Example 5.5 but let*

$$\nu_{12} = 0.8, \text{ and } \nu_{21} = 0.5.$$

So $\alpha_1 - \frac{1}{2}\sigma_1^2 N^2$ *and* $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2$ *are the same as in Example 5.5 but the values for* π_1 *and* π_2 *exchange, i.e.* $\pi_1 = 0.3846$ *and* $\pi_2 = 0.6154$.

Noting that

$$\left(\alpha_1 - \frac{1}{2}\sigma_1^2 N^2\right)\pi_1 + \left(\alpha_2 - \frac{1}{2}\sigma_2^2 N^2\right)\pi_2 = 0.08,$$

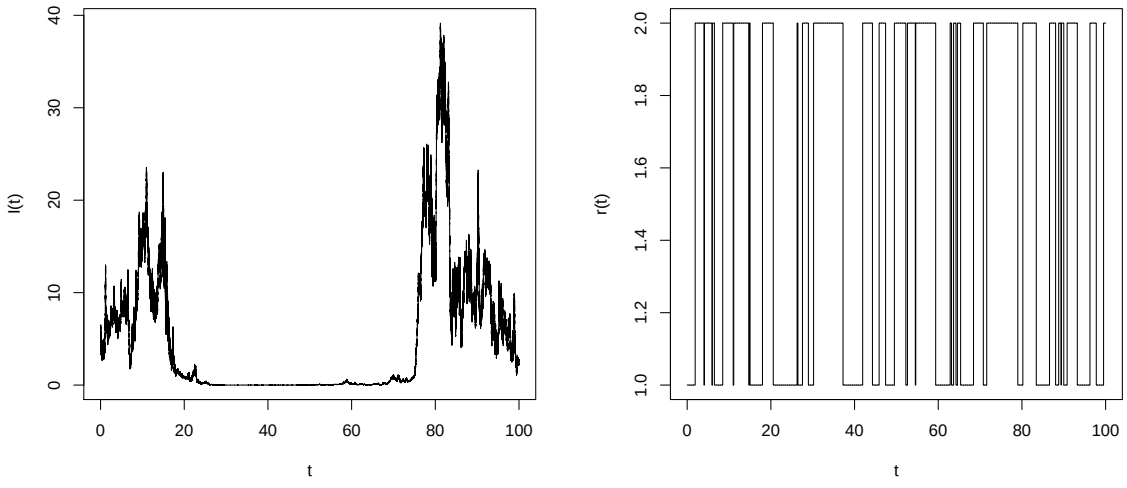
we can therefore conclude, by Theorem 5.13 and Definition 5.10, that the SDE (5.2.1) is stochastically permanent.

The computer simulation in Figure 5.5(a) supports this result clearly, illustrating persistence of the disease. In this case, as in the situation in Example 5.5, we have $\alpha_1 - \frac{1}{2}\sigma_1^2 N^2 < 0$ *with* $\sigma_1^2 < \frac{\beta_1}{N}$ *while* $\alpha_2 - \frac{1}{2}\sigma_2^2 N^2 > 0$ *and this means that the first subsystem dies out while the second one persists (see Theorems 3.2 and 3.6). But in this case,* $\nu_{12} > \nu_{21}$ *which means that the transition rate from subsystem one (extinctive) to subsystem two (persistent) is relatively faster than from subsystem two to subsystem one and therefore the overall system is persistent. This illuminates the important effect of the rates of switching between the states of the overall system.*

5.5 Summary

In this chapter, we have combined the effects of both white and telegraph noise on the classical SIS epidemic model and set up the SDE SIS model with Markovian switching. We have shown the existence of a unique global positive solution for the SIS model incorporating both types of noises and also established conditions for extinction and persistence of the disease. For the SDE SIS model with Markovian switching, a threshold value T_0^{MC} was defined for almost sure extinction and stochastic permanence. Theorems 5.3, 5.6 and 5.8 show that if $T_0^{MC} < 1$, and with other three different conditions, the disease will die out. Theorem 5.13 shows that if $T_0^{MC} > 1$, the disease will persist. We have illustrated our theoretical results with computer simulations.

We now conduct statistical inference for the SDE SIS model (3.2.4) in the next chapter.



(a) (b)

Figure 5.5: (a) Computer simulation of the path $I(t)$ for the SDE SIS model with Markovian switching (5.2.1) using the EM method with step size $\Delta t = 0.001$, using parameter values from Example 5.14 and initial value $I(0) = 5$; (b) the corresponding Markov chain $r(t)$, using the exponential distribution for the switching times of $r(t)$, with $r(0) = 1$.

Chapter 6

Parameter Estimation for the Stochastic SIS Epidemic Model

6.1 Introduction

In this chapter we estimate the parameters in the stochastic SIS epidemic model as formulated in (3.2.3) as an SDE model. Given that $S(t) + I(t) = N$, it is sufficient to study the SDE for $I(t)$ (3.2.4), therefore in this chapter we will concentrate on this SDE only. This SDE is non-linear, and pseudo-Maximum Likelihood Estimation and least squares parameter estimation will be applied. A short literature review regarding the estimation methods in SDEs has been included in section 1.4. Application of Approximate Likelihood Estimation and least squares parameter estimation to SDEs have been widely discussed in previous research. However, variance estimation has not been discussed in most of the papers, which is the main contribution of our research in this chapter. We obtain not only the point estimators but also the interval estimators for parameters and the joint confidence regions taking the correlation among the parameters and the overall degree of confidence into account. Also, we investigate the factors which influence the width of the confidence intervals and the area of the confidence regions both analytically and in our simulation examples.

The organisation of this chapter is as follows: In section 6.2 we apply the least squares estimation approach to our problem and obtain the point estimators, interval estimators and confidence regions for the model parameters β, $\eta = \mu + \gamma$ and σ^2. We consider the cases of parameter estimation for both one sample of data and multiple samples. We also investigate the factors which influence the width of the confidence intervals and the areas of the confidence regions. Simulation examples are given to illustrate our theory. In section 6.3 we discuss the pseudo-MLE method. We obtain the maximum likelihood estimators and exact and approximate confidence regions, and again consider the case of multiple samples. Also we compare the pseudo-MLEs with the least squares estimators both analytically and in our simulation examples. In section 6.4 we summarise the findings in the chapter.

6.2 Least Squares Estimation

Least squares estimation and approximate least squares estimation method have been discussed in [75] and [89]. In this section, we apply least squares estimation in a different way. We use the EM scheme to approximate the path of the process such that the discretised form of the process can be rearranged as a regression model. Then the regression theory can immediately be applied to estimate the model parameters. In this section point estimators and $100(1-\alpha)\%$ confidence intervals as well as $100(1-\alpha)\%$ joint confidence regions will be obtained for our model parameters. Simulation examples will be given to illustrate our theory.

6.2.1 Regression Model

Let $\{I_k\}_{k=0}^n$ be observations from process (3.2.4). Given a stepsize Δt and setting $I_0 = I(0)$, the EM scheme produces the following discretisation over small intervals $[k\Delta t, (k+1)\Delta t]$

$$I_{k+1} - I_k = I_k(\beta N - \mu - \gamma - \beta I_k)\Delta t + \sigma(N - I_k)I_k\Delta W_k, \tag{6.2.1}$$

where $\Delta W_k = W_{k+1} - W_k$.

Equation (6.2.1) can be rewritten as

$$y_{k+1} = \eta u_{k+1} + c + \sigma Z_{k+1}, \tag{6.2.2}$$

where $y_{k+1} = \frac{I_{k+1}-I_k}{I_k(N-I_k)\sqrt{\Delta t}}$, $\eta = \mu + \gamma$, $u_{k+1} = -\frac{\sqrt{\Delta t}}{N-I_k}$, $c = \sqrt{\Delta t}\beta$ and $Z_{k+1} \sim N(0,1)$. We can get the observations $(y_i, u_i)_{i=1}^n$ if data points $\{I_k\}_{k=0}^n$ and stepsize Δt are provided. We then write the model as $y_i = \eta u_i + c + \varepsilon_i$ $(i = 1, 2, ..., n)$, where $\varepsilon_i \sim N(0, \sigma)$. This looks like a simple linear regression model but with the difference that $y = (y_1, y_2, ..., y_n)$ is a random variable instead of a response variable which is conditional on $u = (u_1, u_2, ..., u_n)$. However we still can use the regression theory to estimate η and β since estimation is based on the least squares method, i.e. to minimise $\sum_{i=1}^n (y_i - \eta u_i - c)^2$, which is not affected by whether y is a random variable or not.

Rawlings (1998) [81] discusses multiple linear regression in the general matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{6.2.3}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The calculations work equally well for (6.2.2), which can be written in the matrix

form (6.2.3), where $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ remain the same while $\mathbf{X}$ and $\boldsymbol{\theta}$ become

$$\mathbf{X} = \begin{pmatrix} \sqrt{\Delta t} & u_1 \\ \sqrt{\Delta t} & u_2 \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_n \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \beta \\ \eta \end{pmatrix}.$$

6.2.2 Point Estimators

We use the formulae in the multiple linear regression theory to derive the estimators for η and β as

$$\begin{aligned} \begin{pmatrix} \hat{\beta} \\ \hat{\eta} \end{pmatrix} &= \hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) \\ &= \frac{1}{n\Delta t \sum u_k^2 - \Delta t \left(\sum u_k\right)^2} \begin{pmatrix} \sqrt{\Delta t} \sum u_k^2 \sum y_k - \sqrt{\Delta t} \sum u_k \sum u_k y_k \\ n\Delta t \sum u_k y_k - \Delta t \sum u_k \sum y_k \end{pmatrix}. \end{aligned} \tag{6.2.4}$$

Here $\sum$ represents $\sum_{k=0}^{n-1}$ as does the $\sum$ below.

Then we have point estimators as

$$\hat{\beta} = \frac{\sum u_k^2 \sum y_k - \sum u_k \sum u_k y_k}{n\sqrt{\Delta t} \sum u_k^2 - \sqrt{\Delta t} \left(\sum u_k\right)^2} \tag{6.2.5}$$

and

$$\hat{\eta} = \frac{n \sum u_k y_k - \sum u_k \sum y_k}{n \sum u_k^2 - \left(\sum u_k\right)^2}, \tag{6.2.6}$$

which are equal to

$$\hat{\beta} = \frac{\sum \frac{1}{(N-I_k)^2} \sum \frac{I_{k+1}-I_k}{I_k(N-I_k)} - \sum \frac{1}{N-I_k} \sum \frac{I_{k+1}-I_k}{(N-I_k)^2 I_k}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \left(\sum \frac{\sqrt{\Delta t}}{N-I_k}\right)^2} \tag{6.2.7}$$

and

$$\hat{\eta} = \frac{\sum \frac{I_{k+1}-I_k}{I_k(N-I_k)} \sum \frac{1}{N-I_k} - n \sum \frac{I_{k+1}-I_k}{(N-I_k)^2 I_k}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \left(\sum \frac{\sqrt{\Delta t}}{N-I_k}\right)^2}. \tag{6.2.8}$$

We consider a time interval of total length T divided into n subintervals each of length Δt so $n\Delta t = T$. Hence as $n \to \infty$ and $\Delta t \to 0$ with $n\Delta t = T$, the sums approach the integrals, i.e.

$$\sum_{k=0}^{n-1} \frac{\Delta t}{(N-I_k)^2} \to \int_0^T \frac{1}{(N-I)^2}\, dt$$

$$\sum_{k=0}^{n-1} \frac{I_{k+1}-I_k}{I_k(N-I_k)} \to \int_{I(0)}^{I(T)} \frac{1}{I(N-I)}\, dI \quad \text{etc.}$$

Hence we have that as $n \to \infty$, $\hat{\beta}$ and $\hat{\eta}$ tend to

$$\hat{\beta} = \frac{\int_0^T \frac{1}{(N-I(t))^2}\,dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)I}\,dI - \int_0^T \frac{1}{N-I(t)}\,dt \cdot \int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I}\,dI}{T\int_0^T \frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T \frac{1}{N-I(t)}\,dt\right)^2}$$

and

$$\hat{\eta} = \frac{\int_{I(0)}^{I(T)} \frac{1}{(N-I)I}\,dI \cdot \int_0^T \frac{1}{N-I(t)}\,dt - T\int_{I(0)}^{I(T)} \frac{1}{(N-I)^2 I}\,dI}{T\int_0^T \frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T \frac{1}{N-I(t)}\,dt\right)^2}.$$

6.2.3 Variance of Estimated Parameters

Confidence interval estimators of parameters give more information than simple point estimators. To obtain interval estimators for the parameters β and η, we need to calculate the variance of $\hat{\boldsymbol{\theta}}$ using the formula

$$var(\hat{\boldsymbol{\theta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2, \tag{6.2.9}$$

where σ^2 can be estimated using the residual mean square

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})}{n-p}, \tag{6.2.10}$$

where p is the number of parameters and is 2 in this case. Equation (6.2.10) can be simplified as

$$\hat{\sigma}^2 = \frac{\mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\theta}}}{n-2} \tag{6.2.11}$$

if we substitute $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$ in (6.2.10).

Then equation (6.2.11) can be written as

$$\hat{\sigma}^2 = \frac{1}{n-2}\left(\sum y_k^2 - \left(\sqrt{\Delta t}\sum y_k\right)\hat{\beta} - \left(\sum y_k u_k\right)\hat{\eta}\right). \tag{6.2.12}$$

Substituting (6.2.4) in (6.2.12) we get

$\hat{\sigma}^2 =$

$$\frac{n\sum y_k^2 \sum u_k^2 - \sum y_k^2 (\sum u_k)^2 - \sum u_k^2 (\sum y_k)^2 - n(\sum y_k u_k)^2 + 2\sum u_k \sum y_k \sum y_k u_k}{(n-2)\left(n\sum u_k^2 - (\sum u_k)^2\right)}, \tag{6.2.13}$$

which is

$$\begin{aligned}
&\frac{1}{(n-2)\left(n\sum\frac{\Delta t}{(N-I_k)^2}-\Delta t\left(\sum\frac{1}{N-I_k}\right)^2\right)}\\
&\left(n\sum\frac{(I_{k+1}-I_k)^2}{I_k{}^2(N-I_k)^2}\sum\frac{1}{(N-I_k)^2}-\sum\frac{(I_{k+1}-I_k)^2}{I_k{}^2(N-I_k)^2}\left(\sum\frac{1}{N-I_k}\right)^2\right.\\
&-\sum\frac{1}{(N-I_k)^2}\left(\sum\frac{I_{k+1}-I_k}{I_k(N-I_k)}\right)^2-n\left(\sum\frac{I_{k+1}-I_k}{I_k(N-I_k)^2}\right)^2\\
&\left.+2\sum\frac{1}{N-I_k}\sum\frac{I_{k+1}-I_k}{I_k(N-I_k)}\sum\frac{I_{k+1}-I_k}{I_k(N-I_k)^2}\right).
\end{aligned}\tag{6.2.14}$$

Theorem 6.1 *The estimator $\hat{\sigma}^2=\hat{\sigma}_n^2$ in (6.2.12) (we write it this way to show the dependence on n) is an asymptotically unbiased estimator for σ^2 in (6.2.2), i.e.*

$$\hat{\sigma}_n^2\to\sigma^2\quad a.s.$$

as $n\to\infty$.

Proof.

$$\hat{\sigma}^2=\frac{RSS}{n-p}=\frac{1}{n-2}\sum\left(y_k-\hat{y_k}\right)^2,\tag{6.2.15}$$

where RSS is the sum of squares of residuals for model (6.2.2) and p is the number of parameters to be estimated.

After substituting for $\hat{\beta}$ and $\hat{\eta}$ using (6.2.5) and (6.2.6)

$$\begin{aligned}
y_k-\hat{y_k}=y_k&\\
&-\frac{\sum_i u_i^2\sum_i y_i-\sum_i u_i\sum_i u_iy_i}{\Delta}\\
&-\frac{(n\sum_i u_iy_i-\sum_i u_i\sum_i y_i)u_k}{\Delta},
\end{aligned}$$

where $\Delta=n\sum_i u_i^2-(\sum_i u_i)^2$ and $\sum_i$ represents $\sum_{i=0}^{n-1}$ here and throughout the rest of the chapter.

Since $y_k=\beta\sqrt{\Delta t}+\eta u_k+\sigma Z_k$,

$$\begin{aligned}
y_k-\hat{y_k}=\beta\sqrt{\Delta t}+\eta u_k+\sigma Z_k&\\
&-\frac{\sum_i u_i^2\sum_i(\beta\sqrt{\Delta t}+\eta u_i+\sigma Z_i)-\sum_i u_i\sum_i u_i(\beta\sqrt{\Delta t}+\eta u_i+\sigma Z_i)}{\Delta}\\
&-\frac{[n\sum_i u_i(\beta\sqrt{\Delta t}+\eta u_i+\sigma Z_i)-\sum_i u_i\sum_i(\beta\sqrt{\Delta t}+\eta u_i+\sigma Z_i)]u_k}{\Delta}.
\end{aligned}$$

Therefore $\hat{\sigma}^2$ can be simplified as

$$\frac{\sigma^2}{n-2}\sum\left(Z_k-\frac{u_kn\sum_i u_iZ_i-u_k\sum_i Z_i\sum_i u_i+\sum_i u_i{}^2\sum_i Z_i-\sum_i u_i\sum_i u_iZ_i}{\Delta}\right)^2,$$

which is equal to

$$\frac{\sigma^2}{n-2}\sum\Bigg(Z_k{}^2+\left(\frac{nu_k\sum_i u_iZ_i-u_k\sum_i Z_i\sum_i u_i}{\Delta}\right)^2+\left(\frac{\sum_i u_i{}^2\sum_i Z_i-\sum_i u_i\sum_i u_iZ_i}{\Delta}\right)^2$$
$$-2Z_k\frac{nu_k\sum_i u_iZ_i-u_k\sum_i Z_i\sum_i u_i}{\Delta}-2Z_k\frac{\sum_i u_i{}^2\sum_i Z_i-\sum_i u_i\sum_i u_iZ_i}{\Delta}$$
$$+2\frac{(u_kn\sum_i u_iZ_i-u_k\sum_i Z_i\sum_i u_i)\,(\sum_i u_i{}^2\sum_i Z_i-\sum_i u_i\sum_i u_iZ_i)}{\Delta^2}\Bigg).$$

This can be simplified as

$$\frac{\sigma^2}{n-2}\Bigg(\sum Z_k{}^2+\frac{1}{\Delta^2}\Big(\sum u_k{}^2\left(\sum Z_k\right)^2\left(\sum u_k\right)^2-n^2\left(\sum u_kZ_k\right)^2\sum u_k{}^2$$
$$-n\left(\sum u_k{}^2\right)^2\left(\sum Z_k\right)^2+n\left(\sum u_kZ_k\right)^2\left(\sum u_k\right)^2$$
$$-2\left(\sum u_k\right)^3\sum Z_k\sum u_kZ_k+2n\sum u_k\sum u_kZ_k\sum u_k{}^2\sum Z_k\Big)\Bigg),$$

which equals

$$\frac{\sigma^2}{n-2}\left(\sum Z_k{}^2+\frac{-n\left(\sum u_kZ_k\right)^2-\sum u_k{}^2\left(\sum Z_k\right)^2+2\sum u_k\sum u_kZ_k\sum Z_k}{n\sum u_k{}^2-\left(\sum u_k\right)^2}\right).$$

This can be written as

$$\frac{\sigma^2}{n-2}\left(\sum Z_k{}^2-A\cdot\frac{\left(\sum u_kZ_k\right)^2}{\sum u_k{}^2}-A\cdot\frac{\left(\sum Z_k\right)^2}{n}+B\cdot\frac{\sum u_kZ_k}{\sqrt{\sum u_k{}^2}}\cdot\frac{\sum Z_k}{\sqrt{n}}\right),\qquad(6.2.16)$$

where

$$A=\frac{n\sum u_k^2}{n\sum u_k^2-\left(\sum u_k\right)^2},\ B=2\cdot\frac{\sqrt{n}\sqrt{\sum u_k^2}\sum u_k}{n\sum u_k^2-\left(\sum u_k\right)^2}.$$

Note that

$$\frac{\sum Z_k}{\sqrt{n}}\sim N(0,1)\ \text{ and }\ \frac{\sum u_kZ_k}{\sqrt{\sum u_k^2}}\sim N(0,1),$$

since $Z_k\sim N(0,1)$ and the Z_k are independent. Moreover $\frac{(\sum Z_k)^2}{n}$ and $\frac{(\sum u_kZ_k)^2}{\sum u_k^2}$ have mean 1 and variance 2.

Therefore

$$\frac{\sigma^2}{n-2}\left(-A\cdot\frac{\left(\sum u_kZ_k\right)^2}{\sum u_k{}^2}-A\cdot\frac{\left(\sum Z_k\right)^2}{n}+B\cdot\frac{\sum u_kZ_k}{\sqrt{\sum u_k{}^2}}\cdot\frac{\sum Z_k}{\sqrt{n}}\right)\to 0\ \ a.s$$

as $n\to\infty$.

Also Z_k^2 has mean 1 and variance 2 since $Z_k\sim N(0,1)$. Therefore $\frac{1}{n}\sum Z_k^2\to N(1,\frac{2}{n})$ as $n\to\infty$ by the Central Limit Theorem.

Hence $\hat{\sigma}^2 \to \sigma^2$ with probability one as $n \to \infty$ as required. Using $\hat{\sigma}^2$ to estimate σ^2 in (6.2.9) we have

$$var(\hat{\boldsymbol{\theta}}) = var\begin{pmatrix}\hat{\beta}\\ \hat{\eta}\end{pmatrix} = \frac{1}{n\Delta t\sum u_k^2 - \Delta t\left(\sum u_k\right)^2}\begin{pmatrix}\sum u_k^2 & -\sqrt{\Delta t}\sum u_k\\ -\sqrt{\Delta t}\sum u_k & n\Delta t\end{pmatrix}\hat{\sigma}^2. \tag{6.2.17}$$

6.2.4 Interval Estimation

By the standard least squares theory, if the number of observations n is large, the distribution of the parameter estimators will be almost normal. Then the approximate $100(1-\alpha)\%$ confidence intervals (CIs) for β and η respectively are

$$\hat{\beta}\pm z_{\alpha/2}\sqrt{var(\hat{\beta})} = \frac{\sum u_k^2\sum y_k - \sum u_k\sum u_k y_k}{n\sqrt{\Delta t}\sum u_k^2 - \sqrt{\Delta t}(\sum u_k)^2} \pm z_{\alpha/2}\sqrt{\frac{\sum {u_k}^2\hat{\sigma}^2}{n\Delta t\sum u_k^2 - \Delta t(\sum u_k)^2}} \tag{6.2.18}$$

and

$$\hat{\eta} \pm z_{\alpha/2}\sqrt{var(\hat{\eta})} = \frac{n\sum u_k y_k - \sum u_k\sum y_k}{n\sum u_k^2 - (\sum u_k)^2} \pm z_{\alpha/2}\sqrt{\frac{n\Delta t\hat{\sigma}^2}{n\Delta t\sum u_k^2 - \Delta t(\sum u_k)^2}}, \tag{6.2.19}$$

where $\hat{\sigma}^2$ is the estimation of σ in (6.2.14) and $z_{\alpha/2}$ is the upper $\alpha/2$ value of the standard normal random variable, e.g. $z_{0.025} = 1.96$ for a 95% CI.

We notice that as $n \to \infty$, the $100(1-\alpha)\%$ CIs tend to

$$\begin{aligned}&\frac{\int_0^T \frac{1}{(N-I(t))^2}\,dt\cdot\int_{I(0)}^{I(T)}\frac{1}{(N-I)I}\,dI - \int_0^T\frac{1}{N-I(t)}\,dt\cdot\int_{I(0)}^{I(T)}\frac{1}{(N-I)^2 I}\,dI}{T\int_0^T\frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T\frac{1}{N-I(t)}\,dt\right)^2}\\ &\pm z_{\alpha/2}\sqrt{\frac{\int_0^T\frac{1}{(N-I(t))^2}\,dt\cdot\sigma^2}{T\int_0^T\frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T\frac{1}{N-I(t)}\,dt\right)^2}}\end{aligned} \tag{6.2.20}$$

and

$$\begin{aligned}&\frac{\int_{I(0)}^{I(T)}\frac{1}{(N-I)I}\,dI\cdot\int_0^T\frac{1}{N-I(t)}\,dt - T\int_{I(0)}^{I(T)}\frac{1}{(N-I)^2 I}\,dI}{T\int_0^T\frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T\frac{1}{N-I(t)}\,dt\right)^2}\\ &\pm z_{\alpha/2}\sqrt{\frac{T\sigma^2}{T\int_0^T\frac{1}{(N-I(t))^2}\,dt - \left(\int_0^T\frac{1}{N-I(t)}\,dt\right)^2}},\end{aligned} \tag{6.2.21}$$

respectively.

Theorem 6.2 *The asymptotic widths of the CIs for both β and η, which are*

$$2 \times z_{\alpha/2} \sqrt{\frac{\int_0^T \frac{1}{(N-I(t))^2}\, dt \cdot \sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2}\, dt - \left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2}}$$

and

$$2 \times z_{\alpha/2} \sqrt{\frac{T\sigma^2}{T \int_0^T \frac{1}{(N-I(t))^2}\, dt - \left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2}},$$

are strictly decreasing as T increases.

Proof. Considering first the width of the CI for β,

$$\frac{\int_0^T \frac{1}{(N-I(t))^2}\, dt}{T \int_0^T \frac{1}{(N-I(t))^2}\, dt - \left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2} = \frac{1}{T - \frac{\left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2}{\int_0^T \frac{1}{(N-I(t))^2}\, dt}}. \tag{6.2.22}$$

Then the derivative of the denominator is equal to

$$\begin{aligned}
&\frac{d}{dT}\left(T - \frac{\left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2}{\int_0^T \frac{1}{(N-I(t))^2}\, dt}\right) \\
&= 1 - \frac{2\int_0^T \frac{1}{N-I(t)}\, dt \frac{1}{N-I(T)} \int_0^T \frac{1}{(N-I(t))^2}\, dt - \frac{1}{(N-I(T))^2}\left(\int_0^T \frac{1}{N-I(t)}\, dt\right)^2}{\left(\int_0^T \frac{1}{(N-I(t))^2}\, dt\right)^2} \\
&= \frac{\left(\int_0^T \frac{1}{(N-I(t))^2}\, dt - \frac{1}{N-I(T)} \int_0^T \frac{1}{N-I(t)}\, dt\right)^2}{\left(\int_0^T \frac{1}{(N-I(t))^2}\, dt\right)^2} \geq 0.
\end{aligned}$$

Given a sample path $I(t)$ defined on the interval $[0, T]$ with $I(0) > 0$, we deduce that $I(T) \in (0, N)$ in Theorem 3.1. The only way that the denominator of (6.2.22) is not strictly increasing is if

$$\int_0^T \frac{1}{(N - I(t))^2}\, dt = \frac{1}{N - I(T)} \int_0^T \frac{1}{N - I(t)}\, dt$$

on an interval $[T, T + \varepsilon]$ for some $\varepsilon > 0$.

So if ΔT is small enough

$$(N - I(T + \Delta T)) \int_0^{T+\Delta T} \frac{1}{(N - I(t))^2}\, dt = \int_0^{T+\Delta T} \frac{1}{N - I(t)}\, dt \tag{6.2.23}$$

and

$$(N - I(T)) \int_0^T \frac{1}{(N - I(t))^2}\, dt = \int_0^T \frac{1}{N - I(t)}\, dt. \tag{6.2.24}$$

Subtracting (6.2.24) from (6.2.23) we have

$$\begin{aligned}
&[(N - I(T + \Delta T)) - (N - I(T))] \int_0^{T+\Delta T} \frac{1}{(N - I(t))^2}\, dt \\
&+ (N - I(T))\left(\int_0^{T+\Delta T} \frac{1}{(N - I(t))^2}\, dt - \int_0^{T} \frac{1}{(N - I(t))^2}\, dt \right) \\
&= \int_0^{T+\Delta T} \frac{1}{N - I(t)}\, dt - \int_0^{T} \frac{1}{N - I(t)}\, dt,
\end{aligned}$$

which is equal to

$$\begin{aligned}
&(-I(T + \Delta T) + I(T))\left(\int_0^{T} \frac{1}{(N - I(t))^2}\, dt + \frac{\Delta T}{(N - I(T))^2} + o(\Delta T) \right) \\
&+ (N - I(T))\left(\frac{\Delta T}{(N - I(T))^2} + o(\Delta T) \right) = \frac{\Delta T}{N - I(T)} + o(\Delta T).
\end{aligned}$$

This equals

$$\begin{aligned}
&- I(T)\Big((\beta N - \eta - \beta I(T))\Delta T + \sigma(N - I(T))[B(T + \Delta T) - B(T)] + o(\Delta T) \Big) \\
&\int_0^{T} \frac{1}{(N - I(t))^2}\, dt = o(\Delta T).
\end{aligned}$$

Dividing by $\sqrt{\Delta T}$ we have

$$-I(T)\left((\beta N - \eta - \beta I(T))\sqrt{\Delta T} + \sigma(N - I(T))\frac{B(T + \Delta T) - B(T)}{\sqrt{\Delta T}} \right) = o(\sqrt{\Delta T}).$$

Letting the time step ΔT be very small and choosing $\varepsilon_0 > 0$, $\exists\ \Delta T_0 \leq 1$ such that for $\Delta T < \Delta T_0$ the $o(\sqrt{\Delta T})$ term is between $-\varepsilon_0 \sigma I(T)(N - I(T))\sqrt{\Delta T}$ and $+\varepsilon_0 \sigma I(T)(N - I(T))\sqrt{\Delta T}$, hence must lie between $-\varepsilon_0 \sigma I(T)(N - I(T))$ and $\varepsilon_0 \sigma I(T)(N - I(T))$.

Hence the term

$$\frac{B(T + \Delta T) - B(T)}{\sqrt{\Delta T}} \in \left(\frac{\beta N - \eta - \beta I(T)}{\sigma(N - I(T))} - \varepsilon_0, \frac{\beta N - \eta - \beta I(T)}{\sigma(N - I(T))} + \varepsilon_0 \right).$$

But $\frac{B(T+\Delta T)-B(T)}{\sqrt{\Delta T}} \sim N(0, 1)$ so the probability that it lies in the above interval tends to zero as $\varepsilon_0 \to 0$. Hence

$$P\left(\int_0^{T} \frac{1}{(N - I(t))^2}\, dt = \frac{1}{N - I(T)} \int \frac{1}{N - I(t)}\, dt \ \text{ on } [T, T + \varepsilon] \text{ for some } \varepsilon > 0 \right) = 0.$$

So the denominator of (6.2.22) is strictly increasing and the width of the CI for β is strictly decreasing in T. Similarly we can prove the case for η.

In the same way as for the simple linear SDE case, the asymptotic widths of the CIs here do not depend on the size of time step Δt but only on the total time period T, and are decreasing as T increases.

6.2.5 Joint Confidence Region

We have obtained univariate CIs for each parameter β and η in the last section. However individual CIs do not take into account the correlation among the parameters. Also, they do not reflect the overall degree of confidence. Joint confidence regions take both issues into account. So we will obtain a joint confidence region for β and η in this section.

A $100(1-\alpha)\%$ joint confidence region for the general regression model (6.2.3) is obtained from the following inequality [81]

$$(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \leq p\hat{\sigma}^2 F_{\alpha,p,\nu} \tag{6.2.25}$$

where $F_{\alpha,p,\nu}$ is the value of the F-distribution with degrees of freedom p and ν that leaves probability α in the upper tail, p is the number of parameters and ν is the degrees of freedom associated with $\hat{\sigma}^2$.

Our case only involves two parameters so the $100(1-\alpha)\%$ joint confidence region for β and η can be written as

$$\left(\begin{pmatrix}\hat{\beta}\\ \hat{\eta}\end{pmatrix}-\begin{pmatrix}\beta\\ \eta\end{pmatrix}\right)^T (var(\hat{\beta},\hat{\eta}))^{-1}\left(\begin{pmatrix}\hat{\beta}\\ \hat{\eta}\end{pmatrix}-\begin{pmatrix}\beta\\ \eta\end{pmatrix}\right) \leq 2F_{\alpha,2,n-2}. \tag{6.2.26}$$

After substituting (6.2.17) in (6.2.26), it can be easily calculated as

$$n\Delta t\left(\hat{\beta}-\beta\right)^2 + 2\sqrt{\Delta t}\sum u_k(\hat{\beta}-\beta)(\hat{\eta}-\eta) + \sum {u_k}^2\left(\hat{\eta}-\eta\right)^2 \leq 2\hat{\sigma}^2 F_{\alpha,2,n-2}, \tag{6.2.27}$$

where $\hat{\beta}$ and $\hat{\eta}$ are known in (6.2.5) and (6.2.6).

We compute that

$$4\Delta t\left(\sum u_k\right)^2 - 4n\Delta t\sum u_k^2 = 4\Delta t\left(\left(\sum u_k\right)^2 - n\sum u_k^2\right). \tag{6.2.28}$$

Defining the vectors in $\mathbb{R}^n$

$$\mathbf{a} = (1,1,...,1) \quad \text{and} \quad \mathbf{b} = (u_1,u_2,...,u_n)\,.$$

then $|\mathbf{a}|^2|\mathbf{b}|^2 \geq |\mathbf{a}\cdot\mathbf{b}|^2$ as $|\mathbf{a}\cdot\mathbf{b}| = |\mathbf{a}||\mathbf{b}|cos\theta$, where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. Then we have $|\mathbf{a}|^2 = n$, $|\mathbf{b}|^2 = \sum u_k^2$ and $|\mathbf{a}\cdot\mathbf{b}| = (\sum u_k)^2$. So (6.2.28) is strictly negative since the angle between $\mathbf{a}$ and $\mathbf{b}$ is not 0. Therefore the boundary of the $100(1-\alpha)\%$ joint confidence region is an ellipsoid.

Substituting the definition of u_k in (6.2.27), a $100(1-\alpha)\%$ joint confidence region can be calculated as

$$n\Delta t\left(\hat{\beta}-\beta\right)^2 - 2\sqrt{\Delta t}\sum\frac{\sqrt{\Delta t}}{N-I_k}(\hat{\beta}-\beta)(\hat{\eta}-\eta) + \sum\frac{\Delta t}{(N-I_k)^2}\left(\hat{\eta}-\eta\right)^2 \leq 2\hat{\sigma}^2 F_{\alpha,2,n-2}, \tag{6.2.29}$$

where $\hat{\beta}$ and $\hat{\eta}$ are as in (6.2.7) and (6.2.8).

Example 6.3 *With the same units as in the examples in Chapter 3, we assume that the parameters are given by* $T = 1,\ I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma^2 = 0.03$ *for the model (3.2.4).* $T = 1$ *here represents one year.*

We simulate $I(t)$ *using the above parameters by the EM method with a very small step size* $\Delta t = 0.001$ *and save these* $I(t)$ *as our true dataset. Then we sample every 10th data point in the dataset to obtain the sample for our parameter estimation, so* $n = 100$ *observations and* $\Delta t = 0.01$ *for our sample.*

With the sample we obtained, with $\alpha = 0.05$ *we find the boundary of the* 95% *joint confidence region for* β *and* η *by using (6.2.27) and the univariate* 95% *CIs for each of them by using (6.2.18) and (6.2.19), and also the point estimators by using (6.2.5) and (6.2.6), which are shown in Figure 6.1(a). The ellipsoid in the figure represents the* 95% *joint confidence region while the grey lines represent the univariate CIs. We see that most of the ellipsoid lies in the square which represents the univariate CIs, but the area of the ellipsoid is much smaller than the square. It indicates the advantage of the joint confidence region that it drops out many of the extreme values in the univariate CIs and is more efficient. We see that the ellipsoid is centered at the point estimates of* β *and* η*. The red point which represents the true value of* $\beta = 0.5$ *and* $\eta = 45$ *lies in the ellipse.*

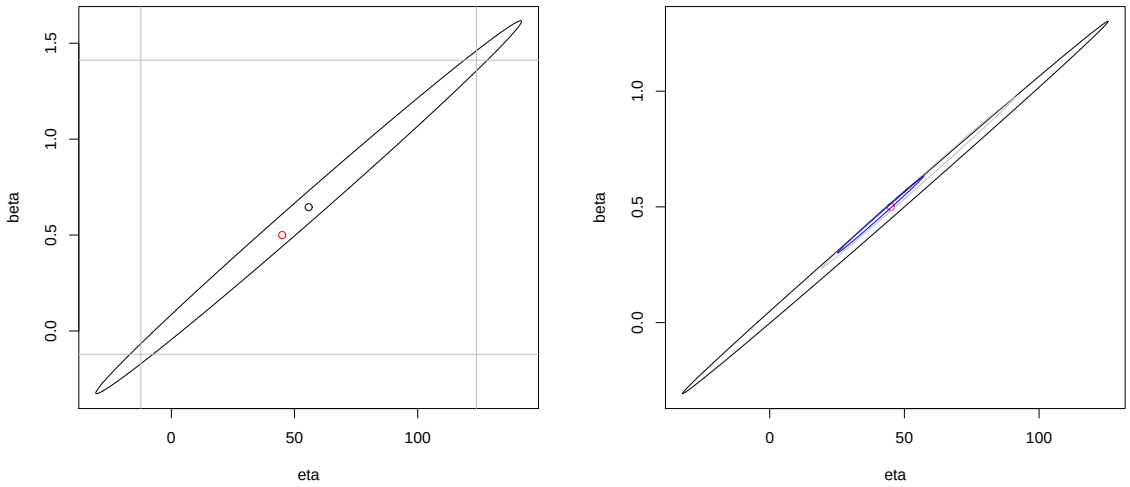
Example 6.4 *Assume that the parameters are given by* $I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma^2 = 0.03$ *for the model (3.2.4), as in Example 6.3.*

In order to see the influence of different interval lengths T *on the* 95% *joint confidence region, we now vary the value of the interval length* T *as* $T = 5$, $T = 20$ *and* $T = 50$ *and use the same method as in Example 6.3 to simulate a dataset for each* T *and sample from each of them. When we increase* T *we increase the number of observations* n *in proportion to* T *to keep* Δt *fixed. We then obtain the three* 95% *joint confidence regions for the different values of* T*, which are shown in Figure 6.1(b). We see that the area of the* 95% *joint confidence region becomes smaller with larger* T *(larger sequence of observations* n*). The red point which represents the true value of* $\beta = 0.5$ *and* $\eta = 45$ *lies in all the ellipses.*

6.2.6 Estimation from Improved Regression Model with More DataSets

The CIs for both β and η are dependent on the sample path. If more datasets are available, we can expand the original regression model to get better parameter estimation.

Assuming that we have m datasets each of size n, we can put all these datasets in



(a) (b)

Figure 6.1: (a) is the 95% joint confidence region for β and η obtained using the parameter values in Example 6.3 with $T = 1$. The ellipsoid in the figure represents the 95% joint confidence region, while the grey vertical and horizontal lines represent the univariate CIs for each of β and η. The black point marked in the figure is the point estimate for β and η, and the red point represents the true values of $\beta = 0.5$ and $\eta = 45$; (b) is the 95% joint confidence region for β and η using the parameter values in Example 6.4, with $T = 5$ (black), $T = 20$ (grey) and $T = 50$ (blue). The red point represents the true value of $\beta = 0.5$ and $\eta = 45$.

the regression model (6.2.2), so that $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ become

$$\mathbf{Y}=\begin{pmatrix} y_{11}\\ y_{12}\\ \vdots\\ y_{1n}\\ y_{21}\\ y_{22}\\ \vdots\\ y_{2n}\\ \vdots\\ y_{m1}\\ y_{m2}\\ \vdots\\ y_{mn}\end{pmatrix},\quad \mathbf{X}=\begin{pmatrix} \sqrt{\Delta t} & u_{11}\\ \sqrt{\Delta t} & u_{12}\\ \vdots & \vdots\\ \sqrt{\Delta t} & u_{1n}\\ \sqrt{\Delta t} & u_{21}\\ \sqrt{\Delta t} & u_{22}\\ \vdots & \vdots\\ \sqrt{\Delta t} & u_{2n}\\ \vdots & \vdots\\ \sqrt{\Delta t} & u_{m1}\\ \sqrt{\Delta t} & u_{m2}\\ \vdots & \vdots\\ \sqrt{\Delta t} & u_{mn}\end{pmatrix},\quad \boldsymbol{\theta}=\begin{pmatrix}\beta\\ \eta\end{pmatrix},\quad \boldsymbol{\varepsilon}=\begin{pmatrix} \varepsilon_{11}\\ \varepsilon_{12}\\ \vdots\\ \varepsilon_{1n}\\ \varepsilon_{21}\\ \varepsilon_{22}\\ \vdots\\ \varepsilon_{2n}\\ \vdots\\ \varepsilon_{m1}\\ \varepsilon_{m2}\\ \vdots\\ \varepsilon_{mn}\end{pmatrix}.$$

using the same formula as in (6.2.4) we have

$$\begin{aligned}\begin{pmatrix}\hat\beta\\ \hat\eta\end{pmatrix}&=\hat{\boldsymbol{\theta}}=(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})\\ &=\frac{1}{mn\Delta t\sum\sum u_{ij}^2-\Delta t\left(\sum\sum u_{ij}\right)^2}\begin{pmatrix}\sqrt{\Delta t}\sum\sum u_{ij}^2\sum\sum y_{ij}-\sqrt{\Delta t}\sum\sum u_{ij}\sum\sum u_{ij}y_{ij}\\ mn\Delta t\sum\sum u_{ij}y_{ij}-\Delta t\sum\sum u_{ij}\sum\sum y_{ij}\end{pmatrix},\end{aligned}\tag{6.2.30}$$

where $\sum\sum=\sum_{i=1}^{m}\sum_{j=0}^{n-1}$ and similarly below.

In the same way, we can get

$$\begin{aligned}\hat\sigma^2&=\frac{\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\theta}}}{mn-2}\\ &=\frac{1}{mn-2}\left(\sum\sum y_{ij}^2-\Big(\sqrt{\Delta t}\sum\sum y_{ij}\Big)\hat\beta-\Big(\sum\sum y_{ij}u_{ij}\Big)\hat\eta\right)\\ &=\frac{1}{(mn-2)\big(mn\Delta t\sum\sum u_{ij}^2-\Delta t\left(\sum\sum u_{ij}\right)^2\big)}\cdot\Bigg(mn\Delta t\sum\sum y_{ij}^2\sum\sum u_{ij}^2\\ &\quad-\Delta t\sum\sum y_k^2\Big(\sum\sum u_{ij}\Big)^2-\Delta t\sum\sum u_{ij}^2\Big(\sum\sum y_{ij}\Big)^2\\ &\quad-mn\Delta t\Big(\sum\sum y_{ij}u_{ij}\Big)^2+2\Delta t\sum\sum u_{ij}\sum\sum y_{ij}\sum\sum y_{ij}u_{ij}\Bigg).\end{aligned}\tag{6.2.31}$$

When proving $\hat\sigma^2$ is an asymptotically unbiased estimator of σ^2, the procedure is similar to the one we used before. We use an equation similar to (6.2.15),

$$\hat\sigma^2=\frac{1}{mn-2}\sum\sum\left(y_{ij}-\hat{y_{ij}}\right)^2.$$

After almost identical working to that used before, we can simplify $\hat{\sigma}^2$ as in (6.2.16), except that the $\sum$ now represents $\sum_{i=1}^{m}\sum_{j=0}^{n-1}$ and the denominator under σ^2 is $mn-2$. We know that $n\to\infty$ implies $mn\to\infty$.

So following almost the same procedure for the proof as before, we can prove that $\hat{\sigma}^2\to\sigma^2$ with probability one as $n\to\infty$.

Using formula (6.2.9) and $\hat{\sigma}$ in (6.2.31) to estimate σ we have

$$var(\hat{\boldsymbol{\theta}})=var\begin{pmatrix}\hat{\beta}\\ \hat{\eta}\end{pmatrix}=\frac{1}{mn\Delta t\sum\sum u_{ij}^2-\Delta t\left(\sum\sum u_{ij}\right)^2}\begin{pmatrix}\sum\sum u_{ij}^2-\sqrt{\Delta t}\sum\sum u_{ij}\\ -\sqrt{\Delta t}\sum\sum u_{ij}mn\Delta t\end{pmatrix}\hat{\sigma}^2. \tag{6.2.32}$$

If the number of observations is large, the $100(1-\alpha)\%$ CIs for β and η estimated from the regression model with m datasets are

$$\begin{aligned}&\hat{\beta}\pm z_{\alpha/2}\sqrt{var(\hat{\beta})}\\ &=\frac{\sum\sum u_{ij}^2\sum\sum y_{ij}-\sum\sum u_{ij}\sum\sum u_{ij}y_{ij}}{mn\sqrt{\Delta t}\sum\sum u_{ij}^2-\sqrt{\Delta t}\left(\sum\sum u_{ij}\right)^2}\pm z_{\alpha/2}\sqrt{\frac{\sum\sum {u_{ij}}^2\hat{\sigma}^2}{mn\Delta t\sum\sum u_{ij}^2-\Delta t\left(\sum\sum u_{ij}\right)^2}}\end{aligned} \tag{6.2.33}$$

and

$$\begin{aligned}&\hat{\eta}\pm z_{\alpha/2}\sqrt{var(\hat{\eta})}\\ &=\frac{mn\sum\sum u_{ij}y_{ij}-\sum\sum u_{ij}\sum\sum y_{ij}}{mn\sum\sum u_{ij}^2-\left(\sum\sum u_{ij}\right)^2}\pm z_{\alpha/2}\sqrt{\frac{mn\Delta t\hat{\sigma}^2}{mn\Delta t\sum\sum u_{ij}^2-\Delta t\left(\sum\sum u_{ij}\right)^2}},\end{aligned} \tag{6.2.34}$$

respectively.

As $n\to\infty$, the $100(1-\alpha)\%$ CIs tend to

$$\begin{aligned}&\frac{\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt\cdot\sum\int_{I_i(0)}^{I_i(T)}\frac{1}{(N-I_i)I_i}\,dI_i-\sum\int_0^T\frac{1}{N-I_i(t)}\,dt\cdot\sum\int_{I_i(0)}^{I_i(T)}\frac{1}{(N-I_i)^2I_i}\,dI_i}{mT\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt-(\sum\int_0^T\frac{1}{N-I_i(t)}\,dt)^2}\\ &\pm z_{\alpha/2}\sqrt{\frac{\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt\cdot\sigma^2}{mT\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt-(\sum\int_0^T\frac{1}{N-I_i(t)}\,dt)^2}}\end{aligned} \tag{6.2.35}$$

and

$$\begin{aligned}&\frac{\sum\int_{I_i(0)}^{I_i(T)}\frac{1}{(N-I_i)I_i}\,dI_i\cdot\sum\int_0^T\frac{1}{N-I_i(t)}\,dt-mT\sum\int_{I_i(0)}^{I_i(T)}\frac{1}{(N-I_i)^2I_i}\,dI_i}{mT\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt-(\sum\int_0^T\frac{1}{N-I_i(t)}\,dt)^2}\\ &\pm z_{\alpha/2}\sqrt{\frac{mT\sigma^2}{mT\sum\int_0^T\frac{1}{(N-I_i(t))^2}\,dt-(\sum\int_0^T\frac{1}{N-I_i(t)}\,dt)^2}},\end{aligned} \tag{6.2.36}$$

respectively. Here $\sum$ represents $\sum_{i=1}^{m}$.

Example 6.5 *Assume that the parameters are given by* $I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma^2 = 0.03$ *for the model (3.2.4), as previously.*

In this example we compare the following three methods in terms of the efficiency of interval estimation. Method 1: One observer is assigned to record $I(t)$ *at one location four times more densely than the comparison during* T*. Method 2: Two observers are assigned to record* $I(t)$ *with the same time steps as the comparison at four locations during* T *and these four samples are combined for estimation. Method 3: one observer is assigned to record* $I(t)$ *with the same time steps as the comparison during time period* $4T$*. To achieve this purpose we design the experiment as follows:*

We obtain three datasets as in Example 6.3, five times. The first four datasets use the model parameters above and $T = 25$*, while the fifth dataset uses* $T = 100$*. Then we sample every twentieth data point in the first dataset to obtain sample A, so* $n = 1,250$ *and* $\Delta t = 0.02$ *for this case. We use sample A as the benchmark. For Method 1, we obtain sample B by sampling every fifth data point in the first dataset, so* $n = 5,000$ *and* $\Delta t = 0.005$ *for this case. We then use (6.2.18) and (6.2.19) to obtain the* 95% *CIs for* β *and* η*. For Method 2, we sample every twentieth data point in the second to fourth datasets to get samples C, D, E and combine them with sample A to obtain four samples each of* $n = 1,250$ *and* $\Delta t = 0.02$*. For samples A, C, D and E combined together we have* $n = 5,000$ *observations in total and* $\Delta t = 0.02$*. We then use estimators from the regression model with more datasets using (6.2.33) and (6.2.34) to obtain the* 95% *CIs for* β *and* η *(with* $\alpha = 0.05,\ z_{\alpha/2} = 1.96$*). For Method 3, we sample every twentieth data point in the fifth data set to obtain sample F so* $n = 5,000$ *and* $\Delta t = 0.02$ *for this case. The results are displayed in Table 6.1.*

We see from Table 6.1 that Method 1 (sample B), using a sample from one location with denser observations, does not give smaller CIs for both β *and* η*, while Method 2 (samples A, C, D and E), using more samples at different locations, decreases the width of the CIs significantly and improves the efficiency of estimation. Method 3 (sample F), using a sample with longer observations at one location, also gives narrower CIs. Therefore we conclude from this example that both Methods 2 and 3 improve the efficiency of estimation. We have repeated our simulations with different model parameter values and the conclusions are the same.*

By substituting (6.2.32) in (6.2.26), we can easily work out the joint $100(1-\alpha)\%$ confidence region for β and η for the regression model with m datasets of size n as

$$mn\Delta t(\hat{\beta}-\beta)^2 + 2\sqrt{\Delta t}\sum\sum u_{ij}(\hat{\beta}-\beta)(\hat{\eta}-\eta) + \sum\sum {u_{ij}}^2(\hat{\eta}-\eta)^2 \le 2\hat{\sigma}^2 F_{\alpha,2,mn-2}, \tag{6.2.37}$$

where $\hat{\beta}$ and $\hat{\eta}$ are given in (6.2.30).

We compute that

$$4\Delta t\left(\sum\sum u_{ij}\right)^2 - 4mn\Delta t\sum\sum u_{ij}^2 = 4\Delta t\left(\left(\sum\sum u_{ij}\right)^2 - mn\sum\sum u_{ij}^2\right). \tag{6.2.38}$$

As for the regression model with one dataset in section 6.2.5, we can prove that (6.2.38) is strictly negative. Therefore the boundary of the $100(1-\alpha)\%$ joint confidence region for the regression model with m datasets is still an ellipse.

Table 6.1: CIs for Example 6.5; results are repeated three times.

	CI for β	Width of CI	CI for η	Width of CI
Sample A ($n = 1,250$, $\Delta t = 0.02$)	(0.36, 1.12)	0.76	(32.06, 106.58)	74.52
	(0.32, 0.78)	0.46	(24.78, 71.42)	42.64
	(0.15, 0.60)	0.45	(10.84, 54.63)	43.79
Sample B ($n = 5,000$, $\Delta t = 0.005$)	(0.41, 1.13)	0.72	(36.77, 107.53)	70.76
	(0.31, 0.72)	0.41	(26.16, 65.84)	39.68
	(0.19, 0.62)	0.43	(14.35, 56.29)	41.94
Sample A, C, D and E ($4 \times n = 1,250$, $\Delta t = 0.02$)	(0.36, 0.68)	0.32	(30.98, 62.92)	31.94
	(0.39, 0.62)	0.23	(33.68, 56.15)	22.47
	(0.38, 0.59)	0.22	(32.62, 53.50)	20.88
Sample F ($n = 5,000$, $\Delta t = 0.02$)	(0.43, 0.76)	0.33	(37.72, 71.73)	34.01
	(0.33, 0.63)	0.30	(28.28, 58.07)	29.79
	(0.38, 0.60)	0.22	(33.12, 54.34)	21.22

Example 6.6 *Assume that the parameters are given by* $T = 1,\ I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,\ m = 10$ *and* $\sigma^2 = 0.03$ *for the model (3.2.4).*

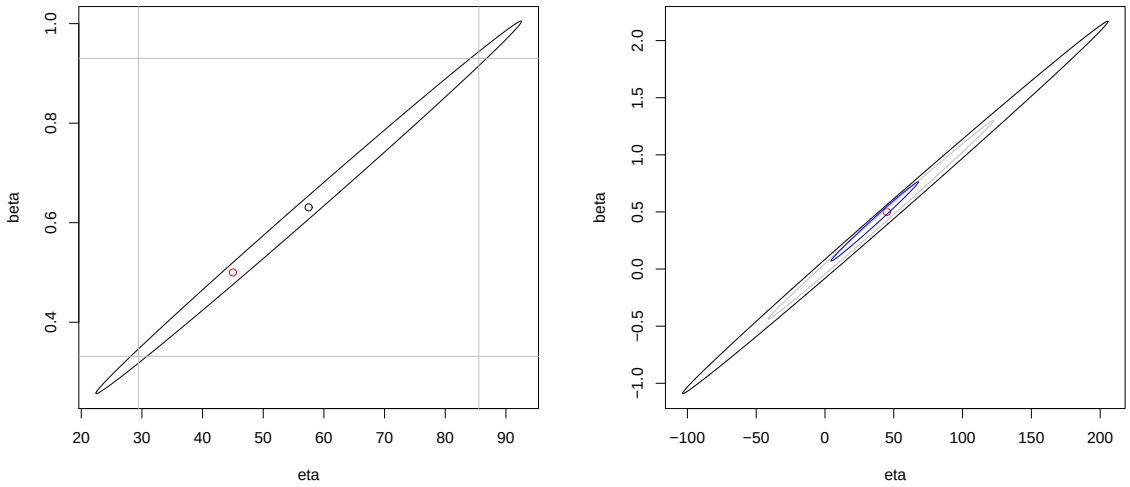
We simulate $I(t)$ *using the above parameters by the EM method with a very small step size* $\Delta t = 0.001$, $m = 10$ *times and save these* $I(t)$ *as ten sets of true data. Then we sample every tenth data point in each dataset to obtain ten samples for our parameter estimation, so* $n = 100$ *and* $\Delta t = 0.01$ *for each of our samples.*

We find the boundary of the 95% *joint confidence region for* β *and* η *using (6.2.38) and the univariate* 95% *CIs for each of them using (6.2.33) and (6.2.34), and also the point estimates using (6.2.30). These are shown in Figure 6.2(a).*

We see that most of the ellipse lies in the square which represents the univariate CIs, but the area of the ellipse is much smaller than that of the square. This indicates the advantage of the joint confidence region, i.e. it does not include many of the extreme values in the univariate CIs and is more efficient. Also we see that the ellipse is centered at the point estimate of β *and* η*. The red point which represents the true value of* $\beta = 0.5$ *and* $\eta = 45$ *lies in the ellipse.*

Example 6.7 *Assume that the model parameters are given by* $T = 1,\ I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100$ *and* $\sigma^2 = 0.03$ *for the model (3.2.4).*

In order to examine the influence of different m *on the* 95% *joint confidence region, we vary the value of* m *as* $m = 1$, $m = 2$ *and* $m = 5$ *and use the same method as in Example 6.6 to simulate datasets for each* m *and sample from each of them. We then obtain three* 95% *joint confidence regions for the different* m*, which are shown in Figure 6.2(b). We see that the area of the* 95% *joint confidence region becomes smaller as* m *becomes larger. Also the red point which represents the true value of* $\beta = 0.5$ *and* $\eta = 45$ *lies in all the ellipses.*



(a) (b)

Figure 6.2: (a) shows the 95% joint confidence region for β and η from (6.2.37) and $m = 10$ datasets generated using the parameter values in Example 6.6. The ellipse in the figure represents the 95% joint confidence region, while the grey lines represent the univariate 95% CIs for each of β and η. The black point marked in the figure is the point estimate for β and η and the red point represents the true value of $\beta = 0.5$ and $\eta = 45$; (b) shows the 95% joint confidence region for β and η using the parameter values in Example 6.7, with $m = 1$ (black), $m = 2$ (grey) and $m = 5$ (blue). The red point represents the true value of $\beta = 0.5$ and $\eta = 45$.

6.3 Pseudo Maximum Likelihood Estimation

In this context, the explicit expressions for MLEs for $\boldsymbol{\phi} = (\beta, \eta, \sigma^2)$ are not attainable, primarily because it is very difficult to find the corresponding likelihood function. Therefore we are unable to use the exact Maximum Likelihood Method. An approximation scheme, the pseudo-likelihood method, will be applied here to obtain estimators for β, η and σ^2. The pseudo-likelihood method has been discussed in [17, 48]. We apply the method in the same way but on a different model. We use the Euler method, which approximates the path of the process, so that the discretised form of the process has a likelihood that is useable and so can be maximised with respect to the parameter values.

6.3.1 Pseudo MLE

The Euler scheme discretises the process as (6.2.1). The increments $I_{k+1} - I_k$ are conditionally independent Gaussian random variables with mean $I_k(\beta N - \eta - \beta I_k)\Delta t$ and variance $\sigma^2 {I_k}^2 (N - I_k)^2 \Delta t$. Therefore the transition density of the process can be written as

$$\begin{aligned} &p(I_{k+1}, (k+1)\Delta t | I_k, k\Delta t) \\ &= \frac{1}{\sqrt{2\pi\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t}} \exp\left(-\frac{1}{2} \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t} \right), \end{aligned} \tag{6.3.1}$$

where $p(I_{k+1}, (k+1)\Delta t | I_k, k\Delta t)$ represents the conditional probability density that $I[(k+1)\Delta t] = I_{k+1}$ given that $I(k\Delta t) = I_k$. Then a pseudo-likelihood is obtained as

$$L_n(\boldsymbol{\phi}) = \prod_{k=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t}} \exp\left(-\frac{1}{2} \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t} \right) \right). \tag{6.3.2}$$

Taking the logarithm of (6.3.2) we have the log pseudo-likelihood

$$\begin{aligned} l_n(\boldsymbol{\phi}) = &-\frac{1}{2} \sum [\ln(2\pi\Delta t) + \ln \sigma^2 + \ln I_k^2 + \ln \left(N - I_k\right)^2] \\ &-\frac{1}{2} \sum \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t}. \end{aligned} \tag{6.3.3}$$

The corresponding partial derivatives with respect to β, η and σ^2 are

$$\frac{\partial l_n(\boldsymbol{\phi})}{\partial \beta} = -\sum \frac{I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t}{\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t} \cdot (-I_k N \Delta t + I_k^2 \Delta t), \tag{6.3.4}$$

$$\frac{\partial l_n(\boldsymbol{\phi})}{\partial (\eta)} = -\sum \frac{I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t}{\sigma^2 {I_k}^2 \left(N - I_k\right)^2 \Delta t} \cdot I_k \Delta t, \tag{6.3.5}$$

$$\frac{\partial l_n(\boldsymbol{\phi})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{(\sigma^2)^2} \cdot \frac{1}{2\Delta t} \cdot \sum \frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{{I_k}^2 \left(N - I_k\right)^2}. \tag{6.3.6}$$

By setting all the partial derivatives equal to zero and solving these simultaneously, we find $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ where the pseudo-likelihood function changes direction. We find that $\hat{\beta}$, $\hat{\eta}$ have the same expressions as the least squares estimators in (6.2.7) and (6.2.8), while $\hat{\sigma^2}$ is almost the same as the least squares estimator (6.2.14) except that it has n in the denominator instead of $(n-2)$. We notice that $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ are a unique solution to the partial derivative equations derived from (6.3.4), (6.3.5) and (6.3.6), and the likelihood function (6.3.2) tends to zero at the boundary. Since the values of the likelihood function are always positive, we conclude that the turning point $(\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$ maximises the pseudo-likelihood function (6.3.2). Therefore $\hat{\boldsymbol{\phi}} = (\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$ are the pseudo-MLEs for (3.2.4).

In the following sections we construct joint confidence regions for the pseudo-MLEs that we have obtained.

6.3.2 Exact Joint Confidence Region

We know that the MLEs are exactly the same as the least squares estimators, except for a minor difference in the estimation of σ^2. If we want to find a joint $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta} = (\beta, \eta)$ then we have already found this in the least squares case in (6.2.27) and (6.2.37) (for both $m = 1$ and $m > 1$) by obtaining an exact 95% confidence region for θ as

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \left(var(\hat{\beta}, \hat{\eta}) \right)^{-1} \sigma^2 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \sigma^2 \chi^2_{\alpha,2}, \tag{6.3.7}$$

where $\chi^2_{\alpha,2}$ is the upper α point of the χ^2 distribution on two degrees of freedom, and then estimating σ^2 by $\hat{\sigma}^2$. Note that we use $\hat{\sigma}^2$ in (6.2.14) instead of $\hat{\sigma}^2$ from the pseudo-MLE since the least squares estimator for σ^2 is unbiased and is slightly better than the pseudo-MLE. Arnold (1998) [4] argues that if plug-in estimates are used for the variance, it is sensible to change the distribution from ${\chi_2}^2$ to $2F_{2,n-2}$ [24], to balance out the loss of accuracy because of the substitution that increases the area of the region. We replace σ^2 by $\hat{\sigma}^2$ and therefore it is more sensible to use $2F_{2,n-2}$ here. Then it will lead to the same analytic form of the $100(1-\alpha)\%$ joint confidence region for β and η as the least squares case in (6.2.26).

We already know the exact confidence region for β and η but we did not obtain the confidence region for all three pseudo-MLEs. In the following sections we construct large sample $100(1-\alpha)\%$ joint confidence regions for all three pseudo-MLEs and for β and η as well for purposes of comparison. There are two ways to construct the asymptotic joint confidence region. The first method is based on the assumption that the pseudo-MLEs are approximately multivariate normally distributed, while the second is based on the likelihood ratio test statistic.

6.3.3 Asymptotic joint confidence regions based on the approximate multivariate normality of pseudo-MLEs

We can regard one data point as $\mathbf{X} = (x_0, x_1, ..., x_n)$, which is a complete run with the initial data I_0 and the transition probability as in (6.3.1). If we obtain m data points $\mathbf{X} = (x_0, x_1, ..., x_n)$ all with the same initial value and the same transition probability, then our m observations are independently and identically distributed and all with the

pseudo-likelihood function as given in (6.3.2). Within this framework, we can apply the asymptotic maximum likelihood theory.

If $m = 1$ (i.e. we have only one run) or m is very small this is not very helpful as the asymptotic theory requires the number of observations (m here) to be very large in order to be valid. In this case we use the exact confidence region as in (6.2.26) for the pseudo-MLE case. If m is large, we can then use the asymptotic pseudo-MLE theory and the likelihood ratio test which we will introduce in the next section.

First we find a joint confidence region for $\boldsymbol{\phi} = (\beta,\, \eta,\, \sigma^2)$. It is well known that the multivariate pseudo-MLEs, the $\hat{\boldsymbol{\phi}}$ in our case, are asymptotically normally distributed [4, 72], i.e.

$$\boldsymbol{\phi}_{(m)} \sim N^{(3)}\left(\boldsymbol{\phi}, \frac{1}{m}\Sigma(\boldsymbol{\phi})\right) \quad \text{approximately,} \tag{6.3.8}$$

where

$$\Sigma^{-1}(\boldsymbol{\phi}) = \sigma^{ij}(\boldsymbol{\phi}) = -E\left\{\frac{\partial^2}{\partial\boldsymbol{\phi}_i \partial\boldsymbol{\phi}_j} \ln f(\mathbf{X};\boldsymbol{\phi})\right\}, \tag{6.3.9}$$

the Fisher information matrix.

Here

$$f(\mathbf{X};\boldsymbol{\phi}) = \prod_{k=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2 {I_k}^2 (N-I_k)^2\,\Delta t}} \exp\left(-\frac{1}{2}\frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 {I_k}^2 (N-I_k)^2\,\Delta t}\right)\right),$$

so that

$$\begin{aligned}\ln f(\mathbf{X};\boldsymbol{\phi}) = &-\frac{1}{2}\sum[\ln(2\pi\Delta t) + \ln\sigma^2 + \ln I_k^2 + \ln(N-I_k)^2] \\ &-\frac{1}{2}\sum\frac{[I_{k+1} - I_k - I_k(\beta N - \eta - \beta I_k)\Delta t]^2}{\sigma^2 {I_k}^2 (N-I_k)^2\,\Delta t}.\end{aligned}$$

The associated quadratic form

$$U = \sum_{i=1}^{3}\sum_{j=1}^{3} m\sigma^{ij}(\boldsymbol{\phi})(\hat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i)(\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j) \tag{6.3.10}$$

has an approximate chi-square distribution with three degrees of freedom for large m.

Because $\hat{\boldsymbol{\phi}}$ is a strongly consistent estimate of $\boldsymbol{\phi}$, the statistics U will still have an asymptotic chi-square distribution with $\sigma^{ij}(\boldsymbol{\phi})$ being substituted by $\sigma^{ij}(\hat{\boldsymbol{\phi}})$.

This will give a three dimensional confidence region for $\boldsymbol{\phi} = (\beta,\, \eta,\, \sigma^2)$. To actually evaluate this asymptotic confidence region for our case is very complicated. The equation (6.3.9) is very difficult to calculate since it involves the approximation of $E\left(\frac{1}{(N-I_k)^2}\right)$, and also it will bring in extra error from the approximation, so we do not use this confidence region in our examples.

On the other hand we could assume that σ is known and that we are trying to estimate $\boldsymbol{\theta} = (\beta,\, \eta)$. This is parallel to the estimation procedure that we used in the least squares

problem (estimating σ by $\hat{\sigma}$ and getting a two dimensional confidence region for β and η). Then

$$\boldsymbol{\theta}_{(m)} \sim N^{(2)}\left(\boldsymbol{\theta}, \frac{1}{m}\Sigma(\boldsymbol{\theta})\right) \quad \text{approximately,}$$

where

$$\Sigma^{-1}(\boldsymbol{\theta}) = \sigma^{ij}(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}_i \partial\boldsymbol{\theta}_j} \ln f(\mathbf{X};\boldsymbol{\theta})\right\}. \tag{6.3.11}$$

The associated quadratic form

$$U' = \sum_{i=1}^{2}\sum_{j=1}^{2} m\sigma^{ij}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \tag{6.3.12}$$

has an approximate chi-square distribution with two degrees of freedom for large m.

Note that $\hat{\boldsymbol{\theta}}$ is the MLE $\hat{\boldsymbol{\theta}}(\sigma) = (\hat{\beta}(\sigma), \hat{\eta}(\sigma))$ with σ known and solves

$$\frac{\partial}{\partial\beta}\ln L_n(\boldsymbol{\theta}) = 0 \quad \text{and} \quad \frac{\partial}{\partial\eta}\ln L_n(\boldsymbol{\theta}) = 0.$$

Here $L_n(\boldsymbol{\theta})$ is given by (6.3.2) except that σ is regarded as known.

If σ is actually unknown, we can substitute σ by its least squares estimator $\hat{\sigma}$. Then the distribution for that statistic U' is $2F_{2,mn-2}$ [24]. We should use $mn-2$ here rather than $m-2$ as the estimator $\hat{\sigma}^2$ is the average of $mn-2$ sums of squares. Also we should use the least squares estimator, not the pseudo-MLE for $\hat{\sigma}$, for the same reason as in section 6.3.2, although the results using the pseudo-MLE will be very close. If m is large, $2F_{2,mn-2}$ will be approximately the same as a chi-square distribution with two degrees of freedom and the asymptotic confidence region will then approach the exact confidence region.

We are unable to work out this asymptotic confidence region numerically for the same reason as in the three dimensional case.

6.3.4 Joint confidence regions based on the likelihood ratio statistic

Another approximate confidence region is based on the likelihood ratio test statistic [4]. Suppose that we have m independent observations $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m$ with common density $f(\mathbf{X}|\boldsymbol{\phi})$. Then we can approximate the $100(1-\alpha)\%$ confidence region for $\boldsymbol{\phi}$ by

$$\{\boldsymbol{\phi} : -2\log R_n(\boldsymbol{\phi}) < \chi^2_{\alpha,3}\}. \tag{6.3.13}$$

Here

$$R_n(\boldsymbol{\phi}) = \frac{L_m(\boldsymbol{\phi})}{L_m(\hat{\boldsymbol{\phi}})},$$

where the vector $\hat{\boldsymbol{\phi}}$ contains the pseudo-MLEs for $\boldsymbol{\phi}$, the parameters,

$$L_m(\boldsymbol{\phi}) = \prod_{j=1}^{m} L_{n,j}(\boldsymbol{\phi}) \tag{6.3.14}$$

and $L_{n,j}(\boldsymbol{\phi}) =$

$$\prod_{k=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}}\exp\left(-\frac{1}{2}\frac{[I_{k+1,j}-I_{k,j}-I_{k,j}(\beta N-\eta-\beta I_{k,j})\Delta t]^2}{\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}\right)\right).$$

So a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\phi}$ is

$$\sum_{j=1}^{m}\sum_{k=1}^{n}\left(\ln\frac{2}{\sqrt{2\pi\hat{\sigma}^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}}-\ln\frac{2}{\sqrt{2\pi\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}}\right.$$
$$\left.+\frac{[I_{k+1,j}-I_{k,j}-I_{k,j}(\beta N-\eta-\beta I_{k,j})\Delta t]^2}{\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}\right)-mn<\chi^2_{\alpha,3}.$$

Again if σ^2 is assumed known, a similar argument shows that a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is

$$\sum_{j=1}^{m}\sum_{k=1}^{n}\left(\frac{[I_{k+1,j}-I_{k,j}-I_{k,j}(\beta N-\eta-\beta I_{k,j})\Delta t]^2}{\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}\right.$$
$$\left.-\frac{[I_{k+1,j}-I_{k,j}-I_{k,j}(\hat{\beta} N-\hat{\eta}-\hat{\beta} I_{k,j})\Delta t]^2}{\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}\right)<\chi^2_{\alpha,2}. \tag{6.3.15}$$

Here again $\hat{\boldsymbol{\theta}}$ is the MLE $\hat{\boldsymbol{\theta}}(\sigma)=(\hat{\beta}(\sigma),\hat{\eta}(\sigma))$ with σ known, and solves

$$\frac{\partial}{\partial\beta}\ln L_m(\boldsymbol{\theta})=0 \quad \text{and} \quad \frac{\partial}{\partial\eta}\ln L_m(\boldsymbol{\theta})=0.$$

In these equations $L_m(\boldsymbol{\theta})$ is given by (6.3.14) but regarded as a function of $\boldsymbol{\theta}=(\beta,\eta)$ with σ known rather than as a function of $\boldsymbol{\phi}=(\beta,\eta,\sigma)$.

Again if we replace the unknown σ by $\hat{\sigma}$ (the least squares estimator), then the distribution should be $2F_{\alpha,2,mn-2}$.

Then (6.3.15) can be written as

$$\sum_{j=1}^{m}\sum_{k=1}^{n}\left(\frac{[I_{k+1,j}-I_{k,j}-I_{k,j}(\beta N-\eta-\beta I_{k,j})\Delta t]^2}{\sigma^2 {I_{k,j}}^2\left(N-I_{k,j}\right)^2\Delta t}-(mn-2)\right)<2F_{\alpha,2,mn-2},$$

which is equivalent to

$$\frac{1}{\hat{\sigma}^2}\left(\sum_{j=1}^{m}\sum_{k=1}^{n}\frac{\left(I_{k+1,j}-I_{k,j}\right)^2}{I_{k,j}\left(N-I_{k,j}\right)^2\Delta t}+\beta^2 mn\Delta t+\eta^2\sum_{j=1}^{m}\sum_{k=1}^{n}\frac{\Delta t}{\left(N-I_{k,j}\right)^2}+\eta\beta\sum_{j=1}^{m}\sum_{k=1}^{n}\frac{-2\Delta t}{N-I_{k,j}}\right.$$
$$\left.+\beta\sum_{j=1}^{m}\sum_{k=1}^{n}\frac{-2(I_{k+1,j}-I_{k,j})}{I_{k,j}(N-I_{k,j})}+\eta\sum_{j=1}^{m}\sum_{k=1}^{n}\frac{2(I_{k+1,j}-I_{k,j})}{I_{k,j}\left(N-I_{k,j}\right)^2}\right)-(mn-2)<2F_{\alpha,2,mn-2}.$$

This can be simplified as

$$\sum_{j=1}^{m}\sum_{k=1}^{n}y_{k,j}^2+\beta^2 nm\Delta t+\eta^2\sum_{j=1}^{m}\sum_{k=1}^{n}u_{k,j}^2+\eta\beta 2\sqrt{\Delta t}\sum_{j=1}^{m}\sum_{k=1}^{n}u_{k,j}+\beta(-2\sqrt{\Delta t})\sum_{j=1}^{m}\sum_{k=1}^{n}y_{k,j}$$
$$+\eta(-2)\sum_{j=1}^{m}\sum_{k=1}^{n}u_{k,j}y_{k,j}-(mn-2)\hat{\sigma}^2<2\hat{\sigma}^2F_{\alpha,2,mn-2},$$

with u_k and y_k defined in (6.2.2).

This can be written as

$$mn\Delta t\left(\beta-\hat{\beta}\right)^2+\sum_{j=1}^{m}\sum_{k=1}^{n}u_{j,k}^2\left(\eta-\hat{\eta}\right)^2+2\sqrt{\Delta t}\sum_{j=1}^{m}\sum_{k=1}^{n}u_{k,j}(\beta-\hat{\beta})(\eta-\hat{\eta})<2\hat{\sigma}^2F_{\alpha,2,mn-2}-D, \tag{6.3.16}$$

where

$$D=-mn\Delta t\hat{\beta}^2-\sum_{j=1}^{m}\sum_{k=1}^{n}u_{j,k}^2\hat{\eta}^2-2\sqrt{\Delta t}\sum_{j=1}^{m}\sum_{k=1}^{n}u_{k,j}\hat{\beta}\hat{\eta}+\sum_{j=1}^{m}\sum_{k=1}^{n}y_{k,j}^2-(mn-2)\hat{\sigma}^2.$$

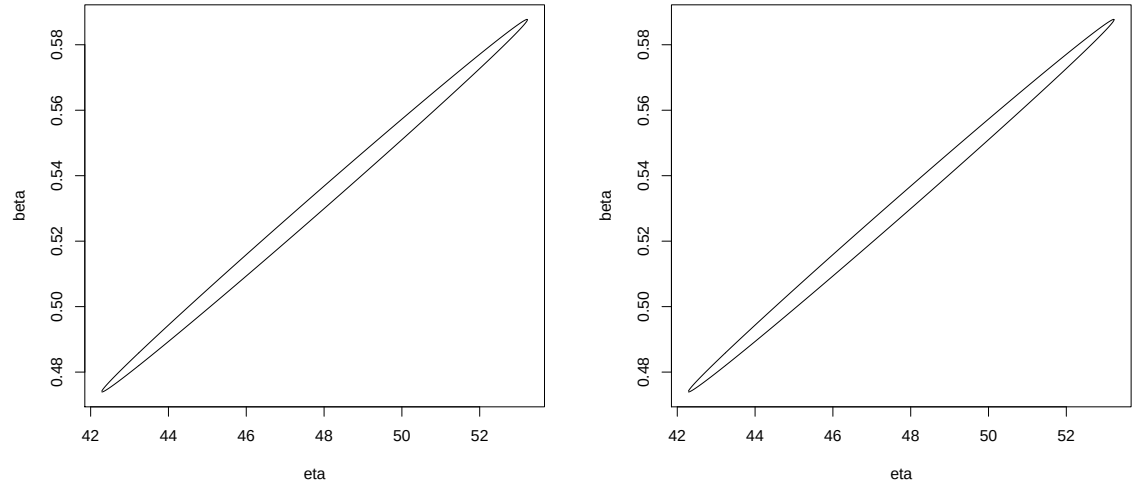
The region (6.3.16) has the same form as the exact confidence region (6.2.37) apart from the substraction of a constant D on the right hand side. We have shown that (6.2.38) is strictly negative and therefore the $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ (6.3.16) is an ellipse centered at the pseudo-MLE $\hat{\beta}$ and $\hat{\eta}$. We numerically compare the exact 95% confidence region for $\boldsymbol{\theta}$ with the asymptotic confidence region obtained by using the likelihood ratio test in the following example, and establish the size of the difference D in this case.

Example 6.8 *Assume that the parameters are given by* $T=5,\ I(0)=10,\ \beta=0.5,\ \mu=20,\ \gamma=25,\ N=100,\ m=100$ *and* $\sigma^2=0.03$ *for the model (3.2.4).*

We use the same method as in Example 6.6 to simulate $m=100$ *datasets and sample from them. With the samples we obtained, we calculate both the* 95% *joint confidence region for* β *and* η *(6.2.37) and the asymptotic confidence region obtained using the likelihood ratio test (6.3.16). These confidence regions are shown in Figure 6.3. The two confidence regions are almost identical and the difference between them can hardly be seen. We calculated* D*, the difference between the two confidence regions as shown in (6.3.16), which is very small,* -2.373×10^{-12} *in this case.*

6.4 Summary

In this chapter we have applied the pseudo-MLE and the least squares method to estimate the parameters in the stochastic SIS model. For the least squares method, we started with the case in which only one dataset is available and then improved our method by considering the case where more than one dataset is available. We have obtained the point estimators, $100(1-\alpha)\%$ CIs and $100(1-\alpha)\%$ joint confidence regions for β and η for both cases. We also investigated which factors influence the width of the CIs and the areas of the confidence regions. Theorem 6.2 states that the asymptotic widths of the CIs for both β and η strictly decrease as the total time period T increases and do not depend on the size of the time step Δt. Example 6.5 shows that a sample from one location with denser observations does not give narrower CIs, while using more than one sample taken at different locations and getting a sample with a longer period of observation at one location decreases the width of CIs significantly and improves the efficiency of estimation. Examples 6.4 and 6.7 show that the area of the confidence region decreases with increasing total time period T and increasing number of samples m.



(a) (b)

Figure 6.3: (a) shows the exact 95% joint confidence region for β and η (6.2.37) using the parameter values in Example 6.8; (b) shows the approximate likelihood ratio based confidence region using (6.3.16).

We have also obtained pseudo-MLEs which are almost the same as the point estimators from the least squares case, with a minor difference in the estimators of σ^2. For obtaining the confidence region for the pseudo-MLEs we considered the following two cases: When the number of samples m is small, we obtained the exact confidence region for β and η in the pseudo-MLE case based on the least squares method. When m is large, we used the asymptotic MLE theory and the likelihood ratio test to obtain the large sample confidence regions for both the three dimensional case (using all three pseudo-MLEs) and the two dimensional case (estimating β and η assuming that σ is known). We only calculated numerically the asymptotic confidence region based on the likelihood ratio test for β and η. Example 6.8 shows that the numerical asymptotic confidence region using the likelihood ratio test for β and η is almost identical to the exact confidence region.

Comparing the least squares estimation method and the pseudo-MLE method, we find that although the pseudo-MLE is more popular for parameter estimation for SDEs, least squares estimation gave the same point estimators and joint confidence region as the pseudo-MLE and is easier to apply. In our case least squares estimation is advantageous. Most of the contents of this chapter have been formed into a paper and submitted to the journal of Statistical Inference for Stochastic Processes.

The Bayesian approach is another popular way to estimate the parameters for SDEs. We will apply the Bayesian estimation to our stochastic SIS model in the next chapter.

Chapter 7

Bayesian Estimation of Parameters for the SDE SIS Epidemic Model

7.1 Introduction

In this chapter, we still consider the SDE SIS model (3.2.3) we derived in Chapter 3. As for the previous chapter we concentrate on the the SDE for $I(t)$ (3.2.4) and apply Bayesian estimation to this SDE. A short literature review regarding Bayesian inference in stochastic epidemic modelling has been conducted in section 1.4, where we found that although many researchers have contributed to the Bayesian inference for the stochastic compartmental model by using the MCMC technique, the noise is introduced in a different way. For the case where Bayesian inference is applied to general SDEs, strong results have been obtained by using the MCMC technique. However the computational cost can be significant. We use a different method for our problem where analytic results will be obtained and is easier to apply.

The organisation of this chapter is as follows: In section 7.2, we set up a conjugate prior and obtain the joint posterior distribution for the model parameters. We maximise this joint posterior distribution with respect to the parameters and obtain the Bayesian estimators. We compare the Bayesian estimators with the point estimators obtained by the least squares approach both analytically and also in simulations. In section 7.3 we obtain confidence intervals for the Bayesian estimators by deriving the marginal distributions from the joint posterior density. In section 7.4 we obtain the joint confidence region for the Bayesian estimators. In both sections 7.3 and 7.4 we compare the results with the least squares case using simulation examples. In section 7.5 we discuss Bayesian estimation for the case where more datasets are available, and a summary is given in section 7.6.

7.2 Bayesian Estimation

In practice, we often have some information about parameters before data is collected. The Bayesian approach is advantageous in this situation since it can include previous

information in the model in the form of a prior distribution.

There are several ways to set a prior distribution [34]. For mathematical convenience, we construct a prior distribution that leads to a simple posterior distribution for our model. The property that the posterior distribution has the same form of distribution as the prior distribution is called conjugacy. In order to obtain a neat theoretical expression for the posterior distribution, we will derive the Bayesian estimators for our problem with a conjugate prior. In practice, if the information that is available contradicts the conjugate prior, it will be necessary to use a more realistic prior distribution. However the posterior distribution may then be very difficult to obtain. In this case, an approximating scheme such as the MCMC method can be applied to obtain an approximate posterior distribution.

Before applying the Bayesian technique, we require the likelihood function for our parameters. However the exact likelihood function is very difficult to derive. An approximation solution, the pseudo-likelihood, which has been obtained in (6.3.2) will be used here.

We now derive the posterior density for our parameters based on the pseudo-likelihood function (6.3.2). We know that the conjugate prior for the two-parameter univariate normal sampling model $N(\mu, \sigma^2)$ has the product form $p(\sigma^2)p(\mu|\sigma)$, where the marginal distribution of σ^2 is scaled inverse-χ^2 and the conditional distribution of μ given σ^2 is normal [34]. In a similar way we can set up the prior distribution for our model and find the posterior distribution.

Theorem 7.1 *We choose the following prior distribution on the parameters*

$\beta|\lambda \sim N(\mu_{10}, (k_{10}\lambda)^{-1})$, *where* $\lambda = \frac{1}{\sigma^2}$,

$\eta|\beta, \lambda \sim N(\mu_{20} + \xi_0\beta, (k_{20}\lambda)^{-1})$,

$\lambda \sim \Gamma(\alpha_0, \beta_0)$.

Also, we suppose that our data is I_0, I_1, ..., I_n *and then set* $x_1 = I_1 - I_0$, $x_2 = I_2 - I_1$, ..., $x_n = I_n - I_{n-1}$. *The posterior density of the parameters* (η, β, λ) *given the data has the form*

$$\lambda^{\bar{\alpha}_0 - 1} e^{-\bar{\beta}_0 \lambda} \cdot \lambda exp\left(-\frac{\lambda}{2}\left(\bar{k}_{10} \left(\beta - \bar{\mu}_{10}\right)^2 + \bar{k}_{20} \left(\eta - \bar{\mu}_{20} - \bar{\xi}_0\beta\right)^2 \right) \right), \tag{7.2.1}$$

where

$$\bar{\alpha}_0 = \alpha_0 + \frac{n}{2}, \tag{7.2.2}$$

$$\bar{\beta}_0 = \beta_0 + \frac{\psi}{2}, \tag{7.2.3}$$

$$\psi = k_{10}\mu_{10}^2 + k_{20}\mu_{20}^2 + \sum \frac{{x_{k+1}}^2}{{I_k}^2 \left(N - I_k\right)^2 \Delta t} - \frac{\left(k_{20}\mu_{20} - \sum \frac{x_{k+1}}{I_k (N - I_k)^2}\right)^2}{k_{20} + \sum \frac{\Delta t}{(N - I_k)^2}} - \bar{k}_{10}\bar{\mu}_{10}^2, \tag{7.2.4}$$

$$\bar{k}_{10} = k_{10} + k_{20}{\xi_0}^2 + n\Delta t - \frac{\left(k_{20}\xi_0 + \sum \frac{\Delta t}{N - I_k}\right)^2}{k_{20} + \sum \frac{\Delta t}{(N - I_k)^2}}, \tag{7.2.5}$$

$$\bar{\mu}_{10} = \frac{\bar{m}_{10}}{\bar{k}_{10}}, \tag{7.2.6}$$

$$\bar{m}_{10} = k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum \frac{x_{k+1}}{I_k(N-I_k)} + \frac{\left(k_{20}\xi_0 + \sum \frac{\Delta t}{N-I_k}\right)\cdot\left(k_{20}\mu_{20} - \sum \frac{x_{k+1}}{I_k(N-I_k)^2}\right)}{k_{20} + \sum \frac{\Delta t}{(N-I_k)^2}}, \tag{7.2.7}$$

$$\bar{k}_{20} = k_{20} + \sum \frac{\Delta t}{(N-I_k)^2}, \tag{7.2.8}$$

$$\bar{\mu}_{20} = \frac{k_{20}\mu_{20} - \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{k_{20} + \sum \frac{\Delta t}{(N-I_k)^2}}, \tag{7.2.9}$$

and

$$\bar{\xi_0} = \frac{k_{20}\xi_0 + \sum \frac{\Delta t}{N-I_k}}{k_{20} + \sum \frac{\Delta t}{(N-I_k)^2}}. \tag{7.2.10}$$

Here $\sum$ *represents* $\sum_{k=0}^{n-1}$ *as does* $\sum$ *below.*

Proof. The posterior distribution of the parameters (η, β, λ) given the data D is

$$\begin{aligned} P(\eta,\beta,\lambda|D) &= P(D|\eta,\beta,\lambda)P(\eta|\beta,\lambda)P(\beta|\lambda)P(\lambda) \\ &\propto \lambda^{\frac{n}{2}} exp\left(-\frac{\lambda}{2}\sum \frac{(x_{k+1} - \beta I_k(N-I_k)\Delta t + \eta I_k \Delta t)^2}{{I_k}^2(N-I_k)^2\Delta t}\right) \\ &\cdot \lambda exp\left(-\frac{\lambda}{2}\Big(k_{10}(\beta-\mu_{10})^2 + k_{20}(\eta - \mu_{20} - \xi_0\beta)^2\Big)\right)\cdot \lambda^{\alpha_0 - 1}e^{-\beta_0\lambda}, \end{aligned} \tag{7.2.11}$$

which is equal to

$$\begin{aligned} &\lambda^{\alpha_0 + \frac{n}{2} - 1}e^{-\beta_0\lambda}\cdot\lambda exp\left(-\frac{\lambda}{2}\left((k_{10} + k_{20}\xi_0^2 + n\Delta t)\beta^2 + \left(k_{20} + \sum \frac{\Delta t}{(N-I_k)^2}\right)\eta^2\right.\right. \\ &- 2\beta\left(k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum \frac{x_{k+1}}{I_k(N-I_k)}\right) \\ &- 2\eta\left(k_{20}\mu_{20} + k_{20}\xi_0\beta - \sum \frac{x_{k+1}}{I_k\,(N-I_k)^2} + \beta\sum \frac{\Delta t}{N-I_k}\right) \\ &\left.\left. + k_{10}\mu_{10}^2 + k_{20}\mu_{20}^2 + \sum \frac{x_{k+1}^2}{{I_k}^2\,(N-I_k)^2\,\Delta t}\right)\right). \end{aligned}$$

This can be written as

$$\lambda^{\alpha_0+\frac{n}{2}-1}e^{-\beta_0\lambda}\cdot\lambda exp\Bigg(-\frac{\lambda}{2}\Bigg($$
$$\left(k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}\right)\cdot\left(\eta-\frac{k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\beta-\frac{k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)^2}}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\right)^2$$
$$+\beta^2\left(k_{10}+k_{20}\xi_0^2+n\Delta t-\frac{\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)^2}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\right)-2\beta\Bigg(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}$$
$$+\sum\frac{x_{k+1}}{I_k(N-I_k)}+\frac{\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)\cdot\left(k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right)}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\Bigg)$$
$$+k_{10}\mu_{10}^2+k_{20}\mu_{20}^2+\sum\frac{{x_{k+1}}^2}{{I_k}^2(N-I_k)^2\Delta t}-\frac{\left(k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right)^2}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\Bigg)\Bigg),$$

which equals

$$\lambda^{\alpha_0+\frac{n}{2}-1}e^{-\beta_0\lambda}\cdot\lambda exp\left(-\frac{\lambda}{2}\left(\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\psi\right)\right).$$

This can easily be simplified as (7.2.1).

We see that (7.2.1) has the same parametric form as the prior distribution so our prior density is a conjugate prior for this problem and gives the density of the posterior estimators.

Note that we shall show that $\bar{k}_{20}$, $\bar{\alpha}_0$, $\bar{\beta}_0$ and $\bar{k}_{10}$ are positive, so that the posterior distribution (7.2.1) is proper.

Clearly $\bar{k}_{20}$ and $\bar{\alpha}_0$ are strictly positive.

In order to show that $\bar{k}_{10}>0$, it is sufficient to show that

$$(k_{20}\xi_0^2+n\Delta t)\left(k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}\right)\geq\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)^2,$$

or equivalently

$$k_{20}\xi_0^2\sum\frac{\Delta t}{(N-I_k)^2}+n\Delta t k_{20}+n\Delta t\sum\frac{\Delta t}{(N-I_k)^2}-2\sum\frac{\Delta t}{N-I_k}k_{20}\xi_0-\left(\sum\frac{\Delta t}{N-I_k}\right)^2\geq 0,$$

which is equal to

$$k_{20}\Delta t\sum\left(\frac{\xi_0}{N-I_k}-1\right)^2+\Delta t^2\left(n\sum\frac{1}{(N-I_k)^2}-\left(\sum\frac{1}{N-I_k}\right)^2\right)\geq 0. \quad (7.2.12)$$

Note that in this section $\sum$ represents $\sum_{k=0}^{n-1}$ for I_k and $\sum_{k=1}^{n}$ for u_k and y_k.

The first term in (7.2.12) is clearly positive and the second term is also positive by the Cauchy-Schwartz inequality. So the result follows.

To show that $\bar{\beta}_0 > 0$, it is sufficient to show that $\psi \geq 0$. We shall do this by using the least squares theory. Indeed it turns out that the posterior Bayesian estimator is a least squares estimator with two extra terms added to take account of the prior distribution. We still use the multiple linear regression model in the general matrix form (6.2.3), where $\boldsymbol{\varepsilon}$ remains the same while $\mathbf{Y}$, $\mathbf{X}$ and $\boldsymbol{\theta}$ become

$$\mathbf{Y} = \begin{pmatrix} \sqrt{k_{10}}\mu_{10} \\ \sqrt{k_{20}}\mu_{20} \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \sqrt{k_{10}} & 0 \\ -\xi_0\sqrt{k_{20}} & \sqrt{k_{20}} \\ \sqrt{\Delta t} & u_1 \\ \sqrt{\Delta t} & u_2 \\ \vdots & \vdots \\ \sqrt{\Delta t} & u_n \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \beta \\ \eta \end{pmatrix}.$$

We can treat $\mathbf{X}$ as the data and $\mathbf{Y}$ as the response variable in a new least squares regression problem.

We use the formulae in the multiple linear regression theorem again, as in (6.2.4) to derive $\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$, where

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} k_{10} + k_{20}\xi_0^2 + n\Delta t & -k_{20}\xi_0 + \sum u_k\sqrt{\Delta t} \\ -k_{20}\xi_0 + \sum u_k\sqrt{\Delta t} & k_{20} + \sum u_k^2 \end{pmatrix}$$

and

$$\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum y_k\sqrt{\Delta t} \\ k_{20}\mu_{20} + \sum u_k y_k \end{pmatrix}.$$

So

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{Det(\mathbf{X}^T\mathbf{X})} \begin{pmatrix} k_{20} + \sum u_k^2 & k_{20}\xi_0 - \sum u_k\sqrt{\Delta t} \\ k_{20}\xi_0 - \sum u_k\sqrt{\Delta t} & k_{10} + k_{20}\xi_0^2 + n\Delta t \end{pmatrix},$$

where

$$\begin{aligned} Det(\mathbf{X}^T\mathbf{X}) &= (k_{10} + k_{20}\xi_0^2 + n\Delta t)\left(k_{20} + \sum u_k^2\right) - \left(-k_{20}\xi_0 + \sum u_k\sqrt{\Delta t}\right)^2 \\ &= (k_{10} + k_{20}\xi_0^2 + n\Delta t)\left(k_{20} + \sum \frac{\Delta t}{(N - I_k)^2}\right) - \left(k_{20}\xi_0 + \sum \frac{\Delta t}{N - I_k}\right)^2. \end{aligned}$$

Hence

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{\left(k_{20}\xi_0 + \sum \frac{\Delta t}{(N-I_k)^2}\right)\bar{k}_{10}} \begin{pmatrix} k_{20} + \sum u_k^2 & k_{20}\xi_0 - \sum u_k\sqrt{\Delta t} \\ k_{20}\xi_0 - \sum u_k\sqrt{\Delta t} & k_{10} + k_{20}\xi_0^2 + n\Delta t \end{pmatrix}.$$

Therefore

$$\begin{aligned} \hat{\beta} = \frac{1}{\left(k_{20}\xi_0 + \sum \frac{\Delta t}{(N-I_k)^2}\right)\bar{k}_{10}} \Bigg(&\left(k_{20} + \sum u_k^2\right)\left(k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum y_k\sqrt{\Delta t}\right) \\ &+ \left(k_{20}\xi_0 - \sum u_k\sqrt{\Delta t}\right)\left(k_{20}\mu_{20} + \sum u_k y_k\right) \Bigg), \end{aligned}$$

which is equal to

$$\frac{1}{\bar{k}_{10}}\Bigg(\left(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}+\sum\frac{x_{k+1}}{I_k(N-I_k)}\right)$$
$$+\frac{\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)\left(k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right)}{k_{20}\xi_0+\sum\frac{\Delta t}{(N-I_k)^2}}\Bigg)=\bar{\mu}_{10}.$$

Similarly

$$\hat{\eta}=\frac{1}{\left(k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}\right)\bar{k}_{10}}\Bigg(\left(k_{20}\xi_0-\sum u_k\sqrt{\Delta t}\right)\left(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}+\sum y_k\sqrt{\Delta t}\right)$$
$$+\left(k_{10}+k_{20}\xi_0^2+n\Delta t\right)\left(k_{20}\mu_{20}+\sum u_k y_k\right)\Bigg),$$

which is equal to

$$\frac{1}{\bar{k}_{10}}\frac{k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\left(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}+\sum\frac{x_{k+1}}{I_k(N-I_k)}\right)$$
$$+\frac{1}{\left(k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}\right)\bar{k}_{10}}\left(k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right)\left(\bar{k}_{10}+\frac{\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)^2}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\right).$$

This can be rearranged as

$$\frac{\bar{\xi}_0}{\bar{k}_{10}}\left(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}+\sum\frac{x_{k+1}}{I_k(N-I_k)}\right)+\frac{\bar{\mu}_{20}}{\bar{k}_{10}}\left(\bar{k}_{10}+\bar{\xi}_0\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)\right),$$

which is equal to

$$\bar{\mu}_{20}+\frac{\bar{\xi}_0}{\bar{k}_{10}}\Bigg(k_{10}\mu_{10}-k_{20}\xi_0\mu_{20}+\sum\frac{x_{k+1}}{I_k(N-I_k)}$$
$$+\left(k_{20}\xi_0+\sum\frac{\Delta t}{N-I_k}\right)\frac{k_{20}\mu_{20}-\sum\frac{x_{k+1}}{I_k(N-I_k)}}{k_{20}+\sum\frac{\Delta t}{(N-I_k)^2}}\Bigg)=\bar{\mu}_{20}+\frac{\bar{\xi}_0}{\bar{k}_{10}}\bar{m}_{10}=\bar{\mu}_{20}+\bar{\xi}_0\bar{\mu}_{10}.$$

Now we define the estimator of σ^2 to be the same as in (6.2.11) but adjusted for dimension, so

$$\hat{\sigma}^2=\frac{\mathbf{Y}^T\mathbf{Y}-\mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\theta}}}{n},$$

where

$$\mathbf{Y}^T\mathbf{Y}=k_{10}\mu_{10}^2+k_{20}\mu_{20}^2+\sum\frac{x_{k+1}^2}{I_k^2\left(N-I_k\right)^2\Delta t},$$

and

$$\mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\theta}} = \left(k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum y_k\sqrt{\Delta t}\right)\bar{\mu}_{10} + \left(k_{20}\mu_{20} + \sum u_k y_k\right)(\bar{\mu}_{20} + \bar{\xi}_0\bar{\mu}_{10}),$$

which is equal to

$$\left(k_{10}\mu_{10} - k_{20}\xi_0\mu_{20} + \sum \frac{x_{k+1}}{I_k(N-I_k)}\right)\bar{\mu}_{10} + \left(k_{20}\mu_{20} - \sum \frac{x_{k+1}}{I_k\,(N-I_k)^2}\right)(\bar{\mu}_{20} + \bar{\xi}_0\bar{\mu}_{10}).$$

This can be written as

$$(\bar{m}_{10} - \bar{\xi}_0\bar{\mu}_{20}\bar{k}_{20})\bar{\mu}_{10} + \bar{\mu}_{20}\bar{k}_{20}(\bar{\mu}_{20} + \bar{\xi}_0\bar{\mu}_{10}) = \bar{k}_{10}\bar{\mu}_{10}^2 + \bar{k}_{20}\bar{\mu}_{20}^2.$$

Hence $\psi = \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\theta}} = n\hat{\sigma}^2 \geq 0$. So $\bar{\beta}_0$ is strictly positive.

Now we have proved $\bar{k}_{10}$, $\bar{k}_{20}$, $\bar{\alpha}_0$ and $\bar{\beta}_0$ are strictly positive so the posterior distribution (7.2.1) is a proper distribution. Therefore the proof is complete.

Now we examine the connection between the Bayesian estimators and the least squares estimators. We expect that the two estimators will be very close if the sample size n is very large, since the larger the available datasets are, the smaller the influence of the prior information on the posterior density. The connection between the two estimation approaches is stated in the following theorem:

Theorem 7.2 *The Bayesian estimators for $\boldsymbol{\varphi} = (\beta, \eta, \sigma^2)$, which maximise the joint posterior distribution, are*

$$\hat{\beta} = \bar{\mu}_{10}, \tag{7.2.13}$$

$$\hat{\eta} = \bar{\mu}_{20} + \bar{\xi}_0\bar{\mu}_{10}, \tag{7.2.14}$$

$$\hat{\sigma}^2 = \frac{\bar{\beta}_0}{\bar{\alpha}_0}, \tag{7.2.15}$$

respectively. Furthermore, $\hat{\beta}$ and $\hat{\eta}$ will tend to the least squares estimators (6.2.7) and (6.2.8) respectively as n becomes infinitely large, while $\hat{\sigma}^2$ will tend to a form almost the same as the least squares estimators (6.2.14) except that n rather than $n-2$ appears in the denominator as n becomes infinitely large.

Proof. To obtain the Bayesian estimators for $\boldsymbol{\varphi} = (\beta, \eta, \sigma^2)$, we take the partial derivatives of the logarithm of the joint posterior density (7.2.1) with respect to β, η and λ.

The logarithm of the function is

$$l(\boldsymbol{\varphi}) = \bar{\alpha}_0 log\lambda - \bar{\beta}_0\lambda - \frac{\lambda}{2}(\bar{k}_{10}(\beta - \bar{\mu}_{10})^2 + \bar{k}_{20}(\eta - \bar{\mu}_{20} - \bar{\xi}_0\beta)^2).$$

The corresponding partial derivatives are

$$\frac{\partial l(\boldsymbol{\varphi})}{\partial \beta} = -\lambda\bar{k}_{10}(\beta - \bar{\mu}_{10}) + \lambda\bar{k}_{20}\bar{\xi}_0(\eta - \bar{\mu}_{20} - \bar{\xi}_0\beta), \tag{7.2.16}$$

$$\frac{\partial l(\boldsymbol{\varphi})}{\partial \eta} = -\lambda\bar{k}_{20}(\eta - \bar{\mu}_{20} - \bar{\xi}_0\beta), \tag{7.2.17}$$

$$\frac{\partial l(\boldsymbol{\varphi})}{\partial \lambda} = \frac{\bar{\alpha}_0}{\lambda} - \bar{\beta}_0 - \frac{1}{2}(\bar{k}_{10}(\beta - \bar{\mu}_{10})^2 + \bar{k}_{20}(\eta - \bar{\mu}_{20} - \bar{\xi}_0\beta)^2). \tag{7.2.18}$$

By setting all the partial derivatives equal to zero and solving these simultaneously, we find $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ where the joint posterior density changes direction. We find that the Bayesian estimators $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ are as in (7.2.13), (7.2.14) and (7.2.15). We notice that $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ are a unique solution to the partial derivative equations (7.2.16), (7.2.17) and (7.2.18), and the posterior density (7.2.1) tends to zero at the boundary. Since the values of the posterior density are always positive, we conclude that the turning point $(\hat{\beta}, \hat{\eta}, \hat{\sigma}^2)$ must maximise the posterior distribution (7.2.1).

Now we prove that $\hat{\beta}$, $\hat{\eta}$ and $\hat{\sigma}^2$ tend to the corresponding least squares estimators when n becomes large. We prove this for the $\hat{\beta}$ case first. When n becomes large (7.2.13) tends to

$$\frac{\sum \frac{x_{k+1}}{I_k(N-I_k)} \sum \frac{\Delta t}{(N-I_k)^2} - \sum \frac{\Delta t}{N-I_k} \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{n\Delta t \sum \frac{\Delta t}{(N-I_k)^2} - \left(\sum \frac{\Delta t}{N-I_k}\right)^2},$$

which equals

$$\frac{\sum \frac{x_{k+1}}{I_k(N-I_k)} \sum \frac{1}{(N-I_k)^2} - \sum \frac{1}{N-I_k} \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \Delta t \left(\sum \frac{1}{N-I_k}\right)^2}.$$

This is exactly as in (6.2.7) for the least squares model.

For the case of $\hat{\eta}$, when n becomes large (7.2.14) tends to

$$-\frac{\sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{\sum \frac{\Delta t}{(N-I_k)^2}} + \frac{\sum \frac{x_{k+1}}{I_k(N-I_k)} \sum \frac{1}{(N-I_k)^2} - \sum \frac{1}{N-I_k} \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \Delta t \left(\sum \frac{1}{N-I_k}\right)^2} \cdot \frac{\sum \frac{\Delta t}{N-I_k}}{\sum \frac{\Delta t}{(N-I_k)^2}},$$

which is equal to

$$\frac{\sum \frac{x_{k+1}}{I_k(N-I_k)} \sum \frac{1}{N-I_k} - n \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{n \sum \frac{\Delta t}{(N-I_k)^2} - \Delta t \left(\sum \frac{1}{N-I_k}\right)^2}.$$

This is exactly as in (6.2.8) for the least squares model.

As for the case of $\hat{\sigma}^2$, when n is infinitely large (7.2.15) tends to

$$\frac{1}{n}\left(\sum \frac{{x_{k+1}}^2}{{I_k}^2 (N-I_k)^2 \Delta t} - \frac{\left(\sum \frac{x_{k+1}}{I_k(N-I_k)^2}\right)^2}{\sum \frac{\Delta t}{(N-I_k)^2}} - \frac{\left(\sum \frac{x_{k+1}}{I_k(N-I_k)} - \frac{\sum \frac{\Delta t}{N-I_k} \sum \frac{x_{k+1}}{I_k(N-I_k)^2}}{\sum \frac{\Delta t}{(N-I_k)^2}}\right)^2}{n\Delta t - \frac{\left(\sum \frac{\Delta t}{N-I_k}\right)^2}{\sum \frac{\Delta t}{(N-I_k)^2}}}\right),$$

which is equal to

$$\frac{1}{n\left(n\sum\frac{\Delta t}{(N-I_k)^2}-\Delta t\left(\sum\frac{1}{N-I_k}\right)^2\right)}$$
$$\left(n\sum\frac{{x_{k+1}}^2}{{I_k}^2(N-I_k)^2}\sum\frac{1}{(N-I_k)^2}-\sum\frac{{x_{k+1}}^2}{{I_k}^2(N-I_k)^2}\left(\sum\frac{1}{N-I_k}\right)^2\right.$$
$$-\sum\frac{1}{(N-I_k)^2}\left(\sum\frac{x_{k+1}}{I_k(N-I_k)}\right)^2-n\left(\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right)^2$$
$$\left.+2\sum\frac{1}{N-I_k}\sum\frac{x_{k+1}}{I_k(N-I_k)}\sum\frac{x_{k+1}}{I_k(N-I_k)^2}\right).$$

This is exactly as in (6.2.14) for the least squares model, but with n on the denominator instead of $n-2$.

We make the following remarks.

Remark 7.3 *When we set up the prior distribution, we have chosen to specify the distribution of β conditional on λ, then the distribution of η on β, λ. It is of concern that this is apparently asymmetric. However we notice that*

$$\begin{aligned}&k_{10}(\beta-\mu_{10})^2+k_{20}(\eta-\mu_{20}-\xi_0\beta)^2\\&=\beta^2(k_{10}+k_{20}\xi_0^2)+\eta^2k_{20}-2\beta\eta\xi_0k_{20}+(-2\mu_{10}k_{10}+2\mu_{20}k_{20}\xi_0)\beta-2\mu_{20}k_{20}\eta\\&+k_{10}\mu_{10}^2+k_{20}\mu_{20}^2,\end{aligned}$$

which can be written as

$$(k_{10}+k_{20}\xi_0^2)\left(\beta-\frac{\eta\xi_0k_{20}}{k_{10}+k_{20}\xi_0^2}+\frac{\mu_{20}k_{20}\xi_0-\mu_{10}k_{10}}{k_{10}+k_{20}\xi_0^2}\right)^2+\eta^2\left(k_{20}-\frac{\xi_0^2k_{20}^2}{k_{10}+k_{20}\xi_0^2}\right)$$
$$+\eta\left(-2\mu_{20}k_{20}+\frac{2\xi_0k_{20}}{k_{10}+k_{20}\xi_0^2}(\mu_{20}k_{20}\xi_0-\mu_{10}k_{10})\right)+k_{10}\mu_{10}^2+k_{20}\mu_{20}^2-\frac{(\mu_{20}k_{20}\xi_0-\mu_{10}k_{10})^2}{k_{10}+k_{20}\xi_0^2},$$

which can be simplified as

$$(k_{10}+k_{20}\xi_0^2)\left(\beta-\frac{\eta\xi_0k_{20}}{k_{10}+k_{20}\xi_0^2}+\frac{\mu_{20}k_{20}\xi_0-\mu_{10}k_{10}}{k_{10}+k_{20}\xi_0^2}\right)^2+\eta^2\frac{k_{10}k_{20}}{k_{10}+k_{20}\xi_0^2}$$
$$-2\eta\frac{k_{10}k_{20}}{k_{10}+k_{20}\xi_0^2}(\mu_{10}\xi_0+\mu_{20})+\frac{k_{10}k_{20}}{k_{10}+k_{20}\xi_0^2}(\mu_{10}\xi_0+\mu_{20})^2,$$

which is equal to

$$(k_{10}+k_{20}\xi_0^2)\left(\beta-\eta\frac{\xi_0k_{20}}{k_{10}+k_{20}\xi_0^2}+\frac{\mu_{20}k_{20}\xi_0-\mu_{10}k_{10}}{k_{10}+k_{20}\xi_0^2}\right)^2+\frac{k_{10}k_{20}}{k_{10}+k_{20}\xi_0^2}(\eta-(\mu_{10}\xi_0+\mu_{20}))^2,$$

which is

$$\hat{k}_{10}(\beta-\hat{\xi}_0\eta+\hat{\mu}_{10})^2+\hat{k}_{20}(\eta-\hat{\mu}_{20})^2, \tag{7.2.19}$$

where $\hat{k}_{10} = k_{10} + k_{20}\xi_0^2$, $\hat{\xi}_0 = \frac{\xi_0 k_{20}}{k_{10}+k_{20}\xi_0^2}$, $\hat{\mu}_{10} = \frac{\mu_{20}k_{20}\xi_0 - \mu_{10}k_{10}}{k_{10}+k_{20}\xi_0^2}$, $\hat{k}_{20} = \frac{k_{10}k_{20}}{k_{10}+k_{20}\xi_0^2}$ *and* $\hat{\mu}_{20} = (\mu_{10}\xi_0 + \mu_{20})$.

This is of the same form with β *and* η *reversed. To express it in another way, it is the same to choose to specify the prior distribution for* η *conditional on* λ *first and then the distribution of* β *conditional on* η, λ. *Therefore the apparent asymmetry is not a problem. Also in this case, the prior distribution on the parameters will be*

$$\eta|\lambda \sim N\left(\hat{\mu}_{10}, \left(\hat{k}_{10}\lambda\right)^{-1}\right),$$

$$\beta|\eta,\lambda \sim N\left(\hat{\mu}_{20} + \hat{\xi}_0\beta, \left(\hat{k}_{20}\lambda\right)^{-1}\right),$$

$$\lambda \sim \Gamma(\alpha_0, \beta_0),$$

where $\hat{\mu}_{10}$, $\hat{k}_{10}$, $\hat{\mu}_{20}$, $\hat{\xi}_0$ *and* $\hat{k}_{20}$ *are defined in (7.2.19).*

Remark 7.4 *If we use the posterior distribution from one stage as the prior distribution for the next stage we should obtain the same answer as if we had used the same prior distribution but observed both stages sequentially.*

We consider the simplest case first. Suppose that we have two data points D_1, D_2, *and the prior distribution on* η, β *and* λ *is* $P(\eta,\beta,\lambda)$ *and the likelihood for the first stage (the probability of observing* D_1 *given* η, β *and* λ*) is*

$$P(D_1|\eta,\beta,\lambda).$$

Then the posterior distribution for η, β *and* λ *is*

$$P(\eta,\beta,\lambda|D_1) = c_1 P(D_1|\eta,\beta,\lambda)P(\eta,\beta,\lambda),$$

where c_1 *is a normalising constant.*

Using this as the prior for the next stage, the posterior distribution for η, β *and* λ *at the second stage is*

$$P(\eta,\beta,\lambda|D_2) = c_2 P(D_2|\eta,\beta,\lambda)P(D_1|\eta,\beta,\lambda)P(\eta,\beta,\lambda),$$

where c_2 *is another normalising constant.*

This is equal to

$$c_2\lambda^{\frac{1}{2}} exp\left(-\frac{\lambda}{2}\frac{(x_2 - \beta I_1(N-I_1)\Delta t + \eta I_1\Delta t)}{I_1^2\,(N-I_1)^2}\Delta t\right)$$
$$\cdot\,\lambda^{\frac{1}{2}} exp\left(-\frac{\lambda}{2}\frac{(x_1 - \beta I_0(N-I_0)\Delta t + \eta I_0\Delta t)}{I_0^2\,(N-I_0)^2}\Delta t\right)\cdot P(\eta,\beta,\lambda),$$

which is equal to

$$c_2\lambda exp\left(-\frac{\lambda}{2}\sum_{k=1}^{2}\frac{(x_k - \beta I_{k-1}(N-I_{k-1})\Delta t + \eta I_{k-1}\Delta t)}{I_{k-1}^2\,(N-I_{k-1})^2}\Delta t\right)\cdot P(\eta,\beta,\lambda),$$

where $x_k = I_k - I_{k-1}$.

This is equal to

$$c_2 P(D_1, D_2|\eta, \beta, \lambda) P(\eta, \beta, \lambda),$$

which is exactly the same posterior distribution as if we had started with the same prior distribution and observed both stages together.

Another way to see this is to look at the equations which define $\bar{k}_{10}$, $\bar{k}_{20}$, $\bar{\mu}_{10}$, $\bar{\mu}_{20}$, $\bar{\xi}_0$, $\bar{\beta}_0$, *i.e.*

$$\bar{k}_{20} = k_{20} + \sum_{k=1}^{n} \frac{\Delta t}{(N - I_{k-1})^2},$$

$$\bar{k}_{20}\bar{\xi}_0 = k_{20}\xi_0 + \sum_{k=1}^{n} \frac{\Delta t}{N - I_{k-1}},$$

$$\bar{k}_{10} + \bar{k}_{20}\bar{\xi}_0^2 = k_{10} + k_{20}\xi_0^2 + n\Delta t,$$

$$\bar{k}_{10}\bar{\mu}_{10} - \bar{k}_{20}\bar{\mu}_{20}\bar{\xi}_0 = k_{10}\mu_{10} - k_{20}\mu_{20}\xi_0 + \sum_{k=1}^{n} \frac{x_k}{I_{k-1}(N - I_{k-1})}$$

and

$$\bar{\beta}_0 + \frac{1}{2}\bar{k}_{10}\bar{\mu}_{10}^2 + \frac{1}{2}\bar{k}_{20}\bar{\mu}_{20}^2 = \beta_0 + \frac{1}{2}k_{10}\mu_{10}^2 + \frac{1}{2}k_{20}\mu_{20}^2 + \frac{1}{2}\sum_{k=1}^{n} \frac{x_k^2}{I_{k-1}^2 (N - I_{k-1})^2 \Delta t}.$$

We get the same result if we start with k_{10}, k_{20}, μ_{10}, μ_{20} *and* β_0 *and add two observations sequentially as if we start with the same values and add two observations simultaneously.*

Example 7.5 *With the same units as in the examples in Chapter 3, we assume that the parameters are given by* $I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma = 0.03$ *for the model (3.2.4).*

We simulate $I(t)$ *using the above parameters by the EM method with a very small step size,* $\Delta t = 0.001$, *and save these* $I(t)$ *as our true dataset. Then we sample every tenth datapoint in the dataset to obtain the sample for our parameter estimation, so* $n = 100$ *observations and* $\Delta t = 0.01$ *for our sample.*

Also we assume two sets of parameters for the prior distribution defined in Theorem 7.1, which are:

Prior A: $\mu_{10} = 0.5$, $k_{10} = 0.05^2$, $\mu_{20} = 40$, $\xi_0 = 10$, $k_{20} = 0.005^2$, $\alpha_0 = 10^3$, $\beta_0 = 1$.

Prior B: $\mu_{10} = 0.3$, $k_{10} = 0.05^2$, $\mu_{20} = 30$, $\xi_0 = 10$, $k_{20} = 0.005^2$, $\alpha_0 = 10^3$, $\beta_0 = 1$.

We see that prior A gives mean values for β *and* η *which are the same as the true values* $\beta = 0.5$ *and* $\eta = 45$, *while prior B gives mean values* $\beta = 0.3$ *and* $\eta = 33$ *which are quite different from the true values (see Theorem 7.1 for the mean values).*

We vary the value of T as $T = 1$, $T = 20$, $T = 100$ and $T = 200$ and use the same method mentioned above to simulate a dataset for each T and sample from each of them. When we increase T we increase n in proportion to T, keeping Δt fixed. For each T, we obtain the least squares estimates using (6.2.7), (6.2.8) and (6.2.14) and the Bayesian estimates using (7.2.13)-(7.2.15) for the three parameters and for both priors. The results are shown in Table 7.1.

We see that using a good prior (prior A) helps to give more accurate parameter estimates when T is small and the least squares estimates are not good. Also the Bayesian estimates for β and η with the different prior distributions become closer as T increases, which means that the prior distribution has less influence on the Bayesian estimates when T is large.

Parameter estimates based on different sample paths will be different. In order to examine the robustness of our estimation, we simulate our true datasets 60 times for each T using the above method and sample from each of them. We obtain 60 least squares estimates and Bayesian estimates for each parameter for each T and then calculate the mean value and the standard deviation for each parameter. The results are shown in Table 7.2.

We see from the mean values that the conclusion is consistent with the one sample case above. We also observe that the standard deviations for the least squares estimates are much larger than those for the Bayesian estimates when T is small. This difference between the standard deviations becomes smaller when T becomes large, but overall the Bayesian estimates are more robust (less variable) than the least squares estimates.

Table 7.1: Least squares and Bayesian parameter estimates from one sample for Example 7.5: the true values of β, η and σ are 0.5, 45 and 0.03 respectively.

Time	Least squares estimates	Prior A	Prior B
T=1	$\hat{\beta} = 2.95$	0.44	0.37
	$\hat{\eta} = 282.17$	46.11	34.33
	$\hat{\sigma} = 0.028$	0.031	0.031
T=20	$\hat{\beta} = 0.79$	0.49	0.39
	$\hat{\eta} = 75.31$	45.51	34.50
	$\hat{\sigma} = 0.032$	0.032	0.032
T=100	$\hat{\beta} = 0.53$	0.50	0.40
	$\hat{\eta} = 48.25$	45.43	35.83
	$\hat{\sigma} = 0.032$	0.032	0.032
T=200	$\hat{\beta} = 0.52$	0.52	0.49
	$\hat{\eta} = 47.20$	46.63	43.64
	$\hat{\sigma} = 0.032$	0.032	0.032

Table 7.2: Statistical summary of parameter estimates from 60 samples for Example 7.5.

Time	Parameter	Least squares estimates		Prior A		Prior B	
		mean	sd.	mean	sd.	mean	sd.
T=1	β	1.25	1.03	0.51	0.037	0.39	0.034
	η	116.46	99.44	45.31	1.41	34.33	1.17
	σ	0.030	0.0024	0.032	0.00011	0.032	0.00011
T=20	β	0.61	0.22	0.50	0.030	0.41	0.027
	η	55.42	21.35	45.31	2.77	36.10	2.25
	σ	0.032	0.00055	0.032	0.00028	0.032	0.00028
T=100	β	0.53	0.10	0.51	0.030	0.44	0.032
	η	48.37	10.39	45.77	2.91	39.22	2.98
	σ	0.032	0.00031	0.032	0.00026	0.032	0.00026
T=200	β	0.53	0.068	0.51	0.037	0.47	0.037
	η	47.45	6.73	46.15	3.64	41.59	3.57
	σ	0.032	0.00018	0.032	0.00017	0.032	0.00017
T=800	β	0.50	0.028	0.50	0.024	0.49	0.023
	η	45.39	2.82	45.31	2.39	43.70	2.30
	σ	0.032	0.000096	0.032	0.000094	0.032	0.000094
T=1200	β	0.50	0.026	0.51	0.023	0.49	0.023
	η	45.66	2.51	45.58	2.26	44.40	2.22
	σ	0.032	0.000094	0.032	0.000092	0.032	0.000092

7.3 Confidence Intervals

In this section we obtain $100(1-\alpha)\%$ confidence intervals (CIs) for the Bayesian estimators by deriving the marginal distributions from the joint posterior distribution. The following theorem shows the marginal distributions for the Bayesian estimators.

Theorem 7.6 *The marginal distributions of the joint posterior density (7.2.1) for the parameters (β, η, λ) have the following forms:*

$$\beta \sim \bar{\mu}_{10} + \sqrt{\frac{\bar{\beta}_0}{\bar{k}_{10}\bar{\alpha}_0}} t_{2\bar{\alpha}_0}, \tag{7.3.1}$$

$$\eta \sim \hat{\mu}_{20} + \sqrt{\frac{\bar{\beta}_0}{\hat{k}_{20}\bar{\alpha}_0}} t_{2\bar{\alpha}_0} \tag{7.3.2}$$

and

$$\lambda \sim \Gamma(\bar{\alpha}_0, \bar{\beta}_0) \tag{7.3.3}$$

Proof. The joint posterior density ignoring the normalising constant has the form (7.2.1). The joint density can be decomposed as

$\beta|\lambda \sim N(\bar{\mu}_{10}, \left(\bar{k}_{10}\lambda\right)^{-1})$,

$\eta|\beta, \lambda \sim N(\bar{\mu}_{20} + \bar{\xi}_0\beta, \left(\bar{k}_{20}\lambda\right)^{-1})$,

$\lambda \sim \Gamma(\bar{\alpha}_0, \bar{\beta}_0)$.

So it is clear that (7.3.3) holds. To obtain the marginal distribution for β, we integrate (7.2.1) with respect to λ and η over their domains, i.e.

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \lambda^{\bar{\alpha}_0-1}\cdot\lambda exp\bigg(-\frac{\lambda}{2}\Big(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\Big)\bigg)d\lambda d\eta.$$

We consider the inner integral first (with respect to λ). We set $Z = A\lambda$ where $A = \frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\right)$. Then the inner integral becomes

$$\int_0^{\infty} exp(-Z)\left(\frac{Z}{A}\right)^{\bar{\alpha}_0}\frac{1}{A}dZ = A^{-\bar{\alpha}_0-1}\int_0^{\infty} exp(-Z)Z^{\bar{\alpha}_0}dZ \propto A^{-\bar{\alpha}_0-1} \tag{7.3.4}$$

since $\int_0^{\infty} exp(-Z)Z^{\bar{\alpha}_0}dZ$ is an unnormalised gamma integral. Now we consider the outer integral, which is

$$\int_{-\infty}^{\infty} A^{-\bar{\alpha}_0-1}d\eta = \int_{-\infty}^{\infty}\left(\frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\right)\right)^{-\bar{\alpha}_0-1}d\eta,$$

which is equal to

$$\left(\frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2\right)\right)^{-\bar{\alpha}_0-1}\int_{-\infty}^{\infty}\left(1+\frac{\bar{k}_{20}}{2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\right)^{-\bar{\alpha}_0-1}d\eta.$$

This can be written as

$$\frac{1}{B}\left(\frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2\right)\right)^{-\bar{\alpha}_0-1}\int_{-\infty}^{\infty}\left(1+\frac{C^2}{2\bar{\alpha}_0+1}\right)^{-\bar{\alpha}_0-1}dC, \tag{7.3.5}$$

where $B = \sqrt{\frac{\bar{k}_{20}(2\bar{\alpha}_0+1)}{2\bar{\beta}_0+\bar{k}_{10}(\beta-\bar{\mu}_{10})^2}}$ and $C = B\cdot(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta)$. This is proportional to

$$\frac{1}{B}\left(\frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2\right)\right)^{-\bar{\alpha}_0-1}$$

since $\int_{-\infty}^{\infty}\left(1+\frac{C^2}{2\bar{\alpha}_0+1}\right)^{-\bar{\alpha}_0-1}dC$ is an unnormalised $t_{2\bar{\alpha}_0+1}$ integral where $2\bar{\alpha}_0+1$ is the degrees of freedom of the t distribution.

Omitting the constants, this is proportional to

$$\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2\right)^{-\bar{\alpha}_0-\frac{1}{2}} \propto \left(1+\frac{\bar{k}_{10}}{2\bar{\beta}_0}\left(\beta-\bar{\mu}_{10}\right)^2\right)^{-\bar{\alpha}_0-\frac{1}{2}}.$$

Then (7.3.1) follows.

In a similar way, we can get the marginal distribution for η as

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \lambda^{\bar{\alpha}_0-1}\cdot\lambda exp\bigg(-\frac{\lambda}{2}\Big(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\Big)\bigg)d\lambda d\eta.$$

The inner integral is the same as the β case so it can be simplified as

$$\int_{-\infty}^{\infty} \left(\frac{1}{2} \left(2\bar{\beta}_0 + \bar{k}_{10} \left(\beta - \bar{\mu}_{10} \right)^2 + \bar{k}_{20} \left(\eta - \bar{\mu}_{20} - \bar{\xi}_0 \beta \right)^2 \right) \right)^{-\bar{\alpha}_0 - 1} d\beta.$$

This can be written as

$$\int_{-\infty}^{\infty} \left(\frac{1}{2} \left(2\bar{\beta}_0 + \hat{k}_{10} (\beta - \hat{\xi}_0 \eta + \hat{\mu}_{10})^2 + \hat{k}_{20} \left(\eta - \hat{\mu}_{20} \right)^2 \right) \right)^{-\bar{\alpha}_0 - 1} d\beta,$$

where $\hat{k}_{10}$, $\hat{\xi}_0$, $\hat{\mu}_{10}$, $\hat{k}_{20}$ and $\hat{k}_{10}$ are defined in remark 7.3 (7.2.19) with k_{10}, k_{20}, ξ_0, μ_{10} and μ_{20} in the expression being changed to $\bar{k}_{10}$, $\bar{k}_{20}$, $\bar{\xi}_0$, $\bar{\mu}_{10}$ and $\bar{\mu}_{20}$.

Following the same procedure as in the β case, we have that the marginal distribution for η is proportional to

$$\left(1 + \frac{\hat{k}_{20}}{2\bar{\beta}_0} \left(\eta - \hat{\mu}_{20} \right)^2 \right)^{-\bar{\alpha}_0 - \frac{1}{2}}.$$

Then (7.3.2) follows.

Example 7.7 *Assume that the parameters are given by* $I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma = 0.03$ *for model (3.2.4), as in Example 7.5.*

We assume the two sets of priors, Prior A and Prior B are as in Example 7.5.

We vary the value of T *as* $T = 1$, $T = 20$, $T = 100$ *and* $T = 200$ *and use the same method mentioned in Example 7.5 to simulate a dataset for each* T *and sample from each of them. When we increase* T *we increase* n *in proportion to* T*, keeping* Δt *fixed. With the sample obtained, with* $\alpha = 0.05$ *we find the Bayesian* 95% *CIs for* β *and* η *using (7.3.1) and (7.3.2) for both priors and for each* T *by using the appropriate percentage points from the* $t_{2\bar{\alpha}_0}$ *distribution. For comparison, we also obtain the* 95% *CIs for* β *and* η *by the least squares method using (6.2.18) and (6.2.19) for each* T*. The results are shown in Table 7.3.*

We see that both the good prior (prior A with mean values at the true values of the parameters) and less good prior (prior B with mean values further away from the true values of the parameters) help decrease the width of the 95% *CI when* T *is small, and the* 95% *CIs for the parameters using the least squares method are very wide. The* 95% *CIs for* β *and* η *with different prior distributions become closer as* T *increases, which means that the prior distribution has less influence on the Bayesian interval estimates when* T *is large.*

7.4 Joint Confidence Region

We have obtained a joint confidence region for β and η using least squares estimation in section 6.2.5. For comparison purposes, we obtain a joint confidence region for only β and η instead of all three parameters in the Bayesian case.

Table 7.3: Least squares and Bayesian confidence intervals for Example 7.7.

		Least squares method	Prior A	Prior B
T=1	β	(0.07, 1.69)	(0.34, 0.64)	(0.26, 0.59)
	width of CI	1.62	0.30	0.33
	η	(4.51, 147.41)	(29.29, 53.86)	(21.86, 49.12)
	width of CI	142.90	24.57	27.26
T=20	β	(0.20, 0.99)	(0.34, 0.58)	(0.27, 0.53)
	width of CI	0.79	0.24	0.26
	η	(16.61, 94.56)	(30.09, 53.58)	(23.16, 48.78)
	width of CI	77.95	23.49	25.62
T=100	β	(0.34, 0.62)	(0.37, 0.56)	(0.34, 0.53)
	width of CI	0.28	0.19	0.19
	η	(29.35, 56.85)	(32.42, 50.94)	(28.72, 48.18)
	width of CI	27.50	18.52	19.46
T=200	β	(0.40, 0.57)	(0.40, 0.55)	(0.38, 0.53)
	width of CI	0.17	0.15	0.15
	η	(34.62, 51.66)	(35.26, 49.35)	(33.31, 47.81)
	width of CI	17.04	14.09	14.50

Theorem 7.8 *A* $100(1-\alpha)\%$ *joint confidence region for* β *and* η *is*

$$\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\leq\bar{r}, \tag{7.4.1}$$

where

$$\bar{r}=2\bar{\beta}_0(\alpha^{-\frac{1}{\bar{\alpha}_0}}-1). \tag{7.4.2}$$

Proof. The proof process is as follows: We first obtain the equation for the contour of the unnormalised joint density for β and η by setting it equal to a constant. We show that the contour of the joint posterior density is an ellipse centered at the Bayesian estimators. We then standardise the ellipse, to centre it at the origin and make its major and minor axes coincide with the Cartesian axes. We then obtain the area of the ellipse as πab where a and b are the lengths of the semi-major and semi-minor axes of the standardised ellipse, which makes the integral $\int\int f(\beta,\eta)\ dregion$ easier to calculate. After solving this integral and normalising the joint posterior for β and η we obtain the height of the distribution so that the volume of the distribution which lies between the height and the top of the distribution is $1-\alpha$. Then the $100(1-\alpha)\%$ joint confidence region is the region enclosed by the contour of the joint posterior density at that height.

We obtain the unnormalised joint posterior for β and η by integrating (7.2.1) with respect to λ over its domain. We can easily find from (7.3.4) that the unnormalised joint posterior for β and η is

$$f(\beta,\eta)=\left(\frac{1}{2}\left(2\bar{\beta}_0+\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2\right)\right)^{-\bar{\alpha}_0-1}.$$

If we set the unnormalised joint posterior density for β and η, $f(\beta,\eta)$, equal to a constant r, this defines a contour of $f(\beta,\eta)$, which is

$$\bar{k}_{10}\left(\beta-\bar{\mu}_{10}\right)^2+\bar{k}_{20}\left(\eta-\bar{\mu}_{20}-\bar{\xi}_0\beta\right)^2=r', \tag{7.4.3}$$

where $r'=2r^{\bar{\alpha}_0+1}-2\bar{\beta}_0$ and with the condition that $r'>0$, which is $r>\bar{\beta}_0^{-\bar{\alpha}_0-1}$.

Equation (7.4.3) can be expanded as

$$\bar{k}_{10}\left(\beta^2-2\bar{\mu}_{10}\beta+\bar{\mu}_{10}^2\right)+\bar{k}_{20}\left(\eta^2+\bar{\mu}_{20}^2+\bar{\xi}_0\beta^2-2\bar{\mu}_{20}\eta-2\bar{\xi}_0\beta\eta+2\bar{\xi}_0\bar{\mu}_{20}\beta\right)=r',$$

which is equal to

$$(\bar{k}_{10}+\bar{k}_{20}\bar{\xi}_0^2)\beta^2+(-2\bar{\mu}_{10}\bar{k}_{10}+2\bar{\xi}_0\bar{\mu}_{20}\bar{k}_{20})\beta+\bar{k}_{20}\eta^2-2\bar{\mu}_{20}\bar{k}_{20}\eta-2\bar{\xi}_0\bar{k}_{20}\beta\eta+\bar{\mu}_{10}^2\bar{k}_{10}+\bar{\mu}_{20}^2\bar{k}_{20}=r'. \tag{7.4.4}$$

We check that

$$\left(-2\bar{\xi}_0\bar{k}_{20}\right)^2-4\bar{k}_{20}(\bar{k}_{10}+\bar{k}_{20}\bar{\xi}_0^2)=4\bar{\xi}_0^2\bar{k}_{20}^2-4\bar{k}_{10}\bar{k}_{20}-4\bar{\xi}_0^2\bar{k}_{20}^2=-4\bar{k}_{10}\bar{k}_{20},$$

which is strictly negative, so the region of $f(\beta,\eta)$ is an ellipse.

Also equation (7.4.4) can be written as

$$A_1(\beta-\hat{\beta})^2+2B_1(\beta-\hat{\beta})(\eta-\hat{\eta})+C_1\left(\eta-\hat{\eta}\right)^2-r'=0, \tag{7.4.5}$$

where

$$A_1=\bar{k}_{10}+\bar{k}_{20}\bar{\xi}_0^2, \tag{7.4.6}$$

$$B_1=-\bar{\xi}_0\bar{k}_{20}, \tag{7.4.7}$$

$$C_1=\bar{k}_{20}, \tag{7.4.8}$$

and $\hat{\beta}$ and $\hat{\eta}$ are the Bayesian estimators for β and η as in (7.2.13) and (7.2.14). We notice that this ellipse is centered at the Bayesian estimators $\hat{\beta}$ and $\hat{\eta}$.

Now we convert the ellipse (7.4.5) to the standard form. We first do the conversion $\beta=\beta'+\hat{\beta}$ and $\eta=\eta'+\hat{\eta}$, so equation (7.4.5) becomes

$$A_1\beta'^2+2B_1\beta'\eta'+C_1\eta'^2-r'=0. \tag{7.4.9}$$

We then do the conversion $\beta'=\beta''cos\phi+\eta''sin\phi$ and $\eta'=\eta''cos\phi-\beta''sin\phi$, where $\phi=\frac{1}{2}arctan\left(\frac{2B_1}{C_1-A_1}\right)$, and equation (7.4.9) becomes

$$A_1\left(\beta''cos\phi+\eta''sin\phi\right)^2+2B_1(\beta''cos\phi+\eta''sin\phi)(\eta''cos\phi-\beta''sin\phi)+C_1\left(\eta''cos\phi-\beta''sin\phi\right)^2=r',$$

which is equal to

$$\begin{aligned}&\beta''^2(A_1cos^2\phi-2B_1cos\phi sin\phi+C_1sin^2\phi)+\eta''^2(A_1sin^2\phi+2B_1sin\phi cos\phi+C_1cos^2\phi)\\&+\beta''\eta''(2A_1cos\phi sin\phi+2B_1cos^2\phi-2B_1sin^2\phi-2C_1cos\phi sin\phi)-r'=0,\end{aligned} \tag{7.4.10}$$

where the coefficient of $\beta''\eta''$ is equal to

$$(A_1 - C_1)sin2\phi + 2B_1cos2\phi = (A_1 - C_1)cos2\phi\left(tan2\phi + \frac{2B_1}{A_1 - C_1}\right) = 0.$$

Therefore (7.4.10) becomes

$$A_2\beta''^2 + B_2\eta''^2 - r' = 0, \tag{7.4.11}$$

where

$$A_2 = A_1cos^2\phi - 2B_1cos\phi sin\phi + C_1sin^2\phi \tag{7.4.12}$$

and

$$B_2 = A_1sin^2\phi + 2B_1cos\phi sin\phi + C_1cos^2\phi. \tag{7.4.13}$$

We check that A_2 can be written as

$$A_1\left(cos^2\phi - 2\frac{B_1}{A_1}cos\phi sin\phi + \frac{B_1^2}{A_1^2}sin^2\phi\right) - \frac{B_1^2}{A_1}sin^2\phi + C_1sin^2\phi,$$

which is equal to

$$A_1\left(cos\phi - \frac{B_1}{A_1}sin\phi\right)^2 + \left(C_1 - \frac{B_1^2}{A_1}\right)sin^2\phi,$$

where $A_1 = \bar{k}_{10} + \bar{k}_{20}\bar{\xi}_0^2 > 0$ and

$$C_1 - \frac{B_1^2}{A_1} = \bar{k}_{20} - \frac{\bar{\xi}_0^2\bar{k}_{20}^2}{\bar{k}_{10} + \bar{k}_{20}\bar{\xi}_0^2} = \frac{\bar{k}_{10}\bar{k}_{20} + \bar{\xi}_0^2\bar{k}_{20}^2 - \bar{\xi}_0^2\bar{k}_{20}^2}{\bar{k}_{10} + \bar{k}_{20}\bar{\xi}_0^2} = \frac{\bar{k}_{10}\bar{k}_{20}}{\bar{k}_{10} + \bar{k}_{20}\bar{\xi}_0^2} > 0.$$

Therefore $A_2 > 0$ and similarly $B_2 > 0$.

Equation (7.4.10) then can be written as

$$\frac{\beta''^2}{\left(\sqrt{B_2r'}\right)^2} + \frac{\eta''^2}{\left(\sqrt{A_2r'}\right)^2} = 1. \tag{7.4.14}$$

We see that the semi-major and the semi-minor axis lengths of the ellipse in (7.4.14) are $\sqrt{B_2r'}$ and$\sqrt{A_2r'}$ and the area of (7.4.14) is $\pi r'\sqrt{A_2B_2}$.

From (7.2.1), (7.4.3) and (7.4.11), we see that the joint density for (β, η, λ) is equivalent to

$$\lambda^{\bar{\alpha}_0 - 1}e^{-\bar{\beta}_0\lambda} \cdot \lambda exp\left(-\frac{\lambda}{2}\left(\frac{\beta''^2}{B_2} + \frac{\eta''^2}{A_2}\right)\right).$$

It is straightforward to obtain the normalised joint density for (β, η, λ), which is

$$\sqrt{\frac{1}{2\pi B_2}}\sqrt{\frac{1}{2\pi A_2}}\frac{\bar{\beta}_0^{\bar{\alpha}_0}}{\Gamma(\bar{\alpha}_0)}\lambda^{\bar{\alpha}_0 - 1}e^{-\bar{\beta}_0\lambda} \cdot \lambda exp\left(-\frac{\lambda}{2}\left(\frac{\beta''^2}{B_2} + \frac{\eta''^2}{A_2}\right)\right).$$

This can be written as

$$\frac{1}{2\pi\sqrt{A_2B_2}}\frac{\bar{\beta}_0^{\bar{\alpha}_0}}{\Gamma(\bar{\alpha}_0)}\lambda^{\bar{\alpha}_0}exp(-\lambda A^*), \tag{7.4.15}$$

where $A^* = \bar{\beta}_0 + \frac{\beta''^2}{2B_2} + \frac{\eta''^2}{2A_2} = \bar{\beta}_0 + \frac{r'}{2}$.

Hence the normalised joint marginal density for (β, η) is

$$f^*(\beta,\eta) = \frac{1}{2\pi\sqrt{A_2B_2}} \frac{\bar{\beta}_0^{\bar{\alpha}_0}}{\Gamma(\bar{\alpha}_0)} \frac{\Gamma(\bar{\alpha}_0+1)}{A^{*(\bar{\alpha}_0+1)}} = \frac{\bar{\alpha}_0}{2\pi\sqrt{A_2B_2}} \frac{\bar{\beta}_0^{\bar{\alpha}_0}}{A^{*(\bar{\alpha}_0+1)}}. \tag{7.4.16}$$

So the $100(1-\alpha)\%$ joint confidence region for β and η is given by

$$1-\alpha = \int\int f^*(\beta,\eta) dregion = \int_0^{\bar{r}} f^*(\beta,\eta)\pi\sqrt{A_2B_2}dr'.$$

Substituting (7.4.16) and A^* in (7.4.15) into the above equation we have

$$1-\alpha = \frac{\bar{\alpha}_0}{2\pi\sqrt{A_2B_2}}\bar{\beta}_0^{\bar{\alpha}_0}\int_0^{\bar{r}}\left(\frac{r'}{2}+\bar{\beta}_0\right)^{-\bar{\alpha}_0-1}\pi\sqrt{A_2B_2}dr' = \frac{\bar{\alpha}_0\bar{\beta}_0^{\bar{\alpha}_0}}{2}\int_0^{\bar{r}}\left(\frac{r'}{2}+\bar{\beta}_0\right)^{-\bar{\alpha}_0-1}dr'.$$

Let $r'' = \frac{r'}{2}$ and so $dr' = 2dr''$. The above equation becomes

$$1-\alpha = \bar{\alpha}_0\bar{\beta}_0^{\bar{\alpha}_0}\int_0^{\frac{\bar{r}}{2}}\left(r''+\bar{\beta}_0\right)^{-\bar{\alpha}_0-1}dr'' = \bar{\beta}_0^{\bar{\alpha}_0}\left(\bar{\beta}_0^{-\bar{\alpha}_0} - \left(\bar{\beta}_0+\frac{\bar{r}}{2}\right)^{-\bar{\alpha}_0}\right).$$

After solving this last equation, we obtain the height $\bar{r}$ of the distribution so that the volume of the distribution which lies between the height and the top of the distribution is $1-\alpha$, which is defined in (7.4.2).

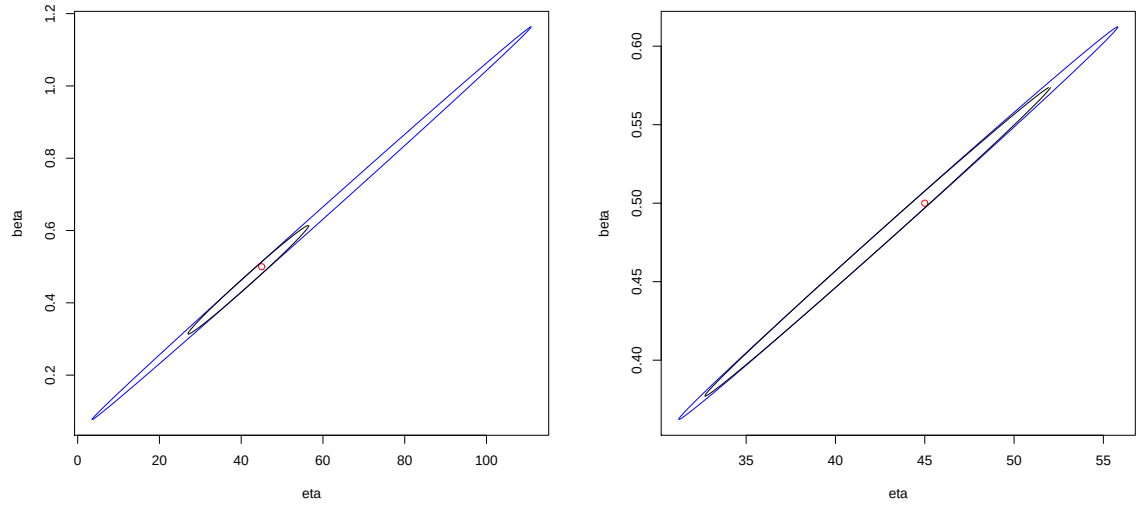
Therefore a $100(1-\alpha)\%$ joint confidence region for β and η can be easily obtained as the form (7.4.1).

Example 7.9 *Assume that the parameters are given by* $I(0) = 10,\ \beta = 0.5,\ \mu = 20,\ \gamma = 25,\ N = 100,$ *and* $\sigma = 0.03$ *for the model (3.2.4), as in Example 7.5.*

Also we assume two sets of priors, Prior A and Prior B, as in Example 7.5.

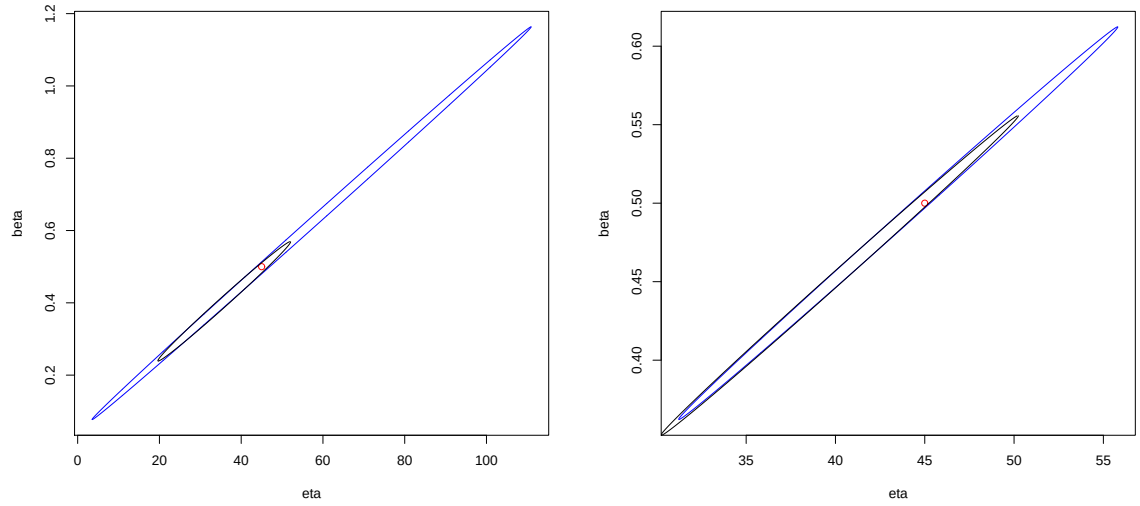
We vary the value of T *as* $T = 20$ *and* $T = 200$ *and use the same method as in Example 7.5 to simulate a dataset for each* T *and sample from each of them. When we increase* T *we increase* n *in proportion to* T*, keeping* Δt *fixed. With the sample obtained, with* $\alpha = 0.05$ *we find the Bayesian* 95% *confidence region for* β *and* η *by (7.4.1) for both priors and for each* T*. For comparison, we also obtain the* 95% *confidence region for* β *and* η *by the least squares method using (6.2.29) for each* T*. The results are shown in Figure 7.1 and Figure 7.2.*

We see that in both cases using a prior does help decrease the size of the 95% *confidence regions when* T *is small, and the* 95% *confidence region for the parameters using the least squares method is quite large. The* 95% *confidence regions for* β *and* η *with the different prior distributions become closer as* T *increases, which means that the prior distribution has less influence on the Bayesian confidence region when* T *is large.*



(a) (b)

Figure 7.1: (a) shows the 95% joint confidence region for β and η using the parameter values in Example 7.9 with prior A and $T = 20$. The results of both the Bayesian method (black) and least squares method (blue) are shown; (b) the 95% joint confidence region for β and η using the parameter values in Example 7.9 with prior A and $T = 200$. The results of both the Bayesian method (black) and least squares method (blue) are shown. The red point in both figures represents the true values $\beta = 0.5$ and $\eta = 45$.



(a) (b)

Figure 7.2: (a) shows the 95% joint confidence region for β and η using the parameter values in Example 7.9 with prior B and $T = 20$. The results of both the Bayesian method (black) and least squares method (blue) are shown; (b) the 95% joint confidence region for β and η using the parameter values in Example 7.9 with prior B and $T = 200$. The results of both the Bayesian method (black) and least squares method (blue) are shown. The red point in both figures represents the true values $\beta = 0.5$ and $\eta = 45$.

7.5 Bayesian estimation with m datasets

In this section we will discuss the Bayesian estimation in the case where we can repeat the experiment and obtain more datasets. Assuming that we have m datasets each of size n, a pseudo-likelihood for the parameters $\boldsymbol{\varphi} = (\beta, \eta, \sigma^2)$ will become

$$L_n(\boldsymbol{\varphi}) = \prod_{i=1}^{m}\prod_{j=0}^{n-1}\left(\frac{1}{\sqrt{2\pi\sigma^2 {I_{ij}}^2 \left(N - I_{ij}\right)^2 \Delta t}} exp\left(-\frac{1}{2}\frac{[I_{i,j+1} - I_{ij} - I_{ij}(\beta N - \eta - \beta I_{ij})\Delta t]^2}{\sigma^2 {I_{ij}}^2 \left(N - I_{ij}\right)^2 \Delta t}\right)\right), \tag{7.5.1}$$

modifying (6.3.2).

If the prior distribution in Theorem 7.1 stays unchanged and we substitute this likelihood into $P(D|\eta, \beta, \lambda)$ in (7.2.11), we see clearly that for the m dataset case, the steps for all the proofs stay the same except that all instances of n change to mn and all the sums of the x_k and I_k terms run from $k = 0$ to $m(n-1)$ instead of from $k = 0$ to $n-1$. Then Theorem 7.1 can easily be rewritten for the m dataset case.

Then if m datasets are available, the Bayesian estimators in Theorem 7.2, the Bayesian interval estimators in Theorem 7.6 and the Bayesian joint confidence region in Theorem 7.8 have the same expressions except that the definitions for $\bar{\alpha}_0$, $\bar{k}_{10}$, $\bar{\mu}_{10}$, $\bar{k}_{20}$, $\bar{\mu}_{20}$ and $\bar{\xi}_0$ in Theorem 7.1 are changed.

7.6 Summary

In this chapter we have applied the Bayesian method to estimate the parameters in the stochastic SIS model. We set up a conjugate prior and obtained the joint posterior distribution for the model parameters in Theorem 7.1. We then maximised the joint posterior distribution with respect to the parameters β, η and σ^2 and obtained the Bayesian estimators. We compared the Bayesian estimators with the least squares estimators in Theorem 7.2, which states that the Bayesian estimators will tend to the corresponding least squares estimators when the sample size is infinitely large. The Bayesian estimators and least squares estimators were also compared in numerical examples using two sets of priors. Example 7.5 shows that a good prior helps the estimation when the sample size is small, and the prior distribution has less influence on the Bayesian estimates when the sample size (the observation period here) is large.

We obtained the $100(1-\alpha)\%$ CIs for the Bayesian estimators by deriving the marginal distributions for these. The CIs using the Bayesian approach were compared with those from the least squares method in Example 7.7, which shows that both good and less good priors help to decrease the width of the 95% CIs when the sample size is small. The 95% CIs with the different prior distributions become closer as the sample size increases and the prior has less influence.

In Theorem 7.8 we also obtained the $100(1-\alpha)\%$ joint confidence region for β and η. Example 7.9 shows that both good and less good priors help to decrease the size of the

95% confidence region when the sample size is small, and the 95% confidence regions for β and η with different prior distributions become closer as the sample size increases. The case of Bayesian estimation when more datasets are available is discussed briefly in section 7.5. Most of the contents of this chapter have been formed into a paper and submitted to the Journal of Mathematical Biology.

In the next chapter, we will draw conclusions for this thesis and discuss further research work.

Chapter 8

Conclusions and Discussion

This thesis has discussed three stochastic versions of the SIS epidemic model and the statistical inference to the SDE SIS model. The deterministic SIS epidemic model is one of the simplest possible epidemic models, which has applications to the transmission of real-life diseases, such as pneumococcus, gonorrhea and tuberculosis. It is important to include the effect of environmental noise in the SIS epidemic model. In Chapter 1, we have reviewed four commonly used types of noise and the effect of these on the deterministic model. The effect of white noise in compartmental epidemic models has been extensively discussed. However, to our knowledge there is no existing literature which discusses this issue for the SIS epidemic model. This thesis was intended to fill this gap. The effect of telegraph noise and the combination of white noise and telegraph noise has been examined in biological models in previous papers and the significant effect of these two types of environmental noise was revealed. This was the motivation to incorporate these two types of noise into the SIS epidemic model. Furthermore, statistical inference is vital in epidemic modelling since the parameter estimates can be used to characterise the infection process and also provide information on key indicators of disease spread which will be of great assistance to the implementation of any disease control policy. That was the motivation for us to conduct parameter estimation for the SDE SIS model with white noise introduced in the second part of this thesis.

Estimation methods have been reviewed in Chapter 1. Three estimation methods, least squares estimation, pseudo-MLE method and the Bayesian approach have been applied to the SDE SIS model. Literature review shows that although least squares and the pseudo-MLE method have been widely discussed, variance estimation for these two methods has not been discussed previously, while we have discussed this in detail in this thesis. As for the Bayesian approach, although strong results have been obtained for the general SDE problems by using the MCMC technique in the existing literature, for the particular model we deal with, it was possible for our alternative approach to be applied where analytic results have been obtained.

In Chapter 2 we summarised some theory which was essential for the research carried out in this thesis. We started with basic probability theory. We then defined the stochastic process, which is an important concept since the solutions of the SDEs are stochastic processes. Brownian motion was then introduced, which is a mathematical representation of randomness, and its properties were discussed. Afterwards, stochastic integrals which

are a component of SDEs were defined and the important Itô formula was then introduced. Some fundamental concepts for Markov chains were then reviewed and the generalised Itô formula was given. We then defined a general version of the SDE and SDE with Markovian switching and also specified the corresponding existence and uniqueness criteria for the solutions. We also constructed the explicit solution for the general linear SDE. Finally we provided definitions of various types of stochastic stability which have been examined for the stochastic models we derived.

In Chapter 3, we derived an SDE version of the classical SIS epidemic model, with white noise introduced in the disease transmission term by a well established method, parameter perturbation. We proved that the SDE had a unique positive global solution and established conditions for extinction and persistence of disease. A key parameter, the basic reproduction number R_0^S, was defined and it was shown that if $R_0^S \leq 1$, under mild extra conditions the disease will die out, while if $R_0^S > 1$ then the disease will persist. The threshold R_0^S is less than the corresponding deterministic version R_0^D, which means the conditions for $I(t)$ to become extinct are weaker in the SDE SIS model. For the persistent case, the effect of the noise intensity σ on the SDE SIS model has been examined and it was shown that the level which $I(t)$ fluctuates around strictly decreases with increasing noise intensity. The range for the level was derived as well, which stayed beneath the deterministic equilibrium. We also showed for the persistent case that the model has a unique stationary distribution and derived expressions for its mean and variance. We made a conjecture about the disease behaviour if $R_0^S \leq 1$ and the other extinction conditions are not satisfied. The simulation results confirmed our conjecture. We have also illustrated our theoretical results with computer simulation in this chapter, including two realistic models for gonorrhea amongst homosexuals and pneumococcus amongst young children.

In Chapter 4, we introduced telegraph noise to the SIS epidemic model and derived the stochastic SIS model with Markovian switching. We established the explicit solution for the stochastic SIS model and also the conditions for extinction and persistence of the disease. A threshold value T_0^S was defined for the Markovian switching model to examine almost sure persistence or extinction. We started with the special case where the Markov chain has only two states and then generalised our results to the general case where the Markov chain has a finite number of states. It is shown that if $T_0^S < 1$, the disease will die out, while if $T_0^S > 1$ the disease will persist. From the definition for T_0^S and R_0^D, we can further illustrate that if not all subsystems are extinctive, then relatively faster transition rates from a persistent subsystem to an extinctive subsystem, such that $T_0^S < 1$, will make the overall system become extinct. This reveals the important roles of the Markov chain in extinction. We also showed that if $T_0^S > 1$ the number of infectious individuals will enter a certain domain $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ in finite time with probability one and will stay in the interval once it has entered, where $(0 \vee (\alpha_1/\beta_1)$ is the smallest and $\alpha_M/\beta_M)$ is the largest long-term endemic level of disease in each of the M separate SIS models between which the Markov chain switches. Moreover, the number of infectious individuals can take any value up to the boundaries of $(0 \vee (\alpha_1/\beta_1),\ \alpha_M/\beta_M)$ but never reach them. We have not been able to prove extinction for the case when $T_0^S = 1$, but the simulation result shows extinction of the disease after a long period of time. Again we have illustrated our theoretical results with computer simulations, including an example with realistic parameter values for *S.Pneumoniae* amongst children.

In Chapter 5, we take a further step of combining the effect of both white and telegraph noise in the SIS epidemic model. We set up the SDE SIS model with Markovian switching and showed the existence of a unique global positive solution for this model. A threshold value T_0^{MC} was defined for almost sure extinction and stochastic permanence. It was shown that if $T_0^{MC} < 1$, and with other three different conditions, the disease will die out, while if $T_0^{MC} > 1$, the disease will persist. We have illustrated our theoretical results with computer simulations. The extinction conditions we have shown for this model are rather strong, so we hope to improve the results in a later stage of research.

The three stochastic SIS models that we have developed have applications in different situations. The stochastic SIS model with white noise introduced is used in situations where there is environmental influence on the transmission coefficient, while the stochastic SIS model with telegraph noise introduced is used in situations where there is abrupt change on all model parameters. The stochastic SIS model incorporating the above mentioned two types of noise can be used to describe both situations. For the stochastic SIS model with white noise introduced, unlike the telegraph noise case, we only introduced the noise in one parameter. This is mainly because the disease transmission coefficient is more likely to be affected by environmental influences compared to the other model parameters, i.e. the death rate and the cure rate. A threshold value, which is used to indicate whether the system will be extinctive or persistent, was derived for each stochastic model. The threshold value for the stochastic model with white noise introduced combines the influence of noise intensity into the original threshold value for the deterministic model, while the threshold value for the stochastic SIS model with telegraph noise introduced combines the influence of the Markov chain into the original deterministic threshold value. For the stochastic SIS model incorporating both types of noise, the threshold value combines both influences. Regarding the techniques we used to show extinction and persistence of the system for the three models, unlike the technique we used for the first two models, the M-matrix technique was applied to the stochastic SIS model with the combined types of noise. This is because, by applying the M-matrix technique, we can show persistence of the system without additional conditions.

In Chapter 6, we applied the pseudo-MLE and the least squares method to estimate the parameters in the SDE SIS model. For the least squares method, we started with the case in which only one dataset is available and then improved our estimation by considering the case where more datasets are available. We have obtained the point estimators, $100(1-\alpha)\%$ CIs and $100(1-\alpha)\%$ joint confidence regions for β and η for both cases. We also investigated the factors which influence the efficiency of the estimation. By doing so, we showed that the asymptotic widths of the CIs for both β and η strictly decrease as the total time period T increases and do not depend on the size of the time step Δt. Simulation examples also show that the efficiency of the estimation is not improved with denser observations (smaller time step) but is improved with a larger numbers of samples m or one sample with a longer observation period T. Pseudo-MLEs are also obtained, which are almost the same as the point estimators from the least squares case, with a minor difference in the estimators of σ^2. For obtaining the confidence region for the pseudo-MLEs, the following two cases were considered: When the number of samples m is small, the exact confidence region for β and η was obtained based on the variance estimation from the least squares results. When m is large, the asymptotic MLE theory

and the likelihood ratio test approach were applied to obtain the large sample confidence regions for both the three dimensional case (using all three MLEs) and the two dimensional case (estimating β and η assuming that σ is known). We only calculated numerically the asymptotic confidence region based on the likelihood ratio test for β and η, which shows that the numerical asymptotic confidence region using the likelihood ratio test for β and η is almost identical to the exact confidence region.

In Chapter 7, we have applied the Bayesian method to estimate the parameters in the stochastic SIS model. We set up a conjugate prior and obtained the joint posterior distribution for the model parameters. We then maximised the joint posterior distribution with respect to the parameters and obtained the Bayesian estimators. The Bayesian estimators were compared with the least squares estimators analytically, which showed that the Bayesian estimators will tend to the corresponding least squares estimators when the sample size is infinitely large. The $100(1-\alpha)\%$ CIs for the Bayesian estimators were obtained by deriving the marginal posterior distributions. The $100(1-\alpha)\%$ joint confidence region for β and η has also been obtained by integrating the joint posterior distribution. Numerical simulations have been conducted to compare the Bayesian estimation and the least squares estimation. The simulation results show that both good and less good priors help to improve the efficiency of estimation and the estimation results from the two methods become closer as the sample size increases, when the prior has less influence. The case of Bayesian estimation when more datasets are available is discussed briefly as well.

Comparing the three estimation methods, the Bayesian method is advantageous when reliable prior information is available and the sample size is small. The least squares estimation is generally better than the pseudo-MLE method in our case since the least squares estimation gave the same point estimators and joint confidence region as the pseudo-MLE and is easier to apply, although approximating MLE methods are more preferred in the existing literature.

Due to the random nature of the population system, a lot of attention has been paid to incorporating stochastic noise into deterministic systems in recent research. The introduction of the stochastic part to the model complicates the system significantly but the randomness incorporated may possibly explain the real-life situation better. Three different types of noise have been incorporated in the deterministic SIS model in this thesis. The model is non-linear, which means the standard stochastic theory cannot be directly applied. The methodology that we used in developing the stochastic model and examining the asymptotic properties, including extinction and persistence, can be applied to other biological or epidemic models. Particularly, telegraph noise and a combination of white and telegraph noise have not been discussed in epidemic models in the existing literature. It would be interesting to see how these two types of noise can affect other epidemic models. There are other types of environmental noise including the one we discussed in section 3.8, which we can pursue further for the SIS epidemic model. Stochastic SIS models with demographic noise have been developed recently [1]. Compared to our stochastic SIS models, the procedure for deriving a stochastic SIS model with demographic noise is more complicated and the existence of a positive solution has not been shown. Also, compartmental models with varying total population size have become one of the important research areas. Our techniques, which were used to develop a stochastic SIS

model with white noise, can be applied to the SIS model with varying total population size and we expect more extinction conditions related to population size $N(t)$.

Three parameter estimation methods have been applied to the SDE SIS model in this thesis. The methods were applied to this specific model but all are based on the fact that the noise introduced follows a normal distribution. Therefore the methodology we applied in the statistical inference part of the thesis can be applied to other SDEs. It is hoped that this thesis has made a useful contribution in this respect.

Bibliography

[1] Allen, E., *Modelling with Itô Stochastic Differential Equations*, Dordrecht: Springer-Verlag, 2007.

[2] Anderson, D.R., Optimal exploitation strategies for an animal population in a Markovian environment: a theory and an example, *Ecology* 56, 1281-1297, 1975.

[3] Anderson, W.J., *Continuous-Time Markov Chains*, Berlin: Springer, 1991.

[4] Arnold, B.C. and Shavelle, R.M., Joint confidence sets for the mean and variance of a Normal distribution, *The American Statistician*, 52(2), 133-140, 1998.

[5] Arnold, I., *Stochastic Differential Equations: Theory and Applications*, New York: Wiley, 1972.

[6] Anderson, R.M. and May, R.M., Population biology of infectious diseases I, *Nature*, 180, 61-367, 1979.

[7] Bailey, N.T.J., Some stochastic models for small epidemics in large populations, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 13(1), 9-19, 1964.

[8] Barlett, M.S., Deterministic and stochastic models for recurrent epidemics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 4, 81-108, 1956.

[9] Basak, G.K., Bisi, A. and Ghosh, M.K., Stability of a random diffusion with linear drift, *Journal of Mathematical Analysis and Applications*, 202, 604-622, 1996.

[10] Bishwal, J.P.N., *Parameter Estimation in Stochastic Differential Equations*, Berlin/Heidelberg: Springer-Verlag, 2008.

[11] Benenson, A.S., *Control of Communicable Diseases in Man*, Fifteenth Edition, Washington: American Public Health Association, 1990.

[12] Boys, R.J. and Giles, P.R., Bayesian inference for stochastic epidemic models with time-inhomogeneous removal rates, *Journal of Mathematical Biology*, 55, 223-247, 2007.

[13] Brauer, F., Allen, L.J.S., Van den Driessche, P. and Wu, J., *Mathematical Epidemiology* Lecture Notes in Mathematics, No. 1945, Mathematical Biosciences Subseries, New York: Springer-Verlag, 2008.

[14] Brugger, S.D., Hathaway, L.J. and Mühlemann, K., Detection of *Streptococcus Pneumoniae* strain cocolonization in the nasopharynx, *Journal of Clinical Microbiology*, 47(6), 1750-1756, 2009.

[15] Busenberg, S., Cooke, K. and Hsieh, Y.H., A model for HIV in Asia, *Mathematical Biosciences*, 128, 185-210, 1995.

[16] Busenberg, S. and van den Driessche, P., Analysis of a disease transmission model in a population with varying size, *Journal of Mathematical Biology*, 28, 257-270, 1990.

[17] Cao, J. and Hu, L., Asymptotic properties of a pseudo-MLE for CIR model, Proceedings of the 2010 International Conference on IIGSS-CPS, Nanjing, China, Vol. 1: Advances on Probability and Statistics, 28-31, 206-209, 2010.

[18] Caswell, H. and Cohen, J.E., Red, white and blue: Environmental variance spectra and coexistence in metapopulations, *Theoretical Population Biology*, 176, 301-316, 1995.

[19] Chen, G. and Li, T., Stability of a stochastic delayed SIR model, *Stochastics and Dynamics*, 9(2), 231-252, 2009.

[20] Coffey, T.J., Enright, M.C., Daniels, M., Morona, J.K., Morona, R., Hryniewicz, W., Paton, J.C. and Spratt, B.G., Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus Pneumoniae*, *Molecular Microbiology*, 27, 73-83, 1998.

[21] Dalal, N., Greenhalgh, D. and Mao, X., A stochastic model of AIDS and condom use, *Journal of Mathematical Analysis and Applications*, 325, 36-53, 2007.

[22] Derrick, W.R. and van den Driessche, P., A disease transmission model in a nonconstant population, *Journal of Mathematical Biology*, 31, 494-512, 1993.

[23] Ding, Y., Xu, M. and Hu, L., Asymptotic behaviour and stability of a stochastic model for AIDS transmission, *Applied Mathematics and Computation*, 204, 99-108, 2008.

[24] Douglas, J.B., Confidence regions for parameter pairs, *The American Statistican*, 41(1), 43-45, 1993.

[25] Du, N.H., Kon, R., Sato, K. and Takeuchi, Y., Dynamical behaviour of Lotka-Volterra competition systems: non autonomous bistable case and the effect of telegraph noise, *Journal of Computational Applied Mathematics* 170, 399–422, 2004.

[26] Engen, S. and Lande, R., Population dynamic models generating the lognormal species abundance distribution, *Mathematical Biosciences*, 132, 169-183, 1996.

[27] Farrington, P., What is the reproduction number for pneumococcal infection, and does it matter? In: *4th International Symposium on Pneumococci and Pneumococcal Diseases, May 9-13, 2004 at Marina Congress Center, Helsinki, Finland*, 2004.

[28] Feng, Z., Huang, W. and Castillo-Chavez, C., Global behaviour of a multi-group SIS epidemic model with age-structure, *Journal of Differential Equations*, 218(2), 292-324, 2005.

[29] Foley, P., Predicting extinction times from environmental stochasticity and carrying capacity, *Conservation Biology*, 8(1), 124-137, 1994.

[30] Fox, G.A., Life-history evolution and demographic stochasticity, *Ecology and Evolution*, 7, 1-14, 1993.

[31] Friedman, A., *Stochastic Differential Equations and Their Applications*, New York: Academic Press, 1976.

[32] Gamerman, D., *Markov Chain Monte Carlo*, London: Chapman and Hall, 1997.

[33] Gard, T.C., *Introduction to Stochastic Differential Equations*, New York: Marcel Dekker, 1988.

[34] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., *Bayesian Data Analysis*, London: Chapman and Hall, 1995.

[35] Gilpin, M.E., *Predator-Prey Communities*, Princeton and London: Princeton University Press, 1975.

[36] Golightly, A. and Wilkinson, D.J., Bayesian inference for nonlinear multivariate diffusion models observed with error, *Computational Statistics and Data Analysis*, 52, 1674-1693, 2008.

[37] Gopalsamy, K., *Stability and Oscillations in Delay Differential Equations of Population Dynamics*, Dordrecht: Kluwer Academic, 1992.

[38] Grafton, R.Q., Kampas, T. and Lindenmayer, D., Marine reserves with ecological uncertainty, *Bulletin of Mathematical Biology*, 67, 957-971, 2005.

[39] Gray, A., Greenhalgh, D., Hu, L., Mao, X. and Pan, J., A stochastic differential equation SIS epidemic model, *SIAM Journal of Applied Mathematics*, 71(3), 876-902, 2011.

[40] Gray, A., Greenhalgh, D., Mao, X. and Pan, J., The SIS epidemic model with Markovian switching, *Journal of Mathematical Analysis and Applications*, 394(2), 496-516, 2012.

[41] Greenhalgh, D., An epidemic model with density-dependent death rate, *IMA Journal of Mathematics Applied in Medicine and Biology*, 7, 1-26, 1990.

[42] Greenhalgh, D., Some threshold and stability results for epidemic models with a density dependent death rate, *Theoretical Population Biology*, 42, 130-151, 1992.

[43] Greenhalgh, D., Some results for an SEIR epidemic model with density dependence in the death rate, *IMA Journal of Mathematics Applied in Medicine and Biology*, 9, 67-106, 1992.

[44] Has'minskii, R.Z., *Stochastic Stability of Differential Equations*, Moscow: Sijthoff and Noordhoff, 1980.

[45] Hayakawa, Y., O'Neill, P.D., Upton, D. and Yip, P.S.F., Bayesian inference for a stochastic epidemic model with uncertain numbers of susceptibles of several types, *Australian and New Zealand Journal of Statistics*, 45, 491-502, 2003.

[46] Hethcote, H. W., The mathematics of infectious diseases, *Society for Industrial and Applied Mathematics*, 42, 599C653, 2000.

[47] Hethcote, H.W. and Yorke, J.A., *Gonorrhea Transmission Dynamics and Control*, Lecture Notes in Biomathematics 56, Berlin: Springer-Verlag, 1994.

[48] Iacus, S.M., *Simulation and Inference for Stochastic Differential Equations with R Examples*, New York: Springer, 2008.

[49] Iannelli, M., Milner, F. A. and Pugliese, A., Analytical and numerical results for the age-structured SIS epidemic model with mixed inter-intracohort transmission, *SIAM Journal of Mathematical Analysis* 23(3), 662–688, 1992.

[50] Keeling, M.J. and Ross, J.V., On methods for studying stochastic disease dynamics, *Journal of the Royal Society Interface*, 5, 171-181, 2008.

[51] Kermack, W.O. and McKendrick, A.G., Contributions to the mathematical theory of epidemics. Part I, *Proceedings of the Royal Society Series A*, 115, 700-721, 1927.

[52] Kloeden, P. E. and Platen, E., *Numerical Solution of Stochastic Differential Equations*, Berlin and New York: Springer, 1992.

[53] Kristensen, N.R., Madsen, H. and Young, P.C., Parameter estimation in stochastic grey-box model, *Automatica*, 40, 225-237, 2004.

[54] Kurtz, T., Solutions or ordinary differential equations as limits of pure jump Markov process, *Journal of Applied Probability*, 7, 49-58, 1970.

[55] Kurtz, T., Limit theorems for sequences of jump Markov processes approximating ordinary differential processes, *Journal of Applied Probability*, 8, 344-356, 1971.

[56] Lajmanovich, A. and Yorke, J.A., A deterministic model for gonorrhea in a nonhomogeneous population, *Mathematical Biosciences*, 28, 221-236, 1976.

[57] Lamb, K.E., Greenhalgh, D. and Robertson, C., A simple mathematical model for genetic effects in pneumococcal carriage and transmission, *Journal of Computational and Applied Mathematics*, 235, 1812-1818, 2011.

[58] Lande, R., Engen, S. and Saether, B.-E., Spatial scale of population synchrony: Environmental correlation versus dispersal and density regulation, *The American Naturalist*, 154(3), 271-281, 1999.

[59] Li, J., Ma, Z. and Zhou, Y., Global analysis of an SIS epidemic model with a simple vaccination and multiple endemic equilibria, *Acta Mathematica Scienta* 26, 83–93, 2006.

[60] Li, X., Gray, A., Jiang, D. and Mao, X., Sufficient and necessary conditions of stochastic permanence and extinction for stochastic logistic populations under regime switching, *Journal of Mathematical Analysis and Applications*, 376, 11-28, 2011.

[61] Lipsitch, M., Vaccination against colonizing bacteria with multiple serotypes, *Proceedings of the National Academy of Sciences*, 94, 6571-6576, 1997.

[62] Loève, M., *Probability Theory*, New York: D. Van Nostrand Company Inc., 1963.

[63] Lu, Q., Stability of SIRS system with random perturbations, *Physica A*, 388, 3677-3686, 2009.

[64] Luo, Q. and Mao, X., Stochastic population dynamics under regime switching, *Journal of Mathematical Analysis and Applications*, 334, 69-84, 2007.

[65] Luo, Q. and Mao, X., Stochastic population dynamics under regime switching II, *Journal of Mathematical Analysis and Applications*, 355, 577-593, 2009.

[66] Mao, X., *Stability of Stochastic Differential Equations with Respect to Semimartingales*, New York: Longman Scientific and Technical, 1991.

[67] Mao, X., *Exponential Stability of Stochastic Differential Equations*, New York: Marcel Dekker, 1994.

[68] Mao, X., *Stochastic Differential Equations and Applications*, 2nd Edition, Chichester, UK: Horwood, 2007.

[69] Mao, X. and Yuan, C., *Stochastic Differential Equations with Markovian Switching*, London: Imperial College Press, 2006.

[70] Mao, X., Yin, G. and Yuan, C., Stabilization and destabilization of hybrid systems of stochastic differential equations, *Automatica*, 43, 264-273, 2007.

[71] McCormack, R.K. and Allen, L.J.S., Stochastic SIS and SIR multihost epidemic models. *Proceedings of the Conference on Differential and Difference Equations and Applications*, R.P. Agarwal and K. Perera, Eds., Hindawi Publishing Corporation, 775-786, 2006.

[72] Mood, A.M., *Introduction to the Theory of Statistics*, New York: McGraw-Hill, 1950.

[73] Nasell, I., Moment closure and the stochastic logistic model, *Theoretical Population Biology*, 63, 159-168, 2003.

[74] Nasell, I., An extension of the moment closure method, *Theoretical Population Biology*, 64, 233-239, 2003.

[75] Nielsen, J.N., Madsen, H. and Young, P.C., Parameter estimation in stochastic differential equations: an overview, *Annual Reviews in Control*, 24, 83-94, 2000.

[76] Nold, A., Heterogeneity in disease transmission modelling, *Mathematical Biosciences*, 52, 227-240, 1980.

[77] O'Neill, P.D., A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov Chain Monte Carlo methods, *Mathematical Biosciences*, 180, 103-114, 2002.

[78] O'Neill, P.D. and Roberts, G.O., Bayesian inference for partially observed stochastic epidemics, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 62, 121-129, 1999.

[79] Padilla, D.K. and Adolph, S.C., Plastic inducible morphologies are not always adaptive: the importance of time delays in a stochastic environment, *Ecology and Evolution* 10, 105-117, 1996.

[80] Peccoud, J. and Ycart, B., Markovian modeling of gene-product synthesis, *Theoretical Population Biology*, 48(2), 222-234, 1995.

[81] Rawlings, J.O., *Applied Regression Analysis: a Research Tool*, Belmont, CA: Wadsworth, 1988.

[82] Roberts, G.O. and Stramer, O., On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm, *Biometrika*, 88(3), 603-621, 2001.

[83] Slatkin, M., The dynamics of a population in a Markovian environment, *Ecology*, 59, 249–256, 1978.

[84] Spagnolo, B., Fiasconaro, A. and Valenti, D., Noise induced phenomena in Lotka-Volterra systems, *Fluctuation and Noise Letters*, 3, L177-L185, 2003.

[85] Streftaris, G. and Gibson, G.J., Bayesian inference for stochastic epidemics in closed populations, *Statistical Modelling*, 4, 63-75, 2004.

[86] Takeuchi, Y., *Global Dynamical Properties of Lotka-Volterra Systems*, Singapore: World Scientific Publishing Company, 1996.

[87] Takeuchi, Y., Du, N.H., Hieu, N.T. and Sato, K., Evolution of predator-prey systems described by a Lotka-Volterra equation under random environment, *Journal of Mathematical Analysis and Applications*, 323, 938–957, 2006.

[88] Thieme, H.R., Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations, *Mathematical Biosciences*, 111, 99-130, 1992.

[89] Timmer, J., Parameter estimation in nonlinear stochastic differential equations, *Chaos, Solitons and Fractals*, 11, 2571-2578, 2000.

[90] Tornatore, E., Buccellato, S.M. and Vetro, P., Stability of a stochastic SIR system, *Physica A*, 354, 111-126, 2005.

[91] Trottier, H. and Philippe, P., Deterministic modeling of infectious diseases: applications to measles and other similar infections, *The Internet Journal of Infectious Diseases*, 2(1), 2002.

[92] Van den Driessche, P. and Watmough, J., A simple SIS epidemic model with backward bifurcation, *Journal of Mathematical Biology*, 40, 525-540, 2000.

[93] Weir, A., *Modelling the impact of vaccination and competition on pneumococcal carriage and disease in Scotland*, Unpublished Ph.D. Thesis, University of Strathclyde, Glasgow, Scotland, 2009.

[94] Yorke, J.A., Hethcote, H.W. and Nold, A., Dynamics and control of the transmission of gonorrhea, *Sexually Transmitted Diseases*, 5, 51-56, 1978.

[95] Young, P., Parameter estimation for continuous-time models - a survey, *Automatica*, 17(1), 23-39, 1981.

[96] Zhang, Q., Arnaoutakis, K., Murdoch, C., Lakshman, R., Race, G., Burkinshaw, R. and Finn, A., Mucosal immune responses to capsular pneumococcal polysaccharides in immunized preschool children and controls with similar nasal pneumococcal colonization rates, *Pediatric Infectious Diseases Journal*, 23, 307-313, 2004.