

**Bayesian Inference from Discretely Observed Epidemic
Model Using Multidimensional Diffusion Approximation
Approach**

By

Aliu, Abbas Hassan

(06/68FE006)

PGD, MSc. Statistics (Ilorin)

**Thesis Submitted to the department of Statistics, University
of Ilorin, Ilorin for the award of Doctor of Philosophy (Ph.D)
in Statistics**

August, 2018.

Certification

This is to certify that this thesis is an original work carried out by ALIU, Abbas Hassan with Matriculation Number 06/68FE006 in the Department of Statistics, Faculty of Physical Sciences, University of Ilorin. The thesis has been read and approved as meeting the requirements for the award of Doctor of Philosophy of the University of Ilorin, Ilorin, Nigeria.

.....

Prof. R.A. Ipinyomi

(Supervisor)

.....

Date

.....

Dr. G. M. Oyeyemi

(Dept. Postgraduate-coordinator)

.....

Date

.....

Prof. O. M. Bamigbola

(Internal-External Examiner)

.....

Date

.....

Dr. A. O. Adejumo

(Head of Department and Chief Examiner)

.....

Date

.....

(External Examiner)

.....

Date

Dedication

This thesis is dedicated to my late mother's and my family.

Acknowledgements

Glory is to Almighty Allah, the most Gracious, the most merciful, for sparing my life, blessed me with the knowledge and wisdom to conceive, nurse and seeing me executing this ideal of attaining this level or feet in life, and for the beginning of good thing in my life and bringing it into this level of my academic career. I would like to thank a number of people who have accompanied me during the writing of this thesis.

First and foremost, I am extremely grateful to my supervisors, Prof. R.A. Ipinyomi and Dr. A.A. Abiodun, for their generous advice, unwavering support and unbounded enthusiasm, without which this thesis would not have come into fruition. I also express my gratitude to all lecturers in the Department of Statistics, University of Ilorin, for their respective input at one time or the other. I thank them for teaching me the most important lesson. Thank you.

My family has been a constant source of support, and I greatly acknowledge their personal way of understanding my work. I owe an immeasurable debt to my beloved brother Mr. Olushola I. Aliu whose served as father from day one of my academic career and the rest of my family for all the love, kindness and support (both emotionally and financially) they have given me. For the numerous

times I have bored them as they enthusiastically ask what I'm doing, and for convincing me to pursue this deranged venture in the first place. I pray that you will all reap the fruits of your labour.

Finally, to those friends and colleagues in the department of Statistics university of Ilorin, especially those who have made my time there so enjoyable, i pray that Almighty God will continue rewarding you His best rewards in this world.

Thanks and God bless.

Abbas Hassan **ALIU**.

August, 2018.

Abstract

Most epidemic data discretely observed are subjected to environmental influence. The dynamic of such data can well be described by discretised-version of Stochastic Differential Equations (SDEs). However, estimating the parameter for such a process proves to be challenging in practice. The difficulty that underlies this was the general intractability of the transition density, resulting into the likelihood functions that are not in closed forms. A direct implementation of Bayesian method of estimation results into convergence problems. Indeed, the literature has not adequately addressed these challenges for multi-dimensional cases. The aim of the study was to find the estimation procedure which would not suffer from these challenges. The objectives of these study were to: (i)

derive epidemic models for explaining the behaviour of SDEs models, (ii) propose an improved diffusion bridge sampler capable of overcoming convergence problems, (iii) propose Bayesian data-augmentation method of parameters estimation which would not suffer from convergence, and (iv) compare the performance of the proposed methods with existing methods using simulated and real-life datasets.

The exact likelihood functions of SDEs are very rare in practice. A numerical approximation approach through discretization of the process using Euler-Maruyama scheme was adopted. Data were simulated by introducing augmented values between every two consecutive time points to allow Euler scheme converge to the true continuous-time SDEs. Then, Bayesian data-augmentation method of estimation was performed for such problems via Markov Chain Monte Carlo (MCMC) method of sampling. Posterior means (PM), posterior credible interval (PCI), acceptance probability, trace plot, Autocorrelation function (ACF) plot and the density plot were use to assess the performance of the estimators under different number of intermediate sub-interval data point.

The findings of the study were that:

- (i) the evolution of SDEs (diffusion process) was found to be the most suitable trajectory that mirrored the exact dynamic epidemic models' influence;
- (ii) amongst the diffusion sampler considered, the proposed diffusion bridge sampler was found to be better when the number of augmented values tend to infinity;
- (iii) the proposed method of parameter estimate was found not to suffer from convergence problem; and
- (iv) results of assessment from epidemic outbreak in the simulation study and in two different real-life datasets revealed that the proposed method of parameter estimate was better than existing methods.

The study concluded that modeling from discretely-observed Stochastic Differential Equations (SDEs) using Bayesian data-augmentation approach provided alternative method of estimation to obtained parameter estimates of any natural dynamic phenomena that experience random variations. It is therefore recommended that the epidemiologists should be encouraged to apply diffusion approximation approaches whenever epidemic models are subjected to environmental influence.

Table of contents

Title page	i
Certification	ii

Dedication	iii
Acknowledgment	iv
Abstract	v
Table of contents	vi
Bibliography	ix
List of Tables	x
List of Figures.....	x
List of Algorithms	xii
List of Abbreviations	xiii
Chapter One	1
Background of the study	1
1.0 Introduction	1
1.1 Stochastic Differential Equations	2
1.2 Diffusion process	3
1.3 Epidemiology and epidemiological data	4
1.3.1Epidemiological Data.....	5
1.4 The Deterministic Model	6
1.5 The Stochastic Model.....	6
1.6 Compartment models	7
1.6.1 SIR-Epidemic Model	8
1.6.2 SEIR-Epidemic model.....	9
1.6.3 SEQIHR-Epidemic Model	11
1.7 Itô Calculus and Itô's formula	12

1.8 Bayesian Framework	13
1.9 Statement of the problem	14
1.10 Aim and Objectives of the study	15
1.12 Significant of the study	16
Chapter Two	18
Literature Review	18
2.1 History of Epidemic Modelling	18
2.2 Models for infectious diseases	21
2.3 Stochastic differential equations (SDEs)	22
2.3.1 Brownian Motion	24
2.3.2 Brownian Motion	26
2.3.3 Itô process	26
2.3.4 Discretisation of Continuous-time process	26
2.3.5 Euler-Maruyama Numerical Scheme	27
2.3.6 Milstein Numerical Scheme	27
2.3.7 Runge-Kutta Numerical Scheme	28
2.4 Methods of Parameter Estimation under Diffusion models	29
2.4.1. Approximate Maximum Likelihood Estimation Approach	29
2.4.2. Analytical Approximation of the Likelihood Function	30
2.4.3. The Kolmogorov Forward Equation	32
2.4.4. Simulated Maximum Likelihood Estimation	33
2.4.5. Local Linearisation	34
2.4.6. Alternatives to Maximum Likelihood Estimation	35

2.4.7. Estimating Functions	35
2.4.8. Martingale Estimating Functions	36
2.4.9. Polynomial Estimating Functions	37
2.4.10. Estimating Functions based on Eigen-functions	38
2.4.11. Simple Estimating Functions	38
2.4.12. Generalised Method of Moments	39
2.4.13. Simulated Moments Estimation	40
2.4.14. Indirect Inference	41
2.4.15. Efficient Method of Moments	42
2.4.16. Exact Algorithm	43
2.5 Bayesian Approach	44
2.5.1 Prior distribution	46
2.5.2 Objective priors versus Subjective priors.....	46
2.5.3 Proper and Improper priors	48
2.5.4 Conjugate priors	48
2.5.5 Jeffreys' prior	49
2.5.6 Maximum Entropy (MAXENT) prior.....	49
2.5.7 Reference Priors.....	50
2.5.8 Bayesian Point Estimation.....	50
2.5.9 Bayesian Interval Estimation	51
2.6 Sampling methods	53
2.6.1 Markov Chain Monte Carlo	53
2.6.2 The Metropolis-Hastings algorithm	54

2.6.3 Metropolis Method (A symmetric proposal)	55
2.6.4 Random walk Metropolis	55
2.6.5 The independence sampler	56
2.6.6 Gibbs sampling	57
2.6.7 Componentwise transitions (Metropolis-within-Gibbs)	58
2.6.8 Blocking	58
2.6.9 Model convergence	59
2.6.10 Burn –in.....	61
Chapter Three	62
Methodology	62
3.0 Epidemic Data.....	62
3.1 Derivation of epidemic models in diffusion approximation processes	63
3.1.1 Derivation of stochastic SIR epidemic model	64
3.1.2 Cholesky Decomposition	66
3.1.3 Derivation of stochastic SEIR epidemic model	68
3.1.4 Derivation of stochastic SEQIHR epidemic model	70
3.2 Model Parameters Estimation	73
3.2.1 Likelihood Function	74
3.2.2 Bayesian Approach	75
3.2.3 Data Augmentations.....	76
3.3 Path Update of Diffusion process.....	78
3.3.1 Generalization to Several Observation Times.....	80
3.3.2 Single site updating	81

3.3.3 Block updating	82
3.3.4 Convergence problems.....	84
3.4 Diffusion Bridge Constructs	85
3.4.1 Euler Scheme proposal Sampler	87
3.4.2 Modified Diffusion Bridge proposal	88
3.4.3 Lindström Bridge (Regularized sampler)	89
3.4.4 Guided proposals	90
3.5 Improved Diffusion Bridge Sampler.....	92
3.5.1 Modified Guided proposal	92
3.6 Diffusion Epidemic Model with Latent Data	95
3.6.1 Procedures for the path update	96
3.7 Observation with measurement error	98
3.8 Summary of bridge constructs	100
3.9 Parameter update for diffusion processes	101
3.9.1 Modified Innovation Scheme	103
3.10 R-package	105
3.11 Discussion	105
Chapter Four	107
Analysis of Results	107
4.0 Simulation study	107
4.1 Behaviour of Diffusion epidemic Model.....	107
4.2 Diffusion Bridge for Modified Guided proposal	109
4.2.1 Bridge Construct performance	111
4.3 Parameter estimation	113
4.3.1 Simulation study on SEIR-Epidemic Model	120

4.4 Application to real life datasets.....	125
4.4.1 Application I: Abakaliki smallpox data	125
4.4.2 Ebola outbreak in the Democratic Republic of Congo	129
4.5 Findings	133
Chapter Five	137
Summary, Conclusion & Recommendations	137
5.1 Summary	137
5.2 Conclusion	139
5.3 Contribution to Knowledge's	141
5.4 Recommendations	141
5.5 Suggestions for Further study.....	141
Bibliography.....	143
Appendix.....	156

List of Tables

Table 3.1 Summary of the infinitesimal means and covariance matrix of the Epidemic models under study.....	73
Table 3.2 Summaries of infinitesimal mean and covariance for Diffusion Bridge proposal without measurement error	100
Table 3.3(a) Summaries of infinitesimal means of Diffusion Bridge with measurement error	100
Table 3.3(b) Summaries of infinitesimal variance of Diffusion Bridge with measurement Error	101

Table 4.1 Empirical acceptance probabilities based on a run of 10000 MCMC iterations using four different proposal schemes (<i>MDB, LB, GP, MGP</i>).....	112
Table 4.2 Parameter estimate for posterior means and credible intervals for different number of imputed values (<i>m=2,5,15 and 50</i>) and their acceptance probability (AP).	115
Table 4.3 Parameter estimate for posterior means and 95% posterior credible intervals for different number of imputed values (<i>m =2,5,15 and 50</i>) and their acceptance Probability(AP).....	121
Table 4.4 Smallpox cases in Abakaliki, Nigeria during 1967, taken from Thompson and Foege (1968). Compounds listed are those before the move of cases 7 and 8, and 2 other uninfected individuals, on day 25 from compound 1 to compound 2.....	126
Table 4.5 Parameter estimate for posterior means and credible intervals for different number of imputed values (<i>m =7 and 21</i>) and their acceptance probability(AP).....	128
Table 4.6 Parameter estimate for posterior means and credible intervals for 7days number of imputed values (<i>m=7,14 and 21</i>) and their acceptance probability (AP).....	132

List of Figures

Figure 1.0 Shows a flow diagram/chart for the SIR epidemic Model.....	9
Figure 1.1 Shows a flow diagram/chart for the SEIR epidemic Model.....	10
Figure 1.2 Shows a flow diagram/chart for the SEQIHR epidemic Model	11
Figure 2.0 Shows three trajectories of one-dimension sample Brownian motion.....	25
Figure 2.1 Shows the well mixing trace plot (Stationarity) and bad mixing trace plot.....	60
Figure 2.2 Shows highest present of autocorrelation (a) and reducible or low present of Autocorrelation (b) plot.....	60
Figure 3.1 Shows an illustration of one-dimensional path segment consisting of discrete data points $x_0, \dots, x_m$ at times $t_0, \dots, t_m$. The observed data x_0, x_m is labelled with red circle, the imputed data $x_1, \dots, x_{m-1}$ with circle	79
Figure 3.2 Shows an illustration of d-dimensional path consisting of the initial value and m discrete observations at times $\tau_0, \dots, \tau_M$ (labelled with red circles) and imputed data (labelled with white circles).....	80
Figure 3.3 Shows an illustration of two-dimensional path with an observable component X_t and a latent component L_t . The observed part of the path consists of the initial value at time τ_0 and m discrete observations at times $\tau_0, \dots, \tau_M$ (labelled with red circles) and latent or imputed data (labelled with white circles).....	96
Figure 4.1 Show evolution of (a) number of SIR, (b) fraction of SIR epidemic model.....	108
Figure 4.2 Shows the evolution of (a) ODEs and (b) SDEs (diffusion process) trajectories.....	109

Figure 4.3 Shows 10000 path proposals for the SIR-Epidemic model on $(\beta, \gamma) =$

$(0.330, 0.178)$ with time interval $[0,1]$ fulfilling the SIR diffusion epidemic
conditioned on $(s_0, i_0) = (71,79)$ and $(s_T, i_T) = (96,71)$ imputed, with
subintervals of $m = (5,50,200)$ generated using Modified Guided diffusion
Bridges (MGP).....110

Figure 4.4 Shows 10000 path proposals for the SIR-Epidemic diffusion model using $(\beta, \gamma) =$

$(0.33, 0.178)$ with time intervals $[0, 1]$, fulfilling the SDE initial conditioned of
 $(s_0, i_0) = (50,50)$ & $(s_T, i_T) = (70,30)$ imputed, with subintervals of
 $m = (5,20)$ generated from three different Diffusion Bridges (MBD, GP, and
MGP).....111

Figure 4.5 Shown acceptance probabilities plot for fully observed based on a run of 1000

MCMC iterations for four different proposal schemes.....112

Figure 4.6 Shown acceptance probabilities plot for partially observed based on a run of 1000

MCMC iterations for four different proposal schemes.....113

Figure 4.7 Show the trace plots for the parameter of interest $\vartheta = (\beta, \gamma)$ of SIR-epidemic

model. The true parameter value for the parameters was indicated by the red
horizontal line.....116

Figure 4.8 Show estimation of the posterior density transmission rate (β) and recovery rate

(γ) . Density estimation takes into account the full Markov chain, i.e. without
thinning, after having discarded a 500 iterations as burn-in periods.....117

Figure 4.9 Shows Autocorrelation plots of the two parameter of interest β and γ . The

Calculation of the autocorrelation takes into account the full Markov chain, i.e.
without thinning, after having discarded a 500 burn-in periods.....119

Figure 4.10 Shows the trace plots for the parameter of interest β , γ and α of diffusion SEIR-epidemic model. The true parameter value for the parameters was indicated by the red horizontal line.....	122
Figure 4.11 Shows estimation of the posterior density (β , γ , α). Density estimation takes into account the full Markov chain, i.e. without thinning, after discarded a 500 burn-in periods.....	123
Figure 4.12 Show autocorrelation plots of (β , γ , α). The autocorrelation takes into account the full Markov chain, i.e. without thinning, after having discarded a 500 burn-in periods.....	124
Figure 4.13 Shows the trace plots, histogram and posterior density and autocorrelation plot of the parameters estimated value from the innovation scheme couple with the newly proposed diffusion bridge approach (<i>MGP</i>).....	128
Figure 4.14 Shows daily counts of symptom onset of Ebola case by date. Data from the 1995 Ebola outbreak in the Democratic Republic of Congo recoded from March 1 (corresponding to day 1 on the x-axis) to July 16. Sources: Khan et al.(1999).....	130
Figure 4.15 Shows daily counts of death from Ebola. Data from the 1995 Ebola outbreak in The Democratic Republic of Congo recoded from March 1 (corresponding to day 1 on the x-axis) to July 16. Sources: Khan et al. (1999).....	130
Figure 4.16 Shows the trace plots, histogram and posterior density and autocorrelation plot of three parameters estimated value from the innovation scheme couple with newly diffusion bridge proposed (<i>MGP</i>).....	133

List of Algorithms

Algorithm 2.1 The Metropolis-Hastings algorithm	54
Algorithm 2.2 The Gibbs Sampler algorithm	57
Algorithm 2.3 Componentwise Transitions algorithm	58
Algorithm 3.1 Cholesky Decomposition	67
Algorithm 3.2 Traditional MCMC Bayesian Augmentation Approach	77
Algorithm 3.3 Single site updating Diffusion Bridge approach	82
Algorithm 3.4 Blocking updating Diffusion Bridge approach	84
Algorithm 3.5 Sample path proposed Diffusion Bridge	86
Algorithm 3.6 Traditional parameters estimate for diffusion process	102
Algorithm 3.7 Proposed Modified Innovation parameters estimate for diffusion process...	107

List of Abbreviations

PM	Posterior Means
ACF	Autocorrelation Function
PCI	Posterior Credible Interval
MCMC	Markov Chain Monte Carlo
SDE	Stochastic Differential Equation
ODE	Ordinary Differential Equation
MDB	Modified Diffusion Bridge
DB	Diffusion Bridge
MH	Metropolis Hastings

SIS	Susceptible Infection Susceptible
SIR	Susceptible Infection Recovery
SEIR	Susceptible Exposed Infection Recovery
SEQIHR	Susceptible Exposed Quarantine Infection Hospitalised Recovery
MLE	Maximum Likelihood Estimation
LB	Lindström Bridge
GP	Guided Proposal
MGP	Modified Guided Proposal
AP	Acceptance Probability
AR	Acceptance Rate
GMM	Generalised Method of Moments
EMM	Efficient Method of Moments
SME	Simulated Moments Estimation
EA	Exact Algorithm
FTC	Faith Tabernacle Church
WHO	World Health Organization
DRC	Democratic Republic of Congo
DT	Observed data at time t
DA	Data Augmentation

EM	Expectation-Maximizations
HPDR	Highest Posterior Density Region
CPDR	Central Posterior Density Region
PCI	Posterior Credible Interval
N	Normal (or Gaussian) Distribution
R	R(software environment for Statistical computing)
Pdf	Probability Density Function

Chapter One

Background of the study

1.0 Introduction

Throughout history, interest has been laid in understanding and modeling the dynamics of systems evolving through time. Instances include are not limited to the growth of populations, but also in other research areas, such as, the interactions between certain species, the spread of epidemics diseases and many other dynamic processes. Initially, the dynamics of these systems were captured through the use of deterministic approaches via ordinary differential equations (ODEs); for example, in describe the spread of epidemics disease through a population using ODE proposed by Kermack and McKendrick (1927). However, the evolution of these systems is not entirely predictable and is subject to random variation.

The deterministic nature of the ordinary differential equation description is unable to capture the random variation of the process and has been proved to be unsatisfactory means through which to capture the true dynamics of such systems. Hence an alternative modeling framework is required, which can account for random behaviour.

In epidemiological field, the attempt to mathematical model epidemic processes leads to the formulation of a system of differential equations. However, this formulation based upon a host of approximations which can both conceptual and mathematical. For example, most epidemic models did not take into the account the environmental fluctuations or random terms. The addition of a random term may serve to account for certain approximations. There are plenty of examples of models that would be well served by the inclusion of a random term in the usual differential equation. However, the stringent mathematical definition of such a random differential equation is not simple, and it was not until the middle of the 20th century that stochastic differential equations were given a solid mathematical foundation.

Recent advances in stochastic differential equations (SDEs) enable a lot of researchers to introduce randomness into deterministic model in such a way to reveal the effect of environmental variability in the system of differential equations or environmental fluctuations in parameters. The idea here is to capture unknown influences such as changing behaviours, public interventions, seasonal effects and social cycles just to mentioned a few. Since in reality many phenomena of such nature are affected by stochastic noise, and therefore, Stochastic Differential Equations may be required.

It is interesting to note that after a deterministic system of ordinary differential equations has been formulated for the population dynamics, one can derive several different kinds of stochastic models that take into account the random nature of the population processes in a consistent manner.

1.1 Stochastic Differential Equations (SDEs)

Stochastic differential equations (SDEs) have advantages over some of the other stochastic formulations in that they can be easily derived directly from the deterministic system of ordinary differential equations and have a relatively simple form (Øksendal B., 2003). Models defined by stochastic differential equations allow for the representation of random variability in dynamical systems. The relevance of this class of models is growing in many applied research areas and is already a standard tool to model. Stochastic differential equation models play a prominent role in a range of application areas, including biology, chemistry, epidemiology, mechanics, microelectronics, economics, and financial statistics (Black & Scholes, 1973;Merton, 1976; Cox et al., 1985a; Bibby & Sørensen, 2001;Elerian et al., 2001;Eraker, 2001;Chiarella et al., 2009).

As with ordinary differential equations (ODEs), stochastic differential equations can be solved numerically in the absence of an analytic solution. Performing inference however,

when no analytic solutions exist is not trivial since transition densities will not be available in closed form. However, parameter estimation under stochastic differential equation faced several challenges for discretely observed data (Sørensen, 2004; Jimenez et al., 2005). Several methods of frequentist procedures to infer on the parameters for such stochastic processes have been considered in the literatures. Most techniques struggle when inter-observation times are large. A complete understanding of stochastic differential equation theory requires familiarity with advanced probability and stochastic processes. These processes are often referred to as a diffusion process.

1.2 Diffusion Processes

Diffusion processes governed by stochastic differential equations (SDEs) are as well known as tool for modelling continuous time data from a wide range of areas. A continuous time stochastic process with (almost surely) continuous sample paths which has the Markov property is called diffusion. Diffusion processes are known to solve stochastic differential equations (SDEs) and, as a consequence, the problem of how to efficiently estimate parameters governing stochastic differential equations has arisen. The stochastic process is referred to as a diffusion process (if it is a Markov process) in which the infinitesimal mean and variance exist. The simplest and most fundamental diffusion process is Brownian motion $B(t)$, which is sometimes called the Wiener process $W(t)$.

On the other hand, a Markov process with transition densities $\pi(s, x; t, y)$ is called diffusion if the following three limits exist:

- (1) For any for all $\varepsilon > 0$, $s \geq 0$ and $x \in R$

$$\lim_{\substack{t \downarrow s \\ t-s \rightarrow 0}} \frac{1}{t-s} \int_{|x-y| > \varepsilon} \pi(s, x; t, y) dy = 0 \quad (1.0)$$

Condition (1.0) tells that, it is very unlikely that the process $X(t)$ undergoes large changes in a short period of time.

- (ii) There exist functions $\alpha(x, t)$ and $\beta(x, t)$ such that for all $\varepsilon > 0$, $t \in [0, T]$ and $x \in (-\infty, \infty)$

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x) \pi(s, x; t, y) dy = \alpha(s, x) \quad (1.1)$$

$\alpha(s, x)$ represents the average vector of the random process $X(t)$

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x)^2 \pi(s, x; t, y) dy = \beta^2(s, x) \quad (1.2)$$

While condition (1.2), $\beta^2(s, x)$ measures the local magnitude of the fluctuations of $X(t) - X(s)$ about the mean values. Where α and β are well defined functions.

Diffusions have become an increasingly important tool for the statistical analysis of real world phenomena. However, approximation of a given dynamic system is often done heuristically in the literature, leading to diffusions that do not correctly mirror the true dynamics of the original process. Furthermore, the statistical inference for diffusions typically turns out to be demanding in real data situations. Hence, the statistical estimation of complex diffusion models as applied to real datasets is not widely spread. These issues and others are addressed in this thesis.

Consequently, there is considerable interest in trying to efficiently estimate diffusion parameters from partial and discrete-time observations. The likelihood-based inference of such observations can be problematic, since the closed form transition densities are rarely available.

1.3 Epidemiology & Epidemiological data

An epidemic is an outbreak of an infectious disease affecting a disproportionately large number of individuals in a population, community, or region within a short period of time. Epidemics of infectious diseases have shaped the history of humankind. They have directly affected economy, politics and demography, the course of wars, social behaviours and religious beliefs (McNeill, 1976, Cunha, 2004, Smallman-Raynor & Cliff, 2004, Oldstone, 2010). Over the last two decades, a number of epidemic models for virus spread through human population have been proposed, based on either the classical Susceptible-Infected-

Removed (SIR) model developed by Kermack & McKendrick (1927), or the classical Susceptible-Exposed-Infected-Removed (SEIR) model developed by Rvachev & Longini (1985).

Modeling of the spread of transmissible infectious diseases is still the subject of a good deal of attention in mathematical epidemiology. Modeling of infectious diseases is a tool to investigate the mechanisms for outbreak and spread of diseases, also to predict the future cause in order to control an epidemic. Modeling the infectious disease enables scientists to understand risky factors associated with transmission, and through these, develop viable control strategies for potential future outbreaks could be determined (Rob Deardon et al., 2010).

It is well known that infectious diseases exhibit environmental fluctuations in parameters of the epidemic model on their transmission dynamics. A model for assessing such effects of a random noise in the system of differential equations of epidemic is still facing many challenges' in the field of epidemiological research work. As more knowledge about the epidemiology and biology of the diseases become available, it is reasonable to include environmental fluctuation in the diseases transmission models.

1.3.1 Epidemiological Data

Most epidemic data are discretely observed and undergo stochastic transition rate. Stochastic epidemic models allow more realistic description of the transmission of diseases as compared to deterministic epidemic models (Becker, 1989; Anderson and Britton, 2000).

Since, epidemic data discretely observed are subjected to environmental influence. The dynamic of such data can well be described by discretised-version of Stochastic Differential Equations (SDEs). However, estimating the parameter for such a process proves to be challenging in practice. Various statistical methods have been proposed to estimate

parameters of epidemiological data, the more sophisticated of which rely on the calculation of a likelihood functions precisely using maximum likelihood estimation. Likelihood-based method of parameter estimation for such epidemics models poses many challenges, because, available data were not in closed form (Sorensen, 2004).

Indeed, the literature has not adequately addressed these challenges for multi-dimensional cases. The two broadly classified in modeling the dynamic of epidemics are deterministic epidemic models and stochastic models.

1.4 The Deterministic Model

Deterministic models have been more widely used, because they are simpler to analyse, especially when numerical solutions are adequate. With deterministic models, we know with certainty the values of the parameters and variables in the model, as there are no random fluctuations in value. The system can be completely defined at any time using the initial conditions that had specified. In epidemiology, the deterministic epidemic model was first explicitly defined by Kermack & McKendrick (1927). Deterministic differential equation is solved to explain how a system changes or evolves, when the change occurs and the effect of the starting point to the initial solution and so forth.

1.5 The Stochastic Model

A stochastic model takes into account some element of variability and randomness in at least one of the parameters or variables of the model. Stochastic models appear to be a more intuitive way to explain many natural phenomena. Stochastic differential equations have advantages over some of the other stochastic formulations, in that they can be easily derived directly from the deterministic system of ordinary differential equations and have a relatively simple form (Øksendal B., 2003). Stochastic differential equation (SDE) models play a prominent role in a range of application areas, including biology, chemistry, epidemiology,

mechanics, microelectronics, economics, and finance (Robinson, 1959, Black & Scholes, 1973, Merton, 1976, Cox et al., 1985a, Bibby & Sørensen, 2001, Elerian et al., 2001, Eraker, 2001, Chiarella et al., 2009). Stochastic differential equations consist of both a deterministic and stochastic part, and capture the dynamics of a system through a solution which fluctuates around the deterministic solution.

A complete understanding of stochastic differential equation theory requires familiarity with advanced probability and stochastic processes. The stochastic process is referred to as a diffusion process if it is a Markov process in which the infinitesimal mean and variance exist. Hence, essential property of the stochastic model is that it undergoes markov property.

1.6 Compartment Models

Another important aspect when considered dynamic models are compartmental models. In a compartment model, all objects involved in a system of interest are arranged in a finite number of compartments, i.e. in groups of objects that are defined through certain specified properties (Jacquez, 1972). The compartments are mutually disjoint. The elements of each compartment are assumed to be homogeneous and well-mixed. Interaction between different compartments happens through the exchange of objects which is described by transition equations.

In this study, we considered compartmental systems that are always *closed*; i.e. there is no flow of objects to and from the compartmental environment. In general, under compartmental model, the early stages of an outbreak case of infectious number will be very small, random variations alone can cause an epidemic to die out and it is particularly important to include this randomness in models for emerging infectious diseases. Stochastic models could be a more appropriate way of modeling epidemics in many circumstances. Stochastic models are able to take care of randomness of infectious contacts occurring in the latent and infectious periods. Another important is that, stochastic epidemic models can provide an additional

degree of realism in comparison with deterministic counterparts. The two adopted compartmental epidemic models used in the study are Susceptible-Infectious-Recovery (SIR) and Susceptible-Exposed-Infectious-Recovered (SEIR) epidemic models. These two models have been widely applied to study the spread of many infectious diseases such as smallpox, measles, cholera etc (Rvachev and Longini, 1985).

1.6.1 SIR-Epidemic Model

We start by defining a simple standard SIR (Susceptible–Infective–Removed) epidemic model. This model considers a so called SIR-Epidemic where individuals at first are susceptible. If they get infected they immediately become infectious and remain so until the Recover assuming immunity during the rest of the outbreak.

In this study, we assuming that the population is closed in which individuals mix uniformly in the community, all individuals are equally susceptible to the disease and equally infectious if they get infected during the outbreak. We also assume that the time scale is short enough that demographic effects (births, deaths and migration) may be neglected and the recovered individuals are assumed to acquired permanent immunity(no transfer from R class back to S compartment). Let $(S_{(t)}, I_{(t)}, R_{(t)})$ denote the numbers of susceptible, infectious and recovered individuals at time t . Because the population is closed with size N , we have $S_{(t)} + I_{(t)} + R_{(t)} = N$ for all time. Also, at the start of the epidemic we assume that $\{S_{(0)}, I_{(0)}, R_{(0)}\} = \{N-1, 1, 0\}$ i.e. that there is one initial infective and no immune individuals. The model is Markovian implying that it may equivalently be defined by its jump rates. This epidemic model has two transitions: $S_{(t)} \rightarrow I_{(t)}$ and $I_{(t)} \rightarrow R_{(t)}$. Likewise, with two parameters': transmissions rate say, β and recovery rate say, γ ; this follows a model proposed by Kermack and McKendrick (1927). The conceptual descriptions of the model can be represented by a flow diagram.



Figure 1.0 Shows a flow diagram/chart for the **SIR Epidemic Model**

Figure 1.0 shows the flow diagram for the SIR model uses black arrows to show the movement between the S and I classes and the I and R classes. The mathematical equations' describing this model is as follow:

$$\begin{aligned}
 \frac{dS_t}{dt} &= -\frac{\beta S_t I_t}{N} \\
 \frac{dI_t}{dt} &= \frac{\beta S_t I_t}{N} - \gamma I_t \\
 \frac{dR_t}{dt} &= \gamma I_t
 \end{aligned} \tag{1.5}$$

This model takes into consideration the number of people infected due to direct contact with an infected individual at time t : $\frac{\beta S_t I_t}{N}$, where $\beta > 0$ is the contact rate, $\gamma > 0$ is the recovery rate, and $N = S(t) + I(t) + R(t)$ is the total population size.

There are many extensions of the above SIR-model. For example, it is sometimes assume non-random called the continuous time Reed-First epidemic model. Another extension is where the disease has a latent period, i.e. where there is a period between when an individual gets infected and until he/she become infectious. Such models are often referred to as SEIR-epidemic model.

1.6.2 SEIR-Epidemic model

A popular way to model epidemic systems is to consider exposed epidemic models: SEIR model, which has been widely used to control the spread of diseases. This type of epidemic model classifies individuals according to their epidemiological status over time. At any point in time an individual can be classified as either susceptible to infection (S), infected but not infectious (E), infected and infectious (I), or removed from the population (R). We assume that individuals can only move through the states in that order. Assuming a homogeneous closed population of size N (i.e. with negligible births, deaths or migration), and

homogeneous mixing within the population, the rate of transition of an individual from $S \rightarrow E$ can be modelled as being proportional to the proportion of infective individuals in the population. The rate of transition from $E \rightarrow I$ and $I \rightarrow R$ are constant. These assumptions are not necessary for either likelihood- or simulation-based methodologies. Figure 1.1 depicted the flow chart of SEIR epidemic model.



Figure 1.1 Shows a flow diagram/chart for the **SEIR Epidemic Model**

Differential equations describing the dynamics of an SEIR epidemic model have the following form:

$$\begin{aligned}
 \frac{dS_t}{dt} &= -\frac{\beta S_t I_t}{N} \\
 \frac{dE_t}{dt} &= \frac{\beta S_t I_t}{N} - \gamma E_t \\
 \frac{dI_t}{dt} &= \gamma E_t - \alpha I_t \\
 \frac{dR_t}{dt} &= \alpha I_t
 \end{aligned} \tag{1.6}$$

Similarly, where $\beta > 0$ is the contact rate, $\gamma > 0$ is the exposed rate, α is the recovery and

$N = S(t) + E(t) + I(t) + R(t)$ is the total population size.

Epidemic models such as SEIR models have been widely used, over the decades, to control the spread of diseases such as yellow fever, smallpox, measles, pandemic influenza, diphtheria, plague, cholera, SARS [Chowell et al, 2004; Lipsitch et al. 2003; and Wang & Ruan, 2004].

Different constructions of the SIR and SEIR model can be found in the literature, for example Anderson & Britton (2000) for an overview.

1.6.3 SEQIHR-Epidemic Model

SEQIHR model adopts the autonomous quarantine/isolation model studied in Safi and Gumel (2010), for the transmission dynamics of an infectious disease, in the presence of Quarantine of exposed individuals and Isolation of infected individuals with clinical symptoms of the diseases. The model is given by the following system of differential-equations:

$$\begin{aligned}
 \frac{dS}{dt} &= \Pi + \psi R(t) - \lambda_t S(t) - \mu S(t) \\
 \frac{dE}{dt} &= \lambda_t S(t) - [\gamma_E + \theta + \mu] E(t) \\
 \frac{dQ}{dt} &= \theta E(t) - [\gamma + \mu] Q(t) \\
 \frac{dI}{dt} &= \gamma_E E(t) - [\sigma + \alpha_I + \mu] I(t) \\
 \frac{dH}{dt} &= \gamma_Q Q(t) + \sigma I(t) - [\alpha_H + \mu] H(t) \\
 \frac{dR}{dt} &= \alpha_I I(t) + \alpha_H H(t)
 \end{aligned} \tag{1.7}$$

where
$$\lambda_t = \beta \frac{I(t) + \eta_e E(t) + \eta[\varepsilon_1 Q(t) + \varepsilon_2 H(t)]}{N(t)} \tag{1.8}$$

Where, susceptible denoted by $S(t)$, exposed (infected, but not yet show clinical symptoms of the disease; $E(t)$, infected with symptoms $I(t)$, quarantined $Q(t)$, (it is assumed that individuals in the $Q(t)$ class are infected but do not display clinical symptoms of the disease), hospitalized $H(t)$ and recovered $R(t)$ individuals.

The quarantine-Isolation model that made-up of the aforementioned compartmental SEQIHR above can be described by the flow chart in figure 1.2.

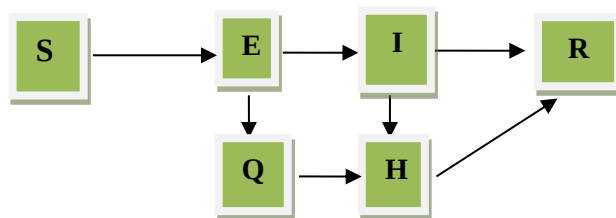


Figure 1.2 shows a flow diagram/chart for the **SEQIHR** Model

In equation (1.7), the parameter Π represents the rate of recruitment into the population; ψ is the rate of loss of infection acquired immunity. β is the effective contact rate and μ is the natural death rate. Exposed individuals are quarantine at a rate θ . Quarantined and infectious individuals are hospitalized at the rate γ_Q and σ respectively. The parameters α_I and α_H represent the recovery rates of infectious and hospitalized individuals, respectively. Finally, γ_E and γ_Q represent the infection rate in infectious class and isolation class respectively.

From the deterministic SEQIHR, it is possible to express R in terms of S, E, Q, I and H because $N = S + E + Q + I + H + R$. Therefore, the model can be represented in the simpler form:

$$\begin{aligned}
\frac{dS}{dt} &= -\lambda_t S(t) \\
\frac{dE}{dt} &= \lambda_t S(t) - [\gamma_E + \theta] E(t) \\
\frac{dQ}{dt} &= \theta E(t) - \gamma_Q Q(t) \\
\frac{dI}{dt} &= \gamma_E E(t) - [\sigma + \alpha_I] I(t) \\
\frac{dH}{dt} &= \gamma_Q Q(t) + \sigma I(t) - \alpha_H H(t)
\end{aligned} \tag{1.8}$$

Where $\lambda > 0$ represent the contact rate, γ_E is the exposed rate of infection, γ_Q represent the isolating infection rate θ is the quarantined rate, σ represent infection rate from isolating individual and the parameters α_I and α_H represent the recovery rates of infectious and hospitalized individuals, respectively.

1.7 Itô Calculus and Itô's formula

Inclusion of white noise as a source of randomness in differential equations leads to difficulties in the application of classical calculus. The need for a new integral definition arises, resulting in stochastic calculus with the Itô calculus as a prominent representative. Kiyoshi Itô (1942) extended the methods of classical calculus to stochastic processes, which

Brownian motion is one, and it is after him that Itô calculus is named. Hence, Itô formula is a method to apply nonlinear transformations to stochastic differential equations (SDEs).

1.8 Bayesian Framework

This particular school of thought takes its name from the English 18th century statistician and theologian Thomas Bayes, whose famous theorem on conditional probabilities lies at the heart of the paradigm. A core concept of the Bayesian paradigm is its interpretation of probability, where, philosophical finesse aside, prior knowledge of the object of one's studies is taken into account. More practically, one associates probability distributions with the parameters of the likelihood, effectively making them random variables. This distribution may then be used to assign relative probabilities to different regions of the parameter space in an effort to quantify one's knowledge about the parameters. Phrased in a mathematical way, it is supposed that X has a distribution that is parameterised by θ and assign the a priori density $\pi(\theta)$ to θ before any observation has been made. This prior distribution may be chosen based on previous knowledge, purely personal beliefs, or in a manner such that its influence on the final inference is as limited as possible. Now, data are needed. The data yield a likelihood $\pi(x|\theta)$ or $L(\theta|x)$ that gives information about the parameters, whereby prior beliefs about the parameters can be updated. One's knowledge is updated with Bayes' theorem

$$\pi(\theta|x) \propto \pi(\theta)\pi(x|\theta) = \pi(\theta)L(\theta|x) \quad (1.3)$$

Equation (1.3) is the proportional posterior density of parameters given data. This may then be used for inference about θ . This approach is fundamentally different from the frequentist approach, where parameters are considered to be unknown but fixed. Naturally, there is a proportionality constant, which has been omitted above, equalling $1/\int \pi(\theta)L(\theta|x)d\theta$. This

constant is often difficult to compute, and for certain purposes not necessary and the numerical methods hereafter described are often used for estimation of this constant.

The Bayesian method of inference is particularly suited to epidemic models, as the parameters of interest are usually defined in terms of individuals, this make more researchers to consider the distribution of these parameters over the whole model.

1.9 Statement of the problem

Recently, there has been interest in stochastic models for epidemics. Modeling with epidemic data have been the major concern of the researchers in the field of epidemiology. The reason for these was due to the inferential problems while trying to estimate the parameters of the models. Most epidemic data were partially observed, had some components that not observed throughout the outbreak of the epidemic diseases. Even when observed, some elements of the components may be observed frequently than other, affected by environmental influence. Consequent to these, there is need for an appropriate method and procedure to conquer these statistical inferential challenges. Almost existing methods in literature based their techniques on the Likelihood-based methods of parameters estimate and these poses many challenges (Sorensen, 2004). Mostly epidemiological details concern outbreak data were rarely obtained. Even when observed, their likelihoods were not in closed form, exhibit environmental fluctuations in the parameters of the models.

There are many comprehensive alternative frequentist methodologies considered in literature. However, the application of such methods becomes problematic either when: inter-observation times are very large, measured with non-equidistance times or some components of epidemic models are latent measured with error and lastly consist of environmental variations or fluctuations. As more knowledge about the epidemiology and biology of the diseases become available, it is reasonable to include these environmental fluctuations in the diseases transmission models. In this study, we considered stochastic differential equation for

explaining the behaviour of such dynamic epidemic models. In addition, stochastic epidemic models allow more realistic description of the transmission of disease as compared to deterministic epidemic models. Even though, statistical inferences for assessing parameter estimate under stochastic differential equations still facing some challenges in the literature, here we consider alternative Bayesian inference method of estimations for such estimation, through the use of diffusion approximation processes. However, stochastic differential equations through diffusion approximation process provide a natural framework for modelling intrinsic stochasticity. The likelihood-based methods of estimations in literature that undergo differential equations are typically not in closed form had high level of serial correlation that resulting into convergence problems. These challenges on parameters method of estimate are addressed in this study. Therefore, the motivation behind this study is to provide Bayesian methods for constructing efficient Markov Chain Monte Carlo (MCMC) algorithms so as to reduce the serial dependence and obtain more reliable results.

1.10 Aim and Objectives of the study

The aim of this work was to find the estimation procedure which would not suffer from convergence problems.

The objectives were to:

- (i) derive epidemic models for explaining the behaviour of SDEs models,
- (ii) propose an improved diffusion bridge sampler capable of overcoming convergence problems,
- (iii) propose Bayesian data-augmentation method of parameters estimation which would not suffer from convergence,
- (iv) compare the performance of the proposed methods with existing methods using simulated and real-life data sets.

1.11 Significant of the study

The analysis of outbreak epidemic data can be more effective when it is based on a model for the actual process which generates the data. Such analysis can be useful for diseases which occur due to re-emerging infectious diseases. The re-emerging infectious diseases outbreak such as Ebola, smallpox, monkey-pox, Lassa fever virus just to mentioned a few, poses an alarming risk to the countries in Africa and beyond.

This poses an urgent threat not only in Africa but also in the international community. This is of particular interest at the moment, because, the World recently has faced the danger of a highly pathogenic diseases. It is therefore important to construct a model for which we can draw inference regarding its unknown quantities so as to be able to give answers to important scientific questions regarding the underlying processes of an outbreak.

However, several reasons suggested that diffusion process described as solutions to stochastic differential equations are to be preferred when stochastic models and their analysis are possible. Modeling with such processes often poses several challenges whenever inferential issue had been raised to obtain parameters estimate of the models in the literature.

For examples, if unknown parameters appear in the diffusion coefficient of any diffusion process, direct implementation of statistical estimation results into a Markov chain that is reducible. The reason for this is the serial correlation that exists between the data points and the diffusion coefficient. This was noted in a simulation study in Elerian et al. (2001) and was justified theoretically in Roberts and Stramer (2001). Likewise, diffusion process required efficient diffusion bridges sampler to generate the realisation discrete-time dataset, which can be difficult, especially in the multidimensional case.

These and many others problems that involving parameters estimate of the stochastic models using diffusion processes (stochastic differential equations) are the main focus of this

research work. Apart from providing estimates for the parameters of interest which are largely responsible for driving the dynamics of the disease, models also have the ability to answer questions which refer to the current state of the outbreak.

These can guide the conduction of effective control policies to prevent a major spread. Also suggest optimal plans for controlling a future outbreak.

Chapter Two

Literature Review

2.1 History of Epidemic Modeling

Modeling of epidemic diseases is a tool to investigate the behaviour of epidemic outbreak and the spread of diseases, in order to predict the future cause of any outbreak so as to control such epidemic disease. This would enables epidemiologist to understand the risk factors associated with the transmission of such diseases in order to develop viable control strategies for potential future outbreaks (Rob Deardon et al., 2010).

The first mathematical model for the spread of infectious diseases is generally accredited to the work of Daniel Bernoulli (1760), then, epidemic modeling has not received much attention until the beginning of the 20th century. Early works under mathematical model of deterministic disease include En'ko (1889), Hamer (1906), Ross (1915) and Kermack and McKendrick (1927). Historical accounts on the development of mathematical epidemiology can be found in Bailey (1975), Dietz (1967), Anderson and May (1991) and Daley and Gani (1999). In the early stages of epidemic modeling, the spread of diseases was generally formulated as a deterministic process. According to Bailey (1975), the first author who included a random component in an epidemic model was McKendrick (1926), but that particular approach was only continued twenty years later. Instead, the class of *chain binomial models*, independently introduced by Lowell Reed and Wade Hampton Frost, (Abbey (1952) or Costa Maia (1952) and Greenwood (1931). An overview about chain binomial models is given in Becker (1989) and Daley and Gani (1999).

McKendrick (1926) was the first epidemiologist that proposed stochastic modeling of epidemic disease. This model did not received much attention at the first instant, because a

model which attracted more attention at that time, even though it was never published was the chain-binomial model of Reed and Frost put forward in a series of lectures in 1928.

Barlett (1949) studied the stochastic version of the deterministic model of Kermack-McKendrick (1926). Since then, the effort put into stochastic modeling infectious diseases has more or less exploded. Detailed statistics on disease counts go back to John Graunt (1620-1674) who recorded weekly death counts in London together with their causes.

In subsequent years, both stochastic and deterministic models were refined and their mathematical analysis was extended. For example, the class of susceptible–infective–removed (SIR) models emerged as the most prominent description of the spread of infectious disease epidemics.

The comprehension of disease dynamics, the development of mathematical tools progressed, the general framework of modeling the spread of diseases changed, as well as increased number of mobility of humans, which raises the risk of fast spreading epidemics, poses modern epidemiology models.

Modern epidemiological models take into account environmental variations, social behaviour, the effect of intervention such as vaccination or isolation, and many other aspects.

One of the earliest epidemic models was contained in a paper by Hamer (1906). Hamer postulated that the probability of a new infection in the next discrete time-step was proportional to the product of the number of susceptible and the number of infective. Ross (1908) translated this ‘mass action principle’ to the continuous time setting.

Herwaarden et. al. (1995) indicated that an endemic in a deterministic system can disappear in its corresponding stochastic system due to stochastic fluctuations. Mao et. al. (2002) showed that a sufficiently small noise can suppress explosions in population dynamics; Nasell (2002) developed stochastic models to show that some stochastic models are a better

approach to describing epidemics for a large range of realistic parameter values in comparison with their deterministic counterparts.

Allen (2007) discusses modeling the SIS and SIR by constructing stochastic differential equation approximations to the continuous time Markov chain models. McCormack and Allen (2006) apply the stochastic differential equation approximations proposed by Allen (2007) to an SIS multi-host epidemic model and explore the stochastic and deterministic models numerically. Chen and Li (2009) study stochastic differential equation (SDE) version of the SIR model both with and without delay, where the SDE is introduced in different styles.

Allen (2008) provided a great introduction to the methods of derivation for various types of stochastic models including stochastic differential equation (SDE) epidemic models. There are fair amount of literature regarding SDE epidemic models. There are different possible approaches to including random effects in the model, both from biological and mathematical perspectives.

This poses an alternative methods in which many researchers have been proposed to estimate parameter for epidemiological systems, the more sophisticated of which rely on the calculation of a likelihood function (for examples: Becker, 1976; Anderson and Britton, 2000, Fearnhead and Meligkotsidou, 2004).

Likelihood based inference for epidemic models poses many challenges, because available epidemic data are often censored or incomplete. In this case the unobserved data must often be inferred from the observed data. Alternatively, approximations can be made in order to match the model structure to the form of the data.

A popular and particularly effective solution is to use the Bayesian approach to estimate parameters, via modern numerical techniques, such as, Markov Chain Monte Carlo (MCMC)

algorithms (Gibson and Renshaw, 1998; O'Neill and Roberts, 1999, Streftaris and Gibson, 2004; Lekone and Finkenstädt, 2006), Sequential Monte Carlo (SMC) routines (Doucet et al., 2001; Liu, 2001) and other numerical approach like importance sampling.

2.2 Models for infectious diseases

The spread of diseases is still a serious concern in both the developed and developing world. The elimination of epidemic disease is desirable not only from a humane viewpoint but also regarding economic factors such as manpower and public health costs. Even for diseases with relatively mild courses, it is generally favourable to invest in prevention rather than cure.

Hilary (2007) proposed stochastic models for two different epidemic models: the Susceptible-Infective-Recovered (SIR) model and the Reed-Frost Epidemic Model. He tried to estimate the infection rate for the Reed-Frost Epidemic model using both classical maximum likelihood methods and Gibbs Sampling. He also used Approximate Bayesian Computation to estimate the rate of infection and recovery for the SIR model. He could not apply maximum likelihood and Gibbs Sampling methods to the SIR model as its likelihood is not readily computed. In his results, the maximum likelihood and Gibbs Sampling methods for the Reed-Frost model agree, but the model does not appear to fit the data well. When considering the SIR model Approximate Bayesian Computation accurately returns parameter values when he has complete data. It performs less well when he has incomplete data. Further work needs to be carried out before he can say whether Approximate Bayesian Computation is useful when considering real-life data sets, which often have incomplete data.

Moneim (2005) studied two classes of epidemic models. These models are the standard SIR and SEIR models with time-varying periodic contact rate. The importance of the latent period was his target. Moneim wanted to know when the latent period can be ignored and when it must be taken into account, these are the main points of his simulation. The comparison of the simulation results of his two models shows that the latent period is affecting the pattern of

the dynamics of the disease. His paper addressed how model predictions are affected by the assumed form of the seasonally varying transmission rate and whether or not a latent class is included.

Assessment of key mechanisms in epidemiological process cannot be achieved without the application of mathematical models. Moreover, real life situations can particularly be rendered by the consideration of random events. Unfortunately, analytic intractability of SDEs governing most multidimensional diffusion processes precludes a closed-form expression for the transition densities. Consequently, inferring the parameters of the SDE using observations taken at discrete times is a challenging problem. Several methods to overcome this difficulty include simulated maximum likelihood estimation (Pedersen, 1995; Durham and Gallant, 2002), closed-form expansion of the transition density (Ait-Sahalia, 2002, 2008; Picchini et al., 2010; Stramer et al., 2010; Picchini and Ditlevsen, 2011), exact simulation approaches (Beskos et al., 2006, 2009, 2013; Sermaidis et al., 2013) and Bayesian imputation approaches (Elerian et al., 2001; Eraker, 2001; Roberts and Stramer, 2001; Golightly and Wilkinson, 2008; Stramer and Bognar, 2011; Kou et al., 2012; Schauer et al., 2016) suffering from estimation problems.

2.3 Stochastic differential equations (SDEs)

Several reasons suggested that stochastic models are to be preferred when their analysis is possible. First, the most natural way to describe the spread of disease is stochastic. In fact, an important part in stochastic modeling lies in showing what deterministic model and the stochastic model converges to, when the population size becomes large. Secondly, there are phenomena which are genuinely stochastic and do not satisfy a law of large numbers. Further, stochastic model with SDE approaches are very reliable when consider extinction of epidemic diseases, since extinction occurs when the epidemic process deviates from the expected level. Finally, Knowledge about uncertainty in estimates target parameters of any

dynamic model requires a stochastic approach, and estimation is not of much use without some knowledge of its uncertainty (Anderson and Britton (2000)).

Stochastic differential equation (SDE) was first initiated and developed by Ito (1942). The theory of (SDE's) provides a useful tool to introduce stochasticity into models and to characterize the evolution of many processes, especially in finance, Mathematics, and almost all sciences, such as Physics, Chemistry, Biology, and Medical science. In many cases, the solutions are not given explicitly due to non-differentiability character of realizations of the Brownian motion; therefore numerical approximations are used to study the properties of these models. Unfortunately, analytic intractability of SDEs governing most multidimensional diffusion processes precludes a closed-form expression for the transition densities. Consequently, inferring the parameters of the SDE using observations taken at discrete times is a challenging problem. Methods to overcome this difficulty in existing work include simulated maximum likelihood estimation (Pedersen, 1995; Durham and Gallant, 2002), closed-form expansion of the transition density (Ait-Sahalia, 2002, 2008; Picchini et al., 2010; Stramer et al., 2010; Picchini and Ditlevsen, 2011), exact simulation approaches (Beskos et al., 2006, 2009, 2013; Sermaidis et al., 2013) and Bayesian imputation approaches (Elerian et al., 2001; Eraker, 2001; Roberts and Stramer, 2001; Golightly and Wilkinson, 2008; Stramer and Bognar, 2011; Kou et al., 2012; Schauer et al., 2016).

Stochastic epidemic models allow more realistic description of the transmission of disease as compared to deterministic epidemic models (Becker, 1979; Anderson and Britton, 2000). However, parameter estimation is challenging for discretely observed data for stochastic models (Sørensen, 2004; Jimenez et al., 2005). Several methods of frequentist procedures to infer on the parameters have been considered. The difficulty that underlies these approaches is the general intractability of the transition density for discrete-time observations. Most techniques struggle when inter-observation times are large. In another way, real datasets in

real life situations may well be of low frequency type, where public health reporting considers new infections per week. Even when high-frequency measurements available, it may be advantageous to work with a thinned dataset when consider SDE approach, for instance, asset price data can be corrected for certain microstructure effects this way (Jones, 1999).

Various methods for inferential problems under stochastic differential equations (SDEs) have been developed. The Hermite polynomial expansion approach proposed by Aït-Sahalia (2002, 2008) performed poorly if the data are sparsely sampled (Stramer and Yan, 2007b). Moreover, the approach has some restrictions which could limit its application, especially for multivariate models (Lindström, 2012). Sarkka and Sottinen (2008) proposed an approach which uses an alternative SDE as an important process and the Girsanov theorem to help evaluate the likelihood ratios of two SDEs. However, the diffusion coefficient of their model was stated-independently, whereas, generally SDE models allow not a state-dependent diffusion coefficient. Recent developments have mainly been focused on Bayesian approaches (Eraker, 2001; Golightly and Wilkinson, 2005, 2006, 2011; Donnet et al., 2010), feasibly applicable in the discretely observed model and the data that are sparsely sampled.

2.3.1 Brownian Motion

The simplest and most fundamental diffusion process is Brownian motion $\{B_t\}$. Brownian motion also referred to as the Wiener process which was first observed by the botanist Robert Brown in 1827, who examined the motion of pollen grains suspended in water, under the lens of the microscope (Brown, 1827). In 1900, Louis Bachelier mathematically defined Brownian motion and proposed it as a model for asset price movements (Bachelier, 1964), whilst the governing laws were independently stated by Albert Einstein (Einstein, 1905). Existence of Brownian motion was ultimately proved by Norbert Wiener (Wiener, 1923).

Formally, the stochastic process $\{B(t)\}$, $t \in [0, \infty)$ is a standard Brownian motion if $B(t)$ depends continuously on t , $B(t) \in (-\infty, \infty)$ and $B(t)$ is a diffusion process satisfying the following assumptions:

- (i) $B(0) = 0$, with probability 1;
- (ii) $\mathbb{E}[B(t)] = 0$ and $\text{Var}[B(t)] = \sigma^2 t$;
- (iii) $B(t)$ has stationary, independent increments. i.e.

For $0 \leq t_0 < t_1 < t_2 < \infty$, $B(t_1) - B(t_0)$ and $B(t_2) - B(t_1)$ are independent ,

Clearly, the transition probability density function of the Brownian motion is of the Gaussian density:

$$p_\theta(t, y | x) = \frac{1}{\sqrt{2\pi}\beta^2(x, \theta)} \exp\left(\frac{-1}{2} \frac{(y - x - \alpha(x, \theta)t)^2}{\beta^2(x, \theta)t}\right) \quad (2.1)$$

However we can generate samples of B_t , which are distributed according to the centred Gaussian distribution with variance t (Folland, 1999; Revuz and Yor, 1994). Figure 2.0 shows an example of one-dimensional Brownian motion paths for three difference trajectories generated from the same initial point.

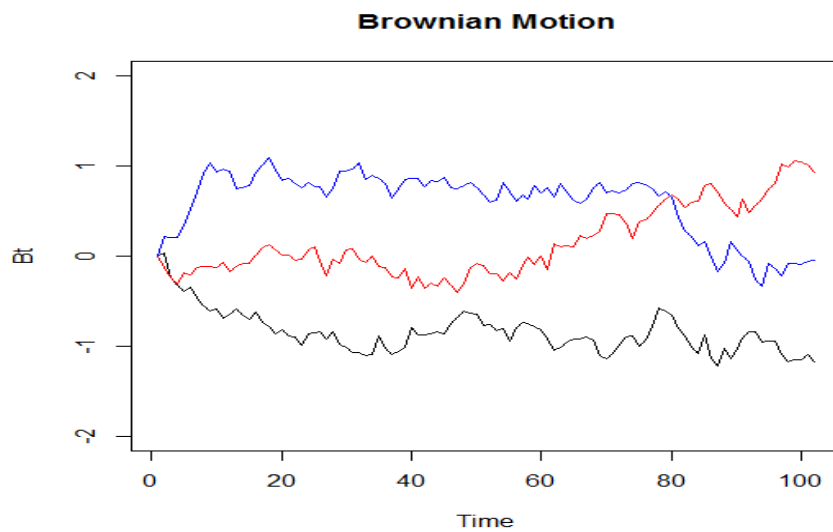


Figure 2.0 shows three trajectories of one-dimension sample Brownian motion.

2.3.2 Brownian Bridge

If the standard Brownian motion is further conditioned on some end point $B_t = x_T$, then the conditioned process $(B_\tau)_{\tau \in [0, t]}$ is called a Brownian bridge. Like Brownian motion, Brownian bridges are Gaussian processes, but without independent increments.

2.3.3 Itô process

An Itô process is a stochastic process $\{X_t\}_{t \geq 0}$, which satisfy

$$X_t = X_0 + \int_0^t \alpha(X_s, s) ds + \int_0^t \sqrt{\beta(X_s, s)} dW_s, \quad X_0 = x_0 \quad (2.2)$$

Here the process can be expressed as a Riemann integral plus an Itô stochastic integral.

Where X_0 is a random variable which represents the initial condition of the process, and α, β functions are non-anticipative random process. The second integral is known as Itô integral which involves the unknown solution, i.e.

$$\int_0^t \sqrt{\beta(X_s, s)} dW_s = \sum_k^{m-1} \beta_k^{1/2} (W_{tk+1} - W_{tk}) \quad (2.3)$$

Equation (2.2) can express in parameterised version of Ito Stochastic differential equation of the form:

$$dX_t = \alpha(X_t, \theta) dt + \sqrt{\beta(X_t, \theta)} dW_t, \quad X_0 = x_0. \quad (2.4)$$

where $\alpha(X_t, \theta)$ is the drift, $\beta(X_t, \theta)$ is the diffusion coefficient and W_t is the wiener process.

2.3.4 Discretisation of continuous-time process

The transition probability in most diffusion processes are intractable, unless the diffusion process is analytically explicitly solvable, which is rarely the case. Here the idea is to discretized diffusion process that would allow numerical solution approaches to solve this intractability as early proposed by Allen, (2007).

An obvious way to obtain numerical approximation methods is to employ a truncated version of the Itô-Taylor expansion (Kloeden & Platen, 1991)

$$X_t = X_{t_0} + \alpha(X_{t_0}, t_0) \int_{t_0}^t ds + \beta(X_{t_0}, t_0) \int_{t_0}^t dB_s + R \quad (2.5)$$

with a remainder term R with i^{th} component

$$R^{(i)} = \sum_{j=1}^m \sum_{l=1}^m \sum_{r=1}^d \beta_{rj}(X_{t_0}, t_0) \left(\frac{\partial \beta_{il}}{\partial x^{(r)}}(X_{t_0}, t_0) \right) \int_{t_0}^t \int_{t_0}^s dB_u^{(j)} dB_s^{(l)} + \dots, \quad (2.6)$$

where the sub- and superscripts denote the components of $\beta \in \mathbb{R}^{d \times m}$, $x \in \mathbb{R}^d$ and $B \in \mathbb{R}^m$.

The simplest stochastic numerical method is the Euler-Maruyama method (Kloeden and Platen, 1992 ;Burrage K. et al., 2000).

2.3.5 Euler-Maruyama Numerical Scheme

The stochastic Euler scheme was first investigated by Maruyama in the early 1950s and is often called the Euler-Maruyama scheme. It seems to be consistent with the Itô stochastic calculus because the noise term approximates the Ito stochastic integral over a discretization subinterval $[t_k, t_{k+1}]$ as $\Delta t = t_{k+1} - t_k$, then, equation (3.3) becomes.

$$X_{t+1} = X_t + \alpha(X_t, \theta) \Delta t + \sqrt{\beta(X_t, \theta)} \Delta W_t, \quad X_0 = x_0 \quad (2.7)$$

$$\text{where, } \Delta W_t = W_{t_{k+1}} - W_{t_k} \sim N(0, \Delta t), \quad k = 0, 1, \dots, K.$$

Other standard numerical methods not considered in this study are:

2.3.6 Milstein Numerical Scheme

Milstein numerical solution was first derived by Milstein (1995), who used the Itô formula to expand an integrand involving the solution in one of the error terms of the Euler-Maruyama scheme. The iterative repetition of this idea underlies the systematic derivation of stochastic Taylor expansions and numerical schemes of arbitrarily high strong and weak orders, as expounded in Kloeden and Platen (1999). The Milstein numerical solution is of this form:

$$X_{t+1} = X_t + \alpha(X_t, \theta) \Delta t + \sqrt{\beta(X_t, \theta)} \Delta W_t + \sqrt{\frac{1}{2} \beta(X_t, \theta) \beta(X_t, \theta) \{(\Delta W_t)^2 - \Delta t\}} \quad (2.8)$$

2.3.7 Runge-Kutta Numerical Scheme

The Milstein scheme above involves the derivatives. If it happens that the derivatives do not exist, then it leads to difficulties in implementing. For that case we need another derivative free method. Then by consider implicit schemes which avoid the use of derivatives.

This was obtained from the corresponding implicit strong Taylor schemes by replacing the derivatives there by finite differences expressed in terms of appropriate supporting values called Runge-Kutta schemes (Kloeden and Platen, 1992). Its numerical solution is of the form:

$$X_{t+1} = X_t + [\gamma f(X_{t+1}) + (1-\gamma)f] \Delta t + L \Delta t W_t + \frac{1}{2} \frac{1}{\sqrt{\Delta t}} [L(\bar{X}_t - L)][(\Delta W_t)^2 - \Delta t] \quad (2.9)$$

With the vector supporting value

$$\bar{X}_{t+1} = X_t + f \Delta t + L \sqrt{\Delta t} \quad \text{and } \gamma \in [0, 1]$$

Another numerical method that can be used to simulate the SDE are: local linearization methods which includes Ozaki and Shoji-Ozaki scheme the filtering based methods which involve the recursive estimation of the SDE realizations by assuming Gaussian approximations of the distributions of the realizations.

For nonlinear SDEs Jazwinski, 1970 considered the Extended Kalman Filter forms a Gaussian approximation to distribution of states and measurements using a Taylor series expansion (Jazwinski, (1970)).

The idea of computing sufficient statistics using Kalman filter cannot be used in simulating the nonlinear SDE because Kalman filter is only for linear dynamic model. Jazwinski (1970) and Sarkka (2013) reviewed Kalman filter and developed ‘Extended Kalman filter’ that deal with nonlinear stochastic differential equations.

A number of further approximation techniques are proposed by different authors, especially in the physics literature and most often in the context of physical processes.

2.4 Methods of Parameter Estimation under Diffusion models

Several methods of estimation to infer the parameters estimate in the discrete-time version of stochastic differential equations have been considered in the literature. Here we revealed some existing method of parameter estimate through frequentist approaches and later considered Bayesian approaches.

2.4.1. Approximate Maximum Likelihood Estimation (MLE) Approach

Lo W. Andrew (1988) considered maximum likelihood estimation (MLE) method of estimation on generalised Ito processes with discretely sampled data. The exact log-likelihood function of the parameter for a discretely-observed diffusion process is usually unknown. Approximate maximum likelihood estimation would, however, be possible if an appropriate approximation of the transition density was available. To implement this idea is to introduce the Euler scheme as follows.

$$X_{k+1} = X_k + \alpha(X_k, \theta)\Delta t_k + \beta(X_k, \theta)N(0, \Delta t_k I) \quad (2.10)$$

where $k = 0, \dots, n-1$, for approximately sampling the process $(X_{t_k})_{k \in \mathbb{N}_0}$ at discrete time points $t_1 < \dots < t_n$ for given parameter θ and initial value $X_{t_0} = x_0$.

This scheme becomes more accurate as the maximum distance between two consecutive time instants tends to zero. Hence, for small Δt_k , we can assume X_{k+1} conditional on X_k to be approximately normally distributed.

The conditional mean and variance can be obtained as

$$\textbf{Mean:} \quad E(X_{k+1} | X_k = x_k) = x_k + \alpha(x_k, \theta)\Delta t_k \quad (2.11)$$

$$\text{and } \textbf{Variance:} \quad \text{Var}(X_{k+1} | X_k = x_k) = \beta(x_k, \theta)\Delta t_k \quad (2.12)$$

The probability density $\pi_\theta(\Delta t_k; x_k, x_{k+1})$ thus, can be approximated by a Gaussian density with mean and variance according to (2.11) and (2.12) above. In case the coefficient of diffusion β does not depend on parameters of interest θ , the result of log-likelihood function corresponds

to the Riemann-Itô approximation of the log-likelihood based on continuous observation. As a general convention, additive constants not depending on suppressed in the log-likelihood function.

$$\ell_n^{Euler}(\theta) = \sum_{k=0}^{n-1} \alpha'(x_k, \theta) \beta^{-1}(x_k)(x_{k+1} - x_k) - \frac{1}{2} \sum_{k=0}^{n-1} \alpha'(x_k, \theta) \beta^{-1}(x_k)(x_k, \theta) \Delta t_k \quad (2.13)$$

Maximisation of the approximated log-likelihood function leads to an *approximate* or *naive maximum likelihood estimator*, sometimes also referred to as *quasi maximum likelihood estimator* (e. g. Honoré, 1997). This estimator has good asymptotic properties in case of decreasing time step. For example, Florens-Zmirou (1989) and Yoshida (1992). The more realistic case, however, is that the time step is fixed. Lo (1988) provides a simple example where the maximum likelihood estimator is inconsistent for fixed observation intervals. More generally, Florens- Zmirou (1989) shows for ergodic diffusion processes with constant diffusion coefficient that the naive maximum likelihood estimator for the drift parameter has an asymptotic bias of the order of the equidistant fixed time step. This deficiency is not due to the Gaussian nature of the approximated transition density but because of the generally misspecified mean and variance of this normal density.

Unfortunately, in many applications in epidemic data the time steps Δt_k are rather large. The fairly simple maximum likelihood approach considered in this section is hence not expected to yield satisfactory results in those cases. More advanced estimation procedures are required in order to address this problem.

2.4.2. Analytical Approximation of the Likelihood Function

Another advanced approach considered in this review was originated by Aït-Sahalia (2002) and it involves the expansion of the transition density in a Gram-Charlier series. The result is a closed-form expression which is shown to converge to the true likelihood as more and more correction terms are included. The method works for one-dimensional diffusion processes

under fairly weak regularity conditions; see Aït-Sahalia paper (2002) for details. Suppose the target process X satisfies the SDE

$$dX_t = \alpha(X_t, \theta)dt + \beta(X_t, \theta)dW_t, \quad X_0 = x_0, \quad (2.14)$$

for $t \geq 0$. The original process X is hence transformed to an appropriate process Z . The approximation of the transition density of Z can then be transferred to the transition density of X . The transformation from X to Z takes place in two invertible steps as follows.

The first operation transforms the diffusion X to a diffusion Y with unit diffusion coefficient. This is done with Lamperti's transformation. Then Y fulfils the stochastic differential equation of the form:

$$dY_t = \left(\frac{\alpha(X_t, \theta)}{\beta(X_t, \theta)} - \frac{1}{2} \frac{\partial \beta}{\partial x}(X_t, \theta) \right) dt + dW_t, \quad Y_0 = y_0 \quad (2.15)$$

for $t \geq 0$,

where $\partial/\partial x$ denotes differentiation with respect to the state variable. Let π_X and π_Y denote the transition densities of X and Y , respectively. Aït-Sahalia (2002) demonstrates that the tails of π_Y are thin enough for the considered expansion. Overall, however, the density $\pi_Y(\Delta t; y_0, y)$ is still not suitable as the function is peaked around $y = y_0$ for small Δt . Hence, one performs a second transformation as:

$$Z_t = \frac{Y_t - y_0}{\sqrt{t}} \quad (2.16)$$

for all $t \geq 0$.

Naturally, the initial value of this new process equals $z_0 = 0$.

Aït-Sahalia shows that for fixed time interval (Δt) the transition density $\pi_Z(\Delta t; z_0, z)$ of Z fulfils the necessary criteria; specifically, it can appropriately be expanded in a convergent series around a standard normal density. Aït-Sahalia however demonstrates that the approximation is sufficiently accurate already for a small number of terms. An extension of the above approximation procedure for multi-dimensional diffusion processes is described by Aït-Sahalia (2008). It is applicable whenever the process can be transformed to one

dimensional diffusion process. One common example under analytical approximation of the likelihood function is the Kolmogorov backward and forward equation.

2.4.3. The Kolmogorov Forward Equation

This Numerical solution make used of the Kolmogorov forward equation which uniquely determines the transition density of a diffusion process with respect to a given initial condition. Poulsen (1999) makes use of this description and approximates the transition density by numerically solving this deterministic partial differential equation of the Kolmogorov forward equation which uniquely determines the transition density of a diffusion process with respect to a given initial condition. This approach has already been pursued by Lo (1988), who applied this idea to particular jump-diffusion processes but does not develop a general procedure.

The following considerations assume a one-dimensional diffusion process whose transition density fulfils the Kolmogorov forward equation

$$\frac{\partial \pi(t; x_0, x)}{\partial t} = \frac{\partial (\alpha(x, \theta) \pi(t; x_0, x))}{\partial x} + \frac{1}{2} \frac{(\beta^2(x, \theta) \pi(t; x_0, x))}{\partial x^2} \quad (2.17)$$

for $t \geq 0$ and $x_0, x \in X$. The diffusion is assumed stationary and ergodic.

Indications for handling multi-dimensional diffusion processes are given in the paper by Poulsen (1999). By the product rule, Equation (2.17) is identical with

$$\frac{\partial \pi(t; x_0, x)}{\partial t} = a(x, \theta) \pi(t; x_0, x) + b(x, \theta) \frac{\partial (\pi(t; x_0, x))}{\partial x} + c(x, \theta) \frac{\partial^2 (\pi(t; x_0, x))}{\partial x^2} \quad (2.18)$$

where $a(x, \theta) = \frac{\partial a(x, \theta)}{\partial x} + \frac{1}{2} \frac{\partial^2 \beta^2 a(x, \theta)}{\partial x^2}$

$$b(x, \theta) = -a(x, \theta) + 2\beta(x, \theta) \frac{\partial (x, \theta)}{\partial x}, \text{ and } c(x, \theta) = \frac{1}{2} \beta^2 a(x, \theta).$$

These coefficients are known as functions of x and θ since α and β are known in parametric form. Poulsen (1999) approximates Equation (2.18) by application of the *Crank-Nicolson method* (Crank & Nicolson, 1947). Some more detail on this is given in Poulsen (1999).

Note that this numerical procedure determines several more values of the transition density than actually needed for the approximation of $\ell_n(\theta)$. Poulsen (1999) shows that obtained approximation $\widehat{\ell}_n(\theta)$ of the log-likelihood function $\ell_n(\theta)$ satisfies

$$\widehat{\ell}_n(\theta) = \ell_n(\theta) + h^2 g_n^{(1)}(\theta, x_0, \dots, x_n) + o(h^2) g_n^{(2)}(\theta, x_0, \dots, x_n), \quad (2.19)$$

Where $g_n^{(1)}$ and $g_n^{(2)}$ appropriate functions and $h > 0$ is chosen such that the computing time for the approximated log-likelihood is at most of order n/h^2 .

2.4.4. Simulated Maximum Likelihood Estimation (SMLE)

This estimation describes an approach by Pedersen (1995a) and Santa-Clara (1995) which is known as *simulated maximum likelihood estimation (SMLE)*. It is based on the observation that by the Chapman-Kolmogorov equation the transition density can be expressed as

$$\pi(s, x, t, y) = \int \pi(s, x, t - \Delta, y) \pi(t - \Delta, z, t, y) dz = E(\pi(t - \Delta, X_{t-\Delta}, t, y) | X_0 = x) \quad (2.20)$$

for all $x, y \in X, t > s \geq 0$ and $0 < \Delta < t - s$. For small Δ , usually $\Delta \ll t - s$.

The function $\pi(t - \Delta, \cdot, t, \cdot)$ can be replaced by a Gaussian density, and hence an approximation of the above expectation can be obtained by Monte Carlo integration through repeated (approximate) simulation of $X_{t-\Delta} | \{X_s = x\}$. In the following, if we considered the work by Pedersen (1995a) who defines the first-order approximation

$$\pi_\theta^{(1)}(s, x, t, y) = \phi(y | x + (t - s)\alpha(x, \theta), (t - s)\beta(x, \theta)) \quad (2.21)$$

of π_θ like in the naive maximum likelihood approach above. Once again, the notation

$\phi(z | \mu, \psi)$ refers to a multivariate Gaussian density with mean μ & variance ψ evaluated at z .

As further refinements of $\pi_\theta^{(1)}$, Pedersen introduces for numbers $N \geq 2$

$$\pi_\theta^{(N)}(s, x, t, y) = \int \dots \int \prod_{k=0}^{N-1} \pi_\theta^{(1)}(\tau_k, \xi_k, \tau_{k+1}, \xi_{k+1}) d\xi_1 \dots d\xi_{N-1}, \quad (2.22)$$

where $\tau_k = s + k(t - s)/N$ for $k = 0, \dots, N$, $\xi_0 = x$ and $\xi_N = y$.

There are numbers of suggestions on how to further improve the SMLM approach. One idea is to change the first-order approximation (2.21) in order to reduce the estimation bias without

having to increase the number of subintervals N . Elerian (1998), for example, replaces the Euler scheme by the Milstein scheme. Durham and Gallant (2002) consider the application of a higher order Itô-Taylor expansion as in Kessler (1997) and the local linearisation method. They furthermore suggested various variance-reduction techniques, like the use of antithetic variates. The latter is also applied by Brandt and Santa-Clara (2001). Stramer and Yan (2007) investigated the trade-off between increasing the number of auxiliary time points and increasing the number of simulated diffusion paths.

2.4.5. Local Linearisation

The idea of the *local linearisation method* (Shoji & Ozaki, 1998a) is to approximate the considered diffusion process by a linear one. Like all linear diffusions, the resulting approximation is explicitly solvable (Kloeden & Platen, 1999), i.e. its transition density is available and can serve as an approximation to the true transition density of the original process. The local linearisation constitutes an improvement of the Euler scheme: While the Euler approximation sets the drift and diffusion coefficient piecewise *constant*, the local linearisation method considers them piecewise *linear*.

Under one-dimensional diffusions process, the local linearisation for the stochastic process X of interest fulfils the stochastic differential equation.

$$dX_t = \alpha(X_t, \theta)dt + \beta(\theta)dW_t, \quad X_0 = x_0. \quad (2.23)$$

In case the diffusion coefficient β depends on the state variable, the process can be converted to this form (2.23), using Lamperti's transform. The local linearisation method usually allows the drift to also depend on time. However, is on time-homogeneous diffusion processes.

For X satisfying the stochastic differential equation (2.23), Itô's formula from (2.7) yields

$$d\alpha(X_t, \theta) = \left(\frac{\partial \alpha}{\partial x} \right) (X_t, \theta) dX_t + \frac{1}{2} \left(\frac{\partial^2 \alpha}{\partial x^2} \right) (X_t, \theta) \beta^2(\theta) dt \quad (2.24)$$

where $\partial/\partial x$ denotes differentiation with respect to the state variable. Assume that $\partial^2\alpha/\partial x^2$ are constant in X_t for $t \in [s, s + \Delta s)$, where $s \geq 0$ and $\Delta s > 0$. Then

$$\alpha(X_t, \theta) - \alpha(X_s, \theta) = \left(\frac{\partial \alpha}{\partial x} \right) (X_s, \theta) (X_t - X_s) + \frac{1}{2} \left(\frac{\partial^2 \alpha}{\partial x^2} \right) (X_s, \theta) \beta^2(\theta) (t - s) \quad (2.25)$$

for all $t \in [s, s + \Delta s)$. With this, the drift function of X can be written as

$$\alpha(X_t, \theta) = C_s^{(1)} X_t + C_s^{(2)} t + C_s^{(3)} \quad (2.26)$$

For appropriate constants $C_s^{(1)}$, $C_s^{(2)}$, and $C_s^{(3)}$ that depend only on s and X_s .

The resulting Stochastic Differential Equation is now of the form:

$$dX_t = (C_s^{(1)} X_t + C_s^{(2)} t + C_s^{(3)}) dt + \beta(\theta) dW_t \quad (2.27)$$

for $t \in [s, s + \Delta s)$ has a linear drift function and a constant diffusion coefficient. As indicated above, an explicit solution to such stochastic differential equation is generally available. See Shoji (1998), Shoji and Ozaki (1998a) or Kloeden and Platen (1999) for the specific solution of (2.27). Shoji and Ozaki (1998b) describe this linearisation procedure also for multi-dimensional processes. The method is, however, only applicable where a transformation to constant diffusion coefficient is possible.

2.4.6. Alternatives to Maximum Likelihood Estimation

Alternatives to Maximum Likelihood Estimation approaches do not try to set up or approximate the likelihood function of the parameter. Instead, they match certain statistics of the model with that of the data. These statistics may be the moments of the process or their sample analogues or some functions derived from auxiliary models. The model parameter is then estimated by the candidate who produces the best conformity.

2.4.7. Estimating Functions

An estimating function for the parameter θ is a function $G_n(\theta; \mathbf{x}_0, \dots, \mathbf{x}_n)$ which depends on the parameter and the data. Let $\mathbf{X}$ be an ergodic diffusion which is the solution of the SDE (2.3), and assume one has observations $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n$ of $\mathbf{X}$ at times $0 = t_0 < t_1 < \dots < t_n$.

When dependence on the observations is clear, we simply write $G_n(\theta)$. An estimate $\hat{\theta}_n$ for θ is

obtained as a solution of $G_n(\theta) = 0$. If we let θ_0 denotes the true parameter value. One usually requires that the estimating function is (at least asymptotically) *unbiased*, that means

$$E_\theta[G_n(\theta)] = 0, \quad (2.28)$$

and that the parameter is *uniquely identifiable*, i.e. $E_{\theta(0)}[G_n(\theta)] = 0 \Leftrightarrow \theta = \theta_0$.

The most prominent representative for an unbiased estimating function is the score function

$$S_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} = \sum_{k=0}^{n-1} \frac{\partial}{\partial \theta} \log \pi(\Delta t_k; x_k, x_{k+1}) \quad (2.29)$$

where the equation $S_n(\theta) = 0$ is solved by the maximum likelihood estimator. However, the score function is usually not available. In that case, one tries to imitate it by adapting the general form

$$G_n(\theta, x_0, \dots, x_n) = \sum_{k=0}^{n-1} g(\theta, \Delta t_k; x_k, x_{k+1}) \quad (2.30)$$

In all estimating functions introduced above, the function g , in turn, is of type

$$g(\theta, \Delta t; x, y) = \sum_{j=1}^J a_j(\theta, \Delta t; x) h_j(\theta, \Delta t; x, y) \quad (2.31)$$

for some $J \in \mathbb{N}$, where the a_j are called the *weights* of the functions h_j . The most common estimating functions that appear in the literature can be categorised in the non-disjoint classes.

2.4.8. Martingale Estimating Functions

Martingale estimating functions appear as a natural choice of an estimating function since the score function possesses the martingale property as well. Martingale estimating functions satisfy the martingale property of the form:

$$\mathbb{E}_\theta(G_n(\theta; X_{t_0}, \dots, X_{t_n}) | \mathcal{F}_{n-1}) = G_{n-1}(\theta; X_{t_0}, \dots, X_{t_{n-1}}) \quad (2.32)$$

for all $n \in \mathbb{N}$, where $\mathcal{F}_n$ is the σ -algebra generated by $\{X_{t_0}, \dots, X_{t_n}\}$. If the function g in (2.29) attains the form (2.30), the condition (2.31) is fulfilled if and only if

$$\mathbb{E}_\theta(h_j(\theta, \Delta t, x, X_{\Delta t}) | X_0 = x) = 0 \quad (2.33)$$

for all $j \in \{1, \dots, J\}$, $\theta \in \Theta$, $\Delta t \in \mathbb{R}_+$ and $x \in X$. An obvious choice is

$$h_j(\theta, \Delta t, x, y) = f_j(y, \theta) - \mathbb{E}_{\theta}(f_j(X_{\Delta t}, \theta) | X_0 = x) \quad (2.34)$$

for *base functions* f_j which are regular enough such that the expected values exist. Then (2.34) is trivially fulfilled.

Furthermore, the well-known martingale theory allows for immediate asymptotic results as $n \rightarrow \infty$. Unbiasedness is directly implied by (2.33). For given h_j , $j = 1, \dots, J$, the weights α_j can be chosen in an optimal way such that the resulting estimator of θ has smallest asymptotic variance within the class of estimating functions satisfying (2.29) with the specified h_j (Sørensen, 2007, Bibby et al., 2009). The choice of the h_j is, however, more subtle. For asymptotic properties of martingale estimating functions, and in particular for asymptotic variances, which are typically of sandwich type and depend on the true parameter, see Sørensen (2008) or Bibby et al. (2009).

2.4.9. Polynomial Estimating Functions

Polynomial estimating functions employ the martingale estimating function of the form (2.34) with f_j being a polynomial. They hence form a subgroup of the above martingale estimating functions. Unlike the true score function, polynomial estimating functions do not require knowledge of the whole transition density but only of the first few conditional moments. They are hence more robust to misspecification. When the conditional moments are analytically not available, they may be obtained by Monte Carlo simulation or MCMC. Kessler and Paredes (2002) describe the impact of such simulation on the resulting estimator of parameter of interest θ .

A *linear estimating function* utilises (2.34) with $J = 1$, h_1 as in (0) and $f_1(y, \theta) = y$,

$$h_1(\theta, \Delta t, x, y) = y - \mathbb{E}_{\theta}(X_{\Delta t} | X_0 = x) \quad (2.35)$$

This estimating function is appropriate when the diffusion coefficient does not depend on parameter of interest θ ; otherwise, a higher order polynomial should be employed.

A quadratic estimating function uses $J = 2$, h_1 as above and

$$f_2(\mathbf{y}, \theta) = h_1(\theta, \Delta t, x, y) h_1'(\theta, \Delta t, x, y)$$

$$h_2(\theta, \Delta t, x, y) = h_1(\theta, \Delta t, x, y) h_1'(\theta, \Delta t, x, y) - \text{var}(X_{\Delta t} | X_0 = x) \quad (2.36)$$

When the diffusion process is one-dimensional, higher order polynomial estimating functions usually employ $f_j(\mathbf{y}, \theta) = y^{k_j}$ for suitable $k_j \in \mathbb{N}_0$.

2.4.10. Estimating Functions based on Eigen-functions

The class of estimating functions based on eigen-functions has been investigated by Kessler and Sørensen (1999). This estimating function is explicit in the sense that the h_j are known in explicit form. Kessler and Sørensen (1999) determine the optimal weights for this function and prove consistency and asymptotic normality for the resulting estimators. Let G_θ denote the infinitesimal generator of a one-dimensional diffusion process X which solves the SDE (2.7). A twice differentiable function $\eta(x, \theta)$ is an *eigen-function* of G_θ with *eigen-value* $\lambda(\theta) \in \mathbb{R}_0$ if

$$G_\theta \eta(x, \theta) = -\lambda(\theta) \eta(x, \theta) \quad (2.37)$$

for all $x \in X$.

Kessler and Sørensen (1999) show that under mild regularity conditions one has

$$\mathbb{E}(\eta(X_{\Delta t}, \theta) | X_0 = x) = \exp(-\lambda(\theta) \Delta t) \eta(x, \theta). \quad (2.38)$$

Hence, if $\eta_1, \dots, \eta_K$ are eigen-functions of G_θ with eigen-values $\lambda_1, \dots, \lambda_K$, the choice (2.33) with $J = K$ and $h_j(\theta, \Delta t, x, y) = \eta_j(y, \theta) - \exp(-\lambda_j(\theta) \Delta t) \eta_j(x, \theta)$. i.e. h_j being as in (2.34) with $f_j(y, \theta) = \eta_j(y, \theta)$, yields a martingale estimating function. For eigenfunctions of the generator of a multi-dimensional diffusion process see Bibby and Sørensen (1995).

2.4.11. Simple Estimating Functions

In the class of simple estimating functions, the function g in estimating functions (2.34) depends on one state variable only, in particular

$$g(\theta, \Delta t, x, y) = \tilde{g}(\theta, x) \text{ or } g(\theta, \Delta t, x, y) = \bar{g}(\theta, y) \quad (2.39)$$

for some functions $\tilde{g}$ and $\bar{g}$.

Simple estimating functions have the advantage that they are often explicitly available. On the other hand, they do not take into account the dependence structure between successive observations.

That completes the collection of those classes of estimating functions which are considered as an alternative to MLE explained in 2.4.1 above. Different estimating functions can be combined and used for different parameter components, as for example for drift parameters and diffusion parameters. See Bibby and Sørensen (2001) for several examples. Jacobsen (2001) investigates the problem of finding optimal or asymptotically optimal estimating functions in the sense that the resulting parameter estimate has smallest variance within a certain class of functions.

2.4.12. Generalised Method of Moments (GMM)

Related to the theory of estimating functions in the previous estimation procedure is the *generalised method of moments (GMM)* as developed by Hansen (1982). This approach, considers functions $\psi_1, \dots, \psi_J$ that depending on the parameter θ and the process X , where

$$\mathbb{E}(\psi_j(\theta; X)) = 0 \quad (2.40)$$

for $j = 1, \dots, J$. Equations (2.36) for all j are called the *moment conditions*.

These are employed in order to estimate the parameter of interest θ . For d -dimensional parameter θ , one requires

$J \geq p$ moment conditions. The functions ψ_j usually depend on only one or two state variables in addition to the parameter. The generalised method of moment (GMM) assumed in the following.

Consider a one-dimensional ergodic diffusion process with true parameter θ_0 starting in x_0 at time t_0 , and assume that there are observations $x_1, \dots, x_n$ at times $t_1 < \dots < t_n$ of this process which are independent identically distribution draws from its invariant measure π_{θ_0} . Denote by X_s and X_t two independent random variables with density π_{θ_0} . The expected value of $\psi_j(\theta;$

X_s, X_t) for $\theta \in \Theta$ is usually not available but can be approximated through the *method of moments estimator* thus:

$$E(\psi_j(\theta; X)) = \frac{1}{n} \sum_{k=0}^{n-1} \psi_j(\theta; x_k, x_{k+1}) \quad (2.41)$$

The *GMM estimator* $\hat{\theta}_n$ is obtained as the minimiser of a norm of this expression. In particular,

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{k=0}^{n-1} \psi(\theta; x_k, x_{k+1}) \right)' C_n \left(\frac{1}{n} \sum_{k=0}^{n-1} \psi(\theta; x_k, x_{k+1}) \right) \quad (2.42)$$

where $\psi = (\psi_1, \dots, \psi_J)^1$ and $C_n \in \mathbb{R}^{J \times J}$ is a positive semi-definite weight matrix. Under certain conditions (Hansen, 1982) deduced that $\hat{\theta}_n$ is consistent and asymptotically normal. Hansen (1982) determines the optimal weight matrix yielding an asymptotically efficient estimator, but as this matrix involves the unknown parameter θ_0 , it cannot be used in practice.

2.4.13. Simulated Moments Estimation (SME)

Simulated moments estimation (*SME*) was carried out by Duffie and Singleton (1993) bases on the same idea as the just considered generalised method of moments. In the previous estimation procedure (2.42) was minimised. Due to the moment conditions (2.40) above is exactly the same as minimising a norm of

$$\frac{1}{n} \sum_{k=0}^{n-1} \psi_j(\theta; x_k, x_{k+1}) - E(\psi_j(\theta; X_s, X_t)) \quad (2.43)$$

for some $t > s \geq 0$. The generalised method of moments could hence theoretically be extended to functions $\psi_j(\theta; X_s, X_t)$ with non-zero expectation as long as this expectation is known. When it is unknown, the SME is still appropriate as it replaces the analytical expected value in (2.41) by the sample mean based on $m + 1$ additionally simulated values $z_0, z_1, \dots, z_m$ from the invariant density π . The *simulated moment's estimator* is obtained as the minimiser of the norm of

$$\frac{1}{n} \sum_{k=0}^{n-1} \psi_j(\theta; x_k, x_{k+1}) - \frac{1}{m} \sum_{i=0}^{m-1} \psi_j(\theta; z_i, z_{i+1}) \quad (2.44)$$

in an analogous manner as in above. Duffie and Singleton (1993) supply conditions for the consistency and asymptotic normality of the simulated moment's estimator.

2.4.14. Indirect Inference

In indirect inference, the parameter of a model of interest is estimated indirectly via the parameter of an auxiliary model (Gourieroux, et al., 1993). This is convenient whenever statistical inference for the original model is complicated and the auxiliary model is chosen such that its parameter can easily be estimated.

Let $\mathcal{M}_\theta$ denote the original model with parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, and A_ρ be the auxiliary model with parameter $\rho \in \mathbb{R} \subseteq \mathbb{R}^h$, where $h \geq p$. Both $\mathcal{M}_\theta$ and A_ρ are assumed to be known in parametric form. Suppose one has observations $\mathbf{x}_0, \dots, \mathbf{x}_n$ from $\mathcal{M}_{\theta_0}$ for some unknown $\theta_0 \in \Theta$. The objective is the estimation of θ_0 . If simulation from $\mathcal{M}_\theta$ is (at least approximately) possible, this can be performed by indirect inference as follows:

- (i) Obtain an estimator $\hat{\rho}_n^{obs}$ of ρ by treating $\mathbf{x}_0, \dots, \mathbf{x}_n$ as observations from the auxiliary model A_ρ . Under regularity conditions, this estimator $\hat{\rho}_n^{obs}$ tends to a parameter $\rho_0 = g(\theta_0)$ for some invertible unknown function g as $n \rightarrow \infty$.
- (ii) Determine θ such that simulated data from $\mathcal{M}_\theta$ associated with A_ρ leads to an estimate of ρ that is close to $\hat{\rho}_n^{obs}$: For fixed θ , (approximately) simulate K datasets $\mathbf{z}^{(\theta,k)} = \{z_0^{(\theta,k)}, \dots, z_n^{(\theta,k)}\}$, $k = 1, \dots, K$, from $\mathcal{M}_\theta$. Denote by $\hat{\rho}_n^{(\theta,k)}$ the estimator of ρ that is obtained when $\mathbf{z}^{(\theta,k)}$ is treated as observed from A_ρ . Then, if certain assumptions are fulfilled, $\hat{\rho}_n^{(\theta,k)}$ tends to $g(\theta)$ for all k and $n \rightarrow \infty$.

The indirect estimator $\hat{\theta}_n$ of θ is now chosen such that $\hat{\rho}_n^{(sim)}(\theta) = K^{-1} \sum_{k=1}^K \hat{\rho}_n^{(\theta,k)}$ is close to $\hat{\rho}_n^{obs}$.

In particular,

$$\hat{\theta}_n = \underset{\theta \in \Theta}{argmin} (\hat{\rho}_n^{obs} - \hat{\rho}_n^{sim}(\theta))' D_n (\hat{\rho}_n^{obs} - \hat{\rho}_n^{sim}(\theta)) \quad (2.45)$$

where $\mathbf{D}_n$ is a positive definite matrix converging to a deterministic positive definite matrix $\mathbf{D}$ as $n \rightarrow \infty$.

Under fairly general assumptions, the indirect estimator $\hat{\theta}_n$ is a consistent and asymptotically normal estimator of parameter θ . Gouriéroux et al. (1993) further suggested asymptotic properties and fields of application of the estimator. In case of approximate simulation, they hence refer to the above method as *quasi-indirect inference*. Monte Carlo experiments, however, show good performance of this approach for moderate sample sizes. Furthermore, the authors remark that estimation results do not improve if the Milstein instead of the Euler scheme is employed for the approximate simulation. Overall, the resulting estimator is asymptotically unbiased also for quasi-indirect inference (Gouriéroux et al., 1993, Broze et al., 1998).

2.4.15. Efficient Method of Moments (EMM)

A conceptually similar approach to the indirect inference in the previous estimation procedure is the *efficient method of moments (EMM)* (Gallant & Tauchen, 1996). In this technique, the criterion of indirect inference method in (2.41) is replaced by

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} Q(\theta)' E_n Q(\theta) \quad (2.46)$$

where

E_n , is an appropriate positive definite matrix and function $Q(\theta)$ is define as:

$$Q(\theta) = \frac{1}{K} \sum_{k=1}^K \frac{\partial q}{\partial p} \left(\hat{\rho}_n^{obs}; z_0^{(\theta,k)}, \dots, z_n^{(\theta,k)} \right) \quad (2.47)$$

q is the log-likelihood function of ρ in the auxiliary model. For certain choices of $\mathbf{D}_n$ and E_n , the estimators $\hat{\theta}_n$ obtained through (2.38) and (2.46) are asymptotically equivalent (Gouriéroux et al., 1993). The EMM is generally more efficient than indirect inference as the

computationally possibly demanding calculation of the maximum likelihood estimator $\hat{\rho}_n^{(\theta,k)}$ in Equation (2.42) is not required. For details on the EMM, see Gallant and Tauchen (1996).

2.4.16. Exact Algorithm (EA)

The Exact Algorithm (EA) enables exact simulation of diffusion paths without any time discretisation error. By now, the algorithm is available in the different variants such as Exact Algorithm 1 (EA1) Beskos & Roberts,(2005), EA2 (Beskos, Papaspiliopoulos, & Roberts, 2006) and EA3 (Beskos, Papaspiliopoulos, & Roberts, 2008). Its implementation is easy and computationally efficient. A drawback, however, is its limited applicability. The EA1 is the earliest and simplest version of the Exact Algorithm discusses the simulation of diffusion paths and then the estimation of parameters by consider a one-dimensional diffusion process $X = (X_t)_{t \geq 0}$ with unit diffusion coefficient, i. e. satisfying an stochastic differential equation (SDE) or diffusion process.

$$dX_t = \alpha(X_t, \theta)dt + dW_t \quad , X_0 = x_0 \quad (2.48)$$

EA1 applies only to this class of diffusions. However, transformation of a general one-dimensional diffusion (2.48) can be obtained with Lamperti's transform. The EA1 aims to draw exact time-discrete skeletons of the diffusion process on the time interval $[0, T]$, where $T \in \mathbb{R}_+$ is fixed. The algorithm is based on the rejection sampling scheme (Grimmett & Stirzaker, 2001).

The difficulty that underlies these approaches is the general intractability of the transition density for discrete-time observations. Most techniques struggle when inter-observation times are large. In general, the application of the above methods becomes problematic as soon as a diffusion process is only partial discretely observed, where some components of the state vector are latent, or observations are measured with error.

A powerful approach to overcome this problem is to estimate the model parameters in a Bayesian framework. Such procedure is able to handle the difficulties aforementioned and

enables the statistical analysis of the complex applications. Large samples or Stationarity of the underlying process are not required under Bayesian approaches. As an appreciated by-product, the technique also estimates the sample path at intermediate observation times. This approach gives room for introducing missing data, such that, the union of missing values and observations forms either high-frequency or low-frequency dataset. This facilitates approximation of the transition density and hence enables parametric inference even for large inter-observation times. Moreover, the techniques are suitable for irregularly spaced observation intervals, multivariate diffusions with possibly latent components and for observations that are subject to measurement error.

2.5 Bayesian Approach

Bayesian approaches have received much attention in the recent literature, due to its wide applicability. In statistics, Bayesian approach is a method of inference in which Baye's rule is used to update the probability estimate for a hypothesis as additional evidence is required (Gilks et al., 1996; Brooks, 1998). The idea behind Bayesian inference is that the likelihood and prior are combined using Bayes' theorem to compute the posterior distribution. Bayesian methods combine prior evidence on the parameters contained in the density $\pi(\theta)$ with the likelihood $\pi(x | \theta)$ to produce the entire posterior density $\pi(\theta | x)$ of θ . From which, one may extract any information of a parameter, as with maximum likelihood estimations.

Bayesian inference about θ is primarily based on the posterior distribution of θ . There are various ways in which you can summarize this distribution. For example, you can report your findings through point estimates. You can also use the posterior distribution to construct hypothesis tests or probability statements.

Bayesian methodology provides a useful means to infer unobserved or missing data, since it treats all unknown parameters and data alike as random variables, for which full posterior distributions can be estimated.

The Bayesian frameworks propose is to address the inference difficulty, especially where our model is of multidimensional diffusions, where some components are partially observed. This makes model parameters more difficult to estimate, because less information is available compared to the fully observed cases.

Bayesian approaches to inference in the field of epidemiology have been growing in popularity over the past two decades (O’Neil and Roberts, 1999; O’Neil, 2002). One common approach in the literature for Bayesian estimation of diffusion models, studied independently by Jones (1999), Eraker (2001), and Elerian et al. (2001), is to consider estimating diffusion models on the basis of discrete measurements as a classic missing-data problem. The idea is to introduce augmented data points between every two consecutive observed data points so that the likelihood can be well approximated. An Euler-Maruyama scheme is then applied by approximating the transition densities over the induced discretisation as Gaussian. Computationally intensive algorithms, such as Markov chain Monte Carlo (MCMC), are then used to integrate over the uncertainty associated with the missing/unobserved data.

Roberts and Stramer (2001) first applied this idea successfully for parameter estimation in a nonlinear diffusion model. The Bayesian imputation frameworks propose is to address the inference difficulty, especially where our model is of multidimensional diffusions, especially when some space components are partially observed. This makes model parameters more difficult to estimate, because less information is available compared to the fully observed cases.

There are various potential advantages of Bayesian inference on the system for modelling epidemic processes. Firstly, since it is based on simulation models it provides a natural

method for imputing missing data. Secondly, the simulated data are matched to the observed data based on some form of summary (or sufficient) statistic. This allows key features of the epidemic to be used to drive the model fit, even though information on individual events or event times may be incomplete or unobserved. Thirdly, the use of a Bayesian framework allows prior information about the parameters to be easily incorporated into the model. This is particularly useful in epidemic systems where there is often a high degree of correlation between parameters, but for which prior information about some of the sub-processes are sometimes known.

2.5.1 Prior Distributions

In Bayesian inference, a prior probability distribution often called simply the prior of an uncertain parameter θ or latent variable is a probability distribution that expresses uncertainty about the parameter θ before the current data are examined. When applying Bayes' theorem, the prior is multiplied by the likelihood function leads to the posterior distribution. Prior distributions have traditionally belonged to one of two interpretations of probability, viz: Objective priors and Subjective priors. Within Bayesian inference, these two interpretation of probability leads to Objective Bayesian inference (Berger, 2006) and Subjective Bayesian inference (Anscombe and Aumann 1963; Goldstein 2006). Differences in the interpretation of probability are best explored outside of this study.

2.5.2 Objective Priors versus Subjective Priors

Bayesian probability measures the degree of belief that you have in a random event. By this definition, probability is highly subjective. It follows that all priors are subjective priors. Not everyone agrees with this notion of subjectivity when it comes to specifying prior distributions. There has long been a desire to obtain results that are objectively valid. Within the Bayesian paradigm, this can be somewhat achieved by using prior distributions that are “objective” (that have a minimal impact on the posterior distribution). Such distributions are

called objective or noninformative priors. However, while noninformative priors are very popular in some applications, they are not always easy to construct (DeGroot and Schervish (2002); Berger (2006) and Goldstein (2006)).

This study explained the four categories of prior distributions viz: Informative, weakly informative, least information, and uninformative priors.

(a) Informative priors

The prior distributions are refers to informative priors when prior/previous information is available about the parameter θ , and included in the prior distribution examined in the current data. This is often available in the model from epidemic data, as a result of previous outbreak of such epidemic used to available. The posterior distribution of θ from the previous models may be used as the prior distribution of θ as was done in our analysis in chapter four.

(b) Weakly informative priors (WIP)

In weakly informative prior (WIP), distributions use prior information by introduce additional information in order to solve an ill-posed problem or to prevent over-fitting. This would provide enough prior information to [prevent results that contradict our knowledge or problems. In addition, allow WIPs to use less prior information than is actually available.

(c) Least informative priors (LIP)

The term least informative priors (LIP) is used to describe a class of prior in which the goal is to minimize the amount of subjective information content, and to use a prior that is determined solely by the model and observed data (to let the data speak for themselves). This type of prior is often favoured by objective Bayesians.

(d) Uninformative or Noninformative Priors

Roughly speaking, a prior distribution is noninformative if the prior is “flat” relative to the likelihood function. Thus, a prior $\pi(\theta)$ is noninformative if it has minimal impact on the posterior distribution of θ . Other names for the noninformative prior are vague, diffuse, and

flat prior. Many statisticians favour noninformative priors because they appear to be more objective. However, it is unrealistic to expect that noninformative priors represent total ignorance about the parameter of interest. In some cases, noninformative priors can lead to improper posteriors (nonintegrable posterior density). In addition, noninformative priors are often not invariant under transformation; that is, a prior might be noninformative in one parameterization but not necessarily noninformative if a transformation is applied.

2.5.3 Proper and Improper Priors

It is important for the prior distribution to be proper. A prior distribution $\pi(\theta)$ is said to be improper if:

$$\int \pi(\theta) d\theta = \infty \quad (2.49)$$

For example, an unbounded uniform prior distribution is an improper prior distribution because, $\pi(\theta) \propto 1$, for $-\infty < \theta < \infty$. Improper priors are often used in Bayesian inference since they usually yield noninformative priors and proper posterior distributions. Improper prior distributions can lead to posterior impropriety. To determine whether a posterior distribution is proper, you need to make sure that the normalizing constant $\int \pi(y|\theta)\pi(\theta)d\theta$ is finite for all y . If an improper prior distribution leads to an improper posterior distribution, inference based on the improper posterior distribution is invalid.

2.5.4 Conjugate Priors

A prior is said to be a conjugate prior for a family of distributions if the prior and posterior distributions are from the same family, which means that the form of the posterior has the same distributional form as the prior distribution. For example, if the likelihood is binomial, $y \sim \text{Bin}(n, \theta)$, a conjugate prior on θ is the beta distribution; it follows that the posterior distribution of θ is also a beta distribution. Other commonly used conjugate prior/likelihood combinations include the normal/normal (self-conjugate), gamma/Poisson, gamma/gamma, and gamma/beta cases. The development of conjugate priors was partially driven by a desire

for computational convenience; conjugacy provides a practical way to obtain the posterior distributions. All probability distributions in the exponential family have conjugate priors Robert (2007).

2.5.5 Jeffreys' Prior

A very useful prior is Jeffrey's' prior or Jeffrey's rule (Jeffreys (1961)). It satisfies the local uniformity property: a prior that does not change much over the region in which the likelihood is significant and does not assume large values outside that range. It is based on the Fisher information matrix. Jeffreys' prior is defined as

$$\pi(\theta) \propto |I(\theta)|^{1/2} \quad (2.50)$$

where $| \cdot |$ denotes the determinant and $I(\theta)$ is the Fisher information matrix based on the likelihood function $p(y|\theta)$

$$I(\theta) = -E \left[\frac{\partial^2 \log p(y|\theta)}{\partial \theta^2} \right] \quad (2.51)$$

Jeffreys' prior is locally uniform and hence noninformative. It provides an automated scheme for finding a noninformative prior for any parametric model $p(y|\theta)$. Another appealing property of Jeffreys' prior is that it is invariant with respect to one-to-one transformations. The invariance property means that if you have a locally uniform prior on θ and $\phi(\theta)$ is a one-to-one function of θ , then $p(\phi(\theta)) = \pi(\theta) \times |\phi'(\theta)|^{-1}$ is a locally uniform prior for $\phi(\theta)$. This invariance principle carries through to multidimensional parameters as well. While Jeffreys' prior provides a general recipe for obtaining noninformative priors, it has some shortcomings: the prior is improper for many models, and it can lead to improper posterior in some cases; and the prior can be cumbersome to use in high dimensions.

2.5.6 Maximum Entropy (MAXENT) prior

A MAXENT prior, proposed by Jaynes (1968), is a prior distribution that is selected among other candidate distributions as the prior of choice when it has the maximum entropy

(MAXENT) in the considered set, given constraints on the candidate set. More entropy is associated with less information, and the least informative prior is preferred as a MAXENT prior. The principle of minimum cross-entropy generalizes MAXENT priors from mere selection to updating the prior given constraints while seeking the maximum, possible entropy.

2.5.7 Reference Priors

When reference priors are used, the analysis is called reference analysis, and the posterior is called the reference posterior. Introduced by Bernardo (1979), reference priors do not express personal beliefs. Instead, reference priors allow the data to dominate the prior and posterior (Berger, Bernardo, and Dongchu 2009). Reference priors are estimated by maximizing the expected intrinsic discrepancy between the posterior distribution and prior distribution. This maximizes the expected posterior information about data when the prior density is $\pi(y)$. In some sense, $\pi(y)$ is the least informative prior about data (Bernardo 2005). Reference priors are often the objective prior of choice in multivariate problems, since other rules (e.g., Jeffreys rule) may result in priors with problematic behavior. Subjective Bayesian criticisms of reference priors are that the concepts of regularization and stabilization are not taken into account, results that contradict knowledge are not prevented, a numerical approximation algorithm may become stuck in low-probability or at regions, and it may not be desirable to let the data speak fully.

2.5.8 Bayesian Point Estimation

Once the posterior distribution or density $\pi(\theta | y)$ has been obtained, Bayesian point estimates of the model parameters θ can be calculated or obtained.

The three most commonly used point estimates are as follows:

- (a) The posterior mean (PM) of parameter θ is

$$E(\theta | y) = \int \int \theta dF(\theta | y) d\theta = \begin{cases} \int \theta \pi(\theta | y) d\theta & \text{continuous} \\ \sum_{\theta} \theta \pi(\theta | y) & \text{discrete} \end{cases}$$

(b) The posterior mode of parameter θ is

Mode ($\theta | y$) = any value $m \in \mathfrak{R}$ which satisfies

$$\pi(\theta = m | x) = \max_{\theta} \pi(\theta | x)$$

$$\lim_{\theta \rightarrow m} \pi(\theta | x) = \text{Sup} \pi(\theta | x)$$

or the set of all such values.

(c) The posterior median of parameter θ is

Median ($\theta | y$) = any value m of θ such that

$$\pi(\theta \leq m | y) \geq \frac{1}{2}$$

$$\text{and } \pi(\theta \geq m | y) \geq \frac{1}{2}$$

or the set of all such values.

2.5.9 Bayesian Interval Estimation

The second type of inference we consider is interval estimation. We wish to find an interval (l, u) that has a predetermined probability of containing the parameter. There are many ways to construct a Bayesian interval estimate, but the three most common ways are defined as follows:

(a) The $(1-\alpha)$ posterior credible interval (PCI)

A range of values that has a known high posterior probability, $(1-\alpha)$, of containing the parameter is known as a Bayesian credible interval. The Bayesian set estimates are called credible sets, which is also known as credible intervals. This is analogous to the concept of confidence intervals used in classical statistics. Given a posterior distribution $\pi(\theta|y)$, A is a credible set for θ if

$$\pi(\theta \in A | y) = \int_A \pi(y | \theta) d\theta \quad (2.52)$$

For example, you can construct a 95% credible set for θ by finding an interval, A , over which

$$\int_A \pi(y | \theta) d\theta = 0.95.$$

You can construct credible sets that have equal tails. A $100(1-\alpha)\%$ equal-tail interval corresponds to the $100(\alpha/2)th$ and $100(1 - \alpha/2)th$ percentiles of the posterior distribution. Some statisticians prefer this interval because it is invariant under transformations.

Bayesian credible interval summarizes our beliefs about the parameter values that could credibly be believed given the observed data that occurred. Another frequently used Bayesian credible set is called the highest posterior density region (HPDR) and central posterior density region (CPDR) interval.

- (b) The $1-\alpha$ (or $100(1-\alpha)\%$) highest posterior density region (HPDR)

The highest posterior density region for parameter θ is the smallest set S such that

$$\pi(\theta \in S | y) \geq 1 - \alpha \text{ and } \pi(\theta_1 | y) \geq \pi(\theta_2 | y) \text{ if } \theta_1 \in S \text{ and } \theta_2 \notin S.$$

- (c) The $1-\alpha$ (or $100(1-\alpha)\%$) central posterior density region (HPDR)

The $1-\alpha$ central posterior density region (HPDR) for a scalar parameter θ may be defined as the shortest single interval $[a, b]$ such that:

$$\pi(\theta < a | y) \leq \alpha/2 \text{ and } \pi(\theta > b | y) \leq \alpha/2$$

In this study, we adopted the posterior mean (PM) under point estimation approach and posterior credible interval (PCI) on Bayesian interval estimation to carry out the simulate assessment results.

Bayesian approaches are interested in exploring an alternative approach in which the explicit calculation of the full joint likelihood is replaced with Monte Carlo approximations based on simulating from the underlying model, an idea dating back at least to Diggle and Gratton (1984). These ideas have recently been extended to the Bayesian framework, in which approximate posterior distributions for the unknown parameters can be generated in a similar vein (Tavaré et al., 1997).

2.6 Sampling methods

This research work solely focus on simulation method of sampling, specifically, Markov chain Monte Carlo (MCMC) method of sampling, where a posterior density known only up to a normalising constant. For multi-dimensional distributions $\pi(\cdot)$, samples of each component will be simulated values from the marginal density of that component.

2.6.1 Markov Chain Monte Carlo (MCMC)

The markov property is an assumption on the conditional probability distribution that allows the future state to be independent of the past state history. A finite state stochastic process with this property is called a Markov chain. In a discrete time setting, the Markov property is

$$\begin{aligned}\mathbb{P}(X_{(tn+1)}) = j | X_{(t0)} = i_0, X_{(t1)} = i_1, \dots, X_{(tn)} = i) \\ = \mathbb{P}(X_{(tn+1)}) = j | X_{(tn)} = i\end{aligned}\tag{2.53}$$

Markov Chain Monte Carlo (MCMC) estimation approximates the posterior distribution for parameter θ or latent variables, $Y = X_{[0,T]}$, given observations X , through samples obtained by generating a sequence of Markov Chain $\{\theta^{(g)}, Y^{(g)}\}_{g=1}^G$ from the posterior distribution $\pi(\theta, Y | X)$. Without any optimization, one can estimate variables of interests by summarizing the statistics of these simulated samples. For example, the posterior mean estimate of $f(\theta, Y)$ defined by

$$\mathbb{E}[f(\theta, Y) | X] = \int f(\theta, Y) \pi(\theta, Y | X) d\theta dY\tag{2.54}$$

$$\text{This is obtained as } \frac{1}{G} \sum_{g=1}^G f(\theta^{(g)}, Y^{(g)})\tag{2.55}$$

Markov Chain Monte Carlo (MCMC) is a very powerful tool employed to approximately draw samples from a specific distribution π say, which is often called as target distribution. Most of the MCMC algorithms used in practice satisfy the conditions which ensure

convergence to the invariant distribution π . From a statistical perspective, the convergence in distribution of the Markov Chain to target distribution π is exploited to estimate expectations under the invariant measure. Surprisingly, there are many ways to construct a MCMC with stationary distribution. Perhaps the simplest is the Metropolis-Hastings (M-H) algorithm (Hastings, 1970).

2.6.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm is a general algorithm used to construct Markov chains with stationary distribution $\pi(\theta)$. The method was first suggested by Metropolis, Rosenbluth, Teller & Teller (1953) and later adapted by Hastings (1970). As such algorithms do not require a closed form representation of the target distribution $\pi(\theta)$, they offer great utility in the Bayesian inference, as posterior distributions are often non-standard. Central to Metropolis-Hastings is the idea of a proposal density, denoted $q(\cdot | \cdot)$, which is some arbitrary transition kernel. It can be advantageous to have a proposal density which is easy to simulate from, however it need not (necessarily) have $\pi(\theta | D_T)$ as its stationary distribution.

Algorithm 2.1 shows the general procedure to construct the Metropolis-Hastings algorithm.

Algorithm 2.1 The Metropolis-Hastings algorithm

1. Initialise the iteration counter to $i = 1$, and choose initial value $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})^t$

Where $\theta^{(0)}$ is chosen from somewhere in the support of $\pi(\theta | D_T)$.

2. Sample a candidate value, θ^* , from a proposal distribution $q(\theta^* | \theta^{(i-1)})$ or the transition kernel.

3. Evaluate the acceptance probability α as:

$$\alpha(\theta^*, \theta^{(i)}) = \min \left\{ 1, \frac{\pi(\theta^* | D_T) \pi(\theta^{(i)}) q(\theta^{(i)} | \theta^*)}{\pi(\theta^{(i)} | D_T) \pi(\theta^*) q(\theta^* | \theta^{(i)})} \right\}$$

4. Sample U from a $U(0,1)$ distribution and set $\theta^{(i+1)} = \theta^*$ if $U < \alpha(\theta^*, \theta^{(i)})$, and $\theta^{(i+1)} = \theta^{(i)}$ otherwise.

5. Set $i = i + 1$, go to step 2 and repeat for the required number of samples.

Hence at each stage, a new value is generated from the proposal distribution and is either accepted (and the chain moves) or rejected (and the chain stays where it is). Note that $\pi(\cdot)$ only enters the acceptance probability as a ratio and therefore the normalising constant of the density does not need to be known.

Naturally, the mixing properties of the M-H algorithm will depend crucially on the choice of proposal distribution and naive choices of proposal distribution q may lead to long mixing times. In particular, a good choice of $q(\cdot | \cdot)$ will lead to a chain which converges rapidly and mixes well; i.e. it moves often and well around the support of posterior distribution $\pi(\theta | D_T)$.

2.6.3 Metropolis Method /A symmetric proposal

The simplest scheme to implement is the Metropolis method which is based on the use of a symmetric proposal with $q(\theta^* | \theta) = q(\theta | \theta^*)$. Clearly, using such a proposal distribution results in the simplification of the acceptance probability, with above acceptance probability in algorithm 2.1 $\alpha(\theta^*, \theta^{(i)})$.

$$\alpha(\theta^*, \theta^{(i)}) = \frac{\pi(\theta^*)}{\pi(\theta^{(i)})} \quad (2.56)$$

A special case of the Metropolis method often considered in the sampler is the random walk metropolis.

2.6.4 Random walk Metropolis

A frequently used idea is a random walk proposal, which is independent of the imputed and observed data. The random walk proposal, in contrast, is always available, easy to implement and not problem-specific apart from the domain of the parameter.

If it is possible to use a random walk as the proposal distribution $q(\cdot | \cdot)$ by consider the second step (step 2) of Algorithm 2.1 above, then the proposal kernel $q(\cdot | \cdot)$ takes the form

$$\theta^* = \theta^{(i)} + \omega \quad (2.57)$$

where ω are independent identically distributed random variates known as innovations.

Typically, the common choices for the distribution innovation ω include the uniform distribution with zero mean and some specified interval width, and the Gaussian distribution with zero mean and some variance. In this instance the Metropolis-Hastings algorithm is known as a random walk sampler (or random walk Metropolis). The variance of the Gaussian random variates innovation ω often determine the mixing of the chain; if there is too low of

variance of innovation, the chain will explore the space slowly, as many proposed values will be accepted. Too large innovation variance; few proposed values will be accepted. Reflecting the correlation within θ in the covariance structure of innovation ω is an important aspect in ensuring the chain efficiently explores the space. If the target distribution is Gaussian, a general set of sufficient conditions for which the optimal acceptance probability is 0.234. Gelman et al (1996) and Roberts et al. (1997) suggested that the variance of innovation ω under random walk proposal should be given by

$$\frac{2.38^2 \text{Var}(\theta | D_T)}{p} \quad (2.58)$$

where $\text{Var}(\theta | D_T)$ is the variance matrix of the target distribution $\pi(\theta | D_T)$. Typically though, $\text{Var} \pi(\theta | D_T)$ will not be available and hence an estimate from one or more pilot runs should be used. We note that, for large p , sampling θ^* from a multivariate Normal distribution may be expensive. In these instances, an alternative approach is to take the components of $\omega = (\omega_1, \dots, \omega_p)'$ as independent identically distributed (univariate) Normal random variates. For example, $\omega_i \sim N(0, \sigma_i^2)$, where

$$\sigma_i^2 = \frac{2.38^2}{p} \text{Var}(\theta | D_T) \quad (2.59)$$

Here, the proposed value θ at time say s , is generated via $\theta^* = \theta_{(s-1)} + \omega$.

2.6.5 The independence sampler

As the name suggests, an independence sampler (or independence chain) proposes a new value θ^* independently of the current value $\theta^{(i)}$. Hence, $q(\theta^* | \theta^{(i)}) = f(\theta^*)$ for some density $f(\cdot)$. Whilst the form of such a proposal may appear to disagree with the Markovian structure of the chain, both $\theta^{(i)}$ and θ^* feature in the acceptance probability, meaning a proposal still depends upon the current state, and thus, the Markov property is preserved. Using such a proposal distribution leads to an acceptance probability of the form

$$\alpha(\theta^*, \theta^{(i)}) = \min \left\{ 1, \frac{\pi(\theta^* | D_T)}{\pi(\theta^{(i)} | D_T)} \times \frac{f(\theta)}{f(\theta^*)} \right\} \quad (2.60)$$

Clearly, we can increase the acceptance probability by making $f(\cdot)$ and $\pi(\cdot | D_T)$ as similar as possible. It is worth noting that in the context of an independence sampler (and in contradiction to the above on random walk Metropolis), the higher the acceptance probability, the better. Tierney (1994) suggests the avoidance of densities $f(\cdot)$ with thin tails.

2.6.6 Gibbs sampling

The Gibbs sampler (or generically Gibbs sampling) originated in the field of image processing. It was introduced by Geman and Geman (1984) before being generalised and brought to the interest of the larger statistical community by Gelfand and Smith (1990). In essence the Gibbs sampler is a MCMC scheme in which the full conditional distributions are used to form the transition kernel and is only meaningful in problems where the state is multivariate or certain components of parameters of interest θ in multidimensional problems. The Gibbs sampler is given by Algorithm 2.2.

Algorithm 2.2 The Gibbs Sampler Algorithm

1. Initialise the iteration counter to $j = 1$ and initialise the state of the chain with $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})'$
 2. Obtain a new value $\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_p^{(j)})'$ from $\theta^{(j-1)}$ by successive generation of values

$$\begin{aligned} \theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}, D_T) \\ \theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}, D_T) \\ &\vdots \\ \theta_d^{(j)} &\sim \pi(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}, D_T) \end{aligned}$$
 3. Change counters j to $j+1$, and return to step 2.
-

The chain approaches its equilibrium state as the number of iterations increases, and once the chain has converged, a value of $\theta^{(j)}$ is a sample from $\pi(\theta | D_T)$. Thus the Gibbs sampler is a way to sample from $\pi(\theta | D_T)$ when direct sampling is costly, complicated or indeed impossible, but sampling from $\pi(\theta_j | \theta_{-j}, D_T)$ is possible. For details of other versions of the Gibbs sampler see, for example Gamerman and Lopes (2006).

2.6.7 Componentwise transitions (*Metropolis-within-Gibbs*).

In practice, the construction of a suitable proposal density could be difficult. However, for many problems of interest, it may be possible to sample from the full conditional distributions for a subset of θ . Let the full conditional distribution for the j^{th} component of θ be denoted by $\pi(\theta_j | \theta_1, \theta_2, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p, D_T) = \pi(\theta_j | \theta_{-j}, D_T)$ $j = 1, \dots, p$.

Then, the algorithm for componentwise transitions is given by Algorithm 2.3.

Algorithm 2.3 Componentwise Transitions/ Metropolis-within-Gibbs

1. Initialise the iteration counter to $j=1$ and initialise the chain with $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})^T$
 2. Obtain a new value $\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_p^{(j)})^T$ from $\theta^{(j-1)}$ by successive generation of values

$$\theta_1^{(j)} \sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_p^{(j-1)}, D_T) \text{ Using a Metropolis-Hastings step with proposal } q_1(\theta_1^* | \theta_1^{i-1})$$

$$\theta_2^{(j)} \sim \pi(\theta_2 | \theta_1^{(j)} \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)}, D_T) \text{ Using a Metropolis-Hastings step with proposal } q_2(\theta_2^* | \theta_2^{i-1})$$

$$\vdots$$

$$\theta_p^{(j)} \sim \pi(\theta_p | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j)}, D_T) \text{ Using a Metropolis-Hastings step with proposal } q_p(\theta_p^* | \theta_p^{i-1})$$
 3. Change counters j to $j+1$, and return to step 2.
-

Note that Algorithm 2.2 is, in fact, just a special case of Algorithm 2.3

If the full conditional distribution for the j^{th} component of θ is available to sample from directly, the resulting acceptance probability is one, it is for this reason that this method is also referred to as *Metropolis-within-Gibbs*. If the full conditional distributions are completely known and can be sampled from all components of θ , we obtain the Gibbs sampler.

2.6.8 Blocking

Blocking is a strategy used to improve the convergence (and indeed mixing) of the chain, although it comes at a higher computational cost. Given that the components of parameter of interest θ can take the form of scalars, vectors or matrices, it can be useful to block certain components together in multidimensional problems. Such a strategy is known as a block update and makes use of multivariate simulation techniques. As discussed in Gamerman and Lopes (2006) it is not the case that the larger the block update, the faster the convergence. Indeed for highly multidimensional problems a large block update is likely to be highly detrimental. Instead, components of θ should be blocked together such that the correlation

between the blocks is low. Any conditionally independent components should be updated on their own (a single-block update).

In general, for more than two decades, Markov Chain Monte Carlo (MCMC) has offers an appealing approach to handling some difficult types of analyses. However, epidemiologists are still largely unfamiliar with either the implementation of MCMC or with interpretation of the resultant output. However, when faced with complex models, which are increasingly common in epidemiologic inquiry, the researcher may find benefit in MCMC. MCMC procedures provide a clear benefit for Bayesian analysis. Unlike deterministic, a maximum-likelihood algorithm, MCMC is a stochastic procedure that repeatedly generates random samples that characterize the distribution of parameters of interest. This is distinct from commonly practiced asymptotic maximum-likelihood techniques, which are typically used to characterize the sampling distribution of an estimator.

2.6.9 Model convergence

Having considered MCMC as the sampling method adopted in the research work. The next step in a MCMC analysis is to assess model convergence. Unlike classical maximum likelihood techniques, MCMC convergence of likelihood estimators happens in the background not through the use of a pre-specified optimization algorithm (such as the Newton–Raphson algorithm).

However, judgments about the convergence of an MCMC procedure require the user to evaluate several diagnostics, the most common of which are visual summaries. The visual plots adopted in this research work which are routinely used to assess model convergence are:

(i) Trace plots

One way to see that the chain has converged is to see a well chain mixing or moving around the parameter space. Trace plots document the magnitude of the sample drawn (y-axis) at each iteration (x-axis) of the Markov Chain Monte Carlo (MCMC) procedure. Once the chain

has identified the stationary distribution of samples, the samples that are drawn will appear to have been randomly sampled from the same region of the y-axis.

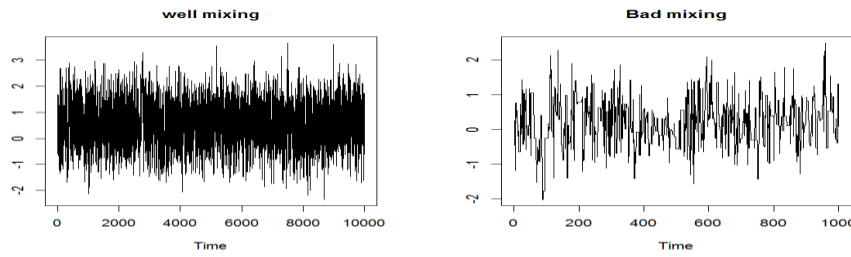


Figure 2.1 shows the well mixing trace plot (Stationarity) and bad mixing trace plot

(ii) Autocorrelation plots

These document the correlation (y-axis) of samples at each step of the chain with previous estimates of that same variable, lagged by some number of iterations (x-axis). Ideally, the autocorrelation declines rapidly, so that it is eventually possible to be confident that the samples from the stationary distribution can be thought of as random, and not reliant on initial values in the chain. If the autocorrelation fails to diminish, a pattern in the sample draws in the trace plot will become more discernible, suggesting that at the very least, the posterior distribution has not been well characterized by the drawn samples.

Autocorrelation as a function of lag are defined for the parameters θ as follows:

$$\rho_{lag} = \frac{\sum_i^{N-lag} (\theta_i - \bar{\theta})(\theta_{i+lag} - \bar{\theta})}{\sum_i^N (\theta_i - \bar{\theta})^2} \quad (2.61)$$

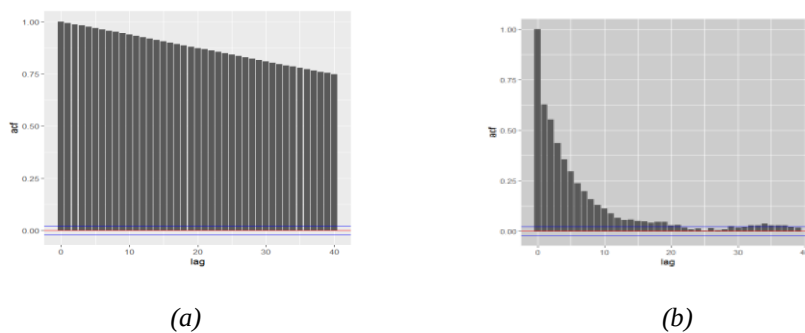


Figure 2.3 shows highest present of autocorrelation (a) and reducible or low autocorrelation (b) plot

(iii) Density plot

As its name implies, this plot is a summary of the sampled values that define the stationary distribution of values, which approximates the posterior distribution of interest. The peak of the density (the maximum a posterior (MAP) estimate) is the mode of the distribution, which is the value with the most support from the data and the specified prior. It should be noted that the density plot is not a formal means of diagnosing model convergence, as are the autocorrelation and trace plots. However, unexpected peaks or strange shapes in the posterior density can be a sign of poor model convergence.

2.6.10 Burn-in period

A ‘burn-in’, which is a grace period during which the Markov Chain wanders to the most probable region of the sample space (discard period). In our own work, we often specify a starting point of 10,000 iterations with a burn-in of 1000 iterations, and increase the values from these as necessary.

Chapter Three

Methodology

3.0 Epidemic Data

Most epidemic data discretely observed. The best way to consider the dynamic of such data can well be described by discretized-version of stochastic differential equation. Diffusion process is described as a solution to such differential equation (Stochastic process). Description of the stochastic processes in terms of diffusion approximation processes enables convenient simulation of the random course of the process, even for large populations. Such models turned out to be unsatisfactory in some applications. In the same way, the difficulty that underlies these are the general intractability of the transition density, making the likelihood function not to be in closed form. Special cases are the Poisson process; the paths of such *Lévy-driven* processes may experience jumps. In this chapter, diffusion approximations for such epidemic models (the standard SIR epidemic model, SEIR exposed epidemic model and SEIQIHR quarantine-isolation epidemic model) were derived in a more reasonable way, which solved the problem of unsatisfactory application and jumps experience occurred in the Levy processes.

Apart from providing estimates for the parameters of interest which are largely responsible for driving the dynamics of the diseases, models also have the ability to answer questions which refer to the progress of the disease based on the current state of the outbreak. It is therefore important to construct models for which we can draw inference regarding its unknown quantities so as to be able to give answers to important scientific questions regarding the underlying processes of an outbreak.

3.1 Derivation of Epidemic Models in Diffusion Approximation Processes

We begin with the general multidimensional stochastic differential equations (SDEs) model satisfies Itô process, where some state variables are unobserved. Such epidemic that, $D_T = (X_{t0}, X_{t1}, \dots, X_{tT})^T$ denote a d-dimensional state variable vector at time $t \geq 0$. The multidimensional Itô formula encounter here have a situation in higher dimensional case.

$$dX_i(t) = \alpha_i(t, X_i(t))dt + \sum_{j=1}^d \beta_{ij}(t, X_i(t))dW_j(t) \quad (3.0)$$

with $W_1, \dots, W_d$ independent Wiener processes (Brownian motion). An Itô process $\{X_t; t \geq 0\}$, which satisfy this stochastic process X_t is of the form:

$$X_t = X_0 + \int_0^t \alpha(X_s, s)ds + \int_0^t \sqrt{\beta(X_s, s)}dW_s \quad (3.1)$$

This can be expressed as a Riemann integral plus an Itô stochastic integral. The second integral is known as Itô integral which involves the unknown solution, i.e.

$$\int_0^t \sqrt{\beta(X_s, s)}dW_s = \sum_k^{m-1} \beta_k^{1/2} (W_{tk+1} - W_{tk}) \quad (3.2)$$

In many applications, the differential equations further depends on parameter of interest θ from a parameter space Θ . Leading to a parameterised family of d-dimensional Itô diffusion processes satisfied by a stochastic differential equation of the form:

$$dX_t = \alpha(X_t, \theta)dt + \sqrt{\beta(X_t, \theta)}dW_t, \quad X_0 = x_0. \quad (3.3)$$

where X_0 is the initial value at time t_0 and θ is an unknown parameter vector of length d.

In matrix notation we have the stochastic differential equations or Diffusion process of the form:

$$dX_t = \alpha(X_t, \theta)dt + \beta^{1/2}(X_t, \theta)dW_t$$

$$dX = \begin{pmatrix} dX_1 \\ \vdots \\ dX_d \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_1(X_{1t}, \theta) \\ \vdots \\ \alpha_d(X_{dt}, \theta) \end{pmatrix}}_{\text{Drift}} dt + \begin{pmatrix} \sqrt{\beta_{11}} & \cdots & \sqrt{\beta_{1d}} \\ \vdots & \ddots & \vdots \\ \sqrt{\beta_{d1}} & \cdots & \sqrt{\beta_{dd}} \end{pmatrix} \begin{pmatrix} W_{1t} \\ \vdots \\ W_{dt} \end{pmatrix} \quad (3.4)$$

Diffusion coefficient has assumption of symmetric square root of the form $\Sigma^{1/2} = \beta$ such that $\beta\beta^T = \Sigma$. If Σ is put in canonical form, such that $\Sigma = \beta^T D \beta$ and $\Sigma^{1/2} = \beta^T D^{1/2} \beta$.

The process $X_{(t)}$ is called a strong solution of the Stochastic Differential Equation, if for all time $t > 0$, $X(t)$ is a function $f(t, (W_{(s)}, s < t))$ of the given wiener process W_t , integral in (3.1) exist and the integral equation satisfied the following conditions.

(1) A Lipschitz condition in x and y

$$\exists k, \forall x \in \mathcal{R}^n, \forall y \in \mathcal{R}^n, \forall t \geq 0 \quad (3.5)$$

$$\|\alpha(t, x) - \alpha(t, y)\| + \|\beta(t, x) - \beta(t, y)\| \leq k\|x - y\|$$

(2) A linear growth condition

$$\exists k, \forall x \in \mathcal{R}^n, \forall t \geq 0 \quad (3.6)$$

$$\|\alpha(t, x)\| + \|\beta(t, x)\| \leq k(1 + \|x\|)$$

(3) x_0 is a constant.

Then there exists a unique strong solution X to be stochastic differential equation (3.4) with continuous trajectories and there exists a constant C such that.

$$\mathbb{E}(\|X_t\|^2) \leq C e^{ct} (1 + \|x_0\|^2). \quad (3.7)$$

3.1.1 Derivation of Stochastic SIR Epidemic Model

First, consider the classical SIR epidemic model (1.1) in chapter one. The two possibilities changes fraction number of susceptible and infectious in the vector $X_t = (s_t, i_t)^T$ for a small

change in time interval Δt : assuming at most one change can occur. It is possible to express fraction number of recovery component r_t in term of $r_t = n - s_t - i_t$, because, $n = s_t + i_t + r_t$

Where, s_t stands for fraction number of susceptible (those who can contract the disease), i_t stands for fraction number of infectious (those who have the disease and can infect others) and the parameters of interest: transmission rate β and infectious rate γ that characterize the propagation of the disease and can also be used to served as control in order to stop the epidemic. The transition probability of SIR epidemic model is given as:

$$\pi(s+i, t+j) = \begin{cases} \frac{\beta s_t i_t}{N} \\ \gamma i_t \\ 0 \end{cases} \quad (3.8)$$

Suppose, $X_t = (s_t, i_t)^T$ represent $(X_1, X_2)^T$. The compartment changes in small time period Δt are:

$$\begin{cases} \Delta X_1 = (1 \ 0)' \text{ with } \frac{\beta X_1 X_2}{N}, \\ \Delta X_2 = (0 \ 1)' \text{ with } \gamma X_2 \end{cases}$$

Then, to derive the infinitesimal means, we apply Allen (2007) approach i.e.

$$\begin{aligned} E(\Delta X) &= \sum_{i=1}^n \pi(\Delta X_i) \\ E[\Delta X] &= \sum_{i=1}^2 \pi_i(\Delta X_i) \\ &= \frac{\beta X_1 X_2}{N} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \gamma X_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{-\beta X_1 X_2}{N} \\ \frac{\beta X_1 X_2}{N} - \gamma X_2 \end{pmatrix} \Delta t \end{aligned} \quad (3.9)$$

Likewise, the infinitesimal covariance matrix $\mathbb{E}[(\Delta X)(\Delta X)^T]$ of the transition probabilities:

$$\begin{aligned}
E[(\Delta X)(\Delta X)^T] &= \sum_{i=1}^n \pi(\Delta X)_i (\Delta X)_i^T \\
E[(\Delta X)(\Delta X)^T] &= \sum_{i=1}^2 \pi_i (\Delta X)_i (\Delta X)_i^T = \pi_1 (\Delta X)_1 (\Delta X)_1^T + \pi_2 (\Delta X)_2 (\Delta X)_2^T \\
&= \frac{\beta X_1 X_2}{N} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \gamma X_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} \\
&= \frac{\beta X_1 X_2}{N} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \gamma X_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\beta X_1 X_2}{N} & 0 \\ 0 & \gamma X_2 \end{pmatrix}
\end{aligned} \tag{3.10}$$

Since the diffusion coefficient matrix has assumption of symmetric square root of the form $\Sigma^{1/2} = \beta$, such that, $\beta\beta^T = \Sigma$. Hence, Σ is put in canonical form, such that $\Sigma = \beta^T D \beta$ and $\Sigma^{1/2} = \beta^T D^{1/2} \beta$.

Following the assumption of invertible symmetric positive definite matrix, apply the Cholesky Decomposition approach to determine the newly diffusion matrix, arrived at Covariance matrices of 2×2 .

$$= \begin{pmatrix} \frac{\beta X_1 X_2}{N} & 0 \\ -\frac{\beta X_1 X_2}{N} & \gamma X_2 \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} \tag{3.11}$$

3.1.2 Cholesky Decomposition

The Cholesky decomposition of a symmetric and positive semi-definite $n \times n$ matrix say Σ is $\Sigma = \beta\beta'$ with β lower triangular. We have $\Sigma \approx LDU$ with the diagonal elements of D non-negative since Σ is positive semi-definite.

Algorithm 3.1 Cholesky Decomposition

The Cholesky factor, a lower triangular matrix say β such that $\Sigma = \beta\beta'$

1. Initialise $B_1 = \sqrt{a_{11}}$, For $j = 1, \dots, n$

$$B_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} B_{jk}^2 \right)^{1/2}, \quad B_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} B_{ik} B_{jk} \right) / B_{jj}$$

2. For $k = 2, \dots, n$

$$(2.1) \text{ Solve } B_{k-1} l_k = a_k$$

$$(2.2) \quad l_{kk} = \sqrt{a_{kk} - l_k' l_k}$$

$$(2.3) \quad B_k = \begin{pmatrix} B_{k-1} & 0 \\ l_k' & l_{kk} \end{pmatrix}$$

Where B_{k-1} denotes the $k-1 \times k-1$ upper left corner of B ;

a_k is the first $k-1$ entries in column k of Σ ;

l_k is the first $k-1$ entries in column k of B^1 ;

a_{kk} and l_{kk} are the kk entries of Σ and B respectively.

Equation (3.10) provide the stochastic differential equation that involves the square root of a symmetric positive definite matrix Σ such that $\beta = \Sigma^{1/2}$.

By substituting infinitesimal means from equation (3.9) and infinitesimal covariance in equation (3.11) above, we arrived at newly epidemic model (stochastic model) of the form:

$$\begin{pmatrix} dX_1 \\ dX_2 \end{pmatrix} = \begin{pmatrix} -\frac{\beta X_1 X_2}{N} \\ \frac{\beta X_1 X_2}{N} - \gamma X_2 \end{pmatrix} \Delta t + \begin{pmatrix} \frac{\beta X_1 X_2}{N} & 0 \\ -\frac{\beta X_1 X_2}{N} & \gamma X_2 \end{pmatrix}^{1/2} \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix}$$

Replace $X_t = (s_t, i_t)^T$ by $(X_1, X_2)^T$,

$$\begin{pmatrix} ds_t \\ di_t \end{pmatrix} = \begin{pmatrix} -\beta N s_t i_t \\ \beta N s_t i_t - \gamma i_t \end{pmatrix} dt + \begin{pmatrix} -\frac{1}{N} \sqrt{\beta N s_t i_t} & 0 \\ \frac{1}{N} \sqrt{\beta N s_t i_t} & -\frac{1}{N} \sqrt{\gamma N i_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix} \quad (3.12)$$

3.1.3 Derivation of Stochastic SEIR Epidemic Model

Similarly, consider SEIR epidemic model in equation (1.2) in the chapter one. Assuming at most, one change can occur, then, the three possibilities in changes in the vector $X_t = (s_t, e_t, i_t)'$ for a small time interval Δt . It is possible to express r_t in other variables $r_t = n - s_t - e_t - i_t$ because, $n = s_t + e_t + i_t + r_t$.

where s_t , e_t , and i_t are random variables representing the fraction number of susceptible, exposed/latent, and infective individuals at time t , and dt is a vanishingly small time period such that only one event can occur in $[t, t+dt]$. Here β is the transmission parameter, γ^{-1} is the mean incubation period and α^{-1} is the mean infectious period.

The stochastic SEIR epidemic model for infectious disease is based on three dimensional processes, with the first component S_t represents the susceptible individuals, the second component E_t represents the exposed individuals and the third component I_t represents the infection individual infected at time t .

If $X_t = (s_t, e_t, i_t)'$ represent $(X_1, X_2, X_3)'$. The transition probabilities and compartment changes in small time period is of the form:

$$p(s+k, i+j, l+m) = \begin{cases} \frac{\beta X_1 X_3}{N} \Delta t & (k, j, m) = (-1, 1, 0) \\ \gamma X_2 \Delta t & (k, j, m) = (0, -1, 1) \\ \alpha X_3 \Delta t & (k, j, m) = (0, 0, -1) \\ 1 - \left(\frac{\beta X_1 X_3}{N} + \gamma X_2 + \alpha X_3 \right) \Delta t & (k, j, m) = (0, 0, 0) \\ 0 & \text{otherwise} \end{cases} \quad (3.13)$$

Hence, the infinitesimal means can be derived thus:

$$E(\Delta \bar{X}) = \sum_{i=1}^3 p_i(\Delta \bar{X}_i) = p_1(\Delta \bar{X}_1) + p_2(\Delta \bar{X}_2) + p_3(\Delta \bar{X}_3)$$

$$\begin{aligned}
&= p_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + p_3 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{-\beta X_1 X_3}{N} \\ \frac{\beta X_1 X_3}{N} \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ -\gamma X_2 \\ \gamma X_2 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ 0 \\ -\alpha X_3 \end{pmatrix} \Delta t \\
&= \begin{pmatrix} \frac{-\beta X_1 X_3}{N} \\ \frac{\beta X_1 X_3}{N} - \gamma X_2 \\ \gamma X_2 - \alpha X_3 \end{pmatrix} \Delta t \tag{3.14}
\end{aligned}$$

The infinitesimal covariance matrix can be derived from

$$\begin{aligned}
E[(\Delta X)(\Delta X)^T] &= \sum_{i=1}^3 p_i (\Delta X)_i (\Delta X)_i^T = p_1 (\Delta X)_1 (\Delta X)_1^T + p_2 (\Delta X)_2 (\Delta X)_2^T + p_3 (\Delta X)_3 (\Delta X)_3^T \\
&= p_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 1 \end{pmatrix} + p_3 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \end{pmatrix} \\
&= p_1 \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} + p_3 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

Equation (3.4) provide the stochastic differential equation that involves the square root of a symmetric positive definite matrix Σ such that $\beta = \Sigma^{1/2}$.

Similar to the Cholesky Decomposition approach above, we determine the newly diffusion matrix with wiener process $W_{(t)}$ $n \times 1$ vector. If Σ is put in canonical form, $\Sigma = \beta^T D \beta$ and $V^{1/2} = \beta^T D^{1/2} \beta$ such that the square root of $\Sigma = \beta$ i.e. $\beta \beta^T = \Sigma$. The matrix β is $n \times m$ and $W_{(t)}$ is.

$$= \begin{pmatrix} -\frac{\beta X_1 X_3}{N} & 0 & 0 \\ \frac{\beta X_1 X_3}{N} & -\gamma X_2 & 0 \\ 0 & \gamma X_2 & \gamma X_2 + \alpha X_3 \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix} \quad (3.15)$$

Replace $X_t = (s_t, e_t, i_t)'$ by $(X_1, X_2, X_3)'$ we arrived at SEIR diffusion epidemic model of the form:

$$\begin{pmatrix} ds_t \\ de_t \\ di_t \end{pmatrix} = \begin{pmatrix} -\beta N s_t i_t \\ \beta N s_t i_t - \gamma e_t \\ \gamma e_t - \alpha i_t \end{pmatrix} dt + \begin{pmatrix} -\frac{1}{N} \sqrt{\beta N s_t i_t} & 0 & 0 \\ \frac{1}{N} \sqrt{\beta N s_t i_t} & -\frac{1}{N} \sqrt{\gamma N e_t} & 0 \\ 0 & \frac{1}{N} \sqrt{\gamma N e_t} & -\frac{1}{N} \sqrt{\alpha N i_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix} \quad (3.16)$$

3.1.4 Derivation of Stochastic SEQIHR Epidemic Model

Another complex epidemic model that is very applicable introduced in chapter one is the quarantine-isolation model in equation (1.3 and 1.4). Also, assuming at most one change can occur. Then the five possibilities in changes in the process are $X_t = (s_t, e_t, q_t, i_t, h_t)'$ for a small time interval Δt . It is possible to express r_t in other variables $r_t = n - s_t - e_t - q_t - i_t - h_t$ because $n = s_t + e_t + q_t + i_t + h_t + r_t$.

Likewise, if we let $X_t = (s_t, e_t, q_t, i_t, h_t)'$ represent $(X_1, X_2, X_3, X_4, X_5)'$. The transition probabilities and compartment changes in small time period is of the form:

$$p(s+k, i+j, l+m, p+q, t+v) = \begin{cases} \lambda_t X_{1t} \Delta t & (k, j, m, q, v) = (-1, 1, 0, 0, 0) \\ \gamma_e X_{2t} \Delta t & (k, j, m, q, v) = (0, -1, 0, 1, 0) \\ \theta X_{2t} \Delta t & (k, j, m, q, v) = (0, -1, 1, 0, 0) \\ \gamma_Q X_{3t} \Delta t & (k, j, m, q, v) = (0, 0, -1, 0, 1) \\ \sigma X_{4t} \Delta t & (k, j, m, q, v) = (0, 0, 0, -1, 1) \\ \alpha_I X_{4t} \Delta t & (k, j, m, q, v) = (0, 0, 0, -1, 0) \\ \alpha_H X_{5t} \Delta t & (k, j, m, q, v) = (0, 0, 0, 0, -1) \\ ,0 & \text{Otherwise} \end{cases} \quad (3.17)$$

And the infinitesimal means can be derived in this form:

$$\begin{aligned} E(\Delta \vec{X}) &= \sum_{i=1}^7 p_i(\Delta \vec{X}_i) = p_1(\Delta \vec{X}_1) + p_2(\Delta \vec{X}_2) + p_3(\Delta \vec{X}_3) + p_4(\Delta \vec{X}_4) + p_5(\Delta \vec{X}_5) + p_6(\Delta \vec{X}_6) + p_7(\Delta \vec{X}_7) \\ &= p_1 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + p_3 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + p_4 \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} + p_5 \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} + p_6 \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} + p_7 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} -\lambda_t X_1 \\ \lambda_t X_1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ -\gamma_e X_2 \\ 0 \\ \gamma_e X_2 \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ \theta X_2 \\ \theta X_2 \\ 0 \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ 0 \\ -\gamma_Q X_3 \\ 0 \\ \gamma_Q X_3 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\sigma X_4 \\ \sigma X_4 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\alpha_I X_4 \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\alpha_H X_5 \end{pmatrix} \Delta t \\ &= \begin{pmatrix} -\lambda_t X_1 \\ \lambda_t X_1 - \gamma_e X_2 + \theta X_2 \\ \theta X_2 + \gamma_Q X_3 \\ \gamma_e X_2 - \sigma X_4 - \alpha_I X_4 \\ \gamma_Q X_3 + \sigma X_4 - \alpha_H X_5 \end{pmatrix} \Delta t = \begin{pmatrix} -\lambda_t S_{(t)} \\ \lambda_t S_{(t)} - \gamma_e E_{(t)} + \theta E_{(t)} \\ \theta E_{(t)} + \gamma_Q Q_{(t)} \\ \gamma_e E_{(t)} - \sigma I_{(t)} - \alpha_I I_{(t)} \\ \gamma_Q Q_{(t)} + \sigma I_{(t)} - \alpha_H H_{(t)} \end{pmatrix} \Delta t \end{aligned} \quad (3.18)$$

Similarly, the infinitesimal covariance matrix for the diffusion coefficient matrix is thus:

$$E[(\Delta X)(\Delta X)^T] = \sum_{i=1}^7 p_i(\Delta X)_i (\Delta X)_i^T$$

$$= p_1(\Delta X)_1(\Delta X)_1^T + p_2(\Delta X)_2(\Delta X)_2^T + p_3(\Delta X)_3(\Delta X)_3^T + p_4(\Delta X)_4(\Delta X)_4^T + p_5(\Delta X)_5(\Delta X)_5^T + p_6(\Delta X)_6(\Delta X)_6^T + p_7(\Delta X)_7(\Delta X)_7^T$$

$$= p_1 \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + p_3 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + p_4 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix} + p_5 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} + p_6 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + p_7 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} P_1 & -P_1 & 0 & 0 & 0 \\ -P_1 & P_1 + P_2 + P_3 & P_3 & -P_2 & 0 \\ 0 & P_3 & P_3 + P_4 & 0 & -P_4 \\ 0 & -P_2 & 0 & P_2 + P_5 + P_6 & -P_5 \\ 0 & 0 & -P_4 & -P_5 & P_4 + P_5 + P_7 \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_t X_1 & \lambda_t X_1 & 0 & 0 & 0 \\ -\lambda_t X_1 & \lambda_t X_1 + (\gamma_Q + \theta) X_2 & \theta X_2 & -\gamma_e X_2 & 0 \\ 0 & \theta X_2 & \theta X_2 + \gamma_e X_3 & 0 & -\gamma_Q X_3 \\ 0 & -\gamma_e X_2 & 0 & \gamma_e X_2 + (\sigma + \alpha_I) X_4 & -\sigma X_4 \\ 0 & 0 & -\gamma_Q X_3 & -\sigma X_4 & \gamma_Q X_3 + \sigma X_4 + \alpha_H X_5 \end{pmatrix}$$

Equation (3.4) provide the stochastic differential equation that involves the square root of a symmetric positive definite matrix Σ such that $\beta = \Sigma^{1/2}$.

Similar to the Cholesky Decomposition approach above, we determine the newly diffusion matrix with wiener process $W_{(t)}$ $n \times 1$ vector. If Σ is put in canonical form, $\Sigma = \beta^T D \beta$ and $V^{1/2} = \beta^T D^{1/2} \beta$ such that the square root of $\Sigma = \beta$ i.e. $\beta \beta^T = \Sigma$. The matrix β is $n \times m$ and $W_{(t)}$ is.

$$= \begin{pmatrix} \lambda_t X_1 & \lambda_t X_1 & 0 & 0 & 0 \\ -\lambda_t X_1 & \lambda_t X_1 + (\gamma_Q + \theta) X_2 & \theta X_2 & -\gamma_e X_2 & 0 \\ 0 & \theta X_2 & \theta X_2 + \gamma_e X_3 & 0 & -\gamma_Q X_3 \\ 0 & -\gamma_e X_2 & 0 & \gamma_e X_2 + (\sigma + \alpha_I) X_4 & -\sigma X_4 \\ 0 & 0 & -\gamma_Q X_3 & -\sigma X_4 & \gamma_Q X_3 + \sigma X_4 + \alpha_H X_5 \end{pmatrix}^{1/2} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \\ dW_4 \\ dW_5 \end{pmatrix} \quad (3.19)$$

Replace $X_t = (s_t, e_t, q_t, i_t, h_t)'$ by $(X_1, X_2, X_3, X_4, X_5)'$ we arrived at SEQIHR diffusion epidemic model of the form:

$$\begin{pmatrix} ds_t \\ de_t \\ dq_t \\ dh_t \\ di_t \end{pmatrix} = \begin{pmatrix} \psi r_t - \lambda s_t \\ \lambda s_t - (\gamma_e + \theta) e_t \\ \theta e_t - \gamma_q q_t \\ \gamma_e e_t - (\sigma + \alpha_i) i_t \\ \gamma_q q_t - \sigma i_t - \alpha_h h_t \end{pmatrix} dt + \begin{pmatrix} -\sqrt{\lambda s_t} & 0 & 0 & 0 & 0 \\ \sqrt{\lambda s_t} & -\sqrt{(\gamma_e + \theta) e_t} & 0 & 0 & 0 \\ 0 & \sqrt{\theta e_t} & -\sqrt{\gamma_q q_t} & 0 & 0 \\ 0 & \sqrt{\gamma_e e_t} & 0 & \sqrt{(\sigma + \alpha_i) i_t} & 0 \\ 0 & 0 & \sqrt{\gamma_q q_t} & \sqrt{\sigma i_t} & -\sqrt{\alpha_h h_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \\ dW_{4t} \\ dW_{5t} \end{pmatrix}$$

Table 3.1 Summary of the infinitesimal means and covariance matrix of the Epidemic models under study.

Model	Infinitesimal means	Infinitesimal covariance matrix
SIR	$\begin{pmatrix} -\beta N s_t i_t \\ \beta N s_t i_t - \gamma i_t \end{pmatrix} dt$	$\begin{pmatrix} -\frac{1}{N} \sqrt{\beta N s_t i_t} & 0 \\ \frac{1}{N} \sqrt{\beta N s_t i_t} & -\frac{1}{N} \sqrt{\gamma i_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix}$
SEIR	$\begin{pmatrix} -\beta N s_t i_t \\ \beta N s_t i_t - \gamma e_t \\ \gamma e_t - \alpha i_t \end{pmatrix} dt$	$\begin{pmatrix} -\frac{1}{N} \sqrt{\beta N s_t i_t} & 0 & 0 \\ \frac{1}{N} \sqrt{\beta N s_t i_t} & -\frac{1}{N} \sqrt{\gamma e_t} & 0 \\ 0 & \frac{1}{N} \sqrt{\gamma e_t} & -\frac{1}{N} \sqrt{\alpha i_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \end{pmatrix}$
SEQIHR	$\begin{pmatrix} \psi r_t - \lambda s_t \\ \lambda s_t - (\gamma_e + \theta) e_t \\ \theta e_t - \gamma_q q_t \\ \gamma_e e_t - (\sigma + \alpha_i) i_t \\ \gamma_q q_t - \sigma i_t - \alpha_h h_t \end{pmatrix} dt$	$\begin{pmatrix} -\sqrt{\lambda s_t} & 0 & 0 & 0 & 0 \\ \sqrt{\lambda s_t} & -\sqrt{(\gamma_e + \theta) e_t} & 0 & 0 & 0 \\ 0 & \sqrt{\theta e_t} & -\sqrt{\gamma_q q_t} & 0 & 0 \\ 0 & \sqrt{\gamma_e e_t} & 0 & \sqrt{(\sigma + \alpha_i) i_t} & 0 \\ 0 & 0 & \sqrt{\gamma_q q_t} & \sqrt{\sigma i_t} & -\sqrt{\alpha_h h_t} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \\ dW_{3t} \\ dW_{4t} \\ dW_{5t} \end{pmatrix}$

3.2 Model Parameters Estimation

In practice, the paths of a diffusion process cannot be observed continuously in time, due to the extremely wiggly trajectories. Estimation of the parameters of interest θ will hence be based on observed states $\mathbf{X} = \{ \mathbf{x}_1, \dots, \mathbf{x}_n \}$ at discrete times $t_1 < \dots < t_n$, as well as on the starting value $\mathbf{x}_0$ at time $t_0 = 0$. The focus of interest for these discrete-time observations now lies on the transition density $\pi(s, \mathbf{x}, t, \mathbf{y})$ with respect to Lebesgue measure, which is defined

$$\text{by} \quad \pi(X_t \in A | X_s = \mathbf{x}) = \int_A \pi(s, \mathbf{x}, t, \mathbf{y}) d\mathbf{y} \quad (3.20)$$

for all measurable sets A , $t > s \geq 0$ and $\mathbf{x}, \mathbf{y} \in X$.

As diffusion processes are Markovian, the log-likelihood function of θ with respect to Lebesgue measure would now becomes

$$\ell_n(\theta) = \sum_{k=0}^{n-1} \log \pi(\Delta t_k; x_k, x_{k+1}) \quad (3.21)$$

$$\text{with } \Delta t_k = t_{k+1} - t_k$$

Under regularity conditions, Dacunha-Castelle and Florens-Zmirou (1986) prove consistency, asymptotic normality and asymptotic efficiency of the corresponding maximum likelihood estimator as n tends to infinity ($n \rightarrow \infty$).

However, the transition probability and hence the log-likelihood function are intractable unless the diffusion process is analytically explicitly solvable, which is rarely the case.

Allen,(2007) suggested a very popular ideas of discretized the diffusion models using Euler-Maruyama of section (2.3.4) method of numerical solutions of discretisation approach on continuous times processes. Now the diffusion model (3.4) becomes

$$X_{t+1} = X_t + \alpha(X_t, \theta)\Delta t + \sqrt{\beta(X_t, \theta)}\Delta W_t, \quad X_0 = x_0. \quad (3.22)$$

$$\Delta W_t = W_{t_{i+1}} - W_{t_i} \sim N_d(0, I\Delta t)$$

For a small time increment Δt , (3.22) converged to Stochastic differential equation (3.4).

Here, we assumed non-equidistant observations D_T at discrete times, resulting in a dataset

$$D_T = (X_{t_0}, X_{t_1}, \dots, X_{t_T})' \quad (3.23)$$

were $t_0 = 0, t_T = t_n = T$ and $t_{k+1} - t_k = \Delta$, for $k = 0, \dots, T - 1$, aiming to make estimate of the parameter vector θ governing the model (3.4), the likelihood function would therefore be considered for the discretisation (3.22).

3.2.1 Likelihood Function

The likelihood function of the numerical solution of diffusion process (3.22) is of the form:

$$L(\theta | D_T) = \prod_{k=0}^{T-1} \pi(x_{t_{k+1}} | x_{t_k}, \theta) \quad (3.24)$$

were, $\pi(x_{t_{k+1}} | x_{t_k}, \theta)$ denotes the transition density of the process X_t from $X_{t_k} = x_{t_k}$ to $X_{t_{k+1}} = x_{t_{k+1}}$. We also assumed that the transition density $\pi(x_{t_{k+1}} | x_{t_k}, \theta)$ is not explicitly

known and this is indeed another problem of parameters estimate under study. Dealing with such process poses the likelihood functions which are not available in closed form. As earlier explained in the previous chapter; the maximum likelihood estimation (MLE) would be intractable. We therefore considered Bayesian inferential method of estimation.

3.2.2 Bayesian Approach

In the Bayesian approach, the hidden aspects of the discretely-observed epidemic models are treated as additional unknown parameters. The joint posterior density of theses parameters and the path of the diffusion processes given the observations are then investigated.

We based our estimation steps in exploring the posterior distribution of parameters of interest θ conditionally on observations of the process $\pi(\theta|D_T)$.

The idea behind this is that the likelihood (3.24) and prior $\pi(\theta)$ are combined using Bayes' theorem to compute the posterior distribution.

$$\pi(\theta | D_T) = \frac{\pi(\theta) \times L(\theta | D_T)}{\int \pi(\theta) \times L(\theta | D_T) d\theta} \quad (3.25)$$

Posterior density in equation (3.25) reflects our updated beliefs about θ after observing data D_T . From the equation (3.25), the denominator can be regarded as a constant of proportionality (as it does not depend upon θ).

The joint posterior density for parameters and the path diffusion processes now become

$$\begin{aligned} \pi(\theta, x_\tau | D_T) &\propto \pi(\theta) \pi(x_\tau | \theta) \pi(D_T | x_T, \theta) \\ \pi(\theta, \mathbf{x}_\tau | D_T) &= \pi(\theta) \pi(\mathbf{x}_\tau | \theta) \prod_{i=0}^{T-1} \prod_{k=0}^{m-1} \pi(\Delta \tau; x_{i,k+1} | x_{i,k}, \theta) \end{aligned} \quad (3.26)$$

Where $\pi(\theta)$ denotes independence prior distribution for θ and the $\prod \pi(x_{tk+1} | x_{tk}, \theta)$ is the Likelihood function (3.24). Thus, the posterior is proportional to the product of the prior $\pi(\theta)$ and the likelihood $L(\theta|D_T)$.

The posterior distribution above can applicable to the data under study after adopted Simulated maximum likelihood estimation suggested by Pedersen (1995) as method of

parameter estimation from discretely observed diffusion process, here we approximate the true transition density π_0 of the diffusion process by adopted the Euler scheme or any other higher-order numerical schemes. This is eligible only if inter-observation times of the observed data $\mathbf{x}$ are small. Since such requirement is usually not fulfilled in many applications, additional data $\mathbf{X}^* = \mathbf{x}^*$ (or data augmentation) at intermediate time points is introduced. To this end, a Markov chain Monte Carlo (MCMC) approach is employed to construct a Markov chain $\{\theta^{(i)}, \mathbf{X}^*\}_{i=1, \dots, L}$ of length L whose elements are samples from the joint posterior density $\pi(\theta, \mathbf{X}^* | \mathbf{D}_T)$ of the parameter θ and the imputed data $\mathbf{X}^*$ conditioned on the sample path observations $\mathbf{D}_T$. The Markov chain $\{\theta^{(i)}\}_{i=1, \dots, L}$ is then regarded as a draw from the marginal density $\pi(\theta | \mathbf{D}_T)$. The imputed or augmented data $\{\mathbf{X}^{*(i)}\}_{i=1, \dots, L}$ is a convenient by-product of the estimation procedure known as sample paths of a diffusion process.

3.2.3 Data Augmentations (DA)

Additional data ($X_{\tau}^* = x_{\tau}^*$) introduced at intermediate time points is refers to as augmented data. The introduction of this additional data reduces the distance between every two consecutive time points. This would enable the approximation of the Euler transition density converges to the true diffusion process. Bayesian data-augmentation approach via MCMC sampling allows more convenient and efficient parameter estimation for discretely observed diffusion process (Aliu, et al., 2017). The time-step for additional data at intermediate time points now need special consideration so that the likelihood can be well approximated. The time-step interval $[0, T]$ is partitioned into M sub-intervals via grid points $0 = \tau_0 < \tau_1 < \dots < \tau_M = T$ (each sub-interval has length $\Delta\tau = T/M$) such that the resulting partition is sufficiently fine for some discrete approximation of the diffusion X^* to be sufficiently accurate. This unobserved data points of the process X^* are treated as missing data.

The resulting jointly posterior distribution now becomes:

$$\pi(\theta, x^* | D_T) \propto \pi(\theta) \prod_{i=0}^{T-1} \prod_{k=0}^{M-1} \pi(\Delta \tau; x_{i,k+1} | x_{i,k}, \theta) \quad (3.27)$$

where $\Delta \tau = \tau_{k+1} - \tau_k$ with $x_{i,0} = x_i$, $x_{i,M} = x_{i+1}$, $x^* = (x_0, \dots, x_{T-1})$, also where $x_i = (x_{i,1}, \dots, x_{i,M-1})$ is a discrete-time skeleton of (3.27) between x_i and x_{i+1} , and $\pi(\Delta \tau, x_{i,k+1} | x_{i,k}, \theta)$ is some approximation of the transition density. By adopting naive maximum likelihood (2.4.1) in chapter two, the probability density $\pi_\theta(\Delta t_k; \mathbf{x}_k, \mathbf{x}_{k+1})$ can thus be approximated by a Gaussian density with means and covariance according to (2.8) and (2.9).

$$\pi(\Delta \tau; x_{i,k+1} | x_{i,k}, \theta) = \phi(x_{i,k+1}; x_{i,k} + \alpha(x_{i,k}, \theta) \Delta \tau, \beta(x_{i,k}, \theta) \Delta \tau) \quad (3.28)$$

where $\phi(\cdot; \alpha, \beta)$ denotes the multivariate Gaussian distribution with infinitesimal means vector $\alpha(X_\tau, \theta)$ and infinitesimal covariance matrix $\beta(X_\tau, \theta)$. Therefore, the resulting algorithm proceeds by alternating between simulation of θ conditional on the augmented data x^* (as parameter update), and simulation of the missing data blocks conditional on θ (as path update). Updating the augmented data x_k between x_k and x_{k+1} , for $k = 0, \dots, M-1$, requires generating the intermediate points according to their conditional distribution given the end, start points or both.

A description for such algorithm can be found in Algorithm 3.2:

Algorithm 3.2 Traditional MCMC Bayesian Augmentation Approach

1. Initialise all unknowns. Use linear interpolation to initialise the X_k . Set counter (c) := 1
 2. Draw $X_{(c)} \sim \pi(\cdot | \theta_{c-1}, D_T)$.
 3. Draw $\theta_{(c)} \sim \pi(\cdot | X_{(c)})$.
 4. If the desired number of simulations has been performed then stop, otherwise, set Counter := $c + 1$ and return to step 2.
-

The full conditionals required in steps 2 and 3 are proportional to $\pi(\theta, X | D_T)$ in equation (3.28). Steps 2 and 3 are performed via a Metropolis-Hastings (MH).

Attention is now turned to the task of performing step 2, that is, to update the augment data conditional on the current parameter values and the observed data.

Various diffusion bridges mechanism to generate discrete-time realisations of x_k^* conditional on x_0 and x_T over the interval $[0, T]$ have been proposed in the literature. The best diffusion bridge mechanism to generate discrete times trajectories for non-linear multidimensional diffusion processes still pose great challenges in the literature. This and many are the objectives of this research work.

3.3 Path Update of Diffusion process

By considering the task of generating discrete-time realisations of a multidimensional diffusion process satisfying a stochastic epidemic models parameters of interest θ that conditioned on observation taken at a fixed future time point X_T . Such realisations are typically termed diffusion bridges.

Initially, observations x_{t_k} are available at evenly spaced times $t_0, t_1, \dots, t_n$ with intervals of length $\Delta t^* = t_{k+1} - t_k$. As it is typically unrealistic to assume that Δ^* is sufficiently small to be used as a time step in the likelihood function (3.24), we put $\Delta t = \Delta t^*/m$ for some positive integer $m > 1$. Then, by choosing m to be sufficiently large ensures that the discretization bias is arbitrarily small, but also introduces $m - 1$ missing values in between every pair of observations, which must be integrated out of the problem. In order to provide a framework for dealing with these missing values, the entire time interval (t_0, t_n) is divided into $mT + 1$ equidistant points $0 = t_0 < t_1 < \dots < t_n = T$ (where $n = mT$) such that X_t is observed at times $t_0, t_m, t_{2m}, \dots, t_n$. Altogether there are $(m-1)T$ missing values which are substituted with simulations X_{t_k} . Stacking all augmented data (both missing and observed) in matrix form gives a skeleton path,

$$D_T = (x_{t_0}, x_{t_1}, \dots, x_{t_{m-1}}, x_{t_m}, x_{t_{m+1}}, \dots, x_{t_{m-1}}, x_{t_n}) \quad (3.29)$$

Simple diagrams that illustrate one-dimensional path that generate realisation discrete data point is illustrated in figure 3.1 below.

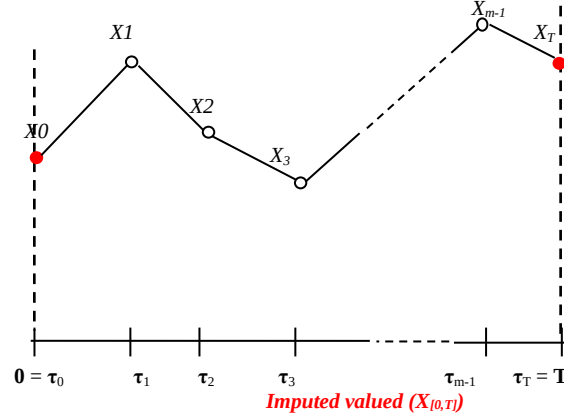


Figure 3.1 Show an illustration of one-dimensional path segment consisting of discrete data points $x_0, \dots, x_m$ at times $t_0, \dots, t_m$. The observed data x_0, x_m is labelled with red circle, the imputed data $x_1, \dots, x_{m-1}$ with circles.

Within this framework, we have data $D_T = (x_{t0}, x_{tm}, \dots, x_m)$. Hence, by adopting a fully Bayesian approach, we formulate the joint posterior for parameters and missing data as in (3.20) or in a simpler way as

$$\pi(\theta, X | D_T) \propto \pi(\theta) \prod_{k=0}^{n-1} \pi(X_{tk+1} | X_{tk}, \theta) \quad (3.30)$$

where $\pi(\theta)$ is the prior density for parameters and,

$\pi(\cdot | X_{tk}, \theta)$ is the Euler density given by equation (3.28) above.

Here, we generalised the density (3.27) into several observation times gives us the likelihood of the entire discretely observed diffusion path as

$$\prod_{i=0}^{M-1} \prod_{k=0}^{m_i-1} \pi(t_{i,k+1} - t_{i,k}, X_{i,k}, X_{i,k+1}) \approx \prod_{i=0}^{M-1} \prod_{k=0}^{m_i-1} \pi^{Euler}(X_{i,k+1} | X_{i,k}, \theta) \quad (3.31)$$

The likelihood now is applicable to any d-dimensional concepts, and is to be deployed accordingly in all occurring acceptance probabilities of the Markov Chain under study.

The two sampling strategies that already existing in the literature for the inherent problem associated with their mixing properties are single site and block updating.

3.3.1 Generalisation to Several Observation Times

The augmented-data (imputed value) considered are easily extendable to the general case (d-dimensional) $d \geq 2$, where more observations times are available than just the starting and the end point of a sample path of a diffusion process.

Suppose there are M complete observations M complete observations $\mathbf{x}_{\tau_1}, \dots, \mathbf{x}_{\tau_M}$ at times

$$0 = \tau_0 < \tau_1 < \dots < \tau_M = T \quad (3.32)$$

in addition to the initial value $\mathbf{x}_{\tau_0}$. For $i = 0, \dots, M-1$, divide each inter-observation interval $[\tau_i, \tau_{i+1}]$ into m_i sufficiently small subintervals with boundaries

$$\tau_i = t_{i,0} < t_{i,1} < \dots < t_{i,m_i-1} < t_{i,m_i} = \tau_{i+1} \quad (3.33)$$

Impute auxiliary data at the newly introduced time points. In the following, observations on $[\tau_i, \tau_{i+1}]$ are labelled $X_{i,0} = \mathbf{x}_{\tau_i}$ and $X_{i,m_i} = \mathbf{x}_{\tau_{i+1}}$, and the imputed data is referred to as $X_{i,1}, \dots, X_{i,m_i-1}$. Thus, one has observed data $D_T = \{X_{0,0}, X_{1,0}, \dots, X_{M-1,0}, X_{M-1,m_{M-1}}\}$ and overall imputed data $X_{[0,T]} = \{X_{i,1}, \dots, X_{i,m_i-1} \mid i = 0, \dots, M-1\}$. The update scheme above is adapted to the generalised setting as is illustrated in figure (3.2).

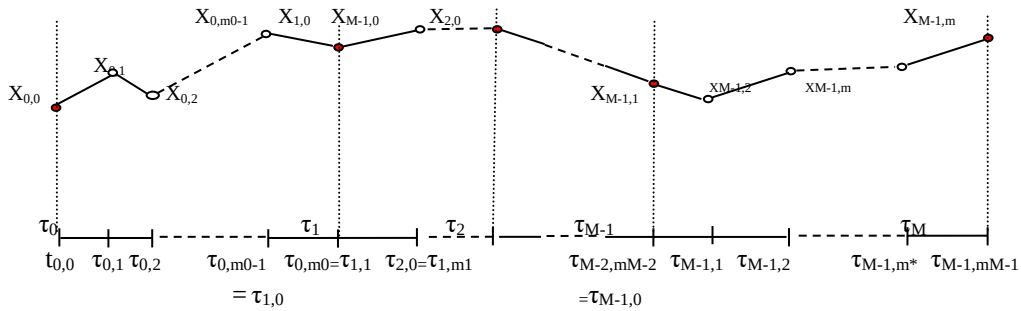


Figure 3.2 Shows an illustration of d -dimensional path consisting of the initial value and M discrete observations at times $\tau_0, \dots, \tau_M$ (labelled with red circles) and imputed data (labelled with white circles).

Now the likelihood function becomes:

$$L(\theta | x_{i,k}) = \prod_{i=0}^{M-1} \prod_{k=0}^{m_i-1} \pi^{Euler}(X_{i,k+1} | X_{i,k}, \theta) \quad (3.34)$$

3.3.2 Single Site Updating

Under single site sampling update, Eraker (2001) and Golightly & Wilkinson (2005) samples the skeleton path update $\pi(X | \theta, D_T)$ indirectly. They implement a Gibbs sampler to update each column X_k of $\mathbf{X}^* | \mathbf{D}_T$ conditional on its neighbours and the parameter value θ . The full conditional distribution of X_k (with k not an integer multiple of m) given the neighbours was proposed thus:

$$\begin{aligned} \pi(X_{k+1} | X_k, X_m) &\propto \pi(X_{k+1} | X_k, \theta) \times \pi(X_m | X_{k+1}, \theta) \\ &= \phi(X_{k+1}; X_k + \alpha(X_k, \theta)\Delta t, \beta(X_k, \theta)\Delta t) \times \phi(X_m; X_{k+1} + \alpha(X_{k+1}, \theta)\Delta t, \beta(X_{k+1}, \theta)\Delta t). \end{aligned} \quad (3.35)$$

For multidimensional diffusions, direct sampling of this distribution is not possible and a Metropolis Hasting step is used. A new $\mathbf{X}_k^*$ is drawn from a suitable proposal density. Follow Eraker (2001) the proposal density was given thus

$$q(X_{k+1}^* | X_k, X_m, \theta) \equiv \phi\left(X_{k+1}^*; \frac{1}{2}(X_k + X_m), \frac{1}{2}\beta(X_k, \theta)\Delta t\right) \quad (3.36)$$

If the iteration counter is at the initial condition say c , then X_k is the value obtained at iteration c , and X_m is the value obtained at iteration $c - 1$. Hence, if the current state of the chain is X_k , then a proposed value X_{k+1}^* is accepted with probability

$$\alpha(X_{k+1}^*, X_{k+1}) = \min \left\{ 1, \frac{\pi(X_{k+1}^* | X_k, X_m, \theta)}{\pi(X_{k+1} | X_k, X_m, \theta)} \times \frac{q(X_{k+1} | X_k, X_m, \theta)}{q(X_{k+1}^* | X_k, X_m, \theta)} \right\} \quad (3.37)$$

And we set $X_{k+1}^{(c)} := X_{k+1}^*$, otherwise we set $X_{k+1} = X_k$. Note that, density $\pi(\cdot | X_k, X_m, \theta)$ needs only to be known up to a multiplicative constant, since the acceptance probability only involves ratios of density. Hence, instead we sample $\pi(\cdot | X_{k+1}, X_m, \theta)$. A description of the algorithm for single site updating can be found in Algorithm 3.3.

Algorithm 3.3 Single site Updating Diffusion Bridge Approach

1. Initialise all unknowns. Use linear interpolation to initialise the X_k . Set counter(c) = 1.
 2. Update $X^{(c)} \mid \theta_{(c-1)}, D_T$ as follows:
 - 2.1 Propose $X_k^* \sim q(X_k^* \mid X_{k-1}, X_{k+1}, \theta)$ for each k not an integer multiple of m .
 - 2.2 Set $X_k^{(c)} := X_k^*$ with probability as in (3.26) otherwise store the current value X_k .
 Draw $\theta_{(c)} \sim \pi(\cdot \mid X^{(c)})$.
 4. If the desired number of simulations has been performed then stop, otherwise, set $c = c + 1$ and return to step 2.
-

Under multidimensional diffusions processes, Elerian et al. (2001) show that an algorithm which updates one column of $\mathbf{X}$ at a time leads to poor mixing, due to high correlation amongst the latent data. Consequently, it is recommended that $\mathbf{X}^* \mid \mathbf{D}_T$ is updated in blocks of random size. Here, we consider the simplest blocking algorithm whereby the missing values are updated in blocks of size $m - 1$, between every pair of observations.

3.3.3 Block Updating

Consider consecutive observations x_{tk} and x_{tM} (where we let $M = k + m$) corresponding to columns $\mathbf{X}_k$ and $\mathbf{X}_M$ in multidimensional process $\mathbf{X}$. Between these two observations, we have $m-1$ missing values, $\mathbf{X}_{k+1}, \dots, \mathbf{X}_{M-1}$ for which the full posterior distribution conditioned on observation as:

$$\pi(X_{k+1}, \dots, X_{M-1} \mid X_k, \mathbf{X}_M, \theta) \propto \prod_{j=k}^{M-1} \pi(X_{k+1} \mid X_k, \theta) \quad (3.38)$$

This distribution aim to sample this density for $k = 0, m, \dots, n-m$ in turn, thereby generating a sample from the path density $\pi(X \mid \theta, D_T)$.

Through adopted Metropolis Hastings step explained in the previous chapter, we construct a Gaussian approximation to the density of X_{k+1} (for $j = k, \dots, M-2$) conditional on start X_{tk} and the end-point X_{tm} of the interval. We also construct the joint density of X_{tM} and X_{tk+1} by

combining the Euler transition density in (3.24) with an approximation of $\pi(X_M | X_{k+1}, \theta)$. By conditioning the resulting distribution on the end-point X_M gives:

$$\tilde{\pi}(X_{k+1} | X_k, X_M, \theta) = \phi(X_{k+1}; X_k + \alpha^*(X_k, \theta)\Delta t, \beta^*(X_k, \theta)\Delta t) \quad (3.39)$$

Where $\alpha^*(X_k, \theta)\Delta t$ and $\beta^*(X_k, \theta)\Delta t$ are the proposed mean and covariance needed in the proposed diffusion bridge under study. Hence, for each $k = 0, m, \dots, n - m$, we sample $\pi(X_{k+1}, \dots, X_{M-1} | X_k, X_M, \theta)$ by proposing X_{k+1}^* for $j = k, \dots, M - 2$ via recursive draws from the density in (3.39).

Note, sampling from (3.39) gives a skeleton path of a diffusion bridge, conditioned on start at X_k and finish at X_M .

If the current state of the chain is $X_k, \dots, X_{M-1}$ then we accept the move with acceptance probability given by:

$$\alpha(X_{k+1}^*, X_{k+1}) = \min \left\{ 1, \left[\prod_{k=0}^{M-1} \frac{\pi(X_{k+1}^* | X_k^*, \theta)}{\pi(X_{k+1} | X_k, \theta)} \right] \times \left[\prod_{k=0}^{M-2} \frac{\tilde{\pi}(X_{k+1} | X_k, X_M, \theta)}{\tilde{\pi}(X_{k+1}^* | X_k^*, X_M, \theta)} \right] \right\} \quad (3.40)$$

This acceptance probability tends to a finite limit as $m \rightarrow \infty$. The Gibbs sampler with block updating then has the following algorithmic form:

Algorithm 3.4 Blocking Update Diffusion Bridge Approach

1. Initialise all unknowns. Use linear interpolation to initialise the X_k . Set counter (c) = 1.
 2. Update $X^{(c)} | \theta_{(c-1)}, D_T$ as follows. For $k = 0, m, \dots, n - m$:
 - 2.1 Propose $X_k^* \sim q(X_{k+1}^* | X_k, X_M, \theta)$ for $k = 0, \dots, M - 2$.
 - 2.2 Accept and store the move with probability given by (3.29) otherwise store the current value of the chain.
 3. Draw $\theta_{(c)} \sim \pi(\cdot | X_{(c)})$.
 4. If the desired number of simulations has been performed then stop, otherwise, set $c = c + 1$ and return to step 2.
-

This study adopted the block updating method to overcome the dependence within the latent process conditional on the model parameters, though, this does not overcome the more fundamental convergence issue.

3.3.4 Convergence Problem

The theoretical and computational application of Bayesian method of estimation becomes problematic, specifically in multi-dimensional diffusion models, due to the following convergence problems:-

- Direct implementation of Bayesian method of estimation results into a Markov chain that is reducible, if unknown parameters appear in the diffusion coefficient of any diffusion process.
- Diffusion bridges samplers resulting into decreasing in acceptance probability of convergence, when the imputed values (m) between any two consecutive time point increases ($m \rightarrow \infty$).

As the discretization gets finer, that is, as imputed value (m) increases, it is possible to make very precise inference about the diffusion coefficient of the process via the quadratic variation. Consider a complete data sample path X on $[0, T]$. However, there is one notorious convergence problem caused by a close link between the model parameters and the quadratic variation of the diffusion path. A practical solution for this problem has not been proven for multidimensional diffusion processes. This open question is addressed in this study. The sample path X gives the integral of the diffusion coefficient through the quadratic variation

$$\langle X, X \rangle_T = \int_0^T \beta^2(X_t, \theta) dt \quad (3.41)$$

This means that the parameter of interest θ can be deduced from X and consequently a scheme which imputes $X^*|\mathcal{D}_T$ and then updates θ will be reducible; since X confirms θ and θ is in turn determined by the quadratic variation, the scheme will not converge. This dependence (between the quadratic variation and diffusion coefficient) was highlighted as a problem by Roberts & Stramer (2001) and results in long mixing times of MCMC algorithms such as the single site Gibbs sampler, though the problem is less noticeable for $m \leq 5$. Elerian et al (2001) overcome this dependence in the context of univariate diffusions by transforming the missing data, giving a partially non-centred parameterisation which leads to an irreducible algorithm even in the limit as $m \rightarrow \infty$. However, for a multi-dimensional diffusion satisfying (3.4), finding such a transformation requires an invertible function, $g : \mathbb{R}^u \rightarrow \mathbb{R}^u$ such that

$$\beta(X_t, \theta)[\beta(X_t, \theta)]^T = \beta^{-1}(X_t, \theta) \quad (3.42)$$

Equation (3.31) is almost always impossible to solve in practice, except we ensure that the diffusion matrix assumed invertible, especially for multidimensional diffusions as was considered in the section (3.1.1).

3.4 Diffusion Bridge Constructs

Suppose X_t is a d -dimensional diffusion with time dependent drift $\alpha : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and diffusion coefficient $\beta : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ governed by the stochastic differential equation

$$d\mathbf{X}_t = \alpha(\mathbf{X}_t, \theta)dt + \beta(\mathbf{X}_t, \theta)d\mathbf{W}_t, \quad X_0 = x_0 \quad (3.43)$$

where W is a standard d -dimensional Brownian motion. When the process X is conditioned to hit a point $x_T \in \mathbb{R}^d$ at time $T > 0$, the resulting process X_t^* on $[0, T]$ is called the diffusion bridge from x_0 to x_T . A stochastic bridge is a random sample path, X_t , that starts from some known value initially say x_0 , but always lands at a fixed value at some future time-point, say x_T at time T . In that setting the fact that the transition densities are unavailable implies that the likelihood of the data is not accessible. A successful approach initiated by Roberts and

Stramer (2001) is to circumvent this problem by viewing the continuous segments between the observed data points as missing data. Computational algorithms can then be designed that augment the discrete-time data by (repeatedly) simulating the diffusion bridges between the observed data points. This statistical application of simulation algorithms for diffusion bridges are one of our set objective in chapter one.

To be able to carry out these procedures in practice, simulating paths from the proposal process has to be relatively easy and, up to a normalizing constant. It is therefore desirable to have proposals that are “close” to the target in an appropriate sense. In this section we considered and construct such proposals for the multi-dimensional setting.

Our aim is to generate discrete-time realisations of $\mathbf{x}_t^*$ conditional on x_0 and x_T on a partition time interval $[0, T]$.

For almost all choices of $\alpha(X_t, \theta)$ and $\beta(X_t, \theta)$ there is no direct way of simulating $x_{[0,T]}$.

Instead, one can first generate a proposal bridge $x_{[0,T]}^*$ and accept with Metropolis-Hasting

acceptance-probability $\min \left\{ 1, \left(\frac{\pi_\theta(x_{[0,T]}^*)}{\pi_\theta(x_{[0,T]})} \right) \right\}$. In sampling diffusion path, a description of the

algorithm can be found in Algorithm 3.5

Algorithm 3.5 Sample path proposed Diffusion Bridge

1. Initialize a counter $i = 1$, skeleton path $x_{r0}^{(i)}$ and $x_m^{(i)}$ according to the equation $q(x_{[0,T]}; x_{rk+1} | x_{rk}, \theta)$
2. Iterate: update the current value of $x_{[0,T]}^*$ according to $x_{r0}^{(i)}$
3. Generate a proposal skeleton path $q(x_{[0,T]} | x_0, \theta)$ and compute

$$\alpha(x_{(0,T)}^*, x_{(0,T)}) = \min \left\{ 1, \frac{\pi(x_{(0,T)}^* | x_T, \theta) q(x_{(0,T)} | x_{(0,T)}^*, x_T, \theta)}{\pi(x_{(0,T)} | x_T, \theta) q(x_{(0,T)}^* | x_{(0,T)}, x_T, \theta)} \right\}.$$

4. Let $x_{[0,T]}^* = x_{rk+1}^*$ if else, $x_{[0,T]}^* = x_{rk}$
 5. Set counter $= i+1$ and return to step 2.
-

The feasibility and efficiency of the algorithm is crucially determined by choice of the transition kernel $\mathbf{q}$ (proposal process) conditioned on x_0 , x_T and other given parameters. **3.7.1**

3.4.1 Euler Scheme proposal Sampler (E)

The Euler scheme or myopic simulation is by applying the Euler-Murayama approximation over each interval of length $\Delta\tau$ leads to a proposal density of the form:

$$q_{Euler}(x_{[0,T]}^* | x_0, x_T, \theta) \sim \phi(x_{[0,T]}^* | x_\tau^* + \mu(x_\tau^*, \theta)\Delta\tau, \psi(x_\tau^*, \theta)\Delta\tau) \quad (3.44)$$

In all, $\phi(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\psi})$ is refers to multivariate Gaussian density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\psi}$ evaluated at $\mathbf{x}$. And our focus is on the choice of mean vector $\boldsymbol{\mu}(\cdot)$ and covariance matrix $\boldsymbol{\psi}(\cdot)$.

With acceptance probability

$$\begin{aligned} \alpha(x_{tk}^*, x_{tk}) &= \min \left\{ 1, \frac{\prod_{k=0}^{m-1} \pi^{Euler}(x_{tk+1}^* | x_{tk}^*, \theta) q_{Euler}(\Delta\tau; x_{tk}, x_{tk+1})}{\prod_{k=0}^{m-1} \pi^{Euler}(x_{tk+1} | x_{tk}, \theta) q_{Euler}(\Delta\tau; x_{tk}^*, x_{tk+1}^*)} \right\} \\ &= \min \left\{ 1, \frac{\prod_{k=0}^{m-1} \pi^{Euler}(x_{tk+1}^* | x_{tk}^*, \theta) \times \prod_{k=0}^{m-2} \pi^{Euler}(x_{tk+1} | x_{tk}, \theta)}{\prod_{k=0}^{m-1} \pi^{Euler}(x_{tk+1} | x_{tk}, \theta) \times \prod_{k=0}^{m-2} \pi^{Euler}(x_{tk+1}^* | x_{tk}^*, \theta)} \right\} \\ &= \min \left\{ 1, \left(\frac{\pi^{Euler}(x_{tk+1}^* | x_{tk}^*, \theta)}{\pi^{Euler}(x_{tk+1} | x_{tk}, \theta)} \right) \right\} \end{aligned} \quad (3.45)$$

The acceptance probability for this proposal bridge is inefficient due to large numbers of rejections, leading to low acceptance probability of MCMC samplers.

This strategy is likely to work well provided that the observation x_0 is not particularly informative, that is, when the measurement error dominates the intrinsic stochasticity of the process. However, as covariance ψ is reduced, the MH acceptance probability decreases. A related approach can be found in Pedersen (1995), where it is assumed that x_T is known. If the acceptance probability of the diffusion bridge is of this form:

$$\alpha(x_{k+1}^*, x) = \min \left\{ 1, \prod_{k=0}^{T-1} \frac{\pi(x_{k+1}^* | x_k^*, \theta) q(x | x^*, x_T, \theta)}{\pi(x_{k+1} | x_k, \theta) q(x^* | x, x_T, \theta)} \right\} \quad (3.46)$$

By apply the transition probability; the Euler proposal for Diffusion Bridge will have the acceptance probability of the form:

$$\alpha(x_k^*, x_k) = \min \left\{ 1, \frac{\pi^{Euler}(x_{k+1}^* | x_k^*, \theta) \times \pi^{Euler}(x_{k+1} | x_k, \theta)}{\pi^{Euler}(Y_{k+1} | Y_k, \theta) \times \pi^{Euler}(x_{k+1}^* | x_k^*, \theta)} \right\} \quad (3.47)$$

This strategy is likely to work well, provided that the observation x_T is not particularly informative, that is, when the measurement error dominates the intrinsic stochasticity of the process. The approach poses acceptance probability that is reducible when the numbers of imputed value tend to infinity ($m \rightarrow \infty$).

3.4.2 Modified Diffusion Bridge proposal (MDB)

Another importance sampler is called the modified Diffusion Bridge (MDB) sampler, which is originally proposed by Durham and Gallant (2002) for the univariate case, and later modified, by Golightly and Wilkinson (2006) for multidimensional case.

Here the diffusion bridge is conditioned on the proposal distribution q_{MDB} of x_{k+1}^* on the preceding value x_k ($k = 0$) and on the end point x_T of the path segment to be updated.

Then the modified diffusion bridge (MDB) sampler draws $x_{[0, T]}$ from the density

$$q_{MDB}(x_{k+1} | x_k, x_T, \theta) \propto \pi(x_T | x_{k+1}, \theta) \times \pi(x_{k+1} | x_k, \theta) \quad (3.48)$$

$$\approx \phi(x_T | x_{k+1} + \mu(x_{k+1}, \theta) \Delta \tau, \psi(x_{k+1}, \theta) \Delta \tau) \quad \times \quad \cdot \phi(x_{k+1} | x_k + \mu(x_k, \theta) \Delta \tau, \psi(x_k, \theta) \Delta \tau_k)$$

where $\Delta \tau = \tau_{k+1} - \tau_k$

The corresponding proposal for $x_{[0, T]}$ is to iteratively by draw

$$x_{k+1}^* \sim \phi \left(x_k + \frac{x_T - x_k}{\tau_T - \tau_k} \Delta \tau_k, \frac{\tau_T - \tau_{k+1}}{\tau_T - \tau_k} \beta(x_k, \theta) \Delta \tau_k \right) \quad (3.49)$$

where $\phi(x; \mu, \psi)$ is a multivariate normal density with infinitesimal mean vector μ :

$$\mu = x_k + \frac{x_T - x_k}{\tau_T - \tau_k} \Delta \tau_k \text{ and covariance matrix } \psi: \psi = \frac{\tau_T - \tau_{k+1}}{\tau_T - \tau_k} \beta(x_k, \theta) \Delta \tau_k.$$

The acceptance probability

$$\alpha(x_{(0,T)}^*, x_{(0,T)}) = \min \left\{ 1, \left(\prod_{k=0}^{T-1} \frac{\pi^{Euler}(x_{k+1}^* | x_k, \theta)}{\pi^{Euler}(x_{k+1}^* | x_k, \theta)} \right) \frac{q_{MDB}(x_{(0,T)} | x_0^*, x_T^*, \theta)}{q_{MDB}(x_{(0,T)}^* | x_0, x_T, \theta)} \right\} \quad (3.50)$$

Where, $q_{MDB}(x_{(0,T)}^* | x_0, x_T, \theta) = \prod_{k=0}^{T-2} q_{MB}(x_{k+1}^* | x_k, x_T, \theta)$

$$= \prod_{k=0}^{T-2} \phi \left(x_{k+1}^* | x_k^* + \frac{x_T^* - x_k^*}{\tau_T - \tau_k} \Delta \tau_k, \frac{\tau_T - \tau_{k+1}}{\tau_T - \tau_k} \beta(x_k^*, \theta) \Delta \tau_k \right)$$

For equidistant time intervals

$$q_{MDB}(x_{k+1}^* | x_k^*, \theta) = \phi \left(x_{k+1}^* | x_k^* + \frac{x_T^* - x_k^*}{T - k}, \frac{T - k - 1}{T - k} \beta(x_k^*, \theta) \Delta \tau_k \right) \quad (3.51)$$

Unfortunately, the MDB is only efficient when the infinitesimal mean (drift) of (3.24) is approximately constant. When this is not the case (so that realisations of the SDE started from the same point exhibit strong and similar non-linearity over the inter observation time) the modified diffusion bridge is likely to be unsatisfactory.

3.4.3 Lindström Bridge (LB)

A bridge construct that combines the Euler-scheme (E) sampler with the Modified Diffusion Bridge (MDB) has been proposed in Lindstrom (2012) to overcome the Euler scheme limitation. This is often called Regularized sampler. Here we give the explicit form of this regularized sampler. For sparsely sampled data, the MDB sampler can produce a poor approximation (Lindstrom, 2012). Let $\alpha(X_{\tau_k})$ and $\beta(X_{\tau_k})$ be the mean and the variance of the Pedersen sampler and $\mu_{MDB}(X_{\tau_k})$ and $\psi_{MDB}(X_{\tau_k})$ be the mean and the variance of the MDB sampler. The joint distribution of $X_{\tau_{k+1}}$ conditional on x_{τ_k} and observed data ($q_{LB}(X_{\tau_{k+1}} | X_{\tau_k}, X_0)$) is then approximated by

$$q_{LB}(x_{(0,T)}^* | x_0, x_T, \theta) = \phi(x_{\tau_{k+1}}^*; \mu_{LB}^*(X_{\tau_k}), \psi_{LB}^*(X_{\tau_k})) \quad (3.52)$$

where $\mu_{LB}^*(X_{\tau_k}) = \omega_k^\gamma \mu_{MDB}(X_{\tau_k}) + (1 - \omega_k^\gamma) \alpha(X_{\tau_k})$ and $\psi_{LB}^*(X_{\tau_k}) = \omega_k^\gamma \psi_{MDB}(X_{\tau_k}) + (1 - \omega_k^\gamma) \beta(X_{\tau_k})$

where $\omega_k^\gamma = \frac{(\tau_k + 1 - \tau_k)(T - \tau_k)}{(\tau_k + 1 - \tau_k)(T - \tau_k) + \gamma(T - \tau_{k+1})^2}$ with $\gamma > 0$.

The acceptance probability

$$\alpha(x_{(0,T)}^*, x_{(0,T)}) = \min \left\{ 1, \left(\prod_{k=0}^{T-1} \frac{\pi^{Euler}(x_{k+1}^* | x_k^*, \theta)}{\pi^{Euler}(x_{k+1} | x_k, \theta)} \right) \frac{q_{LB}(x_{(0,T)} | x_0^*, x_T^*, \theta)}{q_{LB}(x_{(0,T)}^* | x_0, x_T, \theta)} \right\} \quad (3.53)$$

The performance of the regularized sampler depends on the parameter γ . A large γ will make the regularized sampler similar to the Pedersen sampler and a smaller γ will make it similar to the MDB sampler. In practice, Lindström (2012) suggests that $\gamma \in [0.01, 1]$, given that these values have proved successful in simulation experiments. Note also that for a fixed γ , if $T - \tau_{k+1} \gg \Delta_\tau$ then $w_k^\gamma \simeq 0$ and the Euler sampler dominates. However, as τ_{k+1} approaches T , w_k^γ approaches 1 and the Lindström Bridge (LB) is dominated by the Modified Diffusion Bridge. Whilst the Lindström Bridge (LB) attempts to account for nonlinear dynamics by combining the Modified Diffusion Bridge (MDB) with the Euler approach, having to specify a model-dependent tuning parameter is unsatisfactory, since different choices of γ will lead to different properties of the proposed bridges. Moreover, the link between the regularised sampler and the continuous-time conditioned process is unclear.

3.4.4 Guided proposals (GP)

Delyon and Hu (2006) extended the work of Durham and Gallant (2002) to a continuous time setup and derived another innovative proposal process taking the drift to the target diffusion into account. The basic idea in Delyon and Hu consists of superimposing an additional term to the drift of the unconditioned diffusion to guide the process towards the end point x_T . Such proposals are termed guided proposals and have the advantage that only forward simulation of a stochastic differential equation (SDE) is required.

More precisely, in Delyon and Hu, the drift of the proposal that hit $x_T \in R^d$ at time T equals

$$\tilde{\alpha}(\tilde{X}_t, \theta) = \lambda \alpha(X_t, \theta) + \left(\frac{x_T - x_k}{T - t_k} \right), \text{ here } \lambda \text{ is either 0 or 1 is chosen. If } \lambda = 0 \text{ we have } \left(\frac{x_T - x_k}{T - t_k} \right) \text{ as}$$

guiding term matched with the drift of the SDE for a Brownian bridge, which indeed has drift 0. However, this proposal has the drawback, that is, independent of the drift $\alpha(X_t, \theta)$ of the diffusion. If $\lambda = 1$, the guiding term depends on $\alpha(X_t, \theta)$ and this results in a mismatch

between the drift and guiding term. In both cases, there would be substantial mismatch between the proposals and true bridges, rendering the low acceptance probability in a Metropolis-Hastings sampler.

Schauer et al. (2016) do take the drift of the target diffusion into account, in a way different from Delyon and Hu (2006), but these proposals can only partially substantially reduce the mismatch of drift and guiding term without considered the low acceptance probability of the guiding term to pull the process towards the endpoint in the right manner. An improved version of diffusion bridge proposal that would take cares such problems and also exploit the added flexibility are addressed in this thesis.

The basic idea is to replace the generally the conditional process $\{X_t, t \in [0, T]\}$ of intractable transition density $\pi(x_T | x_t)$ that appears in the dynamics of the target bridge by the transition density $\pi(x_T | x_t)$ of a diffusion process $(\tilde{X})$ for which it is known in closed form, a class of linear processes $\{\tilde{X}_t, t \in [0, T]\}$ satisfying

$$d\tilde{X}_t = \tilde{\alpha}(\tilde{X}_t, \theta)dt + B(\tilde{X}_t, \theta)dt + \sqrt{\tilde{\beta}(\tilde{X}_t, \theta)}dW_t, \quad \tilde{X}_0 = x_0 \quad (3.54)$$

With the transition density $\pi(x_T | x_t)$.

here $B(\tilde{X}_t)$ and $\tilde{\beta}(\tilde{X}_t)$ are $d \times d$ matrices and $\tilde{\alpha}(\tilde{X}_t)$ is a d – vector. By defining the process $(\tilde{X}_t)$ as the solution of the SDE in (3.15) with

$$\tilde{\alpha}(\tilde{X}_t)dt = \alpha(X_t) + B(X_t)\nabla_{x_t}\log \pi(x_T | x_t)$$

(3.55) Given a tractable approximation of $\pi(x_T | x_t)$, the Euler-Maruyama approximation can be applied over the discretisation of $[0, T]$ to give a proposal density of the form:

$$\tilde{\alpha}(X_t) = \alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi(x_T | x_t)\{F'\pi(x_T | x_t)\psi(x_T | x_t)\pi'(x_T | x_t) + \Sigma\}^{-1}(x_T, -F'\eta_T) \quad (3.56)$$

where

$$\mu_{GP}(x_{\tau k}) = \alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi(x_T | x_t)\{F'\pi(x_T | x_t)\psi(x_T | x_t)\pi'(x_T | x_t)\}^{-1}(x_T, -F'\eta_T) \quad (3.57)$$

$$\text{and } \psi_{GP}(x_{\tau k}) = \beta(x_{\tau k}, \theta)\Delta\tau \quad (3.58)$$

The acceptance probability

$$\alpha(x_{(0,T)}^*, x_{(0,T)}) = \min \left\{ 1, \left(\prod_{k=0}^{T-1} \frac{\pi^{Euler}(x_{k+1}^* | x_k^*, \theta)}{\pi^{Euler}(x_{k+1} | x_k, \theta)} \right) \frac{q_{GP}(x_{(0,T)} | x_0^*, x_T^*, \theta)}{q_{GP}(x_{(0,T)}^* | x_0, x_T, \theta)} \right\} \quad (3.59)$$

Under all these updated scheme procedure above, the proposal mechanism via MCMC becomes degenerate as imputed values become large ($\mathbf{m} \rightarrow \infty$) Roberts and Stramer (2001).

To overcome this, we consider a new improved version of diffusion bridge proposal, through modification on just recently proposal proposed by Schauer et.al, (2016), here known as Modified Guided proposal (MGP)

3.5 Improved Diffusion Bridge sampler

3.5.1 Modified Guided Proposal (MGP)

By considering the so-called guided proposals, according to the terminology suggested in Papaspiliopoulos and Roberts (2012). This means that our proposal is realizations of a process X° that solves an SDE of the form (3.4) as well, but with a drift and diffusion coefficient term adapted in order to force the process X° to hit the point x_T at time T .

The approach is inspired by the ideas in Clark (1990), Durham and Gallant (2004) and Schauer et al. (2016). The approach allows more flexibility in constructing an appropriate guiding term to both the drift and diffusion coefficient functions of the process. This is particularly aimed at finding procedures with higher acceptance probabilities in the multi-dimensional case.

Recall that, under weak assumptions the target diffusion bridge X^* is characterized as the solution to the Stochastic Differential Equations

$$dX_t^* = \alpha^*(X_t^*, \theta)dt + \beta^*(X_t^*, \theta)dW_t, \quad X_0^* = x_0 \quad (3.60)$$

here $\alpha^*(x_t^*, \theta) = \alpha(x_t, \theta) + \tilde{\alpha}(x_t, \theta) \nabla_x \log \pi(t, x_0; T, x_T)$, $\beta^*(x_t^*, \theta) = \frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_t, \theta)$

and $\tilde{\alpha}(x_t, \theta) = \beta(x_t, \theta)(\beta(x_t, \theta))^T$

The guiding term $\tilde{\alpha}(x_t, \theta) \nabla_x \log \pi(t, x_0; T, x_T)$ is added to the original drift to direct X^* towards x_T , likewise, $\frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_t, \theta)$ for diffusion coefficient from the current position $X_t^* = x_0$ in just the right manner.

Equation (3.59) contains unknown transition densities of the original process X_t , it cannot be employed directly for simulation. We suggest replacing this unknown density by one coming from an auxiliary diffusion process (X_t°) with known transition densities $\tilde{\pi}(x_T | x_0)$.

$$dX_t^\circ = \alpha^\circ(X_t^\circ, \theta)dt + \beta^\circ(X_t^\circ, \theta)dW_t, \quad X_0^\circ = x_0^\circ \quad (3.61)$$

where $\alpha^\circ(x_t, \theta) = \alpha(x_t, \theta) + \tilde{\alpha}(x_t, \theta) \nabla_x \log \tilde{\pi}(t, x_0; T, x_T)$,

$\beta^\circ(x_t, \theta) = \frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_t, \theta)$ and $\tilde{\pi}(s, x_0; t, x_T)$ is the transition density of a diffusion process

$\tilde{X}$ which above expression is known in closed form. These make our proposed bridges different from those defined in Schauer et al (2016). Because, $\tilde{\alpha}(x_t, \theta)$ of the original process appears in the drift and another guide $\frac{T - \tau_{k+1}}{T - \tau_k}$ appears in diffusion coefficient of the proposal process X_t° .

The approach contains two main theoretical results. In the first we give conditions under which the process X_t° is indeed a valid proposal process in the sense that its distribution say $\mathbf{P}^\circ$ (Borel measure on $C([0, T], \mathbb{R}^d)$) is equivalent to the law $\mathbf{P}^*$ of the target process X_t^* and when derive expression for the Radon-Nikodym derivative:

$$\frac{d\mathbf{P}^*}{d\mathbf{P}^\circ}(X_t^\circ) \propto \exp\left(\int_0^T G(s, X_s^\circ)ds\right)$$

$$\text{If } \psi_{(\theta)} = \exp\left(\int_0^T G(s, X_s^\circ) ds\right)$$

$$\text{then, } \frac{d\mathbf{P}^*}{d\mathbf{P}^\circ}(X_t^\circ) = \frac{\tilde{\pi}(0, x_0; T, x_T)}{\tilde{\pi}(0, X_t^\circ; T, x_T)} \psi_{(\theta)}$$

$$\frac{d\mathbf{P}^*}{d\mathbf{P}^\circ}(X_t^\circ) = \frac{\tilde{\pi}(0, x_0; T, x_T) \pi(0, X_t^\circ; T, x_T)}{\pi(0, x_0; T, x_T) \tilde{\pi}(0, X_t^\circ; T, x_T)} \psi_{(\theta)} \quad (3.62)$$

here the functional does not depend on unknown or inaccessible objects. Using the Euler-Maruyama approximation gives the variance of $X_{\tau k+1} | x_{\tau k}, x_T$ in the guided proposal process as $\psi_{GP}(x_{\tau k}) \Delta \tau = \beta(x_{\tau k}) \Delta \tau$. We investigate the effect of using this variance of modified diffusion bridge construct by taking $\psi_{GP}(x_{\tau k}) = \psi_{MDB}(x_{\tau k})$. We observed a worthwhile increase in performance. If x_T is known, used of the MDB variance in place of Euler scheme variance $\beta(x_{\tau k})$ comes at almost no additional computational cost. We denote this construct by MGP.

The propose Bridge sampler now is of the form:

Infinitesimal mean;

$$\mu_{MGP}(x_{\tau k}) = \alpha(x_t, \theta) + \tilde{\alpha}(x_t, \theta) \nabla x \log \tilde{\pi}(s, x_0; t, x_T) \quad (3.63)$$

and infinitesimal covariance

$$\psi_{MGP}(x_{\tau k}) = \frac{T - \tau_{k+1}}{T - \tau_k} \beta(x_t, \theta) \Delta \tau \quad (3.64)$$

And the proposal density $q(\cdot | \cdot)$ becomes:

$$q_{MGP}(x_{(0,T)} | x_0^*, x_T^*, \theta) = \phi(x_{(0,T)}^*; \mu_{MGP}(x_{\tau k}), \psi_{MGP}(x_{\tau k})) \quad (3.65)$$

with acceptance probability

$$\alpha(x_{(0,T)}^*, x_{(0,T)}) = \min \left\{ 1, \left(\frac{\prod_{k=0}^{T-1} \pi^{Euler}(x_{k+1}^* | x_k^*, \theta)}{\prod_{k=0}^{T-1} \pi^{Euler}(x_{k+1} | x_k, \theta)} \right) \frac{q_{MGP}(x_{(0,T)} | x_0^*, x_T^*, \theta)}{q_{MGP}(x_{(0,T)}^* | x_0, x_T, \theta)} \right\} \quad (3.66)$$

3.6 Diffusion Epidemic models with Latent Data

It is often occurs in epidemic data that the state variables of some components are not fully observable. To that end, the path update diffusion bridge needs to be modified for the latent data. This would lead to the extension of the estimation schemes which alternately perform a path and a parameter update of diffusion process to a latent data. Here, if considered the assumption set for observation times, by assume equidistant observation times with partially observed data of the form:

Define $D_T = \{ X_k, k \in \{0, \dots, K\} | t_k \in \{\tau_0, \dots, \tau_m\} \}$ such that,

$$D_T^{(i)} = \{X_{t_0}^{(i)}, X_{t_1}^{(i)}, \dots, X_{t_m}^{(i)}\}^T \quad (3.67)$$

Choose an update interval (t_0, t_T) such that $|\{t_1, \dots, t_{T-1}\} \cap D_T| \leq 1$.

Let first considered 2-dimensional process $X_t^{(1)}$ represent dimension $d_1 > 0$ and $X_t^{(2)}$ dimension $d_2 \geq 0$.

$$X_t^{(i)} = (X_t^{(1)}, X_t^{(2)})^T \in D_T, \quad i = 1 \text{ and } 2 \quad (3.68)$$

For observed part: $X_t^{(1)} \in \mathfrak{R}^{d_1}$, $d_1 > 0$. And unobserved part: $X_t^{(2)} \in \mathfrak{R}^{d_2}$, $d_2 \geq 0$.

Such that $d_1 + d_2 = d$.

For the simplicity notation, the infinitesimal mean given thus:

$$\alpha(X_t^{(i)}, \theta) = \mu^{(i)} = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} = \begin{pmatrix} \alpha(X_t^{(1)}, \theta) \\ \alpha(X_t^{(2)}, \theta) \end{pmatrix} \quad (3.69)$$

Where $\alpha_i(X_t^{(i)}, \theta) = X_t^{(i)} + \alpha_{(i)}(X_t^{(i)}, \theta) + \Sigma_{ij}(\Sigma_{jj})^{-1} \times X_t^{(j)} - E(X_t^{(j)})$

and the infinitesimal covariance matrix as:

$$\Sigma^{(i)} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} = \begin{pmatrix} \beta_i(X_t^{(11)}, \theta) & \beta_i(X_t^{(12)}, \theta) \\ \beta_i(X_t^{(21)}, \theta) & \beta_i(X_t^{(22)}, \theta) \end{pmatrix} \quad (3.70)$$

where $\beta_i(X_t^{(i)}, \theta) = \beta_i(X_t^{(i)}, \theta)\Delta t - \Sigma_{ij}(\Sigma_{jj})^{-1} \times \Sigma_{ij}$ each infinitesimal mean and covariance can now be express in multivariate form as

$$\mu_{\tau k+1}^{(i)} = (x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_\tau) + \beta(x_{\tau k}, \theta)F\Delta_k[F'\beta(x_{\tau k}, \theta)F\Delta_k + \Sigma]^{-1}y_k - (F'(x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_\tau) \quad (3.71)$$

$$\Sigma_{\tau k+1}^{(ij)} = \beta(x_{\tau k}, \theta)\Delta_\tau - \beta(x_{\tau k}, \theta)F\Delta_k[F'\beta(x_{\tau k}, \theta)F\Delta_k + \Sigma]^{-1}F'\beta(x_{\tau k}, \theta)\Delta_k \quad (3.72)$$

The path of latent update can be representing in figure 3.3 below:

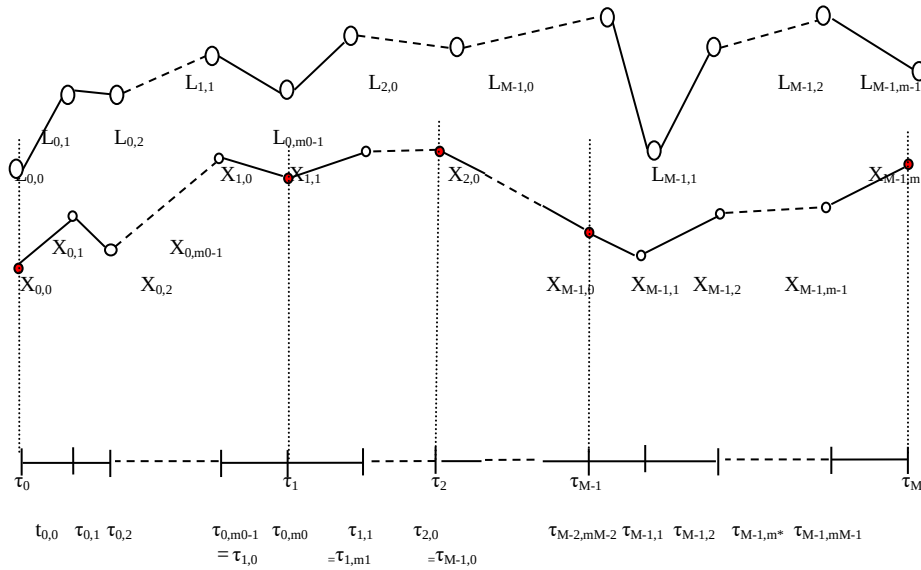


Figure 3.3 shows an illustration of two-dimensional path with an observable component X_t and a latent component L_t . The observed part of the path consists of the initial value at time τ_0 and M discrete observations at times $\tau_0, \dots, \tau_M$ (labelled with red circles) and latent or imputed data (labelled with white circles).

3.6.1 Procedures for the path update

The following situation may now occur for diffusion path update.

- (i) If one has $|\{t_1, \dots, t_{T-1}\} \cap D_T| = 0$ implies no observation time in the interior of $[t_0, t_T]$.

In this situation, there is no imputation problem in the interior. In this case, direct posterior distribution is required to solve the problem.

- (ii) If one has $|\{t_1, \dots, t_{T-1}\} \cap D_T| = 1$ implies exactly one imputed observation time say t_r in the

interior of $[t_0, t_T]$. Hence, this has two strategies imputed values in the interior (t_r and t_T). Then, (a) Propose $x_r^{*(2)} | x_0, x_r^{(1)}, x_T, \theta$ (intermediate observation time (t_r)). Then, generate two conditionally independent proposals on $[t_0, t_r]$ and $[t_r, t_T]$ conditional on $x_0, x_r^{(1)}, x_r^{*(2)}, x_T, \theta$

(b) Propose $x_{rk+1}^* | x_{rk}^*, x_r^{(1)}, x_T, \theta$ or $x_{rk+1}^* | x_{rk}^*, x_r^{(1)}, \theta$ then, update the latent path, which involves drawing from:

(i) Apply Euler approximation of $\pi(x_{rk+1} | x_{rk}, x_r^{(1)}, \theta)$, as was proposed in the previous full observed times diffusion bridge approach.

(ii) Apply Euler Approximation of $\pi(x_0^{(2)} | x_0^{(1)}, x_T, \theta)$, by considering

$$x_0^{(2)} | x_T, \theta \sim N_d(x_T - \mu_{rk}^{(2)} \Delta_{0T}, \Sigma_{rk}^{(2)} \Delta_{0T}) \quad (3.73)$$

where, $\mu_{rk}^{(2)} = \alpha(x_{rk}^{(2)}, \theta)$; $\Sigma_{rk}^{(2)} = \beta(x_{rk}^{(2)}, \theta) \beta'(x_{rk}^{(2)}, \theta)$ and $\Delta_{0T} = t_T - t_0$

Hence, $x_0^{(2)} | x_0^{(1)}, x_T, \theta \sim N_d(\eta_{rk}, \psi_{rk})$

where

$$\begin{aligned} \eta_{rk} &= x_T^{(2)} - \mu_{rk}^{(2)} \Delta_{0T} + \Sigma_{rk}^{21} (\Sigma_{rk}^{11})^{-1} (x_0^{(1)} - x_T^{(1)} + \mu_{rk}^{(1)} \Delta_{0T}) \\ \psi_{rk} &= \left(\Sigma_{rk}^{22} - \Sigma_{rk}^{21} (\Sigma_{rk}^{11})^{-1} \Sigma_{rk}^{12} \right) \Delta_{0T} \end{aligned}$$

With acceptance probability $\alpha\{x_0^{*(2)}, x_{(0,T)}^*, (x_0^{(2)}, x_{(0,T)})\}$ of the form:

$$= \min \left\{ 1, \frac{\pi(x_0^{*(2)}, x_{(0,T)}^* | x_0^{(1)}, x_T, \theta) \times q(x_0^{(2)}, x_{(0,T)} | x_0^{(1)}, x_T, \theta)}{\pi(x_0^{(2)}, x_{(0,T)} | x_0^{(1)}, x_T, \theta) \times q(x_0^{*(2)}, x_{(0,T)}^* | x_0^{(1)}, x_T, \theta)} \right\} \quad (3.74)$$

Where, $\pi(x_0^{*(2)}, x_{(0,T)}^* | x_0^{(1)}, x_T, \theta) \propto \left(\prod_{k=0}^{T-1} \pi^{Euler}(x_{k+1}^* | x_k^*, \theta) \right) \pi(x_0^* | \theta)$

and, $q(x_0^{*(2)}, x_{(0,T)}^* | x_0^{(1)}, x_T, \theta) = \left(\prod_{k=0}^{T-2} q(x_{k+1}^* | x_k^*, x_k^{(1)}, \theta) \right) \times q(x_0^{*(2)} | x_T^{(1)}, x_T, \theta) \quad (3.75)$

where $x_0^* = (x_0^{(1)}, x_0^{*(2)})'$ and $\pi(x_0^* | \theta) \propto \pi(x_0^{*(2)} | x_0^{(1)}, \theta)$

(iii) Apply Euler Approximation of $\pi(x_T^{(2)} | x_{T-1}, x_T^{(1)}, \theta)$

Similarly, $x_T | x_{T-1}, \theta \sim N_d(x_{T-1} + \mu_{T-1}\Delta_{T-1}, \Sigma_{T-1}\Delta_{T-1})$ as in (i) above

Implies that: $x_T^{(2)} | x_{T-1}, x_T^{(1)}, \theta \sim N_d(\eta_T, \psi_T)$

$$\eta_T = x_{T-1}^{(2)} + \mu_{T-1}^{(2)}\Delta_{T-1} + \Sigma_{T-1}^{21}(\Sigma_{T-1}^{11})^{-1}(x_T^{(1)} - x_{T-1}^{(1)} - \mu_{T-1}^{(1)}\Delta_{T-1})$$

Where, $\psi_T = \left(\Sigma_{T-1}^{22} - \Sigma_{T-1}^{21}(\Sigma_{T-1}^{11})^{-1}\Sigma_{T-1}^{12} \right)\Delta_{T-1}$

with acceptance probability $\alpha\{x_{(0,T)}^*, x_T^{*(2)}, (x_{(0,T)}, x_T^{(2)})\}$ of this form:

$$= \min \left\{ 1, \frac{\pi(x_{(0,T)}^*, x_T^{*(2)} | x_0, x_T^{(1)}, \theta) \times q(x_{(0,T)}, x_T^{(2)} | x_0, x_T^{(1)}, \theta)}{\pi(x_{(0,T)}, x_T^{(2)} | x_0, x_T^{(1)}, \theta) \times q(x_{(0,T)}^*, x_T^{*(2)} | x_0, x_T^{(1)}, \theta)} \right\} \quad (3.76)$$

where, $\pi(x_{(0,T)}^*, x_T^{*(2)} | x_0, x_T^{(1)}, \theta) \propto \prod_{k=0}^{T-1} \pi^{Euler}(x_{k+1}^* | x_k^*, \theta)$, and

$$q(x_{(0,T)}^*, x_T^{*(2)} | x_0, x_T^{(1)}, \theta) = \left(\prod_{k=0}^{T-2} q(x_{k+1}^* | x_k^*, x_k^{(1)}, \theta) \right) \times q(x_k^{*(2)} | x_{k-1}^{*(2)}, x_k^{(1)}, \theta) \quad (3.77)$$

All the above Euler approximations density can be adapted to obtain the diffusion bridge under the partially observed times process with epidemic latent data (incomplete data).

3.7 Observation with measurement error

It is importance to known that in practice observation are often measured with error, that is, one has for all time t_k

$$Y_k = F'X_k + \varepsilon_k, \quad \varepsilon_k | \Sigma \sim N(0, \Sigma_k) \quad (3.78)$$

where $X_k \in \mathbb{R}^{d_1}$ is the observable part of X_{t_k} , Y_k is the measurement of X_k , F is a constant $d \times d_0$ identity matrix, and ε_k is the observation error with zero mean and positive definite covariance matrix Σ_k .

The Σ_k are either assumed known from empirical data, or their estimation is included in the inference procedure for parameter path of the diffusion model. Then, the full joint posterior distribution handle observation errors becomes.

$$\pi(x_{[0,T]}, \theta, \Sigma | x_0, x_T, Y_k) \quad (3.79)$$

For simplicity, if assume that the process is known exactly at time $t = 0$. Suppose there are $m+1$ observation imputed times

$$0 = \tau_0 < \tau_1 < \dots < \tau_{m-1} < \tau_m = T \quad (3.80)$$

Also if we let $\mathfrak{D} = \{k \in (0, \dots, n) | \tau_k \in (\tau_0, \dots, \tau_m)\}$ be the set of indices of observation times, this implies that, observations are assumed to be subject to measurement error, the vector X_k has to be update for $k \in \mathfrak{D}$. The posterior density of the diffusion path $(x_{[0,T]})$ conditional on the observations and the parameter of interest now becomes:

$$\begin{aligned} \pi(x_{[0,T]} | x_0, x_T, y_T, \theta, \Sigma) &\propto \pi(y_T | x_T, \Sigma) \pi(x_T | \theta) \\ &= \prod_{k=0}^{n-1} \phi(y_{k+1} | x_{k+1}, \Sigma) \left(\prod_{k=0}^{m-1} \pi(x_{\tau_{k+1}} | x_{\tau_k}, \theta) \right) \pi(x_T | \theta) \end{aligned} \quad (3.81)$$

And the posterior density of the parameter θ conditioned on both the estimated and observed path also becomes:

$$\pi(\theta | Y_k, x_0, x_T) \propto \pi(\theta) \left(\prod_{k=0}^{n-1} \phi(y_{k+1} | x_{k+1}, \Sigma) \right) \left(\prod_{k=0}^{m-1} \pi(x_{\tau_{k+1}} | x_{\tau_k}, \theta) \right) \quad (3.82)$$

,where, $\pi(\theta)$ is independence priori distribution.

Now, the path proposal distributions in (3.81) have to be adjusted to the new setting. In particular, the observable parts X_k need to be update for $k \in \mathfrak{D}$ in consideration from the previous path update.

The new path update with measurement errors can now computed as:

$$x_{\tau_{k+1}} | x_{\tau_k}, \theta \sim N \left\{ \begin{pmatrix} x_{\tau_k} + \alpha(x_{\tau_k}, \theta) \Delta_\tau \\ F'(x_{\tau_k} + \alpha(x_{\tau_k}, \theta) \Delta_\tau) \end{pmatrix}, \begin{pmatrix} \beta(x_{\tau_k}, \theta) \Delta_\tau & \beta(x_{\tau_k}, \theta) F \Delta_k \\ F' \beta(x_{\tau_k}, \theta) \Delta_k & F' \beta(x_{\tau_k}, \theta) F \Delta_k + \Sigma \end{pmatrix} \right\} \quad (3.83)$$

3.8 Summary of bridge constructs

All the diffusion bridges discussed above are formed using an Euler-Maruyama approximation, taking the form $N(x_{\tau k+1}; x_{\tau k} + \mu(x_{\tau k}, \theta)\Delta\tau, \psi(x_{\tau k}, \theta)\Delta\tau)$ or indeed $\mu^*(x_{\tau k}, \theta)$ and $\psi^*(x_{\tau k}, \theta)$ in the special case of known x_0 and x_T . Summaries of $(\mu(x_{\tau k}, \theta), \psi(x_{\tau k}, \theta))$ and $(\mu^*(x_{\tau k}, \theta), \psi^*(x_{\tau k}, \theta))$ for each construct are presented in Table 3.1 and 3.2 below.

Table 3.2: Summaries of infinitesimal mean and variance for Diffusion Bridge proposal without measurement error

Bridge construct	Infinitesimal Mean	Infinitesimal Variance
EM	$x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_{\tau k}$	$\beta(x_{\tau k}, \theta)\Delta_{\tau k}$
MDB	$x_{\tau k} + \frac{x_T - x_{\tau k}}{T - \tau_k}\Delta_{\tau k}$	$\frac{x_T - x_{\tau k}}{T - \tau_k}\beta(x_{\tau k}, \theta)\Delta_{\tau k}$
LB	$w_k^y(\Delta_{\tau k} + \frac{x_T - x_{\tau k}}{T - \tau_k}\Delta_{\tau k}) + (1 - w_k^y)\alpha(x_{\tau k}, \theta)\Delta_{\tau k}$	$w_k^y\frac{x_T - x_{\tau k}}{T - \tau_k} + (1 - w_k^y)\alpha(x_{\tau k}, \theta)$
GP	$\alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi'(x_T x_t)\mathbf{A}(x_T, -\eta_T)$	$\beta(x_{\tau k}, \theta)$
MGP	$\alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi'(x_T x_t)\mathbf{A}(x_T, -\eta_T)$	$\frac{x_T - x_{\tau k}}{T - \tau_k}\beta(x_{\tau k}, \theta)$

$\mathbf{A} = \{F'\pi(x_T|x_t)\psi(x_T|x_t)\pi'(x_T|x_t)\}^{-1}$

Table 3.3(a): Summaries of Infinitesimal mean of Diffusion Bridge proposal with measurement error

Bridge construct	Infinitesimal Mean
EM	$(x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_{\tau}) + \beta(x_{\tau k}, \theta)\mathbf{F}\Delta_k[F'\beta(x_{\tau k}, \theta)\mathbf{F}\Delta_k + \Sigma]^{-1}y_k - (F'(x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_{\tau}))$
MDB	$x_{\tau k} + \frac{x_T - x_{\tau k}}{\Delta_k}\Delta_{\tau}$
LB	$w_k^y(x_{\tau k} + \frac{x_T - x_{\tau k}}{\Delta_k}\Delta_{\tau}) + (1 - w_k^y)(x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_{\tau}) + \beta(x_{\tau k}, \theta)\mathbf{F}\Delta_k[F'\beta(x_{\tau k}, \theta)\mathbf{F}\Delta_k + \Sigma]^{-1}y_k - (F'(x_{\tau k} + \alpha(x_{\tau k}, \theta)\Delta_{\tau}))$
GP	$\alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi'(y_T x_t)F(F'\pi(y_T x_t)\psi_{\tau k+1}^*(y_T x_t)\pi'(y_T x_t)F\Delta_{\tau k} + \Sigma)^{-1}(y_T - F'\eta\tau)$
MGP	$\alpha(x_{\tau k}, \theta) + \beta(x_{\tau k}, \theta)\pi'(y_T x_t)F(F'\pi(y_T x_t)\psi_{\tau k+1}^*(y_T x_t)\pi'(y_T x_t)F\Delta_{\tau k} + \Sigma)^{-1}(y_T - F'\eta\tau)$

Table 3.3(b) Summaries of Infinitesimal Covariance of Diffusion Bridge proposal with measurement error

$$\psi_{EM}^* = \beta(x_{\tau k}, \theta)\Delta_{\tau} - \beta(x_{\tau k}, \theta)F\Delta_k[F'\beta(x_{\tau k}, \theta)F\Delta_k + \Sigma]^{-1}F'\beta(x_{\tau k}, \theta)\Delta_k$$

Bridge construct	Infinitesimal Covariance
EM	$\beta(x_{\tau k}, \theta)\Delta_{\tau} - \beta(x_{\tau k}, \theta)F\Delta_k[F'\beta(x_{\tau k}, \theta)F\Delta_k + \Sigma]^{-1}F'\beta(x_{\tau k}, \theta)\Delta_k$
MDB	$\left(\frac{\Delta_{k+1}}{\Delta_k}\right)\beta(x_{\tau k}, \theta) - \beta(x_{\tau k}, \theta)F(F'\beta(x_{\tau k}, \theta)F\Delta_{\tau k} + \Sigma)^{-1}F'\beta(x_{\tau k}, \theta)\Delta_{\tau k}$
LB	$\omega_k^y(\beta(x_{\tau k}, \theta) - \beta(x_{\tau k}, \theta)F(F'\beta(x_{\tau k}, \theta)F\Delta_{\tau k}^y + \Sigma)^{-1}F'\beta(x_{\tau k}, \theta)\Delta_{\tau k} + (1 - \omega_k^y)\psi_{EM}^*$
GP	$\beta(x_{\tau k}, \theta)\Delta_{\tau} - \beta(x_{\tau k}, \theta)F\Delta_k[F'\beta(x_{\tau k}, \theta)F\Delta_k + \Sigma]^{-1}F'\beta(x_{\tau k}, \theta)\Delta_k$
MGP	$\left(\frac{\Delta_{k+1}}{\Delta_k}\right)\beta(x_{\tau k}, \theta) - \beta(x_{\tau k}, \theta)F(F'\beta(x_{\tau k}, \theta)F\Delta_{\tau k} + \Sigma)^{-1}F'\beta(x_{\tau k}, \theta)\Delta_{\tau k}$

3.9 Parameter update for diffusion processes

One of the major objectives of this study is to consider an improved method of parameters estimate in diffusion approximation process, especially, when inter-observation times are large or non-equidistant or for multi-dimensional diffusion processes or when some components of the state vector are latent or measured with error. Unfortunately, almost epidemic dataset in the field of epidemiology possess at least one of these properties. In addition, the convergence challenge occurred whenever obtaining parameter estimate under diffusion approximation approach would be our major concerned in this study.

Handling unknown parameters appearing in diffusion coefficient on discrete time's observations had greater challenges that received a lot of attention over the past decade. Because direct implementation of data augmentation result into markov chain that is reducible, especially in multidimensional cases. Here we restrict attention to estimation within the Bayesian inference.

As pointed out by Roberts and Stramer (2001), that the Bayesian data augmentation algorithm degenerates whenever parameters of interest θ appears in the diffusion coefficient,

as the quadratic variation of the full data forces the conditional distribution for the next iteration on θ to be degenerate at the current value. Hence, iterates of θ remain stuck at their initial value. The problem was solved in continuous-time observed setting proposed by Chib et al. (2004).

In this study, we consider situation where data are discretely observed with some components completely unobserved (as latent). This makes unknown parameters estimate more difficult, because less information is available compared to fully observed cases.

The jointly posterior distribution of diffusion process given thus:

$$\pi(\theta, x | D_T) \propto \pi(\theta) \prod_{i=0}^{T-1} \prod_{k=0}^{M-1} \pi(\Delta \tau; x_{i,k+1} | x_{i,k}, \theta) \quad (3.84)$$

A description of the algorithm that shows this jointly posterior distribution of diffusion problems can be found in Algorithm 3.6.

Algorithm 3.6 Traditional parameter estimate for diffusion process

1. Initialise the iteration counter $i = 1$ and initialise the chain with $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})'$, where $\theta^{(0)}$ is chosen from somewhere in the support of $\pi(\theta | D_n)$.
 2. Propose a new value θ^* using the transition kernel $q(\theta^* | \theta^{(i-1)})$.
 3. Evaluate the acceptance probability $\alpha(\theta^*, \theta) = \min \left\{ 1, \frac{\pi(\theta^* | D_T) q(\theta | \theta^*)}{\pi(\theta | D_T) q(\theta^* | \theta)} \right\}$.
 4. Set $\theta^{(i)} = \theta^*$ with probability $\alpha(\theta^*, \theta)$, otherwise set $\theta^{(i)} = \theta^{(i-1)}$.
 5. Set $i = i + 1$ and return to step 2.
-

Challenges:

Algorithm 3.6 for sampling realisation diffusion path becomes degeneracy, especially when numbers of imputed value exceed 5, as explored in the literature. The reason is not beyond serial correlation that exists between missing paths (skeleton) of the diffusion bridge and the parameters θ of the model.

To decouple this dependence that exists between the missing paths of the diffusion realisation sample path and the parameter of the model, we redesign algorithm above to cater for this autocorrelation through the existing innovation scheme proposed by Chib et al (2004).

Chib et al adopted this innovation scheme for continuous-time in their work, but here we modified it into discrete-time observation and named it with Modified innovation scheme.

3.9.1 Modified Innovation Method

The idea here is base on decouple the dependence that exists between the missing paths of the diffusion bridge and the parameters θ of the model. Rather than updating θ conditional on the discretized diffusion bridge, we proposed updating θ conditional on the parameter-free Brownian motion (independent Wiener process) driving the discretized diffusion bridge. Golightly & Wilkinson (2010) seize the general concept of Chib et al. (2004) and applied it to a large class of diffusion processes. Golightly & Wilkinson (2010) concepts never justified the new approach consider in this research work.

Let $W_t = (\Delta w_{\tau_1}, \Delta w_{\tau_2}, \dots, \Delta w_{\tau_{m-1}}, \Delta w_{\tau_{m+1}}, \dots, \Delta w_{\tau_{nm-1}})$ denote the Brownian increment innovations. Following the same approaches conceptually, we assumed that the diffusion coefficient of the model (β) is invertible $\beta = \beta\beta^T$. Since, we are considered the irreversible type of diffusion process. The Back-transformation happen via

$$\begin{aligned} \Delta W_t &= \beta(X_t, \theta)^{-\frac{1}{2}} \{\Delta X_t - \alpha(X_t, \theta)\Delta t\} \quad , w_0 = 0 \\ &= \beta(X_t, \theta)^{-\frac{1}{2}} \left\{ X_{t,k+1} - X_{t,k} - \frac{x_{tk+1} - X_{t,k}}{t_{k+1} - t} \Delta t \right\} \end{aligned} \quad (3.85)$$

$$W_{t,k+1} - W_{t,k} = \beta(X_t, \theta)^{-\frac{1}{2}} \left\{ X_{t,k+1} - X_{t,k} - \frac{x_{tk+1} - X_{t,k}}{t_{k+1} - t} \Delta t \right\} \quad (3.86)$$

The above equation means that there is a deterministic link between the target process X_t and the parameter-free Brownian motion process W_t .

If X_t^* denotes the missing paths given the parameters θ , the main idea consists of finding a function h and a process W_t such that $X_t = h(W_t, \theta)$ for all $t \in [0, T]$.

More precisely, we construct a Markov chain for $(\theta, (w_t)_{1 \leq t \leq n})$, where each w_t is an innovation process corresponding to the bridge X_t^* connecting observation x_{t-1} to x_t .

This function is invertible in its first argument and there is another function h^{-1} such that $h^{-1}(X_t, \theta) = W_t$. The connection between X_t^* and W_t can be exploited in the parameter update of the MCMC scheme.

Then, update θ conditioned on $W_t^{imputed}$, x_0 , and x_T rather than θ conditioned on $X_t^{imputed}$, x_0 , and x_T . These decouple the tight dependence between θ and the diffusion bridge. By Using the connection between X_t and W_t exploited in the parameter update of the MCMC scheme in Algorithm 3.6 above resulting into new entirely modified innovation scheme. A description of the algorithm can now be found in Algorithm 3.7.

Algorithm 3.7 Proposed Modified Innovation parameters estimate for diffusion process

1. Initialise the iteration counter $i = 1$ and initialise the chain with $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})^T$, where $\theta^{(0)}$ is chosen from somewhere in the support of $\pi(\theta|D_n)$, and given x_0, x_T and X^{imp} .
 2. Propose a new value θ^* using the transition kernel $q(\theta^*|\theta^{(i-1)})$.
 3. Compute $W^{imp} = h^{-1}(X^{imp}, \theta)$.
 4. Evaluate the acceptance probability $\alpha(\theta^*, \theta)$

$$= \min \left\{ 1, \frac{\pi(\theta^*|W^{imp}, x_0, x_T) q(\theta|\theta^*) \pi(\theta^*)}{\pi(\theta|W^{imp}, x_0, x_T) q(\theta^*|\theta) \pi(\theta)} \right\}.$$
 4. Set $\theta^{(i)} = \theta^*$ with probability $\alpha(\theta^*, \theta)$, otherwise set $\theta^{(i)} = \theta^{(i-1)}$.
 5. If θ^* was accepted, replace X^{imp} by $X^{imp*} = h(W^{imp}, \theta^*)$.
 6. Set $i = i + 1$ and return to step 2.
-

Note that, the path correction in step 5 is new in comparison with the update scheme from the literature. By conditioning the parameter update on this transformation, the acceptance probability for the parameter update becomes:

$$\begin{aligned}
\alpha(\theta^*, \theta) &= \min \left\{ 1, \frac{\pi(\theta^* | w_0, \dots, w_m) q(\theta | \theta^*) \pi(\theta^*)}{\pi(\theta | w_0, \dots, w_m) q(\theta^* | \theta) \pi(\theta)} \right\} \\
&= \min \left\{ 1, \left(\prod_{k=0}^{m-1} \frac{\pi(w_{k+1} | w_k, \theta^*)}{\pi(w_{k+1} | w_k, \theta)} \right) \frac{q(\theta | \theta^*) \times \pi(\theta^*)}{q(\theta^* | \theta) \times \pi(\theta)} \right\} \\
&= \min \left\{ 1, \left(\prod_{k=0}^{m-1} \frac{J(\pi(w_{k+1}, \theta^*))}{J(\pi(w_{k+1}, \theta))} \right) \left(\prod_{k=0}^{m-1} \frac{\pi(X_{k+1}^* | X_k^*, \theta^*)}{\pi(X_{k+1} | X_k, \theta)} \right) \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta | \theta^*)}{q(\theta^* | \theta)} \right\} \quad (3.87)
\end{aligned}$$

where, $X_0^* = x_0$, $X_k^* = h(W_k, \theta)$ for $k = 1, \dots, m$ and

$$J(h(w_{k+1}, \theta)) = \left| \frac{\partial h(w_{k+1}, \theta)}{\partial w_{k+1}} \right| \text{ the Jacobian determinant of } h.$$

Here, the acceptance probability nominators and denominators differ in both parameter and sample path. That is, the critical situation where parameters and sample path are correlated should not occur, specifically if the imputed value becomes large $m > 5$ or interval close to zero $\Delta \rightarrow 0$.

3.10 R package

All algorithms used and sampling of target distribution on estimator for our propose methods would be implemented in R package. R is a general purpose statistical package used in many fields of research. This package includes support for a wide variety of modern statistical and graphical methods.

3.11 Discussion

In this Chapter we present a general framework for Derivation of epidemic models in exploits diffusion processes and Bayesian augmentation method of estimation of discretely observed diffusion processes, those satisfactory deals with both aforementioned challenges (Low acceptance probability, the autocorrelation exist between the Diffusion Bridge and parameters of the model that lead to degeneracy of the Markov chain).

We proposed the posterior distribution of the Euler-Maruyama approximation to diffusions and design algorithms for easier practical computation. These algorithms can be applied to a broad class of multidimensional diffusion processes which are observed (with or without error) only at discrete time points. Our new approaches (Diffusion Bridge sampler and Modified Innovation Scheme) avoid the need for re-parameterisation techniques that are not always available and/or justified. With these viable approaches, practitioners have the option to choose the algorithm with the best convergence and mixing properties for their model.

Chapter Four

Analysis and Results

4.0 Simulation Study

As part of effort towards achieving tangible methods of parameter estimate using diffusion bridge samplers, the simulation results served as a template to understand the behaviour of the proposed methods under different data augmentation scenario. To validate the results of

the simulation study, real-life datasets on the outbreak of smallpox and Ebola diseases were studied.

In all our results, the initial samples obtained before convergence reached in all the output results are discarded as burn-in period. To ensure clear understanding of the results after converged reached, we presenting output results through visual displace.

4.1 Behaviour of Diffusion epidemic Model

(a) Standard Deterministic SIR and Diffusion Approximation SIR Epidemic Model

Figure 4.1 (a) shows evolution of numbers of susceptible, infective and removed individuals in the standard deterministic ODEs SIR model (1.5) for initial value $\{S_0, I_0\} = \{999, 1\}$ with true parameter values transmission rate $\beta = 0.330$, recovery rate $\gamma = 0.178$ and population size $N = 1000$. (b) shows evolution of fractions of susceptible, infective and removed individuals in the SDEs diffusion approximate SIR epidemic model derived in (3.12) that described in the discrete-time observed stochastic differential equation for initial value $\{s_0, i_0\} = \{1, 0.01\}$ having parameters (transmission rate $\beta = 0.330$, recovery rate $\gamma = 0.178$ and population size $N = 100$).

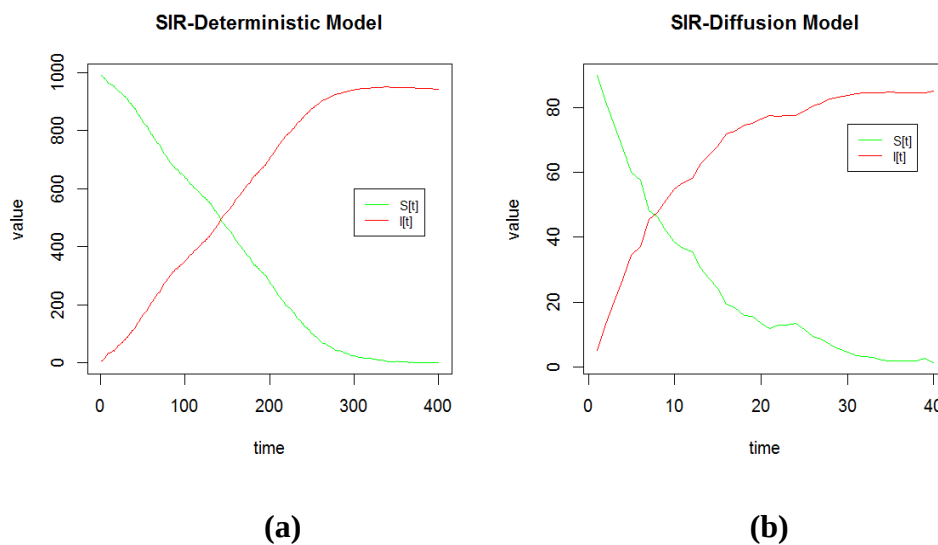


Figure 4.1 show evolution of (a) number of SIR, (b) fraction of SIR epidemic model

(b) Standard Deterministic SEIR and Diffusion Approximation SEIR Epidemic Model

Figure 4.2 (a) shows evolution of numbers of susceptible, infective and removed individuals in the standard deterministic ODEs SEIR model (1.6) for initial value $\{S_0, E_0, I_0\} = \{999, 0, 1\}$ with true parameter values transmission rate $\beta = 0.330$, Infection rate $k = 0.160$, recovery rate $\gamma = 0.200$ and population size $N = 1000$. (b) shows evolution of fractions of susceptible, infective and removed individuals in the SDEs diffusion approximate SIR epidemic model derived in (3.12) that described in the discrete-time observed stochastic differential equation that having parameters (transmission rate $\beta = 0.330$, Infection rate $k = 0.160$, recovery rate $\gamma = 0.200$ and population size $N = 100$).

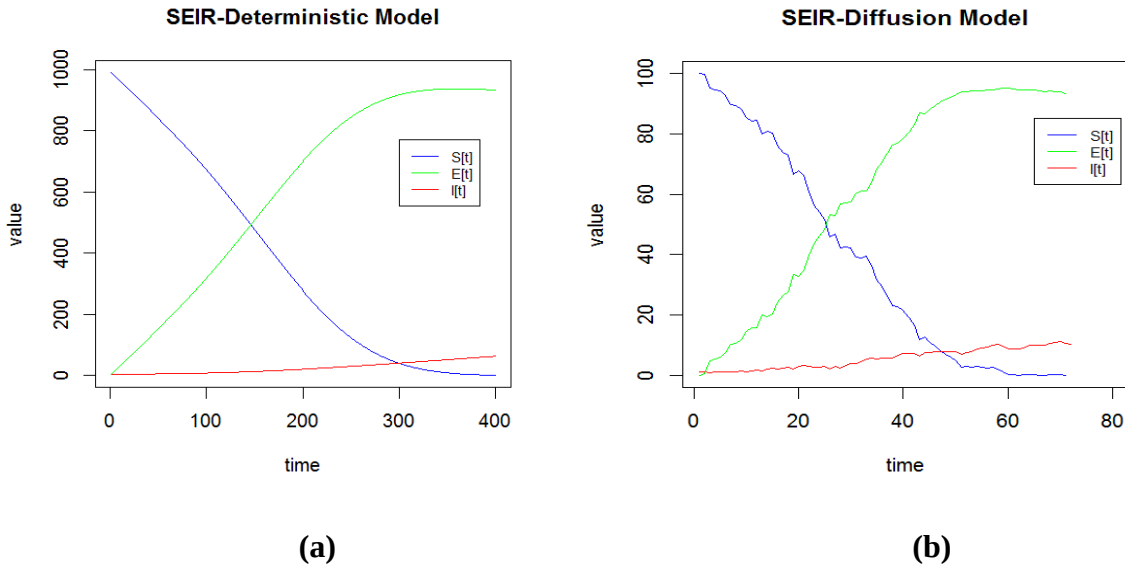
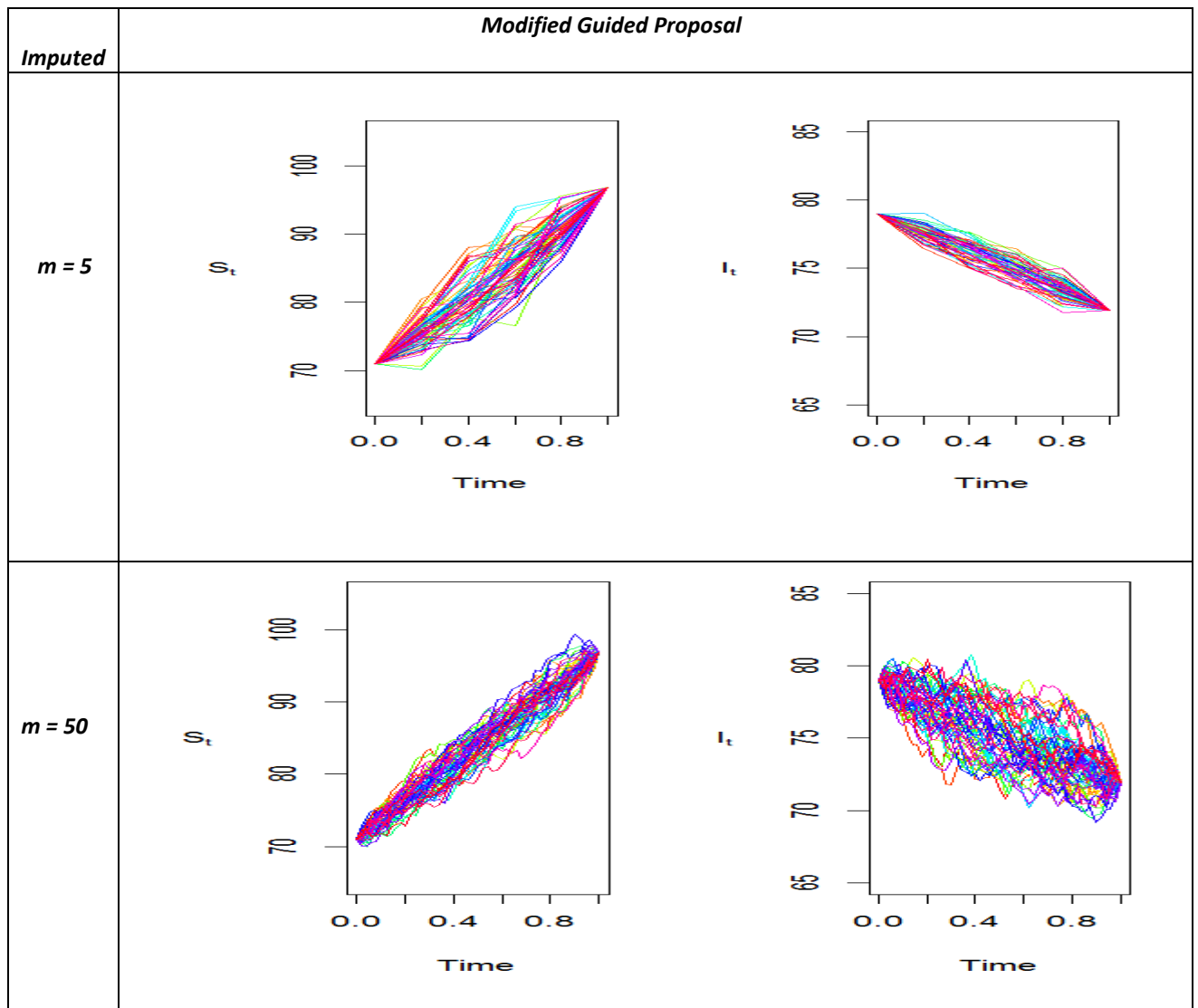


Figure 4.2 shows the evolution of (a) ODEs and (b) SDEs (diffusion process) trajectories. The evolution of the diffusion approximation epidemic models trajectories in figures 4.1 and 4.2 above mirrored the exact dynamic epidemic models in ordinary differential equations.

4.2 Diffusion Bridge for Modified Guided Proposal

We present the output result of newly improved diffusion bridge mechanism (Modified Guided proposal) using block update approach in section 3.6.2 on stochastic diffusion SIR epidemic model derived in section 3.1.1. We generate 100 path proposals, fulfilling the SDE

initial conditioned of $x_0 = (71,79)'$ and end – point $x_T = (96,71)'$ with time interval $[0,1]$, that imputed with subintervals of $m = (5, 50, 200)$ on the model parameter value $\theta = (0.330, 0.178)$. Diffusion bridge realisation sample paths for three different imputed values were show in the figure 4.1 and the acceptance probability for MCMC convergence are 0.85, 0.83 and 0.79 meaning that 85.0%, 83.0% and 79.0% acceptance rate of every 1000 generated paths respectively would be kept for the three different imputed values used.



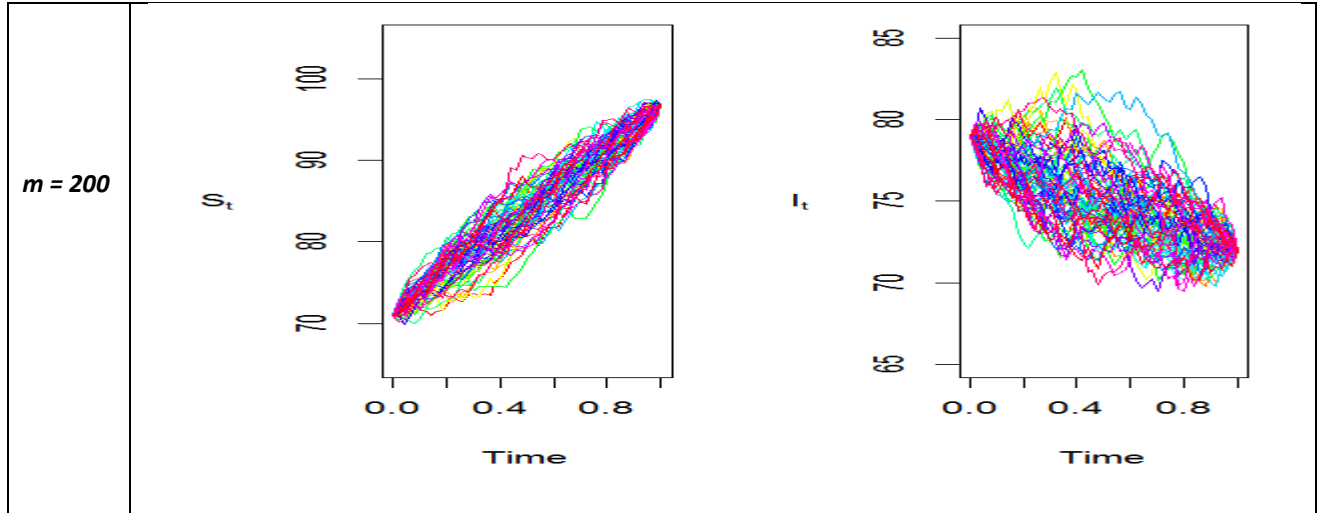
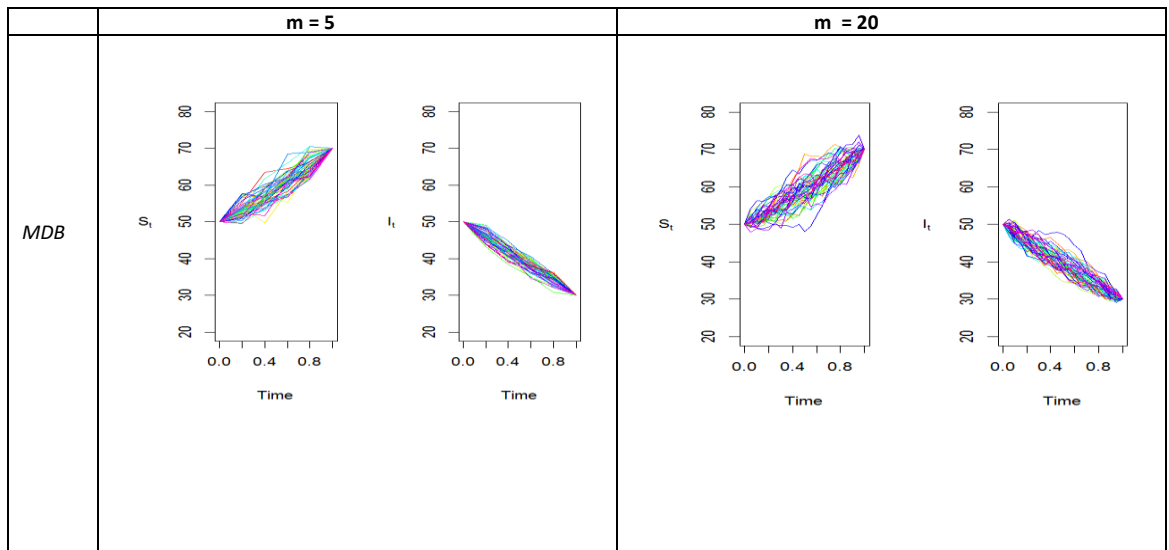


Figure 4.3 10000 path proposals for the SIR-Epidemic model on $(\beta, \gamma)^T = (0.330, 0.178)$ with time interval $[0,1]$ fulfilling the SIR diffusion epidemic conditioned on $(s_0, i_0)^T = (71, 79)^T$ and $(s_T, i_T) = (96, 71)$ imputed, with subintervals of $m = (5, 50, 200)$ generated using Modified Guided diffusion Bridges (MGP).

4.2.1 Diffusion Bridges comparison performance

We compare the accuracy and efficiency of the improved bridge sampler (Modified Guided Proposal, (MGP)) with the two different existing bridges: Modified Diffusion Bridge (MDB), and Guided Proposal (GP). The results of accepted probability for fully observed and partially observed were depicted in table 4.1 and in the figures 4.5 and 4.6 after we extracted the trajectories of sample paths for two different augmented data $m = 5$ and 20 in figure 4.4 below.



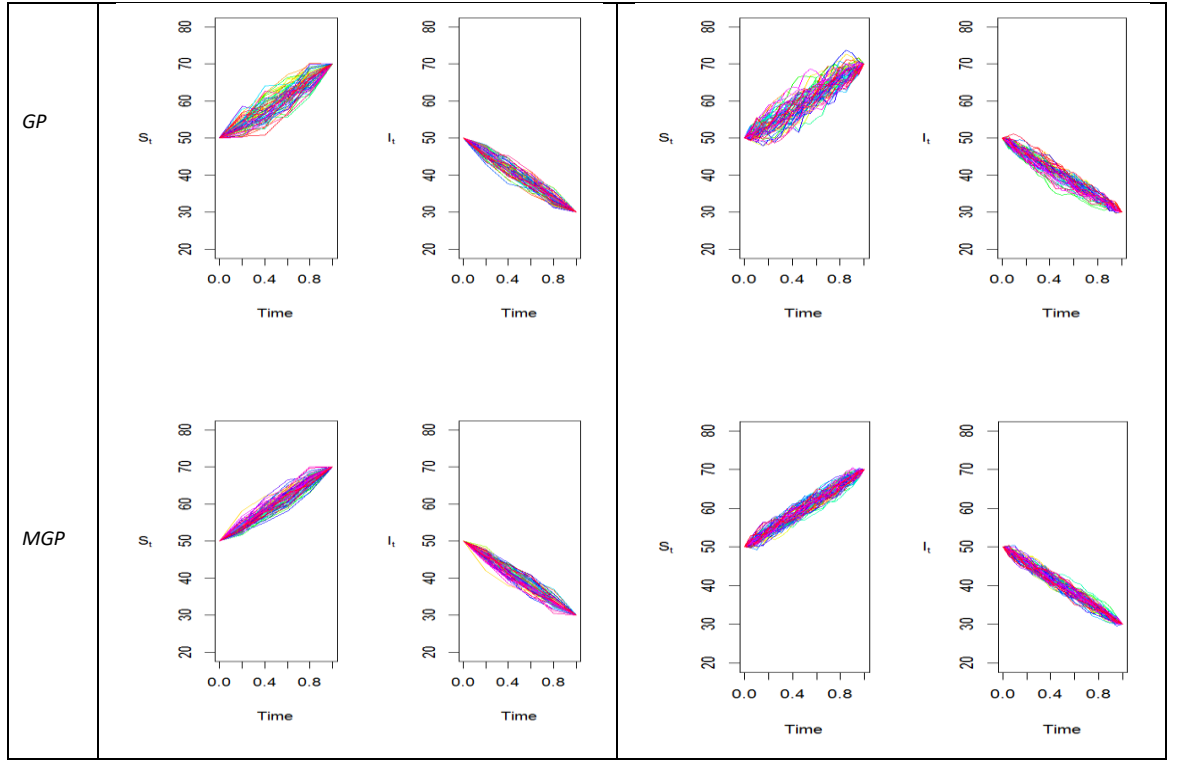


Figure 4.4 shows 10000 path proposals for the SIR-Epidemic diffusion model using $(\beta, \gamma) = (0.33, 0.178)$ with time intervals $[0, 1]$, fulfilling the SDE initial conditioned of $s_0, i_0 = (50, 50)$ & $(s_T, i_T = (70, 30))'$ imputed, with subintervals of $m = (5, 20)$ generated from three different Diffusion Bridges (MDB, GP, and MGP).

The newly improved bridge sampler is based on the simple construction of a process that starts from initial conditioned x_0 at time $t = 0$ and ends in x_T at time $T = 1$, where $x_0 = (s_0, i_0) = (50, 50)$ and the final point $x_T = (s_T, i_T) = (70, 30)$.

The quality of any bridge sampler is determined by the acceptance probability/rate of the proposed candidate. We assessing statistically efficiency by means acceptance probabilities as a proxy for overall efficiency of each bridge, with high probability preferred.

Table 4.1 Shows acceptance probabilities for fully & partially observed data based on a run of 1000 MCMC iterations schemes

Scheme	m = 2		m = 5		m = 20		m = 50	
	Fully observed	Partially observed	Fully observed	Partially observed	Fully observed	Partially observed	Fully observed	Partially observed
MDB	0.672	0.500	0.603	0.443	0.501	0.321	0.304	0.320
LB	0.546	0.421	0.502	0.322	0.403	0.300	0.230	0.200
GP	0.689	0.634	0.650	0.521	0.606	0.389	0.500	0.350
MGP	0.777	0.652	0.750	0.599	0.743	0.504	0.742	0.505

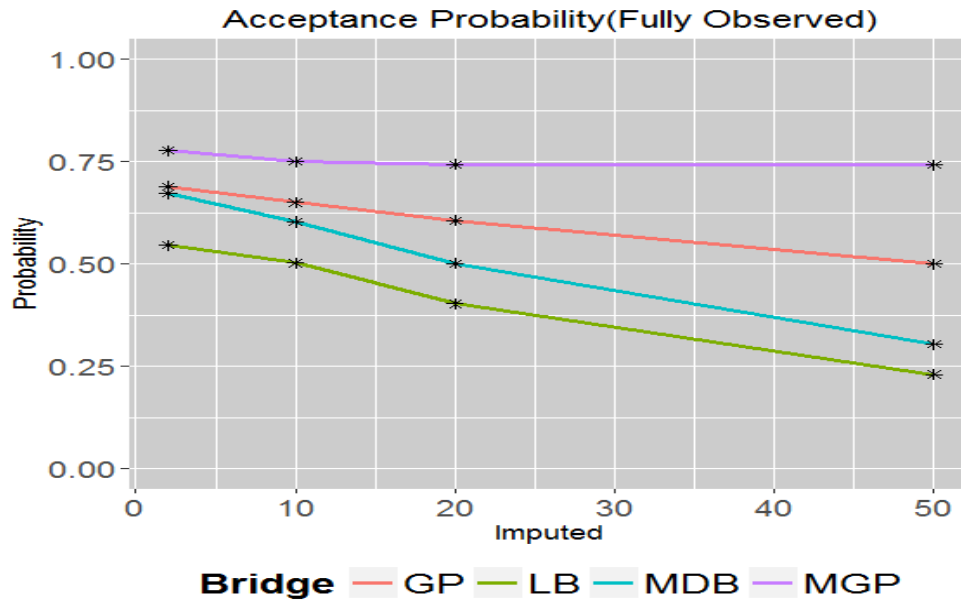


Figure 4.5 shown acceptance probabilities plot for fully observed based on a run of 1000 MCMC iterations for four different proposal schemes.

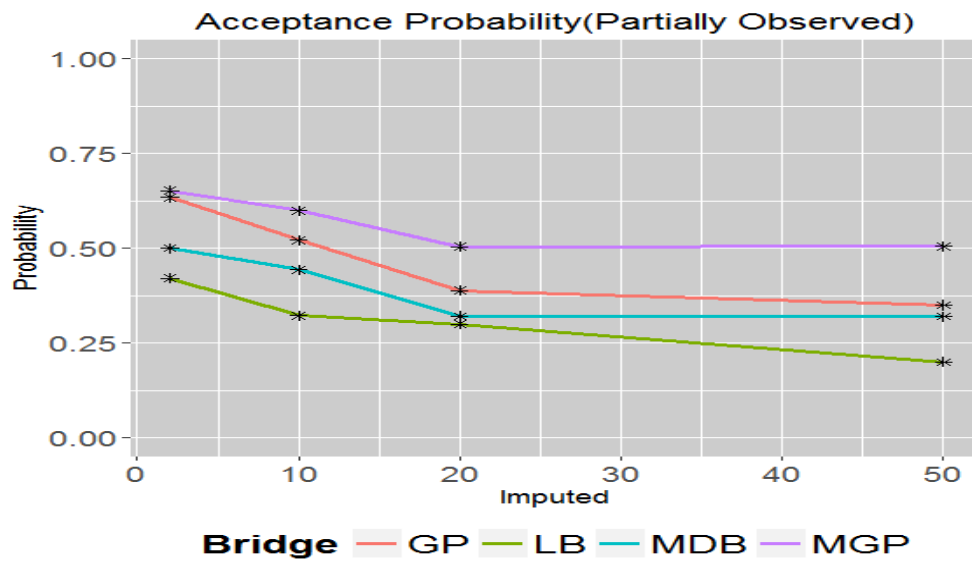


Figure 4.6 shown acceptance probabilities plot for partially observed based on a run of 1000 MCMC iterations for four different proposal schemes.

The results from table 4.1 and figures 4.5 and 4.6 show the efficient performance of the propose bridge together with three other different diffusion bridges. Results from fully observed and partially observed data from multidimensional SIR epidemic diffusion model show that the newly/improved proposed diffusion bridge under study performed better.

4.3 Parameter Estimation

Here we apply the proposed method of parameter estimation explained in (3.87) on the diffusion SIR- epidemic model derived in (3.1.1), which adapted the likelihood of (3.24) under the modified innovation algorithm (3.84) designed in the previous chapter.

To start with, both the fraction of susceptibles (s) and the fraction of infections (i) are assumed known at the beginning of the outbreak of epidemic, observed at time $t = 0$. In particular, the parameter update is performed according to Algorithm (3.7) simultaneously with path update, sampling with the realisation of diffusion bridges. Here, we present evidence on the performance of estimator of $\theta = (\beta, \gamma)$ of SIR-epidemic model, initialised with true parameters set at random $0.001 \leq \beta < 0.850$ and $0.010 \leq \gamma < 0.650$ initial value, that follow standard uniform distribution having tuning parameters $c = 0.5$ and $r = 0.5$ respectively, this was observed within time interval of $[0, 199]$ with length 0.001. We specified initial conditioned as fraction of susceptible and fraction of infectious ($s_0 = 0.99$, $i_0 = 0.01$).

Under sample paths we considered four (4) different augmented data point in between every time point as $m \in \{2, 5, 15 \text{ and } 50\}$ data point subintervals intermediate between every two observation times, run under MCMC sampling over 100000 iterations. We considered 500 first iterations discarded as a burn-in periods. A priori beliefs were specified independently for the two parameters of interest using gamma distribution: $\beta \sim \text{Gamma}(10, 20)$ and $\gamma \sim \text{Gamma}(10, 20)$ for the transmission rate and recovery rate respectively.

The new parameter candidates β^* and γ^* were drawn according to Gaussian random walk proposal density:

$$\beta^* \sim N(\beta, c = 0.09) \text{ and } \gamma^* \sim N(\gamma, r = 0.09)$$

where c and r represent the specified tuning parameters for the variance adopted after several attempts to produce satisfactory results.

We believed with a priori probability having 95% HPDR region that the true parameters lie on as follow regions: $(0.001 \leq \beta \leq 0.800)$ and $(0.010 \leq \gamma \leq 0.600)$ for the parameters of interest respectively.

The Table 4.2 contains summarise of the posterior means (estimated values), 95% posterior credible interval (PCI) and acceptance probability (AP) of each parameter of interest (β, γ) .

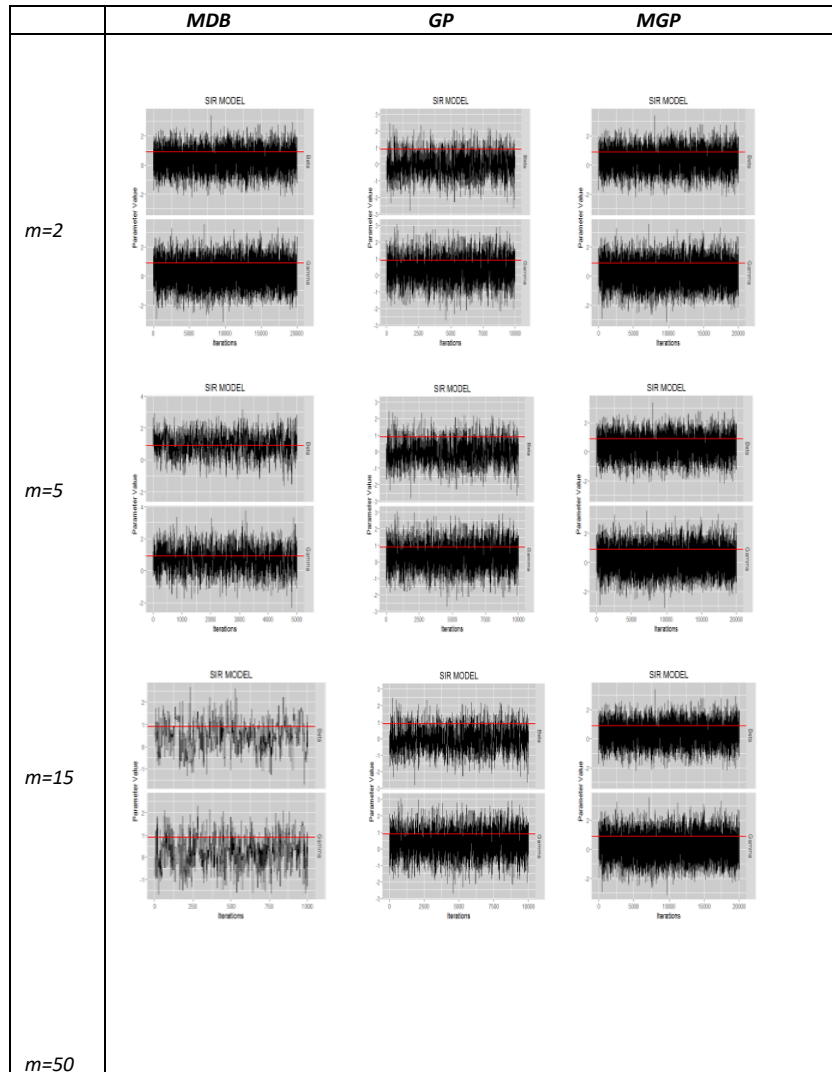
Similarly, figure 4.7 to 4.9 contains corresponding visual presentation for these posterior estimated values trace plots, density plots and autocorrelation function respectively.

Table 4.2 Shows parameter estimates for posterior means and posterior credible intervals for varying number of imputed values ($m = 2, 5, 15$ and 50) and their acceptance probability (AP).

	Proposal	β	γ	AP
m = 2	MGP	0.349(0.200,0.467)	0.189(0.101,0.280)	0.90
	MDB	0.350(0.001,0.768)	0.230(0.065,0.500)	0.66
	GP	0.362(0.003,0.658)	0.200(0.070,0.450)	0.68
m = 5	MGP	0.339(0.301,0.407)	0.186(0.099,0.280)	0.87
	MDB	0.375(0.001,0.785)	0.183(0.020,0.550)	0.60
	GP	0.398(0.020,0.645)	0.177(0.015,0.550)	0.65
m = 15	MGP	0.340(0.219,0.597)	0.179(0.100,0.209)	0.83
	MDB	0.394(0.007,0.700)	0.199(0.014,0.777)	0.53
	GP	0.440(0.003,0.608)	0.241(0.014,0.550)	0.58
m = 50	MGP	0.340(0.202,0.467)	0.158(0.103,0.189)	0.77
	MDB	0.666(0.031,0.708)	0.200(0.016,0.478)	0.42
	GP	0.400(0.033,0.698)	0.176(0.015,0.540)	0.50

These results were presented for comparison. The comparison is intended to highlight the inherent problems of degeneracy of the existing methods (MDB and GP) and the proposed method (MGP). Moreover, the posterior density plot in figure (4.7) had shown that running

each augmented (m) MCMC scheme until convergence produced estimate which are consistent with one another.



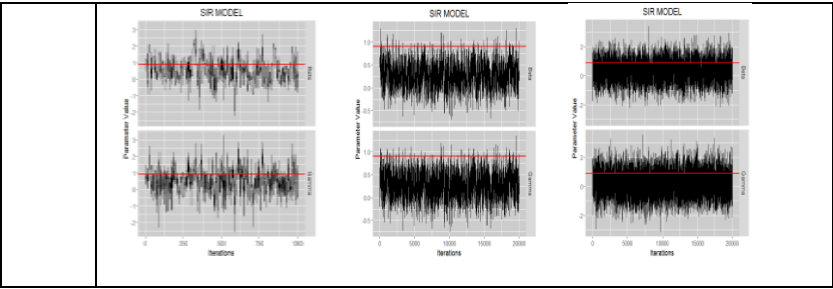
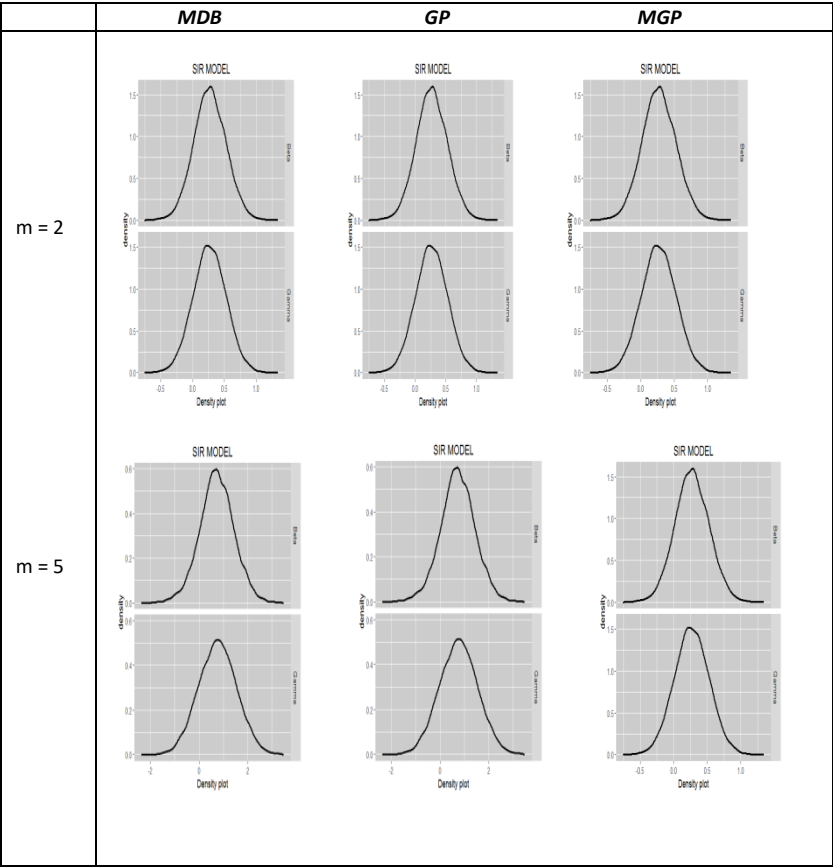


Figure 4.7 Show the trace plots for the parameter of interest $\{\theta = (\beta, \gamma)'\}$ of SIR-epidemic model. The true parameter value for the parameters was indicated by the red horizontal line.



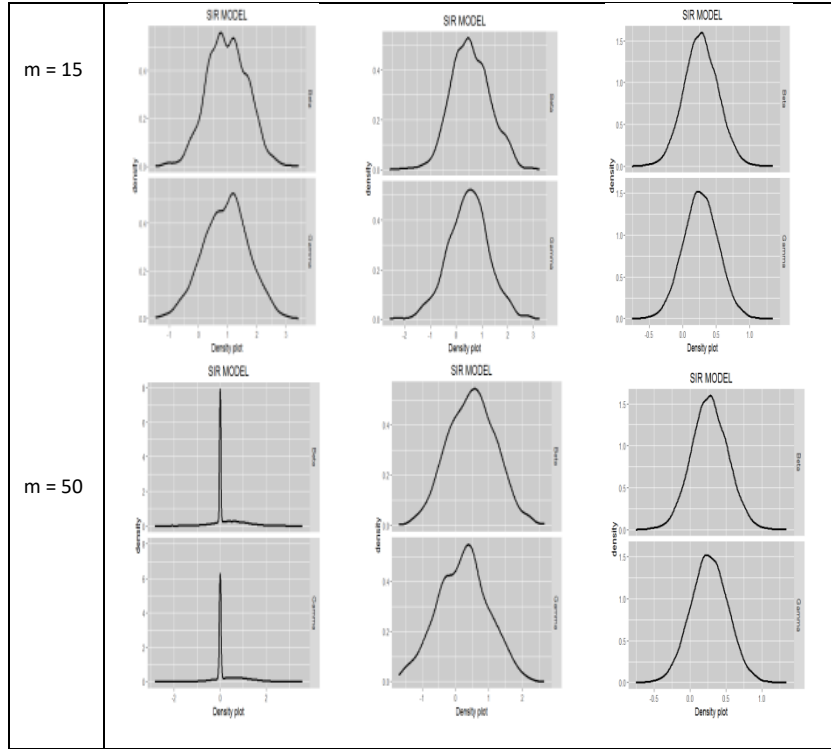
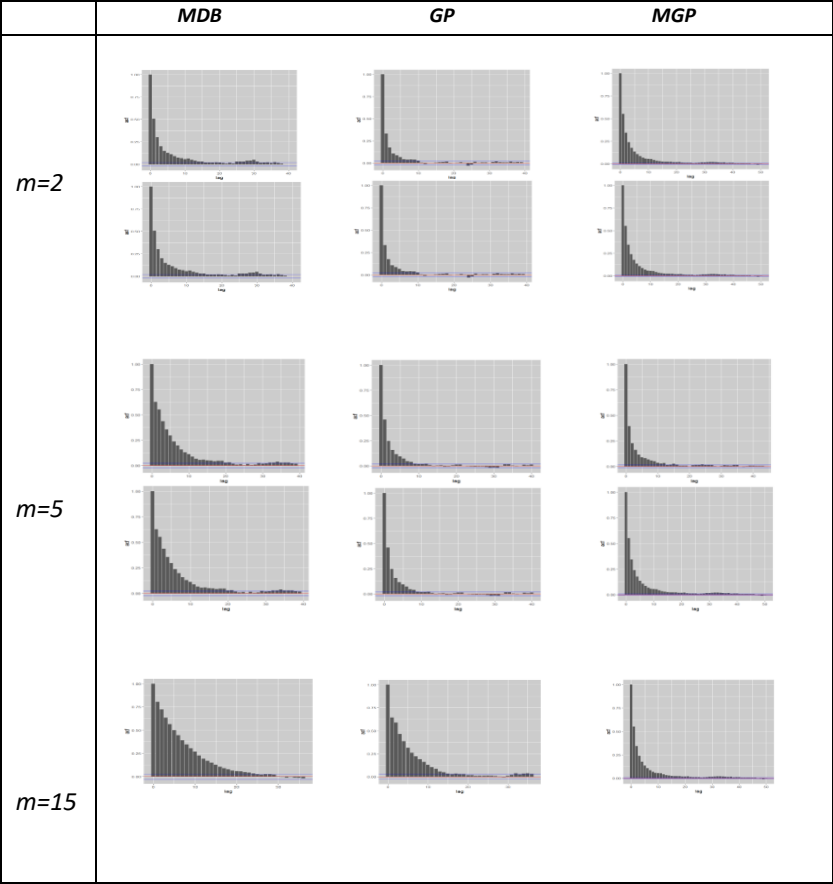


Figure 4.8 Show estimation of the posterior density transmission rate (β) and recovery rate (γ). Density estimation takes into account the full Markov chain, i.e. without thinning, after having discarded a 500 iterations as burn-in periods.

From figure (4.8), the density plot of the two parameters of interest showed that there were notable decreases in discretisation bias when the number of augmented value (m) increases under the proposed method (MGP). The little difference from the results of $m = 2$ and $m = 5$ from the rest suggested that, little behaviour would be gained by running each scheme under the imputed value of $m < 5$.

Another useful way to assessing the relative performance of MCMC scheme is to study the sample auto-correlation functions for each parameter. The serials correlations for the parameters of interest (β, γ) having varying augmented value m were showed in figure 4.9. Unlike other two existing methods used in the above evaluation, it can be seen vividly that as number of augmented value m increases, the proposed method, revealed that the auto-correlation die down very quickly.



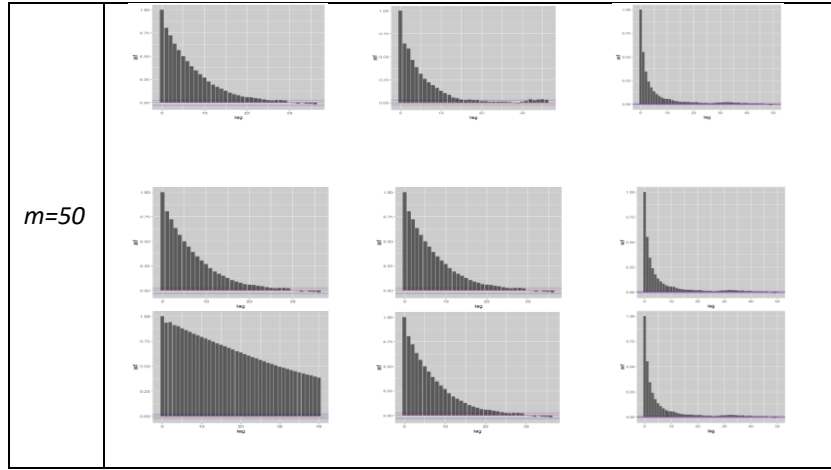


Figure 4.9 Shows Autocorrelation plots of the two parameter of interest β and γ . The Calculation of the autocorrelation takes into account the full Markov chain, i.e. without thinning, after having discarded a 500 burn-in periods.

4.3.1 Simulation study on SEIR Diffusion Epidemic Model

Diffusion SEIR epidemic model have been derived in section 3.1.2 of chapter three. Here we adapted this diffusion SEIR epidemic model in (3.10) to illustrate the proposed parameter estimate method. Just like in the previous illustration, we set the imputed value $m \in \{2, 5, 15 \text{ and } 50\}$ intermediate subintervals in between every two observation times were run for 100000 iterations, and the first 500 iterations were discarded as burn-in periods. A priori beliefs were specified for each parameter, adopting independent gamma prior $\beta \sim \text{Gamma}(10, 20)$, $\gamma \sim \text{Gamma}(10, 20)$ and $\alpha \sim \text{Gamma}(10, 20)$ for the transmission rate β , infectious rate γ and the recovery rate α respectively, having fraction number of initial conditional ($s_0 = 0.90$, $e_0 = 0.09$, and $i_0 = 0.01$).

The new candidates update β^*, γ^* and α^* were drawn according to independent Gaussian random walk proposal density.

$$\beta^* \sim N(\beta, c = 0.006), \gamma^* \sim N(\gamma, r = 0.005) \text{ and } \alpha^* \sim N(\alpha, k = 0.009).$$

where c, r and k represent the specified tuning parameters for each variance of the set parameters of interest adopted after several attempts to produce satisfactory results.

We believed with a prior probability having 95% HPDR that the true parameter lies between the set regions as follow: $(0.001 \leq \beta \leq 0.800)$, $(0.002 \leq \gamma \leq 0.750)$ and $(0.015 \leq \alpha \leq 0.650)$ for the parameters of interest respectively.

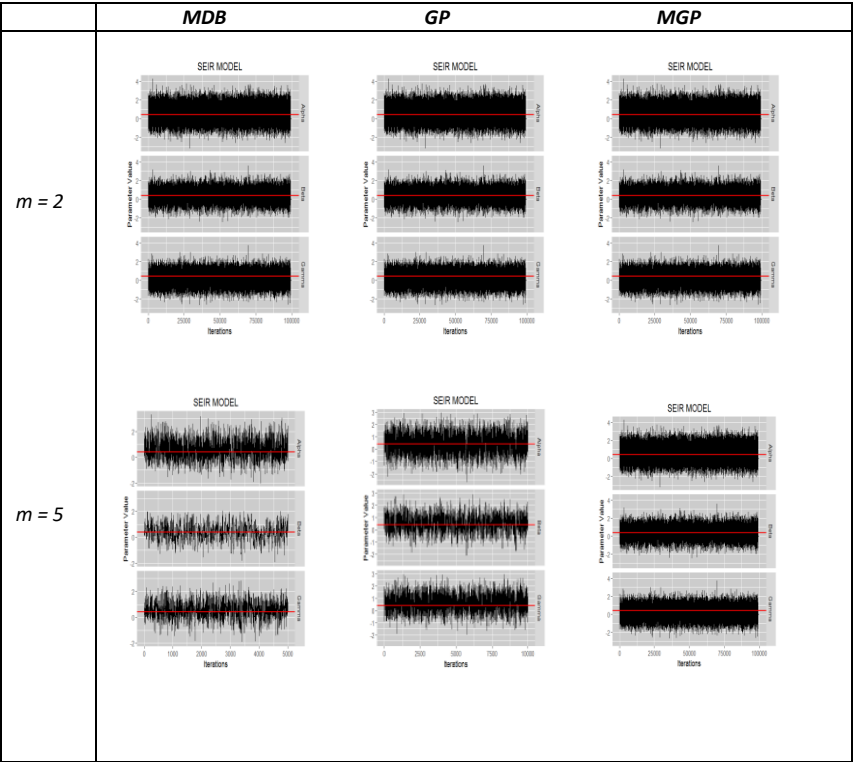
Table 4.3 contains summarise of the posterior means, 95% posterior credible interval and acceptance probability for these parameter of interests (β, γ and α).

Figure 4.10 to 4.12 contains corresponding posterior estimates trace plots, density plots and autocorrelation function for the three parameters β, γ and α respectively.

Table 4.3 Show parameter estimate for posterior means and 95% posterior credible intervals (PCI) of different number of imputed values ($m = 2, 5, 15$ and 50) and their acceptance probability (AP).

	Proposal	β	γ	α	AP
m = 2	MGP	0.649(0.250,0.670)	0.298(0.200,0.580)	0.200(0.199,0.490)	0.78
	MDB	0.650(0.073,0.780)	0.200(0.002,0.650)	0.299(0.017,0.600)	0.60
	GP	0.662(0.002,0.800)	0.190(0.051,0.690)	0.250(0.015,0.650)	0.66
m = 5	MGP	0.634(0.219,0.687)	0.299(0.210,0.585)	0.207(0.196,0.300)	0.73
	MDB	0.675(0.016,0.788)	0.197(0.039,0.660)	0.256(0.019,0.570)	0.56
	GP	0.562(0.002,0.800)	0.200(0.065,0.697)	0.198(0.025,0.650)	0.60
m = 15	MGP	0.640(0.119,0.657)	0.289(0.211,0.580)	0.209(0.193,0.310)	0.70
	MDB	0.665(0.017,0.768)	0.199(0.054,0.670)	0.304(0.052,0.650)	0.52
	GP	0.602(0.030,0.758)	0.201(0.050,0.650)	0.208(0.022,0.640)	0.56

m = 50	MGP	0.600(0.120,0.688)	0.280(0.217,0.583)	0.221(0.172,0.350)	0.70
	MDB	0.603(0.010,0.766)	0.201(0.040,0.691)	0.312(0.024,0.650)	0.30
	GP	0.559(0.032,0.758)	0.202(0.049,0.650)	0.309(0.015,0.650)	0.44



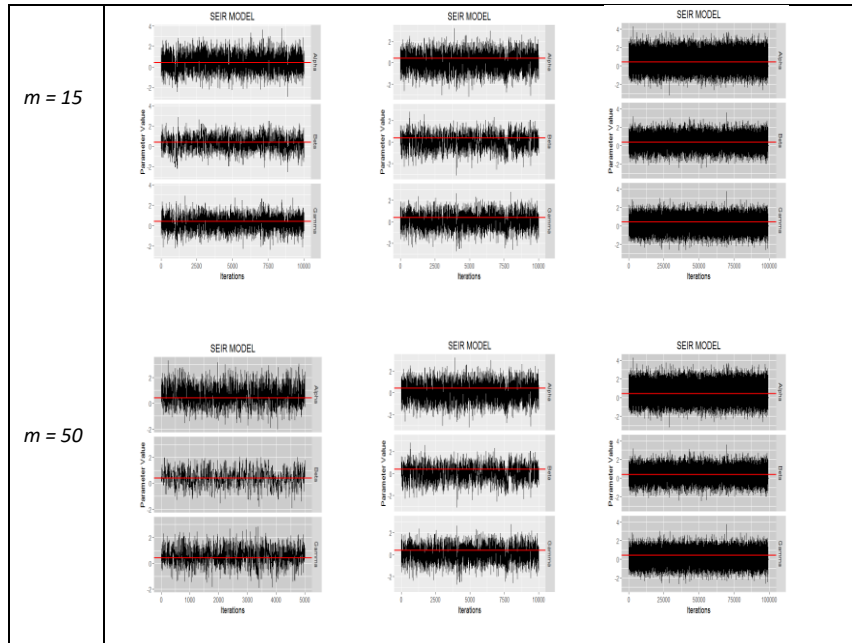
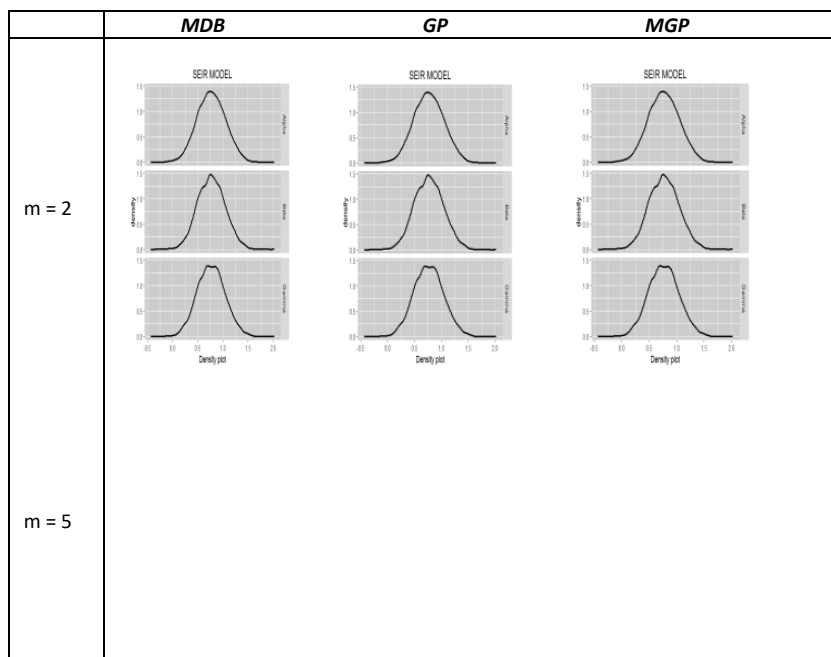


Figure 4.10 Shows the trace plots for the parameter of interest β , γ and α of diffusion SEIR-epidemic model. The true parameter value for the parameters was indicated by the red horizontal line.



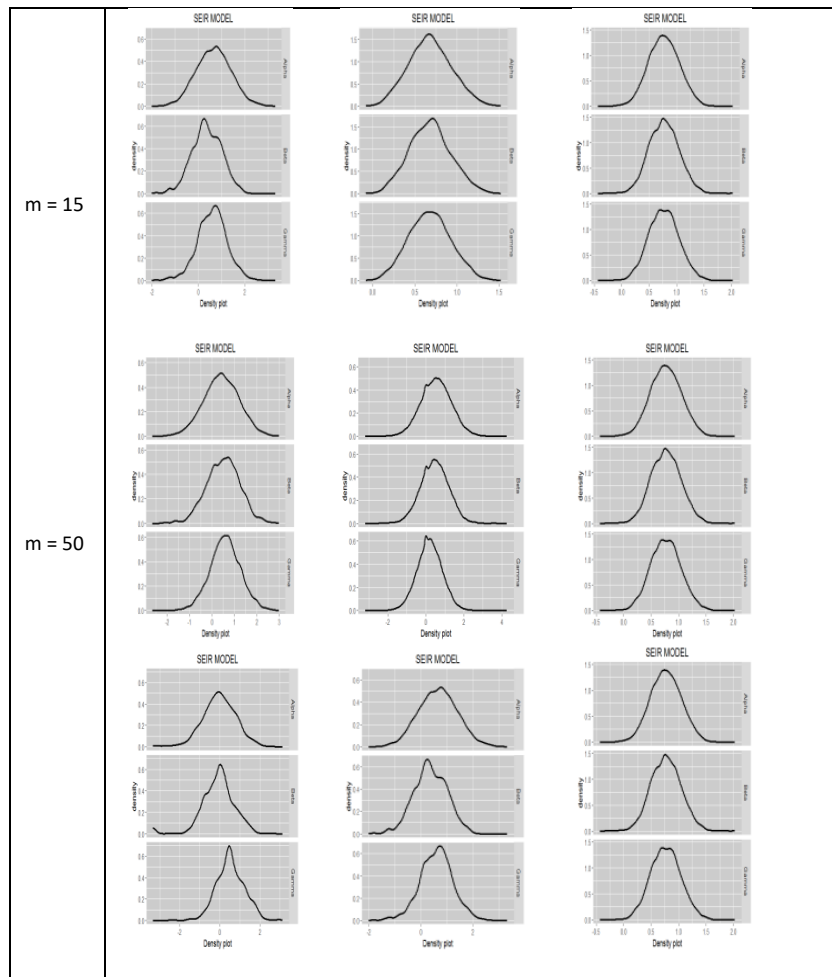
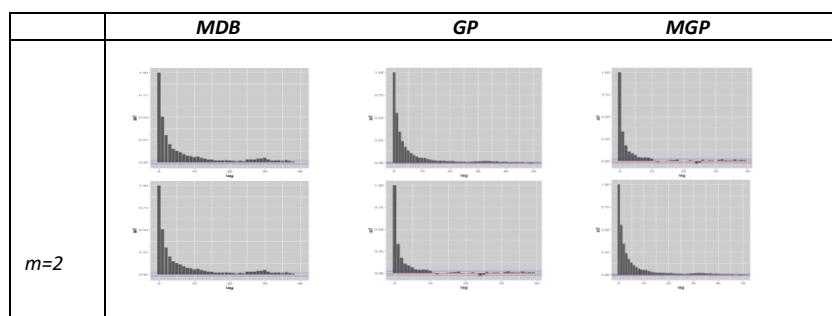


Figure 4.11 Shows estimation of the posterior density $(\beta, \gamma, \alpha)'$. Density estimation takes into account the full Markov chain, i.e. without thinning, after having discarded a 500 burn-in periods.



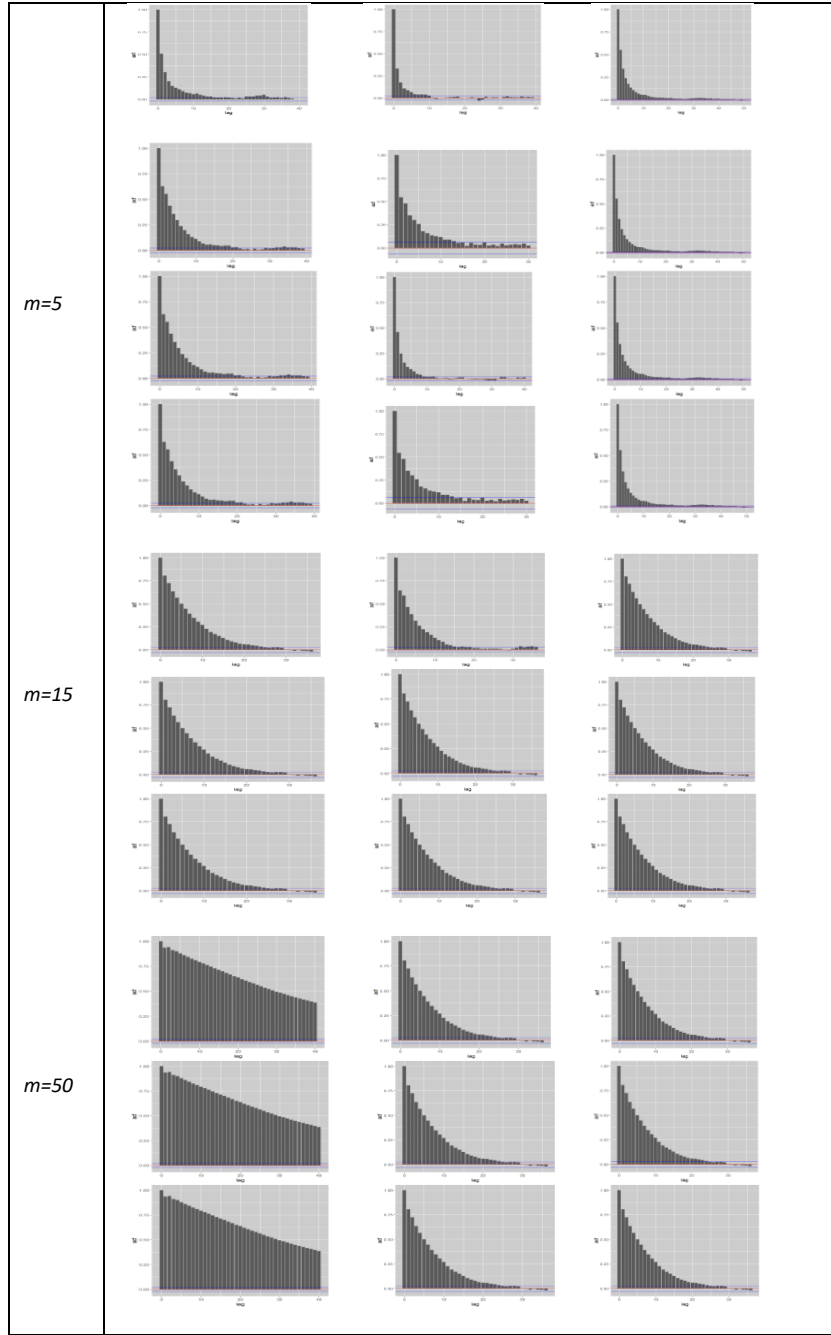


Figure 4.12 Show autocorrelation plots of $(\beta, \gamma, \alpha)'$. The autocorrelation takes into account the full Markov chain, i.e. without thinning, after having discarded a 500 burn-in periods.

4.4 Application to real life datasets

4.4.1 Application I: Abakaliki smallpox data

To ensure that the proposed methods feasibly work well in any dynamics epidemic models, we applied the proposed methods into an outbreak of smallpox infectious disease, which data was collected from the outbreak disease in the commercial town of Abakaliki, Easter part of Nigeria.

In 1967, an outbreak of smallpox occurred in the Nigerian town of Abakaliki. The vast majority of cases were members of the Faith Tabernacle Church (FTC), a religious organisation whose members refused vaccination. A World Health Organization (WHO) report (Thompson and Foege, 1968) describes the outbreak, with information on not only the time series of case detections but also their place of dwelling (compound), vaccination status, and FTC membership. The outbreak has inherent historical interest as it occurred during the WHO smallpox eradication programme initiated in 1959. The outbreak is described in detail in Thompson and Foege (1968) and Eichner and Dietz (2003).

Epidemic data: there were 32 infectious cases in total, 30 of which were FTC members. All of the infected individuals lived in compounds, which were typically one-storey dwellings built around a central courtyard, and capable of housing several families. The FTC members frequently visited one another and were somewhat isolated from the rest of the community. Although FTC members refused vaccination, many of them had been vaccinated prior to joining FTC as described in table 4.4 below.

Table 4.4 Show outbreak of Smallpox cases in Abakaliki, Nigeria during 1967, taken from Thompson and Foege (1968). Compounds listed are those before the move of cases 7 and 8, and 2 other uninfected individuals, on day 25 from compound 1 to compound 2.

Case number	Day of onset of rash	Compound	Confession	Vaccination
-------------	----------------------	----------	------------	-------------

0	0	1	FTC	No
1	13	1	FTC	No
2	20	1	FTC	No
3	22	1	FTC	No
4	25	1	FTC	No
5	25	1	FTC	No
6	25	1	FTC	No
7	26	2	FTC	Yes
8	30	2	FTC	Yes
9	35	1	FTC	No
10	38	4	FTC	No
11	40	5	FTC	No
12	40	1	FTC	No
13	42	1	FTC	No
14	42	1	FTC	No
15	47	1	FTC	No
16	50	5	FTC	No
17	51	2	FTC	No
18	55	1	FTC	No
19	55	2	FTC	No
20	56	6	Non	Yes
21	56	5	FTC	No
22	57	2	FTC	No
23	58	7	FTC	No
24	60	4	FTC	No
25	60	2	FTC	No
26	61	2	FTC	No
27	63	8	Non	Yes
28	66	3	FTC	No
29	66	9	FTC	No
30	71	5	FTC	No
31	76	2	FTC	Yes

FTC = Faith Tabernacle Church

Table 4.4 contains details of the 32 infectious cases of smallpox disease recorded during the outbreak, specifying the date of onset of rash, compound identifier, FTC membership status and vaccination status.

We applied this epidemic datasets (Abakaliki smallpox data) on the proposed methods, through the adopted modified innovation method of parameters estimate in algorithm (3.7) and the proposed modified guided proposal we reached reasonable output results. To the best of my knowledge, the only previous analysis of the full Abakaliki data is that of Eichner and Dietz (2003), where the authors define a stochastic individual-based transmission model that considers not only the case detection times but also the other aspects of the data.

We assume that the epidemic had initiated by a single exposed individual $e_0 = 1$ at time $t = 0$. In addition, FTC individuals have contacts at rate per day with other FTC individuals and contacts at rate per day with anybody (including FTC individuals) in the population. This assumption is made to ensure that all individuals have the same average number of contacts per day outside of their own compound.

Data were simulated from the diffusion SEIR epidemic model (3.16), having initialised parameters of random set $0 \leq \beta \leq 1, 0 \leq \gamma \leq 1$ and $0 \leq \alpha \leq 1$ at the initial value. Every 100th point was recorded to give a sample path of 32 observations on the interval $[0, 76]$.

We set the imputed value for a week ($m = 7days$) and three weeks ($m = 21days$) intermediate subintervals in between every two observation times, MCMC were run for 100000 iterations, and the first 500 iterations were discarded as burn-in periods. Although, the simulated Markov Chains have length of 100000, but we have been thinned by factor 50 to avoid effect of serial correlation.

A priori beliefs were specified for parameter independently using gamma prior distribution $\theta \sim \text{Gamma}(10, 20)$ for the three parameters of interest under study, the initial condition $X_0 = (s_0 = 1/200000, e_0 = 0.01 \text{ and } i_0 = 0.01)$ i.e. we believed that the total population is approximate 200000 sources from Thompson and Foege, 1968. Adopted the same new candidates updated used in the previous illustration 4.3.1, the new candidates update posals β^*, γ^* and α^* were drawn according to independent Gaussian random walk proposal density

$$\beta^* \sim N(\beta, c = 0.006), \gamma^* \sim N(\gamma, r = 0.005) \text{ and } \alpha^* \sim N(\alpha, k = 0.009).$$

where c, r and k represent the specified tuning parameters for each variance of the set parameters of interest adopted after several attempts to produce satisfactory results.

Table 4.5 Show Parameter estimate for posterior means and 95% posterior credible intervals (PCI) for different number of augmented values ($m = 7$ and 21 days) and their acceptance probability (AP).

	Proposal	β	γ	α	AP
$m = 7$	MGP	0.358(0.150,0.567)	0.808(0.688,0.847)	0.152(0.119,0.211)	0.63
$m = 21$	MGP	0.350(0.009,0.690)	0.781(0.591,0.980)	0.200(0.100,0.301)	0.60
Eichner and Dietz		0.335(0.005,0.627)	0.816(0.644,0.922)	0.157(0.010,0.389)	---

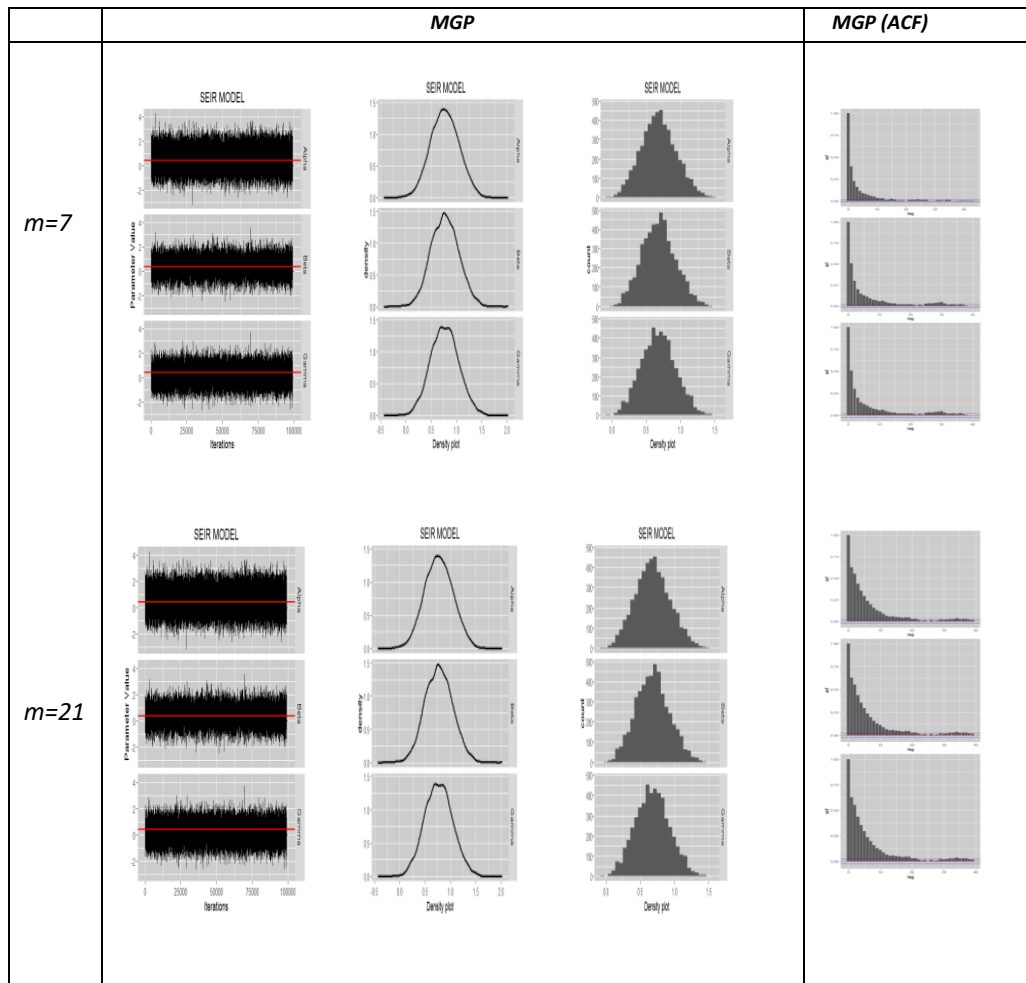


Figure 4.13 Shows the trace plots, histogram and posterior density and autocorrelation plot of the parameters estimated value from the innovation scheme couple with the newly proposed diffusion bridge approach (MGP).

For comparative purpose, we compare the results of the proposed methods with that of approximation likelihood method of Eichner and Dietz model. The proposed methods estimates are fairly similar to those of Eichner and Dietz. The posterior credible intervals

(PCI) and that of confidence intervals (CI) that of Eichner and Dietz are also quite similar. The difference can be explained by uncertainty fluctuation or variations adopted in the proposed methods.

4.4.2 Application II:

4.4.2 Ebola outbreak in the Democratic Republic of Congo

Just like the previous application, we applied the proposed methods to illustrate the performance of the epidemic dataset on the outbreak of Ebola disease in the Democratic Republic of Congo 1995, well documented by Khan et al. (1999).

Chowell et al. (2004) considered part of these datasets to estimate the parameters of SEIR model under deterministic approach. Similarly, Phenyio and Barbel (2006) applied the same dataset on stochastic SEIR model. Though, Phenyio & Barbel could not considered environmental fluctuation effect in their research work as was seriously considered in this study.

The dataset consist of two time series viz: the daily recoded counts of symptom onset of Ebola cases by date, from March 1 to July 16, this accounting for a total of 291 cases, and the daily counts of deaths from Ebola cases, also this accounting for a total of 236 deaths. It was also documented that the first case became ill on January 6, 1995, the last case died on July 16 making the total of 316 cases identified, resulting in a rate of 81% fatality.

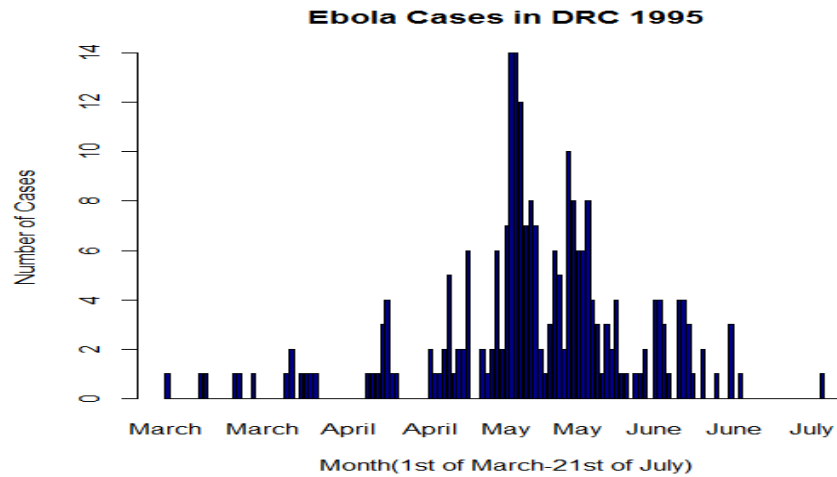


Figure 4.14 Shows daily counts of symptom onset of Ebola case by date. Data from the 1995 Ebola outbreak in the Democratic Republic of Congo recoded from March 1 (corresponding to day 1 on the x-axis) to July 16. **Sources:** Khan et al. (1999).

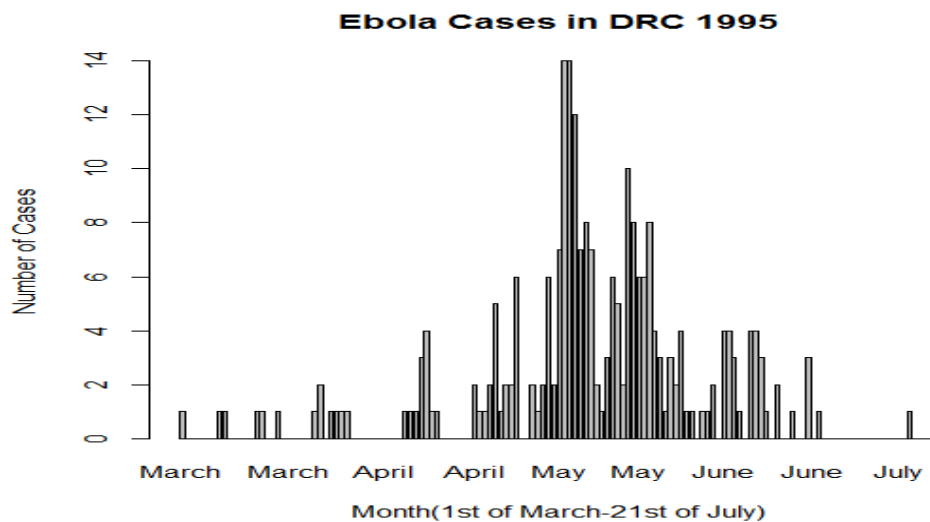


Figure 4.15 Shows daily counts of death from Ebola. Data from the 1995 Ebola outbreak in the Democratic Republic of Congo recoded from March 1 (corresponding to day 1 on the x-axis) to July 16. **Sources:** Khan et al. (1999).

The epidemic lasted for about 200 days, with control intervention introduced 130 days after the start of the epidemic. The exact starting time and evolution of the epidemic prior to March 1 is unobserved. Furthermore, from the total number of 316 identified cases, it can be

deduced that the dates of symptom onset for 25 cases and the dates of removed from the infectious class for 80 cases are not reported in the given time series (*Khan et al. ,1999*).

We initialised the sampler with parameters random constraints $0 < \beta < 1$, $0 < \gamma < 1$ and $0 < \alpha < 1$ having initial state $(s_0, e_0, i_0)^T = (1/5364500, 0.05, 0.01)^T$ represent the fraction of susceptible, exposed and infectious individuals respectively. Assuming a closed population of size $N \approx 5,364,500$ as adjusted from population size of the Bandundu-Kikwit region in 1984 population census, located on the banks of the Kwilu River ~ 475 km east of Kinshasa, the capital of Democratic Republic of Congo.

In order to reflect the observation interval in a real data situation, observations are provided only at time step on the time interval $[0, 138\text{days}]$.

We choose independent gamma priors for the three parameters of interest i.e. $\pi(\theta) \sim \text{Gamma}(10, 20) = \beta \sim \text{Gamma}(10, 20)$, $\gamma \sim \text{Gamma}(10, 20)$, and $\alpha \sim \text{Gamma}(10, 20)$ for the transmission rate, infectious rate and recovery rate respectively.

We applied MCMC algorithm (3.7) described in the previous chapter to simulated data. In the second sampling we pretending that the fraction of susceptible is unobserved, and running the MCMC chain for 100000, thinned by factor 50. For varying augmented dat over a week (7 days), two weeks (14 days) and three weeks (21 days): **$m \in (7, 14 \text{ and } 21 \text{ days})$** imputed time points for paths updated.

By adopted the sampling strategies used in the simulation study above, we applied the new candidates or proposals β^*, γ^* and α^* were drawn according to independently Gaussian random walk proposal density that **c, r** and **k** represent the specified tuning parameters for each variance of the set parameters of interest adopted after several attempts to produce satisfactory results.

$\beta^* \sim N(\beta, c = 0.006)$, $\gamma^* \sim N(\gamma, r = 0.005)$ and $\alpha^* \sim N(\alpha, k = 0.009)$.

Table 4.6 Shows parameter estimate for posterior means and posterior credible intervals for number of augmented values (**$m = 7, 14$ and 21 days**) and their acceptance probability (AP).

	Proposal	β	γ	α	AP
m = 7	MGP	0.394(0.250,0.568)	0.524(0.514,0.653)	0.304(0.221,0.588)	0.71
	MDB	0.385(0.013,0.768)	0.564(0.500,0.750)	0.344(0.024,0.759)	0.52
	GP	0.395(0.013,0.768)	0.534(0.504,0.640)	0.294(0.014,0.609)	0.58
m = 14	MGP	0.394(0.249,0.568)	0.524(0.515,0.658)	0.304(0.222,0.608)	0.67
	MDB	0.385(0.013,0.708)	0.564(0.504,0.750)	0.344(0.024,0.779)	0.42
	GP	0.395(0.013,0.668)	0.534(0.501,0.640)	0.294(0.014,0.609)	0.58
m = 21	MGP	0.394(0.244,0.568)	0.524(0.514,0.655)	0.304(0.232,0.608)	0.61
	MDB	0.385(0.013,0.788)	0.564(0.500,0.750)	0.344(0.024,0.790)	0.42
	GP	0.395(0.013,0.768)	0.534(0.500,0.740)	0.294(0.014,0.609)	0.48

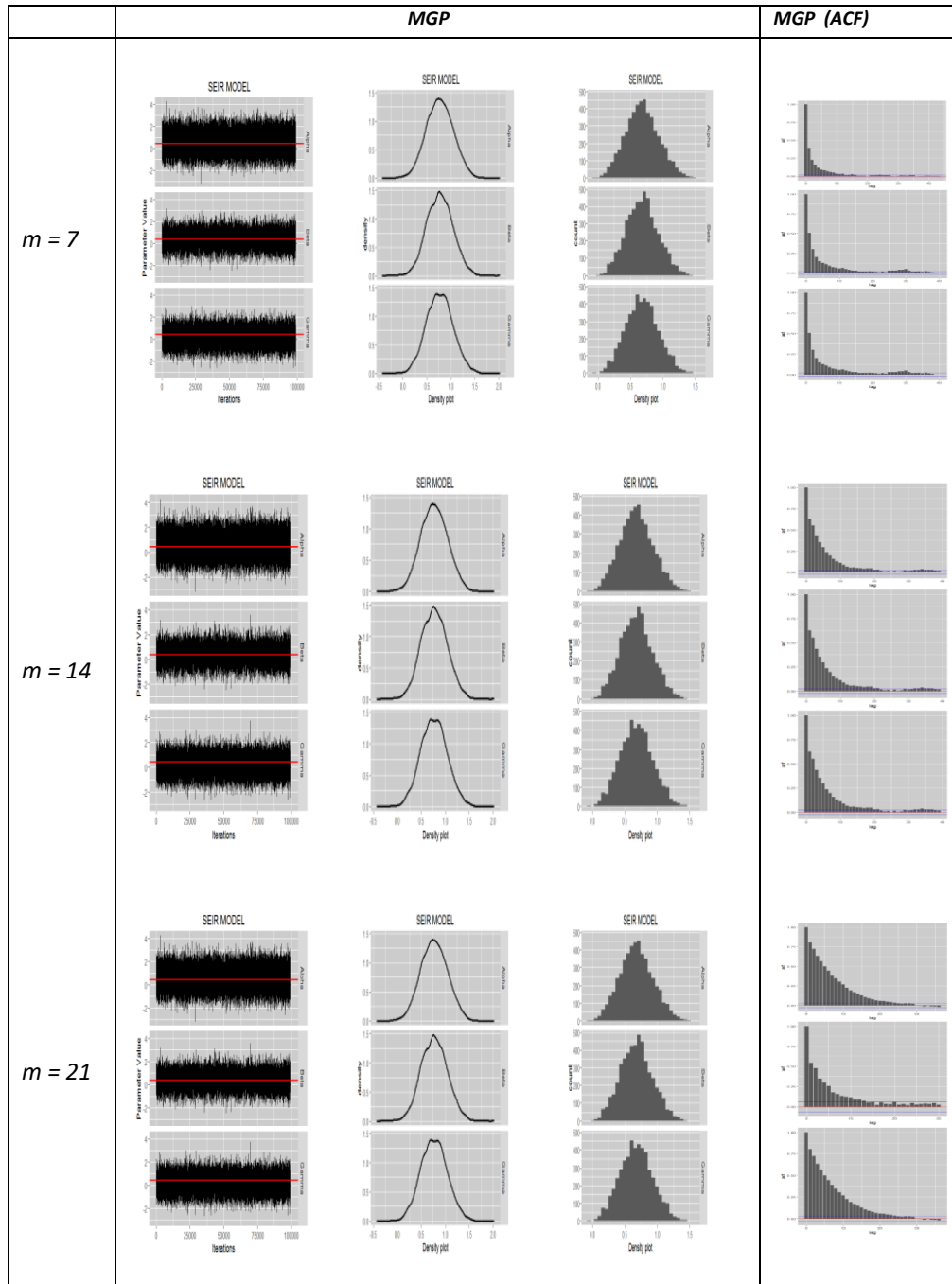


Figure 4.16 Shows the trace plots, histogram and posterior density and autocorrelation plot of three parameters estimated value from the innovation scheme couple with newly diffusion bridge proposed (MGP).

Assessment: Convergence of the Markov chain was assessed using a series of runs testing for stationary via the trace plots and inspect the autocorrelation function, the posterior density of

the parameters of interest reported in figure 4.16 with their respective posterior means (95% posterior credible intervals of the parameters) reported in table 4.6.

4.5 Findings

Based on the output results of this study, the following findings were obtained:

- (i) Table 4.1 shown the acceptance probabilities that based on a run of 10000 MCMC iterations using four different proposal schemes (*MDB*, *LB*, *GP*, *MGP*). From the result the rate or percentage of acceptance probability of *MGP* do not reduced, even when the value of data augmented increase. This implies that, the use of *MGP* diffusion bridge sampler through adapted MCMC algorithm do not deteriorated irrespective of the value of data imputed in between data point.
- (ii) Figure 4.1 Show evolutions of exact number of SIR using Ordinary Differential Equations (ODEs) method and the fraction of SIR epidemic model using Diffusion method. The evolution of Stochastic Differential Equations (Diffusion Approximation Approach) was found to be the most suitable trajectory of SIR epidemic model. Similarly, Figure 4.2 also showed evolution of ODEs and SDEs (diffusion process) trajectories of SEIR epidemic model. The results from these two figures imply that modelling with SDEs mirrored the exact dynamic epidemic models;
- (iii) Table 4.2 showed results of parameter estimate (posterior means and credible intervals) for different number of data augmented (*m=2, 5, 15 and 50*) between data point. From the table, the proposed methods *MGP* had closet Posterior Credible Interval, likewise, the posterior means values get closer to the true distribution value as 95% highest posterior density region, *MGP* credible interval for the parameter of interest get narrower, this implies that the MCMC iteration mixing very well, even when *m* values > 5.

- (iv) Figure 4.3 and Figure 4.4 Showed results of sample realisation trajectories path proposals for the SIR-Epidemic model on set true parameter value $(\beta, \gamma) = (0.330, 0.178)$, having set time interval $[0,1]$ that fulfilling the SIR diffusion epidemic conditioned on the initial number of susceptible and infectious $(s_0, i_0) = (71, 79)$ and the final value $(s_T, i_T) = (96, 71)$ imputed, with subintervals of $\mathbf{m} = (5, 50, 200)$ generated from three different Diffusion Bridges, Modified Diffusion Bridge (MDB), Guided Proposal GP, and Modified Guided diffusion Bridges (MGP). From the figures we found out that, amongst the diffusion bridge samplers considered, the newly proposed diffusion bridge sampler was offer significant better mixing especially, when the number of augmented values tend to infinite;
- (v) Table 4.5 showed Parameter estimate for posterior means and credible intervals for different number of imputed values ($\mathbf{m} = 7 \text{ and } 21\text{days}$) and their respective acceptance probability (AP). The comparison results obtained from the proposed method and that of the approximation likelihood method of Eichner and Dietz model clearly shown that, the result of the methods of parameter estimate values were fairly similar. The little different in the result can be explained by uncertainty environmental influences considered in this study;
- (vi) results from Figure 4.5 and 4.6 Showed the acceptance probabilities plot for fully observed based on a run of 1000 MCMC iterations for four different proposal for varying stochastic data augmentation $\mathbf{m}$ adopted for observed and unobserved component of epidemic data. From the results the proposed method of parameter estimate (Modified Innovation method) was not found to suffer from convergence problems. This implies that, application of the newly method on MCMC scheme

would give out reliable results and simplify iterative optimisation than existing methods of sampling algorithms;

- (vii) Apart from concentrating on the posterior estimates values, we also pay attention on the behaviour of the trace plots. Figure 4.7, 4.10 and 4.13 showed the trace plots for the parameter of interest θ . The true parameter value for the parameters was indicated by the red horizontal line. Results from these plots were clearly revealed that the proposed method (*MGP*) shown MCMC that mixed well, even when augmented data are large. The results of MCMC deteriorated were not accounted for in all iterations that used the proposed method.

In general, the extensive simulation study presented in this chapter and the results obtained from various iteration offers a useful guide on the choice of MCMC algorithms for discretely observed diffusion epidemic models. The main finding of this simulation study is that, whenever the values of the augmented dataset increases the existing methods algorithms performed badly. It is therefore recommended that the use of proposed algorithm should be considered instead. Through this (proposed methods) we found out that the MCMC algorithms do not suffer from the convergence problems, even when unknown parameters appear in the diffusion coefficient within a discrete time framework, especially in the multidimensional cases.

Chapter Five

Summary, Conclusion and Recommendation

5.1 Summary

Epidemic models have been increasingly used to evaluate the potential impact of different control measures and to guide public health policy decisions. Deterministic models for epidemic in humans and animals have a vast literature. Although these modeling can sometimes be sufficient to model the mean behaviour of the underlying stochastic system and guide towards parameters estimates, but they do not allow the quantification of the uncertainty associated to model parameters estimates.

Stochastic models can be used to infer such complex epidemic parameters and provide estimates of their variability. In this study, we start by derived the stochastic models using diffusion process so as to analysing the general stochastic epidemic model, which describe the spread of diseases assuming a closed population with homogeneous mixing and later describe how Bayesian inference is so important used to estimate the epidemiological parameters of the model influence by environmental variability.

In this study, we derived diffusion approximation process for simulating multidimensional representation of epidemic models, in such a way that the desired discrete times observations are discretized using Euler Maruyama Scheme. We considered improved version of guided proposal diffusion bridge sampler to arrive at newly propose diffusion bridge method i.e. modified guided bridge sampler, applied to generate realisation sample path (missing data) in between two consecutive observation times. These propose diffusion bridge method capable of dealing with discretisation bias which often caused a parameter estimate updated challenges in discretisation method of diffusion bridges.

This research work cover stochastic model as well as Bayesian Inference as an alternative to statistical frequentist method of estimation mostly used to infer diffusion approximation

processes for discretely observed data in the literature. We first presented the existing MCMC algorithms which have been proposed in the literature for discretely observed diffusion processes. Several methods of frequentist procedures to infer on the parameters estimates are been considered in the literature. Most techniques struggle when inter-observation times are large or when observations are measure with errors. The difficulty that underlies these is the general intractability of the transition density for discrete-time observations, making the likelihood function not to be in closed form. Some of these methods have been explained in the literature review, such as, the Hermite polynomial expansion approach proposed by Aït-Sahalia (2008), this approach performs poorly if the data are sparsely sampled (Stramer and Yan, 2007b). Sarkka and Sottinen (2008) also proposed another approach which used an alternative SDE as an importance process and the Girsanov theorem to help evaluate the likelihood ratios of the SDEs. This approach also performs poorly if the data were not equidistant in times. Others similar approaches were explained and reviewed in chapter two. All these approaches have some restrictions which could limit its application, especially for multivariate models (Lindström, 2012).

Recent developments have mainly been focused on Bayesian approaches (Donnet et al., 2010; Golightly and Wilkinson, 2011), which feasibly applicable in the model and the data that are sparsely sampled. However, these approaches still facing two major challenges: the first challenge is that of coupling or serial correlation that exist if unknown parameters appear in the diffusion coefficient, direct implementation of data-augmentation results into a Markov chain that is reducible. The second challenge focused on the proposal diffusion bridges mechanism constructed to generate sample realisation data (imputed) in between the two consecutive observations. In this research work we are able to solve these two challenges by developed newly diffusion bridge mechanism and modified on the existence parameters estimation method, specifically for multidimensional diffusion approximation processes.

5.2 Conclusion

In most emerging epidemic outbreaks the distribution of the generation time is not known and inference methods are needed. Especially when consider epidemic data that involving environmental influence. Statistical inference for uncertainty has to be performed during the growing state of the outbreak. The main difficulty in estimating parameters for such epidemic models is that some of the state process is only partially observed and the observed quantities may be suggested in a wide range of time (e.g. weekly, or monthly). Therefore, the likelihood may become very difficult to evaluate.

Stochastic differential equations are natural choice to model the time evolution of dynamic systems of such problems, even when subjected to environmental influences or random variation.

In this study, we highlights the importance of modeling with Stochastic Differential Equations which are capable to capture all the variability and uncertainties which may happen in the dynamic system of any natural phenomena. A convenient and appropriate approach to determine solution to stochastic differential equations is provided by diffusion process. Unfortunately, the statistical inference used under diffusion process to determining the quantities or parameters estimation when exact likelihood-based estimations are impossible suffers from convergence problems.

Practical solutions for these problems have not yet been achieved, especially for multi-dimensional diffusion processes. In this study, we adopted Bayesian data-augmentation method of statistical estimation via Markov Chain Monte Carlo (MCMC) approaches. With MCMC it is possible to examine and assess the distribution of unknown parameters of interest. We constructed a new proposal Diffusion bridge sampler in conjunction with the reparameterisation method (Modified Innovation scheme) to address these challenges which

are often cause problems while evaluating unknown parameters of such processes. Finally we applied these newly approaches to simulation and real life dataset.

The findings from the results were encourages, suggesting that the new approaches are capable and suitable to address the challenges in evaluate parameters estimate when epidemic data are discretely-observed influence by random variation.

5.3 Contribution to Knowledge

In this research work, we have been able to address and solved the two major challenges on parameters estimate for diffusion process, specifically, in multi-dimensional cases, using Bayesian data-augmentation approach.

These challenges are:

Direct implementation of Bayesian method of estimation results into a Markov chain that is reducible, if unknown parameters appear in the diffusion coefficient. This was first noted in a simulation study in Elerian et al. (2001) and was justified theoretically in Roberts and Stramer (2001).

Also, Diffusion bridges samplers resulting into awful MCMC convergence, if the number of imputed values, the augmented data between every successive time point increases ($\mathbf{m} \rightarrow \infty$).

By proposed Modified Guided proposal (MGP) as the first proposed method in this study to capture the convergence challenges often available under diffusion bridges samplers which often resulting into awful MCMC convergence.

Likewise, in this study, we present Bayesian inference algorithm (3.7) through Modified Innovation MCMC scheme to deal with deteriorating of the chain if unknown parameters appear in the diffusion coefficient

All in all, this study addressed the challenges often causing awful inferential issue in diffusion processes, as well as adds to the growing literature on parameter estimation of diffusion approximation models using Bayesian inference framework as unique methodological development that would have many potential applications in many discipline. While our focus and the examples are geared toward epidemic models, the approaches can certainly be applied to a wider range of problems in other research area.

5.4 Recommendations

Based on the contributions of this study, diffusion processes can more easily and more efficiently be applied for modelling and estimation purposes in any dynamic modelling. Future perspectives of this work mainly concern the utilisation of the developed methods to further areas of applications, the variety of which is manifold. Practitioners should feel encouraged to dare apply stochastic differential approaches (diffusion processes) in their research areas. The practitioners should likewise make use of Bayesian inference as a good alternative method of parameters estimate whenever exact likelihood-based methods of estimations are impossible.

This research work provides general guidelines for the setup of appropriate diffusion modelling and it supplies adequate information for their statistical inference.

5.5 Suggestions for Future Study

Modeling with stochastic process using multidimensional diffusion approximation approach (or stochastic differential equations) gives out viable, reasonable results, especially when deal with real life phenomena. In the course of this study, certain assumptions and some additional component in epidemic models (such as demographic effect) were relaxed, both from the epidemiological models and the inferential methods used to carry out the parameter estimate, especially; the population under study for epidemic models was considered to be in closed.

No special consideration on the effect of demographic changes for the dynamic system of the models. Another issue which should carefully been considered in the future is the effect of the number of additional latent components (when unobserved components are more than two states), over the convergence of the set MCMC scheme.

Likewise, the diffusion processes under study basically based on the discretely observed data. Assumption under continuous-time was not considered in the course of the study. To further evaluate the impact and the effect of theses restricted part of the methods, our future work would focus more on theses restricted assumptions in order to evaluate their impact on the behaviour of the proposed methods.

In addition to the above claim, all algorithms used for the propose methods have been implemented in R package. Their provision as a software package is one of the future projects. Such advancements will help improving the comprehension and prediction of epidemic outbreaks and any other related research study.

Overall, the combined application of diffusion modelling and statistical inference promises to supply new insight in many exciting areas of epidemics in the future. This research work has demonstrated the potential of this approach.

Bibliography

- Abbey, H. (1952).** An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24, 201-233.
- Aït-Sahalia, Y. (2002).** Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70, 223-262.
- Aït-Sahalia, Y. (2008).** Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36, 906-937.
- Allen, L. (2006).** *An Introduction to Stochastic Processes with Applications to Biology*. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Allen, E. J. (2007).** Modeling with Itô Stochastic Differential Equations. Springer publisher , Dordrecht, the Netherlands.
- Allen, L.J.S. (2008).** An introduction to stochastic epidemic models. In mathematical epidemiology, Springer. Publisher, pages 81-130.
- Aliu, A. Hassan., Abiodun A.A., and Ipinyomi R.A. (2017).** Statistical inference for discretely observed diffusion epidemic models. International journal of mathematical research, vol. 6(1), 29 – 35. DOI: 10.18488/journal.24.2017.61.29.35
- Anderson, R., & May, R.(1991).** *Infectious Diseases of Humans*. Oxford: Oxford University Press.
- Anderson, H., & Britton, T. (2000).** *Stochastic Epidemic Models and Their Statistical Analysis*.
- Bailey, N. (1957).** *The Elements of Stochastic Processes*. New York: John Wiley & Sons.
- Bailey, N. (1975).** *The Mathematical Theory of Infectious Diseases* (2nd ed.). London: Charles Griffin & Co. Ltd.
- Barbour, A. (1974).** On a functional central limit theorem for Markov population processes. *Advances in Applied Probability*, 6, 21-39.
- Bartlett, M. (1949).** Some evolutionary stochastic processes. *Journal of the Royal Statistical Society, Series B (Methodological)*, 11, 211-229.
- Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006).** Exact and Computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodology*. 68(3), 333–382. With discussions and a reply by the authors.
- Becker N. (1979).** The uses of epidemic models. *Biometrics* 35(1), 295-305.
- Becker, N.(1989).** *Analysis of Infectious Disease Data*. London: Chapman & Hall.
- Beskos, A., & Roberts, G. (2005).** Exact simulation of diffusions. *Annals of Applied Probability*, 15, 2422-2444.
- Beskos, A., Papaspiliopoulos, O., Roberts, G., & Fearnhead, P.(2006).** Exact and Computationally efficient likelihood-based estimation for discretely observed diffusion processes (with comments). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 68, 333-382.
- Beskos, A., Papaspiliopoulos, O., & Roberts, G.(2006).** Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12, 1077-1098.

- Beskos, A., Papaspiliopoulos, O., & Roberts, G. (2008).** A factorisation of diffusion measure and finite sample path constructions. *Methodology and Computing in Applied Probability*, 10, 85-104.
- Berger, J. O. (2006).** “The Case for Objective Bayesian Analysis,” *Bayesian Analysis*, 3, 385–402, <http://ba.stat.cmu.edu/journal/2006/vol01/issue03/berger.pdf>.
- Beskos, A., Papaspiliopoulos, O., & Roberts, G.(2009).** Monte-Carlo maximum likelihood estimation for discretely observed diffusion processes. *Annals of Statistics*, 37, 223-245.
- Bibby, B., Jacobsen, M., & Sørensen, M.(2009).** Estimating functions for discretely sampled diffusion-type models. In Y. Aït-Sahalia & L. Hansen (Eds.), *Handbook of Financial Econometrics* (p. 203-268). Amsterdam: North-Holland.
- Bibby, B., & Sørensen, M.(1995).** Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1, 17-39.
- Bibby, B., & Sørensen, M.(2001).** Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scandinavian Journal of Statistics*, 28, 99-112.
- Black, F., & Scholes, M.(1973).** The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637-654.
- Bladt, M. and Sørensen, M. (2014).** Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli* 20(2), 645–675.
- Bladt, M. and Sørensen, M. (2015).** Simulation of multivariate diffusion bridges. To appear in *Journal of the Royal Statistical Society, series B*.
- Boys, R., Wilkinson, D., & Kirkwood, T.(2008).** Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18, 125-135.
- Box, G. E. P. and Tiao, G. C. (1973).** *Bayesian Inference in Statistical Analysis*, Wiley Classics Library Edition, published 1992, New York: John Wiley & Sons.
- Brandt, M., & Santa-Clara, P.(2001).** Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Working Paper 274, National Bureau of Economic Research*.
- Brauer, F. (2009).** Mathematical epidemiology is not an oxymoron. *BMC Public Health*, 9,S2.
- Broze, L., Scaillet, O., & Zakoïan, J.-M. (1998).** Quasi-indirect inference for diffusion processes. *Econometric Theory*, 14, 161-186.
- Brooks, S. P. and Roberts, G. O. (1998).** “Assessing Convergence of Markov Chain Monte Carlo Algorithms,” *Statistics and Computing*, 8, 319–335.
- Ball, F. G. and Lyne, O. D. (2001).** Stochastic multitype SIR epidemics among a population partitioned into households. *Adv. Appl. Probab.* 33, 99–123.
- Ball, F. G., Sirl, D. J. and Trapman, P. (2010).** Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. Appl. Probab.* 41, 765–796.
- Britton, T., Deijfen, M., Lageröas, A. N. and Lindholm, M. (2008).** Epidemics on random graphs with tunable clustering. *J. Appl. Probab.* 45, 743–756.
- Capasso V. And Sero G. (1978).** A generalization of the Kermack-McKendrick deterministic epidemic model. *Mathematical Biosciences*, 42(1): 43-61.

- Chen, G. and Li, T. (2009).** Stability of stochastic delayed sir model. *Stochastic and Dynamics*, 9(02):231-252.
- Capasso, V., & Morale, D.(2009).** Stochastic modelling of tumour-induced angiogenesis. *Journal of Mathematical Biology*, 58, 219-233.
- Castillo-Chavez, C. Castillo-Garsow and A. A. Yakubu, (2003).** Mathematical Models of Isolation and Quarantine. *JAMA* 290 (2003) 2876-2877.
- Centers for Disease Control (CDC), (2003a).** Ebola hemorrhagic fever. (<http://www.cdc.gov/ncidod/dvrd/spb/mnpages/dispages/ebola.htm>), accessed on August 24,2003.
- Chowell, G., Hengartner, N.W., Castillo-Chavez, C., Fenimore, P.W., and Hyman, J.M.(2004).** The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *Journal of Theoretical biology* 229, 119-126.
- Chowell, G. and Nishiura, H. (2014).** Transmission dynamics and control of Ebola virus disease (EVD): A review. <http://www.biomedcentral.com/1741-7015/12/196> 30 September, 2014.
- Chowell, G, P. W. Fenimore, M. A. Castillo-Garsow and C. Castillo-Chavez,(2003).** SARS Outbreaks in Ontario, Hong Kong and Singapore: The Role of Diagnosis and Isolation as a Control Mechanism. *Journal of Theoretical Biology* 224 (2003) 1-8.
- Chowell G, C. Castillo-Chavez, P. W. Fenimore, C. M. Kribs-Zaleta, L. Arriola and J. M. Hyman, (2004).** Model Parameters and Outbreak Control for SARS. *Emerging Infectious Diseases* 10 (2004) 1258-1263.
- Chen, M.-H., & Shao, Q.-M.(1999).** Monte Carlo estimation of Bayesian credible and hpd-intervals. *Journal of Computational and Graphical Statistics*, 8, 69-92.
- Chen, W.-Y., & Bokka, S.(2005).** Stochastic modeling of nonlinear epidemiology. *Journal of Theoretical Biology*, 234, 455-470.
- Chiarella, C., Hung, H., & Tô, T.-D.(2009).** The volatility structure of the fixed income market under the HJM framework: A nonlinear filtering approach. *Computational Statistics and Data Analysis*, 53, 2075-2088.
- Chib, S., Pitt, M., & Shephard, N.(2004).** Likelihood based inference for diffusion driven state space models. *Working Paper, Nuffield College, University of Oxford*.
- Chib, S., & Shephard, N.(2002).** Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes: Comment. *Journal of Business and Economic Statistics*, 20 (3), 325-27.
- Clark, J. M. C. (1990).** The simulation of pinned diffusions. In *Decision and Control, 1990. Proceedings of the 29th IEEE Conference on*, pp. 1418–1420. IEEE. Van der Meulen and Schauer/Bayesian estimation for diffusions using guided proposals 37
- Costa Maia, J. d. O. (1952).** Some mathematical developments on the epidemic theory formulated by reed and frost. *Human Biology*, 24, 167-200.
- Cox, J., Ingersoll, J., & Ross, S.(1985a).** An intertemporal general equilibrium model of asset prices. *Econometrica*, 53, 363-384.
- Cox, J., Ingersoll, J., & Ross, S.(1985b).** A theory of the term structure of interest rates. *Econometrica*, 53, 385-408.

- Crank, J., & Nicolson, E.(1947).** A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proceedings of the Cambridge Philosophical Society*, 43, 50-67.
- Cunha, B.(2004).** Historical aspects of infectious diseases, part I. *Infectious Disease Clinics of North America*, 18 (1), xi-xv.
- Dargatz, C.(2006).** A diffusion approximation for an epidemic model. *Discussion Paper 517 of SFB 386, Ludwig-Maximilians University Munich*.
- Dargatz, C., Georgescu, V., & Held, L.(2007).** Stochastic modelling of the spatial spread of influenza in Germany. *Austrian Journal of Statistics*, 35, 5-20.
- Daley, D., & Gani, J.(1999).** *Epidemic Modelling: An Introduction*.
- Daley, D., & Kendall, D.(1965).** Stochastic rumours. *IMA Journal of Applied Mathematics*, 1, 42-55.
- DeGroot, M. H. and Schervish, M. J. (2002).** Probability and Statistics, 3rd Edition, Reading, MA: Addison-Wesley.
- Delyon, B. and Hu, Y. (2006).** Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* 116(11),1660-1675.
- Durham, G. B. and Gallant, A. R. (2002).** Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* 20(3), 297–338. With comments and a reply by the authors.
- Dacunha-Castelle, D., & Florens-Zmirou, D. (1986).** Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19, 263-284.
- Diekmann, O., & Heesterbeek, J.(2001).** *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Chichester: John Wiley & Sons.
- Diekmann, O., J.A.P. Heesterbeek, and J.A.J. Metz (1990).** On the definition and the Computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.*, 28:365–382.
- Dietz, K.(1967).** Epidemics and rumours: A survey. *Journal of the Royal Statistical Society, Series A (General)*, 130, 505-528.
- Dobson, A., & Carper, E. (1996).** Infectious diseases and human population history. *BioScience*, 46, 115-126.
- Donnet, S, Foulley, J.L, Samson, A. (2010).** Bayesian Analysis of growth Curves using mixed models defined by stochastic differential equations. *Biometrics* 66(3), 733-741.
- Donnelly. C. A., A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, et al.,(2003).** Epidemiological Determinants of Spread of Causal Agent of SARS in Hong Kong. *Lancet* 361 (2003) 1761-1766.
- Duffie, D., & Singleton, K. (1993).** Simulated moments estimation of Markov models of asset prices. *Econometrica*, 61, 929-952.
- Eichner, M., Dietz, K., (2003).** Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemic.* 158 (2), 110–117.
- Elerian, O.(1998).** A note on the existence of a closed form conditional transition density for the Milstein scheme. *Working Paper, Nuffield College, University of Oxford*.

- Elerian, O., Chib, S., & Shephard, N.(2001).** Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69, 959-993.
- En'ko, P. (1889).** On the course of epidemics of some infectious diseases. *Vrach. St.Petersburg*, 10, 1008-1010, 1039-1042, 1061-1063.
- Eraker, B.(2001).** MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, 19, 177-191.
- Eraker, B.(2002).** Comment on 'Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes'. *Journal of Business and Economic Statistics*, 420, 327-329.
- Eraker, B.(2004).** Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance*, 59, 1367-1404.
- Eraker, B., Johannes, M., & Polson, N. (2003).** The impact of jumps in volatility and returns. *The Journal of Finance*, 58, 1269-1300.
- Eto, A., Saito, T., Yokote, H., Kurane, I., Kanatani, Y., (2015).** Recent advances in the study of live attenuated cell-cultured smallpox vaccine LC16m8. *Vaccine* 33(45), 6106–6111.
- Fuchs, C. (2013).** Inference for diffusion processes. Springer, Heidelberg. With applications in life sciences, With a foreword by Ludwig Fahrmeir.
- Ferguson, N.M., Donnelly, C.A. and Anderson, R.M. (2001a).** The Foot-and- Mouth epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science* 292, 1155-1160.
- Fred Brauer. (2011).** Mathematical Models for communicable diseases. An introduction to mathematical modelling in epidemiology; strategic and tactical models July 21, 2011 pp.22-25.
- Folland, G. B. (1999).** Real Analysis: Modern Techniques and Their Applications. John Wiley and Sons, 2nd edition.
- Fearnhead, P.(2006).** The stationary distribution of allele frequencies when selection acts at unlinked loci. *Theoretical Population Biology*, 70, 376-386.
- Fearnhead, P.(2008).** Computational methods for complex stochastic systems: A review of some alternatives to MCMC. *Statistics and Computing*, 18, 151-171.
- Fearnhead, P., Papaspiliopoulos, O., & Roberts, G. (2008).** Particle filters for Partially observed diffusions. *Journal of the Royal Statistical Society, Series B*, 70, 755-777.
- Florens-Zmirou, D. (1989).** Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20, 547-557.
- Florens-Zmirou, D.(1993).** On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30, 790-804.
- Gallant, A., & Tauchen, G.(1996).** Which moments to match? *Econometric Theory*, 12, 657-681.
- Gamerman, D., & Lopes, H.(2006).** *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). Boca Raton, London, New York: Chapman & Hall.
- Gani, R., Leach, S., (2001).** Transmission potential of smallpox in contemporary publications. *Nature* 414, 748–751.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990).** "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *Journal of the American Statistical Association*, 85, 972–985.

- Gelman, A., Meng, X.L., Stern, H., (1996).** Posterior predictive assessment of model fitness Via realized discrepancies. *Stat. Sin.* 6 (4), 733–760.
- Geman, S. and Geman, D. (1984).** “Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gibson, G.J., Renshaw, E., (1998).** Estimating parameters in stochastic compartmental models using Markov Chain methods. *IMA J. Math. Appl. Med. Biol.* 15, 19–40.
- Gibson, M., & Bruck, J.(2000).** Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104, 1876-1889.
- Gibson, M., & Mjolsness, E.(2001).** Modeling of the activity of single genes. In H. Bolouri & J. Bower (Eds.), *Computational Modeling of Genetic and Biochemical Networks* (p.1-48). Cambridge: MIT Press.
- Gilks, W., Richardson, S., & Spiegelhalter, D. (Eds.). (1996).** *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.
- Gillespie, D.(2007).** Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58, 35-55.
- Gloter, A., & Jacod, J.(2001a).** Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM: Probability and Statistics*, 5, 225-242.
- Gloter, A., & Jacod, J.(2001b).** Diffusions with measurement errors. II. Optimal estimators. *ESAIM: Probability and Statistics*, 5, 243-260.
- Goldstein, M. (2006).** “Subjective Bayesian Analysis: Principles and Practice,” *Bayesian Analysis*, 3, 403–420, <http://ba.stat.cmu.edu/journal/2006/vol01/issue03/goldstein.pdf>.
- Godambe, V.(1991).** *Estimating Functions*. Oxford: Oxford University Press.
- Golightly, A., & Wilkinson, D. (2005).** Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, 61, 781-788.
- Golightly, A., & Wilkinson, D.(2006a).** Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, 16, 323-338.
- Golightly, A., & Wilkinson, D.(2006b).** Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*, 13, 838-851.
- Golightly, A., & Wilkinson, D. (2008).** Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Analysis*, 52, 1674-1693.
- Golightly, A. and Wilkinson, D. J. (2010).** Learning and Inference in Computational Systems Biology, chapter Markov chain Monte Carlo algorithms for SDE parameter estimation, pp. 253–276. MIT Press.
- Golightly, A., & Wilkinson, D. (2010).** Markov chain Monte Carlo algorithms for SDE parameter estimation. In N. Lawrence, M. Girolami, M. Rattray, & G. Sanguinetti (Eds.), *Introduction to Learning and Inference for Computational Systems Biology*. MIT Press (in press).
- Golightly, A., Henderson, D.A., Sherlock, C., (2014).** Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Stat. Comput.* 2014,1–17.

- Gumel AB (2010).** Global asymptomatic dynamics of a model for quarantine and isolation. *Discrete Continuous Dynamic System Ser B* 14:209-231.
- Gumel. A .B, S. G. Ruan, T. Day, J. Watmough, F. Brauer, P. Van den Driessche, et al.,(2004).** Modelling Strategies for Controlling SARS Outbreaks. *Proceedings of the Royal Society of London C B* 271 (2004) 2223-2232.
- Greenwood, M.(1931).** On the statistical measure of infectiousness. *The Journal of Hygiene*, 31, 336-351.
- Grimmett, G., & Stirzaker, D.(2001).** *Probability and Random Processes* (3rd ed.). Oxford: Oxford University Press.
- Halloran, M.E., Longini, I.M., Nizam, A., Yang, Y., (2002).** Containing bioterrorist smallpox. *Science* 298, 1428–1432.
- Hamer, W. H. (1906).** Epidemic disease in England. *The Lancet*, 1:733{739.
- Heesterbeek, J.A.P. (1992) R0, Thesis, CWI, Amsterdam.
- Hansen, L.(1982).** Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029-1054.
- Hastings, W. K. (1970).** “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, 57, 97–109.
- Heesterbeek, J., & Roberts, M. (2007).** The type-reproduction number t in models for infectious disease control. *Mathematical Biosciences*, 206, 3-10.
- Henderson, D.A., Arita, I., (2014).** The smallpox threat: a time to reconsider global policy. *Biosecurity Bioterrorism: Biodefense Strategy Pract. Sci.* 12 (3), 117–121.
- Hethcote, H.(1994).** A thousand and one epidemic models. In S. Levin (Ed.), *Frontiers in Mathematical Biology* (p. 504-515). Berlin: Springer.
- Hethcote, H.(2000).** The mathematics of infectious diseases. *SIAM Review*, 42, 599-653.
- Heyde, C.(1997).** *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer.
- Honoré, P. (1997).** Maximum likelihood estimation of non-linear continuous-time term Structure models. *Working paper 1997-7, Department of Finance, Aarhus School of Business*.
- Isham, V.(2004).** Stochastic models for epidemics. *Research Report No. 263, Department of Statistical Science, University College London*.
- Itô, K., & McKean, H. (1974).** *Diffusion Processes and their Sample Paths*. Berlin Heidelberg, New York: Springer.
- Ito, K. (1942).** On stochastic processes. infinitely divisible laws of probability.
- Jazwinski, A. H. (1970).** *Stochastic Processes and Filtering Theory*. Academic Press.
- Jacobsen, M.(2001).** Discretely observed diffusions: Classes of estimating functions and small - optimality. *Scandinavian Journal of Statistics*, 28, 123-149.
- Jacquez, J.(1972).** *Compartmental Analysis in Biology and Medicine*. Amsterdam: Elsevier.
- Jaynes E (1968).** Prior Probabilities." *IEEE Transactions on Systems Science and Cybernetics*, 4(3), 227-241.
- Jeffreys, H.(1961).** *Theory of Probability*. Princeton: Prentice Hall.

- Jimenez, J., Biscay, R., & Ozaki, T. (2006).** Inference methods for discretely observed continuous-time stochastic volatility models: A commented overview. *Asia-Pacific Financial Markets*, 12, 109-141.
- Jensen, A. C., Ditlevsen, S., Kessler, M. and Papaspiliopoulos, O. (2012).** Markov chain Monte Carlo approach to parameter estimation in the FitzHugh-Nagumo model. *Phys. Rev. E* 86, 041114.
- Jensen, C., Anders (2014).** Statistical Inference for Partially Observed Diffusion Processes. Ph.d. Thesis University of Copenhagen.
- Jones, C.S. (1999).** Bayesian estimation of continuous-time finance models. Unpublished paper. Simon school of Business University of Rochester 233.
- Jones, C.(1998).** A simple Bayesian method for the analysis of diffusion processes. *Working Paper*, University of Pennsylvania.
- Kermack, W. and Mckendrick, A. (1927).** A contribution to the Mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A* 115, 700-772.
- Kloeden, P. E. and Platen, E. (1999).** Numerical Solution to Stochastic Differential Equations. Springer.
- Kloeden, P. E. and Platen, E. (1992).** *Numerical Solution of Stochastic Differential equations*. Springer, New York.
- Kalogeropoulos, K. (2007).** Likelihood-based inference for a class of multidimensional diffusions with unobserved paths. *Journal of Statistical Planning and Inference*, 137, 3092-3102.
- Kalogeropoulos, K., Dellaportas, P., & Roberts, G.(2007).** Likelihood based inference for correlated diffusions (in submission).
- Kalogeropoulos, K., Roberts, G., & Dellaportas, P.(2010).** Inference for stochastic volatility models using time change transformations. *Annals of Statistics (to appear)*.
- Kass, R., & W Raftery, A.(1996).** Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.
- Kessler, M.(1997).** Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24, 211-229.
- Kessler, M., & Paredes, S.(2002).** Computational aspects related to martingale estimating functions for a discretely observed diffusion. *Scandinavian Journal of Statistics*, 29, 425-440.
- Kessler, M., & Sørensen, M.(1999).** Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5, 299-314.
- Khan,A.S, Tshioko,F.K, Heymann,D.L, (1995).** The re-emergence of Ebola hemorrhagic fever, Democratic Republic of Congo, J.inf.Dis.179, 576-586.
- Kloeden, P., & Platen, E. (1991).** Stratonovich and Itô stochastic Taylor expansions. *Mathematische Nachrichten*, 151, 33-50.
- Lindstrom, E. (2012).** A regularized bridge sampler for sparsely sampled diffusions. *Statistics and computing* 22(2), 615-623.
- Liu, J. S. (2001).** Monte Carlo Strategies in Scientific Computing, Springer-Verlag.

- Lipsitch. M, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, et al.,(2003).** Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* 300 (2003) 1966-1970.
- Lipsitch. M, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, et al.,(2003).** Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* 300 (2003) 1966-1970.
- Lin, M., Chen, R. and Mykland, P. (2010).** On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* 105(490), 820–838.
- Lo, A.(1988).** Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4, 231-247.
- Meltzer MI, Atkins CY, Santibanez S, Knust B, Petersen BW, Ervin ED, Nichol ST, Damon IK, Washington ML. (2014).** Estimating the future number of cases in the Ebola epidemic liberia and sierra leone, 2014–2015. *MMWR Surveill Summ* 2014,63:1–14.
- Mao X.; Marion G. And Renshaw E. (2002).** Environmental Brownian noise suppresses explosions in population dynamics stochastic processes and Their applications, 97(1): 95-110.
- Milstein, G. N. (1995).** Numerical Integration of Stochastic Differential Equations. Kluwer Academic.
- McKendrick, A.(1926).** Application of mathematics to medical problems. *Proceedings of the Edinurgh Mathematical Society*, 44, 98-130.
- McNeil, D.(1973).** Diffusion limits for congestion models. *Journal of Applied Probability*, 10, 368-376.
- McNeill, W.(1976).** *Plagues and People*. New York: Anchor.
- Merton, R.(1973).** Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4, 141-183.
- Merton, R.(1976).** Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3, 125-144.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953).** “Equation of State Calculations by Fast Computing Machines,” *Journal of Chemical Physics*, 21, 1087–1092.
- Nåsell, I.(2002).** Stochastic models of some endemic infections. *Mathematical Biosciences*, 179, 1-19.
- Nielsen, J., Madsen, H., & Young, P.(2000).** Parameter estimation in stochastic differential equations; an overview. *Annual Reviews in Control*, 24, 83-94.
- Nancy Hernandez-Ceron, Zhilan Feng and Carios Castillo-Chavez (2013).** Discrete Epidemic Models with arbitrary stage distributions and applications to disease control. Published by: National Institutes of Health Public access. *Bull Math Biol.* 2013 October; 75(10):1716-1764.doi:10.1007/s11538-013-9866-x.
- Øksendal, B.(2003).** *Stochastic Differential Equations. An Introduction with Applications* (6th ed.). Berlin, Heidelberg: Springer.

- Oldstone, M. (2010).** *Viruses, Plagues, and History: Past, Present and Future*. Oxford, New York: Oxford University Press.
- O'Neill, P. D. and Becker, N. G. (2001).** Inference for an epidemic where susceptibility varies. *Biostatistics*, 2(1):99{108.
- Papaspiliopoulos, O., & Roberts, G.(2009).** Importance sampling techniques for estimation of diffusion models. In *Statistics for Stochastic Differential Equations*. London: Chapman & Hall (to appear).
- Papaspiliopoulos, O. and Roberts, G. (2012).** Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations, Monographs on Statistics and Applied Probability*, p. 311–337. Chapman and Hall.
- Papaspiliopoulos, O., Roberts, G. O. and Stramer, O. (2013).** Data Augmentation for Diffusions. *J. Computation. Graph. Statist.* 22(3), 665–688.
- Pedersen, A. R. (1995).** Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* 1(3), 257-279.
- Pedersen, A.(1995a).** A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22, 55-71.
- Pedersen, A.(1995b).** Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, 1, 257-279.
- Poulsen, R. (1999).** Approximate maximum likelihood estimation of discretely observed diffusion processes. *Working Paper 29, Center for Analytical Finance, Aarhus*.
- Phenyo E. Lekone and Barbel F. Finkenstadt (2006).** Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics journal* vol. 62. Pp 1170-1177.
- Riley, S, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, et al.,(2003).** Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science* 300 (2003) 1961-1966.
- Rob Deardon, Stephen P. Books and Bryan T. (2010).** Inference for individual-level models of Infectious disease in large populations. Canada Foundation for Innovation (CFI), and the Natural Sciences & Engineering Research Council of Canada (NSER).
- Ross, R. (1915).** An application of the theory of probabilities to the study of a priori Pathometry, i. *Proc. Roy. Soc. London*, A92:204{230.
- Ruan. S and Wang W. (2004).** Simulating the SARS Outbreak in Beijing with Limited Data. *Journal of Theoretical Biology* 227 (2004) 369-379.
- Ruan S. And Wang W. (2003).** Dynamic behavior of an epidemic model with a nonlinear incidence rate. *Journal of Differential equations*, 188(1): 135-163.
- Ramshaw, J.(1985).** Augmented Langevin approach to fluctuations in nonlinear irreversible processes. *Journal of Statistical Physics*, 38, 669-680.
- Revuz, D., & Yor, M.(1994).** *Continuous Martingales and Brownian Motion*. Springer.
- Robert, C., & Casella, G.(2004).** *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.
- Roberts, G., & Rosenthal, J. (1997).** Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2, 13-25.
- Roberts, G., & Stramer, O.(2001).** On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88, 603-621.

- Roberts, G., & Tweedie, R.(2010).** *Understanding MCMC*. Berlin: Springer (to appear).
- Roberts, M., & Heesterbeek, J.(2003).** A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society of London, Series B*, 270, 1359-1364.
- Robinson, E.(1959).** A stochastic diffusion theory of price. *Econometrica*, 27, 679-684.
- Rvachev, L., & Longini, I.(1985).** A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75, 3-22.
- Roberts and Rosenthal, J. S. (2001).** Handbook of Markov Chain Monte Carlo (Chapman & Hall/CRC Handbooks of Modern Statistical Methods), chapter Optimal proposal distributions and adaptive MCMC. Chapman and Hall/CRC, 1 edition.
- Santa-Clara, P. (1995).** Simulated likelihood estimation of diffusions with an application to the short term interest rate. *Working Paper, Anderson Graduate School of Management, UCLA*.
- Sarkka, S. Souinen, T. (2008).** Application of Girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems. *Bayesian Analysis* 3(3), 555-584.
- Schauer, M., Van der Meulen, F. H. and Van Zanten, J. H. (2016).** Guided proposals for simulating multi-dimensional diffusion bridges. Accepted for publication in *Bernoulli*. ArXiv e-prints 1311.3606 .
- Schneider, K., Dargatz, C., Dobay, A., Schmid, V., Leonhardt, H., & Schermelleh, L.(2010).** Analysis of cell cycle dependent dynamics of Dnmt1 by FRAP and kinetic modeling (*in preparation*).
- Sermaidis, G., Papaspiliopoulos, O., Roberts, G. O., Beskos, A. and Fearnhead, P. (2013).** Markov chain Monte Carlo for exact inference for diffusions. *Scand. J. Stat.* 40(2), 294–321.
- Shoji, I.(1998).** Approximation of continuous time stochastic processes by a local linearization method. *Mathematics of Computation*, 67, 287-298.
- Shoji, I., & Ozaki, T.(1998a).** Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16, 733-752.
- Shoji, I., & Ozaki, T.(1998b).** A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, 85, 240-243.
- Smallman-Raynor, M., & Cliff, A.(2004).** Impact of infectious diseases on war. *Infectious Disease Clinics of North America*, 18, 341-368.
- Sørensen, H.(2001).** Discretely observed diffusions: Approximation of the continuous-time score function. *Scandinavian Journal of Statistics*, 28, 113-121.
- Sørensen, H.(2004).** Parametric inference for diffusion processes observed at discrete points in time: A survey. *International Statistical Review*, 72, 337-354.
- Sørensen, M.(2007).** Efficient estimation for ergodic diffusions sampled at high frequency. *CREATES Research Paper No. 2007-46*.
- Sørensen, H. (2004).** Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey. *Internat. Statist. Rev.* 72, 337–354.
- Sørensen, M. (2008).** Parametric inference for discretely sampled stochastic differential equations. *CREATES Research Paper No. 2008-18*.

- Stramer, O. Bognar, M. and Schneider P. (2010).** Bayesian inference of discretely sample markov processes with closed form likelihood expansions. *The journal of financial econometrics* 8:450-480. 232.
- Stramer, O. and Bognar, M. (2011).** Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Anal.* 6(2), 231–258.
- Stramer, O., & Yan, J. (2007).** Asymptotic of an efficient Monte Carlo estimation for the transition density of diffusion processes. *Methodology and Computing in Applied Probability*, 9, 483-496.
- Stratonovich, R. (1989).** Some Markov methods in the theory of stochastic processes in nonlinear dynamical systems. In F. Moss & P. McClintock (Eds.), *Noise in Nonlinear Dynamical Systems. Theory of Continuous Fokker-Planck Systems* (Vol. 1, p. 16-71). Cambridge: Cambridge University Press.
- Tanner, M., & Wong, W. (1987).** The calculation of posterior distributions by data augmentation (with comments). *Journal of the American Statistical Association*, 82, 528-546.
- Thompson, D., Foege, W., (1968).** Faith Tabernacle Smallpox Epidemic, Abakaliki, Nigeria. World Health Organization, 3, pp. 1–9.
- Tian, T., Burrage, K., Burrage, P., & Carletti, M. (2007).** Stochastic delay differential equations for genetic regulatory networks. *Journal of Computational and Applied Mathematics*, 205, 696-707.
- Tian, T., Xu, S., Gao, J., & Burrage, K.(2007).** Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics*, 23, 84-91.
- Tierney, L. (1994).** Markov Chains for exploring posterior distributions. *The Annals of Statistics*, 22, 1701-1728.
- Tuckwell, H. (1987).** Diffusion approximations to channel noise. *Journal of Theoretical Biology*, 127, 427-438.
- van Kampen, N.(1981a).** Itô versus Stratonovich. *Journal of Statistical Physics*, 24, 175-187.
- van Kampen, N.(1981b).** The validity of nonlinear Langevin equations. *Journal of Statistical Physics*, 25, 431-442.
- van der Meulen, F. and Schauer, M. (2016).** Bayesian estimation of incompletely observed diffusions. ArXiv e-prints .
- Van der Meulen, F. H., Schauer, M. and Van Zanten, J. H. (2014).** Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* 71, 615–632.
- Wang, F. (1977).** Gaussian approximation of some closed stochastic epidemic models. *Journal of Applied Probability*, 14, 221-231.
- Wiener, N.(1923).** Differential space. *Journal of Mathematics and Physics*, 2, 131-174.
- Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. (2015).** Improved bridge constructs for stochastic differential equations. ArXiv eprints .
- Wong, E., & Zakai, M. (1965).** On the convergence of ordinary integrals to stochastic integrals. *The Annals of Mathematical Statistics*, 36, 1560-1564.

- World Health Organization (WHO), (2003a).** Ebola hemorrhagic fever: disease outbreaks. (<http://www.who.int/disease-outbreak-news/disease/A98.4.htm>), accessed on October 17, 2003.
- World Health Organization.(2014).** *Ebola Virus Disease, West Africa -U pdate on 27 July 2014.*
- World Health Organization.(2014).** *Ebola virus disease update - West Africa, 08 August 2014.*
- Wang W, Ruan S (2004).** Simulating the SARS Outbreak in Beijing with Limited data. *J.theor Biol* 227:369-379.
- Yuan S.; Li B.; and Zhang W. (2011).** Analysis on an epidemic model with a ratio-dependent nonlinear incidence rate. *International journal of Biomathematics*, 4(02): 227-239.
- Yoshida, N.(1992).** Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41, 220-242.

Appendix

#####

A1: R Code for Simulation Study on Diffusion Bridges

#####

##SIR Epidemic Diffusion Bridges##

#Drift and Diffusion

alpha=function(x,theta)

{

a=c(-theta[1]*x[1]*x[2],theta[1]*x[1]*x[2]-theta[2]*x[2])

a

}

beta=function(x,theta)

{

b=matrix(c((theta[1]*x[1]*x[2]),(theta[1]*x[1]*x[2]),0,(theta[2]*x[2])),nrow=2,ncol=2))

b

}

#supplimentary

matfun=function(A,fun){

s=svd(A)

(s\$u%*%diag(fun(s\$d))%*%s\$v)

}

rmvn=function(mean,V){

eps=rnorm(dim(V)[1])

sq=matfun(V,sqrt)

mean+(sq%*%eps)

}

lmvnpdf=function(x,mean,V){

```

detv=det(V)

Vinv=solve(V)

-0.5*log(detv)-0.5*t(x-mean)%*%Vinv%*%(x-mean)
}

#Target and proposal

llike=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

for(i in 1:m){

mean=xmat[i,]+alpha(xmat[i,],theta)*dt

ll=ll+lmvnpdf(xmat[i+1,],mean,beta(xmat[i,],theta)*dt)

}

ll

}

#MDB likelihood (proposal)

likMDB=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

mean=((inter-t)*xmat[i,]+x1*dt)/(inter-t+dt)

var=((inter-t)/(inter-t+dt))*beta(xmat[i,],theta)*dt

ll=ll+lmvnpdf(xmat[i+1,],mean,var)

}

ll

```

```

}

##LB

llikLB=function(xmat,theta,dt,gamma){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

al=alpha(xmat[i,],theta)

bet=beta(xmat[i,],theta)

mean=xmat[i,]+al*dt+((dt)/((inter-t+dt)+gamma*((inter-t)^2)/(dt)))*(x1-xmat[i,]-al*(inter-t*dt))

var=bet*dt-((dt^2)/((inter-t+dt)+gamma*((inter-t)^2)/(dt)))*bet

ll=ll+lmvnpdf(xmat[i+1,],mean,var)

}

ll

}

##GP

llikGP=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

bet=beta(xmat[i,],theta)

al=alpha(xmat[i,],theta)

```

```

out=LV_solved((inter-t+dt),theta,xmat[i,])

mu_t=out[[1]]

psi=out[[2]]

G=out[[3]]

mean=a1+bet%*%t(G)%*%solve(G%*%psi%*%t(G)%*%(x1-mu_t))

mean=xmat[i,]+mean*dt

var=bet*dt

ll=ll+lmvnpdf(xmat[i+1,],mean,var)

}

ll

}

##MGP

llikMGP=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

bet=beta(xmat[i,],theta)

a1=alpha(xmat[i,],theta)

out=LV_solved((inter-t+dt),theta,xmat[i,])

mu_t=out[[1]]

psi=out[[2]]

G=out[[3]]

mean=a1+bet%*%t(G)%*%solve(G%*%psi%*%t(G)%*%(x1-mu_t))

mean=xmat[i,]+mean*dt

```

```

var=((inter-t)/(inter-t+dt))*bet*dt
ll=ll+lmvnpdf(xmat[i+1,],mean,var)
}
ll
}

#path proposals(single path)
MDBpath=function(x0,x1,inter,theta,m){
  dt=inter/m
  cpath=matrix(0,ncol=length(x0),nrow=m+1)
  cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)
  cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)
  prop=cpath
  for(i in 1:(m-1)){
    t=i*dt
    mean=((inter-t)*prop[i,]+x1*dt)/(inter-t+dt)
    var=((inter-t)/(inter-t+dt))*beta(prop[i,],theta)*dt
    prop[i+1,]=rmvn(mean,var)
  }
  prop
}

##
LBpath=function(x0,x1,inter,theta,m,gamma){
  dt=inter/m
  cpath=matrix(0,ncol=length(x0),nrow=m+1)
  cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)
  cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)
  prop=cpath

```

```

for(i in 1:(m-1)){
  t=i*dt
  al=alpha(prop[i,],theta)
  bet=beta(prop[i,],theta)
  mean=prop[i,]+al*dt+((dt))/(((inter-t+dt)+gamma*((inter-t)^2)/(dt))*(x1-prop[i,])-al*(inter-t+dt)
  var=bet*dt-((dt^2))/(((inter-t+dt)+gamma*((inter-t)^2)/(dt)))*bet
  prop[i+1,]=rmvn(mean,var)
}
prop
}

#Metropolis-Hastings samplers

MDBMH=function(x0,x1,inter,theta,m,it,plot=FALSE,x1l,x1u,x2l,x2u){
  count=0
  col=rainbow(it)
  dt=inter/m
  cpath=matrix(0,ncol=length(x0),nrow=m+1)
  cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)
  cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)
  for(j in 1:it){
    prop=MDBpath(x0,x1,inter,theta,m)
    if(j==1){lap=1}
    else{lap=llike(prop,theta,dt)-llike(cpath,theta,dt)+likMDB(cpath,theta,dt)-likMDB(prop,theta,dt)}
    if(log(runif(1))<lap){
      cpath=prop
      count=count+1
    }
    if(plot==T){
      if(count==1){

```

```

split.screen(c(1,2))

screen(1)

par(mar=c(5.2,4.9,1.1,1.1))

par(cex.axis=1)

plot(ts(prop[,1],start=0,deltat=dt),ylim=c(x1l,x1u),ylab="",xlab="",main="",col=col[j])

mtext('Time',side=1,line=3.5,cex=1)

mtext(expression("*S[t]*"),side=2,line=3.1,cex=1,las=1)

screen(2)

par(mar=c(5.2,4.9,1.1,1.1))

par(cex.axis=1)

plot(ts(prop[,2],start=0,deltat=dt),ylim=c(x2l,x2u),ylab="",xlab="",main="",col=col[j])

mtext('Time',side=1,line=3.5,cex=1)

mtext(expression("*I[t]*"),side=2,line=3.1,cex=1,las=1)

}

else

{

screen(1,F)

par(mar=c(5.2,4.9,1.1,1.1))

plot(ts(prop[,1],start=0,deltat=dt),col=col[j],ylim=c(x1l,x1u),axes=FALSE,ylab="",xlab=")

screen(2,F)

par(mar=c(5.2,4.9,1.1,1.1))

plot(ts(prop[,2],start=0,deltat=dt),col=col[j],ylim=c(x2l,x2u),axes=FALSE,ylab="",xlab=")

}

}

}

}

close.screen(all=TRUE)

```

```

count/it
}

LBMH=function(x0,x1,inter,theta,m,gamma,it,plot=FALSE,x1l,x1u,x2l,x2u){
count=0
col=rainbow(it)
dt=inter/m
cpath=matrix(0,ncol=length(x0),nrow=m+1)
cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)
cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)
for(j in 1:it){
prop=LBpath(x0,x1,inter,theta,m,gamma)
if(j==1){lap=1}
else{lap=llike(prop,theta,dt)-llike(cpath,theta,dt)+llikLB(cpath,theta,dt,gamma)-
llikLB(prop,theta,dt,gamma)}
if(log(runif(1))<lap){
cpath=prop
count=count+1
if(plot==T){
if(count==1){
split.screen(c(1,2))
screen(1)
par(mar=c(5.2,4.9,1.1,1.1))
par(cex.axis=1)
plot(ts(prop[,1],start=0,deltat=dt),ylim=c(x1l,x1u),ylab="",xlab="",main="")
mtext('Time',side=1,line=3.5,cex=1)
mtext(expression("*x[1]*"),side=2,line=3.1,cex=1,las=1)
screen(2)
par(mar=c(5.2,4.9,1.1,1.1))

```

```

par(cex.axis=1)

plot(ts(prop[,2],start=0,deltat=dt),ylim=c(x2l,x2u),ylab="",xlab="",main=")

mtext('Time',side=1,line=3.5,cex=1)

mtext(expression("*x[2]*"),side=2,line=3.1,cex=1,las=1)

}

else

{

screen(1,F)

par(mar=c(5.2,4.9,1.1,1.1))

plot(ts(prop[,1],start=0,deltat=dt),ylim=c(x1l,x1u),axes=FALSE,ylab="",xlab=")

screen(2,F)

par(mar=c(5.2,4.9,1.1,1.1))

plot(ts(prop[,2],start=0,deltat=dt),ylim=c(x2l,x2u),axes=FALSE,ylab="",xlab=")

}

}

}

}

close.screen(all=TRUE)

count/it

}

#####

##Metropolis-Hastings samplers for median values for T=1

##MDB

MDBMH(x0=c(71,79),x1=c(96,71),inter=1,theta=c(0.05,0.00025,0.003),m=5,it=100,plot=TRUE,x1l=65
,x1u=105,x2l=65,x2u=85)

##LB

```

```
LBMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,gamma=0.01,it=100,plot=TRUE,x1l=65,x1u=105,x2l=65,x2u=85)
```

```
##GP
```

```
GPMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,it=100,plot=TRUE,x1l=65,x1u=105,x2l=65,x2u=85)
```

```
##MGP
```

```
MGPMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,it=100,plot=TRUE,x1l=65,x1u=105,x2l=65,x2u=85)
```

```
#####
```

A2: R Code for Simulation Study on Diffusion Bridges

```
#####
```

##SEIR Epidemic Diffusion Bridges##

```
#drift and diffusion
```

```
alpha=function(x,theta)
```

```
{
```

```
a=c(-theta[1]*x[1]*x[3],theta[1]*x[1]*x[3]-theta[2]*x[2],theta[2]*x[2]-theta[3]*x[3])
```

```
a
```

```
}
```

```
beta=function(x,theta)
```

```
{
```

```
b=matrix(c(theta[1]*x[1]*x[3],-theta[1]*x[1]*x[3],0,-theta[1]*x[1]*x[3],theta[1]*x[1]*x[3]+theta[2]*x[2],-theta[2]*x[2],0,-theta[2]*x[2],theta[2]*x[2]+theta[3]*x[3]),nrow=3,ncol=3)
```

```
b}
```

```
#supplimentary
```

```
matfun=function(A,fun){
```

```

s=svd(A)

(s$u%*%diag(fun(s$d))%*%s$v)

}

rmvn=function(mean,V){
  eps=rnorm(dim(V)[1])
  sq=matfun(V,sqrt)
  mean+(sq%*%eps)
}

lmvnpdf=function(x,mean,V){
  detv=det(V)
  Vinv=solve(V)
  -0.5*log(detv)-0.5*t(x-mean)%*%Vinv%*(x-mean)
}

##LNA

LV_solved=function(t,parms,x_0){
  x_start=c(G=matrix(c(1,0,0,1),ncol=2),psi=matrix(c(0,0,0,0),ncol=2),eta=matrix(c(x_0[1],x_0[2]),ncol=1))
  times=c(0,t)
  out=lsoda(x_start,times,LVMod,parms)
  G=matrix(c(out[2,2],out[2,3],out[2,4],out[2,5]),ncol=2)
  psi=matrix(c(out[2,6],out[2,7],out[2,8],out[2,9]),ncol=2)
  eta=c(out[2,10],out[2,11])
  list(eta,psi,G)
}

LVMod=function(t,x,parms)
{
  G=matrix(c(x[1],x[2],x[3],x[4]),ncol=2)
  psi=matrix(c(x[5],x[6],x[7],x[8]),ncol=2)

```

```

eta=matrix(c(x[9],x[10]),ncol=1)

mat=matrix(c(parms[1]-parms[2]*eta[2],parms[2]*eta[2],-parms[2]*eta[1],parms[2]*eta[1]-
parms[3]),nrow=2,ncol=2)

dG=mat%%G

dpsi=solve(G)%%beta(eta,parms)%%solve(t(G))

deta=alpha(eta,parms)

res=c(dG,pdsi,deta)

list(res)

}

#Target and proposal

llike=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

for(i in 1:m){

mean=xmat[i,]+alpha(xmat[i,],theta)*dt

ll=ll+lmvnpdf(xmat[i+1,],mean,beta(xmat[i,],theta)*dt)

}

ll

}

#MDB likelihood (proposal)

likMDB=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

mean=((inter-t)*xmat[i,]+x1*dt)/(inter-t+dt)

```

```

var=((inter-t)/(inter-t+dt))*beta(xmat[i,],theta)*dt

ll=ll+lmvnpdf(xmat[i+1,],mean,var)

}

ll

}

##LB

llikLB=function(xmat,theta,dt,gamma){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

for(i in 1:(m-1)){

t=i*dt

al=alpha(xmat[i,],theta)

bet=beta(xmat[i,],theta)

mean=xmat[i,]+al*dt+((dt)/((inter-t+dt)+gamma*((inter-t)^2)/(dt)))*(x1-xmat[i,]-al*(inter-t*dt))

var=bet*dt-((dt^2)/((inter-t+dt)+gamma*((inter-t)^2)/(dt)))*bet

ll=ll+lmvnpdf(xmat[i+1,],mean,var)

}

ll

}

##GP

llikGP=function(xmat,theta,dt){

ll=0

m=length(xmat[,1])-1

x1=xmat[m+1,]

inter=m*dt

```

```

for(i in 1:(m-1)){
  t=i*dt
  bet=beta(xmat[i,],theta)
  al=alpha(xmat[i,],theta)
  out=LV_solved((inter-t+dt),theta,xmat[i,])
  mu_t=out[[1]]
  psi=out[[2]]
  G=out[[3]]
  mean=al+bet%*%t(G)%*%solve(G%*%psi%*%t(G)%*%(x1-mu_t))
  mean=xmat[i,]+mean*dt
  var=bet*dt
  ll=ll+lmvnpdf(xmat[i+1,],mean,var)
}
ll
}

##MGP
llikMGP=function(xmat,theta,dt){
  ll=0
  m=length(xmat[,1])-1
  x1=xmat[m+1,]
  inter=m*dt
  for(i in 1:(m-1)){
    t=i*dt
    bet=beta(xmat[i,],theta)
    al=alpha(xmat[i,],theta)
    out=LV_solved((inter-t+dt),theta,xmat[i,])
    mu_t=out[[1]]

```

```

psi=out[[2]]
G=out[[3]]
mean=a1+bet%*%t(G)%*%solve(G%*%psi%*%t(G)%*%(x1-mu_t))
mean=xmat[i,]+mean*dt
var=((inter-t)/(inter-t+dt))*bet*dt
ll=ll+lmvnpdf(xmat[i+1,],mean,var)
}
ll
}

#path proposals(single path)
MDBpath=function(x0,x1,inter,theta,m){
dt=inter/m
cpath=matrix(0,ncol=length(x0),nrow=m+1)
cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)
cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)
cpath[1:(m+1),3]=seq(x0[3],x1[3],length.out=m+1)
prop=cpath
for(i in 1:(m-1)){
t=i*dt
mean=((inter-t)*prop[i,]+x1*dt)/(inter-t+dt)
var=((inter-t)/(inter-t+dt))*beta(prop[i,],theta)*dt
prop[i+1,]=rmvn(mean,var)
}
prop
}

##

LBpath=function(x0,x1,inter,theta,m,gamma){

```

```

dt=inter/m

cpath=matrix(0,ncol=length(x0),nrow=m+1)

cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)

cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)

cpath[1:(m+1),3]=seq(x0[3],x1[3],length.out=m+1)

prop=cpath

for(i in 1:(m-1)){

t=i*dt

al=alpha(prop[i,],theta)

bet=beta(prop[i,],theta)

mean=prop[i,]+al*dt+((dt))/((inter-t+dt)+gamma*((inter-t)^2)/(dt))*(x1-prop[i,])-al*(inter-t+dt)

var=bet*dt-((dt^2))/((inter-t+dt)+gamma*((inter-t)^2)/(dt))*bet

prop[i+1,]=rmvn(mean,var)

}

prop

}

#Metropolis-Hastings samplers

MDBMH=function(x0,x1,inter,theta,m,it,plot=FALSE,x1l,x1u,x2l,x2u,x3l,x3u){

count=0

col=rainbow(it)

dt=inter/m

cpath=matrix(0,ncol=length(x0),nrow=m+1)

cpath[1:(m+1),1]=seq(x0[1],x1[1],length.out=m+1)

cpath[1:(m+1),2]=seq(x0[2],x1[2],length.out=m+1)

cpath[1:(m+1),3]=seq(x0[3],x1[3],length.out=m+1)

for(j in 1:it){

prop=MDBpath(x0,x1,inter,theta,m)

```

```

if(j==1){lap=1}

else{lap=llike(prop,theta,dt)-llike(cpath,theta,dt)+llikMDB(cpath,theta,dt)-llikMDB(prop,theta,dt)}

if(log(runif(1))<lap){

cpath=prop

count=count+1

if(plot==T){

if(count==1){

par(5.2,4.9,1.1,1.1)

plot(ts(prop[,1],start=0,deltat=dt),ylim=c(x1l,x1u),ylab='X(t)',xlab='',main='Modified Diffusion Bridge
for SEI-process',col=col[j])

mtext('Time',side=1,line=3.5,cex=1)

lines(ts(prop[,2],start=0,deltat=dt),ylim=c(x2l,x2u),ylab='X(t)',xlab='',main='',col=col[j])

lines(ts(prop[,3],start=0,deltat=dt),ylim=c(x3l,x3u),ylab='X(t)',xlab='',main='',col=col[j])

}

else

{

lines(ts(prop[,1],start=0,deltat=dt),ylim=c(x1l,x1u),axes=FALSE,ylab='',xlab='',col=col[j])

lines(ts(prop[,2],start=0,deltat=dt),ylim=c(x2l,x2u),axes=FALSE,ylab='',xlab='',col=col[j])

lines(ts(prop[,3],start=0,deltat=dt),ylim=c(x3l,x3u),axes=FALSE,ylab='',xlab='',col=col[j])

}

}

}

close.screen(all=TRUE)

count/it

}

#####

##Metropolis-Hastings samplers for median values for T=1

```

```
##MDB
```

```
MDBMH(x0=c(80,80,80),x1=c(96.816,71.926,51.675),inter=1,theta=c(0.00065,0.03,0.03),m=7,it=100,  
plot=TRUE,x1l=50,x1u=100,x2l=50,x2u=100,x3l=50,x3u=100)
```

```
legend(0,100,legend=c('St','Et','It'),col=c(2,5,1),lty=1,cex=0.5)
```

```
##LB
```

```
LBMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,gamma=0.01,it=100,plot  
=TRUE,x1l=65,x1u=105,x2l=65,x2u=85)
```

```
##GP
```

```
GPMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,it=100,plot=TRUE,x1l=6  
5,x1u=105,x2l=65,x2u=85)
```

```
##MGP
```

```
MGPMH(x0=c(71,79),x1=c(96.816,71.926),inter=1,cvec=c(0.5,0.0025,0.3),m=5,it=100,plot=TRUE,x1l=  
65,x1u=105,x2l=65,x2u=85)
```

```
MDBMH(x0=c(80,80,80),x1=c(96.816,71.926,51.675),inter=1,theta=c(0.00065,0.03,0.03),m=7,it=100,  
plot=TRUE,x1l=50,x1u=100,x2l=50,x2u=100,x3l=50,x3u=100)
```

```
legend(0,100,legend=c('St','Et','It'),col=c(2,5,1),lty=1,cex=0.5)
```

```
#####
```

A3: R Code for Simulation Study on SIR Epidemic Parameter Estimates

```
#####
```

```
library(mvtnorm)
```

```
ff=function(y,mu,sigma){
```

```
  likelihood=sum(dmvnorm(y,mu,sigma,log=T)+dgamma(mu[1],10,20,log=T)+dgamma(mu[2],  
  10,20,log=T))
```

```
  return(likelihood)
```

```

        ff
    }

    sigma=matrix(c(1,0,1,0),2,2)

    T=1000000

    s=0.009;rej=0;rej2=0

    THETA=matrix(0,T,2)

    theta1=runif(1,0,1)

    theta2=runif(1,0,1)

    mu=c(theta1,theta2)

    THETA[1,]=mu

    for(t in 2:T){

        THETA[t,1]=rnorm(1,theta1,s)

        new.theta1=THETA[t,];theta1=THETA[t-1,]

        ratio=ff(new.theta1,mu,sigma)-ff(theta1,mu,sigma)

        alpha=min(1,ratio)

        u=runif(T)

        if(log(u)<alpha)THETA[t,]=new.theta1

        else{THETA[t,]=theta1;rej=rej+1}

        THETA[t,2]=rnorm(1,theta2,s)

        new.theta2=THETA[t,];theta2=THETA[t-1,]

        ratio=ff(new.theta2,mu,sigma)-ff(theta2,mu,sigma)

        alpha=min(1,ratio)

        u=runif(T)

        if(log(u)<alpha)THETA[t,]=new.theta2

        else{THETA[t,]=theta2;rej2=rej2+1}

    }

    library(coda)

```

```

colnames(THETA)=c('Beta','Gamma')

summary(mcmc(THETA))

Rej=rej/T

AccepRate=1-Rej

cat('Acceptance Rate',AccepRate,'\n')

effectiveSize(THETA)

acf(THETA[,2],100,main='Beta',col=3)

legend(75,1,legend='m = 5')

library(ggplot2)

Iterations=1:T

parameterBeta=THETA[,1];parameterGamma=THETA[,2]

B=rep('Beta',T);G=rep('Gamma',T)

Variables=c(B,G)

parameter=c(parameterBeta,parameterGamma)

data=data.frame(Iterations,parameter,Variables)

plt1=ggplot(data,aes(Iterations,parameter))+geom_line()+geom_hline(yintercept=0.4,colour=2,size=
0.8)+facet_grid(Variables~.)+ylab('Parameter Value')+ggtitle('SEIR
MODEL')+theme(panel.background=element_rect(fill='grey80'))

parameter=data.frame(parameterBeta,parameterGamma)

plt2=ggplot(data,aes(parameter))+geom_histogram(bins=30)+facet_grid(Variables~.)+xlab('Density
plot')+ggtitle('SIR MODEL')+theme(panel.background=element_rect(fill='grey80'))

plt3=ggplot(data,aes(parameter))+geom_line(stat='density',size=1)+facet_grid(Variables~.)+xlab('De
nsity plot')+ggtitle('SIR MODEL') +theme(panel.background= element_rect (fill='grey80'))

qacf <- function(x, conf.level = 0.95) {

ciline <- qnorm((1 - conf.level)/2)/sqrt(length(x))

p<- acf(x, plot = FALSE)

bACF<- data.frame(lag=p$lag, acf=p$acf)

q <- ggplot(bACF, aes(lag,acf))+geom_bar(stat = "identity",position = "identity")+ylab('acf')

q <- q + geom_hline(yintercept = -ciline, color = "blue",size = 0.2)

```

```

q <- q + geom_hline(yintercept = ciline, color = "blue",size = 0.2)
q <- q + geom_hline(yintercept = 0, color = "red", size = 0.3)
return(q)
}

q1=qacf(parameterBeta)+theme(panel.background=element_rect(fill='grey80'))
q2=qacf(parameterGamma)+theme(panel.background=element_rect(fill='grey80'))

```

```
#####
```

A4: R Code for Simulation Study on SEIR Epidemic Parameter Estimates

```
#####
```

```

library(mvtnorm)

ff=function(y,mu,sigma){

    likelihood=sum(dmvnorm(y,mu,sigma,log=T)+dgamma(mu[1],10,20,log=T)+dgamma(mu[2],
    10,20,log=T)+dgamma(mu[3],10,20,log=T))

    return(likelihood)

    ff

}

sigma=matrix(c(1,0,0,0,1,0,0,0,1),3,3)

T=1000000

s=0.009;rej=0;rej2=0;rej3=0

THETA=matrix(0,T,3)

theta1=runif(1,0,1)

theta2=runif(1,0,1)

theta3=runif(1,0,1)

mu=c(theta1,theta2,theta3)

THETA[1,]=mu

for(t in 2:T){

    THETA[t,1]=rnorm(1,theta1,s)

```

```

new.theta1=THETA[t,];theta1=THETA[t-1,]
ratio=ff(new.theta1,mu,sigma)-ff(theta1,mu,sigma)
alpha=min(1,ratio)
u=runif(T)
if(log(u)<alpha)THETA[t,]=new.theta1
else{THETA[t,]=theta1;rej=rej+1}
THETA[t,2]=rnorm(1,theta2,s)
new.theta2=THETA[t,];theta2=THETA[t-1,]
ratio=ff(new.theta2,mu,sigma)-ff(theta2,mu,sigma)
alpha=min(1,ratio)
u=runif(T)
if(log(u)<alpha)THETA[t,]=new.theta2
else{THETA[t,]=theta2;rej2=rej2+1}
THETA[t,3]=rnorm(1,theta3,s)
new.theta3=THETA[t,];theta3=THETA[t-1,]
ratio=ff(new.theta3,mu,sigma)-ff(theta3,mu,sigma)
alpha=min(1,ratio)
u=runif(T)
if(log(u)<alpha)THETA[t,]=new.theta3
else{THETA[t,]=theta3;rej3=rej3+1}
}
library(coda)
colnames(THETA)=c('Beta','Gamma','Alpha')
summary(mcmc(THETA))
Rej=rej/T
AccepRate=1-Rej
cat('Acceptance Rate',AccepRate,'\n')

```

```

effectiveSize(THETA)

acf(THETA[,3],100,main='Beta',col=3)

legend(75,1,legend='m = 5')

library(ggplot2)

Iterations=1:T

parameterBeta=THETA[,1];parameterGamma=THETA[,2];parameterAlpha=THETA[,3]

B=rep('Beta',T);G=rep('Gamma',T);A=rep('Alpha',T)

Variables=c(B,G,A)

parameter=c(parameterBeta,parameterGamma,parameterAlpha)

data=data.frame(Iterations,parameter,Variables)

plt1=ggplot(data,aes(Iterations,parameter))+geom_line()+geom_hline(yintercept=0.4,colour=2,size=
0.8)+facet_grid(Variables~.)+ylab('Parameter Value')+ggtitle('SEIR
MODEL')+theme(panel.background=element_rect(fill='grey80'))

parameter=data.frame(parameterBeta,parameterGamma,parameterAlpha)

plt2=ggplot(data,aes(parameter))+geom_histogram(bins=30)+facet_grid(Variables~.)+xlab('Density
plot')+ggtitle('SEIR MODEL')+theme(panel.background=element_rect(fill='grey80'))

plt3=ggplot(data,aes(parameter))+geom_line(stat='density',size=1)+facet_grid(Variables~.)+xlab('De
nsity plot')+ggtitle('SEIR MODEL')+theme(panel.background=element_rect(fill='grey80'))

qacf <- function(x, conf.level = 0.95) {

ciline <- qnorm((1 - conf.level)/2)/sqrt(length(x))

p<- acf(x, plot = FALSE)

bACF<- data.frame(lag=p$lag, acf=p$acf)

q <- ggplot(bACF, aes(lag,acf))+geom_bar(stat = "identity",position = "identity")+ylab('acf')

q <- q + geom_hline(yintercept = -ciline, color = "blue",size = 0.2)

q <- q + geom_hline(yintercept = ciline, color = "blue",size = 0.2)

q <- q + geom_hline(yintercept = 0, color = "red", size = 0.3)

return(q)

}

q1=qacf(parameterBeta)+theme(panel.background=element_rect(fill='grey80'))

```

```
q2=qacf(parameterGamma)+theme(panel.background=element_rect(fill='grey80'))
```

```
q3=qacf(parameterAlpha)+theme(panel.background=element_rect(fill='grey80'))
```

```
#####
```

A5: R Code for Simulation Study on Acceptance Probability Plots

```
#####
```

Acceptance Probability

```
library(ggplot2)
```

```
Bridge=c('MDB','LB','GP','MGP','MDB','LB','GP','MGP','MDB','LB','GP','MGP','MDB','LB','GP','MGP')
```

```
Acc=c(0.70,0.60,0.759,0.80,0.689,0.605,0.753,0.826,0.50,0.450,0.648,0.856,0.320,0.23,0.557,0.860)
```

```
m=c(2,2,2,2,5,5,5,5,20,20,20,20,50,50,50,50)
```

```
data=data.frame(m,Acc,Bridge)
```

```
AR=ggplot(data,aes(m,Acc,colour=Bridge,group=Bridge))+geom_line(size=1)+geom_point(shape=8,colour='black')+ylim(0,1)+ggtitle('Acceptance Probability')+theme(panel.background=element_rect(fill='grey80'))+theme(axis.text=element_text(size=12))+theme(legend.text=element_text(size=15),legend.title=element_text(size=15,face='bold'))
```

```
AR+theme(legend.position='bottom')
```