Sultan Moulay Slimane University
Faculty of Sciences and Techniques

Doctoral Thesis

Parameter Inference of HIV Stochastic Diffusion Epidemic Models with application to Morocco data-set

---

Presented by:
Abdellah Abou-Bakre
Date of final oral examination:
22/ 07 / 2019

Department of Mathematical Sciences
Laboratory of Mathematics and Applications

specialty: MATHEMATICS
Option : PROBABILITIES AND STATISTICS

The dissertation is approved by the following members of the Final Oral Committee:

Mr. Ahmed ZEGHAL Faculty of science and techniques, Sultan Moulay Slimane University Beni Mellal	**Professor**	**President**
Mr. Taib ZIAD University of Technology of Chalmers, Sweden	**Professor**	**Rapporteur**
Mr. Tewfik KERNANE University of Science and Technology Houari Boumediene, Algiers, Alger	**Professor**	**Rapporteur**
Mr. Khalid HILAL Faculty of science and techniques, Sultan Moulay Slimane University Beni Mellal	**Professor**	**Examiner**
Mr. Hamid El MAROUFY Faculty of science and techniques, Sultan Moulay Slimane University Beni Mellal	**Associate Professor**	**Supervisor**
Mr. Abdelaziz QAFFOU Superior School of Technology, Sultan Moulay Slimane University Beni Mellal	**Assistant Professor**	**Invited**

To my mother, my father, my wife JIHAD, *my little daughter* JOUD, *my brothers, my sisters, my family and my friends.*

Preface

> "*All epidemiology, conceived as it is with the variation of disease from time to time and from place to place, must be considered mathematically, however many variables are implicated, if it is to be considered scientifically at all*".(Sir Ronald Ross[1], Ross [106])

Mathematical epidemiology has a long history dating back to Daniel Bernoulli's model of smallpox in 1760 (see Bernoulli [18]). Much of the basic theory was developed as early as 1900 and has been progressing steadily since that time. Recently, models for evaluating the effect of control measures have been used to inform policy development, particularly with regard to the foot and mouth disease outbreak in Great Britain in 2001. Since the first cases of AIDS (Acquired Immune Deficiency Syndrome) epidemic caused by the Human Immunodeficiency Virus (HIV), were identified in 1981, it has become one of the most urgent public-health problems in developing countries. A general interest of the use of mathematical models is to predict the course of an infectious disease and compare the effects of different control strategies.

Mathematical epidemiology differs from most sciences because it does not lend itself to experimental validation of models. Experiences are generally impossible and would probably be unethical. This places great importance on mathematical models as a possible tool for comparing planning strategies for an expected epidemic or pandemic and for dealing with a real-time epidemic.

This thesis consist of HIV dynamic model in an heterosexual population. Firstly, we formulate the stochastic diffusion approximations process associated to the discrete model using the convergence of the master equation. Our main aim is to infer the model parameters of interest. To deal with this task, we use two approaches for estimating these parameters. The one approach consist to use a contrast function associated to the likelihood derived from the approximation of the continuous process using Euler-Maruyama scheme. The consistency and asymptotic normality of the estimator are well established. In the other contribution, we use a Bayesian approach with MCMC methods adopting the data-augmentation technique. We prove that the posterior distribution follows to a GIG density, and we give an algorithm to estimate the model parameters. The theoretical results are illustrated by numerical simulations. A real application to Morocco's case will be discussed.

This dissertation is composed by six chapters, conclusions and future directions and some supplementary proofs in the appendices; the first chapter is a general introduction to the topic of

[1]Sir Ronald Ross (13 May 1857 – 16 September 1932) was a British medical doctor who received the Nobel Prize for Physiology or Medicine in 1902 for his work on the transmission of malaria. His discovery of the malarial parasite in the gastrointestinal tract of a mosquito in 1897 proved that malaria was transmitted by mosquitoes, and laid the foundation for the method of combating the disease.

this thesis. The necessary mathematical background varies from chapter to chapter, but the essential materials are given in Chapter 2.The third chapter describes the studied model and gives the detailed technique to approximate this model by a diffusion process. The fourth and fifth chapters constitute our main contribution to parameters estimation. While the sixth chapter is concerned with numerical simulations and real application to Morocco's data-set.

Acknowledgements

Firstly, I would like to express my immense thankful and sincere gratitude to my supervisor Prof. Hamid El Maroufy for the continuous support of my doctorate study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I would like to express my gratitude to professor Ahmed Zeghal for the honor he brought to this work by accepting to chair the oral examination committee.

I am indebted to professor Taib Ziad of the Chalmers University of Technology (Sweden), professor Tewfik Kernane of Algiers university (Algeria) and professor Abdelkrim Merbouha of Sultan Moulay Slimane university (Morocco) for the interest they have shown to my work and the honor of reporting and sitting on the examination Board.

I would like to thank professor Khalid Hilal for agreeing to examine my work and to be a part of the examination committee of my doctorate.

I am thankful to professor Abdelaziz Qaffou to be a member of the oral committee of my doctorate.

Many thanks to Mr. Redouane Totss and Mr. Hamadi Ghafir for their proofreading of this thesis and for their linguistic support.

I am very thankful to the lab of mathematics and Application (LMA lab), in which I was a member during the years of preparing and writing my thesis. I thank my fellow lab-mates for the stimulating discussions, for the times we were working together, and for all the fun we have had during the years of my thesis.

Obviously, none of this would have been possible without the support of my family. I would like to thank my parents, my brothers, my sisters and especially my wife for supporting me spiritually and morally throughout writing this thesis and in my life in general.

Contents

List of Figures

List of Tables

Symbols and Abbreviations

Symbols

Ω	the universal set or state space or sample space
$\mathcal{F}$	σ-algebra on Ω
$(\Omega, \mathcal{F})$	measurable space
$\mathbb{P}$	probability measure on $(\Omega, \mathcal{F})$
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
$(\mathcal{F}_t)_{t\geq 0}$	filtration
$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$	filtered probability space
$\mathbb{E}$	mathematical expectation
$\pi(.)$	posterior density
$p(.)$	prior density
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^d$	the set of ordered d-tuples $(x_1, x_2, \ldots, x_d)$ for $d \in \mathbb{N}^\star$
$\mathbb{N}$	the set of natural integers
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and covariance matrix σ^2
$\delta(A)$	the Dirac delta function $\delta(A)$ equal 1 if $0 \in A$ and 0 otherwise
$\chi_A(x)$	indicator function of the set A
$L^2_{\mathbb{P}}(\Omega)$	the class of real-valued square integrable functions
$\mathcal{B}$	the Borel σ-algebra
$(B_t)_{t\geq 0}$	Brownian motion
$\sim$	equivalent
$\propto$	proportional
$I(\theta)$	Fisher information matrix

Abbreviations

SDE	stochastic differential equation
ODE	ordinary differential equation
a.s.	almost surely
MCE	minimun contrast estimator
MCMC	Markov Chain Monte Carlo
MLE	maximum likelihood estimator
HIV	human immunodeficiency virus
AIDS	acquired immune deficiency syndrome

Chapter 1

General introduction

1.1 Generalities

From the earliest times to the present, epidemics[1] and pandemics[2] have affected human history in multiple ways: demographically, culturally, politically, financially, and biologically. Humans have never known a time in history when epidemics did not loom large. The Black Death or plague burst in Europe in 1348, and is estimated to have killed over 25 million people in just five years (see, Siettos and Russo [112]). The pandemic influenza virus of 1918–1919 swept through America, Europe, Asia, and Africa smashing the globe: the death toll was around 40 million people. In the next decades: the 1957 and the 1963 influenza pandemics resulted to two and one million deaths respectively (World Health Organization CDC [27]). In the last decades emerging and re-emerging epidemics such as Acquired Immune Deficiency Syndrome (AIDS), measles, malaria, Influenza and tuberculosis cause death to millions of people each year. According to the UNAIDS [126] report on the global AIDS epidemic, an estimated more than 36.9 million people, including 1.8 millions children, were living with Human Immunodeficiency Virus (HIV) worldwide at the end of 2017, while the related deaths is around one million and new infections were 1.8 millions.

To describe and explain the behavior or results of what happening in the real world; or to predict the future behaviors or results, the mathematical modeling is the key tool for doing these tasks.

“ Modeling is an attempt to describe, in a precise way, an understanding of the elements of a system of interest, their states, and their interactions with other elements. The model should be sufficiently detailed and precise so that it can in principle be used to simulate the behavior of the system on a computer. Often the most basic aim is to make clear the current state of knowledge regarding a particular system, by attempting to be precise about the elements involved and the interactions between them. Doing this can be a particularly effective way of

[1] An epidemic is generally considered to be an unexpected, widespread rise in disease incidence at a given time.

[2] A pandemic is best thought of as a very large epidemic. Pandemic can be either discrete events or persistent pandemics (see, McMillen [88] for a detailed definitions and history of pandemics)

highlighting gaps in understanding. In addition, having a detailed model of a system allows people to test that their understanding of a system is correct, by seeing if the implications of their models are consistent with observed experimental data[1] ". Mathematical modeling can play an important role in helping to quantify possible disease control strategies by focusing on the important aspects of a disease, determining threshold quantities for disease survival, and evaluating the effect of particular control strategies.

The very first epidemiological model was formulated by Daniel Bernoulli in 1766, to analyze the mortality due to smallpox[2] in England and evaluating the impact of variolation on human life expectancy (see Bernoulli [18]). The last century has seen a rapid development and emergence of epidemic theory. The Kermack-McKendrick epidemic model introduced by Kermack et al. [72], which describes the relationship between susceptible, infected and immune individuals in a population. Kermack et al. [72] derived the celebrated threshold theorem, which is one of the key results in epidemiology. It predicts, depending on the transmission potential of the infection, the critical fraction of susceptibles in the population that must be exceeded if an epidemic is to occur. Following up the work of Kermack et al. [72], Bartlett [16] published a classic work, in which he examined models and data to expose the factors that determine disease persistence in large populations. It can be said that the first reference book on mathematical modeling of epidemiological systems was published by Bailey [13], which led to the recognition of the importance of modeling in public health decision-making.

After the publication of Bailey's book Bailey [13], the mathematical theory of epidemic models was progressively developed, but gained importance in the eighteenth of last century, with the advent of HIV epidemics. Since then, a very large number of Mathematical models have been created and employed to help explain, to study and to make predictions of epidemics behaviors.

1.2 Mathematical modeling of infectious diseases

Mathematical modeling of infectious disease is simplified by using the compartmental models techniques. The population is divided into non-intersecting compartments, with the assumption that every individual in the same compartment has the same characteristics. The first compartmental models proposed by Kermack et al. [72], a model known as the SIR epidemic model; a compartment model with three states:(Susceptible $\rightarrow$ Infectious $\rightarrow$ Removed). A susceptible

[1] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. 2nd ed. Boca Raton: CRC Press, Dec. 14, 2011. 363 pp.

[2] Smallpox was an infectious disease caused by one of two virus variants, variola major and variola minor. The last naturally occurring case was diagnosed in October 1977 and the World Health Organization (WHO) certified the global eradication of the disease in 1980. The risk of death following contracting the disease was about 30%, with higher rates among babies. Often those who survived had extensive scarring of their skin and some were left blind.

individual becomes a disease transmitter by changing to the infectious state. When an infectious individual is cured or in other ways cannot contribute to the spread of the disease anymore (e.g.dies or is isolated), it is regarded as recovered.

The disease spreads in a population, it divides the population into exclusives classes; There are many classes depending on the structure and the transmission way of the disease. We find in the literature:

- The class of individuals who are healthy but can contract the disease. These are called susceptible individuals or susceptibles. The size of this class is usually denoted by S.
- The class of individuals who have contracted the disease and are now sick with it and also infectious. The size of the class of infectious is denoted by I.
- The class of individuals who have recovered and cannot contract the disease again are called removed/recovered individuals. The class of recovered individuals is usually denoted by R.
- For many important infections there is a significant incubation period during which the individual has been infected but is not yet infectious themselves. During this period the individual is in compartment E (for exposed).
- For some disease, a class for vaccinated individuals is considered, it is usually denoted by V.

We focus our study, in the present thesis, on a particular SIR model in which the classes S and I are divided into two sub classes. Thus, we present here the general SIR model.

1.3 General SIR model

The SIR model is one of the simplest compartmental models, and many models are derivations of this basic form. The model consists of three compartments: S for the number of susceptibles, I for the number of infectives, and R for the number of recovered (or immune). This model is reasonably predictive for infectious diseases which are transmitted from human to human, and where recovery confers lasting resistance, such as measles, mumps, rubella and AIDS.

To formulate a model, we have to make assumptions to simplify reality. The first assumption for Kermack et al. [72] model is that infected individuals are also infectious. The second assumption of the model is that the total population size remains constant (equal N). The number of individuals in each of these classes changes with time t, that is, $S(t)$, $I(t)$, and $R(t)$ are real variables of time t. The total host population size N is the sum of the sizes of these three classes:

$$N = S(t) + I(t) + R(t).$$

In the literature of SIR model there are two approaches as presented in Bailey [13] to modeling the epidemic, the *deterministic model* and *stochastic model* (see, Brauer et al. [20] and Martcheva [85] for a survey of stochastic epidemic models). In the present thesis, we are interested to formulate a stochastic model of the dynamic of HIV/AIDS in a *closed dynamic heterosexual community* and estimate the keys parameters.

1.3.1 Basic deterministic model

In deterministic basic general epidemic model; we consider that the community is being homogeneously mixed and only susceptible individuals can get infected and, after having been infectious for some time, an individual recovers and becomes completely immune for the remainder of the study period. Finally, we assume there are no births, deaths, immigration or emigration during the study period; the community is said to be closed. A consequence of the assumptions is that individuals can only make two moves: from S to I and from I to R. Then, the deterministic basic general epidemic model is defined by the following set of differential equations:

$$\begin{cases} \dfrac{dS}{dt} = -\beta SI \\ \dfrac{dI}{dt} = \beta SI - \mu I \\ \dfrac{dR}{dt} = \mu I \end{cases} \tag{1.1}$$

Where the term βSI in Equation (1.1) comes from the fact that susceptibles must have contact with infectives in order to get infected, so the number of susceptibles who become infected and move to the class I is βSI, where β is the *infection rate*. The number of infectives who recover or die and leave the infected class I to recover class R is μI with μ is the *recovery rate*.

The ratio $R_0 = \dfrac{\beta}{\mu}$ is hence of fundamental importance and can be interpreted as the average number of new infections caused by an infectious individual before recovering. The ratio is often referred to as the *basic reproduction number* (a term with its origin in demography – the average number of individuals that one individual reproduces). When $R_0 > 1$ the epidemic takes off and when $R_0 < 1$ there is no (big) epidemic (the reader is referred to Bailey [13] for more descriptions and details).

1.3.2 Stochastic model

"Deterministic models are those in which there is no element of chance or uncertainty. As such, they can be thought to account for the mean trend of a process only. Stochastic models, on the other hand, account not only for the mean trend but also for the variance structure around it. In an epidemiological context, there are two main kinds of stochasticity: demographic and environmental. Demographic stochasticity reflects the fact that while all individuals may be

subject to the same possible events with the exact same probabilities, chance events may result in differences in the fates of individuals. When a phenomenon is the sum of a large number of small individual effects (as disease propagation in large population), the weak law of large numbers diminishes the effects of demographic stochasticity and a deterministic model becomes appropriate. In contrast, when the population is small, random events cannot be neglected and a stochastic model is necessary. Environmental stochasticity refers to the situation where there is variation in the probability associated with an event. Consequently, some parameters of stochastic models may be uncertain and characterized by a probability distribution instead of a constant value. For fixed starting values, a deterministic model will always produce the same result whereas a stochastic model will produce many different outputs, depending on the actual values the random variables take"[1]. A generalization of the initial simple deterministic epidemic model is given by the stochastic epidemic models. Needless to say, both deterministic and stochastic epidemic models have their important roles to play. Here, we give a brief to the simple stochastic SIR epidemic models in a closed population (see, O'Neill and Roberts [91], Brauer et al. [20], and El Maroufy et al. [38, 36] for more details).

Mathematically the model is defined as follows, we consider a closed population of $N + a$ individuals. At time t, there are $S(t)$ susceptibles, $I(t)$ ineffectives and $R(t) = N + a - S(t) - I(t)$ removed individuals with N and a are positive integers. At time $t = 0$ the population only contains susceptible and infected individuals with $S(0) = N$, $I(0) = a$. The epidemic process is thus completely determined by the bi-variate process $\{(S(t), I(t)), t \geq 0\}$, which is supposed to be a continuous-time Markov chain on the state space,

$$E = \{(s, i), 0 \leq s \leq N, \quad 0 \leq i \leq (N - s) + a\}$$

with the following transitions and associated probabilities from time t to $t + \delta t$:

$$\begin{array}{ccccc} & \text{Transition} & & & \text{Probability} \\ (s,i) & \longrightarrow & (s-1, i+1) & & \frac{\beta}{N}(s+1)(i-1)\delta t + o(\delta t) \\ (s,i) & \longrightarrow & (s, i-1) & & \mu(i+1)\delta t + o(\delta t) \\ (s,i) & \longrightarrow & (s,i) & & -\left(\frac{\beta}{N} si + \mu i\right)\delta t + o(\delta t) \end{array} \tag{1.2}$$

all other transitions having probability $o(\delta t)$, and the parameter β being known as the infection rate and μ as the removal rate.

For $(s, i) \in E$, we define

$$p_{(s,i)}(t) = Prob\{S(t) = s, I(t) = i\}.$$

[1] M. Tibayrenc. *Encyclopedia of Infectious Diseases: Modern Methodologies*. John Wiley & Sons, July 31, 2007. 807 pp.

It follows directly from Eq. (1.2) that these transition probabilities satisfy the set of Kolmogorov equations:

$$\frac{\partial p_{(s,i)}(t)}{\partial t} = p_{(s+1,i-1)}(t)\frac{\beta}{N}(i-1)(s+1) + p_{(s,i+1)}\mu(i+1) - p_{(s,i)}(t)\left(\frac{\beta}{N}is + \mu i\right). \tag{1.3}$$

for $(s,i) \in E$, with $p_{(s,i)}(t) \equiv 0$ if $(s,i) \notin E$ and $p_{(N,a)}(0) = 1$.

A Markov process with the above described dynamics determines the general stochastic epidemic.

Instead of using S and I, we use the normalized process $x(t) = S(t)/N$ and $y(t) = I(t)/N$. By setting $f(x,y) = N\beta xy = (\beta/N)SI$ and $g(x,y) = N\mu y = \mu I$ the Kolmogorov's equations (1.3) become

$$\frac{\partial}{\partial t}p(x,y,t) = f(x+\varepsilon, y-\varepsilon)p(x+\varepsilon, y-\varepsilon, t) + g(x, y+\varepsilon)p(x, y+\varepsilon, t) - [f(x,y)+g(x,y)]\,p(x,y,t), \tag{1.4}$$

where $\varepsilon = 1/N$ and $p(x,y,t) = p(x,y)(t)$. By subtracting and adding terms to Eq. (1.4) and letting $\varepsilon \longrightarrow 0$, we establish, by setting $\mathbf{z} = (x,y)$ (see Fuchs [46] for a rigorous proof), that

$$\frac{\partial}{\partial t}p(\mathbf{z},t) = -\frac{\partial}{\partial \mathbf{z}}\left[U(\mathbf{z},\theta)p(\mathbf{z},t)\right] + \frac{1}{2}\frac{\partial^2}{\partial \mathbf{z}^2}\left[\Sigma(\mathbf{z},\theta)p(\mathbf{z},t)\right], \tag{1.5}$$

with $\theta = (\beta;\mu)$, $U(\mathbf{z},\theta) = \begin{pmatrix} -\beta xy \\ \beta xy - \mu y \end{pmatrix}$ and $\Sigma(\mathbf{z},\theta) = \frac{1}{N}\begin{pmatrix} \beta xy & -\beta xy \\ -\beta xy & \beta xy + \mu y \end{pmatrix}$. Eq. (1.5) is the Fokker–Planck equation associated to the diffusion process $(x(t), y(t))$ which is solution, according to Øksendal [78] (see also Kloeden and Platen [75] and Fuchs [46]), to the nonlinear bi-variate Itô stochastic differential equation:

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} -\beta xy \\ \beta xy - \mu y \end{pmatrix} dt + \sigma(x,y)\begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} \tag{1.6}$$

where $\sigma(x,y) = \frac{1}{\sqrt{N}}\begin{pmatrix} \sqrt{\beta xy} & 0 \\ -\sqrt{\beta xy} & \sqrt{\beta y} \end{pmatrix}$ and W_1, W_2 are two independent Brownian motions.

Another approach to derive the stochastic diffusion process Eq. (1.6) is, using the infinitesimal mean and covariance (see Brauer et al. [20]), since the stochastic SIR epidemic model is a time homogeneous, diffusion process[1].

[1] The reader is referred to the book of Brauer et al. [20] for the explicit details of calculus.

Let $\Delta S(t+\Delta t)-S(t)$ and $\Delta I(t+\Delta t)-I(t)$ which supposed normally distributed approximately. Let $\Delta X(t)\,(\Delta S, \Delta I)^T$. Then, the expectation of $\Delta X(t)$ to order Δt is

$$\mathbb{E}\left(\Delta X(t)\right) = \begin{pmatrix} -\frac{\beta}{N}SI \\ \frac{\beta}{N}SI - \mu I \end{pmatrix} \Delta t$$

The covariance matrix of $\Delta X(t)$ is

$$\begin{aligned} V(\Delta X(t)) =& \mathbb{E}\left(\Delta X(t)\left[\Delta X(t)\right]^T\right) - \mathbb{E}\left(\Delta X(t)\right)\mathbb{E}\left(\Delta X(t)\right)^T \\ \approx& \mathbb{E}\left(\Delta X(t)\left[\Delta X(t)\right]^T\right) \end{aligned}$$

because the elements in the second term are $o([\Delta t]^2)$. Then the covariance matrix of $\Delta X(t)$ to order Δt is

$$V\left(\Delta X(t)\right) = \begin{pmatrix} \frac{\beta}{N}SI & -\frac{\beta}{N}SI \\ -\frac{\beta}{N}SI & \frac{\beta}{N}SI + \mu I \end{pmatrix} \Delta t.$$

The random vector $X(t+\Delta t)$ can be approximated as follows:

$$X(t+\Delta t) = X(t) + \Delta X(t) \approx X(t) + \mathbb{E}(\Delta X(t)) + V(\Delta X(t)). \tag{1.7}$$

The covariance matrix is symmetric and positive definite, then, it has a unique square root $B\sqrt{\Delta t} = \sqrt{V}$. The system of equations Eq. (1.7) are an Euler approximation to a system of Itô SDEs. For sufficiently smooth coefficients, the solution $X(t)$ of Eq. (1.7) converges to the solution of the following system of Itô SDEs:

$$dX(t) = \begin{pmatrix} dS \\ dI \end{pmatrix} = \begin{pmatrix} -\frac{\beta}{N}SI \\ \frac{\beta}{N}SI - \mu I \end{pmatrix} dt + \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix}.$$

where W_1 and W_2 are two independent Brownian motions and $B = (B_{ij})$.

1.3.3 Parameter inference for epidemic models

One often wishes of the epidemiology modeling is to statistically infer model parameters. In case the whole path is observed, parametric inference for diffusion type processes is well developed using maximum likelihood estimations. A powerful technique to overcome this problem is to estimate the model parameters in a Bayesian framework. A well-known approach is based on the idea to introduce auxiliary data points as additional observations. Most Bayesian inference methods rely on Markov chain Monte Carlo (MCMC) techniques which alternately update the auxiliary data and the model parameter. The Bayesian approach of epidemic model has been

treated by several authors, in particular see, Britton [21], Eraker [41], Britton and O'Neill [22], Demiris and O'Neill [32], El Maroufy et al. [36], and Qaffou et al. [100] and references therein.

Most of the proposed models, in the literature about the mathematical modeling of HIV/AIDS epidemics, may be seen as extensions of existing classical epidemic models, by including necessary modifications so as to consider different features of the transmission mechanism of HIV. Among those, we may mention the substantial variability of the infection rates for different subpopulations at risk (homosexuals, heterosexuals, intravenous drug users, etc.), the long incubation period before the exhibition of symptoms, the variability of the infection with respect to the evolution of the infection in each individual, etc. Mathematical models for the spread of HIV/AIDS has been discussed by many authors, see in particular Isham [69], Perelson and Nelson [94], and Perelson [95] and see the book of Tan and Wu [119] for a general review of some deterministic and stochastic models. Recently, many authors steal study HIV/AIDS models, see, for example, Punyacharoensin et al. [99] and Rivadeneira et al. [102], together with the substantial bibliography they contain. However, in the present dissertation, we are focused in stochastic modeling and inference of the model parameters of the HIV/AIDS epidemic in an heterosexual population.

1.4 An overview of the thesis

In this thesis, we consider a dynamic of HIV in a closed heterosexual population. We model it by a stochastic multidimensional SIR epidemic model. This dissertation consider two approaches, contrast function and Bayesian method, used to study dynamic of HIV epidemic model. The focus is on assimilating observed data by estimating relevant model parameters. The rest of the dissertation is organized as follows.

Chapter 2, introduce some necessary concepts and mathematical objects for a good understanding of this manuscript. Then, we give a short introduction to stochastic differential equations and Itô diffusion processes. Furthermore, we give a brief introduction to Bayesian theory and MCMC methods.

In the Chapter 3, we model the epidemic of HIV by a multidimensional SIR epidemic model, which is a four-dimensional continuous-time Markovian jump process. In which each individual can find himself at a given time in one of five mutually exclusive health states named S_F (S_M) susceptible female (male), I_F (I_M) infected female (male) and $R(t)$ the AIDS cases.

We approximate the Markov jump process by a multidimensional Itô diffusion process, exploiting the transition from a state to another by the Master equation and using *Fokker-Planck equation* to construct the diffusion process.

The Chapter 4, represents our first approach to estimate the model parameters of interest, constructed in Chapter 3. We present the minimum contrast estimator (MCE) corresponding to our model. The consistency and normal asymptotic of the MCE estimator are well discussed.

In Chapter 5, we present the second major approach to estimate parameters of the stochastic model of HIV dynamic in a closed heterosexual population, that is Bayesian approach. In this chapter we begin by a short introduction to Bayesian inference for non linear diffusion model using the augmented data method. We update the path of data in the one hand, and on the other hand, we find the posterior distribution of parameters . We close the chapter by presenting an algorithm which summarize the given procedure.

In Chapter 6, we give the numerical simulations for both approaches discussed in Chapters 4 and 5. We simulate data of a discrete Markovian stochastic process, using the Gillespie algorithm, which generates a statistically correct trajectory of a stochastic equation. In a first step, we validate the model by using the simulated data. In a second step we apply these methods presented in Chapters 4 and 5, to real application of data from Morocco.

A summary of the major contributions of this dissertation is presented in the concluding chapter. Some proofs of some key results used as tools in the Chapters 4 and 5, are given in Appendices A to C. Appendix D contain the HIV/AIDS data-set of sexually active population in Morocco, which will be applied in the real application.

Chapter 2

Preliminaries

In this chapter, we introduce briefly some necessary concepts and mathematical objects for a good understanding of this dissertation. Then, we give a short introduction to stochastic differential equations and Itô diffusion processes. Furthermore, we give a brief introduction to Bayesian theory and MCMC methods. The note of this chapter is essentially based on the following documents Kloeden and Platen [75], Karatzas and Shreve [71], Øksendal [78], Kutoyants [82], Klebaner [74], and Fuchs [46].

2.1 Aspects of General Probability Theory

In this section, we recall some concepts from the general probability theory and stochastic processes, which lead us to define an Itô stochastic integral. For more details, the reader is referred to the book of Øksendal [78, Chap. 2].

2.1.1 Stochastic processes

Definition 2.1 (Stochastic process)**.** *A stochastic process can be defined as a parameterized collection (family)* $(X_t)_{t\in\mathbb{T}}$ *of random variables on the same probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. Therefore, the stochastic process* $(X_t)_{t\in\mathbb{T}}$*, can be written as a function:*

$$\begin{aligned} X: \ \mathbb{T}\times\Omega &\longrightarrow \mathbb{R}^d \\ (t,\omega) &\longmapsto X(t,\omega) \end{aligned}$$

with $d \geq 1$*.*

But generally, we omit the dependency on ω in the notation $(X_t)_{t\in\mathbb{T}}$. The parameter space $\mathbb{T}$ is usually the half-line $[0;\infty)$, but it may also be an interval $[a;b]$, the non-negative integers and

even subsets of $\mathbb{R}^n$ for $n \geq 1$. Note that for each $t \in \mathbb{T}$ fixed we have a random variable

$$\omega \longrightarrow X_t(\omega); \quad \omega \in \Omega.$$

On the other hand, fixing $\omega \in \Omega$ we can consider the function

$$t \longrightarrow X_t(\omega); \quad t \in \mathbb{T}.$$

which is called a path of X_t.

Definition 2.2 (Filtration)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family $(\mathcal{F}_t)_{t\geq 0}$ of sub σ-algebras of $\mathcal{F}$. In other words, for each t, $\mathcal{F}_t$ is a σ-algebras included in $\mathcal{F}$ and if $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. A probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t)_{t\geq 0}$ is called a filtered probability space.*

Definition 2.3. *A stochastic process $(X_t)_{t\in\mathbb{T}}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if, for every $t \in \mathbb{T}$, the random variable X_t is $\mathcal{F}_t$-measurable.*

A stochastic process X is always adapted to its natural filtration $(B_t)_{t\geq 0}$ (the last notation meaning that $\mathcal{F}_t$ is the smallest σ-algebra with respect to which all the variables $(X_s, s \leq t)$ are measurable). $\mathcal{F}_t^X$ is hence the smallest filtration to which X is adapted.

An important class of stochastic processes is the *Martingale*.

Definition 2.4 (Martingale)**.** *A stochastic process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a martingale with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ (and with respect to $\mathbb{P}$) if:*

i– X_t *is $\mathcal{F}_t$-measurable for all $t \geq 0$,*

ii– $\mathbb{E}[X_t] < \infty$ *for all $t \geq 0$,*

iii– $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ *for all $s \leq t$.*

2.1.2 Brownian Motion

Brownian motion is the random movement of particles in a fluid resulting from the impact of molecules of the surrounding medium. It was named for the Scottish botanist Robert Brown[1], the first to study such fluctuations. The mathematical description of Brownian motion is a relatively simple probability calculation, of importance in many fields, physics, chemistry, biology and others. A modern model is the Wiener process, named in honor of Norbert Wiener[2], who described the function of a continuous-time stochastic process.

[1] Robert Brown (21 December 1773 – 10 June 1858) was a Scottish botanist and palaeobotanist who made important contributions to botany largely through his pioneering use of the microscope. His contributions include one of the earliest detailed descriptions of the cell nucleus and cytoplasmic streaming; the observation of Brownian motion;

[2] Norbert Wiener (November 26, 1894 – March 18, 1964), was an American mathematician and philosopher.

Definition 2.5 (Brownian Motion)**.** *A real-valued $\mathcal{F}$-adapted process $(B_t)_{t\geq 0}$ is a Brownian motion, also called a Wiener process, if it satisfies the following conditions:*

i− $B_0 = u$ *almost surly for a fixed* $u \in \mathbb{R}$,

ii− *All paths* $(B_t)_{t\geq 0}$ *are almost surly continuous,*

iii− *All paths have independent and stationary increments,*

iv− $B_t \sim \mathcal{N}\left(0, \sigma^2 t\right)$, *for all* $t \geq 0$ *and constant volatility parameter* $\sigma \in \mathbb{R}_+$.

The assertion *iii*− in the last definition means that the increments of the Brownian motion are statistically independent on non-overlapping intervals, *i.e.* $B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots$, $0 \leq t_0 < t_1 < t_2 < \ldots$, are pairwise independent. And stationary increments means that, the probability distribution function for $B_{s+t} - B_s$ is fixed (the same) for all $s \in \mathbb{T}$ such that $s + t \in \mathbb{T}$.

The process $(B_t)_{t\geq 0}$ is called standard Brownian motion when $u = 0$ and $\sigma = 1$. A vector-valued process is said to be $d - dimensional$ (standard) Brownian motion if its d components are mutually independent $one - dimensional$ (standard) Brownian motions. The existence of such process was first proven by Wiener [128]. The probability law induced by standard Brownian motion is thus called Wiener measure.

Note that a Brownian motion $(B_t)_{t\geq 0}$ is a martingale with respect to its natural filtration $\mathcal{F}_t$ generated by $\{B_s,\ s \leq t\}$; $\mathcal{F}_t = \sigma\left(B_s,\ \ 0 \leq s \leq t\right)$. (The reader is referred to Øksendal [78, Chap 3], for the proof).

2.2 Itô Integral and Stochastic Differential Equations

The beginnings of the theory of stochastic integration were motivated and intertwined with the theory of Markov processes, in which Kolmogorov[1] played a fundamental role. In the fifties of the 20th-century, Kiyosi Itô[2], developed the theory of stochastic integration and stochastic differential equations by developing the Itô calculus.

Throughout the rest of this dissertation, $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ is considered as a filtered probability space with sample space Ω, σ-algebra $\mathcal{F}$, $(\mathcal{F}_t)_{t\geq 0}$ the natural filtration and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. The Borel σ-algebra of Lebesgue subsets of $\mathbb{R}$ will be denoted by $\mathcal{B}$.

[1] Andrey Nikolaevich Kolmogorov (25 April 1903 – 20 October 1987) was a Soviet mathematician who made significant contributions to the mathematics of probability theory, topology, intuitionistic logic, turbulence, classical mechanics, algorithmic information theory and computational complexity.

[2] Kiyosi Itô, (September 7, 1915 – 10 November 2008) was a Japanese mathematician. He pioneered the theory of stochastic integration and stochastic differential equations, now known as the Itô calculus. Its basic concept is the Itô integral, and among the most important results is a change of variable formula known as Itô's lemma or formula.

2.2.1 Itô Integral

The Itô (Stochastic) integral can be defined in similar manner to the Riemann–Stieltjes integral (see for example, Surhone et al. [118]), that is as a *limit in probability* of Riemann sums; such a limit does not necessarily exist pathwise. Now, we describe the class of functions for which the Itô integral will be defined :

Definition 2.6. *(Øksendal [78, Defenition 3.1.4.]) Let define the class $\mathcal{V} = \mathcal{V}(S,T)$ of functions*

$$f(t,\omega) : [0;\infty) \times \Omega \longrightarrow \mathbb{R}$$

such that

i− $(t,\omega) \longrightarrow f(t,\omega)$ *is jointly $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ denotes the Borel σ- algebra on $[0;\infty)$.*

ii− $f(t,\omega)$ *is $\mathcal{F}_t$-adapted,*

iii− $\int_S^T \mathbb{E}\left[(f(u,\omega))^2\right] du < \infty.$

Here, $\mathbb{E}(.)$ is the expectation operator with respect to the probability measure $\mathbb{P}$. Now, we give the definition (Øksendal [78, Defenition 3.1.6]) of Itô Integral,

Definition 2.7 (The Itô Integral)**.** *Let $f \in \mathcal{V}(\mathcal{S},\mathcal{T})$. Then the Itô integral of f, from S to T, is defined by*

$$\int_{t_0}^{t} f(s,\omega)dB_s(\omega) = \lim_{n\to\infty} \int_{t_0}^{t} \phi_n(s,\omega)dB_s(\omega) \qquad (\textit{limit in} \quad \mathcal{L}^2_{\mathbb{P}}(\Omega)), \tag{2.1}$$

where $\{\phi_n\}$ is a sequence of elementary functions[1] *such that*

$$\mathbb{E}\left[\int_S^T \left(f\left(s,\omega\right) - \phi_n\left(s,\omega\right)\right)^2 ds\right] \longrightarrow 0, \qquad as \quad n \to \infty. \tag{2.2}$$

Note that such a sequence $\{\phi_n\}$ satisfying (2.2) exists (see, Øksendal [78] and Klebaner [74] for more details). Moreover, the limit in (2.1) exists and does not depends on the actual choice of $\{\phi_n\}$. Furthermore, the Itô integral satisfies the following properties:

Corollary 2.1. *Let $f, g \in \mathcal{V}(\mathbb{T})$*

i− $\mathbb{E}\left[\left(\int_{t_0}^{t} f(s,\omega)dB_s\right)^2\right] = \mathbb{E}\left[\int_s^t f^2(s,\omega)ds\right]$ *(The Itô isometry).*

ii− $\mathbb{E}\left[\int_{t_0}^{t} f(s,\omega)dB_s\right] = 0,$

[1] A function $\phi_n \in \mathcal{V}$ is called elementary function, if it has the form $\phi_n = \sum_j \xi_j(\omega)\chi_{[t_j,t_{j+1})}(t)$, where χ is the indicator function of the interval $[t_j, t_{j+1})$. Since $\phi_n \in \mathcal{V}$ each function ξ_j must be $\mathcal{F}_{t_j}$-measurable.

iii− $\int_{t_0}^{t} f(s,\omega)dB_s$ *is adapted process to the filtration* $(\mathcal{F}_t)_{t\geq 0}$.

An important property of the Itô integral is that it is a martingale, then, it verifies the property of the Definition 2.4 (see, Øksendal [78] and Klebaner [74], for more details).

2.2.2 Stochastic differential equations

Generally, a *stochastic differential equation (SDE)* is a differential equation with terms containing both deterministic and stochastic differentials and whose solution is a stochastic process. SDEs are a powerful and natural tool for the modeling of complex systems that change roughly in continuous time. Application areas include econometrics and finance (Aït-Sahalia [6], Aït-Sahalia and Jacod [8], Eraker and Wang [42], Aït-Sahalia and Hurd [7], Aït-Sahalia et al. [9], and Eraker and Wu [43]), physics (Ramshaw [101], Tuckwell and Williams [124], and Seifert [111]), biology (Leung [83], Elf and Ehrenberg [40], and Sjöberg et al. [113]), systems biology (Golightly and Wilkinson [59, 58] and Golightly Andrew and Wilkinson Darren J. [60]), medicine (Walsh [127] and Capasso and Morale [23]), epidemiology (Barbour [15] and Alonso et al. [3]), population biology (Ferm et al. [44]), genetics (Tian et al. [122]) and many other authors are used the SDEs in other fields.

Definition 2.8 (Itô process)**.** *Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *be a filtered probability space, then an Itô process* $(X_t)_{t\geq 0}$ *is a process which satisfies:*

$$X_t(\omega) = X_s(\omega) + \int_s^t \boldsymbol{\mu}\left(u, X_t\right) du + \int_s^t \boldsymbol{\sigma}\left(u, X_t\right) dB_u, \tag{2.3}$$

for any $[s,t] \subseteq \mathbb{T}$ *, where the functions* $\boldsymbol{\mu}$ *and* $\boldsymbol{\sigma}$ *are jointly* $\mathcal{B} \times \mathcal{F}_{\sqcup}$*-measurable,* $\mathcal{F}$*-adapted and satisfies the following criteria:*

$$\int_s^t \left|\boldsymbol{\mu}\left(u, X_t\right)\right| du < \infty \qquad \text{and} \qquad \int_s^t \left(\boldsymbol{\sigma}\left(u, X_t\right)\right)^2 du < \infty.$$

An Itô process is a stochastic process that can be, formally, written as

$$dX_t = \boldsymbol{\mu}\left(X_t, t\right) dt + \boldsymbol{\sigma}\left(X_t, t\right) dB_t. \tag{2.4}$$

Equipped with the definition of the Itô integral, a stochastic process X is a solution of the stochastic differential equation (2.4), if and only if, X satisfies the stochastic integral equation (SDE) (2.3) almost surely. Then X is an Itô process and one can prove that it is Markovian.

The existence and uniqueness of solutions to the stochastic differential equations (2.4) is guaranteed by the following theorem (Øksendal [78, Theorem 5.2.1.]),

Theorem 2.1 (Existence and uniqueness theorem for stochastic differential equations)**.** *Let* $T > 0$ *and* $\boldsymbol{\mu}(\cdot) : [0,T] \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $\boldsymbol{\sigma}(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \longrightarrow \mathbb{R}^{n\times m}$ *be measurable functions*

satisfying

$$\|\boldsymbol{\mu}(t,x)\|^2 + \|\boldsymbol{\sigma}(t,x)\|^2 \leq C\left(1+\|x\|^2\right), \text{ (Linear growth)} \tag{2.5}$$

$$\|\boldsymbol{\mu}(t,x) - \boldsymbol{\mu}(t,y)\| + \|\boldsymbol{\sigma}(t,x) - \boldsymbol{\sigma}(t,y)\| \leq D\|x-y\|, \quad \text{(Lipschitz continuity)} \tag{2.6}$$

for $x, y \in \mathbb{R}^n, t \in [0,T]$ and some constants C and D.

Let Z be a random variable which is independent of the σ-algebra $\mathcal{F}_\infty = \sigma(B_s, s \geq 0)$ and such that $\mathbb{E}|Z|^2 < \infty$.

Then the stochastic differential equation

$$dX_t = \boldsymbol{\mu}(X_t, t)\, dt + \boldsymbol{\sigma}(X_t, t)\, dB_t, \quad t \in [0,T], X_0 = Z \tag{2.7}$$

has a unique t-continuous solution $X_t(\omega)$ with the property that $X_t(\omega)$ is adapted to the filtration $\mathcal{F}_t^Z = \sigma(Z, B_s, 0 \leq s \leq t)$ and

$$\mathbb{E}\left[\int_0^T |X_t|^2\, dt\right] < \infty \tag{2.8}$$

The solution $(X_t)_{t\geq 0}$ found above is called a strong solution, because the version $(B_t)_{t\geq 0}$ of Brownian motion is given in advance and the solution $(X_t)_{t\geq 0}$ constructed from it is $\mathcal{F}_t^Z$-adapted.

However, it is generally not possible to find an explicit solution of an SDE. An explicit solution to a family of stochastic differential equations is well discussed in Kouritsin and Deli [76].

2.2.3 Itô formula for diffusion processes

The Itô's formula is for stochastic calculus what the Newton-Leibnitz formula[1] is for (the classical) calculus. Not only does it relate differentiation and integration, it also provides a practical method for computation of stochastic integrals. The Itô formula is very useful for evaluating Itô integrals and it serves as the stochastic calculus counterpart of the chain rule. The 1-dimensional Itô formula is given in following theorem (see, Klebaner [74, Theorem 4.16], Øksendal [78, Theorem 4.1.2]).

Theorem 2.2. *Let $(X_t,\ t \geq 0)$ be an Itô process given by*

$$dX_t = \boldsymbol{\mu}(X_t, t)\, dt + \boldsymbol{\sigma}(X_t, t)\, dB_t,$$

[1] The formula expressing the value of a definite integral of a given integrable function f over an interval as the difference of the values at the endpoints of the interval of any primitive F of the function f:

$$\int_a^b f(x)dx = F(b) - F(a)$$

It is named after I. Newton and G. Leibniz, who both knew the rule expressed by the above equation. It is also known as "*Fundamental theorem of calculus*".

Let $g : \mathbb{R} \times [0, \infty) \longrightarrow \mathbb{R}$ be a twice continuously differentiable on $\mathbb{R} \times [0, \infty)$. Then the process $Y_t = g(X_t, t)$ is again an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(X_t, t)dt + \frac{\partial g}{\partial x}(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(X_t, t)\left(dX_t\right)^2, \tag{2.9}$$

where $\left(dX_t\right)^2 = dX_t dX_t$ is computed according to the rules

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0 \quad \textit{and} \quad dB_t \cdot dB_t = dt.$$

Using the above rules Eq. (2.9) becomes

$$\begin{aligned} dY_t =& \frac{\partial g}{\partial t}(X_t, t)dt + \frac{\partial g}{\partial x}(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(X_t, t)\boldsymbol{\sigma}^2\left(X_t, t\right)dt \\ =& \left(\frac{\partial g}{\partial t}(X_t, t) + \frac{\partial g}{\partial x}(X_t)\boldsymbol{\mu}\left(X_t, t\right) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(X_t, t)\boldsymbol{\sigma}^2\left(X_t, t\right)\right)dt + \frac{\partial g}{\partial x}(X_t)\boldsymbol{\sigma}\left(X_t, t\right)dB_t. \end{aligned}$$

For the general situation in higher dimensions, let $(B_t)_{t\geq 0})$ denote d-dimensional Brownian motion and $(X_t)_{t\geq 0}$ an n-dimensional Itô process, Then the general Itô formula is given by the following theorem (Øksendal [78, Theorem 4.2.1])

Theorem 2.3 (The general Itô formula)**.** *Let*

$$dX_t = \boldsymbol{\mu}\left(X_t, t\right)dt + \boldsymbol{\sigma}\left(X_t, t\right)dB_t,$$

be an n-dimensional Itô process, as defined in Definition 2.8. Let $g(x, t) = (g_1(x, t), \ldots, g_d(x, t))$ be a $\mathcal{C}^2$ map from $\mathbb{R}^n \times [0, \infty)$ into $\mathbb{R}^d$. Then the process

$$Y(\omega, t) = g(X_t, t)$$

is again an Itô process, whose component number k, Y_k, is given by

$$dY_k = \frac{\partial g_{(k)}}{\partial t}(X, t)dt + \sum_{i=1}^{d}\frac{\partial g_k}{\partial x_i}(X, t)dX_i + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 g_k(X, t)}{\partial x_i \partial x_j}dX_i dX_j, \tag{2.10}$$

where $dB_t^i dB_t^j = \delta_{ij}dt$, $dt \cdot dB_t^i = dB_t^i \cdot dt = 0$ and δ_{ij} is the Kronecker delta.

For more details and proof, the reader is referred to e.g. Øksendal [78, Chap4].

2.3 Approximation of solutions to stochastic differential equations

Generally, it may not be possible to determine an explicit or closed-form solution to the Itô's stochastic differential equation

$$dX_t = \boldsymbol{\mu}\left(X_t, t\right) dt + \boldsymbol{\sigma}\left(X_t, t\right) dB_t, \quad X_0 = x_0.$$

Hence, the employment of numerical methods for calculating approximations to problems such as (2.4) is preferred. In this regard, we must use a discrete-time approximation to iteratively approximate a solution to (2.4). This method utilizes a recursive algorithm that produces the values of a discrete-time approximation at the given discretization points of a finite subinterval $[t_0, t] \subset \mathbb{T}$. While the approximation is made only at the discretization points, we will always view a discrete-time approximation as a "continuous-time process" defined on $[t_0, t]$ In this dissertation, we present two schemes to approximate (2.4), that is Euler–Maruyama and Milstein approximations. The reader is referred to Panik [92, Chap.7] for more details of approximating methods.

2.3.1 The Euler–Maruyama approximation

The simplest discrete-time recursive routine used to approximate an Itô process of the form (2.4) is the Euler–Maruyama approximation scheme. Given the time discretization $t_0 < t_1 < t_2 < \ldots < t_N = t$ of $[t_0, t]$, the Euler–Maruyama approximation is a continuous-time stochastic process $X = (X_t)_{t\in[t_0,t]}$ satisfying the iterative scheme

$$\begin{aligned} X_{i+1} &= X_i + \boldsymbol{\mu}\left(X_i, t_i\right)\left(t_{i+1} - t_i\right) + \boldsymbol{\sigma}\left(X_i, t_i\right)\left(B_{t_{i+1}} - B_{t_i}\right) \\ &= Xi + \boldsymbol{\mu}\left(X_i, t_i\right)\Delta_{t_i} + \boldsymbol{\sigma}\left(X_i, t_i\right)\Delta B_{t_i}, \qquad i = 1, 2, \ldots, N-1, \end{aligned} \tag{2.11}$$

where $X_{ti} \equiv X_i$, $X_0 = x_0$, $\Delta_{t_i} = t_{i+1} - t_i$, and $\Delta B_{t_i} = B_{t_{i+1}} - B_{t_i}$. To obtain a "good" approximate solution, the time increments $\Delta_i, i = 0, 1, 2, \ldots,$ should be "sufficiently small". Particularly, under a regime of equidistant discretization times, we have $\Delta_{t_i} \equiv \Delta = (t - t_0)/N$.

2.3.2 The Milstein approximation

The Milstein scheme is an amelioration of the Euler–Maruyama method by introducing a correction to the stochastic increment in (2.11), by introducing the term

$$\frac{1}{2}\boldsymbol{\sigma}(X_i, t_i)\frac{\partial \boldsymbol{\sigma}}{\partial x}\left[\Delta {B_i}^2 - \Delta_i\right]$$

from the Itô–Taylor expansion. Therefore, the Milstein scheme appears as

$$X_{i+1} = X_i + \boldsymbol{\mu}(X_i, t_i)\Delta_i + \boldsymbol{\sigma}(X_i, t_i)\Delta B_i + \frac{1}{2}\boldsymbol{\sigma}(X_i, t_i)\frac{\partial \boldsymbol{\sigma}}{\partial x}\left[\Delta {B_i}^2 - \Delta_i\right]. \tag{2.12}$$

2.3.3 Strong and weak convergence of approximation schemes[1]

Let $Y = (Y_t)_{t\in[t_0,t]}$ be the Euler–Maruyama approximation of Itô process $X = (X_t)_{t\in[t_0,t]}$. Since $X(t)$ and $Y(t)$ are both random variables, it is reasonable to use an expression such as $\mathbb{E}[X_t - Y_t]$ to measure the degree of precision (error) of the approximation. More specifically, a time-discretized approximation (Y_t) of a continuous-time process (X_t) converges with *strong order* γ to the solution (X_t) at time t if there exists a constant C (not depending on Δ) such that

$$\mathbb{E}|X_t - Y_t| \leq C\Delta^\gamma \tag{2.13}$$

for N chosen large enough so that $\Delta = (t - t_0)/N\,(0, 1)$, where X_t is the true solution at time t and Y_t the approximation. The strong order of convergence criterion (2.13) indicates the rate at which the "mean endpoint error" decreases as $\Delta \to 0$.

The *weak convergence* is a less restrictive criterion of the (2.13), which considers the rate of decrease of the "error of means". Specifically, a discrete-time approximation (Y_t) of a continuous time process (X_t) converges weakly of order β to the solution (X_t) at time t if there exists a continuously differentiable polynomial function h and a constant C_h (independent of Δ) such that

$$|\mathbb{E}(h(X_t)) - \mathbb{E}(h(Y_t))| \leq C_h\Delta^\beta \tag{2.14}$$

Given these error bounds, one can prove that the strong order of convergence for the Euler–Maruyama scheme is $\gamma = \frac{1}{2}$ (if $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ satisfy uniform growth and Lipschitz conditions), however, for Milstein scheme the strong order of convergence is $\gamma = 1$. The Euler–Maruyama routine converges with weak order $\beta = 1$.

2.4 Transition density and likelihood function

2.4.1 Markov process and transition density

Let $(X_t)_{t\geq 0}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The σ-algebra $\mathcal{F}_t = \sigma(X_s, 0 \leq s < t)$, $t \geq 0$, is the history of the process up to and including time t. The real-valued, $\mathcal{F}_t$-adapted stochastic process $(X_t)_{t\geq 0}$ is called a *Markov process*[2] if the following

[1] The reader is referred to Iacus [67] and Panik [92], as well as to the famous book of Kloeden and Platen [75] for more information and rigorous studies of such schemes .

[2] Markov process is named for the Russian mathematician Andrey Andreyevich Markov (1856 – 1922).

property holds almost surly; *Markov property*

$$\mathbb{P}\left(X_t \in B|\mathcal{F}_s\right) = \mathbb{P}\left(X_t \in B|X_s\right) \quad \text{for all} \quad 0 \leq s \leq t < \infty \text{ and Borel set } B.$$

So given X_s, one can predict the probabilities of future values X_t just as well as if we knew the entire history of the process prior to time s. The process only knows X_s and is not aware of how it got there so that the future depends on the past only through the present, that is, once the present is known, the past and future are independent.

The Itô process $(X_t)_{t\geq 0}$ solution of the SDE (2.4) is a Markov process and its *transition density* $p(s, x, t, y)$ is defined by

$$p(s, x; t, A) = \mathbb{P}(X_t \in A|X_s = x) = \int_A p(s, x; t, y)dy, \tag{2.15}$$

for all $\mathcal{F}$-measurable sets $A \subseteq \mathbb{R}^n$. $p(s, x; t, y)$ is the density of a Markov process $(X_t)_{t\geq 0}$ for going from state $x \in \mathbb{R}^n$ at time $s \geq 0$ to $y \in \mathbb{R}^n$ at time $t > s$. For $s = t$, we define

$$p(s, x; t, y) = \delta\left(x - y\right)$$

where δ denotes the Dirac delta function. If $(X_t)_{t\geq 0}$ is homogeneous in time, *i.e.* the transition density depends on s and t solely through their difference $t - s$, we also write $p(t - s; x, y)$.

2.4.2 Itô diffusion process

A *diffusion process* is defined as a Markov process whose transition probability function p meets the following three properties for all $x \in \mathbb{R}$ $(\mathbb{R}^n)$, $s \geq 0$ and $\varepsilon > 0$ one has uniformly

$$\lim_{t\to s} \frac{1}{t - s} \int_{\|y-x\|>\varepsilon} p(s, x, t, y)dy = 0. \tag{2.16}$$

$$\boldsymbol{\mu}(s, x) = \lim_{t\to s} \frac{1}{t - s} \int_{\|y-x\|\leq\varepsilon} p(s, x, t, y)(y - x)dy, \tag{2.17}$$

$$\boldsymbol{\Sigma}(s, x) = \lim_{t\to s} \frac{1}{t - s} \int_{\|y-x\|\leq\varepsilon} p(s, x, t, y)(y - x)(y - x)^T dy, \tag{2.18}$$

where the limits (2.17) and (2.18) exist.

The vector-valued function $\boldsymbol{\mu}$ in (2.17) is called the *drift* and the symmetric and positive semi-definite matrix-valued function $\boldsymbol{\Sigma}$ is called the *diffusion matrix*. A matrix $\boldsymbol{\sigma}$ with $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is called the *diffusion coefficient*, where $\boldsymbol{\sigma}^T$ is the transpose of $\boldsymbol{\sigma}$. Such a decomposition exists due to the positive semi-definiteness of $\boldsymbol{\Sigma}$, but is not necessarily unique. However, the particular choice of the diffusion coefficient does not influence the distribution of the process X as long as it is a square root of the diffusion matrix.

An $\mathcal{F}$-adapted process X satisfying the Itô SDE

$$dX_t = \boldsymbol{\mu}(X_t, t)dt + \boldsymbol{\sigma}(X_t, t)dB_t, \quad x_{t_0} = x_0,$$

is an Itô diffusion with drift $\boldsymbol{\mu}$ and diffusion matrix $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$ if the coefficients $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ fulfill the Lipschitz condition (2.6) and linear growth (2.5) and are continuous in time.

2.4.3 Fokker–Plank equation associated to an Itô diffusion

Let $\mathbf{X}$ be an Itô diffusion process

$$dX_t = \boldsymbol{\mu}(X_t, t)dt + \boldsymbol{\sigma}(X_t, t)dB_t, \quad x_{t_0} = x_0.$$

Under some regularities on the coefficients $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are fulfilled to ensure the existence and uniqueness as defined in Theorem 2.1. If in addition $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ have two partial derivatives with respect to x, which are bounded and satisfy a Hölder condition with respect to x, then the transition density $p = p(t, x, y)$ as a function in y and t, satisfies the following partial differential equations (PDF),

$$-\frac{\partial p}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\boldsymbol{\sigma}(y,t)^2 p\right) - \frac{\partial}{\partial x}\left(\boldsymbol{\mu}(y,t)p\right) = 0. \tag{2.19}$$

The Eq. (2.19) is a PDE in the forward variables (y, t) and is therefore called the *forward equation*, also known as *Fokker-Plank equation*, diffusion equation, or Kolmogorov's forward equation. Explicitly, it can be written as,

$$\frac{\partial p(s,x,t,y)}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\boldsymbol{\mu}_i(t,y)p(s,x,t,y)) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i x_j}(\boldsymbol{\Sigma}_{i,j}(t,y)p(s,x,t,y)),$$

for fixed y and t, where $x, y \in \mathcal{X}$ and $t > s \geq 0$, and i, j denote the respective components of x, y, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$, where $\boldsymbol{\sigma}^T$ is the transposed vector of $\boldsymbol{\sigma}$. We remark that equation uniquely determines the transition density p, and hence diffusion processes are already completely defined by their instantaneous mean and variance $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Furthermore, if the transition density of a stochastic process fulfills the Fokker–Plank equation (2.4.3), then it is an Itô diffusion process.

2.4.4 Likelihood function for discretely observed processes

The method of maximum likelihood is a technique of data reduction, it requires the knowledge of the form of probability density function. The statistical estimation of the possibly vector-valued parameter θ from an open set $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}^\star$ is the objective of the maximum likelihood (ML) method. The transition density Eq. (2.15) further depends on a parameter θ from a parameter space Θ. For discrete observations $x_{t_1}, \ldots, x_{t_n}$ and given starting value x_{t_0} at time

points $t_0 < \ldots < t_n$, as diffusion processes are Markovian, the *likelihood function* of θ is given by

$$\mathcal{L}_n\left(\theta; x_{t_1}, \ldots, x_{t_n}\right) = \prod_{k=0}^{n-1} p_\theta\left(t_k, x_{t_k}, t_{k+1}, x_{t_{k+1}}\right) = \prod_{k=0}^{n-1} p_\theta\left(t_{k+1} - t_k; x_{t_k}, x_{t_{k+1}}\right). \tag{2.20}$$

The *log-likelihood* function of θ is then given by

$$l_n = \log \mathcal{L}_n\left(\theta; x_{t_1}, \ldots, x_{t_n}\right) = \sum_{k=0}^{n-1} \log p_\theta\left(t_{k+1} - t_k; x_{t_k}, x_{t_{k+1}}\right). \tag{2.21}$$

The logarithmic likelihood function (2.21) expresses the probability of the observed random sample as a function of θ. So with θ treated as a variable in (2.21). The ML method is based upon the *principle of maximum likelihood*: select as an estimate of θ that value of the parameter, $\hat{\theta}_{ML}$, that maximizes the probability of observing the given random sample. So to find $\hat{\theta}_{ML}$, we need only maximize (2.21) with respect to θ. Then, the *maximum likelihood estimators* (MLE) are defined to be estimators that maximize the likelihood function $\hat{\theta}_{ML} = \arg\max_\theta \mathcal{L}_n\left(\theta; x_{t_1}, \ldots, x_{t_n}\right)$. That is, an MLE of θ after observing a sample $x_{t_1}, \ldots, x_{t_n}$ is any θ at which $\mathcal{L}_n$ achieves its maximum, if there are any such θ.

In general, for a particular class of probability densities, as exponential parametric families, under some strict conditions, the MLE, is consistent and asymptotically normal (see, Schervish [110, Chap. 7]). But for more general parametric families, the proofs of these properties are more complicated.

The statistical inference for continuously observed diffusion processes on a finite interval is based on the likelihood of the diffusion and obtained using the Girsanov formula (see e.g. Liptser and Shiryaev [84] and Kutoyants [82]. Discretely observed diffusion processes are discrete time Markov processes and thus their likelihood depends on the transition densities of the diffusion $\mathbb{P}(X(t_k) \in A | X(t_k - 1) = x)$. Since the dependence with respect to the parameters θ of these transition densities is not explicit, then, the likelihood is intractable (see for example, El Maroufy et al. [38], the authors tried to find a closed form of the transition probability, but in a very restrictive case, of the SIR model.). Thus, other approaches have been proposed (the reader is referred to Fuchs [46, Chap. 6] for a survey of alternative methods to ML approach). Since, we are interested only by stochastic diffusion processes, then, instead of using the likelihood function, we use a contrast processes with associated minimum contrast estimators (MCE). They have to satisfy a series of conditions to lead to good estimators. have adopted in this dissertation the terminology of contrast processes and minimum contrast estimators. The details of this alternative approach is the subject of the Chapter 4.

2.5 Bayesian inference for diffusion with discrete observation[1]

2.5.1 Bayes theorem

Bayesian statistics was named after Thomas Bayes[2], who formulated a specific case of Bayes' theorem in his paper published in 1763. Many Bayesian methods were developed by later authors, but the term was not commonly used to describe such methods until the 1950s. During much of the 20th century, Bayesian methods were unfavorable with many statisticians due to philosophical and practical considerations. Many Bayesian methods required a lot of computation to complete, and most methods that were widely used during the century were based on the frequentist interpretation. However, with the advent of powerful computers and new algorithms like Markov chain Monte Carlo(MCMC), Bayesian methods have seen increasing use within statistics coming into the 21st century.

Let π generally denote all posterior densities, p all prior densities and q all proposal densities. Let θ be parameter to be estimated and x be the given observed data. Then the Bayes' theorem is given by the following formula:

Theorem 2.4 (Bayes' Theorem)**.** *Let θ and x be two events such that $P(x) > 0$, then*

$$\pi(\theta|x) = \frac{p(\theta)\pi(x|\theta)}{p(x)}. \tag{2.22}$$

In Bayesian statistics, most of the terms in Bayes' rule have special names. Some of them even have more than one name, with different scientific communities preferring different terminology. Here is a list of the various terms and the names we will use for them:

- $\pi(\theta|x)$ is the *posterior probability*. It describes how certain or confident we are that hypothesis θ is true, given that we have observed data x. Calculating posterior probabilities is the main goal of Bayesian statistics.
- $p(\theta)$ is the *prior probability*, which describes how sure we were that θ was true, before we observed the data x.
- $\pi(x|\theta)$ is the *likelihood*. If we assume that θ is true, this is the probability that we would have observed data x.
- $p(x)$ is the *marginal likelihood*. This is the probability that we would have observed data x, whether θ is true or not.

[1] the contents of this section are inspired by the two following works, Robert [103] and Robert and Casella [104].

[2] Thomas Bayes (c. 1701 – 7 April 1761) was an English statistician, philosopher and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: Bayes' theorem. Bayes never published what would become his most famous accomplishment; his notes were edited and published after his death by Richard Price.

This can be written in three ways as follows:

$$\begin{aligned} \pi(\theta|x) &= \frac{p(\theta)\pi(x|\theta)}{p(x)} \\ \pi(\theta|x) &\propto p(\theta)\pi(x|\theta) \\ \texttt{posterior} &\propto \texttt{prior} \star \texttt{likelihood}. \end{aligned}$$

given the prior distribution $p(\theta)$ and the likelihood $p(x|\theta)$. Then we can get the posterior distribution by the equation

$$\pi\left(\theta \mid x\right) = \frac{\pi\left(x \mid \theta\right) p\left(\theta\right)}{\int_{\Theta} \pi\left(x \mid \theta\right) p\left(\theta\right) d\theta}. \tag{2.23}$$

2.5.2 Markov chain Monte Carlo methods

Markov chain Monte Carlo (MCMC) methods are methods for sampling probability distribution functions or probability density functions (*pdfs*). These *pdfs* may be either probability mass functions on a discrete space, or probability densities on a continuous space. MCMC methods don't require that we have a full analytic description of the properly normalized *pdfs* for sampling to proceed; they only require that we are able to compute ratios of the *pdfs* at pairs of locations. This makes MCMC methods ideal for sampling *posterior pdfs* in probabilistic inferences:

In a probabilistic inference, the posterior density $\pi(\theta \,|\, x)$, or density of parameters θ given data x, is constructed from the likelihood $\pi(x \,|\, \theta)$ and the prior density $p(\theta)$ of parameters, it is often known as "Bayes rule" (2.23),

$$\pi(\theta \,|\, x) = \frac{1}{Z}\, \pi(x \,|\, \theta)\, p(\theta).$$

In these contexts, the constant Z, sometimes written as $p(x)$, is known by the names "evidence", "*marginal likelihood*", "Bayes integral" and "prior predictive probability". It is usually *extremely hard* to calculate the *marginal likelihood* [1]. That is, we often know the function $\pi(\theta \,|\, x)$ up to a constant factor; we can compute ratios of two *pdfs* at pairs of points, but not the precise value at any individual point.

In addition to this normalization-insensitive property of MCMC, in its simplest forms it can be run without computing any derivatives or integrals of the function, and in its simplest forms it is *extremely easy to implement*. For all these reasons, MCMC is ideal for sampling posterior *pdfs* in the real situations in which scientists find themselves. In this thesis, our study is restricted to using *Gibbs Sampler* and *Metropolis-Hastings* algorithms.

[1] The factor Z is often difficult to compute, because the likelihood (or the prior) can have extremely complex structure, with multiple arbitrarily compact modes, arbitrarily positioned in the (high dimensional) parameter space θ.

2.5.3 Gibbs sampling algorithm

Gibbs sampling allows us to sample from a joint multivariate distribution using only the conditional distributions (see Wilkinson [129] for a simple example). Gibbs sampling was first proposed by Geman and Geman [47], who used it to study image-processing models. Since then, Gibbs sampling has become a very popular MCMC method with many fields Carter and Kohn [24], Gilks et al. [53], Arminger and Muthén [5], Porteous et al. [97], and Damlen et al. [30]. A first explanation of Gibbs sampler contain theory and examples (see,Casella and George [25] and the references therein) and the reader can find a review of Gibbs sampler and other MCMC methods in the paper of Smith and Roberts [114]. An algorithm for Gibbs sampler is provided in Algorithm 1, which generates samples of θ and x iteratively, from the conditional distributions $\pi(\theta \mid x)$ and $\pi(x \mid \theta, y)$, respectively.

Algorithm 1 Gibbs Sampler

1: Given an observed-data y.
2: Initialize x by sampling $x \sim \pi(. \mid y)$
3: **repeat**
4: Sample $\theta \sim \pi(\theta \mid x)$ using current x ▷ $\pi(\theta \mid x)$ is known analytically
5: Sample $x \sim \pi(x \mid y, \theta)$ using current θ
6: Store θ as a sample
7: **until** the desired number of samples for θ is reached

2.5.4 The Metropolis Hastings algorithm

Metropolis-Hastings is a Markov chain Monte Carlo sampling method, proposed firstly by Metropolis et al. [89], who used it to do calculations in the field of statistical mechanics, and later generalized by Hastings [63]. Some descriptions of the Metropolis-Hastings sampling may be found in Hitchcock [65] and Chib and Greenberg [28].

Metropolis-Hastings have been used extensively in many variations and by many authors in different fields, cite here Jeliazkov [70], Geweke and Tanizaki [50], Roberts and Stramer [105], Cauchemez et al. [26], Demiris and O'Neill [32], Pratola [98], and Adaszewski et al. [2].

It is usually not possible to sample from the true conditional distribution, $\pi(\theta|x)$, such condition is unfavorite for the Gibbs Sampler. In contrast, Metropolis-Hastings algorithm is aptly suitable, in which we consider a *proposal function* $q(.|\theta)$ from which we generate samples of θ iteratively and then accept the proposal sample with an acceptance probability. Given the posterior density $\pi(\theta|x)$ and a proposal density $q(.|\theta)$, then the Metropolis-Hastings is provided by the Algorithm 2. The reader is referred to Hastings [63] and Robert and Casella [104].

Algorithm 2 Metropolis-Hastings algorithm (see: Robert and Casella [104])

1: Given an observed-data x.
2: Initialize $\theta^{(0)}$
3: **for** $i = 1, \ldots, N$ **do**
4: Sample a candidate $\widetilde{\theta} \backsim q\left(\theta \mid \theta^{(i-1)}, x\right)$ (using a Gibbs sampler step given in Algorithm 1)
5: Calculate $\alpha = \min\left(1, \frac{\pi\left(\widetilde{\theta}|x\right)}{\pi\left(\theta^{(i-1)}|x\right)} \frac{q\left(\theta^{(i-1)} \mid \widetilde{\theta}|x\right)}{q\left(\widetilde{\theta} \mid \theta^{(i-1)}, x\right)}\right)$
6: Accept or reject $\widetilde{\theta}$ with probability α
7: Update $\theta^{(i)} = \begin{cases} \widetilde{\theta} & \text{with probability} \alpha \\ \theta^{(i-1)} & \text{else} \end{cases}$
8: Store $\theta^{(i)}$ as a sample
9: **end for**

2.5.5 Bayesian inference for epidemic diffusion processes

Various alternative methods have been proposed to estimate parameters for epidemiological systems, the more sophisticated of which rely on the calculation of a likelihood function (e.g. Becker [17], Andersson and Britton [4], and Ionides et al. [68]). Likelihood based inference for epidemic models poses many challenges, not least 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 (e.g. using discrete-time models).

A popular and particularly effective solution is to use the Bayesian paradigm to estimate parameters, using numerical techniques such as Markov Chain Monte Carlo (MCMC) algorithms (e.g. Gibson and Renshaw [51], O'Neill and Roberts [91], Eraker [41], Streftaris and Gibson [116], El Maroufy et al. [36], and Qaffou et al. [100]. 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. However, as the population gets larger and/or the process gets more intricate, the likelihood can become mathematically or computationally intractable (e.g. Deardon et al. [31]).

There are various potential advantages of this system for modeling 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 through the use of metrics, usually 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 (such as the length of the infectious period) are sometimes known (e.g. through experimental studies or historical data).

Another major motivation for using these approaches for inference in epidemic models is that in many situations simulation algorithms are much faster to program and perform (Gillespie [55]) than repeated calculation of the likelihood function. Speed is particularly important in the face of an ongoing epidemic, where model predictions would need to be regularly updated and refined as new data emerge, and having a reliable but quick model-fitting algorithm is essential. The ability to combine elements of model and parameter uncertainty into model predictions is also an attractive property.

Chapter 3

Mathematical Modeling of HIV epidemic model

3.1 Introduction

The HIV dynamic is very complex and there is no other human infection, in a similar mode of transmission, that has the same epidemiological characteristics. For example, the incubation period, after HIV infection, is known that is long. During this period, individuals remain in good health and can transmit the disease to others without knowing it. Moreover, even if the disease is known as a sexually transmitted disease, but it is also transmitted by infected mothers to their babies and through the sharing of infected needles, which is common among injecting drug users. All of these factors made it difficult to understand the spread of this epidemic in the population.

Mathematical models based on the mechanism of HIV transmission might help scientific community to better understand how the disease spreads in the community. By developing such mathematical models, we can, in a certain way, predict its spread in different populations and evaluate the potential effectiveness of different approaches for bringing the epidemic under control, and thus help to devise effective strategies to minimize the destruction caused by this epidemic.

Since the first cases of the HIV/AIDS were recognized in the late 80s, mathematical models for the spread of this epidemic have been widely studied; see for example May and Anderson [86], Dietz [33], Tan and Zhu [121], Kremer and Morcom [77], Mode and Sleeman [90], and Tan and Wu [119] and the reference therein. However, this field of study is still challenging. Majority of the articles have focused on only a single population of constant size, while some works have concerned on variable population size in epidemic dynamics Dietz [33] and May et al. [87]. Note that, many models have only focused on a single homosexual population Tan and Xiang [120]. Whilst, heterosexual contact is the predominant mode of transmission in much of the world UNAIDS [126]. In this context, we are interested, in this dissertation, on modeling of the

spread of HIV/AIDS epidemic in a heterosexual closed population with size N; the population is divided into compartments. The epidemic of HIV is modeled by a multidimensional SIR process, which is a four-dimensional continuous-time Markovian jump process. In which each individual can find himself at a given time in one of five mutually exclusive health states named S_F (S_M) susceptible female (male), I_F (I_M) infected female (male) and $R(t)$ the AIDS cases, as presented in Fig. 3.1. Our model is motivated mostly by the work of Sani et al. [109], who used time dependence to approximate the diffusion model; while, here we use the convergence of the master equation and Fokker–Plank equation to derive the diffusion process.

The rest of this chapter is organized as follow; in section 2 we give the description of the adopted model. In section 3 we formulate the discrete space Markov chain and the deterministic model. While, in the section 4 we give a detailed formulation of the SDE epidemic model. A brief conclusion is in the last section.

3.2 Model description

We consider a closed and well mixed heterosexual population of size N. We assume that the infection can be made only by heterosexual contact. A single female or male selects her/his partner randomly from the whole population. We denote by $S_F(t)$, $I_F(t)$, $S_M(t)$, $I_M(t)$ and $Z(t)$ respectively, the sizes of susceptible females, infected females, susceptible males, infected males, and AIDS cases at time t. The population can be divided on compartments as presented in Fig. 3.1. The rate λ_F (λ_M) of infection of a female (male) susceptible is assumed to be proportional to the fraction of infected males (females) in their sub-population: $\lambda_F = \beta\dfrac{I_M(t)}{N_M(t)}$ $\left(\lambda_M = \beta\dfrac{I_F(t)}{N_F(t)}\right)$, λ_F and λ_M are called the forces of infection, where $N_F(t) = S_F(t) + I_F(t)$ and $N_M(t) = S_M(t) + I_M(t)$. The parameter β is the product of the contact rate k and the probability p that a successive number of contacts leads to infection ($k = \dfrac{1}{T}$ per unit time; $p = 1 - (1 - h)^{CT}$) where T is the time interval per partnership, C is the average number of sexual contact per partnership and h is the probability that one sexual contact between a susceptible and an infected individual leads to infection. We assume that all individuals, including AIDS people, leave the random mixing sexually active population at rate μ (due to natural death or for reasons other than dying). An individual at stage AIDS dies from the disease at rate δ. The individuals that leave the system are replaced by inflow of susceptible, at a proportion $0 \leq \alpha \leq 1$ for females and $(1 - \alpha)$ for males. Thus, the inflow rates for susceptible females and males are $B_F = \alpha(\mu N + \delta Z)$ and $B_M = (1-\alpha)(\mu N + \delta Z)$ respectively. The infected individuals develop AIDS at rate γ.

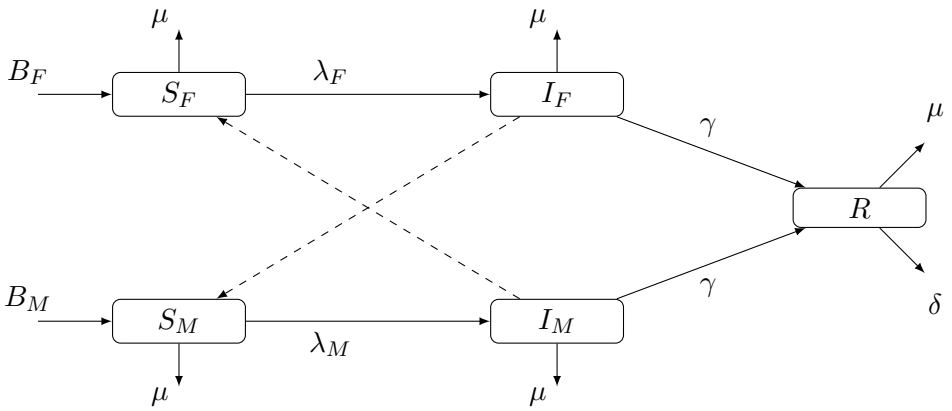

FIGURE 3.1: State diagram for the model of HIV. λ_F and λ_M are forces of infection, B_F and B_M are the inflow rates for susceptible females and males, μ is the mortality rate, γ is the removal rate, δ is the mortality rate from the disease . Dashed arrows indicate that infections are made via sexual contact only.

3.3 Deterministic model and formulation of discrete space Markov chain

The situation as described in Fig. 3.1, can be viewed as a multidimensional SIR model. Since, we have considered a closed population that is, $S_F(t) + I_F(t) + S_M(t) + I_M(t) + Z(t)$ is constant and equal to N for all t. The dynamic of epidemic is then, completely determined by the discrete process $\mathbf{Y}(t) = (S_F(t), I_F(t), S_M(t), I_M(t))$, which is supposed to be a continuous time Markov process with discrete space $\mathcal{D}_N = \Big\{(a, b, c, d) \in \mathbb{N}^4; a+b+c+d \leq N\Big\}$.

For a small time interval $[t, t + \Delta t]$ only one of the following events occurs; Birth of a susceptible female (male), infection of a susceptible female (male), natural death of a susceptible female (male) and a recovery or natural death of an infected female (male), all these events are resumed in the following schema:

$$
\begin{array}{ccc}
k - l & p_{k-l,k}\Delta t & k \\
(S_F - 1, I_F, S_M, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F, S_M - 1, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F + 1, I_F - 1, S_M, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F, S_M + 1, I_M - 1) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F + 1, I_F, S_M, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F, S_M + 1, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F + 1, S_M, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F, S_M, I_M + 1) & \longrightarrow & (S_F, I_F, S_M, I_M) \\
(S_F, I_F, S_M, I_M) & \longrightarrow & (S_F, I_F, S_M, I_M)
\end{array}
$$

$$k - l \quad p_{k-l,k}\Delta t \quad k$$

TABLE 3.1: All events in the pattern of the HIV epidemic, that can occur in a small time step.

Explicitly the process $\mathbf{Y}$ jumps from state $k-l$ at time t to state $k = (S_F, I_F, S_M, I_M)$ at time $t + \Delta t$ with transition $p_{k-l,k}\Delta t$. The rates $p_{k-l,k} = \lim_{\Delta t \to 0} \mathbb{P}(Y_{t+\Delta t} = k | Y_t = k - l)$ are given in details in Eq. (3.1), which resume all jumps.

$$p_{k-l,k} = \begin{cases} \alpha(\mu N + \delta(Z+1)), & l = e_1, \\ \dfrac{\beta I_M (S_F + 1)}{S_M + I_M}, & l = -e_1 + e_2, \\ \mu(S_F + 1), & l = -e_1, \\ (\mu + \gamma)(I_F + 1), & l = -e_2, \\ (1-\alpha)(\mu N + \delta(Z+1)) & l = e_3 \\ \dfrac{\beta I_F (S_M + 1)}{S_F + I_F}, & l = -e_3 + e_4, \\ \mu(S_M + 1), & l = -e_3, \\ (\mu + \gamma)(I_M + 1), & l = -e_4, \\ 1 - \left((\mu N + \delta Z) + \dfrac{\beta(I_M S_F + I_F S_M)}{S_M + I_M} + \mu\,(S_F + S_M) + (\mu+\gamma)\,(I_F + I_M) \right) & l = 0, \\ 0, & otherwise. \end{cases} \tag{3.1}$$

With (e_1, e_2, e_3, e_4) is the standard basis of $\mathbb{R}^4$.

The deterministic approximation of the corresponding multidimensional SIR model of the dynamic of HIV, as described in Fig. 3.1, and using the transition rates Eq. (3.1) is given by the following set of ordinary equations ODE (see, May et al. [87] and Sani et al. [109]):

$$\begin{cases} \dfrac{ds_F}{dt} = & \alpha(\mu + \delta z) - \mu s_F - \beta \dfrac{i_M}{s_M + i_M} s_F \\ \dfrac{di_F}{dt} = & \beta \dfrac{i_M}{s_M + i_M} s_F - (\mu + \gamma) i_F \\ \dfrac{ds_M}{dt} = & (1-\alpha)(\mu + \delta z) - \mu s_M - \beta \dfrac{i_F}{s_F + i_F} s_M \\ \dfrac{di_M}{dt} = & \beta \dfrac{i_F}{s_F + i_F} s_M - (\mu + \gamma) i_M \end{cases} \tag{3.2}$$

Since, the number of infective peoples is a stochastic quantity and the process posses the Markov property, then, the transitions from one state to another are well described by the master equation. Which can be obtained directly from Eq. (3.1). Let $\mathbb{P}(\mathbf{Y}(t) = k)$ be the probability for the process $\mathbf{Y}$ to be at state k at time t. Then, applying the Markov property, the difference equation satisfied by the probability $\mathbb{P}(\mathbf{Y}(t) = k)$ can be expressed in terms of the transition

probabilities. Then, the forward master equation is given by the sum over all jumps l,

$$\begin{aligned}
\frac{d\mathbb{P}(\mathbf{Y}(t)=k)}{dt} &= \sum_l \mathbb{P}(\mathbf{Y}_t = k|\mathbf{Y}_{t-\Delta t} = k-l)\mathbb{P}(\mathbf{Y}_{t-\Delta t} = k-l) \\
&= \alpha\left(\mu N+\delta(Z+1)\right)\mathbb{P}(S_F-1, I_F, S_M, I_M)+\frac{\beta I_M(S_F+1)}{S_M+I_M}\mathbb{P}(S_F+1, I_F-1, S_M, I_M) \\
&+ \mu(S_F+1)\mathbb{P}(S_F+1, I_F, S_M, I_M) + (\mu+\gamma)(I_F+1)\mathbb{P}(S_F, I_F+1, S_M, I_M) \\
&+(1-\alpha)\left(\mu N+\delta(Z+1)\right)\mathbb{P}\left(S_F, I_F, S_M-1, I_M\right)+\frac{\beta I_F(S_M+1)}{S_F+I_F}\mathbb{P}\left(S_F, I_F, S_M+1, I_M-1\right) \\
&+ \mu(S_M+1)\mathbb{P}\left(S_F, I_F, S_M+1, I_M\right) + (\mu+\gamma)(I_M+1)\mathbb{P}\left(S_F, I_F, S_M, I_M+1\right) \\
&- \Big(\alpha(\mu N+\delta Z) + \frac{\beta I_M S_F}{S_M+I_M} + \mu S_F + (\mu+\gamma)I_F + (1-\alpha)(\mu N+\delta Z) \\
&+\frac{\beta I_F S_M}{S_F+I_F} + \mu S_M + (\mu+\gamma)I_M\Big)\mathbb{P}\left(S_F, I_F, S_M, I_M\right). \qquad (3.3)
\end{aligned}$$

3.4 Formulation of SDE Epidemic Model

Instead of tracking the states by the discrete process $\mathbf{Y}$ in Eq. (3.3); we will shift to the continuous states process $\mathbf{X}$ by normalizing the process $\mathbf{Y}$, for large size of population (N sufficiently large), we have $\mathbf{X}(t) = \frac{\mathbf{Y}(t)}{N} = \left(x_1(t) = \frac{S_F(t)}{N}, y_1(t) = \frac{I_F(t)}{N}, x_2(t) = \frac{S_M(t)}{N}, y_2(t) = \frac{I_M(t)}{N}\right)$ and set $z(t) = 1-(x_1+y_1+x_2+y_2)$. The approximation of pure Markov jump processes by diffusion, has been widely discussed by some authors, who use martingale characterizations of Markov processes, convergence of solutions of stochastic equations and the theory of semi-group, let cite Kurtz [79] and Pollard [96] and recently this approach is well discussed in Øksendal [78] and Fuchs [46]. The Kolmogorov's Eq. (3.3) becomes,

$$\begin{aligned}
\frac{1}{N}\frac{\partial}{\partial t}\mathbb{P}(\mathbf{X}_t = \mathbf{x}) &= \alpha\left(\mu+\delta(z+\varepsilon)\right)\mathbb{P}(x_1-\varepsilon, y_1, x_2, y_2)+\frac{\beta y_2}{x_2+y_2}(x_1+\varepsilon)\mathbb{P}(x_1+\varepsilon, y_1-\varepsilon, x_2, y_2) \\
&+ \mu(x_1+\varepsilon)\mathbb{P}(x_1+\varepsilon, y_1, x_2, y_2) + (\mu+\gamma)(y_1+\varepsilon)\mathbb{P}(x_1, y_1+\varepsilon, x_2, y_2) \\
&+(1-\alpha)\left(\mu+\delta(z+\varepsilon)\right)\mathbb{P}\left(x_1, y_1, x_2-\varepsilon, y_2\right)+\frac{\beta y_1 x_2}{x_1+y_1}\mathbb{P}(x_1, y_1, x_2+\varepsilon, y_2-\varepsilon) \\
&+ \mu(x_2+\varepsilon)\mathbb{P}(x_1, y_1, x_2+\varepsilon, y_2) + (\mu+\gamma)(y_2+\varepsilon)\mathbb{P}(x_1, y_1, x_2, y_2+\varepsilon) \\
&-\Big((\mu+\delta z)+\frac{\beta y_2 x_1}{x_2+y_2}+\mu(x_1+x_2)+(\mu+\gamma)(y_1+y_2)+\frac{\beta y_1 x_2}{x_1+y_1}\Big)\mathbb{P}(x_1, y_1, x_2, y_2), \qquad (3.4)
\end{aligned}$$

where $\varepsilon = \frac{1}{N}$. By regrouping, adding and removing terms in Eq. (3.4), it can be written:

$$\begin{aligned}
\frac{1}{N}\frac{\partial}{\partial t}\mathbb{P}(\mathbf{X}_t = \mathbf{x}) &= \alpha\Big((\mu+\delta(z+\epsilon))\mathbb{P}(x_1-\epsilon,y_1,x_2,y_2) - (\mu+\delta z)\mathbb{P}(x_1,y_1,x_2,y_2)\Big)\\
&+\beta\frac{y_2}{x_2+y_2}\Big((x_1+\epsilon)\,\mathbb{P}\,(x_1+\epsilon,y_1-\epsilon,x_2,y_2) - (x_1+\epsilon)\,\mathbb{P}\,(x_1+\epsilon,y_1,x_2,y_2)\\
&-x_1\mathbb{P}\,(x_1,y_1-\epsilon,x_2,y_2)+x_1\mathbb{P}\,(x_1,y_1,x_2,y_2)\Big)+\beta\frac{y_2}{x_2+y_2}(x_1+\epsilon)\mathbb{P}\,(x_1+\epsilon,y_1,x_2,y_2)\\
&+\beta\frac{y_2}{x_2+y_2}\Big(x_1\mathbb{P}\,(x_1,y_1-\epsilon,x_2,y_2)-2x_1\mathbb{P}\,(x_1,y_1,x_2,y_2)\Big)+\mu(x_1+\epsilon)\mathbb{P}(x_1+\epsilon,y_1,x_2,y_2)\\
&+\mu x_1\mathbb{P}(x_1,y_1,x_2,y_2)+(\mu+\gamma)\Big((y_1+\epsilon)\mathbb{P}(x_1,y_1+\epsilon,x_2,y_2)-y_1\mathbb{P}(x_1,y_1,x_2,y_2)\Big)\\
&+(1-\alpha)\Big((\mu+\delta(z+\epsilon))\mathbb{P}(x_1,y_1,x_2-\epsilon,y_2) - (\mu+\delta z)\mathbb{P}(x_1,y_1,x_2,y_2)\Big)\\
&+\frac{\beta y_1}{x_1+y_1}\Big((x_2+\epsilon)\mathbb{P}(x_1,y_1,x_2+\epsilon,y_2-\epsilon) - (x_2+\epsilon)\mathbb{P}(x_1,y_1,x_2+\epsilon,y_2)\\
&-x_2\mathbb{P}(x_1,y_1,x_2,y_2-\epsilon)+x_2\mathbb{P}(x_1,y_1,x_2,y_2)\Big)+\beta\frac{y_1}{x_1+y_1}\Big((x_2+\epsilon)\mathbb{P}(x_1,y_1,x_2+\epsilon,y_2)\\
&+x_2\mathbb{P}(x_1,y_1,x_2,y_2-\epsilon) - 2x_2\mathbb{P}(x_1,y_1,x_2,y_2)\Big)+\mu\Big((x_2+\epsilon)\mathbb{P}(x_1,y_1,x_2+\epsilon,y_2)\\
&-x_2\mathbb{P}(x_1,y_1,x_2,y_2)\Big)+(\mu+\gamma)\Big((y_2+\epsilon)\mathbb{P}(x_1,y_1,x_2,y_2+\epsilon)-y_2\mathbb{P}(x_1,y_1,x_2,y_2)\Big).
\end{aligned} \tag{3.5}$$

This passage to continuum allows us to use standard tools of calculus. Then, the Taylor expansions up to the second order for multivariate functions (see, Azencott [12]). Let $N \to \infty$, thus $\epsilon \to 0$, We have then, the approximations of expressions between big brackets, are given in the following approximations. For the expression between big brackets in the first line on the right side of the Eq. (3.5) above is approximated by,

$$\begin{aligned}
&(\mu+\delta(z+\epsilon))\mathbb{P}(x_1-\epsilon,y_1,x_2,y_2) - (\mu+\delta z)\mathbb{P}(x_1,y_1,x_2,y_2)\\
&\qquad\simeq (\mu+\delta z)\left(-\epsilon\frac{\partial}{\partial x_1}+\frac{1}{2}\epsilon^2\frac{\partial}{\partial x_1}\right)\mathbb{P}(x_1,y_1,x_2,y_2).
\end{aligned}$$

The expression between big brackets in the second and third lines has approximated by,

$$\begin{aligned}
&(x_1+\epsilon)\mathbb{P}\,(x_1+\epsilon,y_1-\epsilon,x_2,y_2) - (x_1+\epsilon)\mathbb{P}(x_1+\epsilon,y_1,x_2,y_2) - x_1\mathbb{P}(x_1,y_1-\epsilon,x_2,y_2)\\
&\qquad + 2x_1\mathbb{P}(x_1,y_1,x_2,y_2) \simeq -\epsilon^2\frac{\partial^2}{\partial x_1\partial y_1}x_1\mathbb{P}(x_1,y_1,x_2,y_2).
\end{aligned}$$

The same thing for the expression, inside big brackets in the fourth line, we have,

$$\begin{aligned}
&((x_1+\epsilon)\mathbb{P}(x_1+\epsilon,y_1,x_2,y_2) + x_1\mathbb{P}(x_1,y_1-\epsilon,x_2,y_2) - 2x_1\mathbb{P}(x_1,y_1,x_2,y_2))\\
&\qquad\simeq\left(\epsilon\left(\frac{\partial}{\partial x_1}-\frac{\partial}{\partial y_1}\right)+\epsilon^2\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial y_1^2}\right)\right)x_1\mathbb{P}(x_1,y_1,x_2,y_2).
\end{aligned}$$

For the fifth line, the expression has the following approximations,

$$(x_1+\epsilon)\mathbb{P}(x_1+\epsilon,y_1,x_2,y_2)-x_1\mathbb{P}(x_1,y_1,x_2,y_2)\simeq\left(\epsilon\frac{\partial}{\partial x_1}+\frac{1}{2}\epsilon^2\frac{\partial}{\partial x_1}\right)\mathbb{P}(x_1,y_1,x_2,y_2).$$

The fifth between expression is approximated by,

$$(y_1+\epsilon)\mathbb{P}(x_1,y_1+\epsilon,x_2,y_2)-y_1\mathbb{P}(x_1,y_1,x_2,y_2)\simeq\left(\epsilon\frac{\partial}{\partial y_1}+\frac{1}{2}\epsilon^2\frac{\partial}{\partial y_1}\right)y_1\mathbb{P}(x_1,y_1,x_2,y_2)$$

For the remaining expressions in Eq. (3.5), they have the same development by exchanging x_1, y_1 and α by x_2, y_2 and $1-\alpha$ respectively.

Integrate these approximations into the Eq. (3.5), let $\epsilon=N^{-1}$, then, the Eq. (3.5) becomes,

$$\begin{aligned}\frac{1}{N}\frac{\partial}{\partial t}\mathbb{P}(x_1,y_1,x_2,y_2)=\frac{1}{N}\Bigg[&\frac{\partial}{\partial x_1}\left(-\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2x_1}{x_2+y_2}\right)+\frac{\partial}{\partial y_1}\left((\mu+\gamma)y_1+\frac{\beta y_2x_1}{x_2+y_2}\right)\\&+\frac{\partial}{\partial x_2}\left(-(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1x_2}{x_1+y_1}\right)+\frac{\partial}{\partial y_2}\left((\mu+\gamma)y_2+\frac{\beta y_1x_2}{x_1+y_1}\right)\Bigg]\mathbb{P}(x_1,y_1,x_2,y_2)\\&+\frac{1}{2N^2}\Bigg[\frac{\partial^2}{\partial x_1^2}\left(\alpha(\mu+\delta z)+\mu x_1\right)+\frac{\partial^2}{\partial y_1^2}\left((\mu+\gamma)y_1\right)+\frac{\partial^2}{\partial x_2^2}\Big((1-\alpha)(\mu+\delta z)+\mu x_2\Big)\\&+\frac{\partial^2}{\partial y_2^2}\left((\mu+\gamma)y_2\right)-2\frac{\partial^2}{\partial x_1\partial y_1}\left(\frac{\beta y_2x_1}{x_2+y_2}\right)-2\frac{\partial^2}{\partial x_2\partial y_2}\left(\frac{\beta y_1x_2}{x_1+y_1}x_2\right)\Bigg]\mathbb{P}(x_1,y_1,x_2,y_2)\end{aligned}\tag{3.6}$$

simplifying by N^{-1} in two sides of the equality in Eq. (3.6) and writing it in vectorial form using the arguments: $\frac{\partial}{\partial\mathbf{x}}=\left(\frac{\partial}{\partial\mathbf{x_i}}\right)_{1\le i\le 4}$ and $\frac{\partial^2}{\partial\mathbf{x}^2}=\left(\frac{\partial^2}{\partial\mathbf{x_i}\partial\mathbf{x_j}}\right)_{1\le i,j\le 4}$, the Eq. (3.6) can be rewritten as

$$\frac{\partial}{\partial t}\mathbb{P}(\mathbf{X}_t=\mathbf{x})=-\frac{\partial}{\partial\mathbf{x}}\left[\xi(\mathbf{x})\mathbb{P}(\mathbf{x},t)\right]+\frac{1}{2}\frac{\partial^2}{\partial\mathbf{x}^2}\left[\frac{1}{N}\mathbf{\Sigma}(\mathbf{x})\mathbb{P}(\mathbf{x},t)\right],\tag{3.7}$$

where,

$$\xi(\mathbf{x})=\begin{pmatrix}\alpha(\mu+\delta z)-\mu x_1-\dfrac{\beta y_2x_1}{x_2+y_2}\\-(\mu+\gamma)y_1+\dfrac{\beta y_2x_1}{x_2+y_2}\\(1-\alpha)(\mu+\delta z)-\mu x_2-\dfrac{\beta y_1x_2}{x_1+y_1}\\-(\mu+\gamma)y_2+\dfrac{\beta y_1x_2}{x_1+y_1}\end{pmatrix},\tag{3.8}$$

$$\boldsymbol{\Sigma}(\mathbf{x})=\begin{pmatrix} \alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2 x_1}{x_2+y_2} & -\frac{\beta y_2 x_1}{x_2+y_2} & 0 & 0 \\ \frac{-\beta y_2 x_1}{x_2+y_2} & (\mu+\gamma)y_1+\frac{\beta y_2 x_1}{x_2+y_2} & 0 & 0 \\ 0 & 0 & (1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1 x_2}{x_1+y_1} & \frac{-\beta y_1 x_2}{x_1+y_1} \\ 0 & 0 & \frac{-\beta y_1 x_2}{x_1+y_1} & (\mu+\gamma)y_2+\frac{\beta y_1 x_2}{x_1+y_1} \end{pmatrix}. \tag{3.9}$$

Since $\boldsymbol{\Sigma}$, as defined in Eq. (3.9), positive definite then the Eq. (3.7) is the forward diffusion. This equation corresponds to a diffusion process with drift vector ξ and diffusion matrix $\frac{\boldsymbol{\Sigma}}{N}$, *i.e.* the intensive Markov jump process $(X_t,\ t \geq 0)$ can be approximated by a diffusion (see, Øksendal [78] and Fuchs [46] for more details) satisfying the SDE :

$$dX_t = \begin{pmatrix} \alpha(\mu+\delta z)-\mu x_1-\frac{\beta y_2 x_1}{x_2+y_2} \\ -(\mu+\gamma)y_1+\frac{\beta y_2 x_1}{x_2+y_2} \\ (1-\alpha)(\mu+\delta z)-\mu x_2-\frac{\beta y_1 x_2}{x_1+y_1} \\ -(\mu+\gamma)y_2+\frac{\beta y_1 x_2}{x_1+y_1} \end{pmatrix} dt + \frac{1}{\sqrt{N}}\boldsymbol{\sigma}(X_t,\theta)\begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \\ dW_4 \end{pmatrix}$$

So, that

$$dX_t = \xi(X_t,\theta)dt + \frac{1}{\sqrt{N}}\boldsymbol{\sigma}(X_t,\theta)dW(t), \tag{3.10}$$

where $\boldsymbol{\sigma}$ is such that $\boldsymbol{\sigma}\boldsymbol{\sigma}^T = \boldsymbol{\Sigma}$, here $\boldsymbol{\sigma}^T$ is transposed matrix of the matrix $\boldsymbol{\sigma}$. A cholesky decomposition yields $\boldsymbol{\sigma} = \begin{pmatrix} \boldsymbol{\sigma}_1 & 0 \\ 0 & \boldsymbol{\sigma}_2 \end{pmatrix}$ with

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} \sqrt{\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2 x_1}{x_2+y_2}} & 0 \\ \frac{-\frac{\beta y_2 x_1}{x_2+y_2}}{\sqrt{\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2 x_1}{x_2+y_2}}} & \sqrt{\frac{(\mu+\gamma)y_1\left[\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2 x_1}{x_2+y_2}\right]+\frac{\beta y_2 x_1}{x_2+y_2}\left[\alpha(\mu+\delta z)+\mu x_1\right]}{\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2 x_1}{x_2+y_2}}} \end{pmatrix}$$

$$\boldsymbol{\sigma}_2 = \begin{pmatrix} \sqrt{(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1 x_2}{x_1+y_1}} & 0 \\ \frac{-\frac{\beta y_1 x_2}{x_1+y_1}}{\sqrt{(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1 x_2}{x_1+y_1}}} & \sqrt{\frac{(\mu+\gamma)y_2\left[(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1 x_2}{x_1+y_1}\right]+\frac{\beta x_2 y_1}{x_1+y_1}\left[(1-\alpha)(\mu+\delta z)+\mu x_2\right]}{(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1 x_2}{x_1+y_1}}} \end{pmatrix}.$$

The stochastic components W_1, W_2, W_3 and W_4 are standard independent Brownian motions, representing the stochastic effect in disease transmission and recovers.

Hence, we had well modeled the dynamic of HIV in a heterosexual population, studied in this dissertation, presented in Fig. 3.1 by a stochastic diffusion process Eq. (3.10). We recall that, our main goal is to illustrate an efficient method to estimate $\theta = (\mu, \beta, \gamma, \delta)$, which is the goal

of two following chapters.

Different diffusion approximation techniques are investigated in literature, let introduce: the convergence of the Master Equation, this technique has been used *e.g.* by Goel and Richter-Dyn [57] and Gillespie [54]. The convergence of the infinitesimal generator of the jump process (see the book Kurtz [79]). The Langevin approach has been studied by many authors Ramshaw [101], Golightly and Wilkinson [59], Wilkinson [129], Tian et al. [122], Golightly and Wilkinson [58], and Seifert [111]. The Kramers-Moyal expansion has been applied by Strumik and Macek [117], Hufnagel et al. [66], and Capasso and Morale [23]. The reader is referred to the book of Fuchs [46, Chap. 4] for a review.

3.5 Conclusion

In this chapter, we have been introduced and described the adopted model, which correspond to the transmission of the HIV dynamic in a closed heterosexual population. This model has been presented in its deterministic form firstly. Finally, we obtained the SDE epidemic model, using convergence of the master equations. Such modeling could be adopted for a family of other infectious diseases.

Chapter 4

Minimum contrast estimator of HIV epidemic model's parameters [2]

4.1 Introduction

Parameter estimation for diffusion processes with small noise based on continuous-time observations, case where the whole path is observed has mainly developed in the literature. The asymptotic statistical theory has been well studied by Kutoyants [81] and Kutoyants [80], Yoshida [131] showed the validity of asymptotic expansions for statistical estimators; see also Sørensen and Uchida [115] and the reference therein. Since the process $\{X_t, t \geq 0\}$ is Markovian, hence if the transition densities of $\{X_t, t \geq 0\}$ are known, one can use the log-likelihood function for estimation of θ. In the context of the SDE's, the maximum likelihood estimate (MLE) is known to have usual good properties (see, Dacunha-Castelle and Florens-Zmirou [29] and Bibby et al. [19]).

Consider the diffusion process $\{X_t, t \geq 0\}$ defined by the stochastic differential equation

$$dX_t = \xi(X_t, \theta)dt + \varepsilon\boldsymbol{\sigma}(X_t, \theta)dW(t), \quad t \geq 0 \tag{4.1}$$

where $\{W_t, t \geq 0\}$ is a four-dimensional standard Brownian, $\varepsilon = \dfrac{1}{\sqrt{N}}$ (N is the population size) , $\theta \subset \Theta$ is the unknown parameter which is to be estimated on the basis of observations of the process $\{X_t, t \geq 0\}$ at times $0 = t_0 < t_1 < ... < t_n$ and ξ and $\boldsymbol{\sigma}$ are known smooth functions as defined in Eqs. (3.8) and (3.9). Unfortunately, the transition densities of $\{X_t, t \geq 0\}$ are usually unknown except in some very special cases. In recent years, the more realistic case of parametric estimation for discretely observed diffusion processes has also been studied by many researchers; see Pedersen [93], Andersson and Britton [4], Aït-Sahalia [10], and Bibby et al. [19].

Statistical methods associated to discrete data have been developed in the asymptotics of a small diffusion coefficient by Kessler [73]. The asymptotic properties of estimators were largely studied

[2] Part of this chapter appears in Abou-Bakre and El Maroufy [1]

over the two last decades (*e.g.* Genon-Catalot [48]; Sørensen and Uchida [115]; Gloter and Sørensen [56]). Sørensen and Uchida obtained consistent and normally asymptotic estimators of drift and diffusion coefficients. Uchida [125] obtained consistent and normal asymptotic estimators of both drift and diffusion coefficients. In this chapter, we treat a particular case of multidimensional diffusion processes defined by the SDE (2.4) for which the model parameters are in both drift and diffusion coefficients. We prove consistency and asymptotic-normality of estimators of the parameter θ in both drift and diffusion coefficients.

This chapter is organized as follows. After this introductory section, we formulate and establish the contrast function and the corresponding minimum estimator in the second section. The third section treats the asymptotic properties of the estimator, that is the consistency and asymptotic normality of the minimum contrast estimator. In the fourth section, we present the proofs of some essential results. Finally, in the last section we present numerical results and simulations.

4.2 Minimum contrast estimator for diffusion

The statistical inference for continuously observed diffusion processes on a finite interval is based on the likelihood of the diffusion which is obtained using the Girsanov formula (see *e.g.* Kutoyants [82] and Liptser and Shiryaev [84]). Discretely observed diffusions are discrete time Markov processes and thus their likelihood depends on the transition probabilities densities of the diffusion. However, in our case, these densities are unavailable in simple closed form (see, El Maroufy et al. [38, 37]), the likelihood is then intractable. As an alternative we will adopt the approach given by Kessler [73], based on approximation of the log-likelihood. They are often called contrast function with their associated minimum contrast estimators (MCE).

From now on, the following notations will be used, let Θ be a compact subset of $\mathbb{R}^4$, θ_0 the true value of parameter $\theta = (\mu, \beta, \gamma, \delta)$ belongs to $\mathring{\Theta}$ the interior of Θ. Let $\varepsilon = \dfrac{1}{\sqrt{N}}$ and X_t^0 solution of the ordinary differential equation, $dX_t^0 = \xi(X_t^0, \theta_0)dt$. Let $\mathbb{P}_{\theta_0}$ the law of the solution of the process (2.4), $W(t)$ be a $four-$dimensional standard Brownian on the filtered probability space $\left(\Omega, \mathcal{F}_t, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{\theta_0}\right)$ where $\mathcal{F}$ is the natural filtration of X. Let make the following assumptions:

Assumption 1. *i. Equation (2.4) has an unique and strong solution in* $[0, T]$.

ii. $\forall m > 0, \sup_t \mathbb{E}(|X_t|^m) < \infty$.

Assumption 2. *i.* $\xi(x, \theta) \in \mathcal{C}^3(\mathbb{R}^4 \times \bar{\Theta}; \mathbb{R}^4)$ *and* $\boldsymbol{\sigma}(x, \theta) \in \mathcal{C}^2(\mathbb{R}^4 \times \bar{\Theta}; \mathbb{R}^4 \otimes \mathbb{R}^4)$.

ii. $\inf_{x,\theta} \det[\boldsymbol{\sigma}\boldsymbol{\sigma}^T] > 0$ *and* $[\boldsymbol{\sigma}\boldsymbol{\sigma}^T]^{-1}(x, \theta) \in \mathcal{C}^2(\mathbb{R}^4 \times \bar{\Theta}; \mathbb{R}^4 \otimes \mathbb{R}^4)$.

Assumption 3. $\theta \neq \theta_0 \implies \xi(x, \theta) \neq \xi(x, \theta_0)$ *and* $\boldsymbol{\sigma}(x, \theta) \neq \boldsymbol{\sigma}(x, \theta_0)$.

The uniqueness and existence of a strong solution of the equation (2.4) are guaranteed by two conditions, local Lipschitz (2.6) and linear growth (2.5) of both coefficients drift and diffusion, for

more details (see the proof in Appendix A). In order to obtain an estimator of the model parameter θ, we construct the contrast function based on a Gaussian approximation of the transition density in the same way as Kessler [73] and Sørensen and Uchida [115]. Thereby, using an n-sample of time equidistant observations of the process X, $\left\{X_{t_k}, k = 1 \dots n; t_{k+1} = t_k + \Delta, \Delta = \frac{1}{n}\right\}$. Then, we have the following contrast function

$$U_{\varepsilon,n}(\theta) = \sum_{k=1}^{n} \left[\log\left(\det\left[\mathbf{\Sigma}_{k-1}\right]\right) + \frac{1}{\Delta} \varepsilon^{-2} \mathbf{P}_k(\theta)^T \mathbf{\Sigma}_{k-1}^{-1} \mathbf{P}_k(\theta) \right], \tag{4.2}$$

where $\mathbf{P}_k(\theta) = X_{t_k} - X_{t_{k-1}} - \frac{1}{n}\xi(X_{t_{k-1}}, \theta)$ and $\mathbf{\Sigma}_k = \mathbf{\Sigma}(X_{t_k}, \theta) = \boldsymbol{\sigma}^T \boldsymbol{\sigma}(X_{t_k}, \theta)$. After a tedious algebraic calculus, and let

$\Delta_1{}^2 = (\mu+\gamma) y_1 \left(\alpha(\mu + \delta z) + \mu x_1 + \dfrac{\beta y_2 x_1}{x_2 + y_2} \right) + \dfrac{\beta y_2 x_1}{x_2 + y_2} \left(\alpha(\mu + \delta z) + \mu x_1\right)$ and

$\Delta_2{}^2 = (\mu+\gamma) y_2 \left((1-\alpha)(\mu + \delta z) + \mu x_2 + \dfrac{\beta y_1 x_2}{x_1 + y_1} x_2 \right) + \dfrac{\beta y_1 x_2}{x_1 + y_1} \left((1-\alpha)(\mu + \delta z) + \mu x_2\right)$.

The explicit form of the contrast function (4.2), is finally given by the following expression:

$$\begin{aligned} U_{n,\varepsilon}(\theta) = \frac{1}{2} \sum_{k=1}^{n} \left[\log\left(\Delta_1\right) + \log\left(\Delta_2\right)\right] + n\varepsilon^{-2} \sum_{k=1}^{n} \Bigg[& P_1{}^2 \left(\frac{(\mu+\gamma) y_1{}^{k-1} + \beta \frac{{y_2}^{k-1} {x_1}^{k-1}}{{x_2}^{k-1} + {y_2}^{k-1}}}{\Delta_1{}^2} \right) \\ & + 2 P_1 P_2 \left(\frac{\beta \frac{{y_2}^{k-1}}{{x_2}^{k-1} + {y_2}^{k-1}} {x_1}^{k-1}}{\Delta_1{}^2} \right) + P_2{}^2 \left(\frac{(1-\alpha)(\mu + \delta z^{k-1}) + \mu {x_2}^{k-1} + \beta \frac{{y_1}^{k-1} {x_2}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}}}{\Delta_1{}^2} \right) \\ & + P_3{}^2 \left(\frac{(\mu+\gamma) y_2^{k-1} + \beta \frac{x_1^{k-1}}{{x_1}^{k-1} + y_1^{k-1}} {y_2}^{k-1}}{\Delta_2{}^2} \right) + 2 P_3 P_4 \left(\beta \frac{\frac{{y_1}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}} {x_2}^{k-1}}{\Delta_2{}^2} \right) \\ & + P_4{}^2 \left(\frac{\alpha(\mu + \delta z^{k-1}) + \mu {x_2}^{k-1} + \beta \frac{{y_1}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}} {x_2}^{k-1}}{\Delta_2{}^2} \right) \Bigg], \end{aligned} \tag{4.3}$$

where the arguments P_1, P_2, P_3 and P_4 are given by:

$$\begin{aligned} P_1 &= \Delta x_1^k - \frac{1}{n} \left(\alpha(\mu + \delta z^{k-1}) - \mu {x_1}^{k-1} - \frac{\beta {y_2}^{k-1} {x_1}^{k-1}}{{x_2}^{k-1} + {y_2}^{k-1}} \right), \\ P_2 &= \Delta y_2^k - \frac{1}{n} \left(-(\mu+\gamma) {y_2}^{k-1} + \frac{\beta {y_1}^{k-1} {x_2}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}} \right), \\ P_3 &= \Delta x_2^k - \frac{1}{n} \left((1-\alpha)(\mu + \delta z^{k-1}) - \mu {x_2}^{k-1} - \frac{\beta {y_1}^{k-1} {x_2}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}} \right), \\ P_4 &= \Delta y_2^k - \frac{1}{n} \left((\mu+\gamma) {y_2}^{k-1} + \frac{\beta {y_1}^{k-1} {x_2}^{k-1}}{{x_1}^{k-1} + {y_1}^{k-1}} \right). \end{aligned}$$

4.3 Consistency and asymptotic normality of MCE

Let define the following arguments $B(x,\theta_0,\theta)=\xi(x,\theta_0)-\xi(x,\theta)$,

$$\mathbf{I}(\theta_0)=\left(\int_0^1\left(\frac{\partial}{\partial\theta_i}\xi(X_s^0,\theta_0)\right)\mathbf{\Sigma}^{-1}(X_s^0,\theta_0)\left(\frac{\partial}{\partial\theta_j}\xi(X_s^0,\theta_0)\right)ds\right)_{1\leq i,j\leq 4}$$

and

$$\begin{aligned}U(\theta_0,\theta)=&\int_0^1\log\det\mathbf{\Sigma}(X_s^0,\theta)ds+\int_0^1 tr\left[\mathbf{\Sigma}(X_s^0,\theta_0)\mathbf{\Sigma}^{-1}(X_s^0,\theta)\right]ds\\&+\frac{\varepsilon^2}{n}\int_0^1 B^T(X_s^0,\theta_0,\theta)\mathbf{\Sigma}^{-1}(X_s^0,\theta)B(X_s^0,\theta_0,\theta)ds.\end{aligned}$$

Let $\hat{\theta}_{\varepsilon,n}$ be a minimum contrast estimator defined by $\hat{\theta}_{\varepsilon,n}=\arg\min_{\theta\in\bar{\Theta}}U_{\varepsilon,n}(\theta)$[1]. We have the following result.

Theorem 4.1. *Suppose that the Assumption 1-3 are fulfilled. Then we have, when $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$*

i. $\hat{\theta}_{\varepsilon,n}\longrightarrow\theta_0$ *in $\mathbb{P}_{\theta_0}$-probability*

ii. $\varepsilon^{-1}\left(\hat{\theta}_{\varepsilon,n}-\theta_0\right)\longrightarrow\mathcal{N}(0,\mathbf{I}^{-1}(\theta_0))$, *in distribution under $\mathbb{P}_{\theta_0}$ if the matrix $\mathbf{I}(\theta_0)$ is non-singular.*

The first statement of the theorem means that the MCE estimator is consistent and converge to the true value of the parameter θ as $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$, while the second property means the normal asymptotic property of the estimator, which let build the confidence intervals. In order to prove the previous statements we will need the following Lemmas,

Lemma 4.1. *Suppose that Assumption 1-3 hold true. Then,*

$$\sup_{\theta\in\bar{\Theta}}\left|\frac{1}{n}U_{\varepsilon,n}(\theta)-U(\theta,\theta_0)\right|\longrightarrow 0,$$

in $\mathbb{P}_{\theta_0}-$probability, as $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$.

Proof. The detailed proof is given in Uchida [125, Poposition 3.1] □

Lemma 4.2. *If Assumption 1-3 are fulfilled. Then,*

$$-\varepsilon\frac{\partial}{\partial\theta}U_{\varepsilon,n}(\theta_0)\longrightarrow\mathcal{N}(0,4\mathbf{I}(\theta_0)),$$

in distribution, under $\mathbb{P}_{\theta_0}$, as $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$.

[1]The notation arg max is an abbreviation of "the argument of the minimum". It is the set of points of a given argument for which the given function attains its lowest value.

Proof of Lemma 4.2. Our proof starts with observation that for $i = \{1, 2, 3, 4\}$

$$\begin{aligned}\frac{\partial}{\partial\theta_i}U(\theta_0) = &\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\log\left(\det\boldsymbol{\Sigma}_{k-1}(\theta_0)\right) + 2\epsilon^{-2}n\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\mathbf{P}_k^T(\theta_0)\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\mathbf{P}_k(\theta_0)\\ &+\epsilon^{-2}n\sum_{k=1}^{n}\mathbf{P}_k^T(\theta_0)\left[\frac{\partial}{\partial\theta_i}\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\right]\mathbf{P}_k(\theta_0),\end{aligned}$$

It follows that

$$\begin{aligned}-\epsilon\frac{\partial}{\partial\theta_i}U(\theta_0) =& -\epsilon\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\log\left(\det\boldsymbol{\Sigma}_{k-1}(\theta_0)\right) + 2\epsilon^{-1}\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\mathbf{P}_k(\theta_0)\\ &-\epsilon\sum_{k=1}^{n}\frac{\mathbf{P}_k^T(\theta_0)}{\epsilon\sqrt{\Delta}}\left[\frac{\partial}{\partial\theta_i}\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\right]\frac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}}\\ =&\underbrace{-\epsilon\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\log\left(\det\boldsymbol{\Sigma}_{k-1}(\theta_0)\right)}_{T_1^i} + \underbrace{2\sqrt{\Delta}\sum_{k=1}^{n}\left[\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\right]\frac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}}}_{T_2^i}\\ &\underbrace{-\epsilon\sum_{k=1}^{n}\frac{\mathbf{P}_k^T(\theta_0)}{\epsilon\sqrt{\Delta}}\left[\frac{\partial}{\partial\theta_i}\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\right]\frac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}}}_{T_3^i}\\ =& -\epsilon T_1^i + 2\sqrt{\Delta}T_2^i - \epsilon T_3^i.\end{aligned} \tag{4.4}$$

Since the differentiability of $\boldsymbol{\Sigma}^{-1}$ with respect to θ and $\inf_{x,\theta}(\det\boldsymbol{\Sigma}) > 0$ (Assumption 2 on page 40), and the fact that the process $\dfrac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}}$ is a Gaussian process with mean 0 and $\boldsymbol{\Sigma}$ as covariance matrix, hence the arguments T_1^i, T_2^i and T_3^i in Eq. (4.4) are bounded with respect to the probability $\mathbb{P}_{\theta_0}$, for each i, as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$.

So in $\mathbb{P}_{\theta_0}$-probability as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$,

$$T_2^i = \sum_{k=1}^{n}\left[\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\right]\frac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}},$$

has a normal distribution with *zero* mean and covariance matrix

$$\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0).$$

Finally,

$$2\sqrt{\Delta}T_2^i \longrightarrow \mathcal{N}\left(0, 4\Delta\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\boldsymbol{\Sigma}_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)\right), \tag{4.5}$$

as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$ in distribution.
By Eq. (4.4), we have, as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$.

$$-\epsilon \frac{\partial}{\partial \theta_i} U(\theta_0) \longrightarrow \lim_{\epsilon \longrightarrow 0, n \longrightarrow \infty} 2\sqrt{\Delta} T_2^i$$

and Eq. (4.5) implies that,

$$T_2^i \longrightarrow \mathcal{N}(0, \mathbf{I}(\theta_0)).$$

This finishes the proof of Lemma 4.2.

□

Lemma 4.3. *If Assumption 1-3 are fulfilled. Then,*

$$-\varepsilon^2 \frac{\partial^2}{\partial \theta_i \partial \theta_j} U_{\varepsilon,n}(\theta_0) \longrightarrow 2\mathbf{I}(\theta_0)_{i,j},$$

in $\mathbb{P}_{\theta_0}$ *– probability, as* $\varepsilon \longrightarrow 0$ *and* $n \longrightarrow \infty$.

Proof of Lemma 4.3. We begin seeing that for each $i, j = \{1, 2, 3, 4\}$

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_i \partial \theta_j} U(\theta_0) = & \sum_{k=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log\left(\det \mathbf{\Sigma}_{k-1}(\theta_0)\right) + 2\epsilon^{-2} n \sum_{k=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbf{P}_k^T(\theta_0) \mathbf{\Sigma}_{k-1}^{-1}(\theta_0) \mathbf{P}_k(\theta_0) \\
& + 2\epsilon^{-2} n \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \mathbf{P}_k^T(\theta_0) \left[\frac{\partial}{\partial \theta_j} \mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\right] \mathbf{P}_k(\theta_0) \\
& + 2\epsilon^{-2} n \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \mathbf{P}_k^T(\theta_0) \mathbf{\Sigma}_{k-1}^{-1}(\theta_0) \frac{\partial}{\partial \theta_j} \mathbf{P}_k(\theta_0) \\
& + \epsilon^{-2} n \sum_{k=1}^{n} \mathbf{P}_k^T(\theta_0) \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\right] \mathbf{P}_k(\theta_0).
\end{aligned}$$

The main idea of the proof is to separate the previous sum as follows

$$\begin{aligned}
\epsilon^2 \frac{\partial^2}{\partial \theta_i \partial \theta_j} U(\theta_0) = & \underbrace{\epsilon^2 \left[\sum_{k=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log\left(\det \mathbf{\Sigma}_{k-1}(\theta_0)\right) + \sum_{k=1}^{n} \frac{\mathbf{P}_k^T(\theta_0)}{\epsilon\sqrt{\Delta}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\right] \frac{\mathbf{P}_k(\theta_0)}{\epsilon\sqrt{\Delta}}\right]}_{T_1^{i,j}} \\
& + \underbrace{2 \sum_{k=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \xi(X_{t_{k-1}}, \theta_0)^T \mathbf{\Sigma}_{k-1}^{-1}(\theta_0) \mathbf{P}_k(\theta_0)}_{T_2^{i,j}} \\
& + \underbrace{2 \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \xi(X_{t_{k-1}}, \theta_0)^T \left[\frac{\partial}{\partial \theta_j} \mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\right] \mathbf{P}_k(\theta_0)}_{T_3^{i,j}}
\end{aligned}$$

$$+ 2\underbrace{\frac{1}{n}\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_j}\xi(X_{t_{k-1}},\theta_0)}_{T_4^{i,j}}$$

$$= \epsilon^2 {T_1}^{i,j} + 2{T_2}^{i,j} + 2{T_3}^{i,j} + 2\frac{1}{n}{T_4}^{i,j}. \tag{4.6}$$

Putting $\Delta = \frac{1}{2}$, The Lemma B.1 ensures that arguments ${T_2}^{i,j}$ and ${T_3}^{i,j}$ tend to 0 as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Since ${T_1}^{i,j}$ defined in Eq. (4.6), are bounded in $\mathbb{P}_{\theta_0}$-probability, then, $\epsilon^2 {T_1}^{i,j} \longrightarrow 0$. For the term ${T_4}^{i,j}$, since the conditions of Lemma B.1 are guaranteed for $\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_j}\xi(X_{t_{k-1}},\theta_0)$, it follows that, for each i and j

$$2\frac{1}{n}{T_4}^{i,j} = 2\frac{1}{n}\sum_{k=1}^{n}\frac{\partial}{\partial\theta_i}\xi(X_{t_{k-1}},\theta_0)^T\mathbf{\Sigma}_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_j}\xi(X_{t_{k-1}},\theta_0) \longrightarrow 2\left(\mathbf{I}(\theta_0)\right)_{i,j}$$

as $\epsilon \longrightarrow 0$ and $n \longrightarrow \infty$. This completes the proof. □

4.4 Proof of Theorem 4.1

Proof. For the first statement of Theorem 4.1, which treats the consistency of estimator $\hat{\theta}_{\varepsilon,n}$, we have, from a version of Lemma 17 in Genon-Catalot [48],

$$\log\det[\mathbf{\Sigma}(X_t^0,\theta)] + tr\left[[\mathbf{\Sigma}(X_t^0,\theta_0)][\mathbf{\Sigma}^{-1}(X_t^0,\theta)]\right] \geq \log\det[\mathbf{\Sigma}(X_t^0,\theta)] + d,$$

where d is the dimension of the process, here $d = 4$, with equality if,

$$\mathbf{\Sigma}(X_t^0,\theta_0) = \mathbf{\Sigma}(X_t^0,\theta).$$

By assumption A_2, we obtain,

$$\int_0^1 B^T(X_t^0,\theta,\theta_0)\left[\mathbf{\Sigma}^{-1}(X_t^0,\theta)\right]B(X_t^0,\theta,\theta_0)ds > 0,$$

with equality if and only if $\xi(X_t^0,\theta) = \xi(X_t^0,\theta_0)$. Thus, it follows from Assumption 3 that,

$$U(\theta,\theta_0) \geq U(\theta_0,\theta_0),$$

with equality if $\theta = \theta_0$. Therefore for any $\eta > 0$,

$$\inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta,\theta_0) > U(\theta_0,\theta_0). \tag{4.7}$$

Moreover from the definition of $\hat{\theta}_{\varepsilon,n}$ we have,

$$U_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) = \inf_{\theta\in\bar{\Theta}} U_{\varepsilon,n}(\theta),$$

and the fact $\theta_0 \in \bar{\Theta}$ then, for any $\eta > 0$ and $\varepsilon \longrightarrow 0$, $n \longrightarrow \infty$,

$$\mathbb{P}_{\theta_0}\left[\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) \leq \bar{U}_{\varepsilon,n}(\theta_0) + \eta\right] \longrightarrow 1, \tag{4.8}$$

where $\bar{U}_{\varepsilon,n}(\theta) = \dfrac{1}{n}U_{\varepsilon,n}(\theta)$.

From Eq. (4.7), for every $\eta > 0$, there exist $\eta^{'} > 0$ such that

$$\inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta,\theta_0) > U(\theta_0,\theta_0) + \eta^{'}.$$

Furthermore, for every η there exists $\eta^{'} > 0$ such that

$$\left|\hat{\theta}_{\varepsilon,n} - \theta_0\right| \geq \eta \text{ then,} U(\hat{\theta}_{\varepsilon,n},\theta_0) \geq \inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta,\theta_0) > U(\theta_0,\theta_0) + \eta^{'}.$$

Hence,

$$\begin{aligned}
\mathbb{P}_{\theta_0}\left(\left|\hat{\theta}_{\varepsilon,n} - \theta_0\right| \geq \eta\right) \leq \quad & \mathbb{P}_{\theta_0}\left(U(\hat{\theta}_{\varepsilon,n},\theta_0) > U(\theta_0,\theta_0) + \eta^{'}\right) \\
& \leq \mathbb{P}_{\theta_0}\left(\left|U(\hat{\theta}_{\varepsilon,n},\theta_0) - \bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n})\right| \geq \frac{\eta^{'}}{3}\right) + \mathbb{P}_{\theta_0}\left(\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) - \bar{U}_{\varepsilon,n}(\theta_0) \geq \frac{\eta^{'}}{3}\right) \\
& + \mathbb{P}_{\theta_0}\left(\left|\bar{U}_{\varepsilon,n}(\theta_0) - U(\theta,\theta_0)\right| \geq \frac{\eta^{'}}{3}\right).
\end{aligned}$$

Since $\left|U(\hat{\theta}_{\varepsilon,n},\theta_0) - \bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n})\right| \geq \dfrac{\eta^{'}}{3}$ it implies that, $\sup\limits_{\theta\in\bar{\Theta}}\left|U(\theta,\theta_0) - \bar{U}_{\varepsilon,n}(\theta)\right| \geq \dfrac{\eta^{'}}{3}$

and we have $\left|U(\theta,\theta_0) - \bar{U}_{\varepsilon,n}(\theta_0)\right| \geq \dfrac{\eta^{'}}{3}$ thus, $\sup\limits_{\theta\in\bar{\Theta}}\left|U(\theta,\theta_0) - \bar{U}_{\varepsilon,n}(\theta)\right| \geq \dfrac{\eta^{'}}{3}$.

Thus,

$$\mathbb{P}_{\theta_0}(\left|\hat{\theta}_{\varepsilon,n} - \theta_0\right| \geq \eta) \leq 2\mathbb{P}_{\theta_0}\left(\sup_{\theta\in\bar{\Theta}}\left|\bar{U}_{\varepsilon,n}(\theta) - U(\theta,\theta_0)\right| \geq \frac{\eta^{'}}{3}\right) + \mathbb{P}_{\theta_0}\left(\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) - \bar{U}_{\varepsilon,n}(\theta_0) \geq \frac{\eta^{'}}{3}\right).$$

Using Eq. (4.8) and the *Lemma* 4.1, we have

$$\mathbb{P}_{\theta_0}\left(\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) - \bar{U}_{\varepsilon,n}(\theta_0) \geq \frac{\eta^{'}}{3}\right) \longrightarrow 0$$

and

$$\mathbb{P}_{\theta_0}\left(\sup_{\theta\in\bar{\Theta}}\left|\bar{U}_{\varepsilon,n}(\theta_0)-U(\theta,\theta_0)\right|\geq\frac{\eta^{'}}{3}\right)\longrightarrow 0,$$

as $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$. This completes the first statement of Theorem 4.1.

Now let prove the normality property of the MCE estimator. We consider the following arguments:

$S_{\varepsilon,n}=\varepsilon^{-1}(\hat{\theta}_{\varepsilon,n}-\theta_0)$, $C_{\varepsilon,n}(\theta_0)=\varepsilon^2\left(\dfrac{\partial^2}{\partial\theta_i\partial\theta_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leq i,j\leq 4}$,

$\Lambda_{\varepsilon,n}=-\varepsilon\left(\dfrac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\theta_0)\right)_{1\leq i\leq 4}$ and $D_{\varepsilon,n}=\displaystyle\int_0^1 C_{\varepsilon,n}\left(\theta_0+u(\hat{\theta}_{\varepsilon,n}-\theta_0)\right)du.$

A first order Taylor expansion with integral remainder for $\left(\dfrac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\theta)\right)$ at point θ_0, if $\hat{\theta}_{\varepsilon,n}\in B(\theta_0;\rho)=\{\theta\in\Theta:|\theta-\theta_0|\leq\rho\}$ yields,

$$\left(\frac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n})\right)-\left(\frac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\theta_0)\right)=\sum_{j=1}^{4}\left(\int_0^1\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}U_{\varepsilon,n}(\theta_0+t(\hat{\theta}_{\varepsilon,n}-\theta_0))\right)dt\right)_{i,j}(\hat{\theta}_{\varepsilon,n}-\theta_0).$$

Since $\hat{\theta}_{\varepsilon,n}$ is a minimum, then it is a root of the function $\dfrac{\partial}{\partial\theta}U_{\varepsilon,n}$ and multiplying by ε, we obtain

$$-\varepsilon\left(\frac{\partial U_{\varepsilon,n}}{\partial\theta_i}(\theta_0)\right)=\sum_{j=1}^{4}\left(\int_0^1\varepsilon^2\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}U_{\varepsilon,n}(\theta_0+t(\hat{\theta}_{\varepsilon,n}-\theta_0))\right)dt\right)_{i,j}\times\varepsilon^{-1}(\hat{\theta}_{\varepsilon,n}-\theta_0),$$

which can be written as

$$-\varepsilon\left(\frac{\partial}{\partial\theta}U_{\varepsilon,n}(\theta_0)\right)=\int_0^1\varepsilon^2\left(\frac{\partial^2}{\partial\theta^2}U_{\varepsilon,n}(\theta_0+t(\hat{\theta}_{\varepsilon,n}-\theta_0))\right)dt\times\varepsilon^{-1}(\hat{\theta}_{\varepsilon,n}-\theta_0).$$

Using previous arguments, we may get

$$D_{\varepsilon,n}S_{\varepsilon,n}=\Lambda_{\varepsilon,n}. \tag{4.9}$$

Now the consistency of $\hat{\theta}_{\varepsilon,n}$, leads to for sufficiently small $\rho>0$, we have

$$\left(\hat{\theta}_{\varepsilon,n}\in B(\theta_0;\rho)\right)\subset\left(\hat{\theta}_{\varepsilon,n}\in\Theta\right),\ \text{for}\rho>0,$$

then

$$\mathbb{P}_{\theta_0}(\hat{\theta}_{\varepsilon,n}\in\Theta)\geq\mathbb{P}_{\theta_0}(\hat{\theta}\in B(\theta_0;\rho))\longrightarrow 1,$$

as $\varepsilon\longrightarrow 0$ and $n\longrightarrow\infty$. Therefore

$$\chi_{\left(\hat{\theta}_{\varepsilon,n}\in\Theta\cap B(\theta_0;\rho)\right)}\longrightarrow 1$$

in $\mathbb{P}_{\theta_0}$ probability as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$.

Moreover, there exists a sequence $(B(\theta_0; \rho_{\varepsilon,n}))_{\varepsilon,n}$ such that,

$$\rho_{\varepsilon,n} \longrightarrow 0 \quad \text{and} \quad \mathbb{P}_{\theta_0}(\hat{\theta}_{\varepsilon,n} \in B(\theta_0; \rho_{\varepsilon,n})) \longrightarrow 1,$$

as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. It follows that,

$$\chi_{\left(\hat{\theta}_{\varepsilon,n} \in \Theta \cap B(\theta_0; \rho_{\varepsilon,n})\right)} \longrightarrow 1$$

in $\mathbb{P}_{\theta_0}$ probability as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Thus

$$|D_{\varepsilon,n} - C_{\varepsilon,n}(\theta_0)|\, \chi_{\left(\hat{\theta}_{\varepsilon,n} \in \Theta \cap B(\theta_0; \rho_{\varepsilon,n})\right)} \leq \sup_{\theta \in B(\theta_0; \rho_{\varepsilon,n})} |C_{\varepsilon,n}(\theta) - C_{\varepsilon,n}(\theta_0)|\,.$$

$C_{\varepsilon,n}$ is continuous with respect to θ since ξ and $\mathbf{\Sigma}$ are $\mathcal{C}^2$ on θ, then,

$$\sup_{\theta \in B(\theta_0; \rho_{\varepsilon,n})} |C_{\varepsilon,n}(\theta) - C_{\varepsilon,n}(\theta_0)| \longrightarrow 0.$$

Therefore, $D_{\varepsilon,n} \longrightarrow C_{\varepsilon,n}(\theta_0)$ in $\mathbb{P}_{\theta_0}$-probability as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Hence Lemma 4.3, ensure that

$$D_{\varepsilon,n} \longrightarrow 2\mathbf{I}(\theta_0), \tag{4.10}$$

in $\mathbb{P}_{\theta_0}$-probability as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Now we define

$$\eta_{\varepsilon,n} = D_{\varepsilon,n} - 2\mathbf{I}(\theta_0). \tag{4.11}$$

The Eq. (4.10) leads to $\eta_{\varepsilon,n} \longrightarrow 0$, as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Combining Eq. (4.9) with Eq. (4.11), we obtain

$$\left(\eta_{\varepsilon,n} + 2\mathbf{I}\left(\theta_0\right)\right) S_{\varepsilon,n} = \Lambda_{\varepsilon,n}.$$

As $S_{\varepsilon,n}$ is bounded in $\mathbb{P}_{\theta_0}$ by Guy et al. [62, Proposition 4.2] and Slutsky's theorem Gut [61, Theorem 11.4], gives

$$\eta_{\varepsilon,n} S_{\varepsilon,n} \longrightarrow 0$$

in $\mathbb{P}_{\theta_0}$. It follows, using Lemma 4.2, that

$$2\mathbf{I}(\theta_0) S_{\varepsilon,n} \longrightarrow \mathcal{N}(0, 4\mathbf{I}(\theta_0)), \tag{4.12}$$

in distribution under $\mathbb{P}_{\theta_0}$ as $\varepsilon \longrightarrow 0$ and $n \longrightarrow \infty$. Finally, since $2\mathbf{I}(\theta_0)$ is not singular, from Eq. (4.12), we get,

$$\varepsilon^{-1}(\hat{\theta}_{\varepsilon,n} - \theta_0) \longrightarrow \mathcal{N}(0, \mathbf{I}^{-1}(\theta_0)), \tag{4.13}$$

which completes the proof of the second statement in Theorem 4.1. □

4.5 Conclusion

This chapter has been concerned with parameter estimation of dynamic transmission of HIV virus in a closed heterosexual population. We have inferred the main parameter of interest for our model using a minimum contrast estimator, for which the consistency and normal asymptotic properties, using rigorous mathematical tools, have been established.

The inference, established in this chapter, is thus likely to be applicable to a rather wider range of models, in which population is partitioned into more than two sub populations with general mechanism of infection, more than what we have considered here. As well as, the approach established here can be applied to various heterogeneous stochastic epidemic model with jumps or driven by Levy processes. Application and simulation for this chapter are remained to the sixth chapter.

Chapter 5

Bayesian inference of the parameters of HIV diffusion

5.1 Introduction

Over the last years there has been an increasable research for performing Bayesian inferential methods of the spread of infectious diseases among populations using stochastic epidemic models. On one hand, stochastic inference for epidemic models is generally complicated, this is due to the highly dependent in data which is often discretely observed while the underlying true process is time continuous. On the other hand, discretely observed diffusions are discrete-time Markov processes, and thus, their likelihood depends on the transition probabilities densities of the diffusion. However, these densities are unavailable in simple closed form (see, El Maroufy [35]). In the case of continuously observed data, parameter estimates can be obtained for complete data (see, Becker [17]), and in the case of incomplete data, estimates have also been developed some novel methods (see, Andersson and Britton [4] and references therein). Recently, the Bayesian inference of the SIR model has been treated by several authors using MCMC methods (see, Britton and O'Neill [22] and Qaffou et al. [100] and references therein).

This chapter is concerned with performing Bayesian statistical inference for parameters of stochastic epidemic controlled by the stochastic differential equation Eq. (2.4). In chapter 4, we used a minimum contrast estimator to estimate the parameters of interest. However, in this chapter, we present our second contribution to study the estimation of interest model parameters . Thus, we use a Bayesian approach to do this task, especially the data augmentation method, which is performed by the application of MCMC techniques (see, Eraker [41], El Maroufy et al. [36], and Qaffou et al. [100]).

5.2 Bayesian inference for non linear diffusion model

Let consider the inference for an Itô diffusion process of type,

$$d\mathbf{X}_t = \xi(\mathbf{X}_t, \theta)dt + \boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{X}_t, \theta)dW(t), \quad t \geq 0, \tag{5.1}$$

where ξ and Σ are defined in Eq. (3.8) and Eq. (3.9), respectively.

We assume that the process $\mathbf{X} = (X_t,\ t \geq 0)$ will be observed at a finite integer of times. The purpose is to infer for the unknown parameter vector $\theta\,(\alpha, \beta, \gamma, \delta)$ on the basis of partial and discrete observations $\mathbf{X}^{obs} = \{X_{t_k}, k = 0, \ldots, M\}$ of the process $\mathbf{X}$. In this chapter, we assume the equi-probability of the recruitment of new female and male susceptible, which means that $\alpha = \dfrac{1}{2}$. To perform the inference on parameter θ, the first idea behind is to attempt to approximate the true transition density p_θ of the diffusion process using a numerical scheme, such as the Euler–Maruyama scheme. However, this is eligible only if inter-observation times of the observed data $\mathbf{X}^{obs}$ are small enough. Such a requirement is not generally satisfied in our case considering the observation of HIV though.

To overcome this limitation, we will augment the observed data $\mathbf{X}^{obs}$, using the Eraker approach developed by Eraker [41]. The method consists of imputing intermediate points between each pair of observations. Furthermore, to infer θ, we will employ a MCMC approach – a Gibbs sampler in particular – to construct a Markov chain $\{\theta^{(i)}, \mathbf{X}^{imp}\}_{i=1,\ldots,L}$ of length L whose elements are samples form joints posterior density $\pi(\theta, \mathbf{X}^{imp}|\mathbf{X}^{obs})$ of parameter θ and imputed data $\mathbf{X}^{imp}$ conditional on observations $\mathbf{X}^{obs}$. The Markov chain $\{\theta^{(i)}\}_{i=1,\ldots,L}$ is regarded as a draw from the marginal density $\pi(\theta|\mathbf{X}^{obs})$.

5.2.1 Bayesian Data Augmentation[1]

The objective is to obtain a sequence of Monte Carlo samples $\{\theta^{(i)}\}_{i=1}^{L}$, which is a sample from the marginal posterior density, since $\{\theta^{(i)}\}_{i=1}^{L}$ is implicitly a sample from the marginal posterior $\pi(\theta|\hat{X}^{obs})$. Let π generically denote all posterior densities. Analogously, let p generically denote all prior densities and q all proposal densities. To approximate the posterior density of parameter θ, based on observations $\mathbf{X}^{obs}$ of a diffusion process we have,

$$\pi(\theta|X_{t_1}, \ldots, X_{t_M}) \propto \pi(X_{t_1}, \ldots, X_{t_M}|\theta)p(\theta),$$

Since diffusion processes possess the Markov property, such complete observations divide a sample path into segments that are mutually independent conditioned on θ. The likelihood of θ

[1] The reader is referred to Fuchs [46, Chap. 7], for a detailed description of this method.

factorize as

$$\pi(X_{t_1},\ldots,X_{t_M}|\theta)=\pi(X_{t_1}|\theta)\prod_{i=1}^{M}\pi(X_{t_i}|X_{t_{i-1}},\theta). \tag{5.2}$$

The data augmentation method, consists of imputing $m-1$ latent data points between each pair of observations $\mathbf{X}^{obs}$. To ensure that discretization bias is arbitrary small we put $\Delta t=\frac{1}{m}$ and $N=mM$ for a chosen positive integer m. Therefore $[0,T]$ is divided into $N+1$ equidistant points $t_0=0<t_1<\ldots<t_m<t_{m+1}<\ldots<t_N=T$, then the diffusion process is in state X_{t_k} at time t_k which is only known on times t_j when j is an integer multiple of m and all points $X_{t_k}, k\neq j$ are treated as missing data.

Let denote by $\hat{X}$ the $4\times(N+1)$ matrix obtained by stacking all elements of augmented data (observed and missing), that is

$$\hat{X}=\begin{pmatrix}\hat{x_1}_{t_0} & \hat{x_1}_{t_1} & \ldots & \hat{x_1}_{t_m} & \hat{x_1}_{t_m+1} & \ldots & \hat{x_1}_{t_N}\\ \hat{x_2}_{t_0} & \hat{x_2}_{t_1} & \ldots & \hat{x_2}_{t_m} & \hat{x_2}_{t_m+1} & \ldots & \hat{x_2}_{t_N}\\ \hat{y_1}_{t_0} & \hat{y_1}_{t_1} & \ldots & \hat{y_1}_{t_m} & \hat{y_1}_{t_m+1} & \ldots & \hat{y_1}_{t_N}\\ \hat{y_2}_{t_0} & \hat{y_2}_{t_1} & \ldots & \hat{y_2}_{t_m} & \hat{y_2}_{t_m+1} & \ldots & \hat{y_2}_{t_N}\end{pmatrix}.$$

Let $\hat{X}_i$ denote the i^{th} column of $\hat{X}$ (if i is a multiple of m, $\hat{X}_i$ is an observed data).

Conditioning on the first observation, the joint posterior density is given by:

$$\pi(\hat{X},\theta)\propto\prod_{i=1}^{N}\pi(\hat{X}_i|\hat{X}_{i-1},\theta)p(\theta), \tag{5.3}$$

where

$$\pi(\hat{X}_i|\hat{X}_{i-1},\theta)=|\Sigma_{i-1}^{-1}|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\hat{\mathbb{P}}_i'\left(\Sigma_{i-1}^{-1}\Delta t\right)\hat{\mathbb{P}}_i\right\} \tag{5.4}$$

with $\hat{\mathbb{P}}_i=\hat{X}_i-\hat{X}_{i-1}-\xi(\hat{X}_{i-1},\theta)\Delta t$.

T construct the Markov chain $\{\theta^{(i)},\mathbf{X}^{imp}\}_{i=1,\ldots,L}$ the following two steps, which represent the Gibbs sampler given by the Algorithm 1, as described in page 25, are alternately executed:

$$\begin{aligned}&\text{Path Update:}\quad \text{Draw } \mathbf{X}^{imp(i)}\propto\pi\left(\mathbf{X}^{imp(i)}|\theta^{(i-1)},\mathbf{X}^{obs}\right).\\ &\text{Parameter Update:}\quad \text{Draw } \theta^{(i)}\propto\pi\left(\theta^{(i)}|\mathbf{X}^{imp(i)},\mathbf{X}^{obs}\right).\end{aligned} \tag{5.5}$$

5.2.2 Path update

The first step in Gibbs sampler is the update of the whole path. Because the number of unobservable (missing data and parameter) is large, it is not possible to obtain independent samples of these quantities directly from (5.3). We use Gibbs sampler with a block strategy instead of single site to overcome the poor mixing due to high correlation amongst the latent data(see,

Elerian et al. [39]), which is based on sampling a block of elements at same time from the posterior instead of whole path and keeping the others constant as conditioning elements. The latent data are updated in blocks of size m. Consider times t_j and t_{j^+}, where j is an integer multiple of m and $j^+ = j + m$, the corresponding observation $\hat{X}^j$ and $\hat{X}^{j^+}$ which treated as fixed (observed points), the full conditional for the latent path in (t_j, t_{j^+}) is

$$\pi(\hat{X}^{j+1}, \ldots, \hat{X}^{j^+-1} | \hat{X}^j, \hat{X}^{j^+}, \theta) = \prod_{i=j}^{j^+-1} \pi(\hat{X}^{i+1} | \hat{X}^i, \theta). \tag{5.6}$$

The first step in Gibbs sampler is to appropriately perform the path update. Thus, direct sampling from the posterior distribution of the latent data given the observed data and parameter is not possible in our case, hence we use a M-H algorithm Hastings [63] and Robert and Casella [104], as given by Algorithm 2 on page 26, for the general implementation of this step.

At each iteration, as a first step in path update, we choose a block of size m that is a time interval $\left[t_j, t_{j^+}\right]$ in which the path will be updated. Having decided about the block update strategy, simulate each column of the block $\hat{X}^\star_{(t_j,t_{j^+})} = \left\{\hat{X}^\star_{t_j}, \ldots, \hat{X}^\star_{t_{j^+}}\right\}$ using a Gaussian proposal with mean $\hat{X}^\star_k + \mu(\hat{X}^\star_k, \theta)\Delta t_k$ and variance matrix $\Sigma(\hat{X}^\star_k, \theta)\Delta t_k$;

$$q(\hat{X}^\star_{k+1} | \hat{X}^\star_k, \theta) = \mathcal{N}\left(\hat{X}^\star_k + \mu(\hat{X}^\star_k, \theta)\Delta t_k, \Sigma(\hat{X}^\star_k, \theta)\Delta t_k\right), \tag{5.7}$$

for $k = t_{j+1}, \ldots, t_{j^+-1}$, with $\hat{X}^\star_{t_j} = \hat{X}_{t_j}$ and $\hat{X}^\star_{t_{j^+}} = \hat{X}_{t_{j^+}}$ are treated constants (observed data). The acceptance probability $\alpha(\hat{X}^\star_{(t_j,t_{j^+})}, \hat{X}_{(t_j,t_{j^+})})$ to choose $\hat{X}^\star_{(t_j,t_{j^+})}$ instead of $\hat{X}_{(t_j,t_{j^+})}$ is given by:

$$\alpha\left(\hat{X}^\star_{(t_j,t_{j^+})}, \hat{X}_{(t_j,t_{j^+})}\right) = 1 \wedge \frac{\pi\left(\hat{X}^\star_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})} | \hat{X}^{obs}, \theta\right) q\left(\hat{X}_{(t_j,t_{j^+})} | \hat{X}^\star_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})}, \hat{X}^{obs}, \theta\right)}{\pi\left(\hat{X}_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})} | \hat{X}^{obs}, \theta\right) q\left(\hat{X}^\star_{(t_j,t_{j^+})} | \hat{X}_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})}, \hat{X}^{obs}, \theta\right)}, \tag{5.8}$$

where $\hat{X}_{-(t_j,t_{j^+})} = \hat{X} \setminus \hat{X}_{(t_j,t_{j^+})}$ is the the complement of $\hat{X}_{(t_j,t_{j^+})}$. The Markov property leads

$$\begin{aligned}
\frac{\pi\left(\hat{X}^\star_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})} | \hat{X}^{obs}, \theta\right)}{\pi\left(\hat{X}_{(t_j,t_{j^+})}, \hat{X}_{-(t_j,t_{j^+})} | \hat{X}^{obs}, \theta\right)} &= \prod_{k=t_j+1}^{t_{j^+}-1} \frac{\pi\left(\hat{X}^\star_{k+1} | \hat{X}^\star_k, \theta\right)}{\pi\left(\hat{X}_{k+1} | \hat{X}_k, \theta\right)} \\
&= \prod_{k=t_j+1}^{t_{j^+}-1} \frac{p_\theta\left(\Delta_{t_k}, \hat{X}^\star_{k+1}, \hat{X}^\star_k\right)}{p_\theta\left(\Delta_{t_k}, \hat{X}_{k+1} | \hat{X}_k\right)},
\end{aligned}$$

the time step Δt_k is now supposed to be small enough such that the Euler scheme is adopted, then p_θ may be replaced by

$$\pi\left(\hat{X}_{k+1}|\hat{X}_k,\theta\right)=\mathcal{N}\left(\hat{X}_k+\mu(\hat{X}_k,\theta)\Delta t_k,\Sigma(\hat{X}_k,\theta)\Delta t_k\right).$$

The proposal density for $\hat{X}^\star_{(t_j,t_{j^+})}$, using Euler scheme for $k=j,\dots,j^+-2$
$\hat{X}_{k+1}\sim\mathcal{N}\left(\hat{X}_k+\mu(\hat{X}_k,\theta)\Delta t_k,\Sigma(\hat{X}_k,\theta)\Delta t_k\right)$ with $\hat{X}_j=X_j$, becomes

$$q\left(\hat{X}^\star_{(t_j,t_{j^+})}|\hat{X}_{t_j},\hat{X}_{t_{j^+}},\theta\right)=\prod_{k=t_j+1}^{t_{j^+-2}}q\left(\hat{X}^\star_{k+1}|\hat{X}^\star_k,\theta\right)=\prod_{k=t_j+1}^{t_{j^+-2}}\pi\left(\hat{X}^\star_{k+1}|\hat{X}^\star_k,\theta\right)\tag{5.9}$$

Integrating Eq. (5.9) into Eq. (5.8), the acceptance probability to choose the proposal $\hat{X}^\star_{(t_j,t_{j^+})}$ is reduced to the following expression

$$\begin{aligned}\alpha(\hat{X}^\star_{(t_j,t_{j^+})},\hat{X}_{(t_j,t_{j^+})})=&1\wedge\left(\prod_{k=t_{j+1}}^{t_{j^+-1}}\frac{\pi\left(\hat{X}^\star_{k+1}|\hat{X}^\star_K,\theta\right)}{\pi\left(\hat{X}_{k+1}|\hat{X}_k,\theta\right)}\right)\left(\prod_{k=t_{j+1}}^{t_{j^+-2}}\frac{\pi\left(\hat{X}_{k+1}|\hat{X}_K,\theta\right)}{\pi\left(\hat{X}^\star_{k+1}|\hat{X}^\star_k,\theta\right)}\right)\\=&1\wedge\frac{\pi\left(\hat{X}^\star_{t_{j^+}}|\hat{X}^\star_{t_{j^+-1}},\theta\right)}{\pi\left(\hat{X}_{t_{j^+}}|\hat{X}_{t_{j^+-1}},\theta\right)}.\end{aligned}\tag{5.10}$$

This situation is summarized in the following algorithm:

Algorithm 3 Path update

1: Given an initial path $\hat{X}$.
2: Choose j multiple of m, set $j^+=j+m$, the time interval $\left[t_j,t_{j^+}\right]$ of size m, whose the path will be updated in its interior
3: put $\hat{X}^\star_{t_j}=\hat{X}_{t_j}$ and $\hat{X}^\star_{t_{j^+}}=\hat{X}_{t_{j^+}}$
4: **for** $k=t_{j+1},\dots,t_{j^+-1}$ **do**
5: sample $q(\hat{X}^\star_{k+1}|\hat{X}^\star_k,\theta)$ according to Eq. (5.7)
6: Accept $\hat{X}^\star_{(t_j,t_{j^+})}$ with probability α given by Eq. (5.10)
7: **end for**

5.2.3 Parameter update and posterior distribution

The last step in Gibbs sampler is the update of parameter $\theta=(\beta,\delta,\gamma)$ conditioning on its current state and the augmented data. Due to the form of likelihood function derived in (5.2), a family of independent gamma distributions is seen as a natural set of conjugate priors in the context of epidemic model, where the parameters are positive (see Demiris and O'Neill [32]). This choice of prior distributions is convenient in terms of Bayesian inference due to conjugacy (see,

O'Neill and Roberts [91]). Furthermore, the flexibility of the gamma distribution means that it is frequently used in practice as a prior distribution for rate parameters in epidemic models (see, Cauchemez et al. [26] and Qaffou et al. [100]). When this type of priors is considered, we obtain the Generalized Inverse Gaussian GIG distribution as a posterior distribution. In the following proposition, we prove this claim for which the proof is left in details in Appendix C.

Proposition 5.1. *If β, δ and γ follow independent gamma distributions: $\pi(\beta) \propto \Gamma(m_\beta, \lambda_\beta)$, $\pi(\delta) \propto \Gamma(m_\delta, \lambda_\delta)$ and $\pi(\gamma) \propto \Gamma(m_\gamma, \lambda_\gamma)$, then*

$$\pi(\beta|\hat{X}_i, \delta, \gamma) \propto \beta^{m_\beta - N - 1} \exp\left\{-\frac{1}{2}\left[\frac{M_\beta}{\beta} + \left(M'_\beta + 2\lambda_\beta\right)\beta\right]\right\}; \tag{5.11}$$

$$\pi(\delta|\hat{X}_i, \beta, \gamma) \propto \delta^{m_\delta - N - 1} \exp\left\{-\frac{1}{2}\left[\frac{M_\delta}{\delta} + \left(M'_\delta + 2\lambda_\delta\right)\delta\right]\right\}; \tag{5.12}$$

and

$$\pi(\gamma|\hat{X}_i, \delta, \beta) \propto \gamma^{m_\gamma - N - 1} \exp\left\{-\frac{1}{2}\left[\frac{M_\gamma}{\gamma} + \left(M'_\gamma + 2\lambda_\gamma\right)\gamma\right]\right\}, \tag{5.13}$$

where M_k, M'_k for $k = \{\beta, \delta, \gamma\}$ are presented in Appendix C.

5.2.4 Implementation and simulation

To implement the described Bayesian estimation presented in this chapter, we summarize the procedures in the following main algorithm:

Algorithm 4 Main Algorithm

1: Initialize all unknown: $\theta = (\beta, \delta, \gamma)$ and the path $\hat{\mathbf{X}}$ using linear interpolation between observed values of $\mathbf{X}^{obs}$.
2: **repeat**
3: Update the path $\hat{\mathbf{X}}^\star$ using Algorithm 3 and Eq. (5.10),
4: Update β^h using (5.11),
5: Update δ^h using (5.12),
6: Update γ^h using (5.13),
7: **until** the desired number of samples for θ is reached.

5.3 Conclusion

This chapter has been concerned with Bayesian estimation of parameters for HIV/AIDS dynamic in a closed heterosexual population. A stochastic diffusion approximation, which developed in Chapter 3 is adopted. We are essentially concerned with the Bayesian analysis of nonlinear, discretely observed stochastic diffusion. The MCMC methods have been used to infer the parameters of interest. We have proposed a Gamma distribution as prior density for parameters,

which seems to be natural for rate parameters in epidemic models. We prove that the posterior distributions of model parameters converge to a limit distribution, which are distributed according to a Generalized Inverse Gaussian GIG density.

The inference, established here, can be likely applicable to some other compartmental models. This work represent an alternative method to overcome discretely observed process considered in Abou-Bakre and El Maroufy [1].

Chapter 6

Numerical simulations with real application to Morocco's case

In this chapter, we illustrate the methods and results presented in previous chapters. Using simulated data in a first step to validate the modeling procedure and applying these methods to real data from Morocco case.

6.1 Simulation using MCE

6.1.1 Simulation

To illustrate the approach presented in Chapter 4, we consider discrete observations $\{\mathbf{X}_k, k = 1, \ldots, n\}$ coming from simulation of the original Markov chain $(\mathbf{X}_t, t \geq 0)$, with transitions given by Eq. (3.1). We use the exact Gillespie algorithm developed by Gibson and Bruck [52]. Instead of describing a general Gillespie algorithm, we only describe it as it applies to the problem of parameter estimation in the given framework.

In our case, the implementation of the algorithm is based on two main steps. The first is the draw of the waiting time τ until the occurrence of the next event, which is exponentially distributed $\exp(\lambda)$ with rate $\lambda = \sum_{k=1}^{11} \lambda_k$. We have the rates $\lambda_1 = \alpha(\mu N + \delta(Z+1))$, $\lambda_2 = (1-\alpha)(\mu N + \delta(Z+1))$ for birth of a new susceptible, $\lambda_3 = \beta \frac{I_M S_F}{S_M + I_M}$, $\lambda_4 = \beta \frac{I_F S_M}{S_F + I_F}$ infection of a susceptible, $\lambda_5 = \mu S_F$, $\lambda_6 = \mu S_M$ for death of a susceptible, $\lambda_7 = \mu I_M$, $\lambda_8 = \mu I_F$ for death of an infected, $\lambda_9 = \gamma I_M$, $\lambda_{10} = \gamma I_F$ recovery of an infected and $\lambda_{11} = \mu Z$ death of a recovered. In the second step, events are randomly selected according to the probabilities $p_k = \frac{\lambda_k}{\lambda}$. The Algorithm 5 is implemented with true values of parameters, until a stopping criterion is fulfilled.

We run for 1000 iterations, the non-linear simulation using 500 observations, for three cases $N = 1000$, $N = 10000$ and $N = 100000$ where N is the initial total population size. For

Algorithm 5 Gillespie algorithm

1: At $t = 0$, initialize the population sizes $(S_f(0), I_f(0), S_m(0), I_m(0)) = (s_f, i_f, s_m, i_m)$.
2: **repeat**
3: Calculate $\lambda = \sum_{k=1}^{11} \lambda_k$.
4: Choose τ from the exponential distribution $\exp(\lambda)$.
5: Choose the event $k = 1, ..., 11$ with probability $p_k = \frac{\lambda_k}{\lambda}$.
6: Update time $t \leftarrow t + \tau$
7: Update and Store $(S_f(t), I_f(t), S_m(t), I_m(t))$ according to the event k chosen in step 5.
8: **until** $t \leq T$

each case, two situations are considered depending on the value of the reproduction number ($R_0 = \frac{1}{\mu + \gamma}$), $R_0 \leq 1$ and $R_0 > 1$. Considering Table 6.1 which summarize the estimates of minimum contrast estimator and standard deviation of parameter $\theta = (\mu, \beta, \delta, \gamma)$. We see that the estimations are close to the true values and the standard deviation decreases as the population size increases. The histograms in Figures 6.1 and 6.2 reveal the asymptotic normality property of the estimators.

TABLE 6.1: Minimum contrast estimation and standard deviation of parameters μ, β, δ and γ in three cases: $N = 1000$, $N = 10000$ and $N = 100000$.

	$N = 1000$		$N = 10000$		$N = 100000$	
true values	MCE	SD	MCE	SD	MCE	SD
$\mu = 0.02$	$\hat{\mu} = 0.0201$	0.0034	$\hat{\mu} = 0.0199$	0.0033	$\hat{\mu} = 0.0200$	0.0033
$\beta = 0.05$	$\hat{\beta} = 0.0481$	0.0118	$\hat{\beta} = 0.0498$	0.0094	$\hat{\beta} = 0.0498$	0.0052
$\delta = 0.10$	$\hat{\delta} = 0.1048$	0.0361	$\hat{\delta} = 0.1005$	0.0094	$\hat{\delta} = 0.1001$	0.0053
$\gamma = 0.08$	$\hat{\gamma} = 0.0866$	0.0308	$\hat{\gamma} = 0.0797$	0.0167	$\hat{\gamma} = 0.0798$	0.0073
$\mu = 0.02$	$\hat{\mu} = 0.0190$	0.0049	$\hat{\mu} = 0.0201$	0.0033	$\hat{\mu} = 0.0200$	0.0033
$\beta = 0.50$	$\hat{\beta} = 0.5217$	0.1074	$\hat{\beta} = 0.4983$	0.0457	$\hat{\beta} = 0.4991$	0.0133
$\delta = 0.10$	$\hat{\delta} = 0.1412$	0.0975	$\hat{\delta} = 0.1005$	0.0084	$\hat{\delta} = 0.0999$	0.0049
$\gamma = 0.08$	$\hat{\gamma} = 0.0746$	0.0367	$\hat{\gamma} = 0.0806$	0.0244	$\hat{\gamma} = 0.0794$	0.0105

6.1.2 Real application: Morocco's case

We consider the case of Morocco. We construct a database in Appendix D of HIV/AIDS in dynamic Morocco which contains susceptible and infected males and females and AIDS cases. This database is obtained by combining statistics and data from Moroccan High Commission of

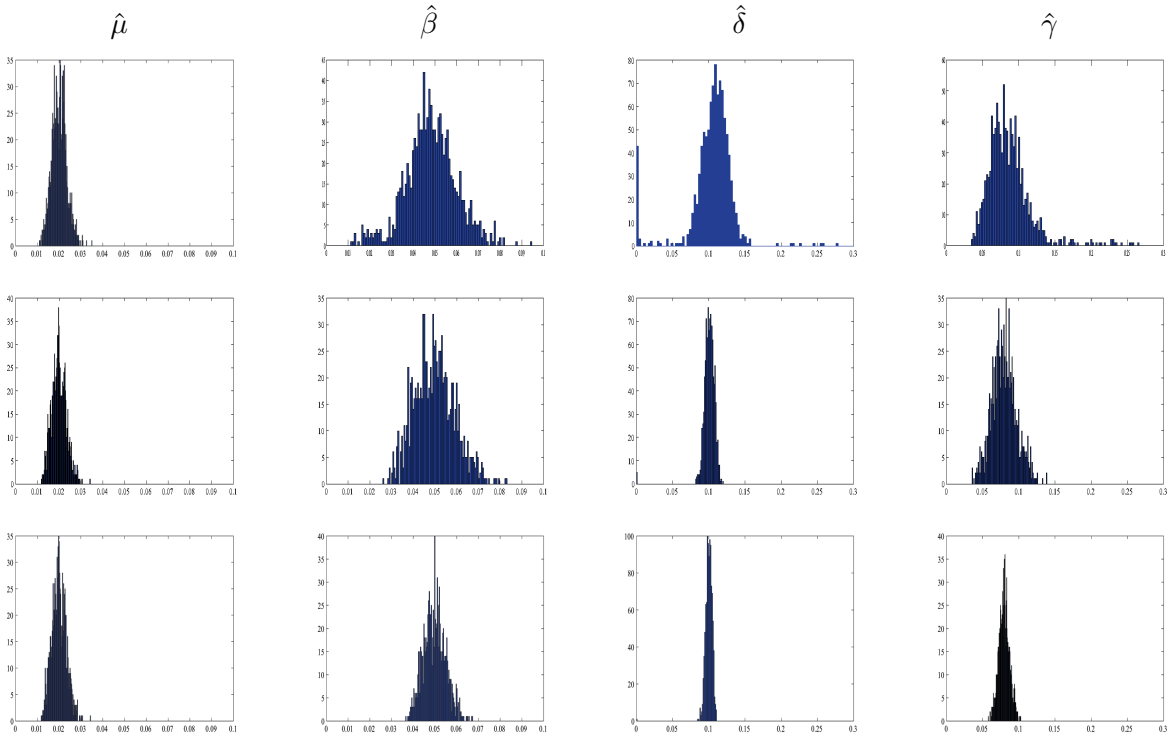

FIGURE 6.1: Frequency histograms for posterior densities estimate of all parameters estimators $\hat{\mu}, \hat{\beta}, \hat{\delta}$ and γ, for $R_0 \leq 1$ with $N = 1000$ (first row), $N = 10000$ (second row) and $N = 100000$ (third row)

Planing HCP[1], Ministry of Public Health with the support of UNAIDS [2], the database of World Bank [3] and the SPECTRUM software [4]. We restraint our study the sexually active population only.

Using this database and by minimizing the contrast function Eq. (4.3), we obtain in Table 6.2 the parameters estimations.

Considering these results, we can say that in Morocco, the mean time between HIV diagnosis and the onset of AIDS is about six years and a half $\frac{1}{\hat{\gamma}}$, we remark that this mean is improved from one and a half in 2011 (see, El Hia et al. [34]) to six and a half in 2014 which means that most diagnoses occur in early stages compared to 2011. Also, the results obtained show that there is a good improvement of the mean time from AIDS diagnosis to death for the Moroccan

[1] *Population du Maroc selon l'age et le sexe - Population marocaine selon l'age et le sexe - Données ouvertes - Maroc* available at http://www.data.gov.ma/data/fr/dataset/population-du-maroc-selon-l-age-et-le-sexe/resource/ba4d6e16-5901-42c5-9aff-1e308ca266c8

[2] *Mise en oeuvre de la déclaration politique sur le VIH/sida, Rapport National 2014.* available at http://files.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2014countries/MAR_narrative_report_2014.pdf. *Mise en œuvre de la déclaration politique sur le VIH/sida. Rapport National 2015.* available at

[3] *Maroc | Data* available at https://donnees.banquemondiale.org/pays/maroc

[4] Available and downloadable at http://www.avenirhealth.org/software-spectrum.php

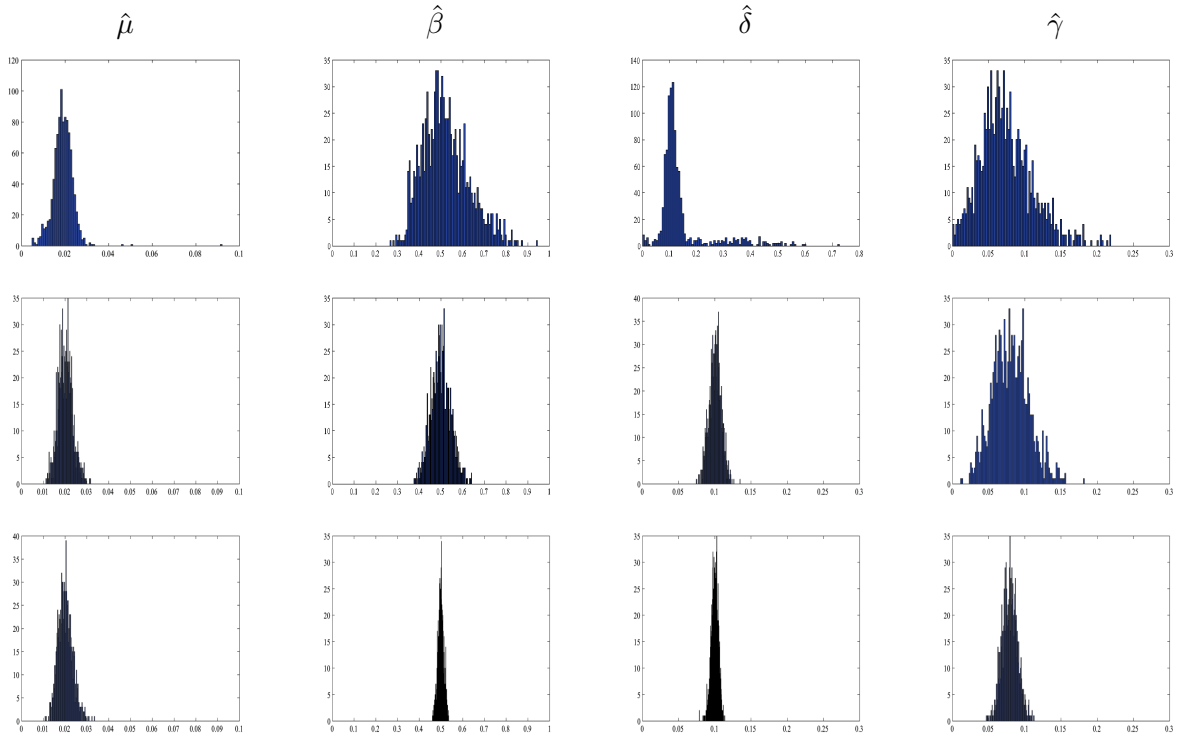

FIGURE 6.2: Frequency histograms for posterior densities estimate of all parameters estimators $\hat{\mu}, \hat{\beta}, \hat{\delta}$ and γ, for $R_0 > 1$ and $N = 1000$ (first row), $N = 10000$ (second row) and $N = 100000$ (third row)

case, which is about ten years in 2014 compared to six years in 2011. For natural death rate we obtain that $\hat{\mu} = 0.0140$ which is very close to the real rate $\mu = 0.0145$ according to World Bank data bank [14]. However, an estimate of the basic reproductive number $R_0 = \dfrac{\beta}{\mu + \gamma} \approx 1.4882$ which means that the number of secondary cases, which an infected person would produce in a completely susceptible Moroccan population, is about one and a half, that is to say, the epidemic will grow even in Morocco.

TABLE 6.2: Minimum contrast estimation and standard deviation of parameters μ, β, δ and γ

parameter	$\hat{\mu}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$
estimation	0.0140	0.2481	0.0999	0.1527
STD	0.0060	0.0612	0.0049	0.0096

6.2 Simulation using Bayesian approach

6.2.1 Simulation

The MCMC scheme is applied to the diffusion stochastic epidemic model with known parameters. The observations $\{X_k, k = 1, \ldots, n\}$ come from the simulation of the original Markov chain $(X_t, t \geq 0)$, which is defined by its transitions rates given by Eq. (3.1). We use the exact Gillespie algorithm given by Gibson and Bruck [52] with true values of parameters as given in the Table 6.1, see the appendix 5 for the detailed algorithm. For each data set the MCMC sampler is run for 20000 iterations with $m = 1$, $m = 2$ and $m = 5$. Looking at Fig. 6.3 which gives plots of the MCMC chains, Fig. 6.4 which summarizes the posterior distributions of parameters and the Table 6.3 which gives posterior mean and variance of the parameter $\theta = (\beta, \delta, \gamma)$; we observe that the estimates get closer to the true values as the number of latent data (augmented data) increases. The histograms in Fig. 6.4 reveal the convergence of the algorithm towards a limit distributions.

TABLE 6.3: Posterior mean and Posterior variance for β, δ, γ and R_0 for $m = 1$, $m = 2$ and $m = 5$.

$m = 1$								
	true value of β	$\hat{\beta}$	true value of δ	$\hat{\delta}$	true value of γ	$\hat{\gamma}$	R_0	$\hat{R}_0$
mean	0.5	0.5537	0.5	0.4996	0.5	0.4976	0.5	0.5552
variance		0.1290		0.0846		0.0188		
mean	1	1.0006	0.5	0.4753	0.5	0.4983	1	1.0277
variance		0.2541		0.1150		0.0051		
mean	2	1.8343	0.5	0.5199	0.5	0.4934	2	1.8102
variance		0.3385		0.1054		0.0046		
$m = 2$								
	true value of β	$\hat{\beta}$	true value of δ	$\hat{\delta}$	true value of γ	$\hat{\gamma}$	R_0	$\hat{R}_0$
mean	0.5	0.5040	0.5	0.5252	0.5	0.5070	0.5	0.4883
variance		0.0903		0.0885		0.0188		
mean	1	1.0378	0.5	0.5620	0.5	0.4997	1	0.9774
variance		0.2311		0.0833		0.0042		
mean	2	2.1406	0.5	0.5177	0.5	0.5101	2	2.0827
variance		0.2999		0.0598		0.0009		
$m = 5$								
	true value of β	$\hat{\beta}$	true value of δ	$\hat{\delta}$	true value of γ	$\hat{\gamma}$	R_0	$\hat{R}_0$
mean	0.5	0.5008	0.5	0.5154	0.5	0.5004	0.5	0.4930
variance		0.0282		0.0896		0.0568		
mean	1	0.9990	0.5	0.5295	0.5	0.5016	1	0.9689
variance		0.2450		0.1273		0.0051		
mean	2	2.0392	0.5	0.5001	0.5	0.5047	2	2.0293
variance		0.3517		0.1001		0.0047		

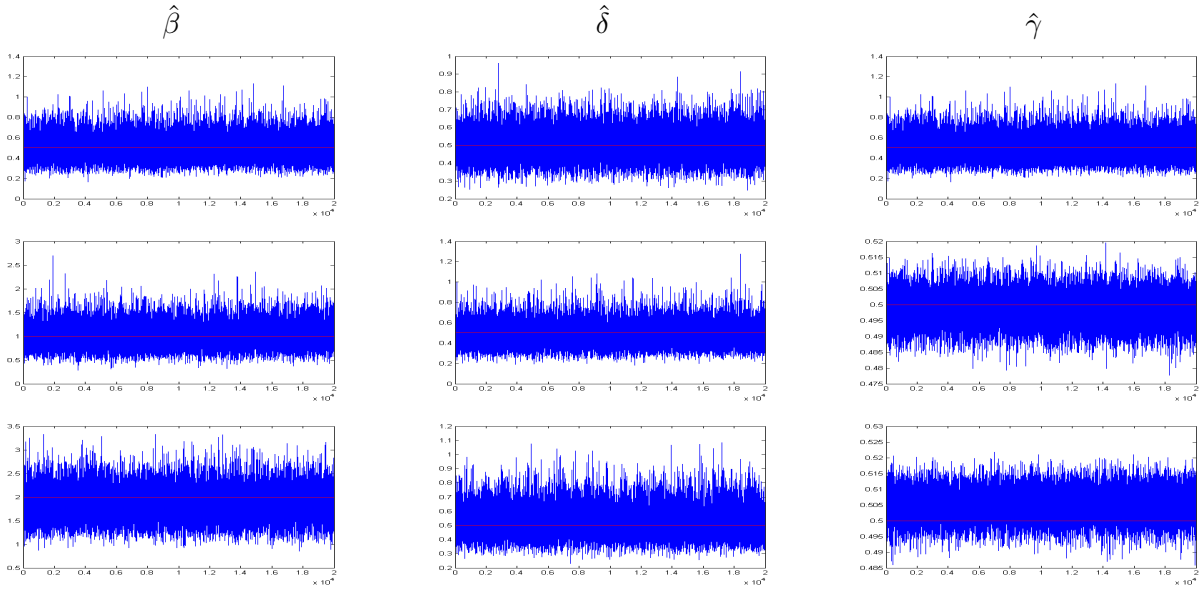

FIGURE 6.3: Trace plots of the MCMC chains for the parameters after 20000 iterations. The red lines show the true values of the parameters $\beta = \{0.5, 1, 2\}$, which correspond to $R_0 < 1, R_0 = 1, R_0 > 1$ and $\delta = \gamma = 0.5$.

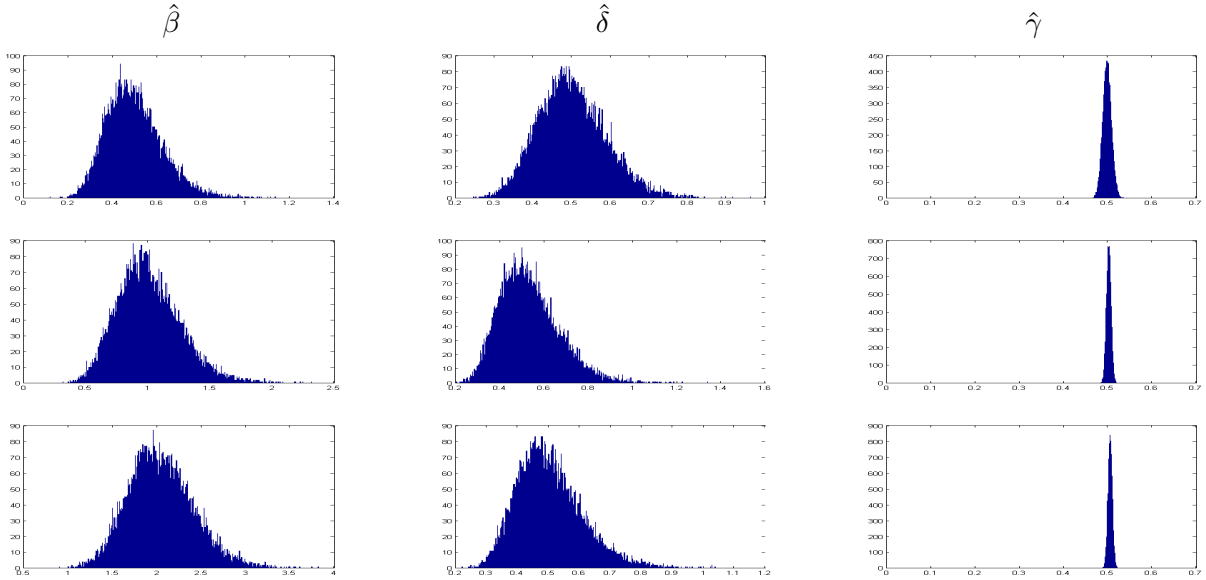

FIGURE 6.4: Posterior densities for the parameters of the model (after 20000 iterations), in three cases: $R_0 < 1$(first row), $R_0 = 1$(second row) and $R_0 > 1$(third row).

6.2.2 Real application: Morocco's case

We consider the case of Morocco; the database considered here (see, Appendix D) are data of HIV/AIDS in dynamic Morocco which contains susceptible and infected males and females and AIDS cases in sexually active population.

The results presented in Table 6.4, give an estimation of parameters. The estimated basic reproductive number $\hat{R}_0$ bigger than one, which means that the number of secondary cases which an infected person would produce in a completely susceptible Moroccan population is more than two, that is to say, the epidemic will keep growing in Morocco.

TABLE 6.4: Posteriors mean and variance for $\hat{\beta}$, $\hat{\delta}$, $\hat{\gamma}$ and $\hat{R}_0$ in cases of $m = 1$, $m = 2$, $m = 5$ and $m = 10$, for data of Morocco, with 20000 iterations.

m		$\hat{\beta}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{R}_0$
1	mean	$0,9619$	$0,2186$	$0,2288$	$2,1502$
	variance	$0,0790$	$0,0002$	$0,0048$	
2	mean	$0,9781$	$0,2145$	$0,2187$	$2,2578$
	variance	$0,0355$	$0,0003$	$0,0032$	
5	mean	$0,9795$	$0,2145$	$0,2166$	$2,2722$
	variance	$0,0352$	$0,0003$	$0,0031$	
10	mean	$0,9798$	$0,2145$	$0,2154$	$2,2791$
	variance	$0,0355$	$0,0003$	$0,0032$	

The Fig. 6.5 and Fig. 6.6 give the exact proportions of observed number the infected males and females in the sexually active Moroccan population from 1985 to 2014, and the projection of these subpopulations from 2015 up to 2020 using the estimated value of parameters given in Table 6.4. As illustrated in Fig. 6.5 and Fig. 6.6, for example, the exact proportion of the observed number of infected females in morocco during the years 2015 and 2016 are 4.757×10^{-4} and 4.833×10^{-4} respectively, whereas the predicted values corresponding to these proportions are respectively 4.703×10^{-4} and 4.842×10^{-4}. Then, we can affirm that the model presented here, is well adapted to the study of transmission of HIV in heterosexual population.

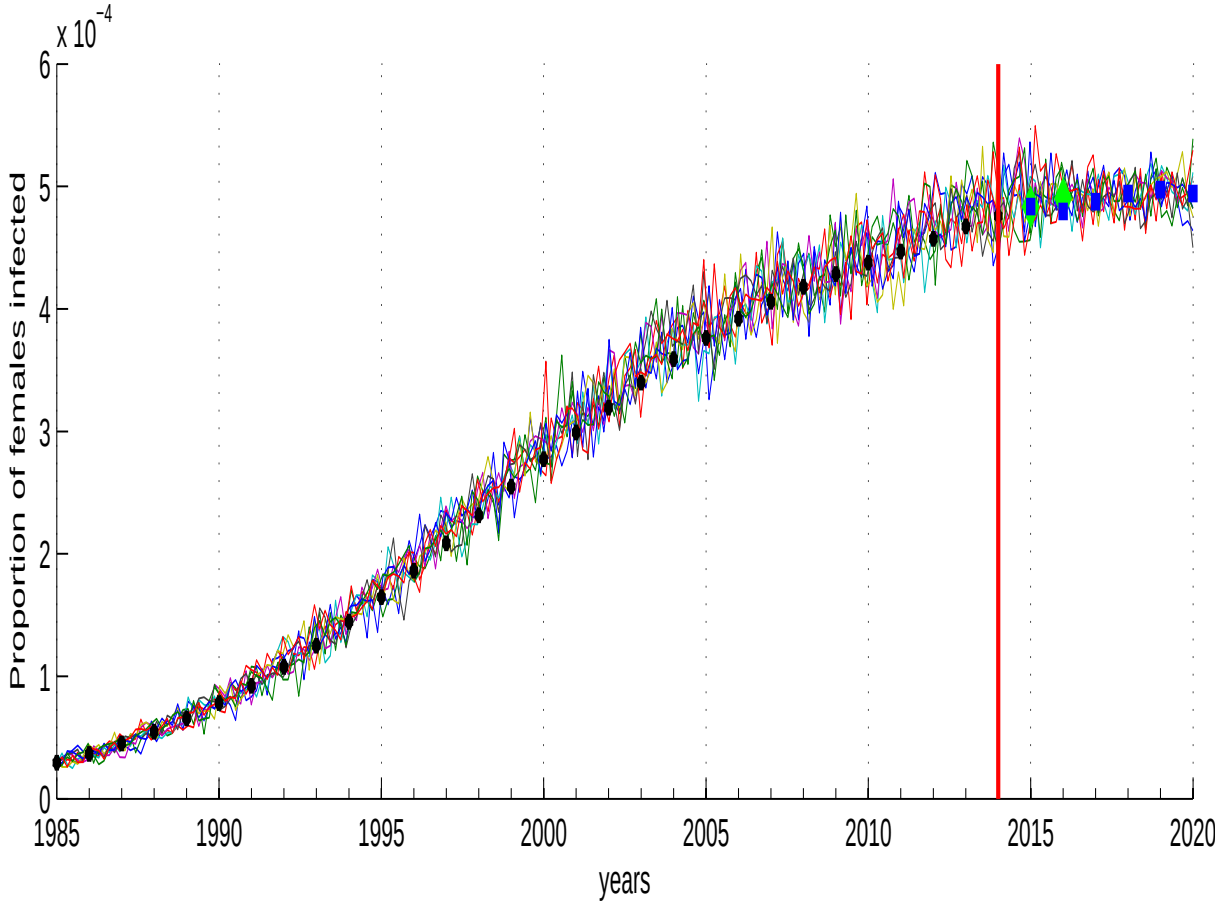

FIGURE 6.5: Plots of 10 trajectories of females infected proportion from 1985 to 2020. The black stars design the true observed data of females from 1985 up to 2014. The green diamonds design real data for 2015 and 2016. Well the blue squares give the mean predicted value of the proportion of infected females. (For interpretation of the references to color of these illustrations, the reader is referred to the electronic version of this dissertation.)

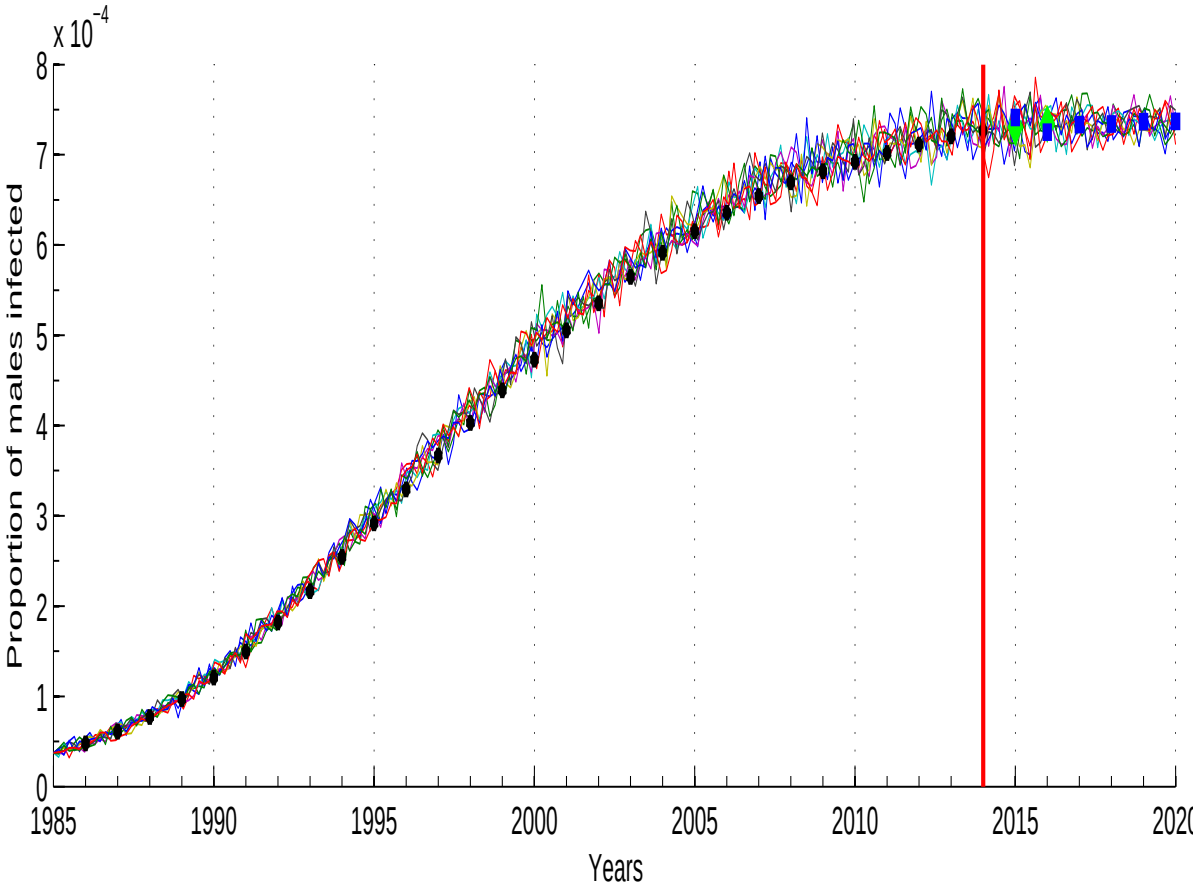

FIGURE 6.6: Plots of 10 trajectories of males infected proportion from 1985 to 2020. The black stars design the true observed data of males from 1985 up to 2014. The green diamonds design real data for 2015 and 2016. Well the blue squares give the mean predicted value of the proportion of infected males. (For interpretation of the references to color of these illustrations, the reader is referred to the electronic version of this dissertation.)

Conclusions and future directions

At the end of this thesis, we hope that the contributing results given in this dissertation be helpful to more knowledge of the modeling of HIV/AIDS dynamic in particular, and for other epidemics which can be modeled by a compartmental model. Which give us some insights of future perspectives.

After presenting the essential preliminaries notions, which give us the helpful tools and necessary ingredients to the comprehension of this dissertation. Foremost, we have been modeled the HIV/AIDS dynamic in a closed heterosexual population by a multidimensional SIR model, using the convergence of its associated master equation to approximate this model by a stochastic diffusion process.

The research studies of this thesis, which focuses on model parameter estimation, lead to two contributions. The first one is parameter estimation by using a contrast function and minimizing the minimum contrast estimator MCE, for which we proved the consistency and asymptotic normality properties. Numerical simulations are well presented; we simulate the data using a Gillespie algorithm which gives exact simulations, the results of simulation ensure the validity of the constructed model. The results obtained in this contribution have been published in Abou-Bakre and El Maroufy [1]. In the second contribution of this thesis, we use Bayesian approach to infer the model parameters of interest. In this contribution, we have used the augmented data method proposed by Eraker [41] and contributing as do El Maroufy et al. [36] and Qaffou et al. [100], since the data is discretely observed in time. To deal with that, an MCMC schemes are developed and used to estimate the model parameters in this case. We prove that the posterior distributions of model parameters converge to a limit distribution, which are distributed according to a Generalized Inverse Gaussian GIG density. Also, a numerical simulation is discussed for this contribution which prove the convergence of the distributions of model parameters to a GIG distribution. The developed results in this contribution are submitted for a possible publication in an international journal.

The main goal of this thesis is to apply the methods and techniques, developed in chapter 3 and 4, to real data and application. Therefore, we have applied the methodology, algorithms and results obtained in this dissertation, to a real data-set for Morocco's case, for which we can

affirm that the HIV and AIDS will persist and still increasing in the population since we find that the reproduction number R_0 is bigger than one.

The results presented in this dissertation provide some guidelines on our future directions and perspectives. A natural first perspective is, to study the complete model with major complexities and more realistic assumptions, that are: a population in which both types of transmission heterosexual and homosexual, non closed population which means mobility and migration of individuals, other ways of contamination rather than only sexual including drug users, taking treatment and vaccination in consideration. As a second perspective is, to study the stochastic optimal control of this model. Furthermore, generalize the study to other epidemics like H1N1, tuberculosis, Ebola and others.

Appendix A

To examine the first statement in Assumption 1, there are two conditions ensuring the existence and uniqueness of the solution to the Eq. (4.1) that are local Lipschitz and the linear growth conditions of both coefficients drift and diffusion (see for example, Liptser and Shiryaev [84] and Kutoyants [82] for details).

To verify the local Lipschitz for ξ, let x_1 and x_2 be in $\mathbb{R}^4$, we have after some calculus

$$||\xi(x_1) - \xi(x_2)|| \leq K||x_1 - x_2||,$$

where $K = \max\{K_1, K_2, K_3, K_4\}$ with $K_1 = 2\beta^2 + (\alpha\delta + \mu)^2$, $K_2 = (\mu + \gamma)2 + (\alpha\delta)^2$, $K_3 = 2\beta^2 + ((1-\alpha)\delta + \mu)^2$ and $K_2 = (\mu + \gamma)2 + ((1-\alpha)\delta)^2$.

For the linear growth condition of ξ coefficient we obtain after a simple calculus:

$$||\xi(x, \theta)||^2 \leq K' \left(1 + ||x||^2\right),$$

where $K' = \max\{K'_1, K'_2, K'_3, K'_4\}$ with $K'_1 = \mu^2 + 2\beta^2 + (\alpha\mu\delta)^2$, $K'_2 = (\mu + \gamma)2 + (\alpha\mu\delta)^2$, $K'_3 = \mu^2 + 2\beta^2 + ((1-\alpha)\mu\delta)^2$ and $K'_4 = (\mu + \gamma)2 + ((1-\alpha)\mu\delta)^2$.

For the second statement in Assumption 1, since the process X_t is such that $|X_t| \leq 1$, for all $t \geq 0$, then for all $m > 0$ we have: $\sup_t |X_t|^m < \infty$. Therefore Assumption 1 is assured.

For Assumption 2; looking at the drift and diffusion coefficients Eqs. (3.8) and (3.9), which are constructed in Chapter 3. We observe that all components of these coefficients are well defined, and built by the usual functions, then they admit continuous derivatives up to superior order. thereby the Assumption 2 is satisfied.

We calculate the determinant of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T = \boldsymbol{\Sigma}$, and $\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{\sigma})^2$ where $\det(\boldsymbol{\sigma}) = \det(\boldsymbol{\sigma}_1)\det(\boldsymbol{\sigma}_2)$ with $\det(\boldsymbol{\sigma}_1) > 0$ and $\det(\boldsymbol{\sigma}_2) > 0$. We calculate the inverse of $\boldsymbol{\Sigma}$, which is equal to:

$$\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\sigma}^{-1})^T \boldsymbol{\sigma}^{-1}$$

$$= \begin{pmatrix} \frac{(\mu+\gamma)y_1 + \frac{\beta y_2}{x_2+y_2}x_1}{\Delta_1^2} & \frac{\beta \frac{y_2}{x_2+y_2}x_1}{\Delta_1^2} & 0 & 0 \\ \frac{\frac{\beta y_2}{x_2+y_2}x_1}{\Delta_1^2} & \frac{\alpha(\mu+\delta z)+\mu x_1+\frac{\beta y_2}{x_2+y_2}x_1}{\Delta_1^2} & 0 & 0 \\ 0 & 0 & \frac{(\mu+\gamma)y_2+\frac{\beta y_1}{x_1+y_1}x_2}{\Delta_2^2} & \frac{\frac{\beta y_1}{x_1+y_1}x_2}{\Delta_2^2} \\ 0 & 0 & \frac{\frac{\beta y_1}{x_1+y_1}x_2}{\Delta_2^2} & \frac{(1-\alpha)(\mu+\delta z)+\mu x_2+\frac{\beta y_1}{x_1+y_1}x_2}{\Delta_2^2} \end{pmatrix},$$

where

$$\Delta_1^2 = (\mu+\gamma)y_1\left(\alpha(\mu+\delta z)+\mu x_1+\beta\frac{y_2}{x_2+y_2}x_1\right)+\beta\frac{y_2}{x_2+y_2}x_1\left(\alpha(\mu+\delta z)+\mu x_1\right)$$

$$\Delta_2^2 = (\mu+\gamma)y_2\left((1-\alpha)(\mu+\delta z)+\mu x_2+\beta\frac{y_1}{x_1+y_1}x_2\right)+\beta\frac{y_1}{x_1+y_1}x_2\left((1-\alpha)(\mu+\delta z)+\mu x_2\right)$$

We check that all elements of the above matrix admit second derivatives with respect to θ and x, because it is constrained by product of usual functions. Thus $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\sigma}^{-1})^T\boldsymbol{\sigma}^{-1} \in \mathcal{C}^2(R^4 \times \bar{\Theta})$

For the Assumption 3 is well done due to the drift and diffusion coefficients form, that ensure the identifiability of these coefficients with respect to θ and x. Therefor ξ and $\boldsymbol{\sigma}$ are identifiable.

Appendix B

Let introduce the following Lemma:

Lemma B.1. *Let $f \in \mathcal{C}^{1,1}$ on $\Theta \times \mathbb{R}^p$ a differentiable with continuous derivatives. If the Assumptions $A_1 - A_3$ are fulfilled, then under $\mathbb{P}_\theta$*

i.

$$\frac{1}{n}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) \longrightarrow \int_0^1 f(X_s^0, \theta)ds,$$

as $\epsilon \to 0$ and $n \to \infty$ uniformly in θ.

ii.

$$\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta)P_k(\theta_0) \longrightarrow 0$$

as $\epsilon \to 0$ and $n \to \infty$ uniformly in θ.

To proof the result in the Lemma B.1, we need the following lemma, which their proof is given in Sørensen and Uchida [115, Lemma 3].

Lemma B.2. *Let $f \in \mathcal{C}^{2,2}$ on $\Theta \times \mathbb{R}^p$ a differentiable with continuous derivatives. If the Assumptions $A_1 - A_3$ are fulfilled, then under $\mathbb{P}_\theta$*

i.

$$\epsilon^{-2}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta)P_k^i P_k^j(\theta_0) \longrightarrow \int_0^1 f(X_s^0, \theta)\mathbf{\Sigma}^{ij}(X_s^0, \theta_0)ds,$$

uniformly in $\theta \in \bar{\Theta}$ as $\epsilon \to 0$ and $n \to \infty$.

ii.

$$\epsilon^{-2}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta)P_k^i P_k^j(\theta_0) \longrightarrow \int_0^1 f(X_s^0, \theta)\mathbf{\Sigma}^{ij}(X_s^0, \theta_0)ds + M^2\int_0^1 f(X_s^0, \theta)B^i B^j(X_s^0, \theta_0, \theta)ds,$$

uniformly in $\theta \in \bar{\Theta}$ as $\epsilon \to 0$ and $n \to \infty$.

Proof of Lemma B.1. To prove this Lemma, we need a Taylor stochastic expansion, developed in Azencott [12], of the process X_t, we have the approximations:

$$X_t = x_\theta(t) + \epsilon g_\theta(t) + \epsilon^2 R^{\epsilon,n}(t) \tag{B.1}$$

where $x_\theta(t)$ is the solution of the deterministic equation $dx_\theta(t) = \xi(x_\theta(t))dt$ verifying $x_\theta(0) = x_0$, the remainder R satisfies for $\epsilon \to 0 \quad \sup_{t\leq 1} |\epsilon R^{\epsilon,n}(t)| \to 0$, and g is a continuous martingale satisfies:

$$dg_\theta(t) = \frac{\partial \xi}{\partial x}(\theta, x_\theta(t)) g_\theta(t)dt + \boldsymbol{\sigma}(\theta, x_\theta(t))dW_t, \quad \text{with } g_\theta(0) = 0.$$

The approximate process of Eq. (B.1) in discrete times $t_k, k \in \{1, \ldots, n\}$ is given by:

$$X_{t_k} \simeq x_\theta(t_k) + \epsilon g_\theta(t_k). \tag{B.2}$$

Since X_t^0, the solution of the deterministic process correspond to the case of $\epsilon = 0$ in Eq. (4.1), is indistinguishable to the deterministic process x_θ.
We have

$$\sup_{t\leq 1} |X_t - X_t^0| = \sup_{t\leq 1} |\epsilon g_\theta(t)|, \tag{B.3}$$

which goes to zero as $\epsilon \longrightarrow 0$. As $f \in \mathcal{C}^{1,1}$, then there exist a C such that $\sup_{t\leq 0} \frac{\partial}{\partial x} f(X_t^0) \leq C$, thus

$$\sup_{t\leq 1} |f(X_t) - f(X_t^0)| \leq \sup_{t\leq 1} C|X_t - X_t^0|,$$

therefore for all θ

$$\sup_{t\leq 1} |f(X_t) - f(X_t^0)| \longrightarrow 0.$$

Moreover, $\sup_\theta \frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta)$ is bounded in P_{θ_0}-probability as $\epsilon \to 0$ and $n \to \infty$, while f is C^1 on θ. Then

$$\sup_{\epsilon,n} \mathbb{E}_\theta \left[\sup_\theta \frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta) \right] < \infty.$$

Therefore the family $\left(\frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta) \right)_{n\mathbb{N}^\star}$, is tight uniformly[1] with θ.
Thus for every θ

$$\frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta) \longrightarrow \int_0^1 f(X_s^0, \theta) ds,$$

as $n \to \infty$, uniformly in $\theta \in \bar{\Theta}$.

[1] A sequence of random variables $X_1, X_2, \ldots$ is uniformly tight if for every $\varepsilon > 0$ there exists a compact K such that $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$ for every K.

To prove the second assertion, it is enough to check the conditions of the following in Genon-Catalot and Jacod [49, Lemma 9].

Lemma B.3. *Let χ_i and U be two random variables with χ_i being $\mathcal{G}_i$-measurable, If*

(i) $\sum_{i}^{n} \mathbb{E}(\chi_i|\mathcal{G}_{i-1}) \longrightarrow U$ *in* $\mathbb{P}$*-probability,*

(ii) $\sum_{i}^{n} \mathbb{E}\left[(\chi_i)^2|\mathcal{G}_{i-1}\right] \longrightarrow 0$ *in* $\mathbb{P}$*-probability,*

Then,

$\sum_{i}^{n} \chi_i^n \longrightarrow U$ *in* $\mathbb{P}$*-probability.*

where the filtration $\mathcal{G}_t$ generated by the Brownian process $\mathcal{G}_t = \sigma(B_s, s \leq t)$.

Let $\chi_k = f(X_{t_{k-1}}, \theta)\mathbf{P}_k(\theta_0)$, we want to show that $\chi_k \longrightarrow 0$, then it must verifies the two conditions of the Lemma Lemma B.3.

To establish these conditions let recall the a result from Florens-Zmirou [45, Lemma1]

Lemma B.4. . *Let $f \in C^{2(s+1)}$ thus,*

$$\mathbb{E}(f(X_{K\Delta})|\mathcal{F}_{k-1}) = \sum_{l=0}^{s} \frac{\Delta^l}{l!}\mathcal{A}^l f(X_{(k-1)\Delta}) + \int_0^{\Delta}\int_0^{u_1}\dots\int_0^{u_s} \mathbb{E}\left[(\mathcal{A}^{s+1}f)(X_{(K-1)\Delta+u_s+1})|\mathcal{F}_{k-1}\right] du_1 \dots du_{s+1},$$

where Δ is a fixed step size and $\mathcal{A} = \xi\dfrac{\partial}{\partial x} + \sigma^2\dfrac{\partial^2}{\partial x^2}$ is the infinitesimal generator associated to the diffusion process $(X_t)_{t\geq 0}$. Putting $\Phi(x, y) = x - y$ and $\Delta = \dfrac{1}{n}$, so we have

$$\mathbb{E}(\Phi(X_{t_K}, X_{t_{K-1}})|\mathcal{F}_{k-1}) = \Phi(X_{t_{K-1}}, X_{t_{K-1}}) + \frac{1}{n}\xi(X_{t_{k-1}}) + \int_0^{\frac{1}{n}}\int_0^{u_1} \mathbb{E}\left[(\mathcal{A}^2\Phi(X_{t_K+u_2}, X_{t_{K-1}})|\mathcal{F}_{k-1}\right] du_1 du_2.$$

Lemma 1 in Kessler [73], shows that the integral remainder in Lemma B.4 is bounded in probability and goes to *zero*. Using the said Lemma we have,

$$\int_0^{\frac{1}{n}}\int_0^{u_1} \mathbb{E}\left[(\mathcal{A}^2\Phi(X_{t_K+u_2}, X_{t_{K-1}})|\mathcal{F}_{k-1}\right] du_1 du_2 = R(\frac{1}{n}, X_{t_{k-1}}),$$

where R satisfies $\exists C > 0$, such that $R(\dfrac{1}{n}, X_{t_{k-1}}) \leq \dfrac{1}{n}C\left(1 + |X_{t_{k-1}}|\right)^C$.

Then we have

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[\mathbf{P}_k(\theta_0)|\mathcal{F}_{k-1}\right] =& \mathbb{E}_{\theta_0}\left[(X_{t_k} - X_{t_{k-1}})|\mathcal{F}_{k-1}\right] - \frac{1}{n}\xi(X_{t_{k-1}}, \theta_0) \\ =& \frac{1}{n}\xi(X_{t_{k-1}}, \theta_0) - \frac{1}{n}\xi(X_{t_{k-1}}, \theta_0) + R(\frac{1}{n}, X_{t_{k-1}}) \\ \leq& \frac{C2^C}{n^2}.\end{aligned}$$

It follows that,

$$\mathbb{E}_{\theta_0}\left[f(X_{t_{k-1}})\mathbf{P}_k(\theta_0)|\mathcal{G}_{k-1}\right] = f(X_{t_{k-1}})\mathbb{E}_{\theta_0}\left[\mathbf{P}_k(\theta_0)|\mathcal{G}_{k-1}\right] \longrightarrow 0,$$

as $n \to \infty$ and $\epsilon \to 0$, which prove the first condition.

For the second condition, we see that

$$\mathbb{E}_{\theta_0}\left[\left(f(X_{t_{k-1}})\mathbf{P}_k(\theta_0)\right)^2|\mathcal{G}_{k-1}\right] = f^2(X_{t_{k-1}})\mathbb{E}_{\theta_0}\left[\mathbf{P}_k^2(\theta_0)|\mathcal{G}_{k-1}\right].$$

But,

$$\mathbf{P}_k(\theta_0)^2 = (X_{t_k} - X_{t_{k-1}})^2 - \frac{2}{n}(X_{t_k} - X_{t_{k-1}})\xi(X_{t_{k-1}}, \theta_0) + \frac{1}{n^2}\xi^2(X_{t_{k-1}}, \theta_0),$$

then,

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[\mathbf{P}_k^2(\theta_0)|\mathcal{F}_{k-1}\right] =& \mathbb{E}_{\theta_0}\left[(X_{t_k} - X_{t_{k-1}})^2|\mathcal{F}_{k-1}\right] + \xi(X_{t_{k-1}}, \theta_0)\mathbb{E}_{\theta_0}\left[(X_{t_k} - X_{t_{k-1}})|\mathcal{F}_{k-1}\right] \\ &+ \frac{1}{n^2}\xi^2(X_{t_{k-1}}, \theta_0).\end{aligned} \tag{B.4}$$

Furthermore, using Lemma B.4, we have

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[(X_{t_k} - X_{t_{k-1}})^2|\mathcal{G}_{k-1}\right] =& (X_{t_{k-1}} - X_{t_{k-1}})^2 + \frac{1}{n}\mathcal{A}\left[(X_{t_k} - X_{t_{k-1}})^2\right] + \frac{1}{2n^2}\mathcal{A}^2\left[(X_{t_k} - X_{t_{k-1}})^2\right] \\ &+ \int_0^{\frac{1}{n}}\int_0^{u_1}\int_0^{u_2}\mathbb{E}\left[(\mathcal{A}^3\left[(X_{t_{k-1}+u_s+1} - X_{t_{k-1}})^2\right])|\mathcal{F}_{k-1}\right]du_1du_2du_3 \\ =& \frac{\epsilon^2}{n}\mathbf{\Sigma}(X_{t_{k-1}}, \theta_0) + \frac{1}{n^2}\xi^2(X_{t_{k-1}}, \theta_0) + R(\frac{1}{n^3}, X_{t_{k-1}})\end{aligned} \tag{B.5}$$

By Eq. (B.5), the expectation $\mathbb{E}_{\theta_0}\left[\mathbf{P}_k^2(\theta_0)|\mathcal{F}_{k-1}\right]$ in Eq. (B.4) goes to *zero*, as $n \to \infty$, hence

$$\mathbb{E}_{\theta_0}\left[\left(f(X_{t_{k-1}})\mathbf{P}_k(\theta_0)\right)^2|\mathcal{F}_{k-1}\right] \longrightarrow 0.$$

This completes the proof. □

Appendix C

The joint posterior density is given by the equation (5.3). Or the component inside the exponential operator in (5.4) can be rewritten as function of β under

$$\begin{aligned}\hat{\mathbb{P}}_i =& \beta \begin{pmatrix} x_{1i}y_{2i} \\ -x_{1i}y_{2i} \\ x_{2i}y_{1i} \\ -x_{2i}y_{1i} \end{pmatrix} \Delta t + \begin{pmatrix} \dfrac{\Delta Y_i(1)}{\Delta t} + \mu x_{1i} - \alpha(\mu + \delta z_i) \\ \dfrac{\Delta Y_i(2)}{\Delta t} + (\mu + \gamma)\, y_{1i} \\ \dfrac{\Delta Y_i(3)}{\Delta t} + \mu x_{2i} - (1-\alpha)(\mu + \delta z_i) \\ \dfrac{\Delta Y_i(4)}{\Delta t} + (\mu + \gamma)\, y_{2i} \end{pmatrix} \Delta t \\ =& \beta \begin{pmatrix} x_{1i}y_{2i} \\ -x_{1i}y_{2i} \\ x_{2i}y_{1i} \\ -x_{2i}y_{1i} \end{pmatrix} \Delta t + \begin{pmatrix} A(1)_i \\ A(2)_i \\ A(3)_i \\ A(4)_i \end{pmatrix} \Delta t.\end{aligned}$$

Let $\Delta_i = \beta C_{1i} + C_{2i}$ with $C_{1i} = x_1 y_2 \left((\mu + \gamma) y_1 + (\alpha(\mu + \delta z) + \mu x_1)\right)$ and $C_{2i} = (\mu + \gamma) y_1 \left(\alpha(\mu + \delta z) + \mu x_1\right)$. Then,

$$\begin{aligned}\hat{\mathbb{P}}_i' \left(\Sigma_{i-1}^{-1} \Delta t\right) \hat{\mathbb{P}}_i =& \frac{1}{\beta C_{1i} + C_{2i}} \left(K_{0i} + K_{1i}\beta + K_{2i}\beta^2\right) \\ &+ \frac{1}{\beta C_{1i}^{'} + C_{2i}^{'}} \left(K_{0i}^{'} + K_{1i}^{'}\beta + K_{2i}^{'}\beta^2\right),\end{aligned}$$

where

$$\begin{aligned} K_{0i} &= (\mu + \gamma)\, y_{1i} A(1)_i^2 \Delta t + \left(\alpha(\mu + \delta Z_i) + \mu x_{i1}\right) A(2)_i^2 \Delta t, \\ K_{1i} &= 2(\mu + \gamma) y_{1i} x_{1i} y_{2i} A(1)_i - 2 x_{1i} y_{2i} \left(\alpha(\mu + \delta z_i) + \mu x_{i1}\right) A(2)_i + \left(A(1)_i + A(2)_i\right)^2 \Delta t, \\ K_{2i} &= \left((\mu + \gamma)\, y_{1i} (x_{1i} y_{2i})^2 + \left(\alpha(\mu + \delta z_i) + \mu x_{i1} (x_{1i} y_{2i})^2\right)\right) \Delta t, \end{aligned}$$

and the constants $C_{1i}^{'}$, $C_{2i}^{'}$, $K_{0i}^{'}$, $K_{1i}^{'}$ and $K_{2i}^{'}$ are obtained using the above constants by exchanging x_{i1}, y_{1i} and α respectively by x_{i2}, y_{i2} and $1 - \alpha$, and exchanging $A(1)_i$,$A(2)_i$ by $A(3)_i$

and $A(4)_i$ respectively. If $C_{1i} = 0$ (C'_{1i}=0) that correspond to the case when we have no female (male) infected. In this situation the study is reduced to a simple SIR model see for example Qaffou et al. [100].
Let the following assumption hold true from now on: "the numbers of infected people female and male are not equal to zero" which means that $y_{1_i} \neq 0$ and $y_{2_i} \neq 0$ for all i.

In this case, if $p(\beta) \propto \beta^{m_\beta - 1} \exp(-\lambda_\beta \beta)$, then we obtain

$$\pi(\beta|\hat{Y}_i, \delta, \gamma) \propto \beta^{m_\beta-1} \exp(-\lambda_\beta\beta) \prod_{i=1}^{N} \left(C_{1i}C'_{i1}\right)^{-\frac{1}{2}} (\beta + C_\beta)^{-1}$$
$$\times \exp\left\{-\frac{1}{2}\left[\frac{1}{\beta+C_\beta}\left(\frac{K_{0i}}{C_{1i}}+\frac{K_{1i}}{C_{1i}}\beta+\frac{K_{2i}}{C_{1i}}\beta^2\right)+\frac{1}{\beta+C_\beta}\left(\frac{K'_{0i}}{C'_{i1}}+\frac{K'_{1i}}{C'_{i1}}\beta+\frac{K'_{2i}}{C'_{i1}}\beta^2\right)\right]\right\},$$

where $C_\beta = \min\limits_{1\leq i\leq N}\left\{\frac{C_{2i}}{C_{1i}}, \frac{C'_{2i}}{C'_{1i}}\right\}$. Let use the same notation for the new constants $K_{ij} = \frac{K_{ij}}{C_{1i}}$ and $K'_{ij} = \frac{K'_{ij}}{C'_{1i}}$. Then the following result is obtained:

$$\pi(\beta|\hat{Y}_i, \delta, \gamma) \propto \exp(\lambda_\beta C_\beta) h(\beta)^{m_\beta - N - 1} \exp\left\{-\frac{1}{2}\left[\frac{M_\beta}{h(\beta)} + \left(M'_\beta + 2\lambda_\beta\right) h(\beta)\right]\right\}, \tag{C.1}$$

where $h(\beta) = \beta + C_\beta$, $M_\beta = \sum\limits_{i=0}^{N}\left(K_{0i} + K_{2i}C_\beta^2 - K_{1i}C_\beta\right) + \sum\limits_{i=0}^{N}\left(K'_{0i} + K'_{2i}C_\beta^2 - K'_{1i}C_\beta\right)$ and $M'_\beta = \sum\limits_{i=1}^{N} K_{2i} + K'_{2i}$.

We remark that $h(\beta)$ has a GIG distribution form and since $h(\beta) = \beta + C_\beta$ is a linear function with $C_\beta > 0$ and their Jacobian equals to 1. Then, Eq. (C.1) becomes:

$$\pi(\beta|\hat{Y}_i, \delta, \gamma) \propto \exp(\lambda_\beta C_\beta) \beta^{m_\beta - N - 1} \exp\left\{-\frac{1}{2}\left[\frac{M_\beta}{\beta} + \left(M'_\beta + 2\lambda_\beta\right) \beta\right]\right\}.$$

In the same manner, we construct the posterior distribution of the parameter δ we can see that:

$$\hat{\mathbb{P}}_i = \delta \begin{pmatrix} -\alpha z_i \\ 0 \\ -(1-\alpha) z_i \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} \frac{\Delta Y_i(1)}{\Delta t} + \mu x_{1_i} - \alpha\mu \\ \frac{\Delta Y_i(2)}{\Delta t} + (\mu+\gamma)\, y_{1_i} \\ \frac{\Delta Y_i(1)}{\Delta t} + \mu x_{2_i} - \mu(1-\alpha) \\ \frac{\Delta Y_i(1)}{\Delta t} + (\mu+\gamma)\, y_{2_i} \end{pmatrix} \Delta t$$

$$=\delta \begin{pmatrix} -\alpha z_i \\ 0 \\ -(1-\alpha)z_i \\ 0 \end{pmatrix} \Delta t + \begin{pmatrix} F(1)_i \\ F(2)_i \\ F(3)_i \\ F(4)_i \end{pmatrix} \Delta t,$$

if $\delta \propto \delta^{m-1} \exp(-\lambda\delta)$ and $z_i \neq 0$ then,

$$\pi(\delta|\hat{Y}_i, \gamma, \beta) \propto \exp(\lambda_\delta C_\delta) h(\delta)^{m_\delta - N - 1} \ \exp\left\{-\frac{1}{2}\left[\frac{M_\delta}{h(\delta)} + \left(M_\delta^{'} + 2\lambda_\delta\right) h(\delta)\right]\right\},$$

where $C_\delta = \min\limits_{1\leq i\leq N}\left\{\frac{C_{2i}}{C_{1i}}, \frac{C_{2i}^{'}}{C_{1i}^{'}}\right\}$, $M_\delta = \sum\limits_{i=0}^{N}\left(E_{0i} + E_{2i}C_\delta^2 - E_{1i}C_\delta\right) + \sum\limits_{i=0}^{N}\left(E_{0i}^{'} + E_{2i}^{'}{C_\delta}^2 - E_{1i}^{'}C_\delta\right)$ and $M_\delta^{'} = \sum\limits_{i=1}^{N} E_{2i} + E_{2i}^{'}$ with

$$\begin{aligned}
E_i &= C_{1i}\delta + C_{2i} \\
&= \alpha z_i\left((\mu+\gamma)y_{1i} + \beta x_{1i}y_{2i}\right)\delta + (\mu+\gamma)y_{1i}\left(\alpha\mu + \mu x_{1i} + \beta x_{1i}y_{2i}\right) + \beta x_{1i}y_{2i}\left(\alpha\mu + \mu x_{1i}\right), \\
E_{0i} &= \left((\mu+\gamma)y_{1i} + \beta x_{1i}y_{2i}\right)F(1)_i^2 + 2\beta x_{1i}y_{2i}F(1)_i F(2)_i + \left(\alpha\mu + \mu x_{1i} + \beta x_{1i}y_{2i}\right)F(2)_i^2, \\
E_{1i} &= \alpha z_i\left(F(2)_i^2 - 2(F(1)_i\left((\mu+\gamma)y_{1i} + \beta x_{1i}y_{2i}\right) + F(2)_i\beta x_{1i}y_{2i})\right), \\
E_{2i} &= \alpha^2 z_i^2\left((\mu+\gamma)y_{1i} + \beta x_{1i}y_{2i}\right),
\end{aligned}$$

where the constants with prime symbol ($'$) are obtained from the above constants by interchanging x_{i1}, y_{1i} and x_{i2}, y_{i2}, α and $1-\alpha$ and interchanging $F(1)_i$, $F(2)_i$ and $F(3)_i$, $F(4)_i$.

The same techniques are used for γ;

$$\hat{\mathbb{P}}_i = \gamma \begin{pmatrix} 0 \\ y_{1i} \\ 0 \\ y_{2i} \end{pmatrix} \Delta t + \begin{pmatrix} \frac{\Delta Y(1)}{\Delta t} + \mu x_{1i} - \alpha(\mu + \delta z) \\ \frac{\Delta Y_i(2)}{\Delta t} + \mu y_{1i} \\ \frac{\Delta Y_i(1)}{\Delta t} + \mu x_{2i} - (1-\alpha)(\mu + \delta z_i) \\ \frac{\Delta Y_i(1)}{\Delta t} + \mu y_{2i} \end{pmatrix} \Delta t \quad = \gamma \begin{pmatrix} 0 \\ y_{1i} \\ 0 \\ y_{2i} \end{pmatrix} \Delta t + \begin{pmatrix} B(1)_i \\ B(2)_i \\ B(3)_i \\ B(4)_i \end{pmatrix} \Delta t$$

if $\gamma \propto \gamma^{m_\gamma - 1} \exp(-\lambda_\gamma\gamma)$ then, we have after tedious calculus that

$$\pi(\gamma|\hat{Y}_i, \delta, \beta) \propto \ \exp(\lambda_\gamma C_\gamma)\gamma^{m_\gamma - N - 1} \ \exp\left\{-\frac{1}{2}\left[\frac{M_\gamma}{\gamma} + \left(M_\gamma^{'} + 2\lambda_\gamma\right)\gamma\right]\right\},$$

where $C_\gamma = \min_{1\le i\le N}\left\{\frac{C_{2i}}{C_{1i}}, \frac{C'_{2i}}{C'_{1i}}\right\}$, $M_\gamma = \sum_{i=0}^{N}\left(D_{0i}+D_{2i}C_\gamma^2 - D_{1i}C_\gamma\right)+\sum_{i=0}^{N}\left(D'_{0i}+D'_{2i}{C_\gamma}^2 - D'_{1i}C_\gamma\right)$ and $M'_\gamma = \sum_{i=1}^{N} D_{2i} + D'_{2i}$ with

$$
\begin{aligned}
D_i &= C_{1i}\gamma + C_{2i} \\
&= \mu y_{1i}\left(\alpha(\mu+\delta z_i)+\mu x_{1i}+\beta x_{1i}y_{2i}\right)\gamma + \left(\mu y_{1i}+\beta x_{1i}y_{2i}\right)\left(\alpha(\mu+\delta z_i)+\mu x_{1i}\right)+\beta x_{1i}y_{2i} \\
D_{0i} &= (\mu y_{1i}+\beta x_{1i}y_{2i})B(1)_i^2 + \beta x_{1i}y_{2i}(2B(1)_i^2 + B(1)_iB(2)_i) \\
D_{1i} &= 2y_{1i}(B(1)_i + B(2)_i) + B(1)_i^2 \\
D_{2i} &= y_{1i}(\mu y_{1i}+\beta x_{1i}y_{2i})
\end{aligned}
$$

the constants D'_{0i}, D'_{1i}, $D_2{}'_i$ and D'_i are obtained using the above constants by interchanging x_{i1}, y_{1i} and x_{i2}, y_{i2}, α and $1-\alpha$, $B(1)_i$, $B(2)_i$ and $B(3)_i$, $B(4)_i$.

Appendix D

The constructed database, is obtained by combining statistics and data from Moroccan High Commission for Planing HCP [64], Ministry of Public Health with the support of UNAIDS [107, 108], the database of World Bank [14] and the SPECTRUM software [11]. We restrict the data to the sexually active population only.

year	$\mathbf{S_F}$	$\mathbf{S_M}$	$\mathbf{I_F}$	$\mathbf{I_M}$	**AIDS**
1986	5370920	5370920	35	110	1
1987	5515825	5515825	78	248	9
1988	5663588	5663588	132	419	14
1989	5814875	5814875	199	632	20
1990	5970053	5970053	282	894	26
1991	6128126	6128126	383	1214	28
1992	6290202	6290202	509	1612	30
1993	6457337	6457337	721	2041	44
1994	6629538	6629538	973	2555	77
1995	6806267	6806267	1266	3150	57
1996	6987697	6987697	1623	3874	66
1997	7171473	7171473	2044	4729	92
1998	7350417	7350417	2719	5553	93
1999	7515002	7515002	3518	6530	165
2000	7649765	7649765	4427	7641	112
2001	7767033	7767033	5390	8818	129
2002	7871873	7871873	6362	10006	150
2003	7971494	7971494	7367	11139	205
2004	8069812	8069812	8340	12236	271
2005	8167432	8167432	9255	13269	289
2006	8264123	8264123	10144	14270	291
2007	8357674	8357674	10999	15235	369
2008	8447147	8447147	11886	16122	416

year	$\mathbf{S_F}$	$\mathbf{S_M}$	$\mathbf{I_F}$	$\mathbf{I_M}$	**AIDS**
2009	8533958	8533958	12732	16968	412
2010	8621216	8621216	13564	17801	461
2011	8705203	8705203	14315	18552	350
2012	8784900	8784900	14929	19166	433
2013	8858900	8858900	15528	19764	634
2014	8926659	8926659	16109	20346	513

TABLE D.1: Morocco's database of HIV/AIDS for sexually active population, since 1986 to 2014.

Bibliography

[1] A. Abou-Bakre and H. El Maroufy. "Parameter inference for HIV stochastic diffusion model in closed heterosexual population". In: *Mathematical Methods in the Applied Sciences* 41.18 (2018), pp. 9081–9091 (cit. on pp. 39, 57, 69).

[2] S. Adaszewski et al. "Simultaneous estimation of population receptive field and hemodynamic parameters from single point BOLD responses using Metropolis-Hastings sampling". In: *NeuroImage* 172 (May 15, 2018), pp. 175–193 (cit. on p. 25).

[3] D. Alonso, A. J. McKane, and M. Pascual. "Stochastic amplification in epidemics". In: *Journal of the Royal Society Interface* 4.14 (June 22, 2007), pp. 575–582 (cit. on p. 15).

[4] H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. New York: Springer-Verlag, 2000 (cit. on pp. 26, 39, 51).

[5] G. Arminger and B. O. Muthén. "A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the metropolis-hastings algorithm". In: *Psychometrika* 63.3 (Sept. 1, 1998), pp. 271–300 (cit. on p. 25).

[6] Y. Aït-Sahalia. "Estimating and Testing Continuous-Time Models in Finance: The Role of Transition Densities". In: *Annual Review of Financial Economics* 1.1 (Dec. 5, 2009), pp. 341–359 (cit. on p. 15).

[7] Y. Aït-Sahalia and T. R. Hurd. "Portfolio Choice in Markets with Contagion". In: *Journal of Financial Econometrics* 14.1 (Jan. 1, 2016), pp. 1–28 (cit. on p. 15).

[8] Y. Aït-Sahalia and J. Jacod. "Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data". In: *Journal of Economic Literature* 50.4 (Dec. 2012), pp. 1007–1050 (cit. on p. 15).

[9] Y. Aït-Sahalia et al. "Estimation of the Continuous and Discontinuous Leverage Effects". In: *Journal of the American Statistical Association* 112.520 (Oct. 2, 2017), pp. 1744–1758 (cit. on p. 15).

[10] Y. Aït-Sahalia. "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach". In: *Econometrica* 70.1 (2002), pp. 223–262 (cit. on p. 39).

[11] Avenir Health. *software-spectrum*. 2018. URL: http://www.avenirhealth.org/software-spectrum.php (visited on 08/12/2018) (cit. on pp. 61, 81).

[12] R. Azencott. "Formule de Taylor stochastique et developpement asymptotique d'integrales de Feynmann". In: *Séminaire de Probabilités XVI, 1980/81 Supplément: Géométrie Différentielle Stochastique*. Ed. by J. Azéma and M. Yor. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1982, pp. 237–285 (cit. on pp. 34, 74).

[13] N. T. J. Bailey. *The Mathematical Theory of Infectious Diseases and Its Applications*. Griffin, 1975. 413 pp. (cit. on pp. 2, 4).

[14] world bank. *Maroc | Data*. URL: https://donnees.banquemondiale.org/pays/maroc (visited on 08/12/2018) (cit. on pp. 61, 62, 81).

[15] A. D. Barbour. "On a functional central limit theorem for Markov population processes". In: *Advances in Applied Probability* 6.1 (Mar. 1974), pp. 21–39 (cit. on p. 15).

[16] M. S. Bartlett. "The Critical Community Size for Measles in the United States". In: *Journal of the Royal Statistical Society: Series A (General)* 123.1 (1960), pp. 37–44 (cit. on p. 2).

[17] N. G. Becker. "On a general stochastic epidemic model". In: *Theoretical Population Biology* 11.1 (Feb. 1, 1977), pp. 23–36 (cit. on pp. 26, 51).

[18] D. Bernoulli. "Essai d'une nouvelle analyse de la mortalite causee par la petite verole, et des avantages de l'inoculation pour la prevenir". In: *Histoire de l'Academie Royale des Sciences. Avec les Mémoires de Mathematique & de Physique. Paris* (1760), pp. 1–45 (cit. on pp. v, 2).

[19] B. M. Bibby, M. Jacobsen, and M. Sørensen. "CHAPTER 4 - Estimating Functions for Discretely Sampled Diffusion-Type Models". In: *Handbook of Financial Econometrics: Tools and Techniques*. Ed. by Y. Aït-sahalia and L. P. Hansen. Vol. 1. Handbooks in Finance. San Diego: North-Holland, Jan. 1, 2010, pp. 203–268 (cit. on p. 39).

[20] F. Brauer, P. v. d. Driessche, and J. Wu, eds. *Mathematical Epidemiology*. Mathematical Biosciences Subseries. Berlin Heidelberg: Springer-Verlag, 2008 (cit. on pp. 4–6).

[21] T. Britton. "Estimation in multitype epidemics". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.4 (Jan. 1, 1998), pp. 663–679 (cit. on p. 8).

[22] T. Britton and P. D. O'Neill. "Bayesian Inference for Stochastic Epidemics in Populations with Random Social Structure". In: *Scandinavian Journal of Statistics* 29.3 (Sept. 2002), pp. 375–390 (cit. on pp. 8, 51).

[23] V. Capasso and D. Morale. "Stochastic modelling of tumour-induced angiogenesis". In: *Journal of Mathematical Biology* 58.1 (Jan. 2009), pp. 219–233 (cit. on pp. 15, 37).

[24] C. K. Carter and R. Kohn. "On Gibbs sampling for state space models". In: *Biometrika* 81.3 (Sept. 1, 1994), pp. 541–553 (cit. on p. 25).

[25] G. Casella and E. I. George. "Explaining the Gibbs Sampler". In: *The American Statistician* 46.3 (Aug. 1, 1992), pp. 167–174 (cit. on p. 25).

[26] S. Cauchemez et al. "A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data". In: *Statistics in Medicine* 23.22 (2004), pp. 3469–3487 (cit. on pp. 25, 56).

[27] W. H. O. CDC. *Avian influenza : assessing the pandemic threat*. 2005. URL: http://apps.who.int/iris/handle/10665/68985 (cit. on p. 1).

[28] S. Chib and E. Greenberg. "Understanding the Metropolis-Hastings Algorithm". In: *The American Statistician* (Feb. 27, 2012) (cit. on p. 25).

[29] D. Dacunha-Castelle and D. Florens-Zmirou. "Estimation of the coefficients of a diffusion from discrete observations". In: *Stochastics* 19.4 (Dec. 1, 1986), pp. 263–284 (cit. on p. 39).

[30] P. Damlen, J. Wakefield, and S. Walker. "Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.2 (1999), pp. 331–344 (cit. on p. 25).

[31] R. Deardon et al. "Inference for individual-level models of infectious diseases in large populations". In: *Statistica Sinica* 20.1 (2010), pp. 239–261 (cit. on p. 26).

[32] N. Demiris and P. D. O'Neill. "Bayesian inference for stochastic multitype epidemics in structured populations via random graphs". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.5 (2005), pp. 731–745 (cit. on pp. 8, 25, 55).

[33] K. Dietz. "On the transmission dynamics of HIV". In: *Mathematical Biosciences* 90.1 (July 1, 1988), pp. 397–414 (cit. on p. 29).

[34] M. El Hia et al. "Modelling the spread of HIV / AIDS in Morocco". In: *International Journal of Computer Science Issues* 9.6 (2012), pp. 230–235 (cit. on p. 61).

[35] H. El Maroufy. "Note on the threshold theorem of a heterogeneous SIR epidemic". In: *Statistics & Probability Letters* 83.1 (Jan. 1, 2013), pp. 292–296 (cit. on p. 51).

[36] H. El Maroufy et al. "Bayesian inference for nonlinear stochastic SIR epidemic model". In: *Journal of Statistical Computation and Simulation* 86.11 (July 23, 2016), pp. 2229–2240 (cit. on pp. 5, 8, 26, 51, 69).

[37] H. El Maroufy, D. Kiouach, and T. Ziad. "Final outcome probabilities for SIR epidemic model". In: *Communications in Statistics - Theory and Methods* 45.8 (2016), pp. 2426–2437 (cit. on p. 40).

[38] H. El Maroufy, L. Omari, and Z. Taib. "Transition Probabilities for Generalized SIR Epidemic Model". In: *Stochastic Models* 28.1 (2012), pp. 15–28 (cit. on pp. 5, 22, 40).

[39] O. Elerian, S. Chib, and N. Shephard. "Likelihood Inference for Discretely Observed Nonlinear Diffusions". In: *Econometrica* 69.4 (July 2001), pp. 959–993 (cit. on p. 54).

[40] J. Elf and M. Ehrenberg. "Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation". In: *Genome Research* 13.11 (Nov. 2003), pp. 2475–2484 (cit. on p. 15).

[41] B. Eraker. "MCMC Analysis of Diffusion Models With Application to Finance". In: *Journal of Business & Economic Statistics* 19.2 (Apr. 1, 2001), pp. 177–191 (cit. on pp. 8, 26, 51, 52, 69).

[42] B. Eraker and J. Wang. "A non-linear dynamic model of the variance risk premium". In: *Journal of Econometrics* 187.2 (Aug. 2015), pp. 547–556 (cit. on p. 15).

[43] B. Eraker and Y. Wu. "Explaining the negative returns to volatility claims: An equilibrium approach". In: *Journal of Financial Economics* 125.1 (July 2017), pp. 72–98 (cit. on p. 15).

[44] L. Ferm, P. Lötstedt, and A. Hellander. "A Hierarchy of Approximations of the Master Equation Scaled by a Size Parameter". In: *Journal of Scientific Computing* 34.2 (Feb. 1, 2008), pp. 127–151 (cit. on p. 15).

[45] D Florens-Zmirou. "Approximate Discrete-Time Schemes for Statistics of Diffusion Processes. Statistics". In: *Statistics (Ber).* 4.20 (1989), pp. 547–C557 (cit. on p. 75).

[46] C. Fuchs. *Inference for Diffusion Processes: With Applications in Life Sciences*. Berlin Heidelberg: Springer-Verlag, 2013 (cit. on pp. 6, 11, 22, 33, 36, 37, 52).

[47] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (Nov. 1984), pp. 721–741 (cit. on p. 25).

[48] V. Genon-Catalot. "Maximum contrast estimation for diffusion processes from discrete observations". In: *Statistics* 21.1 (1990), pp. 99–116 (cit. on pp. 40, 45).

[49] V. Genon-Catalot and J. Jacod. "On the estimation of the diffusion coefficient for multi-dimensional diffusion processes". In: *Annales de l'I.H.P. Probabilités et statistiques* 29.1 (1993), pp. 119–151 (cit. on p. 75).

[50] J. Geweke and H. Tanizaki. "Bayesian estimation of state-space models using the Metropolis–Hastings algorithm within Gibbs sampling". In: *Computational Statistics & Data Analysis* 37.2 (Aug. 28, 2001), pp. 151–170 (cit. on p. 25).

[51] G. J. Gibson and E. Renshaw. "Estimating parameters in stochastic compartmental models using Markov chain methods". In: *Mathematical Medicine and Biology: A Journal of the IMA* 15.1 (Mar. 1, 1998), pp. 19–40 (cit. on p. 26).

[52] M. Gibson and J. Bruck. "An Efficient Algorithm for Generating Trajectories of Stochastic Gene Regulation Reaction". In: *The journal of physical chemistry A* 104.9 (1998), pp. 1876–1889 (cit. on pp. 59, 63).

[53] W. R. Gilks, N. G. Best, and K. K. C. Tan. "Adaptive Rejection Metropolis Sampling within Gibbs Sampling". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 44.4 (1995), pp. 455–472 (cit. on p. 25).

[54] D. T. Gillespie. "Approximating the master equation by Fokker–Planck-type equations for single-variable chemical systems". In: *The Journal of Chemical Physics* 72.10 (May 15, 1980), pp. 5363–5370 (cit. on p. 37).

[55] D. T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25 (Dec. 1, 1977), pp. 2340–2361 (cit. on p. 27).

[56] A. Gloter and M. Sørensen. "Estimation for stochastic differential equations with a small diffusion coefficient". In: *Stoch. Process. their Appl.* 119.3 (2009), pp. 679–699 (cit. on p. 40).

[57] N. S. Goel and N. Richter-Dyn. *Stochastic Models in Biology*. Caldwell, N.J: The Blackburn Press, Apr. 2004. 269 pp. (cit. on p. 37).

[58] A. Golightly and D. J. Wilkinson. "Bayesian inference for nonlinear multivariate diffusion models observed with error". In: *Computational Statistics & Data Analysis* 52.3 (Jan. 1, 2008), pp. 1674–1693 (cit. on pp. 15, 37).

[59] A. Golightly and D. J. Wilkinson. "Bayesian inference for stochastic kinetic models using a diffusion approximation". In: *Biometrics* 61.3 (2005), pp. 781–788 (cit. on pp. 15, 37).

[60] Golightly Andrew and Wilkinson Darren J. "Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo". In: *Interface Focus* 1.6 (Dec. 6, 2011), pp. 807–820 (cit. on p. 15).

[61] A. Gut. *Probability: A Graduate Course*. 1st ed. New York: Springer-Verlag, 2005. XXIV, 608 (cit. on p. 48).

[62] R. Guy, C. Laredo, and E. Vergu. "Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient". In: *Stoch. Process. their Appl.* 124.1 (2014), pp. 51–80 (cit. on p. 48).

[63] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1, 1970), pp. 97–109 (cit. on pp. 25, 54).

[64] Haut Commissariat au Plan. *Population du Maroc selon l'age et le sexe - Population marocaine selon l'age et le sexe - Données ouvertes - Maroc*. 2014. URL: http://www.data.gov.ma/data/fr/dataset/population-du-maroc-selon-l-age-et-le-sexe/resource/ba4d6e16-5901-42c5-9aff-1e308ca266c8 (visited on 08/12/2018) (cit. on pp. 61, 81).

[65] D. B. Hitchcock. "A History of the Metropolis–Hastings Algorithm". In: *The American Statistician* 57.4 (Nov. 1, 2003), pp. 254–257 (cit. on p. 25).

[66] L. Hufnagel, D. Brockmann, and T. Geisel. "Forecast and control of epidemics in a globalized world". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.42 (Oct. 19, 2004), pp. 15124–15129 (cit. on p. 37).

[67] S. M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer Series in Statistics. New York: Springer-Verlag, 2008 (cit. on p. 19).

[68] E. L. Ionides, C. Bretó, and A. A. King. "Inference for nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 103.49 (Dec. 5, 2006), pp. 18438–18443 (cit. on p. 26).

[69] V. Isham. "Mathematical Modelling of the Transmission Dynamics of HIV Infection and AIDS: a Review". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 151.1 (1988), pp. 5–30 (cit. on p. 8).

[70] I. Jeliazkov. "Marginal Likelihood From the Metropolis–Hastings Output AU - Chib, Siddhartha". In: *Journal of the American Statistical Association* 96.453 (Mar. 1, 2001), pp. 270–281 (cit. on p. 25).

[71] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. 2nd ed. Graduate Texts in Mathematics. New York: Springer-Verlag, 1998 (cit. on p. 11).

[72] W. O. Kermack, A. G. McKendrick, and T. W. Gilbert. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Aug. 1, 1927), pp. 700–721 (cit. on pp. 2, 3).

[73] M. Kessler. "Estimation of an Ergodic Diffusion from Discrete Observations". In: *Scandinavian Journal of Statistics* 24.2 (1997), pp. 211–229 (cit. on pp. 39–41, 75).

[74] F. C. Klebaner. *Introduction to stochastic calculus with applications*. 2nd ed. London ; Hackensack, N.J: Imperial College Press, 2005. 416 pp. (cit. on pp. 11, 14–16).

[75] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Berlin Heidelberg: Springer-Verlag, 1992 (cit. on pp. 6, 11, 19).

[76] M. Kouritsin and L. Deli. "On explicit solutions to stochastic differential equations". In: *Stochastic Analysis and Applications* 18.4 (Jan. 2000), pp. 571–580 (cit. on p. 16).

[77] M. Kremer and C. Morcom. "The effect of changing sexual activity on HIV prevalence". In: *Mathematical Biosciences* 151.1 (July 1, 1998), pp. 99–122 (cit. on p. 29).

[78] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Universitext. Berlin Heidelberg: Springer-Verlag, 2003 (cit. on pp. 6, 11, 13–17, 33, 36).

[79] T. G. Kurtz. *Approximation of Population Processes*. Society for Industrial and Applied Mathematics, Jan. 1981 (cit. on pp. 33, 37).

[80] J. A. Kutoyants. *Identification of dynamical systems with small noise*. Dordrecht: Kluwer, 1994 (cit. on p. 39).

[81] Y. A. Kutoyants. *Parameter estimation for stochastic processes*. Berlin: Heldermann, 1984 (cit. on p. 39).

[82] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer London, 2004. 482 pp. (cit. on pp. 11, 22, 40, 71).

[83] H. K. Leung. "Expansion of the master equation for a biomolecular selection model". In: *Bulletin of Mathematical Biology* 47.2 (Mar. 1, 1985), pp. 231–238 (cit. on p. 15).

[84] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes*. Vol. 6. Springer Berlin Heidelberg, 2001. 427 pp. (cit. on pp. 22, 40, 71).

[85] M. Martcheva. *An Introduction to Mathematical Epidemiology*. Texts in Applied Mathematics. Springer US, 2015 (cit. on p. 4).

[86] R. M. May and R. M. Anderson. "Transmission dynamics of HIV infection". In: *Nature* 326.6109 (Mar. 12, 1987), pp. 137–142 (cit. on p. 29).

[87] R. M. May, R. M. Anderson, and A. R. McLean. "Possible demographic consequences of HIV/AIDS epidemics. I. assuming HIV infection always leads to AIDS". In: *Mathematical Biosciences* 90.1 (July 1, 1988), pp. 475–505 (cit. on pp. 29, 32).

[88] C. W. McMillen. *Pandemics: A Very Short Introduction*. Very Short Introductions. Oxford, New York: Oxford University Press, Dec. 8, 2016. 176 pp. (cit. on p. 1).

[89] N. Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1, 1953), pp. 1087–1092 (cit. on p. 25).

[90] C. J. Mode and C. K. Sleeman. *Stochastic Processes in Epidemiology: HIV/AIDS, Other Infectious Diseases and Computers*. Singapore: World Scientific Publishing Co Pte Ltd, June 9, 2000. 768 pp. (cit. on p. 29).

[91] P. D. O'Neill and G. O. Roberts. "Bayesian inference for partially observed stochastic epidemics". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162.1 (Jan. 1, 1999), pp. 121–129 (cit. on pp. 5, 26, 56).

[92] M. J. Panik. *Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modeling*. 1 edition. Wiley, Mar. 15, 2017. 294 pp. (cit. on pp. 18, 19).

[93] A. R. Pedersen. "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes". In: *Bernoulli* 1.3 (1995), pp. 257–279 (cit. on p. 39).

[94] A. Perelson and P. Nelson. "Mathematical Analysis of HIV-1 Dynamics in Vivo". In: *SIAM Review* 41.1 (Jan. 1, 1999), pp. 3–44 (cit. on p. 8).

[95] A. S. Perelson. "Modelling viral and immune system dynamics". In: *Nature Reviews Immunology* 2.1 (Jan. 2002), p. 28 (cit. on p. 8).

[96] D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. New York: Springer-Verlag, 1984 (cit. on p. 33).

[97] I. Porteous et al. "Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation". In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '08. event-place: New York, NY, USA. ACM, 2008, pp. 569–577 (cit. on p. 25).

[98] M. T. Pratola. "Efficient Metropolis–Hastings Proposal Mechanisms for Bayesian Regression Tree Models". In: *Bayesian Analysis* 11.3 (Sept. 2016), pp. 885–911 (cit. on p. 25).

[99] N. Punyacharoensin et al. "Mathematical models for the study of HIV spread and control amongst men who have sex with men". In: *European Journal of Epidemiology* 26.9 (Sept. 20, 2011), p. 695 (cit. on p. 8).

[100] A. Qaffou, H. El Maroufy, and T. Kernane. "Bayesian parameter inference for stochastic SIR epidemic model with hyperbolic diffusion". In: *Communications in Statistics - Simulation and Computation* 46.9 (Oct. 21, 2017), pp. 6907–6922 (cit. on pp. 8, 26, 51, 56, 69, 78).

[101] J. D. Ramshaw. "Augmented Langevin approach to fluctuations in nonlinear irreversible processes". In: *Journal of Statistical Physics* 38.3 (Feb. 1, 1985), pp. 669–680 (cit. on pp. 15, 37).

[102] P. S. Rivadeneira et al. “Mathematical Modeling of HIV Dynamics After Antiretroviral Therapy Initiation: A Review”. In: *BioResearch Open Access* 3.5 (Oct. 1, 2014), pp. 233–241 (cit. on p. 8).

[103] C. Robert. *Le choix bayésien: Principes et pratique*. Statistique et probabilités appliquées. Paris: Springer-Verlag, 2006 (cit. on p. 23).

[104] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. 2nd ed. Springer Texts in Statistics. New York: Springer-Verlag, 2004 (cit. on pp. 23, 25, 26, 54).

[105] G. O. Roberts and O. Stramer. “On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm”. In: *Biometrika* 88.3 (Oct. 1, 2001), pp. 603–621 (cit. on p. 25).

[106] R. Ross. *The Prevention Of Malaria*. John Murray;London, 1911 (cit. on p. v).

[107] Royaume du Maroc ministère de la santé. *Mise en oeuvre de la déclaration politique sur le VIH/sida, Rapport National 2014*. 2014 (cit. on pp. 61, 81).

[108] Royaume du Maroc ministère de la santé. *Mise en œuvre de la déclaration politique sur le VIH/sida. Rapport National 2015*. 2015 (cit. on pp. 61, 81).

[109] A. Sani, D. P. Kroese, and P. K. Pollett. “Stochastic models for the spread of HIV in a mobile heterosexual population”. In: *Mathematical Biosciences* 208.1 (July 1, 2007), pp. 98–124 (cit. on pp. 30, 32).

[110] M. J. Schervish. *Theory of Statistics*. Springer Series in Statistics. New York: Springer-Verlag, 1995 (cit. on p. 22).

[111] U. Seifert. “Stochastic thermodynamics: principles and perspectives”. In: *The European Physical Journal B* 64.3 (Aug. 1, 2008), pp. 423–431 (cit. on pp. 15, 37).

[112] C. I. Siettos and L. Russo. “Mathematical modeling of infectious disease dynamics”. In: *Virulence* 4.4 (May 15, 2013), pp. 295–306 (cit. on p. 1).

[113] P. Sjöberg, P. Lötstedt, and J. Elf. “Fokker–Planck approximation of the master equation in molecular biology”. In: *Computing and Visualization in Science* 12.1 (Jan. 1, 2009), pp. 37–50 (cit. on p. 15).

[114] A. F. M. Smith and G. O. Roberts. “Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 55.1 (1993), pp. 3–23 (cit. on p. 25).

[115] M Sørensen and M Uchida. “Small-diffusion asymptotics for discretely sampled stochastic differential equations”. In: *Bernoulli* 9.6 (2003), pp. 1051–1069 (cit. on pp. 39–41, 73).

[116] G. Streftaris and G. J. Gibson. “Bayesian inference for stochastic epidemics in closed populations”. In: *Statistical Modelling* 4.1 (Apr. 1, 2004), pp. 63–75 (cit. on p. 26).

[117] M. Strumik and W. M. Macek. “Statistical analysis of transfer of fluctuations in solar wind turbulence”. In: *Nonlinear Processes in Geophysics* 15.4 (July 28, 2008), pp. 607–613 (cit. on p. 37).

[118] L. M. Surhone, M. T. Tennoe, and S. F. Henssonow. *Riemann-Stieltjes Integral*. VDM Publishing, Aug. 2010. 106 pp. (cit. on p. 14).

[119] W. Y. Tan and H. Wu. *Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention*. World Scientific, 2005. 610 pp. (cit. on pp. 8, 29).

[120] W. Y. Tan and Z. Xiang. "A state space model for the HIV epidemic in homosexual populations and some applications". In: *Mathematical Biosciences* 152.1 (Aug. 15, 1998), pp. 29–61 (cit. on p. 29).

[121] W. Y. Tan and X. Zhu. "A stochastic model of the HIV epidemic for heterosexual transmission involving married couples and prostitutes: II. The chain multinomial model of the HIV epidemic". In: *Mathematical and Computer Modelling* 26.11 (Dec. 1, 1997), pp. 17–92 (cit. on p. 29).

[122] T. Tian et al. "Stochastic delay differential equations for genetic regulatory networks". In: *Journal of Computational and Applied Mathematics*. Special issue on evolutionary problems 205.2 (Aug. 15, 2007), pp. 696–707 (cit. on pp. 15, 37).

[123] M. Tibayrenc. *Encyclopedia of Infectious Diseases: Modern Methodologies*. John Wiley & Sons, July 31, 2007. 807 pp. (cit. on p. 5).

[124] H. C. Tuckwell and R. J. Williams. "Some properties of a simple stochastic epidemic model of SIR type". In: *Mathematical Biosciences* 208.1 (July 2007), pp. 76–97 (cit. on p. 15).

[125] M. Uchida. "Estimation for Discretely Observed Small Diffusions Based on Approximate Martingale Estimating Functions". In: *Scandinavian Journal of Statistics* 31.4 (2004), pp. 553–566 (cit. on pp. 40, 42).

[126] UNAIDS. *UNAIDS data 2018*. July 26, 2018. URL: http://www.unaids.org/en/resources/documents/2018/unaids-data-2018 (visited on 01/23/2019) (cit. on pp. 1, 29).

[127] J. B. Walsh. "A Stochastic Model of Neural Response". In: *Advances in Applied Probability* 13.2 (1981), pp. 231–281 (cit. on p. 15).

[128] N. Wiener. "Differential-Space". In: *Journal of Mathematics and Physics* 2.1 (Oct. 1923), pp. 131–174 (cit. on p. 13).

[129] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, Apr. 18, 2006 (cit. on pp. 25, 37).

[130] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. 2nd ed. Boca Raton: CRC Press, Dec. 14, 2011. 363 pp. (cit. on p. 2).

[131] N. Yoshida. "Estimation for diffusion processes from discrete observation". In: *Journal of Multivariate Analysis* 41.2 (May 1992), pp. 220–242 (cit. on p. 39).