Sultan Moulay Slimane University
Faculty of Sciences and Technics
Center for Doctoral Studies in Mathematics and Applied Physics

Doctoral Thesis

Bayesian Modelling of Regime Switching Models for Disease Progression

---

Presented by:
EL HOUCINE HIBBAH

25 October 2019

Laboratory of Mathematics and Applications
Option : Statistics and Probabilities

The dissertation is approved by the following members of the final oral committee:

Mr. Khalid HILAL FST, Sultan Moulay Slimane University, Béni Mellal	**Professor**	**President**
Mr. Abdelali KAAOUACHI EST, Mohammed Premier University, Ouajda	**Professor**	**Rapporteur**
Mr. Abdelkrim MERBOUHA FST, Sultan Moulay Slimane University, Béni Mellal	**Professor**	**Rapporteur**
Mr. Tewfik KERNANE Sciences and Technology University, USTHB, Algeria	**Professor**	**Rapporteur**
Mr. Hamid EL MAROUFY FST, Sultan Moulay Slimane University, Béni Mellal	**Hability Professor**	**Supervisor**
Mr. Abdelkader STOUTI FST, Sultan Moulay Slimane University, Béni Mellal	**Professor**	**Co-supervisor**
Mr. Abelaziz QAFFOU EST, Sultan Mouly Slimane University, Béni Mellal	**Assistant Professor**	**Invited**

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible .

First and foremost I would like to thank my supervisor, Professor Hamid El Maroufy for giving me the opportunity to work on this challenging and interesting project. He has always made himself available for answering any of my questions. There is no occasion that he has not been humble and modest to help me and advise me in any tiny step during the last years of this thesis, even though his many responsibilities and duties.

Beside my supervisor, I would like to thank the rest of the thesis committee: Professor, Tewefik Kernane, Professor, Merbouha Abdelkarim, Professor, Kaaouachi Abdelali for the honor of reporting this work. My sincere thanks goes to Professor Khalid Hilal for the honor of accepting to chair this committee.

I am indebted to Professor Abdelkader Stouti for offering me the opportunity to start this research project as a co-supervisor.

I am thankful to Professor Abdelaziz Qaffou to be invited as a member of oral examination.

My colleagues at the laboratory of Mathematics and Applications have enriched my graduate life in many ways especially with their theoretical ideas and computational expertises and research ideas.

Throughout my life I have been fortunate to have had excellent teachers at every stage of my formal education. I can not mention all of them, but it was because of their example and encouragement that I have got to this stage of my life.

I owe my deepest thanks to my family - my mother, father, and brothers who have always stood by me throughout my career.

Lastly, thank you all and thank God!

Preface

Markovian processes are useful in describing changes in patient health condition over time (see for example, Bartolomeo *et al.* [10], Newton *et al.* [105]) . Nevertheless, we can not observe the progression of the disease directly; because *the patient health condition is hidden or unrevealed* and we must appeal for observations such as *bio-markers'observations* to achieve our aim. The necessity for the markers is justified by their correlation to the disease development and to their possibility to be measured and observed. In this thesis, the usefulness of the *bio-markers'observations* is shown here by a version of the hidden Markov models: the Markov switching models (MSM); especially the*Markov switching autoregressive models*, where each observation conditional on the current hidden state will depend on the previous observation according to an autoregressive process of first order. In fact, since their introduction by Hamilton [67]; regime switching models do not stop from shining, due to their double role in capturing the phenomenon behaviour in long term; as well as in capturing its possible short term switch. Hence, we get models that are well suited for state dependent dynamics.

Our aim is first to specify and describe the *Markov switching auto-regressive model* for disease progression in both the discrete and the continuous time (since disease observations are obtained at irregularly different time points), and to estimate the parameters in a Bayesian context with the corresponding simulations and interpretations. Then, we address the stochastic version of the MSM: *the hybrid switching diffusion* (HSD) process.

This thesis consists in describing and estimating the parameters of the MSM and the HSD processes using the *Bayesian Markov chain Monte Carlo (MCMC)* methods. The MCMC algorithm for each model will be provided in details. The theoretical results are illustrated by numerical simulations. A real application is provided for the HSD.

This dissertation is composed of four chapters, a conclusion and perspectives; the first chapter is an introduction to the thesis. The second chapter describes the necessary background about the Bayesian inference and MCMC methods; as well as it contains a preliminary on stochastic differential equations. The third chapter constitutes our contribution to the Bayesian estimation of the MSM in both the discrete and the continuous time. While the fourth chapter is concerned with the MCMC methods for the HSD process.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
ARCH	Autoregressive Conditional Heteroscedasticity model
ARHMM	Autoregressive Hidden Markov Model
ALSC	Lalla Salma Foundation for Cancer Prevention and Treatment
BIC	Bayesian information criterion
CA	Cancer Antigen.
CD4	Cluster of Differentiation 4.
CEA	Carcino Embryonic Antigen.
COPD	the Chronic Obstructive Pulmonary Disease
CTMSAR	Continuous Time Markov Switching Autoregressive model
CTMSAR(1)	Continuous Time Markov Switching Autoregressive model of first order
DCMM	Double chain Monte Carlo
DEM	Differential equation model
DLCT	Danish Lung Cancer Trial
EM	Expectation-Maximization
ESS	Effective Sample Size
FEV1	The volume of air that can be forced out in one second after taking a deep breath
FFBS	Forward Filtering Backward Smoothing
GARCH	Generalized Autoregressive Conditional Heteroscedasticity model
HER	Human Epidermal Receptor.
HIV	Human Immunodeficiency Viruses.
HMM	Hidden Markov Model
HSD	Hybrid Switching Diffusion
MAR(1)HMM	Multivariate Autoregressive Hidden Markov Model of first order
MCMC	Markov Chain Monte Carlo
MH	Metropolis Hastings
MHA	Metropolis Hastings Algorithm
ML	Maximum Likelihood
MSAR	Markov Switching Autoregressive Model
MSM	Markov Switching Model
PMCMC	Particle Markov Chain Monte Carlo
PSRF	Potential Scale Reduction Factor
RSM	Regime Switching Model
RSAR	Regime Switching Autoregressive Model
RWMHA	Random Walk Metropolis Hastings Algorithm
SDE	Stochastic Diffusion equation
SDP	Switching Diffusion Process
SDRSD	State Dependent Regime Switching Diffusion
TNM	Tumor, node, metastasis

List of Notations

Notation	Meaning
$(\Omega, \mathscr{F}, P)$	A probability space.
$\xrightarrow{d}$	Convergence in distribution.
$\xrightarrow{p}$	Convergence in probability.
$\xrightarrow{a.s}$	Almost sure convergence.
exp	Exponential or exponential matrix.
$\mathscr{E}(\mu)$	Exponential distribution with parameter μ.
$\mathscr{N}(\mu, \sigma^2)$	Normal density with mean μ and variance σ^2.
$\mathscr{LN}(\mu, \sigma^2)$	Log-normal density with parameters μ and σ^2.
$\mathscr{B}(\alpha, \beta)$	Beta density with parameters α and β.
$\mathscr{G}(\alpha, \beta)$	Gamma density with parameters α and β.
$\mathscr{IG}(\alpha, \beta)$	Inverse Gamma density with parameters α and β.
$\mathscr{D}(u_1, \ldots, u_n)$	Dirichlet density with parameters $(u_1, \ldots, u_n)$.
Π	The probability transition matrix.
Q	The transition intensity matrix.
χ	The indicator function.
P	The distribution or the posterior distribution.
$\propto$	Proportional.
$\mathbb{R}$	Real number set.
$\mathbb{N}$	Integer number set.
$\mathbb{E}$	Expectation.
$\mathbb{C}ov$	Covariance.
$\mathbb{V}ar$	Variance.
a	Number of discrete Markov states.
Θ	Parameter space.
N	The number of subjects.
n_i	The number of individual observations by subject.
Tr	Trace matrix.
$\mathscr{U}$	The uniform distribution.

Synthèse

Les modèles de Markov sont largement utilisés dans les sciences de la santé; où le suivi de la progression des maladies est d'une grande importance pour les médecins et les patients. Ils offrent une manière très efficace de décrire le comportement de transition à travers les états d'un patient, dans le but de comparer les traitements pour différents médicaments et de choisir le traitement le plus convenable afin d'améliorer l'état sanitaire du malade (voir, Chunling et Tsokos [36] et les références y figurant).

Malheureusement, on ne peut observer l'état du malade directement. En fait l'état sanitaire du malade est souvent caché. Par conséquent, nous faisons recours à des éléments intermédiaires; qui one une forte corrélation avec les états du malade. Ces éléments sont quantifiables et peuvent être mesurés; ils sont présents dans le corps ou le sang des malades, ils sont appelés marqueurs biologiques. Ces marqueurs ont prouvés qui'ils ont une forte dépendance avec l'évolution de l'état du malade; particulièrement pour les deux important maladies sujets de cette thèse: le cancer du sein, et la maladie pulmonaire obstructive chronique (COPD). Par exemple, pour le cancer du sein; les antigènes tumoraux 15-3 ($CA15-3$) et l'antigène carcino-embryonnaire (CEA) ont été associés à l'évolution du cancer du sein métastatique, d'un stade à l'autre (voir, Laessig et *al* [90]).

Notre motivation provient alors du fait que les stades de la maladie sont cachés, ils seront donc modélisés par des états Markoviens de premier ordre. Les biomarqueurs sont observés; ils seront modélisés par des observations continues; telles que les observations autorégressives, les observations ARMA et les observations par diffusion (par exemple, voir Kim et Kim [81]). Ainsi, la modélisation sera effectuée par des versions du modèle de Markov caché (HMM) tels que les modèles avec changement de régimes Markoviens, ou leurs variantes stochastiques; les processus de diffusion avec changement de régimes Markoviens.

Dans un premier temps, nous nous intéresserons dans cette thèse au modèle autorégressive avec changement de régimes Markoviens de premier ordre (MSAR(1)). Semblable aux modèles HMM, le MSAR(1) a des états cachés Markoviens de premier ordre. Mais contrairement aux modèles HMM, où les observations sont indépendantes; nous verrons qu'étant donné l'état caché actuel, l'observation actuelle dépend de la précédente via un processus autorégressif de premier ordre. Il convient de souligner que notre modélisation concerne de nombreuses personnes avec un nombre de temps de suivi différent d'un patient à l'autre.

Dans cette thèse, nous commencerons par discuter le cas des temps discrets du modèle MSAR(1); appelé model de Markov caché autorégressif multi-varié de premier ordre (MAR(1)HMM). Ce modèle est similaire au modèle multi-varié à double chaînes de Markov (DCMM); où les observations sont considérées comme discrètes et Markoviennes. Cependant, dans notre cas, les observations sont continues et autorégressives de premier ordre.

Nous aborderons ensuite la version des temps continues du MSAR(1); le modèle autorégressif à temps continu de premier ordre (CTMSAR(1)). Le cadre temporel continu est mieux convenable à l'analyse des phénomènes, car il prend en compte les intervalles de temps irréguliers. Bien que nous ayons des paramètres autorégressifs similaires pour le deux cas discret et continu, nous nous concentrons sur les probabilités de transition du processus de Markov dans le cas discret et sur les intensités de transition pour le cas continu.

Cependant, la plupart du temps nous devons nous rapprocher la densité du processus avec changement de régimes Markoviens (MSM). De plus, nous aimerions ajouter plus de dynamique stochastique au processus d'observations, ou simplement avoir des petits intervalles de temps nécessaires pour toute approximation. Ainsi, nous obtenons la version stochastique du MSM; qui est le processus de diffusion hybride avec changement de régimes (HSD).

Dans le processus HSD, nous avons deux composantes: les observations qui sont supposées provenir d'un processus de diffusion qui dépend du processus Markovien; et le processus avec changement de régimes qui'est régi par une matrice Markovienne d'intensités, supposée être dépendente du processus de diffusion, conservatrice et irréductible.

Concernant les méthodes d'estimation, on peut utiliser les méthodes classiques telles que la méthode du maximum de vraisemblance (ML) où les méthodes d'espérance-maximisation (EM) ; en raison de leur efficacité. Néanmoins, pour les modèles MSM, la fonction de vraisemblance est insoluble la plupart du temps, en plus d'un nombre de paramètres tres élevé, ça rend l'inférence peu pratique. Nous utilisons donc les méthodes Bayésiennes de Monte Carlo avec chaîne de Markov (MCMC) ; qui supposent que la fonction de vraisemblance ne soit connue que proportionnellement. De plus pour les modèles MSM, les états cachés sont manquants et sont ainsi considérés comme des quantités aléatoires; et MCMC permet d'augmenter la fonction de vraisemblance avec les états manquants grâce à l'outil d'augmentation de données (par exemple, voir Hobert [72]). Enfin, notre choix pour les méthodes MCMC découle du fait que l'inférence Bayésienne permet d'introduire des connaissances à priori dans la fonction de vraisemblance.

Dans le contexte MCMC, l'objectif est de construire une distribution Markovienne stationnaire pour les paramètres et les états cachés. En fait, l'algorithme procède en alternant la simulation des états cachés et le calcul des paramètres. Les états cachés sont calculés conjointement à l'aide d'un algorithme progressif-rétrogressif (FFBS). Pour bien comprendre cet algorithme, on peut se référer à Chib [31]. Quant à l'inférence des paramètres, nous utilisons soit l'échantillonnage de Gibbs lorsqu'on obtient une densité à postériori connue à partir de laquelle on peut échantillonner; sinon, nous faisons appel à l'algorithme Metropolis Hastings (MHA). Néanmoins, l'estimation Bayésienne pour le HSD processus est un autre défi à relever ; car la plupart du temps, nous devons approximer la densité de diffusion. En fait, nous sommes généralement confrontés à des données à basse fréquence. Pour surmonter ce problème, nous devons introduire des

données virtuelles ou latentes entre les observations successives via le mécanisme d'imputation des données (par exemple, voir Elerian et al [46]).

Par conséquent, notre algorithme MCMC alternera entre la simulation des données imputées, les états cachés et le calcul de l'estimation des paramètres. Bien que la simulation des paramètres et des états cachés soit semblable au cas de MSAR(1), la nouveauté ici réside dans le fait que pour les données imputées, nous optons pour l'algorithme d'imputation temporelle aléatoire telle qu'il est décrit par Blackwell *et al.* [16] au lieu de l'algorithme d'imputation temporelle fixe. De plus, dans notre cas, la matrice d'intensités est mise à jour après l'actualisation des observations par les nouvelles données imputées, car la matrice d'intensités dépend du processus de diffusion.

Enfin, nous testons notre algorithme MCMC par des études de simulation pour le cas du MSM, tandis que le processus HSD est illustré par une application à la maladie COPD.

Le but de cette thèse concerne l'utilisation des méthodes MCMC pour estimer les processus MSM ou leur variante stochastique; les modèles HSD, pour l'étude de la progression des maladies.

Au chapitre 2, nous passons en revue les principes de base de l'inférence Bayésienne et des règles de convergence des methodes MCMC; avec une description des principaux algorithmes considérés tels que les algorithmes Gibbs Sampler et MHA. Nous terminerons par un préliminaire sur les équations différentielles stochastiques.

Au chapitre 3, et après un exposé préalable sur les modèles HMM et la spécificité des modèles MSAR(1), nous décrivons brièvement le problème d'idenfication des paramètres rencontré dans les modèles MSM. Ensuite, nous décrirons notre MSM à la fois dans le cadre du temps discret et du temps continu. Plus tard, nous fournissons les détails de l'inférence Bayésienne pour les paramètres et les états de Markov cachés. Dans ce chapitre, nous verrons combien il est important d'évaluer les probabilités de transition dans le cas des temps discrets, tandis que nous nous concentrerons sur les intensités de transition dans le cadre de temps continu. Ensuite, l'inférence est validée par une simulation.

Au chapitre 4, nous nous intéressons aux méthodes MCMC pour les processus HSD. Le défi ici est que nous imputons des données latentes entre des observations successives pour résoudre le problème des intervalles de temps grands et pour permettre une approximation aisée de la densité de diffusion. Nous verrons comment l'algorithme MCMC alterne entre la computation des paramètres, les états cachés et les données imputées. Ensuite, nous donnons un cas de simulation avec une application à la progression de la maladie COPD.
Dans le dernier chapitre, nous fournissons une conclusion et des perspectives.

Enfin, dans les annexes, nous passons en revue quelques propriétés de base de convergence des variables aléatoires dans l'annexe A; en annexe B, nous fournissons quelques preuves importantes relatives à la condition de la balance détaillée nécessaire à la convergence de l'algorithme MCMC. L'annexe C fournit les détails concernant la propriété de Chapman Kolmogorov pour les modèles Markoviens, et

en annexe D, nous passons en revue la formulation de la fonction de vraisemblance pour les processus de diffusion avec changement de régime indépendant, ainsi que l'algorithme progressif-rétrogressif (FFBS) de ce processus stochastique. En annexe E, nous présentons la forme numérique de Peng: pour l'exponentielle de la matrice d'intensités de transition. Annexe F donne un extrait des observations du marqueur FEV1 concernant la progression de la maladie COPD. Enfin, les annexes G et H fournissent respectivement les programmes de simulation Matlab pour les processus MAR(1)HMM et HSD.

Chapter 1

Introduction

Summary

In this introduction, we provide in section 1 the motivations behind our thesis; especially the need for modelling and characterizing the disease progression. In section 2, we go through the models used such as the Markov switching models or their stochastic variants the switching diffusion processes. While, in section 3, we discuss the reason for choosing the Bayesian MCMC approach as a method of estimation. Finally, we give an overview of the thesis.

1.1 Motivations

The aim of this thesis is to develop some kinds of latent variable models; so as to address the problems encountered in slow developed disease progression such as controlling and monitoring the progression of a disease in a patient; or addressing the response of patients to specific treatment, therapy or simply due to natural development. Some of the diseases that are slow developed and cause many difficulties are: the breast cancer and the chronic obstructive pulmonary disease (COPD).

In fact, breast cancer is a major global public health problem. It is the most commonly diagnosed cancer and is one of the leading causes of death in women worldwide. In arabic countries, breast cancer has become the most common cancer among women (see Donnelly *et al.* [41]), with one million new cases in the world constituting 18% of all female cancer cases (see for example McPherson *et al.* [101]).

In Morocco, the interest of studying breast cancer has clearly increased during the last decade; characterized by the creation of Lalla Salma foundation (ALSC). The reason behind this interest is that breast cancer represents 39.9% of all cancer women with a median of 45 years; making south Mediterranean breast cancer patients younger than those of French (as reported by Slaoui *et al.* [130]). Hence, this magnitude of the epidemics has lead us to look for modelling the disease progression, since the major challenge faced by medical doctors, biologists and scientists is to unravel the stage of the cancer during its development. In fact, if we determine

concisely the hidden state of the disease during the follow up, we can address the response of the patients to the treatment or to the therapy. Hence, comes the necessity to find some easy observed parameters that are correlated to the disease and would help us to follow the patient at any time. These quantities in breast cancer are called bio-markers; they are proteins or serums in the body of patients and are observable and could be measured.

Indeed, many studies have shown the correlation between the disease stage and the quantity of the bio-markers, among them: Samy *et al.* [122] has found that the levels of the bio-markers CEA[1] and $CA15-3$[2] increased significantly in stage II compared with stage I. In [90], the markers CEA and $CA\,15-3$ have been shown to be useful tools in the follow-up and the therapy monitoring for breast cancer patients; where a clear correlation is observed between CEA and/or $CA\,15-3$ and breast cancer progression. Also, the combination of three markers ($CA\,15-3$, CEA, $Her2$[3]) has enhanced the sensitivity of detection of advanced metastatic breast cancer (see Pedersen *et al.* [109]).

While to the COPD disease; a respiratory condition characterized by airflow limitation, it has become a major challenge; because it is the third leading cause of mortality in the world (see Lozano *et al.* [98]); with a prevalence rate of 3.5% in North africa, and a serious public health problem in Morocco as shown in a study by El-Rhazi *et al.* [48]. By the way, doctors and scientists need markers to manage treatments as well as to follow the patients'health states. One of the famous markers in this field is FEV1 (the volume of air that can be forced out in one second after taking a deep breath). This marker is widely used in COPD studies as it is associated with mortality in COPD and it is also considered an important variable to evaluate the severity of COPD (see Almagro *et al.* [5] and the references therein).

Luckily that we have a data set about FEV1 observations for COPD progression to apply our methods on. Nevertheless, it was difficult for us to get data sets for bio-markers in breast cancer, and consequently we will limit ourselves to the simulation studies for the breast cancer development.

1.2 Models

Driven by the unraveled nature of the disease stage that are correlated to the bio-markers'observations, modelling is carried out through the family of the hidden Markov models (HMM) or their extensions (Markov switching autoregressive models, Markov switching GARCH models, switching regression models). The modelling could be brought too by the stochastic regime switching diffusion models. Indeed, the HMM have been proven to be useful tools in many fields ; including speech processing, biology and disease progression (see for example, Benmiloud and

[1] Carcinoembryonic antigen (CEA) is a protein found in many types of cells but associated with tumors and the developing fetus.

[2] Protein found on epithelial cells that is part of a larger protein called $MUC1$. Measuring the amount of $CA\,15-3$ in the blood may be useful for the follow-up of breast cancer.

[3] A protein involved in normal cell growth and that may cause cancer cells to grow more quickly.

Piczunski [12], Boys and Handerson [19], Chantel *et al.* [28]). Similar to regular HMM, we will use discrete Markovian hidden states in our modelling. But, by contrast to HMM, the independence constraint between successive observations is relaxed; since we use extensions to incorporate the dependence between the successive observations, suggesting an autoregressive dependence of first order between continuous observations. An illustration of the model can be found in [50].

These kinds of models can be even generalized to regime switching ARMA models such as in [81]. Clearly, the first model considered is an extension of the multivariate double chain Markov model (DCMM) developed by Fitzpatrick and Marchev [52]; in which the relationship between the successive observations is taken into account, but the DCMM supposes discrete values for the observations and not continuous ones as we will consider in our autoregressive hidden Markov model of first order (AR(1)HMM); one of the process that belongs to the family of Markov switching models (MSM). In fact, the coexistence of the autoregressive mechanism and the regime switching elements in MSM adds more dynamics to the observed process, because it produces switches in the observations that a time series can not handle. Thus, MSM's have been extremely popular in many areas such as: stock market returns (example, Pagan and Schwert [107]), business cycle (example, Chauvet and Piger [29], monetary policy (example, Rubio-Ramirez *et al.* [119]), or in studying the changing dependency structures of multivariate time series in health management (example, Liu *et al.* [96]).

The choice for MSM's raises too from their ability to draw inference about the transition probabilities between the hidden states of the disease based on the observations. At the first glance an AR(1)HMM will be considered in discrete time or for observations with equidistant time intervals.
Later on, our focus will be on a continuous time Markov switching autoregressive model (CTMSAR); a more appropriate framework for irregularly spaced observations in patients'analysis than the discrete time steps. Unlike discrete time situation, the focus in continuous time would be of estimating the transition intensity rates'matrix instead of the transition probabilities'matrix. Consequently the calculation of the mean sojourn time in a disease stage will be very easy.

While MSM's fit many phenomena analysis, what if our disease is subject to thermal fluctuations or if its rate of change is uncertain? In this case we include a random term to our processes. Hence, our work is deepened by adopting a modelling with a regime switching diffusion process (SDP). In fact, diffusion models are useful tools for quantifying the dynamics of continuous time phenomena. Moreover, they include some randomness, dynamics and unpredictable factors into modelling. Also, stochastic differential equations can deal with low frequency data or when only a small number of observations is available. From another side, diffusion processes have been a focus of research in the past years, and they can describe any complex situation by dealing effectively with randomness. Their applications range from genetics, finance, to biology (see for example, Chen *et al.* [30], Eraker [49], Golightly and Wilkinson [62]). By adding the regime shifts to the diffusion process, we get the SDP that can capture additional variability in the dynamics for disease

behaviour.

While one can consider the switching states to be independent from the diffusion process as in [34, 65], we will opt for the dependency of the latent states on the observations as it is reported in [27, 23]; to obtain the so-called hybrid switching diffusion process (HSD) or the state dependent regime switching diffusion process (SDRSD). This mutual dependence between the regime switching component and the diffusion component is new in application, modelling as well as in statistical inference.

1.3 Methods

The maximum likelihood (ML) is usually the preferred method of estimation because of its efficiency gain. However, the likelihood is hard to evaluate analytically especially in MSM's or their analogues the SDP's; where the likelihood can be known only proportionally. In addition, with the high dimensional parameter space in MSM, sampling is impractical. Consequently, we adopt Markov chain Monte Carlo (MCMC) methods as tools for computation; since they can deal with a large number of parameters and can solve any related problem when the likelihood needs to be known up to a constant.

From another side, RSM's have the extra new parameters of the unknown hidden states to be evaluated, and MCMC methods appeal for the data augmentation mechanism for this matter (see for example, Tanner and Wong [134]); where the likelihood is augmented with the hidden states as extra unknown parameters. The MCMC algorithm renders the sampling very easy because it uses the posterior densities to alternate between sampling the hidden states and sampling the parameters. Furthermore, in the case of diffusion processes, the observations are only available at discrete and sparse time intervals. Consequently, it is necessary to impute virtual or latent data between successive observations; in order to get small time intervals and hence we can well approximate the diffusion transition density. Again data augmentation by the new imputed observations is inevitable; where MCMC methods have been proven to be very useful (see for example, Eraker [49], Golightly and Wilkinson [63]).

1.4 An overview of the thesis

The aim of this thesis is concerned with MCMC methods for estimating the MSM processes or the stochastic SDP for disease progression; where:

In chapter 2, we go through the basic principles of the Bayesian inference and MCMC convergence rules; with the description of the main algorithms considered such as the Gibbs Sampler and the MHA, and we will finish by a preliminary on stochastic differential equations.

In chapter 3, and after a preliminary on HMM and discussing the difference between the Markov switching autoregressive models and other Markov switching

models, we see how we can handle the label switching problem in MSM. Then, we will describe our MSM in both the discrete time and the continuous time framework. Later on, we provide the details of the Bayesian inference for both the parameters and the hidden Markov states; where the hidden states would be evaluated in block via a FFBS algorithm. In this chapter, we will see how it is important to evaluate the transition probabilities in the discrete time case, while we will focus on the transition intensities in the continuous time framework. Then, the inference is validated by a simulation study.

In chapter 4, our focus will be on MCMC methods for the HSD processes. The challenge here is that we impute latent data between successive observations to overcome the problem of dispersed time intervals. By the way the diffusion density can be easily approximated. We will see how the MCMC algorithm alternates between the computation of the parameters, the hidden states and the imputed data. The exception in our work, is that we opt for the random time data imputation approach; where the times of imputation and the imputed data proposed are accepted or rejected using a MHA. Also, in this chapter we will see how the parameters of the state dependent transition intensity matrix are computed. Then, we give a simulation case with an application to COPD disease progression.

In the last chapter, we provide a conclusion and perspectives.

Finally, in the appendices, we go through some basic properties of convergence of the random variables in appendix A; in appendix B, we provide some important proofs relative to the detailed balance condition necessary to the convergence of the MCMC algorithm. Appendix C gives the details of the Chapman Kolmogorov property for Markov models; and in the appendix D, we review the likelihood formulation for state independent regime switching diffusion process as well as the FFBS algorithm of this stochastic process. While in appendix E, we present Peng numerical form for the exponential of the transition intensity matrix. Appendix F gives an extract about FEV1 observations in COPD disease progression. Lastly, appendices G and H provide respectively Matlab simulation programs for the MAR(1)HMM and the HSD processes.

Chapter 2

Preliminaries: Bayesian Inference, Markov Chain Monte Carlo, and Stochastic Differential Equations

Summary

In this chapter, and after a brief history about the MCMC methods in section 1; we introduce briefly some necessary concepts, principles in the Bayesian MCMC inference in section 2 and section 3. Then, we define and discuss the properties of the two main important algorithms in the Bayesin MCMC context; the MHA and the Gibbs sampler in section 4 and section 5. In later sections, an overview is given about some important concepts in the MCMC algorithm simulation such as: The burn-in, the starting point determination for the MCMC algorithm, as well as the convergence assessment and precautions raising in MCMC simulations. Finally, we give some notions about stochastic differential equations; especially the stochastic diffusion equations.

2.1 History of Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) are technics that simulate a Markov chain whose invariant states follow a given (target) probability in a very high dimensional state space. It is particularly useful for the evaluation of the posterior distributions in complex Bayesian models and for the evaluation of numerical integration.

MCMC methods were present for almost as long as Monte Carlo technics exist in practice; but without impact or real usage until late 1980's, due to the lack of computing machinery or the impossibility to use them in practice. MCMC algorithms date back to the same time as the development of regular Monte Carlo (MC) methods with the stem of the Metropolis algorithm in the early 1950's as the first MCMC algorithm called the Metropolis algorithm; published by Metropolis *el al.* [102] in 1953 to be the first MCMC algorithm. This algorithm was later generalized by Hastings (1970) to overcome the curse of dimensionality met by regular Monte Carlo methods.

Nevethless the real revolution of the MCMC methods was highlighted by the work of Gelfand and Smith [57] to be the starting point for an intensive use of MCMC methods in Bayesian methods and statistical computing; through the use of computing via the main algorithms of the Gibbs sampler and the Metropolis Hastings algorithm.

Meanwhile, Tanner and Wong [134] address the concept of data augmentation though not having good theoretical argument to be well accepted in MCMC applications. Then comes the paper of Tierney [135]; who brought all the theories and laid out all of the assumptions needed to analyze the Markov chains and then developed their properties, in particular: the convergence of ergodic averages and the central limit theorems. Even after this revolution, MCMC methods continue to impress and provide meaningful solutions, and thus continue to expand the applications of statistics in most complex models at little cost. Robert and Casella [115] give a short but consistent history about MCMC methods.

2.2 Bayesian inference

MCMC is widely related to the Bayesian inference. In the Bayesian inference, the parameters are considered random quantities; where we are interested in the posterior densities of the paremters after the observations. However, the Bayesian approach allows us to include any prior knowledge or believe in the posterior distribution of unknown parameters given the observations (see for example, Congdon [37, Chapter. 1]).

2.2.1 Bayes rule

Bayes theorem provides the probability about a parameter of interest θ given the observation y as follows:

$$\begin{aligned} p(\theta|y) &= p(y|\theta) \times \pi(\theta)|p(y) \\ &\propto p(y|\theta) \times \pi(\theta). \end{aligned}$$

Here the normalizing constant $p(y)$ is omitted, and the posterior of θ will be evaluated up to a constant. The first ingredient $p(y/\theta)$ gives the likelihood function of the observations given the parameters . The second ingredient $\pi(\theta)$ is called the prior density as it contains the probability distribution of θ before the observation of y; and the inference should be based on the probability distribution of θ after observing the value of y: $p(\theta/y)$ called the posterior distribution of θ given y, in direct opposition to the prior distribution $\pi(\theta)$ (the distribution before observation) . We should clarify that the dimension of θ could be high, that is why we need MCMC methods to explore the posterior of θ.

2.2.2 Bayesian inference versus frequentist approach

There are many difference between Bayesian and frequentist inference:

In the frequentist approach, the parameters are fixed and we can not make probabilistic statements about them. The outcome is usually a point estimate with standard errors, and the parameters estimates are based on data. However, sometimes the inference is intractable for complicated models due to the unavailability of the likelihood.

On the other hand, in the Bayesian approach, the parameters are random variables and the uncertainty is expressed in probability statements or distributions. Also, we use prior knowledge to locate a value of the parameter space to update our degree of belief, and to generate samples from the posterior distributions to estimate the quantities of interest.

The Bayesian analysis combines prior information with data and can directly estimate any function of parameters or any quantity of interest, and is very useful when the likelihood is only available up to constant.

Briefly, when we have some knowledge about the phenomenon under study, it is recommended to incorporate this preexisting information in the analysis; and the

Bayesian approach includes this kind of information through the prior distribution (for more details see for example Lin [94], Sharma [125], Baio [8], Samaniego [120] and the references therein).

2.2.3 Prior distributions[1]

The prior distribution expresses one's uncertainty about the unknown parameter before the observation is taken into account. It attributes uncertainty rather than randomness to the uncertain quantity. It is often the purely subjective assessment of an experienced experts. Hence, we could make many choices concerning the priors; we may have a *conjugate prior*: if the posterior distribution is in the same family as the prior distribution. Here the prior and the posterior distribution are called conjugate distributions; for example: the Binomial and the Beta distributions are conjugate.

But sometimes the experimenter has only a few clue about the data generating process. In these circumstances, it might be preferable to use a *vague or non-informative prior*; which does not favour any of the possible values that the parameter can take on, called also flat, dispersed and it translates the ignorance about the parameter before observing the evidence; a typical example could be a uniform prior distribution.

Meanwhile, when the information gathered from the previous studies, past experiences or expert opinions can be combined in the prior through the specification of the hyper-parameters of the prior distribution; we called the prior an *informative prior*. This informative prior can dominate the likelihood and has a clear impact on the likelihood. Another kind of prior is the *improper prior*, where a prior $P(.)$ is said to be improper for the parameter θ on the parameter sapce Θ, if its measure dP is with an infinite mass: $\int_\Theta dP = \infty$. If an improper prior leads to an improper posterior, the inference based on the posterior is invalid.

The priors are chosen in this thesis depending the phenomenon in hand. For example, when the parameter values are real we choose a normal distribution, but when the parameter is positive such as the variance we adopt a Gamma prior. Another example is the Dirichlet prior chosen when we have to deal with discrete values. Some of the prior distributions that are adopted in this thesis are:
Normal distribution: a real variable X is said to follow a Gaussian probability density function g with mean μ and variance σ^2: $X \sim \mathcal{N}(\mu, \sigma^2)$ if:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2},\ x \in \mathbb{R}.$$

Beta distribution: a variable X with values $x \in (0,1)$ is said to follow a Beta probability density function g: $X \sim \mathscr{B}(\alpha, \beta)$ for some $\alpha > 0$ and $\beta > 0$ if:

$$g(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ x \in (0,1).$$

[1] The contents of this section are inspired from the book of: Robert and Casella [114].

$\Gamma(.)$ is the gamma integral.
Gamma distribution: a variable X with values $x > 0$ is said to follow a Gamma probability density function g: $X \sim \mathscr{G}(\alpha, \beta)$ for some $\alpha > 0$ and $\beta > 0$ if:

$$g(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)},\ x \in \mathbb{R}+.$$

Inverse Gamma distribution: Inverse Gamma distribution can be obtained from the Gamma distribution since, if $X \sim \mathscr{G}(\alpha, \beta)$ then $Y = \frac{1}{X}$ has an $\mathscr{IG}(\alpha, 1/\beta)$ distribution.
Dirichlet distribution: the probability density of the Dirichlet distribution for variables $\mathbf{X} = (x_1, \ldots, x_n)$ with parameters $\mathbf{u} = (u_1, \ldots, u_n)$ is defined by

$$g(\mathbf{X}) = \mathscr{D}(\mathbf{X},\ \mathbf{u}) = \frac{\Gamma(\sum_{i=1}^{n} u_i)}{\prod_{i=1}^{n} \Gamma(u_i)} \prod_{i=1}^{n} x_i^{u_i - 1} \text{ when } x_1, \ldots, x_n \geq 0; \sum_{i=1}^{n} x_i = 1 \text{ and}$$

$u_1, \ldots, u_n > 0$.
Dirichlet's distribution is an exponential natural family.[2] We choose in this thesis the Beta distribution to sample random number from the Dirichlet distribution; by using the marginal distribution of X_i for $i = 1, \ldots, n$ and generating recursively from $X_i \sim \mathscr{B}(u_i, \sum_{k=1}^{n} u_k - u_i)$. The reason behind using the Beta approach is well explained by Kim and Nelson [83, Chapter. 1].

2.3 Markov chain Monte Carlo and the ergodic theorem

2.3.1 Markov chains

A *Markov chain* is a sequence of random variables generated from a transition kernel such that the next state of the variable is generated only from the current state of the chain. Hence, a Markov Chain consists of a countable (possibly finite) set $\mathbb{S}$ (called the state space) together with a countable family of random variables $X_0, X_1, X_2, \ldots$ with values in $\mathbb{S}$ such that $p[X_{t+1} = s | X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0] = p[X_{t+1} = s | X_t = s_t]$. This Markov chain is called *irreducible*; if all the states of the chain are accessible from each other. Another important property of the Markov chain is *aperiodicity*: a Markov chain is called *aperiodic*, if the movement of the Markov chain is not subject to regular periodic or if the period of return times equal to 1. Also, a Markov chain can be a Harris recurrent chain, if it returns to a particular part of the state space an unbounded number of times. For more information on the properties of Markov chains, the reader is advised to consult the book of Gilks *et al.* [61]. Also, to be familiar with basics properties of convergence of random variables that are necessary to understand the following, see Appendix A.

[2] The reader is referred to the book of Robert [113] for the explicit details of the exponential family.

2.3.2 The ergodic theorem and the Markov chain Monte Carlo principle

The goal of the MCMC in a Bayesian context is to construct a Markov chain that converges to a stationary distribution (the target posterior distribution). Then, the Markov chain is simulated for a number of iterations. If the transition kernel of the Markov chain is aperiodic and irreducible, then convergence to the distribution of interest is guaranteed by the ergodic theorem (see theorem 4.4, Tierney [136]). The asymptotic properties show how we can draw a sample from a Markov chain with equilibrium distribution $\pi(x)$; or to estimate the expected value with respect to $\pi(x)$, of a function $f(x)$ of interest. In fact, if $X_1, X_2, ..., X_t, ...$ is a realization from an appropriate chain, typically available asymptotic results include :

$$\begin{aligned} &X^t \xrightarrow[t\longrightarrow\infty]{d} \pi(x) \text{ and} \\ &\frac{1}{t}\sum_{i=1}^{t} f(X^i) \xrightarrow[t\longrightarrow\infty]{a.s.} \mathbb{E}_\pi\{f(X)\} \text{ almost surely.} \end{aligned} \tag{2.1}$$

As stated in [117], successive X_t will be correlated due to the Markov properties[3] To take advantage of the first asymptotic result, suitable spacings will be required between realizations used to form the sample, or parallel independent runs of the chain might be considered. The ergodic average of a function f is the result of the central limit theorem, where for finite irreducible chain $(X_{t,\, t\in\mathbb{R}^+})$ with stationary distribution π and with a transition kernel Q satisfies for any Q integrable function f:

$$\begin{aligned} &\frac{1}{\sqrt{N}}\left(\sum_{t=1}^{N}(f(X_t) - E_\pi(f(X)))\right) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2), \text{ with} \\ &\sigma_f^2 = \gamma_0 + 2\sum_{k=0}^{\infty}\gamma_k, < \infty,\ \gamma_k = \mathbb{C}ov_\pi(f(X_0), f(X_k)). \end{aligned}$$

2.3.3 Stationarity and reversibility

Getting to the unique invariant distribution in the MCMC algorithm can be achieved, if the Markov chain is reversible or in other words it satisfies the detailed balance condition defined as follows:

Definition 2.1. *A Markov chain with transition kernel Q satisfies the detailed balance condition if there exists a function π satisfying $Q(y,x)\pi(y) = Q(x,y)\pi(x)$ for every (x, y).*

This detailed balance condition guarantees the reversibility of the Markov chain and the existence of a stationary limiting distribution as given by the following theorem from Robert and Casella [114, Chapter. 6]:

[3] One can see a classical book reviewing the basic definitions, concepts pertaining to Markov chains: Freedman [53].

Theorem 2.1. *Suppose that a Markov chain with transition function Q satisfies the detailed balance condition with π a probability density function, then:*

i/ The density π is the invariant density of the chain.

ii/ The chain is reversible.

Proof: (see Appendix B.)
Regular conditions of irreducibility and aperiodicity for the MCMC algorithms can be verified if the detailed balance is respected especially for the two important algorithms of the MCMC methods: the Gibs sampler and the MHA.

2.4 The Metropolis Hastings algorithm

When the conditional posterior $\pi(\theta)$ of a parameter θ is only known up to a constant or is un-normalized, the MHA is called for, as in [68]. The idea is that we can draw iteratively a value of a parameter $\theta^{(g+1)}$ at iteration $g+1$ based on the value at the iteration g by drawing candidate new value from a conditional proposal density $Q(\theta^{(g+1)}/\theta^{(g)})$. The algorithm proceeds for large number of iteration G until convergence as follow:

Algorithm 2.1

For $g = 1 : G$,

i) Initialize $\theta^{(0)}$,

ii) $g = g + 1$,

iii) Draw $\theta^{(*)}$ from $Q(\theta^{(*)}/\theta^{(g)})$,

iv) Accept $\theta^{(g+1)}$ with probability $\alpha(\theta^{(g)}, \theta^{(*)}) = \min\left\{\frac{\pi(\theta^{(*)})Q(\theta^{(g)}/\theta^{(*)})}{\pi(\theta^{(g)})Q(\theta^{(*)}/\theta^{(g)})}, 1\right\}$,

(a) Draw a uniform random variable u from $\mathscr{U}(0,1)$

(b) If $u < \alpha(\theta^{(g)}, \theta^{(*)})$: $\theta^{(g+1)} = \theta^{(*)}$, else: $\theta^{(g+1)} = \theta^{(g)}$.

2.4.1 Convergence properties for the Metropolis Hastings algorithm

The MHA verifies the detailed balance condition or the reversibility condition as shown by Roberts and Rosenthal [116]. The proof is provided in Appendix B.

Recall that we draw a sample according to $Q(\theta^{(*)}/\theta^{(g)})$, and then accept/reject according to $\alpha(\theta^{(g)}, \theta^{(*)})$. In other word, our transition kernel is $T(\theta^{(g)}, \theta^{(*)}) = \alpha(\theta^{(g)}, \theta^{(*)})Q(\theta^{(*)}/\theta^{(g)})$, that verifies $\pi(\theta^{(g)})T(\theta^{(g)}, \theta^{(*)}) = \pi(\theta^{(*)})T(\theta^{(*)}, \theta^{(g)})$,

which is the detailed balance condition. In other words, the MHA leads to a stationary distribution $\pi(\theta)$. Since the Markov chain of the MHA has; by construction, an invariant probability distribution π; if it is also an aperiodic Harris recurrent chain, then the ergodic theorem can be applied. Aperiodicity follows from the reversibility of the MHA of the transition kernel with invariant distribution π. Furthermore, the property of irreducibility of MHA chain follows from sufficient conditions such as positivity of the conditional density Q. More details about the convergence properties of the MHA can be found in [114, Chapter. 6].

2.4.2 Special cases of the Metropolis Hastings algorithm

There are many versions of the MHA depending on the choice of the proposal density (for more explanations, see Roberts and Rosenthal [116]), among them:

- *Symmetric Metropolis Algorithm:* here $Q(x, y) = Q(y, x)$, and the acceptance probability is simplified to $\alpha(x, y) = \min\left\{\frac{\pi(y)}{\pi(x),}\, 1\right\}$.

- *Random walk Metropolis-Hastings:* here $Q(x,\, y) = Q(y\text{-}x)$. The direction and distance of the new point from the current point is independent of the current point.

- *Independence sampler:* here $Q(x, y) = Q(y)$, i.e. $Q(x, .)$ does not depend on x.

- *Langevin algorithm:* here the proposal is generated from $Y_{n+1} \sim \mathcal{N}(X_n + (\delta/2) \bigtriangledown \log(\pi(X_n)), \delta)$, for some small δ, where $\bigtriangledown$ is the gradient of the target probability density function $\pi(.)$. The Langevin diffusion proposal tilts the draws toward the mode, unlike the random walk which moves symmetrically.

2.5 The Gibbs sampler

When the full conditional of the parameter is a known density, we can sample from it directly; this is called Gibbs sampler, as in [117]:
Let $\pi(x) = \pi(x_1, ..., x_k)$, $x \in \mathbb{R}^n$, denotes a joint density, and let $\pi(x_i/x_{-i})$ denotes the induced full conditional densities for each of the components x_i, given values of the other components $x_{-i} = (x_j;\, j \neq i)$, $i = 1, ..., k$, $1 < k \leq n$.
The Gibbs sampler algorithm proceeds by drawing successively from the full conditional distributions $\pi(x_i/x_{-i})$, $i = 1, ..., k$, as follows:

Algorithm 2.2

Pick arbitrary starting values $(x_1^0, ..., x_k^0)$;
x_1^1 from $\pi(x_1/x_0^{-1})$;
x_2^1 from $\pi(x_2/x_1^1, x_3^0, ..., x_k^0)$;
x_3^1 from $\pi(x_3/x_1^1, x_2^1, ..., x_k^0)$;
.
.
.
x_k^1 from $\pi(x_k/x_{-k}^1)$.

This completes a transition from $x^0 = (x_1^0, ..., x_k^0)$ to $x^1 = (x_1^1, ..., x_k^1)$, and so on we produce a sequence x^0, x^1, ..., $x^t, ...$. The transition kernel to move from x^t to x^{t+1} is given by

$$K_G(x^t, x^{t+1}) = \prod_{l=1}^{k} \pi(x_1^{t+1}/x_j^t,\, j > l,\, x_j^{t+1},\, j < l).$$

The components could be sub-vectors; the components that are more correlated should be sampled in block to have a good mixing of the chain and to avoid slow convergence, but at the expense of drawing from a multivariate conditional distribution. The Gibbs sampler is a special case of the MHA with the acceptance probability equal to 1.
Convergence properties such as irreducibility, recurrence and aperiodicity are satisfied by most Gibbs samplers. Hence, the law of large numbers holds for the Markov chain governing the transition, and the central limit theorem can apply (see for example, Tierney [136]).

2.6 Burn-in

Burn-in describes the practice of throwing away some iterations at the beginning of an MCMC run. You start at an initial state x_0, then you run the Markov chain for B steps (the burn-in period), during which you throw away all the sampled data (since the chain may be still too far from the converging chain). This is because one has no theoretical analysis of the Markov chain dynamics that tells where the good starting points are. In other word one does not know how much burn-in is required to get to a good starting point. For example, suppose we have M draws from an MCMC output about a parameter θ: $(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)}, ..., \theta^{(M)})$, and we want the expectation of any scalar function g: $\mu = \mathbb{E}\{g(\theta)\}$, then we discard the burn-in B first samples; and by application of central limit theorem; this quantity can be estimated by $\frac{1}{M-B} \sum_{t=B+1}^{M} g(\theta^{(t)})$.

2.7 Starting point determination

Not much has been said about the starting values especially for rapid mixing chains; since for irreducible chains, the algorithm will eventually converge no matter what initial values you select as stated by Gilks [61, Chapter. 6]. However, taking time to select good initial values will speed up convergence, most importantly the time spent to detect convergence will be reduced. By the way, if we choose to run several replications in parallel, the distribution of starting points should be over-dispersed with respect to the target distribution.

Several approaches have been proposed for generating initial values for MCMC samplers (see for example, Brooks [20]). For example, we can set the hyper-parameters to fixed values, or we can take the ML or the method of moments estimates as starting points for the chain. Alternatively, when the informative priors are available, the prior might also be used to generate suitable starting values. However, more rigorous methods proposed to sample from mixture of t-distributions, via a simple mode-finding algorithm to generate suitable starting values. Another way takes advantage of the simulated annealing algorithm to sample initial values. A discussion by many important authors about starting values in MCMC context is provided in [80].

2.8 Convergence assessment

It is worth mentioning that in practice no general exact tests for convergence exist, instead the tests for convergence should more formally be called tests for lack of convergence. Indeed, from our theory of Markov chains, we expect our chains to eventually converge to the stationary distribution, which is also our target distribution. However, there is no guarantee that our chain has converged after M draws. But there are several tests we can do; both visual and statistical (as inspired from Brook and Roberts [21], and Sinharay [129]), to see if the chain appears to be converged.

2.8.1 Visual inspection

One way to see if the chain has converged is to see how well our chain is mixing, or moving around the parameter space. Trace-plots inform us about the movement of the parameter simulation with iteration; if the chain is taking a long time to move around the parameter space then it will take longer to converge. If the model has converged, the trace plot will move like a spiral repeatedly around the mode of the distribution. Density plots are also used. Density plot permits to discover if the draws comes from the same distribution. Sometimes non-convergence is reflected in a multi-modal distribution. Another way to assess convergence is to assess the autocorrelations between the draws of our Markov chain; the autocorrelation between the draws should decay after a few lags for better ergodicity.

2.8.2 Formal convergence diagnostic

While there are many statistical tests, the most important ones are:
A- Gelman and Rubin Multiple Sequence Diagnostic:
Computed for each parameter as follows:

Algorithm 2.3

---

i) Run $m \geq 2$ chains of length $2n$ from over-dispersed starting values.

ii) Discard the first n draws in each chain.

iii) Calculate the within-chain variance $W = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{j=1}^{n} (\theta_{ij} - \overline{\theta_j})^2$.

iv) Calculate the between-chain variance $B = \frac{m}{n-1} \sum_{j=1}^{m} (\overline{\theta_j} - \overline{\overline{\theta}})^2$

v) Estimate the variance of the stationary distribution as a weighted average of W and B by:
$\hat{var}(\theta) = (1 - \frac{1}{n})W + \frac{1}{n}B$.

vi) Calculate the potential scale reduction factor: $\hat{R} = \sqrt{\frac{\hat{var}(\theta)}{W}}$

vii) $\hat{R}$ approaches to 1 as the chain converges.

B- The Geweke diagnostic:
The Geweke diagnostic takes two non-overlapping parts (usually the first 0.1 and last 0.5 proportions) of the Markov chain and compares the means of both parts, using a difference of means test to see if the two parts of the chain are from the same distribution (null hypothesis). The test statistic is a standard Z-score[4] with the standard errors adjusted for autocorrelation. The idea is to mimic the simple two-sample test of means. If the mean of the first 10% is not significantly different from the last 50%, then we conclude that the target distribution converged somewhere in the first 10% of the chain. The test proceeds as follows: Take two samples X_1 of size n_1 and X_2 of size n_2, and run the t-test of mean (unequal variances) with $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} \frac{S_2^2}{n_2}}}$. Where the degree of freedom for the student test for T is: $\min(n_1 - 1, n_2 - 1)$.

C- The effective sample size (ESS):
The effective sample size indicates how long we must run our chains to obtain

[4] A z-score describes the position of a raw score in terms of its distance from the mean when measured in standard deviation units.

reasonably accurate estimates of the posterior. This issue is complicated by the fact that we can not obtain independent draws from the posterior distribution and must settle instead for correlated samples. The ESS is defined as (see for example, Sharma [125]):

$$ESS = \frac{N}{1 + 2\sum_{k=0}^{\infty} \rho(k)}.$$

where N is the number of samples and $\rho(k)$ is the autocorrelation at lag k. If your samples are independents, your ESS equals the actual sample size. If the correlation at lag k decreases extremely slowly, so slowly that the sum in the denominator diverges, your effective sample size is zero. In one word, the lower the autocorrelation, the greater the amount of information contained in a given number of draws from the posterior.

D- Other tests:

There are many other formal diagnostic tests such as: the Heidelberger-Welsh test used to check the stationarity of the chain. Also, Raftery-Lewis diagnostics provides us with how many iterations are necessary to estimate the posterior for a given quantity.

Visual and formal diagnostics should be performed together to check stationarity for the variable of interest. For more details on diagnostics assessment refer to Brooks and Roberts [21].

2.9 Precautions raising in Markov chain Monte Carlo simulation

Many precautions must be considered when choosing MCMC as a tool for simulation; one of the issue is to find classes of proposal kernels that produce chains that converge and mix quickly especially for the MHA. Furthermore, much research has been conducted into the properties of the random walk Metropolis Hastings algorithm (RWMHA) when it comes to the choice of the scaling step and its relation to the acceptance rate. It has been shown that the optimal acceptance rate for proposals tends to $0,234$ and the scale move should be fixed according to this rate, because we should have a balance between the distance of proposed moves and the chances of acceptance. Increasing the former will reduce the autocorrelation in the chain if the proposal is accepted; but if it is rejected, the chain will not move at all, so the autocorrelations will be high (for more details see for example, Livingstone and Girolami [97]).

Also, when the target density is multi-modal, the MHA explores well only within any one of the modal, and the algorithm can recourse to the so called simulated tempering in order to flatten out the distribution to yield good estimates from the original target density as discussed by Craiu and Rosenthal [38].

From another side, when confronted with a poor behaviour of the chain, you have other options to improve your algorithm by for example re-parameterizing

the Gibbs sampler, using auxiliary variables, working first with simpler models or trying to fit well behaved data (see for example, Kass *et al.* [80]). Moreover, you should check you MCMC efficiency by comparison with other methods such as ML or EM methods.

In [117], Roberts and Smith discussed the implementation and the convergence issues such as deciding whether to adopt for several runs of the chain starting from a wide range of initial values is necessary. They also addressed the issue of how long the chain should be run for, and whether this can be done in advance or needs to be determined by some kind of sequential stopping rule.

Finally, MCMC depends on the efficiency of random variate generation and on any effort to reduce the variance of the estimates, where even the classical Monte Carlo variance reduction ideas are still relevant.

2.10 Preliminary on stochastic diffusion processes

2.10.1 Stochastic diffusion process definition, existence and uniqueness

2.10.1.1 Definition

A stochastic diffusion process $(X_{t,t\geq 0})$ is a solution to a stochastic differential equation (SDE). It is a continuous-time Markov process that has continuous sample paths (trajectories) . It is governed by the following SDE:

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t))dW_t. \tag{2.2}$$

Where $(W_{t,t\geq 0})$ is a Wiener process, $\mu(.)$ and $\sigma(.)$ are linear or non-linear functions called respectively the drift coefficient or the deterministic component, and the diffusion coefficient or the stochastic component. If $\mu(.)$ and $\sigma(.)$ do not depend on t, the process is called time-homogeneous.

2.10.1.2 Existence and uniqueness

To ensure the existence and the uniqueness of a solution of the process 2.2 for $0 \leq t \leq T$ where $x \in \mathbb{R}$ and T is fixed, the followings are sufficient conditions:

- $\mid \mu(x,t) \mid^2 + \mid \sigma(x,t) \mid^2 \leq C(1+\mid x \mid^2)$, for some constant C. This is the growth condition.
- $\mid \mu(x,t) - \mu(y,t) \mid + \mid \sigma(x,t) - \sigma(x,t) \mid \leq D(\mid y - x \mid)$, for some constant D. This is the Lipschitz condition.

2.10.2 Stochastic diffusion process discretization

Rarely can we find an explicit form for diffusion processes. When no explicit form is available, we can use discrete approximations of the continuous process. Different

schemes are available, the most important are the Euler Maruyama scheme and the Milstein scheme.

2.10.2.1 The Euler Maruyama scheme

The Euler Maruyama scheme approximates the process $\{X_t,\, t \geq 0\}$ for small time interval $\Delta_j = t_{j+1} - t_j$ at the discrete time-points t_j, $1 \leq j \leq N$ by the recursion:

$$X_{t_{j+1}} = X_{t_j} + \mu(X_{t_j}) + \sigma(X_{t_j})\Delta W_j,\ X_{t_0} = x_0.$$

Where $\Delta W_j = \sqrt{\Delta_j}.Z_j$, with Z_j being standard normal variables with mean 0 and variance 1 for all j. This procedure supposes constant drift and diffusion functions between time steps, consequently the approximation improves for smaller time steps.

2.10.2.2 The Milstein scheme

It can be derived through stochastic Taylor expansions. The Milstein scheme improves the accuracy of the approximation for $\{X_t,, t \geq 0\}$ by adding a second term from Itô formula and the discretization is of the form:

$$X_{t_{j+1}} = X_{t_j} + \mu(X_{t_j}) + \sigma(X_{t_j})\Delta W_j + \frac{1}{2}\sigma(X_{t_j})\sigma^{'}(X_{t_j})((\Delta W_j)^2 - \Delta_j).$$

The Euler term plus a second term. For more introductory notions and principles on SDE, one can refer to Bachar *et al.* [7].

2.11 Preliminary on hybrid switching diffusion model

Hybrid switching diffusion processes (HSD) or state-dependent regime switching diffusion processes have received much attention lately due to their many applications in fields such as biology, finance,...

2.11.0.1 Definition

A Hybrid switching diffusion process (see for example, Shao [124]) is a two-component process $(X_t, S_t)_{t\geq 0}$, where $(X_{t,\, t\geq 0})$ describes the continuous dynamics that satisfies a stochastic diffusion equation (SDE), and depends on a Markovien regime switching process $(S_{t,\, t\geq 0})$. The regime switching process $(S_{t,\, t\geq 0})$ is governed by a Markovien transition intensity matrix Q that depends on $(X_{t,\, t\geq 0})$. Hence, we have:

$$d(X_t) = \mu(X_t, S_t)dt + \sigma(X_t, S_t)dW_t;\ X_0 = x \in \mathbb{R}^d, \tag{2.3}$$

Where $(W_{t,t\geq 0})$ is a Brownian motion in $\mathbb{R}^d$, $d \geq 1$, $\sigma\, is\, d \times d$ matrix, and μ is a vector in $\mathbb{R}^d$. While (S_t) is a continuous time Markov chain on the state space $\mathbb{S} = \{1, 2, ..., a\}$; with $a \in \mathbb{N}^*$ satisfying:

$$\mathbb{P}(S_{t+\delta} = l | S_t = k,\, X_t = x) = \begin{cases} q_{kl}(x)\delta + o(\delta), & \text{if } q \neq l. \\ 1 + q_{kk}(x)\delta + o(\delta), & \text{if } q = l. \end{cases}$$

with $\delta > 0$, and the matrix $Q = (q_{kl}(x))$ is assumed to be irreducible and conservative for each $x \in \mathbb{R}^d$, so for every $k \in \mathscr{S}$: $q_k(x) = q_{kk}(x) = -\sum_{l \neq k} q_{kl}(x) < \infty$. If $q_{kl}(x)$ is independent of x, the process is called state-independent regime switching diffusion process.

2.11.0.2 Existence and Uniqueness of HSD process

The novelty of our model is due to the continuous-state dependence between $(S_{t,t\geq 0})$ and $(X_{t,t\geq 0})$; $(S_{t,t\geq 0})$ is a Markov chain only for a fixed x. Unlike the usual diffusion processes represented by SDE's, we have a Markovian two-component process $(X_t, S_t)_{t\geq 0}$. Sufficient conditions for the existence of a solution of this process is assured under the following conditions (as demonstrated by Yin *et al.* [144]):

- The function $Q(.)$ is bounded and continuous.
- The functions $\mu(.,.)$ and $\sigma(.,.)$ satisfy for some $K > 0$ and $K_0 > 0$, for all x, $z \in \mathbb{R}^d$, and $s \in \mathbb{S}$:
  - $\mid \mu(x,s) \mid^2 \leq K(1 + \mid x \mid^2)$, $\mid \sigma(x,s) \mid^2 \leq K(1 + \mid x \mid^2)$, and
  - $\mid \mu(x,s) - \mu(z,s) \mid \leq K_0(\mid x - z \mid)$, $\mid \sigma(x,s) - \sigma(z,s) \mid \leq K_0(\mid x - z \mid)$,

Later, by using an Euler approximation for the state-dependent switching diffusion process by generating a sequence of independent and identically distributed random variables with normal distribution to approximate the Brownian motion, and based on a truncated Taylor series to approximate the exponential of the transition rate matrix $Q(x)$ for every fixed x, Yin *et al.* [144] showed the existence of a unique solution for the HSD process.

Some other important terms such as the likelihood and the FFBS algorithm for regime switching diffusion processes are provided in Appendix D of this manuscript.

2.11.0.3 Data imputation for diffusion processes

When the transition density of the diffusion process is known, parameters'estimation is straightforward using classical methods such as ML approach. However, in most cases the likelihood is intractable and we need to approximate the transition density; where data is available at only discrete time points. Moreover, these discrete time intervals need to be too small for a good approximation of the transition density. Thus, we recourse to the Bayesian data imputation mechanism, which constitutes in imputing auxiliary data between successive observations of the low frequency observed diffusion process. By the way, our likelihood will be augmented by the new imputed data X^{imp} beside the observed data X^{obs} (see for example, Eraker [49], Golightly and Wilkinson [63]). Let suppose, we impute $\{X_k^{imp}, k = 1, ..., m\}$ between two successive observations with the times $\{t_k, k = 1, ..., \}$ for the interval time (a, b) with $\Delta t_{k+1} = t_{k+1} - t_k$ and $\Sigma = \sigma\sigma^{'}$. As summarized in [56], there are

many imputation schemes among them:
The Euler proposal: where we propose every new imputed data from:

$$X^*_{k+1} \sim \mathscr{N}\left(X^*_k + \mu(X^*_k, \theta)\Delta t_{k+1},\, \Sigma(X^*_k, \theta)\Delta t_{k+1}\right),$$
$$\text{for } k = 1,\, ...,\, b-2.$$

The double-sided Euler proposal: we update X^{imp} on (a, b) from the left to the right, where every X^*_{k+1} is updated conditioned on the already updated preceding X^*_k and the value X_{k+2}, and we propose every new imputed data from:

$$X^*_{k+1} \sim \mathscr{N}\left(X^*_k + \frac{X_{k+2} - X^*_k}{t_{k+2} - t_k}\Delta t_{k+1},\, \frac{t_{k+2} - t_{k+1}}{t_{k+2} - t_k}\Sigma(X^*_k, \theta)\Delta t_{k+1}\right),$$
$$\text{for } k = 1,\, ...,\, b-2,\ \text{and } X^*_a = X_a.$$

The modified bridge proposal: we update X^{imp} on (a, b) by updating every X^*_{k+1} conditioned on the preceding X^*_k and on the right end point X_b, and we propose every new imputed data from:

$$X^*_{k+1} \sim \mathscr{N}\left(X^*_k + \frac{X_b - X^*_k}{t_b - t_k}\Delta t_{k+1},\, \frac{t_b - t_{k+1}}{t_b - t_k}\Sigma(X^*_k, \theta)\Delta t_{k+1}\right),$$
$$\text{for } k = 1,\, ...,\, b-2,\ \text{and } X^*_a = X_a.$$

The diffusion bridge proposal: we update X^{imp} on (a, b) by updating every X^*_{k+1} in a similar way as Euler scheme, but we propose from:

$$X^*_{k+1} \sim \mathscr{N}\left(X^*_k + \frac{X_b - X^*_k}{t_b - t_k}\Delta t_{k+1},\, \Sigma(X^*_k, \theta)\Delta t_{k+1}\right),$$
$$\text{for } k = 1,\, ...,\, b-2,\ \text{and } X^*_a = X_a.$$

Other proposal can be considered such as the Gaussian and the Student-t proposals.

Chapter 3

Bayesian Estimation of Discrete and Continuous Time Markov Switching Autoregressive Models

Summary

In this chapter, and after the introduction, we give the specificities of the Markov switching autoregressive models against other Markov switching models in section 2; then a preliminary on HMM is provided in section 3, before discussing the label

switching problem encountered in MSM in section 4. Section 5 is dealing with the model specification, the likelihood formulation, the Bayesian inference and the simulation study of the multivariate autoregressive hidden Markov model of first order (MAR(1)HMM) in the discrete time. While in section 6, we provide the model description, the likelihood formulation, the MCMC inference and the simulation study for the continuous time Markov switching autoregressive model of first order (the CTMSAR(1)) . Then, we finish by a conclusion. The contents of this chapter are the results of a **book chapter by El-Maroufy *et al.* [47].**

3.1 Introduction

The Markov switching model (MSM) belongs to the family of the state space models. Like the state space models, the MSM has two equations: the measurement equation and the transition equation. When the transition component is discrete and Markovien, we obtain the MSM. The Markov switching model is an extension of HMM by including lagged observations (see, Lu *et al.* [99]). In other words, the measurement equation supposes dependence between observations. If the current observable random variables depend on their historical values as well as on the current hidden state variable through an autoregressive relationship, we get the Markov switching autoregressive model (MSAR).

These kinds of models are very acquainted to model disease behaviour such as breast cancer; where it was found that there is a correlation between the disease development, recurrence or prediction and observable bio-markers (see for example, Samy *et al.* [122], Shigeru *et al.* [126], Dorit *et al.* [42], Walter *et al.* [142]). This gives an opportunity of using an HMM to predict the stage of the disease based on bio-markers'data; by the way, we can address the effectiveness of the treatments in their influence on the transition of the cancer from one state to another. This is because in HMM, we have two constituents: the Markovian hidden process convenient to represent the breast cancer stages, and the observations process given by the bio-marker data. Consequently, using these processes would allow to learn about the disease transition rates, such as for example how the disease progresses from primary breast cancer to advanced cancer stage.

However, the model we consider here is a variation of the regular HMM, since we will use extensions to incorporate dependence among successive observations, suggesting autoregressive dependence among continuous observations. Consequently, we have relaxed the conditional independence assumption from a standard HMM, because we would like to add some dynamics to the patient disease progression and also because in reality the current patient bio-marker observation is dependent on the past one. In fact, the autoregressive assumption in HMM has shown its advantage over regular HMM that can not catch the strong dependence between successive observations (see for example, Ailliot and Monbet [2]). A similar model to ours can be found in [50].

Our main goal in this chapter will be to compute the parameters of the model.

However; as detailed in chapter 2, parameters'estimation is very challenging for HMM family models, since the likelihood is not available in a closed form most of the time. Thus, we call for a Markov Chain Monte Carlo (MCMC) procedure instead of a maximum likelihood based approach. This choice rises from the fact that the Bayesian analysis uses prior knowledge about the process being measured; it allows direct probability statements and an approximation of posterior distributions for the parameters. Instead in the maximum likelihood approach, we can not have declared prior or have exact distribution for the parameters when the likelihood is intractable or when we have missing data (see for example, Lindley [95], Bolstad [18], Gelman [59]).

Since the realisation of HMM includes two separate entities: the parameters and the hidden states; the Bayesian computation is carried out after augmenting the likelihood by the missing hidden states (see for example, Tanner and Wong [134], Hobert [72]). In our case, the hidden states are sampled using a Gibbs sampler adopting a joint estimation of the hidden states or a block update of the states (instead of a single update of each state separately) by the mean of a FFBS algorithm. After estimating the hidden states, we can compute the autoregressive parameters and the transition probabilities of the Markov chain. Here, the parameters are sampled by Gibbs sampler from their posterior densities after specifying conjugate priors for the parameters.

Hence, we can construct an MCMC algorithm to get an output from each parameter, where the MCMC algorithm will alternate between simulating the hidden states and the parameters. Finally, we can obtain the posterior statistics for all the parameters such as the means, standard deviations and confidence intervals using the MCMC output; after assessing the convergence of the MCMC algorithm. To be acquainted with the basic theory, notions and algorithm of MCMC refer to Brooks [22].

3.2 Markov switching autoregressive models versus other Markov switching models

The Markov switching autoregressive models were first proposed in [67] to describe econometric time series; they are a generalization of both HMM and autoregressive models, and will be effective in representing multiple heterogeneous dynamics such as the disease progression dynamics, and can be even generalized to a regime switching ARMA models such as in [81]. Furtheremore, the formulation of our model can be seen as an extension of the multivariate double chain Markov model (DCMM) developed by Fitzpatrick and Marchev [52], where there are two discrete Markov chains of first order; the first Markov chain is observed and the second one is hidden. In contrast to this DCMM, our multivariate autoregressive hidden Markov model of first order MAR(1)HMM will lead to continuous observations, with each observation conditional on the hidden process will depend on the previous observation according to an autoregressive process of first order. This dynamic is promising for

continuous observed disease bio-markers.

The study of MSAR are considered in both the discrete time framework as well as the continuous time case (CTMSAR(1)) . Nevertheless, we should note that the continuous time latent Markov models are more familiar in describing disease progression; from human immunodeficiency virus, cervical cancer, breast cancer, to genomic tests (see for example, Chantel *et al.* [28], Kirby and Spiegelhalter [84], Chunling and Tsokos [36], Mitra and Gupta [103]). Furthermore, the CTMSAR(1) is a more appropriate framework for irregularly spaced time observations than the discrete time steps. Unlike discrete time situation, the focus in continuous time would be of estimating the transition intensity rate matrix instead of the transition probability matrix; an estimation that would have recourse to disease markers'observations (see for example, Bureau *et al.* and the references therein).

3.3 Preliminary on hidden Markov models

HMM have been proven to be useful tools for tackling numerous concrete problems in many fields; but some possible applications of HMM are in: speech processing, biology, disease progression, economics and gene expression (see for example, Benmiloud and Piczunski [12], Boys and Handerson [19], Chantel *el al.* [28], Albert and Chib [4], Korolkiewickz and Elliot [87], Zeng and Frias [148]). For a complete review of HMM, the reader is referred to Zucchini and Mcdonalds [149], in which properties and definitions of HMM are presented in an intelligible way; with both the classical estimation by ML method, EM algorithm and the Bayesian inference are addressed.

Since the model suggested is an extension of the HMM, we will describe the HMM in more details: an HMM is a stochastic process $\{X_t, Y_t\}_{t=0}^{T}$, where $\{X_t\}_{t=0}^{T}$ is a hidden Markov chain (unobservable) and $\{Y_t\}_{t=0}^{T}$ is a sequence of observable independent random variables such that Y_t depends only on X_t for the time $t = 0, 1, ..., T$. Here the process $\{X_t\}_{t=0}^{T}$ evolves independently of $\{Y_t\}_{t=0}^{T}$ and is supposed to be an homogeneous finite Markov chain with probabilities transition matrix Π of dimension $a \times a$, where a indicates the number of the hidden states and $\Pi_0 = (\Pi_{01}, ..., \Pi_{0a})$ is the initial state distribution.

We denote the probability density function of $Y_t = y_t$ given $X_t = k$ for $k \in \{1, ..., a\}$ with $P_{x_t}(y_t, \theta_k)$, where θ_k refers to the parameters of P when $X_t = k$. We suppose further that the processes $Y_t|X_t$ and $Y_{t'}|X_{t'}$ are independent for $t \neq t'$. Let $\Theta = (\theta_1, ...\theta_a)$ and $\theta = (\Pi_0, \Pi, \Theta)$, then the HMM can be described as follows: first, the likelihood of the observations and the hidden states can be decomposed to

$$P(y_0, ..., y_T, x_0, ..., x_T, \theta) = P(y_0, ..., y_T|x_0, ..., x_T, \theta)P(x_0, ..., x_T, \theta),$$

Since $\{X_t\}_{t=0}^{T}$ is a Markov chain,

$$P(x_0, ..., x_T, \theta) = \Pi_0(x_0) \prod_{t=1}^{T} \Pi(x_t|x_{t-1}),$$

Under the conditional independence of the observations given the hidden states:

$$P(y_0, ..., y_T|x_0, ..., x_T, \theta) = P_{x_0}(y_0|\theta_{x_0}) \prod_{t=1}^{T} P_{x_t}(y_t|\theta_{x_t})$$

Consequently the likelihood function for the hidden states and the observations is given by

$$P(y_0, y_1, ..., y_T, x_0, x_1, ..., x_T, \theta) = \Pi(x_0)P_{x_0}(y_0|\theta_{x_0}) \prod_{t=0}^{T} \Pi(x_t|x_{t-1})P_{x_t}(y_t|\theta_{x_t}).$$

3.4 The label switching problem in Markov switching models

One of the problems that arises when estimating the MSM or the HMM in a Bayesian context is the so called label switching problem; which means the likelihood in MSM is invariant to the permutations of the labels, by the way we get a problem of identification when estimating the parameters that depend on the Markov states as detailed by Sperrin *et al.* [132]. Suppose, we have n observations $y = (y_1, ..., y_n)$ taken from a a component MSM with the parameters depending on the hidden states $\Theta = (\Theta_1, \ldots, \Theta_a)$. Let Pm_a denotes the set of all permutations on 1, 2, ..., a. Then the likelihood is the same for all the permutations $\nu \in Pm_a$: $L(y, \Theta_1, \ldots, \Theta_a) = L(y, \Theta_{\nu(1)}, \ldots, \Theta_{\nu(a)})$. Moreover, if exchangeable[1] priors are used, the posterior of each parameter will have $a!$ symmetric modes. Consequently, the posterior inference will be meaningless. Jasra *et al.* [77] has summarized many methods to overcome the labels switching issue, among them we could cite:

3.4.1 Identification constraints on the parameters:

Identifiability constraint on the parameter space is simple and works well in many situations especially when the priors are exchangeable; however it could circumvent the parameters from exploring freely during MCMC iterations as it is recommended in MCMC principle. Indeed, with constraints; we come with the implications of truncating the support of the posterior in terms of simulation. However, since identifiability constraints can be imposed after MCMC run; we can simulate from the unconstrained posterior distribution and then impose the constraint. Furthermore, identifiability constraints use exchangeable priors that are weakly informative.

3.4.2 The loss function algorithm:

This is a relabelling algorithm which it is applied to the parameters after obtaining the MCMC output, perhaps the best known algorithm is based on decision theoretic

[1] containing no component-specific information

arguments of Hurn *et al.* [73]. Samples from MCMC output are post-processed according to some loss function criterion. The algorithm is well justified. However, it is computationally intensive for large data sets or high dimensional parameter space.

3.4.3 The random permutation algorithm:

This algorithm proceeds while performing MCMC iterations by choosing a random permutation ν of the labels from the $a!$ parameters. This permutation is applied to the parameters as well as to the hidden states: the new $\Theta = \nu(\Theta)$ and the new sequence of the hidden state is also permuted using the permutation ν (see for example, Fruhwirth-Schnatter [55, 54]).

3.4.4 The probabilistic relabelling algorithm:

The probabilistic relabelling algorithm of Sperrin *et al.* [132] involves the calculation of the likelihood of the permutations via an EM-type algorithm where the observations correspond to the available data and the permutations are the missing data for the EM algorithm.

There are many other methods. Overall, each method should be evaluated separately in term of performance, sensitivity before use.

3.5 The Multivariate autoregressive hidden Markov model of first order

In this section, we first describe the MAR(1)HMM and we formulate the likelihood. Later, we give the details of the Bayesian inference for both the parameters and the hidden Markov states that would be evaluated in block through a FFBS algorithm. In this section, we will see how it is important to evaluate the transition probabilities in the discrete time case, before validating the inference by a simulation study.

3.5.1 Model description and specification

The MAR(1)HMM model we consider is a variation of the regular hidden Markov model; where conditionally on the latent states, the observations are not independent like it is the case for a regular HMM. Instead, the current observation is allowed to depend on the previous observation according to an autoregressive model of first order. The figure 3.1 illustrates the dynamics of the model. As in an HMM model, the latent states evolve according to a discrete first order time homogeneous Markov process.

We consider data of n continuous random variables observed over time, each of potentially different lengths, that is for each individual $i = 1, 2, ..., n$, we observe a vector $y_{i,.} = (y_{i,u_i}, ..., y_{i,m_i})^T$, with $u_i < m_i$.

Define $u_0 = \min_{1\leq i\leq n}\{u_i\}$ and $M = \max_{1\leq i\leq n}\{m_i\}$ and note that the times u_i and m_i may vary over the entire observation period from u_0 to M with the restriction that $u_i - m_i \geq 1$, for $i = 1, 2, ..., n$.
We assume for $i = 1, 2, ..., n$ for integer time $t = u_i, ..., m_i$, that the random variable $Y_{i,t}$ taking non-negative values depends only on the states X_t and the previous observation $Y_{i,t-1}$ and based on the model developed by Farcomeni and Arima [50], we get the following model:

$$Y_{i,t}|_{X_t=x_t} = \beta^{(x_t)}Y_{i,t-1} + \mu^{(x_t)} + \varepsilon_{i,t}. \tag{3.1}$$

The choice of the autoregressive part of the model is motivated by the fact that successive bio-marker observations are most of the time correlated for many diseases, unlike the hypothesis of independence between observations in HMM's.
We interpret x as the vector of the hidden health states of the patients; in the case of breast cancer these states would be localized or advanced metastatic breast cancer for example; while y is the vector of the bio-markers observed and measured for the patients. The $\varepsilon_{i,t}$ are normal variables with mean 0 and variance σ^2 such that $\varepsilon_{i,t}$ and $\varepsilon_{i',t'}$ are uncorrelated, for $(i,t) \neq (i',t')$.

The parameters $\beta^{(x_t)}$ and $\mu^{(x_t)}$ are parameters taking values in $\mathbb{R}$ for each hidden state x_t and $\sigma^2 \in \mathbb{R}^+$.
Similar to Fitzpatrick and Marchev [52], the transition matrix of the Markov chain Π is time homogeneous with dimension $a \times a$ where a is the number of hidden states, and $\Pi = (\Pi_{gh}, g = 1, ..., a;\ h = 1, ..., a)$ where $\Pi_{gh} = P(X_t = h|X_{t-1} = g)$, for $g = 1, 2, ..., a$; $h = 1, 2, ..., a$ and $t = u_{0+1}, ..., M$. We let the first state X_{u_0} to be selected from a discrete distribution with vector of probabilities $r = (r_1, ..., r_a)$. Also we consider the time of initial observation u_i, the initial observed state y_{i,u_i}, and the number of consecutive time points that were observed $m_i - u_i + 1$. Let $\mu = (\mu^{(1)}, ..., \mu^{(a)})$, $\beta = (\beta^{(1)}, ..., \beta^{(a)})$, and $\theta = (\mu, \beta, \sigma^2, r, \Pi)$ be the set of all parameters in the model. We suppose that the individuals; Y_i behave independently conditionally on X. Therefore for $i = 1, ..., n$,

$$P(y_{i,.}|y_{i,u_i}, x, \theta) = \prod_{t=u_i+1}^{m_i} P(y_{i,t}|y_{i,t-1}, x_t, \Theta),$$

and

$$P(x|\theta) = P(x_{u_0}) \prod_{t=u_0+1}^{M} P(x_t|x_{t-1}, \Pi),$$

where

$$P(x_t|x_{t-1}, \Pi) = P(X_t = x_t|X_{t-1} = x_{t-1}, \Pi) = \Pi_{x_{t-1},x_t}.$$

Then the likelihood density for the observations of all individuals $y = (y_{1,.}, ..., y_{n,.})$ given first time vector of observations $y_0 = (y_{1,u_1}, ..., y_{n,u_n})$, x and θ is

$$P(y|y_0, x, \theta) = \prod_{i=1}^{n} P(y_{i,.}|y_{i,u_i}, x, \theta),$$

The joint mass of each $y_{i,.}$ and x given y_{u_i} and θ can be written as follow:

$$P(y_{i,.}, x|y_{i,u_i}, \theta) = P(y_{i,.}|y_{i,u_i}, x, \theta) \times P(x|y_{u_i}, \theta).$$

Using the Markov property of the hidden process, we have after simplification

$$P(x|y_{i,u_i}, \theta) \propto P(y_{i,u_i}|x, \theta)P(x|\theta) = P(y_{i,u_i}|x_{u_i}, \theta) r_{x_{u_0}} \Pi_{x_{u_0}, x_{u_0+1}} \times ... \times \Pi_{x_{M-1}, x_M}.$$

$$\text{In addition } P(y_{i,.}|y_{i,u_i}, x, \theta) = \prod_{t=u_i+1}^{m_i} P(y_{i,t}|y_{i,t-1}, x, \theta), \text{ consequently,}$$

$$P(y_{i,.}, x|y_{i,u_i}, \theta) \propto r_{x_{u_0}} P(y_{i,u_i}|x_{u_i}, \theta) \prod_{t=u_0+1}^{M} \Pi_{x_{t-1}, x_t} \prod_{t=u_i+1}^{m_i} P(y_{i,t}|y_{i,t-1}, x, \theta).$$

Finally, under the hypothesis of normal error distribution for the autoregressive parameters of the model (4.1) and the Chapman-Kolmogorov property (see appendix C), the joint distribution of $y_{i,.}$ and x given y_{i,u_i} and θ can be simplified to:

$$\begin{aligned} P(y_{i,.}, x|y_{i,u_i}\theta) \quad &\propto \quad P(y_{i,u_i}|x_{u_i}, \theta) \prod_{h=1}^{a} r_h^{\chi_{\{x_{u_0}\}}(h)} \prod_{t=u_0+1}^{M} \prod_{g=1}^{a} \prod_{h=1}^{a} \Pi_{g,h}{}^{\chi_{\{x_t, x_{t-1}\}}(g,h)} \\ &\times \quad \prod_{t=u_i+1}^{m_i} \prod_{h=1}^{a} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(h)} - \beta^{(h)} y_{i,t-1}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(h)}, \end{aligned}$$

where ϕ denotes the density of a standard normal distribution $\mathcal{N}(0, 1)$ and $\chi_{\{A\}}(x)$ is the usual indicator function of a set A. Finally the joint distribution of y and x has the following form:

$$\begin{aligned} P(y, x|y_0, \theta) \quad &\propto \quad \prod_{h=1}^{a} r_h^{\chi_{\{x_{u_0}\}}(h)} \prod_{t=u_0+1}^{M} \prod_{g=1}^{a} \prod_{h=1}^{a} \Pi_{g,h}{}^{\chi_{\{x_t, x_{t-1}\}}(g,h)} \\ &\times \quad \prod_{i=1}^{n} \prod_{l=1}^{a} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,u_i} - \mu^{(l)}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(l)} \\ &\times \quad \prod_{i=1}^{n} \prod_{t=u_i+1}^{m_i} \prod_{h=1}^{a} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(h)} - \beta^{(h)} y_{i,t-1}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(h)}. \end{aligned} \tag{3.2}$$

3.5.2 Bayesian estimation of the model parameters

We will use a Bayesian approach to estimate the model parameters. Inference in the Bayesian framework is obtained through the posterior density, which is proportional to the prior multiplied by the likelihood. The posterior distribution for our model, as in most cases can not be derived analytically and we will approximate it through Markov chain Monte Carlo methods (MCMC); specifically designed for working

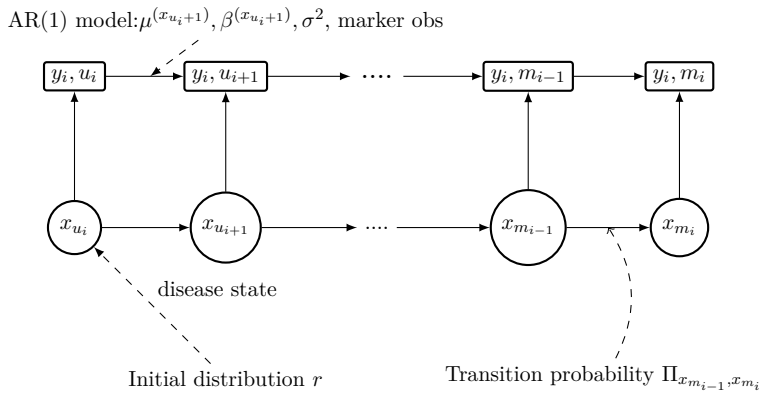


Figure 3.1: Diagram of MAR(1)HMM: u_i and m_i vary between minimum u_0 and Maximum M.

with the augmented likelihood with the hidden states.

In fact, as mentioned in chapter 2, MCMC methods start by specifying the prior density for the parameters θ. Since the data Y is available, the general sampling methods work recursively by alternating between simulating the full conditional posterior distribution of x given y and θ, and computing the posterior of θ given x and y.

A-Prior distributions:

Under the assumption of independence between the parameters, the prior density for $\theta = (\mu, \beta, \sigma^2, r, \Pi)$, could be written as

$$P(\theta) = P(r)P(\Pi)P(\mu)P(\beta)P(\sigma^2)$$

r are the parameters of a multinomial distribution, hence the natural choice for the prior would be a Dirichlet distribution $r \sim \mathbb{D}(\alpha_{01}, ..., \alpha_{0a})$. Later on, $\sum_{j=1}^{a} \Pi_{ij} = 1$, we assume that $\Pi_i \sim \mathbb{D}(\delta_{i1}, ..., \delta_{ia})$ for each row i of the transition matrix. This choice of the Dirichlet prior can be even the default $\mathbb{D}(1, ..., 1)$ as recently discussed by Tuyl *et al.* [140]. In fact, a Dirichlet prior is justified because the posterior density of each row of the transition matrix is proportional to the density of a Dirichlet distribution, and hence choosing a Dirichlet prior would give a posterior Dirichlet. This can be justified as follow for a given set of parameters $\lambda = (\lambda_1, ..., \lambda_a)$ from a discrete or from a multinomial density:
$\pi(x_1, ..., x_a, \lambda_1, ..., \lambda_a) = \frac{n!}{x_1!...x_a!}\lambda_1^{x_1}...\lambda_a^{x_a}$ for the non negative integers $x_1, ..., x_a$, with $\sum_{i=1}^{a} x_i = n$. This probability mass function can be expressed, using the gamma

function Γ,

$$\pi(x_1,...,x_a,\lambda_1,...,\lambda_a) = \frac{\Gamma\left(\sum_{i=1}^{a} x_i + 1\right)}{\prod_{i=1}^{a}\Gamma(x_i+1)} \prod_{i=1}^{a} \lambda_i^{x_i}.$$

This form shows its resemblance to the Dirichlet distribution, and by supposing the prior $\lambda \propto \mathbb{D}(\alpha_0,...,\alpha_a)$, the posterior is

$$P(\lambda|x) \propto P(\lambda)P(x|\lambda) \propto \prod_i \lambda_i^{x_i} \prod_i \lambda_i^{\alpha_i - 1} \propto \prod_i \lambda_i^{x_i+\alpha_i-1} \propto \mathbb{D}(x_1+\alpha_1,...,x_a+\alpha_a).$$

Furthermore, concerning the priors for parameters of the autoregressive model, we suppose for $h = 1,...,a$: $\mu^{(h)} \sim \mathcal{N}(\alpha_h, \tau_h)$, $\beta^{(h)} \sim \mathcal{N}(b_h, c_h)$, and an inverse gamma ($\mathbb{IG}$) prior for $\sigma^2 \sim \mathbb{IG}(\varepsilon, \zeta)$. $\alpha_h, \tau_h, b_h, c_h, \varepsilon, \zeta$ are hyper-parameters to be specified. For more details on the Bayesian inference and prior selection in HMM, the reader is referred to Cappe and Ryden [25]. In our case, prior distributions for the autoregressive parameters were proposed by Sampietro [121] for a mixture of autoregressive models, who points out that they are conventional prior choices for mixture of autoregressive models.

B- Sampling the posterior distribution for the hidden states

Chib [31] developed a method for the simulation of the hidden states jointly in the case of a univariate HMM. We will describe his full Bayesian algorithm for the univariate HMM, before a generalization to our MAR(1)HMM.

Chib's algorithm for the univariate hidden Markov model for the estimation of the states:

Suppose we have an observed process $Y_n = (y_1,....,y_n)$ and the hidden states $X_n = (x_1,....,x_n)$, θ are the parameters of the model. We adopt for simplicity $X_t = (x_1,....,x_t)$ the history of the states up to time t and $X^{t+1} = (x_{t+1},....,x_n)$ the future from $t+1$ to n. We use the same notation for Y_t and Y^{t+1}.

For each state $x_t \in \{1,2,....,a\}$ for $t = 1,2,....,n$, the hidden model can be described by a conditional density given the hidden states $\pi(y_t|Y_{t-1}, x_t = k) = \pi(y_t|Y_{t-1}, \theta_k)$, $k = 1,....,a$, with x_t depending only on x_{t-1} and having transition matrix Π and initial distribution Π_0 and the parameters for $\pi(.)$ are $\theta = (\theta_1,....,\theta_a)$.

Chib shows that it is preferable to simulate the full latent data $X_n = (x_1,....,x_n)$ from the joint distribution of $x_1,....,x_n|Y_n,\theta$, in order to improve the convergence property of the MCMC algorithm because instead of n additional blocks if each state is simulated separately, only one additional block is required. First, the joint posterior density of the hidden states given the observations and the parameters could be written as:

$$P(X_n|Y_n,\theta,\Pi) = P(x_n|Y_n,\theta)P(x_{n-1}|Y_n,x_n,\theta,\Pi) \times ... \times P(x_1|Y_n,X^2,\theta,\Pi).$$

For sampling, it is sufficient to consider the sampling of x_t from $P(x_t|Y_n,X^{t+1},\theta,\Pi)$. Moreover, $P(x_t|Y_n,X^{t+1},\theta,\Pi) \propto P(x_t|Y_t,\theta,\Pi)P(x_{t+1}|x_t,\Pi)$. This expression has

2 ingredients: the first is $P(x_{t+1}|x_t, \Pi)$, which is the transition matrix from the Markov chain. The second is $P(x_t|Y_t, \theta, \Pi)$ that would be obtained by recursively starting at $t = 1$. The mass function $P(x_{t-1}|Y_{t-1}, \theta, \Pi)$ is transformed into $P(x_t|Y_t, \theta, \Pi)$, which is in turn transformed into $P(x_{t+1}|Y_{t+1}, \theta, \Pi)$ and so on. For $k = 1, ..., a$, the update is:

$$P(x_t = k|Y_t, \theta, \Pi) = \frac{P(x_t = k|Y_{t-1}, \theta, \Pi)\pi(y_t|y_{t-1}, \theta_k)}{\sum_{l=1}^{a} P(x_t = l|Y_{t-1}, \theta, \Pi)\pi(y_t|y_{t-1}, \theta_l)}.$$

These calculations are initialized at $t = 0$, by setting $P(x_1|Y_0, \theta)$ to be the stationary distribution of the Markov chain. Precisely, the simulation proceeds for $k = 1, ..., a$, recursively by first simulating $P(x_1 = k|Y_0, \theta)$, from the initial distribution $\Pi_0(k)$ and $P(x_1 = k|Y_1, \theta, \Pi) \propto P(x_1 = k|Y_0, \theta, \Pi)\pi(y_1|Y_0, \theta_k)$. Then we get by forward calculation $P(x_t = k|Y_{t-1}, \theta) = \sum_{l=1}^{a} \Pi_{lk} P(x_{t-1} = l|Y_{t-1}, \theta)$, for each $t = 2, ..., n$, where Π_{lk} is the transition probability and
$P(x_t = k|Y_t, \theta) \propto P(x_t = k|Y_{t-1}, \theta, P)\pi(y_t|Y_{t-1}, \theta_k)$. The last term in the forward computation $P(x_n = k|Y_n, \theta)$ would serve as a start for the backward pass, and we get recursively for each $t = n - 1, ..., 1$;
$P(x_t = k|Y_n, X^{t+1}, \theta) \propto P(x_t|Y_{t-1}, \theta, \Pi)P(x_{t+1}|x_t = k, \Pi)$, which permits the obtention of $X_n = (x_1, ..., x_n)$.

Simulating the hidden states for the MAR(1)HMM:

Returning to our model, and adopting notations and algorithm developed by Fitzpatrick and Marchev [52]: f will denote the observation density for the MAR(1)HMM, and for
$u_0 < t < M$; $x_{-t} = (x_{u_0}, ..., x_t)$, $x^t = (x_t, ..., x_M)$, $y(t) = (y_{i,t}, i = 1, 2, ..., n)$, $y_{,t} = \bigcup_{i:u_i<t} \{y_{i,u_i}, ..., y_{i,\min\{t,m_i\}}\}$, and $y^t = \bigcup_{i:t<m_i} \{y_{i,\max\{t+1,u_i\}}, ..., y_{i,m_i}\}$. The posterior distribution of the hidden state could be written as:

$$P(x_{-M}|y_{,M}, \theta) = P(x_M|y_{,M}, \theta) \times ... \times P(x_{u_0}|y_{,M}, x^{u_0+1}, \theta)$$

So we could sample the whole sequence of states by sampling from $P(x_t|y_{,M}, x^{t+1}, \theta)$. Hence, as illustrated in figure 3.2, the estimation of the hidden states is performed recursively by first initializing

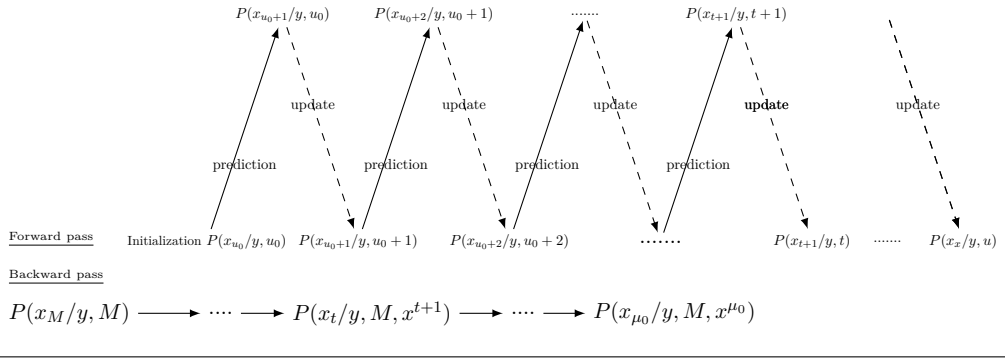
$$P(x_{u_0}|y_{,u_0}, \theta) \propto P(y_{,u_0}|x_{u_0})P(x_{u_0}|r); y_{,u_0} = \{y_{i,u_i,}, u_i = u_0, i = 1, ..., n\}.$$
$$P(x_{u_0+1} = k|y_{,u_0}, \theta) = \sum_{l=1}^{a} \Pi_{lk} P(x_{u_0} = l|Y_{,u_0}); k = 1, ..., a.$$
$$P(x_{u_0+1} = k|y_{,u_0+1}, \theta) \propto P(x_{u_0+1} = k|y_{,u_0}, \theta) f(y(u_0)|y_{,u_0}, \theta_k).$$

We perform a similar calculation for every state at time t, and we conclude by calculating

$$P(x_M = k|y_{,M-1}, \theta) = \sum_{l=1}^{a} \Pi_{lk} P(x_{M-1} = l|Y_{,M-1}, \theta), \text{ and}$$
$$P(x_M = k|y_{,M}, \theta) \propto P(x_M = k|y_{,M-1}, \theta) f(y(M)|y_{,M-1}, \theta_k).$$

Later on, we get $P(x_M = k|y_{,M}, \theta)$, which permits the simulation of $P(x_M|y_{,M}, \theta)$. Finally, by backward calculation we simulate from the probabilities $P(x_t|y_{,M}, x^{t+1}, \theta) \propto P(x_{t+1}|x_t, \Pi)P(x_t|y_{,t}, \theta)$ for each time $t = M-1, ..., u_0$. These backward probabilities would permit the simulation of the latent states.

Figure 3.2: Forward-backward algorithm



C- Sampling from $P(\theta|x, y)$

We will simulate the transition rate matrix Π, as well as the auto-regressive parameters.

Sampling Π :

Under the prior assumption of a Dirichlet prior for each row of the transition matrix $P(\Pi_i) \propto \mathbb{D}(\delta_{i1}, ..., \delta_{ia})$, and the independence assumption between these rows, the posterior distribution for Π_i can be developed using (3.4) as follow: Let n_{ij} denotes the number of single transitions from state i to state j, so

$$\begin{aligned} P(\Pi_i|y, x) &\propto P(\Pi_i) \prod_{t=u_0+1}^{M} \prod_{j=1}^{a} \Pi^{\chi_{\{x_{t-1}, x_t\}}(i,j)} \propto P(\Pi_i) \prod_{j=1}^{a} \Pi_{ij}^{n_{ij}} \propto \prod_{j=1}^{a} \Pi_{ij}^{\delta_{ij}+n_{ij}-1} \\ &\propto \mathbb{D}(\delta_{i1} + n_{i1}, ..., \delta_{ia} + n_{ia}). \end{aligned}$$

Sampling the posterior distribution for the initial distribution

Let $n_{0l} = \chi_{x_{u_0}}(l)$, for $l = 1, ..., a$. Using (3.4), and under a Dirichlet prior $\mathbb{D}(\delta_{01}, ..., \delta_{0a})$ for the parameter r, we obtain

$$P(r|x, y) \propto P(r) \prod_{l=1}^{a} r_l^{\chi_{\{x_{u_0}\}}(l)} \propto \prod_{l=1}^{a} r_l^{\delta_{0l}+n_{0l}-1} \propto \mathbb{D}(\delta_{01} + n_{01}, ..., \delta_{0a} + n_{0a}).$$

Sampling the posterior distribution for the autoregressive parameters μ, β, σ^2:

When a complete conditional distribution is known such as the normal distribution or the beta distribution, we use the Gibbs sampler to draw the random variable. This is the case for our model. Let us define

$n_{u_i}(l) = \sum_{i=1}^{n} \chi_{\{x_{u_i}=l\}}, n_l = \sum_{i=1}^{n} \sum_{t=u_i+1}^{m_i} \chi_{\{x_t=l\}}, N = \sum_{l=1}^{a} n_l, n_{0l} = \chi_{\{x_{u_0}\}}(l)$. So for $l = 1, 2, ..., a$; by supposing $\mathcal{N}(\alpha_l, \tau_l)$ as prior distribution and using (3.4); the conditional posterior distribution of $\mu^{(l)}$ is :

$$\begin{aligned} P(\mu^{(l)}|y,x) &\propto P(\mu^{(l)}) \prod_{i=1}^{n} \left[\frac{1}{\sigma}\phi\left(\frac{y_{i,u_i}-\mu^{(l)}}{\sigma}\right)\right]^{\chi_{\{x_{u_i}\}}(l)} \\ &\times \prod_{i=1}^{n}\prod_{t=u_i+1}^{m_i}\left(\frac{1}{\sigma}\phi(\frac{y_{i,t}-\mu^{(l)}-\beta^{(l)}y_{i,t-1}}{\sigma})\right)^{\chi_{\{x_t\}}(l)}. \\ &\propto \exp\frac{-1}{2}\left\{\frac{(\mu^{(l)}-\alpha_l)^2}{\tau_l} + \sum_{i=1,x_{u_i}=l}^{n}\left(\frac{y_{i,u_i}-\mu^{(l)}}{\sigma}\right)^2 + \right. \\ &\qquad \left. \sum_{i=1}^{n}\sum_{t=u_i+1,x_t=l}^{m_i}\left(\frac{y_{i,t}-\mu^{(l)}-\beta^{(l)}y_{i,t-1}}{\sigma}\right)^2\right\}. \end{aligned}$$

Then

$$\begin{aligned} &\mu^{(l)}/y,x \sim \mathcal{N}(\tilde{\alpha_l}, \tilde{\tau_l}), \\ &\text{with inverse variance} \\ &\tilde{\tau_l}^{-1} = \frac{n_{u_i}(l)+n_l}{\sigma^2} + \frac{1}{\tau_l}, \\ &\text{and mean} \\ &\tilde{\alpha_l} = \tilde{\tau_l}\left(\frac{\sum_{i=1,x_{u_i}=l}^{n} y_{i,u_i} + \sum_{i=1}^{n}\sum_{t=u_i+1,x_t=l}^{m_i}(y_{i,t}-\beta^{(l)}y_{i,t-1})}{\sigma^2} + \frac{\alpha_l}{\tau_l}\right). \end{aligned}$$

For $\beta^{(l)}, l = 1, ..., a$, and similar to $\mu^{(l)}$, $\mathcal{N}(b_l, c_l)$ was proposed as prior choice to obtain:

$$P(\beta^{(l)}|y,x) \propto P(\beta^{(l)}) \prod_{i=1}^{n}\prod_{t=u_i+1}^{m_i}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,t}-\mu^{(l)}-\beta^{(l)}y_{i,t-1}}{\sigma}\right)\right]^{\chi_{\{x_t\}}(l)},$$

Therefore

$$\begin{aligned} &\beta^{(l)}/y,x \sim \mathcal{N}(\tilde{b_l}, \tilde{c_l}), \\ &\text{with inverse variance} \\ &\tilde{c_l}^{-1} = \frac{1}{c_l} + \frac{\sum_{i=1}^{n}\sum_{t=u_i+1,x_t=l}^{m_i} y_{i,t-1}^2}{\sigma^2}, \\ &\text{and mean} \\ &\tilde{b_l} = \tilde{c_l}\left(\frac{b_l}{c_l} + \frac{\sum_{i=1}^{n}\sum_{t=u_i+1,x_t=l}^{m_i}(y_{i,t}-\mu^{(l)})y_{i,t-1}}{\sigma^2}\right). \end{aligned}$$

For the posterior distribution of σ^2, by supposing $\mathbb{IG}(\varepsilon, \zeta)$ as prior we deduce from (3.4)

$$\begin{aligned} P(\sigma^2|y,x) &\propto (\sigma^2)^{-(\varepsilon+1)} \exp\left(-\frac{\zeta}{\sigma^2}\right) \prod_{i=1}^{n} \left[\frac{1}{\sigma}\phi\left(\frac{y_{i,u_i} - \mu^{(x_{u_i})}}{\sigma}\right)\right] \\ &\times \prod_{i=1}^{n} \prod_{t=u_i+1}^{m_i} \left[\frac{1}{\sigma}\phi\left(\frac{y_{i,t} - \mu^{(x_t)} - \beta^{(x_t)} y_{i,t-1}}{\sigma}\right)\right], \end{aligned}$$

consequently

$$\begin{aligned} &\sigma^2/y, x \sim \mathbb{IG}(\tilde{\varepsilon}, \tilde{\zeta}), \\ &\text{with parameters} \\ &\tilde{\varepsilon} = \frac{n_{u_i} + N}{2} + \varepsilon, \\ &\text{and} \\ &\tilde{\zeta} = \frac{\sum_{i=1}^{n} (y_{i,u_i} - \mu^{(x_{u_i})})^2 + \sum_{i=1}^{n} \sum_{t=u_i+1}^{m_i} (y_{i,t} - \mu^{(x_t)} - \beta^{(x_t)} y_{i,t-1})^2}{2} + \zeta. \end{aligned}$$

Finally, the algorithm is ran for $d = 1, ..., D$ [2] iterations by alternating between the following steps, where in each step we compute a conditional posterior for the given parameter:

Algorithm 3.1

---

i) For $h = 1, 2, ..., a$, give reference values for the hyper-parameters $\alpha_h, \tau_h, a_h, b_h$, δ_{0h}, and δ_{ih} for $i = 1, 2, ..., a$.

ii) Initialization (Step $d = 1$ of the MCMC iterations):
Initialize $\Pi^{(1)}$, $r^{(1)}$, $\mu_{(1)}^{(h)}$, $\beta_{(1)}^{(h)}$ and $\sigma^{2(1)}$., for $h = 1, 2, ..., a$.

iii) Simulation of the hidden states:

(a) Initialization of forward simulation:
$P(x_{u_0}^{(d)}|y_{,u_0}, \theta) \propto P(y_{,u_0}|x_{u_0}^{(d)}) P(x_{u_0}^{(d)}|r^{(d)})$,
with $y_{,u_0} = \{y_{i,u_i,}, u_i = u_0, i = 1, ..., n\}$.

(b) Forward simulation: For $k = 1, ..., a$ and $t = u_0 + 1, ..., M$:
$P(x_t^{(d)} = k|y_{,t-1}, \theta) = \sum_{l=1}^{a} \Pi_{lk}^{(d)} P(x_{t-1}^{(d)} = l|Y_{,t-1}, \theta)$ and
$P(x_t^{(d)} = k|y_{,t}, \theta) = \frac{P(x_t^{(d)}=k|y_{,t-1},\theta_k) f(y(t)|y_{,t-1},\theta_k)}{\sum_{l=1}^{a} P(x_t^{d}=l|y_{,t-1},\theta) f(y(t)|y_{,t-1},\theta_l)}$.

(c) Initialization of backward simulation: For $k = 1, ..., a$, given
$P(x_M^{(d)} = k|y_{,M}, \theta)$ from forward simulation, we get $P(x_M^{(d)}|y_{,M}, \theta)$.

[2] D is sufficiently a large number of iterations for the MCMC algorithm to converge.

(d) Backward simulation: For $k = 1, ..., a$ and $t = M - 1, ..., u_0$: $P(x_t^{(d)}|y_{,M}, x^{t+1(d)}, \theta) \propto P(x_{t+1}^{(d)}|x_t^{(d)}, \pi)P(x_t^{(d)}|y_{,t}, \theta)$.

iv) Estimation of the initial distribution and the transition distribution

(a) for $l = 1, ..., a$, $k = 1, ..., a$. Calculate $n_{0l} = \chi_{\{x_{u_0}^{(d)}\}}(l)$ and $n_{kl} = \Sigma_{t=u_0+1}^{M} \chi_{\{x_{t-1}^{(d)}, x_t^{(d)}\}}(k, l)$.

(b) Sample $(r_1^{(d+1)}, ..., r_a^{(d+1)}) \propto \mathbb{D}(\delta_{01} + n_{01}, ..., \delta_{0a} + n_{0a})$.

(c) For $i = 1, ..., a$; sample $(\Pi_{i1}^{(d+1)}, ..., \Pi_{ia}^{(d+1)}) \propto \mathbb{D}(\delta_i 1 + n_{i1}, ..., \delta_i a + n_{ia})$.

v) Simulation of μ: For $l = 1, ..., a$,

(a) $\tilde{\tau_l}^{-1} = \frac{n_{u_i}(l) + n_l}{\sigma_{(d)}^2} + \frac{1}{\tau_l}$.

(b) $\tilde{\alpha_l} = \tilde{\tau_l} \left(\frac{\sum_{i=1, x_{u_i}=l}^{n} y_{i,u_i} + \sum_{i=1}^{n} \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i,t} - \beta_{(d)}^{(l)} y_{i,t-1})}{\sigma_{(d)}^2} + \frac{\alpha_l}{\tau_l} \right)$.

(c) Simulate $\mu_{(d+1)}^{(l)}/y, x \sim \mathscr{N}(\tilde{\alpha_l}, \tilde{\tau_l})$.

vi) Simulation of β: For $l = 1, ..., a$,

(a) $\tilde{c_l}^{-1} = \frac{1}{c_l} + \frac{\sum_{i=1}^{n} \sum_{t=u_i+1, x_t=l}^{m_i} y_{i,t-1}^2}{\sigma_{(d)}^2}$.

(b) $\tilde{b_l} = \tilde{c_l} \left(\frac{b_l}{c_l} + \frac{\sum_{i=1}^{n} \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i,t} - \mu_{d+1}^{(l)}) y_{i,t-1}}{\sigma_{(d)}^2} \right)$.

(c) Simulate $\beta_{(d+1)}^{(l)}/y, x \sim \mathscr{N}(\tilde{b_l}, \tilde{c_l})$.

vii) - Simulation of σ^2:

(a) $\tilde{\varepsilon} = \frac{n_{u_i} + N}{2} + \varepsilon$.

(b) $\tilde{\zeta} = \frac{\sum_{i=1}^{n} (y_{i,u_i} - \mu_{(d+1)}^{(x_{u_i})})^2 + \sum_{i=1}^{n} \sum_{t=u_i+1}^{m_i} (y_{i,t} - \mu_{(d+1)}^{(x_t)} - \beta_{(d+1)}^{(x_t)} y_{i,t-1})^2}{2} + \zeta$.

(c) Simulate $\sigma_{(d+1)}^2/y, x \sim \mathbb{IG}(\tilde{\varepsilon}, \tilde{\zeta})$.

3.5.3 Simulation study

We apply our results to the breast cancer model discussed earlier. The main reason behind our work is that the progression of breast cancer can not be seen directly unless we use observations related to the disease that could characterize its progression; these observations here are quantities which could be measured, they are called bio-markers; where the word bio-marker is used to designate any objective indication of a biological process or disease condition including during treatment and should be measurable. Furthermore, bio-markers are increasingly used in the management of breast cancer patients. One example is reported in [90] stating that "there is correlation between elevation of CEA and/or $CA\,15-3$ and disease progression, in breast cancer patients".

Also in our MAR(1)HMM, we have autoregressive dependence among the observations so as to add more dynamics to the model; unlike conventional HMM's where the successive observations given the Markov process are independent. We used the classification of breast cancer in three states: local, where the disease is confined within the breast; the regional phase, when the lymph nodes are involved; and the distant stage, where the cancer is found in other parts of the body. We restrict ourselves to these three stages unlike other stages classifications that divide the progression in more than three stages such as the TNM (tumor, node, metastasis) system.

By lack of finding data about breast cancer bio-markers, We will confine ourselves to simulate an MAR(1)HMM model for observation time $M = 50$, and a number of individuals $n = 100$, $a = 3$ for the number of the possible Markov states; with the observation time length for each individual selected uniformly between 2 and M. Let $\mu = (\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = (12, 24, 36)$, since markers such as $CA\,15-3$, increase as the disease advances toward metastatic breast cancer. In addition, $CA\,15-3$ increases rapidly between successive observations , thus we take in the simulation $\beta = (\beta^{(1)}, \beta^{(2)}, \beta^{(3)}) = (0.2, 0.4, 0.8)$.
The algorithm of simulation works as follows:

Algorithm 3.2

---

i) For each individual $i = 1, ..., n$, choose m_i the length of observation for the individual i.

ii) Generate each discrete disease state x_t using the transition probability matrix $\Pi = (0.7, 0.2, 0.1; 0.1, 0.6, 0.3; 0.2, 0.3, 0.5)$ for $t = u_{0+1}, ..., M$.

iii) Generate the observations $y_{i,t}$ for all individuals using our model 4.1.

We choose as prior for $\sigma^2 \sim \mathbb{IG}(0.001, 0.001)$. Also a $\mathbb{D}(1, ..., 1)$ prior is supposed for each row of Π. While, we suppose non informative priors for the μ's and the

β's.

Having the hidden states and the observations, we ran our algorithm for 8000 MCMC iterations or until convergence. We should report that the simulation of the Dirichlet posterior was carried out following Kim and Nelson [83, page. 22] and Krozlig [88, page. 155]) who claimed that the parameters of the posterior Dirichlet should be simulated using the beta distribution approach. Appendix G provides the Matlab program of the data generation as well as the MCMC program for the parameters'simulation.

MCMC algorithm convergence can be assessed either graphically or statistically using convergence tools described in chapter 2. Hence, graphical convergence was assessed by analysing MCMC iteration mixing plots that are shown in Figure 3.3. Also, the autocorrelation sample graphs were checked as illustrated in Figure 3.4, and finally we inspect the histograms of the posterior densities for the parameters of the models in figure 3.5. All parameters show good mixing of chains, autocorrelations that decay immediately after a few lags and perfect posterior densities'fitting.

Also, the Gelman [58] potential scale reduction factor (PSRF) was computed and plot for all the parameters. The PSRF measurement uses more than two MCMC chains (3 chains in this works are considered) and it is measured for each parameter of the model. it should show how the chains have forgotten their initial values and that the output from all chains is indistinguishable. It is based on a comparison of within chain and between-chain variances, and is similar to a classical analysis of variance; when the PSRF is high (perhaps greater than 1.1 or 1.2), then we should run our chains out longer to improve convergence to the stationary distribution. The PSRF for each parameter declines to 1 as the number of iterations approaches infinity to confirm convergence. All the parameters have shown a PSRF less than 1.1 as the number of iteration increases and by the way a good sign of convergence (figure 3.6).

In addition and to overcome the label switching or in other words the identifiability problem, we used well separated hyper-parameters for the priors. Even when we start from different initial values for the parameters, our algorithm converges immediately after few iterations.

Finally, let us point out that after checking for convergence, we can use the results of the ergodic theorem reported in equation 2.1 to compute different posterior statistics such as the posterior means.

Table 3.1 shows how the posterior mean values estimated from our algorithm are very close to the true ones. Even though the efficiency of the MCMC methods to approximate well the parameters of the MAR(1)HMM, we have limited ourselves here to the case where we have equidistant intervals which is not well fitting to the disease studies that suppose irregular time intervals as we will see in the next section.

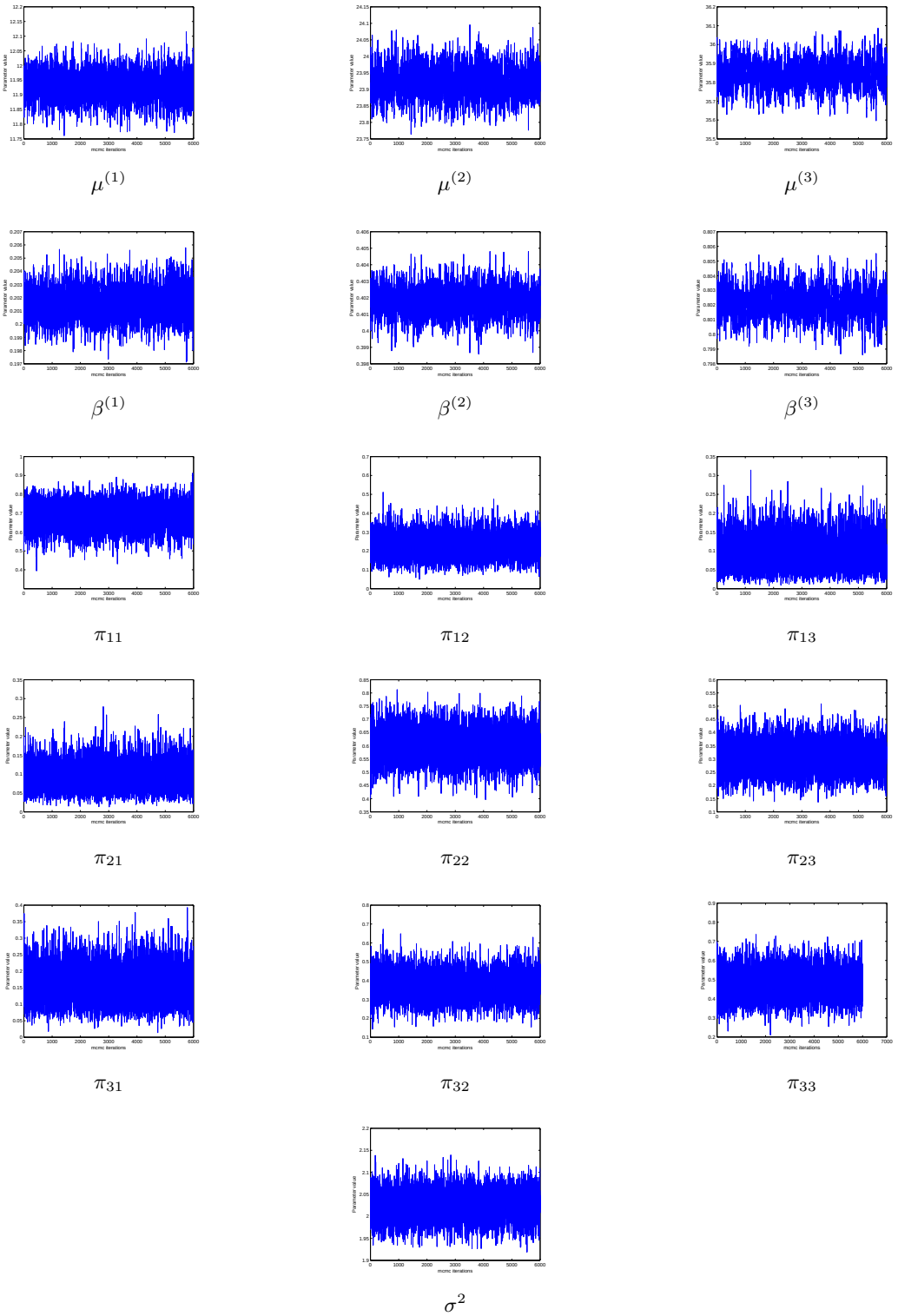


Figure 3.3: Markov chain mixing for each parameter through MCMC algorithm simulation for the MAR(1)HMM model.

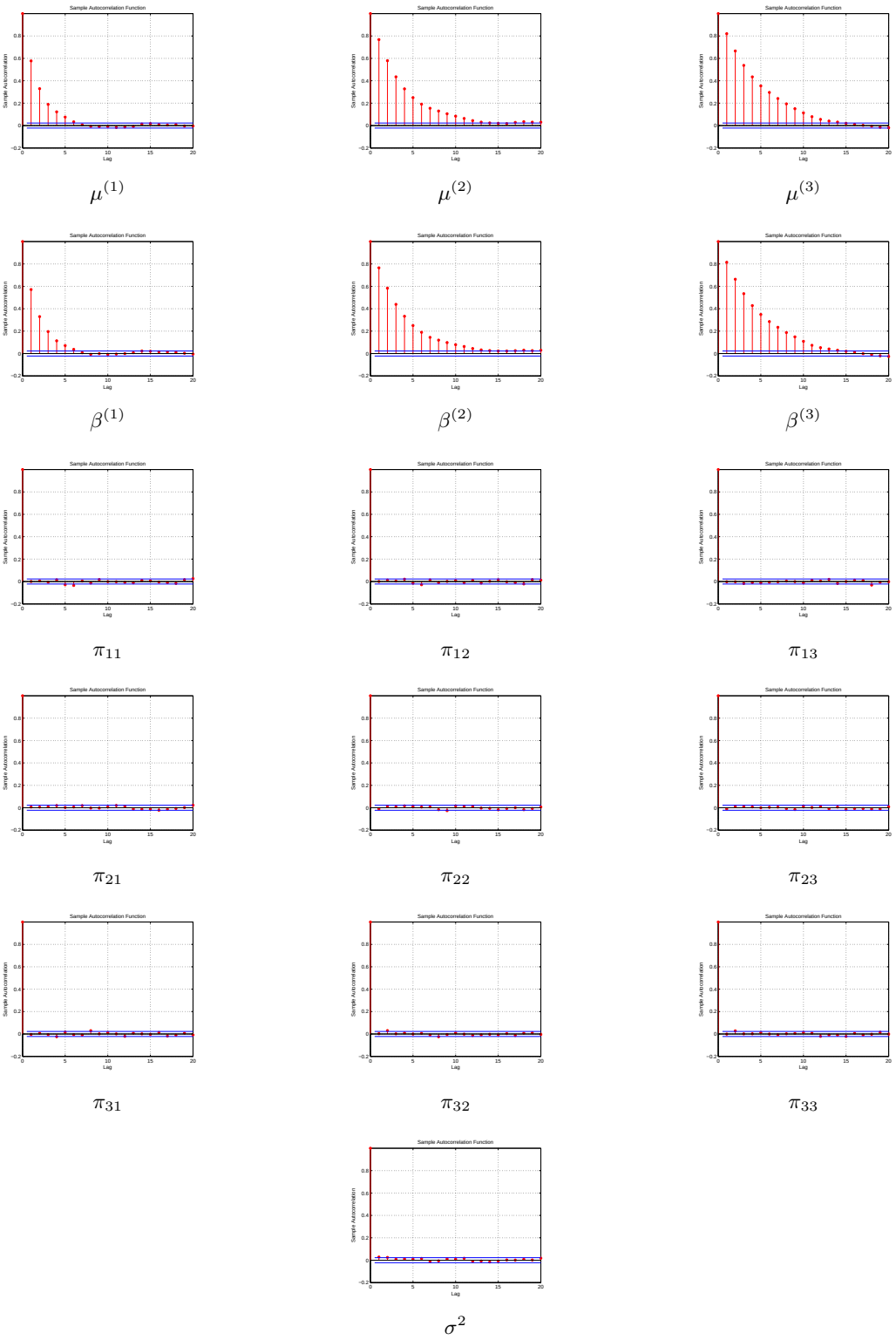


Figure 3.4: Autocorrelation sample plots for the parameters of the MAR(1)HMM model.

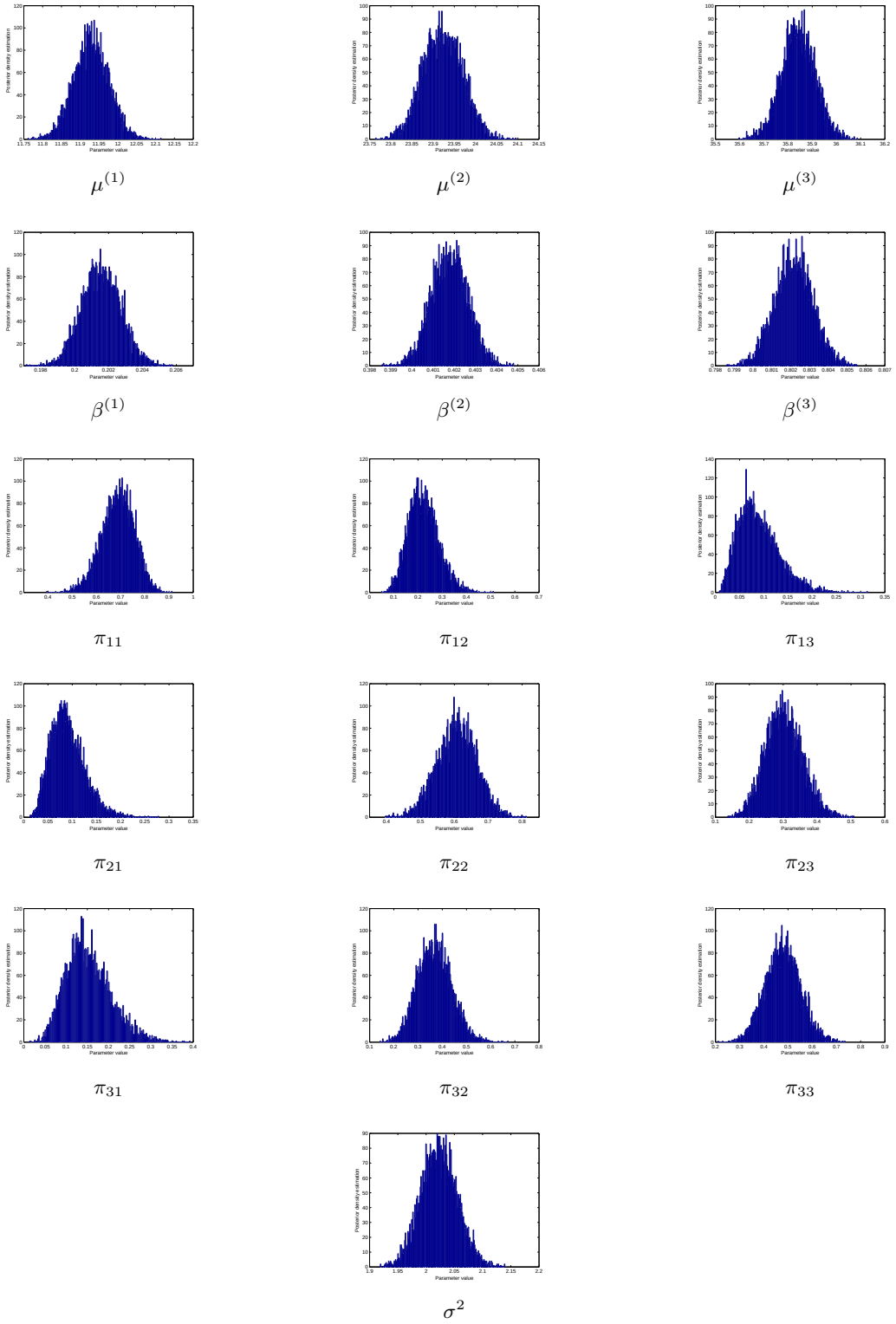


Figure 3.5: Posterior densities for the parameters of the MAR(1)HMM model (after 8000 iterations).

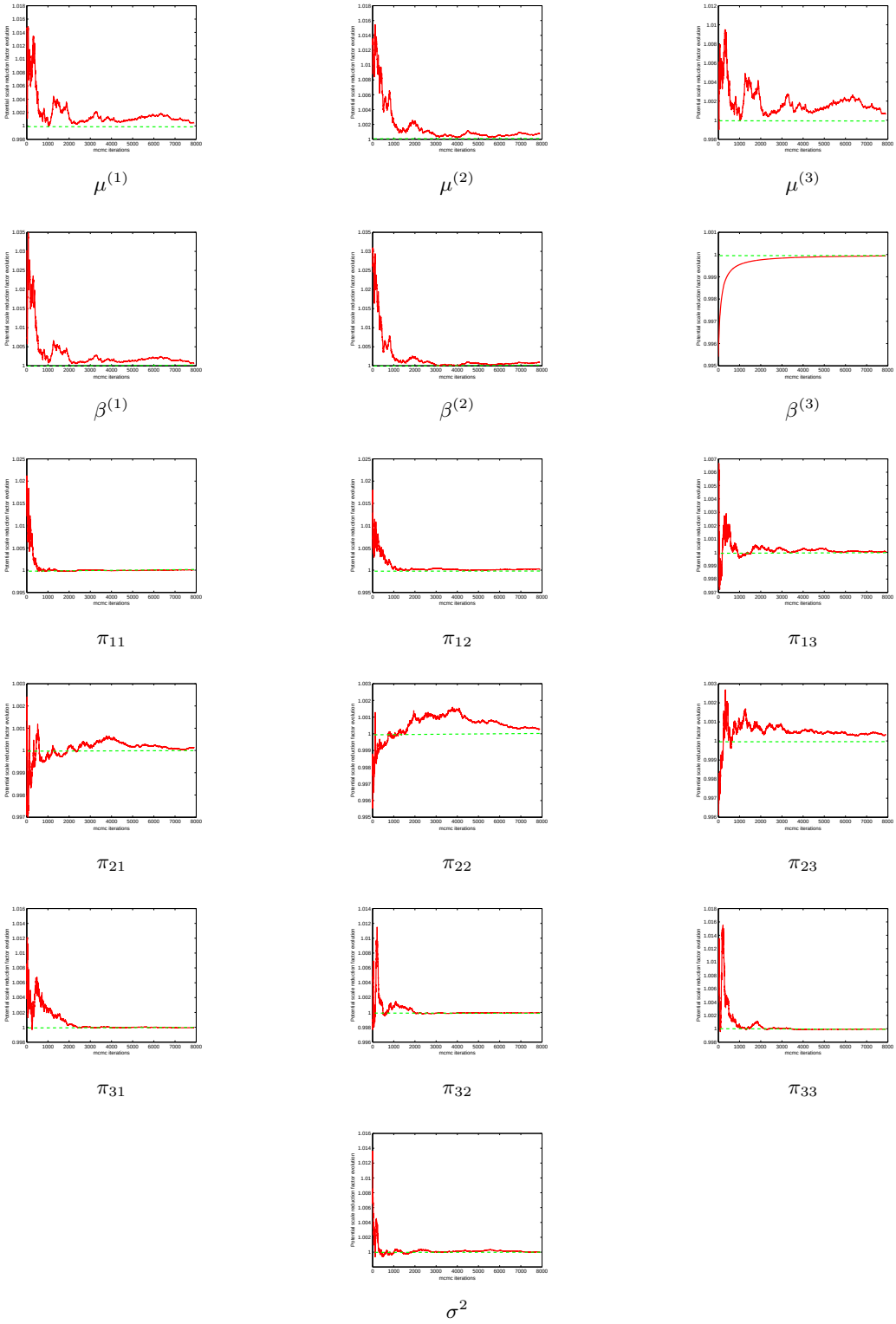


Figure 3.6: Potential scale reduction factor convergence to less than 1.02 with more iterations for all the parameters of the MAR(1)HMM model.

Parameter	True value	Posterior statistics		
		Mean	Standard deviation	Credible Interval (95%)
μ^1	12	11.929	0.047	(11.851-12.005)
μ^2	24	24.013	0.46	(23.97-24.041)
μ^3	36	36.01	0.21	(35.98-36.02)
β^1	0.2	0.2016	0.0012	(0.1997-0.2035)
β^2	0.4	0.4018	0.0009	(0.4004-0.4032)
β^3	0.8	0.8022	0.0010	(0.8005-0.8038)
π_{11}	0.7	0.688	0.068	(0.5715-0.797)
π_{12}	0.2	0.223	0.062	(0.129-0.332)
π_{13}	0.1	0.090	0.042	(0.032-0.17)
π_{21}	0.1	0.091	0.035	(0.041-0.154)
π_{22}	0.6	0.607	0.059	(0.507-0.701)
π_{23}	0.3	0.302	0.055	(0.214-0.397)
π_{31}	0.2	0.153	0.053	(0.075-0.250)
π_{32}	0.3	0.368	0.071	(0.257-0.488)
π_{33}	0.5	0.479	0.073	(0.358-0.599)
σ^2	2	2.023	0.032	(1.970-2.077)

Table 3.1: Posterior inference for the parameters of the MAR(1)HMM model.

3.6 The continuous time Markov switching autoregressive model of first order

In this section, we describe the CTMSAR(1) model, we formulate the likelihood. Later, we give the details of the Bayesian inference for both the parameters and the hidden Markov states but by taking into account the irregular time intervals. In this section, we will focus on the evaluation of the transition intensity matrix in the continuous time framework, and again we finish by a simulated example.

Briefly, the difference between the continuous time framework and the discrete case is that here the irregular time intervals are taken into account in the evaluation of the likelihood as well as when computing the posterior of the hidden states and the formulation of the posteriors of the parameters. Furthermore, while the transition probabilities are estimated using a Dirichlet posterior in the discrete case. Here, we will focus on the transition intensities that are computed using a MHA. For more details on the Bayesian estimation for the CTMSAR(1) model, one can refer to the work by **Hibbah and El-Maroufy** [71].

3.6.1 Model description and notations

We consider a continuous time Markov switching autoregressive model of first order (CTMSAR(1)); where given the hidden Markov process, the observations at irregular time points depends, as in the distrete case, on the previous ones through an autoregressive process of first order, see the diagram in figure 3.7 for an illustration.

We have n individuals with n_i the number of follow-up time for each individual i such that $1 \leq i \leq n$. We consider t_{ij} the clinical visit at time j for individual i; and the latent Markov disease state will be denoted x_{ij} for individual i at time t_{ij}. There are a different and discrete states for the Markov process (disease stages). These states are unknown and should be computed based on the observations. In breast cancer for example, a can be equal to 4 states: Benign breast cancer, primary breast cancer, local breast cancer and advanced breast cancer. The observations could be the concentration of the serum $CA\,15-3$ in the blood measured over a certain period of follow-up for some patients. The exception for this model is that the times of visits and the length of the follow-up periods are different from one patient to another.

Since the autoregressive observations are depending on the Markovian process, we get the following CTMSAR(1) model:

$$Y_{i,j}|X_{ij} = x_{ij} = \beta^{(x_{ij})}Y_{i,j-1} + \mu^{(x_{ij})} + \varepsilon_{i,j}. \tag{3.3}$$

The $\varepsilon_{i,j}$ are normal variables with mean 0 and variance σ^2 such that $\varepsilon_{i,j} \perp \varepsilon_{i\prime,j\prime}, (i,t) \neq (i\prime, j\prime)$.
The parameters $\beta^{(x_{ij})}$ and $\mu^{(x_{ij})}$ are real parameters and σ^2 is positive. we adopt a multinomial distribution for the initial distribution $r = (r_1, ..., r_a)$ for the first state of the Markov chain. Hence, a patient can be in any disease situation in his first visit. We denote $\mu = (\mu^{(1)}, ..., \mu^{(a)})$, and $\beta = (\beta^{(1)}, ..., \beta^{(a)})$. For example, $\mu^{(1)}$ and $\beta^{(1)}$ represent respectively the values for the autoregressive parameters corresponding to the state $a = 1$.
We should point out that since we place ourselves in the continuous time, the Markov process is usually represented by a transition intensity matrix of dimension $a \times a$ (noted here by $Q(t)$ for the time t), instead of the transition probability matrix $\Pi(t)$ as it is the case for the discrete time Markov chains. Using Chapman-Kolmogorov equation, these two matrices have the following relation:

$$\Pi(t) = \exp(Qt).$$

Let $\Theta = (\mu, \beta, \sigma^2, r, Q)$ be the set of all parameters in the model and $\theta = (\mu, \beta, \sigma^2)$ be the set of the autoregressive parameters. We adopt for individual i , $x_i^j = (x_{ij}, ..., x_{in_i})$ the future of the hidden states from time t_{ij} to time t_{in_i}, and $x_i^{-j} = (x_{i1}, ..., x_{ij})$ the past of the hidden states from time t_{i1} to time t_{ij}. The hidden states are not observed and must be estimated too. We use the same notations for the observations y_i^j and y_i^{-j}. Also, we should notice that we follow the convention that when we condition on the hidden states or Q, we intuitively condition on the observation times since we place ourselves in the continuous time case.
We suppose a continuous time homogeneous Markov matrix Π of first order with the states taking discrete values in the set $\{1, ..., a\}$, and under this Markov assumption, the distribution for the vector x of the hidden states given the parameters of the

model is expressed as:

$$\begin{aligned} P(x|\Theta &= P(x_{11},...,x_{1n_1},...,x_{n1},...,x_{nn_n}|\Theta) = \prod_{i=1}^{n} P(x_{i1},...,x_{in_i}|\Theta) \\ &= \prod_{i=1}^{n} P(x_{i1}|r) \prod_{j=2}^{n_i} P(x_{ij}|x_{ij-1},Q) \end{aligned}$$

Since we assume an autoregressive dependence between successive time observations given the hidden states concerning each individual i, we could write for the vector $y_{i,.} = (y_{i,1},...,y_{i,n_i})$:

$$P(y_{i,.}|x,\Theta) = P(y_{i,1}|x_{i1},\theta) \prod_{j=2}^{n_i} P(y_{i,j}|y_{i,j-1},x_{ij},\theta).$$

The likelihood formulation is very important and it is the beginning for any analysis and computation. One exception here is that the hidden states are unknown, thus we augment the density for the observations by the hidden states as in [134], consequently

$$\begin{aligned} P(y,x|\Theta) &= P(y|x,\Theta)P(x|Q), \text{ and} \\ P(y|x,\Theta) &= P(y_{1,.},...,y_{n,.}|x,\Theta) \\ &= \prod_{i=1}^{n} P(y_{i,1}|x_{i1},\theta) \prod_{j=2}^{n_i} P(y_{i,j}|y_{i,j-1},x_{ij},\theta). \end{aligned}$$

Under the hypothesis of normal errors distribution for the autoregressive parameters of the model and the Chapman-Kolmogorov property, the joint distribution of y and x given the parameters can be expressed as:

$$\begin{aligned} P(y,x|\Theta) &= \prod_{i=1}^{n} \prod_{h=1}^{a} r_h^{\chi_{\{x_{i1}\}}(h)} \\ &\times \prod_{j=2}^{n_i} \prod_{g=1}^{a} \prod_{h=1}^{a} (\Pi_{g,h}(t_{ij} - t_{ij-1}))^{\chi_{\{x_{ij-1},x_{ij}\}}(g,h)} \\ &\times \prod_{i=1}^{n} \prod_{h=1}^{a} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,1} - u^{(h)}}{\sigma} \right) \right]^{\chi_{\{x_{i1}\}}(h)} \\ &\times \prod_{j=2}^{n_i} \prod_{h=1}^{a} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,j} - u^{(h)} - \beta^{(h)} y_{i,j-1}}{\sigma} \right) \right]^{\chi_{\{x_{ij}\}}(h)}. \end{aligned}$$

Where ϕ denotes the density of a standard normal distribution, and $\chi_{\{A\}}(x)$ is the usual indicator function of a set A. Finally and based on the exponential expression

for the transition probabilities in continuous time, we obtain the likelihood:

$$\begin{aligned} P(y,x|\Theta) &= \prod_{i=1}^{n}\prod_{h=1}^{a} r_h^{\chi_{\{x_{i1}\}}(h)} \\ &\times \prod_{j=2}^{n_i}\prod_{g=1}^{a}\prod_{h=1}^{a}\left\{[\exp(\,Q(t_{ij}-t_{ij-1}))]_{g,h}\right\}^{\chi_{\{x_{ij-1},x_{ij}\}}(g,h)} \\ &\times \prod_{i=1}^{n}\prod_{h=1}^{a}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,1}-u^{(h)}}{\sigma}\right)\right]^{\chi_{\{x_{i1}\}}(h)} \\ &\times \prod_{j=2}^{n_i}\prod_{h=1}^{a}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,j}-u^{(h)}-\beta^{(h)}y_{i,j-1}}{\sigma}\right)\right]^{\chi_{\{x_{ij}\}}(h)}. \end{aligned} \tag{3.4}$$

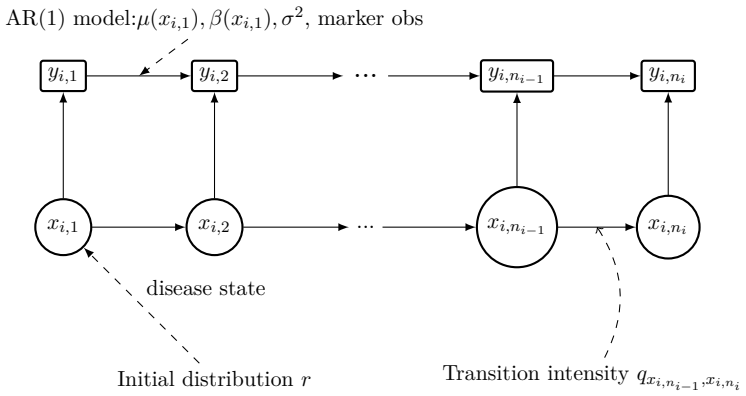


Figure 3.7: Diagram of CTMSAR(1): Times $t_{i,j}$ vary between minimum $t_{i,1}$ and t_{i,n_i}.

3.6.2 Exponential matrix and the conditions of existence of a valid generator

The likelihood formulation necessitates the computation of the exponential of the transition probability matrix $\Pi(t)$,which can not be normally obtained in a simple and explicit closed form. However, Peng [110] developed a theorem that provided a numerical form for Π in the case of finite state space Markov model using the eigenvalues of its infinitesimal generator Q (the details of the theorem is provided in Appendix E). Meanwhile, There are some major issues for empirical matrices such as the embeddability (or the existence at the time t of a generator that from its exponential we obtain the transition matrix); added to the uniqueness of the embedding and the effects of data/sampling error, since in reality we have only discrete times. Davies [39] discussed when a transition can be embeddable and proved

how to approximate a non-embeddable Markov matrix by an embeddable one.

Also, Israel *et al.* [75] identifies conditions under which a valid generator exists and proposes the post-adjustment of the elements to obtain a valid generator. In fact, many methods have been proposed for adjustment of the transition matrix to obtain the true generator. While to the computational implementation for the matrix exponential, Al-Mohy and Higham [6] adopted an efficient scaling and squaring algorithm for this matter.

3.6.3 Bayesian inference

Direct estimation of regime switching autoregressive (RSAR) processes is very difficult because the likelihood is known only up to a constant. we will appeal instead for the Bayesian inference by the means of Markov Chain Monte Carlo (MCMC) methods, because it is very acquainted to solve the issue of complicated formulation of the likelihood. In fact, the Bayesian inference is based on the posterior density of the model parameters; where the posterior density of a certain block of the parameters is proportional to the likelihood of the data times the prior density of that parameters'block. Moreover, the latent Markov states are not observed and should be computed, thus the full data likelihood is augmented by the hidden states as new variables. The Bayesian algorithm should alternate between the simulation of the parameters from their conditional posterior given the hidden states and the computation from the posterior of the hidden state given the parameters. If a closed form of the conditional probability distribution of one parameter is known, we use Gibbs sampler. Otherwise, we appeal for MHA to compute the desired parameter.

We will assume the independence between the parameters $\Theta = (r,\, Q,\, \mu,\, \beta,\, \sigma^2)$, consequently the prior density could be written as

$$P(\Theta) = P(r)P(Q)P(\mu)P(\beta)P(\sigma^2).$$

r are the parameters of a multinomial distribution, hence the natural choice for the prior would be a Dirichlet distribution $r \sim D(\alpha_{01}, ..., \alpha_{0a})$. Furthermore, concerning the priors for parameters of the autoregressive model, we suppose for $[h = 1, ..., a]$: $\mu^{(h)} \sim \mathcal{N}(\alpha_h, \tau_h)$, $\beta^{(h)} \sim \mathcal{N}(b_h, c_h)$ and $\sigma^2 \sim IG(\varepsilon, \zeta)$. For more details on how those priors are chosen refer to Cappe and Ryden [25]. For the transition intensity matrix; we consider in this work a component-wise sampling for each element q_{ij}, $i \neq j$; where we consider an exponential prior as in the work of Siekmann *et al.* [127]; $P(q_{ij}) \sim \frac{\exp(Tr(Q))}{C}$. C is a constant to be chosen and Tr is trace matrix. This prior choice guarantees that the rate transitions can not have arbitrarily large values. Also, we should mention that we could also appeal for the block-wise sampling of the whole transition intensity matrix as in [100]; where the block $\{q_{ij}, i \neq j\}$ was updated from the prior $\prod_{i=1}^{a} \prod_{j=1, j\neq i}^{a} q_{ij}^{e_{ij}-1} \exp\left(-\frac{q_{ij}}{f_{ij}}\right)$, for the hyper-parameters e_{ij} and f_{ij}.

$\alpha_{01}, ..., \alpha_{0a}$, α_h, τ_h, b_h, c_h, ε, ζ are hyper-parameters to be specified carefully.

A- Sampling the posterior distribution for the hidden states:

We will extend here the block update of the hidden states for the MAR(1)HMM) by **El-Maroufy *et al.*** [47] in the discrete case to the continuous time case; where we simulate the full latent data $x_{i,.} = (x_{i1}, ..., x_{in_i})$ for each individual.

With $y_{i,.} = (y_{i1}, ..., y_{in_i})$, we write the full conditional for each patient i, $i = 1, ..., n_i$ as follows

$$\begin{aligned} P(x_{i,.}|y_{i,.}, \Theta) &= P(x_{in_i}|y_{i,.}, \Theta)P(x_{in_{i-1}}|y_{i,.}, x_{in_i}, \Theta) \times ... \\ &\times \; P(x_{ij}|y_{i,.}, x_i^{j+1}, \Theta) \times ...P(x_{i1}/y_{i,.}, x_i^2, \Theta). \end{aligned}$$

Thus to implement this sampling scheme, it is sufficient to sample each individual hidden state x_{ij} from the term $P(x_{ij}|y_{i,.}, x_i^{j+1}, \Theta)$. Moreover,

$$P(x_{ij}|y_{i,.}, x_i^{j+1}, \Theta) \propto P(x_{ij}|y_i^{-j}, \Theta)P(x_{ij+1}|x_{ij}, Q). \tag{3.5}$$

This simplification contains two terms: $P(x_{ij+1}|x_{ij}, Q)$ which is just the transition probability. While the first term is evaluated through forward calculation by first initializing $P(x_{i1}|y_{i,1}, \Theta) \propto P(y_{i,1}|x_{i1}, \theta)P(x_{i1}|r)$. Then for $j = 2, ..., n_i$:

$$P(x_{ij} = k|y_i^{-(j-1)}, \Theta) \propto \sum_{l=1}^{a} \pi_{lk}(t_{ij} - t_{ij-1})P(x_{ij-1}|y_i^{-(j-1)}, \Theta) \text{ and}$$
$$P(x_{ij} = k|y_i^{-j}, \Theta) \propto P(x_{ij} = k|y_i^{-(j-1)}, \Theta)f(y_{i,j}|y_{ij-1}, \theta_k),$$

for each $k \in \{1, ..., a\}$. f is the density for the observations obtained from the formulation in (3.4). At the end of this forward procedure we obtain $P(x_{in_i}|y_i^{-n_i}, \Theta)$ that would serve as start for backward simulation for finally getting for $j = n_i - 1, ..., 2, 1$: $P(x_{ij}/y_{i,.}, x_i^{j+1}, \Theta)$ using 3.5. These last backward computations would permit to find the hidden states $(x_{i1}, ..., x_{in_i})$ for $i = 1, ..., n$. The algorithm is summarized in the following steps:

Algorithm 3.3

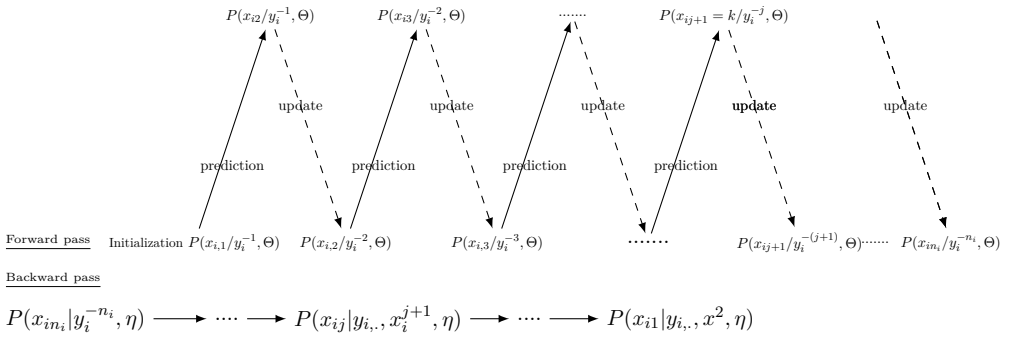
- **Forward filtering:**
  - i) Initialization of the first state from $P(x_{i1}|y_{i,1}, \Theta) \propto P(y_{i,1}|x_{i1}, \theta)P(x_{i1}|r)$.
  - ii) For $j = 2, ..., n_i$: Simulate $P(x_{ij} = k|y_i^{-(j-1)}, \Theta) \propto \sum_{l=1}^{a} \pi_{lk}(t_{ij} - t_{ij-1})P(x_{ij-1}|y_i^{-(j-1)}, \Theta)$.
  - iii) Simulate $P(x_{ij} = k|y_i^{-j}, \Theta) \propto P(x_{ij} = k|y_i^{-(j-1)}, \eta)f(y_{i,j}|y_{ij-1}, \theta_k)$.
  - iv) Finish forward pass by computing $P(x_{in_i}|y_i^{-n_i}, \Theta)$.
- **Backward smoothing:**
  - i) Initialize from $P(x_{in_i}|y_i^{-n_i}, \Theta)$.

ii) for $j = n_i - 1, ..., 2, 1$: Calculate $P(x_{ij}|y_{i,.}, x_i^{j+1}, \Theta)$ using (3.5).

iii) Use those last probabilities to sample the hidden states.

The continuous time FFBS algorithm for the hidden states takes into account the irregular time intervals, and each time we have to compute the transition matrix from the exponential of the transition intensity matrix. Figure 3.8 illustrates all the steps of the algorithm.

Figure 3.8: FFBS algorithm



B- Sampling from $P(\theta|x, y)$

We will start by detailing the steps for computing the rates of the transition intensity, then we pass to the simulation of the autoregressive parameters.

Sampling Q

We would adopt to simulate each element $q_{ij}, i \neq j$ separately . Using (3.4) and the fact that the posterior for Q is proportional to the prior multiplied by the likelihood; by keeping only the terms in the likelihood that depends on the q_{ij} we obtain :

$$P(q_{ij}|y, x) \propto \pi(q_{ij}) \times \prod_{i=1}^{n} \prod_{j=2}^{n_i} \prod_{g=1}^{a} \prod_{h=1}^{a} \left\{ [\exp(\, Q(t_{ij} - t_{ij-1}))]_{g,h} \right\}^{\chi_{\{x_{ij-1}, x_{ij}\}}(g,h)}.$$

Hence by first supposing the exponential prior for each $q_{ij}, i \neq j$: $\pi(q_{ij}) \sim \frac{\exp(Tr(Q))}{C}$.
Let us also pose

$$g(q_{ij}/y, x) = \prod_{i=1}^{n} \prod_{j=2}^{n_i} \prod_{g=1}^{a} \prod_{h=1}^{a} \left\{ [\exp(\, Q(t_{ij} - t_{ij-1}))]_{g,h} \right\}^{\chi_{\{x_{ij-1}, x_{ij}\}}(g,h)}, \text{ we get}$$

$$P(q_{ij}|y, x) \propto \pi(q_{ij}) g(q_{ij}|y, x).$$

This is a non standard distribution and we recourse to MHA to draw the transition rates q_{ij}. So, a proposal distribution should be fixed to evaluate the MHA ratio. Since symmetric distribution like normal one is not efficient for sampling positive values of the $q_{i,j}$, a Gamma $\mathscr{G}(\delta_1, \delta_2)$ distribution is proposed to keep

operating on real positive numbers. Moreover, the hyper-parameters δ_1 and δ_2 could be chosen arbitrarily, but here they are chosen based on maximum likelihood estimation for the q_{ij} (refer to Li and Chan [92]). Consequently, we pose $\delta_1 = 0.001 + \sum_{l=1}^{n}\sum_{k=2}^{n_i} \chi_{x_{kl-1},x_{kl}}(i,j)$ and $\delta_2 = 0.01 + T_i$, where T_i is the time spent in state i. This choice is to ensure q_{ij} draws with approximately a mean of $\frac{\delta_1}{T_i}$ and accelerate the convergence, by the way the MHA sampling ratio; as adapted by **Hibbah and El-Maroufy** [71] would be:

$$\alpha = \min\left(1, \frac{g(\hat{q_{ij}})\pi(\hat{q_{ij}})}{g(q_{ij})\pi(q_{ij})} \times co\right), \tag{3.6}$$

Where co is the proposal factor ratio between the Gamma proposal density of the new value and the preview value for the q_{ij}. Precisely, the MHA proceeds as follows:

Algorithm 3.4

---

i) Calculate the hyper-parameters $\delta_1 = 0.001 + \sum_{l=1}^{n}\sum_{k=2}^{n_i} \chi_{x_{kl-1},x_{kl}}(i,j)$ and $\delta_2 = 0.01 + T_i$ for the Gamma distribution.

ii) Propose a new q_{ij}^* from $\Gamma(\delta_1, \delta_2)$.

iii) Compute the exponential prior for the new q_{ij}^* and the old q_{ij} from $\frac{\mathscr{E}(Tr(Q))}{C}$.

iv) Compute the ratio co between the Gamma proposal density for the old q_{ij} and the Gamma proposal density for the new q_{ij}^*.

v) Calculate the proportional posterior likelihood g for the new q_{ij}^* and the old q_{ij}.

vi) Compute

$$\alpha = \min\left(1, \frac{g(\hat{q_{ij}})\pi(\hat{q_{ij}})}{g(q_{ij})\pi(q_{ij})} \times co\right),$$

vii) Draw a uniform random $\mathscr{U}(0,1)$ to decide on the rejection or the acceptance of the new value q_{ij}^* by comparing it to α.

This MHA ratio requires the computation of a matrix exponential which is a notorious computation, even though the availability of the explicit form of Peng [110]. So, we appeal for a combination of a scaling and squaring, and Pade approximation as described by Al-Mohy and Higham [6]; a method that gives good precision and it is available in many softwares.

Sampling the posterior distribution for initial distribution $P(r/x,y)$

Let $n_{0l}=\sum_{i=1}^{n}\chi_{x_{i1}}(l)$, for $l=1,...,a$. From (3.4), we have

$$\begin{aligned}
P(y,x|\Theta) &= \prod_{i=1}^{n}\prod_{h=1}^{a} r_h^{\chi_{\{x_{i1}\}}(h)} \\
&\times \prod_{j=2}^{n_i}\prod_{g=1}^{a}\prod_{h=1}^{a}\left\{[\exp(Q(t_{ij}-t_{ij-1}))]_{g,h}\right\}^{\chi_{\{x_{ij-1},x_{ij}\}}(g,h)} \\
&\times \prod_{i=1}^{n}\prod_{h=1}^{a}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,1}-u^{(h)}}{\sigma}\right)\right]^{\chi_{\{x_{i1}\}}(h)} \\
&\times \prod_{j=2}^{n_i}\prod_{h=1}^{a}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,j}-u^{(h)}-\beta^{(h)}y_{i,j-1}}{\sigma}\right)\right]^{\chi_{\{x_{ij}\}}(h)}.
\end{aligned}$$

By keeping only the first term that depends on r from the joint mass, and under a Dirichlet prior $D(\alpha_{01},...,\alpha_{0a})$ for the parameter r ($P(r)=\prod_{h=1}^{a} r_h^{\alpha_{0h}-1}$), we obtain:

$$\begin{aligned}
P(r|x,y) &\propto \prod_{h=1}^{a} r_h^{\alpha_{0l}-1}\prod_{i=1}^{n}\prod_{h=1}^{a} r_h^{\chi_{\{x_{i1}\}}(h)} \\
&\propto \prod_{h=1}^{a} r_l^{\alpha_{0h}+n_{0h}-1}. \\
&\propto \mathscr{D}(\alpha_{01}+n_{01},...,\alpha_{0a}+n_{0a}).
\end{aligned}$$

Thus, we could draw the initial distribution r from this Dirichlet distribution using Gibbs sampler.

The posterior distributions for the autoregressive parameters μ,β,σ^2:

When a complete posterior conditional distribution is known such as the normal distribution or the beta distribution, we use the Gibbs sampler to draw the random variable, this is the case for the autoregressive parameters. let us define $n_{0l}=\sum_{i=1}^{n}\chi_{\{x_{i1}=l\}}, n_0=\sum_{l=1}^{a} n_{0l}, N_l=\sum_{i=1}^{n}\sum_{j=2}^{n_i}\chi_{\{x_{ij}=l\}}, N=\sum_{l=1}^{a} N_l$. By supposing $\mathscr{N}(\alpha_l,\tau_l)$ as a prior distribution for $\mu^{(l)}$ for $l=1,2,....,a.$, and using (3.4), the conditional posterior distribution of $\mu^{(l)}$ is the prior multiplied by the likelihood:

$$\begin{aligned}
P(\mu^{(l)}|y,x) &\propto P(\mu^{(l)})\prod_{i=1}^{n}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,1}-u^{(l)}}{\sigma}\right)\right]^{\chi_{\{x_{i1}\}}(l)} \\
&\times \prod_{i=1}^{n}\prod_{j=2}^{n_i}\left(\frac{1}{\sigma}\phi(\frac{y_{i,j}-u^{(l)}-\beta^{(l)}y_{i,j-1}}{\sigma})\right)^{\chi_{\{x_{ij}\}}(l)} \\
&\propto \exp\frac{-1}{2}\left\{\frac{(\mu^{(l)}-\alpha_l)^2}{\tau_l}+\sum_{i=1,x_{i1}=l}^{n}\left(\frac{y_{i,1}-\mu^{(l)}}{\sigma}\right)^2+\right. \\
&\qquad \left.\sum_{i=1}^{n}\sum_{j=2,x_{ij}=l}^{n_i}\left(\frac{y_{i,j}-\mu^{(l)}-\beta^{(l)}y_{i,j-1}}{\sigma}\right)^2\right\}.
\end{aligned}$$

Then,

$$\mu^{(l)}|y,x \sim \mathcal{N}(\alpha_{1l},\tau_{1l}). \tag{3.7}$$

with inverse variance

$${\tau_{1l}}^{-1} = \frac{n_{0l}+N_l}{\sigma^2}+\frac{1}{\tau_l}.$$

and mean

$$\alpha_{1l} = \tau_{1l}\left(\frac{\sum\limits_{i=1,x_{i1}=l}^{n} y_{i,1} + \sum\limits_{i=1}^{n}\sum\limits_{j=2,x_{ij}=l}^{n_i}(y_{i,j}-\beta^{(l)}y_{i,j-1})}{\sigma^2}+\frac{\alpha_l}{\tau_l}\right).$$

For $\beta^{(l)}, l=1,...,a$, we use the same technique as for $\mu^{(l)}$, we found that under $\mathcal{N}(b_l,c_l)$ prior, the conditional posterior distribution (likelihood $\times$ the prior) of σ^2 gives after keeping only the terms in the conditional posterior that are dependent on $\beta^{(l)}$:

$$\begin{aligned} P(\beta^{(l)}|y,x) &\propto P(\beta^{(l)})\prod_{i=1}^{n}\prod_{j=2}^{n_i}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,j}-u^{(l)}-\beta^{(l)}y_{i,j-1}}{\sigma}\right)\right]^{\chi_{\{x_{ij}\}}(l)} \\ &\propto \exp\frac{-1}{2}\left\{\frac{(\beta^{(l)}-b_l)^2}{c_l}+\right. \\ &\qquad \left.\sum_{i=1}^{n}\sum_{j=2,x_{ij}=l}^{n_i}\left(\frac{y_{i,j}-\mu^{(l)}-\beta^{(l)}y_{i,j-1}}{\sigma}\right)^2\right\}. \end{aligned}$$

Therefore,

$$\beta^{(l)}|y,x \sim \mathcal{N}(b_{1l},c_{1l}). \tag{3.8}$$

with inverse variance

$$c_{1l}^{-1} = \frac{1}{c_l}+\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=2,x_{ij}=l}^{n_i} y_{i,j-1}^2}{\sigma^2}.$$

and mean

$$b_{1l} = c_{1l}\left(\frac{b_l}{c_l}+\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=2,x_{ij}=l}^{n_i}(y_{i,t}-\mu^{(l)})y_{i,j-1}}{\sigma^2}\right).$$

For the conditional posterior distribution (likelihood $\times$ the prior) of σ^2, by supposing $IG(\varepsilon,\zeta)$ as prior; we deduce from (3.4) after discarding terms in the conditional posterior that are independent from σ^2:

$$\begin{aligned} P(\sigma^2|y,x) &\propto (\sigma^2)^{-(\varepsilon+1)}\exp\left(-\frac{\zeta}{\sigma^2}\right)\prod_{i=1}^{n}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,1}-u^{(x_{i1})}}{\sigma}\right)\right] \\ &\times \prod_{i=1}^{n}\prod_{j=2}^{n_i}\left[\frac{1}{\sigma}\phi\left(\frac{y_{i,j}-u^{(x_{ij})}-\beta^{(x_{ij})}y_{i,j-1}}{\sigma}\right)\right]. \end{aligned}$$

Consequently,

$$\sigma^2|y,x \sim IG(\varepsilon_1,\zeta_1). \tag{3.9}$$
$$\text{with parameters } \varepsilon_1 = \varepsilon + \frac{n_0+N}{2}.$$
$$\text{and } \zeta_1 = \zeta + \frac{\sum_{i=1}^{n}(y_{i,1}-\mu^{(x_{i1})})^2 + \sum_{i=1}^{n}\sum_{j=2}^{n_i}(y_{i,j}-\mu^{(x_{ij})}-\beta^{(x_{ij})}y_{i,j-1})^2}{2}.$$

Finally the MCMC algorithm can be summarized as follows:

Algorithm 3.5

---

i) for $h = 1, 2, ..., a$, give reference values for the hyper-parameters α_h, τ_h, b_h, c_h, and ε, and ζ.

ii) Initialization (Step $d = 1$ of the MCMC iterations):
Initialize $Q^{(1)}$, $r^{(1)}$, $\mu_{(1)}$, $\beta_{(1)}$ and $\sigma^{2(1)}$.

iii) Computation of the transition matrix π using formula $\Pi(t) = \exp(Qt)$ for each observation time t.

iv) Simulation of the hidden states by FFBS algorithm using formula (3.5.)

v) Estimation of the initial distribution:
$P(r|x,y) \propto \mathscr{D}(\alpha_{01}+n_{01}, ..., \alpha_{0a}+n_{0a})$,
such that $n_{0l} = \sum_{i=1}^{n} \chi_{\{}x_{i1}\})(l)$, for $l = 1, ..., a$.

vi) Simulation of the transition intensity matrix Q using (3.6).

vii) Simulation of μ using the distribution in (3.7).

viii) Simulation of β using the distribution in (3.8)

ix) Simulation of σ^2 using the distribution in (3.9).

3.6.4 Simulation study

Before providing our simulation algorithm, we should point out that our focus is to assess the usefulness of a bio-marker in disease development. Potential of bio-marker applications include the prediction of the response to specific therapeutic interventions such as the serum $HER-2$ neu that predicts response to a treatment called trastuzumab in breast cancer. Also, the bio-marker can predict the cancer recurrence or disease progression in the future. An example of a prognostic cancer

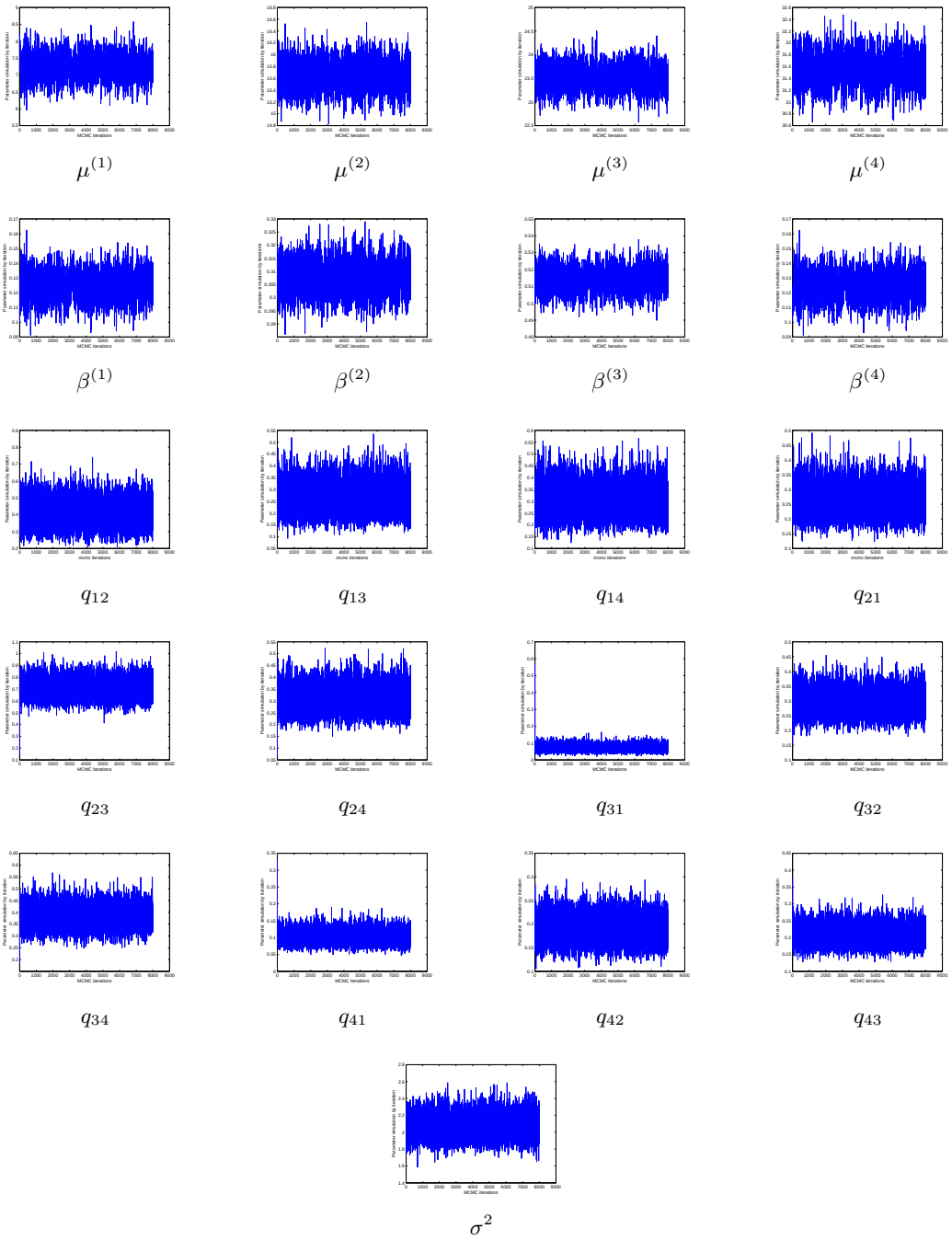
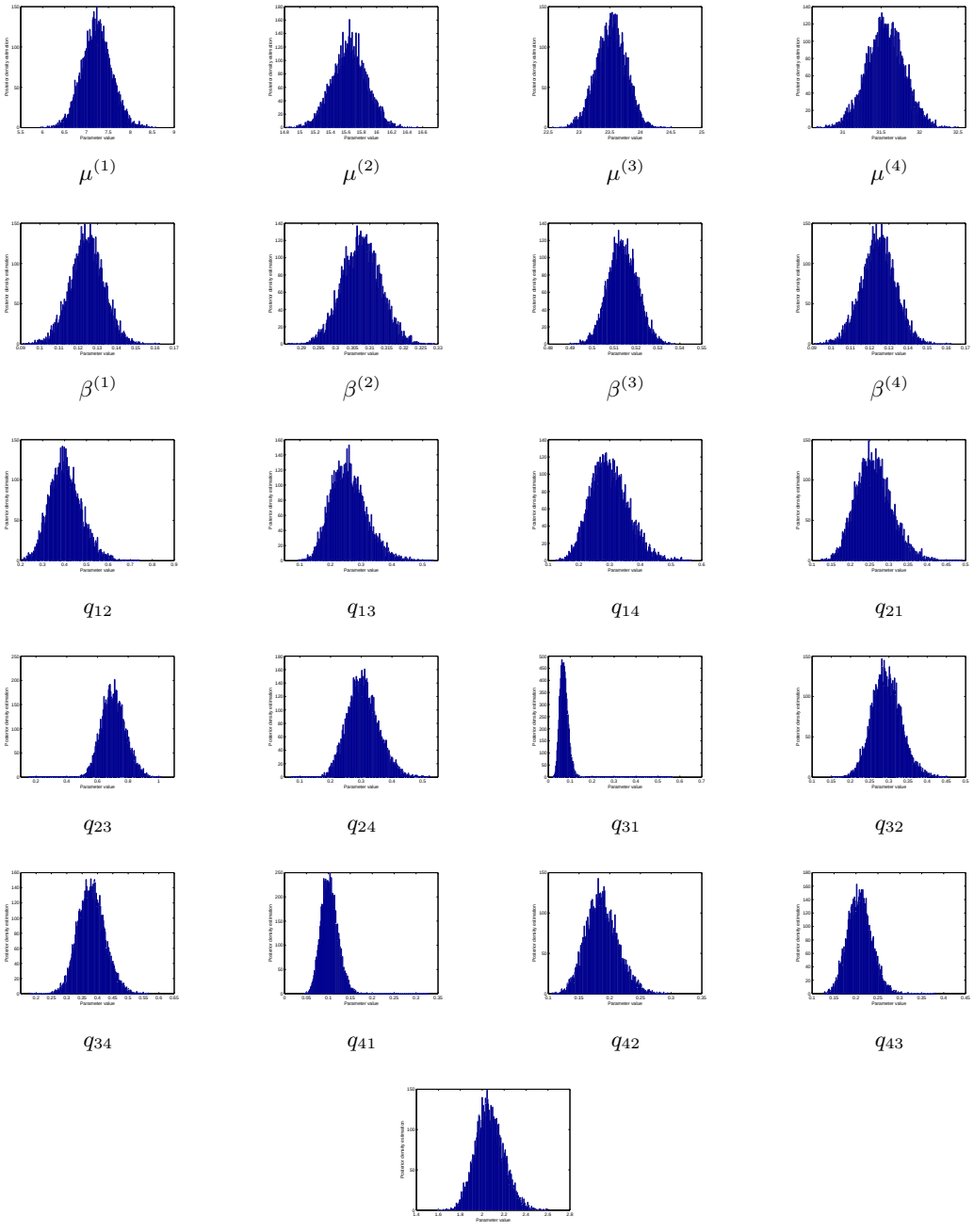


Figure 3.9: 8000 MCMC iteration plots for the parameters of the CTMSAR(1) model



$\mu^{(1)}$ $\mu^{(2)}$ $\mu^{(3)}$ $\mu^{(4)}$

$\beta^{(1)}$ $\beta^{(2)}$ $\beta^{(3)}$ $\beta^{(4)}$

q_{12} q_{13} q_{14} q_{21}

q_{23} q_{24} q_{31} q_{32}

q_{34} q_{41} q_{42} q_{43}

σ^2

Figure 3.10: Posterior density plots for the parameters of the CTMSAR(1) model

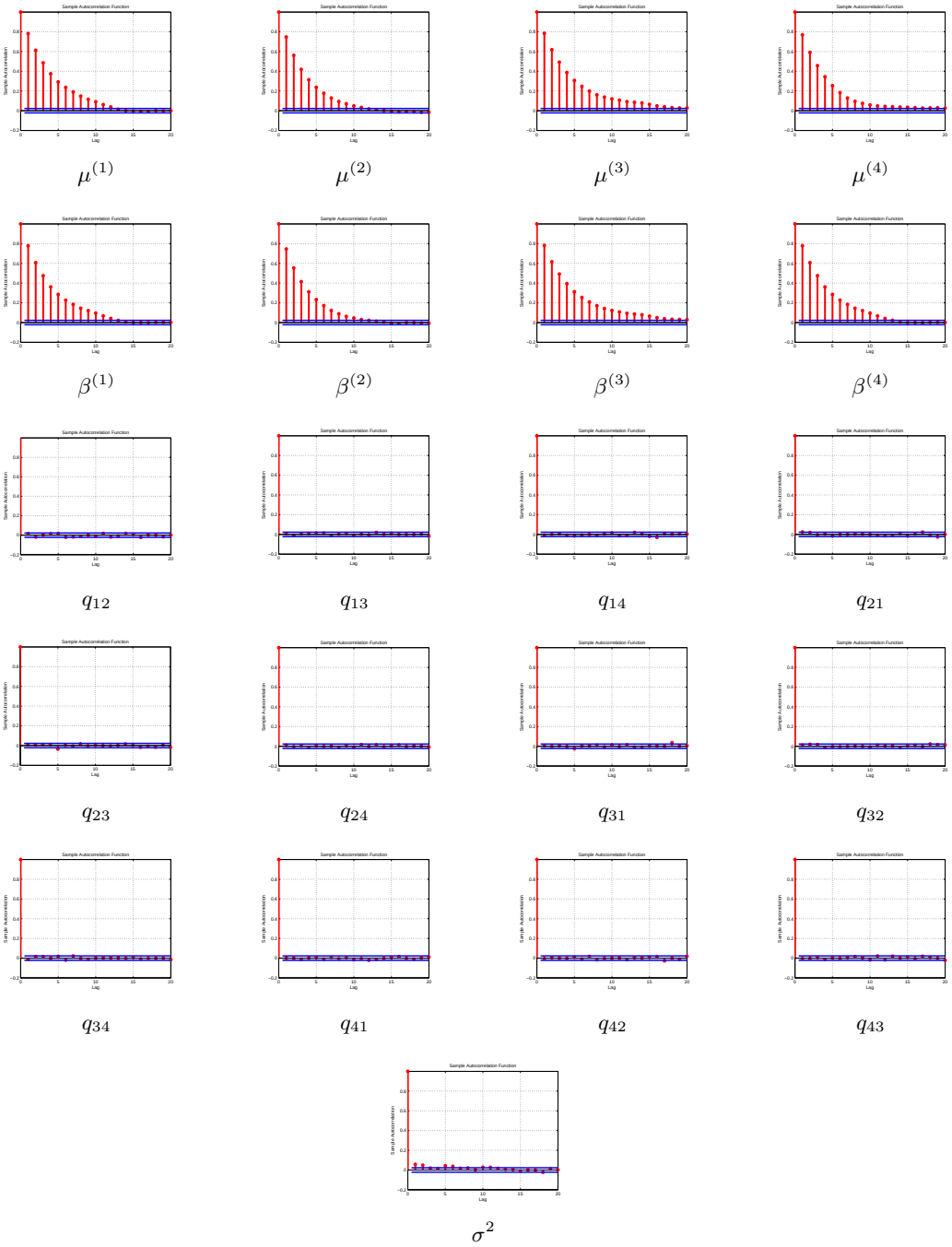
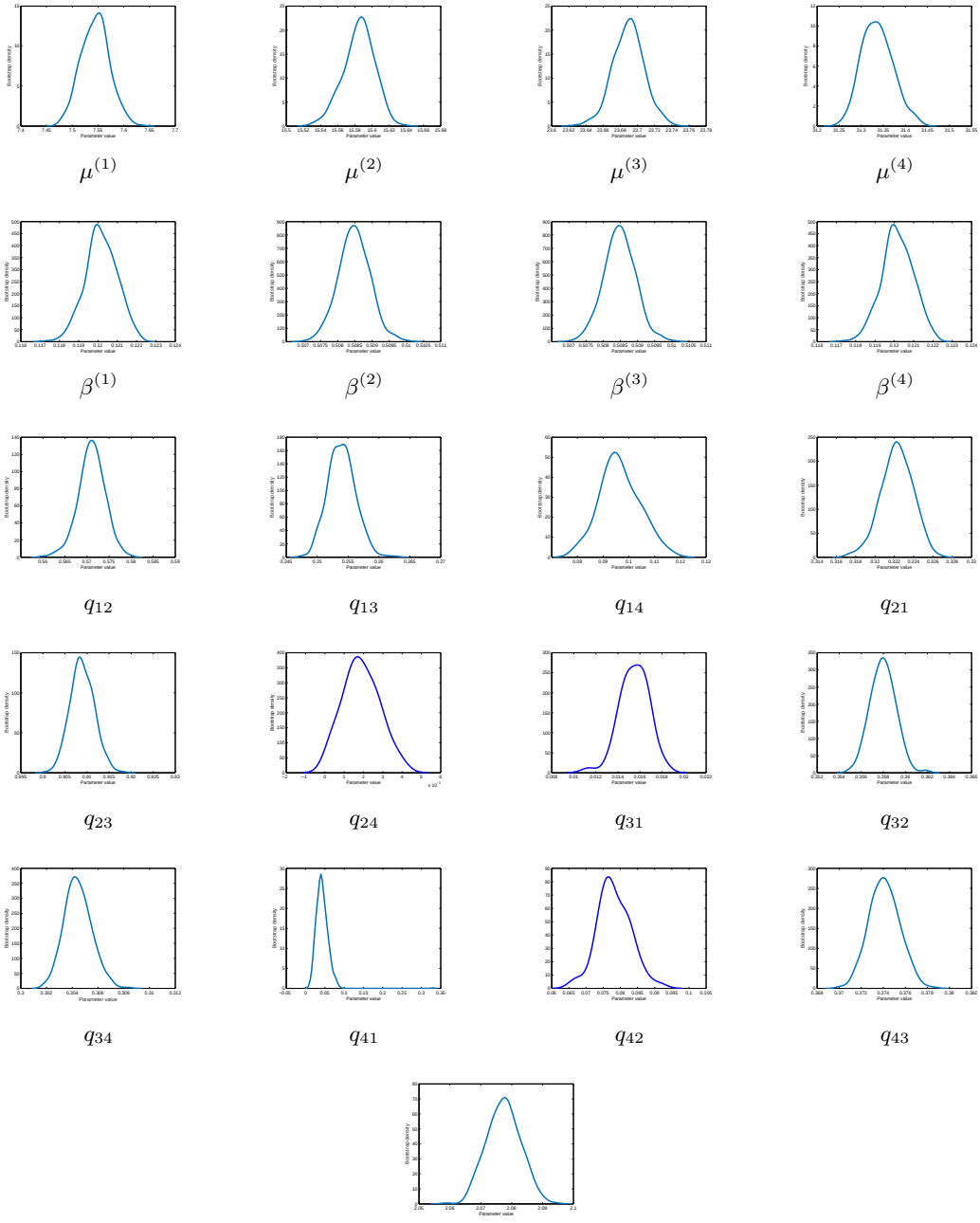


Figure 3.11: Autocorrelation sample plots for the parameters of the CTMSAR(1) model



$\mu^{(1)}$ $\mu^{(2)}$ $\mu^{(3)}$ $\mu^{(4)}$

$\beta^{(1)}$ $\beta^{(2)}$ $\beta^{(3)}$ $\beta^{(4)}$

q_{12} q_{13} q_{14} q_{21}

q_{23} q_{24} q_{31} q_{32}

q_{34} q_{41} q_{42} q_{43}

σ^2

Figure 3.12: 1000 bootstrap kernel density for the parameters of the CTMSAR(1) model

Table 3.2: MCMC estimation for the parameters of the CTMSAR(1) model (8000 iterations)

		Posterior computations		
Parameters	True value	Mean	Standard deviation	Credible interval (95%)
μ^1	8	8.03	0.68	(7.98-8.04)
μ^2	16	15.65	0.23	(15.21-16.09)
μ^3	24	24.01	0.25	(23.98-24.03)
μ^4	32	31.56	0.24	(31.07-32.04)
β^1	0.1	0.12	0.009	0.11-0.14)
β^2	0.3	0.31	0.006	(0.3-0.32)
β^3	0.5	0.51	0.007	(0.5-0.53)
β^4	0.7	0.0.71	0.007	(0.7-0.73)
q_{12}	0.4	0.4	0.07	(0.28-0.56)
q_{13}	0.25	0.25	0.058	(0.16-0.39)
q_{14}	0.35	0.3	0.062	(0.19-0.43)
q_{21}	0.3	0.26	0.048	(0.18-0.36)
q_{23}	0.7	0.71	0.078	(0.56-0.87)
q_{24}	0.4	0.31	0.051	(0.21-0.41)
q_{31}	0.1	0.072	0.02	(0.04-0.11)
q_{32}	0.3	0.29	0.061	(0.22-0.37)
q_{34}	0.4	0.38	0.044	(0.3-0.47)
q_{41}	0.1	0.1	0.02	(0.07-0.14)
q_{42}	0.18	0.19	0.026	(0.14-0.24)
q_{43}	0.2	0.21	0.028	(0.16-0.27)
σ^2	2	2.07	0.13	(1.83-2.34)

bio-marker is the $21-gene$ recurrence tamoxifen-treated breast cancer. Another class of bio-markers is the diagnostic bio-marker; that assesses the disease condition (for more details see, Goossens *et al.* [64]).

Our model considers the disease stages; the x_{ij} as unknown parameters, since the goal will be to predict the states at any time based on bio-marker measurements (the observations y_{ij}). Here, we used the FFBS algorithm to compute the states. Another benefit of our process is to compute the transition intensities q_{ij} that are important tools to compare if the transition intensities have changed between two patient groups under different treatments for example; without forgetting the use of the $1/q_{ii}$ to assess the mean sojourn time in a state. As to the autoregressive parameter; for example if we get $\mu = (8, 16, 24, 32)$ and $\beta = (0.1, 0.3, 0.5, 0.7)$, this means that autoregressive intercept for stage 1 is 8; which provides an indication on the value of the marker in stage 1. Furthermore, having $\beta = 0.5$ informs us about the mean amplitude of progression between successive observations in state 3.

To assess how the MCMC algorithm can accurately estimate the model param-

Table 3.3: Bootstrap inference (1000 scenarios) for the parameters of the CTM-SAR(1) model (5000 MCMC iterations in each scenario).

	Bootstrap inference					
Parameters	True value	Mean	Squared bias	Variance	MSE	Credible Interval (95%)
μ^1	8	7.5444	0.2076	$7.4195e^{-04}$	0.2083	(7.4905-7.5987)
μ^2	16	15.5846	0.1726	$3.2893e^{-04}$	0.1729	(15.5471-15.6174)
μ^3	24	23.6894	0.0965	$3.2930e^{-04}$	0.0968	(23.6516- 23.7263)
μ^4	32	31.3366	0.4401	0.0012	0.4413	(31.2731- 31.4142)
β^1	0.1	0.1202	0.0408	$6.7260e^{-07}$	0.0408	(0.1185- 0.1217)
β^2	0.3	0.3090	$8.1000e^{-05}$	$2.4841e^{-07}$	$8.1248e^{-05}$	(0.3082-0.3100)
β^3	0.5	0.5085	$7.2250e^{-05}$	$2.0959e^{-07}$	$7.2460e^{-05}$	(0.5076-0.5094)
β^4	0.7	0.7175	$3.0625e^{-04}$	$8.2046e^{-07}$	$3.0707e^{-04}$	(0.7155-0.7192)
q_{12}	0.586	0.5696	$2.6896e^{-04}$	$8.2847e^{-04}$	0.0011	(0.5642-0.5771)
q_{13}	0.254	0.2540	0.000	$4.9238e^{-06}$	$4.9238e^{-06}$	(0.2499-0.2584)
q_{14}	0.0710	0.0623	$7.6260e^{-05}$	$5.3807e{-}05$	$1.3007e^{-04}$	(0.0495-0.0784)
q_{21}	0.3261	0.3223	$1.4440e^{-05}$	$2.6672e^{-06}$	$1.7107e^{-05}$	(0.3188-0.3253)
q_{23}	1.0230	0.9089	0.0130	$7.3389e^{-06}$	0.0130	(0.9038-0.9145)
q_{24}	0.0020	0.0019	$1.4991e^{-08}$	$8.5465e^{-07}$	$8.6964e^{-07}$	(0.0003-0.0038)
q_{31}	0.0170	0.0168	$4.6404e^{-08}$	$1.9864e^{-06}$	$2.0328e^{-06}$	(0.0138-0.0200)
q_{32}	0.349	0.3580	$8.1000e^{-05}$	$1.3123e^{-06}$	$8.2312e^{-05}$	(0.3559-0.3601)
q_{34}	0.342	0.3044	0.0014	$1.1073e^{-06}$	0.0014	(0.3026-0.3067)
q_{41}	0.0310	0.0344	$1.1558e^{-05}$	$3.1584e^{-06}$	$1.4717e^{-05}$	(0.0311-0.0387)
q_{42}	0.0860	0.0830	$8.9084e^{-06}$	$1.3435e^{-05}$	$2.2344e^{-05}$	(0.0749- 0.0911)
q_{43}	0.345	0.3741	$8.4681e^{-04}$	$1.8927e^{-06}$	$8.4870e^{-04}$	(0.3714-0.3768)
σ^2	2	2.0774	0.0060	$2.9301e^{-05}$	0.0060	(2.0674-2.0881)

eters, we generate observations y_{ij} and their times t_{ij} for 50 patients with a number of observations chosen randomly between 10 and 12 for each patient. We consider $a = 4$ Markov states with the following transition rate matrix

$$Q = \begin{pmatrix} -1 & 0.4 & 0.25 & 0.35 \\ 0.3 & -1.4 & 0.7 & 0.4 \\ 0.1 & 0.3 & -0.8 & 0.4 \\ 0.1 & 0.18 & 0.2 & -0.48 \end{pmatrix},$$

The autoregressive vectors are $\mu = (8, 16, 24, 32)$, $\beta = (0.1, 0.3, 0.5, 0.7)$, and we let $\sigma^2 = 2$. The algorithm of generation works as follows:
For each patient i, we generate the hidden state x_{ij} and the observation time t_{ij} following Banks *et al.* [9], When a Markov state is generated, we use model (4.1) to simulate the observation y_{ij}. Consequently, the algorithm of generation proceeds by:

Algorithm 3.6

i) For each patient $i = 1, .., 50$, choose the number of visits n_i between 10 and 12.

ii) Initialize first time visit $t_{i1} = 0$.

iii) Initialize first state x_{i1} from initial distribution $r = (0.3, 0.4, 0.3)$.

iv) Initialize first marker observation y_{i1} from the distribution $\mathcal{N}(\mu^{(x_{i1})}, \sigma^{2(x_{i1})})$.

v) for $j = 2, .., n_i$:

(a) Compute the next time visit t_{ij} from an exponential random distribution with mean $\frac{1}{q_{ii}}$ with $i = x_{ij-1}$.

(b) Generate the state x_{ij} using the transition intensities Q_{ij} from the line x_{ij-1} of the transition intensity matrix Q.

(c) Generate the observation y_{ij} from the relation $y_{ij} = \mu^{(x_{ij})} + \beta^{(x_{i1})} y_{ij-1} + \mathcal{N}(0, \sigma^{2(x_{i1})})$.

We choose concisely the hyper-parameters for the priors for each parameter and we ran the MCMC algorithm for 8000 iterations. Before giving our results, we should mention that in MSAR like models there is an identifiability problem known as label switching. Namely, a permutation of the state labels may result in the same value for the likelihood. In our case, we have not encountered this issue because we have selected well separated hyper-parameters in the prior specification.

Our MCMC has computed consistent posterior values comparable to the true ones for all the parameters (see table 3.2). Furthermore, the algorithm checking for the convergence has shown a good mixing of chain through MCMC iterations for each parameter (figure 3.9). Also, all the parameters have good posterior density plots that are asymptotically normal (figure 3.10). Furthermore, the autocorrelation plots show that they decay just after a few lags (figure 3.11); this proves that the MCMC algorithm has overcome the problem of correlations between successive MCMC iterations for all the parameters. All these figures provide us with a good sign of convergence for the algorithm.

After we have showed that our MCMC algorithm works well for one scenario, we have run a bootstrap simulation for 1000 samples, each sample consists of 5000 MCMC iterations. This time we used a transition rate matrix presented in the table 2.3 of the work of Ma [100]:

$$Q = \begin{pmatrix} -0.911 & 0.586 & 0.254 & 0.071 \\ 0.326 & -1.351 & 1.023 & 0.002 \\ 0.017 & 0.349 & -0.708 & 0.342 \\ 0.031 & 0.086 & 0.345 & -0.462 \end{pmatrix},$$

The others parameters remained as before. Table 3.3 shows the bootstrap inference such as the mean value, the squared bias, the mean squared error (MSE), and the 95% credible interval for all the parameters of the model. Also figure 3.12 provides

the bootstrap kernel density for each parameters of the model. we can see that the bootstrap inference as well as the density shapes are of reasonable approximation to the true values.

3.7 Conclusion

A Markov chain Monte Carlo method is adopted to estimate the parameters of an MSAR model for both the discrete and the continuous time cases. Data augmentation of the unknown parameters with the hidden states has played an important role in the estimation process. Since the Markov states are unobserved, they have been also computed in each iteration of the MCMC algorithm by the famous block update FFBS scheme; where we have extended the method of Chib [31] to a set of individual observations with different lengths and even in the continuous time case.

The parameters are simulated using Gibbs sampler when the full conditional posterior of the parameter is a known one, otherwise MHA is called for.

Developing this sound computational methods for the MSAR would allow to learn about the hidden states at any time and to provide us with a mean to predict future observations using the autoregressive structure (example, Krozlig [88]). Furthermore, we would like to point out that our model can easily be extended to include missing observations, as we should only add an extra step in each MCMC iteration to estimate the missing observations.

From another side, we can estimate the autoregressive model for different values of the autoregressive order; $p \geq 1$, by evaluating the Bayesian information criterion to select the best order that fits the observations of the model. Also, with practical marker observations one could learn about the patient stage of the disease at any time and its development. Furthermore, with observations about a disease progression one can assess how a given medicine would influence the transition of the disease from a state $i \in \{1, ..., a\}$ to a state $j \in \{1, ..., a\}$ by using the transition rate q_{ij} or the transition probabilities of Π . Hence, we may compare between two treatment groups to see how the transitions evolve from one group to another depending on the treatment. In addition, MSAR models in continuous time allow to calculate the mean sojourn time in a state $i \in \{1, ..., a\}$ by using the parameter q_{ii}.

Finally, the methods used here can be extended to the more dynamics switching diffusion model for disease progression by exploring the MCMC methods for diffusion process especially with its data imputation mechanism to overcome low frequency observed data or data with large time intervals. Also, the switching diffusion processes add more stochastic dynamics to the study of any phenomena. That is what we will see in chapter 4.

Chapter 4

Bayesian Inference for Continuous Time Hybrid Switching Diffusion Process

Summary

This chapter is organized as follows: After the introduction, in section 2, we give the mathematical description of the hybrid switching diffusion process. Section 3 provides the Bayesian inference for the latent imputed data, the hidden states, and the parameters including the parameters of transition rate matrix. In section 4, we see more details on how priors are chosen and how posteriors densities are computed with an illustration by an important non-linear HSD process with an application to pneumonic disease evolution. Finally we give a discussion and a conclusion. The contents of this chapter are the results of an article by **Hibbah *et al.*** [70].

4.1 Introduction

Understanding phenomena development such as disease progression is very challenging because it incorporates behavioural switching in continuous time;

which requires us to apply a methodology that is suitable to data at irregular times. Indeed, one should appeal to a rich class of continuous time models. These models are usually taken to be diffusion processes such as the Ornstein-Uhlenbeck models (example, Dunn and Gipson [44]) . Even though the usefulness of diffusion processes and the many advantages they can offer, they can not handle all the dynamics caused by some complicated phenomena. In many cases, phenomena change behaviour or evolution depending on intermediate indicators or variables. For example in economics, the states can be the economic situations (growth, stability, crisis) and the developed variable can be options pricing or any economic factors that are changing attitude due to different economic situations. Also, in ecology (see for example, Blackwell *et al.* [16]) , animal movement has been shown to change behaviour depending of the state of whether the animal is resting at a nest, foraging or heading for a nest. Our modelling can be well present in disease progression such as HIV progression (example, Guihenneuc-Jouyaux *et al.* [66]) , where an observed bio-marker called CD4 counts decline is varying from one HIV disease stage to another.

Thus, we recourse for diffusion processes but with the extension of regime switching. By adding the regime switching characteristics, the regime switching model (RSM) will be combining two components: a continuous diffusion observation component and an unobserved component most of the time discrete and supposed to be Markovian. Consequently, RSM are usually supposed to be Markovian and hence called Markov switching models (MSM). Their underlying idea is the switching mechanism who is supposed to be latent (unobserved or unknown). The simplest model supposes a two-dimensional Markov processes, where the first is continuous and real valued and depending on the second, and the second has discrete values and acts monotonously from the first process. These models have attracted a lot of attention in diverse fields or sciences such as in population dynamics (example, Li *et al.* [93]), in pattern recognition (example, Fink [51]), or in ecology (see for example, Blackwell [14, 15]) .

However in our case, we will go a little further and direct our attention toward a more advanced switching diffusion process that supposes the dependence of the rate of the switching or of the transition on the observations or covariates. A fundamental theoretical description and practical significance related to this modelling is provided in [13]. A recent example with the Bayesian estimation methodology is presented in [16], where animal movement is an adaptive movement since it depends and follows the habitat (resting or foraging). These kinds of models are called HSD models and highlight the coexistence of continuous dynamics and discrete events: one component describes the continuous dynamics, whereas the other is a switching process representing discrete events with the switching part depending on the continuous dynamics. The exception to RSM is that the rates of transition for the discrete states in HSD process depend on the continuous dynamics (see for example, Ghosh [60], Krystul and Blom [89]).

In this chapter, we will adopt the Markov chain Monte Carlo (MCMC) methods described in chapter 2 to estimate the parameters of a HSD model. While we

could find classical methods to compute the parameters of a RSM as in [65, 76] where a ML or EM procedures are adopted, likelihood inference is very challenging for HSD. In fact, the diffusion component is non-linear most of the time and most of the diffusion phenomena of practical interest are non-linear in their nature (example, Ogawa [106]) . Furthermore, the transition rates of the regime switching are dependent on the diffusion observations (see for example, Berman [13], Blackwell[16]). Moreover, the observations are available only at dispersed discrete times or with measurement errors (for more details, see Golightly and Wilkinson [63]) . This leads us to impute virtual data between successive observations to find an efficient approximation to the unavailable transition density of the HSD using an Euler approximation in the case of HSD as in [108]. The method of imputation supposes augmenting the likelihood with the imputed data. In the Bayesian context, the procedure alternates between updating the imputed data and updating the parameters.

Unfortunately the update of the imputed data can suffer from poor mixing such as the single site update of Eraker [49], or has a long mixing time of MCMC algorithm such as the block update of Durham and Gallant [45] if the number of imputation data is so large. The challenge of our modelling is that we have to augment the likelihood with the hidden states beside the imputed data and the parameters. Also, unlike regular state independent Markov switching diffusion process, our model supposes the dependence of the transition rate matrix on the diffusion process.

Furthermore, instead of fixed times imputation, we opt for the algorithm of random time point update as described in [111] and applied to HSD in a Bayesian context by Blackwell [16] . These will avoid any discretization error, where the times and the imputed data proposed are accepted or rejected using a MHA. Thus, we will appeal for the Bayesian inference through MCMC methods to sample the posterior distribution for the parameters, the latent imputed data and the hidden states for the uni-dimensional regime switching process; since the discrete states of the switches are supposed unknown in this model. We will extend the Bayesian inference and data imputation of Fuchs [56, Chapter. 7] for non-linear diffusion to uni-dimensional HSD using the random time intervals for the imputation as in [16].

4.2 Model description and notations

We consider a continuous time HSD process of the following form

$$\begin{aligned} dX(t) &= \mu(X(t), S(t))dt + \sigma(X(t), S(t))dW_t. \\ X(t) &= X_0, \text{ and } S(t) = S_0, \end{aligned} \tag{4.1}$$

with $P(S(t+\delta t) = j | S(t) = i, X(s), S(s), s \leq t) = q_{ij}(X(t))\delta t + o(\delta t)$, $i \neq j$, $\mu()$ and $\sigma(.)$ are appropriate real valued functions satisfying certain regularity conditions. W is a uni-dimensional standard Brownian motion. $S(.)$ is a switching process

taking discrete values in $\{1, .., a\}$ with a an integer and with the dynamics of $S(.)$ depending on $X(t)$. Specifically, the spontaneous transition rate matrix $Q(x,t) = (q_{kl}(x,t))_{1\leq k,l\leq a}$ for the switching process has the following properties:

i) $q_{kl}(x,t) \geq 0$, for $k \neq l$, for any x and time t.

ii) $\sum_l q_{kl}(x,t) = 0$, for any x and time t.

Yin and Zhu [145] have summarized the difference between continuous state dependent switching diffusion process and Markovian switching models in term of properties of solutions of the process and numerical procedures.

We suppose we have N individuals with n_i the number of observation times for each individual i such that the trajectory for each individual i is given by $x_i = (x_{i1}, ..., x_{in_i})$. We consider t_{ij} the times of observations for individual i, and the latent switching states will be denoted s_{ij} for individual i at time t_{ij}. Time points are usually different between subjects, also n_i may differ between individuals. Θ is the set of the parameters in the model, while θ is the set of all parameters except the parameters related to the transition rate matrix. Hence, by applying the Markov property and the conditional probability and Bayes rules, the likelihood could be written as:

$$\begin{aligned}
L(X,S;\Theta) &\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij},s_{ij}|x_{ij-1},s_{ij-1},\Theta)\\
&\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij},s_{ij-1},\Theta)P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)\\
&\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij-1},\Theta)P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)\\
&\propto \text{Applying exponential matrix to the switching process gives:}\\
L(X,S;\Theta) &\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij-1},\Theta)\\
&\times \left[\exp\left(Q(x_{ij-1})\Delta t_{ij}\right)\right]_{s_{ij-1},s_{ij}}
\end{aligned}$$

$P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)$ is the element (s_{ij-1},s_{ij}) of the transition probability matrix evaluated at the diffusion location x_{ij-1}. As discussed in section 2 of chapter 3, the transition matrix is obtained from the exponential of the generator Q evaluated at x_{ij-1}, and we obtain the elements $\left[\exp\left(Q(x_{ij-1})\Delta t_{ij}\right)\right]_{s_{ij-1},s_{ij}}$, with $\Delta t_{ij} = t_{ij} - t_{ij-1}$. This exponential approximation supposes constant x_t for a short period of time Δt_{ij}.

4.3 Bayesian inference

Direct estimation of HSD processes is very difficult because the transition density of the diffusion component is most of the time unavailable in a closed form and

we need to approximate it. Thus, to overcome the issue of low frequency data and consequently the density approximation can be used, we need to introduce intermediate virtual data between successive observations. Thus, we appeal for the Bayesian data imputation (see for example, Elerian *et al.* [46], Eraker [49], Roberts and Stramer [118]) . Hence, an $m_{ij}-1$ missing observations are imputed between successive observations x_{ij} and x_{ij-1} to obtain for each individual i a vector $X_i = (x_{i1}, X_{i1}^2, ..., X_{i1}^{m_{i1}}, x_{i2}, X_{i2}^2 ..., x_{in_{i-1}}, X_{in_{i-1}}^2, ..., X_{in_{i-1}}^{m_{in_i}}, x_{in_i})$, with $X_{ij}^1 = x_{ij}$ and X_{ij}^k will have corresponding S_{ij}^k at the time point t_{ij}^k (the k^{th} imputation time after observation x_{ij}) . Since we have three categories of variables to estimate: the parameters Θ, the latent switching state S and the imputed (latent) data X, our imputed joint posterior density is formulated as follows:

$$\begin{aligned} P(\Theta, X, S|x) \quad &\propto \quad P(\Theta) \prod_{i=1}^{N} (P(X_{i1}^1, S_{i1}^1|\Theta) \\ &\times \quad \prod_{j=1}^{n_i-1} P(X_{ij+1}^1|X_{ij}^{m_{ij}}, S_{ij}^{m_{ij}}, \Theta) \left[\exp\left(Q(X_{ij}^{m_{ij}})\Delta t_{ij+1}^1\right)\right]_{S_{ij}^{m_{ij}}, S_{ij+1}^1} \\ &\times \quad \prod_{j=1}^{n_i-1} \prod_{k=1}^{m_{ij-1}} P(X_{ij}^{k+1}|X_{ij}^k, S_{ij}^k, \Theta) \left[\exp\left(Q(X_{ij}^k)\Delta t_{ij}^{k+1}\right)\right]_{S_{ij}^k, S_{ij}^{k+1}}, \end{aligned}$$

To evaluate the conditional posterior, we use a numerical Euler approximation (example, Kloeden and Platen [85]) for the small time t_{ij}^k (the k^{th} imputation time between observation x_{ij} and x_{ij+1}, with the exception that $\Delta t_{ij+1}^1 = t_{ij+1}^1 - t_{ij}^{m_{ij}}$ and t_{ij}^1 is the time at the observation x_{ij}) , and we have:

$$\pi^{Euler}(X_{ij}^{k+1}|X_{ij}^k, S, \Theta) \approx \phi(X_{ij}^{k+1}|X_{ij}^k + \mu(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1}, \sigma(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1}),$$

With S_{ij}^k the regime switching state value for individual i at time imputation t_{ij}^k, with $\Delta t_{ij}^{k+1} = t_{ij}^{k+1} - t_{ij}^k$ and $\phi(z|\nu, \Lambda)$ is a normal distribution with mean ν and variance Λ. $P(\Theta)$ is the prior distribution for the set of parameters Θ. The MCMC algorithm will generate a Markov chain targeting the augmented posterior $P(\Theta, X, S|x)$ under some mild regularity conditions. In fact, with the data augmentation procedure of Tanner and Wong [134], the inference may proceed by alternating between the simulation of the parameters conditional on the augmented data, and the simulation of the augmented data given the observed data and the current state of the model parameters. A discussion of convergence issues related to non-linear diffusion processes with latent data is provided in [63]. Such issues include the update choice for the imputed data and that the m_{ij} should not be so large in order to avoid the slow mixing of the targeted chain in the MCMC algorithm. Hence, we perform our simulation by alternating between drawing from the following conditional posteriors given all other quantities (denoted by .):

i) Draw the latent(imputed) observations from $P(X|.)$

ii) Draw the latent switching states from $P(S|.)$

iii) Draw the parameters from $P(\Theta|.)$

4.3.1 Sampling the latent data

While it is possible to update each latent (imputed) data separately, we will adopt a block update for the whole sequence of imputed data for each individual observation X_i. Since it is difficult to come with an analytical form for the posterior distribution of the imputed data, a MHA (see for example, Fuchs [56, Chapter. 7]) is utilized to draw the new imputed data X_i^{new} from an old X_i^{old}, for $i = 1, ..., n$. Though for the imputation of auxiliary data, we opt for the algorithm of random time point update as described in [111] and applied to HSD in a Bayesian context by Blackwell *et al.* [16]. This will avoid any discretization error.

Suppose that the transition rates are all bounded above. Let the times will be generated from a Poisson process with parameter $\kappa > \max_{i,x,t}\{-Q_{ii}(x,t)\}$. The algorithm of data imputation proceeds for each individual $i = 1, ..., N$ as follow:

Algorithm 4.1

---

i) For $j = 1, ..., n_{i-1}$

ii) For each interval $[t_{ij}, t_{ij+1}]$, the times are generated from a Poisson process with parameter κ, and we obtain the $t_{ij}^{new,k}$ time points for $k = 1, ..., m_{ij}$.

iii) for $k = 1$: $X_{ij}^{new,1} = X_{ij}^{old,1} = x_{ij}$ and $S_{ij}^{new,1} = S_{ij}^{old,1}$.

iv) for each $k = 2, ..., m_{ij}$,

   (a) we determine whether there is a switch in time with probability $\frac{-Q_{S_{ij}^{new,k-1},S_{ij}^{new,k-1}}(X_{ij}^{new,k-1})}{\kappa}$.

   (b) If there is a switch, we sample the new state $S_{ij}^{new,k} = C$ with probability $\frac{Q_{S_{ij}^{new,k-1},C}(X_{ij}^{new,k-1})}{-Q_{S_{ij}^{new,k-1},S_{ij}^{new,k-1}}(X_{ij}^{new,k-1})}$

   (c) We propose the new imputed $X_{ij}^{new,k}$ following the last update of time $t_{ij}^{new,k}$ and $S_{ij}^{new,k-1}$.

v) We ran a MHA to decide on the acceptance of the new time proposal $T_i^{new} = \{t_{ij}^{new,k},\ k = 1, ..., m_{ij},\ j = 1, ..., n_{i-1}\}$ as well as the proposed imputed data X_i^{new} .

So by introducing $m_{ij} - 1$ new observations between two successive observations x_{ij-1} and x_{ij}, the MHA acceptance ratio will be:

$$\zeta(X_i^{new}, T_i^{new}; X_i^{old}, T_i^{old}) = 1 \wedge \frac{P(X_i^{new}|D,S,\Theta)\psi(X_i^{old}|X_i^{new},D,S,\Theta)}{P(X_i^{old}|D,S,\Theta)\psi(X_i^{new}|X_i^{old},D,S,\Theta)}.$$

Where P denotes the posterior density, ψ the proposal density and D the observed, not the imputed data. Due to the Markov property in HSD, we have

$$\begin{aligned}
&\frac{P(X_i^{new}|D,S,\Theta)}{P(X_i^{old}|D,S,\Theta)} \\
&= \prod_{j=1}^{n_i-1} \frac{P(X_{ij+1}^1|X_{ij}^{new,m_{ij}},S,\Theta)}{P(X_{ij+1}^1|X_{ij}^{old,m_{ij}},S,\Theta)} \prod_{j=1}^{n_i-1}\prod_{k=1}^{m_{ij-1}} \frac{P(X_{ij}^{new,k+1}|X_{ij}^{new,k},S,\Theta)}{P(X_{ij}^{old,k+1}|X_{ij}^{old,k},S,\Theta)} \\
&= \prod_{j=1}^{n_i-1} \frac{P(\Delta t_{ij+1}^{new,1},X_{ij}^{new,m_{ij}},X_{ij+1}^1)}{P(\Delta t_{ij+1}^{old,1},X_{ij}^{old,m_{ij}},X_{ij+1}^1)} \prod_{j=1}^{n_i-1}\prod_{k=1}^{m_{ij-1}} \frac{P(\Delta t_{ij}^{new,k+1},X_{ij}^{new,k},X_{ij}^{new,k+1})}{P(\Delta t_{ij}^{old,k+1},X_{ij}^{old,k},X_{ij}^{old,k+1})}
\end{aligned}$$

$\Delta t_{ij}^{k+1} = t_{ij}^{k+1} - t_{ij}^k$ is very small to permit an Euler approximation for P by

$$P^{Euler}(X_{ij}^{k+1}|X_{ij}^k,S,\Theta) \approx \phi(X_{ij}^{k+1}|X_{ij}^k + \mu(X_{ij}^k,S_{ij}^k,\Theta)\Delta t_{ij}^{k+1}, \sigma(X_{ij}^k,S_{ij}^k,\Theta)\Delta t_{ij}^{k+1}),$$

with S_{ij}^k the regime switching state value for individual i at time imputation t_{ij}^k , and $\phi(z|\nu,\Lambda)$ is a normal distribution with mean ν and variance Λ.
While to the choice of the proposal density ψ, one could choose an Euler proposal or a double-sided Euler proposal, but due to the dependency between X^{new} and X^{old}, we will adopt the modified bridge proposal ψ_{MB} of Durham and Gallant [45]; supposed to overcome the issue of dependence between successive draws, and as adopted in the Bayesian context by Chib and Shephard [33]. Note that $X_{ij}^{new,1} = X_{ij}^{old,1} = x_{ij}$ and $S_{ij}^{new,1} = S_{ij}^{old,1}$, and the proposed path will be accepted with probability:

$$\begin{aligned}
\zeta(X_i^{new},T_i^{new};X_i^{old},T_i^{old}) = 1 \wedge &\left(\prod_{j=1}^{n_i-1} \frac{P(X_{ij+1}^1|X_{ij}^{new,m_{ij}},S,\Theta)}{P(X_{ij+1}^1|X_{ij}^{old,m_{ij}},S,\Theta)}\right) \times \\
&\left(\prod_{j=1}^{n_i-1}\prod_{k=1}^{m_{ij-1}} \frac{P(X_{ij}^{new,k+1}|X_{ij}^{new,k},S,\Theta)}{P(X_{ij}^{old,k+1}|X_{ij}^{old,k},S,\Theta)}\right) \times \frac{\psi_{MB}(X_i^{old}|X_{i1},X_{in_i},S,\Theta)}{\psi_{MB}(X_i^{new}|X_{i1},X_{in_i},S,\Theta)},
\end{aligned} \tag{4.2}$$

where

$$\frac{\psi_{MB}(X_i^{old}|X_{i1},X_{in_i},S,\Theta)}{\psi_{MB}(X_i^{new}|X_{i1},X_{in_i},S,\Theta)} = \prod_{j=1}^{n_i-1}\prod_{k=1}^{m_{ij-1}} \frac{\psi_{MB}(X_{ij}^{old,k+1}|X_{ij}^1,X_{ij+1}^1,S,\Theta)}{\psi_{MB}(X_{ij}^{new,k+1}|X_{ij}^1,X_{ij+1}^1,S,\Theta)}$$

with

$$\begin{aligned}
\psi_{MB}(X_{ij}^{new,k+1}|X_{ij}^1,X_{ij+1}^1,S,\Theta) = \\
&\phi(X_{ij}^{new,k+1}|X_{ij}^{new,k} + \frac{X_{ij}^{m_{ij}} - X_{ij}^{new,k}}{t_{ij+1}^{new,1} - t_{ij}^{new,k}}\Delta t_{ij}^{new,k+1}, \ldots \\
&\frac{t_{ij+1}^{new,1} - t_{ij}^{new,k+1}}{t_{ij+1}^{new,1} - t_{ij}^{new,k}}\sigma(X_{ij}^{new,k},S_{ij}^{new,k},\Theta)\Delta t_{ij}^{new,k+1}).
\end{aligned}$$

4.3.2 Sampling the switching hidden states

We will adapt the block update of the hidden states of Chib [31] as extended to the case of the CTMSAR(1) model by **Hibbah and El-Maroufy [71]** and detailed

in chapter 3. The block updated is applied to the vector of the hidden states for each individual i, $i = 1, 2, ..., N$. We stack the new imputed observations with the original observations in one vector of size N_i (The staked data contains both the observation data x_i and the imputed terms X_{ij}^k to give the new stacked vector $\tilde{X}_i$) $\tilde{X}_{i.} = (\tilde{X}_{i1}, ..., \tilde{X}_{iN_i})$. This new stacked vector of observations for the individual i should have corresponding switching hidden vector $\tilde{S}_{i.} = (\tilde{S}_{i1}, ..., \tilde{S}_{iN_i})$, and the corresponding time vector $\tilde{t}_{i.} = (\tilde{t}_{i1}, ..., \tilde{t}_{iN_i})$. The times contains both the original as well as the new times staked in a new vector and we use the FFBS algorithm to sample the whole sequence of the hidden states. A similar procedure is described in [111]. Let denote $\tilde{X}_i^{-j} = (\tilde{X}_{i1}, ..., \tilde{X}_{ij})$, $\tilde{X}_i^j = (\tilde{X}_{ij}, ..., \tilde{X}_{iN_i})$, $\tilde{S}_i^{-j} = (\tilde{S}_{i1}, ..., \tilde{S}_{ij})$, and $\tilde{S}_i^j = (\tilde{S}_{ij}, ..., \tilde{S}_{iN_i})$. Now, we write the joint conditional distribution for the hidden states as:

$$P(\tilde{S}_{i.}|\tilde{X}_{i.}, \Theta) = \prod_{j=1}^{N_i} P(\tilde{S}_{ij}|\tilde{X}_{i.}, \tilde{S}_i^{j+1}).$$

Hence the states computation is based on the term $P(\tilde{S}_{ij}|\tilde{X}_{i.}, \tilde{S}_i^{j+1})$ which will be evaluated in the backward pass after running the forward filtering, and our algorithm proceeds by:

Algorithm 4.2

---

i) Initialize for the time $j = 1$.

ii) For $j = 2, ..., N_i$, and $k = 1, ..., a$, compute and alternate between :

(a) $P(\tilde{S}_{ij} = k|\tilde{X}_i^{-(j-1)}, \Theta) \propto \sum_{l=1}^{a} P(\tilde{S}_{ij} = k|\tilde{S}_{ij-1} = l, \tilde{X}_{ij-1}) \times$
$P(\tilde{S}_{ij-1}|\tilde{X}_i^{-(j-1)}, \Theta)$

(b) $P(\tilde{S}_{ij} = k|\tilde{X}_i^{-j}, \Theta) \propto P(\tilde{S}_{ij} = k|\tilde{X}_i^{-(j-1)}, \Theta) \times f(\tilde{X}_{ij}|\tilde{X}_{ij-1}), \tilde{S}_{ij-1}, \Theta)$.

Later on, the backward smoothing proceeds by:

i) Initialize for the time $j = N_i$ from the last forward quantity $P(\tilde{S}_{iN_i} = k|\tilde{X}_{i.}, \Theta)$.

ii) For $j = N_i - 1, ..., 2, 1$, and $k = 1, ..., a$, compute and alternate between :

(a) $P(\tilde{S}_{ij+1}|\tilde{S}_{ij}, \tilde{X}_{ij})$ is the element $(\tilde{S}_{ij}, \tilde{S}_{ij+1})$ of the transition probability matrix evaluated at the diffusion location $\tilde{X}_{ij}$. Since the transition matrix is obtained from the exponential of the generator Q evaluated at $\tilde{X}_{ij}$, and we obtain the element $\left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij}, \tilde{S}_{ij+1}}$ for small time $\Delta\tilde{t}_{ij+1}$, assuming constant $\tilde{X}_{ij}$ for a short period of time $\Delta\tilde{t}_{ij+1}$.

(b) $P(\tilde{S}_{ij}|\tilde{X}_{i\cdot}, \tilde{S}_i^{j+1}) \propto P(\tilde{S}_{ij} = k|\tilde{X}_i^{-j}, \Theta) \times P(\tilde{S}_{ij+1}|\tilde{S}_{ij}, \tilde{X}_{ij})$.

(c) Use these last probabilities to draw the hidden states.

Again, let us remind that in the HSD model, the transition rates are related to the observed diffusion component.

4.3.3 Sampling from the posterior of the parameters

Given the estimation of $\tilde{X}$ and $\tilde{S}$ in the previous subsection, the conditional posterior of θ is proportional to the prior $P(\Theta)$ multiplied by the likelihood, which gives:

$$\begin{aligned} P(\theta|\tilde{X}, \tilde{S}) &\propto P(\theta)P(X_{i1}^1, S_{i1}^1|\Theta) \\ &\times \prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)\left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}} \end{aligned}$$

Usually, we do not come with a standard distribution to sample from, by using Gibbs sampler; thus we will appeal to a MHA. After using proportionality and dropping the terms of the transition rate matrix since they are independent of θ. Hence, we come with a MHA (see for example, Fuchs [56, Ch. 7]) with an acceptance probability ratio for every new proposal θ^* as:

$$\zeta(\theta, \theta^*) = 1 \wedge \frac{P(\theta^*)}{P(\theta)} \times \left(\frac{\prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta^*)}{\prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)} \right) \times \frac{\psi(\theta|\theta^*, \tilde{X}, \tilde{S})}{\psi(\theta^*|\theta, \tilde{X}, \tilde{S})}$$

As before, using Euler approximation we have:

$$\begin{aligned} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta) &= P^{Euler}(\tilde{X}_{ij+1}|\tilde{X}_{ij}, S, \theta) \\ &\approx \phi(\tilde{X}_{ij+1}|\tilde{X}_{ij+1} + \mu(\tilde{X}_{ij+1}, \tilde{S}_{ij}, \theta)\Delta\tilde{t}_{ij}, \sigma(\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)\Delta\tilde{t}_{ij+1}), \end{aligned}$$

ψ is a proposal density to draw a new Θ^*, it depends on the form of the likelihood and the hypothesis of the HSD. Sometimes, we could use simply a random walk proposal which is independent of $\tilde{X}$ and $\tilde{S}$ and it is only related to the old draw $\Theta^{(old)}$, and we simply propose from a gaussian distribution for some $\theta_j^{new} \sim \mathscr{N}(\theta_j^{old}, \varepsilon)$. ε is the random walk step that can be adjusted to improve convergence. For positive values, we could propose from the log-normal distribution: $log\,\theta_j^{new} \sim \mathscr{LN}(log\,\theta_j^{old}, \varepsilon)$.

Finally, and since we have a HSD, the parameters will be dependent on the switching process $\tilde{S}$, so for every given parameter θ_j, we will have to estimate $(\theta_j^1, ..., \theta_j^k, ..., \theta_j^a)$, and the posterior for θ_j^k given the hidden state k is

$$P(\theta_j^k|\tilde{X}, \tilde{S}) \propto P(\theta) \times \prod_{i=1}^{N}\prod_{j=1,\tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta). \tag{4.3}$$

4.4 Numerical example with application to disease progression

We will consider here an example of a non-linear HSD process where we give more details on how we can estimate each parameter of the model. In fact, a non-linear drift and a non-linear volatility would allow the HSD process to dispose of elements to represent clearly any complex problem and to be more flexible. Considering the non-linear process of Ait-Sahalia [3] with an extra-addition of the regime switching leads us to this new HSD model, where for each hidden state $k \in \{1, ..., a\}$, we have:

$$d\tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2 + \frac{\alpha_{3,k}}{\tilde{X}_{ij}}\right] d\tilde{t}_{ij+1} + \sigma\tilde{X}_{ij}^{\eta_k} dW,$$

For each one of the hidden switching state $k \in \{1, ..., a\}$, and as argued by Ait-Sahalia [3] for his general form, it can be deduced that a negative $\alpha_{2,k}$ guarantees ergodicity and second order stationarity for a volatility function $\sigma_k(\tilde{X}_{ij}) = \sigma\tilde{X}_{ij}^{\eta_k}$. The first order Euler approximation of our HSD model for a state $\tilde{S}_{ij} = k$ and $\varepsilon \sim \mathcal{N}(0, 1)$, is:

$$\tilde{X}_{ij+1} - \tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2 + \frac{\alpha_{3,k}}{\tilde{X}_{ij}}\right] \Delta\tilde{t}_{ij+1} + \sigma\tilde{X}_{ij}^{\eta_k}\sqrt{\Delta\tilde{t}_{ij+1}}\varepsilon. \quad (4.4)$$

4.4.1 Computation of the parameters of the drift and the volatility terms

To simulate our parameters α_k and σ^2 using the MCMC algorithm, first let us mention that when the posterior density is not known we have to use a MHA as described in parameters update 4.3, otherwise if we come up with a known posterior density to draw from it directly, we use Gibbs sampler as it is the case here. For this reason, let us pose

$$\begin{aligned}
Y_{ij} &= \frac{\tilde{X}_{ij+1} - \tilde{X}_{ij}}{\tilde{X}_{ij}^{\eta_k}}, \\
\beta_k &= (\alpha_{0,k},\, \alpha_{1,k},\, \alpha_{2,k},\, \alpha_{0,k}) \text{ and} \\
y_{ij} &= \left(\frac{\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{\tilde{X}_{ij}\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{\tilde{X}_{ij}^2\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{(1|\tilde{X}_{ij})\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}\right).
\end{aligned}$$

In matrix form, the Euler discretization for our model may be represented as in [131]:
$Y = y\beta_k + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, which is the formulation for a regression model. Consequently, the parameters of the HSD can be easily computed using the Bayesian approach for regime switching regression model. Hence for each $\beta_k, k = 1, ..., a$, the posterior is proportional to the prior multiplied by the likelihood

$$P(\beta_k|\tilde{X}, \tilde{S}) \quad \propto \quad P(\beta) \times \prod_{i=1}^{N} \prod_{j=1, \tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \Theta).$$

As in [55, page. 251]; by supposing a conjugate normal prior for $\beta_k \sim \mathcal{N}_4(b_{0k}, \sigma^2 B_{0K})$, we have

$$\begin{aligned}&\beta_k|. \sim \mathcal{N}_4(b_k, B_K), \text{ where}\\ &B_k = (B_{0K}^{-1} + y'y)^{-1} \text{ and } b_K = B_k(B_{0K}^{-1} b_{0k} + y_k^{'} Y_k),\end{aligned}$$

Y_k and y_k corresponds only to the observations where $\tilde{S}_{ij} = k$.
Similarly, under an inverted conjugate prior for $\sigma^2 \sim \mathscr{IG}(c_0, C_0)$, the posterior density for σ^2 given the observations, the imputed data, the other parameters and the switching states is

$$\begin{aligned}&\sigma^2|. \sim \mathscr{IG}(c, C), \text{ where}\\ &c = (c_0 + \frac{M}{2}) \text{ and } C = (C_0 + \frac{1}{2}\sum_{k=1}^{a} b_{0k}^{'} B_{0K}^{-1} b_{0k} + \frac{1}{2}\sum_{k=1}^{a}(y_k^{'} y_k - b_k^{'} B_K^{-1} b_k)),\end{aligned}$$

M is the number for all individuals observations.

Ridge regression:

One of the problem that can be faced here is the effects of multi-collinearity on the Bayesian regression estimation. Consequently, we recourse to the Bayesian ridge regression approach similar to Tsionas and Tassiopoulos [139] who suppose the same priors as earlier for β_k and σ^2 and we get again similar posteriors as before, the only difference is that in the ridge regression we have $B_{0K} = Diagonal(R_1, R_2, R_3, R_4)$. Moreover, to overcome the issue of choosing fixed value for R_j; we suppose the prior $P(R_1, R_2, R_3, R_4) \sim \prod_{j=1}^{4} R_j^{\frac{A}{2}-1} \exp(-\frac{b}{2} R_j)$. Hence, we alternate between updating the following posteriors, for each $k = 1, ..., a$:

$$\begin{aligned}R_{jj} &\sim Gamma(\frac{A + M_k}{2}, \frac{b + \beta_j^2}{2}),\\ B_{0K} &= Diag(R_1, R_2, R_3, R_4).\\ \beta_k &\sim \mathcal{N}_4(b_k, B_K). \qquad (4.5)\\ \sigma^2 &\sim \mathscr{IG}(c, C). \qquad (4.6)\end{aligned}$$

M_k is the number of observations for $\tilde{S}_{ij} = k$.

Computation of η_k:

To compute η_k in the volatility term, we use again the fact that the posterior is proportional to the prior multiplied by the likelihood. Hence, by supposing a Gamma prior $\mathscr{G}(0.01, 0.01)$, we come with a non standard posterior, and we call for the random walk MHA with the acceptance probability :

$$\zeta(\eta_k^{new}, \eta_k^{old}) = 1 \wedge \left(\frac{P(\eta_k^{new}) \prod_{i=1}^{N} \prod_{j=1, \tilde{S}_{ij}=k}^{N_i - 1} P(\tilde{X}_{ij+1} | \tilde{X}_{ij}, \tilde{S}_{ij}, \eta_k^{new}, \Theta_{-\eta_k})}{P(\eta_k^{old}) \prod_{i=1}^{N} \prod_{j=1, \tilde{S}_{ij}=k}^{N_i - 1} P(\tilde{X}_{ij+1} | \tilde{X}_{ij}, \tilde{S}_{ij}, \eta_k^{new}, \Theta_{-\eta_k})} \right) \times \frac{\eta_k^{new}}{\eta_k^{old}}.$$

$\Theta_{-\eta_k}$ represent all the parameters except η_k. Each η_k must be positive. As before we propose from a log-normal distribution $\mathscr{LN}(\eta_k^{old}, \varepsilon_\eta)$ with step ε_η chosen concisely.

4.4.2 Sampling the posterior dependent parameters on the transition rate matrix:

In this example, we will suppose that the intensities of the transition rate matrix are dependent on the observed diffusion process through the relation of Gompertz model as in [141]. The consideration of the Gompertz model comes from the fact that this function has been used longer in insurance, in biology such as tumor evolution or bacteria growth and in many other fields (example, Tjørve and Tjørve [137]). Also the Gompertz model has interpretable parameters, and we have:

$$q_{kl}(\tilde{X}_{ij}) = \lambda_{kl} \exp(-\gamma_{kl}\tilde{X}_{ij}),\ \lambda_{kl} > 0,\ \gamma_{kl} > 0,$$

for $k \neq l \in \{1, ..., a\}$, $i = 1, ..., N$; $and\, j = 1, ..., N_i$. We will suppose prior independence between the parameters of Θ. While the $\lambda' s$ and the $\gamma' s$ can be updated in block as in [100], here we adopt an approach similar to the approach of Siekmann *et al.* [127], where each element is updated conditional on the other since they are correlated. This single update is comparable to the single update of the transition intensities adopted in chapter 3. Hence by supposing a Gamma prior $\mathscr{G}(0.01,\, 0.01)$ for the λ_{kl}, the posterior of λ_{kl} will be:

$$P(\lambda_{kl}|\tilde{X}, \tilde{S}) \quad \propto \quad P(\lambda_{kl}) \prod_{i=1}^{N} \prod_{j=1}^{N_i - 1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}.$$

As we can see, all the other parameters especially θ are omitted due to the Bayes rule. Consequently, we obtain a non standard posterior and we have to adopt a random walk MHA to draw a new λ_{kl}^{new} (the new MCMC iteration) from an old value λ_{kl}^{old} (the previous MCMC iteration) with an acceptance probability :

$$\zeta(\lambda_{kl}^{new}, \lambda_{kl}^{old}) = 1 \wedge \left(\frac{P(\lambda_{kl}^{new}) \prod_{i=1}^{N} \prod_{j=1}^{N_i - 1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]^{new}_{\tilde{S}_{ij},\tilde{S}_{ij+1}}}{P(\lambda_{kl}^{old}) \prod_{i=1}^{N} \prod_{j=1}^{N_i - 1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]^{old}_{\tilde{S}_{ij},\tilde{S}_{ij+1}}} \right) \times \frac{\lambda_{kl}^{new}}{\lambda_{kl}^{old}}, \quad (4.7)$$

Where we propose the λ_{kl}^{new} from a log-normal distribution $\mathscr{LN}(\lambda_{kl}^{old},\, \varepsilon_\lambda)$ so as to keep operating on real positive values with ε_λ the random walk step.
Similarly, a no close form is obtained for the posterior of γ_{kl}. So, we consider again a random walk MHA. With a Gamma prior $\mathscr{G}(0.01,\, 0.01)$ on γ_{kl} and a log-normal proposal $\mathscr{LN}(\gamma_{kl}^{old},\, \varepsilon_\gamma)$, our MHA acceptance probability is:

$$\zeta(\gamma_{kl}^{new}, \gamma_{kl}^{old}) = 1 \wedge \left(\frac{P(\gamma_{kl}^{new}) \prod_{i=1}^{N} \prod_{j=1}^{N_i - 1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]^{new}_{\tilde{S}_{ij},\tilde{S}_{ij+1}}}{P(\gamma_{kl}^{old}) \prod_{i=1}^{N} \prod_{j=1}^{N_i - 1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]^{old}_{\tilde{S}_{ij},\tilde{S}_{ij+1}}} \right) \times \frac{\gamma_{kl}^{new}}{\gamma_{kl}^{old}}, \quad (4.8)$$

4.4.3 Numerical implementation and simulation:

To assess the accuracy of our finding, we will simulate observations for $N = 50$ individuals with follow up size between 7 and 10 for every individual. We suppose

we have $a = 3$ regime switching states. The parameters of the non-linear HSD are : $\alpha_0 = (0.3, 0.6, 0.9)$, $\alpha_1 = (0.02, 0.04, 0.06)$, $\alpha_2 = (-0.08, -0.06, -0.04)$, $\alpha_3 = (0.01, 0.05, 0.09)$, $\eta = (0.4, 0.5, 0.6)$, $\sigma^2 = 0.04$, and

$$\lambda = \begin{pmatrix} 0 & 0.2 & 0.4 \\ 2 & 0 & 3 \\ 0.3 & 0.9 & 0 \end{pmatrix}, \gamma = \begin{pmatrix} 0 & 0.2 & 0.04 \\ 0.01 & 0 & 4 \\ 1 & 0.8 & 0 \end{pmatrix}$$

The simulation algorithm works with the help of the Euler approximation for every individual i for $i = 1, ..., 50$, and we have:

Algorithm 4.3

---

i) Choose n_i uniformly in $[7, 10]$ for $j = 1$

   (a) Choose the regime switching state $S_{i1} = k$ uniformly for k in $\{1, 2, 3\}$.

   (b) Initialize $X_{i,1} \sim \mathcal{N}(\alpha_{0,k}, \sigma^2)$.

   (c) Compute Q using formulation $Q_{kl}(X_{i1}) = \lambda_{kl} \exp(\gamma_{kl} X_{i1})$, $\lambda_{kl} > 0$, for $k \neq l \in \{1, 2, 3\}$, and $Q_{kk}(X_{i1}) = -\sum_{l \neq k} Q_{kl}(X_{i1})$.

   (d) Compute t_{i2} from $\exp(-Q_{kk}(X_{i1}))$

   For $j = 2, ..., n_i$:

ii) Compute S_{ij} using the line S_{ij-1} of $Q(X_{ij-1})$

iii) Compute X_{ij} for S_{ij-1} using Euler approximation for the model 4.4.

iv) Compute Q using formulation $Q_{kl}(X_{ij}) = \lambda_{kl} \exp(\gamma_{kl} X_{ij})$, $\lambda_{kl} > 0$, for $k \neq l \in \{1, 2, 3\}$, and $Q_{kk}(X_{ij}) = -\sum_{l \neq k} Q_{kl}(X_{ij})$.

v) Compute t_{ij+1} from $\exp(-Q_{kk}(X_{ij}))$, if $j < n_i$.

To assess the efficiency of our methods, we will see how our MCMC algorithm can estimate the true values (values used to generate the simulated data). Before providing the MCMC algorithm, we should point out that we do not opt for the usual regularly spaced point imputation procedure that can drive the Bayesian estimation to break down if the amount of imputation is large (for example, Golightly and Wilkinson [63]). In fact, it has been shown dependence between the unknown parameters in the diffusion and the missing data while adopting this imputation. This can result in slow rates of convergence of naive sampling or could conduct to identifiability problem as in the single update of Eraker [49] or the block update of Durham and Gallant [45].

Thus we call for the random time data imputation that allows exact estimation as in [16]. After the generation of the simulated data and to check the accuracy

of the estimation, we ran the MCMC algorithm for a large number of iterations. During this running there is a burnings period (where the algorithm has not converged yet). The burnings period varies depending on the volatility complexity, the number of parameters in the model, or the number of imputation as well as the latent data. After the burnings period, the MCMC converges and the inference is based on the last iterations of the MCMC algorithm. Appendix H provides the Matlab program of data generation as well as the MCMC program for the parameters'simulation. Our algorithm proceeds for 8000 iterations (number of iterations found in our case to be convenient for the MCMC to converge so as we can draw inference from the MCMC output) as follow:

Algorithm 4.4

---

i) Initialize Θ

ii) for $m = 2, ..., 8000$:

(a) Propose the new times T_i^{new} using a Poisson process with parameter κ.

(b) Propose the new imputed data X_i^{new} Using Euler approximation and modified bridge proposal.

(c) Accept the new proposal of the times as well as the imputed data using MHA (4.2).

(d) Stack the new data and the new times in new vector of data and times: $\tilde{X}$ and $\tilde{T}$.

(e) Simulate the regime switching states $P(\tilde{S}|.)$ using the FFBS algorithm.

(f) Simulate the parameters of the transition rate matrix Q: λ and γ using the random walk MHA (4.7) and (4.8) respectively.

(g) Compute $\beta_k|. \sim \mathcal{N}_4(b_k, B_K)$ from (4.5) for $k = 1, 2, 3$.

(h) Compute $\sigma^2|. \sim \mathcal{IG}(c, C)$ from (4.6).

(i) Simulate η_k using MHA (4.7), for $k = 1, 2, 3$.

Table 4.1 gives the posterior statistics of our algorithm such as the posterior means, the standard deviations and the 95% credible intervals. We can see that for all the parameters, we get posterior means comparable to the true values even though with this complicated process with a large number of parameters. All the parameters have minimal standard deviations. Furthermore; except for few parameters, all the credible intervals cover perfectly the true values. Moreover, the graphical convergence checking has been performed for the model parameters such as the inspection of the trace-plot, the kernel density and the autocorrelation function plot for every parameter. It is revealed that for each parameter, we get a good mixing of the

MCMC chain that converges asymptotically to the true value (figure 4.1) , a perfect density shape that is asymptotically normal (figure 4.2) and autocorrelations that decay immediately after a few lags (figure 4.3). This simulation shows that the MCMC methods are more appropriate and efficient to estimate this complicated process; especially that we have opted for the random time data imputation. Our algorithm works well if we choose different values for this switching one factor model, or ran a bootstrap simulation. Indeed, the algorithm has shown its efficiency.

4.4.4 Application to disease progression:

Many models have been proposed to model disease progression through markers'observations, among them Young *et al.* [146] proposed a deterministic differential equation model (DEM) to assess markers in disease progression such as Alzheimer's disease, Huntington's disease, or Parkinson's disease. Taking the stochastic version of this DEM and adding the regime switching, we get a process similar to the previous regime switching one factor model with the following expression, for $k = 1, 2, 3$:

$$d\tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2\right] d\tilde{t}_{ij+1} + \sigma dW$$

We applied this process to model marker observations from a slow developed disease: COPD (chronic obstructive pulmonary disease). Doctors use stages to address the stage of the COPD (the severity of the disease). Knowing how severe the COPD in a patient helps to choose the best treatment. The idea is to unravel the stages of the disease using one marker or a combination of many markers.

In fact, there are many clinical markers for COPD disease among them: chest hyperinflation, low body mass index (BMI), the use of accessory muscles of respiration, and prolonged expiration. Since, we are interested in one dimensional diffusion process, we take into consideration as a marker the FEV1 (how much air one can exhale from his lungs in one second, measured in liter). We will see how the FEV1 observations allow to estimate the parameters of the HSD process. We have supposed that our model has three hidden states; $a = 3$ (Mild, Moderate, and severe) as classified by Stockley [133].

We extracted our FEV1 observations using data from the Danish Lung Cancer Screening Trial (DLCST) (for more details see, Shaker *et al.* [123]); where 2052 current or ex-smokers aged $50 - 70$ years had FEV1 measured annually for 5 years (2005-2009). An extract from this trial is provided in appendix F, where we have for each subject, five FEV1 measurements with the date of each measurement. While having more than 10 marker observations by patient would give more parameters precision and identification, disposing here of only 5 marker observations by patient is found to be sufficient for giving accurate results. From the database, it can be seen that the values of the FEV1 marker decreases with the severity of the COPD disease (from values that are greater than $3\,L$ in mild stages to approximately less than $1.5\,L$ for severe stages).

We ran our MCMC algorithm to fit the HSD model and make the inference after

the burnings period (inference is based on the last 3000 samples after the burnings). In fact, starting from good initial values for the parameters would help in reducing the time of the burnings period and hence accelerate the MCMC convergence. Such values can be obtained from other estimation methods such as the ML computation.

Also, while the $\alpha's$ and σ are computed here using Gibbs sampler via a ridge regression approach; in the general case when the posterior is not a known one we call for the MHA such as it is the case here for the parameter of the transition rate matrix; the $\lambda's$ and the $\gamma's$.

Furthermore, this MCMC uses the random time data imputation mechanism. Hence, we have seen that between successive observations, it could happen no imputation or 1 imputation, or 5, or more than 10 imputed time points.

Finally, table 4.2 gives the posterior computations for each one of the parameters of the model. It shows how the marker can lead to the estimation of the parameters depending on the hidden states. Moreover, and while here we fitted a model with 3 hidden states one can consider the cases where $a = 4\, or\, 5$, and uses the Bayesian information criterion (BIC) to choose the best model. Also, we could take other form of stochastic differential equations and choose which equation fits well the data using the BIC.

4.5 Discussion

Bayesian approach is very efficient in simulating complicated models such as the non-linear diffusion processes. In fact, one can incorporate any prior information or knowledge in the likelihood through the prior specification and this is possible because the posterior of the parameters is proportional to prior multiplied by the likelihood. Also, and while MCMC algorithm can converge even when starting in dispersed initial values for the parameters, we can take use of classical method inference on data to get a good starting values such as ML and EM algorithms.

Another issue that should be pointed here too is that this model uses observation intervals that are non equi-distant, though we get accurate estimate using the Euler discretization; and why not should we try to improve this accuracy in the future by using the Milstein discretization (see for example, Tse *et al.* [138]). Moreover, the number of observations imputed was chosen using the random time imputation which gives exact simulation. With this way, for small intervals we do not impute any data while for large intervals we could impute data.

Other ideas that can attract attention, is the use of the random walk MHA; that could have many problems such as the moving step. Indeed, a bad choice for the moving step can lead to bad mixing or create high correlated draws. Thus, we should sometimes avoid the random walk MHA and find good proposal density for every new draw for each parameter or adopting more efficient algorithms such as the accept-reject MHA developed in [32]. Hopefully here the ridge regression has allowed to sample many parameters through Gibbs sampler.

Finally, even that the HSD model here adopts an homoscedastic σ^2, it could be

easily extended to σ_k^2 depending on the hidden states $k = 1, .., a$ or we can take the stochastic variance as in the case of ARCH and GARCH model (see for example, Adesi *et al.* [1], Bollerslev *et al.* [17], Chou [35], Nelson [104]).

4.6 Conclusion

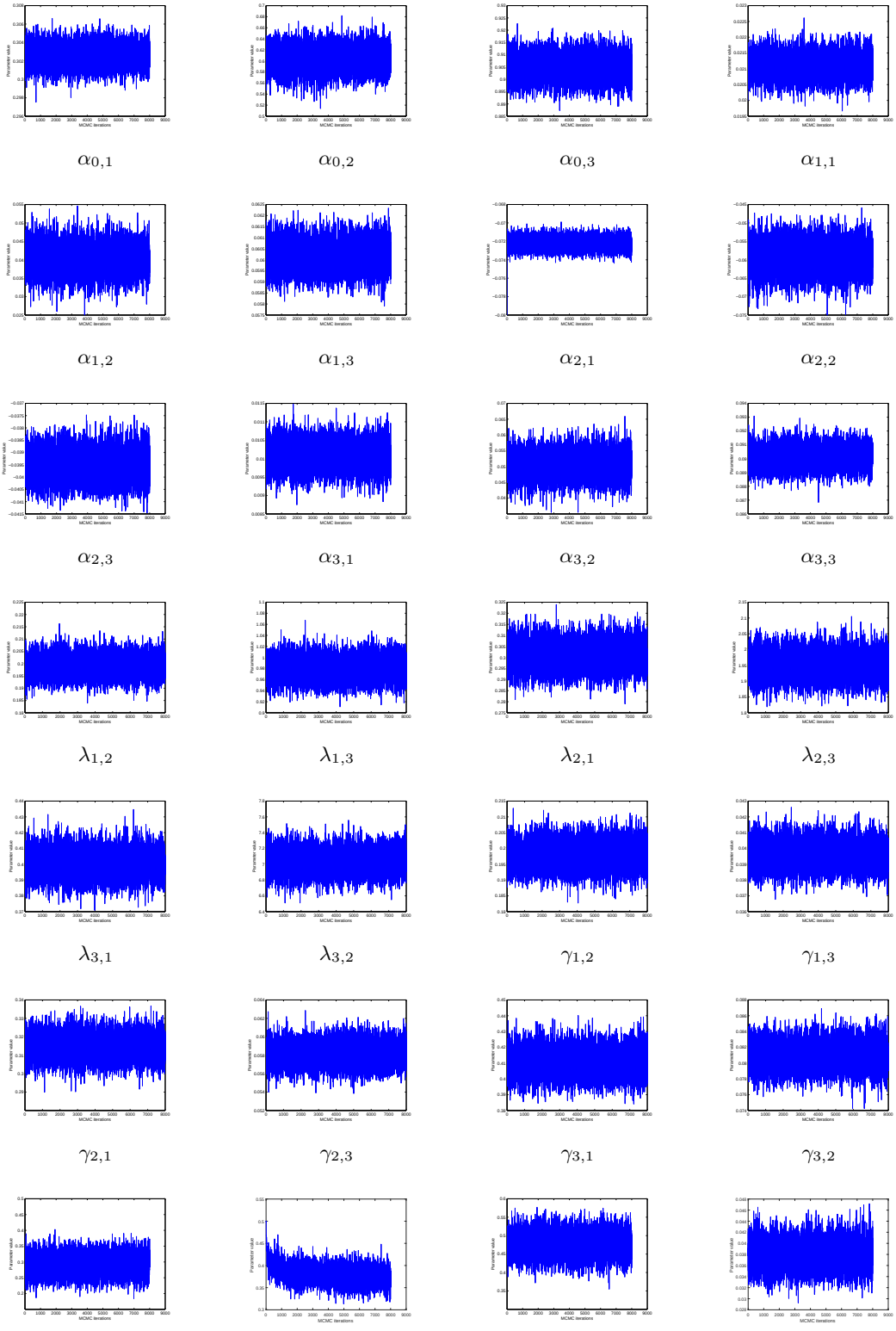
This chapter provides a Bayesian approach for the simulation of a state-dependent switching diffusion process; one of the process usually hard to handle in a classical framework. We used Euler discretization to overcome the issue of dispersed observations as it is the case for most diffusion models. We have adopted the random time data imputation instead of the fixed time data imputation. We have let the transition rate matrix to depend on the diffusion observations, unlike the CTM-SAR(1)model in chapter 3; where the intensities are supposed to be independent of the observation process. Hence, we have lead the estimation of three categories of variables: the imputed data, the hidden switching states, and the parameters of the diffusion process; the sampling of the hidden state has been realized by a FFBS algorithm adapted to the HSD. Overall, even though the complexity of the switching one factor model, the MCMC algorithm has shown its efficiency to estimate the parameters accurately.

Table 4.1: MCMC estimation for the parameters of the model(8000 iterations)

Posterior computations				
Parameters	True value	Mean	Standard deviation	Credible interval (95%)
$\alpha_{0,1}$	0.3	0.3026	0.0011	(0.3004,0.3048)
$\alpha_{0,2}$	0.6	0.6028	0.0201	(0.5636,0.6426)
$\alpha_{0,3}$	0.9	0.9048	0.0047	(0.8956,0.9140)
$\alpha_{1,1}$	0.02	0.0211	$3.5015.10^{-04}$	(0.0204,0.0218)
$\alpha_{1,2}$	0.04	0.0401	0.0036	0.0331,0.0472)
$\alpha_{1,3}$	0.06	0.0602	$6.0245.10^{-04}$	(0.0590,0.0614)
$\alpha_{2,1}$	-0.08	-0.0801	0.0022	(-0.0802,-0.0799)
$\alpha_{2,2}$	-0.06	-0.0593	0.0038	(-0.0671,-0.0522)
$\alpha_{2,3}$	-0.04	-0.0395	$5.575210.^{-04}$	(-0.0406,-0.0384)
$\alpha_{3,1}$	0.01	0.0101	$3.4369.10^{-04}$	(0.0094,0.0107)
$\alpha_{3,2}$	0.05	0.0503	0.0037	(0.0431,0.0577)
$\alpha_{3,3}$	0.09	0.0901	$7.1083.10^{-04}$	(0.0887,0.0915)
λ_{12}	0.2	0.1839	0.0040	(0.1911,0.2070)
λ_{13}	1	0.9108	0.0195	(0.9413,1.0181)
λ_{21}	0.3	0.2790	0.0059	(0.2895,0.3126)
λ_{23}	2	1.8189	0.0392	(1.8738,2.0267)
λ_{31}	0.4	0.3708	0.0080	(0.3846,0.4161)
λ_{32}	7	6.5094	0.1432	(6.7576,7.3138)
γ_{12}	0.2	0.1827	0.0040	(0.1902,0.2056)
γ_{13}	0.04	0.03681	$7.7337.10^{-04}$	(0.0381,0.0411)
γ_{21}	0.3	0.2899	0.0063	(0.3021,0.3268)
γ_{23}	0.06	0.0538	0.0012	(0.0558,0.0604)
γ_{31}	0.4	0.3841	0.0082	(0.3948,0.4264)
γ_{32}	0.08	0.0742	0.0016	(0.0778,0.0842)
η_1	0.3	0.2909	0.0201	(0.2510,0.3298)
η_2	0.4	0.3788	0.0430	(0.3098,0.4694)
η_3	0.5	0.5324	0.0199	(0.4933,0.5716)
σ^2	0.04	0.0377	0.0023	(0.0333,0.0425)

Table 4.2: MCMC output for fitting a HSD Process via FEV1 marker in COPD disease progression

	Posterior computations		
Parameters	Mean	Standard deviation	Credible interval (95%)
$\alpha_{0,1}$	0.0030	0.0084	(-0.0128,0.0172)
$\alpha_{0,2}$	0.0169	0.0231	(-0.0300,0.0618)
$\alpha_{0,3}$	0.0164	0.0127	(-0.0113,0.0402)
$\alpha_{1,1}$	-0.0539	0.0209	(-0.0952,-0.0136)
$\alpha_{1,2}$	-0.0226	0.0656	(-0.1545,0.1105)
$\alpha_{1,3}$	-0.0240	0.0395	(-0.1037,0.0497)
$\alpha_{2,1}$	-0.0264	0.0242	(-0.0730, 0.0202)
$\alpha_{2,2}$	-0.0094	0.0744	(-0.1557,0.1452)
$\alpha_{2,3}$	-0.0120	0.0450	(-0.0991,0.0761)
λ_{12}	0.2657	0.0058	(0.2538,0.2782)
λ_{13}	0.7661	0.0369	(0.7224,0.8706)
λ_{21}	0.0458	7.55×10^{-4}	(0.0435,0.0470)
λ_{23}	0.6294	0.0114	(0.6152,0.6584)
λ_{31}	0.0708	0.0011	(0.0688,0.0733)
λ_{32}	0.2223	0.0034	(0.2144,0.2263)
γ_{12}	0.0122	4.79×10^{-4}	(0.0113,0.0131)
γ_{13}	0.312	0.012	(0.2889,0.3369)
γ_{21}	2.0406	0.081	(1.8855,2.2007)
γ_{23}	0.0087	3.59×10^{-4}	(0.0081, 0.0095)
γ_{31}	4.592	0.187	(4.2384,4.9777)
γ_{32}	3.122	0.126	(2.8812,3.3773)
σ^2	0.0086	7.78×10^{-4}	(0.0073,0.0103)



$\alpha_{0,1}$ $\alpha_{0,2}$ $\alpha_{0,3}$ $\alpha_{1,1}$

$\alpha_{1,2}$ $\alpha_{1,3}$ $\alpha_{2,1}$ $\alpha_{2,2}$

$\alpha_{2,3}$ $\alpha_{3,1}$ $\alpha_{3,2}$ $\alpha_{3,3}$

$\lambda_{1,2}$ $\lambda_{1,3}$ $\lambda_{2,1}$ $\lambda_{2,3}$

$\lambda_{3,1}$ $\lambda_{3,2}$ $\gamma_{1,2}$ $\gamma_{1,3}$

$\gamma_{2,1}$ $\gamma_{2,3}$ $\gamma_{3,1}$ $\gamma_{3,2}$

η_1 η_2 η_3 σ^2

Figure 4.1: 8000 MCMC iteration plots for the parameters of the HSD model

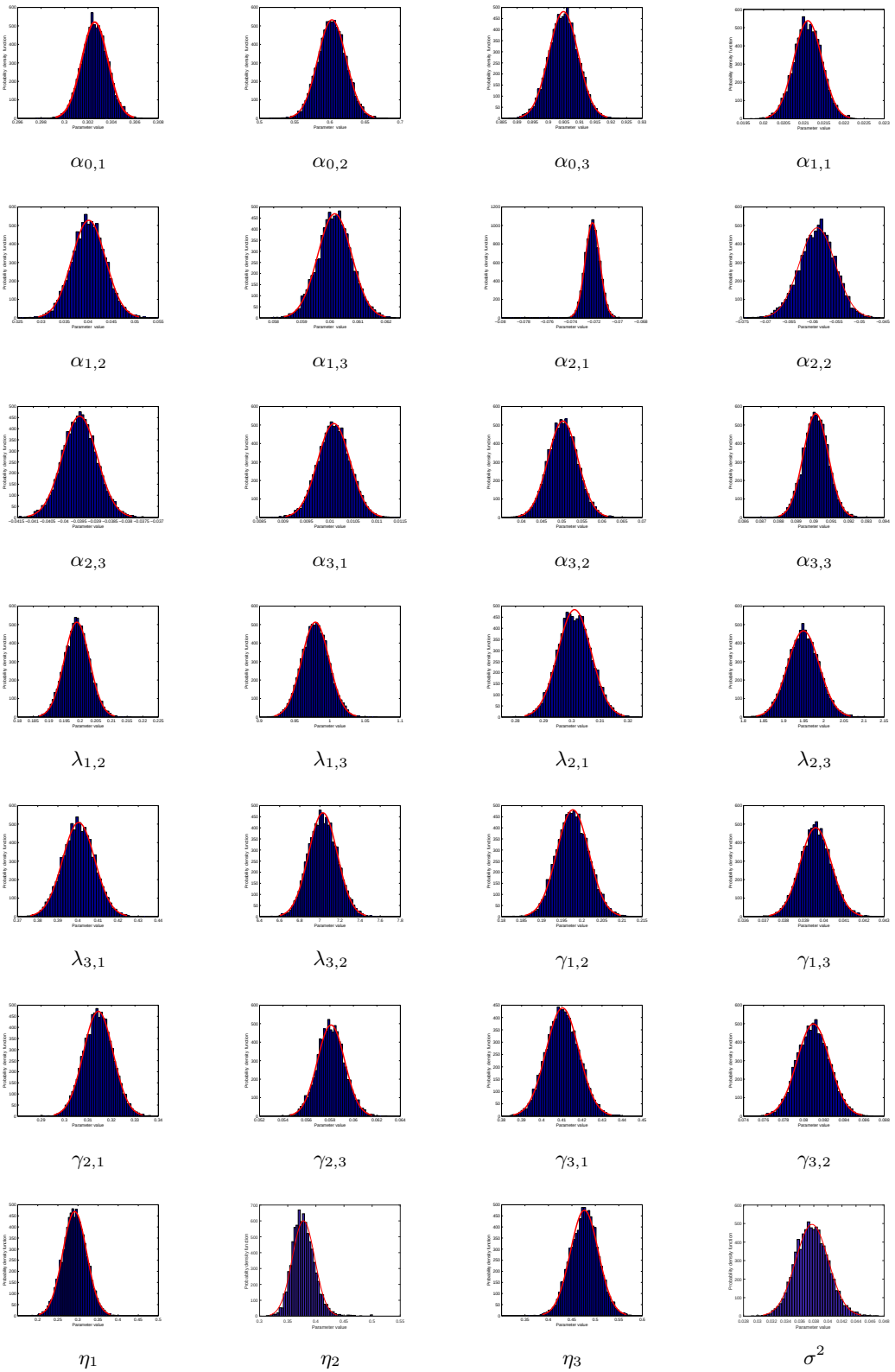


Figure 4.2: Posterior density plots for the parameters of the HSD model

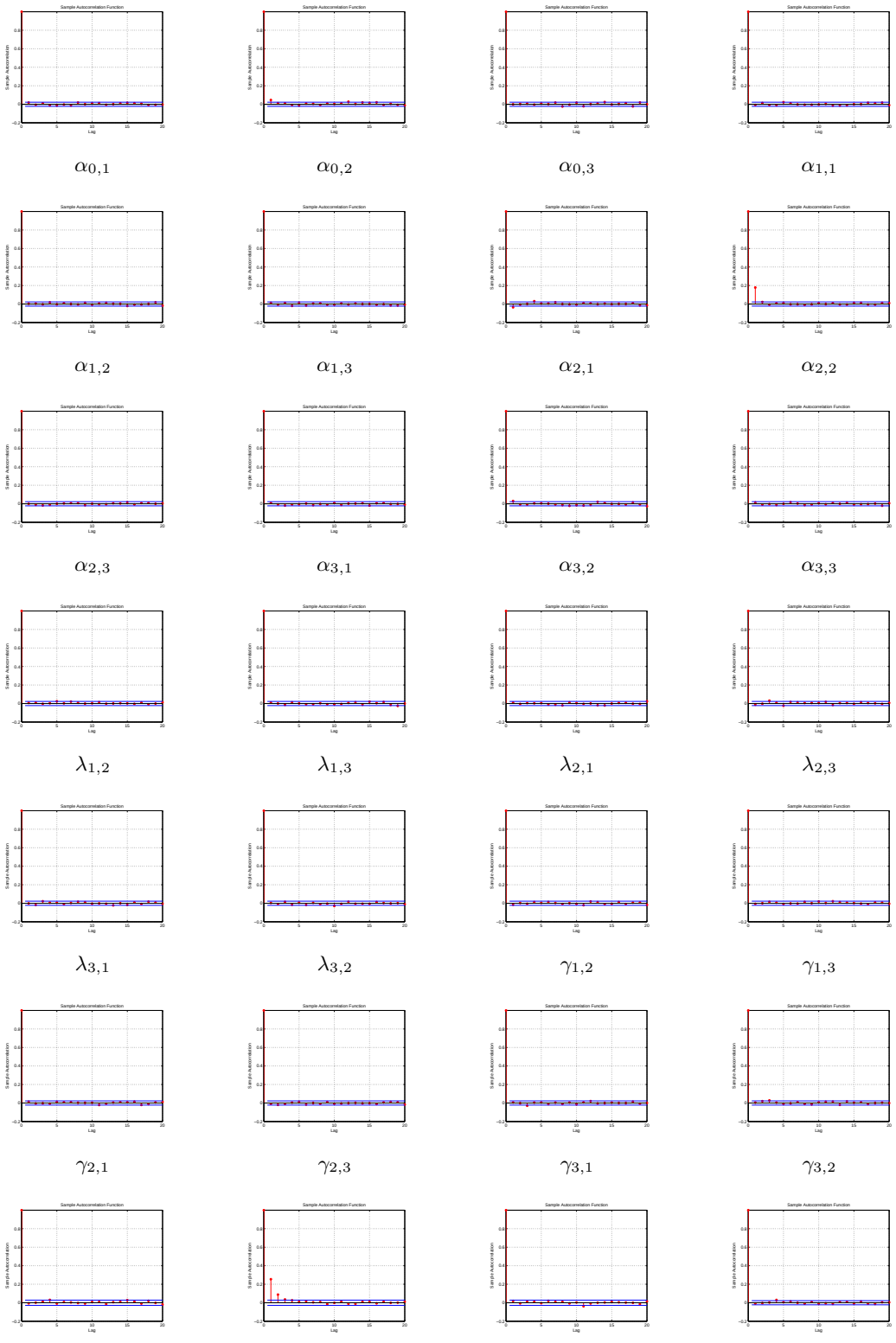


Figure 4.3: Autocorrelation sample plots for the parameters of the HSD model

Chapter 5

Conclusion and Perspectives

MCMC methods were adapted to estimate MSM in both the continuous time and the discrete time cases, where the likelihood is augmented by the hidden states. In fact, the simulation of MSM has many practical advantages especially the transition rates of the Markov process that allow the disease follow-up and monitoring. Also, the Bayesian computation of the latent stages has been realized in block; hence we could predict the patient stage at any time.

While MSM offer many advantages, they do not include some randomness as well as they can not overcome the issue of sparsely observed models; which has conducted us to take into account the hybrid switching process, that allows to impute data between successive observations via data imputation tool. Also, the HSD process provides more stochastic dynamics as well as it provides a mutual correlation between the observed component and the latent process. The data imputation tool here is different from the known fixed time data imputation mechanism, instead we opt for the random time data imputation mechanism where the times imputed are also proposed randomly with data at each MCMC iteration. We give an application of the HSD using the marker FEV1 for COPD disease progression.

For all the processes considered, we have seen how MCMC methods can take advantage of prior specification to draw efficient posterior statistics about the parameters.

Our perspective is to look for modelling latent variable models when the observations are multi-dimensional; so as to analyse the utility of the co-existence of many bio-markers in addressing the behaviour of disease development, with the consideration of correlations between the observations (see for example, Lee *et al.* [91], Bellone [11]).

Future concerns could be the consideration of joint prior for depending parameters, since it results in a better estimation of the parameters compared to independent modeling as in [40].

In addition, working on finding better proposal density when using the MHA would overcome the shortcoming of the RWMHA, especially in regime switching processes as it is our case. Indeed, MCMC methods encounter serious difficulties if the target distribution has isolated modes and by the way the MCMC mixing would be very bad. Hence, HSD process should be evaluated using other MCMC algorithms such as the simulated tempering algorithm; where the density of interest is flattened in order to facilitate movement among the modes (example, Woodard

et al. [143], Zheng [147]).

Moreover, most of the time, the parameters were estimated separately; while efficient computation was observed or could be improved with the Particle Markov chain Monte Carlo (PMCMC) methods that have been developed recently (example, Kantas *et al.* [79]), where we can use particle filters to build efficient proposal distributions in high-dimensions models as it is the case for RSM. Finally, we could have considered MSM with exogenous explanatory variables or with the transitions between states relying on the realizations of underlying time series as well as the current and possibly past states (see for example, Chang *et al.* [27], Kim and Kim [82]).

Appendix A

Basics properties of convergence of random variables

The basic definitions presented in this Appendix are inspired from the book of Koralov and Sinai [86].
Let $(\Omega, \mathscr{F}, P)$ be a probability space with elementary elements $\omega \in \Omega$.

A.1 Convergence in distribution

A sequence of random variables $X_1, X_2, X_3, ...$ converges in distribution to a random variable X, shown by $X_n \xrightarrow{d} X$, if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, for all x at which $F_X(x)$ is continuous. $F_X(.)$ is the cumulative distribution function.

A.2 Convergence in probability

A sequence of real-valued random variables $X_1, X_2, X_3, ...$ converges in probability to a real-valued random variable X, shown by $X_n \xrightarrow{p} X$,
if $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$, for all $\varepsilon > 0$.

A.3 Almost surely convergence

A sequence of random variables $X_1, X_2, X_3, ...$ converges almost surely to a random variable X, shown by $X_n \xrightarrow{a.s} X$,
if $P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$.

A.4 Law of large numbers

Let $X_1, X_2, X_3, ...$ be a sequence of random variables with finite expectations $\mathbb{E}(X_n) = \mu_n$, $n = 1, 2, ...$ Let $Y_n = \frac{(X_1+...+X_n)}{n}$ and $\overline{\mu} = (\mu_1 + ... + \mu_n)/n$. The sequence of random variables X_n satisfies the law of large numbers if $Y_n - \overline{\mu} \xrightarrow{p} 0$. It satisfies the strong law of large numbers if $Y_n - \overline{\mu} \xrightarrow{a.s} 0$.

A.5 Central limit theorem

Let $X_1, X_2, X_3, ...$ denote independent, identically distributed real-valued random variables each with mean μ and variance $\sigma^2 < \infty$ and let $\overline{X}_n = \frac{(X_1+...+X_n)}{n}$, $n =$

1, 2, ..., then $\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1)$.

Appendix B

The detailed balance condition, stationarity, reversibility, and the Metropolis Hastings algorithm

B.1 Proof of the relation between stationarity, reversibility and the detailed balance condition

Proof:
Part (i) follows by noting that, by the detailed balance condition; for any measurable set B, $\int_{\mathscr{Y}} Q(y,B)\pi(y)dy = \int_{\mathscr{Y}}\int_B Q(y,x)\pi(y)dxdy = \int_{\mathscr{Y}}\int_B Q(x,y)\pi(x)dxdy = \int_B \pi(x)dx$, since $\int_B \pi(x)dy = 1$. With the existence of the kernel Q and invariant density π, it is clear that detailed balance and reversibility are the same property.

B.2 Proof for the Metropolis Hastings algorithm and the detailed balance condition

As shown by Roberts and Rosenthal [116]; the MHA verifies the detailed balance condition.
Proof:

$$\text{suppose } \alpha(\theta^{(g)},\theta^{(*)}) < 1 \text{ then } \frac{\pi(\theta^{(g)})Q(\theta^{*}/\theta^{(g)})}{\pi(\theta^{(*)})Q(\theta^{(g)}/\theta^{(*)})} > 1,$$

$$\text{thus } \alpha(\theta^{(*)},\theta^{(g)}) = 1,$$

$$\text{If } \alpha(\theta^{(g)},\theta^{(*)}) < 1 \text{ then } \frac{\pi(\theta^{(g)})Q(\theta^{(*)}/\theta^{(g)})}{\pi(\theta^{(*)})Q(\theta^{(g)}/\theta^{(*)})} > 1 \text{ thus,}$$

$$\alpha(\theta^{(*)},\theta^{(g)}) = 1$$

$$\text{Now suppose } \alpha(\theta^{(g)},\theta^{(*)}) < 1 \text{ and } \alpha(\theta^{(*)},\theta^{(g)}) = 1 \text{ we have,}$$

$$\pi(\theta^{(g)})Q(\theta^{(*)}/\theta^{(g)})\alpha(\theta^{(g)},\theta^{(*)}) = \pi(\theta^{(*)})Q(\theta^{(g)}/\theta^{(*)}),$$

$$\pi(\theta^{(g)})Q(\theta^{(*)}/\theta^{(g)})\alpha(\theta^{(g)},\theta^{(*)}) = \pi(\theta^{(*)})Q(\theta^{(g)}/\theta^{(*)})\alpha(\theta^{(*)},\theta^{(g)}),$$

$$\pi(\theta^{(g)})T(\theta^{(g)},\theta^{(*)}) = \pi(\theta^{(*)})T(\theta^{(*)},\theta^{(g)}).$$

Appendix C

The Chapman Kolmogorov property

Let $\mathbb{S}$ be a continuous time finite Markov state space taking values in $1, 2, ..., a$, with $a \in \mathbb{N}^*$ and Π is the transition probability matrix of the Markov process.
The following properties are inspired from the book by Rausand and Hoyland [112, Chapter. 8].

C.1 The Chapman Kolmogorov equation

By using the Markov property and the law of total probability, the Chapman-Kolmogorov equation could be written as:

$$\begin{aligned} \Pi_{ij}(t+s) &= \sum_{k=0}^{a} \Pi_{ik}(t)\Pi_{kj}(s) \text{ for all } i,\, j \in \mathbb{S},\, t,\, s > 0, \text{ or in matrix form} \\ \Pi(t+s) &= \Pi(t)\,.\,\Pi(s). \end{aligned}$$

C.2 The Chapman Kolmogorov differential equation

let Δt be a small time interval, and let us pose $\dot{\Pi}(\Delta t) = \lim_{\Delta t \to 0} \dfrac{\Pi_{ij}(t+\Delta t) - \Pi_{ij}(t)}{\Delta t}$, the Kolmogorov forward equation could be written as

$$\dot{\Pi}_{ij}(t) = \sum_{k=0}^{a} q_{kj}\Pi_{ij}(t), \text{ for all } i\, j \in \mathbb{S}, \text{ and } t > 0, \text{ or}$$

$$\Pi(t) \cdot Q = \dot{\Pi}(t), \text{ in matrix form,}$$

the q_{kj}'s are the elements of the transition intensity matrix Q of the Markov process evaluated at time $t \geq 0$. Following the same principles, we can obtain the backward equation.

C.3 The Chapman Kolmogorov property and the exponential matrix

For any $t \geq 0$, it may be shown that the solution to the forward or backward equations is $\Pi(t) = \exp(Qt)$. The term $\exp(Qt)$ is called the matrix exponential.

Computationally, this may be evaluated using the Taylor series expansion of the matrix exponential

$$\exp(Qt) = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!}.$$

Appendix D

The likelihood for state independent regime switching diffusion process, and the forward backward algorithm

D.1 Forward filtering backward sampling for states estimation for state independent regime switching diffusion process

The concept of forward filtering backward sampling (FFBS) algorithm is very important in regime switching models since it allows the estimation of a block of the hidden states recursively. Hence, the main of its steps as in [78] for model 2.3; but by supposing that the latent states are independents from the diffusion component: first, the joint posterior of the hidden states for the times $t = 1, 2, ..., T$ is decomposed as

$$P(S|X, \Theta) = \prod_{t=0}^{T-1} P(S_t|S_{t+1}, X^t, \Theta),$$

where $X^t = (X_1, X_2, ..., X_t)$, and Θ is the set of the parameters of the diffusion process. To simulate $P(S|X, \Theta)$, we consider the following steps:

- Run the Kalman filter for $t = 1, 2, ..., T$ to get the moments of $P(S_t|X^t, \Theta)$.
- Sample the last state from $P(S_T|X^T, \Theta)$.
- Sample backward through time : $P(S_t|S_{t+1}, X^t, \Theta)$.

D.2 The likelihood for state independent regime switching diffusion process

Johannes and Polson [78] provide a discretization plus likelihood formulation for state independent regime switching process where for every state $i \in \mathscr{S}$ ($\mathscr{S}$ is the state space for S), model 2.3 with state independence can be written as:

$$d(X_t) = \mu(\Theta, X_t, i)dt + \sigma(\Theta, X_t, i)dW_t,$$

which gives the discretization:

$$X_t = \mu(\Theta_{S_t}, X_{t-1})dt + \sigma(\Theta_{S_t}, X_{t-1})\varepsilon_t,$$

and the joint likelihood is given by :

$$P(X|S,\Theta) = \prod_{t=1}^{T} P(X_t|X_{t-1}, S_{t-1}, \Theta) \text{ with}$$
$$P(X_t|X_{t-1}, S_{t-1}, \Theta) = \mathscr{N}(\mu(\Theta_{S_{t-1}}, X_{t-1}), \sigma(\Theta_{S_{t-1}}, X_{t-1})).$$

Appendix E

Peng numerical form for the exponential of the transition intensity matrix

Peng [110] developed a theorem that provided a numerical form for the transition matrix Π in the case of continuous time finite state space Markov model using the eigenvalues λ_i's with multiplicity d_i of its infinitesimal generator Q, stated as follow:

Theorem E.1. *Let $(X_t)_{t\geq 0}$ be a Markov chain on a finite state space and denote $\lambda = \lambda_1, \lambda_2, ..., \lambda_M$ (may be complex and not distinct) the eigenvalues of its infinitesimal generator Q. Let the minimal polynomial of Q be of the form $f(x) = \prod_{i=0}^{M}(x-\lambda_i)^{d_i}$ where the λ_i, $i \geq 1$ are distinct and $d_i \geq 1$. Then*

$$\pi(t) = \sum_{j=0}^{d_i-1}\left[\frac{R(i,j)}{j!)}(Q-\lambda_i I)^j t^j\right] Exp(\lambda_i t)$$

Where

$$R(i,j) = \left[\prod_{m\neq j}\frac{(Q-\lambda_m I)^{d_m}}{(\lambda_i-\lambda_m)^{d_m}}\right]\left[I + \sum_{n=1}^{d_i-j-1} C_{i,n)}(Q-\lambda_i I)\right]$$

And $-C_{i,0} = 1$, *with*

$$-C_{i.n} = \sum_{\sum_{m\neq i} k_m \leq n;\, k_m \leq d_m;\, m\neq i}\left(\prod_{m\neq i}\frac{1}{(\lambda_i-\lambda_m)^{k_m}}\right) C_{i,(n-\sum_{m\neq i} k_m)}$$

For $1 \leq n \leq d_i - 1$.

Appendix F

An extract about FEV1 observations using data from the Danish Lung Cancer Screening Trial

We provide here an extract from the database about FEV1 observations concerning 2052 current or ex-smokers aged $50-70$ years having FEV1 measured annually for 5 years(DLCST) (more descriptions is given in [123]).

Table F.1: Extract about FEV1 observations using data from the Danish Lung Cancer Screening Trial (DLCST)

Patient	Visit	Visit date	FEV1
4	1	2004-10-05	1,97
4	2	2005-12-01	1,8
4	3	2007-01-03	2,03
4	4	2008-01-07	1,77
4	5	2009-01-19	1,55
6	1	2004-10-06	2,27
6	2	2005-12-01	2,05
6	3	2007-01-10	1,99
6	4	2007-10-29	1,87
6	5	2009-01-19	1,93
7	1	2004-10-06	1,93
7	2	2005-12-01	1,87
7	3	2007-01-24	2,02
7	4	2007-11-07	2,08
7	5	2009-01-19	1,85
8	1	2004-10-06	2,4
8	2	2005-12-01	2,5
8	3	2007-01-24	2,37
8	4	2007-11-07	2,28
8	5	2009-01-19	2,49
9	1	2004-10-06	3,54

Continued on next page

Table F.1 – *Continued from previous page*

Patient	Visit	Visit date	FEV1
9	2	2006-01-26	3,42
9	3	2007-01-24	3,09
9	4	2007-11-21	3,11
9	5	2009-01-19	2,97
11	1	2004-10-06	3,24
11	2	2005-12-01	3,15
11	3	2007-01-24	2,83
11	4	2007-11-07	3,11
11	5	2009-01-20	2,89
12	1	2004-10-06	4,59
12	2	2005-11-30	4,38
12	3	2007-02-28	4,44
12	4	2007-11-26	4,28
12	5	2009-02-02	4,21
13	1	2004-10-06	3,22
13	2	2005-12-01	3,22
13	3	2007-01-10	3,17
13	4	2007-12-05	3,13
13	5	2009-02-18	3,01
14	1	2004-10-06	2,68
14	2	2005-12-01	2,55
14	3	2006-01-03	2,34
14	4	2007-10-09	2,3
14	5	2009-01-19	2,15
18	1	2004-10-06	2,24
18	2	2005-12-01	2,25
18	3	2007-01-08	2,23
18	4	2007-10-09	2,03
18	5	2009-01-19	1,89
2024	1	2005-06-02	2,17
2024	3	2007-09-17	1,72
2024	4	2008-12-06	1,67
2024	5	2009-08-17	1,73
2036	1	2005-06-06	2,88
2036	2	2006-09-04	2,65
2036	3	2007-10-22	2,25
2036	4	2008-11-24	2,17
2036	5	2009-09-14	2,79
2039	1	2005-06-06	2,79
2039	2	2006-08-09	2,59
2039	3	2007-08-29	2,47

Continued on next page

Table F.1 – *Continued from previous page*

Patient	Visit	Visit date	FEV1
2039	4	2008-08-20	2,46
2039	5	2009-08-10	2,56
2044	1	2005-06-06	3,37
2044	2	2006-08-10	3,22
2044	3	2007-08-29	3,08
2044	4	2008-08-20	2,74
2044	5	2009-08-17	2,91
2043	1	2005-06-06	1,05
2043	2	2006-08-10	0,81
2043	3	2007-08-27	0,85
2043	4	2008-09-09	0,83
2043	5	2009-08-20	0,82
2046	1	2005-06-06	4,58
2046	2	2006-12-11	4,2
2046	3	2007-11-08	3,76
2046	4	2008-11-19	4,36
2046	5	2009-09-07	4,18
2063	1	2005-06-07	2,57
2063	2	2006-08-10	2,24
2063	3	2007-09-19	2,12
2063	4	2008-11-05	1,85
21	1	2004-10-06	3,44
21	2	2005-12-01	3,23
21	3	2007-01-24	2,93
21	4	2007-10-29	2,88
21	5	2009-01-20	2,73
22	1	2004-10-06	3,83
22	2	2005-12-01	3,77
22	4	2007-11-05	3,8
22	5	2009-01-29	3,54
24	1	2004-10-07	3,25
24	2	2005-12-01	3,05
24	3	2007-01-15	2,98
24	4	2008-02-11	3,22
24	5	2009-01-19	2,87
25	1	2004-10-07	3,78
25	2	2005-12-01	3,82
25	3	2007-01-29	3,48
25	4	2007-11-12	3,1
25	5	2009-01-26	3,07
26	1	2004-10-07	2,02

Continued on next page

Table F.1 – *Continued from previous page*

Patient	Visit	Visit date	FEV1
26	2	2005-12-01	2,14
26	3	2007-01-29	2,05
26	4	2007-11-12	1,82
26	5	2009-01-26	1,85
2073	1	2005-06-07	3,18
2073	2	2006-09-20	2,81
2073	3	2007-09-26	2,61
2073	4	2008-09-22	2,67
2073	5	2009-09-16	2,02
31	1	2004-10-07	3,46
31	2	2005-12-05	3,26
31	3	2007-01-31	3,32
31	4	2007-11-07	3,23
31	5	2009-01-20	3,1
32	1	2004-10-18	3,88
32	2	2005-11-24	3,97
32	3	2007-01-31	4,1
32	4	2007-11-12	3,9
32	5	2009-01-20	3,77
35	1	2004-10-18	3,71
35	2	2005-12-05	3,45
35	3	2007-01-10	3,3
35	4	2007-10-30	3,18
35	5	2009-01-19	2,97
38	1	2004-10-18	2,04
38	2	2005-12-05	2,01
38	3	2007-01-31	1,96
38	4	2007-11-12	1,82
38	5	2009-01-26	1,8
39	1	2004-10-18	3,08
39	2	2006-03-16	3,65
39	3	2007-01-10	3,5
39	4	2007-10-09	3,42
39	5	2009-02-19	3,01
2112	1	2005-06-09	3,68
2112	2	2006-09-21	3,52
2112	3	2007-09-03	3,61
2112	4	2008-09-01	3,39
2112	5	2009-05-27	3,51
41	1	2004-10-18	1,84
41	2	2005-12-05	1,76

Continued on next page

Table F.1 – *Continued from previous page*

Patient	Visit	Visit date	FEV1
41	3	2007-01-31	1,57
41	4	2007-10-22	1,8
41	5	2009-01-20	1,74
2115	1	2005-06-09	3,63
2115	2	2006-08-17	3,59
2115	3	2007-09-03	3,08
2115	4	2008-09-02	3,36
2115	5	2009-05-27	3,23
2116	1	2005-06-09	3,81
2116	2	2006-08-23	3,59
2116	3	2007-09-24	3,65
2116	4	2008-11-22	3,05
2116	5	2009-06-15	3,21
2119	1	2005-06-09	2,33
2119	2	2006-08-17	2,27
2119	3	2007-08-27	2,12
2119	4	2008-08-25	2,02
2119	5	2009-06-10	1,94
47	1	2004-10-18	3,69
47	2	2005-12-08	3,48
47	3	2007-01-10	3,43
47	4	2007-10-09	3,32
47	5	2009-01-27	3,43
48	1	2004-10-18	3,67
48	2	2005-12-05	3,33
48	3	2007-01-10	3,31
48	4	2007-10-30	3,21
48	5	2009-01-20	3,19
2121	1	2005-06-09	2,95
2121	2	2006-09-18	2,88
2121	3	2007-09-24	2,84

Appendix G

Matlab program for the discrete time Markov switching autoregressive model

```
% This is the simulation program(generation from some true values,
%then mcmc estimation)
%n is the number of individuals
%M maximum observations,UM is is 2*n vector containing the
%first time and
% last time for each individual with n individuals;D :
%MCMC iteration number,y: matrix of observations.
% uzero the initial minimum time, M maximum of all times
clc;clear;
uzero=1;M=50;n=100;N=0;D=8000;UM=zeros(2,n);
y=zeros(n,M);a=3;
for j=1:n
    UM(1,j)=randi([uzero,M-1],1,1);
    UM(2,j)=randi([UM(1,j)+1,M],1,1);
end
 %mi matrix for mu with initialization
mi=zeros(a,D);mi(:,1)=[12 24 36];
%beta matrix with initialization
bta=zeros(a,D);
bta(:,1)=[0.2 0.4 0.8];
%sigma matrix with initialization
sig=zeros(1,D);sig(1,1)=2;
%alphazeros and alpha for initial distribution
%and transition distribution
alphaa=[3 1 2;1 3 2;1 2 3];alphazeros=[1;1;1];
%generation of a sequence of markov chain
r0=[0.1 0.6 0.3];
p0=[0.7 0.2 0.1;0.1 0.6 0.3;0.2 0.3 0.5];
% markov chain generation by simulationMC function
xgen=simulationMC(r0,p0,M);
% Generation of the observations
for i=1:n
```

```
    initialobstime=UM(1,i);endobstime=UM(2,i);
        k=xgen(initialobstime);
        y(i,initialobstime)=mi(k,1)+sqrt(sig(1,1))*...
            randn(1);
     for t=initialobstime+1:endobstime
         N=N+1;k=xgen(t);
         y(i,t)=mi(k,1)+bta(k,1)*y(i,t-1)+sqrt(sig(1,1))*...
             randn(1);
    end
end
% Prior hyperparameters for the MARHMM model
%hyperparameters for sigma priors
epsilon=0.1;zeta=0.1 ;
%hyparameters for autoregessive  mu prior
alphaprime=[2 4 9];tau=[0.2 0.4 0.9];
%hyperparameters for autoregressive beta prior
b=[0.1 0.3 0.7];c=[0.01 0.05 0.09];
% transition parameters pij , intial distribution r,
%number of initial states
%nis,and number of transitions matrix from state to
%state ntss
r=zeros(a,D);r(:,1)=[0.2 0.5 0.3];
pii=zeros(a,a,D);
pii(:,:,1)=[0.5 0.3 0.2;0.2 0.4 0.4;0.1 0.2 0.7];
%fsmatrix saves forward probabilities bsmatrix saves
%backward probabilities
% for chib algorithm
fsmatrix=zeros(a,M,D);bsmatrix=zeros(a,M,D);x=zeros(M,D);
nis=zeros(1,a);mn=zeros(1,a);
%The main algorithm
x(uzero,1)=randsample(a,1,true,r(:,1));
for t=uzero+1:M
    x(t,1)=randsample(a,1,true,pii(x(t-1,1),:,1));
end
for d=1:D
% Forward backward simulation of the hidden states
%initialisation first state
sumfzero=0;
for k=1:a
    sumfzero=sumfzero+...
        obsdensityfzero(k,uzero,n,y,UM,mi,sig,d)*...
        r(k,d);
end
for k=1:a
```

```
 fsmatrix(k,uzero,d+1)=...
      obsdensityfzero(k,uzero,n,y,UM,mi,sig,d)*...
      r(k,d)/sumfzero;
end
for t=uzero+1:M
    %first calculate forward matrix for every k and the
    %time t minus one
    for k=1:a
        fsmatrixminusone(k,t,d+1)=0;
        for l=1:a
            fsmatrixminusone(k,t,d+1)=...
                fsmatrixminusone(k,t,d+1)+...
            pii(l,k,d)*fsmatrix(l,t-1,d+1);
        end
    end
    %now we calculate forward matrix for every k
    %at the time t
    sumfs=0;
    for l=1:a
        sumfs=sumfs+fsmatrixminusone(l,t,d+1)*...
        observationdensityf(l,t,n,UM,y,bta,mi,sig,d);
    end
    for k=1:a
        fsmatrix(k,t,d+1)=fsmatrixminusone(k,t,d+1)*...
        observationdensityf(k,t,n,UM,y,bta,mi,sig,d)/sumfs;
    end
end
%Backward calculation
% we deduce first the simulation of x_M from
%forward calculation
mn=mnrnd(1,fsmatrix(:,M,d+1));
x(M,d+1)=find(mn==1);
%we  simulate  x_t from forward calculation
for t=M-1:-1:uzero
    j=x(t+1,d+1);
    sumbs=0;
    for l=1:a
        sumbs=sumbs+fsmatrix(l,t,d+1)*pii(l,j,d);
    end
    for k=1:a
        bsmatrix(k,t,d+1)=fsmatrix(k,t,d+1)*pii(k,j,d)/...
            sumbs;
    end
    mn=mnrnd(1,bsmatrix(:,t,d+1));
```

```
    x(t,d+1)=find(mn==1);
end
%calculation of the  number of initial states(nis) and
%the number of transition
%from state to state(ntss)
%number initial state nis
nis=zeros(1,a);
%number transition from state to state
l=x(uzero,d+1);
nis(1,l)=nis(1,l)+1;
%nis=2.*nis;
ntss=zeros(a,a);
for t=uzero+1:M
    k=x(t-1,d+1);
    l=x(t,d+1);
    ntss(k,l)=ntss(k,l)+1;
end
%siumulation of initial distribution r by gibbs sampler
r(:,d+1)=drchrnd1(alphazeros(1,:)+nis(1,:));
for k=1:a
    pii(k,:,d+1)=drchrnd1(alphaa(k,:)+ntss(k,:));
end
%Simulation of mu (mi) by gibbs sampler
tau1=zeros(1,a);alpha1=zeros(1,a);
for l=1:a
    summu1=0;summu2=0;nl=0;nul=0;
    for i=1:n
        if (x(UM(1,i),d+1)==l)
            nul=nul+1;
            summu1=summu1+y(i,UM(1,i));
        end
        for t=UM(1,i)+1:UM(2,i)
            if (x(t,d+1)==l)
                summu2=summu2+y(i,t)-bta(l,d)*y(i,t-1);
                nl=nl+1;
            end
        end
     end
    tau1(l)=(nl+nul)/sig(1,d)+1/tau(l);tau1(l)=1/tau1(l);
    alpha1(l)=tau1(l)*(((summu1+summu2)/sig(1,d))+...
        (alphaprime(l)/tau(l)));
    mi(l,d+1)=normrnd(alpha1(l),sqrt(tau1(l)));
 end
%Simulation of beta (bta) by gibbs sampler
```

```
b1=zeros(1,a);c1=zeros(1,a);
for l=1:a
    sumbeta1=0;sumbeta2=0;
    for i=1:n
        for t=UM(1,i)+1:UM(2,i)
            if (x(t,d+1)==l)
                sumbeta1=sumbeta1+(y(i,t-1))^2;
                sumbeta2=sumbeta2+(y(i,t)-mi(l,d+1))*...
                    y(i,t-1);
            end
        end
    end
    c1(l)=1/c(l)+(sumbeta1/sig(1,d));c1(l)=1/c1(l);
    b1(l)=c1(l)*(b(l)/c(l)+(sumbeta2/sig(1,d)));
    bta(l,d+1)=normrnd(b1(l),sqrt(c1(l)));
end
%Simulation of sigma(sig) by gibbs sampler
sumsigma1=0;sumsigma2=0;
for i=1:n
    l=x(UM(1,i),d+1);
    sumsigma1=sumsigma1+(y(i,UM(1,i))-mi(l,d+1))^2;
    for t=UM(1,i)+1:UM(2,i)
        l=x(t,d+1);
        sumsigma2=sumsigma2+...
            (y(i,t)-mi(l,d+1)-bta(l,d+1)*y(i,t-1))^2;
    end
end
sig(1,d+1)=1/(gamrnd(epsilon+n/2+N/2,1/...
    (zeta+(sumsigma1+sumsigma2)/2)));
end
%End the MCMC iterations
pi1=pii(1,:,:);pi1=squeeze(pi1);
pi2=pii(2,:,:);pi2=squeeze(pi2);
pi3=pii(3,:,:);pi3=squeeze(pi3);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Some of the intermediate functions used in this
%simulation are:
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function r = drchrnd1(a)
%this is a one sample draw from dirchlet random
p=length(a);a2=zeros(1,p);r=zeros(1,p);
a2(1)=sum(a(2:p));
r(1)=betarnd(a(1),a2(1),1,1);
for j=2:p-1
```

```
    a2(j)=sum(a(j+1:p));
    r(j)=betarnd(a(j),a2(j),1,1)*(1-sum(r(1:j-1)));
end
r(p)=1-sum(r(1:p-1));
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%draw from an inverse gamma
function p = invgampdf(x,a,b )
%the computation of inverse gamma pdf from gamma pdf
p=b^a/gamma(a).*(1./x).^(a+1).*exp(-b./x);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function obsfzero =...
    obsdensityfzero(k,uzero,n,y,UM,mi,sig,d)
%this is the degenerate observation density
   %function for the time uzero
logobsf=0;
for i=1:n
    z=UM(1,i);
    if z==uzero
       logobsf=logobsf+log(normpdf(y(i,uzero),mi(k,d),...
           sqrt(sig(1,d))));
    end
end
obsfzero=exp(logobsf);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
 function obsf =...
     observationdensityf(k,t,n,UM,y,bta,mi,sig,d)
%this is the observation density
     %function f(y(t)/y,t-1,o_k)
logobsf=0;
for i=1:n
    debut=UM(1,i);
    fin=UM(2,i);
    if  (debut<t)&&(t<=fin)
            logobsf=logobsf+log(normpdf(y(i,t)-...
                mi(k,d)-bta(k,d)*...
                y(i,t-1),0,sqrt(sig(1,d))));
    end
end
obsf=exp(logobsf);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
function states = simulationMC(r,pii,T)
%This function generate a sequence of a markov
%chain form intial distribution r,transtion probability pii,
%for length T %a=length(r);
states=zeros(1,T);
mn=mnrnd(1,r);states(1)=find(mn==1);
for t=2:T
    mn=mnrnd(1,pii(states(t-1),:));states(t)=find(mn==1);
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix H

Matlab simulation program for the hybrid switching diffusion model

```
% M maximum observations number, ni() is 1*n vector
%containing the lengths of obsevation or each individual
%with n individuals; Ni() is the same vector
%but after data imputation .D:MCMC iteration number,
%a is the number of the discrete Markov states, q is
%the transition % matrix for the n individuals and
%M is the maximum observations
%number for each individual before data imputation
%Mx is the the maximum observation for each
% individual after imputation
%s contains the hidden states and the times of switching
%for the states.
% x is the observation matrix before imputation.
% xxim is the observation matrix after imputation.
%sdd is the matrix for the hidden state after imputation.
% First we generate a process from some true values,
%then we ran our MCMC
% for sufficiently a large number of iteration D
%to obtain an output from each parrameter.
rng('shuffle');
M=10;Mx=120;n=30;D=8000;ni=zeros(n,1);a=3;s=zeros(2,n,M);
x=zeros(n,M);
% Generating ni between 7 and M
for i=1:n
    ni(i)=randi([7,M],1,1);
end
%Parameter of the diffusion process with initialization
%alp0 is the MCMC vector for the alpha0.
%alp1 is the MCMC vector for the alpha1.
%alp2 is the MCMC vector for the alpha2.
%alp3 is the MCMC vector for the alpha3.
```

```
%Parameter of the diffusion process with initialization
alp0=zeros(a,D);alp1=zeros(a,D);alp2=zeros(a,D);
alp3=zeros(a,D);
%ph0 is the generation vector for alpha0.
%ph1 is the generation vector for alpha1.
%ph2 is the generation vector for alpha2.
%ph3 is the generation vector for alpha3
ph0=[0.3 ;0.6 ;0.9];ph1=[0.02;0.04;0.06];
ph2=[-0.08;-0.06;-0.04];
ph3=[0.01;0.05;0.09];
alp0(:,1)=[0.29 ;0.61 ;0.87];
alp1(:,1)=[0.018;0.043;0.056];
alp2(:,1)=[-0.084;-0.055;-0.039];
alp3(:,1)=[0.009;0.052;0.1];
b0=[0.3 0.02 -0.08 0.01;0.6 0.04 -0.06 0.05;...
    0.9 0.06 -0.04 0.09];
B0=zeros(a,4,4);
b01=squeeze(b0(1,:));
b02=squeeze(b0(2,:));
b03=squeeze(b0(3,:));
%Hyper-parameters for the prior in the case
%for switching regression.
B0(1,:,:)=[0.01 0 0 0;0 0.01 0 0;0 0 0.01 0;0 0 0 0.01];
B01=squeeze(B0(1,:,:));
B0(2,:,:)=[0.02 0 0 0;0 0.02 0 0;0 0 0.02 0;0 0 0 0.02];
B02=squeeze(B0(2,:,:));
B0(3,:,:)=[0.03 0 0 0;0 0.03 0 0;0 0 0.03 0;0 0 0 0.03];
B03=squeeze(B0(3,:,:));
% eta matrix with initialization
eta=zeros(a,D);eta0=[0.3;0.4;0.5];
eta(:,1)=[0.3;0.4;0.5];
%sigma matrix with initialization
sig=zeros(1,D);sig0=0.02;sig(1,1)=0.017;
% An initial distribution for the hidden states
r=zeros(a,D);r0=[0.3 0.3 0.4];r(:,1)=[0.31;0.31;0.36];
%alphazeros  for the prior of the initial distribution.
alphazeros=[1;1;1];
%Parameters of Gompertz function in the transition rate
%matrix q0
% The lamdas lmda0 and the gammas gma0, parmeter used to
%generate the first  q0
lmda0=[0 0.2 1;0.3 0 2;0.4 7 0];
gma0=[0 0.2 0.04;0.3 0 0.06;0.4 0.08 0];
lmda(:,:,1)=[0 0.18 1.02;0.29 0 1.9;0.37 6.77 0];
```

```
gma(:,:,1)=[0 0.22 0.043;0.27 0 0.055;0.39 0.078 0];
%the q0 saves for all the n individuals and the
%the times before imputation
q0=zeros(a,a,n,M);
%Generation of the diffusion observations and
%the Markov switching
%states
for i=1:n
    %choose the first regime switching state
    s(1,i,1)=samplefromp(r0,1);k= s(1,i,1);s(2,i,1)=0;
    %Initialize the first observation in the diffusion
    %process
    x(i,1)=normrnd((ph0(k)),sig0(1));
    %compute the transition matrix for the initial time ti1
    qadjust=zeros(a,a);
    for k=1:a
      for l=1:a
        if l~=k
         qadjust(k,l)=lmda0(k,l)*exp(-gma0(k,l)*x(i,1));
         qadjust=normalizeq(qadjust,k);
        end
     end
    end
    q0(:,:,i,1)=qadjust(:,:);rr=s(1,i,1);
     %Simulation of the next time ti2 and save it in
     %s(2,i,2)
    er=1/(-q0(rr,rr,i,1));
    s(2,i,2)=exprnd(er);
    %Computation of the hidden states for the time tij
   for j=2:ni(i)
       qminus=zeros(2,a-1);l=1;
       rr=s(1,i,j-1);
     for k=1:a
        if (k~=rr)
           qminus(1,l)=q0(rr,k,i,j-1);qminus(2,l)=k;l=l+1;
        end
     end
     mnqminus=zeros(1,a-1);
     mnqminus(1,:)=qminus(1,:)/(-q0(rr,rr,i,j-1));
     mn=mnrnd(1,mnqminus(1,:));
     f=find(mn,1,'first');s(1,i,j)=qminus(2,f);
     k=s(1,i,j-1);
     %Computation of the observation xij
     x(i,j)=x(i,j-1)+(ph0(k,1)+ph1(k,1)*x(i,j-1)+ph2(k,1)*...
```

```
        (x(i,j-1))^2+...
        ph3(k,1)/x(i,j-1))...
    *(s(2,i,j)-s(2,i,j-1))+sqrt(sig0(1,1))*((x(i,j-1))^...
    (eta0(k,1)))*...
    (sqrt(s(2,i,j)-s(2,i,j-1)));
    %Simulation of the transition rate matrix for the
    %time tij
    qadjust=zeros(a,a);
     for k=1:a
      for l=1:a
        if l~=k
         qadjust(k,l)=lmda0(k,l)*exp(-gma0(k,l)*x(i,j-1));
         qadjust=normalizeq(qadjust,k);
        end
      end
    end
    q0(:,:,i,j)=qadjust(:,:);rr=s(1,i,j);
    %Computation of the next time tij+1
     if j<ni(i)
        er=1/(-q0(rr,rr,i,j));
        s(2,i,j+1)=exprnd(er)+s(2,i,j);
     end
  end
end
%After the generation, we start our MCMC algorithm
%Data imputation for d=1 and initialization of
%the switching states
%xim, st and td are respectively the imputed data,
%the hidden states and the times of imputations
%N1(i)and N(i) are the number of observations including
%the imputed data for individual i for both the
%previous MCMC
%iteration d and the new MCMC iteration d.
%td1 and td are the times of imputation for both the
%previous MCMC iteration
%and the new MCMC iteration.
%xs1 and xs save the positions of the original
%observations after
% imputation .
N1=zeros(n,1);td1=zeros(n,Mx);xs1=zeros(n,M);
N=zeros(n,1);td=zeros(n,Mx);xs=zeros(n,M);
%An intermediate matrix to save the hidden
%states for all the tij.
st=zeros(n,Mx);
```

```
%The new transition rate matrix q for all the times tij
%including the new imputed times
q=zeros(a,a,n,Mx);
%Creation of the times of imputation using
%an exponential random
% s2id saves all diagonal elements for q0 for all the times
s2id=zeros(a,n,M);
for i=1:n
    for j=1:ni(i)
        for k=1:a
            s2id(k,i,j)=-q0(k,k,i,j);
        end
    end
end
%The maximum of the diagonal elements
s2imax=(max(s2id(:))+0.05);

for i=1:n
    xs1(i,1)=1;
    N1(i)=1;
    td1(i,1)=s(2,i,1);
   for j=2:(ni(i))
       xx=exprnd(1/s2imax);
       tdd=s(2,i,j-1)+xx;
       while tdd<s(2,i,j)
       N1(i)=N1(i)+1;
       td1(i,N1(i))=tdd;
       xx=exprnd(1/s2imax);
       tdd=tdd+xx;
       end
      N1(i)=N1(i)+1;
      td1(i,N1(i))=s(2,i,j);
      xs1(i,j)=N1(i);
   end
end
%Creation of the imputed observations and
%the corresponding hidden states for MCMC iteration d=1.
%The matrix of the imputed data xim
xim=zeros(n,Mx);
for i=1:n
   k=1;
   for j=1:N1(i)
     if j==xs1(i,k)
        xim(i,j)=x(i,k);st(i,j)=s(1,i,k);
```

```
        q(:,:,i,j)=q0(:,:,i,k);k=k+1;
      else
        qminus=zeros(2,a-1);ss =1;kk=st(i,j-1);
        for rr=1:a
            if (rr~=kk)
               qminus(1,ss)=q(kk,rr,i,j-1);
               qminus(2,ss)=rr;ss=ss+1;
            end
        end
        mnqminus=zeros(1,a-1);
        mnqminus(1,:)=qminus(1,:)/(-q(kk,kk,i,j-1));
        mn=mnrnd(1,mnqminus(1,:));f=find(mn,1,'first');
        st(i,j)=qminus(2,f);
        kk=st(i,j-1);
        %Compute the imputed data
        xim(i,j)=xim(i,j-1)+(ph0(kk,1)+ph1(kk,1)*...
            xim(i,j-1)+ph2(kk,1)*...
            (xim(i,j-1))^2+ph3(kk,1)/xim(i,j-1))...
        *(td1(i,j)-td1(i,j-1))+sqrt(sig0(1,1))*...
        ((xim(i,j-1))^(eta0(kk,1)))*...
        (sqrt(td1(i,j)-td1(i,j-1)));
       qadjust=zeros(a,a);
          for rr=1:a
              for ss=1:a
                  if ss~=rr
                     qadjust(rr,ss)=lmda(rr,ss,1)*...
                         exp(-gma(rr,ss,1)*...
                         xim(i,j-1));
                     qadjust=normalizeq(qadjust,rr);
                  end
              end
          end
          %Update the transition rate matrix
         q(:,:,i,j)=qadjust(:,:);
      end
    end
end
%xxd saves the observations including the imputed
%observations for all MCMC iterations.
%sdd saves the hidden states for all MCMC iterations
%including the stacked hidden states
%qold the transition rate matrix is saved for
%use in the next MCMC
%iteration
```

```
xxd=zeros(n,Mx,D);sdd=zeros(n,Mx,D);qold=zeros(a,a,n,Mx);
for i=1:n
xxd(i,:,1)=xim(i,:);
sdd(i,:,1)=st(i,:);
  for j=1:N1(i)
    qold(:,:,i,j)=q(:,:,i,j);
  end
end
XIMP=zeros(n,Mx);qim=zeros(a,a,n,Mx);
%The main algorithm
for i=1:n
N(i)=N1(i);td(i,:)=td1(i,:);xs(i,:)=xs1(i,:);
end
sim=zeros(n,Mx);
for d=1:D
%Metropolis hasting algorithm for the imputed data
%XIMP will save the old observations including
%the imputed ones
XIMP=squeeze(xxd(:,:,d));
%First create the proposal times of imputation
for i=1:n
    xs1(i,1)=1;
    N1(i)=1;
    td1(i,1)=s(2,i,1);
   for j=2:(ni(i))
       xx=exprnd(1/s2imax);
       tdd=s(2,i,j-1)+xx;
       while tdd<s(2,i,j)
       N1(i)=N1(i)+1;
       td1(i,N1(i))=tdd;
       xx=exprnd(1/s2imax);
       tdd=tdd+xx;
       end
      N1(i)=N1(i)+1;
      td1(i,N1(i))=s(2,i,j);
      xs1(i,j)=N1(i);
   end
end
%Create the proposed imputed data xim
% this imputation gives rise to the computation
%of intermediate
%transition rate matrix qim and the simulation of
%intermediate hidden states sim.
for i=1:n
```

```
Nom1=1;Den1=1; Nom2=1;Den2=1;l=1;
qold(:,:,i,:)=q(:,:,i,:);
while l<(ni(i))
 km1=xs1(i,l);km2=xs1(i,l+1);km11=xs(i,l);
 km21=xs(i,l+1);
 xim(i,km1)=x(i,l);xim(i,km2)=x(i,l+1);
 sim(i,km1)=sdd(i,km11,d);
 sim(i,km2)=sdd(i,km21,d);
 qim(:,:,i,km1)=qold(:,:,i,km11);
 qim(:,:,i,km2)=qold(:,:,i,km21);
if    km2-km1<=1
      l=l+1;
else
   for j=km1+1:km2-1
       kk=sim(i,j-1);
       %Compute the imputed data using modified
       %browanian proposal
       xim(i,j)=normrnd((xim(i,j-1)+...
            (xim(i,km2)-xim(i,j-1))*...
            ((td1(i,j)-td1(i,j-1))/...
       (td1(i,km2)-td1(i,j-1)))),((sig(1,d))*...
       ((xim(i,j-1))^...
       (2*eta(kk,d)))*...
      (td1(i,j)-td1(i,j-1))*(td1(i,km2)-td1(i,j))/...
      (td1(i,km2)-td1(i,j-1))));
      qadjust=zeros(a,a);
      for rr=1:a
            for ss=1:a
                if ss~=rr
                   qadjust(rr,ss)=lmda(rr,ss,1)*...
                        exp(-gma(rr,ss,1)*xim(i,j-1));
                   qadjust=normalizeq(qadjust,rr);
                end
            end
       end
       %Check whether there is a switch
      qim(:,:,i,j)=qadjust(:,:);
      ratioswitch=min(1,((-qim(kk,kk,i,j-1))/s2imax));
      %If no switch keep the previous hidden state
      u=rand();
      if ratioswitch<=u
          sim(i,j)=sim(i,j-1);
      %Else draw the hidden state sim(i,j)
      else
```

```
        qminus=zeros(2,a-1);ss =1;kk=sim(i,j-1);
        for rr=1:a
          if (rr~=kk)
            qminus(1,ss)=qim(kk,rr,i,j-1);
            qminus(2,ss)=rr;ss=ss+1;
          end
        end
        mnqminus=zeros(1,a-1);
        mnqminus(1,:)=qminus(1,:)/(-qim(kk,kk,i,j-1));
        mn=mnrnd(1,mnqminus(1,:));f=find(mn,1,'first');
        sim(i,j)=qminus(2,f);
      end
      kk=sim(i,j-1);
      Den1=Den1*normpdf(xim(i,j),xim(i,j-1)+...
          (xim(i,km2)-xim(i,j-1))*...
          (td1(i,j)-td1(i,j-1))/...
      (td1(i,km2)-td1(i,j-1)),((sig(1,d))*...
      ((xim(i,j-1))^(2*eta(kk,d)))*...
      (td1(i,j)-td1(i,j-1))*(td1(i,km2)-td1(i,j))/...
      (td1(i,km2)-td1(i,j-1))));
    end
l=l+1;
end
end

xim(i,(N1(i)))=x(i,(ni(i)));
qim(:,:,i,N1(i))=qold(:,:,i,N(i));

 for  j=2:(N1(i))
      k=sim(i,j-1);
      Nom2=Nom2*normpdf(xim(i,j),xim(i,j-1)+...
          (alp0(k,d)+...
          alp1(k,d)*xim(i,j-1)...
          +alp2(k,d)*((xim(i,j-1))^2)+...
          (alp3(k,d)/xim(i,j-1)))...
      *(td1(i,j)-td1(i,j-1)),sig(1,d)*...
      ((xim(i,j-1)^(2*eta(k,d))))*...
      (td1(i,j)-td1(i,j-1)));
 end
l=1;
while l<(ni(i))
      km11=xs(i,l);km21=xs(i,l+1);
      if   km21-km11<=1
           l=l+1;
```

```
            else
              for j=km11+1:km21-1
                k=sdd(i,j-1,d);
                Nom1=Nom1*normpdf(XIMP(i,j),XIMP(i,j-1)+...
                     (XIMP(i,km21)-XIMP(i,j-1))...
                     *(td(i,j)-td(i,j-1))/...
                (td(i,km21)-td(i,j-1)),((sig(1,d))*...
                ((XIMP(i,j-1))^(2*eta(k,d)))*...
                (td(i,j)-td(i,j-1))*(td(i,km21)-td(i,j))/...
                (td(i,km21)-td(i,j-1))));
              end
            l=l+1;
            end
      end
    Nomvector1(i,d)=Nom1;
     for   j=2:(N(i))
            k=sdd(i,j-1,d);
            Den2=Den2*normpdf(XIMP(i,j),XIMP(i,j-1)+...
                 (alp0(k,d)+alp1(k,d)*...
                 XIMP(i,j-1)...
                 +(alp2(k,d)*(XIMP(i,j-1))^2)+...
                 (alp3(k,d)/XIMP(i,j-1)))...
             *(td(i,j)-td(i,j-1)),sig(1,d)*...
             ((XIMP(i,j-1)^(2*eta(k,d))))...
             *(td(i,j)-td(i,j-1)));
      end
       %Finally update the xxd,q, and sim after
       %deciding based on the MHA
        mh=min(1,Nom1*Nom2/Den1*Den2);u=rand();
        if mh>u
            N(i)=N1(i);td(i,:)=td1(i,:);
            q(:,:,i,:)=qim(:,:,i,:);
            xs(i,:)=xs1(i,:);
            xxd(i,:,d+1)=xim(i,:);
        else
            xxd(i,:,d+1)=XIMP(i,:);N1(i)=N(i);
            td1(i,:)=td(i,:);
            q(:,:,i,:)=qold(:,:,i,:);
            sim(i,:)=squeeze(sdd(i,:,d));
        end
end


% Forward backward simulation of the hidden states
```

```
%fsmatrix saves forward probabilities bsmatrix saves
%backward probabilities
for i=1:n
    fsmatrix=zeros(a,Mx);bsmatrix=zeros(a,Mx);
    fsmatrixminusone=zeros(a,Mx);
    %initialisation first state
    sumfzero=0;
    for k=1:a
        sumfzero=sumfzero+...
            obsdensityfzero(k,i,xxd,alp0,sig,d)*r(k,d);
    end
    for k=1:a
        fsmatrix(k,1)=obsdensityfzero(k,i,xxd,alp0,sig,d)*...
            r(k,d)/sumfzero;
    end
    for j=2:N1(i)
        %calculate the transition probability matrix
        %for each time tij
        p=squeeze(q(:,:,i,j))*(td1(i,j)-td1(i,j-1));
        p=expm(p);
        %first calculate forward matrix for every k
        %and the time tij-1
        for k=1:a
            fsmatrixminusone(k,j)=0;
            for l=1:a
                fsmatrixminusone(k,j)=fsmatrixminusone(k,j)+...
                p(l,k)*fsmatrix(l,j-1);
            end
        end

        %now we calculate forward matrix for every k
        %at the time tij
        sumfs=0;k=sim(i,j-1);
         for l=1:a
             sumfs=sumfs+fsmatrixminusone(l,j)*...
             observationdensityf(k,i,j,xxd,alp0,alp1,...
             alp2,alp3,eta,sig,td1,d);
         end
         for l=1:a
             fsmatrix(l,j)=fsmatrixminusone(l,j)*...
             observationdensityf(k,i,j,xxd,alp0,alp1,...
             alp2,alp3,eta,sig,td1,d)...
             /sumfs;
         end
```

```
    end


%Backward calculation
% we deduce first the simulation of sdd_N1i(i) from
forward calculation
mn=mnrnd(1,fsmatrix(:,N1(i)));
sdd(i,N1(i),d+1)=find(mn,1,'first');
%x(i,ni(i),d+1)=randsample(a,1,true,fsmatrix(:,ni(i)));
%we  simulate  sdd_ij from forward calculation
for j=N1(i)-1:-1:1
    ind=sdd(i,j+1,d+1);p=squeeze(q(:,:,i,j+1))*...
        (td1(i,j+1)-td1(i,j));
    p=expm(p);
    sumbs=0;
    for l=1:a
        sumbs=sumbs+fsmatrix(l,j)*p(l,ind);
    end
    for k=1:a
        bsmatrix(k,j)=fsmatrix(k,j)*p(k,ind)/sumbs;
    end
   mn=mnrnd(1,bsmatrix(:,j));
   sdd(i,j,d+1)=find(mn,1,'first');
 end
end

%number of intial states nis
nis=zeros(1,a);
for i=1:n
    for k=1:a
        if (sdd(i,1,d+1)==k)
            nis(1,k)=nis(1,k)+1;
        end
    end
end
% Initial distribution posterior using a dirchlet posterior
r(:,d+1)=drchrnd1(alphazeros(1,:)+nis(1,:));

%Simulation of the parameters of the Gompertz model
%in the transition rate
%matrix
%1)simulation of lamda's
lamdaprime=squeeze(lmda(:,:,d));qn=zeros(a,a);
lamdaproposal=zeros(a,a);
```

```
aq=zeros(a,a);bq=zeros(a,1);nm=1;dn=1;qproposal=q;
qold=q;priornew=1;
 priorold=1;
 for k=1:a
     for l=1:a
         if l~=k
            loglamda=...
                normrnd(log(mean2(lmda(k,l,1:d))),0.02);
            lamdaproposal(k,l)=exp(loglamda);
            priorold=priorold*...
                gampdf(lmda(k,l,d),0.01,0.01);
            priornew=priornew*...
                gampdf(lamdaproposal(k,l),0.01,0.01);
         end
     end
 end
for k=1:a
     for l=1:a
         if l~=k
            for i=1:n
                qproposal(k,l,i,1)=lamdaproposal(k,l)...
                    *exp(-gma(k,l,d)*...
                    xxd(i,1,d+1));
                qn=squeeze(qproposal(:,:,i,1));
                qn=normalizeq(qn,k);
                qproposal(:,:,i,1)=qn(:,:);
                for j=2:N1(i)
                    qproposal(k,l,i,j)=lamdaproposal(k,l)*...
                        exp(-gma(k,l,d)*...
                        xxd(i,j-1,d+1));
                    qn=squeeze(qproposal(:,:,i,j));
                    qn=normalizeq(qn,k);
                    qproposal(:,:,i,j)=qn(:,:);
                end
            end
         end
     end
end

for i=1:n
     for j=2:N1(i)
         pnew=squeeze(qproposal(:,:,i,j))*...
              (td1(i,j)-td1(i,j-1));
         pnew=expm(pnew);
```

```
        pex=squeeze(qold(:,:,i,j))*...
            (td1(i,j)-td1(i,j-1));
        pex=expm(pex);
        nm=nm*pnew(sdd(i,j-1,d+1),sdd(i,j,d+1));
        dn=dn*pex(sdd(i,j-1,d+1),sdd(i,j,d+1));
    end
end
% The Metropolis Hastings algorithm for the
%lamdas (prior*likelihood)
prnew=priornew*nm;
prold=priorold*dn;
mh=min(1,prnew/prold);u=rand();
if u<mh
    lmda(:,:,d+1)=lamdaproposal(:,:);q=qproposal;
else
   lmda(:,:,d+1)=lmda(:,:,d);
end


%Simulation of the parameters of the gompertz
%model in the transition rate
%matrix
%2)simulation of gamma's
gammaprime=squeeze(gma(:,:,d));gammaproposal=zeros(a,a);
priorold=1;priornew=1;
for k=1:a
    for l=1:a
        if l~=k
           loggamma=...
               normrnd(log(mean2(gma(k,l,1:d))),0.04);
           gammaproposal(k,l)=exp(loggamma);
           priornew=priornew*...
               gampdf(gammaproposal(k,l),0.01,0.01);
           priorold=priorold*gampdf(gma(k,l,d),0.01,0.01);
        end
    end
end
nm=1;dn=1;qproposal=q;qold=q;
for k=1:a
    for l=1:a
        if l~=k
           for i=1:n
               qproposal(k,l,i,1)=lmda(k,l,d+1)*...
                   exp(gammaproposal(k,l)*...
```

```
                    xxd(i,1,d+1));
                qn=squeeze(qproposal(:,:,i,1));
                qn=normalizeq(qn,k);
                qproposal(:,:,i,1)=qn(:,:);
                for j=2:N1(i)
                    qproposal(k,l,i,j)=lmda(k,l,d+1)*...
                        exp(gammaproposal(k,l)*...
                        xxd(i,j-1,d+1));
                    qn=squeeze(qproposal(:,:,i,j));
                    qn=normalizeq(qn,k);
                    qproposal(:,:,i,j)=qn(:,:);
                end
            end
        end
    end
end
for i=1:n
    for j=2:N1(i)
        pnew=squeeze(qproposal(:,:,i,j))*...
            (td1(i,j)-td1(i,j-1));
        pnew=expm(pnew);
        pex=squeeze(qold(:,:,i,j))*...
            (td1(i,j)-td1(i,j-1));
        pex=expm(pex);
        nm=nm*pnew(sdd(i,j-1,d+1),sdd(i,j,d+1));
        dn=dn*pex(sdd(i,j-1,d+1),sdd(i,j,d+1));
    end
end
 % The Metropolis Hastings algorithm ratio for
 %the gammas (prior*likelihood)
prnew=priornew*nm;
prold=priorold*dn;
mh=min(1,prnew/prold);u=rand();
if u<mh
    gma(:,:,d+1)=gammaproposal(:,:);q=qproposal;
else
    gma(:,:,d+1)=gma(:,:,d);
end
%Simulation of alpha0,alpha1,alph2,alpha3,sigma
%using a switching regression appraoch:
B1=zeros(4,4);Y=zeros(a,2000);y=zeros(a,4,2000);
kj=zeros(4,4);
A=0.22;B=0.01;
c1=zeros(a,1);C1=zeros(a,1);bta=zeros(a,4);mx=zeros(a,1);
```

```
% bta is the matrix alpha0,alpha1,alph2,alpha3 that depends
%on three hidden
% states.
for k=1:a
 bta(k,:)=[alp0(k,d),alp1(k,d),alp2(k,d),alp3(k,d)];
end
for i=1:n
    for j=2:N1(i)
        k=sdd(i,j-1,d+1);mx(k)=mx(k)+1;
        Y(k,mx(k))=(xxd(i,j,d+1)-xxd(i,j-1,d+1))/...
            ((xxd(i,j-1,d+1)^(eta(k,d)))*...
            sqrt((td1(i,j)-td1(i,j-1))));
        sss=[1,xxd(i,j-1,d+1),(xxd(i,j-1,d+1))^2,1/...
            xxd(i,j-1,d+1)];
        sc=sqrt((td1(i,j)-td1(i,j-1)))/...
            ((xxd(i,j-1,d+1)^(eta(k,d))));
        y(k,:,mx(k))=sss*sc;
    end
end
sumc1=0;sumC1=0;
for k=1:a
   YR=zeros(1,mx(k));yr=zeros(1,4,mx(k));
   YR(1,:)=squeeze(Y(k,1:mx(k)));YRS=squeeze(YR(1,:));
   yr(1,:,:)=squeeze(y(k,:,1:mx(k)));yrs=squeeze(yr(1,:,:));
   yrYR=yrs*YRS';yryr=yrs*transpose(yrs);YRSYRS=YRS*YRS';
% Switching ridge regression start by updating
%the diagonal elements of the
% ridge matrix used to overcome collinarity
    for i=1:4
        %kk(i,i)=gamrnd((A+mx(k)),(B+(bta(k,i))^2));
        kj(i,i)=4*(A+mx(k))/(B+bta(k,i)^2);
        %kj(i,i)=5/(sum(bta(k,:).^2));
    end
  BB0=yryr/sig(1,d)+kj/sig(1,d);
  b0k=squeeze(b0(k,:));bb0k=kj*b0k';
  B1=inv(BB0);
   bb=((sig(1,d)*BB0)\(yrYR+bb0k));
   %bb=BB0\(yrYR+bb0k);
   bta(k,:)=mvnrnd(bb,B1);
   alp0(k,d+1)=bta(k,1);alp1(k,d+1)=bta(k,2);
   alp2(k,d+1)=bta(k,3);alp3(k,d+1)=bta(k,4);
   btas=squeeze(bta(k,:));
   %Simulation of sigma from inverse gamma
   sumC1=sumC1+YRSYRS/2+b0k*(BB0\b0k')/2-bb'*(BB0\bb)/2;
```

```
    sumc1=sumc1+mx(k,1)/2;
end
sumc1=sumc1+A;
sumC1=sumC1+B;
sig(1,d+1)=(sumC1/chi2rnd(sumc1))/12;
%Simulation of eta by metropolis hasting algorithm
etaold=squeeze(eta(:,d));etanew=zeros(a,1);
for k=1:a
    logeta=normrnd(log(mean2(eta(k,1:d))),0.03);
    etanew(k,1)=exp(logeta);
    priornew=gampdf(etanew(k,1),0.01,0.01);
    priorold=gampdf(etaold(k,1),0.01,0.01);
    nom=1;den=1;
    for i=1:n
      for j=2:N(i)
          if sdd(i,j-1,d+1)==k
           nom=nom*normpdf((xxd(i,j,d+1)),xxd(i,j-1,d+1)+...
                (alp0(k,d+1)+(alp1(k,d+1)*...
           xxd(i,j-1,d+1))+(alp2(k,d+1)*((xxd(i,j-1,d+1))^2))+...
           alp3(k,d+1)/xxd(i,j-1,d+1))...
           *(td1(i,j)-td1(i,j-1)),sig(1,d+1)*...
           ((xxd(i,j-1,d+1))^(2*etanew(k,1)))*...
           (td1(i,j)-td1(i,j-1)));
           den=den*normpdf(xxd(i,j,d+1),xxd(i,j-1,d+1)+...
                (alp0(k,d+1)+(alp1(k,d+1)*...
           xxd(i,j-1,d+1))+(alp2(k,d+1)*(xxd(i,j-1,d+1))^2)+...
           (alp3(k,d+1)/xxd(i,j-1,d+1)))...
           *(td1(i,j)-td1(i,j-1)),sig(1,d+1)*...
           ((xxd(i,j-1,d+1)^(2*etaold(k,1))))*...
           (td1(i,j)-td1(i,j-1)));
          end
      end
    end
    nom=nom*priornew*etanew(k,1);
    den=den*priorold*etaold(k,1);
    mh=min(1,nom/den);u=rand();
    if u<mh
       eta(k,d+1)=etanew(k,1);
    else
        eta(k,d+1)=eta(k,d);
    end
end
end
%Our MCMC iterations end here
```

```
%Example how to compute the posterior mean of some quatities
gma12=squeeze(gma(1,2,1:d));
gma13=squeeze(gma(1,3,1:d));
gma21=squeeze(gma(2,1,1:d));
gma23=squeeze(gma(2,3,1:d));
gma31=squeeze(gma(3,1,1:d));
gma32=squeeze(gma(3,2,1:d));
lmda12=squeeze(lmda(1,2,1:d));
lmda13=squeeze(lmda(1,3,1:d));
lmda21=squeeze(lmda(2,1,1:d));
lmda23=squeeze(lmda(2,3,1:d));
lmda31=squeeze(lmda(3,1,1:d));
lmda32=squeeze(lmda(3,2,1:d));
mgma12=mean(gma12);
mgma13=mean(gma13);mgma21=mean(gma21);
mgma23=mean(gma23);
mgma31=mean(gma31);mgma32=mean(gma32);
mlmda12=mean(lmda12);mlmda13=mean(lmda13);
mlmda21=mean(lmda21);mlmda23=mean(lmda23);
mlmda31=mean(lmda31);mlmda32=mean(lmda32);
msig=mean(sig(1,1:d));
alp01=mean(alp0(1,1:d));
alp02=mean(alp0(2,1:d));alp03=mean(alp0(3,1:d));
alp11=mean(alp1(1,1:d));alp12=mean(alp1(2,1:d));
alp13=mean(alp1(3,1:d));
alp21=mean(alp2(1,1:d));alp22=mean(alp2(2,1:d));
alp23=mean(alp2(3,1:d));
alp31=mean(alp3(1,1:d));alp32=mean(alp3(2,1:d));
alp33=mean(alp3(3,1:d));

save hsdsimulation;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Some of the function used in this program
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function r = drchrnd1(a)
%this is a one sample draw from dirchlet random
p=length(a);a2=zeros(1,p);r=zeros(1,p);
a2(1)=sum(a(2:p));
r(1)=betarnd(a(1),a2(1),1,1);
for j=2:p-1
    a2(j)=sum(a(j+1:p));
    r(j)=betarnd(a(j),a2(j),1,1)*(1-sum(r(1:j-1)));
end
r(p)=1-sum(r(1:p-1));
```

```
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function p = invgampdf(x,a,b )
%the computation of inverse gamma pdf from gamma pdf
p=b^a/gamma(a).*(1./x).^(a+1).*exp(-b./x);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function qadjust = normalizeq(q,k)
a=size(q,1);sum=0;
for l=1:a
    if l~=k
       sum=sum+q(k,l);
    end
end
qadjust=q;qadjust(k,k)=-sum;

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function obsfzero = obsdensityfzero(k,i,xxd,alp0,sig,d)
%this is the degenerate observation density function
%for ind i at the time ti1
obsfzero=normpdf(xxd(i,1,d),alp0(k,d),sig(1,d));
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    observationdensityf(k,i,j,xxd,alp0,alp1,...
        alp2,alp3,eta,sig,td1,d)
%this is the observation density function
%f(xxd(i,j)xxd(i,j-1),k) state k
obsf=normpdf(xxd(i,j,d),xxd(i,j-1,d)+(alp0(k,d)+...
    (alp1(k,d)*xxd(i,j-1,d))+...
    (alp2(k,d)*(xxd(i,j-1,d))^2)+(alp3(k,d)/...
    xxd(i,j-1,d)))*(td1(i,j)-td1(i,j-1)),sig(1,d)*...
    ((xxd(i,j-1,d)^(2*eta(k,d))))...
    *(td1(i,j)-td1(i,j-1)));
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function x=samplefromp(p,n)
%Inputs - p is the probability vector of length k
% - n is the number of random
% integers from 1,2, ...,k returned
%Output - a row vector of length n with entries
% from the set {1, 2, ..., k} with
% probabilities specified by p.
k=size(p,2);
```

```
u=rand(1,n);
x=zeros(1,n);
for i=1:k
x=x+i*(sum(p(1:i))-p(i) <=u & u<sum(p(1:i)));
end
 end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Bibliography

[1] Adesi, G.B. , Rasmussen, H., and Ravanelli, C. (2005) . An option pricing formula for the GARCH diffusion model. *Computational Statistics & Data Analysis*. **49(2)**: 287-310.

[2] Ailliot, P. , and Monbet, V. (2012) . Markov-Switching Autoregressive Models for Wind Time Series. *Environmental Modelling & Software*, **30**: 92–101.

[3] Ait-Sahalia, Y. (1996) . Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*. **9**: 385–426.

[4] Albert, J. and Chib, S. (1993) . Bayes inference via gibbs sampling of autogressive times series subject to markov mean and variance shifts. *Journal of Business and Economic Statistics*, **11**: 1-15.

[5] Almagro, P. , Martinez-Camblor, P. , Soriano, J.B. , et al. (2012) . Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study. *PLoS One* **9(2): e89866**.

[6] Al-Mohy, A.H. , and Higham, N.J. (2009) . A New Scaling and Squaring Algorithm for the Matrix Exponential. *SIAM J. Matrix Anal. Appl* **31(3)**: 970–989.

[7] Bachar, M. , Batzel, J.J. , and Ditlevsen, S. (2013) *Stochastic Biomathematical Models with Applications to Neuronal Modeling* . Springer-Verlag Berlin Heidelberg. DOI 10.1007/978-3-642-32157-3 1.

[8] Baio, G. (2012) . *Bayesian methods in health economics* (Boca Raton, FL , Chapman & Hall/CRC biostatistics series, 53).

[9] Banks, H.T. , Broido, A., Canter, B. , Gayvert, K. , Hu, S. , Joyner, M. , and Link, K. (2012) . Simulation algorithms for continuous time Markov chain models. *Simulation and Modeling Related to Computational Science and Robotics Technology* **37**: 3-18.

[10] Bartolomeo, N. , Trerotoli, P. , and Serio, G. (2011) . Progression of liver cirrhosis to HCC: an application of hidden Markov model. *BMC Med. Res. Methodol.*: 11-38.

[11] Bellone, B. (2005) . Classical Estimation of Multivariate Markov-Switching Models using MSVARlib. *Econometrics* **0508017**, EconWPA.

[12] Benmiloud, B. , and Piczunski, W. (1995) . Estimation des parametres dans les chaines de markov cachees et segmentations. Traitement du signal. *Traitement du signal* **, 12(5)**: 433-454.

[13] Berman, S.M. (1994) . A bivariate markov process with diffusion and discrete components. *Communications in Statistics. Stochastic Models* **10(2)**: 271-308. DOI:10.1080/15326349408807297.

[14] Blackwell, P.G. (1997) . Random diffusion models for animal movement. *Ecological Modelling*. **100** 87-102.

[15] Blackwell, P.G. (2003) . Bayesian Inference for Markov Processes with Diffusion and Discrete Components. *Biometrika*. **90(3)** 613-27.

[16] Blackwell, P.G., Niu, M., Lambert, M.S. and LaPoint, S.D. (2016) . Exact Bayesian inference for animal movement in continuous time. *Methods Ecol Evol*. **7**: 184–195. doi:10.1111/2041-210X.12460.

[17] Bollerslev, T. , Chou, R.Y. , and Kroner, K.F. (1992) . ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**(1-2): 5–59.

[18] Bolstad, W. (2007) . *Introduction to Bayesian statistics*. John Wiley, 2nd edition.

[19] Boys, R.J. , and Handerson, D.A. (2004) . A bayesian appraoch to DNA sequence segmentation(with discussion). *Biometrics* **60**: 573-88.

[20] Brooks, S. (1998) . Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47(1)**: 69-100. DOI: 10.1111/1467-9884.00117.

[21] Brooks, S.P. , and Roberts, G. O. (1999) . Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing* **8(4)**: 319-35.

[22] Brooks, S. , Gelman, A. , Jones, G.L. , and Meng, X. (2011) . *Handbook of Markov chain Monte Carlo*. CRC press.

[23] Bu, R. , Giet, L., Hadri, K., and Lubrano, M. (2016). Reducible Diffusions with Time-Varying Transformations with Application to Short-Term Interest Rates. *Economic Modelling* **52**: 266–277.

[24] Bureau, A., Hughes, J.P., and Shiboski, S.C. (2000) . An S-PLUS Implementation of Hidden Markov Models in Continuous Time. *Journal of Computational and Graphical Statistics*, **9**: 621-632.

[25] Cappe, O. , Moulines, E. , and Ryden, T. (2005) . *Inference in Hidden Markov Models*. Springer-Verlag, New York.

[26] Celoux, G. (1998) . Bayesian inference for mixture: The label switching problem. In Payne, R. and Green, P., editors, *Proceedings in Computational Statistics*, **pages 227-232**. Physica, Heidelberg.

[27] Chang, Y. , Choi, Y. , and Park, J. (2015) "Regime Switching Model with Endogenous Autoregressive Latent Factor," Manuscript, Indiana University.

[28] Chantel, G.J. , Richardson, S. , Longini, I.M. , and Brandslund, I. (2000) . Modeling disease progression by a hidden markov process: Application to characterizing CD4 cell decline. *Biometrics* **56**: 733-41.

[29] Chauvet, M. , and Piger, J. (2008) . A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* **26**: 42-49.

[30] Chen, K.C. , Wang, T.Y. , Tseng, H.H. , Huang, C.Y., and Kao, C.Y.(2005) A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics.* **21(12)**: 2883-90.

[31] Chib, S. (1996) . Calculating posterior distributions and model estimates in Markov Mixtures models. *Journal of Econometrics* **75**: 79-98.

[32] Chib, S., and Jeliazkov, I. (2005) . Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. *Statistica Neerlandica.* **59**: 30–44.

[33] Chib S., and Shephard, N. (2002) . [Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes]: **comment**. *J Bus Econom Stat* **20(3)**: 325–27.

[34] Choi, S. (2009) Regime-Switching Univariate Diffusion Models of the Short-Term Interest Rate. *Studies in Nonlinear Dynamics and Econometrics* **13(1)**. ISSN (Online) 1558-3708. DOI: https://doi.org/10.2202/1558-3708.1614.

[35] Chou, R.Y. (1988) . Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics.* **3(4)**: 279–294.

[36] Chunling, C. , and Tsokos, C.P. (2009) . Markov Modeling of Breast Cancer. *Journal of Modern Applied Statistical Methods.* **8(2)**: 626-631.

[37] Congdon, P. (2006) . *Bayesian statistical modelling* (John Wiley & Sons Wiley series in probability and statistics, 2nd ed). ISBN: 0470018755,9780470018750,9780470035931.

[38] Craiu, R.V. , and Rosenthal, J.S. (2013) . Bayesian Computation Via Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application* **1(1)**: 179-201.

[39] Davies, E.B. (2010) . Embeddable Markov Matrices. *Electronic Journal of Probability* **15(47)**: 1474-1486.

[40] Demirhana, H. , and Kalaylioglu, Z. (2015) . Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models. *Journal of Multivariate Analysis.* **135**: 163-174.

[41] Donnelly, T. , Al-Khater, A.H. , Al-Kuwari, M. , et al. (2011) . Study exploring breast cancer screening practices amongst Arabic women living in the State of Qatar. *Avicenna.* **1**. doi: 10.5339/avi.2011.1.

[42] Dorit, L. , Dorothea, N. , Volker, H., Michael, U. , Steffen, K. , Ingo B. , and Petra, S.(2007) . Importance of CEA and CA 15-3 during Disease Progression in Metastatic Breast Cancer Patients. *Anticancer research* **27**: 1963–1968.

[43] Doucet, A., Logothetis, A. , and Krishnamurthy, V. (2000) . Stochastic sampling algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Automa. Control* **45**: 188-202.

[44] Dunn, J.E. , and Gipson, P.S. (1977) . Analysis of Radio Telemetry Data in Studies of Home Range. *Biometrics.* **33(1)**: 85-101.

[45] Durham, G.B., and Gallant, A.R. (2002) . Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Stat.* **20**: 297–338.

[46] Elerian, O. , Chib, S. , and Shephard, N. (2001) . Likelihood inference for discretely observed non-linear diffusion. *Econometrica* **69(4)**: 959-993

[47] **EL-Maroufy, H. , Hibbah, E.H. , Zyad, A. , and Ziad, T. (2017) . Bayesian Estimation of Multivariate Autoregressive Hidden Markov Model with Application to Breast Cancer Biomarker Modeling, In *Bayesian Inference, Javier Prieto Tejedor Ed, IntechOpen.* Available from: https://www.intechopen.com/books/bayesian-inference/bayesian-estimation-of-multivariate-autoregressive-hidden-markov-model-with-application-to-breast-ca.**

[48] El-Rhazi, K. , Nejjari, C. , BenJelloun, M.C. , El-Biaze, M. , Attassi, M. , and Garcia-Larsen, V. (2016). Prevalence of chronic obstructive pulmonary disease in Fez, Morocco: results from the BOLD study. *The International Journal of Tuberculosis and Lung Disease* **20(1)**: 136–141.

[49] Eraker, B. (2001) . MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics* **19**: 177–91.

[50] Farcomeni, A. , and Arima, S. (2013) . A bayesian autoregressive three state hidden markov model for identifying switching monotonic regimes in microarray time course data. *Statistical Application in Genetics and Molecular Biology* **11(3)**.

[51] Fink, G.A. (2008) *Markov Models for Pattern Recognition.* Springer-Verlag, Berlin.

[52] Fitzpatrick, M. , and Marchev, D. (2013) . Efficient bayesian estimation of the multivariate double chain markov model. *Statistics and Computing* **23(4)**: 467-80.

[53] Freedman, D. (1983). *Markov Chains*. Springer-Verlag, New York.

[54] Fruhwirth-Schnatter, S. (2001a) . Markov chain monte carlo estimation of classical and dynamic switching and mixture models.*Journal of the American Statistical Association*, **96**: 194-209.

[55] Fruhwirth-Schnatter, S. (2006) . *Finite Mixture and Markov Switching Models*. Springer Series in Statistics.New York:NY.

[56] Fuchs, C. (2013) . *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer-Verlag, Berlin Heidelberg.

[57] Gelfand, A.E. , and Smith, A.F.M. (1990) . Sampling based approaches to calculating marginal densities. *J. American Statist. Assoc.* **85**: 398–409.

[58] Gelman, A. (1995) . Inference and monitoring convergence. In Gilks, W. , R.S. , and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, **pp. 131**. Chapman and Hall/CRC, London.

[59] Gelman, A., and Shalizi, C.R. (2013) . Philosophy and the practice of bayesian statistics in the social sciences. *British Journal of Mathematical and Statistical Psychology*, **66**: 8-38.

[60] Ghosh, M.K. , Arapostathis, A. , and Marcus, S.I. (1993) . Optimal control of switching diffusions with application to flexible manufacturing systems, *SIAM J. Control Optim.*. **31**: 1183–1204.

[61] Gilks, W. , Richardson, S. , and Spiegelhalter, D.(1996) . "Introducing Markov chain Monte Carlo." In Markov Chain Monte Carlo in Practice, Eds. Gilks. Richardson, and Spiegelhalter. 59-74. London, Chapman& Hall.

[62] Golightly, A. , and Wilkinson, D.J. (2006) . Bayesian sequential inference for stochastic kinetic biochemical network models. *J. Comput. Biol.***13(3)**: 838-51.

[63] Golightly, A. , and Wilkinson, D.J. (2008) . Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis* **52(3)**: 1674-93.

[64] Goossens, N. , Nakagawa, S. , Sun, X. , and Hoshida, Y. (2015) . Cancer biomarker discovery and validation. *Translational Cancer Research,* **4(3)**: 256–269. http://doi.org/10.3978/j.issn.2218-676X.2015.06.04.

[65] Goutte, S. , Ismail, A. and Pham, H. (2017) . Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance* **24(1)**: 38-75.

[66] Guihenneuc-Jouyaux, C. , Richardson, S. , and Longini, IM. Jr. (2000). Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. *Biometrics*. **56(3)**: 733-41.

[67] Hamilton, J. (1989) . A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**: 357-384.

[68] Hastings, W.K. (1970) . Monte Carlo sampling methods using Markov chain and their applications. *Biometrika* **(57)**: 97-109.

[69] Hering, A.S. , Kazor, K. and Kleiber, W. (2015) . A Markov-Switching Vector Autoregressive Stochastic Wind Generator for Multiple Spatial and Temporal Scales. *Resources* **4(1)**: 70-92.

[70] **Hibbah, E.H. , El-Maroufy, H. , Fuchs, C. , and Ziad, T. (2019). An MCMC Computational Approach for A Continuous Time State-dependent Regime Switching Diffusion Process.** ***The Journal of Applied Statistics (DOI: 10.1080/02664763.2019.1677573).***

[71] **Hibbah, E.H. , and El-Maroufy, H. (2019) . Bayesian Continuous Time Markov Switching Autoregressive Process in Disease Progression.** ***Pakistan Journal of Statistics and Operation Research (In Review).***

[72] Hobert, J.P. (2011) . The data augmentation algorithm: theory and methodology. In Steve, B., A. G. G. L. J. and Xiao-Li, M., editors, *The Handbook of Markov Chain Monte Carlo*, **page 253**. Chapman and Hall/CRC.

[73] Hurn, M. , J.A. , and Rober, C. (2003) . Estimating mixtures of regressions. journal of computational and graphical statistics. *Journal of the American Statistical Association*, **79**: 55-79.

[74] Inamura, Y. (2006) . Estimating Continuous Time Transition Matrices From Discretely Observed Data. *Bank of Japan Working Paper Series.*

[75] Israel, R.B. , Rosenthal, J.S , and Wei, J.Z (2000) . Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings. *Mathematical Finance* **11(2)**: 245-265.

[76] Janczura, J. , and Weron, R. (2011) .Efficient estimation of Markov regime-switching models: An application to electricity wholesale market prices. *AStA Adv. Stat. Anal.* **96**: 385-407.

[77] Jasra, A., Holmes, C.C. , and Stephens, D.A. (2005) . Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling. *Statistical Science* **20(1)**: 50-67.

[78] Johannes, M. , Polson, N. (2006) "MCMC methods for financial econometrics." In Handbook of Financial Econometrics Eds. Ait-Sahalia, Y. , Hansen, L.P. Elsevier.

[79] Kantas, N. , Doucet, A. , Singh, S.S. , Maciejowski, J. , and Chopin, N. (2015) . On Particle Methods for Parameter Estimation in State-Space Models. *Statistical Science.* **30(3)**: 328-351, DOI: 10.1214/14-STS511.

[80] Kass, R.E. , Bradley, P. , Carlin, B.P. , Gelman, A. , and Neal, R.M. (1998) . Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician* **52(2)**: 93-100.

[81] Kim, C. , and Kim, J. (2015) . Bayesian inference in regime-switching ARMA models with absorbing states: The dynamics of the ex-ante real interest rate under regime shifts. *Journal of Business and Economic Statistics* **33**: 566-78.

[82] Kim, C.J. , and Kim, J. (2018) . *Non-Markovian Regime-Switching Models* . Available at SSRN: https://ssrn.com/abstract=3159530 or http://dx.doi.org/10.2139/ssrn.3159530.

[83] Kim, C. , and Nelson, C. (1999) . *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press.

[84] Kirby, A.J. ,a nd Spiegelhalter, D.J. (1994) . Statistical modelling for the precursors of cervical cancer. *In Case Studies in Biometry (eds N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse).* New York: Wiley.

[85] Kloeden, P.E. , and Platen, E. (1992) . *Numerical solution of stochastic differential equations.* Applications of Mathematics (New York) 23, Springer-Verlag, Berlin, doi:10.1007/978-3-662-12616-5.

[86] Koralov, L. , and Sinai, Y.G. (2007) . *Theory of probability and random processes.* Springer Berlin Heidelberg New York.

[87] Korolkiewickz, M. and Elliot, J. (2008) . A hidden markov model of credit quality. *Journal of Economic Dynamics and Control*, **32**: 3807-3819.

[88] Krozlig, H.M. (1997) . *Markov-Switching Vector Autoregressions.* Springer-Verlag Berlin: Heidelberg.

[89] Krystul, J. and Blom, H.A.P. (2004) . *Monte Carlo simulation of rare events in hybrid systems.* Technical Report WP8, Deliverable D8.3. HYBRIDGE.

[90] Laessig, D. , Nagel, D. , Heinemann, V. , Untch, M. , Kahlert, S. , Bauerfeind, I. , and Stieber, P. (2007) . Importance of CEA and CA 15-3 during disease progression in metastatic breast cancer patients. *Anticancer Res.* **27(4A)**: 1963-8.

[91] Lee, Y. , Ghosh, D. , Hardison, R. , and Zhang, R.H. (2014). MRHMMs: Multivariate Regression Hidden Markov Models and the variants. *Bioinformatics.* **30(12)**: 1755-1756. https://doi.org/10.1093/bioinformatics/btu070.

[92] Li, Y.P. , and Chan, W. (2006) . Analysis of longitudinal multinomial outcome data. *Biometrical J.* **48(2)**: 319-326.

[93] Li, X. , Jiang, D. , and Mao, X. (2009) . Population dynamical behavior of Lotka-Volterra system under regime switching. *J. Comput. Appl. Math.*. **232**: 427–448.

[94] Lin, M.Y. (2013) . *Bayesian statistics* (School of Public Health, Boston University, Boston, MA, Tech. Rep.

[95] Lindley, D. (2000) . The philosophy of statistics. *The Statistician*, **49**: 293-337.

[96] Liu, H. , Chen, J., Ouyang, Q., and Su, H. (2016) . A Review on Prognostics of Proton Exchange Membrane Fuel Cells. *Vehicle Power and Propulsion Conference (VPPC) 2016 IEEE* **1**: 1–6.

[97] Livingstone, S. , and Girolami, M. (2014) . Information-Geometric Markov Chain Monte Carlo Methods using Diffusions. *Entropy* **16(6)**: 3074-3102. DOI: 10.3390/e16063074.

[98] Lozano, R. , Naghavi, M. , Foreman, K. , *et al.* (2012) . Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**: 2095–2128.

[99] Lu, H.M. , Zeng, D. , and Chen, H. (2012) . Prospective Infectious Disease Outbreak Detection Using Markov Switching Models. *IEEE Transactions on Knowledge and Data Engineering* **22(4)**: 565-577.

[100] Ma, J. (2013) . *A Bayesian approach to longitudinal categorical data in a continuous time Markov chain model. Ph.D. Thesis.* The University of Texas School of Public Health, Houston, TX.

[101] McPherson, K. , Steel, C.M. , and Dixon, J.M. (2000) . ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *B.M.J.* **321(7261)**: 624-8.

[102] Metropolis, N. , Rosenbluth, A. , Rosenbluth, M. , Teller, A. , and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21(6)**: 1087–92.

[103] Mitra , R. , and Gupta, M. (2011) . A continuous-index Bayesian hidden Markov model for prediction of nucleosome positioning in genomic DNA. *Biostatistics* **12(3)**: 462-77.

[104] Nelson, D.B. (1990) . ARCH Models as Diffusion Approximations. *Journal of Econometrics* **45**: 7–38.

[105] Newton, P.K. , Mason, J. , Venkatappa, N. , Jochelson, M.S. , Hurt, B. , Nieva, G. , Comen, E. , Norton, L. and Kuhn, P. (2015). Spatiotemporal Progression of Metastatic Breast Cancer: A Markov chain model highlighting the role of early metastatic sites. *npj Breast Cancer* **1** art.num. 15018.

[106] Ogawa, S. (1992) . Monte Carlo Simulation of Non-linear Diffusion Processes. *Japan J. Indust. Appl. Math.* **9**: 25-33.

[107] Pagan, A. , and Schwert, G. (1990) . Alternative models for conditional stock volatility. *Journal of econometrics* **45**: 267-90.

[108] Parton, A. , and Blackwell, P.G. (2016) . Bayesian Inference for Multistate 'Step and Turn' Animal Movement in Continuous Time. *Journal of Agricultural, Biological, and Environmental Statistics* **22(3)**: 373-392.

[109] Pedersen, A.C. , Sorensen, P.D. , Jacobsen, E.H. , Madsen, J.S. , and Brandslund, I. (2013) . Sensitivity of CA 15-3, CEA and serum HER2 in the early detection of recurrence of breast cancer. *Clin. Chem. Lab. Med.* **51(7)**: 1511-9. doi: 10.1515/cclm-2012-0488.

[110] Peng, N.F. (1996) . Spectral representations of the transition probability matrices for continuous time finite Markov chains. *J. Appl. Prob* **33**: 28-33.

[111] Rao, V. , and Teh, Y.W. (2013) . Fast MCMC Sampling for Markov Jump Processes and Extensions. *Journal of Machine Learning Research* **14**: 3295-3320.

[112] Rausand, M. , and Hoyland, A. (2004) *System Reliability Theory Models, Statistical Methods, and Applications* (Hoboken, New Jersey, John Wiley & Sons, 2nd ed).

[113] Robert, C. (2006). *Le choix bayésien: Principes et pratique*. springer verlag France, Paris.

[114] Robert, C. , and Casella, G. (2004) *Monte Carlo Statistical Methods* (New York, NY, Springer/Springer Texts in Statistics, 2nd ed), ISBN: 0387212396,9780387212395.

[115] Robert, C. , and Casella, G. (2011) . A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science* **26**: 102–15.

[116] Roberts, G.O. , and Rosenthal, J.S. (2004) . General state space Markov chains and MCMC algorithms. *Probability Surveys* **1**: 20-71. ISSN: 1549-5787, DOI: 10.1214/154957804100000024.

[117] Roberts, G.O. , and Smith, A.F.M. (1993) . Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society Series B (Methodological)* **55(1)**: 3–23. doi: 10.2307/2346063.

[118] Roberts, G.O. and Stramer, O. (2001) . On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika*. **88(3)**: 603-621.

[119] Rubio-Ramirez, J. , Waggoner, D. , and Zha, T. (2010) . Markov-switching structural vector autoregressions: theory and application. Review of Economic Studies. *Journal of econometrics* **77**: 665–96.

[120] Samaniego, F.J. *A comparison of the Bayesian and frequentist approaches to estimation* (Springer-Verlag New York/Springer series in statistics).

[121] Sampietro, S. (2005) . Mixture of autoregressive components for modeling financial market volatility. *LIUC papers. Serie Metodi quantitativi*, **16**.

[122] Samy, N., Ragab, H.M. , El-Maksoud, N.A. , and Shaalan, M. (2010) . Prognostic significance of serum Her2/neu, BCL2, CA15-3 and CEA in breast cancer patients: a short follow-up. *Cancer Biomark.* **6(2)**: 63-72. doi: 10.3233/CBM-2009-0119.

[123] Shaker, S.B. , Dirksen, A. , Lo, P. , Skovgaard, L.T. , de-Bruijne, M. , and Pedersen J.H. (2012) . Factors influencing decline in lung density in a Danish lung cancer screening cohort. *Eur Respir J.*. **40(5)** 1142-8. doi: 10.1183/09031936.00207911.

[124] Shao, J. , "Criteria for transience and recurrence of regime-switching diffusion processes," Electron. J. Probab. **20(63):1–15** , ISSN: 1083-6489 doi: 10.1214/EJP.v20-4018.

[125] Sharma, S. (2017) . Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annual Review of Astronomy and Astrophysics*, **55**: 213-59.

[126] Shigeru, I. , Takashi, K. , and Takahiro, H. (1999) . Serum c-erbB-2 Levels in Monitoring of Operable Breast Cancer Patients. *Jpn J Clin Oncol.* **29(7)**: 336-339.

[127] Siekmann, I. , Wagner, L.E.II , Yule, D. , Fox, C. , Bryant, D. , Crampin, E.J. , and Sneyd, J. (2011). MCMC estimation of Markov models for ion channels. *Biophys. J.* **100**: 1919-1929.

[128] Sims, C.A. , and Zha, T. (2006) . Were there regime switches in us monetary policy? *The American Economic Review*: 54-81.

[129] Sinharay, S. (2014) . *Assessing convergence of the Markov chain Monte Carlo algorithms: A review* . CRC press. (ETS RR-03-07). Princeton, NJ: Educational Testing Service (2003a).

[130] Slaoui, M. , Razine, R. , Ibrahimi, A. , Attaleb, M. , Mzibri, M.E. , and Amrani, M. (2014) . Breast cancer in Morocco: a literature review. *Asian Pac. J. Cancer Prev.* **15(3)**: 1067-74.

[131] Sorwar, G. (2011) . Estimating single factor jump diffusion interest rate models. *Applied Financial Economics*. **21(22)**: 1679-1689, DOI: 10.1080/09603107.2011.591729 .

[132] Sperrin, M., Jaki, M.L. , Wit, E. (2010) . *Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. Stat Comput* **20**: 357–366. DOI 10.1007/s11222-009-9129-8.

[133] Stockley, R.A. (2014) . Biomarkers in chronic obstructive pulmonary disease: confusing or useful?. *International Journal of COPD*. **4:9**: 163-177.

[134] Tanner, M.A. , and Wong, W.H. (1987) . The calculation of posterior distributions by data augmentation (with discussion). *T. J. Amer. Statist. Assoc.* **82**: 528-50.

[135] Tierney, L. (1994) . Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**: 1701–86.

[136] Tierney, L.(1996) . "Introduction to general state-space Markov chain theory." In Markov Chain Monte Carlo in Practice, Eds. Gilks. Richardson, and Spiegelhalter. 59-74. London, Chapman& Hall.

[137] Tjørve, K.M.C. and Tjørve, E.T. (2017) . The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS ONE*. **12(6)**: e0178691. https://doi.org/10.1371/journal.pone.017869.

[138] Tse, Y.K. , Yu, J. , and Chang, X.B. (2002) . Estimation of Hyperbolic Diffusion Using Mcmc Method. *Research Collection.* School of Economics. Available at: http://ink.library.smu.edu.sg/soe_research_all/5.

[139] Tsionas, E., and Tassiopoulos, A. (2016) . Bayesian implications of ridge regression and Zellner's g-prior (Tech. Rep.). Athens: Athens University of Economics and Business - Department of Economics. SSRN: http://ssrn.com/abstract=238758.

[140] Tuyl, F. , Gerlach, R. , and Mengersen, K. (2013) . Posterior predictive arguments in favor of the bayes-laplace priors as the consensus prior for the binomial and multinomial parameters. *Bayesian analysis* **4**: 151-158.

[141] Van-Den-Hout, A. , Fox, J.P. , and Klein-Entink, R.H. (2015) . Bayesian inference for an illness-death model for stroke with cognition as a latent time-dependent risk factor. *Statistical Methods in Medical Research* **24(6)**: 769-87. DOI: 10.1177/0962280211426359.

[142] Walter, P.C. , Rainer, N. , Allan, L. , Kim , L. , Suhail, A. , and Christopher, P.P. (2003) . Potential Clinical Utility of Serum HER-2/neu Oncoprotein Concentrations in Patients with Breast Cancer. *Clinical Chemistry* **49(10)**: 1579–1598.

[143] Woodard, D.B. , Schmidler, S.C. , and Huber, M. (2009) . Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *The Annals of Applied Probability* **19(2)**: 617-640, DOI: 10.1214/08-AAP555.

[144] Yin, G. , Mao, X. , Yuan, C. , and Cao, D. (2010) . Approximation methods for hybrid diffusion systems with state-dependent switching processes : numerical algorithms and existence and uniqueness of solutions. *SIAM Journal on Mathematical Analysis* **41 (6)**: 2335-2352.

[145] Yin, G. , and Zhu, C. (2010). Properties of solutions of stochastic differential equations with continuous state-dependent switching. *J. Differential Eqs.* **249**: 2409–2439.

[146] Young, A.L. , Oxtoby, N.P. , Ourselin, S. , Schott J.M. , and Alexander, D.C. (2015) . A simulation system for biomarker evolution in neurodegenerative disease. *Med Image Anal.* **26(1)**: 47-56.doi: 10.1016/j.media.2015.07.004.

[147] Zheng, Z. (2003) . On swapping and simulated tempering algorithms. *Stochastic Processes and their Applications* **104(1)**: 131-154.

[148] Zeng, Y. and Frias, J. (2006) . A novel hmm-based clustering algorithm for the analysis of gene expression time-course data. *Computational Statistics and Data Analysis* **50**: 2472-2494.

[149] Zucchini, W. and MacDonalds, I. (2009) . *Hidden Markov models for times series:An introduction using R.* Springer, New York.

Résumé: Cette thèse a pour objectif de fournir un cadre statistique pour le suivi et la surveillance de la progression de la pathologie. En fait, durant son développement, le stade de la maladie est élucidé et invisible sauf par le biais d'autres facteurs observés tels que les marqueurs biologiques. Nous faisons appel alors aux modèles de Markov latents pour modéliser sa progression, en particulier les modèles autorégressifs avec changements de régimes Markoviens dans les deux cas: temps discret et continu. Plus tard, et dans le but d'avoir un modèle avec plus de dynamique, de hasard et de fluctuations stochastiques; et pour posséder un modèle capable de résoudre le problème des observations dispersées; nous adoptons le processus stochastique de diffusion hybride.

Par conséquent, nous allons étape par étape pour simuler ces différents modèles dans le cadre Bayésien à travers les méthodes MCMC; très convenable à estimer ces modèles dont la fonction de vraisemblance n'est connue que proportionnellement et ayant un nombre important de paramètres, plutôt que d'utiliser les approches classiques telles que l'approche du maximum de vraisemblance. En conséquence, l'interprétation, la modélisation et la simulation des paramètres pour les modèles latents à l'aide des méthodes MCMC seront rendues plus simples.
Mots clés: Progression de la maladie, Processus de diffusion hybride, Modèles avec changement de régimes Markoviens, Méthodes MCMC, Estimation conjointe des états cachés, Imputation temporelle aléatoire de données.

Abstract: This thesis aims to provide a statistical framework for the follow up and for monitoring the disease progression. In fact, one of the difficulties faced when analyzing disease development is that at any time the disease stage is unraveled and unseen; except through other factors that are observed and can inform us about the disease such as biomarkers. Thus, we call for latent Markov models to characterize the disease progression; especially the Markov switching autoregressive model in both the discrete and the continuous time cases. Later on, and to have a model with more dynamics, randomness and stochastic fluctuations; and by the way to own a model that can overcome the problem of sparsely dispersed observations, we adopt the stochastic hybrid switching diffusion process.

Hence, we go step by step to simulate these different models using a Bayesian approach very acquainted to estimate complex models with a large number of parameters and where the likelihood can be known only up to a constant; rather than using classical approaches like the weighted least squares technics, or the method of moments and others. Consequently, the parameters'interpretation, modelling and simulation for latent models through Bayesian methods or their analogue the Markov Chain Monte Carlo approach will be rendered comprehensible.
Key words: Disease progression, Hybrid switching diffusion process, Markov switching model, MCMC methods, Block update for the hidden states, Random time data imputation.