



**Université Sultan Moulay Slimane
Faculté des Sciences et Techniques**

Centre d'Études Doctorales en Mathématiques et Physique Appliquées

THÈSE DE DOCTORAT

**INFÉRENCE BAYÉSIENNE POUR LES MODÈLES ÉPIDÉMIQUES DE
TYPE SIR**

Soutenue et présentée par
ABDELAZIZ QAFFOU

Le 21 Juillet 2017

Structure de recherche : Laboratoire de Mathématiques et Applications
Option : STATISTIQUE ET PROBABILITÉS

devant le jury composé de :

M. Ahmed ZEGHAL FST, Université Sultan My Slimane de Beni Mellal	Professeur	Président
M. Taib ZIAD Université de Technologie Chalmers, Suède	Professeur	Rapporteur
M. Abdelkrim MERBOUHA FST, Université Sultan My Slimane de Beni Mellal	Professeur	Rapporteur
M. Khalid HILAL FST, Université Sultan My Slimane de Beni Mellal	Professeur	Rapporteur
M. Tewfik KERNANE Université des Sciences et de la Technologie USTHB, Algérie	Professeur	Examineur
M. Cherki DAOUI FST, Université Sultan My Slimane de Beni Mellal	Professeur	Examineur
M. Hamid EL MAROUFY FST, Université Sultan My Slimane de Beni Mellal	Professeur Habilité	Directeur de thèse



*A ma mère, mon père, ma femme, ma fille **Marwa**, ma soeur, mes frères,
ma famille et mes amis.*



Remerciements

Ce travail a été réalisé au laboratoire de Mathématique et Applications de la Faculté des Sciences et Techniques de l'Université Sultan Moulay Slimane, sous la direction du professeur Hamid EL MAROUFY qui a voulu assumer avec patience la direction de cette thèse et qui a été tout le temps disponible, généreux et surtout sympathique. Je ne peux qu'admirer ses compétences scientifiques et d'encadrement qu'il veuille bien trouver ici l'expression de ma profonde gratitude.

Mes vifs remerciements vont à l'endroit du professeur Bouchaib MERNARI Président de l'Université Sultan Moulay Slimane, Béni Mellal, pour son soutien à la recherche scientifique au sein des différentes structures de l'Université Sultan Moulay Slimane.

J'adresse mes remerciements au professeur Ahmed ZEGHAL Doyen de la Faculté des Sciences et Techniques, pour l'honneur qu'il a porté à ce travail en acceptant de présider le Jury.

Je suis redevable au professeur Taib ZIAD de l'Université de Technologie de Chalmers (Suède) pour l'intérêt qu'il a témoigné à mon travail et l'honneur qu'il me fait de rapporter et de siéger dans le Jury d'examen.

Je tiens à remercier vivement mes professeurs Abdelkrim MERBOUHA et Khalid HILAL pour avoir accepté de rapporter mon travail et faire partie du Jury de mon doctorat.

J'adresse mes remerciements à Messieurs les professeurs Tewfik KERNANE et Cherki DAOUI pour l'honneur qu'ils m'ont fait d'être membres du Jury.

Un grand merci à Monsieur Khalid CHAOUCH professeur à la faculté des lettres et des sciences humaines Béni-Mellal pour son appuis linguistique pour rédiger la version anglaise de cette thèse.

Mes sincères remerciements vont à Monsieur Elmokhtar ZBAIR ingénieur en chef statisticien au ministère de l'enseignement supérieur, pour son soutien moral tout au long de ce travail.

Je ne pourrai passer sous silence le soutien accordé par ma famille. Je remercie tout spécialement ma soeur et mes frères pour leurs encouragements et leurs aides efficaces. J'exprime une gratitude particulière à mes parents pour avoir bien voulu assumer la lourde tâche de m'élever et éduquer, je n'oublierai jamais leurs bienfaits. Que dieu les bénisse et les récompenne. Enfin, je tiens à remercier de tout mon cœur mon épouse qui a accepté mes longues absences, ses encouragements précieux ont donné leur fruit... .

“All epidemiology, conceived as it is with the variation of disease from time to time and from place to place, must be considered mathematically, however many variables are implicated, if it is to be considered scientifically at all.” (Sir Ronald Ross¹, [Ross 1911])

Avant-propos

C'est avec un grand intérêt que j'ai fait connaissance, il y a quelques années, d'une classe importante des modèles mathématiques qui avaient suscité beaucoup d'intérêt dans le milieu des épidémiologistes et des biomathématiciens. Il s'agit des modèles de type **SIR**, qui sont plus appropriés pour analyser, contrôler et prédire la dynamique des maladies infectieuses complexes au sein d'une population. Leur principale caractéristique est la suivante :

Une population fermée est subdivisée en trois classes d'individus : les susceptibles, les infectés et les éliminés. Chaque infecté reste infectieux pendant une période de temps aléatoire durant laquelle il transmet l'infection aux susceptibles qui dans ce cas deviennent infectés et infectants. Une fois que la période d'infection est terminée, celui-ci est éliminé de manière permanente du processus d'infection par immunisation ou par décès par exemple.

Le seuil de l'épidémie est l'un des concepts les plus importants dans l'étude des modèles de ce type. Bien que ce concept soit théorique, il nous permet d'étudier le comportement qualitatif d'une épidémie et de la contrôler d'une manière efficace. Ce seuil est décrit par le taux de reproduction de base $R_0 = \frac{\beta}{\mu}$ où β est le taux d'infection et μ celui d'élimination.

Le domaine d'investigation de cette dissertation est l'*inférence Bayésienne* de ces trois principaux paramètres précités, en utilisant la technique de *Monte-Carlo* par les chaînes de Markov **MCMC**.

Ce travail est rédigé sous forme d'une monographie :

On commence par une introduction générale qui place le travail dans son contexte historique et scientifique. De plus, notre travail comporte quatre autres chapitres et trois Annexes **A**, **B** et **C** réservées aux démonstrations les plus difficiles.

Dans le chapitre 2, nous présentons des résultats préliminaires, outils et quelques définitions utilisées dans ce mémoire de thèse. Dans le chapitre 3, nous

1. Ronald Ross (13 mai 1857 à Almora, Inde - le 16 septembre 1932 à Londres) était un médecin bactériologiste et entomologiste britannique de l'Armée des Indes britanniques qui parvint à montrer, le 20 août 1897, que la transmission du paludisme des oiseaux se fait par un moustique. En 1902, il a reçu le prix Nobel de physiologie ou médecine pour ses travaux sur la modélisation mathématique du paludisme.

nous intéressons à l'estimation Bayésienne des taux d'infection β et d'élimination μ , dans le cas où ils sont supposés indépendants, on explore leurs distributions a posteriori ainsi que celle des données manquantes, en utilisant la discrétisation de Milstein. Le chapitre suivant traite l'inférence Bayésienne dans le cas où les deux paramètres sont supposés dépendants de loi conjointe distribuée selon la loi gamma bivariée de Kibble introduite par Kibble [Kibble 1941]. Dans le chapitre 5, une inférence Bayésienne du taux de reproduction de base R_0 est considérée, la méthode MCMC est adoptée pour explorer sa distribution a posteriori et celle des données latentes.

Finalement, la thèse est achevée par une conclusion générale suivie d'une liste de bibliographie.

Table des matières

1	Introduction Générale	1
1.1	Généralités	1
1.2	Formulation mathématique du modèle épidémique SIR	3
1.2.1	Modèle déterministe de type SIR	4
1.2.2	Modèle stochastique de type SIR	4
1.3	Contribution	6
2	Préliminaires	11
2.1	Mouvement Brownien	12
2.2	Intégrale d'Itô	13
2.3	Equations différentielles stochastiques	14
2.3.1	Formule d'Itô	14
2.3.2	Existence et unicité des solutions de l'équation d'Itô	15
2.4	Schémas de discrétisation	16
2.4.1	Schéma d'Euler	16
2.4.2	Schéma de Milstein	17
2.4.3	A propos de convergence	18
2.5	Processus de diffusion	18
2.6	Equation de Fokker-Planck	20
2.7	Approximation d'un processus à saut par un processus de diffusion	20
2.8	Inférence Bayésienne	24
2.8.1	Théorème de Bayes	24
2.8.2	Grandes lignes de l'approche Bayésienne	25
2.8.3	Probabilités <i>a priori</i> et <i>a posteriori</i>	25
2.9	Méthode de Monte Carlo par chaînes de Markov	25
2.10	Les Algorithmes MCMC	27
2.10.1	Algorithme de Metropolis-Hastings	28
2.10.2	Échantillonneur de Gibbs	29
2.10.3	Algorithme d'Acceptation Rejet-Metropolis Hastings	30
2.11	Algorithme de Gillespie	31
2.12	Inférence Bayésienne pour les processus de diffusion stochastiques	32
3	Inférence Bayésienne pour le modèle épidémique stochastique SIR avec paramètres indépendants	35
3.1	Introduction	35
3.2	Diffusion hyperbolique	36
3.3	Inférence Bayésienne pour le modèle non linéaire SIR	38
3.4	Simulations MCMC	40
3.4.1	Simulation des données manquantes	40
3.4.2	La distribution a posteriori des paramètres β et μ	41

3.4.3	Algorithme	42
3.4.4	Simulations	42
3.4.5	Application aux données réelles	45
3.5	Conclusion	46
4	Inférence Bayésienne pour le modèle épidémique stochastique SIR avec paramètres dépendants	49
4.1	Introduction	49
4.2	Inférence Bayésienne pour β et μ	50
4.2.1	Distribution a priori des paramètres β et μ	52
4.2.2	Distribution a posteriori des paramètres β et μ	53
4.2.3	Algorithme	53
4.3	Simulation MCMC	54
4.4	Conclusion	55
5	Inférence Bayésienne du nombre de reproduction de base du modèle épidémique SIR	59
5.1	Introduction	59
5.2	Inférence classique de R_0	61
5.3	Inférence Bayésienne de R_0	63
5.3.1	La distribution a priori de R_0	64
5.3.2	La distribution a posteriori de R_0	65
5.3.3	La distribution a posteriori des données manquantes	65
5.3.4	Algorithme d'estimation de R_0	66
5.4	Simulations MCMC	66
5.5	Application aux données réelles	67
5.6	Conclusion	67
6	Conclusion et perspectives	71
A		73
A.1	Preuve de l'équation (3.7) et l'équation (3.8)	73
A.2	Preuve de la Proposition 3.1	74
A.3	Preuve de la Proposition 3.2	78
B		79
B.1	Preuve de la Proposition 4.1	79
C		81
C.1	Preuve de la Proposition 5.1	81
C.2	Preuve de la Proposition 5.2	82
	Bibliographie	85

Liste des tableaux

3.1	Moyenne et variance a posteriori de β , μ et R_0 pour $m = 5$	43
3.2	Moyenne et variance a posteriori de β , μ et R_0 pour $m = 10$	43
3.3	Moyenne et variance a posteriori de β , μ et R_0 pour $m = 15$	43
3.4	Nombre d'individus susceptibles et infectés à des dates de la période terminale de la peste d'Eyam et leurs estimations générées à partir de l'algorithme de Gillespie et de l'approximation de Milstein avec les paramètres $\beta = 4.6264$ et $\mu = 2.9544$	46
3.5	Moyenne et variance a posteriori pour les estimations de β , μ et R_0 , correspondant à une étape de temps de 1 mois, pour les données de peste Eyam, avec 20000 itérations.	46
4.1	Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.1$	54
4.2	Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.5$	55
4.3	Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.9$	55
5.1	Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=5$	67
5.2	Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=10$	67
5.3	Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=15$	68
5.4	Nombre d'individus susceptibles et infectés aux dates de la période terminale de la peste d'Eyam et leurs estimations selon la méthode MCMC et la méthode MV avec les données manquantes $m = 15$	70

Table des figures

1.1	Schéma du modèle de type SIR.	3
3.1	Comparaison entre les approches de Gillespie $\{(**), (**)\}$, de Milstein $\{(-), (-)\}$ et d'Euler $\{(\dots), (\dots)\}$. Pour l'interprétation des références à la couleur dans la légende de cette figure et celles qui seront illustrées dorénavant, le lecteur est renvoyé à la version électronique de ce manuscrit ou le papier publié.	38
3.2	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.1 avec 20000 itérations dans tous les cas.	44
3.3	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.2 avec 20000 itérations dans tous les cas.	44
3.4	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.3 avec 20000 itérations dans tous les cas.	45
4.1	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.1 avec $\rho = 0.1$ et 20000 itérations dans tous les cas.	56
4.2	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.2 avec $\rho = 0.5$ et 20000 itérations dans tous les cas.	56
4.3	Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.3 avec $\rho = 0.9$ et 20000 itérations dans tous les cas.	57
5.1	Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 5$	68
5.2	Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 10$	69
5.3	Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 15$	69

Acronymes

Notation	Signification
AR-MH	Accept-Reject Metropolis-Hastings
BGD	Distribution gamma bivariée
EDS	Equation différentielle stochastique
EKP	Equations de Kolmogorov progressives
EKR	Equations de Kolmogorov retrogrades
EMV	Estimateur de maximum de vraisemblance
GIG	Gaussienne inverse généralisée
KBGD	Distribution gamma bivariée de Kibble
MCMC	Monte Carlo par chaînes de Markov
MGD	Distribution gamma multivariée
MH	Metropolis Hastings
MV	Maximum de vraisemblance
PGCD	Le plus grand commun diviseur
sd	Erreur standard
TV	Variation totale

Notations mathématiques

- Ω : Est un ensemble non vide, il désigne l'espace d'échantillonnage ou d'état
 \mathcal{F} : Tribu sur Ω
 \mathbb{P} : Une mesure de probabilité sur l'espace probabilisable (Ω, \mathcal{F})
 \mathcal{F}_t : Filtration
 \mathbb{E} : Espérance mathématique
 Var : Variance mathématique
 Cov : Covariance mathématique
 σ : Ecart-type
 Σ : Matrice variance-covariance
 ρ : Coefficient de corrélation
 \mathbb{R} : Ensemble des nombres réels
 \mathbb{R}^n : Ensemble des nombres réels de dimension n
 \mathbb{N} : Ensemble des nombres entiers naturels
 \mathcal{N} : La loi normale
 χ^2 : La loi chi-deux
 $\Gamma(.,.)$: La loi gamma
 $\Gamma(.)$: La fonction gamma
 $\mathcal{B}(.,.)$: La fonction bêta
 $I_p(.)$: La fonction de Bessel d'ordre p
 χ : La fonction indicatrice
 \mathcal{H} : Ensemble des processus mesurables et de carré intégrable
 L^2 : Espace des fonctions de carré intégrable
 \mathcal{C}^2 : Ensemble des fonctions deux fois dérivables et à dérivée seconde continue
 ∇ : L'opérateur gradient
 Δ : L'opérateur Laplacien
 Θ : Espace des paramètres
 \mathcal{A} : Opérateur infinitésimale
 $D_{\mathcal{A}}$: Ensemble des fonctions telles que la limite existe, $\forall x \in \mathbb{R}^n$
 $\mathcal{B}(\mathbb{X})$: Désigne la tribu Borélienne sur \mathbb{X} pour tout \mathbb{X} Borélien
 \sim : Equivalent
 \propto : Proportionnel
 $\xrightarrow{\text{Loi}}$: Convergence en loi.

Introduction Générale

Sommaire

1.1 Généralités	1
1.2 Formulation mathématique du modèle épidémique SIR . .	3
1.2.1 Modèle déterministe de type SIR	4
1.2.2 Modèle stochastique de type SIR	4
1.3 Contribution	6

1.1 Généralités

L'époque moderne a connu des déplacements de dizaines de millions d'individus sur des espaces régionaux et continentaux. Lors de ces déplacements, les êtres humains ont souvent amené avec eux, volontairement ou involontairement, des animaux, des plantes, des virus, des bactéries et des germes qui se sont révélés néfastes aux terres d'accueil, et y ont apporté de nouvelles maladies infectieuses et épidémies. Ainsi, les épidémies au fil du temps sont devenues une des causes fondamentales de misères, de mort et de pauvreté.

Malgré le progrès des sciences durant le XX^{ème} siècle, notamment le développement important des traitements et des vaccins efficaces contre les maladies infectieuses et transmissibles, on constate que les épidémies sont loin d'être contrôlées entièrement.

Vu ce contexte, les intéressés ont recours à la modélisation mathématique de ces fléaux afin de les analyser et de les contrôler. La modélisation mathématique de la dynamique des maladies infectieuses s'inscrit dans le cadre de renforcer la coopération interdisciplinaire entre les biomathématiciens et les épidémiologistes en vue d'accroître la capacité de prédiction. C'est ainsi "*qu'on a besoin de développer des modèles qui peuvent nous aider à évaluer les conséquences d'une stratégie*¹". Ainsi, la modélisation mathématique joue un rôle crucial dans le choix des programmes de vaccination.

La modélisation mathématique des épidémies a une longue histoire, elle débuta avec les travaux attribués à Daniel Bernoulli² qui présenta en 1760 un modèle, dont

1. Extrait traduit de l'ouvrage de Bailey [Bailey 1975].

2. Daniel Bernoulli est médecin, physicien et mathématicien suisse 1700-1782, neveu de Jacob Bernoulli, qui fut le premier à découvrir la théorie des probabilités.

le principal objectif était de savoir si la variolisation (l'inoculation du pus d'une personne atteinte de variole) était plus avantageuse ou plus risquée pour les personnes ayant contracté cette maladie. Ses hypothèses étaient simplistes : équiprobabilité d'être infecté par la petite vérole, pour toute personne ne l'ayant jamais eue, sans distinction ni d'âge ni de sexe. Il considéra également que le risque de mortalité dû à une cause autre que la petite vérole était égal au risque de mortalité dû à la maladie.

En dépit des faibles moyens de calcul disponibles à cette époque, Bernoulli démontra que la variolisation permettait d'augmenter l'espérance de vie des individus exposés au risque de maladie. Il détermina la population la plus sensible, exposée à cette maladie, et estima explicitement le taux de mortalité induit par celle-ci. Des détails supplémentaires sur ce modèle peuvent être trouvés dans Bernoulli [Bernoulli 1760], Zeeman [Zeeman 1993], Dietz et Heesterbeek [Dietz 2000] et Sabatier *et al.* [Sabatier 2005].

Cent cinquantes années après, Hamer [Hamer 1906] proposa un modèle dans lequel le nombre de nouveaux cas dépendait explicitement du nombre de cas existants et du nombre de susceptibles dans la population, il appliqua ce qu'on appelle la loi d'action de masse. La loi d'action de masse est une loi qui permet, en chimie, de caractériser la cinétique d'une réaction. Son principe est que la vitesse d'une réaction est proportionnelle aux concentrations de chacune des substances réactantes.

Au début du vingtième siècle, Sir Ronald Ross, après sa découverte du vecteur d'infection de la malaria qu'est le plasmodium, le parasite lui ayant été découvert par Laveran en 1897, et de l'influence des conditions climatiques et géographiques dans l'expansion de cette maladie infectieuse, proposa un modèle mathématique pour cette infection, dont le but était de démontrer que la réduction de la population anophélienne était un moyen de prévenir le paludisme (malaria). En 1911, Ross, dans son ouvrage "*The Prevention of Malaria*" [Ross 1911] (voir aussi [Bacäer 2011]) affirme que l'éradication du paludisme est possible dans une zone, à condition de faire baisser la densité des moustiques dans ladite zone. Il subdivise la population totale en deux compartiments ; les individus sains, mais sensibles ou susceptibles, c'est-à-dire pouvant être contaminées par la maladie et les individus ayant la malaria et pouvant ainsi la transmettre. Une fois contaminés, les hôtes sensibles deviennent immédiatement contaminants, et capables de transmettre la maladie et, une fois guéris, ils redeviennent sensibles et donc peuvent à nouveau contracter la malaria, l'hypothèse d'immunité n'étant pas prise en compte.

Malgré l'importance de la modélisation mathématique des épidémies, elle a été reçue avec réserve par les épidémiologistes et les responsables de la santé publique. Une des raisons à cela réside dans la complexité des modèles proposés et dans le fait qu'ils sont difficiles à comprendre par les non-mathématiciens. D'où les nécessités tant théorique que pratique des modèles simples, pratiques et limpides, et l'épidémiologie mathématique fût progressivement développée. Les contributions les plus importantes sont présentées dans l'ouvrage de Bailey [Bailey 1975], dans lequel, il fit état d'une classe importante de modèles épidémiques simples qui vont connaître un succès retentissant dans l'histoire de traitement mathématique des épidémies. Il s'agit des modèles épidémiques généraux de type SIR (Susceptible → Infecté → Eli-

miné³) (voir, FIGURE 1.1) ; leurs principales caractéristiques sont les suivantes : une population fermée est subdivisée en trois classes, la classe des susceptibles notée **S** (individus exposés mais sains), la classe des infectés notée **I** (individus infectés et infectieux) et la classe des éliminés notée **R** (infectés qui quittent l'état d'infection par immunisation, décès ou mis en quarantaine). Chaque individu reste infectieux pendant une période de temps appelée "**période d'infection**", une fois cette période est terminée, l'individu est éliminé du processus d'infection d'une façon permanente.



FIGURE 1.1 – Schéma du modèle de type SIR.

1.2 Formulation mathématique du modèle épidémique SIR

Deux types de formulations mathématiques pour les modèles de type SIR sont présentés dans la littérature, la formulation déterministe et la formulation stochastique. Dans ce travail, nous considérons que la seconde. Avant d'aller un peu loin, il est utile de présenter le modèle déterministe de base de Kermack-McKendrick [Kermack 1927].

1.2.1 Modèle déterministe de type SIR

Le modèle déterministe de type SIR le plus classique est celui décrit, pour la première fois, par Kermack-McKendrick. Le modèle suppose que la population est constante, que la période d'incubation de l'agent infectieux est instantanée, et que la durée de l'infectiosité est la même que la durée de la maladie. Le modèle suppose aussi que la population est entièrement homogène, sans structure d'âge, spatiale ou sociale.

Ce type de modélisation est adapté aux périodes d'intervalles tellement courtes. La mortalité naturelle et l'émigration sont équilibrées par la naissance et l'immigration ; pendant l'intervalle de temps $[t, t + \delta t]$, le nombre S de susceptible décroît de βSI , le nombre d'infectés I croît de βSI et décroît de μI et le nombre des éliminés R croît de μI où β (resp. μ) représente le taux d'infection (resp. le taux d'élimination). Donc, le modèle est gouverné par le système de trois équations différentielles non linéaires ordinaires suivantes :

³. La traduction anglaise du mot éliminé est "Removed", d'où la notation **R** de la classe des éliminés.

probabilité pour que deux événements aient lieu dans un laps de temps $[t, t + \delta t[$ assez petit est négligeable par rapport à δt , Bartlett [Bartlett 1949] et Bailey [Bailey 1950] établirent que les probabilités $P_{(i,j)}(t)$ pour tout $(i, j) \in \mathbb{S}$ vérifient les égalités suivantes :

$$\begin{aligned} P_{(i,j)}(t + \delta t) &= \left[\frac{\beta(i+1)(j-1)}{n} \delta t + o(\delta t) \right] P_{(i+1,j-1)}(t) \\ &+ [\mu(j+1)\delta t + o(\delta t)] P_{(i,j+1)}(t) \\ &+ \left[1 - \left(\frac{\beta ij}{n} + \mu j \right) \delta t + o(\delta t) \right] P_{(i,j)}(t), \end{aligned} \quad (1.3)$$

avec

$$P_{(i,j)}(t) = 0 \text{ si } (i, j) \notin \mathbb{S} \text{ et } P_{(n,a)}(0) = 1.$$

Ainsi, ces probabilités obéissent aux équations de Kolmogorov progressive (**EKP**), ces équations découlent d'une interprétation rigoureuse des probabilités de transitions infinitésimales en (1.2) et (1.3) et se déduisent comme suit :

En supposant qu'un seul événement a eu lieu dans un intervalle de temps $[t, t + \delta t[$ assez petit, alors (1.3) entraîne l'équation suivante :

$$\begin{aligned} \frac{P_{(i,j)}(t + \delta t) - P_{(i,j)}(t)}{\delta t} &= \frac{\beta(i+1)(j-1)}{n} P_{(i+1,j-1)}(t) + \frac{o(\delta t)}{\delta t} P_{(i+1,j-1)}(t) \\ &+ \mu(j+1) P_{(i,j+1)}(t) + \frac{o(\delta t)}{\delta t} P_{(i,j+1)}(t) \\ &- \left(\frac{\beta ij}{n} + \mu j \right) P_{(i,j)}(t) + \frac{o(\delta t)}{\delta t} P_{(i,j)}(t), \end{aligned}$$

alors, quand $\delta t \rightarrow 0$ on obtient les EKP's :

$$\begin{aligned} P'_{(i,j)}(t) &= \frac{\beta(i+1)(j-1)}{n} P_{(i+1,j-1)}(t) + \mu(j+1) P_{(i,j+1)}(t) \\ &- \left(\frac{\beta ij}{n} + \mu j \right) P_{(i,j)}(t), \end{aligned} \quad (1.4)$$

avec la condition initiale $P_{(n,a)}(0) = 1$.

1.3 Contribution

Récemment, les mathématiciens, les épidémiologistes, les immunologistes ont commencé à collaborer pour créer des modèles susceptibles de prédire l'évolution d'une maladie. Les modèles des maladies infectieuses ont d'abord été utilisés pour comprendre la dynamique temporelle et spatiale⁴ d'une épidémie, puis pour envisager une stratégie thérapeutique ou de lutte contre la maladie. Ces modèles des

4. Les modèles épidémiques spatiaux de type SIR, connaissent à l'heure actuelle une grande popularité et étudiés par assez bien de chercheurs en épidémiologie, voir en l'occurrence Zakary *et al.* [Zakary 2016].

maladies infectieuses sont de plus en plus utilisés pour prévoir un éventail de possibilités futures, afin d'aider et soutenir le développement des connaissances et le processus décisionnel au niveau scientifique, médical et sanitaire. Pour parvenir à ces objectifs, d'une part on a eu recours à de nouvelles approches inférentielles pour valider les modèles qui révèlent la dynamique de ces épidémies, d'autre part, prédire leur évolution et même pronostiquer les conséquences des programmes de vaccinations. L'étape d'estimation des paramètres est cruciale à justifier puisque les résultats en dépendent de façon évidente.

Pour les modèles de type SIR, il y a trois types de paramètres d'intérêt, le taux d'infection β , le taux d'élimination μ et le taux de reproduction de base $R_0 = \frac{\beta}{\mu}$, qui s'interprète comme étant le nombre moyen d'individus à qui un individu infecté transmet la maladie, durant toute sa période d'infection. Ce dernier paramètre est très important dans l'étude de la propagation des épidémies, étant donné qu'il permet de quantifier la virulence d'une maladie infectieuse et donc, de prévoir l'évolution de l'épidémie dans le temps. De plus, en pratique, on s'intéresse souvent à la valeur de R_0 pour juger l'efficacité de certaines mesures visant à ralentir la propagation d'une épidémie quantitativement, comme la vaccination. En effet, on cherche à savoir de quelle façon la valeur de R_0 a été modifiée suite à ces mesures. Enfin, mentionnons (e.g. El Maroufy et Ziad [Maroufy 2010] et Daley et Gani [Daley 1994] et leurs références à ce sujet) qu'une épidémie est sévère si et seulement si le taux de reproduction de base est au dessus de son seuil $R_0 = 1$. Grâce à son succès qui provient de sa formulation simple $R_0 = \frac{\beta}{\mu}$ et du fait qu'il est maintenant considéré, de façon unanime, comme une grandeur clé en épidémiologie, son estimation et celles des deux paramètres qui l'explicitent a fait l'objet d'un certain nombre d'essais dans la littérature. Les premières tentatives ont été avec les travaux de Becker [Becker 1989] et Anderson et May [Andersson 1991], en se basant sur la fonction de vraisemblance explicitée à partir des processus de comptage ; en observant complètement le processus (S, I) , l'estimateur de maximum de vraisemblance (EMV) de $\theta = (\beta, \mu)$ a été obtenu et vérifie les propriétés de consistance et de normalité asymptotique. Andersen *et al.* [Andersen 1993] ont investi les martingales construites à partir des processus de comptage⁵, ils ont montré que si le processus épidémique est observé au moment de son arrêt, l'estimateur de $\theta = (\beta, \mu)$ et celui de $R_0 = \frac{\beta}{\mu}$ sont consistents lorsque la taille de la population initiale est assez large. Ces approches d'estimations n'ont pas attiré assez d'attention ultérieurement ; une raison à cela réside dans le fait que les processus épidémiques sont rarement observés en détail au moment de leur extinction. Récemment, Oh [Oh 2013], en supposant que le processus épidémique n'est observé qu'à des points discrets, exprima la fonction de vraisemblance en fonction des probabilités de transition vérifiant les EKP's (1.4) ; mais une forme explicite simple applicable de ces probabilités de transition n'est pas encore disponible (voir, El Maroufy *et al.* [Maroufy 2012]). Globalement, il existe une vaste littérature sur l'inférence paramétrique classique concernant les modèles

5. Une théorie sur les martingales appliquée aux modèles épidémiques de type SIR est résumée dans l'ouvrage d'Andersson et Britton [Andersson 2000].

épidémiques de type SIR. Ce n'est pas notre objectif ici de faire un sondage sur ce sujet, plus les références susmentionnées, le lecteur intéressé est également référé à des publications récentes et à ses références (Cauchemez et Ferguson [Cauchemez 2008]; Cintrón-Arias *et al.* [Cintrón-Arias 2009]; Corani et Gatto [Corani 2007]; De Meester *et al.* [Meester 2002]; Ross *et al.* [Ross 2006], [Ross 2009]; Zu Dohna et Pineda-Krch [zu Dohna 2010]; Yip et Chen [Yip 1998] et Britton [Britton 1998]).

Cependant, il y avait peu de travaux de recherche sur l'approche Bayésienne pour l'estimation paramétrique des modèles épidémiques. Cette approche a été exploré au cours des années quatre vingt dix. A partir de cette période, les moyens de calcul rapides combinés avec les méthodes de simulation de Monte Carlo par chaînes de Markov (MCMC), ont fortement contribué à l'utilisation de ladite approche de façon correcte, rigoureuse et puissante. Comme dans de nombreux domaines d'applications statistiques, les modèles épidémiques ont considérablement profité des algorithmes MCMC. En particulier, Becker et Britton [Becker 1999], Becker [Becker 1989]; Choi et Rempala [Choi 2012]; Gibson et Renshaw [Gibson 1998]; Lekone et Roberts [Lekone 2004]; Roberts et Stramer [Roberts 2001] ont établi les premiers résultats d'inférence pour les modèles SIR, basés sur les algorithmes MCMC. Aussi, Les travaux qui se concentrent d'avantage sur l'inférence Bayésienne pour les modèles épidémiques de type SIR sont notamment, ceux de Clancy et O'Neill [Clancy 2008]; Demiris et O'Neill [Demiris 2005a]; Demiris et O'Neill [Demiris 2005b]; Demiris et O'Neill [Demiris 2006]; Streftaris et Gibson [Streftaris 2004]; Neal et Wilkinson [Neal 2005]; O'Neill et Roberts [O'Neill 1999]; Hayakawa *et al.* [Hayakawa 2003] et O'Neill [O'Neill 2002]). Plus précisément, les auteurs dérivent des lois a posteriori pour les paramètres β et μ et implémentent des algorithmes plus robustes et parfaits pour la simulation de ces lois. Fearnhead et Meligkotsidou [Fearnhead 2004] proposent une approche un peu différente basée sur l'algorithme progressive-retrograde qui peut permettre d'estimer les paramètres β et μ . Mais ces travaux restent loin d'être appliqués aux données réelles, du fait que les observations en épidémiologie sont généralement recueillies en faible fréquences. C'est pour cette raison que nous devons avoir recours à une méthode Bayésienne alternative plus adaptée et réellement applicable, tel est l'objectif de cette thèse.

Dans cette thèse, l'approche Bayésienne utilisée est différente de celle utilisée dans les références précédemment citées. Ici, on traite ce problème d'estimation en s'inspirant des travaux d'Eraker [Eraker 2001] et Kim *et al.* [Kim 1998] sur la volatilité stochastique des modèles en finance, en approximant le processus épidémique en espace d'états discret par un processus de diffusion. Cela revient à augmenter les données de faibles fréquences par l'introduction de données latentes entre chaque pair d'observations, simulées en utilisant l'approximation des équations différentielles stochastiques qui gouvernent le processus aléatoire par le schéma d'Euler ou bien de Milstein.

L'introduction des données intermédiaires entre chaque paire d'observations implique l'estimation des valeurs manquantes, en plus des paramètres du modèle, où les données manquantes et les paramètres sont traités comme des variables aléatoires. Cette tâche est effectuée par l'inférence dans un cadre Bayésien à l'aide des

méthodes MCMC qui mettent à jour en alternance les données manquantes et les paramètres.

En effet, le résultat d'une modélisation Bayésienne est une loi a posteriori conjointe de l'ensemble des paramètres. Cette loi est le fruit de la combinaison de la connaissance a priori disponible sur les paramètres (connaissance modélisée via une loi dite loi a priori) avec la nouvelle information apportée par les données issues d'une expérimentation. L'étude de la loi a posteriori permet de caractériser l'information actuelle que l'on a sur les paramètres. Ce processus de mise à jour de l'information est la base des approches Bayésiennes considérées ici.

Cette thèse a trois principaux objectifs ; en premier lieu, nous utilisons une densité hyperbolique comme approximation de la loi conditionnelle du processus en vue d'estimer les paramètres supposés des variables aléatoires indépendantes. Ensuite, nous traitons le même problème mais dans le cas où β et μ ont des lois a priori dépendantes. Finalement, le troisième objectif de ce travail est l'estimation classique et Bayésienne du taux de reproduction de base $R_0 = \frac{\beta}{\mu}$.

Conformément à ces trois objectifs, les principaux chapitres de cette thèse peuvent être structurés ainsi ; dans le deuxième chapitre des résultats préliminaires, outils et quelques définitions sont présentés. Dans le troisième chapitre, nous faisons une inférence Bayésienne pour le modèle stochastique épidémique SIR avec paramètres indépendants. L'objectif de ce chapitre est d'estimer les deux paramètres d'intérêt, le taux d'infection et le taux d'élimination, en plus des observations manquantes, en augmentant les données de faibles fréquences par l'introduction de données latentes entre chaque paire d'observations, simulées en utilisant la méthode de discrétisation de Milstein des équations différentielles stochastiques. Cette tâche est effectuée par l'application des méthodes MCMC qui mettent à jour en alternance les données manquantes et les paramètres avec leurs distributions a posteriori.

Dans le chapitre 4, nous présentons l'estimation Bayésienne du modèle stochastique épidémique SIR avec paramètres d'intérêt dépendants. Dans ce cas, on considère la distribution de Kibble gamma bivariée comme une bonne distribution candidate a priori pour modéliser la dépendance entre les paramètres. Un algorithme MCMC est établi pour explorer la distribution a posteriori des paramètres avec des données simulées pour illustrer la méthode proposée.

Dans le chapitre 5, encore l'inférence Bayésienne est utilisée pour l'estimation du taux de reproduction de base R_0 . On dérive la distribution a posteriori de R_0 et celle des données manquantes.

Enfin, la thèse se termine par une conclusion et perspectives générales et des Annexes présentant les résultats mathématiques et les démonstrations des propositions que nous avons utilisé dans ce travail.

Préliminaires

Sommaire

2.1	Mouvement Brownien	12
2.2	Intégrale d'Itô	13
2.3	Equations différentielles stochastiques	14
2.3.1	Formule d'Itô	14
2.3.2	Existence et unicité des solutions de l'équation d'Itô	15
2.4	Schémas de discrétisation	16
2.4.1	Schéma d'Euler	16
2.4.2	Schéma de Milstein	17
2.4.3	A propos de convergence	18
2.5	Processus de diffusion	18
2.6	Equation de Fokker-Planck	20
2.7	Approximation d'un processus à saut par un processus de diffusion	20
2.8	Inférence Bayésienne	24
2.8.1	Théorème de Bayes	24
2.8.2	Grandes lignes de l'approche Bayésienne	25
2.8.3	Probabilités <i>a priori</i> et <i>a posteriori</i>	25
2.9	Méthode de Monte Carlo par chaînes de Markov	25
2.10	Les Algorithmes MCMC	27
2.10.1	Algorithme de Metropolis-Hastings	28
2.10.2	Échantillonneur de Gibbs	29
2.10.3	Algorithme d'Acceptation Rejet-Metropolis Hastings	30
2.11	Algorithme de Gillespie	31
2.12	Inférence Bayésienne pour les processus de diffusion stochastiques	32

Dans ce chapitre, nous nous introduisons les ingrédients et les notions nécessaires pour la bonne compréhension de ce manuscrit et pour formuler les problèmes qui sont traités par la suite. Il est partagé en 12 Sections ; les Sections 2.1-2.6 comportent un bref rapport sur les éléments de bases de la théorie des équations différentielles stochastiques ainsi que les équations de Fokker-Planck associées à un processus de diffusion d'Itô. La Section 2.7 est consacrée à l'approximation des processus Markoviens discrets par des processus de diffusion. Dans les Sections 2.8-2.12 nous évoquons aux lecteurs le paradigme Bayésien et ces outils pratiques, notamment la méthode de MCMC et ces algorithmes les plus connus.

2.1 Mouvement Brownien

C'est le botaniste Robert Brown qui donna son nom au mouvement Brownien en observant vers 1827 les mouvements erratiques de particules de pollen en suspension dans un liquide. Quelques années plus tard en 1905, Albert Einstein mit en évidence les étranges relations que le processus entretenait avec l'équation de la chaleur. Vers 1909, Jean Perrin entreprit son étude expérimentale et Paul Langevin posa la première équation. Mais il faudra attendre 1925 les travaux de Norbert Wiener pour que le mouvement Brownien ait véritablement un sens mathématique, comme modèle d'un bruit blanc. À partir des années 1950, Kiyoshi Itô l'utilisa pour définir l'intégrale stochastique qui porte son nom et fonda les bases du calcul stochastique.

Définition 2.1. *Un processus stochastique est une famille $(X_t)_{t \in \mathbb{R}^+}$ de variables aléatoires indexée par t .*

Définition 2.2. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ est une suite croissante (au sens de l'inclusion) de sous-tribus de \mathcal{F} , i.e.*

$$\forall s, t \geq 0, \text{ si } s \leq t \text{ alors } \mathcal{F}_s \subset \mathcal{F}_t.$$

Dans ce cas, on dit que $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ est un espace probabilisé filtré.

Définition 2.3. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ une filtration et $X = (X_t)_{t \in \mathbb{R}^+}$ un processus aléatoire adapté pour cette filtration, i.e. pour tout $t \geq 0$, X_t est \mathcal{F}_t -mesurable. Supposons en plus que chaque X_t est intégrable, i.e. $\mathbb{E}(|X_t|) < \infty$; on dit que X est une \mathcal{F} -martingale¹ si pour tout t et $h \geq 0$, on a $X_t = \mathbb{E}(X_{t+h} | \mathcal{F}_t)$ ps (presque sûrement).*

Définition 2.4. *On dit qu'un processus $B = (B_t)_{t \in \mathbb{R}^+}$ est un mouvement Brownien standard sur $(\Omega, \mathcal{F}, \mathbb{P})$ s'il vérifie les assertions suivantes :*

- i- $B_0 = 0$, \mathbb{P} -presque sûrement (p.s),
- ii- B est à trajectoires continues; pour tout $w \in \Omega$ la fonction $t \rightarrow X_t(w)$ est continue.
- iii- $\forall t, h \geq 0$, la variable aléatoire $B_{t+h} - B_t$ est indépendante de la tribu du passé avant t , soit $\mathcal{F}_t^B = \sigma(B_u, u \leq t)_{t \in \mathbb{R}^+}$ et $B_{t+h} - B_t \sim \mathcal{N}(0, h)$,
- iv- B est à accroissement indépendants, c'est-à-dire pour tout $n \in \mathbb{N}$ et $t_i, i = 1, \dots, n, 0 \leq t_0 \leq t_1 \leq \dots \leq t_n$ les variables aléatoires $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ et B_{t_0} sont indépendantes.

2.2 Intégrale d'Itô

On considère un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ muni de la filtration \mathcal{F}_t .

1. Les lecteurs trouveront dans l'ouvrage de Klenke [Klenke 2014] un excellent cours de base sur l'espérance conditionnelle et la théorie des martingales.

Définition 2.5. On appelle tribu des prévisibles sur $\Omega \times [0, \infty[$ la plus petite tribu rendant mesurable tous les processus continus adaptés à la filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$. Un processus est prévisible s'il est mesurable par rapport à cette tribu.

Supposons donné un mouvement Brownien standard (parfois appelé processus de Wiener et il est noté, dans ce cas, par W_t) B_t , adapté à la filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ et tel que pour tout $0 \leq s \leq t$, l'accroissement $B_t - B_s$ soit indépendant de \mathcal{F}_s . Sur un intervalle de temps $[a, b]$, on note \mathcal{H} l'ensemble des processus $f(t, \omega)$ définis pour $t \in [a, b]$, \mathcal{F}_t -mesurable et de carré intégrable presque sûrement. Dans ces conditions, si $f \in \mathcal{H}$ et si $a = t_0 < t_1, \dots, < t_n < t_{n+1} = b$ est une subdivision de l'intervalle $[a, b]$, si f est indépendant des incréments $B_{t_{i+1}} - B_{t_i}$, on dit que f est prévisible.

Pour toute fonction f de \mathcal{H} , on définit l'intégrale stochastique d'Itô comme la limite dans L^2 des accroissements $B_t - B_s$. Ainsi, l'intégrale stochastique d'Itô est définie comme suit :

$$\int_a^b f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \sum_{i=0}^n f(t_i, \omega) [B_{t_{i+1}} - B_{t_i}](\omega).$$

Cette définition est cohérente avec les propriétés usuelles de l'intégrale au sens de Riemann et Stieltjes.

On a de plus quelques propriétés complémentaires :

- i- si $f \in \mathcal{H}$ et $\int_a^b \mathbb{E}(f^2(t, \omega)) dt < \infty$, alors $\int_a^b \mathbb{E}(f^2(t, \omega)) dB(t) = 0$,
- ii- si $f, g \in \mathcal{H}$ et $\int_a^b [\mathbb{E}(f^2(t, \omega)) + \mathbb{E}(g^2(t, \omega))] dt < \infty$, alors

$$\mathbb{E} \left(\int_a^b f(t, \omega) dB(t) \right) \left(\int_a^b g(t, \omega) dB(t) \right) = \int_a^b \mathbb{E}(f(t, \omega)g(t, \omega)) dt.$$

Si $f(t, \omega) \in \mathcal{H}$ et pour tout $t \in [a, b]$, telle que $\int_a^t \mathbb{E}(f^2(t)) dt < \infty$, alors l'intégrale stochastique $\int_a^b f(t, \omega) dB_t(\omega)$ est une martingale et ses trajectoires sont presque sûrement continues².

2.3 Equations différentielles stochastiques

Les équations que l'on rencontre en modélisation stochastique sont souvent des équations à un bruit blanc ou à coefficients qui sont eux-mêmes des processus aléatoires. Dans le cas de bruit blanc, on peut modéliser l'équation par un mouvement Brownien et la traiter comme une équation d'Itô. Dans le cas où le processus est Markovien, on a ce qu'on appelle une diffusion.

2. Il sera utile pour le lecteur de consulter l'ouvrage de Comets et Meyer [Comets 2006] qui exhibent les notions fondamentales et les propriétés élémentaires en rapport avec l'intégrale stochastique.

2.3.1 Formule d'Itô

On appelle processus d'Itô un processus stochastique X_t à valeurs dans \mathbb{R} vérifiant l'équation intégrale suivante :

$$X(t_2) - X(t_1) = \int_{t_1}^{t_2} a(t, X_t) dt + \int_{t_1}^{t_2} b(t, X_t) dB_t,$$

qu'on note sous forme différentielle

$$dX_t = a(t, X_t) dt + b(t, X_t) dB_t. \quad (2.1)$$

Si $f(t, x)$ est une fonction de classe $\mathcal{C}^2([0, +\infty[\times \mathbb{R})^3$ par rapport à x , alors $\frac{\partial f(t, X_t)}{\partial x}$ admet une intégrale stochastique par rapport au même processus Brownien donnée par la formule d'Itô suivante (cf. Øksendal [Øksendal 1995, chapitre 4]) :

$$df(t, X_t) = \left(\frac{\partial f(t, X_t)}{\partial t} + a(t) \frac{\partial f(t, X_t)}{\partial x} + \frac{1}{2} b^2(t) \frac{\partial^2 f(t, X_t)}{\partial x^2} \right) dt + b(t) \frac{\partial f(t, X_t)}{\partial x} dB_t.$$

Cette formule permet le calcul de l'espérance d'un processus donné. Elle permet aussi de montrer que certains processus sont des martingales, puisque l'intégrale stochastique l'est.

Pour un processus multidimensionnel X_t à valeurs dans \mathbb{R}^n , on a l'équation suivante

$$dX_t = A(t) dt + B(t) dB_t$$

où $A(t) = (a_1(t), \dots, a_n(t))^t$ et $B(t) = (b_{ij}(t))$ est une matrice d'ordre $n \times m$. Le processus Brownien étant de dimension m , $B_t = (B_1(t), \dots, B_m(t))$.

Soit $f(t, x)$ une fonction de classe \mathcal{C}^2 de $[0, +\infty) \times \mathbb{R}^n$ à valeurs dans \mathbb{R}^p , alors le processus $f(t, X_t)$ admet une dérivée stochastique donnée par

$$df(t, X_t) = \left[\frac{\partial f(t, X_t)}{\partial t} + \sum_{i=1}^n \frac{\partial f(t, X_t)}{\partial x_i} a_i(t) + \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^m \frac{\partial^2 f(t, X_t)}{\partial x_i \partial x_j} b_{ik}(t) b_{jk}(t) \right] dt + \sum_{k=1}^n \sum_{i=1}^m \frac{\partial f(t, X_t)}{\partial x_i} b_{ik}(t) dB_k(t). \quad (2.2)$$

La formule ci-dessus s'appelle la formule d'Itô pour un processus d'Itô multidimensionnel.

Dans le cas, où la fonction f est indépendante du temps et $X_t = B_t$ est un mouvement Brownien sur \mathbb{R}^n , la formule d'Itô se réduit à

$$f(B_t) = f(B_0) + \int_0^t \nabla f(B_s) dB_s + \frac{1}{2} \int_0^t \Delta f(B_s) ds,$$

3. \mathcal{C}^2 est l'ensemble des fonctions deux fois dérivables et à dérivée seconde continue.

où les deux opérateurs ∇ et Δ représentent respectivement, le gradient et le Laplacien. Cette formule permet d'entrevoir le lien qui existe entre la théorie probabiliste et la théorie du potentiel (cas où f est harmonique). Pour plus de détails à ce sujet, nous préconisons aux lecteurs l'ouvrage de Jedrzejewski [Jedrzejewski 2009].

2.3.2 Existence et unicité des solutions de l'équation d'Itô

Soit $T > 0$, $a(t, x)$ une fonction mesurable de $[0, T] \times \mathbb{R}^n$ dans \mathbb{R}^m et $b(t, x)$ une fonction mesurable de $[0, T] \times \mathbb{R}^n$ dans $\mathbb{R}^{n \times m}$, des matrices d'ordre $n \times m$ vérifiant les conditions suivantes :

Condition de croissance : il existe deux constantes C_1 et C_2 positives telles que, pour tout $x \in \mathbb{R}^n$,

$$\begin{aligned}\|a(t, x)\|^2 &\leq C_1^2(1 + \|x\|^2), \\ \|b(t, x)\|^2 &\leq C_2^2(1 + \|x\|^2).\end{aligned}$$

Condition de Lipschitz : il existe deux constantes K_1 et K_2 positives telles que pour tout, $x, y \in \mathbb{R}^n$,

$$\begin{aligned}\|a(t, x) - a(t, y)\| &\leq K_1 \|x - y\|, \\ \|b(t, x) - b(t, y)\| &\leq K_2 \|x - y\|.\end{aligned}$$

Soit X_0 une variable aléatoire indépendante de la tribu $\mathcal{F}_\infty = \sigma(B_s, s \geq 0)$ et $\mathbb{E}|X_0|^2 < \infty$. Alors l'équation différentielle stochastique d'Itô

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \quad X_0 = x_0 \in \mathbb{R}^m, \quad (2.3)$$

dans laquelle W_t désigne ici et désormais un mouvement Brownien standard ou un processus de Wiener, sous les conditions précédentes, admet une solution unique $(X_t)_{t \in \mathbb{R}^+}$ dont presque sûrement toutes les trajectoires $X_t(\omega), \omega \in \Omega$ sont continues. On montre aussi que $(X_t)_{t \in \mathbb{R}^+}$ est un processus Markovien d'état initiale X_0 et de probabilité de transition

$$p(s, x, t, A) = \mathbb{P}(X_t \in A | X_s = x) = \int_A p(s, x, t, y)dy, \quad \forall t > s \geq 0,$$

pour tout Borélien $A \subseteq X$ ($X = \mathbb{R}$ ou $X = \mathbb{R}^n$). $p(s, x, t, y)$ est la densité de probabilité de transition du processus Markovien $(X_t)_{t \in \mathbb{R}^+}$ de l'état $x \in X$ à l'instant s à l'état y à l'instant $t > s$, (cf. Capasso et Bakstein [Capasso 2005]). Si le processus $(X_t)_{t \in \mathbb{R}^+}$ est homogène, la densité ne dépendra que de $t - s$ et on la note dans ce cas $p(t - s, x, y)$.

Dans plusieurs modèles gouvernés par des équations stochastiques de type (2.3), les paramètres a et b dépendent d'un paramètre $\theta \in \Theta$ (où Θ est l'espace des paramètres) qu'on doit estimer.

Pour une série d'observations du processus $(X_t)_{t \in \mathbb{R}^+}$ discrètes $X_{t_0}, X_{t_1}, \dots, X_{t_N}$ aux

instants $0 = t_0, t_1, \dots, t_N$ et du fait que le processus est Markovien, la fonction de vraisemblance de θ s'écrit, voir par exemple Bishwal [Bishwal 2008], comme suit.

$$L(\theta) = \prod_{i=0}^{N-1} p_\theta(t_i, X_{t_i}, t_{i+1}, X_{t_{i+1}}) = \prod_{i=0}^{N-1} p_\theta(t_{i+1} - t_i, X_{t_i}, X_{t_{i+1}}). \quad (2.4)$$

Un estimateur de θ , noté $\hat{\theta}$, est obtenu en maximisant le logarithme de la fonction de vraisemblance :

$$\hat{\theta} = \arg \max \log(L(\theta)).$$

2.4 Schémas de discrétisation

2.4.1 Schéma d'Euler

L'un des schémas les plus utilisés pour l'approximation de la solution d'une équation différentielle stochastique est le schéma d'Euler, à l'origine utilisé pour générer des solutions aux équations différentielles déterministes. L'idée est la suivante : étant donné un processus d'Itô $\{X_t, 0 \leq t \leq T\}$ solution de l'équation différentielle stochastique (2.3), avec une valeur initiale déterministe $X_{t_0} = X_0$ et une discrétisation de l'intervalle $[0, T]$, $0 = t_0 < t_1 < \dots < t_N = T$. L'approximation d'Euler pour X est donnée par le processus stochastique continue Y satisfaisant le schéma itérative suivant :

$$Y_{i+1} = Y_i + a(t_i, Y_i)(t_{i+1} - t_i) + b(t_i, Y_i)(W_{i+1} - W_i), \quad (2.5)$$

pour $i = 0, 1, \dots, N - 1$, avec $Y_0 = X_0$. Nous avons simplifié les notations en posant $Y(t_i) = Y_i$ et $W(t_i) = W_i$. En général le pas $\Delta t = t_{i+1} - t_i$ est considéré constant (i.e. $\Delta t = \frac{1}{N}$). Entre deux instants t_i et t_{i+1} , le processus peut être défini différemment. Une approche naturelle est d'utiliser une interpolation linéaire, de sorte que $Y(t)$ soit définie par :

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

Nous pouvons voir que pour simuler le processus Y , il suffit de simuler un processus Brownien⁴.

2.4.2 Schéma de Milstein

Il est bien évidemment raisonnable de chercher d'avoir une méthode de résolution allant au delà de l'approximation précédente et dans cet état d'esprit, Milstein a proposé une approximation du second ordre qui utilise à nouveau le calcul stochastique différentiel. En utilisant à nouveau un pas de discrétisation temporel constant, on a le schéma itératif suivant pour l'équation différentielle stochastique (2.3) :

$$\begin{aligned} X_{i+1} &= X_i + a(X_i, t_i)(t_{i+1} - t_i) + b(X_i, t_i)(W_{i+1} - W_i) \\ &\quad + \frac{1}{2} b(X_i, t_i) \frac{\partial b}{\partial x} [(W_{i+1} - W_i)^2 - (t_{i+1} - t_i)]. \end{aligned} \quad (2.6)$$

4. Les lecteurs intéressés pourront se reporter à l'ouvrage de Lacus [Lacus 2008].

Cette approximation améliore les instabilités numériques par rapport à la méthode d'Euler. Toutefois, il y a un lien entre les deux méthodes dans le cas où on peut réaliser une transformée de Lamperti de l'équation différentielle stochastique de départ. En effet, dans le cas où l'équation stochastique de départ n'a pas de bruit multiplicatif, comme par exemple dans l'exemple du processus d'Ornstein-Uhlenbeck, la méthode d'Euler a un ordre de convergence fort égale à 1. Or, avec une transformée de Lamperti (si $b(x, t) = b(x)$ est indépendant du temps), on peut transformer l'équation stochastique en une autre équation sans bruit multiplicatif⁵. Ainsi, on peut montrer que le schéma d'Euler de l'équation transformée est identique au schéma de Milstein sur l'équation originale. Dans le cas où la transformée de Lamperti est difficile à obtenir analytiquement, il est utile d'utiliser le schéma de Milstein qui est plus précis.

2.4.3 A propos de convergence

Deux mesures de vitesse de convergence sont habituellement utilisées sur les schémas décrits ci-dessus. Si $Y_n \approx X_{n\Delta t}$ est une approximation d'une trajectoire du processus solution de l'équation différentielle stochastique précédente : la vitesse de convergence forte est donnée par le plus grand γ tel que

$$\mathbb{E}(|X_T - Y_N|) = o(h^\gamma),$$

et la vitesse de convergence faible concerne la convergence des moments et est donnée par le plus grand η tel que

$$|\mathbb{E}(f(X_T)) - \mathbb{E}(f(Y_N))| = o(h^\eta),$$

pour f polynôme ou fonction régulière à support compact. Pour chaque schéma, des conditions de régularité spécifiques devront être imposées à a et b de l'équation différentielle stochastique (2.1), pour que la vitesse propre à la méthode soit effectivement atteinte. L'ordre de convergence forte pour le schéma d'Euler est donné par $\gamma = \frac{1}{2}$ tandis que le schéma de Milstein a un ordre fort de convergence égal à 1.

2.5 Processus de diffusion

Un processus de diffusion⁶ est défini comme étant un processus Markovien dont la fonction de densité de probabilité de transition vérifie les propriétés suivantes ; pour tout $x \in \mathbb{R}^n$, $s \geq 0$ et $\varepsilon > 0$, on a uniformément :

$$\lim_{t \rightarrow s} \frac{1}{t - s} \int_{\|y-x\| > \varepsilon} p(s, x, t, y) dy = 0. \quad (2.7)$$

5. Si le processus X_t est d'Itô gouverné par l'équation (2.1) et $b(t, X_t) > 0$, $\forall(t, X_t)$, alors le processus $\tilde{X}_t = \int_\xi^x \frac{1}{b(t, X_t)} dx|_{x=X_t}$ où ξ est un point de l'espace d'état de X_t , est appelé la transformée de Lamperti de X_t et il est gouverné par l'équation suivante : $d\tilde{X}_t = \tilde{a}(t, \tilde{X}_t)dt + dW_t$, $\tilde{a}(t, \tilde{X}_t)$ à déterminer.

6. Une description complète des processus de diffusion est présentée dans l'ouvrage d'Øksendal [Øksendal 1995].

$$a(s, x) = \lim_{t \rightarrow s} \frac{1}{t - s} \int_{\|y-x\| \leq \varepsilon} p(s, x, t, y)(y-x) dy \text{ existe.} \quad (2.8)$$

$$\Sigma(s, x) = \lim_{t \rightarrow s} \frac{1}{t - s} \int_{\|y-x\| \leq \varepsilon} p(s, x, t, y)(y-x)(y-x)^T dy \text{ existe.} \quad (2.9)$$

L'égalité (2.7) s'appelle condition de Dynkin [Dynkin 1965]) et cela veut dire que des grands sauts sont improbables sur des intervalles de temps courts. La fonction vectorielle $a(s, x)$ dans (2.8) s'appelle la dérive et la matrice Σ dans (2.9), symétrique définie positive, s'appelle matrice de diffusion. La matrice b telle que $bb^T = \Sigma$ s'appelle le coefficient de diffusion.

Une telle décomposition existe due à la propriété "*définie positive*", (*i.e.* pour tout $x_0 \in \mathbb{R}^n$, $\sum_{i,j=1}^n (bb^T)_{i,j} x_i x_j \geq 0$), mais elle n'est pas unique, c'est à dire qu'il pourrait exister une autre matrice c vérifiant $cc^T = \Sigma$, telle que les matrices b et c n'ont pas le même nombre de colonnes. Mais Stroock et Varadhan [Stroock 1997, chapitre 5] montrèrent que le choix de coefficient n'affecte pas la distribution du processus $(X_t)_{t \in \mathbb{R}^+}$ vérifiant l'équation suivante :

$$dX_t = a(t, X_t)dt + b(t, X_t)dB_t, \quad X_0 = x_0, \quad (2.10)$$

où $x_0 \in \mathbb{R}^n$, $a = (a_1, \dots, a_n)$ et $b = (b_1, \dots, b_n)$ telle que $bb^T = \Sigma$, sont des fonctions mesurables remplissant les conditions de croissance et de Lipshitz. Le processus gouverné par (2.10) s'appelle processus de diffusion d'Itô. Si $p(t, x, y)$ désigne la probabilité de transition de X_t , alors les opérateurs U_t qui transforment toute fonction continue bornée f à

$$U_t f(x) = \mathbb{E}_x(f(X_t)) = \mathbb{E}(f(X_t) | X(0) = x) = \int_{\mathbb{R}^n} f(y)p(t, x, y)dy,$$

forment une famille de semi-groupe d'opérateurs $U_{t+s} = U_t U_s$, ce qui traduit l'équation de Chapman-Kolmogorov : (cf. Stroock et Varadhan [Stroock 1997])

$$p(t+s, x, y) = \int p(t, x, u)p(s, u, y)du.$$

Le générateur infinitésimal du semi-groupe est défini par : $\mathcal{A} = \lim_{h \rightarrow 0} \frac{U_h - I}{h}$, soit

$$\mathcal{A}f(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t}, \quad x \in \mathbb{R}^n.$$

L'ensemble des fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ pour lesquelles la limite existe quelque soit x dans \mathbb{R}^n est noté $D_{\mathcal{A}}$. Le générateur infinitésimal pour le processus de diffusion d'Itô homogène, *i.e.* $a(t, X_t) = a(X_t)$ et $b(t, X_t) = b(X_t)$ est définie, voir par exemple Øksendal [Øksendal 1995, Chapitre. 7], de la manière suivante : $\forall f \in \mathcal{C}^2(\mathbb{R}^n)$, alors $f \in D_{\mathcal{A}}$ et

$$\mathcal{A}f(x) = \sum_i a_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (bb^T)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}. \quad (2.11)$$

2.6 Equation de Fokker-Planck

Supposons que la densité de probabilité de transition $p = p(t, x, y)$ du processus de diffusion d'Itô (2.10) est suffisamment régulière de telle sorte que les dérivées dans les équations ci-dessous existent et soient continues. Alors, sous les conditions mentionnées dans la section précédente, p satisfait l'équation de *Kolmogorov progressive* (cf. Karatzas et Shreve [Karatzas 1998] ou Stroock et Varadhan [Stroock 1997]) : Pour tout $(s, x) \in \mathbb{R}^+ \times \mathbb{R}^n$,

$$\frac{\partial p(s, x, t, y)}{\partial t} = - \sum_{i=1}^n \frac{\partial}{\partial x_i} (a_i(t, y) p(s, x, t, y)) + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} (\Sigma_{i,j}(t, y) p(s, x, t, y)), \quad (2.12)$$

et l'équation de Kolmogorov rétrograde : Pour tout $(t, x) \in \mathbb{R}^+ \times \mathbb{R}^n$,

$$- \frac{\partial p(s, x, t, y)}{\partial s} = \sum_{i=1}^n (a_i(s, x) \frac{\partial}{\partial x_i} p(s, x, t, y)) + \frac{1}{2} \sum_{i,j=1}^n \Sigma_{i,j}(s, x) \frac{\partial^2}{\partial x_i \partial x_j} (p(s, x, t, y)). \quad (2.13)$$

Chacune des équations ci-dessus détermine d'une façon unique, sous réserve de certaines conditions (voir, Karatzas et Shreve [Karatzas 1998, chapitre 5]) la densité du processus p . De plus, si la densité de probabilité de transition d'un processus stochastique vérifie l'équation de Kolmogorov progressive alors c'est un processus de diffusion d'Itô vérifiant l'équation stochastique de type (2.10) et l'équation de Kolmogorov s'appelle, dans ce cas, équation de *Fokker-Planck*.

Les équations de Kolmogorov progressive et rétrograde sont des outils important dans l'approximation des processus Markovien à saut et à espace d'état discret par les processus de diffusion comme nous allons le voir dans la section suivante.

2.7 Approximation d'un processus à saut par un processus de diffusion

Généralement, en épidémiologie, on s'intéresse à l'évolution continue du nombre d'individus tels que le nombre des susceptibles, des infectés et des éliminés. Ces chiffres sont des quantités stochastiques d'espaces d'état discret, un sous-ensemble de \mathbb{N} ou un sous-ensemble multidimensionnel équivalent. Si le processus possède la propriété Markovienne, à espaces d'états discrets entraînent des trajectoires discontinues c'est à dire à saut, d'où l'appellation "*processus à saut*". Parfois les sauts sont infiniment petits par rapport à la taille totale du système, dans ce cas une approximation des trajectoires discontinues par celles qui sont continues est intéressante.

Cette Section est dédiée à l'approximation d'un processus épidémique de type SIR à saut par un autre processus à espace d'état continu et à trajectoires presque sûrement continues. Avant d'aller loin, nous rappelons qu'en général, la densité p d'un processus Markovien quelconque d'espace d'état $X \subset \mathbb{R}^n$ vérifie l'équation dif-

férentielle de Chapman-Kolmogorov (voir, Gihman et Skorohod [Gihman 1975]) :

$$\begin{aligned} \frac{\partial p(t, x)}{\partial t} &= \int_{\mathbb{X}} [Q(t, y, x - y)p(t, y) - Q(t, x, y - x)p(t, x)] dy \\ &\quad - \sum_{i=1}^n \frac{\partial}{\partial x_i} a_i(t, x)p(t, x) + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} \Sigma_{i,j}(t, x)p(t, x). \end{aligned} \quad (2.14)$$

Dans cette équation $p(t, x) = p(0, x_0, t, x)$, avec $x_0 = X_0$, Q le taux de transition, $a = (a_1, \dots, a_n)$ le vecteur dérivé et $\Sigma_{i,j}$ la matrice de diffusion sont définis conformément aux relations (2.8) et (2.9), telle que, pour tout $t \geq 0$:

$$Q(t, x, y - x) = \lim_{h \rightarrow 0} \frac{1}{h} p(t, x, t + h, y),$$

et

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_{\|y-x\| \leq \varepsilon} (y_i - x_i)(y_j - x_j)(y_k - x_k)p(t, x, t + h, y) dy = 0.$$

Gardiner [Gardiner 2010], évoqua que lorsque $Q(\cdot, \cdot, z) \equiv 0$, pour tout $z \neq 0$, l'équation (2.14) se réduit à l'équation de Fokker-Planck (2.12), ceci montre selon la section précédente que le processus est une diffusion d'Itô. Autrement dit, si $a \equiv 0$ et $\Sigma \equiv 0$, l'équation (2.14) est réduite à :

$$\frac{\partial p(t, x)}{\partial t} = \int_{\mathbb{X}} [Q(t, y, x - y)p(t, y) - Q(t, x, y - x)p(t, x)] dy,$$

si on considère que p est une probabilité au lieu d'une densité, il est convenable d'écrire l'équation rétrograde précédente sous forme :

$$\frac{\partial p(t, x)}{\partial t} = \sum_{\Delta} [Q(t, x - \Delta, \Delta)p(t, x - \Delta) - Q(t, x, -\Delta)p(t, x)]. \quad (2.15)$$

La somme se fait sur tous les sauts possibles du processus, dans ce cas le processus est purement à saut.

Dans le cas du modèle de type SIR, décrit dans la Section (1.2), il n'y a que deux sauts non nuls possibles du processus Markovien (S,I), i.e. $\Delta_1 = (-1, 1)$ (infection) et $\Delta_2 = (0, -1)$ (élimination) et

$$Q(t, S, I, j) = \begin{cases} \frac{\beta SI}{n} & \text{si } j = 1 \\ \beta I & \text{si } j = 0. \end{cases}$$

Grâce à (2.15) les équations de Kolmogorov (1.4) sont rétablies. Lorsque l'épidémie s'arrête de se transmettre, il ne reste plus d'infectés dans la population, la probabilité devient nulle i.e. $\frac{\beta I}{n} = 0$. Au lieu de considérer les nombres S et I , on les remplace par les proportions suivantes $x = \frac{S}{n}$ et $y = \frac{I}{n}$, $n = X(0)$ et $a = Y(0)$, où X et Y sont respectivement le nombre des susceptibles et des infectés.

2.7. Approximation d'un processus à saut par un processus de diffusion 11

Pour l'élaboration de l'équation de Fokker-Planck du processus de diffusion $z = (x, y)$, nous procédons par une méthode analytique, en posant

$$f(x, y) = n\beta xy = \frac{\beta}{n}XY \quad \text{et} \quad g(x, y) = n\mu y = \mu Y,$$

donc (1.4) se transforme, en posant $P_{(i,j)}(t) = P_{(x,y)}(t) = P(x, y, t)$ et $\varepsilon = \frac{1}{n}$, à

$$\begin{aligned} P'(x, y, t) &= f(x + \varepsilon, y - \varepsilon)P(x + \varepsilon, y - \varepsilon, t) \\ &\quad + g(x, y + \varepsilon)P(x, y + \varepsilon, t) \\ &\quad - [f(x, y) + g(x, y)]P(x, y, t). \end{aligned} \tag{2.16}$$

En ajoutant et retranchant des termes de (2.16) :

$$\begin{aligned} &\frac{1}{2} \overbrace{[f(x + \varepsilon, y)P(x + \varepsilon, y, t) - f(x, y)P(x, y, t)]}^{A_1} \\ &+ \frac{1}{2} \overbrace{[f(x, y)P(x, y, t) - f(x - \varepsilon, y)P(x - \varepsilon, y, t)]}^{A_2} \\ &- \frac{1}{2} \overbrace{\left[\begin{array}{c} f(x, y + \varepsilon)P(x, y + \varepsilon, t) - 2g(x, y + \varepsilon)P(x, y + \varepsilon, t) \\ - f(x, y)P(x, y, t) + g(x, y)P(x, y, t) \end{array} \right]}^{A_3} \\ &- \frac{1}{2} \overbrace{\left[\begin{array}{c} f(x, y)P(x, y, t) - g(x, y)P(x, y, t) \\ - f(x, y - \varepsilon)P(x, y - \varepsilon, t) + g(x, y - \varepsilon)P(x, y - \varepsilon, t) \end{array} \right]}^{A_4} \\ &+ \frac{1}{2} \overbrace{\left[\begin{array}{c} f(x + \varepsilon, y)P(x + \varepsilon, y, t) - 2f(x, y)P(x, y, t) \\ + f(x - \varepsilon, y)P(x - \varepsilon, y, t) \end{array} \right]}^{A_5} \\ &+ \frac{1}{2} \overbrace{\left[\begin{array}{c} f(x, y + \varepsilon)P(x, y + \varepsilon, t) + g(x, y + \varepsilon)P(x, y + \varepsilon, t) \\ - 2f(x, y)P(x, y, t) - 2g(x, y)P(x, y, t) \\ + f(x, y - \varepsilon)P(x, y - \varepsilon, t) + g(x, y - \varepsilon)P(x, y - \varepsilon, t) \end{array} \right]}^{A_6} \\ &- \overbrace{\left[\begin{array}{c} f(x + \varepsilon, y)P(x + \varepsilon, y, t) - f(x + \varepsilon, y - \varepsilon)P(x + \varepsilon, y - \varepsilon, t) \\ - f(x, y)P(x, y, t) + f(x, y - \varepsilon)P(x, y - \varepsilon, t) \end{array} \right]}^{A_7}. \end{aligned}$$

On a

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} A_1 &= \lim_{\varepsilon \rightarrow 0} \left[\frac{\varepsilon f(x + \varepsilon, y)P(x + \varepsilon, y, t) - \varepsilon f(x, y)P(x, y, t)}{\varepsilon} \right] \\ &= \frac{\partial}{\partial x}(\beta xy P(x, y, t)) \quad \text{car } \varepsilon f(x, y) = \beta xy, \end{aligned}$$

de même

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} A_2 &= \lim_{\varepsilon \rightarrow 0} \left[\frac{\varepsilon f(x, y)P(x, y, t) - \varepsilon f(x - \varepsilon, y)P(x - \varepsilon, y, t)}{\varepsilon} \right] \\ &= \frac{\partial}{\partial x}(\beta xy P(x, y, t)), \end{aligned}$$

même pour $\lim_{\varepsilon \rightarrow 0} A_3 = \frac{\partial}{\partial y} [(\beta xy - \mu y)P(x, y, t)]$ et ainsi de suite on continue de la même manière et on trouve $\lim_{\varepsilon \rightarrow 0} A_4 = \frac{\partial}{\partial y} [(\beta xy - \mu y)P(x, y, t)]$ et lorsque ε est assez petit

$$\begin{aligned} A_5 &\simeq \frac{\partial^2}{\partial x^2} [\varepsilon(\beta xy)P(x, y, t)] \\ A_6 &\simeq \frac{\partial^2}{\partial y^2} [\varepsilon(\beta xy + \mu y)P(x, y, t)] \\ A_7 &\simeq \frac{\partial^2}{\partial x \partial y} [\varepsilon(\beta xy)P(x, y, t)]. \end{aligned}$$

Finalement, on trouve :

$$\begin{aligned} \frac{\partial}{\partial t} P(x, y, t) &= \frac{\partial}{\partial x} (\beta xy P(x, y, t)) - \frac{\partial}{\partial y} [(\beta xy - \mu y)P(x, y, t)] \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[\frac{1}{n} (\beta xy) P(x, y, t) \right] \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left[\frac{1}{n} (\beta xy + \mu y) P(x, y, t) \right] \\ &\quad - \frac{\partial^2}{\partial x \partial y} \left[\frac{1}{n} (\beta xy) P(x, y, t) \right]. \end{aligned}$$

Sous forme vectoriel, en posant $z(t) = (x(t), y(t))$, on obtient

$$\frac{\partial}{\partial t} P(x, y, t) = -\frac{\partial}{\partial z} [U(x, y)P(x, y, t)] + \frac{1}{2} \frac{\partial^2}{\partial z^2} [\Sigma(x, y)P(x, y, t)], \quad (2.17)$$

avec

$$U(x, y) = \begin{pmatrix} -\beta xy \\ \beta xy - \mu y \end{pmatrix} \text{ et } \Sigma(x, y) = \frac{1}{n} \begin{pmatrix} \beta xy & -\beta xy \\ -\beta xy & \beta xy + \mu y \end{pmatrix}.$$

L'équation (2.17) s'appelle équation de Fokker-Planck associée au processus de diffusion $z = (x, y)$ qui est solution de l'équation différentielle stochastique suivantes :

$$dz(t) = U(z(t))dt + \sigma(z(t))dW_t \quad (2.18)$$

où

$$\sigma \sigma' = \Sigma; \quad \sigma(x, y) = \frac{1}{\sqrt{n}} \begin{pmatrix} \sqrt{\beta xy} & 0 \\ -\sqrt{\beta xy} & \sqrt{\mu y} \end{pmatrix} \text{ et } U(x, y) = \begin{pmatrix} -\beta xy \\ \beta xy - \mu y \end{pmatrix}.$$

Remarque 2.6.

- i- On peut retrouver (2.17) en développant le deuxième terme de (2.16) à l'ordre 2 autour de (x, y) quand ε est assez petit.
- ii- Aussi les équations de Fokker-Planck peuvent être rétablies, en caractérisant le processus épidémiques de probabilité de transition (1.2) par un processus de

Poisson, voir spécialement l'ouvrage de Ethier et Kurtz [Ethier 2005, Chapitre. 11]), le générateur infinitésimal \mathcal{G}_ε du processus (x, y) est tel que (cf. El Maroufy et al. [Maroufy 2016]),

$$\mathcal{G}_\varepsilon f(x, y) = \varepsilon^{-1} \beta xy (f(x - \varepsilon, y + \varepsilon) - f(x, y)) + \varepsilon^{-1} \mu y (f(x, y - \varepsilon) - f(x, y)).$$

Pour ε assez petit le générateur \mathcal{G}_ε est approché par l'opérateur adjoint \mathcal{A}^* de \mathcal{A} défini en (2.11) et il est donné par (2.17).

2.8 Inférence Bayésienne

L'essence du paradigme bayésien nous provient du révérend Thomas Bayes⁷. Toutes les méthodes de l'approche Bayésienne découlent essentiellement du fameux théorème de Bayes. Dans cette Section nous ferons une brève introduction à l'approche Bayésienne : théorème de Bayes dans sa forme la plus simple, densité *a priori* et *a posteriori*.

2.8.1 Théorème de Bayes

Le théorème de Bayes (*ou originalement appelé probabilité des causes*) est un résultat de base en théorie des probabilités. Etant donné deux événements A et B , il nous permet de trouver la probabilité de A sachant B si l'on connaît la probabilité de A , de B , et de B sachant A . Dans sa forme la plus simple, on a

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A)\mathbb{P}(B | A)}{\mathbb{P}(B)} \text{ avec } \mathbb{P}(A) > 0 \text{ et } \mathbb{P}(B) > 0.$$

2.8.2 Grandes lignes de l'approche Bayésienne

En premier lieu, cette approche considère les paramètres comme des variables aléatoires. Ainsi, ont tous une distribution dite *a priori* de densité $\pi(\theta)$. Le choix des lois a priori est une étape fondamentale dans l'analyse bayésienne. Ce choix peut avoir différentes motivations. Les stratégies sont diverses. Elle peuvent se baser sur des expériences du passé ou sur une intuition. Ensuite, on peut utiliser les règles de Bayes pour déduire une distribution *a posteriori* à partir de la distribution *a priori* et des données observées. Finalement, il est important de comprendre que la distribution *a priori* est subjective. On devrait la choisir avant d'observer les données ; Chaque soucieux est libre de choisir sa propre densité *a priori* en fonction de ses croyances personnelles appuyées par des informations a priori sur les paramètres du modèle (cf. Boldstad [Boldstad 2004]).

7. Thomas Bayes 1701-1761, est un mathématicien statisticien britannique, connu pour avoir formulé la fameuse formule de probabilité conditionnelles.

2.8.3 Probabilités *a priori* et *a posteriori*

D'après ce qui a été décrit plus haut, on peut trouver finalement la distribution de paramètre qu'on veut estimer sachant nos données *i.e.* $\pi(\theta | x)$ en connaissant la distribution des données sachant le paramètre $f(x | \theta)$. Ainsi,

$$\pi(\theta | x) = \frac{f(x | \theta) \pi(\theta)}{\int_{\Theta} f(x | \theta) \pi(\theta) d\theta}.$$

Cette dernière est la fameuse densité *a posteriori*. Elle représente ce qu'on sait par rapport au paramètre en considérant les données observées; c'est aussi la mise à jour de $\pi(\theta)$ après l'observation de notre échantillon.

2.9 Méthode de Monte Carlo par chaînes de Markov

Une chaîne de Markov est une suite aléatoire de variables aléatoires. Notons X_t la valeur d'une variable aléatoire au temps t . Ces variables aléatoires peuvent être vues comme évoluant dans le temps, avec une probabilité de transition dépendante de l'état présent dans la chaîne. On peut définir un noyau de transition comme une fonction qui détermine les transitions. Cette fonction est notée K et est définie sur $E \times \mathcal{B}(E)$ où E est l'espace des états (donc tous les états possibles). Si E est discret, alors le noyau de transition est simplement une matrice où chaque élément est

$$\begin{aligned} P_{i_n, j} &= \mathbb{P}[X_{n+1} = j | X_n = i_n, \dots, X_0 = i_0] \\ &= \mathbb{P}[X_{n+1} = j | X_n = i_n] \end{aligned}$$

$\forall j, i_n, \dots, i_0 \in E$. Si toute fois E est continu, alors le noyau de transition est défini comme $K(x, x')$, où x est l'ensemble de l'état présent et x' est l'ensemble de l'état vers lequel le système transite.

Alors, étant donné un noyau de transition K , une séquence $\{X_n\}_{n \geq 0}$ de variables aléatoire est une chaîne de Markov si pour tout k , on a

$$\begin{aligned} \mathbb{P}[X_{k+1} \in A | x_0, x_1, \dots, x_k] &= \mathbb{P}[X_{k+1} \in A | x_k] \\ &= \int_A K(x_k, x) dx. \end{aligned}$$

Cette propriété est appelée "*la propriété de Markov*". En résumé, l'état suivant dépend uniquement de l'état présent et non de l'histoire de la chaîne.

Une chaîne de Markov peut avoir une distribution stationnaire qu'elle garde pour toujours. Toutefois, pour que la distribution stationnaire soit unique et vrai peu importe le point de départ de la chaîne, cette dernière doit respecter quelques propriétés importantes.

Irréductibilité : On dit que i communique avec j si $\exists n > 0$ tel que la probabilité d'aller à j à partir de i en n pas n'est pas nulle. De plus, on dit que C est une

classe d'équivalence si $\forall i, j \in C$, i communique avec j et j communique avec i . Alors, si une chaîne de Markov ne comporte qu'une seule classe d'équivalence, alors on dit que la chaîne est irréductible. Cette propriété se traduit par,

$$\exists n > 0 \text{ tel que } \mathbb{P}[X_n = j \mid X_0 = i] > 0, \forall i, j.$$

Apériodicité : La période, notée d_i , d'un état i est le plus grand nombre qui divise tout n tel que la probabilité d'aller de i à i en n pas est strictement positive. En d'autres mots, $d_i = \text{PGCD} \left\{ n : P_{ii}^{(n)} > 0, n \geq 1 \right\}$ où le PGCD est le plus grand commun diviseur. Si $d_i = 1$, l'état est dite apériodique. Aussi, si la probabilité d'aller à i à partir de i est non nulle, alors i est apériodique. Alors, si tout les états de la chaîne sont apériodiques, la chaîne est donc dite apériodique. Avec ces deux propriétés, la chaîne de Markov doit converger vers une loi stationnaire. À ceci, on peut ajouter une dernière propriété, il s'agit de la réversibilité. Toutefois, elle n'est pas nécessaire ici. C'est seulement une propriété qui nous aide à construire une chaîne de Markov avec la loi stationnaire f .

Réversibilité : Une chaîne de Markov est dite réversible par rapport à f si la probabilité d'aller de i à j en n pas est exactement la même que d'aller de j à i en même nombre de pas n i.e. $f_i P_{ij}^{(n)} = f_j P_{ji}^{(n)}$ où f_i est la valeur de la distribution stationnaire à l'état i .

Ces propriétés précitées nous permettent de construire une chaîne de Markov ergodique⁸ dont la loi stationnaire converge vers la distribution que nous souhaitons f . Une méthode de *Monte Carlo par Chaînes de Markov* (MCMC) est une méthode de simulation de la distribution f associé au noyau de transition en produisant une Chaîne de Markov ergodique $(X^{(i)}, i \geq 0)$ dont la distribution stationnaire est f (ici, la distribution *a posteriori*) partant d'un point de départ $X^{(0)}$ (Etant donné que la chaîne est ergodique, la valeur de départ est, en principe, sans importance). En fait, nous devons renvoyer les lecteurs à Robert et Casella [Robert 2004] et Robert [Robert 2006] pour une introduction profonde et pédagogique pour comprendre les méthodes MCMC .

En générale, la méthode MCMC permet la génération d'une série de valeurs du paramètre θ d'un modèle en utilisant seulement la loi a posteriori, lorsque cette loi ne peut pas être simulée directement. La chaîne de Markov $(\theta^{(i)}, i \geq 0)$ produite est ergodique, ce qui signifie que $\theta^{(i)}$ de loi limite la loi d'intérêt $\pi(\theta|x)$ pour toute valeur initiale $\theta^{(0)}$. Par conséquent, pour i suffisamment large, la loi $\theta^{(i)}$ est approché par $\pi(\theta|x)$ quelle que soit $\theta^{(0)}$.

En pratique, on doit se demander que signifie un " i suffisamment large", car il détermine le nombre de simulations à effectuer. Le taux de convergence de la densité de

8. On a essayé de minimiser le recours à la théorie des chaînes de Markov, bien que certaines de ces notions comme l'ergodicité ne doit pas être omises. Nous renvoyons les intéressés à Meyn et Tweedie [Meyn 1993] et à Robert et Casella [Robert 2004, Chapitre. 4] pour un exposé plus expéditif de ces notions.

$\theta^{(i)}$ vers la densité limite, apporte une réponse à ce problème (cf. Robert et Tweedie [Robert 2005]), de plus ce taux de convergence dépend de $\theta^{(0)}$, mais il existe des tests de diagnostic voir, par exemple, Robert et Casella [Robert 2004, Chapitre. 12], considérés comme des indicateurs de stationnarité de la chaîne et limitent partiellement cette difficulté. Une fois acquis le principe d'utilisation de la méthode MCMC, la mise en œuvre de ce principe nécessite l'implémentation des algorithmes de génération pour produire de telles chaînes de Markov.

2.10 Les Algorithmes MCMC

Dans cette Section, nous présenterons les trois principaux algorithmes les plus importantes conçues pour créer des chaînes de Markov de loi stationnaire donnée, à savoir les algorithmes de *Metropolis-Hastings* (MH), l'échantillonnage de **Gibbs** et l'algorithme d'*Acceptation Rejet-Metropolis-Hastings* (AR-MH). Ces algorithmes s'appliquent à une grande variété de problèmes et se contentent d'une loi a posteriori connue à une constante près.

2.10.1 Algorithme de Metropolis-Hastings

En 1970, quelques années après les travaux de Metropolis et ses collaborateurs [Metropolis 1953], Hastings [Hastings 1970] développa la version moderne de *l'algorithme de Metropolis-Hastings*. Il a étendu l'application de l'algorithme de Metropolis à des cas plus généraux. La grande différence entre l'algorithme de Metropolis et celui dit Metropolis-Hastings est essentiellement la relaxation de l'hypothèse qui exige une distribution instrumentale (ou loi candidate) $q(\cdot|\theta)$ qui doit être simplement et rapidement simulable, et explicitement formulée à une constante multiplicative indépendante de θ et parfois symétrique, *i.e.* $q(\tilde{\theta} | \theta, x) = q(\theta | \tilde{\theta}, x)$. Etant donnée la densité à postériori $\pi(\theta|x)$ (dans ce context cette loi est appelée loi d'intérêt ou cible), connue à un facteur de normalisation près, et la densité instrumentale. L'algorithme génère la chaîne $(\theta^{(i)}, i \geq 0)$ comme suit :

ALGORITHME MH : ROBERT ET CASELLA [ROBERT 2004, SECTION 7.3]

i- Initialiser $\theta^{(0)}$.

ii- À l'itération i :

a- simuler $\tilde{\theta} \sim q(\theta | \theta^{(i-1)}, x)$

b- calculer $\alpha = \min\left(1, \frac{\pi(\tilde{\theta}|x)}{\pi(\theta^{(i-1)}|x)} \frac{q(\theta^{(i-1)}|\tilde{\theta}|x)}{q(\tilde{\theta}|\theta^{(i-1)},x)}\right)$

c- accepter $\tilde{\theta}$ avec la probabilité α :

$$\theta^{(i)} = \begin{cases} \tilde{\theta} & \text{avec la probabilité } \alpha \\ \theta^{(i-1)} & \text{sinon.} \end{cases}$$

iii- Changer la valeur de i à $i + 1$ et aller en ii.

Selon Robert et Casella [Robert 2004], la chaîne produite par cet algorithme est ergodique, *i.e.* Pour tout g telle que $\mathbb{E}_\pi [\|g(\theta)\|] < \infty$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g(\theta^{(t)}) = \int g(\theta) \pi(\theta) d\theta,$$

de plus on a

$$\lim_{n \rightarrow \infty} \left\| \int K^n(y, \cdot) \mu(d\theta) - \pi \right\|_{TV} = 0,$$

pour toute loi initiale μ , $K^n(y, \cdot)$ est le noyau de la chaîne après n transitions. En particulier $\lim_{t \rightarrow \infty} \mathbb{P}[y^{(t)} \in A] = \int_A \pi(y) dy$. Cet algorithme a une autre propriété importante est stipufiante ; le choix de la loi instrumentale n'influe pas sur la convergence de la suite produite vers la loi d'intérêt. Mais en pratique, ce choix peut facilement empêcher la convergence si les probabilités d'atteindre les points loin du support de la loi a priori π sont trop petites.

2.10.2 Échantillonneur de Gibbs

On voit l'échantillonnage de Gibbs comme étant un cas particulier de l'algorithme de Metropolis-Hastings. On se sert de cet algorithme quand on veut échantillonner des valeurs d'une fonction multidimensionnelle avec un nombre fixe de variables ; $\theta = (\theta_1, \theta_2, \dots, \theta_r)$. En générale la densité conjointe n'est pas nécessairement connue ; toutefois, toutes les distributions conditionnelles $\pi_i(\theta_i | \theta_{-i})$, où $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_r)$, doivent être connues. C'est dans cette optique que cette méthode semble être une bonne méthode pour faire l'échantillonnage d'une distribution *a posteriori* dans l'approche Bayésienne.

ALGORITHME GIBS : Robert et Casella [Robert 2004, Chapitre. 10]

i- Initialiser $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_r^{(0)})$

ii- À l'itération i simuler :

$$\begin{aligned} \theta_1^{(i+1)} &\sim \pi \left(\cdot \mid \theta_2^{(i)}, \dots, \theta_r^{(i)} \right) \\ \theta_2^{(i+1)} &\sim \pi \left(\cdot \mid \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_r^{(i)} \right) \\ &\vdots \\ &\vdots \\ \theta_r^{(i+1)} &\sim \pi \left(\cdot \mid \theta_1^{(i+1)}, \dots, \theta_{r-1}^{(i+1)} \right). \end{aligned}$$

iii- Changer la valeur de i à $i + 1$ et aller en ii-.

Ici, tous les candidats simulés sont acceptés contrairement à l'algorithme traité précédemment (i.e. $\alpha = 1$). La raison pour laquelle $\alpha = 1$ est qu'on génère des valeurs des vraies distributions. Alors nous n'avons pas besoin de corriger avec une certaine probabilité d'acceptation. Comme on peut le voir, la méthode de Gibbs est facile à implémenter. Toutefois, cette méthode a quelques problèmes connus dans certains cas particuliers. Premièrement, pour une distribution à grande dimension, si deux variables sont parfaitement corrélées, alors l'échantillonneur restera pris et ne pourra pas changer les valeurs associées à ces variables. De plus, si une fonction est définie de telle sorte que tous les états ont une probabilité associée très près de zéro et qu'un seul point de l'espace a une très grande probabilité, alors l'échantillonneur prendra trop de temps pour converger. Ceci occasionnera des erreurs dans la simulation (voir, Bégin [Bégin 2010]). Pour l'algorithme de Gibbs, le taux d'acceptation égal à 1 et ne peut s'appliquer si le vecteur paramètre à simuler est de dimension variable, (voir, Robert [Robert 1996]).

2.10.3 Algorithme d'Acceptation Rejet-Metropolis Hastings

L'algorithme d'*Acceptation-Rejet Metropolis-Hastings* (**AR-MH**) est un algorithme hybride entre l'algorithme *Acceptation-Rejet* et l'algorithme *Metropolis-Hasting*. Il a été établi par Chib et Greenberg [Chib 1995]. Soit $\pi(\theta|x) \propto p(y|\theta)\pi(\theta)$ la densité cible et $q(\theta|x)$ la densité proposée (instrumentale). L'algorithme AR-MH est une procédure d'échantillonnage MCMC dans laquelle $p(x|\theta)\pi(\theta) \leq Mq(\theta|x)$, où $M > 0$, n'est pas satisfaite pour certain $\theta \in \Theta$, où Θ est le support de la densité cible non normalisée p , ce qui est nécessaire pour l'algorithme AR. Dans ce cas, on considère $D = \{\theta \in \Theta : p(x|\theta)\pi(\theta) \leq Mq(\theta|x)\}$ et son complémentaire \bar{D} . Supposons que l'état actuel de la chaîne soit θ , alors sachant les données x l'algorithme AR-MH, selon Chib et Jeliazkov [Chib 2005, Section. 3], se fait en deux étapes :

Étape AR : Générer un tirage $\theta' \sim q(\theta|x)$; accepter θ' avec probabilité $\min\{1, p(y|\theta')\pi(\theta')/Mq(\theta|x)\}$. Continuer le processus jusqu'à qu'un tirage au sort θ' soit accepté.

Étape MH : Étant donné la valeur actuelle θ et la valeur proposée θ' , poser

$$\alpha = \begin{cases} 1, & \text{si } \theta \in D \\ \frac{Mq(\theta|x)}{p(x|\theta)\pi(\theta)}, & \text{si } \theta \in \bar{D} \text{ et } \theta' \in D \\ \min \left\{ 1, \frac{p(x|\theta')\pi(\theta')q(\theta|x)}{p(x|\theta)\pi(\theta)q(\theta'|x)} \right\}, & \text{si } \theta \in \bar{D} \text{ et } \theta' \in \bar{D}. \end{cases}$$

Retourner à θ' avec probabilité α . Sinon retourner à θ .

L'implémentation explicite de cet Algorithme est comme suit :

ALGORITHME AR-MH : Bégin [Bégin 2010]

- i- Initialiser $\theta^{(0)}$
- ii- À l'itération i :

- a- simuler $\tilde{\theta} \sim q(\theta|x)$
- b- calculer $\alpha = \pi(\tilde{\theta}|x)/M'q(\theta|x)$, la constante positive M' doit être bien choisie en fonction de M
- c- accepter $\tilde{\theta}$ avec la probabilité α :

$$\theta^{(i)} = \begin{cases} \tilde{\theta} & \text{avec la probabilité } \alpha \\ \theta^{(i-1)} & \text{sinon} \end{cases}$$

- d- s'il y a rejet, on retourne à l'étape a-
- e- sinon, calculer

$$\alpha' = \begin{cases} 1; & \text{si } \pi(\theta^{(i-1)}|x) < M'q(\theta^{(i-1)}|x) \\ \frac{\pi(\theta^{(i-1)}|x)}{M'q(\theta^{(i-1)}|x)}; & \text{si } \pi(\theta^{(i-1)}|x) > M'q(\theta^{(i-1)}|x) \\ & \text{et } \pi(\tilde{\theta}|x) < M'q(\tilde{\theta}|x) \\ \min \left\{ 1, \frac{\pi(\tilde{\theta}|x)}{\pi(\theta^{(i-1)}|x)} \frac{q(\theta^{(i-1)}|x)}{q(\tilde{\theta}|x)} \right\}; & \text{si } \pi(\theta^{(i-1)}|x) > M'q(\theta^{(i-1)}|x) \\ & \text{et } \pi(\tilde{\theta}|x) > M'q(\tilde{\theta}|x) \end{cases}$$

- f- accepter $\tilde{\theta}$ avec la probabilité α' :

$$\theta^{(i)} = \begin{cases} \tilde{\theta} & \text{avec la probabilité } \alpha' \\ \theta^{(i-1)} & \text{sinon.} \end{cases}$$

- iii- Changer la valeur de i à $i + 1$ et aller en ii-.

2.11 Algorithme de Gillespie

Le dernier algorithme que nous allons utiliser dans le cadre de notre travail est celui de Gillespie. Tel qu'il est établi par Gillespie [Gillespie 1977] est un algorithme de simulation exact pour les processus de Markov à pur sauts en temps continu, cet algorithme est adapté aux modèle épidémique SIR ; si nous connaissons l'état du système au temps t , ce qui signifie que le nombre de sujets susceptibles et infectés $(S(t), I(t)) = (s, i)$, alors nous devons considérer les deux événements qui peuvent se produire ; une infection avec le taux $\lambda_1 = si\frac{\beta}{n}$ ou une récupération $\lambda_2 = i\mu$, et les changements correspondants dans chaque classe. La mise en œuvre de l'algorithme repose sur deux étapes principales ; la première est le tirage du temps d'attente τ jusqu'à l'occurrence de l'événement suivant, qui est distribué exponentiellement avec le taux $\lambda = \lambda_1 + \lambda_2$. Dans la deuxième étape, les événements sont choisis au hasard en fonction des probabilités $p_k = \frac{\lambda_k}{\lambda}$, $k = 1, 2$. L'algorithme est implémenté comme suit :

ALGORITHME DE GILLESPIE

- i- À $t = 0$, établir le nombre de population initial $S(0) = s$ et $I(0) = i$.
- ii- Choisir τ à partir de la distribution exponentielle.
- iii- Choisir $k = 1, 2$ selon les probabilités $p_1 = \frac{\lambda_1}{\lambda}$ et $p_2 = \frac{\lambda_2}{\lambda}$.
- iv- Changer le nombre d'individus pour refléter l'événement k tiré à l'étape ii. Définir $t \rightarrow t + \tau$ et aller à l'étape ii.

2.12 Inférence Bayésienne pour les processus de diffusion stochastiques

Le problème réside dans le fait que les probabilités de transitions ne sont pas valables sous forme explicite pour les observations à temps discret. La plupart des techniques se débattent lorsque les temps d'inter-observation sont importants. Jeux de données dans les sciences de vie, cependant, peuvent bien être de type basse fréquence. Des exemples sont des enquêtes sur les plantes avec une évaluation annuelle, des épidémies où les rapports sur la santé publique infections par semaine, ou des mesures coûteuses et donc peu fréquentes en la génétique.

Récemment des méthodes d'inférence Bayésiennes, sont basées sur l'introduction des données manquantes telles que l'union des valeurs manquantes et des observations forment un ensemble de données à haute fréquence. Ceci facilite l'approximation de la densité de transition et permet donc une inférence paramétrique, même pour les grands temps d'inter-observation. En outre, les techniques conviennent pour des intervalles d'observation espacés de façon irrégulière, diffusions multivariées avec des composantes éventuellement latentes et pour des observations qui sont sujets à une erreur de mesure. Ils s'appliquent même lorsque des composantes différentes sur l'espace d'état sont observées de façon non synchrone. Stationnarité et ergodicité de la diffusion n'est généralement pas nécessaire. En tant que méthode Bayésienne, la procédure d'estimation n'est pas nécessairement dépendante de grands échantillons.

L'introduction de données intermédiaires entre les deux observations implique l'estimation des valeurs manquantes en plus des paramètres du modèle, où les deux, les données manquantes et les paramètres sont traités comme des variables aléatoires.

Cette tâche est performée par l'application de la méthode MCMC, qui met à jour alternativement les données imputées et le paramètre du modèle et sont habituellement réalisables dans un temps de calcul modéré. Les considérations suivantes exigent une familiarité avec les méthodes MCMC de base. Des textes introductifs sur ce sujet se trouvent dans Gilks *et al.* [Gilks 1996], Robert et Casella [Robert 2004], Gamerman et Lopes [Gamerman 2006] et C. Fuchs [Fuchs 2006].

Inférence Bayésienne pour le modèle épidémique stochastique SIR avec paramètres indépendants

Sommaire

3.1	Introduction	35
3.2	Diffusion hyperbolique	36
3.3	Inférence Bayésienne pour le modèle non linéaire SIR	38
3.4	Simulations MCMC	40
3.4.1	Simulation des données manquantes	40
3.4.2	La distribution a posteriori des paramètres β et μ	41
3.4.3	Algorithme	42
3.4.4	Simulations	42
3.4.5	Application aux données réelles	45
3.5	Conclusion	46

3.1 Introduction

L'inférence pour les modèles épidémiques est compliquée par le fait que, premièrement, une ou plusieurs variables du modèle peuvent être non observées, et deuxièmement, les données sont souvent disponibles à des moments discrets, tandis que le processus réel sous-jacent est continu dans le temps. Dans le cas où les données sont observées en temps continu, des estimations de paramètres peuvent être obtenues pour des données complètes (voir Becker [Becker 1976]) et dans le cas de données incomplètes, des estimations ont également été développées par Anderson et Britton [Andersson 2000] en utilisant des méthodes de martingale, filtrage direct exact (cf. Fearnhead et Meligkotsidou [Fearnhead 2004]), ou par les méthodes MCMC qui offrent une puissance et une flexibilité énormes par rapport à d'autres approches (voir par exemple, Gibson et Renshaw [Gibson 1998] ; O'Neill et Roberts [O'Neill 1999] ; Streftaris et Gibson [Streftaris 2004] ; Neal et Wilkinson [Neal 2005]). Récemment, l'inférence Bayésienne du modèle SIR a été traitée par plusieurs auteurs en utilisant des méthodes MCMC, comme nous l'avons déjà mentionné dans l'introduction, les approches Bayésiennes utilisées sont différentes de celles présentées

dans ce chapitre, puisque les données ne sont disponibles qu'à des instants discrets. Nous introduisons un ensemble de $m - 1$ données latentes auxiliaires entre chaque paire d'observations. L'idée de relever les données de basse fréquence observées avec des simulations à haute fréquence a été précédemment poursuivie par Pedersen [Pedersen 1995]. Cette tâche est réalisée par l'application de techniques MCMC (voir, Eraker [Eraker 2001] et El Maroufy *et al.* [Maroufy 2015]) qui mettent à jour alternativement les données et sont habituellement réalisables dans un temps de calcul modéré. Nous utilisons le schéma de Milstein qui fournit une approximation avec une précision améliorée qui est meilleure que le schéma d'Euler utilisé par El Maroufy *et al.* [Maroufy 2015]. De plus, nous déclarons que la famille des distributions a priori gamma indépendantes conduit à une famille de distribution a posteriori Gaussienne inverse généralisée (**GIG**). Ce choix de distributions a priori est convenable en terme d'inférence Bayésienne (cf. O'Neill et Roberts [O'Neill 1999]). En outre, la flexibilité de la distribution gamma en pratique nous encourage de la considérer comme distribution a priori des paramètres dans les modèles épidémiques, voir par exemple, Cauchemez *et al.* [Cauchemez 2004] ; Streftaris et Gibson [Streftaris 2004]. Nous développons des méthodes MCMC utilisées pour l'inférence afin d'explorer la distribution a posteriori des paramètres. De plus, l'algorithme **AR-MH** comme il est décrit dans la Sous-Section 2.10.3, donne une convergence plus rapide et exige une connaissance de la densité cible non normalisée et d'une densité de proposition (instrumentale) simple à simuler (cf. Eraker [Eraker 2001] et Golightly et Wilkinson [Golightly 2005]). Donc, nous suggérons dans ce chapitre une distribution normale comme une bonne densité de proposition pour simuler les données manquantes.

Dans la Section 3.2 du présent chapitre, nous établissons la diffusion hyperbolique. L'inférence Bayésienne pour le modèle SIR non linéaire est présentée dans la Section 3.3, tandis que les simulations par la technique MCMC sont présentées dans la Section 3.4 avec application aux données réelles. Les preuves de certaines déclarations mathématiques sont données dans l'Annexe A.

3.2 Diffusion hyperbolique

En général, nous considérons l'équation différentielle particulière suivante de (2.18) :

$$dY_t = U(Y_t, \theta)dt + \sigma^{1/2}(Y_t, \theta)dW_t$$

où Y_t est une variable d'état, W_t est un mouvement Brownien standard défini sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, $U(., .)$ et $\sigma(., .)$ sont des fonctions scalaires connues, et θ un vecteur de paramètres inconnus et nous supposons que les conditions sous lesquelles l'équation différentielle stochastique peut être résolue pour Y_t sont satisfaites (cf. Øksendal [Øksendal 1995]). Il est bien connu, comme mentionné dans Elerian [Elerian 1998], que le schéma de Milstein (2.6) pour la discrétisation fournit une meilleure approximation de la densité de transition que le schéma d'Euler (2.5), voir FIGURE 3.1 qui illustre les trajectoires du modèle SIR. Ce schéma décrit dans Kloeden et Platen [Kloeden 1992] pour les processus généraux est donné, pour Δt

suffisamment petit, par

$$Y_{t+\Delta t} = Y_t + U(Y_t, \theta)\Delta t + \sigma^{1/2}(Y_t, \theta)\Delta W_t + \frac{1}{2}\sigma^{1/2}(Y_t, \theta)\frac{\partial\sigma^{1/2}(Y_t, \theta)}{\partial Y_t}((\Delta W_t)^2 - \Delta t) \quad (3.1)$$

où $\Delta W_t = \sqrt{\Delta t}\xi$ avec ξ est une variable aléatoire normale standard.

Si on note $a = \sigma^{1/2}(Y_t, \theta)\sqrt{\Delta t}$ et $b = \frac{1}{2}\sigma^{1/2}(Y_t, \theta)\frac{\partial\sigma^{1/2}(Y_t, \theta)}{\partial Y_t}\Delta t$, donc, en utilisant l'équation (3.1), on obtient

$$H = Y_{t+\Delta t} - Y_t - U(Y_t, \theta)\Delta t + b = a\xi + b\xi^2 = b\left(\left(\xi + \frac{a}{2b}\right)^2 - \frac{a^2}{4b^2}\right).$$

On pose $\lambda = \frac{a^2}{4b^2}$ et $\bar{H} = \left(\xi + \frac{a}{2b}\right)^2$, on a \bar{H} suit une distribution chi-deux χ^2 non centrée de degré de liberté 1 et un paramètre de non-centralité $\lambda = \frac{a^2}{4b^2}$. Nous montrons (cf. Elerian [Elerian 1998, Theorem 2.1]) que la densité conditionnelle par rapport à Y_t de \bar{H} est

$$f_{\bar{H}}(h) = \frac{1}{2}e^{-\frac{\lambda+h}{2}}\left(\frac{h}{\lambda}\right)^{-\frac{1}{4}}I_{-\frac{1}{2}}(\sqrt{\lambda h}), \quad h \geq 0,$$

où $I_{-\frac{1}{2}}(\omega) = \sqrt{\frac{2}{\omega}}\sum_{j=0}^{\infty}\frac{(\omega/2)^{2j}}{j!\Gamma(j+\frac{1}{2})} = \sqrt{\frac{2}{\pi\omega}}\cosh(\omega)$ étant une fonction de Bessel modifiée du premier type d'ordre $-\frac{1}{2}$ et $\cosh(\omega) = \frac{1}{2}(\exp(\omega) + \exp(-\omega))$ est la fonction cosinus hyperbolique. Nous renvoyons le lecteur à littérature classique pour une présentation détaillée de ce type de fonctions, voir par exemple, Abramowitz et Stegun [Abramowitz 1972] et Lebedev et Silverman [Lebedev 1972].

Or $H = b(\bar{H} - \lambda)$, donc la densité de H est $f_H(h) = \frac{1}{|b|}f_{\bar{H}}\left(\frac{h}{b} + \lambda\right)$, et après un calcul simple, nous établissons que la densité de transition de probabilité approximative est exprimée comme suit : pour $\frac{h}{b} + \lambda \geq 0$ et $b \neq 0$

$$f_H(h) = \frac{1}{2|b|}\left(\frac{h}{\lambda b} + 1\right)^{-\frac{1}{4}}e^{-\left(\frac{h}{2b} + \lambda\right)}I_{-\frac{1}{2}}\left(\sqrt{\lambda\left(\frac{h}{b} + \lambda\right)}\right). \quad (3.2)$$

3.3 Inférence Bayésienne pour le modèle non linéaire SIR

Puisque la distribution inconditionnelle de l'épidémie non linéaire de SIR est inconnue, on ne s'attend pas à ce que les méthodes de maximum de vraisemblance ou de martingale donnent des résultats satisfaisants. Pour cette raison, nous proposons d'utiliser la méthode MCMC pour estimer les paramètres et les données latentes après discrétisation par le schéma de Milstein (2.6). Pour ce faire, considérons le processus suivant :

$$(x(t), z(t)) = \left(x(t), 1 + \frac{a}{n} - x(t) - y(t)\right),$$

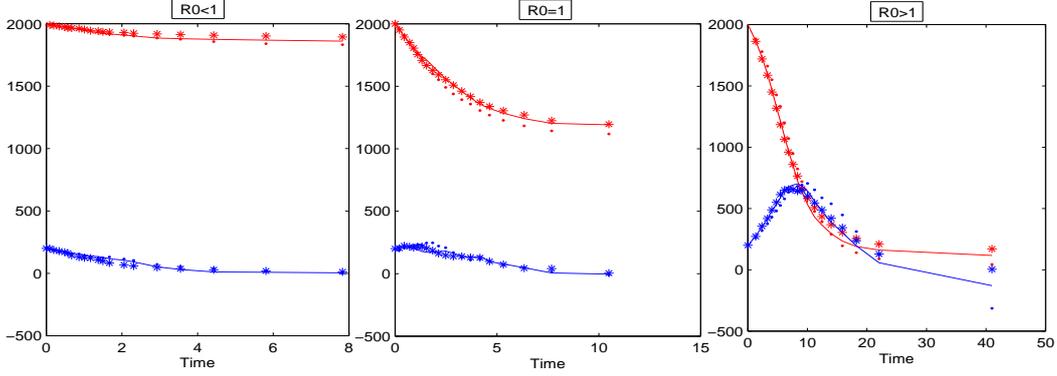


FIGURE 3.1 – Comparaison entre les approches de Gillespie $\{(**), (***)\}$, de Milstein $\{(-), (-)\}$ et d'Euler $\{(\dots), (\dots)\}$. Pour l'interprétation des références à la couleur dans la légende de cette figure et celles qui seront illustrées dorénavant, le lecteur est renvoyé à la version électronique de ce manuscrit ou le papier publié.

en utilisant (2.18) et en appliquant la formule d'Itô (2.2) on obtient

$$\begin{aligned} dx(t) &= U_1(x(t), z(t))dt + \sigma_1^{1/2}(x(t), z(t))dW_t^1, \\ dz(t) &= U_2(x(t), z(t))dt + \sigma_2^{1/2}(x(t), z(t))dW_t^2, \end{aligned} \quad (3.3)$$

où

$$\begin{aligned} U_1(x(t), z(t)) &= -\beta x(t) \left(1 + \frac{a}{n} - x(t) - z(t)\right), \\ U_2(x(t), z(t)) &= \mu \left(1 + \frac{a}{n} - x(t) - z(t)\right), \\ \sigma_1(x(t), z(t)) &= \frac{\beta x(t) \left(1 + \frac{a}{n} - x(t) - z(t)\right)}{n}, \\ \sigma_2(x(t), z(t)) &= \frac{\mu \left(1 + \frac{a}{n} - x(t) - z(t)\right)}{n}. \end{aligned} \quad (3.4)$$

Comme mentionné dans la Section 2.7 et la Section 3.2, nous allons utiliser avec la version discrétisée de (3.3), donnée par l'approximation de Milstein (2.6) :

$$\begin{aligned} x_i &= x_{i-1} + U_1(\mathbf{Y}_{i-1}, \theta)\Delta t + \sigma_1^{1/2}(\mathbf{Y}_{i-1}, \theta)\sqrt{\Delta t}\xi_1 + b_{i-1,1}(\xi_1^2 - 1), \\ z_i &= z_{i-1} + U_2(\mathbf{Y}_{i-1}, \theta)\Delta t + \sigma_2^{1/2}(\mathbf{Y}_{i-1}, \theta)\sqrt{\Delta t}\xi_2 + b_{i-1,2}(\xi_2^2 - 1), \end{aligned} \quad (3.5)$$

où $b_{i,1} = \frac{1}{2}\sigma_1^{\frac{1}{2}}(\mathbf{Y}_i, \theta)\partial_{x_i}\sigma_1^{\frac{1}{2}}(\mathbf{Y}_i, \theta)\Delta t$, $b_{i,2} = \frac{1}{2}\sigma_2^{\frac{1}{2}}(\mathbf{Y}_i, \theta)\partial_{z_i}\sigma_2^{\frac{1}{2}}(\mathbf{Y}_i, \theta)\Delta t$; ξ_1 et ξ_2 sont deux variables aléatoires normales standards indépendantes.

L'approximation discrète implique une densité de transition hyperbolique, qui est une très bonne approximation de cette dernière si l'intervalle de discrétisation Δt est suffisamment petit. D'autre part, nous introduisons un problème de données manquantes, l'intervalle de temps $[0, T]$ est divisé en $N = m \times T$ points équidistants $0 = t_0 < t_1 < \dots < t_{N-1} < t_N = T$, voir par exemple, El Maroufy *et al.* [Maroufy 2015], Eraker [Eraker 2001] et Golightly et Wilkinson [Golightly 2005]. Soit $\mathbf{Y}=(u, v)$, où u représente la partie observée du système et v désigne la partie non observée, où la totalité m d'observations sont manquantes dans l'intervalle de

temps $[t, t + 1[$. Nous notons par $\widehat{\mathbf{Y}}_i = (\widehat{x}_i, \widehat{z}_i)$ les données manquantes, et $\widehat{\mathbf{Y}}$ la matrice obtenue en empilant tous les éléments des données augmentées, c'est-à-dire

$$\widehat{\mathbf{Y}} = \begin{pmatrix} u_0 & \widehat{u}_1 & \cdot & \cdot & \cdot & \widehat{u}_{m-1} & u_m & \widehat{u}_{m+1} & \cdot & \cdot & \cdot & u_N \\ v_0 & \widehat{v}_1 & \cdot & \cdot & \cdot & \widehat{v}_{m-1} & v_m & \widehat{v}_{m+1} & \cdot & \cdot & \cdot & v_N \end{pmatrix}.$$

Conditionnellement à la première observation et sachant que le processus est Markovien, la densité a posteriori¹ est donnée par

$$\pi(\widehat{\mathbf{Y}}, \theta) \propto \prod_{i=1}^N p(\widehat{\mathbf{Y}}_i | \widehat{\mathbf{Y}}_{i-1}, \theta) \pi(\theta), \quad (3.6)$$

où $\widehat{\mathbf{Y}}_i$ est la i^{eme} colonne de $\widehat{\mathbf{Y}}$, π est la densité a priori du vecteur de paramètres θ et p représente la densité de probabilité de transition hyperbolique non normalisée donnée par (3.2). Notons que toutes les densités a posteriori conditionnellement à l'observation sont proportionnelles à (3.6). Pour simplifier, introduisons les quantités suivantes :

$$\begin{aligned} a_{i,1} &= \sigma_1^{\frac{1}{2}}(\mathbf{Y}_i, \theta) \sqrt{\Delta t} = \sqrt{\frac{\beta}{n} x_i \left(1 + \frac{a}{n} - x_i - z_i\right) \Delta t}, \\ b_{i,1} &= \frac{\Delta t}{2} \sigma_1^{\frac{1}{2}}(\mathbf{Y}_i, \theta) \partial_{x_i} \sigma_1^{\frac{1}{2}}(\mathbf{Y}_i, \theta) = \frac{\Delta t}{4n} \beta \left(1 + \frac{a}{n} - 2x_i - z_i\right) = \beta \widetilde{b}_{i,1}, \\ \lambda_{i,1} &= \frac{a_{i,1}^2}{4b_{i,1}^2} = \frac{1}{\beta} \frac{4n x_i (1 + \frac{a}{n} - x_i - z_i)}{\Delta t (1 + \frac{a}{n} - 2x_i - z_i)^2} = \frac{1}{\beta} \widetilde{\lambda}_{i,1}, \\ a_{i,2} &= \sigma_2^{\frac{1}{2}}(\mathbf{Y}_i, \theta) \sqrt{\Delta t} = \sqrt{\frac{\mu}{n} \left(1 + \frac{a}{n} - x_i - z_i\right) \Delta t}, \\ b_{i,2} &= \frac{\Delta t}{2} \sigma_2^{\frac{1}{2}}(\mathbf{Y}_i, \theta) \partial_{z_i} \sigma_2^{\frac{1}{2}}(\mathbf{Y}_i, \theta) = \frac{-\mu \Delta t}{4n} = \mu \widetilde{b}_{i,2} \end{aligned}$$

et

$$\lambda_{i,2} = \frac{a_{i,2}^2}{4b_{i,2}^2} = \frac{1}{\mu} \frac{4n (1 + \frac{a}{n} - x_i - z_i)}{\Delta t} = \frac{1}{\mu} \widetilde{\lambda}_{i,2}.$$

(3.5) implique conditionnellement à $\widehat{\mathbf{Y}}_i$ que x_{i+1} et z_{i+1} sont indépendantes (car W^1 et W^2 sont deux mouvements Browniens indépendants standards).

Donc $p(\widehat{\mathbf{Y}}_{i+1} | \widehat{\mathbf{Y}}_i) = p_{x_{i+1} | \widehat{\mathbf{Y}}_i} \times p_{z_{i+1} | \widehat{\mathbf{Y}}_i}$, où pour Δt suffisamment petit, (voir, Annexe A)

$$p_{x_{i+1} | \widehat{\mathbf{Y}}_i}(h_{i,1}) \simeq A_i \beta^{-\frac{1}{2}} e^{\frac{-\beta}{2 \left(\frac{\Delta x_i}{b_{i,1}} + \widetilde{\lambda}_{i,1}\right)} - \frac{\left(\frac{\Delta x_i}{2b_{i,1}} + \widetilde{\lambda}_{i,1}\right) - \left(\frac{\widetilde{\lambda}_{i,1} \Delta x_i}{b_{i,1}} + \widetilde{\lambda}_{i,1}^2\right)^{1/2}}{\beta}}, \quad (3.7)$$

$$p_{z_{i+1} | \widehat{\mathbf{Y}}_i}(h_{i,2}) \simeq B_i \mu^{-\frac{1}{2}} e^{\frac{-\mu}{2 \left(\frac{\Delta z_i}{b_{i,2}} + \widetilde{\lambda}_{i,2}\right)} - \frac{\left(\frac{\Delta z_i}{2b_{i,2}} + \widetilde{\lambda}_{i,2}\right) - \left(\frac{\widetilde{\lambda}_{i,2} \Delta z_i}{b_{i,2}} + \widetilde{\lambda}_{i,2}^2\right)^{1/2}}{\mu}}, \quad (3.8)$$

1. Aussi considérée comme une fonction de vraisemblance présentée implicitement par la formule (2.4).

avec $\Delta x_i = x_{i+1} - x_i$, $\Delta z_i = z_{i+1} - z_i$, $h_{i,1} = \Delta x_i + \beta x_i(1 + \frac{a}{n} - x_i - z_i)\Delta t + b_{i,1}$ et $h_{i,2} = \Delta z_i - \mu(1 + \frac{a}{n} - x_i - z_i)\Delta t + b_{i,2}$. A_i et B_i sont deux termes qui dépendent de x_i et z_i .

3.4 Simulations MCMC

Nous avons formulé dans (3.7) et (3.8) les densités de transitions conjointes pour les paramètres du modèle ainsi que les données observées et non observées, mais l'intérêt réel est dans la distribution des données manquantes. Comme dans le contexte épidémiologique, le nombre de données manquantes est relativement important, d'où un échantillonneur de Gibbs est le plus approprié pour échantillonner indépendamment les quantités .

3.4.1 Simulation des données maquantes

La première étape de l'échantillonnage de Gibbs implique la simulation des observations manquantes. Nous considérons ici le cas où l'échantillonneur de Gibbs met à jour une colonne de $\hat{\mathbf{Y}}$ dans le temps. Pour ce faire, nous avons besoin de la distribution a posteriori conditionnelle de $\hat{\mathbf{Y}}_i$, d'un simple calcul, donne :

$$\pi(\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1}) \propto p(\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1}),$$

où

$$p(\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1}) = p(x_i/\hat{\mathbf{Y}}_{i-1}) \times p(x_{i+1}/\hat{\mathbf{Y}}_{i+1}) \times p(z_i/\hat{\mathbf{Y}}_{i-1}) \times p(z_{i+1}/\hat{\mathbf{Y}}_{i+1}).$$

Les distributions marginales précédentes sont calculées en utilisant (3.7) et (3.8). À l'itération h de l'échantillonnage de Gibbs, on peut tirer $\hat{\mathbf{Y}}_i \sim \pi(\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1})$. Pour le modèle non linéaire, Eraker [Eraker 2001] suggère que $\hat{\mathbf{Y}}_i$ est mis à jour avec une densité de proposition normale $q(\cdot/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1})$ avec moyenne $\frac{\hat{\mathbf{Y}}_{i-1} + \hat{\mathbf{Y}}_{i+1}}{2}$ et variance $\frac{\Delta t}{2}\sigma^2(\hat{\mathbf{Y}}_{i-1})$. Eraker [Eraker 2001] motive ce choix en prouvant que, selon le schéma d'Euler, la densité de $\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1}$ est proche de $\mathcal{N}\left(\frac{1}{2}(\hat{\mathbf{Y}}_{i-1} + \hat{\mathbf{Y}}_{i+1}), \frac{\Delta t}{2}\sigma^2(\hat{\mathbf{Y}}_{i-1})\right)$ pour Δt suffisamment petit.

Dans le cas du schéma de Milstein et pour la même raison, la distribution normale est encore considérée comme une bonne densité de proposition (instrumentale) comme indiqué dans la proposition suivante, dont la preuve est exploré dans l'annexe A :

Proposition 3.1. *Pour Δt suffisamment petit, nous avons approximativement :*

$$\hat{\mathbf{Y}}_i/\hat{\mathbf{Y}}_{i-1}, \hat{\mathbf{Y}}_{i+1} \xrightarrow{Loi} \mathcal{N}\left(\frac{1}{2}(\hat{\mathbf{Y}}_{i-1} + \hat{\mathbf{Y}}_{i+1}), \frac{1}{2}\sigma^2(\hat{\mathbf{Y}}_{i-1})\Delta t\right).$$

3.4.2 La distribution a posteriori des paramètres β et μ

L'étape suivante dans l'échantillonneur de Gibbs est d'échantillonner $\theta^{(h+1)} = (\beta^{(h+1)}, \mu^{(h+1)})$ conditionnellement à son état actuel $\theta^{(h)}$ et de données augmentées

$\widehat{\mathbf{Y}}^{(h)}$.

Si la densité a posteriori des paramètres n'a pas de forme standard, nous pourrions utiliser l'algorithme de MH, mais ce n'est pas le cas ici. Etant donnée la forme de la fonction de vraisemblance exprimée en (3.7) et (3.8), une famille de distributions gamma indépendantes est considérée comme un ensemble naturel de densités a priori conjugués dans le contexte du modèle épidémique, où les paramètres sont positives (cf. Demiris et O'Neill [Demiris 2005a], [Demiris 2005b]). Ce choix de distributions a priori est commode en terme d'inférence Bayésienne, comme a été considéré par O'Neill et Roberts [O'Neill 1999]. De plus, vue la flexibilité de la distribution gamma, elle est utilisée en pratique comme distribution a priori des paramètres dans les modèles épidémiques (cf. Cauchemez *et al.* [Cauchemez 2004]; Streftaris et Gibson [Streftaris 2004]). Lorsqu'on considère ce type de distributions a priori, on obtient la distribution GIG² comme distribution a posteriori. Dans la proposition suivante, nous prouvons cette affirmation (pour la preuve, voir Annexe A).

Proposition 3.2. *Si β et μ suivent deux distributions gamma indépendantes : $\pi(\beta) \propto \Gamma(m_\beta, \lambda_\beta)$ et $\pi(\mu) \propto \Gamma(m_\mu, \lambda_\mu)$, alors*

$$\pi(\beta \mid \widehat{\mathbf{Y}}) \propto \beta^{-\frac{N-1}{2} + m_\beta - 1} e^{-\frac{1}{2}(A_1\beta + \frac{B_1}{\beta})} \quad (3.9)$$

et

$$\pi(\mu \mid \widehat{\mathbf{Y}}) \propto \mu^{-\frac{N-1}{2} + m_\mu - 1} e^{-\frac{1}{2}(A_2\mu + \frac{B_2}{\mu})}, \quad (3.10)$$

où A_1, A_2, B_1 et B_2 sont présentées en Annexe A. C'est à dire, les distributions a posteriori de β et μ sont des distributions GIG.

3.4.3 Algorithme

- i- Initialiser $\widehat{\mathbf{Y}}_i$, β et μ en utilisant une interpolation linéaire entre les valeurs observées de \mathbf{Y}_i pour initialiser $\widehat{\mathbf{Y}}_i$.
- ii- Pour tout $i = 1, 2, \dots, N - 1$, à l'itération h , simuler $\widehat{\mathbf{Y}}_i^{(h)} \mid (\widehat{\mathbf{Y}}_{i-1}^{(h-1)}, \widehat{\mathbf{Y}}_{i+1}^{(h-1)}, \beta, \mu)$ en utilisant l'algorithme AR-MH avec densité de proposition de loi normale $\mathcal{N}\left(\frac{1}{2}(\widehat{\mathbf{Y}}_{i-1}^{(h)} + \widehat{\mathbf{Y}}_{i+1}^{(h-1)}), \frac{\Delta t}{2} \sum_{i-1}^{(h)}\right)$, avec $\sum_{i-1} = \sigma^2(\widehat{\mathbf{Y}}_{i-1})$.
- iii- Simuler $\beta^{(h)}$ en utilisant (3.9).
- iv- Simuler $\mu^{(h)}$ en utilisant (3.10).
- v- Augmenter la valeur de h et retourner à l'étape ii-.

2. Si les lecteurs ne sont pas familiarisés avec ce type de lois, nous les renvoyons à consulter la référence de Koudou et Ley [Koudou 2014] et sa bibliographie non seulement pour les propriétés et caractérisations des lois GIG's mais aussi pour leurs applications à la statistique.

3.4.4 Simulations

Afin d'étudier la performance de la méthodologie d'estimation, nous avons d'abord généré des données artificielles. Les données sont échantillonnées à partir de la chaîne de Markov double discrète initiale en utilisant l'algorithme de Gillespie exact donné dans la Section 2.11, avec des valeurs réelles de β et μ comme indiqué dans les TABLES 3.1, 3.2 et 3.3. L'estimation est obtenue après 20000 itérations de notre algorithme, avec $m = 5$, $m = 10$ et $m = 15$ (nombre de données manquantes).

	valeurs de β	$\hat{\beta}$	valeurs de μ	$\hat{\mu}$	R_0	\hat{R}_0
moyenne	1	1.0626	1	1.0998	1	0.9661
variance		0.0431		0.0485		
moyenne	0.7	0.6905	0.2	0.2131	3.5	3.2402
variance		0.0492		0.0774		
moyenne	0.3	0.3105	0.5	0.5672	0.6	0.5475
variance		0.0149		0.0086		

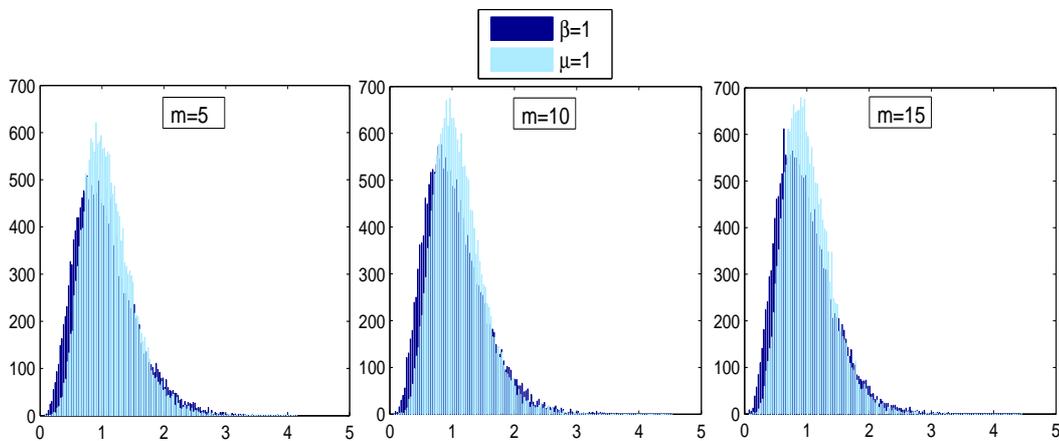
TABLE 3.1 – Moyenne et variance a posteriori de β , μ et R_0 pour $m = 5$.

	valeurs de β	$\hat{\beta}$	valeurs de μ	$\hat{\mu}$	R_0	\hat{R}_0
moyenne	1	1.0594	1	1.1001	1	0.9630
variance		0.066		0.0677		
moyenne	0.7	0.6777	0.2	0.2102	3.5	3.2240
variance		0.0507		0.0796		
moyenne	0.3	0.2961	0.5	0.5374	0.6	0.5509
variance		0.0147		0.0085		

TABLE 3.2 – Moyenne et variance a posteriori de β , μ et R_0 pour $m = 10$.

Les TABLES 3.1, 3.2 et 3.3 et les FIGURES 3.2, 3.3 et 3.4 montrent que les estimations des paramètres sont en général beaucoup plus proches de leurs vraies valeurs que celles dans El Maroufy *et al.* [Maroufy 2015] et la valeur de m augmente considérablement. Les FIGURES 3.2, 3.3 et 3.4 illustrent les histogrammes des densités a posteriori des paramètres, qui révèlent la convergence de l'algorithme vers une distribution limite. Dans les cas réels où les vraies valeurs ne sont pas disponibles, il est préférable de choisir m le plus grand possible, ce qui entraîne un énorme coût de calcul. Mais pour garder l'équilibre entre le coût de calcul et la qualité de l'ajustement, une petite valeur moyenne de m devrait être considérée.

	valeurs de β	$\hat{\beta}$	valeurs de μ	$\hat{\mu}$	R_0	\hat{R}_0
moyenne	1	1.0626	1	1.0998	1	0.9661
variance		0.0431		0.0485		
moyenne	0.7	0.6905	0.2	0.2131	3.5	3.2402
variance		0.0492		0.0774		
moyenne	0.3	0.3105	0.5	0.5672	0.6	0.5475
variance		0.0149		0.0086		

TABLE 3.3 – Moyenne et variance a posteriori de β , μ et R_0 pour $m = 15$.FIGURE 3.2 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.1 avec 20000 itérations dans tous les cas.

3.4.5 Application aux données réelles

Dans cette Section, nous simulerons une flambée de peste dans une population fermée. Les données sont tirées de celle d'Eyam, Derbyshire, en Angleterre, qui a connu une telle flambée suite à la Grande peste de Londres (voir, Raggett [Raggett 1982]). La peste Eyam n'a survécu qu'à 83 individus d'une population initiale de 350 personnes. Les valeurs des sujets susceptibles et infectés dans la peste Eyam sont données dans Raggett [Raggett 1982, TABLE II] de la mi-juin à la mi-octobre 1666 commençant par $S(0) = 254$ et $I(0) = 7$. Nous avons exécuté l'algorithme pour les données de la peste Eyam de la mi-juin à la mi-septembre (avant l'extinction de l'épidémie), comme le montre la TABLE 3.4. Comme l'intervalle de temps des données est $15\frac{1}{2}$ jours, les paramètres sont multipliés par 2 pour obtenir les taux correspondant à 1 mois pour comparer avec les résultats de Raggett calculés dans un temps d'échelle de 1 mois. Par exemple, pour $m = 15$ données simulées latentes, on obtient $\beta = 2.3132$ correspondant à une étape de temps de $15\frac{1}{2}$ jours, c'est $\beta = 4.6264$ par mois et $\mu = 1.4772$ correspondant à $15\frac{1}{2}$ jours c'est $\mu = 2.9544$ par

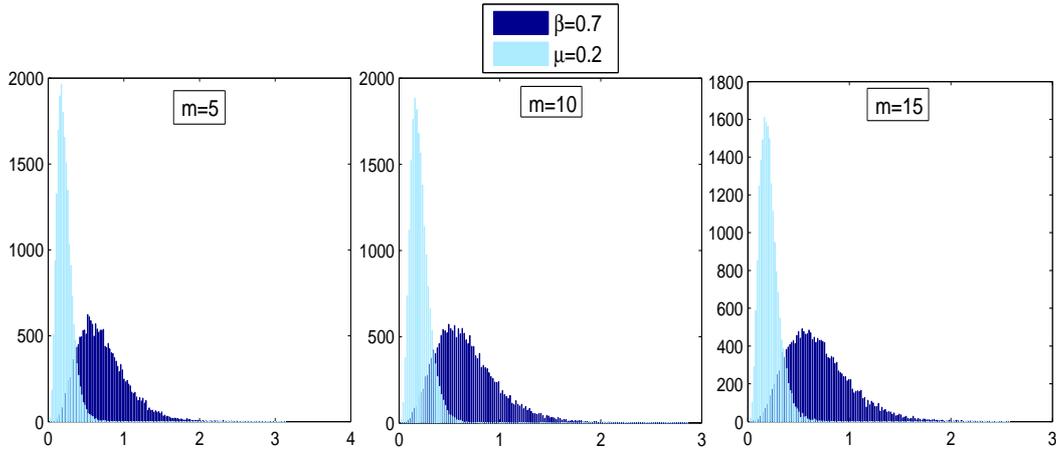


FIGURE 3.3 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.2 avec 20000 itérations dans tous les cas.

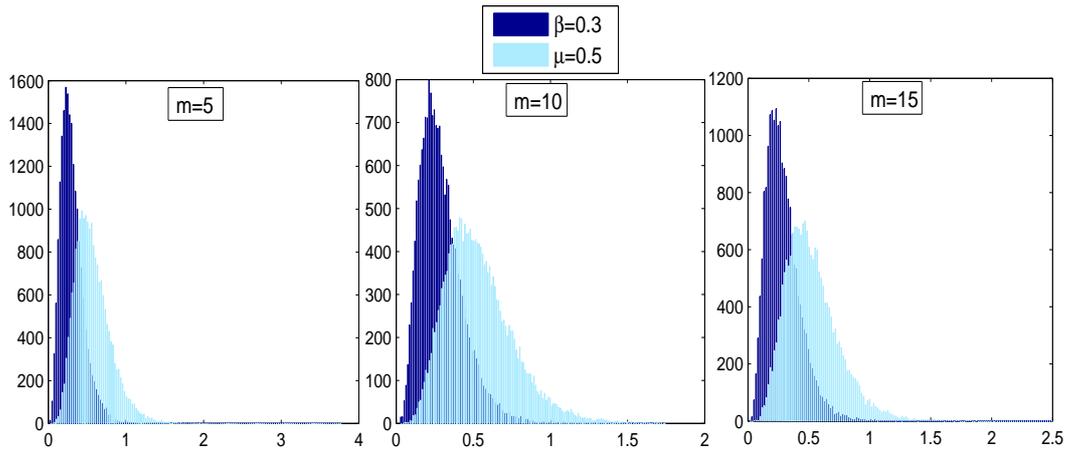


FIGURE 3.4 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 3.3 avec 20000 itérations dans tous les cas.

mois. Les valeurs obtenues par notre algorithme sont très proches des estimations données ci-dessus à partir du modèle déterministe (voir, Raggett [Raggett 1982]). Nous avons également obtenu des estimations pour les données réelles en utilisant les paramètres $\beta = 4.6264$ et $\mu = 2.9544$, données dans la TABLE 3.5 avec $m = 15$, la troisième colonne pour l'algorithme de Gillespie et la quatrième colonne pour l'approximation de Milstein.

	Données réelles		Gillespie		Milstein	
Date(1666)	S(t)	I(t)	S(t)	I(t)	S(t)	I(t)
Juin 19	254	7	254	7	254	7
Juillet 3/4	235	14.5	231.02	14.38	229.27	15.17
Juillet 19	201	22	197.57	21.12	189.10	24.29
Août 3/4	153.5	29	164.59	21.64	146.78	25.83
Août 19	121	20	139.61	17.35	117.12	19.12
Septembre 3/4	108	8	123.52	11.82	100.88	11.25
Septembre 19	97	8	113.84	7.34	92.87	6.87
Octobre 4/5	inconnu	inconnu	108.25	4.35	89.06	2.88
Octobre 20	83	0	105.06	2.49	87.28	1.37

TABLE 3.4 – Nombre d’individus susceptibles et infectés à des dates de la période terminale de la peste d’Eyam et leurs estimations générées à partir de l’algorithme de Gillespie et de l’approximation de Milstein avec les paramètres $\beta = 4.6264$ et $\mu = 2.9544$.

m		a posteriori de β	a posteriori de μ	\hat{R}_0
5	moyenne	4.6716	2.9932	1.5607
	variance	0.0431	0.0485	
10	moyenne	4.6506	2.9722	1.5647
	variance	0.0412	0.0178	
15	moyenne	4.6264	2.9544	1.5659
	variance	0.0403	0.0171	

TABLE 3.5 – Moyenne et variance a posteriori pour les estimations de β , μ et R_0 , correspondant à une étape de temps de 1 mois, pour les données de peste Eyam, avec 20000 itérations.

3.5 Conclusion

Ce chapitre a porté sur l’estimation Bayésienne du modèle épidémique stochastique SIR à partir des données de trajectoire. Les estimateurs proposés sont basés sur la simulation par les méthodes MCMC. De telles méthodes se sont avérées particulièrement bien adaptées pour résoudre les intégrales de grande dimension associées au calcul de densités a posteriori dans des applications antérieures impliquant des observations latentes ou manquantes. En particulier, ceci résulte du fait que la densité a posteriori peut être écrite comme la densité conjointe a posteriori des données augmentées constituées à la fois de parties observables et non observables du système continu.

L’approximation de diffusion du processus de saut approprié est ensuite utilisée

pour estimer les données manquantes entre chaque paire de temps d'observation. Nous avons adopté le schéma de discrétisation de Milstein au cours du processus d'estimation des paramètres du modèle épidémiologique SIR qui donne des valeurs plus précises que le schéma d'Euler dans le contexte Bayésien tel qu'il est adopté par El Maroufy *et al.* [Maroufy 2015]. De plus, pour une meilleure estimation, nous avons proposé la distribution gamma qui est fréquemment utilisée en pratique comme distribution a priori des paramètres des modèles épidémiques. De plus, nous avons montré l'efficacité des méthodes MCMC dans la simulation de données synthétiques à partir du modèle épidémiologique SIR. Par conséquent, les résultats présentés pour les données simulées et réelles sont assez bons que les autres approches standards.

Cette méthode est donc susceptible d'être applicable à une gamme plus large de modèles et à des ensembles de données plus larges et plus récentes que ce que nous avons considéré ici, nous pouvons considérer des maladies infectieuses de notre temps telles que H1N1, Ebola... . D'autres recherches seront consacrées à l'amélioration de l'inférence paramétrique en considérant une dimension de paramètre plus élevée. De plus, nous allons tenter à la généraliser à des modèles épidémiques plus compliqués. Enfin, le chapitre suivant sera axé sur l'estimation Bayésienne des paramètres β et μ dans le cas dépendant.

Inférence Bayésienne pour le modèle épidémique stochastique SIR avec paramètres dépendants

Sommaire

4.1 Introduction	49
4.2 Inférence Bayésienne pour β et μ	50
4.2.1 Distribution a priori des paramètres β et μ	52
4.2.2 Distribution a posteriori des paramètres β et μ	53
4.2.3 Algorithme	53
4.3 Simulation MCMC	54
4.4 Conclusion	55

4.1 Introduction

Le même principe qu’au chapitre 3, puisque les données ne sont disponibles qu’à des moments discrets, on introduit un ensemble de $m - 1$ données latentes entre chaque paire d’observations. L’idée de relever les données de basse fréquence observées avec des fréquences simulées à haute fréquence a été précédemment poursuivie par Pedersen [Pedersen 1995].

Cette tâche est réalisée aussi par l’application des techniques MCMC qui mettent alternativement à jour les données et sont habituellement réalisables dans un temps de calcul modéré. En outre, nous examinons le même modèle épidémique de type SIR dans lequel le taux d’infection dépend du taux d’élimination. Dans un contexte de modélisation, ce mécanisme d’infection reflète le changement de comportement en réponse à l’état perçu de l’épidémie. En particulier, cette approche s’adresse aux épidémies où le nombre actuel d’infections est inconnu, mais où l’on dispose d’informations sur le nombre d’individus éliminés. Les données concernant les maladies sexuellement transmissibles sont souvent de ce type. Des modèles déterministes pour les maladies sexuellement transmissibles qui tiennent compte du changement de comportement ont été proposés par Blythe, Brauer et Castillo-Chavez [Blythe 1997], Brauer, Blythe, Castillo-Chavez et Velasco-Hernandez [Blythe 1995]. En particulier, ces modèles permettent au taux de recrutement dans la partie sexuellement active

de la population de dépendre de l'état actuel perçu de l'épidémie. Dans plusieurs circonstances, les chercheurs doivent traiter les données dépendantes appariées. De telles cas surviennent généralement dans les essais cliniques impliquant l'analyse de survie et l'hydrologie (flux dans deux districts distincts de la même rivière).

Dans ce chapitre, pour modéliser la dépendance entre les paramètres d'infection et d'élimination dans le modèle épidémique SIR, nous adoptons la distribution gamma bivariée de Kibble (**KBGD**) [Kibble 1941] comme distribution conjointe a priori des paramètres β et μ du modèle. La popularité du KBGD chez les praticiens découle en partie du fait qu'elle permet de modéliser la dépendance (conditionnelle aux caractéristiques observées) entre le taux d'infection et le taux d'élimination dans le modèle épidémique SIR. De plus, nous utilisons la densité donnée dans Risken [Risen 1989] comme approximation de la densité de transition de probabilité du processus. Nous développons également des méthodes MCMC utilisées pour explorer la distribution a posteriori des paramètres du modèle.

Le reste du chapitre est structuré comme suit : dans la Section 4.2 nous fournissons l'inférence Bayésienne des paramètres, la Sous-Section 4.2.1 donne la densité a priori des paramètres, la Sous-Section 4.2.2 établit la densité a posteriori des paramètres et l'algorithme dans la Sous-Section 4.2.3, tandis que les simulations MCMC sont présentées dans la Section 4.3 avec application aux données simulées et on conclue par la Section 4.4. Certaines démonstrations mathématiques sont explorées dans l'Annexe B.

4.2 Inférence Bayésienne pour β et μ

Il existe de nombreuses approches d'inférence utilisées pour obtenir une estimation des paramètres du modèle de diffusion donné par (2.18) à partir de données discrètes avec basse fréquence comme estimation par maximum de vraisemblance et techniques de martingale. Malheureusement, dans de nombreuses applications dans les modèles épidémiques, le temps entre les observations consécutives est assez grand. On ne s'attend pas à ce que ces approches considérées donnent des résultats satisfaisants dans ces cas. Des procédures d'estimation plus avancées peuvent être envisagées pour résoudre ce problème. Nous proposons dans ce chapitre la méthode MCMC pour estimer la diffusion après discrétisation, en utilisant la densité de transition donnée dans Risken [Risen 1989, Section 4.7]. La méthode consiste à augmenter les observations de basse fréquence par l'insertion d'un nombre fini de données latentes entre deux observations consécutives. Afin d'appliquer cette approximation à notre modèle, nous considérons la transformation suivante : $\tilde{\mathbf{y}}_{t_i} = (x_{t_i}, z_{t_i}) = (x_{t_i}, 1 + \frac{a}{n} - x_{t_i} - y_{t_i})$ comme données réelles aux instants $0 = t_0, t_1, \dots, t_N = T$.

Par la formule d'Itô (2.2), on trouve $d\tilde{\mathbf{y}}_{t_i} = U(\tilde{\mathbf{y}}_{t_i}, \theta)dt_i + \sigma(\tilde{\mathbf{y}}_{t_i}, \theta)dW_{t_i}$, où

$$U(\tilde{\mathbf{y}}_{t_i}, \theta) = \begin{pmatrix} -\beta x_{t_i}(1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \\ \mu(1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \end{pmatrix}$$

et

$$\sigma(\tilde{\mathbf{y}}_{t_i}, \theta) = \begin{pmatrix} -\sqrt{\frac{\beta x_{t_i}(1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n}} & 0 \\ 0 & \sqrt{\frac{\mu(1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n}} \end{pmatrix}.$$

L'équation de Fokker-Plank et la formule (4.109) dans Risken [Risken 1989] conduisent à la densité de transition suivante :

$$\begin{aligned} p(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \theta) &= p(x_{t_i}, z_{t_i} | x_{t_{i-1}}, z_{t_{i-1}}, \beta, \mu) \\ &= \left[2\sqrt{\pi\Delta t_i} \right]^{-2} |\Sigma_{i-1}|^{-\frac{1}{2}} e^{\frac{-1}{4\Delta t_i}(\Delta x_{t_i} - \mu_{i-1}, \Delta z_{t_i} - \mu'_{i-1})(\Sigma_{i-1})^{-1} \begin{pmatrix} \Delta x_{t_i} - \mu_{i-1} \\ \Delta z_{t_i} - \mu'_{i-1} \end{pmatrix}} \end{aligned} \quad (4.1)$$

où

$$\Sigma_{i-1} = \begin{pmatrix} \frac{\beta}{n} x_{t_{i-1}} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) & 0 \\ 0 & \frac{\mu}{n} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \end{pmatrix},$$

$\Delta x_{t_i} = x_{t_i} - x_{t_{i-1}}$, $\Delta z_{t_i} = z_{t_i} - z_{t_{i-1}}$, $\mu_{i-1} = -\beta x_{t_{i-1}} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \Delta t_i$ et $\mu'_{i-1} = \mu (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \Delta t_i$.

Puisque Σ_{i-1} est diagonale, on déduit immédiatement de (4.1) que

$$p(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \theta) = p_1(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \beta) \times p_2(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \mu) \quad (4.2)$$

avec

$$\begin{aligned} p_1(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \beta) &= \left[2\sqrt{\pi\Delta t_i} \right]^{-1} e^{-\frac{n}{2}\Delta x_{t_i} \left(\frac{\beta}{n} x_{t_{i-1}} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \right)^{-1/2}} \\ &\quad \times e^{\frac{-n}{4\Delta t_i} \left(\frac{\Delta x_{t_i}^2}{\frac{x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})}{\beta} + x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\Delta t_i^2 \beta} \right)} \end{aligned} \quad (4.3)$$

et

$$\begin{aligned} p_2(\tilde{\mathbf{y}}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}}, \mu) &= \left[2\sqrt{\pi\Delta t_i} \right]^{-1} e^{-\frac{n}{2}\Delta z_{t_i} \left(\frac{\mu}{n} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \right)^{-1/2}} \\ &\quad \times e^{\frac{-n}{4\Delta t_i} \left(\frac{\Delta z_{t_i}^2}{\frac{(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})}{\mu} + (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\Delta t_i^2 \mu} \right)}. \end{aligned} \quad (4.4)$$

4.2.1 Distribution a priori des paramètres β et μ

Les distributions gamma bivariées (BGD) sont de bonnes densités a priori candidates de paramètres, ils fournissent un cadre approprié pour modéliser la dépendance entre les paramètres. Cependant, les extensions définissant des distributions gamma multivariées (MGD) sont plus controversées. Ces distributions ont trouvé des applications utiles dans plusieurs domaines. Par exemple, dans

la modélisation des précipitations à deux pluviomètres à proximité (voir, Izawa [Izawa 1965]), données obtenues à partir d'expériences de pluviométrie (cf. Moran [Moran 1969], [Moran 1970]), la dépendance entre l'écoulement annuel et la précipitation aréolaire étudié par Clarke [Clarke 1980], données de vent considérées par Smith et Adelfang [Smith 1981] et la dépendance entre les précipitations et le ruissellement (cf. Mathai et Moschopoulos [Mathai 1991]). Ils ont également trouvé des applications dans la théorie de la fiabilité, les processus de renouvellement et les problèmes de routage stochastique. Nous utilisons ici la distribution KBGD qui a été introduite par Kibble [Kibble 1941], sa densité conjointe est donnée par :

$$f(x, y | \nu, \lambda_1, \lambda_2, \rho) = \frac{(\lambda_1 \lambda_2)^\nu}{(1-\rho)\Gamma(\nu)} \left(\frac{xy}{\rho \lambda_1 \lambda_2} \right)^{\frac{\nu-1}{2}} e^{-\frac{\lambda_1 x + \lambda_2 y}{1-\rho}} I_{\nu-1} \left(\frac{2\sqrt{\rho \lambda_1 \lambda_2 xy}}{1-\rho} \right) \quad (4.5)$$

où $x, y, \lambda_1, \lambda_2 \geq 0$, $0 \leq \rho < 1$, $\nu \geq 0$ et $I_\nu(\cdot)$ est la fonction de Bessel modifiée du premier type d'ordre ν défini comme $I_\nu(\alpha) = \sum_{k=0}^{\infty} \frac{(\alpha/2)^{2k+\nu}}{\Gamma(k+\nu+1)k!}$, $\alpha \geq 0$. Le lecteur pourra consulter l'ouvrage de Lebedev et Silverman [Lebedev 1972, Chapitre 5] pour un traitement approfondi des fonctions de Bessel.

Dans le présent chapitre, nous supposons que le taux d'infection β et le taux d'élimination μ dans le modèle épidémique SIR suivent comme distribution a priori KBGD (4.5). La corrélation entre β et μ est égale à ρ ; donc β et μ sont indépendants si et seulement si $\rho = 0$. En utilisant la représentation en série de la fonction de Bessel, donc (4.5) prend la forme suivante :

$$f(\beta, \mu | \nu, \lambda_1, \lambda_2, \rho) = \sum_{k=0}^{\infty} f(k|\rho) f(\beta | \nu + k, \frac{\lambda_1}{1-\rho}) f(\mu | \nu + k, \frac{\lambda_2}{1-\rho})$$

où $f(k|\rho) = \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} (1-\rho)^\nu \rho^k$, $k \in \mathbb{N}$, est la probabilité de masse de la distribution binomiale négative $BN(\nu, 1-\rho)$ et $f(\beta | \nu + k, \frac{\lambda_1}{1-\rho})$, $f(\mu | \nu + k, \frac{\lambda_2}{1-\rho})$ sont les fonctions de densité de probabilité des distributions gamma $\Gamma(\nu + k, \frac{\lambda_1}{1-\rho})$, $\Gamma(\nu + k, \frac{\lambda_2}{1-\rho})$ respectivement. Donc, on a

$$\begin{aligned} f(\beta, \mu | \nu, \lambda_1, \lambda_2, \rho) &\propto \sum_{k=0}^{\infty} \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} (1-\rho)^\nu \rho^k \beta^{\nu+k-1} e^{-\frac{\lambda_1 \beta}{1-\rho}} \mu^{\nu+k-1} e^{-\frac{\lambda_2 \mu}{1-\rho}} \\ &\propto (1-\rho)^\nu (\beta \mu)^{\nu-1} e^{-\frac{\lambda_1 \beta + \lambda_2 \mu}{1-\rho}} \sum_{k=0}^{\infty} \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} (\rho \beta \mu)^k. \end{aligned} \quad (4.6)$$

4.2.2 Distribution a posteriori des paramètres β et μ

En raison de la forme de la densité de transition dans (4.2) et avec KBGD comme densité a priori de β et μ , le théorème de Bayes produit la densité conjointe a posteriori de β et μ mixte comme le montre la proposition suivante. La preuve est laissée à l'Annexe B :

Proposition 4.1. *Si β et μ suivent une densité a priori KBGD, alors leur distribution a posteriori prend la forme suivante :*

$$p(\beta, \mu | \tilde{y}) \propto \sum_{k=0}^{\infty} c_k GIG(\beta | \frac{-N}{2} + \nu + k, A, B) GIG(\mu | \frac{-N}{2} + \nu + k, C, D) \quad (4.7)$$

où A, B, C, D et c_k sont données en Annexe B, et $GIG(\cdot | p, a, b)$ est la densité de probabilité de la distribution Gaussienne inverse généralisée de paramètres p, a et b .

4.2.3 Algorithme

- i- Initialiser β et μ en utilisant l'interpolation linéaire entre les valeurs observées de \tilde{y}_i pour initialiser \hat{y}_i .
- ii- Pour tout $i = 1, 2, \dots, N - 1$, à l'itération h simuler $\hat{y}_i^{(h)} | (\hat{y}_{i-1}^{(h-1)}, \hat{y}_{i+1}^{(h-1)}, \beta, \mu)$ en utilisant l'algorithme AR-MH avec densité proposée $\mathcal{N}(\frac{1}{2}(\hat{y}_{i-1}^{(h)} + \hat{y}_{i+1}^{(h)}), \frac{\Delta t}{2} \sum_{i-1}^{(h-1)})$.
- iii- Simuler $\beta^{(h)}$ et $\mu^{(h)}$ en utilisant la densité conjointe (4.7) comme une mixture de deux distributions GIG indépendantes.
- iv- Augmenter la valeur de h et revenir à l'étape ii.

4.3 Simulation MCMC

Afin d'étudier la performance de la méthodologie d'estimation, nous avons d'abord généré des données artificielles. Les données sont échantillonnées à partir de la chaîne de Markov double discrète initiale avec des valeurs réelles de β et μ données dans les TABLES 4.1, 4.2 et 4.3. L'estimation est obtenue après 20000 itérations de notre algorithme pour $\nu = 1$ et $\lambda_1 = \lambda_2 = 2$ et pour trois cas $\rho = 0.1$, $\rho = 0.5$ ou $\rho = 0.9$.

En pratique, on ne peut simuler qu'un nombre fini d'itérations de MCMC, le choix de la condition initiale peut donc influencer la qualité de convergence. Il est alors nécessaire de ne pas tenir compte des premières itérations de l'algorithme, mais il est difficile de déterminer la durée de cette période de "chauffe" ("burn-in period"). En dimension plus grande que 1, il n'existe pas de critère rigoureux de convergence permettant de savoir si la chaîne est proche de son régime asymptotique. Il existe en revanche des techniques empiriques. De plus, la vitesse de convergence de ces algorithmes peut être trop faible dans certaines situations.

Les TABLES 4.1, 4.2 et 4.3 et les FIGURES 4.1, 4.2 et 4.3 montrent que les estimations des paramètres sont généralement beaucoup plus proches de leurs vraies valeurs dans les trois cas $\rho = 0.1$, $\rho = 0.5$ et $\rho = 0.9$, c'est à dire lorsque ρ augmente les deux paramètres β et μ sont fortement corrélées et s'ajustent bien avec les vraies valeurs. Les FIGURES 4.1, 4.2 et 4.3 représentent les histogrammes des densités a posteriori des paramètres, qui révèlent la convergence de l'algorithme vers la distribution limite.

$\rho = 0.1$						
	β	$\hat{\beta}$	μ	$\hat{\mu}$	R_0	\hat{R}_0
Moyenne a posteriori	1	0.9013	1	0.8987	1	1.0028
Erreur standard		0.0987		0.1013		0.0028
Moyenne a posteriori	0.3	0.2041	0.8	0.6893	0.375	0.2961
Erreur standard		0.0959		0.1107		0.0789
Moyenne a posteriori	1.3	1.2037	0.8	0.6902	1.625	1.7439
Erreur standard		0.0963		0.1098		0.1189

TABLE 4.1 – Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.1$.

$\rho = 0.5$						
	β	$\hat{\beta}$	μ	$\hat{\mu}$	R_0	\hat{R}_0
Moyenne a posteriori	1	0.9496	1	0.9415	1	1.0136
Erreur standard		0.0504		0.0585		0.0136
Moyenne a posteriori	0.3	0.2791	0.8	0.7584	0.375	0.3680
Erreur standard		0.0209		0.0416		0.0070
Moyenne a posteriori	1.3	1.2688	0.8	0.7699	1.625	1.6480
Erreur standard		0.0312		0.0301		0.023

TABLE 4.2 – Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.5$.

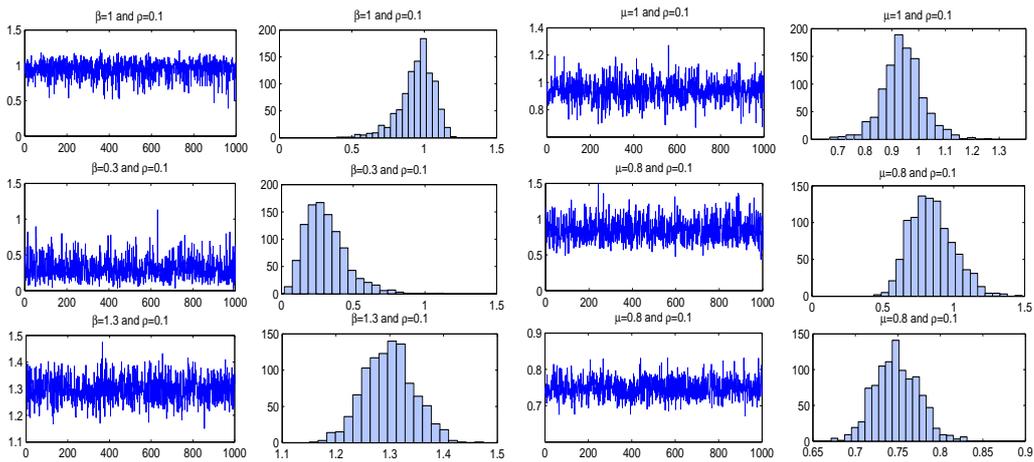
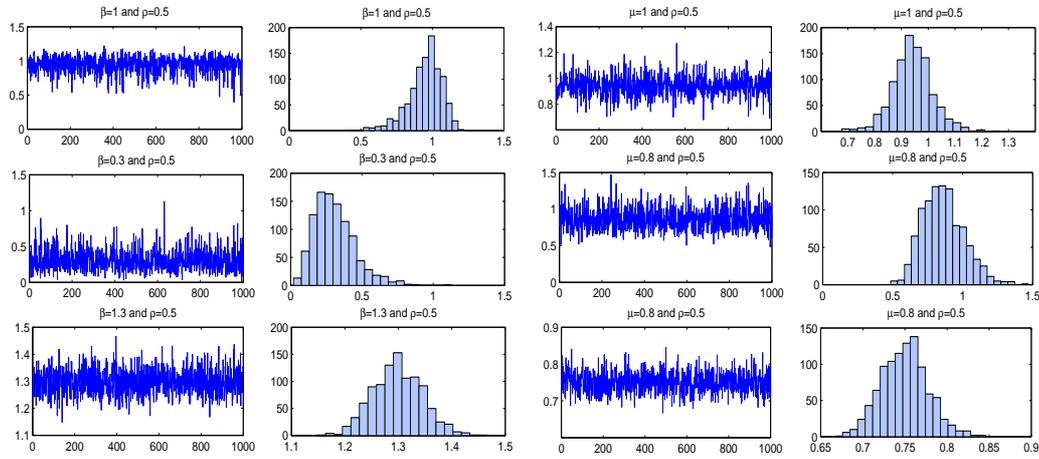


FIGURE 4.1 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.1 avec $\rho = 0.1$ et 20000 itérations dans tous les cas.

$\rho = 0.9$						
	β	$\hat{\beta}$	μ	$\hat{\mu}$	R_0	\hat{R}_0
Moyenne a posteriori	1	0.9828	1	0.9023	1	1.0892
Erreur standard		0.0172		0.0977		0.0892
Moyenne a posteriori	0.3	0.3058	0.8	0.7937	0.375	0.3852
Erreur standard		0.0058		0.0063		0.0102
Moyenne a posteriori	1.3	1.2987	0.8	0.7957	1.625	1.6321
Erreur standard		0.0013		0.0043		0.0071

TABLE 4.3 – Moyenne a posteriori et erreur standard de β , μ et R_0 pour $\rho = 0.9$.FIGURE 4.2 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.2 avec $\rho = 0.5$ et 20000 itérations dans tous les cas.

4.4 Conclusion

Le présent chapitre a décrit l'estimation Bayésienne des paramètres dépendants du modèle épidémique SIR stochastique. L'estimateur proposé dans ce chapitre est encore fondé sur l'inférence basée sur la simulation par les méthodes MCMC. Le cadre d'estimation implique l'introduction d'observations latentes ou manquantes. Nous adoptons l'approximation de la densité de transition donnée dans Risken [Risken 1989] au cours du processus d'estimation des paramètres du modèle épidémique SIR qui donne des valeurs plus précises. De plus, pour une meilleure estimation, nous avons proposé la distribution gamma bivariée de Kibble comme une bonne densité a priori de paramètres, qui peut être écrite sous la forme d'un mélange discret de la distribution binomiale négative et de deux distributions gamma. Ensuite, nous démontrons que la densité a posteriori des paramètres peut être écrite comme un mélange discret de la distribution binomiale négative et de deux distri-

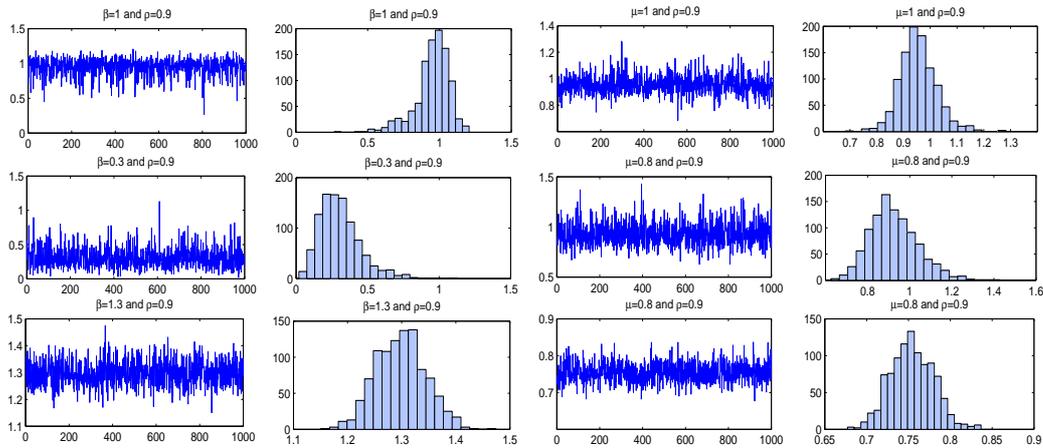


FIGURE 4.3 – Histogramme de fréquences pour les densités a posteriori de β et μ pour les mêmes paramètres exacts dans la TABLE 4.3 avec $\rho = 0.9$ et 20000 itérations dans tous les cas.

butions GIG.

De plus, nous avons montré l'efficacité des méthodes MCMC pour simuler les données synthétiques du modèle épidémique SIR, l'estimation s'améliore au fur et à mesure que ρ augmente. Cela est dû au fait que l'algorithme d'estimation est basé sur une hypothèse de dépendance entre les paramètres.

Cette méthode est donc susceptible d'être appliquée à d'autres modèles épidémiques plus généraux. D'autres recherches seront consacrées à l'amélioration de l'inférence paramétrique en considérant une dimension de paramètre plus élevée et aussi dans le cas où les densités des processus sont hyperboliques. Enfin, le chapitre suivant sera axé sur l'estimation Bayésienne du nombre de reproduction de base R_0 . Ce paramètre a une grande importance, car, si $R_0 \ll 1$, donc une épidémie sera très improbable.

Inférence Bayésienne du nombre de reproduction de base du modèle épidémique SIR

Sommaire

5.1	Introduction	59
5.2	Inférence classique de R_0	61
5.3	Inférence Bayésienne de R_0	63
5.3.1	La distribution a priori de R_0	64
5.3.2	La distribution a posteriori de R_0	65
5.3.3	La distribution a posteriori des données manquantes	65
5.3.4	Algorithme d'estimation de R_0	66
5.4	Simulations MCMC	66
5.5	Application aux données réelles	67
5.6	Conclusion	67

5.1 Introduction

Dans ce dernier chapitre, notre attention porte sur l'inférence Bayésienne du nombre de reproduction de base $R_0 = \frac{\beta}{\mu}$, où β est le taux d'infection et μ est le taux d'élimination. Ce paramètre R_0 est très important dans l'étude de la propagation des épidémies, comme déjà mentionné dans l'introduction de la présente thèse, ce paramètre décrit le nombre moyen de nouvelles infections dues à un individu malade. Comme on peut l'imaginer, si ce nombre est inférieur à 1 alors l'épidémie tendra à s'éteindre, tandis qu'elle pourra persister voire s'étendre à la population entière si $R_0 > 1$.

Pour estimer R_0 , nous avons utilisé deux approches, l'inférence classique et l'inférence Bayésienne. Premièrement, dans l'inférence classique, la densité de transition est rarement connue explicitement, plusieurs approches numériques rendent l'inférence par maximum de vraisemblance possible pour les modèles de diffusion. En fait, Pedersen [Pedersen 1995] a proposé une méthode pour obtenir une approximation de la fonction de vraisemblance par une simulation assez étendue. La méthode de Pedersen a été considérablement améliorée par Durham et Gallant [Durham 2002],

où leur méthode est beaucoup plus efficace. Aussi Poulsen [Poulsen 1999] a obtenu une approximation de la densité de transition en résolvant numériquement une équation différentielle partielle. Pour plus d'informations sur l'estimation des processus de diffusion, le lecteur pourra se référer à Bibby *et al*, pages 203-268 dans l'ouvrage de Ait-Sahalia et Hansen [Ait-Sahalia 2010].

Dans ce chapitre, nous explorons l'EMV de R_0 , et nous prouvons sa convergence. Deuxièmement, dans l'inférence Bayésienne, nous adoptons l'approche proposée dans le chapitre 1, en considérant le modèle épidémique SIR avec R_0 comme paramètre d'intérêt. La méthode consiste toujours à augmenter les observations de basse fréquence par l'insertion de $m - 1$ nombre de données latentes entre deux observations consécutives. Ici, l'objet de l'inférence est R_0 en utilisant la densité de transition d'approximation donnée dans Risken [Riskin 1989], contrairement à l'approximation du schéma d'Euler présentée par El Maroufy *et al*. [Maroufy 2015] ou l'approximation du schéma de Milstein adoptée dans le chapitre 1 (Qaffou *et al*. [Qaffou 2017b]); où l'objet de l'inférence est à la fois les paramètres β et μ séparément. De plus, nous déclarons que la famille des distributions a priori gamma indépendantes conduisent à explorer pour R_0 une distribution a priori *bêta généralisée* (**BG**) du second type et une distribution a posteriori en tant que rapport entre deux distributions GIG indépendantes. Ce choix d'une telle distribution gamma a priori est commode en terme d'inférence Bayésienne pour les modèles de type SIR (cf. O'Neill et Roberts [O'Neill 1999]). De plus, la flexibilité de la distribution gamma la rend très attrayante en pratique, une distribution a priori des paramètres dans les modèles épidémiques (voir Cauchemez *et al*. [Cauchemez 2004], Streftaris et Gibson [Streftaris 2004]). Entre autres choses, nous obtenons une distribution a posteriori des données manquantes.

La suite de ce chapitre est organisée comme suit : dans la Section 5.2 nous effectuons l'inférence classique de R_0 et dans la Section 5.3 nous établissons l'inférence Bayésienne. Dans la Section 5.4, nous illustrons le résultat avec des simulations MCMC ; et l'application aux données réelles de la peste Eyam a été établi dans la Section 5.5. Quelques brèves conclusions sont données à la Section 5.6. Enfin, les preuves de quelques énoncés mathématiques sont données à l'Annexe C.

5.2 Inférence classique de R_0

Nous considérons toujours le modèle décrit par l'équation différentielle stochastique (3.3) avec des coefficients non linéaires. La principale difficulté dans le contexte de l'estimation directe du paramètre $\theta = (\beta, \mu)$ est que la densité de transition du processus $\tilde{\mathbf{y}} = (x, y)$ est inconnue. Par conséquent, la fonction de vraisemblance est également inconnue et EMV qui est connu pour avoir les bonnes propriétés habituelles n'est pas une solution (cf. Dacunha-Castelle [Dacunha-Castelle 1986] et Florens-Zmirou [Florens-Zmirou 1989]). Une façon courante pour surmonter cette difficulté est de fonder l'inférence sur les observations discrètes du processus. Si nous disposons des observations à des moments discrets $0 = t_0, t_1, \dots, t_N = T$, nous

pouvons utiliser une approximation de la densité de transition inconnue $p(\tilde{\mathbf{y}}_i/\tilde{\mathbf{y}}_{i-1}, \theta)$ de $\tilde{\mathbf{y}} = (x, y)$ par une séquence de densités $\tilde{p}(\tilde{\mathbf{y}}_i/\tilde{\mathbf{y}}_{i-1}, \theta)$ qui convergent vers $p(\tilde{\mathbf{y}}_i/\tilde{\mathbf{y}}_{i-1}, \theta)$ quand $\max_{1 \leq i \leq N} |t_i - t_{i-1}| \rightarrow 0$ et ensuite nous définissons la fonction log-vraisemblance approximative

$$l(\theta) = \sum_{i=1}^N \ln \tilde{p}(\tilde{\mathbf{y}}_i/\tilde{\mathbf{y}}_{i-1}, \theta).$$

Pour ce faire, nous utiliserons la densité donnée dans Risken [Risken 1989, Section 4.7] comme approximation analytique de p . Pour appliquer cette approximation à notre modèle, considérons la transformation suivante : $\tilde{\mathbf{y}}_{t_i} = (x_{t_i}, z_{t_i}) = (x_{t_i}, 1 + \frac{a}{n} - x_{t_i} - y_{t_i})$ comme données réelles aux instants $0 = t_0, t_1, \dots, t_N = T$.

Avec la formule d'Itô (2.2), on trouve $d\tilde{\mathbf{y}}_{t_i} = U(\tilde{\mathbf{y}}_{t_i}, \theta)dt_i + \sigma(\tilde{\mathbf{y}}_{t_i}, \theta)dW_{t_i}$, où

$$U(\tilde{\mathbf{y}}_{t_i}, \theta) = \begin{pmatrix} -\beta x_{t_i}(1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \\ \mu(1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \end{pmatrix}$$

et

$$\sigma(\tilde{\mathbf{y}}_{t_i}, \theta) = \begin{pmatrix} -\sqrt{\frac{\beta x_{t_i}(1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n}} & 0 \\ 0 & \sqrt{\frac{\mu(1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n}} \end{pmatrix}.$$

L'équation de Fokker Plank et la formule (4.109) donnée à Risken [Risken 1989] (voir le chapitre précédent, Section 4.2) conduit à la densité de transition suivante :

$$\begin{aligned} \tilde{p}(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \theta) &= \tilde{p}(x_{t_i}, z_{t_i}/x_{t_{i-1}}, z_{t_{i-1}}, \beta, \mu) \\ &= \left(2\sqrt{\pi\Delta t_i}\right)^{-2} |\Sigma_{i-1}|^{-\frac{1}{2}} e^{\frac{-1}{4\Delta t_i}(\Delta x_{t_i} - \mu_{i-1}, \Delta z_{t_i} - \mu'_{i-1})\Sigma_{i-1}^{-1}(\Delta x_{t_i} - \mu_{i-1}, \Delta z_{t_i} - \mu'_{i-1})} \end{aligned} \quad (5.1)$$

où

$$\Sigma_{i-1} = \begin{pmatrix} \frac{\beta}{n}x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) & 0 \\ 0 & \frac{\mu}{n}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}) \end{pmatrix},$$

$\Delta x_{t_i} = x_{t_i} - x_{t_{i-1}}$, $\Delta z_{t_i} = z_{t_i} - z_{t_{i-1}}$, $\mu_{i-1} = -\beta x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\Delta t_i$ et $\mu'_{i-1} = \mu(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\Delta t_i$.

Puisque Σ_{i-1} est diagonale, on déduit immédiatement de (5.1) que

$$\tilde{p}(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \theta) = \tilde{p}_1(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \beta) \times \tilde{p}_2(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \mu), \quad (5.2)$$

où

$$\begin{aligned} \tilde{p}_1(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \beta) &= \left(2\sqrt{\pi\Delta t_i}\right)^{-1} e^{-\frac{n}{2}\Delta x_{t_i}} \left(\frac{\beta}{n}x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\right)^{-\frac{1}{2}} \\ &\quad \times e^{\frac{-n}{4\Delta t_i} \left(\frac{\Delta x_{t_i}^2}{\frac{x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})}{\beta} + x_{t_{i-1}}(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})\Delta t_i^2 \beta} \right)} \end{aligned}$$

et

$$\begin{aligned} \tilde{p}_2(\tilde{\mathbf{Y}}_{t_i}/\tilde{\mathbf{Y}}_{t_{i-1}}, \mu) &= \left(2\sqrt{\pi\Delta t_i}\right)^{-1} e^{-\frac{n}{2}\Delta z_{t_i}} \left(\frac{\mu}{n}\left(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}\right)\right)^{-\frac{1}{2}} \\ &\quad \times e^{\frac{-n}{4\Delta t_i} \left(\frac{\Delta z_{t_i}^2}{\left(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}\right)\mu} + \left(1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}\right)\Delta t_i^2 \mu \right)}. \end{aligned}$$

Après un calcul fastidieux, la log-vraisemblance peut être simplement réécrite comme suit :

$$\begin{aligned} l(\theta) &= l(\beta, \mu) = l_1(\beta) + l_2(\mu) \\ &\propto -\frac{N}{2} \ln(\beta) - \frac{1}{4} \left(\frac{A}{\beta} + C\beta \right) - \frac{N}{2} \ln(\mu) - \frac{1}{4} \left(\frac{B}{\mu} + D\mu \right), \end{aligned} \quad (5.3)$$

où A, B, C et D sont des quantités données dans l'Annexe C.

Avec des calculs simples, égalant le gradient à zéro, $\partial_\theta l(\beta, \mu) = 0$, on obtient

$$\hat{\beta} = \frac{\sqrt{N^2 + AC} - N}{C} \quad \text{et} \quad \hat{\mu} = \frac{\sqrt{N^2 + BD} - N}{D}.$$

Donc, l'EMV de R_0 est $\hat{R}_0 = \frac{\hat{\beta}}{\hat{\mu}}$.

Les estimateurs $\hat{\beta}$ et $\hat{\mu}$ sont consistents et asymptotiquement normaux sous conditions appropriées. La forme (5.3) de la fonction contraste nous permet d'investir le Théorème 1 dans Sørensen et Uchida [Sørensen 2003]. Quand on prend les mêmes hypothèses pour vérifier la consistance de \hat{R}_0 , il suffit de montrer que les cinq hypothèses (A₁)-(A₅) données dans Sørensen et Uchida [Sørensen 2003] sont remplies. On peut se référer à Andersson et Britton [Andersson 2000] pour vérifier (A₁). U et σ sont respectivement polynomiaux et quadratiques en x, z, β et μ ; et $x, z \in [0, 1]$. Par conséquent, les hypothèses (A₂)-(A₄) sont vérifiées. (A₅) est déduit immédiatement de (5.2). Alors $\hat{\theta} = (\hat{\beta}, \hat{\mu})$ est asymptotiquement consistant et possède la propriété de normalité.

Considérons maintenant \hat{R}_0 , nous avons par le développement de Taylor :

$$\hat{R}_0 - R_0 = \nabla_\theta \left(\frac{\beta}{\mu} \right)' (\hat{\theta} - \theta) + o(\|\hat{\theta} - \theta\|),$$

et par l'inégalité de Cauchy-Schwartz, on obtient

$$|\hat{R}_0 - R_0| \leq \left| \nabla_\theta \left(\frac{\beta}{\mu} \right)' (\hat{\theta} - \theta) \right| + |o(\|\hat{\theta} - \theta\|)| \leq \left\| \nabla_\theta \left(\frac{\beta}{\mu} \right)' \right\| \|\hat{\theta} - \theta\| + |o(\|\hat{\theta} - \theta\|)|.$$

Puisque $\|\hat{\theta} - \theta\| \rightarrow 0$ en probabilité quand le nombre d'observation est plus grand et $\left\| \nabla_\theta \left(\frac{\beta}{\mu} \right)' \right\| < +\infty$, donc $|\hat{R}_0 - R_0| \rightarrow 0$ en probabilité. Finalement, \hat{R}_0 est consistant, en plus (voir Rao [Rao 1973, p. 388]) on vérifie que :

$$\widehat{\text{Var}}(\hat{R}_0) \approx \frac{\hat{\beta}^2}{\hat{\mu}^4} \text{Var}(\hat{\mu}) + \frac{1}{\hat{\mu}^2} \text{Var}(\hat{\beta}) - 2 \frac{\hat{\beta}}{\hat{\mu}^3} \text{Cov}(\hat{\beta}, \hat{\mu}). \quad (5.4)$$

Cette variance est investit pour établir l'intervalle de confiance de R_0 .

5.3 Inférence Bayésienne de R_0

La dynamique de la transmission de personne à personne qui concerne la transmission vectorielle peut être résumée par le taux de reproduction de base R_0 qui est considéré comme un paramètre d'intérêt du modèle épidémique SIR, qui s'interprète comme le nombre moyen de nouveaux cas générés par un sujet infectieux dans une population susceptible. Ce paramètre est très important dans l'étude de la propagation des épidémies, il permet d'évaluer la virulence d'une maladie infectieuse. Plus la valeur de $R_0 > 1$ est grande, plus l'épidémie est importante et peut s'étendre. Notre objectif est de développer une méthodologie plus avancée pour l'estimation de R_0 en utilisant une inférence Bayésienne. L'estimation de R_0 , ou des paramètres équivalents dans des modèles plus complexes peut généralement être obtenus via les méthodes MCMC. Nous proposons la méthode MCMC en utilisant la densité de transition donnée dans Risken [Risken 1989, Section 4.7]. La méthode consiste à augmenter les observations de basse fréquence par l'insertion d'un nombre fini de données latentes entre deux observations consécutives. Parmi d'autres choses, on obtient une expression de la densité a posteriori de R_0 .

5.3.1 La distribution a priori de R_0

On suppose que β et μ suivent une distribution a priori gamma indépendante, respectivement $\Gamma(m_\beta, \lambda_\beta)$ et $\Gamma(m_\mu, \lambda_\mu)$. De plus, la flexibilité de la distribution gamma signifie qu'il est fréquemment utilisée en pratique comme distribution a priori des paramètres de modèles épidémiques (voir Cauchemez *et al.* [Cauchemez 2004], Streftaris et Gibson [Streftaris 2004]). Cependant, l'introduction de la distribution a priori pour R_0 est rarement mentionnée dans la littérature, et nous considérons maintenant ceci pour le modèle actuel.

À partir de Clancy et O'Neill [Clancy 2008], nous vérifions¹ que la distribution a priori de $R_0 = \frac{\beta}{\mu}$ est donnée par une distribution bêta généralisée de second type $GB_2(1, \frac{\lambda_\mu}{\lambda_\beta}, m_\beta, m_\mu)$ (voir, McDonald [McDonald 1984]) :

$$\pi(R_0) = \frac{\frac{\lambda_\beta}{\lambda_\mu}}{\mathcal{B}(m_\beta, m_\mu)} \frac{\left(\frac{\lambda_\beta}{\lambda_\mu} R_0\right)^{m_\beta-1}}{\left(1 + \frac{\lambda_\beta}{\lambda_\mu} R_0\right)^{m_\beta+m_\mu}}, \quad (5.5)$$

pour $R_0 > 0$ où $\mathcal{B}(m_\beta, m_\mu)$ est la fonction bêta. La moyenne et la variance a priori sont données par

$$\mathbb{E}(R_0) = \frac{m_\beta \lambda_\mu}{(m_\mu - 1) \lambda_\beta}, \quad \text{et} \quad \text{Var}(R_0) = \frac{m_\beta (m_\beta + m_\mu - 1)}{(m_\beta - 1)^2 (m_\mu - 2)} \left(\frac{\lambda_\beta}{\lambda_\mu}\right)^2. \quad (5.6)$$

La question de savoir si $R_0 > 1$ est souvent intéressante. Le choix des paramètres m_β , m_μ , λ_β et λ_μ est fixé tel que $\mathbb{P}(R_0 > 1) = 0.5$. Cette égalité signifie que R_0 a

1. La preuve est classique, elle se base sur le théorème de transfert appliqué aux vecteurs aléatoires.

une médiane égale à 1, i.e :

$$\frac{\frac{\lambda_\beta}{\lambda_\mu}}{\mathcal{B}(m_\beta, m_\mu)} \int_0^1 \frac{\left(\frac{\lambda_\beta}{\lambda_\mu} R_0\right)^{m_\beta-1}}{\left(1 + \frac{\lambda_\beta}{\lambda_\mu} R_0\right)^{m_\beta+m_\mu}} dR_0 = 0.5, \quad (5.7)$$

avec la transformation $u = \frac{\frac{\lambda_\beta}{\lambda_\mu} R_0}{1 + \frac{\lambda_\beta}{\lambda_\mu} R_0}$, (5.7) devient

$$\frac{1}{\mathcal{B}(m_\beta, m_\mu)} \int_0^{\frac{\lambda_\beta}{\lambda_\beta+\lambda_\mu}} u^{m_\beta-1} (1-u)^{m_\mu-1} du = 0.5.$$

Donc, le choix des paramètres m_β , m_μ , λ_β et λ_μ est prouvé telle que la médiane de $\mathcal{B}(m_\beta, m_\mu)$ est égale à $\frac{\lambda_\beta}{\lambda_\beta+\lambda_\mu}$. Les valeurs initiales de m_β et m_μ peuvent être données en utilisant l'estimation de R_0 combinée avec les équations (5.4) et (5.6). Pour λ_β , nous pouvons choisir en simulant la médiane de $\mathcal{B}(m_\beta, m_\mu)$ pour différentes valeurs de λ_β . Les valeurs de λ_β qui sont sur la droite $y = \lambda_\beta$ sont les valeurs initiales désirées de λ_β .

5.3.2 La distribution a posteriori de R_0

En raison de la forme de la densité de (5.2) et avec GB_2 en (5.5) comme distribution a priori de R_0 , Le théorème de Bayes produit une distribution a posteriori de R_0 comme le rapport de deux distributions GIG indépendantes. Plus formellement, nous avons le résultat suivant dont la preuve est laissée à l'Annexe C :

Proposition 5.1. *Si $R_0 \sim GB_2(1, \frac{\lambda_\mu}{\lambda_\beta}, m_\beta, m_\mu)$, alors sa distribution a posteriori est donnée par le rapport entre deux distributions GIG indépendantes :*

$$\pi(R_0/\tilde{y}) \sim \frac{GIG_{\left(\frac{-N}{2}+m_\beta, \frac{A}{2}, \frac{4\lambda_\beta+C}{2}\right)}}{GIG_{\left(\frac{-N}{2}+m_\mu, \frac{B}{2}, \frac{4\lambda_\mu+D}{2}\right)}}, \text{ pour } R_0 \geq 0 \quad (5.8)$$

où la densité de $GIG_{(p,a,b)}$ est $f(x,p,a,b) = \frac{(\frac{b}{a})^{\frac{p}{2}}}{2\Gamma_p(\sqrt{ab})} x^{p-1} e^{-\frac{1}{2}\left(\frac{a}{x}+bx\right)} \chi_{\{x>0\}}$, $a > 0, b > 0$ et $p \in \mathbb{R}$.

5.3.3 La distribution a posteriori des données manquantes

Pour les processus de diffusion univariés, la méthode MCMC a été étudiée de manière approfondie, par exemple dans les modèles financiers. Roberts et Stramer [Roberts 2001], Elerian [Elerian 1998] et Durham et Gallant [Durham 2002] utilisent des schémas de mise à jour de blocs pour simuler les données latentes. Pour notre modèle bivarié partiellement observé, en raison de la grande dimensionnalité (grand nombre de données manquantes et de paramètres), il est pratique d'utiliser l'échantillonneur de Gibbs présenté dans la Sous-Section 2.10.2, la première étape de cet échantillonnage implique la simulation des données manquantes. Nous considérons

ici le cas dans lequel l'échantillonneur de Gibbs met à jour une seule colonne de $\hat{\mathbf{y}}$ dans le temps avec $\hat{\mathbf{y}}_{t_i} = (\hat{x}_{t_i}, \hat{z}_{t_i})$ sont les données manquantes. À l'itération h de l'échantillonneur de Gibbs, on peut simuler $\hat{\mathbf{y}}_{t_i} \sim \pi(\hat{\mathbf{y}}_{t_i} / \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}})$. Pour le modèle non linéaire, Eraker [Eraker 2001] suggère que $\hat{\mathbf{y}}_{t_i}$ soit mise à jour avec une densité de proposition normale $q(\cdot / \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}})$ avec moyenne $\frac{1}{2}(\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}})$ et variance $\frac{\Delta t}{2}\sigma^2(\hat{\mathbf{y}}_{t_{i-1}})$. Les auteurs justifient ce choix en prouvant que, si Δt est petit, la densité de $\hat{\mathbf{y}}_{t_i} / \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}$ est près de $\mathcal{N}\left(\frac{1}{2}(\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}}), \frac{\Delta t}{2}\sigma^2(\hat{\mathbf{y}}_{t_{i-1}})\right)$. Dans notre cas, le résultat suivant, prouvé dans l'Annexe C, est obtenu.

Proposition 5.2. *Si la distribution a priori de R_0 est une bêta généralisée de second degré $GB_2(1, \frac{\lambda\mu}{\lambda\beta}, m_\beta, m_\mu)$, donc, pour Δt suffisamment petit, on a approximativement :*

$$p(\hat{\mathbf{y}}_{t_i} / \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}, R_0) \propto h_{t_i}^{\frac{m_\beta + m_\mu - 1.5}{2}} e^{-\theta\sqrt{h_{t_i}}}, \quad (5.9)$$

où h_{t_i} et θ sont données en Annexe C.

Cette distribution de probabilité nous permet de générer les données manquantes.

5.3.4 Algorithme d'estimation de R_0

L'algorithme est basé sur les distributions a posteriori calculées dans la proposition 5.1 et 5.2 :

- i- Initialiser toutes les données manquantes entre chaque paire d'observations en utilisant une interpolation linéaire et initialiser R_0 selon le principe de la Section 5.3.1.
- ii- A l'itération h on utilise l'algorithme AR-MH pour extraire l'observation latente en utilisant la densité conjointe donnée par (5.9).
- iii- Simuler R_0 en utilisant (5.8).
- iv- Augmenter la valeur de h et revenir à l'étape ii.

5.4 Simulations MCMC

Pour illustrer la méthodologie présentée dans ce chapitre, nous considérons le modèle épidémique SIR avec des paramètres connus.

Pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m \in \{5; 10; 15\}$, $S = 3000$ (Susceptibles), $I = 200$ (Infectées), $N = 10$ (taille d'échantillon) et 1000 simulations. L'échantillon observé est généré en utilisant l'algorithme de Gillespie donné dans la Section 2.11.

Les TABLES 5.1, 5.2 et 5.3 et les FIGURES 5.1, 5.2 et 5.3 montrent que l'estimation de R_0 est généralement beaucoup plus proche de la vraie valeur, la méthode MCMC fournit des résultats au moins aussi bons que la méthode MV quand $R_0 < 1$

valeurs de R_0	$\hat{R}_0(\text{MCMC})$	sd(MCMC)	$\hat{R}_0(\text{MV})$	sd(MV)
0.5	0.4833	0.0166	0.4998	0.0001
1	1.0112	0.0111	1.0128	0.0127
1.5	1.5660	0.0659	1.5881	0.0880
10	9.9635	0.0364	10.7223	0.7222

TABLE 5.1 – Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=5$.

valeurs de R_0	$\hat{R}_0(\text{MCMC})$	sd(MCMC)	$\hat{R}_0(\text{MV})$	sd(MV)
0.5	0.4715	0.0282	0.4995	0.0004
1	1.0095	0.0094	1.0165	0.0164
1.5	1.5621	0.0621	1.5841	0.0840
10	9.9479	0.0520	10.7234	0.7234

TABLE 5.2 – Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=10$.

(l'épidémie est faible). Quand l'épidémie est forte, c'est à dire $R_0 \geq 1$, l'estimation par MCMC proposé dans ce chapitre surperforme l'estimateur par MV (par exemple, pour $R_0 = 10$ et $m = 15$, $\text{sd}(\text{MCMC})=0.0661$ et $\text{sd}(\text{MV})=0.6856$). Les FIGURES 5.1, 5.2 et 5.3 illustrent les histogrammes de la densité a posteriori du paramètre R_0 qui révèlent la convergence de l'algorithme vers la distribution limite du paramètre R_0 en (5.8).

5.5 Application aux données réelles

Reprenons le même exemple introduit dans la Sous-Section 3.4.5, l'algorithme MCMC fournit l'estimation $\hat{R}_0 = 1.51$ tandis que l'EMV donne la valeur $\hat{R}_0 = 1.57$, avec les données manquantes $m = 15$. Les erreurs moyennes calculées sur la base de l'échantillon observé montrent l'efficacité de l'estimateur MCMC par rapport à l'EMV, (voir, la TABLE (5.4)).

5.6 Conclusion

Dans ce chapitre, nous avons adopté la densité de transition d'approximation donnée dans Risken [Risken 1989, p. 86, formula 4.109] pour estimer le paramètre R_0 pour le modèle épidémique SIR qui donne des valeurs plus précises dans le contexte Bayésien. Nous avons considéré la famille de distribution a priori gamma indépendante qui fournit une distribution a posteriori de R_0 comme un rapport entre deux distributions GIG. Le résultat obtenu montre que la valeur ajoutée d'estimation de R_0 , est que la méthode MV fournit des estimations assez bonne que la

valeurs de R_0	$\hat{R}_0(\text{MCMC})$	sd(MCMC)	$\hat{R}_0(\text{MV})$	sd(MV)
0.5	0.4723	0.0276	0.5011	0.0011
1	1.0079	0.0078	1.0170	0.0169
1.5	1.5666	0.0666	1.5949	0.0948
10	9.9339	0.0661	10.6856	0.6856

TABLE 5.3 – Estimation de R_0 et erreurs standard (sd) avec méthodes MCMC et MV pour $m=15$.

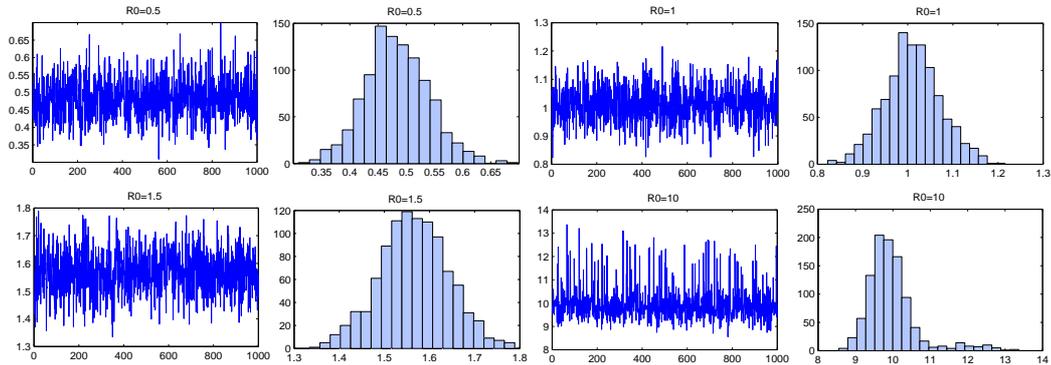


FIGURE 5.1 – Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 5$.

méthode MCMC quand l'épidémie est faible, c'est à dire $R_0 < 1$. Cependant, quand l'épidémie est sévère ($R_0 \geq 1$) l'estimateur MCMC est plus efficace que l'EMV. Nous avons également montré l'efficacité de la méthode MCMC en simulant les données synthétiques et réelles du modèle épidémique SIR. Ensuite, quand une épidémie est forte, nous pouvons utiliser la méthode MCMC comme une bonne méthode pour estimer R_0 .

Cette méthode est donc susceptible d'être appliquée à une gamme plus large de modèles et à des ensembles de données plus larges. Ce que nous avons considéré dans ce chapitre. D'autres recherches seront consacrées à l'amélioration de l'inférence paramétrique en considérant une dimension de paramètre plus élevée.

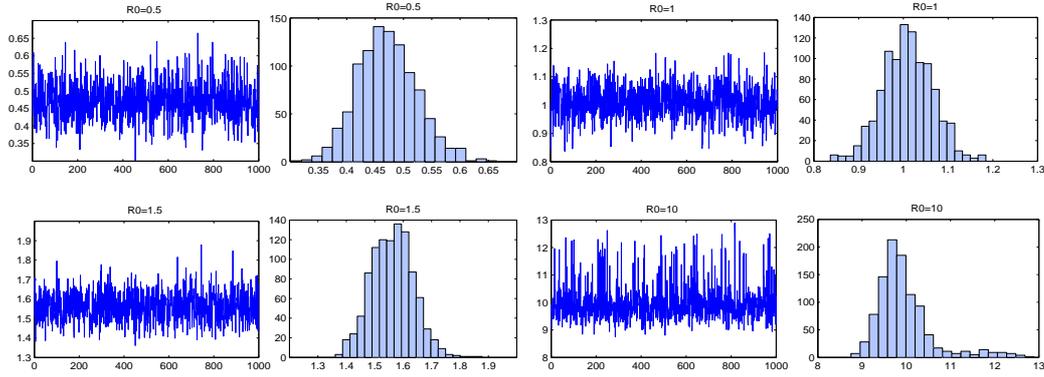


FIGURE 5.2 – Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 10$.

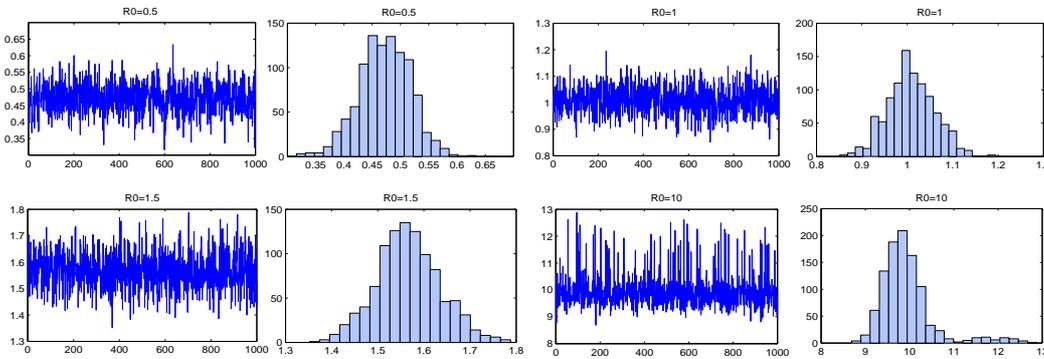


FIGURE 5.3 – Estimations MCMC et histogrammes pour $R_0 \in \{0.5; 1; 1.5; 10\}$, $m = 15$.

	Données réelles		Estimation MCMC		Estimation MV	
	S(t)	I(t)	S(t)	I(t)	S(t)	I(t)
Juillet 3/4	235	14.5	229.29	15.15	212.06	19.93
Juillet 19	201	22	189.14	24.27	131.88	28.81
Août 3/4	153.5	29	146.79	25.83	81.16	7.32
Août 19	121	20	117.10	19.13	74.66	0.36
Septembre 3/4	108	8	100.84	11.27	74.41	0.02
Septembre 19	97	8	92.83	5.87	74.52	0.00
Octobre 4/5	inconnu	inconnu	89.01	2.89	74.57	0.00
Octobre 20	83	0	87.21	1.38	74.59	0.02
	Erreur moyenne		6.59	2.06	44.47	11.59

TABLE 5.4 – Nombre d'individus susceptibles et infectés aux dates de la période terminale de la peste d'Eyam et leurs estimations selon la méthode MCMC et la méthode MV avec les données manquantes $m = 15$.

Conclusion et perspectives

Au terme de cette thèse, nous estimons que les résultats présentés contribueront au développement des méthodes d'inférence statistique Bayésiennes pour l'estimation paramétrique des modèles épidémiques stochastique dans une population fermée et divisée en trois classes d'individus, les susceptibles, les infectés et les éliminés, en ouvrant de nouveaux horizons à la recherche scientifique sur cette thématique émergente.

Après avoir présenté les notions préliminaires et les ingrédients utiles pour la bonne compréhension du présent manuscrit, nous avons présenté des méthodes Bayésiennes pour l'estimation paramétrique du taux d'infection β et du taux d'élimination μ de dits modèles ainsi que du taux de reproduction de base R_0 .

Tout d'abord, nous avons établi des résultats d'estimation Bayésienne du taux d'infection et du taux d'élimination dans le cas où ces deux paramètres ont des lois a priori indépendantes en utilisant les techniques de Monte Carlo par les Chaînes de Markov, est ce en augmentant les données de faibles fréquences par l'introduction de données latentes entre chaque paire d'observations, simulées en utilisant la méthode standard de l'approximation des équations différentielles stochastiques par le schéma de Milstein ; qui donne des approximations plus précises que le schéma d'Euler dans le contexte Bayésien tel qu'il est adopté par El Maroufy *et al.* [Maroufy 2015]. Les résultats ont été pertinents, et nous avons pu remarquer par simulations que les estimations sont très proches des valeurs exactes à partir des quelles nos observations ont été générées, ces simulations acquises par la méthode MCMC ont bien justifiées les aboutissements théoriques. Nous avons prouvé que la loi instrumentale investie pour reproduire les données manquantes ce n'est que la loi normale. De plus, pour une meilleure modélisation du problème, nous avons proposé la distribution gamma, qui est fréquemment utilisée en pratique, comme distribution a priori des paramètres et nous avons montré que les lois a posteriori obtenues pour les paramètres sont des lois particulières ; les lois inverses Gaussiennes généralisées. Les résultats obtenus ont fait l'objet d'une publication (cf. [Qaffou 2017b]).

Par ailleurs, nous avons généralisé les résultats susmentionnés au cas où les deux lois a priori des deux paramètres sont dépendantes. Dans ce cas, nous avons adopté l'approximation de la densité de transition établie par Risken [Riskin 1989] au cours du processus d'estimation des paramètres. De plus, nous avons proposé la distribution gamma bivariée de Kibble comme une véritable densité conjointe a priori ; nous avons montré que la distribution conjointe a posteriori des paramètres ce n'est qu'un mélange discret de la distribution binomiale négative et de deux distributions inverse Gaussiennes généralisées. Les résultats obtenus sont soumis pour publication dans

le journal "Journal of Statistical Computation and Simulation" [Qaffou 2017c].

Enfin, nous avons clôturé ce travail de thèse par l'estimation Bayésienne du nombre de reproduction de base R_0 . Nous avons adopté la densité de transition approchée donnée dans Risken [Risen 1989, formule 4.109]. Nous avons considéré la loi bêta généralisée de second type comme loi a priori pour R_0 qui a abouti à une distribution a posteriori de R_0 comme étant un rapport entre deux distributions inverses Gaussiennes généralisées. Les résultats obtenus montrent que la valeur ajoutée de l'estimation directe de R_0 réside dans le fait que l'algorithme consomme assez moins de temps que les approches adoptées dans les travaux précités. La méthode MCMC a fourni des estimations aussi bonne que la méthode classique basée sur le maximum de vraisemblance lorsque l'épidémie est faible, *i.e.* $R_0 < 1$, cependant pour $R_0 \geq 1$, l'estimation par la méthode MCMC est plus efficace et consistente que l'estimation par le maximum de vraisemblance. Les aboutissements seront incessamment publiés dans la revue "Brazilian Journal of Probability and Statistics" [Qaffou 2017a].

Les résultats présentés dans cette thèse offrent naturellement de nombreuses perspectives. La première est l'adaptation de la méthode Bayésienne établie ici à des modèles épidémiques multivariés de type SIR et l'amélioration de l'inférence paramétrique en considérant une dimension de paramètre plus élevée. Il serait aussi intéressant de voir le comportement de nos modèles et faisant varier la population totale car dans la réalité la démographie n'est pas constante. Aussi, avec une bonne collaboration avec les professionnels de la santé publique, nous pensons pouvoir concevoir, avec nos modèles des outils pour fournir des informations importantes en matière de santé publique pour le contrôle des épidémies.

La deuxième perspective envisageable serait l'étude théorique de la comparaison de ces nouveaux estimateurs développés ici et les estimateurs existants et application aux données **partielles** réelles des épidémies actuelles comme H1N1, Ebola, HIV,...etc.

Finalement, une perspective qui semble être, une continuité logique de ce travail, est de voir l'estimation de R_0 en utilisant une distribution a priori de R_0 convenable dans le cas où les deux paramètres β et μ sont supposés des variables aléatoires dépendantes.

A.1 Preuve de l'équation (3.7) et l'équation (3.8)

Combinant (3.2) et (3.5), on obtient

$$p_{x_{i+1}|\widehat{\mathbf{Y}}_i}(h_{i,1}) = \frac{1}{2|b_{i,1}|} \left(\frac{h_{i,1}}{\lambda_{i,1}b_{i,1}} + 1 \right)^{-\frac{1}{4}} e^{-\left(\lambda_{i,1} + \frac{h_{i,1}}{2b_{i,1}}\right)} I_{-\frac{1}{2}} \left(\sqrt{\lambda_{i,1} \left(\frac{h_{i,1}}{b_{i,1}} + \lambda_{i,1} \right)} \right),$$

et

$$p_{z_{i+1}|\widehat{\mathbf{Y}}_i}(h_{i,2}) = \frac{1}{2|b_{i,2}|} \left(\frac{h_{i,2}}{\lambda_{i,2}b_{i,2}} + 1 \right)^{-\frac{1}{4}} e^{-\left(\lambda_{i,2} + \frac{h_{i,2}}{2b_{i,2}}\right)} I_{-\frac{1}{2}} \left(\sqrt{\lambda_{i,2} \left(\frac{h_{i,2}}{b_{i,2}} + \lambda_{i,2} \right)} \right)$$

avec $h_{i,1} = \Delta x_i + \beta x_i(1 + \frac{a}{n} - x_i - z_i)\Delta t + b_{i,1}$

et $h_{i,2} = \Delta z_i - \mu(1 + \frac{a}{n} - x_i - z_i)\Delta t + b_{i,2}$,

donc

$$p_{x_{i+1}|\widehat{\mathbf{Y}}_i}(h_{i,1}) \propto \frac{\beta^{-\frac{1}{2}} \tilde{\lambda}_{i,1}^{-\frac{1}{2}}}{|\tilde{b}_{i,1}|} \left(\frac{h_{i,1}}{\lambda_{i,1}b_{i,1}} + 1 \right)^{-\frac{1}{2}} e^{-\left(\lambda_{i,1} + \frac{h_{i,1}}{2b_{i,1}}\right)} \times \cosh \left(\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1} \right). \quad (\text{A.1})$$

Puisque $\left(\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1 \right)^{-\frac{1}{2}} = \left(\frac{\Delta x_i}{b_{i,1}\lambda_{i,1}} - \frac{\mu_i \Delta t}{b_{i,1}\lambda_{i,1}} + \frac{\beta}{\lambda_{i,1}} + 1 \right)^{-\frac{1}{2}}$ et $\tilde{b}_{i,1}\tilde{\lambda}_{i,1}$ est indépendant de β et Δt , pour Δt suffisamment petit et le fait que $(1-h)^{-\frac{1}{2}} \simeq e^{\frac{h}{2}}$, on obtient

$$\begin{aligned} \left(\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1 \right)^{-\frac{1}{2}} &\simeq \left(\frac{\Delta x_i}{\tilde{b}_{i,1}\tilde{\lambda}_{i,1}} + 1 - \frac{\beta}{\tilde{\lambda}_{i,1}} \right)^{-\frac{1}{2}} \simeq \left(\frac{\Delta x_i}{\tilde{b}_{i,1}\tilde{\lambda}_{i,1}} + 1 \right)^{-\frac{1}{2}} \left(1 + \frac{\beta}{\frac{\Delta x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1}} \right)^{-\frac{1}{2}} \\ &\simeq \left(\frac{\Delta x_i}{\tilde{b}_{i,1}\tilde{\lambda}_{i,1}} + 1 \right)^{-\frac{1}{2}} e^{\frac{-\beta}{2\left(\frac{\Delta x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1}\right)}}, \end{aligned}$$

c'est à dire $\left(\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1 \right)^{-\frac{1}{2}} \simeq u_i e^{\frac{-\beta}{2\left(\frac{\Delta x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1}\right)}}$ avec $u_i = \left(\frac{\Delta x_i}{\tilde{b}_{i,1}\tilde{\lambda}_{i,1}} + 1 \right)^{-\frac{1}{2}}$.

Puisque $\lambda_{i,1} + \frac{h_{i,1}}{2b_{i,1}} = \frac{\tilde{\lambda}_{i,1}}{\beta} + \frac{\Delta x_i}{2\beta\tilde{b}_{i,1}} + \frac{x_i(1 + \frac{a}{n} - x_i - z_i)\Delta t}{2b_{i,1}} + \frac{1}{2}$,

donc $e^{-\left(\lambda_{i,1} + \frac{h_{i,1}}{2b_{i,1}}\right)} \simeq e^{-1/2} e^{-\frac{\left(\tilde{\lambda}_{i,1} + \frac{\Delta x_i}{2\tilde{b}_{i,1}}\right)}{\beta}}$, pour Δt suffisamment petit.

On considère maintenant le terme $\cosh \left(\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1} \right)$, il est facile de vérifier que

$$\cosh \left(\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1} \right) = \frac{e^{\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1}} + e^{-\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1}}}{2} \simeq \frac{e^{\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1}}}{2}.$$

$$\text{en plus } \lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1} \simeq \frac{\left(\tilde{\lambda}_{i,1} \left(\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)\right)^{\frac{1}{2}}}{\beta} + \frac{\tilde{\lambda}_{i,1}^{\frac{1}{2}}}{2 \left(\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)^{\frac{1}{2}}},$$

$$\text{donc } \cosh\left(\lambda_{i,1} \sqrt{\frac{h_{i,1}}{b_{i,1}\lambda_{i,1}} + 1}\right) \simeq v_i e^{\frac{\left(\frac{\tilde{\lambda}_{i,1} \Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}^2\right)^{\frac{1}{2}}}{\beta}} \text{ avec } v_i = \frac{1}{2} e^{\frac{\tilde{\lambda}_{i,1}^{\frac{1}{2}}}{2 \left(\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)^{\frac{1}{2}}}}.$$

En remplaçant tous les termes dans (A.1), finalement, on obtient

$$p_{x_{i+1}|\hat{\mathbf{Y}}_i}(h_{i,1}) \simeq A_i \beta^{-\frac{1}{2}} e^{\frac{-\beta}{2 \left(\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)} - \frac{\left(\frac{\Delta x_i}{2b_{i,1}} + \tilde{\lambda}_{i,1}\right) - \left(\frac{\tilde{\lambda}_{i,1} \Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}^2\right)^{1/2}}{\beta}}$$

$$\text{avec } A_i = \frac{1}{|b_{i,1}|} \left(\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)^{-\frac{1}{2}} e^{\frac{1}{2} \left(\frac{\tilde{\lambda}_{i,1}}{\frac{\Delta x_i}{b_{i,1}} + \tilde{\lambda}_{i,1}}\right)^{\frac{1}{2}}}.$$

Avec la même technique, nous montrons que

$$p_{z_{i+1}|\hat{\mathbf{Y}}_i}(h_{i,2}) \simeq B_i \mu^{-\frac{1}{2}} e^{\frac{-\mu}{2 \left(\frac{\Delta z_i}{b_{i,2}} + \tilde{\lambda}_{i,2}\right)} - \frac{\left(\frac{\Delta z_i}{2b_{i,2}} + \tilde{\lambda}_{i,2}\right) - \left(\frac{\tilde{\lambda}_{i,2} \Delta z_i}{b_{i,2}} + \tilde{\lambda}_{i,2}^2\right)^{1/2}}{\mu}},$$

$$\text{où } B_i = \frac{1}{|b_{i,1}|} \left(\frac{\Delta z_i}{b_{i,1}} + \tilde{\lambda}_{i,1}\right)^{-\frac{1}{2}} e^{\frac{1}{2} \left(\frac{\tilde{\lambda}_{i,1}}{\frac{\Delta z_i}{b_{i,1}} + \tilde{\lambda}_{i,1}}\right)^{\frac{1}{2}}}.$$

A.2 Preuve de la Proposition 3.1

Soit $\phi_k(\hat{\mathbf{Y}}_t, \theta) = \frac{1}{2} \sigma_k^{1/2}(\hat{\mathbf{Y}}_t, \theta) \frac{\partial \sigma_k^{1/2}(\hat{\mathbf{Y}}_t, \theta)}{\partial \hat{\mathbf{Y}}_t^k}$, pour $k \in \{1, 2\}$. Selon (3.5), on a

$$\hat{Y}_i^k = \hat{Y}_{i-1}^k + \mu_k(\hat{\mathbf{Y}}_{i-1}, \theta) \Delta t + \sigma_k^{1/2}(\hat{\mathbf{Y}}_{i-1}, \theta) \xi_{i-1} \sqrt{\Delta t} + \phi_k(\hat{\mathbf{Y}}_{i-1}, \theta) (\xi_{i-1}^2 - 1) \Delta t, \quad (\text{A.2})$$

et

$$\hat{Y}_{i+1}^k = \hat{Y}_i^k + \mu_k(\hat{\mathbf{Y}}_i, \theta) \Delta t + \sigma_k^{1/2}(\hat{\mathbf{Y}}_i, \theta) \xi_i \sqrt{\Delta t} + \phi_k(\hat{\mathbf{Y}}_i, \theta) (\xi_i^2 - 1) \Delta t, \quad (\text{A.3})$$

où $\xi_i \sim \mathcal{N}(0, 1)$. En combinant (A.2) et (A.3), on a

$$\begin{aligned} \frac{\hat{Y}_i^k - \frac{\hat{Y}_{i-1}^k + \hat{Y}_{i+1}^k}{2}}{\sqrt{\frac{\Delta t}{2}} \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta)} &= \underbrace{\sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) [\mu_k(\hat{Y}_{i-1}, \theta) - \mu_k(\hat{Y}_i, \theta)]}_{A_{\Delta t}} \sqrt{\frac{\Delta t}{2}} \\ &+ \underbrace{\frac{\sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta)}{\sqrt{2}} [\sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) \xi_{i-1} - \sigma_k^{1/2}(\hat{Y}_i, \theta) \xi_i]}_{B_{\Delta t}} \\ &+ \underbrace{\sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) [\phi_k(\hat{Y}_{i-1}, \theta)(\xi_{i-1}^2 - 1) - \phi_k(\hat{Y}_i, \theta)(\xi_i^2 - 1)]}_{C_{\Delta t}} \sqrt{\frac{\Delta t}{2}}. \end{aligned} \quad (\text{A.4})$$

De (A.2), on déduit que

$$\frac{\hat{Y}_i^k - \hat{Y}_{i-1}^k}{\sqrt{\Delta t}} = \mu_k(\hat{Y}_{i-1}, \theta) \sqrt{\Delta t} + \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) \xi_{i-1} + \phi_k(\hat{Y}_{i-1}, \theta) (\xi_{i-1}^2 - 1) \sqrt{\Delta t}, \quad (\text{A.5})$$

$$\mu_k(\hat{Y}_i, \theta) = \mu_k(\hat{Y}_{i-1}, \theta) + (\hat{Y}_i^k - \hat{Y}_{i-1}^k) \frac{\partial \mu_k}{\partial x}(\zeta, \theta), \text{ o}\tilde{\text{A}}^1 \zeta \in (\hat{Y}_{i-1}^k, \hat{Y}_i^k), \quad (\text{A.6})$$

et

$$\sigma_k^{1/2}(\hat{Y}_i, \theta) = \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) + (\hat{Y}_i^k - \hat{Y}_{i-1}^k) \frac{\partial \sigma_k^{1/2}}{\partial x}(\zeta, \theta), \text{ o}\tilde{\text{A}}^1 \zeta \in (\hat{Y}_{i-1}^k, \hat{Y}_i^k). \quad (\text{A.7})$$

Donc (A.5) et (A.6) implique que pour Δt suffisamment petit et $\zeta \in (\hat{Y}_{i-1}^k, \hat{Y}_i^k)$

$$\begin{aligned} A_{\Delta t} &= -\sqrt{\frac{\Delta t}{2}} \sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) \left[\frac{\hat{Y}_i^k - \hat{Y}_{i-1}^k}{\sqrt{\Delta t}} \frac{\partial \mu_k}{\partial x}(\zeta, \theta) \right] \sqrt{\Delta t} \\ &= -\sqrt{\frac{\Delta t}{2}} \sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) \left[\mu_k(\hat{Y}_{i-1}, \theta) \sqrt{\Delta t} + \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) \xi_{i-1} \right. \\ &\quad \left. + \sqrt{\Delta t} \phi_k(\hat{Y}_{i-1}, \theta) (\xi_{i-1}^2 - 1) \right] \frac{\partial \mu_k}{\partial x}(x, \theta)|_{x=\zeta} \sqrt{\Delta t} \\ &= \left[-\frac{\Delta t}{\sqrt{2}} \sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) \mu_k(\hat{Y}_{i-1}, \theta) - \sqrt{\frac{\Delta t}{2}} \xi_{i-1} - \sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) \right. \\ &\quad \left. \frac{\Delta t}{\sqrt{2}} \phi_k(\hat{Y}_{i-1}, \theta) (\xi_{i-1}^2 - 1) \right] \frac{\partial \mu_k}{\partial x}(x, \theta)|_{x=\zeta} \sqrt{\Delta t}. \end{aligned}$$

Par les dérivatives limites de la fonction dérive et la fonction de diffusion, on obtient

$$\begin{aligned}
(\mathbb{E}[A_{\Delta t}]^2)^{1/2} &\leq \left[\frac{(\Delta t)^{3/2}}{\sqrt{2}} \left| \sigma_k^{-1/2}(\hat{\mathbf{Y}}_{i-1}, \theta) \mu_k(\hat{\mathbf{Y}}_{i-1}, \theta) \right| + \frac{\Delta t}{\sqrt{2}} (\mathbb{E}[\xi_{i-1}]^2)^{1/2} \right. \\
&\quad \left. + \frac{(\Delta t)^{3/2}}{2\sqrt{2}} \frac{\partial \sigma_k^{1/2}(\hat{\mathbf{Y}}_{i-1}, \theta)}{\partial \hat{\mathbf{Y}}_{i-1}} (\mathbb{E}[\xi_{i-1}^2 - 1])^{1/2} \right] \frac{\partial \mu_k}{\partial x}(x, \theta)|_{x=\zeta} \\
&\leq \left[\frac{(\Delta t)^{3/2}}{\sqrt{2}} \left| c(\hat{\mathbf{Y}}_{i-1}, \theta) \right| + \frac{\Delta t}{\sqrt{2}} + \frac{(\Delta t)^{3/2}}{2\sqrt{2}} M_1 (\mathbb{E}[\xi_{i-1}^2 - 1])^{1/2} \right] M_0 \\
&\leq \left[\frac{(\Delta t)^{3/2}}{\sqrt{2}} \left| c(\hat{\mathbf{Y}}_{i-1}, \theta) \right| + \frac{\Delta t}{\sqrt{2}} + \frac{(\Delta t)^{3/2}}{2\sqrt{2}} M_1 \sqrt{2} \right] M_0,
\end{aligned}$$

où $M_0 = \sup \frac{\partial \mu_k}{\partial x}(x, \theta)|_{x=\zeta}$, $M_1 = \sup \frac{\partial \sigma_k^{1/2}(\hat{\mathbf{Y}}_{i-1}, \theta)}{\partial \hat{\mathbf{Y}}_{i-1}}|_{x=\zeta}$,

et $c(\hat{\mathbf{Y}}_{i-1}, \theta) = \sigma_k^{-1/2}(\hat{\mathbf{Y}}_{i-1}, \theta) \mu_k(\hat{\mathbf{Y}}_{i-1}, \theta)$.

Finalement quand $\Delta t \rightarrow 0$, on a $A_{\Delta t} \rightarrow 0$ en moyenne quadratique.

Même arguments appliqués à $B_{\Delta t}$, en utilisant (A.7), on a

$$B_{\Delta t} = \frac{1}{\sqrt{2}}(\xi_{i-1} - \xi_i) - \frac{\xi_i}{\sqrt{2}}(\hat{Y}_i^k - \hat{Y}_{i-1}^k) \frac{\partial \sigma_k^{1/2}(x, \theta)}{\partial x}|_{x=\zeta} \sigma_k^{-1/2}(\hat{\mathbf{Y}}_{i-1}, \theta).$$

Donc (A.5) donne

$$\begin{aligned}
B_{\Delta t} - \frac{1}{\sqrt{2}}(\xi_{i-1} - \xi_i) &= -\sqrt{\frac{\Delta t}{2}} \xi_i \frac{\partial \sigma_k^{1/2}(x, \theta)}{\partial x}|_{x=\zeta} \\
&\quad \times \left[c(\hat{\mathbf{Y}}_{i-1}, \theta) \sqrt{\Delta t} + \xi_{i-1} + \frac{1}{2} \frac{\partial \sigma_k^{1/2}(x, \theta)}{\partial x}|_{x=\zeta} (\xi_{i-1}^2 - 1) \sqrt{\Delta t} \right].
\end{aligned}$$

Alors

$$\begin{aligned}
\left[\mathbb{E} \left(B_{\Delta t} - \frac{1}{\sqrt{2}}(\xi_{i-1} - \xi_i) \right)^2 \right]^{1/2} &\leq \frac{\Delta t}{\sqrt{2}} M_1 c(\hat{\mathbf{Y}}_{i-1}, \theta) (\mathbb{E}[\xi_i]^2)^{1/2} + \sqrt{\frac{\Delta t}{2}} M_1 (\mathbb{E}[\xi_{i-1}^2])^{1/2} \\
&\quad + \frac{\sqrt{\Delta t}}{2\sqrt{2}} M_1^2 (\mathbb{E}[\xi_i(\xi_{i-1}^2 - 1)])^{1/2}, \\
&\leq \frac{\Delta t}{\sqrt{2}} M_1 c(\hat{\mathbf{Y}}_{i-1}, \theta) + \sqrt{\frac{\Delta t}{2}} M_1 \sqrt{3} + \frac{\Delta t}{2\sqrt{2}} M_1^2 \sqrt{10}.
\end{aligned}$$

Ça donne $\left[\mathbb{E} \left(B_{\Delta t} - \frac{1}{\sqrt{2}}(\xi_{i-1} - \xi_i) \right)^2 \right]^{1/2} \rightarrow 0$, avec $B_{\Delta t} \rightarrow \mathcal{N}(0, 1)$ en moyenne

quadratique, car $\frac{1}{\sqrt{2}}(\xi_{i-1} - \xi_i) \rightarrow \mathcal{N}(0, 1)$.

En procédant de la même façon pour $C_{\Delta t}$:

$$C_{\Delta t} = \frac{\sqrt{\Delta t}}{2\sqrt{2}} \frac{\partial \sigma_k^{1/2}(\hat{\mathbf{Y}}_{i-1}, \theta)}{\partial \hat{Y}_{i-1}^k} (\xi_{i-1}^2 - 1) - \underbrace{\frac{\sqrt{\Delta t}}{2\sqrt{2}} \frac{\partial \sigma_k^{1/2}(\hat{\mathbf{Y}}_i, \theta)}{\partial \hat{Y}_i^k} \sigma_k^{1/2}(\hat{\mathbf{Y}}_i, \theta) \sigma_k^{-1/2}(\hat{\mathbf{Y}}_{i-1}, \theta) (\xi_i^2 - 1)}_{T_i^k}.$$

En utilisant (A.7) et (A.5), il est simple de voir que

$$\begin{aligned}
T_i^k &= \frac{\sqrt{\Delta t}}{2\sqrt{2}} \frac{\partial \sigma_k^{1/2}(\hat{Y}_i, \theta)}{\partial \hat{Y}_i^k} \left[\sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) + \left(\mu_k(\hat{Y}_{i-1}, \theta) \sqrt{\Delta t} \right. \right. \\
&\quad \left. \left. + \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) \xi_{i-1} + \frac{\sqrt{\Delta t}}{2} \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta) \frac{\partial \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta)}{\partial \hat{Y}_{i-1}^k} (\xi_{i-1}^2 - 1) \right) \right. \\
&\quad \left. \frac{\partial \sigma_k^{1/2}(x, \theta)}{\partial x} \Big|_{x=\zeta} \sqrt{\Delta t} \right] \sigma_k^{-1/2}(\hat{Y}_{i-1}, \theta) (\xi_i^2 - 1),
\end{aligned}$$

donc

$$\begin{aligned}
(\mathbb{E}[T_i^k]^2)^{1/2} &\leq \frac{\sqrt{\Delta t}}{2\sqrt{2}} M_1 \left[1 + c(\hat{Y}_{i-1}, \theta) M_1 \Delta t + \sqrt{\Delta t} M_1 (\mathbb{E}(\xi_{i-1}^2))^{1/2} \right. \\
&\quad \left. + \frac{\Delta t}{2} M_1^2 (\mathbb{E}(\xi_{i-1}^2 - 1)^2)^{1/2} \right] (\mathbb{E}(\xi_i^2 - 1)^2)^{1/2} \\
&\leq \frac{\sqrt{\Delta t}}{2\sqrt{2}} M_1 \left[1 + c(\hat{Y}_{i-1}, \theta) M_1 \Delta t + \sqrt{\Delta t} M_1 + \frac{\Delta t}{2} M_1^2 \sqrt{2} \right] \sqrt{2} \\
&\leq \frac{\sqrt{\Delta t}}{2} M_1 \left[1 + c(\hat{Y}_{i-1}, \theta) M_1 \Delta t + \sqrt{\Delta t} M_1 + \frac{\Delta t}{2} M_1^2 \sqrt{2} \right],
\end{aligned}$$

alors

$$(\mathbb{E}[C_{\Delta t}]^2)^{1/2} \leq \frac{\sqrt{\Delta t}}{2\sqrt{2}} M_1 \sqrt{2} + \frac{\sqrt{\Delta t}}{2} M_1 \left[1 + c(\hat{Y}_{i-1}, \theta) M_1 \Delta t + \sqrt{\Delta t} M_1 + \frac{\Delta t}{2} M_1^2 \sqrt{2} \right]$$

et il suit que $C_{\Delta t} \rightarrow 0$, en moyenne quadratique.

Finalement, à partir des analyses précédentes (voir, Gut [Gut 2005, Theorem

11.1]) on déduit que $\frac{\hat{Y}_i^k - \frac{\hat{Y}_{i-1}^k + \hat{Y}_{i+1}^k}{2}}{\sqrt{\frac{\Delta t}{2} \sigma_k^{1/2}(\hat{Y}_{i-1}, \theta)}} = A_{\Delta t} + B_{\Delta t} + C_{\Delta t} \sim \mathcal{N}(0, 1)$ quand Δt est suffisamment petit.

A.3 Preuve de la Proposition 3.2

Supposons que $\pi(\beta) \propto \beta^{m_\beta-1} e^{-\lambda_\beta \beta}$ pour $\beta \in \mathbb{R}_+^*$, substituons (3.7) dans (3.6) implique

$$\begin{aligned} \pi(\beta \mid \widehat{\mathbf{Y}}) &\propto \prod_{i=1}^{N-1} p_{x_{i+1} \mid \widehat{\mathbf{Y}}_i}(x_{i+1}) \pi(\beta) \\ &\propto \prod_{i=1}^{N-1} A_i \beta^{-\frac{1}{2}} e^{\frac{-\beta}{2 \left(\frac{\Delta x_i + \tilde{\lambda}_{i,1}}{\tilde{b}_{i,1}} \right)} - \left(\frac{\tilde{\lambda}_{i,1} \Delta x_i + \tilde{\lambda}_{i,1}^2}{\tilde{b}_{i,1}} \right)^{1/2}} \beta^{m_\beta-1} e^{-\lambda_\beta \beta}, \\ &\propto \beta^{-\frac{N-1}{2}} \beta^{m_\beta-1} \\ &\quad \times e^{-\frac{1}{2} \left[\left(\sum_{i=1}^{N-1} \frac{1}{\left(\frac{\Delta x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1} \right)} + 2\lambda_\beta \right) \beta + \frac{1}{\beta} \left(2 \sum_{i=1}^{N-1} \left(\frac{\Delta x_i}{2\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1} \right) - \left(\frac{\tilde{\lambda}_{i,1} \Delta x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1}^2 \right)^{1/2} \right) \right]}, \\ &\propto \beta^{-\frac{N-1}{2} + m_\beta - 1} e^{-\frac{1}{2} \left[A_1 \beta + \frac{B_1}{\beta} \right]} \end{aligned}$$

avec

$$\begin{cases} A_1 = 2\lambda_\beta + \sum_{i=1}^{N-1} \frac{1}{\left(\frac{x_{i+1} - x_i}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1} \right)} \\ B_1 = 2 \sum_{i=1}^{N-1} \left[\left(\frac{x_{i+1} - x_i}{2\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1} \right) - \left(\frac{\tilde{\lambda}_{i,1}(x_{i+1} - x_i)}{\tilde{b}_{i,1}} + \tilde{\lambda}_{i,1}^2 \right)^{1/2} \right]. \end{cases}$$

Encore de la même manière, si $\pi(\mu) \propto \mu^{m_\mu-1} e^{-\lambda_\mu \mu}$ pour $\mu \in \mathbb{R}_+^*$, la distribution a posteriori de μ est : $\pi(\mu \mid \widehat{\mathbf{Y}}) \propto \mu^{-\frac{N-1}{2} + m_\mu - 1} e^{-\frac{1}{2} \left(A_2 \mu + \frac{B_2}{\mu} \right)}$, avec

$$\begin{cases} A_2 = 2\lambda_\mu + \sum_{i=1}^{N-1} \frac{1}{\left(\frac{z_{i+1} - z_i}{\tilde{b}_{i,2}} + \tilde{\lambda}_{i,2} \right)} \\ B_2 = 2 \sum_{i=1}^{N-1} \left[\left(\frac{z_{i+1} - z_i}{2\tilde{b}_{i,2}} + \tilde{\lambda}_{i,2} \right) - \left(\frac{\tilde{\lambda}_{i,2}(z_{i+1} - z_i)}{\tilde{b}_{i,2}} + \tilde{\lambda}_{i,2}^2 \right)^{1/2} \right]. \end{cases}$$

Ça complète la démonstration de la proposition 3.2.

B.1 Preuve de la Proposition 4.1

Soit $\tilde{y}_{t_0}, \dots, \tilde{y}_{t_N}$ un échantillon aléatoire de distribué selon la loi (4.5), la densité a posteriori jointe de $\theta = (\beta, \mu)$ est donnée par $p(\theta|\tilde{y}) \propto \prod_{i=1}^N p(\tilde{y}_{t_i}|\tilde{y}_{t_{i-1}}, \theta)\pi(\theta)$ où $p(\tilde{y}_{t_i}|\tilde{y}_{t_{i-1}}, \theta)$ est la densité de transition donnée à (4.2) et $\pi(\theta)$ est la densité a priori de $\theta = (\beta, \mu)$ donnée par (4.6), prenons $y_{t_{i-1}} = 1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}$ dans (4.3) et (4.4), et enlevant tous les termes indépendant de β et μ , on obtient

$$\begin{aligned} p(\theta|\tilde{y}) &\propto \prod_{i=1}^N p_1(\tilde{y}_{t_i}|\tilde{y}_{t_{i-1}}, \beta)p_2(\tilde{y}_{t_i}|\tilde{y}_{t_{i-1}}, \mu)p(\theta) \\ &\propto \prod_{i=1}^N \beta^{-1/2} e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta x_{t_i}^2}{\frac{x_{t_{i-1}} y_{t_{i-1}}}{\beta} + x_{t_{i-1}} y_{t_{i-1}} \Delta t_i^2 \beta} \right)} \mu^{-1/2} e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta z_{t_i}^2}{\frac{y_{t_{i-1}}}{\mu} + y_{t_{i-1}} \Delta t_i^2 \mu} \right)} \\ &\quad \times (1-\rho)^\nu (\beta\mu)^{\nu-1} e^{-\frac{\lambda_1\beta + \lambda_2\mu}{1-\rho}} \sum_{k=0}^{\infty} \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} (\rho\beta\mu)^k. \end{aligned}$$

En prenant :

$$A = \sum_{i=1}^N \frac{n\Delta x_{t_i}^2}{2\Delta t_i x_{t_{i-1}} y_{t_{i-1}}}, \quad B = \sum_{i=1}^N \frac{n}{2} x_{t_{i-1}} y_{t_{i-1}} \Delta t_i + \frac{2\lambda_1}{1-\rho}, \quad C = \sum_{i=1}^N \frac{n\Delta z_{t_i}^2}{2\Delta t_i y_{t_{i-1}}}$$

$$\text{et } D = \sum_{i=1}^N \frac{n}{2} \Delta t_i y_{t_{i-1}} + \frac{2\lambda_2}{1-\rho}.$$

Après simplification, on obtient

$$p(\theta|\tilde{y}) \propto \sum_{k=0}^{\infty} \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} \rho^k \beta^{-\frac{N}{2} + \nu + k - 1} e^{-\frac{1}{2}(\frac{A}{\beta} + B\beta)} \mu^{-\frac{N}{2} + \nu + k - 1} e^{-\frac{1}{2}(\frac{C}{\mu} + D\mu)}.$$

Approximativement, on peut écrire

$$p(\theta|\tilde{y}) \propto \sum_{k=0}^{\infty} \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} \rho^k c_k^1 c_k^2 GIG(\beta | -\frac{N}{2} + \nu + k, A, B) GIG(\mu | -\frac{N}{2} + \nu + k, C, D),$$

$$\text{avec } c_k^1 = \frac{2I_{-\frac{N}{2} + \nu + k}(\sqrt{AB})}{\left(\sqrt{\frac{B}{A}}\right)^{-\frac{N}{2} + \nu + k}} \text{ et } c_k^2 = \frac{2I_{-\frac{N}{2} + \nu + k}(\sqrt{CD})}{\left(\sqrt{\frac{D}{C}}\right)^{-\frac{N}{2} + \nu + k}}$$

sont les constantes de normalisation inverse de $GIG(\beta | -\frac{N}{2} + \nu + k, A, B)$ et $GIG(\mu | -\frac{N}{2} + \nu + k, C, D)$ respectivement. $I_p(\cdot)$ est la fonction de Bessel de second degré d'ordre p .

Quand $\max(\Delta t_i) \rightarrow 0$ on a $AB \rightarrow +\infty$ et $CD \rightarrow +\infty$; prenant le développement asymptotique suivant de I : $I_p(x) \sim_{+\infty} \sqrt{\frac{\pi}{2}} x^{-1/2} e^{-x} (1 + \frac{4p^2-1}{8x} + \dots)$, limitant

au premier terme ($I_p(x) \sim_{+\infty} \sqrt{\frac{\pi}{2x}} e^{-x}$), donc on obtient $c_k^1 \sim \frac{2\sqrt{\frac{\pi}{2\sqrt{AB}}} e^{-\sqrt{AB}}}{(\sqrt{\frac{B}{A}})^{-\frac{N}{2} + \nu + k}}$ et

$$c_k^2 \sim \frac{2\sqrt{\frac{\pi}{2\sqrt{CD}}} e^{-\sqrt{CD}}}{(\sqrt{\frac{D}{C}})^{-\frac{N}{2} + \nu + k}}, \text{ on a aussi } c_k^1 c_k^2 \propto \left(\sqrt{\frac{AC}{BD}} \right)^k.$$

Donc

$$p(\theta|\tilde{y}) \propto \sum_{k=0}^{\infty} f(k|\tilde{\rho}) GIG(\beta | -\frac{N}{2} + \nu + k, A, B) GIG(\mu | -\frac{N}{2} + \nu + k, C, D),$$

où $f(k|\tilde{\rho}) = \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k)} (1-\tilde{\rho})^\nu \tilde{\rho}^k$, $\tilde{\rho} = \sqrt{\frac{AC}{BD}} \rho$, $k \in \mathbb{N}$, est la fonction probabilité de masse de la distribution binomiale négative $BN(\nu, 1-\tilde{\rho})$ et $GIG(\beta | -\frac{N}{2} + \nu + k, A, B)$, $GIG(\mu | -\frac{N}{2} + \nu + k, C, D)$ sont deux distribution GIG.

C.1 Preuve de la Proposition 5.1

En utilisant la densité de transition approximative donné par Risken [Risken 1989], on trouve à partir de (5.1)

$$\begin{aligned} \tilde{p}(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \theta) &\propto \beta^{-1/2} \mu^{-1/2} \\ &\times e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta x_{t_i}^2}{\frac{x_{t_{i-1}}(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})}{\beta} + x_{t_{i-1}}(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})\Delta t_i^2 \beta} \right)} \\ &\times e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta z_{t_i}^2}{\frac{(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})}{\mu} + (1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})\Delta t_i^2 \mu} \right)}. \end{aligned}$$

En remplaçant seulement les termes qui dépendent de $R_0 = \frac{\beta}{\mu}$ et μ , on obtient :

$$\begin{aligned} \tilde{p}(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, \theta) &\propto (R_0\mu)^{-1/2} \mu^{-1/2} \\ &\times e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta x_{t_i}^2}{\frac{x_{t_{i-1}}(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})}{R_0\mu} + x_{t_{i-1}}(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})\Delta t_i^2 R_0\mu} \right)} \\ &\times e^{-\frac{n}{4\Delta t_i} \left(\frac{\Delta z_{t_i}^2}{\frac{(1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})}{\mu} + (1+\frac{a}{n}-x_{t_{i-1}}-z_{t_{i-1}})\Delta t_i^2 \mu} \right)}. \end{aligned}$$

Puisque $\pi(R_0, \mu/\tilde{\mathbf{y}}) \propto p(\tilde{\mathbf{y}}/R_0, \mu)\pi(R_0, \mu) \propto \prod_{i=1}^N \tilde{p}(\tilde{\mathbf{y}}_{t_i}/\tilde{\mathbf{y}}_{t_{i-1}}, R_0, \mu)\pi(R_0, \mu)$, où $\pi(R_0, \mu)$ est la densité jointe de R_0 et μ , donc

$$\pi(R_0, \mu/\tilde{\mathbf{y}}) \propto R_0^{-N/2} \mu^{-N} e^{-\frac{1}{4}(\frac{A}{R_0\mu} + CR_0\mu + \frac{B}{\mu} + D\mu)} (R_0\mu)^{m\beta-1} e^{-\lambda\beta R_0\mu} \mu^{m\mu} e^{-\lambda\mu},$$

avec

$$\begin{aligned} A &= \sum_{i=1}^N \frac{n\Delta x_{t_i}^2}{\Delta t_i x_{t_{i-1}} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})}, B = \sum_{i=1}^N \frac{n\Delta z_{t_i}^2}{\Delta t_i (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}})}, \\ C &= \sum_{i=1}^N n\Delta t_i x_{t_{i-1}} (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}), D = \sum_{i=1}^N n\Delta t_i (1 + \frac{a}{n} - x_{t_{i-1}} - z_{t_{i-1}}). \end{aligned}$$

Donc

$$\pi(R_0, \mu/\tilde{y}) \propto R_0^{-\frac{N}{2}+m_\beta-1} \mu^{-N+m_\beta+m_\mu-1} e^{-\frac{1}{4} \left(\frac{A+B}{\mu} + ((4\lambda_\beta+C)R_0+4\lambda_\mu+D)\mu \right)}.$$

On intègre par rapport à μ , on obtient :

$$\pi(R_0/\tilde{y}) \propto R_0^{-\frac{N}{2}+m_\beta-1} \int_0^{+\infty} \mu^{-\frac{N}{2}+m_\beta-\frac{N}{2}+m_\mu-1} e^{-\frac{1}{2} \left(\frac{A/2+B/2}{\mu} + \left(\frac{4\lambda_\beta+C}{2} R_0 + \frac{4\lambda_\mu+D}{2} \right) \mu \right)} d\mu. \quad (C.1)$$

D'après un résultat de Jorgensen [Jorgensen 1982], on a :

si $X \sim GIG_{p,a,b}$ et $Y \sim GIG_{p',a',b'}$ alors la densité du rapport $\frac{X}{Y}$ est $f_{\frac{X}{Y}}(z) \propto$

$$z^{p-1} \int_0^{+\infty} \nu^{p+p'-1} e^{-\frac{1}{2} \left(\frac{a+a'}{\nu} + (bz+b')\nu \right)} d\nu.$$

En prenant $(p, a, b) = (\frac{-N}{2} + m_\beta, \frac{A}{2}, \frac{4\lambda_\beta+C}{2})$ et $(p', a', b') = (\frac{-N}{2} + m_\mu, \frac{B}{2}, \frac{4\lambda_\mu+D}{2})$, on déduit de (C.1) que $\pi(R_0/\tilde{y})$ est le rapport entre $GIG_{(\frac{-N}{2}+m_\beta, \frac{A}{2}, \frac{4\lambda_\beta+C}{2})}$ et $GIG_{(\frac{-N}{2}+m_\mu, \frac{B}{2}, \frac{4\lambda_\mu+D}{2})}$, ce qui signifie que

$$\pi(R_0/\tilde{y}) \sim \frac{GIG_{(\frac{-N}{2}+m_\beta, \frac{A}{2}, \frac{4\lambda_\beta+C}{2})}}{GIG_{(\frac{-N}{2}+m_\mu, \frac{B}{2}, \frac{4\lambda_\mu+D}{2})}}.$$

C.2 Preuve de la Proposition 5.2

Nous avons

$$p(\hat{\mathbf{y}}_{t_i} | \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}, \beta, \mu) \propto p(\hat{\mathbf{y}}_{t_i}, \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}} | \beta, \mu) \propto p(\hat{\mathbf{y}}_{t_i} | \hat{\mathbf{y}}_{t_{i-1}}, \beta, \mu) p(\hat{\mathbf{y}}_{t_{i+1}} | \hat{\mathbf{y}}_{t_i}, \beta, \mu).$$

En utilisant la densité de transition approximative de Risken [Risen 1989]

$$p(\hat{\mathbf{y}}_{t_i} | \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}, \beta, \mu) \propto |\Sigma_{i-1}|^{-\frac{1}{2}} e^{-\frac{1}{4\Delta t} (\Delta \hat{\mathbf{y}}_{t_i} - U_{i-1})' \Sigma_{i-1}^{-1} (\Delta \hat{\mathbf{y}}_{t_i} - U_{i-1})} \\ \times |\Sigma_i|^{-\frac{1}{2}} e^{-\frac{1}{4\Delta t} (\Delta \hat{\mathbf{y}}_{t_{i+1}} - U_i)' \Sigma_i^{-1} (\Delta \hat{\mathbf{y}}_{t_{i+1}} - U_i)}$$

$$\text{où } U_i = \begin{pmatrix} -\beta x_{t_i} (1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \\ \mu (1 + \frac{a}{n} - x_{t_i} - z_{t_i}) \end{pmatrix} \text{ et } \Sigma_i = \begin{pmatrix} \frac{\beta x_{t_i} (1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n} & 0 \\ 0 & \frac{\mu (1 + \frac{a}{n} - x_{t_i} - z_{t_i})}{n} \end{pmatrix},$$

puisque $U_i = U_{i-1} + o_p(\sqrt{\Delta t})$ et $\Sigma_i = \Sigma_{i-1} + o_p(\sqrt{\Delta t})$ donc, pour $\Delta t \rightarrow 0$, on a

$$p(\hat{\mathbf{y}}_{t_i} | \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}, \beta, \mu) \rightarrow |\Sigma_{i-1}|^{-1} e^{-\frac{1}{4\Delta t} (\Delta \hat{\mathbf{y}}_{t_i} - U_{i-1})' \Sigma_{i-1}^{-1} (\Delta \hat{\mathbf{y}}_{t_i} - U_{i-1})} \\ \times e^{-\frac{1}{4\Delta t} (\Delta \hat{\mathbf{y}}_{t_{i+1}} - U_{i-1})' \Sigma_{i-1}^{-1} (\Delta \hat{\mathbf{y}}_{t_{i+1}} - U_{i-1})}$$

donc

$$p(\hat{\mathbf{y}}_{t_i} | \hat{\mathbf{y}}_{t_{i-1}}, \hat{\mathbf{y}}_{t_{i+1}}, \beta, \mu) \propto |\Sigma_{i-1}|^{-1} e^{-\frac{2}{4\Delta t} (\hat{\mathbf{y}}_{t_i} - \frac{\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}}}{2})' \Sigma_{i-1}^{-1} (\hat{\mathbf{y}}_{t_i} - \frac{\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}}}{2})} \\ \propto e^{-\frac{1}{2\Delta t} (\hat{\mathbf{y}}_{t_i} - \frac{\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}}}{2})' \Sigma_{i-1}^{-1} (\hat{\mathbf{y}}_{t_i} - \frac{\hat{\mathbf{y}}_{t_{i-1}} + \hat{\mathbf{y}}_{t_{i+1}}}{2})}, \quad (C.2)$$

et

$$\begin{aligned}
p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, R_0) &\propto p(\hat{\mathbf{Y}}_{t_i}, R_0 | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}) \\
&= \int_{\mathbb{R}} p(\hat{\mathbf{Y}}_{t_i}, R_0, \mu | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}) d\mu \\
&\propto \int_{\mathbb{R}} p(R_0, \mu) p(\mu) p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, \beta, \mu) d\mu.
\end{aligned} \tag{C.3}$$

En utilisant (C.2) et (C.3), on obtient

$$\begin{aligned}
p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, R_0) &\propto \int_{\mathbb{R}} (R_0 \mu)^{m_\beta - 1} e^{-\lambda_\beta R_0 \mu} \mu^{m_\mu - 1} e^{-\lambda_\mu \mu} \\
&\quad \times e^{-\frac{1}{2\Delta t} (\hat{\mathbf{Y}}_{t_i} - \frac{\hat{\mathbf{Y}}_{t_{i-1}} + \hat{\mathbf{Y}}_{t_{i+1}}}{2})' \Sigma_{i-1}^{-1} (\hat{\mathbf{Y}}_{t_i} - \frac{\hat{\mathbf{Y}}_{t_{i-1}} + \hat{\mathbf{Y}}_{t_{i+1}}}{2})} d\mu \\
&\propto \int_{\mathbb{R}} \mu^{p-1} e^{-\frac{1}{2} (a\mu + \frac{b}{\mu})} d\mu
\end{aligned}$$

avec $p = m_\beta + m_\mu - 1$, $a = 2(\lambda_\beta R_0 + \lambda_\mu)$, $b = \frac{n}{\Delta t R_0 x_{t_{i-1}} (1 + \frac{n}{n} - x_{t_{i-1}} - z_{t_{i-1}})} h_{t_i}$ et $h_{t_i} = \left(x_{t_i} - \frac{x_{t_{i-1}} + x_{t_{i+1}}}{2}\right)^2 + R_0 x_{t_{i-1}} \left(z_{t_i} - \frac{z_{t_{i-1}} + z_{t_{i+1}}}{2}\right)^2$.

Alors, la densité a posteriori $p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, R_0)$ a la forme d'une distribution GIG(p,a,b) avec constante de normalisation $2 \left(\frac{b}{a}\right)^{\frac{p}{2}} I_p(\sqrt{ab})$, où $I_p(\cdot)$ est la fonction de Bessel de troisième degré, donc $p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, R_0) \propto 2 \left(\frac{b}{a}\right)^{\frac{p}{2}} I_p(\sqrt{ab})$.

Prenant le développement asymptotique suivant de $I : I_p(x) \sim_{+\infty} \sqrt{\frac{\pi}{2}} x^{-1/2} e^{-x} (1 + \frac{4p^2 - 1}{8x} + \dots)$, limitant au premier rang et éliminant tous les termes indépendants de x_{t_i} et z_{t_i} , on obtient

$$p(\hat{\mathbf{Y}}_{t_i} | \hat{\mathbf{Y}}_{t_{i-1}}, \hat{\mathbf{Y}}_{t_{i+1}}, R_0) \propto h_{t_i}^{\frac{m_\beta + m_\mu - 1.5}{2}} e^{-\theta \sqrt{h_{t_i}}},$$

avec $h_{t_i} = \left(x_{t_i} - \frac{x_{t_{i-1}} + x_{t_{i+1}}}{2}\right)^2 + R_0 x_{t_{i-1}} \left(z_{t_i} - \frac{z_{t_{i-1}} + z_{t_{i+1}}}{2}\right)^2$

et $\theta = \sqrt{\frac{2n(\lambda_\beta R_0 + \lambda_\mu)}{\Delta t R_0 x_{t_{i-1}} (1 + \frac{n}{n} - x_{t_{i-1}} - z_{t_{i-1}})}}$.

Bibliographie

- [Abramowitz 1972] M. Abramowitz et I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. In 10th printing, pages 443–444. United States Department of Commerce, National Bureau of Standards (NBS), 1972. (Cité en page 37.)
- [Ait-Sahalia 2010] T. Ait-Sahalia et L.P. Hansen. *Handbook of Financial Econometrics-Tools and Techniques*. volume 1, pages 443–444. Elsevier, North Holland, Amsterdam, 2010. (Cité en page 60.)
- [Andersen 1993] P.K. Andersen et Ø. Borgan R.D.Gill N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag New York, 1993. (Cité en page 7.)
- [Andersson 1991] R.M. Andersson et R.M. May. *Infectious diseases of humans : dynamics and control*. Oxford University Press, 1991. (Cité en page 7.)
- [Andersson 2000] H. Andersson et T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Springer-Verlag, New York, 2000. (Cité en pages 7, 35 et 63.)
- [Bacäer 2011] N. Bacäer. *A short history of mathematical population dynamics*. Springer-Verlag, London, 2011. (Cité en page 2.)
- [Bailey 1950] N. T. J. Bailey. *A simple stochastic epidemic*. Biometrika, vol. 37, no. 3/4, pages 193–202, 1950. (Cité en page 5.)
- [Bailey 1975] N. T. J. Bailey. *The mathematical theory of infectious diseases and its applications*. 2nd ed. Hafner Press, New York, 1975. (Cité en pages 1, 3 et 4.)
- [Bartlett 1949] M. S. Bartlett. *Some evolutionary stochastic processes*. Journal of the Royal Statistical Society B11, vol. 72, pages 211–229, 1949. (Cité en pages 4 et 5.)
- [Becker 1976] N.G. Becker. *On a general stochastic epidemic model*. Theoretical Population Biology, vol. 11, pages 23–26, 1976. (Cité en page 35.)
- [Becker 1989] N.G. Becker. *Analysis of infectious disease data. Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1989. (Cité en pages 7 et 8.)
- [Becker 1999] N.G. Becker et T. Britton. *Statistical studies of infectious disease incidence*. J. R. Stat. Soc. Ser. B Stat. Methodol, vol. 61, no. 2, pages 287–307, 1999. (Cité en page 8.)
- [Bégin 2010] J. F. Bégin. *Analyse MCMC de certains modèles de diffusion avec application au marché européen du carbone*. projet de recherche CRSNG, 2010. (Cité en page 30.)

- [Bernoulli 1760] D. Bernoulli. *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir*. Mém. Acad. Roy. Sci., Paris, 1760. (Cité en page 2.)
- [Bishwal 2008] J.P.N. Bishwal. *Parameter Estimation in Stochastic Differential Equations*. Springer-Verlag Berlin Heidelberg, 2008. (Cité en page 16.)
- [Blythe 1995] S.P. Blythe et F. Brauer C. Castillo-Chavez J. Velasco-Hernandez. *Models for sexuality transmitted diseases with recruitment*. Biometrics Unit Technical Report BU-1193-M, Cornell University, 1995. (Cité en page 50.)
- [Blythe 1997] S.P. Blythe et F. Brauer C. Castillo-Chavez. *Demographic recruitment in sexually transmitted disease models*. First World Congress on Computational Medicine, Public Health and Biotechnology Part II Austin, TX, 1994. Ed. Matthew Witten. Series in Mathematical Biology and Medicine, vol. 5, pages 1438–1457, 1997. (Cité en page 50.)
- [Boldstad 2004] W. Boldstad. *Introduction to Bayesian Statistics*. Wiley, 2004. (Cité en page 25.)
- [Britton 1998] T. Britton. *Estimation in multitype epidemics*. J. Roy. Statist. Soc. B, vol. 60, no. 4, pages 663–679, 1998. (Cité en page 7.)
- [Capasso 2005] V. Capasso et D. Bakstein. *An Introduction to Continuous-Time Stochastic Processes : Theory, Models, and Applications to Finance, Biology, and Medicine*. BIRKHÄUSER, Boston, 2005. (Cité en page 16.)
- [Cauchemez 2004] S. Cauchemez et F. Carrat C. Viboud A.J. Valleron P.Y. Böelle. *A Bayesian MCMC approach to study transmission of influenza : application to household longitudinal data*. Statistics in Medicine, vol. 23, no. 22, pages 3469–3487, 2004. (Cité en pages 36, 42, 60 et 64.)
- [Cauchemez 2008] S. Cauchemez et N. M. Ferguson. *Likelihood-based estimation of continuous-time epidemic models from time-series data : application to measles transmission in london*. Journal of the Royal Society Interface 5, vol. 23, pages 885–897, 2008. (Cité en page 7.)
- [Chib 1995] S. Chib et E. Greenberg. *Understanding the Metropolis-Hastings algorithm*. Biostatistics, vol. 49, page 327–335, 1995. (Cité en page 30.)
- [Chib 2005] S. Chib et I. Jeliazkov. *Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation*. Statistica Neerlandica, vol. 59, no. 1, page 30–44, 2005. (Cité en page 30.)
- [Choi 2012] B. Choi et G.A. Rempala. *Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling*. Biostatistics, vol. 13, no. 1, pages 153–165, 2012. (Cité en page 8.)
- [Cintrón-Arias 2009] A. Cintrón-Arias et C. Castillo-Chávez L. M. Bettencourt A. L. Lloyd H. T. Banks. *The estimation of the effective reproductive number from disease outbreak data*. Mathematical Biosciences and Engineering, vol. 6, no. 2, pages 261–282, 2009. (Cité en page 7.)

- [Clancy 2008] D. Clancy et P.D. O'Neill. *Bayesian estimation of the basic reproduction number in stochastic epidemic models*. Bayesian Analysis, vol. 3, no. 4, pages 737–757, 2008. (Cité en pages 4, 8 et 64.)
- [Clarke 1980] R.T. Clarke. *Bivariate gamma distributions for extending annual stream flow records from precipitation*. Water Resources Research, vol. 16, pages 863–870, 1980. (Cité en page 52.)
- [Comets 2006] F. Comets et T. Meyre. *Calcul stochastique et modèles de diffusions, Cours et exercices corrigés*. Dunod, Paris, 2006. (Cité en page 14.)
- [Corani 2007] G. Corani et M. Gatto. *Structural risk minimization : a robust method for density-dependence detection and model selection*. Ecography, vol. 30, no. 3, pages 400–416, 2007. (Cité en page 7.)
- [Dacunha-Castelle 1986] D. Dacunha-Castelle. *Estimation of the coefficients of a diffusion from discrete observations*. Stochastics, vol. 19, pages 263–284, 1986. (Cité en page 61.)
- [Daley 1994] D. J. Daley et J. Gani. *A deterministic general epidemic model in a stratified population*. A Tribute to Peter Whittle (F.P.Kelly), John Wiley, Chichester, pages 117–132, 1994. (Cité en page 7.)
- [Demiris 2005a] N. Demiris et P.D. O'Neill. *Bayesian inference for epidemics with two levels of mixing*. Scand. J. Stat, vol. 32, pages 265–280, 2005. (Cité en pages 8 et 41.)
- [Demiris 2005b] N. Demiris et P.D. O'Neill. *Bayesian inference for stochastic multitype epidemics in structured populations via random graphs*. J. Roy. Statist. Soc. B, vol. 67, pages 731–745, 2005. (Cité en pages 8 et 41.)
- [Demiris 2006] N. Demiris et P.D. O'Neill. *Computation of final outcome probabilities for the generalised stochastic epidemic*. Stat Comput, vol. 16, pages 309–317, 2006. (Cité en page 8.)
- [Dietz 2000] K. Dietz et J.A.P. Heesterbeek. *Bernoulli was ahead of modern epidemiology*. Nature, vol. 408, page 513, 2000. (Cité en page 2.)
- [Durham 2002] G.B. Durham et A.R. Gallant. *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes*. J. Business and Econom. Statist, vol. 20, pages 297–338, 2002. (Cité en pages 60 et 65.)
- [Dynkin 1965] E.B. Dynkin. *Markov Processes*. Springer, Berlin, 1965. (Cité en page 19.)
- [Elerian 1998] O. Elerian. *A note on the existence of a closed-form conditional transition density for the Milstein scheme*. Economics Discussion W18, Nuffield College, Oxford, 1998. (Cité en pages 37 et 65.)
- [Eraker 2001] B. Eraker. *MCMC Analysis of Diffusion Models With Application to Finance*. Journal of Business and Economic Statistics, vol. 19, pages 177–191, 2001. (Cité en pages 8, 36, 39, 41 et 65.)
- [Ethier 2005] S.N. Ethier et T.G. Kurtz. *Markov Processes Characterization and Convergence*. volume 39. New Jersey : John Wiley and Sons Inc, 2005. (Cité en page 24.)

- [Fearnhead 2004] P. Fearnhead et L. Meligkotsidou. *Exact filtering for partially-observed continuous time models*. Journal of the Royal Statistical Society. B, vol. 66, pages 771–789, 2004. (Cité en pages 8 et 35.)
- [Florens-Zmirou 1989] D. Florens-Zmirou. *Approximate discrete-time schemes for statistics of diffusion processes*. Statistics, vol. 20, pages 547–557, 1989. (Cité en page 61.)
- [Fuchs 2006] C. Fuchs. *Inference for diffusion processes with applications in Life Sciences*. Heidelberg : Springer, 2006. (Cité en page 33.)
- [Gamerman 2006] D. Gamerman et H. F. Lopes. *Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference*. London, UK : Chapman and Hall, 2^e édition, 2006. (Cité en page 33.)
- [Gardiner 2010] C. Gardiner. *Stochastic Methods*. Springer-Verlag, 2010. (Cité en page 21.)
- [Gibson 1998] G.J. Gibson et E. Renshaw. *Estimating parameters in stochastic compartmental models using Markov chain methods*. Journal of Mathematics in Applied Medicine and Biology, vol. 15, pages 19–40, 1998. (Cité en pages 8 et 35.)
- [Gihman 1975] I.I. Gihman et A.V. Skorohod. *The Theory of Stochastic Processes, Vol II*. Springer, Berlin, 1975. (Cité en page 21.)
- [Gilks 1996] W. Gilks et S. Richardson D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. New York : Chapman Hall/CRC, 1996. (Cité en page 33.)
- [Gillespie 1977] D. Gillespie. *Exact stochastic simulation of coupled chemical reactions*. Journal of Physical Chemistry, vol. 81, pages 2340–2364, 1977. (Cité en page 31.)
- [Golightly 2005] A. Golightly et D.J. Wilkinson. *Bayesian Inference for Stochastic Kinetic Models Using a Diffusion Approximation*. Biometrics, vol. 61, pages 781–788, 2005. (Cité en pages 36 et 39.)
- [Gut 2005] A. Gut. *Probability : a graduate course*. pages 781–788. Springer-Verlag, 2005. (Cité en page 77.)
- [Hamer 1906] W.H. Hamer. *Global stability of the endemic equilibrium of multigroup SIR epidemic models*. Epidemic disease in England, vol. 1, pages 733–739, 1906. (Cité en page 2.)
- [Hastings 1970] W.k. Hastings. *Monte Carlo Sampling Methods Using Markov Chain and Their Applications*. Biometrika, pages 97–109, 1970. (Cité en page 28.)
- [Hayakawa 2003] Y. Hayakawa et D. Upton P.S.F. Yip P.D. O’Neill. *Bayesian Inference for a Stochastic Epidemic Model with Uncertain Numbers of Susceptibles of Several Types*. Australian and New Zealand Journal of Statistics, vol. 45, no. 3, pages 491–502, 2003. (Cité en page 8.)
- [Izawa 1965] T. Izawa. *Two or multi-dimensional gamma-type distribution and its application to rainfall data*. Papers in Meteorology and Geophysics, vol. 15, page 167, 1965. (Cité en page 52.)

- [Jedrzejewski 2009] F. Jedrzejewski. *Modèles aléatoires et physique probabiliste*. Springer, 2009. (Cit  en page 15.)
- [Jorgensen 1982] B. Jorgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer-Verlag, New York, 1982. (Cit  en page 82.)
- [Karatzas 1998] I. Karatzas et S.E. Shreve. *Methods of Mathematical Finance*. volume 39. Springer-Verlag, New York, 1998. (Cit  en page 20.)
- [Kermack 1927] W. O. Kermack et A. G. McKendrick. *Contributions to the Mathematical Theory of Epidemics. III*. Proceedings of the Royal Society, Series A, vol. 115, pages 700–721, 1927. (Cit  en page 3.)
- [Kibble 1941] W.F. Kibble. *two-variate gamma type distribution*. Sankhya, vol. 5, pages 137–150, 1941. (Cit  en pages i, 50 et 52.)
- [Kim 1998] S. Kim et N. Shephard S. Chib. *Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models*. Review of Economic Studie, 1998. (Cit  en page 8.)
- [Klenke 2014] A. Klenke. *Probability Theory, A comprehensive cours, 2^{end} Edition*. Springer-Verlag London, 2014. (Cit  en page 12.)
- [Kloeden 1992] P.E. Kloeden et E. Platen. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Springer, Berlin, 1992. (Cit  en page 37.)
- [Koudou 2014] A.E. Koudou et C. Ley. *Characterizations of GIG laws : A survey*. Probability Surveys, vol. 11, pages 161–176, 2014. (Cit  en page 42.)
- [Lacus 2008] S.M. Lacus. *Simulation and inference for stochastic differential equations : with R examples*. volume 39. Springer Science, New York, 2008. (Cit  en page 17.)
- [Lebedev 1972] N. Lebedev. *Special functions and their applications*. Translated by R. Silverman, Dover publishing, New York, 1972. (Cit  en pages 37 et 52.)
- [Lekone 2004] P.E. Lekone et B.F. Roberts. *Statistical inference and model selection for the 1861 Hagelloch measles epidemic*. Biostatistics, vol. 5, pages 249–261, 2004. (Cit  en page 8.)
- [Maroufy 2010] H.El Maroufy et Z. Taib. *On the Qualitative Behaviour of SIR Epidemics with Generalized Infection Rate Functions*. Applied Mathematics Information Sciences, vol. 4, no. 3, pages 353–363, 2010. (Cit  en page 7.)
- [Maroufy 2012] H.El Maroufy et L. Omari Z.Taib. *Transition Probabilities for Generalized SIR Epidemic Model*. Stochastic Models, 2012. (Cit  en page 7.)
- [Maroufy 2015] H.El Maroufy et T. Kernane S. Bechket A. Ouddaj. *Bayesian Inference for Non Linear Stochastic SIR Epidemic Model*. Journal of Statistical Computation and Simulation, vol. 86, no. 11, pages 2229–2240, 2015. (Cit  en pages 36, 39, 44, 47, 60 et 71.)
- [Maroufy 2016] H.El Maroufy et D. Kiouach T. Ziad. *Final outcome probabilities for SIR epidemic model*. Communications in Statistics-Theory and Methods, vol. 45, no. 8, pages 2426–2437, 2016. (Cit  en page 24.)

- [Mathai 1991] A.M. Mathai et P.G. Moschopoulos. *On a multivariate gamma*, *Journal of Multivariate Analysis*, vol. 39, pages 135–153, 1991. (Cit  en page 52.)
- [McDonald 1984] B. McDonald. *Some Generalized Functions for the Size Distribution of Income*. *Econometrica*, vol. 52, no. 3, pages 647–663, 1984. (Cit  en page 64.)
- [Meester 2002] L. De. Meester et A. G mez B. Okamura K. Schwenk. *Computation of final outcome probabilities for the generalised stochastic epidemic*. *Acta Oecol-Intern J Ecol*, vol. 23, pages 121–135, 2002. (Cit  en page 7.)
- [Metropolis 1953] N. Metropolis et A.W. Rosenbluth M.N. Rosenbluth A.H. Teller E. Teller. *Equation of State Calculations by Fast Computing Machines*. *Journal of Chemical Physics*, vol. 21, pages 1087–1092, 1953. (Cit  en page 28.)
- [Meyn 1993] S. Meyn et R. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1993. (Cit  en page 27.)
- [Moran 1969] P.A.P. Moran. *Statistical inference with bivariate gamma distributions*. *Biometrika*, vol. 54, pages 385–394, 1969. (Cit  en page 52.)
- [Moran 1970] P.A.P. Moran. *The methodology of rain making experiments*. *Review of the International Statistical Institute*, vol. 38, pages 105–115, 1970. (Cit  en page 52.)
- [Neal 2005] P.J. Neal et G.O. Wilkinson. *Bayesian Inference for Stochastic Kinetic Models Using a Diffusion Approximation*. *Biometrics*, vol. 61, pages 781–788, 2005. (Cit  en pages 8 et 35.)
- [Oh 2013] C. Oh. *Approximation of estimates in the susceptible-infectious-removed epidemic model*. *Journal of Mathematics and Statistics*, vol. 9, no. 1, pages 38–42, 2013. (Cit  en page 7.)
- [ ksendal 1995] B.  ksendal. *Stochastic Differential Equations : with Applications*. Springer-Verlag Heidelberg, New York, 1995. (Cit  en pages 14, 18, 19 et 37.)
- [O’Neill 1999] P.D. O’Neill et G.O. Roberts. *Bayesian inference for partially observed stochastic epidemics*. *J.R.S.S, Series A*, vol. 162, pages 121–129, 1999. (Cit  en pages 8, 35, 36, 41 et 60.)
- [O’Neill 2002] P.D. O’Neill. *A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods*. *Mathematical Biosciences*, vol. 180, pages 103–114, 2002. (Cit  en page 8.)
- [Pedersen 1995] A. Pedersen. *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*. *Scandinavian Journal of Statistics*, vol. 22, pages 55–71, 1995. (Cit  en pages 36, 49 et 60.)
- [Poulsen 1999] R. Poulsen. *Approximate maximum likelihood estimation of discretely observed diffusion processes*. Working Paper 29, Centre for Analytical Finance, Aarhus, 1999. (Cit  en page 60.)

- [Qaffou 2017a] A. Qaffou et H. El Maroufy T. Kernane. *Bayesian Inference of the Basic Reproduction Number for a SIR Epidemic Model*. Brazilian Journal of Probability and Statistics (en révision), 2017. (Cit  en page 72.)
- [Qaffou 2017b] A. Qaffou et H. El Maroufy T. Kernane. *Bayesian Parameter Inference for Stochastic SIR Epidemic Model with Hyperbolic Diffusion*. Communications in Statistics-Simulation and Computation, pages 1–16, 2017. (Cit  en pages 60 et 71.)
- [Qaffou 2017c] A. Qaffou et H. El Maroufy. *Bayesian Inference for SIR Epidemic Model with dependent parameters*. Journal of Statistical Computation and Simulation (Soumis), 2017. (Cit  en page 72.)
- [Raggett 1982] G.F. Raggett. *Modeling the Eyam Plague*. Bulletin of the Institute of Math. and its Appl, vol. 18, pages 221–226, 1982. (Cit  en page 45.)
- [Rao 1973] C.R. Rao. *Linear Statistical Inference and its Applications*. Second-edition, Wiley, New York, 1973. (Cit  en page 63.)
- [Risken 1989] H. Risken. *The Fokker-Plank Equation, Method of Solution and Application*. Second-edition, Springer-Verlag, New York, 1989. (Cit  en pages 50, 51, 55, 60, 61, 64, 67, 72, 81 et 82.)
- [Robert 1996] C. Robert. *M thodes de Monte-Carlo par Cha nes de Markov*. Economica, 1996. (Cit  en page 30.)
- [Robert 2004] C. P. Robert et G. Casella. *Monte Carlo Statistical Methods*. New York : Springer, 2 edition, 2004. (Cit  en pages 27, 28, 29 et 33.)
- [Robert 2005] G. Robert et R. Tweedie. *Understanding MCMC*. Springer-Verlag, New York, 2005. (Cit  en page 27.)
- [Robert 2006] C. Robert. *Le choix Bay sien, Principes et pratique*. Springer-Verlag France, Paris, 2006. (Cit  en page 27.)
- [Roberts 2001] G.O. Roberts et O. Stramer. *On inference for partially observed nonlinear diffusion models using Metropolis-Hastings algorithms*. Biometrika, vol. 88, pages 603–621, 2001. (Cit  en pages 8 et 65.)
- [Ross 1911] R. Ross. *The prevention of malaria*. London : John Murray, 1911. (Cit  en pages 7 et 2.)
- [Ross 2006] J. V. Ross et T. Taimre P. K. Pollett. *On parameter estimation in population models*. Theor. Popul. Biol, vol. 70, no. 638, page 498–510, 2006. (Cit  en page 7.)
- [Ross 2009] J. V. Ross et D. E. Pagendam P. K. Pollett. *On parameter estimation in population models II : multi-dimensional processes and transient dynamics*. Theor. Popul. Biol, vol. 75, no. 638, page 123–132, 2009. (Cit  en page 7.)
- [Sabatier 2005] P. Sabatier et M.A. Dubois D.j. Bicout B. Durant. *Le recours   la mod lisation en  pid miologie animale*. Epidemiol. et Sant  Anim, vol. 47, pages 15–33, 2005. (Cit  en page 2.)

- [Smith 1981] O.E. Smith et S.I. Adelfang. *Gust model based on the bivariate gamma distribution*. Journal of Spacecraft, vol. 18, pages 545–549, 1981. (Cité en page 52.)
- [Sørensen 2003] M. Sørensen et M. Uchida. *Small-diffusion asymptotics for discretely sampled stochastic differential equations*. Bernoulli, vol. 9, pages 1051–1069, 2003. (Cité en page 63.)
- [Streftaris 2004] G. Streftaris et G.J. Gibson. *Bayesian inference for stochastic epidemics in closed populations*. Statistical Modelling, vol. 18, 2004. (Cité en pages 8, 35, 36, 42, 60 et 64.)
- [Stroock 1997] D. W. Stroock et S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag :Berlin, Germany, 1997. (Cité en pages 19 et 20.)
- [Yip 1998] P.S.F. Yip et Q. Chen. *Statistical inference for a multitype epidemic model*. Journal of Statistical Planning and Inference, vol. 71, pages 229–244, 1998. (Cité en page 7.)
- [Zakary 2016] O. Zakary et M. Rachik I. Elmouki. *On the analysis of a multi-regions discrete SIR epidemic model : an optimal control approach*. International Journal of Dynamics and Control, pages 1–14, 2016. (Cité en page 6.)
- [Zeeman 1993] E.C. Zeeman. *Controversy in science : on the ideas of Daniel Bernoulli and René Thom*. Nieuw Arch. Wisk, vol. 11, page 257, 1993. (Cité en page 2.)
- [zu Dohna 2010] H. zu Dohna et M. PinedaKrch. *A SIRS epidemic model with infection-age dependence*. Theoret. Popul. Biol, vol. 78, pages 71–76, 2010. (Cité en page 7.)

Résumé : L'objectif de cette thèse est l'inférence Bayésienne paramétrique pour les modèles **SIR** (*Susceptible-Infecté-Eliminé*) en approchant le processus épidémique discret par une diffusion d'Itô. Notre approche consiste à augmenter les données de faible fréquence par l'introduction de données latentes entre chaque paire d'observations, simulées en utilisant les schémas standards de discrétisation des équations différentielles stochastiques, Euler et Milstein.

Dans ce travail, trois problèmes d'inférence Bayésienne sont considérés. Le premier concerne le cas où les deux paramètres, le taux d'infection et le taux d'élimination sont indépendants ; nous explorons leurs distributions a posteriori ainsi que celle des données manquantes, en approchant la densité de probabilité de transition du processus par une distribution hyperbolique. Le deuxième, traite l'inférence Bayésienne dans le cas où les deux taux sont supposés a priori dépendants, en considérant la loi de Kibble gamma bivariée comme loi a priori. Le troisième consiste à l'estimation Bayésienne du taux de reproduction de base ; sa distribution a posteriori est explorée et celle des données manquantes. Des algorithmes MCMC ont été établis pour simuler les distributions a posteriori.

Mots clés : Modèle **SIR**, Inférence Bayésienne, Distributions a priori et a posteriori, Données manquantes, Méthodes MCMC.

Abstract : The objective of this thesis is the Bayesian parametric inference for the **SIR** (*Susceptible-Infected-Removed*) models approaching the discrete epidemic process by a Itô diffusion. Our approach consists in increasing the low frequency data by introducing latent data between each pair of observations, simulated using the standard discretization schemes of the stochastic differential equations, Euler and Milstein.

In this work, three problems of Bayesian inference are considered. The first concerns the case where the two parameters, the infection rate and the removed rate are independent ; we investigate their a posteriori distributions as well as the missing data, by approaching the transition probability density of the process by a hyperbolic distribution. The second one deals with the Bayesian inference in the case where the two rates are assumed to be a priori dependents, considering the density of Kibble gamma bivariate as a priori distribution. The third concern the Bayesian estimation of the basic reproduction rate ; Its a posteriori distribution is explored and that of the missing data. MCMC algorithms have been established to simulate posterior distributions.

Keywords : **SIR** model, Bayesian inference, Prior and Posterior Distributions, Missing data, MCMC methods.
