Optimisation des méthodes algorithmiques en inférence bayésienne. Modélisation dynamique de la transmission d'une infection au sein d'une population hétérogène

Dorota Gajda

▶ **To cite this version:**

Dorota Gajda. Optimisation des méthodes algorithmiques en inférence bayésienne. Modélisation dynamique de la transmission d'une infection au sein d'une population hétérogène. Médecine humaine et pathologie. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA11T054 . tel-00659618

HAL Id: tel-00659618

https://theses.hal.science/tel-00659618

Submitted on 13 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS XI
FACULTE DE MEDECINE PARIS-SUD

Année 2011

N° attribué par la bibliothèque

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERITE PARIS XI

Spécialité : Santé Publique
Option : Biostatistique

Ecole Doctorale de rattachement : **ED 420 « Santé Publique »**

présentée et soutenue publiquement par

Dorota GAJDA

le 13 Octobre 2011

Optimisation des méthodes algorithmiques en inférence bayésienne. Modélisation dynamique de la transmission d'une infection au sein d'une population hétérogène.

Directeur de thèse : Professeur Chantal GUIHENNEUC-JOUYAUX

Responsable scientifique - co-directeur : Professeur Jean MACCARIO

JURY :

M. Jean-Christophe THALABARD	Président
M. Jean-Michel MARIN	Rapporteur
M. Raphaël PORCHER	Rapporteur
Mme. Sophie DONNET	Examinateur
M. Yann LE STRAT	Examinateur
Mme. Chantal GUIHENNEUC-JOUYAUX	Directeur
M. Jean MACCARIO	Co-directeur

Remerciements

En premier lieu, je tiens à remercier ma directrice de thèse Madame Chantal Guihenneuc-Jouyaux, de m'avoir accordé sa confiance et son soutien permanent, pour sa disponibilité, sa bienveillance et son aide.

Je remercie également mon co-directeur de thèse Monsieur Jean Maccario, pour sa disponibilité, sa patience, la liberté d'action qu'il m'a accordée tout au long de cette thèse ainsi que son aide.

C'était un vrai plaisir sur le plan scientifique et humain d'être encadré par ces deux responsables de valeur.

Je tiens à remercier Monsieur Jean-Christophe Thalabard de m'avoir fait l'honneur d'être président de mon jury de thèse, mes deux rapporteurs Monsieur Raphaël Porcher et Monsieur Jean-Michel Marin pour le temps passé et l'attention portée à ce document et leurs remarques constructives. Je remercie aussi mes examinateurs Madame Sophie Donnet et Monsieur Yann Le Strat d'avoir accepté de participer à ce jury de thèse et d'évaluer mon travail.

Je remercie chaleureusement Madame Pascale Tubert-Bitter et Monsieur Thierry Moreau, directrice et ex-directeur de l'équipe de biostatistiques EQ1 (ex. U780) et également Madame Jacqueline Clavel, directrice de l'équipe EQ6 (ex. U754) de CESP INSERM 1018 qui m'ont accueillie et permis de vivre une expérience humaine et professionnelle enrichissante au sein de leurs unités durant les quatre ans de thèse.

Mes remerciements vont aussi à l'ensemble de membres de ces deux équipes, pour leur bonne humeur et leut gentillesse, leur contribution à l'ambiance agréable, leur aide et leurs bons conseils, le tout contribuant à un cadre de travail idéal.

Je voudrais aussi remercier Monsieur Jean Bouyer le directeur de l'école doctorale pour sa compréhension, sa confiance et son aide dans l'organisation et bon déroulement de la fin de cette thèse.

J'en profite pour remercie le Ministère de l'Enseignement Supérieur et la Recherche et Université Paris II Assas pour le financement de mon projet grâce à l'Allocation de Recherche d'une part et au poste ATER d'autre part.

Un très grand merci à tous les doctorants et anciens doctorants de CESP que j'ai eu l'occasion et le plaisir de rencontrer, et avec avec qui j'ai partagé les aléas de la thèse, pour leur aide, leur soutien et leurs conseils . Je remercie tout particulièrement Helena et Juliette pour leur amabilité, leur serviabilité et leur amitié.

Enfin je souhaite de remercier Sébastien et nos familles, surtout mes parents, ainsi que nos amis pour leurs encouragements chaleureux, leur compréhension, l'intérêt et l'estime portés à mon travail.

Résumé

Ce travail se décompose en deux grandes parties, "Estimations répétées dans le cadre de la modélisation bayésienne" et "Modélisation de la transmission de maladies infectieuses dans une population. Estimation des paramètres.". Les techniques développées dans la première partie sont utilisées en fin de la seconde partie.

La première partie est consacrée à des optimisations d'algorithmes stochastiques très souvent utilisés, notamment dans le contexte des modélisations Bayésiennes. Cette optimisation est particulièrement faite lors de l'étude empirique d'estimateurs des paramètres d'un modèle où les qualités des estimateurs sont évaluées sur un grand nombre de jeux de données simulées. Quand les lois a posteriori ne sont pas explicites, le recours à des algorithmes stochastiques itératifs (de la famille des algorithmes dits de Monte Carlo par Chaîne de Makov) pour approcher les lois a posteriori est alors très couteux en temps car doit être fait pour chaque jeu de données. Dans ce contexte, ce travail consiste en l'étude de solutions évitant un trop grand nombre d'appels à ces algorithmes mais permettant bien-sûr d'obtenir malgré tout des résultats précis. La principale technique étudiée dans cette partie est celle de l'échantillonnage préférentiel.

La seconde partie est consacrée aux études de modèles épidémiques, en particulier le modèle compartimental dit SIS (Susceptible-Infecté-Susceptible) dans sa version stochastique. L'approche stochastique permet de prendre en compte l'hétérogénéité de l'évolution de la maladie dans la population. les approches par des processus Markoviens sont étudiés où la forme des probabilités de passage entre les états est non linéaire. La solution de l'équation différentielle en probabilité n'est alors en général pas explicite. Les principales techniques utilisées dans cette partie sont celles dites de développement de l'équation maîtresse ("master

equation") appliquées au modèle SIS avec une taille de population constante. Les propriétés des estimateurs des paramètres sont étudiées dans le cadre fréquentiste et bayésien. Concernant l'approche Bayésienne, les solutions d'optimisation algorithmique de la première partie sont appliquées.

Mots clés : MCMC, Echantillonnage Pondéré, Estimations Répétées, Modèle Epidémique, Modèle SIS, Equation Maîtresse.

Abstract

This work consists in two parts, " Repeated estimates in bayesian modelling " and " Modelling of the transmission of infectious diseases in a population. Estimation of the parameters. ". Techniques developed in the first part are used at the end of the second part.

The first part deals with optimizations of very often used stochastic algorithms, in particular in the context of Bayesian modelling. This optimization is particularly made when empirical study of estimates based on numerous simulated data sets is done. When posterior distribution of parameters are not explicit, its approximation is obtained via iterative stochastic algorithms (of the family of Markov Chain Monte Carlo) which is computationally expensive because has to be done on each data set. In this context, solutions are proposed avoiding an excess large number of MCMC calls but nevertheless giving accurate results. The Importance Sampling method is used in combination with MCMC in Bayesian simulation study.

The second part deals with epidemic models, in particular the compartimental model SIS (Susceptible-Infectious-Susceptible) in its stochastic version. The stochastic approach allows to take into account the heterogeneousness of disease evolution in the population. Markov Process is particularly studied where transition probability between states is not linear, the solution of the differential equation in probability being then generally not explicit. The main techniques used in this part are the ones based on Master equation applied on SIS model with a constant population size. Empirical properties of parameters estimates are studied in frequentist and Bayesian context with algorithmic optimization presented in the first part.

Key words : MCMC, Importance Sampling, Repeated Estimations, Epidemic Model, SIS

Model, Master Equation.

Liste des travaux issues du travail de thèse

Publications

Gajda D., Guihenneuc-Jouyaux C., Rousseau J., Mengersen K., Nur D., (2010), Use in practice of importance sampling for repeated MCMC for Poisson models., *Electron. J. Statist.*, 4, 361-383.

Gajda D., Guihenneuc-Jouyaux C., Rousseau J., Mélange de distributions comme fonction d'importance dans l'échantillonnage préférentiel combiné avec l'algorithme de Monte Carlo par Chaîne de Markov, *Revue des Nouvelles Technologies de l'Information (RNTI)* S-1 - un numéro spécial JDS2009. Éditeur : Hermann, Paris 2011.

Gajda D., Maccario J., Master equation development in SIS model. *en préparation*

Communications orales et présentations de posters

Gajda D., Guihenneuc-Jouyaux C., Rousseau J., Mengersen K., Nur D., Importance Sampling combiné avec les algorithmes MCMC dans le cas d'estimations répétées., *41èmes Journées de Statistique*, Bordeaux (SFdS 2009) - communication oral.

Gajda D., Guihenneuc-Jouyaux C., Rousseau J., Mengersen K., Nur D., (2010), KL divergence and L1 norm for Importance Sampling combined with MCMC in Bayesian simulation study., ISCB2010, Montpellier - poster session

Table des matières

I Introduction générale **1**

II Estimations répétées dans le cadre de la modélisation bayésienne. **7**

1 Introduction **9**

2 Matériel et Méthodes **11**

2.1 Modélisation et inférence Bayésienne 11

2.1.1 Méthodes de Monte Carlo 13

2.1.2 Méthodes de Monte Carlo par Chaîne de Markov (MCMC) 18

2.2 MCMC combiné avec Importance Sampling 28

2.2.1 Démarche classique avec MCMC 29

2.2.2 Introduction de l'échantillonnage pondéré dans les estimations 30

2.2.3 Stratégie de référence fixée 32

2.2.4 Stratégie de référence choisie 33

2.2.5 Stratégie du mélange 36

2.2.6 Critères de comparaison entre les différentes approximations 44

2.3 Modèles de Poisson 45

2.3.1 Modèles de simulation 46

2.3.2 Modèles d'estimation 47

3 Résultats **47**

3.1 Résultats avec la stratégie de référence fixe 48

3.2 Résultats avec la stratégie de référence choisie 50
3.2.1 Modèle 1 : Modèle de Poisson 51
3.2.2 Modèle 2 : Régression de Poisson 53
3.2.3 Modèle 3 : Régression de Poisson avec extravariabilité 54
3.2.4 Augmentation de la taille de l'échantillon et du nombre de covariables 55
3.3 Stratégie de mélange 59
3.3.1 Résultats 59
3.3.2 Estimations des constantes de normalisation dans la stratégie de mélange 60
3.4 Sélection automatique 62

4 Discussion **65**

III Modélisation de la transmission de maladies infectieuses dans une population. Estimation des paramètres. **69**

5 Introduction **71**

6 Matériel et Méthodes **75**
6.1 Les processus de Markov à temps continu. 75
6.1.1 La formule de Chapman-Kolmogorov 75
6.1.2 Stationnarité 76
6.2 Equation Maîtresse 77
6.2.1 Comportement asymptotique de l'équation maîtresse du processus à espace d'états fini et discret 79
6.3 Processus de naissance et de mort (PNM) 79
6.3.1 Les moments de PNM 79
6.3.2 Solution stationnaire de PNM 80
6.3.3 Solution quasi-stationnaire de PNM 81
6.3.4 Solution générale de PNM 83
6.3.5 Solution approchée de PNM - taux de transition non linéaire 84

6.4 Développement de l'équation maîtresse pour le PNM 85
6.5 Modèle SIS avec la population de taille constante. 88
6.5.1 Modèle SIS déterministe 89
6.5.2 Modèle SIS stochastique 89
6.6 Modèle pour décrire les données 91
6.6.1 Modèle déterminé par le développement de l'équation maîtresse (M) . 92
6.6.2 Modèle simple (S) 94
6.7 Estimation par maximum de vraisemblance 96
6.7.1 Estimation des paramètres 96
6.7.2 Estimation de la variance asymptotique 96
6.8 Estimation bayésienne 99
6.8.1 Estimation des paramètres 99

7 Résultats **100**

7.1 Modélisation 100
7.2 Simulations 107
7.2.1 Choix des paramétrisations 107
7.2.2 Données simulées 107
7.3 Résultats des estimations fréquentistes 108
7.3.1 Problème calculatoire 109
7.3.2 Interprétation des résultats 111
7.3.3 Exemple de résultats de l'estimation de R_0 dans la phase évolutive . . 113
7.3.4 Identifiabilité des paramètres de modèles 116
7.3.5 Résultats avec le paramètre β fixe 118
7.4 Observations sous l'état stationnaire - Estimateurs de R_0 121
7.4.1 Résultats algébriques 121
7.4.2 Application aux données simulées 123
7.5 Résultats des estimations bayésiennes 127
7.5.1 Choix des lois a priori 127

7.5.2 Estimation des paramètres - utilisation de la stratégie de mélange . . . 129
7.5.3 Analyse des résultats 130

8 Discussion **131**

IV Conclusion générale **137**

Bibliographie **141**

Annexe de la modélisation bayésienne **150**

Annexe de la modélisation épidémique **176**

Articles **250**

Table des figures

2.1 Algorithme de Hastings-Metropolis. 20

2.2 Algorithme de Hastings-Metropolis. 21

2.3 Échantillonneur de Gibbs. 22

2.4 Algorithme de Rejection Sampling. 23

2.5 Construction de la fonction enveloppe dans l'algorithme Adaptive Rejection Sampling pour $n = 3$ 24

2.6 Algorithme Adaptive Rejection Sampling. 25

2.7 Slice Sampling 26

2.8 Version simple de Slice Sampling 26

2.9 Inversion pour les variables discrètes. 27

2.10 Démarche classique via MCMC 29

2.11 Démarche via IS avec référence fixe 32

2.12 Démarche via IS avec référence choisie 34

2.13 Modèles de simulation des données. 46

2.14 Modèles d'estimation et lois a priori. 47

3.1 Étude des performances de l'estimation de $E_{[\theta|X^{(k)}]}[g(\theta)]$ pour $g(\theta) = a$ avec la stratégie "référence fixe" (où le premier jeu de données $X^{(1)}$ est la référence fixe) dans le modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$ 51

3.2 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 1 avec $g(\lambda) = \lambda$ pour $\lambda = 1$ (à gauche) et $\lambda = 20$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 52

3.3 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 2 avec $\theta = (a = 0, b = 0.5)$ pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 54

3.4 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$ (en première ligne), $\sigma^2 = 1/4$ (en deuxième ligne) ou $\sigma^2 = 1/2$ (en troisième ligne), pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (au centre), $g(\theta) = \sigma^2$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 56

3.5 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/2$ et $n = 1000$, pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) et $g(\theta) = \sigma^2$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 58

3.6 Densités marginales a posteriori de a (à gauche), b (au centre) et σ^2 (à droite) approchées par MCMC dans le Modèle 3 avec les paramètres $\theta = (a = 0, b = 0.5, \sigma^2 = 1/2)$. Dix densités a posteriori sélectionnées par la méthode de classification PAM sont tracées en gras. 64

7.1 $\alpha = 1.2$, $\beta = 0.3$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires 103

7.2 $\alpha = 1.2$, $\beta = 0.6$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires 104

7.3 $\alpha = 1.2$, $\beta = 0.9$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires 105

7.4 $\alpha = 1.2$, $\beta = 1.1$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires 106

7.5 Les trois scénarios des simulations des données selon la phase : phase initiale « early » (à gauche), phase évolutive « during » (au centre), phase stationnaire « after » (à droit). 108

7.6 Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » - le trait rouge représente la vraie valeur du paramètre. 113

7.7 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » - les losanges rouges représentent les écarts-types des estimations. 114

7.8 PP-plots de R_0 pour les observations de la phase evolutive : « during » 115

7.9 Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus 117

7.10 Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » (β-fixe) 119

7.11 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » (β-fixe) ; (M) en rouge, (S) en bleu 119

7.12 PP-plots de R_0 pour les observations de la phase évolutive : « during » (β-fixe) . 120

7.13 Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0 ; (M) en rouge, (S) en bleu 126

7.14 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0 126

7.15 PP-plots de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0 127

7.16 Lois a priori pour les modèles (M) et (S) 128

7.17 Les paramètres de la loi a priori inverse gamma sur σ^2 128

7.18 Les estimations bayésiennes de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during ». Résultats MCMC en première ligne, IS en deuxième ligne. 131

1 Performance d'inférence pour $g(\theta) = \lambda$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 1 avec $\lambda = 1$ 156

2 Performance d'inférence pour $g(\theta) = \lambda$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 1 avec $\lambda = 20$ 157

3 Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 2 avec $a = 0$ et $b = 0.5$ 158

4 Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 2 avec $a = 0$ et $b = 0.5$ 159

5 Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$ 160

6 Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$ 161

7 Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$ 162

8 Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$ 163

9 Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$ 164

10 Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$ 165

11 Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$ 166

12 Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$ 167

13 Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$ 168

14 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b_1, \ldots, b_{10} = 0.05$ (normal) et $\sigma^2 = 1/2$, pour $g(\theta) = a$, $g(\theta) = b_1, \ldots, b_{10}$ et $g(\theta) = \sigma^2$; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 169

15 Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b_1, \ldots, b_{10} = 0.05$ (normal + bernoulli) et $\sigma^2 = 1/2$, pour $g(\theta) = a$, $g(\theta) = b_1, \ldots, b_{10}$ et $g(\theta) = \sigma^2$; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième. 171

16 $\alpha = 1.2$, $\beta = 0.3$, extinction 178

17 $\alpha = 1.2$, $\beta = 0.6$, extinction 179

18 $\alpha = 1.2$, $\beta = 0.9$, extinction 180

19 $\alpha = 1.2$, $\beta = 1.1$, extinction 181

20 Boxplots des estimations de α pour les observations de la phase initale : « early » 183

21 Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase initale : « early » 183

22 Boxplots des estimations de β pour les observations de la phase initiale : « early » 184

23 Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase initiale : « early » 184

24 Boxplots des estimations de i_0 pour les observations de la phase initiale : « early » 185

25 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase initiale : « early » 185

26 Boxplots des estimations de R_0 pour les observations de la phase initiale : « early » 186
27 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase initiale : « early » 186
28 Boxplots des estimations de α pour les observations de la phase évolutive : « during » 187
29 Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase évolutive : « during » 187
30 Boxplots des estimations de β pour les observations de la phase évolutive : « during » 188
31 Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase évolutive : « during » 188
32 Boxplots des estimations de i_0 pour les observations de la phase évolutive : « during » 189
33 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase évolutive : « during » 189
34 Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » 190
35 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » 190
36 Boxplots des estimations de α pour les observations de la phase stationnaire : « after » 191
37 Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase stationnaire : « after » 191
38 Boxplots des estimations de β pour les observations de la phase stationnaire : « after » 192
39 Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase stationnaire : « after » 192
40 Boxplots des estimations de i_0 pour les observations de la phase stationnaire : « after » 193

41 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase stationnaire : « after » 193
42 Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « after » 194
43 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « after » 194
44 PP-plots de $\theta = \alpha$ pour les observations de la phase initiale : « early » 195
45 PP-plots de $\theta = \beta$ pour les observations de la phase initiale : « early » 195
46 PP-plots de $\theta = i_0$ pour les observations de la phase initiale : « early » 196
47 PP-plots de R_0 pour les observations de la phase initiale : « early » 196
48 PP-plots de $\theta = \alpha$ pour les observations de la phase évolutive : « during » 197
49 PP-plots de $\theta = \beta$ pour les observations de la phase evolutive : « during » 197
50 PP-plots de $\theta = i_0$ pour les observations de la phase evolutive : « during » . . . 198
51 PP-plots de R_0 pour les observations de la phase evolutive : « during » 198
52 PP-plots de $\theta = \alpha$ pour les observations de la phase stationnaire : « after » . . . 199
53 PP-plots de $\theta = \beta$ pour les observations de la phase stationnaire : « after » . . . 199
54 PP-plots de $\theta = i_0$ pour les observations de la phase stationnaire : « after » . . . 200
55 PP-plots de R_0 pour les observations de la phase stationnaire : « after » 200
56 Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase initial : « early » ; modèle (M) - points rouges, modèle (S) - points bleus 216
57 Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus 216
58 Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase stationnaire : « after » ; modèle (M) - points rouges, modèle (S) - points bleus 217
59 Boxplots des estimations de i_0 pour les observations de la phase initiale : « early » (β-fixe) 219

60 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase initiale : « early » (β-fixe) 219
61 Boxplots des estimations de R_0 pour les observations de la phase initiale : « early » 220
62 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase initiale : « early » (β-fixe) 220
63 Boxplots des estimations de i_0 pour les observations de la phase évolutive : « during » (β-fixe) 221
64 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase évolutive : « during » (β-fixe) 221
65 Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » (β-fixe) 222
66 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » (β-fixe) 222
67 Boxplots des estimations de i_0 pour les observations de la phase stationnaire : « after » (β-fixe) 223
68 Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase stationnaire : « after » (β-fixe) 223
69 Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « after » (β-fixe) 224
70 Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « after » (β-fixe) 224
71 PP-plots de $\theta = i_0$ pour les observations de la phase initale : « early » (β-fixe) . . 225
72 PP-plots de R_0 pour les observations de la phase initale : « early » (β-fixe) 225
73 PP-plots de $\theta = i_0$ pour les observations de la phase évolutive : « during » (β-fixe) 226
74 PP-plots de R_0 pour les observations de la phase évolutive : « during » (β-fixe) . 226
75 PP-plots de $\theta = i_0$ pour les observations de la phase stationnaire : « after » (β-fixe) 227
76 PP-plots de R_0 pour les observations de la phase stationnaire : « after » (β-fixe) . 227
77 Résultats d'estimations bayésiennes par MCMC de α pour les observations de la phase évolutive : « during » 234

78 Résultats d'estimations bayésiennes IS de α pour les observations de la phase évolutive : « during » 234
79 Résultats d'estimations bayésiennes par MCMC de β pour les observations de la phase évolutive : « during » 235
80 Résultats d'estimations bayésiennes IS de β pour les observations de la phase évolutive : « during » 235
81 Résultats d'estimations bayésiennes par MCMC de i_0 pour les observations de la phase évolutive : « during » 236
82 Résultats d'estimations bayésiennes IS de i_0 pour les observations de la phase évolutive : « during » 236
83 Résultats d'estimations bayésiennes par MCMC de R_0 pour les observations de la phase évolutive : « during » 237
84 Résultats d'estimations bayésiennes IS de R_0 pour les observations de la phase évolutive : « during » 237
85 Les estimations bayésiennes de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus 242

Liste des tableaux

3.1 Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=1/2) 49

3.2 Erreurs quadratiques moyennes par rapport à la vraie moyenne a posteriori dans le Modèle 1 pour $g(\lambda) = \lambda$ avec $\lambda = 1$ (à gauche) et $\lambda = 20$ (à droite) et $n = 20$. 53

3.3 Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 2 avec $a = 0$, $b = 0.5$ et $n = 20$ pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (à droite) 54

3.4 Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/8$ (en première partie) ou $\sigma^2 = 1/4$ (en deuxième partie) ou $\sigma^2 = 1/2$ (en troisième partie) et $n = 20$ pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) and $g(\theta) = \sigma^2$ (à droite) 57

3.5 Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/2$ et $n = 1000$ pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) et $g(\theta) = \sigma^2$ (à droite) 58

3.6 Erreurs quadratiques moyennes par rapport à la vraie moyenne a posteriori dans le modèle 1 pour $g(\lambda) = \lambda$ avec $\lambda = 20$ et $n = 20$ ensemble avec la stratégie de mélange 59

3.7 Erreurs quadratiques moyennes entre l'estimation du rapport des constantes de normalisation et sa vraie valeur. 61

3.8 Erreurs quadratiques moyennes entre l'estimation de l'espérance a posteriori de θ et sa vraie valeur. 62

3.9 Erreurs quadratiques moyennes dans le Modèle 3 avec la Sélection Automatique avec les paramètres $\theta = (a = 0, b = 0.5, \sigma^2 = 1/2)$ pour $g(\theta) = a$ (en première partie), $g(\theta) = b$ (en deuxième partie) et $g(\theta) = \sigma^2$ (en troisième partie) 64

4.1 Rapport des temps de calculs entre les méthodes combinées avec l'échantillonnage préférentiel et la méthode classique MCMC concernant la régression de Poisson avec $n = 1,000$ et 10 covariables. 68

6.1 Les dérivées partielles 98

7.1 Pramétrisations pour α_0, β_0 et i_0 107

7.2 MC Caractéristiques 109

7.3 Le nombre d'estimations « utilisables » 111

7.4 Efficacité relative d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (95% des estimations) 116

7.5 Efficacité relative d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (β-fixe) 120

7.6 Résultats d'estimation par maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after » - nouveaux estimateurs de R_0 pour les observations dans l'état stationnaire 125

1 Model 1 : Estimation of $E_{[\lambda|X]}(g(\lambda))$ for $\lambda = 1$ 152

2 Modèle 1 : Estimation de $E_{[\lambda|X]}(g(\lambda))$ pour $\lambda = 20$ 153

3 Modèle 2 : Estimation de $E_{[\theta|X]}(g(\theta))$ for $\theta = (a = 0, b = 0.5)$ 153

4 Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=$1/8$) 154

5 Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=$1/4$) 154

6 Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=$1/2$) 155

7 Erreurs quadratique moyennes par rapport à MCMC dans le Modèle 3 avec 10 covaribles normales et avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $g(\theta) = \sigma^2$ (extension) 170

8 Erreurs quadratique moyennes par rapport à MCMC dans le Modèle 3 avec 5 covariables normales et 5 covariables de Bernoulli avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $g(\theta) = \sigma^2$ (extension) 172

9 Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 2 avec 10 covaribles normales avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$ 173

10 Les vraisemblances, la densités a priori et a posteriori, les espérances a posteriori et les rapport des constantes de normalisation des modèles : Normal-Normal, Poisson-Gamma, Binomial-Beta, Gamma-Gamma et Négative Binomial Beta. 174

11 Estimation des rapports des constantes de normalisation via trois méthodes : importance sampling ($\hat{r}_{is}$), reciprocal importance sampling ($\hat{r}_{recip}$) et ratio importance sampling ($\hat{r}_{is}$). Les vraies valeurs sont rapportées dans la colonne dénotée par r_0. 175

12 Le nombre des problèmes selon la catégorie 182

13 Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase initiale du processus « early » 201

14 Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase initiale du processus « early » 202

15 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase initiale du processus « early » 203

16 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early » 204

17 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early » (95% des estimations) 205

18 Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase évolutive du processus « during » 206

19 Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase évolutive du processus « during » 207

20 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase évolutive du processus « during » 208

21 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase évolutive du processus « during » 209

22 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (95% des estimations) 210
23 Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase stationnaire du processus « after » 211
24 Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase stationnaire du processus « after » 212
25 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase stationnaire du processus « after » 213
26 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after » 214
27 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « after » (95% des estimations) 215
28 Le nombre des estimations « utilisables » (β fixe) 218
29 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase initiale du processus « early » (β-fixe) 228
30 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early » (β-fixe) 229
31 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase évolutive du processus « during » (β-fixe) 230
32 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase évolutive du processus « during » (β-fixe) . . . 231
33 Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase stationnaire du processus « after » (β-fixe) 232
34 Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after » (β-fixe) . . 233
35 Résultats d'estimation bayésienne du paramètre α avec les données provenant de la phase évolutive du processus « during » 238
36 Résultats d'estimation bayésienne du paramètre β avec les données provenant de la phase évolutive du processus « during » 239

37 Résultats d'estimation bayésienne du paramètre i_0 avec les données provenant de la phase évolutive du processus « during » 240

38 Résultats d'estimation bayésienne du paramètre R_0 avec les données provenant de la phase évolutive du processus « during » 241

Introduction générale

Ce travail se décompose en deux grandes parties, "Estimations répétées dans le cadre de la modélisation bayésienne" et "Modélisation de la transmission de maladies infectieuses dans une population. Estimation des paramètres.". Les techniques développées dans la première partie sont utilisées en fin de la seconde partie.

La première partie est consacrée à des optimisations d'algorithmes stochastiques très souvent utilisés, notamment dans le contexte des modélisations Bayésiennes.

En effet, le résultat d'une modélisation Bayésienne est une loi a posteriori jointe 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 (son espérance, ses quantiles, sa variabilité) permet de caractériser la connaissance actuelle (donc suite à ces nouvelles données) que l'on a sur les paramètres. Ce processus de mise à jour de la connaissance est la base des approches Bayésiennes. Ces approches pourtant séduisantes n'ont été que relativement récemment fortement utilisées (depuis les années 1990). La raison en est simple et est d'ordre pratique. A part dans le cas particulier du choix de lois dites lois conjuguées, le calcul explicite de la loi a posteriori recherchée est souvent impossible. On doit donc faire appel à des techniques d'approximation. Or, ces techniques n'ont été mises à disposition que dans le début des années 90 à travers le logiciel WinBugs. L'approximation de la loi a posteriori se fait le plus souvent à travers une méthode itérative utilisant des algorithmes stochastiques de la famille des algorithmes dits de Monte Carlo par Chaîne de Makov (MCMC). Cette approximation ne peut se faire qu'après convergence de ces algorithmes, ce qui parfois nécessite un assez grand nombre d'itérations. Lors de l'étude de nouvelles approches statistiques ou lors de la mise en place de nouveaux modèles pour répondre plus finement à des problèmes concrets, il est indispensable d'étudier empiriquement les performances des estimations des paramètres même si des résultats asymptotiques existent. Ces études permettent en effet d'avoir un regard à distance finie et dans des situations réalistes sur les qualités attendues des estimateurs. Pour ce faire, on fait appel bien souvent à des simulations de données dans des situations considérées comme réalistes que ce soit du point de vue des tailles d'échantillon

ou des valeurs sous-jacentes des paramètres. Pour chaque jeu de données simulé, l'inférence Bayésienne nécessite alors l'emploi des algorithmes MCMC. Le nombre de jeux de données simulés devant être important pour pouvoir émettre des conclusions, on comprendra que les temps de calculs deviennent très importants. Dans ce contexte, nous proposons donc d'étudier des solutions évitant un trop grand nombre d'appels à ces algorithmes mais permettant bien-sûr d'obtenir malgré tout des résultats précis. La principale technique étudiée dans cette partie est celle de l'échantillonnage préférentiel.

La seconde partie est consacrée aux études de modèles épidémiques, en particulier le modèle compartimental dit SIS (Susceptible-Infecté-Susceptible).

Le modèle SIS permet de modéliser l'évolution d'une maladie dans une population quand on suppose l'existence de uniquement deux sous-populations : les sujets susceptibles (S) et les sujets infectés (I). Un sujet susceptible (S) peut devenir infecté via un contact infectieux uniquement avec un sujet déjà infecté (I). Ce modèle suppose de plus, qu'un sujet guéri ne développe pas de résistance ou d'immunité et donc redevient susceptible. Il est également possible d'introduire des "sorties" ou des "entrées" dans les compartiments à travers des probabilités de décès ou de naissance. Il s'agit donc de modèles très simples dans leur formulation. Deux grandes approches existent alors : l'approche déterministe ou l'approche stochastique. L'approche déterministe suppose que la maladie évolue dans la population de façon homogène. Une solution moyenne des effectifs des compartiments est alors donnée. Cette approche ne prend pas en compte une éventuelle hétérogénéité et ne fournit qu'un comportement "moyen" sans caractériser les déviations autour de cette moyenne. L'approche stochastique permet de prendre en compte cette hétérogénéité (modèle plus adapté par exemple dans le cas de population de petite taille) en introduisant une fluctuation aléatoire. Plusieurs possibilités existent, nous avons principalement retenu les approches par des processus Markoviens. Dans le cas stochastique, selon la forme des probabilités de passage entre les états, on distingue les processus stochastiques linéaires et non linéaires. Les processus stochastiques linéaires sont caractérisés par des probabilités de transition dépendantes de façon linéaire des états, sinon il s'agit de processus stochastiques non linéaires. Dans le cas linéaire, la solution de l'équation

différentielle en probabilité est en général analytique contrairement au cas non linéaire. Nous nous intéresserons donc particulièrement au cas non linéaire et aux techniques dites de développement de l'équation maîtresse ("master equation") qui consistent à exprimer le processus étudié par une somme d'un phénomène déterministe et d'une fluctuation aléatoire dont la grandeur dépend d'un paramètre qui est défini comme la taille du système. Cette méthode a été développée dans le cas des processus dits de naissance et de mort. Nous les appliquerons au modèle SIS avec une taille de population constante. Les propriétés des estimateurs des paramètres sont étudiées dans le cadre fréquentiste et bayésien. Concernant l'approche Bayésienne, les résultats sur l'optimisation algorithmique de la première partie seront alors utilisés.

Estimations répétées dans le cadre de la modélisation bayésienne.

1 Introduction

L'objet de ce travail est de présenter l'Importance Sampling (échantillonnage pondéré ou préférentiel) IS comme une méthode d'optimisation algorithmique dans le cas de l'étude empirique (basée sur des simulations) d'un estimateur dans le cadre d'une modélisation bayésienne. L'étude de simulations permet d'évaluer certaines propriétés statistiques d'un estimateur via des réplications d'événements aléatoires. Afin de caractériser les performances des estimateurs selon différentes situations et de contrôler les fluctuations aléatoires, ce type d'analyse nécessite de simuler sous différentes paramétrisations beaucoup de jeux de données, puis pour chaque jeu de données, d'estimer les paramètres ou des fonctions de ces paramètres. Le contexte de notre travail est un modèle paramétrique sous lequel les jeux de données ont été simulés pour certaines valeurs des paramètres. Concernant l'estimation dans le contexte Bayésien, des lois a priori ont été spécifiées sur les paramètres, ces lois a priori restent les mêmes quelques soient les jeux de données. La démarche bayésienne, comme abordée dans la vaste littérature (Cf. par exemple Robert (2007)), consiste à combiner l'information a priori des paramètres représentée par des lois a priori avec la source d'information provenant des données à travers la vraisemblance pour obtenir la loi a posteriori des paramètres conditionnelle aux données. Quand la loi a posteriori ou quand les moments de cette loi ne sont pas explicites, une approximation est obtenue par des algorithmes stochastiques basés sur les méthodes dites de Monte Carlo par Chaînes de Markov (MCMC) comme présentées par Hastings (1970) ou Geman et Geman (1984). Ces algorithmes permettent d'obtenir des réalisations Markoviennes de la loi a posteriori recherchée et, via la théorie ergodique, d'obtenir ainsi des estimations de ses moments. D'un point de vue pratique, le recours aux algorithmes itératifs MCMC doit être fait pour chaque jeu de données simulé. L'utilisation répétée des algorithmes itératifs MCMC

peut être très coûteuse en temps calcul.

L'objectif de ce travail est d'étudier et d'améliorer l'efficacité de l'utilisation dans les inférences bayésiennes d'une autre technique basée sur l'Importance Sampling (IS) dans le contexte présenté ci-dessus. Cette méthode nécessite le choix d'une fonction d'importance, choix souvent délicat à faire. Dans le cas particulier de l'étude de différents de jeux de données, notre idée consiste à utiliser l'algorithme MCMC pour un nombre limité de jeux de données présélectionnés et ainsi d'obtenir des réalisations de chacune des lois a posteriori correspondantes. Concernant les autres jeux de données, les estimations ne seront pas faites via l'algorithme MCMC mais de manière plus rapide via IS. La fonction d'importance sera alors choisie comme étant fonction des lois a posteriori obtenues précédemment sur les jeux de données présélectionnés. Ce choix est alors essentiel et différentes stratégies sont étudiées dans ce travail. L'idée d'utilisation simultanée de l'IS a été déjà proposée entre autre par Geyer et Thompson (1992), Gelfand (1992) ou plus récemment par McVinish et al. (2008) néanmoins dans des contextes différents.

Pour un lot de jeux de données présélectionnés parmi tous les jeux de données, nous avons testé trois stratégies de choix de la fonction d'importance (appelée ici loi de " référence ") pour le calcul de l'IS : la première stratégie (appelée "référence fixe") consiste à proposer une seule fonction d'importance pour tous les jeux de données en la choisissant par un simple tirage aléatoire parmi les lois a posteriori présélectionnées, la seconde (appelée "référence choisie") permet l'utilisation d'une fonction d'importance différente pour chaque nouvelle estimation (et donc chaque nouveau jeu de données) en la choisissant sur la base d'un critère parmi les lois a posteriori présélectionnées et la troisième (appelée "stratégie du mélange") utilise à nouveau une seule fonction d'importance pour tous les jeux de données mais comme étant un mélange équiprobable des lois a posteriori présélectionnées . Pour la deuxième stratégie ("référence choisie"), il faut choisir pour chaque nouveau jeu de données une loi a posteriori parmi les lois préselectionnées comme fonction d'importance. L'idée est de faire ce choix non pas aléatoirement mais de manière la plus judicieuse afin d'obtenir des estimations de bonne qualité. Nous proposons alors trois critères de choix : Le premier basé sur la minimisation de la norme L_1 de la différence entre deux lois a posteriori, le deuxième basé sur la minimisation

de la divergence de Kullback-Leibler de ces densités a posteriori et finalement le troisième qui minimise la variance de l'estimateur IS. Les trois stratégies ont été comparées sur la base d'erreurs quadratiques moyennes avec les résultats qui auraient été obtenus via MCMC .

Enfin, pour éviter le choix arbitraire des lois a posteriori présélectionnées, une procédure supplémentaire de sélection automatique a été établie.

Les méthodes évoquées ici ont été étudiées sur trois types de modèles poissonniens : le modèle de Poisson avec un paramètre, la régression de Poisson avec une covariable (deux paramètres qui sont l'ordonnée à l'origine et le coefficient associé à la covariable), et la régression de Poisson avec extravariabilité gaussienne (les deux paramètres précédents et la variance résiduelle).

2 Matériel et Méthodes

2.1 Modélisation et inférence Bayésienne

Dans ce travail, nous nous situons dans le contexte de modélisation paramétrique bayésienne des données où l'objectif est alors d'estimer la loi a posteriori des paramètres du modèle. L'idée de base de la démarche bayésienne est de partir de l'information au préalable disponible sur les paramètres, de la "combiner" à l'information contenue dans les données et ainsi d'en déduire une information "mise à jour" sur ces mêmes paramètres. Les paramètres sont considérés comme des variables aléatoires dont les lois représentent la connaissance disponible sur ces paramètres. Les lois dites a priori correspondent aux informations préalables et les lois dites a posteriori aux informations "mises à jour". Si, de manière générale, θ est le paramètre ou le vecteur des paramètres, les données X sont décrites par le modèle probabiliste $\pi(x|\theta)$, $\theta \in \Theta$, où θ est aléatoire et $\pi(\theta)$ est la loi a priori de θ. Afin de combiner les informations a priori et les nouvelles informations contenues dans les données, on recherche la loi a pos-

teriori $\pi(\theta|x)$ sur les paramètres. Cette loi, d'après le Théorème de Bayes, s'écrit comme le rapport de la loi jointe (θ, x) et de la loi marginale de X

$$\pi(\theta|x) = \frac{\pi(\theta, x)}{\pi(x)} \tag{2.1}$$

$$\text{où} \quad \pi(\theta, x) = \pi(\theta) \cdot \pi(x|\theta) \quad \text{et} \quad \pi(x) = \int_\Theta \pi(x|\theta) \cdot \pi(\theta) d\theta$$

alors l'equation (2.1) peut être presentée sous la forme équivalente

$$\pi(\theta|x) = \frac{\pi(\theta) \cdot \pi(x|\theta)}{\int_\Theta \pi(x|\theta) \cdot \pi(\theta) d\theta} \tag{2.2}$$

où la loi conditionnelle $\pi(x|\theta)$ est la fonction de vraisemblance représentant l'information des données. La loi marginale sur X s'appelle classiquement une constante de normalisation de $\pi(\theta|X)$ car ne dépend pas de θ.

Le résultat d'une approche Bayésienne est donc la loi a posteriori des paramètres mais pour des raisons d'interprétation, on s'intéresse également aux résumés statistiques descriptifs de cette loi a posteriori comme : la moyenne a posteriori, l'écart-type a posteriori, les quantiles a posteriori (intervalle de crédibilité IC), etc. De manière générale, ces résumés statistiques correspondent à l'espérance a posteriori d'une fonction $g(\theta), \theta \in \Theta$ pour les différents choix de la fonction g

$$E_{[\theta|X]}\left[g(\theta)\right] = \int_\Theta g(\theta)\pi(\theta|x)d\theta \tag{2.3}$$

L'expression de la fonction de densité a posteriori (2.2) fait intervenir la loi marginale de X qui résulte d'une intégrale. Cette intégrale, à part dans certains cas comme par exemple lors de l'utilisation de familles de lois a priori conjuguées (Cf. Robert, 2007, Chap.3), peut être difficile ou même impossible à calculer. Par conséquent, la vraie loi a posteriori n'est pas explicite (il n'y a pas de solution analytique). De plus, même si la loi a posteriori est explicite, le calcul des espérances de la forme (2.3) peut également ne pas être explicite car à nouveau, fait intervenir une intégrale bien souvent multidimensionnelle. Afin d'obtenir une estimation de $E_{[\theta|X]}\left[g(\theta)\right]$, le recours aux solutions numériques devient indispensable. Si la loi a posteriori $\pi(\theta|X)$ est connue, on peut utiliser un générateur des nombres aléatoires pour simuler un échantillon iid sous cette loi, puis l'espérance est estimée par la moyenne empirique de g sur cet échantillon. Dans le cas contraire, en faisant un changement de la densité d'échantillonnage, on peut

approcher l'espérance via l'estimateur d'échantillonnage préférentiel (*importance sampling*) qui utilise un échantillon iid généré selon cette nouvelle densité. L'estimateur d'échantillonnage préférentiel sous sa forme auto normalisée (*self-normalized importance sampling*) résout le problème des constantes de normalisation inconnues. Ces techniques sont connues comme les méthodes de Monte Carlo pour l'intégration (Cf. par exemple Robert and Casella, 2004, 2010). Le choix générique d'une densité d'échantillonnage n'est pas toujours évident et la résolution de ce problème d'estimation se fait plutôt par l'utilisation des algorithmes beaucoup plus sophistiqués basés sur les méthodes dites de Monte Carlo par chaînes de Markov (MCMC). Ces algorithmes permettent de simuler un échantillon d'une chaîne de Markov dont la loi stationnaire est la loi a posteriori $\pi(\theta|X)$. Cet échantillon n'est donc plus indépendant mais la dépendance est de type Markovien. La théorie ergodique, assure la convergence des moyennes empiriques des fonctions g sur cette chaîne vers les espérances a posteriori. Les deux types de méthodes sont ensuite décrits.

2.1.1 Méthodes de Monte Carlo

Comme mentionné ci-dessus, les méthodes dites de Monte Carlo sont fondées sur les générateurs de nombres aléatoires. Il s'agit de techniques numériques qui servent a générer des variables aléatoires sous certaines lois de probabilité. La plupart des logiciels de calcul scientifique intègrent dans leurs options les générateurs aléatoires. La description de ces générateurs ne sera pas abordée ici, cependant plus de détails concernant ces techniques se trouve par exemple dans l'ouvrage de Devroye (1986) ou Ripley (1987). Cette partie est consacrée au problème d'évaluation de l'espérance (2.3) via l'utilisation de ces générateurs. Ainsi dans ce qui suit par échantillons iid on comprendra les séquences obtenues selon ces générateurs.
Le problème d'estimation s'écrit de manière générale comme (Robert and Casella, 2004, 2010)

$$E_f\left[g(\theta)\right] = \int_{\Theta} g(\theta) f(\theta) d\theta \tag{2.4}$$

où f est une densité de probabilité et g est une fonction quelconque sur un ensemble Θ. Nous supposons ici que l'intégrale (2.4) existe et est finie. Soit $\{\theta_1, \ldots, \theta_m\}$ un échantillon aléatoire iid de taille m généré sous la densité f alors via la loi des grands nombres cet échantillon

permet d'estimer (2.4) par la moyenne empirique de $g(\theta_j)$, $j = 1, \ldots, m$:

$$\bar{g}_m^{MC} = \frac{1}{m} \sum_{j=1}^{m} g(\theta_j) \tag{2.5}$$

Si $E_f \left[g^2(\theta)\right] < \infty$ alors la variance de l'estimateur $\bar{g}_m^{MC}$ dans l'équation (2.5) est

$$var(\bar{g}_m^{MC}) = \frac{1}{m} \int \left(g(\theta) - E_f \left[g(\theta)\right]\right)^2 f(\theta) d\theta \tag{2.6}$$

pouvant être estimée par

$$v_m^{MC} = \frac{1}{m^2} \sum_{j=1}^{m} \left(g(\theta_j) - \bar{g}_m^{MC}\right)^2 \tag{2.7}$$

Pour m suffisamment grand, le théorème central limite établit la convergence en loi de l'estimateur $\bar{g}_m^{MC}$,

$$\frac{\bar{g}_m^{MC} - E_f \left[g(\theta)\right]}{\sqrt{v_m^{MC}}} \sim \mathcal{N}(0, 1) \tag{2.8}$$

2.1.1.1 Échantillonnage pondéré - Importance Sampling.

Afin d'illustrer les techniques d'échantillonnage pondéré, Robert and Casella (2004) présentent l'exemple de Ripley (1987) consistant en l'évaluation de la probabilité que la variable Y de Cauchy $\mathcal{C}(0, 1)$ soit supérieure à 2. Cette probabilité s'exprime comme une intégrale de la forme (2.4) où $g(y) = \mathbb{I}_{Y>2}(y)$ est la fonction indicatrice de l'ensemble $\{Y > 2\}$ et f la densité de la loi de Cauchy $\mathcal{C}(0, 1)$. Cette intégrale peut ainsi être approchée par une moyenne empirique de la forme (2.5). Cependant cette méthode directe n'est pas nécessairement la plus performante et le changement de la loi d'échantillonnage (donc autre que f) peut permettre de réduire la variance de l'estimateur de la probabilité en question. Robert and Casella (2010) fournissent de plus des exemples subséquents où l'échantillonnage direct sous la densité f n'est pas possible. Dans ces deux cas, la solution est obtenue via la simple égalité suivante faisant intervenir un changement de loi sous l'intégrale (2.4)

$$E_f \left[g(\theta)\right] = \int_\Theta g(\theta) f(\theta) d\theta = \int_\Theta g(\theta) \frac{f(\theta)}{h(\theta)} h(\theta) d\theta \tag{2.9}$$

où h est également une densité de probabilité.

Ainsi, avec un échantillon iid $\{\theta_1, \ldots, \theta_m\}$ généré sous la loi h, on peut approcher (2.9) par

$$\bar{g}_m^{IS} = \frac{1}{m} \sum_{j=1}^{m} g(\theta_j) \frac{f(\theta_j)}{h(\theta_j)} \tag{2.10}$$

La fonction h s'appelle la fonction d'importance. La convergence presque sûre de $\bar{g}_m^{IS}$ vers $E_f[g(\theta)]$ (via la relation (2.9)) est assurée par la loi des grands nombres sous la condition que le support de h inclut le support produit $g \times f$ c'est à dire $\text{supp}(f) \supset \text{supp}(g \times f)$. Afin de simplifier les écritures, les poids $w(\theta) = f(\theta)/h(\theta)$ sont introduits, l'estimateur $\bar{g}_m^{IS}$ s'écrit donc comme

$$\bar{g}_m^{IS} = \frac{1}{m} \sum_{j=1}^{m} g(\theta_j) w(\theta_j) \tag{2.11}$$

Nous garderons désormais cette formulation.

Si $E_h\left[g^2(\theta)w^2(\theta)\right] = E_f\left[g^2(\theta)w(\theta)\right] = \int g^2(\theta)\frac{f^2(\theta)}{h(\theta)}d\theta < \infty$, alors la variance de l'estimateur $\bar{g}_m^{IS}$ est

$$var(\bar{g}_m^{IS}) = \frac{1}{m} \int \left(g(\theta)w(\theta) - E_f[g(\theta)]\right)^2 h(\theta) d\theta \tag{2.12}$$

car $E_h[g(\theta)w(\theta)] = E_f[g(\theta)]$. En utilisant l'échantillon iid $\{\theta_1, \ldots, \theta_m\} \sim h$, la variance (2.12) peut alors être estimée par

$$v_m^{IS} = \frac{1}{m^2} \sum_{j=1}^{m} \left(g(\theta_j)w(\theta_j) - \bar{g}_m^{IS}\right)^2 \tag{2.13}$$

Par le théorème central limite, l'estimateur $\bar{g}_m^{IS}$ converge aussi en loi vers une variable normale

$$\frac{\bar{g}_m^{IS} - E_f[g(\theta)]}{\sqrt{v_m^{IS}}} \sim \mathcal{N}(0,1) \tag{2.14}$$

La condition $E_h\left[g^2(\theta)w^2(\theta)\right] = E_f\left[g^2(\theta)w(\theta)\right] = \int g^2(\theta)\frac{f^2(\theta)}{h(\theta)}d\theta < \infty$ garantit que la variance (2.12) de l'estimateur $\bar{g}_m^{IS}$ est finie. Dans ce cas, la queue de la distribution de la fonction d'importance h doit être "plus lourde" que celle de la densité f et de manière générale les poids w doivent être bornés pour éviter la divergence de la variance. Robert and Casella (2004) démontrent dans leur théorème 3.12 que le choix optimal en terme de minimisation de la variance (2.12) est de prendre comme fonction d'importance

$$h^*(\theta) = \frac{|g(\theta)|f(\theta)}{\int |g(\gamma)|f(\gamma)d\gamma} \tag{2.15}$$

Néanmoins, h^* dépend elle même de l'intégrale que l'on souhaite évaluer $\int g(\theta)f(\theta)d\theta$. Les auteurs suggèrent qu'un bon choix en pratique de la fonction d'importance peut être celui pour lequel $|g|f/h$ est le plus stable possible. De plus, l'estimateur IS $\bar{g}_m^{IS}$ fournit des estimations

médiocres sur le plan de la convergence si la condition $\int \frac{f^2(\theta)}{h(\theta)} d\theta < \infty$ n'est pas satisfaite. Ces deux conditions s'appliquent également à l'estimateur IS "normalisé" (*Self-Normalised Importance Sampling*) décrit ci-après présentant l'avantage de pouvoir s'affranchir du calcul de constantes de normalisation.

2.1.1.2 Échantillonnage préférentiel "normalisé" - Self-Normalised Importance Sampling.

Un autre estimateur de $E_f[g(\theta)]$ est une version "normalisée" de $\bar{g}_m^{IS}$. Il est défini de la manière suivante :

$$\bar{g}_m^{SNIS} = \frac{\sum_{j=1}^m g(\theta_j) w(\theta_j)}{\sum_{j=1}^m w(\theta_j)} \tag{2.16}$$

En effet, on peut démontrer que pour un échantillon iid $\theta_1, \ldots, \theta_m \sim h$, $\frac{1}{m}\sum_{j=1}^m w(\theta_j)$ $= \frac{1}{m}\sum_{j=1}^m f(\theta_j)/h(\theta_j)$ converge presque sûrement vers 1, ainsi $\bar{g}_m^{SNIS}$ (2.16) converge vers $E_f[g(\theta)]$. Même si cet estimateur présente un biais, il est toutefois souvent préférable à $\bar{g}_m^{IS}$ car les estimations obtenues via (2.16) sont plus stables, et peuvent avoir une plus petite variance (Robert and Casella, 2004). De plus, comme suggéré plus haut, ce nouvel estimateur peut être utilisé même si les densités f et h sont connues à une constante près. Soient deux constantes c et d telles que : $f(\theta) = \frac{\tilde{f}(\theta)}{c}$ et $h(\theta) = \frac{\tilde{h}(\theta)}{1/d}$ alors les fonctions $\tilde{f}$ et $\tilde{h}$ sont respectivement identiques aux densités f et h à une constante dite de normalisation près c et $1/d$. Posons $\tilde{w}(\theta) = \frac{\tilde{f}(\theta)}{\tilde{h}(\theta)}$, l'estimateur (2.16) s'écrit de manière équivalente :

$$\bar{g}_m^{SNIS} = \frac{\sum_{j=1}^m g(\theta_j) \tilde{w}(\theta_j)}{\sum_{j=1}^m \tilde{w}(\theta_j)} \tag{2.17}$$

car

$$\frac{\sum g(\theta_j) w(\theta_j)}{\sum w(\theta_j)} = \frac{\sum g(\theta_j) \frac{f(\theta)}{h(\theta)}}{\sum \frac{f(\theta)}{h(\theta)}} = \frac{\sum g(\theta_j) \frac{\tilde{f}(\theta)}{c} / \frac{\tilde{h}(\theta)}{1/d}}{\sum \frac{\tilde{f}(\theta)}{c} / \frac{\tilde{h}(\theta)}{1/d}} = \frac{\sum g(\theta_j) \frac{\tilde{f}(\theta)}{\tilde{h}(\theta)} \frac{1}{cd}}{\sum \frac{\tilde{f}(\theta)}{\tilde{h}(\theta)} \frac{1}{cd}} = \frac{\sum g(\theta_j) \tilde{w}(\theta_j)}{\sum \tilde{w}(\theta_j)}$$

Sous les conditions exposées déjà ci-dessus

$$E_f[w(\theta)] < \infty \quad \text{et} \quad E_f\left[g^2(\theta) w(\theta)\right] < \infty \tag{2.18}$$

qui comme dans Douc et al. (2007) peuvent s'écrire conjointement $E_h\left[(1+g^2(\theta))w(\theta)\right] < \infty$ l'estimateur (2.17) a une variance (Geweke, 1989, Th.2) :

$$\begin{aligned} var(\bar{g}_m^{SNIS}) &= \frac{1}{m}\frac{1}{c^2d^2}\Big\{var_h\left[g(\theta)\tilde{w}(\theta)\right] + E_f\left[g(\theta)\right]^2 var_h\left[\tilde{w}(\theta)\right] \\ &\quad -2E_f\left[g(\theta)\right] cov_h\left[g(\theta)\tilde{w}(\theta);\tilde{w}(\theta)\right]\Big\} \qquad (2.19) \\ &= \frac{1}{m}\frac{1}{c^2d^2}\int \big(g(\theta) - E_f\left[g(\theta)\right]\big)^2\tilde{w}(\theta)^2 h(\theta)d\theta \qquad (2.20) \end{aligned}$$

Nous pouvons voir que l'estimateur $\bar{g}_m^{SNIS}$ (formule 2.17) et sa variance peuvent s'exprimer uniquement en fonction des poids $\tilde{w}$ et donc en fonction de $\tilde{f}$ et $\tilde{h}$. Cette possibilité d'utiliser les densités à une constante près est très utile dans la pratique. En effet, rappelons que la densité f correspond dans la démarche bayésienne à la densité a posteriori des paramètres θ qui fait intervenir la loi marginale des données. Ce dernier terme est le résultat d'une intégrale souvent complexe voire non explicite. Cet estimateur IS "normalisé" permet d'éviter ce calcul.

De plus, on peut également montrer (Cf. annexe (IV)) la convergence asymptotique de $\bar{g}_m^{SNIS}$ vers une loi normale centrée en $E_f\left[g(\theta)\right]$ et de variance $var(\bar{g}_m^{SNIS})$.

On déduit un estimateur v_m^{SNIS} de la variance de $var(\bar{g}_m^{SNIS})$ (2.20) de la manière suivante :

$$v_m^{SNIS} = \frac{\sum_{j=1}^m \big(g(\theta_j) - \bar{g}_m^{SNIS}\big)^2\tilde{w}^2(\theta_j)}{\big(\sum_{j=1}^m \tilde{w}(\theta_j)\big)^2} \qquad (2.21)$$

car

$$\frac{\frac{1}{m}\frac{1}{m}\sum_{j=1}^m \big(g(\theta_j) - \bar{g}_m^{SNIS}\big)^2\tilde{w}^2(\theta_j)}{\frac{1}{m^2}\big(\sum_{j=1}^m \tilde{w}(\theta_j)\big)^2} \xrightarrow{p.s.} \frac{\frac{1}{m}\int \big(g(\theta) - E_f\left[g(\theta)\right]\big)^2\tilde{w}^2(\theta)h(\theta)d\theta}{c^2d^2} \qquad (2.22)$$

Le théorème centrale limite s'applique et on obtient

$$\frac{\bar{g}_m^{SNIS} - E_f\left[g(\theta)\right]}{\sqrt{v_m^{SNIS}}} \sim \mathcal{N}(0,1) \qquad (2.23)$$

Les deux conditions d'existence de la variance (2.18) peuvent être vérifier de deux manières (voir Geweke (1989))

$$w(\theta) < \infty \quad \forall\theta \in \Theta \quad \text{et} \quad var_f\left[g(\theta)\right] < \infty \qquad (2.24)$$

ou

$$\Theta \quad \text{est compact et} \quad f(\theta) < \infty \quad \text{et} \quad h(\theta) > \epsilon > 0 \quad \forall\theta \in \Theta \qquad (2.25)$$

Ces conditions ne sont pas nécessaires pour établir le bon comportement de l'estimateur $\bar{g}_m^{SNIS}$, cependant elles peuvent servir comme un point de départ dans la construction ou dans le choix de la fonction d'importance h.

2.1.2 Méthodes de Monte Carlo par Chaîne de Markov (MCMC)

La famille des méthodes MCMC dites de Monte Carlo par Chaîne de Markov est très large et reste un champ de recherche actuelle. Les analyses effectuées dans ce travail on été faites sur le logiciel WinBUGS et sous R avec le packages BRUGS et R2WinBUGS. Une sélection des algorithmes MCMC utilisés par ses logiciels sera présentée ici mais de manière très générale. L'objectif est plutôt de montrer leurs cotés itératifs sans élaborer tous les aspects techniques.

2.1.2.1 Les éléments théoriques de base

Pour estimer l'intégrale (2.4) l'échantillon $\{\theta_1, \ldots, \theta_n\}$ n'est pas nécessairement constitué de réalisations indépendantes. Il suffit que chaque θ_i soit effectivement tiré selon la loi f - la densité cible, mais une certaine forme de dépendance peut exister. Une manière de le faire est de construire une chaîne de Markov avec f comme densité stationnaire (Gilks et al., 1995b).

$\{\theta_0, \ldots, \theta_n\}$ est une chaîne de Markov si pour tout $t > 0$ la probabilité conditionnelle de θ_t sachant $\{\theta_1, \ldots, \theta_{t-1}\}$ dépend seulement de θ_{t-1}

$$P(\theta_t|\theta_{t-1}, \ldots, \theta_0) = P(\theta_t|\theta_{t-1}) = K(\theta_{t-1}, \theta_t) \tag{2.26}$$

où K est le noyau de transition de la chaîne de Markov. Dans le cas de variables continues, K est une densité conditionnelle et dans le cas discret, K est une matrice de transition entre les états discrets de la chaîne. Ainsi, si l'état actuel est θ_{t-1}, l'état suivant θ_t de la chaîne peut être généré selon le noyau de transition $K(\theta_{t-1}, \theta_t)$. Si f, la densité stationnaire de la chaîne existe, alors, sous le régime stationnaire, $\theta_{t-1} \sim f$ implique que $\theta_t \sim f$. Elle est liée au noyau de transition K via la relation

$$\int_\Theta K(\theta, \omega) f(\theta) d\theta = f(\omega) \tag{2.27}$$

Les méthodes de Monte Carlo par Chaîne de Markov (MCMC) consistent alors à choisir le noyau de transition K de manière que f - la densité recherchée, soit la densité stationnaire

unique de la chaîne. La relation (2.27) est alors satisfaite quelque soit l'état initial θ_0. Robert and Casella (2004) définissent ces méthodes comme des procédures de simulation d'une chaîne de Markov ergodique dont la densité stationnaire est la densité f. Dans ces procédures, le temps t est alors le numéro de l'itération.

Des conditions théoriques (pas développées ici en détail) sur le noyau de transition utilisé sont nécessaires. De manière générale, le noyau K doit permettre des transitions vers tous les états quelque soit l'état actuel afin que le chaîne puisse visiter l'ensemble du support Θ (irréducibilité) c'est-à-dire que la probabilité d'atteindre un sous-ensemble non-vide de Θ soit non-nulle ($K(\theta, \cdot) > 0$). De plus, la chaîne doit pouvoir atteindre tout sous-ensemble non-vide de Θ une infinité de fois (récurrence). Enfin, la loi asymptotique de la chaîne ainsi obtenue doit être la loi stationnaire de f (ergodicité).

Pour que la chaîne puisse atteindre son régime stationnaire puis explorer correctement l'ensemble du support, il peut être nécessaire d'avoir un grand nombre d'itérations (probablement plus grand que pour un échantillon iid). Le problème de la convergence sera abordé ultérieurement.

Parmi les algorithmes MCMC, les deux algorithmes classiques sont l'algorithme de Hastings-Metropolis (Metropolis et al., 1953; Hastings, 1970) et l'échantillonneur de Gibbs (Geman and Geman, 1984). Le dernier sera presenté comme un cas particulier du premier. De plus une méthode adaptative d'échantillonnage (*Derivative-free Adaptive Rejection Sampling*) de Gilks (1992) et le *Slice sampling* de Neal (1997, 2003) sont présentées.

2.1.2.2 Algorithme de Hastings-Metropolis

Pour cet algorithme, la loi selon laquelle on souhaite simuler un échantillon doit être connue à une constante de normalisation près. Soit θ_t un état courant de la chaîne, pour chaque itération suivante l'algorithme échantillonne un candidat y selon une loi instrumentale (loi de proposition) $q(y|\theta_t)$. Le candidat y est accepté avec la probabilité $\alpha(\theta_t, y)$.

$$\alpha(\theta_t, y) = \min\left\{1, \frac{f(y)q(\theta_t|y)}{f(\theta_t)q(y|\theta_t)}\right\} \tag{2.28}$$

Si le candidat y est retenu, il est noté θ_{t+1}, sinon θ_{t+1} prend la même valeur que dans l'itération précédente : $\theta_{t+1} = \theta_t$. La décision d'acceptation ou de rejet du candidat y est prise à l'aide

FIGURE 2.1: Algorithme de Hastings-Metropolis.

---

tirer $y \sim q(y|\theta_t)$
tirer $u \sim \mathcal{U}nif[0,1]$

$$\theta_{t+1} = \begin{cases} y & \text{si} \quad u \leq \alpha(\theta_t, y) \\ \theta_t & \text{sinon} \end{cases}$$

---

d'une variable uniforme $u \sim \mathcal{U}nif[0,1]$. Si $u \leq \alpha(\theta_t, y)$ le candidat est accepté. L'itération $t+1$ de l'algorithme de Hastings-Metropolis pour l'état courant θ_t est présentée dans le figure (2.1).

Si la loi instrumentale choisie est symétrique $q(y|\theta) = q(\theta|y)$ (comme par exemple une marche aléatoire dans Metropolis et al. (1953)) la probabilité d'acceptation devient $\alpha(\theta_t, y) = \min\{1, f(y)/f(\theta_t)\}$. Tierney (1994) propose d'utiliser une loi instrumentale non-conditionnelle : $q(y|\theta) = q(y)$. Toute $q(y|\theta)$ doit néanmoins couvrir le support de la densité f. Il est souvent nécessaire de modifier/adapter les paramètres de variabilité de $q(y|\theta)$ pendant les premières itérations pour que la probabilité d'acceptation ne soit ni trop faible ni trop forte. En effet, une probabilité d'acceptation trop faible suggère que les candidats y proposés sont trop peu probables (trop éloignés du support de f) et une probabilité d'acceptation trop forte témoigne d'un comportement de l'algorithme trop local et donc de la non visite des queues de distribution de f. Les itérations de cette phase préliminaire sont donc écartées.

Cette méthode s'applique en théorie pour les lois multidimensionnelles où θ est donc un vecteur. Cependant, pour des grandes dimensions, le choix de la densité de proposition q n'est pas toujours facile et il peut être calculatoirement plus efficace (Gilks et al., 1995b) d'utiliser la méthode de Hastings-Metropolis successivement pour les différentes composantes θ^i de $\theta = \{\theta^1, \ldots, \theta^k\}$ (*single-component* Hastings-Metropolis methode). θ^i peut aussi être un vecteur ou un simple scalaire. Notons θ^{-i} le vecteur θ privé de la composante θ^i. A chaque itération, l'algorithme actualise une seule composante à la fois. Ainsi à l'entrée de l'itération $t+1$, la $i^{\text{ième}}$ composante est θ_t^i et les composantes restantes $\theta_t^{-i} = \{\theta_{t+1}^1, \ldots, \theta_{t+1}^{i-1}, \theta_t^{i+1}, \ldots, \theta_t^k\}$, les composantes précédentes étant ainsi déjà actualisées. Un candidat y est généré selon une loi

FIGURE 2.2: Algorithme de Hastings-Metropolis.

---

tirer $y \sim q_1(y|\theta_t^1, \theta_t^2, \ldots, \theta_t^k)$
tirer $u \sim \mathcal{U}nif[0,1]$

$$\theta_{t+1}^1 = \begin{cases} y & \text{si} \quad u \leq \alpha(\theta_t^1, \{\theta_t^2, \ldots, \theta_t^k\}, y) \\ \theta_t^1 & \text{sinon} \end{cases}$$

- -

tirer $y \sim q_2(y|\theta_{t+1}^1, \theta_t^2, \ldots, \theta_t^k)$
tirer $u \sim \mathcal{U}nif[0,1]$

$$\theta_{t+1}^2 = \begin{cases} y & \text{si} \quad u \leq \alpha(\theta_t^2, \{\theta_{t+1}^1, \theta_t^3, \ldots, \theta_t^k\}, y) \\ \theta_t^2 & \text{sinon} \end{cases}$$

- -

$\vdots$ $\vdots$ $\vdots$

- -

tirer $y \sim q_k(y|\theta_{t+1}^1, \ldots, \theta_{t+1}^{k-1}, \theta_t^k)$
tirer $u \sim \mathcal{U}nif[0,1]$

$$\theta_{t+1}^k = \begin{cases} y & \text{si} \quad u \leq \alpha(\theta_t^k, \{\theta_{t+1}^1, \ldots, \theta_{t+1}^{k-1}\}, y) \\ \theta_t^k & \text{sinon} \end{cases}$$

---

instrumentale $q_i(y|\theta_t^{-i})$ qui est accepté avec la probabilité

$$\alpha(\theta_t^i, \theta_t^{-i}, y) = \min\left\{1, \frac{f(y|\theta_t^{-i}) q_i(\theta_t^i|\theta_t^{-i})}{f(\theta_t^i|\theta_t^{-i}) q_i(y|\theta_t^{-i})}\right\} \tag{2.29}$$

où $f(\theta^i|\theta^{-i})$ est la densité conditionnelle

$$f(\theta^i|\theta^{-i}) = \frac{f(\theta)}{\int f(\theta) d\theta^i} \tag{2.30}$$

Pour l'ensemble des composantes, la $t+1^{\text{ième}}$ itération de l'algorithme de Hastings-Metropolis est présentée dans le figure (2.2).

2.1.2.3 Échantillonneur de Gibbs

Si la loi instrumentale $q_i(y|\theta^{-i})$ de l'algorithme de Hastings-Metropolis présenté précedemment est la loi conditionnelle particulière $f(y|\theta^{-i})$ définie dans (2.30) alors on obtient un algorithme dit échantillonneur de Gibbs (Geman and Geman, 1984). En remplaçant la loi de proposition q_i par la densité conditionnelle $f(\cdot|\theta^{-i})$ dans l'expression de la probabilité d'acceptation (2.29), alors cette probabilité vaut uniformément 1. L'algorithme de Gibbs génère donc des réalisations sous la loi stationnaire f à partir de ses densités conditionnelles

FIGURE 2.3: Échantillonneur de Gibbs.

---

tirer $\theta_{t+1}^1 \sim f(\theta^1|\theta_t^1, \ldots, \theta_t^k)$
tirer $\theta_{t+1}^2 \sim f(\theta^2|\theta_{t+1}^1, \theta_t^3, \ldots, \theta_t^k)$
$\vdots$
tirer $\theta_{t+1}^k \sim f(\theta^k|\theta_{t+1}^1, \ldots, \theta_{t+1}^{k-1})$

$f(\theta^i|\theta^{-i})$. La $t+1^{ime}$ itération de l'algorithme de Gibbs avec comme valeur courante de la chaîne $\theta_t = \{\theta_t^1, \ldots, \theta_t^k\}$ est présentée dans la figure (2.3)

A chaque itération de l'algorithme, une réalisation de la densité conditionnelle $f(y|\theta^{-i})$ est simulée. Si cette densité a une forme analytique appartenant à la famille des densités conjuguées, une réalisation de cette densité peut être obtenue directement via un générateur aléatoire (Devroye, 1986; Ripley, 1987). Si les constantes de normalisation de $f(y|\theta^{-i})$ sont inconnues alors d'autres techniques doivent être employées. Il en existe beaucoup mais seulement les méthodes sur lesquelles sont fondées les algorithmes de WinBUGS sont décrites brièvement.

Si les densités ne peuvent pas être simulées directement, WinBUGS utilise la sélection suivante de stratégies : si la densité est non-conjuguée mais log-concave un échantillonneur *derivative-free adaptive rejection sampling* de Gilks (1992) est utilisé, si la densité conditionnelle est définie sur un support borné, le *Slice Sampling* de Neal (1997) est utilisé et finalement pour les densités dont le support n'est pas borné, WinBUGS se sert de l'algorithme de Hastings-Metropolis. Pour les variables discrètes, la méthode d'inversion (Ripley, 1987) est utilisée.

2.1.2.4 Méthodes hybrides

WinBUGS profite de la structure hiérarchique du modèle étudié pour construire un échantillonneur de Gibbs. A chaque niveau de la hiérarchie du modèle, la méthode la plus adaptée selon la description ci-dessus et utilisée. L'algorithme de Gibbs ainsi obtenu est donc une version hybride. Les méthodes *derivative-free adaptive rejection sampling* et *Slice Sampling*

FIGURE 2.4: Algorithme de Rejection Sampling.

répéter jusqu'à l'acceptation de x tirer $x \sim q(x)$
tirer $u \sim \mathcal{U}nif[0,1]$
si $u \leq \frac{f(x)}{q(x)}$ accepter x

sont présentées.

Adaptive Rejection Sampling (ARS)

Avant d'introduire cette méthode adaptative, la méthode de base *rejection sampling* est présentée qui peut d'ailleurs être utilisée pour simuler la densité conditionnelle $f(y|\theta^{-i})$. Elle permet de générer un échantillon iid selon la densité f si cette dernière est connue à une constante de normalisation près. Une densité instrumentale q est requise facile à simuler et telle que $q(x) \geq f(x)$ sur ensemble du support de f. La densité q s'appelle une enveloppe. Un candidat x est accepté avec la probabilité $f(x)/q(x)$. La procédure d'échantillonnage continue jusqu'à l'acceptation d'un candidat x. L'algorithme est présenté sur le figure (2.4)

La rapidité et l'efficacité de la méthode dépend du choix de l'enveloppe q. Plus elle est proche de la densité f, plus la probabilité d'acceptation est grande.

Gilks (1992) (Cf. aussi Gilks and Wild, 1992) a proposé une construction adaptative de l'enveloppe q basée sur les sécantes pour les densités univariées log-concaves. Son utilisation est donc assez polyvalente car beaucoup de densités satisfont cette propriété. Il existe une autre version du même algorithme basée sur les tangentes mais elle ne sera pas présentée ici. Pour la première méthode de sécantes, il faut d'abord choisir n ordonnées x_i dans le support de f augmentées de la borne inférieure x_0 et de la borne supérieure x_{n+1} (les bornes peuvent être infinies). Cet ensemble d'ordonnées doit être rangé en ordre croissant soit $S_n = \{x_i; i, j = 0, \ldots, x_{n+1}, x_i < x_j \quad \text{si} \quad i < j\}$. Ensuite pour tout $1 \leq i < j \leq n$, les sécantes $L_{i,j}(x)$ passant par les points $(x_i, \ln f(x_i))$ et $(x_j, \ln f(x_j))$ sont construites. Sur chaque intervalle $[x_i, x_{i+1}[$ on définie une fonction linéaire par morceau $h_i(x)$.

$$h_i(x) = \min\left[L_{i-1,i}(x), L_{i+1,i+2}(x)\right] \tag{2.31}$$

On suppose que $\min(u, v) = u$ si v n'est pas défini. Puisque f est une fonction log-concave, la fonction $H_n(x) = \sum_{i=1}^{n} h_i(x)\mathbb{1}_{[x_i,x_{i+1}[}(x)$ est une enveloppe de la fonction $\ln f$ où $\mathbb{1}_{[x_i,x_{i+1}[}(x)$ est la fonction indicatrice de l'ensemble $[x_i, x_{i+1}[$. La construction de $H_n(x)$ est présentée sur le graphique (2.5)

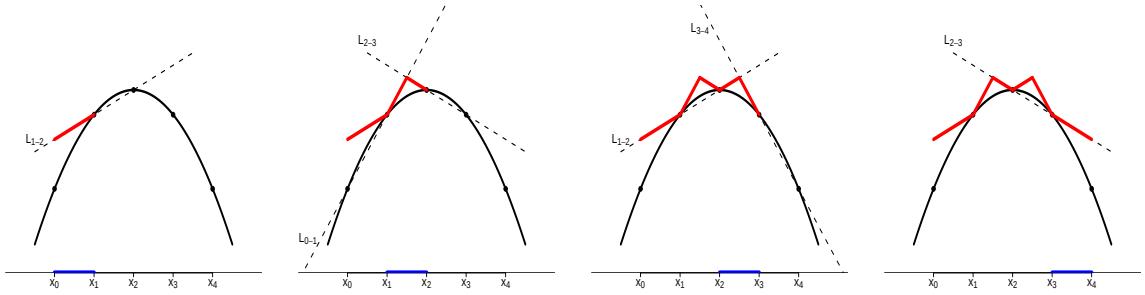

FIGURE 2.5: Construction de la fonction enveloppe dans l'algorithme Adaptive Rejection Sampling pour $n = 3$

L'enveloppe $Q_n(x)$ de la fonction f est donc une fonction exponentielle par morceau $Q_n(x) = \sum_{i=1}^{n} q_i(x)\mathbb{1}_{[x_i,x_{i+1}[}(x)$ où $q_i(x) = \frac{1}{m_i}\exp h_i(x)$ avec $m_i = \int \exp h_i(x)dx$. q_i peut être simulé directement (Devroye, 1986). Pour cette fonction instrumentale Q_n, l'algorithme classique de *rejection sampling* est utilisé, cependant si un candidat est rejeté il est est ajouté à l'ensemble S_n. Le nouveau ensemble $S_{n+1} = S_n \cup x$ est ensuite rangé à nouveau dans l'ordre croissant. La nouvelle enveloppe Q_{n+1} est construite et l'algorithme de *rejection sampling* recommence. Chaque candidat rejeté augmente la probabilité d'acceptation ($\lim_{n\to\infty} \frac{f(x)}{Q_n(x)} = 1$) car l'enveloppe Q_n s'approche de la fonction f. Le schéma de cet algorithme est présenté dans la figure (2.6).

Si le support de f n'a pas de borne inférieure, l'ordonnée x_0 doit être choisie de manière que la sécante $L_{1,2}$ soit croissante. Si la borne supérieure n'existe pas, l'ordonnée x_{n+1} doit être choisie telle que la sécante $L_{n-1,n}$ soit décroissante. Gilks et al. (1995a) précise de plus que si f est une densité conditionnelle $f(y|\theta^{-1})$ (comme dans l'algorithme de Gibbs), l'enveloppe $Q_n(x)$ construite lors de l'itération précédente peut être utilisée et réactualisée à chaque nouveau pas du Gibbs augmentant ainsi la probabilité d'acceptation. Si la densité conditionnelle $f(y|\theta^{-i})$ est multivariée $y = (y^1, y^2, \ldots, y^k)$, la méthode adaptative peut être toujours utilisée car

FIGURE 2.6: Algorithme Adaptive Rejection Sampling.

fixer n and S_n	construire H_n et calculer Q_n
tirer $x \sim Q_n(x)$ (*) tirer $u \sim \mathcal{U}nif[0,1]$	
si $u > \frac{f(x)}{Q_n(x)}$:	- fixer $S_{n+1} = S_n \cup x$ - construire H_{n+1} est calculer Q_{n+1} - revenir au (*)
si $u \leq \frac{f(x)}{Q_n(x)}$:	accepter x

chaque $f(y^m|\theta^{-i})$ est proportionnelle à $f(y|\theta^{-i})$ et donc log-concave par rapport au y^m (Gilks and Wild, 1992).

Slice Sampling

La méthode *Slice Sampling* a été proposée par Neal (1997) (Cf. aussi Neal, 2003) et elle est principalement composée de trois étapes dont deux peuvent être construites avec les différents schémas. Par souci de clarté, seulement le concept général est décrit ici. Si la loi $f(x)$ dont on souhaite obtenir des réalisations est univariée et connue à la constante de normalisation près et si elle est proportionnelle à une fonction $g(x)$, alors une nouvelle variable auxiliaire $u \in \mathbb{R}$ est introduite de manière à ce que la densité jointe de x et u soit uniforme sur l'ensemble $Z = \{(x,u) : 0 < u < g(x)\}$ qui est la surface sous la courbe $g(x)$. La densité jointe $f(x,y)$ s'écrit comme

$$f(x,u) = \begin{cases} \frac{1}{c} & \text{si } 0 < u < g(x) \\ 0 & \text{sinon} \end{cases}$$

avec $c = \int_Z g(x)dx$ tandis que la densité marginale de x est égale à la densité cible $f(x)$

$$f(x) = \int_0^{g(x)} \frac{1}{c} du = \frac{g(x)}{c} \tag{2.32}$$

L'idée de *Slice Sampling* consiste à simuler un couple sous la loi jointe et de ne garder que x. L'échantillonnage direct n'est pas toujours facile donc une chaîne de Markov avec $f(x,u)$ comme densité stationnaire est construite par l'échantillonneur de Gibbs. La $t+1^{\text{ième}}$ itération de l'algorithme est dans la figure (2.7).

FIGURE 2.7: Slice Sampling

---

tirer $u_{t+1} \sim \mathcal{U}nif[0, g(x_t)]$
tirer $x_{t+1} \sim \mathcal{U}nif\,[\{x : 0 \leq u_{t+1} \leq g(x)\}]$

FIGURE 2.8: Version simple de Slice Sampling

---

tirer $u_{t+1} \sim \mathcal{U}nif[0, g(x_t)]$
répéter le tirage $x_{t+1} \sim \mathcal{U}nif[I]$ tant que $u_{t+1} \leq g(x_{t+1})$
où $I = \text{Supp}(f) = [a, b]$, $-\infty < a < b < +\infty$

Le deuxième tirage s'effectue uniformément dans la région $S = \{x : 0 \leq u \leq g(x)\}$ que Neal (1997) appelle le *slice* (tranche). Le tirage de u sous la loi uniforme (étape a) ne pose aucune difficulté pratique tandis que l'échantillonnage uniforme dans S requiert la connaissance de la forme exacte de S ce qui est possible seulement si toutes les solutions de $u = g(x)$ sont calculables. Neal (1997) propose donc de construire autour de x_t un intervalle $I = (L, R)$ contenant tout S ou au moins une bonne partie de S (étape b) selon un des quatre schémas. Dans le premier schéma (1), $L = inf(S)$ et $R = sup(S)$ si S peut être établit analytiquement. Dans le deuxième schéma (2), l'intervalle I correspond au support de f s'il est borné. Les deux autres procédures *stepping out procedure* (3) et *doubling procedure* (4) consistent à choisir aléatoirement un intervalle contenant x_t et à l'élargir itérativement jusqu'à ce que L et R se trouvent en dehors de S.

La dernière étape (c) consiste à simuler x_{t+1} uniformément dans l'intervalle I en vérifiant si le nouveau candidat appartient à l'intersection $S \cap I$ (i). Il est possible de rétrécir l'intervalle I à chaque fois que le candidat n'est pas accepté (ii) via une *shrinkage procedure*. A titre d'exemple pour le cheminement (a),(b) avec (2), (c) avec (i), l'algorithme *Slice Sampling* est présenté dans la figure (2.8).

Dans cet example, l'algorithme n'est pas le plus efficace mais il montre sont coté itératif. Les autres procédures (Neal, 1997, 2003) n'alourdissent pas particulièrement l'algorithme. La méthode se généralise au cas multivarié. Elle était pensée comme une alternative aux

FIGURE 2.9: Inversion pour les variables discrètes.

---

tirer $u \sim \mathcal{U}nif[0,1]$

$$x = \begin{cases} x_0 & u < p_0 \\ x_1 & p_0 \leq u < p_1 \\ \vdots & \\ x_j & \sum_{i=1}^{j-1} p_i \leq u < \sum_{i=1}^{j} p_i \\ \vdots & \end{cases}$$

---

algorithmes de Hastings-Metropolis et de l'échantillonneur de Gibbs mais elle produit des chaînes très corrélées et un grand nombre d'itérations peut être nécessaire pour atteindre la convergence.

Algorithme de Hastings-Metropolis dans Gibbs

A l'intérieur de l'échantillonneur de Gibbs, on peut également utiliser l'algorithme de Hastings-Metropolis (Cf. section (2.1.2.2)) (*Hastings within Gibbs*).

Inversion pour les variables discrètes

Pour les loi discrètes, l'échantillonnage est direct via une inversion de la fonction de répartition. D'après Robert and Casella (2004), on définit pour une fonction non-décroissante F une inverse généralisée F^{-1} comme

$$F^{-1}(u) = \inf \{x : F(x) \geq u\} \tag{2.33}$$

Si $U \sim \mathcal{U}nif[0,1]$ est une variable aléatoire uniforme, la variable $F^{-1}(U)$ suit la loi F. La méthode d'inversion est directe si la forme exacte de F^{-1} est connue. Dans le cas contraire, l'equation $F(x) = u$ doit être résolue numériquement ce qui est peu envisageable pour des variables continues (Devroye, 1986). Pour une variable discrète X telle que la probabilité $P(X = x_i) = p_i$, $i = 0, 1, \ldots$, avec $\sum_i p_i = 1$ et $F(x_j) = \sum_{i=0}^{j} p_i$, la procédure des simulation est présentée dans la figure (2.9).

Le point commun entre toutes ces méthodes est leur convergence théorique vers la densité que l'on souhaite évaluer. La théorie ergodique garantie par la suite la convergence des moments empiriques vers les vrais moments de cette loi. En pratique, il faut toutefois souvent de longues réalisations des chaînes générées par les algorithmes MCMC avant que leur convergence soit atteinte. Différents outils de diagnostique de convergence existent Cowles and Carlin (1996); Brooks and Roberts (1998); Mengersen et al. (1999).

2.2 MCMC combiné avec Importance Sampling

Dans cette section, la méthode d'échantillonnage préférentiel combinée (IS combiné) avec les algorithmes MCMC est présentée dans le cadre d'estimations répétées. Le contexte général de ce type d'étude ainsi que le démarche classique seront présentés avant d'aborder les stratégies combinées mises en place lors de ce travail de thèse.

On parle ici d'estimations répétées si l'intégrale (2.3) doit être évaluée pour un certain nombre de jeux de données : $X^{(1)}, X^{(2)}, \ldots, X^{(K)}$. De plus pour chaque jeu de données, la loi a posteriori est supposée de ne pas être explicite (il n'y a pas de solution analytique). Classiquement, pour le jeu de données k, la solution est apportée par l'utilisation des algorithmes MCMC. Une chaîne de Markov $\theta_1, \ldots, \theta_N$ distribuée asymptotiquement selon la loi a posteriori $\pi(\theta|X^{(k)})$ et par théorème ergodique on obtient

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i) \longrightarrow E_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.34}$$

avec la convergence presque sûre. Puisque la loi a posteriori dépend des observations $X^{(k)}$ alors la chaîne et l'estimation bayésienne d'une fonction $g(\theta)$ seront affectées par ces observations. On constate que pour deux échantillons de données $X^{(1)}$ et $X^{(2)}$ indépendants provenant du même modèle probabiliste, on obtiendra deux chaînes $\theta_1^{(1)}, \ldots, \theta_M^{(1)}$ et $\theta_1^{(2)}, \ldots, \theta_M^{(2)}$ et donc deux estimateurs $\frac{1}{M}\sum_{i=1}^{M} g(\theta_i^{(1)})$ et $\frac{1}{M}\sum_{i=1}^{M} g(\theta_i^{(2)})$. En pratique, on a rarement deux échantillons identiques alors on obtient deux estimateurs différents même si dans les deux cas la même loi a priori $\pi(\theta)$ est choisie. Ce raisonnement favorise une étude des performances des estimateurs des paramètres.

2.2.1 Démarche classique avec MCMC

Quand les calculs analytiques ne sont pas accessibles, un moyen de tester les performances des estimateurs d'un ou plusieurs paramètres est de simuler sous le modèle $\pi(x|\theta)$ différents jeux de données en fixant les paramètres, puis de les estimer. S'il existe une solution analytique de la loi a posteriori, ainsi que des moments de cette loi, il est possible de déterminer la qualité de cette estimation par rapport aux vraies valeurs sinon, on a recours à chaque fois aux algorithmes MCMC. Imaginons que L différentes situations caractéristiques sont intéressantes à étudier (et donc L valeurs possible des paramètres), on doit alors étudier les performances des estimateurs sous ces L modèles probabilistes. Pour chaque modèle l ($l \in 1, ...L$), on simule K réplications de jeux de données afin d'obtenir des résultats qui tiennent compte des fluctuations aléatoires. Ainsi, dans le cadre d'approches bayésiennes, l'étude des performances d'un estimateur nécessite de faire $L \times K$ fois l'algorithme MCMC. Même si cette méthode est très efficace quand elle est répétée $L \times K$ fois, elle peut être coûteuse en temps. Dans le cas de réplications de $K = 101$ jeux de données $X^{(1)}, \ldots, X^{(101)}$ sous un seul modèle (L=1), la démarche se présente sur la figure 2.10.

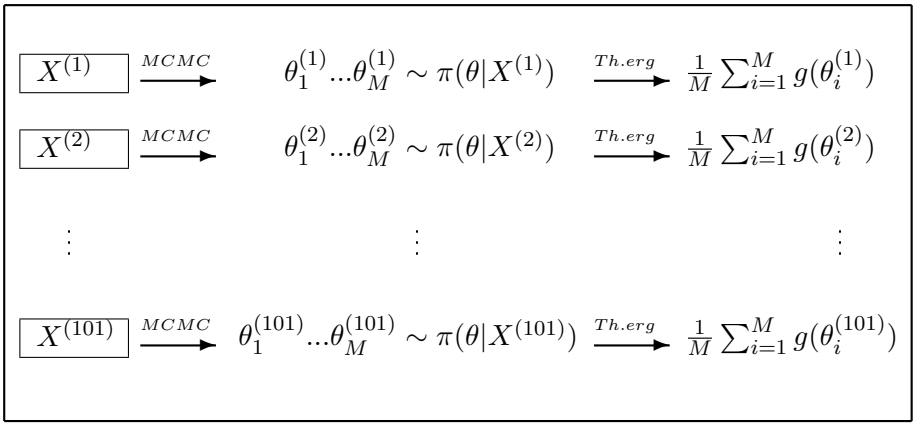

FIGURE 2.10: Démarche classique via MCMC

En raison de ce grand nombre de calculs, on peut vouloir chercher une autre méthode d'estimation qui permettrait de garder la qualité d'estimation obtenue avec MCMC mais en gagnant du temps.

2.2.2 Introduction de l'échantillonnage pondéré dans les estimations

Une méthode permettant d'éviter d'utiliser à chaque fois les algorithmes MCMC est l'échantillonnage pondéré (Importance sampling). Pour deux jeux de données $X^{(m)}$ et $X^{(k)}$ distribués selon la même loi $\pi(x|\theta)$ (donc simulés sous le même modèle probabiliste faisant intervenir les mêmes valeurs des paramètres), l'idée de l'importance sampling dans le cas d'estimation bayésienne de $E_{[\theta|X^{(k)}]}[g(\theta)]$ est basée simplement sur la transformation suivante :

$$\begin{aligned} E_{[\theta|X^{(k)}]}[g(\theta)] &= \int g(\theta)\boldsymbol{\pi(\theta|X^{(k)})}d\theta \\ &= \int g(\theta)\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m)})}\boldsymbol{\pi(\theta|X^{(m)})}d\theta \end{aligned} \tag{2.35}$$

avec $\pi(\theta|X^{(m)})$ non nulle.

Si on a déjà obtenu une réalisation de la loi a posteriori $\pi(\theta|X^{(m)})$, c'est-à-dire une chaîne $\theta_1^{(m)},\ldots,\theta_N^{(m)}$ simulée selon $\pi(\theta|X^{(m)})$ avec un algorithme MCMC alors, par Monte Carlo (théorème ergodique), on obtient l'approximation concernant l'autre jeu de données $X^{(k)}$

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i^{(m)})\frac{\pi(\theta_i^{(m)}|X^{(k)})}{\pi(\theta_i^{(m)}|X^{(m)})} \to E_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.36}$$

Seules les réalisations $\theta_1^{(m)},\ldots,\theta_N^{(m)}$ de la loi a posteriori obtenue uniquement pour le premier jeu $X^{(m)}$ sont utilisées. Par la suite on appellera $X^{(m)}$ **le jeux de données de référence** pour le calcul de l'Importance Sampling et $\frac{\pi(\theta_i^{(m)}|x^{(k)})}{\pi(\theta_i^{(m)}|x^{(m)})}$ les poids.

L'expression dans l'approximation ci-dessus peut être simplifiée pourvu que l'on ait choisi la même loi a priori. Sachant que

$$\begin{aligned} \pi(\theta|x^{(k)}) &= \frac{\pi(x^{(k)}|\theta)\pi(\theta)}{\pi(x^{(k)})} \\ \pi(\theta|x^{(m)}) &= \frac{\pi(x^{(m)}|\theta)\pi(\theta)}{\pi(x^{(m)})} \end{aligned}$$

pour chaque $\theta_i^{(m)} \sim \pi(\theta|x^{(m)})$, $i = 1,\ldots,N$, on peut écrire l'approximation (2.36) comme

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i^{(m)})\frac{\pi(x^{(k)}|\theta_i^{(m)})}{\pi(x^{(m)}|\theta_i^{(m)})}\cdot\frac{\pi(x^{(m)})}{\pi(x^{(k)})} \to E_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.37}$$

Dans l'expression précédente, l'approximation n'est en général pas calculable car, les lois marginales $\pi(x^{(m)})$ et $\pi(x^{(m)})$ (constantes de normalisation) ne sont souvent pas explicites.

On utilisera alors la version normalisée de l'estimateur IS comme présentée dans la section (2.1.1.2). Si on note

$$w_i(k,m) = \frac{\pi(x^{(k)}|\theta_i^{(m)})}{\pi(x^{(m)}|\theta_i^{(m)})} \quad \text{et} \quad \tilde{w}_i(k,m) = \frac{w_i(k,m)}{\sum_{i=1}^{N} w_i(k,m)}$$

et si $\theta_1^{(m)}, \ldots, \theta_N^{(m)}$ est une réalisation markovienne sous la loi $\pi(\theta|X^{(m)})$ alors la convergence presque-sûre est vérifiée pour

$$\sum_{i=1}^{N} g(\theta_i^{(m)})\tilde{w}_i(k,m) \longrightarrow E_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.38}$$

La démonstration de la convergence de cette estimateur se trouve dans l'annexe (IV).

Cet estimateur est aussi consistant, ceci étant la conséquence de la loi forte des grands nombres. Comme vu auparavant dans la section (2.1.1.2), Geweke (1989) étudie l'estimateur IS normalisé dans le cas des réalisations iid. Ses conclusions peuvent être élargies aux cas de réalisations markoviennes par la théorie des chaînes de Markov (Tierney, 1994). Dans la discussion de l'article de Tierney (1994), Doss (1994) démontre que pour une chaîne stationnaire et ergodique qui satisfait les conditions $E_{[\theta|X^{(k)}]}\left[(g(\theta)w(\theta))^2\right] < \infty$ et $E_{[\theta|X^{(k)}]}\left[(w(\theta))^2\right] < \infty$, l'estimateur IS normalisé est asymptotiquement gaussien.

L'algorithme itératif MCMC peut être donc remplacé par la méthode de l'échantillonnage pondéré combiné avec MCMC dont le calcul est généralement plus rapide. Même si le choix de cette méthode est justifié en théorie, on va confronter cette théorie à l'expérience.

Afin d'améliorer la qualité des estimations obtenues avec la méthode de l'importance sampling (IS), nous proposons trois approches différentes où soit le jeux de référence est fixe, soit la référence est choisie parmi un lot de jeux de données présélectionnées selon des critères qui seront précisés, ou bien pour ce lot de jeux de données une fonction d'importance est construite par mélange des densités a posteriori correspondantes.

Avant de les présenter explicitement, il faut ajouter quelques mots de commentaire. Si le calcul des vraies lois et vraies valeurs des espérances a posteriori n'est pas analytiquement accessible, les algorithmes MCMC sont les plus utilisés, mais comme indiqué auparavant, ils peuvent être très longs. Afin de pouvoir déterminer la qualité des estimations d'IS lorsque les vraies intégrales sont inconnues, on calculera les estimations également via les algorithmes

MCMC et, ainsi, les deux estimations (MCMC et IS) seront comparées via un critère de type "erreur quadratique". Cette stratégie consiste à refaire la démarche décrite ci-dessus présentée sur la figure (2.10).

2.2.3 Stratégie de référence fixée

La méthode d'échantillonnage pondéré IS a été présentée pour deux jeux de données $X^{(m)}$ et $X^{(k)}$ arbitraires. On a utilisé une réalisation $\theta_1^{(m)}, \ldots, \theta_N^{(m)}$ de la loi a posteriori $\pi(\theta|X^{(m)})$ obtenue uniquement pour un seul jeu de données $X^{(m)}$ pour estimer l'espérance $E_{[\theta|X^{(k)}]}[g(\theta)]$. Ensuite, la méthode est appliquée pour K jeux de données répliqués sous le même modèle. Elle est donc utilisée pour calculer les estimateurs $E_{[\theta|X^{(k)}]}[g(\theta)]$, où $k = 1, \ldots, m-1, m+1, \ldots, K$ pour m fixé et les différentes fonctions $g(\theta)$ en utilisant systématiquement $X^{(m)}$ comme jeu de données de référence. Cette approche pour $K = 101$ jeux de données et $m = 1$ dans un modèle est illustrée dans le graphique suivant (2.11).

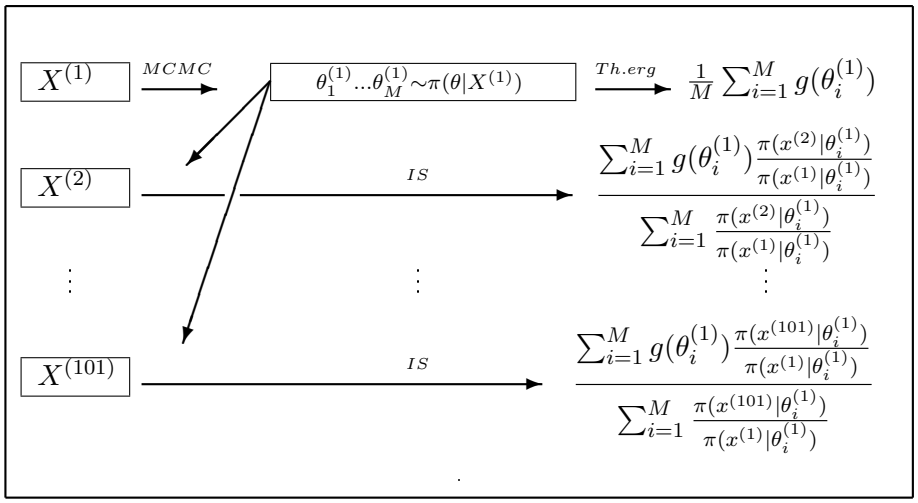

FIGURE 2.11: Démarche via IS avec référence fixe

Pour résumer : l'algorithme MCMC est utilisé sur le jeu de données $X^{(m)}$, une réalisation markovienne $\theta_1^{(m)}, \ldots, \theta_N^{(m)}$ de la loi a posteriori $\pi(\theta|X^{(m)})$ est ainsi obtenue. Elle est utilisée d'un côté pour calculer l'estimateur de $E_{[\theta|X^{(m)}]}$ (qui est égale $\frac{1}{N}\sum_{i=1}^{N} g(\theta_i^{(m)})$), et de l'autre côté pour calculer l'estimateur IS de $E_{[\theta|X^{(k)}]}(g(\theta))$, $k = 1, \ldots, m-1, m+1, \ldots, K$ (toujours par rapport à $X^{(m)}$) qui est exprimé par (2.38). Puis, les performances des estimateurs IS

sont étudiées par comparaison avec les estimateurs obtenus via MCMC et ceci sur chaque jeu de données $X^{(k)}, k = 1, \ldots, m-1, m+1, \ldots, K$ en faisant varier m de 1 à K.

Compte tenu que les vraies valeurs a posteriori sont inconnues, l'approximation de $E_{[\theta|X^{(k)}]}[g(\theta)]$ par IS est jugée inefficace ou insuffisante, si l'estimation via IS est éloignée de l'estimation via MCMC (par exemple, l'estimation via IS en dehors de l'intervalle de crédibilité à 95% obtenu via MCMC), une amélioration est alors recherchée.

2.2.4 Stratégie de référence choisie

Probablement le choix arbitraire de $X^{(m)}$ comme jeu de données de référence n'est pas le meilleur choix pour estimer $E_{[\theta|X^{(k)}]}[g(\theta)]$ pour tous les $k = 1, \ldots, m-1, m+1, \ldots, K$. L'estimation IS peut fluctuer selon le choix de la référence m. L'objectif est donc de trouver un jeux de données de référence grâce auquel l'estimateur IS de $E_{[\theta|X^{(k)}]}[g(\theta)]$ coïncidera le mieux avec celui obtenu avec MCMC pour $X^{(k)}$.

Pour améliorer les estimations via IS, l'idée de base de la deuxième approche est d'augmenter raisonnablement le nombre de jeux de données sur lesquels l'algorithme MCMC est appliqué puis de choisir parmi ces différents jeux de données, le plus adapté pour être le jeu de données de référence utilisé dans le calcul de l'importance sampling d'un nouveau jeu de données. L'idée est donc de présélectionner $M < K$ jeux de données parmi K pour lesquels MCMC sera effectué, ce nombre étant arbitraire et donc contestable mais devant être assez faible pour ne pas trop alourdir la méthode. Ainsi, sur un lot des M premiers jeux de données $X^{(1)}, \ldots, X^{(M)}$ présélectionnés, les réalisations markoviennes des différentes lois a posteriori sont obtenues via MCMC . L'importance sampling de chaque autre jeu de données $X^{(k)}$, $k = M+1, \ldots, K$ est utilisé avec une référence $X^{(m_k)}$ sélectionnée parmi $X^{(1)}, \ldots, X^{(M)}$ pour être le meilleur "candidat" de l'IS de $X^{(k)}$. Dans notre étude, trois critères de sélection (explicités plus bas) seront étudiés et comparés. La figure (2.12) illustre cette approche pour $K = 101$ jeux de données avec $M = 10$.

Comme lors de la première approche, les performances des estimations via IS seront étudiées par comparaison avec les estimations via MCMC considérées comme les estimations de référence mais, cette fois-ci sur chaque $X^{(k)}, k = M+1, \ldots, K$ et parallèlement, une compa-

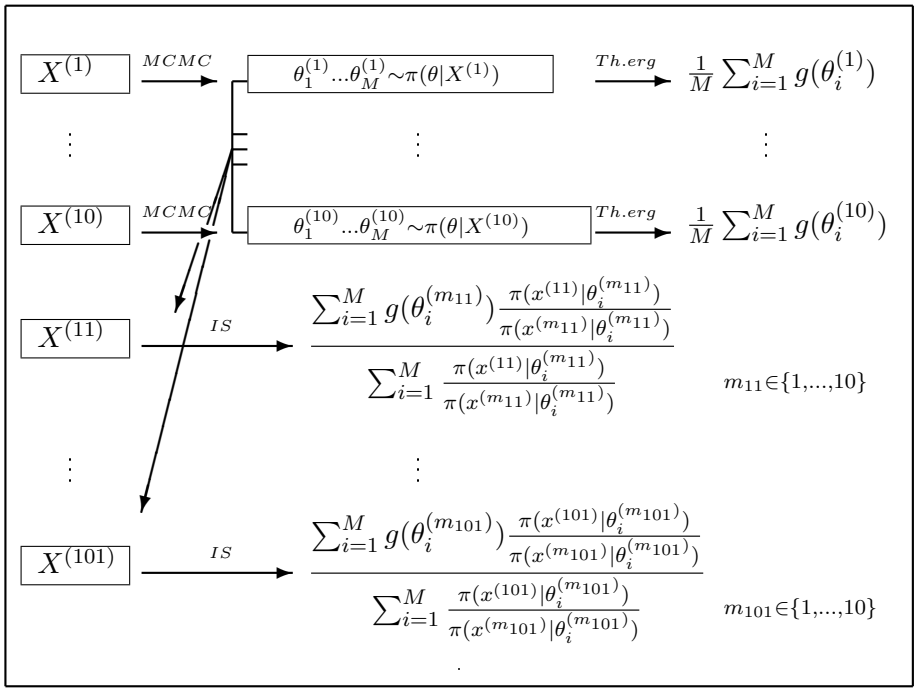

FIGURE 2.12: Démarche via IS avec référence choisie

raison sera également menée entre les deux approches (référence fixe versus référence choisie).

De manière générale, on peut écrire que dans la seconde approche pour $k = M+1, \ldots, K$, l'espérance $E_{[\theta|X^{(k)}]}[g(\theta)]$ est approchée par

$$\sum_{i=1}^{N} g(\theta_i^{(m_k)}) \tilde{w}_i(k, m_k) \longrightarrow E_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.39}$$

où m_k est choisi parmi $\{1, \ldots, M\}$.

2.2.4.1 Critères de sélection des jeux de données de références

Pour $M > 1$, on propose trois critères de choix du jeu de données de référence $X^{(m_k)}$ $m_k \in \{1, \ldots, M\}$. Les différents choix de la fonction d'importance pour l'échantillonnage pondéré sont proposés dans la littérature. Par example dans Cappé et al. (2008) ce choix est basé sur la minimisation de la déviance et dans Douc et al. (2007) sur la minimisation de la variance asymptotique de IS.

Critère minimisant la norme L_1

Soit l'estimateur IS donné comme dans l'équation (2.39). On s'attend à ce que cet estimateur fournisse une bonne estimation de $E_{[\theta|X^{(k)}]}[g(\theta)]$ si la densité a posteriori $\pi(\theta|X^{(m_k)})$ est "proche" de la densité a posteriori $\pi(\theta|X^{(m_k)})$. Il semble donc assez naturel de choisir comme référence, le jeu de données $X^{(m_k)}$ pour lequel la norme L_1 de la différence des densités a posteriori ci-dessus $||\pi(\theta^{(m_k)}|X^{(m_k)}) - \pi(\theta^{(m_k)}|X^{(k)})||_1$ soit la plus petite. Le premier critère proposé ici correspond alors à la plus petite estimation de cette norme et m_k est choisi satisfaisant l'expression suivante

$$\min_{m_k \in \{1,\ldots,M\}} \left\{ \sum_{i=1}^{N} \left| \tilde{w}_i(k, m_k) - \frac{1}{N} \right| \right\}. \tag{2.40}$$

Critère minimisant la divergence de Kullback-Leibler

La plus grande difficulté dans l'application de l'IS est la condition sur le support de la fonction d'importance (Cf. section (2.1.1.1)). Pour traiter cette difficulté, on peut étudier le comportement des queues de distribution de la fonction d'importance et de la densité d'intérêt puisque on souhaite que la première soit plus aplatie. Afin de contrôler le comportement aux queues de la distribution, de nombreux auteurs suggèrent de fonder le choix de la fonction d'importance sur la minimisation de la divergence de Kullback-Leibler (Cf. par exemple, Asmussen et al. (2005) ; Chen and Shao (1997) ; Rubinstein and Kroese (2004a) et Gustafson and Wasserman (1995)). En suivant le même concept, le deuxième critère de choix est fondé sur cette divergence c'est à dire que l'on choisit m_k correspondant à la densité $\pi(\theta|X^{(m_k)})$ pour laquelle la divergence de Kullback-Leibler $KL(\pi(\theta|X^{(K)}), \pi(\theta|X^{(m_k)})) = \int log(\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})})\pi(\theta|X^{(k)})d\theta$ est la plus petite. Le choix de m_k réalise le minimum suivant

$$\min_{m_k \in \{1,\ldots,M\}} \left\{ \sum_{i=1}^{N} \tilde{w}_i(k, m_k) \ln w_i(k, m_k) - \frac{\ln \frac{1}{N} \sum_{i=1}^{N} w_i(k, m_k)}{\frac{1}{N} \sum_{i=1}^{N} w_i(k, m_k)} \right\}. \tag{2.41}$$

Critère minimisant la variance de l'estimateur IS

Le troisième critère est basé sur la variance de $g(\theta)\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}$. On souhaiterait choisir $m_k \in \{1, \ldots, M\}$ qui satisfait l'inégalité suivante

$$Var_{[\theta|X^{(m_k)}]}\left[\frac{g(\theta)\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}\right] < Var_{[\theta|X^{(k)}]}[g(\theta)] \tag{2.42}$$

qui correspond à la variance de l'estimateur IS à gauche et à la variance de l'estimateur obtenu avec MCMC à droite.

Puisque $E_{[\theta|X^{(m_k)}]}[\frac{g(\theta)\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}] = E_{[\theta|X^{(k)}]}[g(\theta)]$, il est donc équivalent d'écrire

$$E_{[\theta|X^{(m_k)}]}\left[\left(\frac{g(\theta)\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}\right)^2\right] < E_{[\theta|X^{(k)}]}[(g(\theta))^2] \tag{2.43}$$

Le choix de $X^{(m_k)}$ est limité aux M jeux de données présélectionnés, alors il n'est pas certain que ces inégalités soient satisfaites pour toutes densités $\pi(\theta|X^{(m_k)})$. Cependant si $E_{[\theta|X^{(m_k)}]}[(\frac{g(\theta)\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})})^2]$ est finie, on peut trouver m_k qui rend la variance de l'estimateur IS la plus petite. Pour cela, quelque soit $k = M+1, \ldots, K$, le troisième critère choisit $m_k \in \{1, \ldots, M\}$ qui correspond à la densité a posteriori $\pi(\theta|X^{(m_k)})$ minimisant la variance de $g(\theta)\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}$.

Comme explique Robert (2007, section 6.2.2), pour minimiser cette variance, $\pi(\theta|X^{(m_k)})$ doit être proportionnelle à $|g(\theta)| \cdot \pi(\theta|X^{(k)})$. Pour cette raison m_k est choisi de manière à ce que la quantité $|g(\theta)|\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}$ soit la plus stable possible. Suivant cette idée, le troisième critère choisit m_k pour lequel l'expression suivante est satisfaite

$$\min_{m_k \in \{1,\ldots,M\}} \left\{ \frac{\max_{i=1,\ldots,N} u(\theta_i^{(m_k)}) - \min_{i=1,\ldots,N} u(\theta_i^{(m_k)})}{\sum_{i=1}^{N} u(\theta_i^{(m_k)})} \right\} \tag{2.44}$$

$$\text{where} \quad u(\theta_i^{(m_k)}) = |g(\theta_i^{(m_k)})| \cdot \pi(X^{(k)}|\theta_i^{(m_k)})/\pi(X^{(m_k)}|\theta_i^{(m_k)})$$

le troisième critère dépend de la fonction g et doit donc être évalué pour chaque fonction g ce qui n'est pas le cas des deux critères précédents.

2.2.5 Stratégie du mélange

La troisième stratégie présentée ici consiste à choisir comme fonction d'importance une densité de mélange. Il s'agit donc d'un mélange des M densités à posteriori présélectionnées pour lesquelles les réalisations a posteriori sont obtenues via l'algorithme MCMC. En effet pour obtenir une bonne estimation via l'échantillonnage préférentiel il est important que le

support de la fonction d'importance couvre le support de la fonction d'intérêt. Lorsque la densité a posteriori est inconnue, il n'est pas évident de savoir si la fonction d'importance satisfait cette condition même si dans le cadre d'estimations répétées étudiées ici, on se restreint aux densités de la même famille. Si le nombre de composantes du mélange est suffisamment grand, il paraît intuitivement plus probable que le mélange couvre mieux le support qu'une seule densité a posteriori. Cependant, dans le cas de lois a posteriori non explicites, la stratégie de mélange nécessite l'estimation des constantes de normalisation. Dans la littérature, il existe différentes méthodes d'estimation de constantes de normalisation ou de leurs rapports. Néanmoins, elles ne sont pas toutes adaptées aux estimations répétées car leur rapidité devient une question importante. Dans un premier temps, l'estimateur d'échantillonnage préférentiel avec un mélange comme fonction d'importance est présenté, puis plusieurs méthodes d'estimations des constantes de normalisations sont discutées. Trois méthodes parmi elles sont choisies et appliquées à cinq modèles dans les familles des lois a priori conjuguées : Normal - Normal, Poisson - Gamma, Binomial - Beta, Gamma -Gamma et Négative Binomial - Beta pour lesquels les constantes de normalisations et les espérances a posteriori des paramètres sont analytiquement disponibles. Il s'agit de modèles simples avec un espace de paramètres unidimensionnel. Les constantes puis les paramètres sont estimées et comparées ainsi aux vraies valeurs.

2.2.5.1 Importance Sampling avec un mélange

Soit $(\theta_1^{(u)}, \ldots, \theta_{n_u}^{(u)})$ l'ensemble des n_u réalisations sous la loi $\pi(\theta|X^{(u)})$ obtenu par exemple via un algorithme MCMC pour $u = 1, \ldots, M$. Pour $k > M$, nous avons alors

$$\frac{1}{|n|}\sum_{u=1}^{M}\sum_{i=1}^{n_u} g(\theta_i^{(u)})\frac{\pi(\theta_i^{(u)}|X^{(k)})}{\pi_{mix}(\theta_i^{(u)})} \to E_{\pi(\theta|X^{(k)})(g(\theta))} \tag{2.45}$$

où $\pi_{mix}(\theta)$ est une densité de mélange des M densités a posteriori de θ

$$\pi_{mix}(\theta) = \sum_{m=1}^{M}\frac{n_m}{|n|}\pi(\theta|X^{(m)}) \tag{2.46}$$

La version normalisée de l'estimateur d'Importance Sampling (2.45) s'écrit comme

$$\sum_{u=1}^{M}\sum_{i=1}^{n_u} w(\theta_i^{(u)})g(\theta_i^{(u)}) \tag{2.47}$$

où w est le poids

$$w(\theta_i^{(u)}) = \frac{\pi(\theta_i^{(u)}|X^{(k)})/\pi_{mix}(\theta_i^{(u)})}{\sum_{u=1}^{M}\sum_{i=1}^{n_u}\pi(\theta_i^{(u)}|X^{(k)})/\pi_{mix}(\theta_i^{(u)})} \tag{2.48}$$

Avec le théorème de Bayes chaque densité a posteriori de mélange (2.46) : $\pi(\theta|X^{(m)})$ s'écrit comme

$$\pi(\theta|X^{(m)}) = \frac{\pi(X^{(m)}|\theta)\pi(\theta)}{c_m}$$

où $c_m = \pi(X^{(m)}) = \int \pi(X^{(m)}|\theta)\pi(\theta)d\theta$ est la constante de normalisation, $\pi(X^{(m)}|\theta) = V(X^{(m)}|\theta)$ est la vraisemblance et $\pi(\theta)$ est la densité a priori.

Ainsi, après simplification des lois a priori $\pi(\theta)$ et des constants c_m dans le numérateur et le dénominateur, le poids (2.48) peut s'écrire maintenant de la manière suivante

$$w(\theta_i^{(u)}) = \frac{\pi(X^{(k)}|\theta_i^{(u)})/\sum_{m=1}^{M}\frac{n_m}{|n|}\frac{\pi(X^{(m)}|\theta_i^{(u)})}{c_m}}{\sum_{v=1}^{M}\sum_{j=1}^{n_v}\pi(X^{(k)}|\theta_j^{(v)})/\sum_{m=1}^{M}\frac{n_m}{|n|}\frac{\pi(X^{(m)}|\theta_j^{(v)})}{c_m}} \tag{2.49}$$

La connaissance des constantes de normalisations est nécessaire par exemple dans l'estimation du facteur de Bayes ou dans l'estimation de rapport de vraisemblances en présence d'observations manquantes. En fait, dans ce type d'applications, il faut souvent estimer des rapports de constantes de normalisation et non les constantes elles-mêmes (cf. Meng and Wong (1996) pour plus d'exemples et certaines méthodes qu'ils proposent).

On peut remarquer qu'en mettant en facteur une des constantes de normalisation (par exemple c_p) au numérateur et au dénominateur dans l'équation (2.49), il est possible de faire apparaître des ratios de constantes de normalisation dans l'expression du poids.

$$w(\theta_i^{(u)}) = \frac{\pi(X^{(k)}|\theta_i^{(u)})/\sum_{m=1}^{M}\frac{n_m}{|n|}\frac{\pi(X^{(m)}|\theta_i^{(u)})}{c_m/c_p}}{\sum_{v=1}^{M}\sum_{j=1}^{n_v}\pi(X^{(k)}|\theta_j^{(v)})/\sum_{m=1}^{M}\frac{n_m}{|n|}\frac{\pi(X^{(m)}|\theta_j^{(v)})}{c_m/c_p}} \tag{2.50}$$

Les ratios c_m/c_p, $m = 1, \ldots, M$ où $c_p \in \{c_1, \ldots, c_M\}$ peuvent être ensuite estimés par les différentes méthodes proposées dans la littérature. Chen and Shao (1997) font une revue des méthodes pour estimer ces ratios. Parmi les méthodes qu'ils mentionnent, il y a importance sampling classique, *bridge sampling* (Meng and Wong, 1996), *path sampling* (Gelman and Meng, 1994), *marginal likelihood* méthode (Chen, 1994) qui est, dans le contexte présenté ici, équivalente à *reciprocal importance sampling*, *reverse logistic regression* (Geyer, 1993) et finalement *ratio importance sampling* dont ils sont les auteurs (Chen and Shao, 1997). Il existe

aussi des méthodes plus récentes pour estimer les constantes de normalisation (de Valpine, 2008; Doss, 2010).

2.2.5.2 Estimation des rapports de constantes de normalisation

Dans l'article (Gajda et al., 2011) nous avons estimé les constantes de normalisation par une méthode itérative *reverse logistic regression* proposée par Geyer (1993) et par une approche connue comme *reciprocal importance sampling* mentionnée entre autre par Gelfand and Dey (1994), DiCiccio et al. (1997) ou Gelman and Meng (1998). La majeure difficulté de la première technique est qu'en étant itérative, elle peut être très longue avant que sa convergence soit atteinte. Sur un exemple de modèle de régression poissonnienne avec dix covariables, présenté dans l'article (Gajda et al., 2011), cette méthode s'est avérée trop longue et la deuxième méthode a donc été privilégiée.

Pour ne pas alourdir les calculs, il est intéressant de choisir une méthode simple pour estimer les rapports de constantes de normalisation. Seulement les méthodes : importance sampling classique, *reciprocal importance sampling* et *ratio importance sampling* sont présentées ici. De plus la méthode *reverse logistic regression* est décrite car utilisée pendant les travaux de thèse (Gajda et al., 2011).

Importance Sampling

c_m, $m = 1, \ldots, M$ sont les constantes de normalisation de M lois a posteriori $\pi(\theta|X^{(m)})$ dont n_m réalisations $(\theta_1^{(m)}, \ldots, \theta_{n_m}^{(m)})$ sont obtenues par les algorithmes MCMC. Ainsi en faisant des lissages de densités à partir de ces réalisations, nous pouvons avoir une certaine idée du support de chacune (il s'agit d'un cas unidimentionnel). En effet, la condition d'utilisation de l'estimateur par importance sampling est que le support de la fonction d'importance couvre le support de la densité selon laquelle on souhaite calculer une intégrale. Donc en comparant graphiquement les lissages des densités, il est possible de choisir comme fonction d'importance une de ces densités a posteriori avec un support plus large - si cela possible. Soit Ω_m le support

de la densité $\pi(\theta|X^{(m)})$ alors on choisit la fonction d'importance $\pi(\theta|X^{(p)})$ telle que $\Omega_m \subseteq \Omega_p$. On peut démontrer (Cf. section (IV) en annexe) qu'en prenant la réalisation $\theta_i^{(p)} \sim \pi(\theta|X^{(p)})$, le rapport des constantes de normalisation c_m/c_p peut être estimé par

$$\frac{c_m}{c_p} \leftarrow \frac{1}{n_p} \sum_{i=1}^{n_p} \frac{\pi(X^{(m)}|\theta_i^{(p)})}{\pi(X^{(p)}|\theta_i^{(p)})} \tag{2.51}$$

Lorsque la dimension de l'espace des paramètres augmente et lorsque le modèle est plus compliqué, le choix "graphique" de la fonction d'importance n'est pas évident.

Reciprocal Importance Sampling

Parmi les différentes méthodes d'estimation de ratios de constantes de normalisation, Chen and Shao (1997) citent la méthode "marginal likelihood" de Chen (1994). L'idée de Chen (1994) est d'utiliser une réalisation d'une densité jointe pour estimer la densité marginale. Sans donner plus de détails sur cette méthode dans le contexte original, il est plus intéressant ici d'utiliser cet estimateur pour estimer les constantes de normalisations (donc les densités marginales $m(x)$ des données X) comme

$$\hat{m}(x) = \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\omega(\theta_i)}{L(x, \theta_i)\pi(\theta_i)} \right]^{-1}$$

où θ_i est une réalisation de la loi a posteriori $\pi(\theta|x)$, la fonction $\omega(\theta)$ est entièrement connue et son support Ω_ω est couvert par le support de la densité a posteriori : $\Omega_\omega \subset \Omega_{\pi(\cdot|x)}$, $L(x, \theta_i)$ et $\pi(\theta_i)$ sont la vraisemblance et la densité a priori respectivement.

Dans le calcul des facteurs de Bayes , DiCiccio et al. (1997) présente l'estimateur "reciprocal importance sampling" proposé par Gelfand and Dey (1994).

Puisque l'estimation du rapport de constantes de normalisation utilisée ici est plus proche du concept de facteur de Bayes, je garde la notion de "reciprocal importance sampling" pour décrire ce type de méthodes. Ainsi l'inverse de chaque constante c_m peut être estimée par

$$\frac{1}{c_m} \leftarrow \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{f(\theta_i^{(m)})}{\pi(\theta_i^{(m)})\pi(X^{(m)}|\theta_i^{(m)})} \tag{2.52}$$

et leur rapport par

$$\frac{c_m}{c_p} \leftarrow \frac{\frac{1}{n_p}\sum_{i=1}^{n_p} \frac{f_p(\theta_i^{(p)})}{\pi(\theta_i^{(p)})\pi(X^{(p)}|\theta_i^{(p)})}}{\frac{1}{n_m}\sum_{i=1}^{n_m} \frac{f_m(\theta_i^{(m)})}{\pi(\theta_i^{(m)})\pi(X^{(m)}|\theta_i^{(m)})}} \tag{2.53}$$

où $\theta_i^{(m)} \sim \pi(\theta|X^{(m)})$ $(m, p \in \{1, \ldots, M\})$ sont des réalisations obtenues via MCMC auparavant. Les fonctions f_m, $m = 1, \ldots, M$ sont les densités arbitraires satisfaisant les conditions de support plus "étroit". Ici, les f_m sont choisies comme les densités de lois normales de moyenne $\hat{\mu}_m = \frac{\sum_{i=1}^{n_m} \theta_i^{(m)}}{n_m}$ et de variance $\hat{\sigma}_m^2 = \frac{\sum_{i=1}^{n_m} \theta_i^{(m)2}}{n_m} - (\hat{\mu}_m)^2$.

Ratio Importance Sampling

Chen and Shao (1997) ont proposé la méthode de "ratio importance sampling" qui consiste à simuler une réalisation θ_i selon une densité arbitraire $h(\theta)$ dont le support est l'union des supports des densités a posteriori pour les quelles le ratio de leurs constantes est recherché. Cet estimateur a la forme suivante :

$$\frac{c_m}{c_p} \leftarrow \frac{\sum_{i=1}^{n} \frac{\pi(X^{(m)}|\theta_i)\pi(\theta_i)}{h(\theta_i)}}{\sum_{i=1}^{n} \frac{\pi(X^{(p)}|\theta_i)\pi(\theta_i)}{h(\theta_i)}} \tag{2.54}$$

où $\theta_i \sim h(\theta)$, $\Omega_h \supseteq \Omega_{\pi(\cdot|X^{(m)})} \cup \Omega_{\pi(\cdot|X^{(p)})}$.

Le choix optimal de la fonction h est une densité qui minimise l'erreur quadratique moyenne relative ;

$$h_{opt}(\theta) \propto |\pi(\theta|X^{(m)})\pi(\theta)/c_m - \pi(\theta|X^{(p)})\pi(\theta)/c_p|$$

Cependant cette densité est inutilisable en pratique car elle dépend des constantes inconnues c_m et c_p. Dans ce cas, les auteurs proposent un schéma en deux étapes pour obtenir un estimateur optimal du ratio $r = c_m/c_p$. La première étape consiste à approcher r par l'estimateur (2.54) en utilisant les réalisations d'une fonction h arbitraire $h(\theta) > 0$. Notons $\hat{\tau}$ l'estimateur de r ainsi obtenu. L'étape suivante consiste à générer une réalisation $\vartheta_{n,i}$ de la densité ψ_n :

$$\psi_n(\theta) \propto |\pi(\theta|X^{(m)})\pi(\theta) - \hat{\tau}\pi(\theta|X^{(p)})\pi(\theta)|$$

à l'aide par exemple de l'algorithme de Metropolise Hasting ou de l'échantillonneur de Gibbs. Finalement le rapport r est estimé via (2.54) avec cette nouvelle réalisation $\vartheta_{n,i} \sim \psi_n$.

Comme solution approximative optimale quant au choix de h, Chen and Shao (1997) proposent d'utiliser des fonctions d'importance "judicieuses" h_m^I et h_p^I de manière à simuler facilement les θ_i sous la densité

$$h_{opt}^I(\theta) \propto |h_m^I(\theta) - h_p^I(\theta)|$$

Dans le cas où les support $\Omega_{\pi(\cdot|X^{(m)})}$ et $\Omega_{\pi(\cdot|X^{(p)})}$ sont disjoints, $h_{opt}^I(\theta)$ est un mélange de deux fonctions $h_{opt}^I(\theta) = (h_m^I(\theta) + h_p^I(\theta))/2$.

Remarquons que dans notre cas, le calcul de l'estimateur IS (2.47) avec les poids (2.50) nécessite la détermination de $M-1$ ratios de constante. Pour des raisons de simplicité, une seule fonction h peut être utilisée pour l'ensemble de ces $M-1$ estimations. Le choix (pas nécessairement optimal) peut être, par extension de la solution approximative optimale de Chen and Shao (1997), un mélange de M "judicieuses" fonctions d'importance h_m. Par exemple, on peut choisir un mélange de densités de lois normales f_m de moyennes et variances $\hat{\mu}_m$ et $\hat{\sigma}_m^2$ comme obtenues par MCMC lors de la méthode de "reciprocal importance sampling" (2.2.5.2).

Lors de l'étude de Chen and Shao (1997), les auteurs ont comparé ces différentes approches via des erreurs quadratiques moyennes relatives. Sur cette base, l'efficacité de l'estimateur "ratio importance sampling" a été jugée la meilleure quand le choix de la fonction h était optimal.

Reverse Logistic Regression

Dans l'article (Gajda et al., 2011), une autre méthode itérative "reverse logistic regression" de Geyer (1993) a été utilisée pour estimer les constantes de normalisation, elle est aussi brièvement présentée ici.

Pour simplifier les notations, posons $h_m(\theta) = \pi(X^{(m)}|\theta)\pi(\theta)$. L'equation (2.46) s'écrit donc maintenant comme

$$\pi_{mix}(\theta) = \sum_{m=1}^{M} \frac{n_m}{|n|} \frac{h_m(\theta)}{c_m} \tag{2.55}$$

L'idée de *reverse logistic regression* commence par la réécriture suivante de π_{mix}

$$\pi_{mix}(\theta) = \sum_{m=1}^{M} h_m(\theta) e^{\eta_m} \tag{2.56}$$

où

$$\eta_m = -\log c_m + \log \frac{n_m}{|n|}$$

Si θ résulte de ce mélange, alors la probabilité que θ provienne de la $u^{\text{ème}}$ composante du mélange est $p_u(\theta, \eta)$ où

$$p_u(\theta, \eta) = \frac{h_u(\theta) e^{\eta_u}}{\sum_{m=1}^{M} h_m(\theta) e^{\eta_m}} \tag{2.57}$$

Geyer(1993) propose d'estimer les η_m, qui déterminent c_m, par maximisation de la log vraisemblance :

$$l(\eta) = \sum_{u=1}^{M} \sum_{i=1}^{n_u} \log p_u(\theta_i^{(u)}, \eta) \tag{2.58}$$

où $\eta = (\eta_1, \ldots \eta_M)$. La dérivée de la log vraisemblance :

$$\frac{\partial l(\eta)}{\partial \eta_r} = n_r - \sum_{u=1}^{M} \sum_{i=1}^{n_u} p_r(\theta_i^{(u)}, \eta) \tag{2.59}$$

et le hessien est donné par :

$$\frac{\partial^2 l(\eta)}{\partial \eta_r^2} = -\sum_{u=1}^{M} \sum_{i=1}^{n_u} p_r(\theta_i^{(u)}, \eta) \left[1 - p_r(\theta_i^{(u)}, \eta)\right] \tag{2.60}$$

$$\frac{\partial^2 l(\eta)}{\partial \eta_r \partial \eta_s} = \sum_{u=1}^{M} \sum_{i=1}^{n_u} p_r(\theta_i^{(u)}, \eta) p_s(\theta_i^{(u)}, \eta) \quad s \neq r \tag{2.61}$$

Puisque la log vraisemblance est convexe, elle peut être maximisée par beaucoup de méthodes itératives comme par exemple celle de Newton-Raphson :

$$\eta^{(k+1)} = \eta^{(k)} - (H^{(k)})^{-1} \upsilon^{(k)}$$

où H est la matrice hessienne et υ le gradient. Comme alternative, nous pouvons utiliser une maximisation itérative en considérant chaque paramètre un par un et en comparant $\xi_r = e^{-\eta_r}$ à

$$\frac{1}{n_r} \sum_{u=1}^{M} \sum_{i=1}^{n_u} \frac{h_r(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m} \tag{2.62}$$

En effet quand $\frac{\partial l(\eta)}{\partial \eta_r} = 0$ et $\xi_r = e^{-\eta_r}$ on obtient

$$\frac{1}{n_r} \sum_{u=1}^{M} \sum_{i=1}^{n_u} \frac{h_r(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m} = \xi_r \tag{2.63}$$

Pour des raisons d'identifiabilité, il est nécessaire d'imposer une contrainte comme $\xi_1 = 1$. En effet, si $d = \xi_1$ et si $\xi_m = \xi_m^* d$ pour $m = 1, \ldots, M$, alors l'équation (2.63) devient

$$\frac{1}{n_r} \sum_{u=1}^{M} \sum_{i=1}^{n_u} \frac{h_r(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m^* d} = \xi_r^* d \tag{2.64}$$

On obtient en simplifiant par d

$$\frac{1}{n_r} \sum_{u=1}^{M} \sum_{i=1}^{n_u} \frac{h_r(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m^*} = \xi_r^* \tag{2.65}$$

Il suffit d'estimer ξ_m^* pour $m = 2, \ldots, M$ en fixant $\xi_1^* = 1$ car on montre facilement que l'estimateur IS défini par (2.47) et (2.48) avec $\pi_{mix}(\theta_i^{(u)}) = \sum_{m=1}^{M} h_m(\theta_i^{(u)}) e^{\eta_m}$ est :

$$\frac{\sum_{u=1}^{M} \sum_{i=1}^{n_u} g(\theta_i^{(u)}) \frac{h_k(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m^*}}{\sum_{u=1}^{M} \sum_{i=1}^{n_u} \frac{h_k(\theta_i^{(u)})}{\sum_{m=1}^{M} h_m(\theta_i^{(u)})/\xi_m^*}} \tag{2.66}$$

où seuls les termes en ξ_m^* interviennent.

Cet algorithme est globalement convergent (Cf. Jensen et al., 1991, Th. 2 et Th. 4), mais il peut converger très lentement amenant à des soucis de temps de calculs surtout dans le cadre des estimation répétés.

2.2.6 Critères de comparaison entre les différentes approximations

Pour quantifier l'efficacité des approximations obtenues via importance sampling avec les différentes stratégies par rapport à celles obtenues via MCMC seul, nous comparons ces estimations à travers les erreurs quadratiques. Les estimateurs obtenus par les deux méthodes (IS ou MCMC) sont notés de la manière suivante :

$$MCMC_{(k)} := \frac{1}{N} \sum_{i=1}^{N} g(\theta_i^{(k)}), \quad IS_{(k)/(m)} := \frac{1}{N} \sum_{i=1}^{M} g(\theta_i^{(m)}) \tilde{w}_i(k, m)$$

où l'indice k identifie le jeu de données pour lequel on estime l'espérance *a posteriori* de $g(\theta)$ et l'indice m indique le jeu de données de référence $X^{(m)}$ utilisé dans le calcul de l'approximation via IS. On souhaite que les estimateurs IS ($IS_{(k)/(m)}$) soient "proches" de ceux obtenus via MCMC ($MCMC_{(k)}$) pour tous les jeux de données $X^{(k)}$. La comparaison est donc basée sur une distance "moyenne" entre les deux méthodes, les erreurs quadratiques moyennes.

Classiquement, l'erreur quadratique moyenne mesure une distance carrée entre estimateur et la vraie valeur ici la vraie espérance a posteriori. Si l'espérance a posteriori n'est pas explicite, nous comparons alors l'estimateur IS avec l'estimateur classique qui est donc celui obtenu via MCMC. L'erreur quadratique moyenne définie par la suite est donc une sorte de distance entre les deux estimateurs.

Concernant l'approche IS où le jeu de données de référence est fixe et identique pour tous les $k = 1, \ldots, m-1, m+1, \ldots, K$ dans le calcul de l'IS, l'erreur quadratique moyenne est définie par :

$$EQ_{(fixe)} = \frac{1}{K-1} \sum_{k \neq m} (IS_{(k)/(m)} - MCMC_{(k)})^2 \tag{2.67}$$

Lorsqu'on se place dans le cas où la référence $X^{(m_k)}$ est choisie parmi $X^{(1)}, ..., X^{(M)}$ selon les différents critères et ceci pour chaque jeu de données $k = M, \ldots, K$, l'erreur quadratique moyenne est définie par :

$$EQ_{(choisie)} = \frac{1}{K-M} \sum_{k \neq m_k} (IS_{(k)/(m_k)} - MCMC_{(k)})^2 \tag{2.68}$$

Enfin, concernant la stratégie du mélange de M densités a posteriori, l'erreur quadratique moyenne s'écrit comme

$$EQ_{(mix)} = \frac{1}{K-M} \sum_{k \neq m} (IS_{(k)/(mix)} - MCMC_{(k)})^2 \tag{2.69}$$

Dans le deuxième cas, $EQ_{(choisie)}$ est calculée pour les trois critères de choix (norme L_1, divergence de Kullback-Leibler et minimisation de la variance) ce qui permet, de plus, d'évaluer l'effet du critère sur l'estimation avec IS par rapport au MCMC.

2.3 Modèles de Poisson

L'ensemble de ce travail est développé dans le cadre des modèles de type poissonnien. Le premier modèle est le modèle de Poisson avec un paramètre λ. Le deuxième est un modèle de régression de Poisson avec un terme constant et une covariable Z (et donc deux coefficients a et b respectivement à estimer). Finalement, le troisième modèle est également un modèle de régression de Poisson mais avec extra-variabilité poissonnienne introduite via un résidu gaussien

ϵ dans la relation loglinéaire. Le premier modèle est un exemple de base qui permet d'avoir une loi a posteriori explicite pourvu que la loi a priori soit choisie conjuguée. Le deuxième modèle est un GLM classique et le troisième modèle prend en compte une sur-dispersion qui dans certaines applications permet d'éviter des biais d'estimations (Breslow, 1984). Sous chaque modèle, les données sont simulées en fixant différentes valeurs des paramètres. La covariable Z est simulée selon une loi gaussienne centrée réduite et le résidu ϵ selon une loi gaussienne centrée de variance σ^2. Puis avec les données ainsi obtenues, les espérances a posteriori des fonctions des paramètres sont estimées.

2.3.1 Modèles de simulation

Le première étape consiste à considérer différents modèles et à simuler différentes réplications sous chaque modèle. La table (2.13) donne les différentes valeurs des paramètres utilisés dans les simulations (et donc les différents modèles envisagés).

FIGURE 2.13: Modèles de simulation des données.

	Modèle 1	Modèle 2	Modèle 3
Modèle	$X_i\|\lambda \sim \mathcal{P}(\lambda)$ $i = 1, \ldots, n$	$X_i\|\lambda \sim \mathcal{P}(\lambda_i)$ $log(\lambda_i) = a + bZ_i$ $Z_i \sim \mathcal{N}(0,1)$ $i = 1, \ldots, n$	$X_i\|\lambda \sim \mathcal{P}(\lambda_i)$ $log(\lambda_i) = a + bZ_i + \epsilon_i$ $Z_i \sim \mathcal{N}(0,1), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$ $i = 1, \ldots, n$
Valeurs des paramètres	$\lambda = 1$ ou 20	$a = 0,\ b = 0.5$	$a = 0,\ b = 0.5$ $\sigma^2 = 1/8$ ou $1/4$ ou $1/2$

Pour chaque modèle et chaque configuration de paramètres, on réplique $K = 101$ jeux de données $X^{(1)}, ..., X^{(101)}$ où un jeu de données $X^{(k)}$ contient n observations $X^{(k)} = (X_1, ..., X_n)$, $k = 1, \ldots, 101$. Dans notre étude, n a été fixé à 20. Les observations X, la covariable Z et les aléas ϵ sont simulés par un générateur pseudoaléatoire de la loi de Poisson et de la loi normale.

2.3.2 Modèles d'estimation

Dans un deuxième temps, les paramètres sont estimés. Les lois a priori considérées dans cette étude sont choisies vagues et sont présentées dans la table (2.14) ainsi que les modèles d'estimation.

FIGURE 2.14: Modèles d'estimation et lois a priori.

	Modèle 1	Modèle 2	Modèle 3
Modèle	$X_i\|\lambda \sim \mathcal{P}(\lambda)$	$X_i\|\lambda \sim \mathcal{P}(\lambda_i)$ $log(\lambda_i) = a + bZ_i$	$X_i\|\lambda \sim \mathcal{P}(\lambda_i)$ $log(\lambda_i) = a + bZ_i + \epsilon_i$ $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
Lois a priori	$\lambda \sim \mathcal{G}(\alpha, \beta)$ $\alpha = \beta = 0.01$	$a, b \sim \mathcal{N}(0, 10^5)$	$a, b \sim \mathcal{N}(0, 10^5)$ $\sigma^2 \sim \mathcal{IG}(0.01, 0.01)$

Pour chacun des 101 jeux de données $X^{(k)}$ simulés comme décrit ci-dessus, les espérances a posteriori $E_{[\theta|X^{(k)}]}[g(\theta)]$ pour chaque fonction g sont estimées par les différentes méthodes d'approximation. Notons que dans le premier modèle, $\theta = \lambda$, dans le deuxième modèle $\theta = (a, b)$ et $\theta = (a, b, \sigma^2)$ dans le troisième modèle.

Les estimateurs de $E_{[\theta|X^{(k)}]}[g(\theta)]$ sont approchés selon les deux types de méthodes pour chaque jeu de données : la méthode classique soit via MCMC pour tous les jeux de données et les trois stratégies combinant MCMC et Importance Sampling.

3 Résultats

La première stratégie ("référence fixe") est explorée tout d'abord brièvement pour un seul choix de jeu de données de référence (le premier jeu de données). Afin d'étudier plus généralement les performances de cette stratégie de "référence fixe", tous les choix possibles

pour le jeu de données de référence ont été considérés. La stratégie de "référence choisie" est ensuite étudiée et comparée à la première. Les résultats ont été obtenus également sur des modèles de régression poissonnienne où le nombre de covariables et la taille de l'échantillon varient. Ensuite, la troisième stratégie dite de "mélange" a été examinée combinant IS avec MCMC via une fonction d'importance construite comme mélange de densités a posteriori présélectionnées. Cette méthode nécessitait l'évaluation de constantes de normalisation, une étude comparant différentes techniques d'approximation de ces constantes (techniques présentées dans la section (2.2.5.2)) est menée. Enfin, les deux dernières stratégies nécessitant la présélection de jeux de données, nous présentons une méthode permettant de faire cette sélection de manière automatique. L'ensemble de ce travail a débouché sur la publication de deux articles (Gajda et al., 2010, 2011).

Dans l'ensemble des estimations pour tous les modèles étudiés, les algorithmes MCMC ont été utilisés avec 50000 itérations avec un temps de chauffe de 5000 itérations. La convergence a été étudiée en vérifiant que si les erreur de Monté Carlo soient inférieurs à 5% des écarts-types à posteriori pour chaque paramètre. Les diagnostics de Gelman et Rubin ainsi que des outils graphiques (trace des paramètres avec bonne mélangeance, autocorrélations des paramètres) ont été aussi utilisés.

3.1 Résultats avec la stratégie de référence fixe

Dans un première temps, la stratégie de référence fixe à été étudiée pour un seul jeux de données de référence, $X^{(1)}$ $(m = 1)$ pour trois types de modèles de Poisson décrits dans la section (2.3). Lors de cette analyse, les fonctions $g(\theta)$ suivantes ont été considérées : $g(\theta) = \theta$, $g(\theta) = \theta^2$ et $g(\theta) = \mathbb{1}_{\theta<\theta_s}$ avec θ_q fixé. Les deux premières fonctions permettent d'approcher l'espérance et l'écart-type a posteriori de θ. La dernière fonction permet d'approcher les valeurs de la fonction de répartition a posteriori de θ pour différents quantiles θ_q choisis $q \in \{0.975, 0.95, 0.75, 0.50, 0.25, 0.05, 0.025\}$). Les estimations de $E_{[\theta|X^{(k)}]}[g(\theta)]$ sont calculées suivant les deux types de méthodes pour chaque jeu de données, la méthode classique MCMC et la méthode combinée (via importance sampling combiné avec MCMC) avec systématique-

ment comme fonction d'importance le premier jeux de donnée. L'ensemble des résultats est présenté dans l'annexe (Cf. tableau (1)-(6)).

Les moyennes, les écarts-types et les proportions moyennes de valeurs inférieures au quantile spécifié sont toujours proches entre la la méthode MCMC et l'IS combiné avec MCMC pour $X^{(1)}$ sauf pour des quantiles d'ordres petits. En effet, les proportions des valeurs inférieures aux quantiles d'ordres les plus petits (25%, 5%, 2.5%) sont souvent un peu plus dispersées pour la méthode combinée. A titre d'exemple, les résultats du modèle 3 (modèle de Poisson avec extra variabilité) sont présentés dans le tableau (3.1) quand l'extra-variabilité est importante.

TABLE 3.1: Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=1/2)

a=0, b=1/2, σ^2=1/2	MCMC $g(\theta) = a$	IS $g(\theta) = a$	MCMC $g(\theta) = b$	IS $g(\theta) = b$
$\hat{E}_{[\theta\|X]}(g(\theta))$ [1]	−0.149	−0.046	0.548	0.515
CI [4]	[-0.803 ;0.350]	[-0.429 ;0.402]	[0.008 ;1.040]	[-0.013 ;0.967]
$\hat{sd}_{[\theta\|X]}(g(\theta))$ [2]	0.340	0.253	0.301	0.232
CI	[0.229 ;0.619]	[0.172 ;0.322]	[0.183 ;0.544]	[0.122 ;0.347]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.975}))$ [3]	0.975	0.994	0.975	0.977
CI		[0.973 ;1.000]		[0.928 ;1.000]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.95}))$ [3]	0.950	0.983	0.950	0.956
CI		[0.942 ;1.000]		[0.862 ;1.000]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.75}))$ [3]	0.750	0.837	0.750	0.731
CI		[0.671 ;0.999]		[0.356 ;0.975]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.50}))$ [3]	0.500	0.606	0.500	0.468
CI		[0.367 ;0.932]		[0.118 ;0.835]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.25}))$ [3]	0.250	0.318	0.250	0.209
CI		[0.089 ;0.719]		[0.018 ;0.380]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.05}))$ [3]	0.050	0.070	0.050	0.025
CI		[0.006 ;0.233]		[0.000 ;0.083]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.025}))$ [3]	0.025	0.041	0.025	0.009
CI		[0.001 ;0.207]		[0.000 ;0.035]

[1]Moyenne des 100 moyennes a posteriori estimées de θ
[2]Moyenne des 100 écarts types a posteriori estimés de θ
[3]Moyenne des 100 proportions a posteriori estimées de θ où θ_q le quantile d'ordre q de $\pi(\theta|X)$
[4]Intervalle contenant 95% des approximations

Les estimations de $E_{[\theta|X^{(k)}]}[g(\theta)]$ avec $g(\theta) = \theta$ pour tout $k = 2, \ldots, 101$ obtenues avec l'approche classique MCMC (en bleu intervalle de crédibilité à 95%) et avec l'approche IS combiné (référence fixe $X^{(1)}$) (en vert) sont représentés sur les graphique (1)-(13) en annexe. A titre d'exemple, la figure (3.1) montre les résultats pour un cas du modèle 3. Globalement, les estimations a posteriori obtenue via MCMC et via IS combiné avec $X^{(1)}$ comme référence

fixe sont très proches pour les modèles étudiés. Cependant, comme on peut le constater en étudiant la figure (3.1), les estimations de $E_{[\theta|X^{(68)}]}[g(\theta)]$ $(k = 68)$ avec $g(\theta) = a$ dans le cas du modèle 3 ($a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$) sont beaucoup moins cohérentes entre MCMC et IS combiné pour certains jeux de données comme par exemple le 68^{me}. Probablement pour ce jeu de données $X^{(68)}$, le jeu de données $X^{(1)}$ n'est pas la meilleure référence fixe pour estimer $E_{[\theta|X^{(68)}]}[g(\theta)]$.

Dans la deuxième stratégie dite "référence choisie", on s'attend à éviter ce genre de problème et donc à améliorer les résultats car le jeux de données de référence est choisi dans un lot de jeux de données présélectionnés.

3.2 Résultats avec la stratégie de référence choisie

Dans cette section, la première stratégie avec toutes les références fixes possibles est comparée aux résultats obtenus avec la stratégie de la référence choisie. Cette stratégie de référence choisie est étudiée pour les trois critères de choix présentés dans la section (2.2.4.1). La meilleure et la pire référence fixe en termes d'erreur quadratique moyenne sont identifiées.

Les résultats de la stratégie "référence choisie" sont présentés et discutés plus en détail pour chaque modèle ci-dessous.

Pour chaque $m = 1, \ldots, K$ on prend successivement $X^{(1)}$ jusqu'au $X^{(K)}$ comme le jeu de données de référence et on calcule les $K - 1$ estimations de $E_{[\theta|X^{(k)}]}$ via IS combiné (avec le résultats de MCMC avec $X^{(m)}$) pour ensemble de $k = 1, \ldots, m-1, m+1, \ldots, K$. Ces estimations sont comparées via une erreur quadratique moyenne (Cf. l'équation (2.67)) avec les estimations obtenues avec MCMC avec $X^{(k)}$. En variant m de 1 à K on obtient K erreurs quadratiques moyennes. Ces erreurs sont ensuite représentées sous la forme d'un box-plot. Indépendamment, la stratégie de référence choisie est utilisée et pour chaque critère l'erreur quadratique moyenne (Cf. l'équation (2.68)) est calculée et reportée sur le même box-plot. Les erreurs quadratiques sous le premier critère, le deuxième et le troisième sont marquées comme les losanges noirs, blancs et gris respectivement.

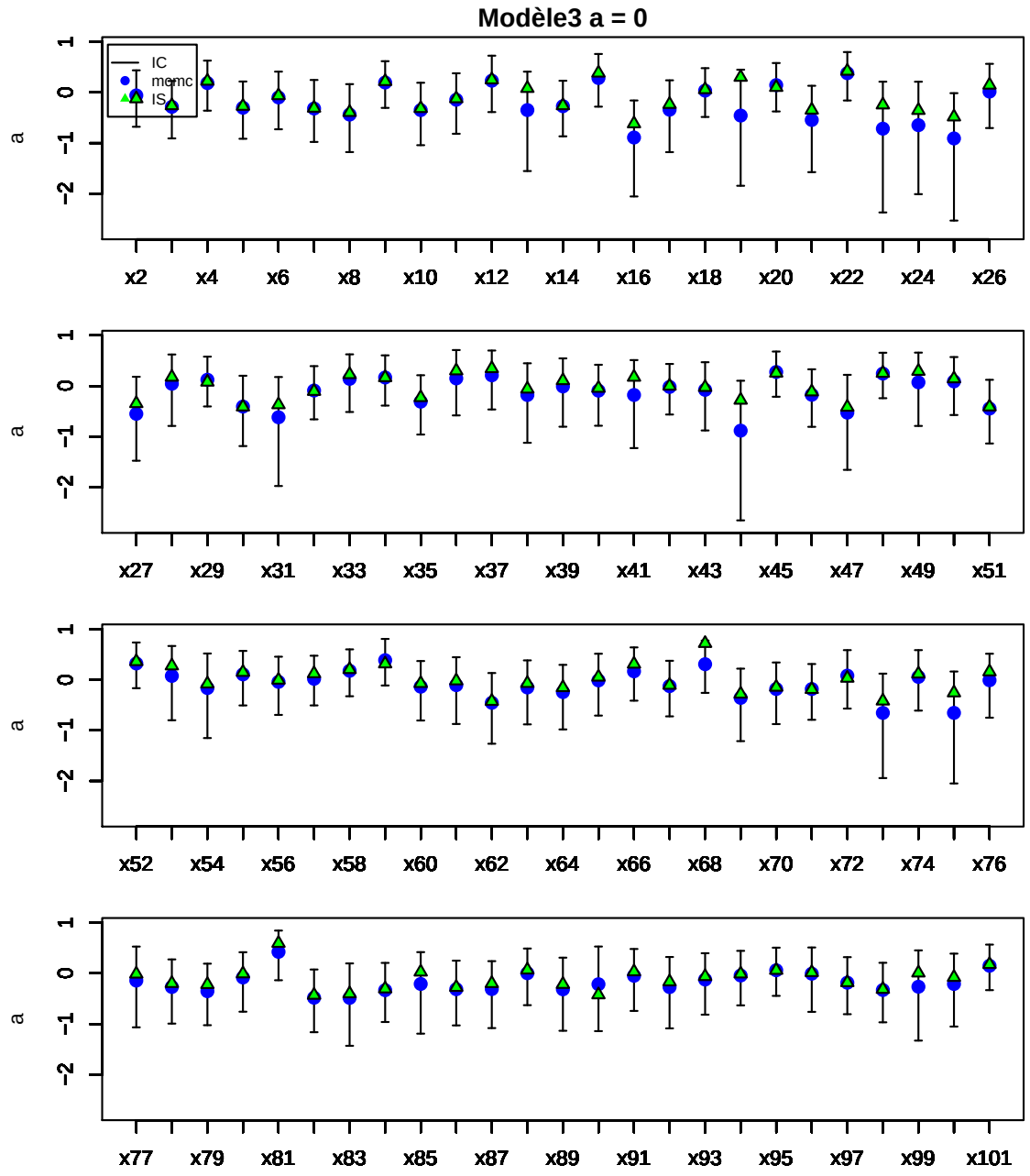

FIGURE 3.1: Étude des performances de l'estimation de $E_{[\theta|X^{(k)}]}[g(\theta)]$ pour $g(\theta) = a$ avec la stratégie "référence fixe" (où le premier jeu de données $X^{(1)}$ est la référence fixe) dans le modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$

3.2.1 Modèle 1 : Modèle de Poisson

Les erreurs quadratiques moyennes de la stratégie de référence fixe avec les erreurs quadratiques moyennes des trois critères de la référence choisie pour le modèle simple de Poisson sont illustrées sur le graphique (3.2) avec la valeur fixée dans les simulations $\lambda = 1$ à gauche

et $\lambda = 20$ à droite. En présence de grandes erreurs quadratiques moyennes dans la stratégie de la référence fixe, qui sur les box-plots sont représentes comme des valeurs aberrantes, la comparaison avec la stratégie de référence choisie n'est pas suffisamment claire. Pour bien distinguer les erreurs quadratiques obtenues pour les critères, un box-plot supplémentaire a été ajouté sans ces valeurs extrêmes.

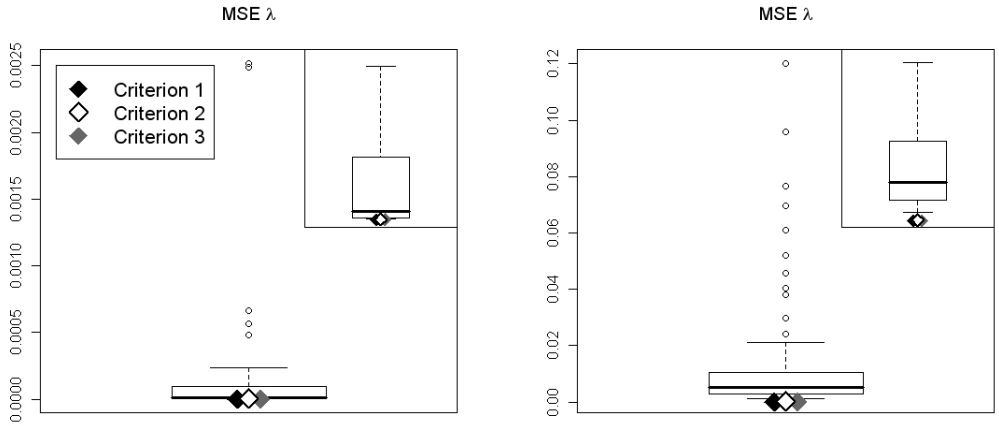

FIGURE 3.2: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 1 avec $g(\lambda) = \lambda$ pour $\lambda = 1$ (à gauche) et $\lambda = 20$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

L'atout de ce modèle simple est l'existence des espérances a posteriori sous une forme explicite. En effet la vraie espérance a posteriori pour un modèle de données $X_i^{(k)} \sim \mathcal{P}(\lambda)$ avec une loi a priori gamma $\mathcal{G}(\alpha, \beta)$ est $(\sum_{i=1}^{n} X^{(k)} + \alpha)/(\beta + n)$ pour tout jeu de données $X^{(k)}, k = 1, \ldots, K$. Ainsi, dans le tableau (3.2), pour chaque stratégie les erreurs quadratiques moyennes sont calculées par rapport aux vraies espérances a posteriori pour $\lambda = 1$ (à gauche) et $\lambda = 20$ (à droite).

Dans toutes les stratégies étudiées, les erreurs quadratiques moyennes sont petites (les valeurs reportées dans le tableau (3.2) sont multipliées par 10^6). Pour $\lambda = 1$, l'erreur quadratique moyenne des estimations calculées avec MCMC est plus petite que celles avec les autres stratégies. Cependant, quand $\lambda = 20$ pour la stratégie de référence choisie avec le premier et le troisième critère, on obtient des erreurs plus petites que celles obtenues avec MCMC. Dans la paramétrisation avec $\lambda = 1$, les stratégies de référence choisie sont meilleures que les stra-

tégies de références fixes quelque soit le jeu de données de référence. Dans la paramétrisation avec $\lambda = 20$, ce n'est pas vrai uniquement pour le critère 2, néanmoins pour les stratégies de référence choisie, on obtient les erreurs quadratiques toujours plus petites que celles dans le pire cas de référence fixe. Dans la stratégie de référence choisie avec le deuxième critère fondé sur la divergence de Kulback-Leibler, les erreurs quadratiques importantes peuvent se produire à cause du logarithme des poids. Pour $\lambda = 20$ le second critère donne une erreur quadratique moyenne égale à 124.95 10^{-6} à cause de deux valeurs extrêmes. Si ces dernières sont éliminées l'erreur devient 53.3 10^{-6} étant comparable avec les deux autres critères.

TABLE 3.2: Erreurs quadratiques moyennes par rapport à la vraie moyenne a posteriori dans le Modèle 1 pour $g(\lambda) = \lambda$ avec $\lambda = 1$ (à gauche) et $\lambda = 20$ (à droite) et $n = 20$

strategy	$\lambda = 1$		$\lambda = 20$	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
MCMC [3]	1.050	0.035	90.327	1.194
La meilleure ref. "fixe" [4]	3.481	0.071	1152.119	90.131
La pire ref. "fixe" [5]	2520.477	83.316	120260.480	10325.171
Stratégie de réf. "choisie" avec le critère 1 [6]	1.253	0.087	42.127	0.669
Stratégie de réf. "choisie" avec le critère 2 [6]	2.148	0.172	124.950	3.958
Stratégie de réf. "choisie" avec le critère 3 [6]	2.306	0.131	44.615	0.671

[1] Erreur quadratique moyenne par rapport à la vraie moyenne a posteriori($\times 10^6$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^6$)
[3] Stratégie avec MCMC ($\times 10^6$)
[4] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^6$)
[5] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^6$)
[6] IS avec la stratégie de réf. "choisie" ($\times 10^6$)

3.2.2 Modèle 2 : Régression de Poisson

Pour le modèle de Poisson avec une covariable Z et deux paramètres a et b, les erreurs quadratiques moyennes (les box-plots avec les résultats en fonction des critères) sont présentées sur la figure (3.3). Les trois critères dans la stratégie de référence choisie fournissent une erreur quadratique moyenne plus petite que la plupart des erreurs obtenues dans la référence fixe avec les différents jeux de données comme référence. Dans ce modèle, la vraie loi a posteriori n'est pas explicite donc les résultats d'estimation avec IS combiné sont comparées avec les

estimations MCMC. Les valeurs numériques des erreurs quadratiques moyennes obtenues dans le meilleur et le pire cas de référence fixe et celles dans la stratégie de référence choisie avec les trois critères sont présentées dans le tableau (3.3).

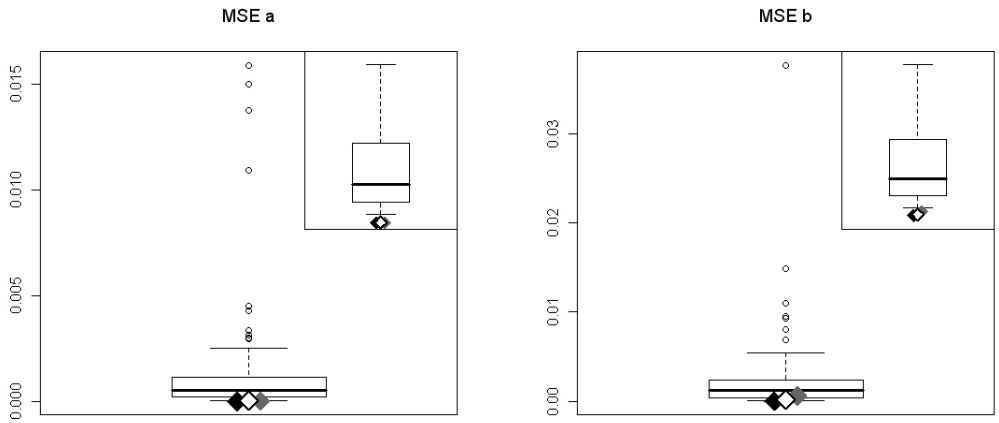

FIGURE 3.3: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 2 avec $\theta = (a = 0, b = 0.5)$ pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

TABLE 3.3: Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 2 avec $a = 0$, $b = 0.5$ et $n = 20$ pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (à droite)

strategy	a		b	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	0.042	0.003	0.046	0.003
La pire ref. "fixe" [4]	15.886	2.930	37.664	2.047
Stratégie de réf. "choisie" avec le critère 1 [5]	0.007	<0.001	0.014	0.001
Stratégie de réf. "choisie" avec le critère 2 [5]	0.058	0.070	0.155	0.020
Stratégie de réf. "choisie" avec le critère 3 [5]	0.026	0.001	0.594	0.001

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

3.2.3 Modèle 3 : Régression de Poisson avec extravariabilité

Les résultats des estimations pour le modèle de Poisson avec extravariablité avec une covariable Z et un résidu ϵ concernent trois paramétres a, b et σ^2 et sont présentes sous forme graphique dans la figure (3.4) et sous forme numérique dans le tableau (3.4) pour $\sigma^2 = 1/8$, $\sigma^2 = 1/4$ et $\sigma^2 = 1/2$ dans la première, deuxième et la troisième ligne respectivement. Ces résultats montrent que la stratégie de référence choisie avec les trois critères donnent les erreurs quadratiques moyennes (calculées par rapport à MCMC car les lois a posteriori ne sont pas explicites) très comparables et toujours plus petites que celles dans le pire cas de référence fixe. Cependant la performance de la stratégie de référence fixe peut être meilleure cas les erreurs quadratiques moyennes dans le meilleur cas de référence fixe sont toujours plus petites que celles pour les critères.

3.2.4 Augmentation de la taille de l'échantillon et du nombre de covariables

Comme première extension, la taille de chaque jeu de données a été augmentée à $n = 1000$ dans le modèle 3 ($a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$) et comme seconde extension, le nombre de covariables est également augmenté à 10 covariables : $\log \lambda_i = a + \sum_{j=1}^{10} b_j Z_{ij} + \epsilon_i$. Les covariables Z_{ij} ont été simulées de deux manières différentes. Dans le premier cas $Z_{ij} \sim \mathcal{N}(0,1)$ pour $i = 1, \ldots, 10$ et dans le deuxième cas $Z_{ij} \sim \mathcal{N}(0,1)$ pour $i = 1, \ldots, 5$ et $Z_{ij} \sim \mathcal{B}ernoulli(0.25)$ pour $i = 6, \ldots, 10$. Les paramètres b_j ont été fixés à 0.05 ($a = 0$ et $\sigma^2 = 1/2$). Dans cette partie, les résultats avec $n = 1000$ sont présentés et ceux concernant les cas avec 10 covariables sont mis en annexe.

Les lois a priori des coefficients b_j ont été choisies identiques ($\mathcal{N}(0, 10^5)$). Les lois a priori des autres paramètres sont inchangées (Cf. tableau (2.14)).

Pour ces trois modèles, les box-plots des erreurs quadratiques sont présentés dans la figure (3.5) pour $n = 1000$ et dans les figures (14) et (15) placées en annexe pour les deux cas avec 10 covariables. Les valeurs numériques des erreurs quadratiques dans le meilleur et le pire cas de référence fixe avec les erreurs quadratiques moyennes pour les critères sont présentées dans

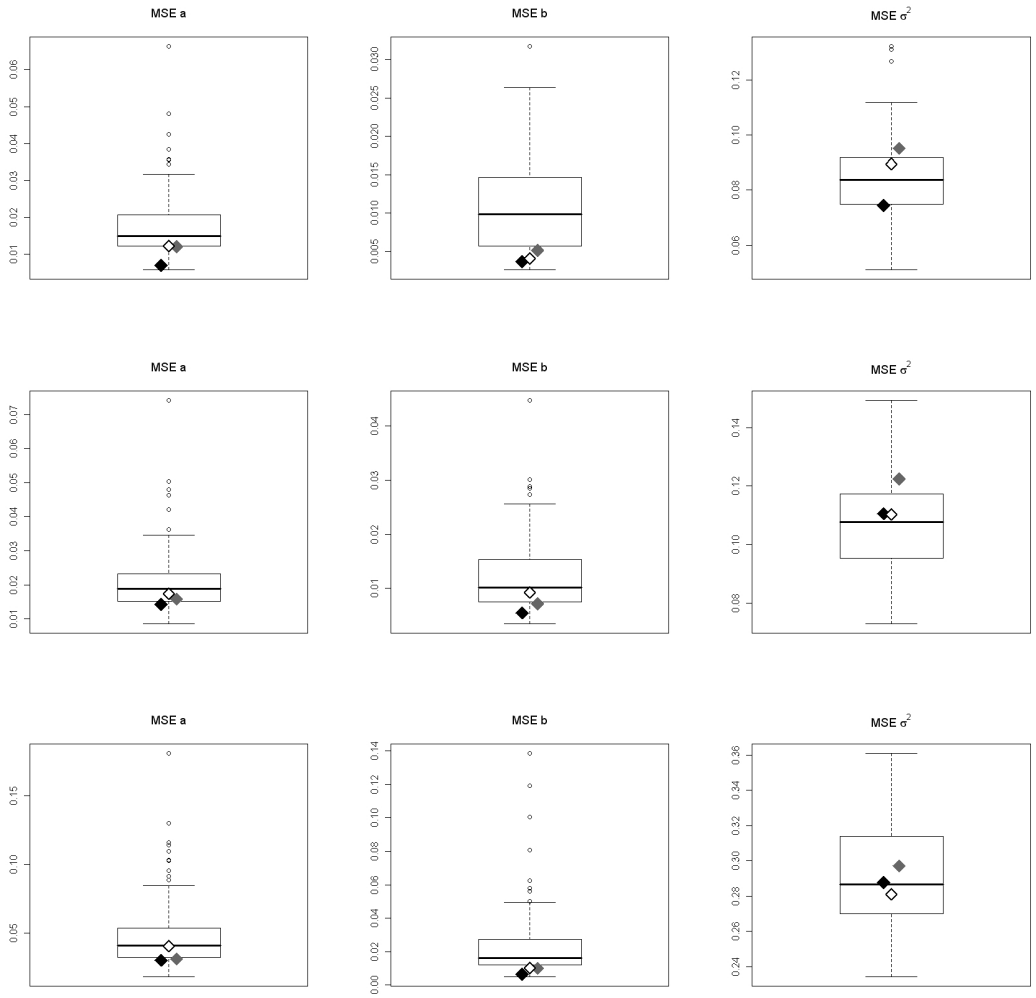

FIGURE 3.4: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$ (en première ligne), $\sigma^2 = 1/4$ (en deuxième ligne) ou $\sigma^2 = 1/2$ (en troisième ligne), pour $g(\theta) = a$ (à gauche) et $g(\theta) = b$ (au centre), $g(\theta) = \sigma^2$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

le tableau (3.5) pour $n = 1000$ et dans les tableaux (7), (8) placés en annexe pour les cas à 10 covariables. Ces erreurs quadratiques sont calculées par rapport aux résultats d'estimation avec MCMC.

La figure (3.5) et le tableau (3.5) peuvent être comparés avec ceux obtenus pour le même modèle mais avec $n = 20$ (Cf. la denière ligne de (3.4) et (3.4)). Comme attendu, l'aug-

TABLE 3.4: Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/8$ (en première partie) ou $\sigma^2 = 1/4$ (en deuxième partie) ou $\sigma^2 = 1/2$ (en troisième partie) et $n = 20$ pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) and $g(\theta) = \sigma^2$ (à droite)

strategy	a		b		σ^2	
$\sigma^2 = 1/8$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	5.791	0.385	0.261	0.222	51.079	4.224
La pire ref. "fixe" [4]	66.384	3.928	31.762	2.814	132.263	8.763
Stratégie de réf. "choisie" avec le critère 1 [5]	7.058	0.579	3.674	0.338	74.385	5.527
Stratégie de réf. "choisie" avec le critère 2 [5]	12.325	1.725	4.125	0.344	89.494	5.963
Stratégie de réf. "choisie" avec le critère 3 [5]	12.122	1.311	5.156	0.564	95.190	6.340
	a		b		σ^2	
$\sigma^2 = 1/4$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	8.531	1.075	3.453	0.381	72.822	4.434
La pire ref. "fixe" [4]	74.240	10.928	44.829	3.199	149.402	12.262
Stratégie de réf. "choisie" avec le critère 1 [5]	14.282	1.672	5.498	0.568	110.520	9.642
Stratégie de réf. "choisie" avec le critère 2 [5]	17.309	2.562	9.282	0.861	110.328	9.902
Stratégie de réf. "choisie" avec le critère 3 [5]	15.785	1.831	7.183	0.568	122.447	12.324
	a		b		σ^2	
$\sigma^2 = 1/2$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	17.827	2.053	4.859	0.396	234.222	18.205
La pire ref. "fixe" [4]	181.103	8.557	138.765	5.907	361.212	33.731
Stratégie de réf. "choisie" avec le critère 1 [5]	29.918	2.782	6.398	0.645	287.724	31.109
Stratégie de réf. "choisie" avec le critère 2 [5]	40.138	3.519	10.198	0.804	280.986	31.820
Stratégie de réf. "choisie" avec le critère 3 [5]	30.869	2.389	9.941	0.761	297.203	27.430

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

mentation du nombre des observations par jeu de données a réduit les valeurs des erreurs quadratiques. Les résultats des trois critères sont toujours meilleurs que le pire cas de référence fixe.

Pour les deux autres modèles, les critères fournissent les meilleurs résultats en terme

d'erreurs quadratiques moyennes que les pires cas de référence fixe. Le meilleur cas de référence fixe donnant des erreurs quadratiques plus petites que celles obtenues avec les critères dans la stratégie de référence choisie est à nouveau retrouvé.

Dans les deux modèles avec 10 covariables, les erreurs quadratiques moyennes sont très comparables dans les estimations des cinq premiers paramètres $b_1 - b_5$.

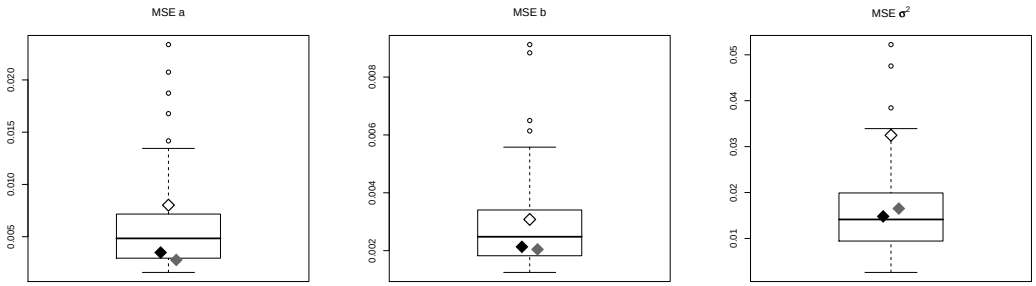

FIGURE 3.5: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/2$ et $n = 1000$, pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) et $g(\theta) = \sigma^2$ (à droite) ; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

TABLE 3.5: Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 3 avec $a = 0$, $b = 0.5$, $\sigma^2 = 1/2$ et $n = 1000$ pour $g(\theta) = a$ (à gauche), $g(\theta) = b$ (au centre) et $g(\theta) = \sigma^2$ (à droite)

strategy	a		b		σ^2	
$\sigma^2 = 1/2$	MSE [1]	$q_{97.5\%}$ [2]	MSE [1]	$q_{97.5\%}$ [2]	MSE [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	1.580	0.077	1.247	0.067	2.600	0.126
La pire ref. "fixe" [4]	23.382	0.520	9.124	0.266	52.230	1.083
Stratégie de réf. "choisie" avec le critère 1 [5]	3.486	0.144	2.131	0.097	14.806	0.506
Stratégie de réf. "choisie" avec le critère 2 [5]	8.019	0.308	3.079	0.139	32.478	0.890
Stratégie de réf. "choisie" avec le critère 3 [5]	2.776	0.169	2.043	0.083	16.514	0.602

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

3.3 Stratégie de mélange

3.3.1 Résultats

La stratégie de mélange est appliquée dans les deux modèles. Le premier est un modèle simple de Poisson (Modèle 1) avec le paramètre de moyenne $\lambda = 20$ fixé dans les simulations des données. Pour ce modèle simple, les constantes de normalisation de l'estimateur (2.47) dans l'expression des poids (2.49) sont calculées par la méthode *reverse logistic regression*. La loi de mélange est construite avec les 10 premières densités a posteriori présélectionnées. Les résultats des estimations ont été comparés aux vraies espérances a posteriori connues pour ce modèle. Les erreurs quadratiques moyennes obtenues pour cette stratégie complètent les résultats obtenus dans la section (3.2.1) et sont présentées dans le tableau (3.6) en gras. En terme d'erreur quadratique, la stratégie de mélange est la meilleure parmi toutes les estimations y compris MCMC classique néanmoins reste comparable avec la stratégie de référence choisie avec le premier et le troisième critère.

TABLE 3.6: Erreurs quadratiques moyennes par rapport à la vraie moyenne a posteriori dans le modèle 1 pour $g(\lambda) = \lambda$ avec $\lambda = 20$ et $n = 20$ ensemble avec la stratégie de mélange

strategy	$\lambda = 20$	
	MSE [1]	$q_{97.5\%}$ [2]
MCMC [3]	90.327	1.194
La meilleure ref. "fixe" [4]	1152.119	90.131
La pire ref. "fixe" [5]	120260.480	10325.171
Stratégie de réf. "choisie" avec le critère 1 [6]	42.127	0.669
Stratégie de réf. "choisie" avec le critère 2 [6]	124.950	3.958
Stratégie de réf. "choisie" avec le critère 3 [6]	44.615	0.671
Stratégie de mélange	**39.914**	**0.506**

[1] Erreur quadratique moyenne par rapport à la vraie moyenne a posteriori($\times 10^6$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^6$)
[3] Stratégie avec MCMC ($\times 10^6$)
[4] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^6$)
[5] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^6$)
[6] IS avec la stratégie de réf. "choisi" ($\times 10^6$)

Le deuxième modèle est un modèle de Poisson linéaire généralisé sans extra-variabilité

comme le modèle 2 mais avec 10 covariables : $log(\lambda_i) = a + \sum_{i=j}^{10} b_j Z_{ij}$ où $Z_{ij} \sim \mathcal{N}(0,1)$ pour $i = 1, \ldots, 10$. Dans les simulations des données, tous les paramètres b_j $(j = 1, \ldots, 10)$ ont été fixés à 0.05 et $a = 0$. Les lois a priori sur tous les paramètres ont été prises vagues gaussiennes $\mathcal{N}(0, 10^5)$. Les constantes de normalisation de l'estimateur (2.47) avec les poids (2.49) ont été estimées par la méthode *reciprocal importance samling* car la méthode *reverse logistic regression* a été jugée trop longue. Les estimations ont été comparés à celles obtenues via MCMC car les vrais espérances a posteriori sont non-explicites. Les résultats de cette stratégie sont présentés dans le tableau (9) en annexe avec les résultats de la stratégie de référence choisie avec les trois critères. Les conclusions restent similaires à celles du première modèle : la stratégie de mélange donne une légère meilleure performance que la deuxième stratégie avec les critères.

3.3.2 Estimations des constantes de normalisation dans la stratégie de mélange

L'analyse de la littérature à été faite pour identifier les méthodes d'estimations de ces constantes afin de cibler celles qui sont les plus adaptées et suffisamment performantes dans le cadre des estimations répétées. Quelques méthodes sont sélectionnées et testées pour 5 modèles avec des lois a priori choisies conjuguées. Ainsi, pour ces modèles, les lois a posteriori sont connues et donc les constantes de normalisation également. Ceci permet l'étude des performances de l'estimateur IS avec un mélange en comparant les résultats aux vraies valeurs.

Les méthodes d'estimations des ratios de constantes de normalisation ont été testées sur 5 modèles appartenant aux familles exponentielles avec des lois a priori conjuguées. Ainsi, les constantes de normalisation sont connues explicitement dans ce cas. Les différents modèles envisagés et leurs caractéristiques (vraisemblance, densité a priori et a posteriori, l'espérance a posteriori et le rapport de constantes de normalisations) sont résumés en annexe dans le tableau (10). Dans chaque cas, seulement un seul paramètre θ est considéré, les autres (hyperparamètres, ...) sont fixés et supposés connus. Sous chaque modèle, 100 jeux de données de taille $n = 20$ ont été générés pour une valeur de paramètre θ fixée à θ_0. Tout d'abord, les réalisations des lois a posteriori ont été obtenues via MCMC pour tous les jeux de données.

TABLE 3.7: Erreurs quadratiques moyennes entre l'estimation du rapport des constantes de normalisation et sa vraie valeur.

Modèle :	Norm-Norm	Pois-Gam	Bin-Beta	Gam-Gam	NégBin-Beta
$\frac{1}{9}\sum_{k=1}^{10}(r_0-\hat{r}_{is})^2$	8.400e-04	4.785e-05	2.327e-01	6.356658e-01	1.763e+04
$\frac{1}{9}\sum_{k=1}^{10}(r_0-\hat{r}_{recip})^2$	8.673e-12	1.108e-08	3.862e-03	4.537e-10	9.241e-01
$\frac{1}{9}\sum_{k=1}^{10}(r_0-\hat{r}_{ris})^2$	2.296e+00	1.130e-05	8.818e-02	3.647e+00	3.579e+01

Ensuite, $M = 10$ jeux de données sont présélectionnés (les 10 premiers) pour lesquels les réalisations via MCMC sont gardées et pour les 90 autres jeux de données, les espérances a posteriori de θ sont estimées par importance sampling soit avec la stratégie de référence choisie selon le critère utilisant la norme L_1 (voir 2.2.4) soit avec la stratégie de mélange (estimateur (2.47)). Pour l'approche avec mélange, les rapports des constantes de normalisation ont été estimés par trois méthodes différentes présentées dans la section (2.2.5.2) : importance sampling classique, *reciprocal importance sampling* et *ratio importance sampling*. Pour l'approche importance sampling classique, le choix de la fonction d'importance a été fait graphiquement étant dans le cas unidimensionnel. Ainsi si la $p^{\text{ième}}$ densité a posteriori est choisie comme la fonction d'importance, les mêmes rapports $r_m = c_m/c_p$, $m = 1, \ldots, p-1, p+1, \ldots, M$ seront estimés également par *reciprocal importance sampling* et *ratio importance sampling*.

Les résultats concernant les estimations des ratios de constantes de normalisation sont en annexe dans le tableau (11) et les erreurs quadratiques moyennes associées à ces résultats se trouvent dans le tableau (3.7). Parmi les trois méthodes comparées, l'approche *reciprocal importance sampling* (colonne $\hat{r}_{recip}$) donne les résultats les plus proches des vrais rapport de constantes de normalisation. Ceci est en accord avec les résultats théoriques présentés par Chen and Shao (1997).

Toutes les estimations ont été ensuite comparées aux vraies valeurs des espérances a posteriori.

Les estimations des rapports ont été ensuite injectées dans les poids (2.50) de l'estimateur IS avec la stratégie de mélange (2.47). Les résultats des estimations des espérances a posteriori de θ sont ensuite comparés avec les vraies valeurs via les erreurs quadratiques moyennes. Ces résultats sont présentés dans le tableau (3.8) ainsi que ceux obtenus directement par MCMC ou par l'estimateur IS avec référence choisie par norme L_1. Ces résultats montrent

TABLE 3.8: Erreurs quadratiques moyennes entre l'estimation de l'espérance a posteriori de θ et sa vraie valeur.

Modèle :	Norm-Norm	Pois-Gam	Bin-Beta	Gam-Gam	NégBin-Beta
Méthode :					
MCMC	7.693e-06	4.677e-04	3.581e-07	3.125e-08	3.248e-07
IS stratégie 2 (L_1)	3.338e-05	1.261e-03	5.409e-07	1.838e-07	6.147e-07
IS mélange avec $\hat{r}_{is}$	4.990e-05	2.741e-04	2.628e-07	2.579e-06	2.068e-07
IS mélange avec $\hat{r}_{recip}$	4.300e-05	1.364e-04	2.110e-07	1.383e-07	5.243e-08
IS mélange avec $\hat{r}_{ris}$	4.905e-05	1.750e-04	2.478e-07	1.605e-05"	1.376e-07

que pour les modèles étudiés, la stratégie de mélange utilisant la méthode *reciprocal importance samling* permet d'obtenir bien souvent les meilleures estimations. Les résultats d'estimations obtenus directement avec MCMC restent seulement meilleurs pour le modèle Normal-Normal et Gamma-Gamma mais les erreurs dans l'ensemble (3.8) sont petites. L'objectif de cette analyse a été le choix d'une méthode pour estimer les rapports de constantes de normalisation intervenant dans l'estimateur IS sous la forme (2.45) et dans le cadre d'estimations répétées. Les résultats suggèrent de choisir la méthode "reciprocal importance sampling" qui a donné les meilleurs résultats en terme d'estimation de l'espérance a posteriori ainsi que d'estimation des rapports de constantes de normalisation. Son point fort est aussi sa simplicité et la possibilité d'estimer individuellement les constantes de normalisation comme leur inverse (2.52).

3.4 Sélection automatique

Le rôle des jeux de données présélectionnés est essentiel dans la stratégie de référence choisie ou la stratégie de mélange. L'extension suivante consiste à présélectionner un lot de M jeux de données de références de manière à ce que les densités a posteriori correspondantes constituent un ensemble varié des fonctions d'importance potentielles. Pour comprendre l'idée de cette sélection que l'on appelle automatique, rappelons que pour le modèle de Poisson simple $\mathcal{P}(\lambda)$ (modèle 1) pour les données $X^{(k)} = (X_1^{(k)}, \ldots, X_n^{(k)})$ et une loi a priori gamma $\mathcal{G}(\alpha, \beta)$ sur le paramètre λ, la loi a posteriori de λ est une loi Gamma $\mathcal{G}(\alpha + \sum_{i=1}^n X^{(k)}, \beta + n)$. Pour chaque jeu de données $X^{(k)}$ de taille n, cette lois a posteriori dépend de la somme $\sum_{i=1}^n X^{(k)}$ des observations. Les lois a posteriori dans ce modèle se distinguent par cette somme. Obtenir différentes lois a posteriori revient à choisir des jeux de données pour lesquels

les sommes des observations diffèrent. Dans cette exemple, on peut indiquer une statistique qui identifie la loi a posteriori. Parmi l'ensemble des jeux de données, on souhaite donc choisir un lot de M (ici 10) dont les sommes constituent une représentation de toutes les sommes des 90 autres jeux de données - toutes les valeurs de la statistique décrivant la loi a posteriori. Ainsi, le choix de la fonction d'importance étant une de ces densités a posteriori, on espère que ce choix sera fait de manière optimale pour chaque jeu de donnée.

L'idée est donc de construire M clusters à partir de K densités a posteriori puis de sélectionner une densité a posteriori dans chaque cluster. La construction des clusters est fondée sur la distance entre les statistiques caractérisant les densités a posteriori.

La méthode de classification la plus adaptée ici est la méthode de "k-medoids" qui partitionne l'ensemble des éléments en M clusters et choisit un élément central de chaque cluster appelé "medoid".

Différentes méthodes de classification existent sous R. La procédure de sélection automatique est illustrées ici avec la méthode *Partitioning Around Medoids* (PAM) proposée par Kaufman and Rousseeuw (1990) documentée par exemple dans Nakache and Confais (2005). La fonction de R s'appelle "PAM" dans le package Cluster de (Maechler et al., 2005). Il existe aussi d'autres méthodes de "k-medoids" comme *Clustering LARge Applications* (CLARA) (Kaufman and Rousseeuw, 1990), *Clustering LARge Applications based on RANdomized Search* (CLARANS) (Ng and Han, 1994) et *Fast Intelligent Subspace Clustering Algorithm using Dimension Voting* (Woo et al., 2004).

Le modèle 1 sert ici seulement comme exemple explicatif mais n'a pas de réel intérêt dans la pratique. La sélection automatique est alors utilisée pour le modèle 3 ($a = 0$, $b = 0.5$ et $\sigma^2 = 1/2, n = 20$) correspondant à des situations réalistes plus usuelles. La densité a posteriori dans ce modèle est inconnue. Comme la méthode PAM nécessite l'utilisation d'une statistique "représentant" les données, le choix arbitraire de la moyenne empirique a été fait ici (d'autres choix peuvent être envisagés). La méthode PAM a été ensuite utilisée sur R avec la distance Euclidienne.

La figure (3.6) illustre les densités marginales a posteriori de chaque paramètre estimées classiquement par MCMC. Les densités noires correspondent aux 10 jeux de données sélection-

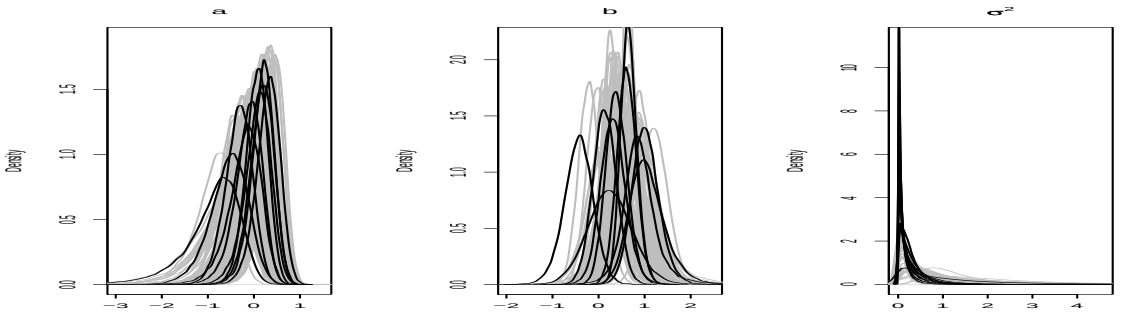

FIGURE 3.6: Densités marginales a posteriori de a (à gauche), b (au centre) et σ^2 (à droite) approchées par MCMC dans le Modèle 3 avec les paramètres $\theta = (a = 0, b = 0.5, \sigma^2 = 1/2)$. Dix densités a posteriori sélectionnées par la méthode de classification PAM sont tracées en gras.

nés par la méthode de classification PAM sur la base des moyennes empiriques. Ce graphique montre assez clairement que les densités présélectionnées représentent bien l'ensemble de densités a posteriori.

Afin d'illustrer l'impact de cette sélection automatique, le lot de 10 densités présélectionnées correspondant aux 10 jeux de données a été utilisé dans la stratégie de référence choisie avec les trois critères. Les estimations avec IS combiné ont été comparées avec les estimations MCMC via les erreurs quadratiques, les résultats étant présentés dans le tableau (3.9). Ces résultats sont très comparables avec ceux rapportés dans le tableau (3.4). Ils restent toujours meilleurs que les résultats obtenus pour le pire référence fixe.

TABLE 3.9: Erreurs quadratiques moyennes dans le Modèle 3 avec la Sélection Automatique avec les paramètres $\theta = (a = 0, b = 0.5, \sigma^2 = 1/2)$ pour $g(\theta) = a$ (en première partie), $g(\theta) = b$ (en deuxième partie) et $g(\theta) = \sigma^2$ (en troisième partie)

stratégie	a		b		σ^2	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
Stratégie de réf. "choisie" avec le critère 1 [3]	25.784	1.790	5.073	0.337	259.049	27.550
Stratégie de réf. "choisie" avec le critère 2 [3]	40.005	3.577	13.133	0.805	279.517	35.416
Stratégie de réf. "choisie" avec le critère 3 [3]	32.721	3.098	6.307	0.449	292.787	30.407

[1 – 2] Cf. la légende du tableau 3.5

[3] IS avec la stratégie de réf. "choisie"

Cette sélection automatique offre l'avantage d'éviter à l'utilisateur le choix arbitraire des 10 jeux de données présélectionnés et peut permettre éventuellement d'écarter les cas de mauvais choix arbitraire.

4 Discussion

Dans les modèles étudiés les stratégies de IS combiné ont montré de bonnes performances par rapport aux résultats MCMC. Globalement les erreurs quadratiques moyennes ont été petites. Les trois critères de la stratégie de référence choisie ont permis d'améliorer le choix de référence utilisée par l'estimateur IS pour chaque nouvelle estimation, dans le sens où les résultats obtenus avec les critères sont toujours meilleurs que le pire cas de référence fixe. Presque pour tous les modèles, ils existent des cas de références fixes qui donnent des résultats meilleurs que ceux obtenus avec les critères. Dans la stratégie de référence fixe le choix de jeu de données de référence est fait en quelque sorte au hasard. En pratique, rien ne garantit d'éviter un "mauvais" choix de référence pour faire toutes les estimations.Le critère basé sur la minimisation de la norme L_1 semble être le plus stable des trois et aussi le plus simple.

Toujours afin d'améliorer les performances des estimateurs, une troisième stratégie basée sur un mélange à été proposée dans le contexte des estimations répétées. La fonction d'importance de l'estimateur IS est construite comme un mélange de densités a posteriori correspondantes au lot de jeux de données présélectionnés. La stratégie a été appliquée d'abord aux deux modèles de Poisson montrant une très bonne performance. Cette stratégie a soulevé le problème d'estimation des constantes de normalisation. Dans un deuxième temps, plusieurs méthodes d'estimation des constantes de normalisation ou de leur rapport ont été étudiées et appliquées dans les estimations répétées des paramètres de modèles de familles conjuguées. Pour ces modèles, les espérances a posteriori sont entièrement connues. Les résultats des estimations ont été ainsi comparés aux vraies espérances. De plus les estimations ont été aussi

effectuées avec MCMC et avec la stratégie de référence choisie avec le première critère (L_1). Les estimations obtenues avec la stratégie de mélange avec entre autre la méthode *reciprocal importance sampling* ont montré une bonne performance par rapport aux résultats de MCMC et IS combiné avec la référence choisie par le premier critère.

Ces analyses peuvent être considérées comme préliminaires et des recherches supplémentaires devront être effectuées.

Une partie importante du chapitre relatif aux méthodes a été consacrée aux algorithmes MCMC pour insister sur leurs caractère itératif qui peut s'avérer long. L'importance sampling dans le cadre des estimations répétés a été présenté comme une alternative plus rapide.

Les méthodes proposées ici ne sont pas à être considérées comme solution universelle. Elles doivent être plutôt utilisées avec une précaution. Les éléments à prendre en compte sont : la structure du modèle, l'évaluation de la vraisemblance et la proportion des jeux de données pour lesquels MCMC est utilisé.

Si le modèle est simple, les algorithmes MCMC vont être très rapides et le gain dans l'utilisation des stratégies avec IS combiné va être infime. L'enjeu des analyses présentées ici étaient surtout d'évaluer la performances des estimations entre les deux types d'approches. Tous les modèles de Poisson étudiés ont une structure hiérarchique très simple donc l'inférence bayésienne via les algorithmes MCMC ne constitue pas un vrai poids calculatoire. La stratégie de référence fixe et sans aucune doute plus rapide mais un mauvais choix de la référence peut amener à de nettement moins bons résultats. La performance des estimations est améliorée par la stratégie de référence choisie mais son prix est un effort supplémentaire dans l'évaluation des critères.

En effet dans les trois critères, il faut évaluer le rapport des vraisemblances $\pi(X^{(k)}|\theta)/\pi(X^{(m_k)}|\theta)$ pour θ parmi toutes les réalisations de M chaînes MCMC

Une fois que toutes les vraisemblances $\pi(X^{(m_k)}|\theta)$ pour $m_k = 1, \ldots, M$ sont évaluées, pour chaque nouveau k il faut encore calculer $\pi(X^{(k)}|\theta)$. Dans la stratégie de mélange, à la même difficulté s'ajoute l'évaluation des constantes de normalisation ou de leur rapport.

Ainsi si la vraisemblance est difficile à évaluer car par exemple non explicitée, son calcul va être long aussi bien dans MCMC que dans les stratégies IS combiné avec MCMC. La rapidité

des stratégies avec IS par rapport à l'utilisation systématique de MCMC va dépendre donc de la durée des autres calculs à l'intérieur de l'échantillonneur MCMC utilisé.

Si un modèle étudié a une structure complexe pour lequel MCMC est long mais la vraisemblance est simple à calculer, on peut probablement prendre M plus grand (tout de même petit par rapport au nombre total K). Si au contraire, la vraisemblance est difficile à calculer, une trop grande valeur de M ne permettra probablement pas aux stratégies combinées d'accélérer les calculs.

La réponse exacte sera certainement fournit par le calcul de la complexité numérique de ces deux approches. Elle dépend du modèle, de la vraisemblance, de la valeur de M et de l'algorithme MCMC choisi. Cela signifie que la construction de l'algorithme MCMC doit être connue. Il devrait être idéalement construit sur mesure pour le modèle étudié. Dans ce cas le calcul de la complexité numérique de l'algorithme s'ajoute à la construction de ce dernier. Ces efforts supplémentaires ne seront probablement pas souhaités par un utilisateur de logiciel.

Par rapport aux algorithmes MCMC, les méthodes proposées ici ne requièrent pas de compétences particulières de programmation donc elles peuvent être utilisées conjointement avec les logiciels d'analyse bayésienne car les stratégies combinées utilisent les réalisations MCMC pour un certain nombre des jeux de données M. Il suffit donc avec la première réalisation de MCMC (pour le jeu de données $X^{(1)}$) d'évaluer le temps (réel) approximatif d'évaluations de trois critères dans la stratégie de référence choisie, une constante de normalisation (ou un rapport des constantes) et l'ensemble des vraisemblances dans la stratégie de mélange pour avoir une idée de la durée des stratégies combinées pour M jeux de données. Cela a un sens seulement si les MCMC sont effectivement longs pour l'ensemble des estimations.

Pour les raisons présentées ci-dessus, le temps (réel) des calculs a été évalué seulement dans les deux modèles avec 10 covariables (sans et avec extra-vérifiabilité).

Déjà pour ces modèles relativement simples, on peut observer une accélération des calculs pour les stratégies combinées.

Une première extension de cette partie consistera à explorer l'utilisation de la stratégie de mélange, tout comme les deux autres stratégies, dans d'autres familles de modèles plus

TABLE 4.1: Rapport des temps de calculs entre les méthodes combinées avec l'échantillonnage préférentiel et la méthode classique MCMC concernant la régression de Poisson avec $n = 1,000$ et 10 covariables.

	Stratégie 1 critère 1 (L_1)	Stratégie 2 critère 2 (KL)	Stratégie 2 critère 3	Stratégie 2 Mélange	Stratégie 3
Modèle 2 (10 cov.)	0.057	0.957	0.771	1.156	0.423
Modèle 3 (10 cov.)	0.036	0.278	0.266	0.652	-[1]

[1] La stratégie de mélange n'a pas été évaluée pour le modèle

compliqués. Une seconde extension sera d'essayer d'améliorer la stratégie de mélange par l'utilisation d'un mélange non-équiprobable où les poids dépendraient des jeux de données dans chaque nouvelle estimation.

Ces concepts devront être analysés avec un soin particulier vis à vis de la duré et de la complexité des calculs qu'ils entraînent.

Modélisation de la transmission de maladies infectieuses dans une population. Estimation des paramètres.

5 Introduction

Cette partie du manuscrit est consacrée au modèle SIS (susceptibles - infectés - susceptibles) dans une population de taille constante, modèle décrit comme une chaîne de Markov. Le modèle SIS est un modèle qui sous certaines conditions permet de décrire l'évolution d'une maladie dans une population dans laquelle on distingue seulement deux types de sous-populations : un groupe des sujets susceptibles et (S) un groupe des sujets infectés (I). Un sujet susceptible peut devenir infecté via un contact infectieux uniquement avec un sujet déjà infecté. La probabilité d'infection dépend alors à la fois du taux des contacts infectieux et des nombres de sujets infectés et susceptibles. De plus chaque sujet infecté guérit avec une même probabilité et devient ainsi à nouveau susceptible. Une fois guéri il n'acquiert aucune résistance ni immunité et sa probabilité de redevenir infecté lors d'un contact infectieux est la même qu'au départ. Les décès et les migrations - toutes causes confondues - sont possibles, cependant chaque individu décédé ou parti est remplacé par un sujet susceptible, ce qui garantit la taille constante de la population. Dans ce modèle la maladie étudiée ne se caractérise pas par un autre état intermédiaire comme une phase latente où des sujets infectés ne sont pas encore infectieux (ou cette phase est négligeable), il y a un seul niveau de susceptibilité et d'infectiosité pour tous les individus.

Le modèle SIS ainsi défini est étudié entre autre par Allen and Burgin (2000). Il s'agit d'un modèle simple d'épidémie qui peut être présenté comme un modèle compartimental à deux compartements : un pour les sujets susceptibles et un pour les sujets infectés. Les transitions vers et depuis chaque état étant données par les différents taux. Un moyen simple pour décrire les évolutions temporaires des effectifs dans deux sous-populations est d'y associer un système des équations différentielles. On parle alors d'un modèle déterministe. Il suppose que

la maladie évolue dans la population de façon homogène et fournit donc une solution moyenne des effectifs. Même s'il donne une information générale d'évolution de la maladie au sein de la population il ne prend pas en compte son hétérogénéité et ses éventuelles déviations de la moyenne. Pour une population de petite taille il semble plus adéquat d'utiliser un modèle stochastique qui saisit mieux cette hétérogénéité (ou si de grandes fluctuations sont attendues).

Il existe différents moyens d'introduire une partie aléatoire dans le modèle déterministe par exemple via l'ajout d'un aléa brownien aux équations différentielles ou une information a priori sur leurs paramètres. Allen and Burgin (2000) présente le modèle SIS sous la forme d'une chaîne de Markov à temps continu. Connaissant, la loi initiale, les taux des transitions et en prenant un petit incrément du temps, nous pouvons déterminer les probabilités instantanées de la chaîne via les calculs des puissances consécutives de la matrice de transition. La probabilité stationnaire est déduite via le calcul des valeurs et vecteurs propres de cette matrice. Si l'incrément de temps est assez petit on peut supposer que se produit une seule transition au maximum dans cet intervalle. Dans ce cas la probabilité instantanée pour chaque état peut être exprimée comme une fonction des probabilités instantanée de ses états voisins à l'instant précédant et des transitions qui mènent depuis et vers lui. C'est une sorte d'équation du bilan de probabilité qui pour un incrément infinitésimal du temps peut être transformée en une équation différentielle. En effet prenant tous les états du processus on obtient un système d'équations différentielles pour les probabilités instantanées. Cette démarche sera expliquée ultérieurement pour le modèle SIS, cependant il faut préciser ici qu'il s'agit de l'équation maîtresse (master équation) proposée sous ce nom par Van Kampen (1981, 2007). Le concept est connu en probabilité dans la théorie des chaînes de Markov, mais la notion d'équation maîtresse est plutôt connue en chimie et en physique où on parle de l'équation maîtresse chimique ou cinétique. Elle est souvent présentée comme la forme différentielle de l'équation de Chapman-Kolmogorov (progressive).

Selon la forme des probabilités de passage on distingue les processus stochastiques linéaires et nonlinéaires. On parle d'un processus stochastique linéaire si ses probabilités de transitions dépendent de façon linéaire de ses états, sinon il s'agit d'un processus stochastiques nonlinéaire. Dans le cas linéaire la solution de l'équation différentielle en probabilité (équation

maîtresse) est généralement analytique via sa transformation par la fonction génératrice de probabilité dont on déduit les probabilités instantanées. Il est aussi possible de déterminer les moments des processus en utilisant la fonction génératrice des moments. Dans la domaine de la biologie populationnelle Kendall (1949) donne les exemples de modèles stochastiques linéaires (les processus de naissance et de mort) décrivant la croissance de la population dont les probabilités sont obtenues de cette manière. Comme exemple de modèle épidémique on peut citer Stollenwerk and Briggs (2000) où l'équation maîtresse explicitée est appliquée pour modéliser une maladie des plantes (modèle SI : susceptibles - infectés).

Dans le cas nonlinéaire la transformation de l'équation maîtresse en équation de la fonction génératrice de probabilité ne permet généralement pas de déterminer la forme exacte de cette dernière. Théoriquement il est possible résoudre les système des équations différentielles directement mais ce peut être très laborieux si le nombre d'états possibles du processus n'est pas petit. Pour déterminer le comportement du processus il est possible d'utiliser des simulations (l'algorithme de Gillespie (1977)). Cette approche peut s'avérer longue car il faut générer un nombre important de trajectoires et un formalisme mathématique général peut être préférable. Dans le cas des probabilités de passage non linéaires on peut utiliser une approximation. De même que pour le modèle stochastique linéaire pour déterminer les moments du processus il est possible d'utiliser une fonction génératrice des moments ou une fonction génératrice des cumulants (Whittle, 1957; Goodman, 1953). Cela peut mener à un système ouvert d'équations différentielles où chaque équation dépend de la solution de la ou des solutions des équations suivantes. C'est le cas par exemple dans Matis and Kiffe (1996) avec un modèle logistique stochastique décrivant la croissance de la population via un processus de naissance et de mort avec les taux de naissance et de mort qui sont des polynômes d'ordre deux, ou bien dans Krishnarajah et al. (2005) pour le modèle SIS avec la probabilité d'une nouvelle infection étant aussi une fonction quadratique du nombre de sujets infectés. Pour clore le système d'équations il faut faire des hypothèses concernant la loi du processus, ses moments ou ses cumulantes - plus précisément les moments ou les cumulants d'ordre supérieur sont annulés. Whittle (1957) et Nåsell (2003) ont utilisé une approximation normale imposant ainsi que les cumulants d'ordre supérieur soient nuls, Keeling (2000) a supposé une loi log-normale

et Nåsell (2003) une loi de Poisson ou une loi binomiale. Il existe aussi des techniques plus récentes qui servent à clore le système d'équations des probabilités comme celles présentées par Singh and Hespanha (2007), Gómez-Uribe and Verghese (2007) ou Goutsias (2007). Le concept générale de ces techniques dite *moment-closure approximations* est bien résumé avec beaucoup des références dans Gillespie (2009).

Une autre idée est d'utiliser le développent de l'équation maîtresse proposé par Van Kampen (2007). L'idée de Van Kampen (2007) consiste à exprimer le processus étudié par une somme d'un phénomène déterministe et une fluctuation aléatoire dont la grandeur dépend d'un paramètre qui est défini comme la taille du système. Pour approcher la loi du processus, on part d'une solution moyenne et on lui ajoute le comportement aléatoire du processus obtenu via le développement de Taylor de l'équation maîtresse au le voisinage de cette solution. Cette technique, expliquée plus en détail dans la partie méthode, est utilisée, parmi de rares exemples, par Chen and Bokka (2005) qui applique le développement de l'équation maîtresse au modèle stochastique SIR (susceptibles - infectés - résistants). Alonso et al. (2007, +supp), propose de l'utiliser dans le même type de modèle, mais finalement il décrit le processus via les équations de Langevin et il déduit la probabilité de processus via la densité spectrale. Alonso and McKane (2002) l'utilise pour le N-patch modèle stochastique. Le rôle de ce modèle est de décrire le nombre (ou plutôt la fraction) de sous-population toujours existante (patchs) constituant la population entière (la métapopulation). L'existence de chaque sous-population dépend dans ce modèle de l'immigration, de la colonisation et du taux de croissance.

Le travail de thèse abordé dans cette partie est basé seulement sur la méthode de développement. La partie concernant des méthodes résume son concept général dans le cas du processus de naissance et de mort non linéaire. La théorie abordée ici provient principalement de l'ouvrage de van Kampen "Stochastic Processes in Physics and Chemistry". Ensuite, les méthodes sont appliquées au modèle SIS avec une taille de population constante. Ce modèle est d'abord décrit sous la forme déterministe puis comme une chaîne de Markov. Les résultats sont utilisés dans un premier temps dans la modélisation de nombre des sujets infectés, puis les deux modèles (un fondé sur la solution d'un modèle SIS déterministe classique et l'autre dont la solution est dérivée de la méthode de développement) sont utilisés pour construire la

méthode d'ajustement. Les propriétés des estimateurs des paramètres sont étudiées dans le cadre fréquentiste et bayésien. Les conclusions et les extensions sont présentées à la fin.

6 Matériel et Méthodes

6.1 Les processus de Markov à temps continu.

Les notions utilisées ici sont telles que établies par van Kampen.

Le processus de Markov est un processus stochastique $Y : \{Y(t), t \in R\}$ tel que pour n instants successifs $t_1 < t_2 < \ldots, < t_n$ où $t_i \in R$, les probabilités conditionnelles sont :

$$P(y_n, t_n | y_1, t_1; \ldots; y_{n-1}, t_{n-1}) = P(y_n, t_n | y_{n-1}, t_{n-1})$$

La probabilité conditionnelle de y_n à l'instante t_n sachant y_{n-1} à t_{n-1} ne dépend d'aucune valeur de Y à un instant antérieur. Un processus de Markov est entièrement déterminé par deux probabilités : une probabilité des différents états instantanée $P(y_1, t_1)$ et la probabilité de transition $P(y_2, t_2 | y_1, t_1)$ car à partir d'elles, les lois marginales de processus peuvent être reconstruites. Par exemple, pour trois instantes $t_1 < t_2 < t_3$

$$\begin{aligned} P(y_1, t_1; y_2, t_2; y_3, t_3) &= P(y_1, t_1; y_2, t_2) P(y_3, t_3 | y_1, t_1; y_2, t_2) = \\ &= P(y_1, t_1) P(y_2, t_2 | y_1, t_1) P(y_3, t_3 | y_2, t_2) \end{aligned} \tag{6.1}$$

Les lois marginales suivantes s'obtiennent par récurrence.

6.1.1 La formule de Chapman-Kolmogorov

L'intégrale de (6.1) par rapport à y_2 est égale à :

$$P(y_1, t_1; y_3, t_3) = P(y_1, t_1) \int P(y_2, t_2 | y_1, t_1) P(y_3, t_3 | y_2, t_2) dy_2$$

ainsi en divisant la dernière équation par $P(y_1, t_1)$ on obtient

$$P(y_3, t_3|y_1, t_1) = \int P(y_2, t_2|y_1, t_1)P(y_3, t_3|y_2, t_2)dy_2 \tag{6.2}$$

qui est l'équation de Chapman-Kolmogorov. Les probabilités de transitions doivent satisfaire cette équation ainsi que la relation suivante :

$$P(y_2, t_2) = \int P(y_2, t_2|y_1, t_1)P(y_1, t_1)dy_1$$

6.1.2 Stationnarité

Si le système dans lequel le processus Y est observé est dans un état d'équilibre, Y est un processus stationnaire. Dans ce cas la probabilité instantanée P ne dépend pas du temps $P(y, t) = P(y)$ et les probabilités de transition du processus markovien ne dépendent que de l'intervalle de temps séparant la transition. La définition classique de stationnarité dit que la loi marginale de processus n'est pas modifiée par un changement de temps donc

$$P(y_1, t_1 + \tau, y_2, t_2 + \tau; \ldots; y_n, t_n + \tau) = P(y_1, t_1, y_2, t_2; \ldots; y_n, t_n)$$

La probabilité de transition s'écrit comme

$$P(y_2, t_2|y_1, t_1) = T_\tau(y_2|y_1) \quad \text{où} \quad \tau = t_2 - t_1$$

et avec cette nouvelle notation l'équation de Chapman-Kolmogorov devient

$$T_{\tau+\tau'}(y_3|y_1) = \int T_{\tau'}(y_3|y_2)T_\tau(y_2|y_1)dy_2 \tag{6.3}$$

où $\tau' = t_3 - t_2$.

En général un processus markovien n'est pas stationnaire. Supposons que Y soit un processus de Markov défini par les probabilités $P(y_0)$ et $T_\tau(y_1|y_0)$. Si par exemple , pour un instant t_0 fixe et une valeur y_0 fixe avec $P(y, t_0) = \delta(y - y_0)$, un processus de Markov Y^* est défini pour $t_1 > t_0$ en posant

$$P^*(y_1, t_1) = T_{t_1-t_0}(y_1|y_0) \quad \text{ou plus généralement} \quad P^*(y_1, t_1) = \int T_{t_1-t_0}(y_1|y_0)p(y_0)dy_0$$

$$P^*(y_2, t_2|y_1, t_1) = T_{t_2-t_1}(y_2|y_1)$$

Le processus Y^* est non-stationnaire car les probabilités instantanées dépendent de l'instant t_0, néanmoins les probabilités de transition dépendent seulement de l'intervalle de temps. Il s'agit donc d'un processus de Markov homogène. On peut s'attendre à ce qu'au bout d'un certain temps ($t_1 \to \infty$) le système atteigne l'équilibre et qu'ainsi $P^*(y_1, t_1) \to P(y_1)$ et $T_{t_1-t_0}(y_1|y_0) \to P(y_1)$.

6.2 Equation Maîtresse

Cette section présente la dérivation de van Kampen de l'équation maîtresse (master equation) à partir de l'équation de Chapman-Kolmogorov (6.3) pour un processus de Markov homogène avec les probabilité de transition décrite par T_τ. L'équation maîtresse est déduite en prenant la limite quand τ' tend vers zéro dans l'équation de Chapman-Kolmogorov. La probabilité de transition $T_{\tau'}(y_2|y_1)$ pour un petit τ' peut être exprimée comme

$$T_{\tau'}(y_2|y_1) = (1 - a_0\tau')\delta(y_2 - y_1) + \tau' W(y_2|y_1) + o(\tau') \tag{6.4}$$

où $W(y_2|y_1)$ est un taux de transition de y_1 à y_2 par unité de temps, le coefficient $1 - a_0\tau'$ est la probabilité d'aucune transition pendant τ' et donc :

$$a_0(u) = \int W(y_2|u)dy_2 \tag{6.5}$$

étant une conséquence de $\int T_\tau(y_2|y_1)dy_2 = 1$. On remplace l'expression pour $T_{\tau'}$ (6.4) dans l'équation (6.3) et on obtient

$$T_{\tau+\tau'}(y_3|y_1) = \left[1 - a_0(y_3)\tau'\right] T_\tau(y_3|y_1) + \tau' \int W(y_3|y_1)T_\tau(y_2|y_1)dy_2$$

En remplaçant a_0 par l'intégrale de l'équation (6.5), et en prenant la limite quand τ' tend vers zéro, on peut transformer la dernière équation en :

$$\frac{d}{d\tau}T_\tau(y_3|y_1) = \int \left\{W(y_3|y_2)T_\tau(y_2|y_1) - W(y_2|y_3)T_\tau(y_3|y_1)\right\} dy_2$$

Cette forme différentielle de l'équation de Chapman-Kolmogorov s'appelle l'équation maîtresse. Sa forme plus commune est

$$\frac{\partial P(y,t)}{\partial t} = \int \left\{W(y|y')P(y',t) - W(y'|y)P(y,t)\right\} dy' \tag{6.6}$$

où $P(y,t)$ est la solution interprétée comme la probabilité de transition $T_{t-t_1}(y|y_1)$ (P^* auparavant) pour $t \geq t_1$ avec la condition initiale $P(y,t_1) = \delta(t-t_1)$.

Pour un processus markovien Y discret prenant les états notés n l'équation maîtresse (6.6) devient

$$\frac{dp_n(t)}{dt} = \sum \{W_{nn'}p_{n'}(t) - W_{n'n}p_n(t)\} \tag{6.7}$$

avec l'interprétation d'une équation du bilan de probabilité de transition depuis et vers chaque état n.

Dans le cas discret (qui peut être généralisé au cas continu), l'équation maîtresse (6.7) s'écrit à l'aide d'une matrice $\mathbb{W}$ définie comme $\mathbb{W}_{nn'} = W_{nn'}$ pour $n \neq n'$ et $\mathbb{W}_{nn} = -\sum_{n' \neq n} W_{n'n}$. Dans ce cas le système d'équations déterminé par l'équation (6.7) s'écrit de manière plus courte comme $\dot{p}(t) = \mathbb{W}p(t)$ où p est la transposée du vecteur $(p_1, \ldots, p_n)$. La solution formelle pour la valeur initiale $p(0)$ est la suivante

$$p(t) = e^{t\mathbb{W}}p(0) \tag{6.8}$$

La méthode via les valeurs et vecteurs propres de $\mathbb{W}$ ne peut être utilisée que si $\mathbb{W}$ est symétrique, sinon il n'est pas certain que toutes les solutions puissent être obtenues en superposant les solutions propres car les éléments de $p(t)$ ne sont pas nécessairement tout positifs dans ce cas.

La solution générale doit être fondée sur les propriétés suivantes de la matrice $\mathbb{W}$

$$\mathbb{W}_{nn'} \geq 0\text{‘} \quad (\text{quand} \quad n \neq n') \quad (a); \quad \sum_n \mathbb{W}_{nn'} = 0 \quad (\forall n') \quad (b) \tag{6.9}$$

Les propriétés (6.9) impliquent que $\mathbb{W}$ a un vecteur propre (gauche) $\psi = (1,1,1,\ldots)$ avec la valeur propre $\lambda_\psi = 0$. Il existe un vecteur propre (droit) ϕ avec la même valeur propre : $\mathbb{W}\phi = 0$ et ϕ est la solution temps-indépendante de l'équation maîtresse. Après normalisation ϕ est la densité de probabilité stationnaire (sous la condition que $\phi_n \geq 0$).

Par la suite on considérera seulement les processus markoviens Y discrets, c'est-à-dire les chaînes de Markov pour lesquelles l'équation maîtresse prend la forme (6.7).

6.2.1 Comportement asymptotique de l'équation maîtresse du processus à espace d'états fini et discret

Dans le cas de processus à espace d'états fini et discret, la propriété fondamentale de l'équation maîtresse est que quand le temps tend vers l'infini toutes les solutions convergent vers la ou les solutions stationnaires. La démonstration peut être fondée entre autre sur le théorème de Perron et Frobenius sur une échelle de temps discrétisée pour laquelle le processus est une chaîne de Markov (Van Kampen, 2007).

6.3 Processus de naissance et de mort (PNM)

Un cas particulier de processus stochastiques sont les processus de Markov à temps continu à valeurs dans un espace d'état fini tels que seules les transitions entre les états adjacents sont autorisées. Il s'agit de processus de naissance et de mort (PNM) où les probabilités de transitions par unité de temps sont $g_n = W(n+1|n)$ le taux de naissance à partir de l'état n (gain) et $r_n = W(n-1|n)$ le taux de mort à partir de l'état n (réduction). L'équation maîtresse correspondante est

$$\dot{p}_n = r_{n+1}p_{n+1} + g_{n-1}p_{n-1} - (r_n + g_n)p_n \tag{6.10}$$

On peut distinguer différents types de processus selon la forme de l'espace d'état : $n \in \mathbb{Z}$, $n \in \mathbb{N}$ ou $n = 0, 1, \ldots, N$ ou bien selon la forme de taux de transition r_n et g_n étant linéaires ou non linéaires (pas simultanément linéaires). Si le processus a des barrières réfléchissantes ou absorbantes l'équation maîtresse peut avoir une autre forme où les taux de transition sont modifiés sur ces bornes.

De l'équation (6.10) on déduit les moments de PNM.

6.3.1 Les moments de PNM

D'abord, pour abréger l'écriture posons pour une fonction $f(n)$ l'opérateur $\mathbb{E}$ définie comme

$$\mathbb{E}f(n) = f(n+1), \quad \mathbb{E}^{-1}f(n) = f(n-1) \tag{6.11}$$

A l'aide de cet opérateur l'équation (6.10) devient

$$\dot{p}_n = (\mathbb{E}-1)r_n p_n + (\mathbb{E}^{-1}-1)g_n p_n \tag{6.12}$$

Les équations des moments d'ordre k de PNM sont obtenues en multipliant l'équation (6.12) par n^k puis calculant la somme par rapport n. Sachant que la relation suivante est généralement satisfaite [1]

$$\sum_{n=0}^{N-1} g(n)\mathbb{E}f(n) = \sum_{n=1}^{N} f(n)\mathbb{E}^{-1}g(n) \tag{6.13}$$

la moyenne $\langle n \rangle$ est

$$\begin{aligned} \frac{d}{dt}\langle n \rangle &= \sum n(\mathbb{E}-1)r_n p_n + \sum n(\mathbb{E}^{-1}-1)g_n p_n \\ &= \sum r_n p_n(\mathbb{E}^{-1}-1)n + \sum g_n p_n(\mathbb{E}-1)n \\ &= -\langle r_n \rangle + \langle g_n \rangle \end{aligned} \tag{6.14}$$

Le deuxième $\langle n^2 \rangle$ puis k-ième $\langle n^k \rangle$ moments sont

$$\frac{d}{dt}\langle n^2 \rangle = 2\langle n(g_n - r_n)\rangle + \langle g_n + r_n \rangle \tag{6.15}$$

$$\frac{d}{dt}\left\langle n^k \right\rangle = \sum_{l=1}^{k} \binom{k}{l} \left\langle n^{k-l}(g_n + (-1)^l r_n) \right\rangle \tag{6.16}$$

Selon la forme de r_n et g_n le système établi par (6.16) peut être plus ou moins facile à résoudre.

6.3.2 Solution stationnaire de PNM

La solution stationnaire p_n^s de l'équation maîtresse décrite par (6.12) correspond à $\dot{p}_n = 0$

$$\begin{aligned} 0 &= (\mathbb{E}-1)r_n p_n^s + (\mathbb{E}^{-1}-1)g_n p_n^s \\ &= (\mathbb{E}-1)\left\{r_n p_n^s - \mathbb{E}^{-1} g_n p_n^s\right\} \end{aligned}$$

Le terme dans l'accolade ne dépend pas du temps, on pose :

$$r_n p_n^s - \mathbb{E}^{-1} g_n p_n^s = -J \tag{6.17}$$

1. Les limites de deux sommes différent mais en pratique il s'agit souvent de termes nuls de la somme.

où $-J$ est le bilan de probabilité du passage de n à $n-1$. Dans le cas où $n = 0, 1, 2, \ldots, N$, les taux de transition vers ou depuis des états en dehors de cet ensemble doivent être nuls. Plus particulièrement $r_0 = g_{-1} = 0$ ainsi que $r_{N+1} = g_N = 0$. Cela garantit que l'équation (6.10) soit correcte. Pour $n = 0$ (6.10) est $\dot{p}_0 = r_1 p_1 - g_0 p_0$ qui dans l'état stationnaire devient $0 = r_1 p_1 - g_0 p_0$. Cette équation à l'aide d'opérateur $\mathbb{E}^{-1}$ s'écrit comme $0 = r_1 p_1 - \mathbb{E}^{-1} g_1 p_1$ impliquant que $J = 0$. La substitution de $J = 0$ dans l'équation (6.17) donne que $r_n p_n^s = g_{n-1} p_{n-1}^s$ et en appliquant cette relation systématiquement on obtient

$$p_n^s = \frac{g_{n-1} g_{n-2} \cdots g_1 g_0}{r_n r_{n-1} \cdots r_2 r_1} p_0^s \tag{6.18}$$

Tous les p^s sont donnés en terme de p_0^s, qui est fixé par la condition de la normalisation comme

$$\frac{1}{p_0^s} = 1 + \sum_{n=1}^{N} \frac{g_0 g_1 \cdots g_{n-1}}{r_1 r_2 \cdots r_n} \tag{6.19}$$

Van Kampen (2007) constate que dans le cas $n \in \mathbb{N}$ le résultat est identique au précédant (6.18). Pour $-\infty < n < \infty$ le résultat est le même seulement pour $n \geq 0$. Pour $n \leq 0$, la solution stationnaire est déduite de $r_n p_n^s = g_{n-1} p_{n-1}^s$ comme

$$p_n^s = \frac{r_{n+1} r_{n+2} \cdots r_{-1} r_0}{g_n g_{n+1} \cdots g_{-2} g_{-1}} p_0^s \tag{6.20}$$

Mais on ne peut plus démontrer que $J = 0$. La solution stationnaire n'est plus unique mais dépend de J.

6.3.3 Solution quasi-stationnaire de PNM

Nåsell (1996) discute la densité quasi-stationnaire dans un cas particulière du PNM - un modèle épidémique SIS (décrit dans la section (6.5) plus loin dans cette thèse). Son analyse se traduit facilement pour un PNM plus général. Pour $n = 0, \ldots, N$ et $g_0 = 0, r_0 = 0$ et $g_N = 0$ le PNM défini ci-dessus par (6.10) a un état absorbant en 0 et ainsi la densité stationnaire (6.18) centrée sur 1 pour $n = 0$ (toute probabilités p_n sont nulles pour $n = 1, \ldots, N$ sauf $p_0^s = 1$). Le modèle prédit l'extinction du processus indépendamment des paramètres définissant les taux de transition r_n et g_n. Cependant le temps d'extinction dépend de ces paramètres et il peut être infiniment long. Durant cette longue période la loi du processus Y (PNM) est proche de

la loi conditionnelle à la non-extinction, une loi quasi-stationnaire (Nåsell, 1996). De même la manière que dans la section (6.2) l'équation maîtresse (6.10) du processus Y peut être écrite sous la forme matricielle $\dot{p}(t) = \mathbb{W}p(t)$ avec p étant un vecteur colonne de coordonnées p_i $(i = 0, \ldots, N)$, la matrice $\mathbb{W}$ étant donnée par

$$\mathbb{W} = \begin{pmatrix} -(r_0 + g_0) & r_1 & 0 & \ldots & 0 \\ g_0 & -(r_1 + g_1) & r_2 & \ldots & 0 \\ 0 & g_1 & -(r_2 + g_2) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \ldots & -(r_N + g_N) \end{pmatrix} \tag{6.21}$$

La probabilité stationnaire p est une Dirac centrée sur 0, la seule barrière absorbante de Y. Pourtant on peut étudier le processus Y conditionnel à la non-extinction avec les probabilités instantanés $q_n(t)$ $(n = 1, \ldots, N)$. Soit $p_Q(t)$ un vecteur composé des coordonnées de $p(t)$ à l'exception de première coordonnée $p_1(t)$. Alors le vecteur $q(t)$ de $q_n(t)$ s'obtient comme

$$q(t) = p_Q(t)/(1 - p_0(t)) \tag{6.22}$$

et en dérivant cette relation et sachant que $\dot{p}_0 = r_1 p_1$ on obtient que

$$\dot{q} = \frac{\dot{p}_Q}{1 - p_0} + r_1 q_1 \frac{p_Q}{1 - p_0} \tag{6.23}$$

ce qui donne un système d'équations

$$\dot{q} = q\mathbb{W}_Q + r_1 q_1 q \tag{6.24}$$

où $\mathbb{W}_Q$ s'obtient an supprimant la première ligne et la première colonne de la matrice $\mathbb{W}$ (6.21). La densité quasi-stationnaire est donc la solution de l'équation (6.24) quand $\dot{q} = 0$, cela dit $\mathbb{W}_Q q = -r_1 q_1 q$. Il s'agit d'un système d'équations où chacune est de la forme

$$g_{n-1} q_{n-1} - (r_n + g_n) q_n + r_{n+1} q_{n+1} = -r_1 q_1 q_n \tag{6.25}$$

avec $q_0 = q_{n+1} = 0$. On constate immédiatement que l'application des opérateurs $\mathbb{E}$ et $\mathbb{E}^{-1}$ (6.11) ne permet pas de déterminer la solution stationnaire car $(\mathbb{E} - 1)\{r_n q_n^s - \mathbb{E}^{-1} g_n q_n^s\} = -r_1 q_1 q_n$. En divisant (6.25) par r_1 et en changeant l'ordre on écrit

$$\frac{r_{n+1}}{r_1} q_{n+1} - \frac{g_n}{r_1} q_n - \frac{r_n}{r_1} q_n + \frac{g_{n-1}}{r_1} q_{n-1} = -q_1 q_n \tag{6.26}$$

Posons

$$f_n = \frac{r_n}{r_1} q_n - \frac{g_{n-1}}{r_1} q_{n-1} \tag{6.27}$$

L'équation (6.26) devient donc en termes de f_n

$$f_{n+1} - f_n = -q_1 q_n \tag{6.28}$$

$f_1 = q_1$ et $f_{N+1} = 0$. En exprimant f_n en fonction de q_n on obtient

$$f_n = q_1 (1 - \sum_{i=1}^{n-1} q_i) \tag{6.29}$$

En remplaçant cette équation dans (6.27) on obtient

$$q_1 (1 - \sum_{i=1}^{n-1} q_i) = \frac{r_n}{r_1} q_n - \frac{g_{n-1}}{r_1} q_{n-1} \tag{6.30}$$

ce qui permet établir une forme de récurrence de q_n

$$q_n = \frac{r_1 q_1}{r_n} (1 - \sum_{i=1}^{n-1} q_i) + \frac{g_{n-1}}{r_n} q_{n-1} \tag{6.31}$$

Pour déterminer la solution stationnaire q^s il faudra donc employer une procédure itérative. Nåsell (1996) établit cette équation dans un cas particulier du processus SIS. Il discute aussi deux formes approximatives, évoquées antérieurement par Kryscio and Lefévre (1989), dans la première le modèle étudié suppose qu'il y ait au moins un individu infecté, dans la deuxième on supprime l'origine ($n = 0$). Ces formes approximatives sont néanmoins toujours données par une récurrence.

6.3.4 Solution générale de PNM

Dans le cas linéaire pour $n \in \mathbb{Z}$ il est possible de résoudre l'équation (6.10) explicitement par exemple à l'aide d'une fonction génératrice de probabilité. Si n est borné, selon la forme des taux de transitions r_n et g_n l'équation (6.10) n'a plus de sens aux bornes donc il peut être nécessaire d'ajouter des équations supplémentaires pour décrire l'évolution de probabilité sur chaque borne (Cf. van Kampen pour voir comment traiter ces cas). Cependant si le taux de naissance à partir de la borne supérieure ainsi que le taux de mort à partir de la borne inférieure sont nuls, on peut supposer que l'équation (6.10) est valide pour tout $n \in \mathbb{Z}$ et chercher sa

solution en prenant les valeurs initiales $p_n(0) = 0$ pour tous les n au-dessous et au-dessus la borne inférieure et supérieure respectivement. Pour ces valeurs de n les probabilités $p_n(t)$ restent nulles pour $t > 0$ car les transitions vers ces états sont exclues.

Dans le cas des taux de transitions non linéaires les solutions, autres que les solutions stationnaires, ne sont généralement pas explicitées. Néanmoins, Picard (1965) démontre que la solution de (6.10) d'un processus de naissance et de mort est de la forme (6.8) où $p(t)$ (ici un vecteur $p_n(t)$) est bien une fonction de probabilité. On note qu'il montre que la matrice $\mathbb{W}$ du PNM a des valeurs propres distinctes et non-positives une étant égale à zéro, la matrice $\mathbb{W}$ étant diagonalisable. On peut donc trouver la solution de $\dot{p}(t) = \mathbb{W}p(t)$ on superposant les solutions propres $p(t) = \sum_{i=0}^{N} c_i e^{\lambda_i t} v_i$ où les constantes c_i sont déterminées pour la densité initiale $p(0)$ comme décrit pour la matrice symétrique dans la section (6.2). Nåsell (1996) remarque que en rangeant les valeurs propres λ_i dans l'ordre décroissant ($0 = \lambda_{(0)} > \lambda_{(1)} > \lambda_{(2)} > \ldots$) dans certains cas où $|\lambda_{(1)}| \ll |\lambda_{(2)}|$ la solution $p(t)$ peut être approchée par

$$p(t) \approx c_{(0)} v_{(0)} + c_{(1)} v_{(1)} = (0 \quad q_1 \cdots q_N)^T = (0 \quad q)^T$$

Cependant, le calcul des valeurs propres de $\mathbb{W}$, ou bien de $e^{t\mathbb{W}}$, même s'il est théoriquement possible, n'est pas une tâche facile. Comme dans les applications ultérieures les taux de transition sont non linéaires, par la suite dans la section (6.3.5) une solution approchée de (6.10) sera présentée pour ce contexte-là.

6.3.5 Solution approchée de PNM - taux de transition non linéaire

Dans un première temps pour trouver la solution approchée de l'équation (6.12) on va considérer que n est continu, $r_n = r(n)$ et $g_n = g(n)$ sont les fonctions continues et différentiables variant peu entre n et $n + 1$, et p_n est aussi une fonction continue et différentiable. La notation $\mathbb{E}$ (idem pour $\mathbb{E}^{-1}$) utilisée ci-dessus pour $n \in \mathbb{Z}$ est un opérateur dont la forme différentielle développée est

$$\mathbb{E}_n = 1 + \frac{\partial}{\partial n} + \frac{1}{2!}\frac{\partial}{\partial n^2} + \ldots, \quad \mathbb{E}_n^{-1} = 1 - \frac{\partial}{\partial n} + \frac{1}{2!}\frac{\partial}{\partial n^2} - \ldots \tag{6.32}$$

développement de Taylor d'un fonction de $n + 1$ au voisinage de n. Ces nouveaux opérateurs remplacent $\mathbb{E}$ et $\mathbb{E}^{-1}$ dans l'équation (6.12). En négligeant les dérivées d'ordre supérieur à 2

l'équation (6.12) est approximée par

$$\frac{d}{dt}P_n(t) = \frac{\partial}{\partial n}\left\{r(n) - g(n)\right\}P_n + \frac{1}{2}\frac{\partial^2}{\partial n^2}\left\{r(n) + g(n)\right\}P_n \tag{6.33}$$

qui est une équation de Fokker-Planck. En général cette équation n'a pas de solution explicite donc le recours aux méthodes d'approximation sera nécessaire.

Au passage on remarque que pour $n = 0, \ldots, N$ la solution générale de l'équation maîtresse (6.10) du PNM peut être approchée à l'aide d'équation aux différences pour un petit incrément de temps Δt :

$$p_n(t + \Delta t) = g_{n-1}p_{n-1}(t)\Delta t + (1 - g_n\Delta t - r_n\Delta t)p_n + r_{n+1}p_{n+1}(t) \tag{6.34}$$

qui sous forme matricielle avec la matrice $\mathbb{W}$ définie dans (6.21) est

$$p(t + \Delta t) = (1 - \mathbb{W}\Delta t)p(t) \tag{6.35}$$

et ainsi en partant de l'instant t_0 l'évolution de la probabilité du processus après le temps $m\Delta t$ pour $m \in \mathbb{N}$ on obtient en calculant la $m^{\text{ième}}$ puissance de $(1 - \mathbb{W}\Delta t)$ comme

$$p(t_0 + m\Delta t) = (1 - \mathbb{W}\Delta t)^m p(t_0) \tag{6.36}$$

6.4 Développement de l'équation maîtresse pour le PNM

Comme alternative van Kampen a proposé un approche systématique du développement dont l'idée essentielle est d'exprimer la valeur du processus n comme la somme d'un phénomène déterministe et d'une fluctuation aléatoire d'ordre qui dépend de la taille du système. Plus précisément, la taille du système noté Ω est un paramètre qui gouverne l'importance relative des fluctuations. En principe dans la collection de N particules, les fluctuations sont d'ordre $N^{1/2}$. Dans la plupart des cas Ω est le nombre d'éléments du système. Plus exactement il faut sélectionner un paramètre présent dans l'équation maîtresse ainsi dans les taux de transitions, qui gouverne l'ampleur des fluctuations et donc l'amplitude des sauts.

Formellement, s'il existe un paramètre Ω, les probabilités des transitions r_n et g_n peuvent être exprimées comme

$$r_n = f(\Omega)\left[\rho_0(\frac{n}{\Omega}) + \Omega^{-1}\rho_1(\frac{n}{\Omega}) + \Omega^{-2}\rho_2(\frac{n}{\Omega}) + \ldots\right] \tag{6.37}$$

$$g_n = f(\Omega)\left[\gamma_0(\frac{n}{\Omega}) + \Omega^{-1}\gamma_1(\frac{n}{\Omega}) + \Omega^{-2}\gamma_2(\frac{n}{\Omega}) + \ldots\right] \tag{6.38}$$

où $\rho_0, \rho_1, \ldots$ et $\gamma_0, \gamma_1, \ldots$ sont des fonctions de n/Ω et qui sinon ne dépend pas de Ω.

Si r_n et g_n ne peuvent pas être exprimés sous cette forme-là, la méthode systématique de développement n'est pas applicable.

Supposons que le paramètre Ω existe et que la probabilité p_n atteint un pic aigü pour une valeur macroscopique $\Omega\phi(t)$ (la solution déterministe) et a une amplitude d'ordre $\Omega^{1/2}$. On pose alors que

$$n = \Omega\phi(t) + \Omega^{1/2}\xi \tag{6.39}$$

où $\phi(t)$ est une fonction déterministe et ξ est une nouvelle variable aléatoire qui va décrire les fluctuations de n. La probabilité p_n est maintenant écrite comme une fonction de ξ

$$p_n(t) = \Pi(\xi, t) \tag{6.40}$$

L'étape suivante consiste à transformer l'équation (6.12) décrivant l'évolution de probabilité p_n d'observer n éléments à l'instant t en probabilité Π d'observer ξ à l'instant t, en remplaçant au même temps n par (6.39). Il faut d'abord transformer opérateurs $\mathbb{E}$ et $\mathbb{E}^{-1}$ (6.11) appliqués à une fonction de n en des opérateurs agissant sur une fonction de ξ.

L'opérateur $\mathbb{E}$ (6.11) appliqué à une fonction transforme n en $n+1$ et alors via la relation $n = \Omega\phi(t) + \Omega^{1/2}\xi$, ξ est transformé en $\xi + \Omega^{-1/2}$. Par analogie, l'opérateur $\mathbb{E}^{-1}$ appliqué à une fonction transforme n en $n+1$ et ξ en $\xi - \Omega^{-1/2}$. En effet $n+1 = \Omega\phi(t) + \Omega^{1/2}\xi + 1 = \Omega\phi(t) + \Omega^{1/2}(\xi + \Omega^{-1/2})$ et $n - 1 = \Omega\phi(t) + \Omega^{1/2}(\xi - \Omega^{-1/2})$. En faisant maintenant le développement de $\mathbb{E}f(\xi)$ et $\mathbb{E}^{-1}f(\xi)$ au voisinage de ξ on obtient

$$\mathbb{E}f(\xi) = f(\xi + \Omega^{-1/2}) = f(\xi) + \Omega^{-1/2}\frac{\partial f}{\partial \xi} + \frac{1}{2}\Omega^{-1}\frac{\partial^2 f}{\partial \xi^2} + \ldots$$

$$\mathbb{E}^{-1}f(\xi) = f(\xi - \Omega^{-1/2}) = f(\xi) - \Omega^{-1/2}\frac{\partial f}{\partial \xi} + \frac{1}{2}\Omega^{-1}\frac{\partial^2 f}{\partial \xi^2} - \ldots$$

De cette manière on obtient les opérateurs suivants

$$\mathbb{E}_\xi = 1 + \Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} + \ldots, \quad \mathbb{E}_\xi^{-1} = 1 - \Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} - \ldots \tag{6.41}$$

La dérivée $\dot{p}_n$ par rapport à t dans (6.12) calculée à partir de l'équation (6.40) est

$$\frac{d}{dt}p_n(t) = \frac{\partial \Pi}{\partial \xi}\frac{d\xi}{dt} + \frac{\partial \Pi}{\partial t}$$

En prenant dans ce calcul n constant (fixé) $d\xi/dt = -\Omega^{1/2}d\phi/dt$ d'où on obtient que

$$\dot{p_n} = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} \tag{6.42}$$

Dans la partie droite de l'équation (6.12) n est remplacé par l'expression (6.39), p_n par Π (6.40) et les operateurs $\mathbb{E}, \mathbb{E}^{-1}$ (6.11) par $\mathbb{E}_\xi$ et $\mathbb{E}_\xi^{-1}$ (6.41) définis ci-dessus.

$$\frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} = (\mathbb{E}_\xi - 1)r(\Omega\phi + \Omega^{1/2})\Pi(\xi,t) + (\mathbb{E}_\xi^{-1} - 1)g(\Omega\phi + \Omega^{1/2})\Pi(\xi,t) \tag{6.43}$$

L'identification des termes avec $\Omega^{1/2}$ et Ω^0 dans la dernière équation permet de trouver la solution de $\phi(t)$ et $\Pi(\xi,t)$ respectivement. En effet dans le développement systématique l'équation (6.43) est écrite d'abord à l'aide de conditions (6.37) et (6.38)

$$\begin{aligned} \frac{\partial \Pi}{\partial t} &- \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} \\ &= f(\Omega)(\mathbb{E}_\xi - 1)\left[\rho_0(\phi + \Omega^{-1/2}\xi) + \Omega^{-1}\rho_1(\phi + \Omega^{-1/2}\xi) + \ldots\right]\Pi(\xi,t) \\ &+ f(\Omega)(\mathbb{E}_\xi^{-1} - 1)\left[\gamma_0(\phi + \Omega^{-1/2}\xi) + \Omega^{-1}\gamma_1(\phi + \Omega^{-1/2}\xi) + \ldots\right]\Pi(\xi,t) \end{aligned} \tag{6.44}$$

Ensuite chaque fonction $\rho_0, \rho_1, \ldots$ et $\gamma_0, \gamma_1, \ldots$ d'argument $\phi + \Omega^{-1/2}\xi$ est approchée par le développement de Taylor au voisinage de ϕ. La démonstration se trouve en annexe (IV).

Finalement le développement systématique de l'équation maîtresse d'un PNM est

$$\begin{aligned} \frac{\partial \Pi}{\partial \tau} &- \Omega^{1/2}\frac{d\phi}{d\tau}\frac{\partial \Pi}{\partial \xi} \\ &= \Omega^{1/2}(\rho_0(\phi) - \gamma_0(\phi))\frac{\partial \Pi}{\partial \xi} + \frac{1}{2!}\Omega^0(\rho_0(\phi) + \gamma_0(\phi))\frac{\partial^2 \Pi}{\partial \xi^2} + \frac{1}{3!}\Omega^{-1/2}(\rho_0(\phi) - \gamma_0(\phi))\frac{\partial^3 \Pi}{\partial \xi^3} \\ &+ \Omega^0(\rho_0'(\phi) - \gamma_0'(\phi))\frac{\partial}{\partial \xi}\xi\Pi + \frac{1}{2!}\Omega^{-1/2}(\rho_0'(\phi) + \gamma_0'(\phi))\frac{\partial^2}{\partial \xi^2}\xi\Pi + \Omega^{-1/2}(\rho_0''(\phi) - \gamma_0''(\phi))\frac{\partial}{\partial \xi}\xi^2\Pi \\ &+ \Omega^{-1/2}(\rho_1(\phi) - \gamma_1(\phi))\frac{\partial \Pi}{\partial \xi} + o(\Omega^{-1}) \end{aligned} \tag{6.45}$$

En prenant ϕ comme une solution de

$$\frac{d\phi}{d\tau} = \gamma_0(\phi) - \rho_0(\phi) \tag{6.46}$$

avec une condition initiale $\phi(0) = y_0/\Omega$ correspondant à $p_n(0) = \delta(Y - y_0)$, les deux termes d'ordre $\Omega^{1/2}$ s'annulent. La solution stationnaire ϕ s'obtient comme la racine de $\rho_0(\phi) - \gamma_0(\phi) = 0$ et il est possible que cette solution ne soit pas unique. En ignorant les autres termes, les termes d'ordre Ω^0 donnent

$$\frac{\partial \Pi(\xi,\tau)}{\partial \tau} = (\rho_0'(\phi) - \gamma_0'(\phi))\frac{\partial}{\partial \xi}(\xi\Pi) + \frac{1}{2!}(\rho_0(\phi) + \gamma_0(\phi))\frac{\partial^2 \Pi}{\partial \xi^2} \tag{6.47}$$

qui est une équation de Fokker-Planck très particulière car sa solution stationnaire est gaussienne si $(\rho_0'(\phi) - \gamma_0'(\phi)) > 0$. Les moments de Π sont obtenus en multipliant l'équation (6.47) par ξ est ξ^2

$$\frac{d\langle\xi\rangle}{d\tau} = -(\rho_0'(\phi) - \gamma_0'(\phi))\langle\xi\rangle \tag{6.48}$$

$$\frac{d\langle\xi^2\rangle}{d\tau} = 2(\rho_0'(\phi) - \gamma_0'(\phi))\langle\xi^2\rangle + (\rho_0(\phi) + \gamma_0(\phi)) \tag{6.49}$$

La variance $\langle\langle\xi^2\rangle\rangle$ satisfait la même équation que le deuxième moment $\langle\xi^2\rangle$. En effet la condition initiale $p_n(0)$ est une mesure de Dirac les fluctuations initiales sont alors nulles donc

$$\langle\xi\rangle_0 = \langle\xi^2\rangle_0 = \langle\langle\xi^2\rangle\rangle_0 = 0$$

Pour obtenir la solution de l'équation maîtresse il faut d'abord trouver la solution $\phi(\tau|y_0)$ de l'équation (6.46) puis les solutions de (6.48) et (6.49) avec $\phi(\tau|y_0)$. La solution de l'équation maîtresse du processus de naissance et de mort Y est ainsi approchée par une gaussienne avec la moyenne $\langle Y\rangle_\tau = \Omega\phi(\tau|y_0) + \Omega^{1/2}\langle\xi\rangle_\tau$ et la variance $\langle\langle Y^2\rangle\rangle_\tau = \Omega\langle\langle\xi^2\rangle\rangle_\tau$.

6.5 Modèle SIS avec la population de taille constante.

La méthode de développement de l'équation maîtresse est appliquée à un modèle de transmission d'un maladie dans une population de taille constante où il y a seulement deux types de sujets : susceptibles et infectés.

6.5.1 Modèle SIS déterministe

Prenons d'abord le modèle SIS déterministe avec taille constante de la population tel que présenté par Allen and Burgin (2000) :

$$S(t+\Delta t) = S(t)[1-\lambda(t)\Delta t] + (\beta\Delta t + \gamma\Delta t)I(t), \tag{6.50}$$

$$I(t+\Delta t) = I(t)[1-\beta\Delta t - \gamma\Delta t] + \lambda(t)\Delta t I(t), \tag{6.51}$$

où S, décrit le processus du nombre de sujets susceptibles et I le processus de nombre du sujets infectés. Δt est un intervalle de temps fixe, $S(0) > 0, I(0) > 0$ et $S(0)+I(0) = S(t)+I(t) = N$. λ est la force d'infection donnée par $\lambda = \frac{\alpha}{N(t)}I(t)$ où α est un taux de contact, $\beta\Delta t$ est un taux de naissance ou de décès. Finalement $\gamma\Delta t$ est le taux de guérison (la proportion de sujets guérissant dans la période Δt).

Ce modèle peut être écrit comme le système d'équations différentielles

$$\frac{dS}{dt} = -\lambda(t)S(t) + (\beta+\gamma)I(t) \tag{6.52}$$

$$\frac{dI}{dt} = \lambda(t)S(t) - (\beta+\gamma)I(t) \tag{6.53}$$

Le taux de naissance et le taux de guérison apparaissent uniquement dans la somme désormais on note $\beta := \beta+\gamma$. Puisque la somme des deux processus est constante $S(t)+I(t) = N$ nous pouvons décrire le modèle avec une seule équation

$$\frac{dI}{dt} = \frac{\alpha}{N}(N-I)I - \beta I \tag{6.54}$$

dont la solution avec la condition initiale $I(0) = I_0$ est

$$I(t) = \frac{\alpha-\beta}{\frac{\alpha}{N} + \frac{\alpha-\beta-\frac{\alpha}{N}I_0}{I_0}e^{-(\alpha-\beta)t}} \tag{6.55}$$

6.5.2 Modèle SIS stochastique

Le modèle SIS stochastique avec une taille de population constante présenté ici est tel que abordé par Allen and Burgin (2000). C'est un modèle SIS stochastique à temps discret formulé comme une chaîne de Markov. $p_i(t) = Pr(I_t = i)$ est la probabilité que la variable

aléatoire I_t prenne la valeur i et les probabilités de transitions sont les suivantes

$$Pr(I_{t+\Delta t} = i+1 \quad | \quad I_t = i) = \frac{\alpha}{N} i(N-i)\Delta t \tag{6.56}$$

$$Pr(I_{t+\Delta t} = i-1 \quad | \quad I_t = i) = \beta i \Delta t \tag{6.57}$$

Le probabilités $p_i(t)$ satisfont l'équation aux différences

$$\begin{aligned} & p_i(t+\Delta t) = \\ & = \frac{\alpha}{N}(i-1)\left(N-(i-1)\right)\Delta t p_{i-1}(t) + \beta(i+1)\Delta t p_{i+1}(t) \\ & + \left(1 - \frac{\alpha}{N} i(N-i)\Delta t - \beta i \Delta t\right) p_i(t), \end{aligned} \tag{6.58}$$

$$p_0(t+\Delta t) = p_0(t) \tag{6.59}$$

Quand Δt tends vers zéro ($\Delta t \to 0$) la dérivée par rapport au temps de p_i est $\dot{p}_i(t) = \lim_{\delta t \to 0} \frac{p_{in}(t+\Delta t) - p_{in}(t)}{\Delta t}$ alors l'équation aux différences peut s'écrire comme l'équation différentielle

$$\begin{aligned} \dot{p}_i(t) & = \frac{\alpha}{N}(i-1)\left(N-(i-1)\right) p_{i-1}(t) + \beta(i+1) p_{i+1}(t) \\ & - \left(\frac{\alpha}{N} i(N-i) + \beta i\right) p_i(t), \end{aligned} \tag{6.60}$$

A l'aide des opérateurs $\mathbb{E}_i$ et $\mathbb{E}_i^{-1}$ définis auparavant l'équation (6.60) s'écrit comme

$$\dot{p}_i(t) = (\mathbb{E}_i^{-1} - 1)\frac{\alpha}{N} i(N-i) p_i(t) + (\mathbb{E}_i - 1)\beta i p_i(t), \tag{6.61}$$

Comme nous pouvons le constater, la probabilité de transition dans l'équation (6.56) est non linéaire. Pour trouver la solution de $p_i(t)$ on utilise la méthode de développement de l'équation maîtresse. Posons dans l'équation (6.61) la taille du système $\Omega = N$ et $i = \Omega\phi + \Omega^{1/2}\xi$ et $p_i(t) = \Pi(\xi, t)$ et , puis remplaçons les opérateurs $\mathbb{E}_i$ et $\mathbb{E}_i^{-1}$ par $\mathbb{E}_\xi$ et $\mathbb{E}_\xi^{-1}$. Avec la relation $\dot{p}_i(t) = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi}$ on obtient

$$\begin{aligned} & \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} \\ & = (\mathbb{E}_\xi^{-1} - 1)\frac{\alpha}{\Omega}(\Omega\phi + \Omega^{1/2}\xi)(\Omega - \Omega\phi + \Omega^{1/2}\xi)\Pi(\xi, t) \\ & + (\mathbb{E}_\xi - 1)\beta(\Omega\phi + \Omega^{1/2}\xi)\Pi(\xi, t) \end{aligned} \tag{6.62}$$

L'étape suivante consiste à utiliser la forme différentielle des opérateurs $\mathbb{E}_\xi$ et $\mathbb{E}_\xi^{-1}$, puis à identifier les termes avec $\Omega^{1/2}$ pour trouver la solution ϕ et les termes avec Ω^0 pour trouver la solution $\Pi(\xi,t)$. On obtient donc la solution ϕ avec

$$\frac{d\phi}{dt} = -\alpha\phi^2(t) - (\beta-\alpha)\phi(t) \tag{6.63}$$

Pour la solution $\Pi(\xi,t)$ on obtient de l'équation suivante :

$$\begin{aligned} \frac{\partial\Pi}{\partial t} &= -(-2\alpha\phi(t) - (\beta-\alpha))\frac{\partial}{\partial\xi}(\xi\Pi) \\ &+ \frac{1}{2}(-\alpha\phi^2(t) + (\beta+\alpha)\phi(t))\frac{\partial^2\Pi}{\partial\xi^2} \end{aligned} \tag{6.64}$$

L'équation (6.64) a la forme de l'équation de Fokker-Planck linéaire (Cf. van Kampen(1997/2007)), dans l'état stationnaire la solution $\Pi(\xi,t)$ est donc gaussienne si $(\alpha > \beta)$. Pour déterminer la forme exacte de cette gaussienne il faut encore trouver sa moyenne et sa variance. La moyenne est la solution de

$$\frac{d\langle\xi\rangle}{dt} = (-2*\alpha\phi(t) - (\beta-\alpha))\langle\xi\rangle \tag{6.65}$$

mais avec la condition initiale $\langle\xi\rangle_0 = 0$ la moyenne est égale $\langle\xi\rangle = 0$
La variance de ξ, notée $\langle\xi^2\rangle$ car elle est égale au moment d'ordre 2, est la solution de l'équation inhomogéne du 1$^{\text{er}}$ ordre :

$$\frac{d\langle\xi^2\rangle}{dt} = (-4\alpha\phi(t) - 2(\beta-\alpha))\langle\xi^2\rangle + (-\alpha\phi^2(t) + (\beta+\alpha)\phi(t)) \tag{6.66}$$

Ainsi ξ suit une loi gaussienne $\xi \sim \mathcal{N}(0, \langle\xi^2\rangle)$ et $i \sim \mathcal{N}(\Omega\phi, \Omega\langle\xi^2\rangle)$. Puisque $i = \Omega\phi + \Omega^{1/2}\xi$ où ϕ est déterministe et ξ est aléatoire l'espérance de i est égale $\Omega\phi$ et sa variance $\Omega\langle\xi^2\rangle$.

6.6 Modèle pour décrire les données

Soit X_t le processus de nombre de sujets infectés, α un taux de contact, $\beta\Delta t$ un taux de naissance ou de décès combiné avec un taux de guérison.

6.6.1 Modèle déterminé par le développement de l'équation maîtresse (M)

Soit $x = (x_j)_{j=1...J}$ sont les nombres de sujets infectés observés aux instantes de temps $(t_1, \ldots, t_J)$

$$x_j = \Omega\phi_\theta(t_j) + \Omega^{1/2}\xi_j \tag{6.67}$$

$$\xi_j \sim \mathcal{N}\left(0, \left\langle \xi^2 \right\rangle_\theta (t_j)\right) \tag{6.68}$$

où $\theta = (\alpha, \beta)$ et un vecteur de paramètres, ϕ_θ et $\left\langle \xi^2 \right\rangle_\theta$ sont les fonctions non linéaires définies comme des solutions des équations différentielles suivantes

$$\frac{d\phi_\theta}{dt} = -\alpha\phi_\theta^2(t) - (\beta - \alpha)\phi_\theta(t) \tag{6.69}$$

$$\phi_\theta(t_0) = \phi_0 \tag{6.70}$$

$$\begin{aligned} \frac{d\left\langle \xi^2 \right\rangle_\theta}{dt} &= (-4\alpha\phi_\theta(t) - 2(\beta - \alpha))\left\langle \xi^2 \right\rangle_\theta \\ &+ (-\alpha\phi_\theta^2(t) + (\beta + \alpha)\phi_\theta(t)) \end{aligned} \tag{6.71}$$

$$\left\langle \xi^2 \right\rangle_\theta (t_0) = 0 \tag{6.72}$$

Supposons que :

$$x_j \sim \mathcal{N}\left(\Omega\phi_\theta(t_j), \Omega\left\langle \xi^2 \right\rangle_\theta (\phi_\theta(t_j), t_j)\right) \tag{6.73}$$

Alors la contribution de x_j observé à l'instant t_j à la vraisemblance est

$$V_\theta(x_j) = \frac{1}{\sqrt{2\pi\Omega\left\langle \xi^2 \right\rangle_\theta (\phi_\theta(t_j), t_j)}} \exp\left\{ \frac{(x_j - \Omega\phi_\theta(t_j))^2}{2\Omega\left\langle \xi^2 \right\rangle_\theta (\phi_\theta(t_j), t_j)} \right\} \tag{6.74}$$

où $\phi_\theta(t_j)$ est la solution de l'équation (6.69) avec la condition initiale $\phi_0 = \frac{x_{j-1}}{\Omega}$ (la condition initiale est construite à partir de l'observation précédente)

Pour alléger l'écriture je note $\left\langle \xi^2 \right\rangle = v$ où v_i signifie $v_j = v(\phi_j, t_j)$ et $\phi_j = \phi(t_j)$ étant les solutions obtenues du développement de l'équation maîtresse. Les paramètres α, β et i_0 sont notés $\theta = (\theta_1, \theta_2, \theta_3)$. Le modèle (M) suppose que la loi des effectifs x_j observés aux instants t_i pour $j = 1, \ldots, J$ est $X_j \sim \mathcal{N}(\Omega\phi_j; \Omega v_j)$

La contribution V_j de la $j^{\text{ème}}$ observation à la vraisemblance est alors

$$V_j = \frac{1}{\sqrt{2\pi\Omega v_j}} \exp\left\{-\frac{(x_j - \Omega\phi_j)^2}{2\Omega v_j}\right\} \tag{6.75}$$

$$\ln V_j = -\frac{1}{2}\ln(2\pi\Omega) - \frac{1}{2}\ln v_j - \frac{(x_j - \Omega\phi_j)^2}{2\Omega v_j}$$

et donc la logvraisemblance l a l'expression

$$l = \sum \ln V_j = -\frac{J}{2}\ln(2\pi\Omega) - \frac{1}{2}\sum_{i=1}^{J}\ln v_j - \frac{1}{2\Omega}\sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)^2}{v_j}$$

La dérivée de la logvraisemblance où $k = 1, 2, 3$

$$\frac{dl}{d\theta_k} = -\frac{1}{2}\sum_{i=1}^{J}\frac{1}{v_j}\frac{dv}{d\theta_k} + \sum_{i=1}^{J}\frac{(x_j - \Omega\phi_j)\frac{d\phi}{d\theta_k}}{v_j} + \frac{1}{2\Omega}\sum_{i=1}^{J}\frac{(x_j - \Omega\phi_j)^2\frac{dv}{d\theta_k}}{v_j^2}$$

La dérivée seconde de la logvraisemblance où $k, l = 1, 2, 3$

$$\begin{aligned}
\frac{d^2l}{d\theta_l d\theta_k} &= -\frac{1}{2}\sum_{i=1}^{J}\frac{\frac{d^2v}{d\theta_l d\theta_k}}{v_j} + \frac{1}{2}\sum_{j=1}^{J}\frac{\frac{dv}{d\theta_l}\frac{dv}{d\theta_k}}{v_j^2} - \Omega\sum_{j=1}^{J}\frac{\frac{d\phi}{d\theta_l}\frac{d\phi}{d\theta_k}}{v_j} \\
&+ \sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)\frac{d^2\phi}{d\theta_l d\theta_k}}{v_j} - \sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)\frac{d\phi}{d\theta_l}\frac{dv}{d\theta_k}}{v_j^2} - \sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)\frac{d\phi}{d\theta_l}\frac{dv}{d\theta_k}}{v_j^4} \\
&+ \frac{1}{2\Omega}\sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)^2\frac{d^2v}{d\theta_l d\theta_k}}{v_j^2} - \frac{1}{\Omega}\sum_{j=1}^{J}\frac{(x_j - \Omega\phi_j)^2\frac{d\phi}{d\theta_l}\frac{dv}{d\theta_k}}{v_j^3}
\end{aligned}$$

L'information de Fisher s'obtient comme l'opposé de l'espérance de la dérivée seconde de la log vraisemblance : $-E(\frac{d^2l}{d\theta_l d\theta_k})$. Les termes avec $(x_j - \Omega\phi_j)$ dans la dérivée seconde dans le calcul de l'information de Fisher s'annulent $E(x_j - \Omega\phi_j) = 0$ car $E(x_j) = \Omega\phi_j$ et les termes avec $(x_j - \Omega\phi_j)^2$ sont modifiés car $E(x_j - \Omega\phi_j)^2$ est la variance de x_j qui vaut v_j. On obtient

$$I_{kl}(\theta) = (\frac{1}{2} - \frac{1}{2\Omega})\sum_{j=1}^{J}\frac{\frac{d^2v}{d\theta_l d\theta_k}}{v_i} + (\frac{1}{\Omega} - \frac{1}{2})\sum_{j=1}^{J}\frac{\frac{dv}{d\theta_l}\frac{dv}{d\theta_k}}{v_j^2} + \Omega\sum_{j=1}^{J}\frac{\frac{d\phi}{d\theta_l}\frac{d\phi}{d\theta_k}}{v_j} \tag{6.76}$$

Le modèle (M) sera utilisé concurremment à un modèle (S) décrit par une seule équation différentielle avec une erreur résiduelle normale, de moyenne 0 et la variance σ^2 constante pour tous temps, décrit ci-dessous.

6.6.2 Modèle simple (S)

Soit $x = (x_j)_{j=1...J}$ sont les nombres de sujets infectés observés aux instantes de temps $(t_1, \ldots, t_J)$

$$x_j = I_\theta(t) + \epsilon_j \tag{6.77}$$

$$\epsilon_j \sim \mathcal{N}(0, \sigma^2) \tag{6.78}$$

où $\theta = (\alpha, \beta)$ et un vecteur de paramètres, $I_\theta(t)$ est une fonctions non linéaires définie comme la solution d'équations différentielle suivante

$$\frac{dI_\theta}{dt} = \frac{\alpha}{N}(N - I_\theta)I_\theta - \beta I_\theta \tag{6.79}$$

$$I_\theta(t_0) = I_0 \tag{6.80}$$

Supposons que :

$$x_j \sim \mathcal{N}\left(I_\theta(t_j), \sigma^2\right) \tag{6.81}$$

avec la contribution de x_j observé à l'instant t_j à la vraisemblance donnée par

$$V_\theta(x_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{(x_j - I_\theta(t_j))^2}{2\sigma^2}\right\} \tag{6.82}$$

où $I_\theta(t_j)$ est la solution de l'équation (6.79) avec la condition initiale $I_0 = x_{j-1}$ (la condition initiale est construite à partir de l'observation précédente). Compte tenu de la relation $I(t_j) = \Omega\phi_j$ on peut écrire $X_j \sim \mathcal{N}(\Omega\phi_j; \sigma^2)$ pour pouvoir utiliser une partie des résultats obtenu pour le modèle **(M)**. Les paramètres de ce modèle α, β, i_0 et σ^2 sont notés analogiquement $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$.

La contribution V_j de la jème effective à la vraisemblance avec une écriture plus légère est

$$V_j = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x_j - \Omega\phi_j)^2}{2\sigma^2}\right\} \tag{6.83}$$

$$\ln V_j = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{(x_j - \Omega\phi_j)^2}{2\sigma^2}$$

et donc la logvraisemblance l est

$$l = \sum \ln V_j = -\frac{J}{2}\ln(2\pi) - \frac{J}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{J}(x_j - \Omega\phi_j)^2$$

La dérivée de la logvraisemblance par rapport à trois premiers paramètres $k = 1, 2, 3$

$$\frac{dl}{d\theta_k} = \frac{\Omega}{\sigma^2}\sum_{j=1}^{J}(x_j - \Omega\phi_j)\frac{d\phi}{d\theta_k}$$

La dérivée de la logvraisemblance par rapport σ^2

$$\frac{dl}{d\sigma^2} = -\frac{J}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{J}(x_j - \Omega\phi_j)^2$$

La dérivée seconde de la logvraisemblance où $k, l = 1, 2, 3$

$$\frac{d^2l}{d\theta_k d\theta_l} = -\frac{\Omega^2}{\sigma^2}\sum_{j=1}^{J}\frac{d\phi}{d\theta_k}\frac{d\phi}{d\theta_l} + \frac{\Omega}{\sigma^2}\sum_{j=1}^{J}(x_j - \Omega\phi_j)\frac{d^2\phi}{d\theta_k d\theta_l}$$

La dérivée seconde de la logvraisemblance par rapport σ^2 et θ_k où $k = 1, 2, 3$

$$\frac{d^2l}{d\sigma^2 d\theta_k} = -\frac{\Omega}{\sigma^4}\sum_{j=1}^{J}(x_j - \Omega\phi_j)\frac{d\phi}{d\theta_k}$$

La dérivée seconde de la logvraisemblance par rapport σ^2

$$\frac{d^2l}{d\sigma^4} = \frac{J}{2}\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\sum_{j=1}^{J}(x_j - \Omega\phi_j)^2$$

L'information de Fisher pour $k, l = 1, 2, 3$ est

$$I_{kl}(\theta) = -E\left(\frac{d^2l}{d\theta_k d\theta_l}\right) = \frac{\Omega^2}{\sigma^2}\sum_{j=1}^{J}\frac{d\phi}{d\theta_k}\frac{d\phi}{d\theta_l}$$

L'information de Fisher pour $l = 4$ (σ^2) et θ_k où $k = 1, 2, 3$ est

$$I_{kl}(\theta) = -E\left(\frac{d^2l}{d\theta_k d\sigma^2}\right) = 0 \tag{6.84}$$

et finalement l'information de Fisher pour $k = 4$ et $l = 4$

$$I_{kl}(\theta) = -E\left(\frac{d^2l}{d\sigma^4}\right) = \frac{J}{2\sigma^4} \tag{6.85}$$

6.7 Estimation par maximum de vraisemblance

6.7.1 Estimation des paramètres

Pour les valeurs observées $x = (x_j)_{j=1...J}$ des nombres de sujets infectés aux instantes de temps $(t_1, \ldots, t_J)$ on veut estimer les paramètres $\theta = (\alpha, \beta)$ et aussi la valeur initiale x_0 qui n'est pas observée ainsi que la valeur σ^2 dans le modèle simple (S). Dans l'estimation nous utilisons la méthode du maximum de vraisemblance.

Pour trouver les valeurs de $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ et $\hat{x_0}$ qui maximisent la log-vraisemblance dans le cas du modèle (M)

$$(\hat{\theta}, \hat{x_0}) = \arg \left\{ \max_{\theta, x_0} \log V \left(x_{1:J}; \theta, x_0, \{\phi_\theta(t_j)\}_{j=1}^J, \left\{ \langle \xi^2 \rangle_\theta (t_j) \right\}_{j=1}^J \right) \right\}$$

et les valeurs de $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$, $\hat{\sigma^2}$ et $\hat{x_0}$ qui maximisent la log-vraisemblance du modèle simple (S)

$$(\hat{\theta}, \hat{\sigma^2}, \hat{x_0}) = \arg \left\{ \max_{\theta, \sigma^2, x_0} \log V \left(x_{1:J}; \theta, \sigma^2, x_0, \{I_\theta(t_j)\}_{j=1}^J \right) \right\}$$

6.7.2 Estimation de la variance asymptotique

6.7.2.1 Estimation à partir de la matrice hessienne

Pour obtenir l'estimation de la variance asymptotique on peut calculer le hessien $\hat{H}$ de la log-vraisemblance en valeurs de paramètres estimés. L'information de Fisher est ainsi approchée par

$$\hat{I}_{kl} \approx -\hat{H}_{kl} \tag{6.86}$$

d'où la variance asymptotique est calculée comme

$$\hat{\sigma^2} = \hat{I}_{kl}^{-1} / J \tag{6.87}$$

6.7.2.2 Estimation à partir de la matrice d'Information de Fisher

L'estimation de la variance asymptotique on peut calculer comme l'inverse de la matrice d'Information de Fisher calculée directement. Pour calculer l'information de Fisher (équations ((6.76), (6.84) et (6.85))) il faut calculer les dérivées : $\frac{d^2 v}{d\theta_l d\theta_l}$, $\frac{dv}{d\theta_k}$ et $\frac{d\phi}{d\theta_k}$ en valeurs des

paramètres estimés pour le modèle (M) et $\frac{d\phi}{d\theta_k}$ pour le modèle (S). Cela nous pouvant faire en approchant les dérivées par hessien et gradient numériques de ϕ et v. Une autre idée est d'écrire un système d'équations différentielles contenant les équations pour ces dérivées (Ramsay2007 les appelle : *sensitivity differential equations*)à partir de système de deux équations différentielles pour ϕ est $\langle \xi^2 \rangle$ (maintenant noté v), puis les résoudre numériquement.

6.7.2.3 *Sensitivity Differential Equations*

Ecrivons le système des équations pour ϕ et v de manier générale comme

$$\begin{cases} \frac{d\phi}{dt} = f(\phi, \theta) \\ \frac{dv}{dt} = h(v, \phi, \theta) \end{cases}$$

où $v = v(\phi, \theta)$ est fonction de ϕ et θ. En dérivant chaque équation par rapport au θ_k on obtient

$$\frac{d}{dt}\left(\frac{d\phi}{d\theta_k}\right) = \frac{\partial f}{\partial \phi}\frac{d\phi}{d\theta_k} + \frac{\partial f}{\partial \theta_k} \tag{6.88}$$

Avec les conditions initiales $\frac{d\phi}{d\theta_k}|_{t=0} = 0$ pour $k = 1, 2$ et $\frac{d\phi}{d\theta_k}|_{t=0} = \frac{1}{\Omega}$ pour $k = 3$ (c-à-d, $\frac{d\phi}{d\alpha}|_{t=0} = 0$, $\frac{d\phi}{d\beta}|_{t=0} = 0$ et $\frac{d\phi}{di_0}|_{t=0} = \frac{1}{\Omega}$)

$$\frac{d}{dt}\left(\frac{d^2\phi}{d\theta_k d\theta_l}\right) = \frac{\partial^2 f}{\partial \phi^2}\frac{d\phi}{d\theta_k}\frac{d\phi}{d\theta_l} + \frac{\partial^2 f}{\partial \theta_l \partial \phi}\frac{d\phi}{d\theta_k} + \frac{\partial f}{\partial \phi}\frac{d^2\phi}{d\theta_k d\theta_l} + \frac{\partial^2 f}{\partial \phi \partial \theta_k}\frac{d\phi}{d\theta_l} + \frac{\partial^2 f}{\partial \theta_k \partial \theta_l} \tag{6.89}$$

Avec les condition initiales pour les dérivées secondes $\frac{d^2\phi}{d\theta_k d\theta_l}|_{t=0} = 0$

$$\frac{d}{dt}\left(\frac{dv}{d\theta_k}\right) = \frac{\partial h}{\partial v}\frac{dv}{d\theta_k} + \frac{\partial h}{\partial \phi}\frac{d\phi}{d\theta_k} + \frac{\partial h}{\partial \theta_k} \tag{6.90}$$

Les conditions initiales étant $\frac{dv}{d\theta_k}|_{t=0} = 0$ pour $k = 1, 2, 3$

$$\begin{aligned} \frac{d}{dt}\left(\frac{d^2v}{d\theta_k d\theta_l}\right) &= \frac{\partial^2 h}{\partial v^2}\frac{dv}{d\theta_k}\frac{dv}{d\theta_l} + \frac{\partial^2 h}{\partial \phi \partial v}\frac{d\phi}{d\theta_l}\frac{dv}{d\theta_k} + \frac{\partial^2 h}{\partial \theta_l \partial v}\frac{dv}{d\theta_k} + \frac{\partial h}{\partial v}\frac{d^2v}{d\theta_k d\theta_l} + \frac{\partial^2 h}{\partial v \partial \phi}\frac{dv}{d\theta_l}\frac{d\phi}{d\theta_k} + \frac{\partial^2 h}{\partial \phi^2}\frac{d\phi}{d\theta_k}\frac{d\phi}{d\theta_l} \\ &+ \frac{\partial^2 h}{\partial \theta_l \partial \phi}\frac{d\phi}{d\theta_k} + \frac{\partial h}{\partial \phi}\frac{d^2\phi}{d\theta_k d\theta_l} + \frac{\partial^2 h}{\partial v \partial \theta_k}\frac{dv}{d\theta_l} + \frac{\partial^2 h}{\partial \phi \partial \theta_k}\frac{d\phi}{d\theta_l} + \frac{\partial^2 h}{\partial \theta_k \partial \theta_l} \end{aligned} \tag{6.91}$$

En tenant compte de forme de fonctions f et h : $f(\phi, \theta = (\alpha, \beta, i_0)) = -\alpha\phi^2 + (\alpha - \beta)\phi$ et $h(v, \phi, \theta = (\alpha, \beta, i_0)) = -4\alpha\phi v + 2(\alpha - \beta)v - \alpha\phi^2 + (\alpha + \beta)\phi$ respectivement, les dérivées partielles dans les équations (6.88)-(6.91) sont résumées dans la table 6.1.

TABLE 6.1: Les dérivées partielles

$\frac{\partial f}{\partial \phi} = -2\alpha\phi - \beta + \alpha$	$\frac{\partial^2 f}{\partial \phi^2} = -2\alpha$	$\frac{\partial^2 f}{\partial \alpha \partial \phi} = -2\phi + 1$	$\frac{\partial^2 f}{\partial \beta \partial \phi} = -1$	$\frac{\partial^2 f}{\partial i_0 \partial \phi} = 0$
$\frac{\partial f}{\partial \alpha} = -\phi^2 + \phi$	$\frac{\partial^2 f}{\partial \alpha^2} = 0$	$\frac{\partial f}{\partial \beta} = -\phi$	$\frac{\partial^2 f}{\partial \beta^2} = 0$	
$\frac{\partial f}{\partial i_0} = 0$	$\frac{\partial^2 f}{\partial i_0^2} = 0$	$\frac{\partial^2 f}{\partial \alpha \partial \beta} = 0$	$\frac{\partial^2 f}{\partial \alpha \partial i_0} = 0$	$\frac{\partial^2 f}{\partial \beta \partial i_0} = 0$
$\frac{\partial h}{\partial v} = -4\alpha\phi - 2\beta + 2\alpha$	$\frac{\partial^2 h}{\partial v^2} = 0$	$\frac{\partial^2 h}{\partial \alpha \partial v} = -4\phi + 2$	$\frac{\partial^2 h}{\partial \beta \partial v} = -2$	$\frac{\partial^2 h}{\partial i_0 \partial v} = 0$
$\frac{\partial h}{\partial \phi} = -4\alpha v - 2\alpha\phi + \beta + \alpha$	$\frac{\partial^2 h}{\partial \phi^2} = -2\alpha$	$\frac{\partial^2 h}{\partial \phi \partial v} = -4\alpha$		
$\frac{\partial h}{\partial \alpha} = -4\phi v + 2v - \phi^2 + \phi$	$\frac{\partial^2 h}{\partial \alpha^2} = 0$	$\frac{\partial^2 h}{\partial \alpha \partial \phi} = -4v - 2\phi + 1$	$\frac{\partial^2 h}{\partial \beta \partial \phi} = 1$	$\frac{\partial^2 h}{\partial i_0 \partial \phi} = 0$
$\frac{\partial h}{\partial \beta} = -2v + \phi$	$\frac{\partial^2 h}{\partial \beta^2} = 0$	$\frac{\partial h}{\partial i_0} = 0$	$\frac{\partial^2 h}{\partial i_0^2} = 0$	
$\frac{\partial^2 h}{\partial \alpha \partial \beta} = 0$	$\frac{\partial^2 h}{\partial \alpha \partial i_0} = 0$	$\frac{\partial^2 h}{\partial \beta \partial i_0} = 0$		

6.7.2.4 Estimation de la variance de R_0

Puisque le taux de reproduction R_0 n'apparaît pas explicitement dans les vraisemblances du modèle (M) et (S), son estimateur (plug-in) est calculé comme le rapport des estimateurs de α et β. La variance de R_0 est calculée via la delta méthode (Casella and Berger, 2002).

$$var(\hat{R}_0) = var\left(\frac{\hat{\alpha}}{\hat{\beta}}\right) = \left(\frac{1}{\hat{\beta}}\right)^2 var(\hat{\alpha}) + \left(\frac{\hat{\alpha}}{\hat{\beta}^2}\right)^2 var(\hat{\beta}) - 2\frac{\hat{\alpha}}{\hat{\beta}^3} cov(\hat{\alpha}, \hat{\beta}) \tag{6.92}$$

6.8 Estimation bayésienne

Pour décrire les données $x = (x_j)_{j=1\ldots J}$, réalisation du processus X_t de nombre de sujets infectés (avec α un taux de contact et β un taux de guérison (confondu avec des naissances ou des décès)), observées aux instants de temps $(t_1, \ldots, t_J)$, les deux modèles sont utilisés. Le premier déterminé par le développement de l'équation maîtresse (M) est décrit dans la section (6.6.1) avec la vraisemblance (6.75) et le second le modèle simple (S) décrit dans la section (6.6.2) avec la vraisemblance (6.75). Les paramètres du modèle (M) sont $\theta = (\alpha, \beta, i_0)$ et ceux du modèle (S) $\theta = (\alpha, \beta, i_0, \sigma^2)$. Dans le cadre bayésien les paramètres sont supposés aléatoires et une loi a priori est associée à chaque paramètre. Soit $\pi(\theta)$ la loi a priori de θ. La loi a posteriori $\pi(\theta|x)$ s'obtient par application du théorème de Bayes.

$$\pi(\theta|x) \propto V_\theta(x)\pi(\theta) \tag{6.93}$$

où $V_\theta(x)$ est la vraisemblance du modèle.

6.8.1 Estimation des paramètres

La solution de l'inférence bayésienne est la loi a posteriori $\pi(\theta|x)$. Dans l'analyse des modèles (M) et (S), l'estimateur de θ est choisi comme minimisant l'erreur quadratique sous la loi jointe $\pi(\theta, x)$ de θ et x

$$\hat{\theta}(x) = \arg\left\{\min_\theta E_{\pi(\theta,x)}\left[(\theta - \hat{\theta}(x))^2\right]\right\} = E_{\pi(\theta|x)}(\theta) \tag{6.94}$$

où l'espérance a posteriori $E_{\pi(\theta|x)}(\theta)$ est donnée par l'intégrale suivante

$$E_{\pi(\theta|x)}(\theta) = \int \theta\pi(\theta|x)d\theta \tag{6.95}$$

Les différentes méthodes d'estimation des espérances a posteriori ainsi que des variances

de ces estimations sont présentées dans la première partie du manuscrit.

7 Résultats

Les résultats du modèle SIS (susceptible - infecté - susceptible) sont présentés soit dans la version déterministe, soit dans la version stochastique via la méthode du développement. Les deux modèles sont utilisés pour décrire le nombre de sujets infectés dans une population de taille constante. Le processus SIS est présenté sous la forme d'un processus de naissance et de mort où les taux d'apparitions de nouveaux cas infectés sont une fonction non linéaire des effectifs de sujets infectés. La méthode de développement de van Kampen est utilisée pour déterminer la moyenne et la variance du processus.

Dans un premier temps, les résultats du développement et la solution déterministe sont comparés aux simulations du PNM générés selon les probabilités de transition. Puis les deux modèles de régression sont construits à partir du modèle déterministe et du modèle stochastique. Une étude de simulations est alors mise en place pour étudier les propriétés des estimateurs des paramètres sous ces deux modèles. Les données utilisées dans les estimations ont été simulées à partir du PNM. Deux types d'inférences ont été effectués : fréquentiste et bayésienne. Pour accélérer les calculs des estimations bayésiennes, la méthode de l'échantillonnage préférentiel a été utilisée.

7.1 Modélisation

Nous pouvons tout d'abord remarquer que la solution déterministe $I(t)$ est égale à l'espérance du processus des sujets infectés $N\phi(t)$. Le nombre moyen de sujets infectés dans l'état stationnaire est alors celui pour lequel les dérivées (6.54) et (6.63) s'annulent. Les deux solutions stationnaires possibles sont : $I^s = 0$ et $I^s = N(1 - \beta/\alpha)$. Pour une valeur initiale

$I(0) = i_0$ strictement positive, la solution générale de (6.54) ou (6.63) est non nulle (cf. équation (6.55)). Si $\alpha > \beta$, la solution converge vers l'état stationnaire non-nul $I = N(1 - \beta/\alpha)$ et si $\alpha < \beta$, vers la solution nulle $I = 0$. L'approximation de la probabilité stationnaire Π^s est gaussienne si $\alpha > \beta$. Seuls les cas pour lesquels $\alpha > \beta$ seront étudiés. Cette condition correspond de plus au cas où le taux de reproduction R_0 est supérieur à 1. Il s'agit d'un paramètre qui pour le modèle SIS est $R_0 = \frac{\alpha}{\beta}$ (Allen and Burgin, 2000; Anderson and May, 1992) et qui, par définition, représente le nombre de cas secondaires d'infectés produit par un sujet infecté introduit dans une population des susceptibles pendant toute la durée de sa période infectieuse (Cf. par exemple Diekmann et al., 1990). Ainsi, si $R_0 > 1$ l'infection persiste et si $R_0 < 1$, elle s'éteint. La valeur $R_0 = 1$ constitue une zone critique pour l'évolution de la maladie car soit elle disparaît soit elle devient pandémique.

Dans le modèle SIS, la zone critique correspond aux valeurs α et β proches : $\alpha \approx \beta$. Même si $R_0 > 1$, et donc $\alpha > \beta$, l'extinction de la maladie est toujours possible dans le modèle SIS stochastique définit comme un PNM. En effet, la probabilité d'extinction précoce peut être importante, selon le choix des paramètres α et β et la valeur initiale $I(0) = i_0$, car elle vaut $(\frac{\beta}{\alpha})^{i_0}$ (Allen and Burgin, 2000; Van Kampen, 2007). L'objectif est de comparer les réalisations des processus stochastiques obtenues via des simulations avec les résultats obtenus à partir du développement de l'équation maîtresse. Pour cela, 1000 trajectoires des processus ont été générées selon l'algorithme de Gillespie (1977) avec les probabilités de transition (6.56) et (6.57) avec un incrément de temps Δt suivant une loi exponentielle de paramètre égal à la somme des taux du modèle. La moyenne et la variance des trajectoires persistantes (cas endémique) ont été comparées à l'espérance et la variance déduites par le développement de l'équation maîtresse $\Omega\phi$ et $\Omega \left\langle \xi^2 \right\rangle$ (Cf. (6.63) et (6.66) respectivement).

Les paramètres α et β ont été choisis de manière à respecter la condition de normalité de la loi stationnaire du processus des infectés et de manière à correspondre aux différentes niveaux de la solution stationnaire $I^s = N(1 - \beta/\alpha)$: $\alpha = 1.2$ avec $\beta = 0.3$ ($I^s = 25$) ou avec $\beta = 0.6$ ($I^s = 5$) ou avec $\beta = 0.9$ ($I^s = 75$). Le quatrième cas de paramétrisation est choisi proche de la zone critique de l'évolution d'épidémie $\alpha \approx \beta$ avec $\alpha = 1.2$ et $\beta = 1.1$. Pour obtenir les différentes probabilités d'extinction $(\frac{\beta}{\alpha})^{i_0}$, les valeurs initiales i_0 du processus sont

fixées à i_0=1, 10, 25, 50, 75 et 99. La taille de la population a été fixée à $N = 100$ ($N = \Omega$). Les résultats sont présentés sur les graphiques (7.1), (7.2), (7.3), et (7.4). Les moyennes et variances des trajectoires qui au bout d'un certain temps ont été absorbées par l'état zéro (l'épidémie disparaît complètement de la population - à la fin de l'épidémie, la population est constituée uniquement de sujets susceptibles) sont représentées sur les graphiques (16), (17), (18) et (19) en annexe. La proportion de ces trajectoires respectent bien la probabilité d'extinction $(\frac{\beta}{\alpha})^{i_0}$.

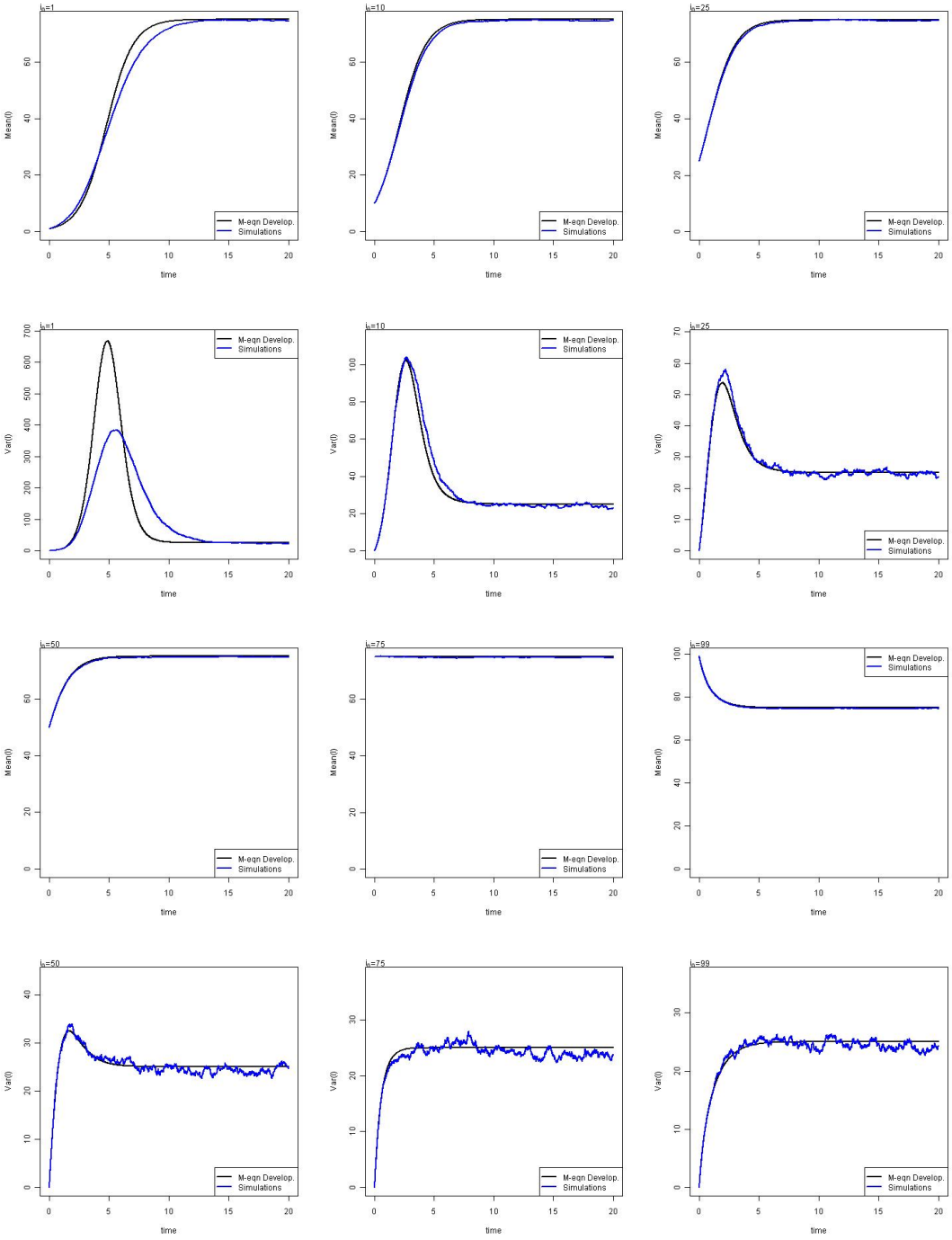

FIGURE 7.1: $\alpha = 1.2$, $\beta = 0.3$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires

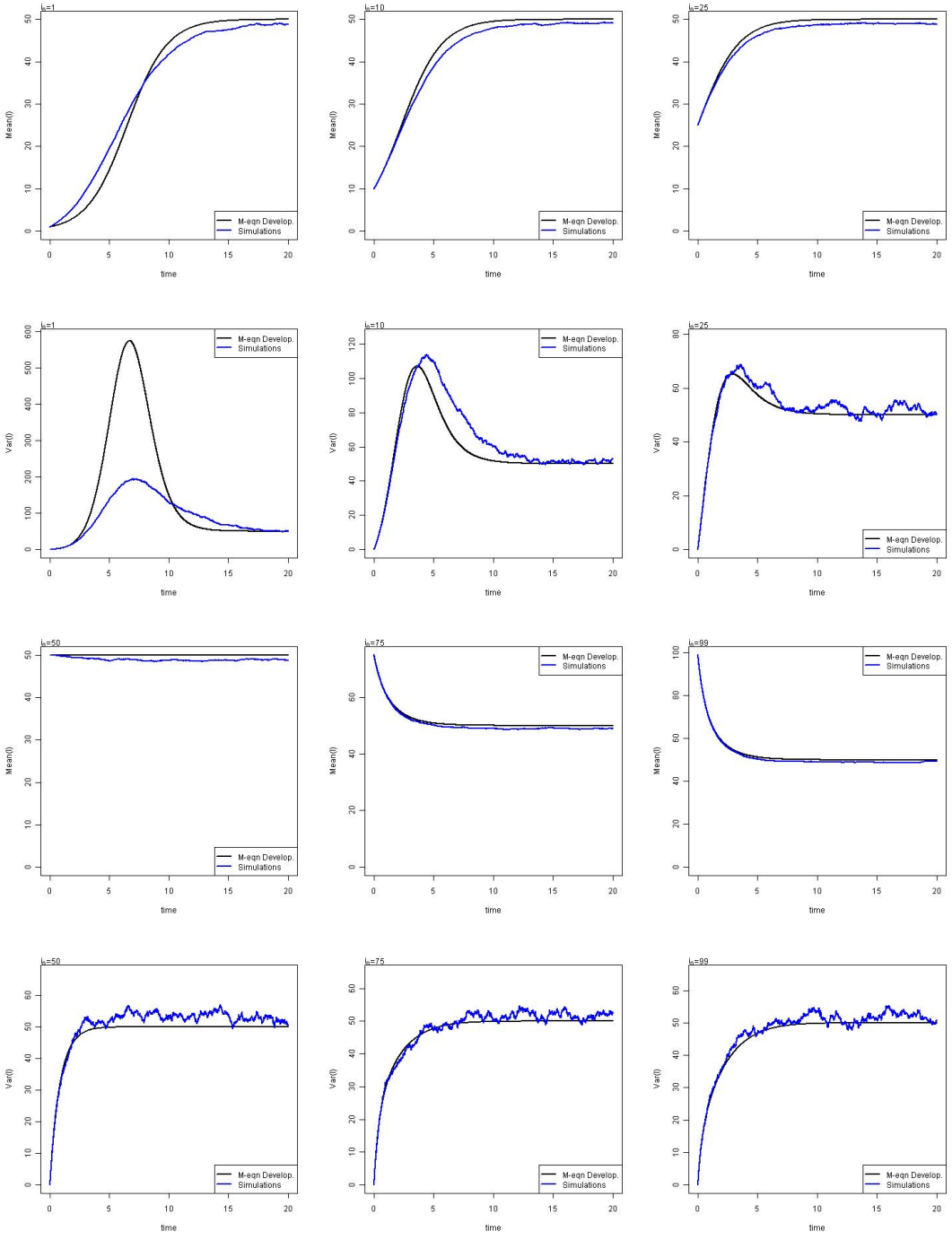

FIGURE 7.2: $\alpha = 1.2$, $\beta = 0.6$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires

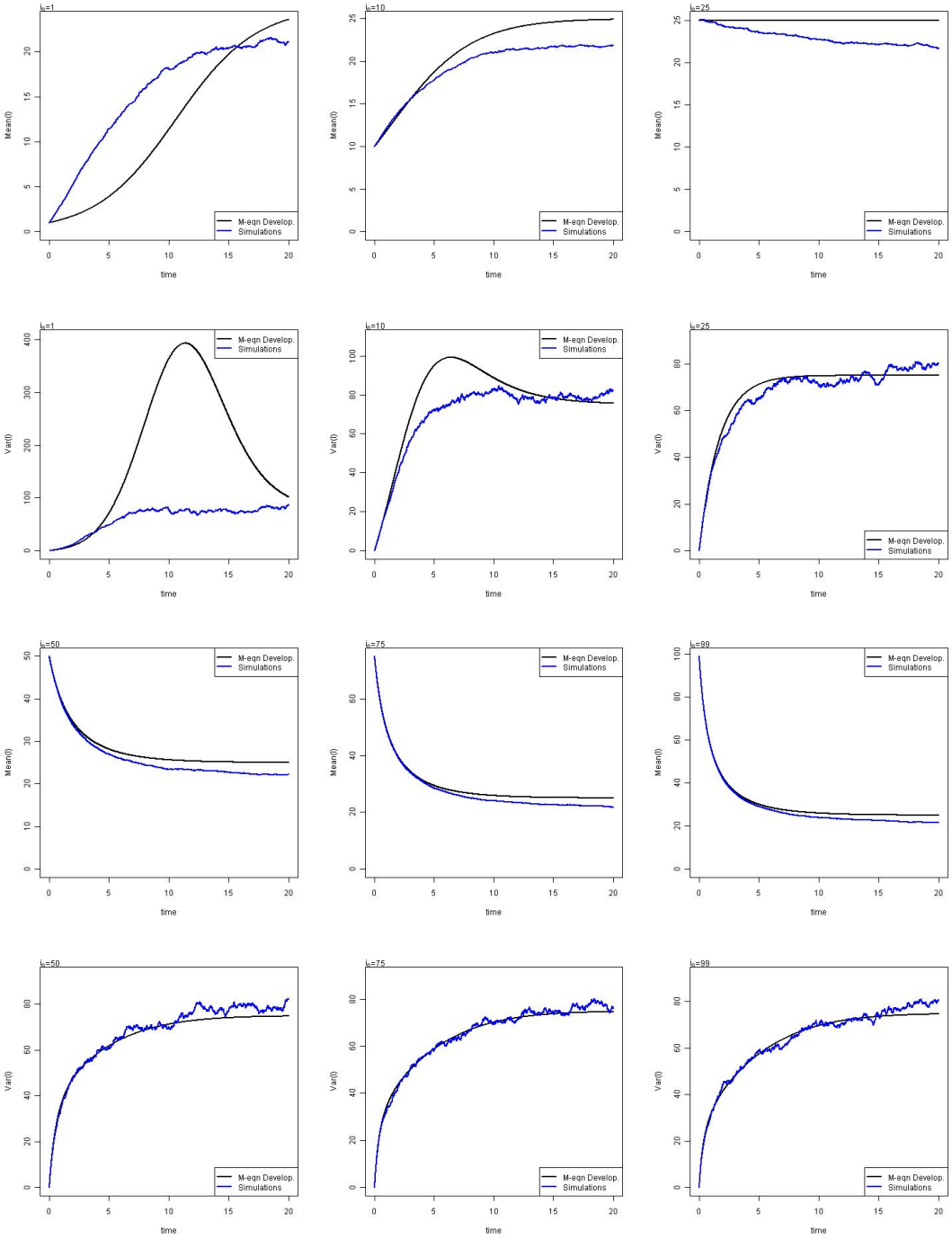

FIGURE 7.3: $\alpha = 1.2$, $\beta = 0.9$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires

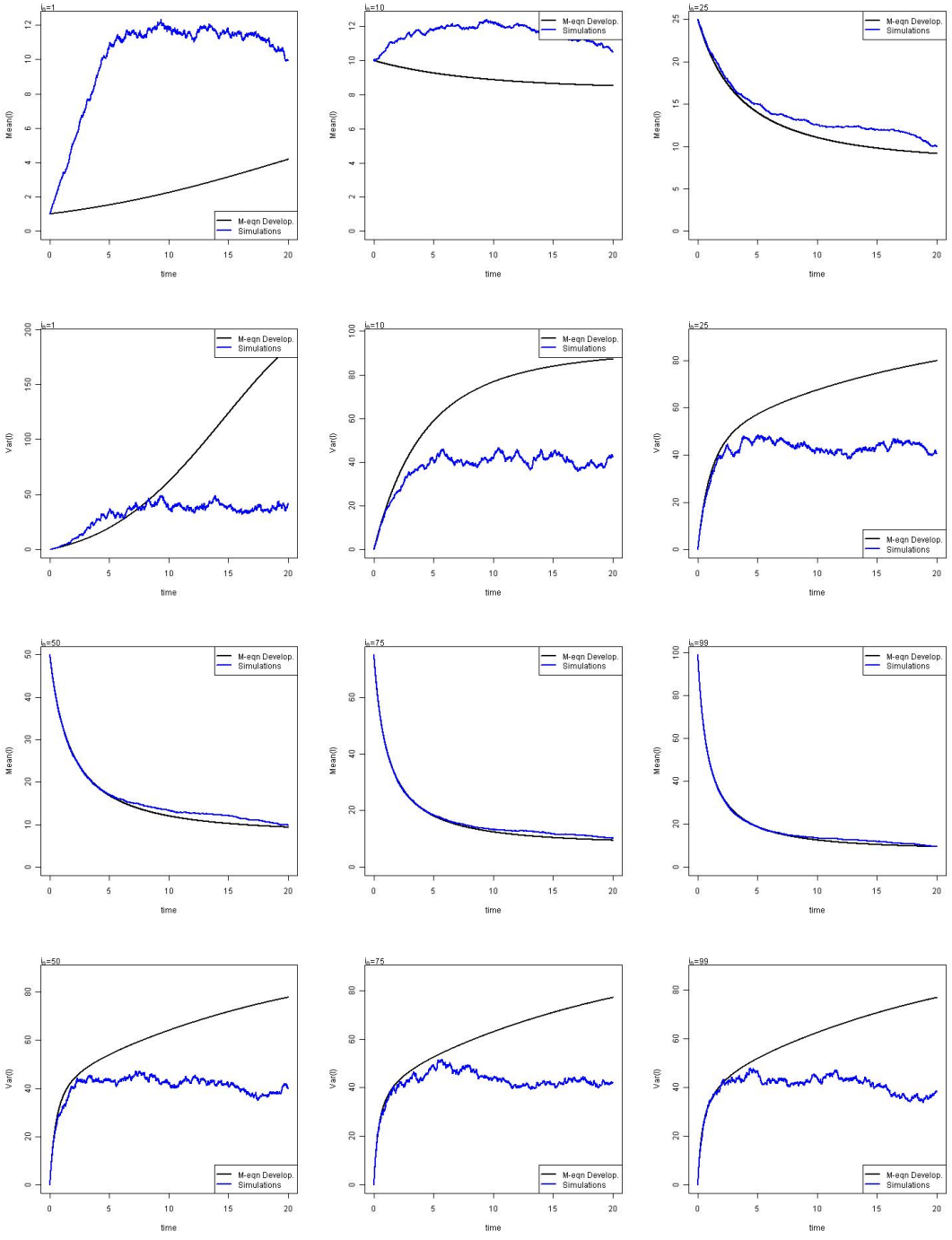

FIGURE 7.4: $\alpha = 1.2$, $\beta = 1.1$, non-extinction. Les graphiques correspondent aux valeurs $i_0 = 1, 10, 25, 50, 75, 99$. Les moyennes sont données sur les lignes impaires, les variances sur les lignes paires

7.2 Simulations

7.2.1 Choix des paramétrisations

Les valeurs des paramètres α, β et de la valeur initiale $i_0 = x_0$ déterminent l'état stationnaire (s) du processus ainsi que la durée de la phase évolutive (t) du processus avant qu'il atteigne son régime stationnaire. Ainsi, pour une population de taille $N = 100$ constante, trois niveaux différents de l'état stationnaire ont été choisis : $s = 25$, $= 50$ et $s = 75$ (ce qui correspond à la situation où 25%, 50% et 75% de la population sont atteints par la maladie) et également, trois durées de la phase évolutive (t) : $t = 1$, $t = 5$ et $t = 20$ (ce qui correspond au cas où la maladie évolue dans la population soit assez rapidement ($t = 1$ an), en vitesse moyenne ($t = 5$ ans) et lentement ($t = 20$ ans)). Les choix des paramétrisations pour α, β et x_0 correspondants aux différentes scénarios sont présentés dans le tableau (7.1).

TABLE 7.1: Pramétrisations pour α_0, β_0 et i_0

		$i_0 = 1$			$i_0 = 10$		
t	s	#	α_0	β_0	#	α_0	β_0
1	25	1	34.8	26.1	10	24	18
	50	2	20.4	10.2	11	15	7.5
	75	3	14.4	3.6	12	12	3
5	25	4	6.8	5.1	13	4.8	3.6
	50	5	4	2	14	3	1.5
	75	6	2.8	0.7	15	2.4	0.6
20	25	7	1.6	1.2	16	1.2	0.9
	50	8	1	0.5	17	0.8	0.4
	75	9	0.72	0.18	18	0.6	0.15

7.2.2 Données simulées

$S = 1000$ jeux de données de $n = 6$ observations ont été simulés selon des processus de Poisson, avec comme probabilités de transition g et r :

$$g_{t+\Delta t} = Pr(X_{t+\Delta t} = x + 1 \quad | \quad X_t = x) = \frac{\alpha}{N} x(N - x)\Delta t \tag{7.1}$$

$$r_{t+\Delta t} = Pr(X_{t+\Delta t} = x - 1 \quad | \quad X_t = x) = \beta x \Delta t \tag{7.2}$$

où Δt a été généré selon une loi exponentielle de paramètre $g_t + r_t$: $\Delta t \sim Exp(g_t + r_t)$. Ainsi, chaque jeu de données $(x_j)_{j=1...J}$ est constitué des réalisations du processus de Poisson aux instants équidistants $(t_1, \ldots, t_J)$ préétablis. Trois situations sont envisagées selon que les observations $(x_j)_{j=1...J}$ proviennent des trois intervalles différents : $[0, 0.5t]$, $[0, t]$ et $[t, 2t]$. Dans le premier cas, il s'agit d'observations provenant de la phase initiale d'évolution d'épidémie (« early »), dans le deuxième cas de la phase entière d'évolution (« during ») et dans le troisième cas du régime stationnaire (« after »). Les différents scénarios sont illustrés sur la figure (7.5).

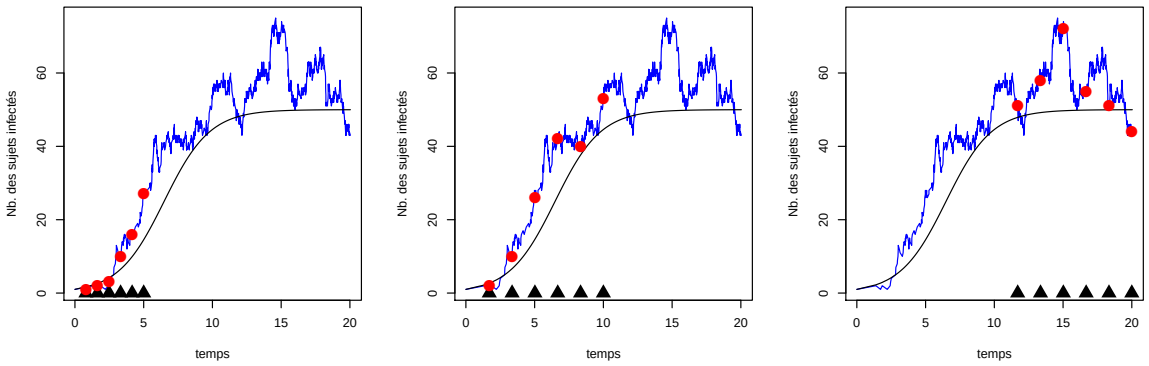

FIGURE 7.5: Les trois scénarios des simulations des données selon la phase : phase initiale « early » (à gauche), phase évolutive « during » (au centre), phase stationnaire « after » (à droit).

7.3 Résultats des estimations fréquentistes

Pour chaque jeu de données, les estimations des paramètres $\hat{\theta}_k$ et des variances asymptotiques correspondantes $\hat{\sigma^2}_k$ ont été calculées numériquement pour les deux modèles (M) et (S). Les valeurs maximisant la vraisemblance $\hat{\theta}_k$ sont calculées numériquement par la méthode d'optimisation [1] sous contrainte [2] de Nelder-Mead (Nelder and Mead, 1965) implémentée sous `R`. Dans la procédure d'optimisation, un solveur d'équations différentielles [3] est inclus pour trouver les valeurs $\{\phi_\theta(t_j)\}_{j=1}^J$, $\left\{\left\langle \xi^2 \right\rangle_\theta (t_j)\right\}_{j=1}^J$ dans le cas du modèle (M) et les valeurs

1. commande `constrOptim`
2. $\theta \in [0.001, N]$
3. package `deSolve` avec `ode`

, $\{I_\theta(t_j)\}_{j=1}^{J}$ dans le cas du modèle (S).

Pour obtenir l'estimation de la variance asymptotique, le hessien numérique $\hat{H}$ à été calculé par différences finies via un pas supplémentaire de la méthode numérique d'optimisation [4] L-BFGS-B (Broyden, Fletcher, Goldfarb and Shanno avec des contraintes (Byrd et al., 1995)). L'inverse de la matrice hessienne est calculé par la commande `qr.solve` sous `R`. Pour comparer les deux modèles, les caractéristiques suivantes (table 7.2) des estimations ont été évaluées.

TABLE 7.2: MC Caractéristiques

MC mean	$\bar{\theta} = \frac{1}{S}\sum_{k=1}^{S}\hat{\theta}_k$	MC sd	$\sqrt{\frac{1}{S-1}\sum_{k=1}^{S}(\hat{\theta}_k - \bar{\theta})^2}$
MC bias	$\bar{\theta} - \theta_0$	MC MSE	$\frac{1}{S}\sum_{k=1}^{S}(\hat{\theta}_k - \theta_0)^2$
MC relative bias	$\frac{\bar{\theta}-\theta_0}{\theta_0}$	Mean of as. variances	$\frac{1}{S}\sum_{k=1}^{S}\hat{\sigma^2}_k$
MC variance	$\frac{1}{S-1}\sum_{k=1}^{S}(\hat{\theta}_k - \bar{\theta})^2$	Mean of as. sd	$\frac{1}{S}\sum_{k=1}^{S}\hat{\sigma}_k$

7.3.1 Problème calculatoire

La méthode d'optimisation avec contrainte a été utilisée pour assurer que l'algorithme reste dans une certaine proximité de la solution théorique (valeurs fixées dans les simulations). Le nombre d'itérations a été fixé à 1000 (500 par défaut de `R`) pour que l'algorithme puisse atteindre sa convergence. Pour les valeurs des paramètres ainsi obtenues, les variances asymptotiques sont ensuite calculées comme décrit dans la section (6.7.2). Cependant, bien qu'il y ait eu diagnostic de convergence, certaines estimations de la variance asymptotique étaient négatives. Ces cas ont été supprimés car ils correspondent à des solutions non-optimales dans la recherche des maximas de la fonction de vraisemblance, matrice hessienne non définie négative pour le maximum de vraisemblance. Seules les estimations pour lesquelles les variances estimées étaient positives ont été analysées.

Les cas où l'algorithme a convergé vers des optima ne respectant pas les contraintes, ces résultats ont été aussi exclus de l'analyse ne correspondant pas aux maximas recherchés.

4. commande `optim`

Tous les résultats pour lesquels l'estimation de α a été plus petite que l'estimation de β sont supprimés également.

Les nombres des estimations (parmi un total de 1000) retenues pour les analyses d'estimateurs des paramètres des deux modèles sont regroupés dans le tableau (7.3. Dans le tableau sont aussi indiques les nombres des jeux de données pour lesquelles les estimations qui ont été obtenues sous les deux modèles à la fois ((M)∩(S)).

Les nombres de cas finalement utilisés en fonction des différentes causes d'élimination sont présentés dans le tableau (12) en annexe. Dans la quasi-totalité des cas, l'algorithme d'optimisation satisfait son critère de convergence. Les variances négatives signifient que le résultat trouvé par l'algorithme n'est pas une vraie solution de maximum de vraisemblance, ou bien qu'il s'agit d'un problème de mauvais conditionnement de la fonction de vraisemblance. En effet, si la surface de la vraisemblance au voisinage de la vraie solution a une courbure très faible, le diagnostic de convergence est satisfait sans que le maximum soit réellement atteint.

Les résultats réalisant les contraintes peuvent indiquer un maximum local qui n'est pas une solution recherchée. Il a été décidé que l'investigation détaillée de ces cas ne rentrait pas dans le cadre de ce travail de thèse.

Des solutions à ce problème d'estimation existent aussi avec les vraisemblances pénalisées (Ramsay et al., 2007) qui font un arbitrage entre le biais et la variance mais leur application n'a pas été envisagée ici. Une solution alternative avec les méthodes bayésiennes est étudiée plus loin pour une partie des paramétrisations.

Les estimations des variances asymptotiques obtenues via l'inverse de la matrice d'information de Fisher (Cf. section (6.7.2.2)) ont été encore plus souvent négatives. Pour calculer l'information de Fisher, il faut approcher numériquement les espérances des dérivées des solutions des équations numériques par rapport aux paramètres. Il faut alors calculer des gradients numériques ou résoudre - toujours numériquement - un système d'équations différentielles (comme décrit par *Sensitivity Differential Equations*) qui est assez compliqué. Dans les deux cas, le cumul des erreurs numériques peut amener à des résultats incohérents avec encore

des variances négatives. L'estimation de la variance asymptotique par le hessien numérique (section 6.7.2) paraît être une solution plus simple du point de vue calculatoire.

TABLE 7.3: Le nombre d'estimations « utilisables »

	Phase :	« early »			« during »			« after »		
#	Modèle :	(M)	(S)	(M)∩(S)	(M)	(S)	(M)∩(S)	(M)	(S)	(M)∩(S)
1		617	389	228	621	478	372	205	223	82
2		471	774	392	767	897	698	408	359	168
3		260	724	223	850	961	818	410	420	181
4		682	708	491	675	731	497	374	363	139
5		534	770	443	813	880	728	397	384	164
6		248	695	214	912	974	895	412	422	177
7		699	728	516	691	675	479	387	355	133
8		523	756	420	888	887	795	381	318	140
9		268	729	244	956	932	893	388	367	145
10		664	340	249	577	405	317	243	331	124
11		569	812	488	819	802	664	415	320	157
12		501	891	477	892	912	814	448	496	230
13		705	633	491	686	658	483	406	405	165
14		590	798	496	842	779	669	361	398	144
15		542	858	495	948	941	892	423	436	184
16		717	630	483	640	358	247	334	246	80
17		658	801	549	874	744	664	346	357	138
18		549	916	540	974	963	941	365	319	112

7.3.2 Interprétation des résultats

La méthode de développement suppose que $\alpha > \beta$ et que i_0 grand pour bien fonctionner (Cf. les résultats de la modélisation dans la section (7.1)). Nous avons étudié les simulations pour des valeurs de α, β et i_0 allant jusqu'à la limite de cette condition, c'est-à-dire quand les valeurs initiales i_0 sont plus petites que l'ordre de fluctuations $N^{1/2}$ (ce qui correspond aux paramétrisations avec i_0=1) et quand le rapport des paramètres α/β est proche de 1 (ainsi $R_0 \approx 1$) - donc lorsque l'état stationnaire est au niveau $s = 25$.

Par ailleurs, comme nous avons vu dans la partie concernant la modélisation, la moyenne et la variance obtenues via la méthode de développement imitent de manière très fidèle la moyenne et la variance des simulations, donc celles du processus SIS. La moyenne du modèle (M) correspond à la moyenne du modèle (S), d'autre part le modèle (M) utilise une équation supplémentaire pour décrire la variance. On s'attend donc à ce que le modèle (M) donne des estimations de paramètres au moins aussi bonnes que le modèle (S) qui suppose une variance

constante.

L'ensemble des résultats numériques est résumé dans les tables (13)-(26) dans l'annexe IV. Toutes les estimations sont illustrées dans les box-plots (20)-(43). Les propriétés asymptotiques de l'estimateur du maximum de vraisemblance ont été étudiées à l'aide des graphiques PP-plot (Cf. (44)-(55)) (probability-probability plot) où en abscisse se trouvent les probabilités cumulées observées, et en ordonnée, les probabilités cumulées théoriques de la loi normale centrée réduite (Wilk and Gnanadesikan, 1968; Michael, 1983).

Les résultats des estimations ne sont pas univoques selon les paramétrisations. Globalement, les paramètres α et β sont mieux estimés sous le modèle (M) surtout dans les scénarios « early » et « during », tandis que les estimations de R_0 sont meilleures en termes de biais et d'efficacité relative sous le modèle (S).

Le paramètre i_0 semble être mieux estimé sous le modèle (S) mais ceci devient de moins en moins vrai quand les observations s'éloignent du début de l'épidémie (cf. scénario « after »). Cependant, pour les paramétrisations avec $i_0 = 10$, les propriétés asymptotiques de l'estimateur de i_0 sont meilleures sous le modèle (M) dans les scénarios « early » et « during » même si, dans l'ensemble des paramétrisations, les estimations avec un biais très important sont nombreuses. Dans l'ensemble, on peut considérer i_0 comme un paramètre de nuisance car le nombre initial d'infectés n'est pas en général un problème épidémiologique considéré comme intéressant.

Les écarts-types asymptotiques sont généralement plus en accord avec les écarts-types des estimations sous le modèle (M) ce qui explique cette légère amélioration dans les propriétés asymptotiques de ce modèle.

Dans le scénario « after », les deux modèles se ressemblent car la variance du modèle (M) devient constante dans l'état stationnaire ce qui est également vrai avec le modèle (S), mais en dépendant des paramètres α et β. Dans ce scénario, l'estimation de i_0 est plutôt mauvaise. Ce résultat n'est pas très surprenant car les observations sont éloignées de l'état initial.

La méthode de développement est la plus performante quand les valeurs de i_0 et R_0 ($\alpha > \beta$) sont grandes. Parmi les paramétrisations étudiées, les valeurs les plus importantes de i_0 et R_0 correspondent aux paramétrisations (#12), (#15) et (#18) dans lesquelles $i_0 = 10$

et $R_0 = 4$. Dans ces trois paramétrisations, le modèle (M) reste meilleur pour l'estimation du paramètre α et du paramètre β dans les scénarios « early » et « during ». Dans l'ensemble des paramétrisations, R_0 est plutôt mieux estimé sous le modèle (S). Dans le scénario « after », les deux modèles sont comparables. Dans ce scénario, les estimateurs de α, β et i_0 ont généralement de mauvaises propriétés asymptotiques selon les PP-plots (non-normalité). Pour R_0, le même scénario donne de bons résultats asymptotiques sous le modèle (M).

7.3.3 Exemple de résultats de l'estimation de R_0 dans la phase évolutive

Le paramètre épidémiologique le plus intéressant est le taux de reproduction R_0 qui caractérise l'impact de la maladie dans la population. Comme exemple, une partie des résultats est présentée pour ce paramètre dans la phase évolutive « during ».

Le figure (7.6), qui est la même que celle reportée dans l'annexe (34), présente les box-plots d'estimations pour l'ensemble des paramétrisations.

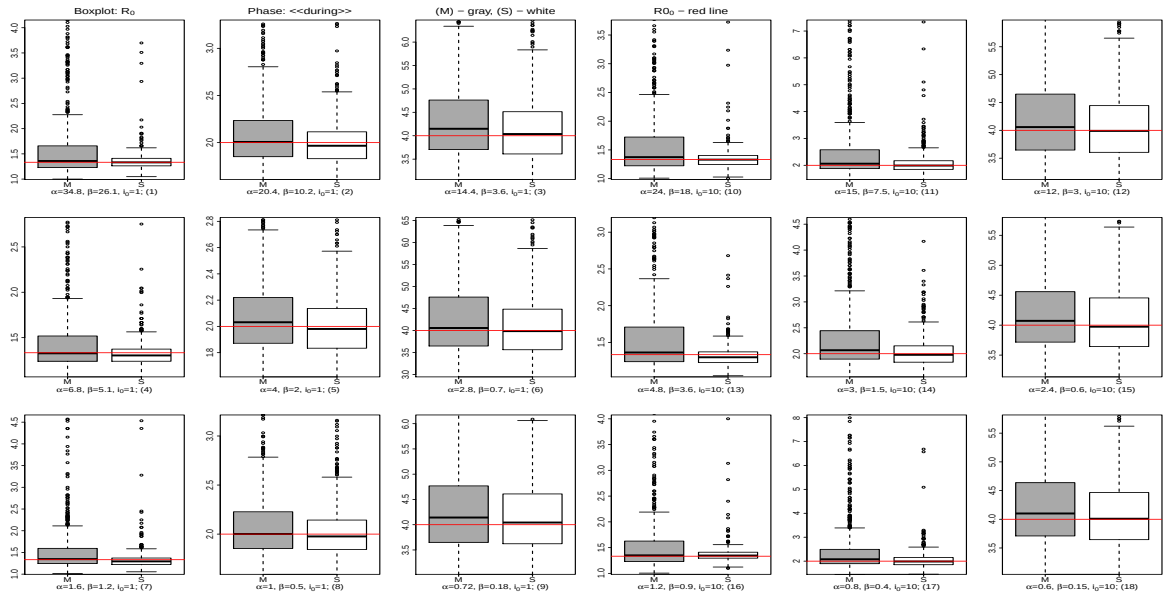

FIGURE 7.6: Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » - le trait rouge représente la vraie valeur du paramètre.

Les box-plots des estimations de R_0 sous le modèle (M) sont représentés en gris et ceux du modèle (S) sont en blanc. La ligne horizontale rouge représente la vraie valeur du rapport

de α et β fixés dans la simulation. Les estimations sous le modèle (S) sont plus proches de la vraie valeur que les estimations sous le modèle (M).

Sur le figure (7.7) (et (35) dans l'annexe) avec une échelle logarithmique, sont présentés les box-plots des estimations des écarts-types asymptotiques avec le même code couleur pour les deux modèles. Pour chaque estimation de R_0, l'écart-type asymptotique a été estimé comme la racine de la variance obtenue par la delta méthode (section (6.7.2.4)). Les losanges rouges reportés sur chaque box-plot représentent les écarts-types des estimations de R_0. Sous les deux modèles, les écarts-types asymptotiques sous-estiment l'écart-type des estimations. Cependant, en écartant les 5% des estimations de R_0 les plus extrêmes, les écarts-types sont diminués et correspondent mieux aux écarts-types asymptotiques (Cf. les résultats du tableau (22) en annexe).

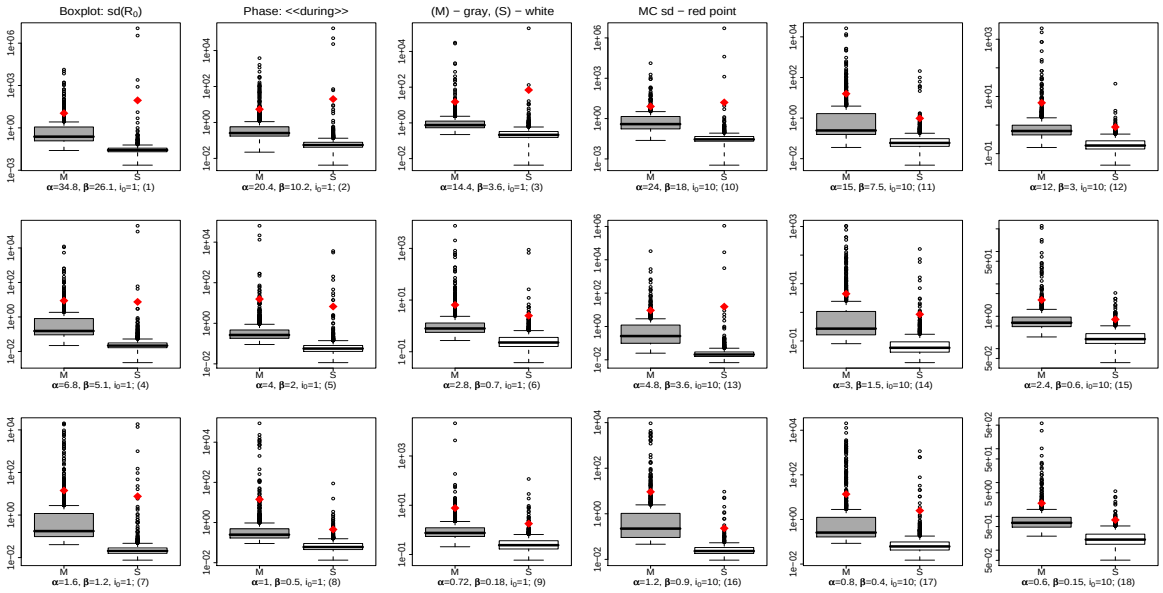

FIGURE 7.7: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » - les losanges rouges représentent les écarts-types des estimations.

Les propriétés asymptotiques des estimations de R_0 sont illustrées sur les PP-plots dans la figure (7.8) (et (51) dans l'annexe). Les probabilités théoriques (ordonnées) en fonction des probabilités observées (de valeur $(\hat{R}_{0k} - \alpha_0/\beta_0)/\hat{\sigma}_k$) (abscisses) sont présentées en rouge pour le modèle (M) et en bleu pour le modèle (S). La distribution des estimations sous le modèle

(M) est plus proche de la loi normale centrée réduite que celle des estimations sous le modèle (S) est plus aplatie. Puisque les estimations sous le modèle (S) ont un plus petit biais (Cf. box-plot (7.6) et d'après les résultats numériques dans le tableau (22) dans l'annexe),l'écart type est donc sous-estimé sous ce modèle.

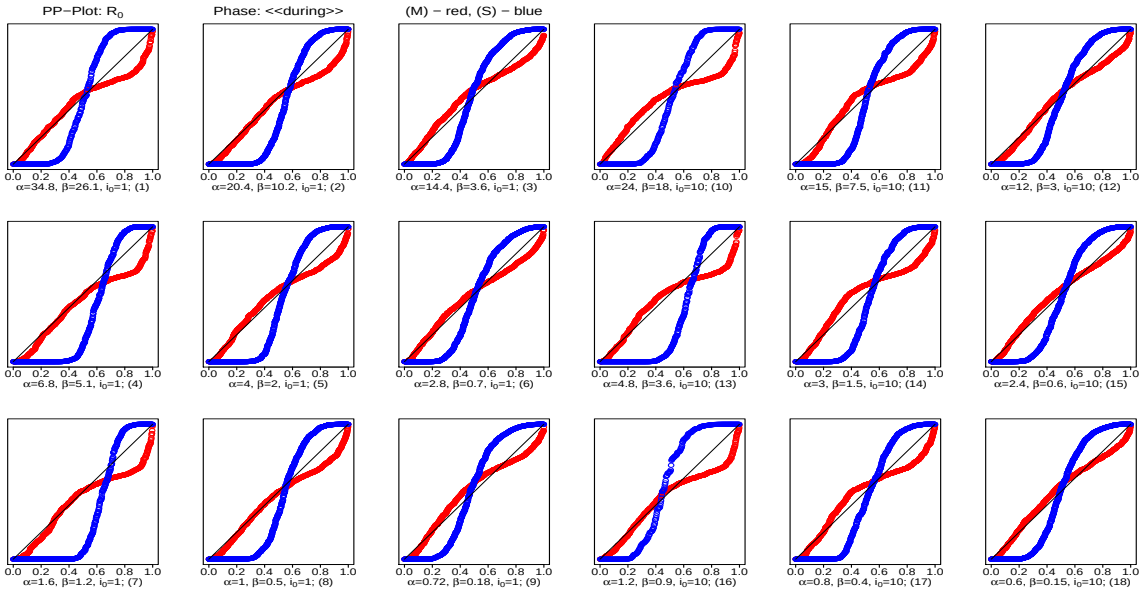

FIGURE 7.8: PP-plots de R_0 pour les observations de la phase evolutive : « during »

Pour prendre en compte à la fois le biais et la précision des estimations, les erreurs quadratiques moyennes ont été calculées. Afin de comparer les estimations sous les deux modèles, l'efficacité relative (RE) est déduite comme le rapport de l'erreur quadratique moyenne des estimations du modèle (M) divisé par l'erreur quadratique moyenne des estimations du modèle (S) (Cf. tableau (7.2)). Ainsi les valeurs supérieur à 1 (RE>1) sont en faveur du modèle (S). Ces résultats extraits du tableau (22) sont présentés dans le tableau suivant

Les valeurs de RE sont toutes plus grandes que 1 et suggèrent que le modèle (S) est meilleur en terme d'estimation du paramètre R_0. Sur les box-plots 7.6, on constate un nombre important d'estimations au-delà de l'extrémité de la moustache supérieure. Elles influencent les résultats numériques (7.2) du tableau (21) qui ne sont pas cohérentes avec les conclusions fournies par les box-plots. Pour cette raison, 5% des estimations les plus extrêmes de R_0 ont été exclues de l'analyse (2.5% des plus petites et 2.5% des plus grandes). Les résultats dans le

TABLE 7.4: Efficacité relative d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (95% des estimations)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE [1]	-	**200.60**	-	**21.58**	-	**2.52**	-	**92.95**	-	**128.90**	-	**2.43**
	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE	-	**123.22**	-	**4.38**	-	**3.20**	-	**93.77**	-	**38.00**	-	**1.61**
	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE	-	**1001.88**	-	**11.30**	-	**1.61**	-	**304.76**	-	**878.10**	-	**1.77**

[1] Les caractéristiques présentées dans le tableau sont évaluées pour 95% des estimations.
5% des estimations extrêmes ont été écartées.
En annexe (Cf. tableau(21)) on donne les RE calculés sur l'ensemble des résultats.

tableau (7.4) sont donc évalués sur 95% des estimations obtenues pour ce paramètre. On peut remarquer que les résultats des estimations pour les paramètres α, β (et i_0) ont été moins sensibles aux estimations extrêmes, les conclusions n'ont pas été modifiées.

7.3.4 Identifiabilité des paramètres de modèles

Pour expliquer l'incohérence entre les résultats des estimations de α et β par rapport aux résultats des estimations de R_0, on peut faire l'hypothèse que les bonnes estimations d'un des deux paramètres α ou β correspondent aux mauvaises estimations de l'autre paramètre. Pour vérifier cette idée, les estimations de α ont été représentées en fonction des estimations de β sur les graphiques avec une échelle log-log. Pour le scénario correspondant aux observations dans la phase évolutive, ce graphe se trouve sur la figure (7.9) et, pour les autres scénarios, en annexes (56)-(58).

Sur ces graphiques, une droite d'équation $\alpha = R_0\beta$ a été tracée avec la vraie valeur de R_0. Dans le plupart des estimations, α et β sont alignés sur cette droite sans forcement corres-

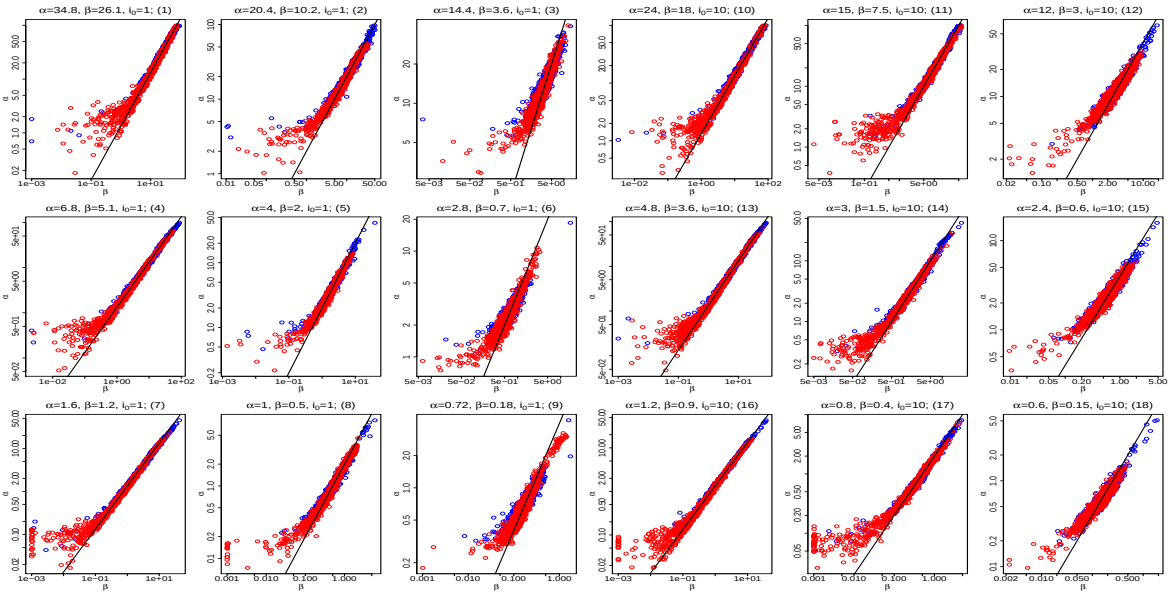

FIGURE 7.9: Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus

pondre au valeurs exactes α_0 et β_0. Ceci montre le problème d'identifiabilité des paramètres pour les modèles (M) et (S). Il s'avère impossible d'estimer ces deux paramètres α et β à la fois car il existe une corrélation quasi-fonctionnelle qui lie α et β.

D'autre part, on remarque surtout pour le modèle (M), qu'il y a un nombre important d'estimations de α et β qui s'écartent de la droite. Cela peut indiquer une sur-estimation de α ou une sous-estimation de β. Cependant, dans l'estimation du paramètre R_0, ce sont les trop petites valeurs de β qui sont la cause dans la sur-estimation de R_0. Ce phénomène est bien plus prononcé pour le modèle (M) que pour le modèle (S). Ceci suggère un problème de conditionnement de la fonction de vraisemblance comme fonction du paramètre β, mal identifié dans les deux modèles.

Pourtant la nature du paramètre β et son interprétation comme taux de guérison permet de supposer qu'il est connu ou au moins facile à estimer par le biais d'autres méthodes. Dans la section suivante, sont présentés les résultats d'estimations par maximum de vraisemblance avec le paramètre β supposé connu et fixe.

7.3.5 Résultats avec le paramètre β fixe

Dans un deuxième temps, seuls les paramètres α, i_0 et R_0 ont été estimés, le paramètre β étant fixé à sa vraie valeur β_0 utilisée dans la simulation des données. En fait, parmi les paramètres α et R_0, il suffit d'estimer un des deux car l'estimation de l'autre est déduite de la relation $\hat{R_0} = \hat{\alpha}/\beta_0$ et la variance par $var(\hat{R_0}) = var(\hat{\alpha})/\beta_0^2$. Les résultats des estimations sont donc équivalents à la constante multiplicative près. Seuls les résultats pour R_0 sont présentés ici.

Les mêmes méthodes d'optimisation ont été utilisées. Les critères d'exclusion décrits dans la section (7.3.1) ont été appliqués pour écarter les estimations susceptibles de ne pas réaliser les maxima recherchés. Les nombres d'estimations retenues dans l'analyse sont présentés dans le tableau (28) en annexe. Lors de cette analyse secondaire, beaucoup plus d'estimations ont pu être utilisées. Les résultats des estimations de R_0 et i_0 et leurs écarts-types sont présentés dans les box-plots (59)-(70) en l'annexe. Les propriétés asymptotiques sont illustrées sur les PP-plots (73)-(76) en annexe. Les résultats numériques des estimations sont mis dans les tableaux (29)-(34) également placés en annexe.

Globalement, les résultats des estimations sont très comparables pour les deux modèles. Les estimations du paramètre R_0 sont néanmoins meilleures en termes de biais et d'efficacité relative sous le modèle (S). Les écarts-types asymptotiques correspondent mieux aux écarts-types des estimations sous le modèle (M) (surtout dans les paramétrisations avec $i_0 = 10$) mais les propriétés asymptotiques (PP-plots) sont comparables pour les deux modèles. L'efficacité relative est systématiquement en faveur du modèle (S) mais reste pourtant proche de un. Les conclusions sont aussi valables pour le paramètre i_0 mais il faut souligner que dans le scénario avec les observations dans la phase stationnaire « after », les estimations de i_0 sont plutôt médiocres.

A titre d'exemple, les résultats des estimations du paramètre R_0 dans la phase évolutive « during » sont présentés ici.

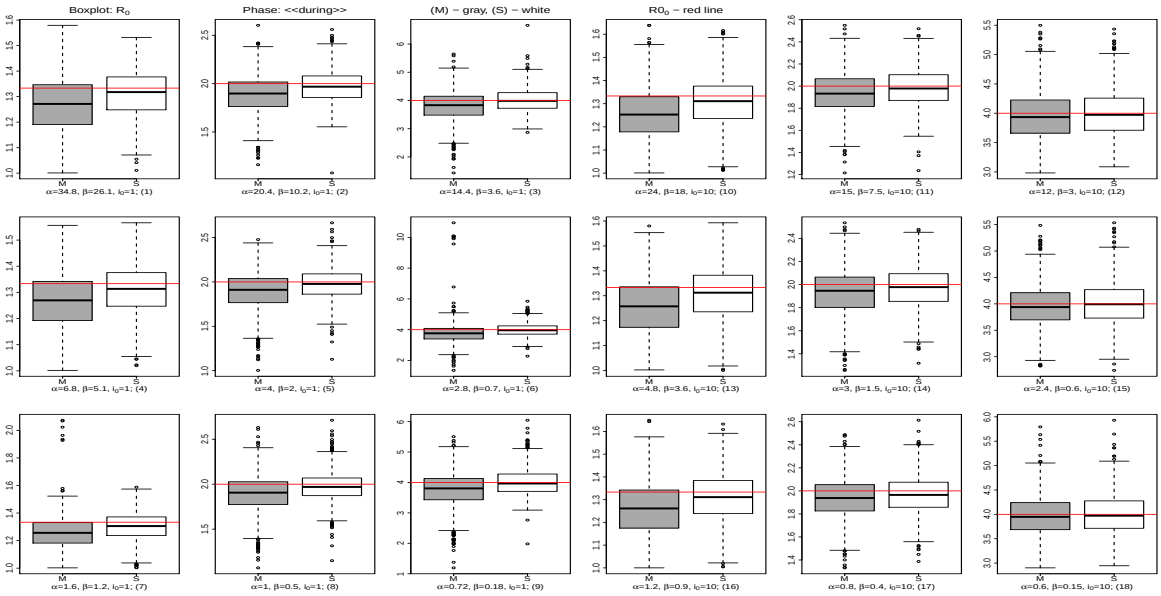

FIGURE 7.10: Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » (β-fixe)

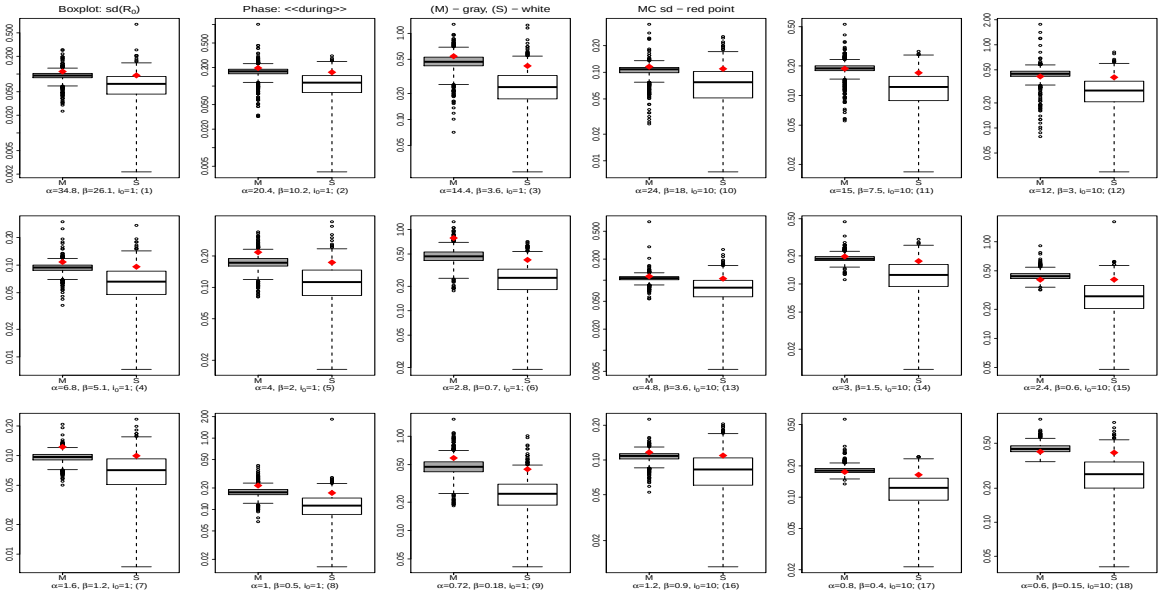

FIGURE 7.11: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » (β-fixe) ; (M) en rouge, (S) en bleu

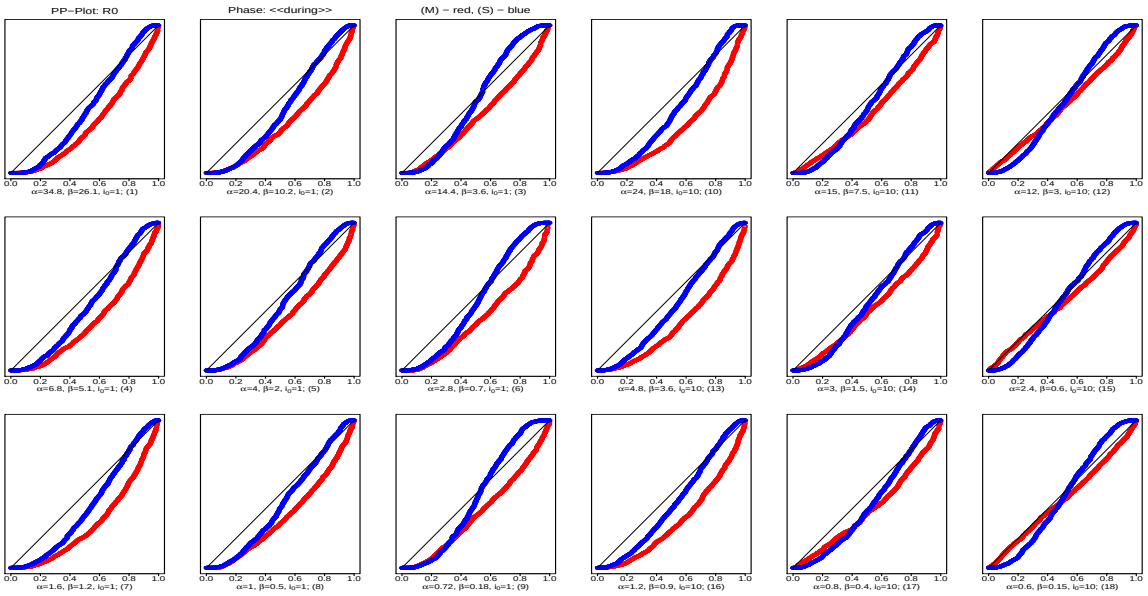

FIGURE 7.12: PP-plots de R_0 pour les observations de la phase évolutive : « during » (β-fixe)

TABLE 7.5: Efficacité relative d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE	-	**1.73**	-	**1.76**	-	**1.90**	-	**1.53**	-	**1.33**	-	**1.06**
	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE	-	**1.64**	-	**1.94**	-	**3.79**	-	**1.59**	-	**1.38**	-	1.00
	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
RE	-	**1.87**	-	**1.96**	-	**2.01**	-	**1.53**	-	**1.24**	-	**1.05**

7.4 Observations sous l'état stationnaire - Estimateurs de R_0

7.4.1 Résultats algébriques

Si les observations proviennent d'un état stationnaire, on peut étudier sous cette forme le modèle SIS stochastique de la section (6.5.2). En comparant à zéro les dérivées $\frac{\partial \phi}{\partial t}$ et $\frac{\partial \langle \xi^2 \rangle}{\partial t}$, dans les équations (6.63) et (6.66) respectivement, on obtient les solutions stationnaires du processus SIS stochastique. Pour noter qu'il s'agit de modèles sous état stationnaire, un indice s est ajouté.

$$\phi^s = 1 - \frac{\beta}{\alpha}, \quad \langle \xi^2 \rangle^s = \frac{\beta}{\alpha} \tag{7.3}$$

Dans ce cas, le processus du nombre de sujets susceptibles est gaussien de moyenne $\Omega(1 - \frac{\beta}{\alpha})$ et variance $\Omega\frac{\beta}{\alpha}$ où $\Omega = N$ est la taille de la population. Puisque la solution du modèle SIS déterministe $I(t)$ (6.55) et la solution (6.63) de l'équation $\phi(t)$ sont liées via la relation $I(t) = N\phi(t)$, les deux modèles de données (M) et (S) de la section (6.6) peuvent s'écrire dans l'état stationnaire de manière beaucoup plus simple :

$$\text{(M)} : \quad X_i \sim \mathcal{N}\left(N(1 - \frac{\beta}{\alpha}), N\frac{\beta}{\alpha}\right), \qquad \text{(S)} : \quad X_i \sim \mathcal{N}\left(N(1 - \frac{\beta}{\alpha}), \sigma^2\right) \tag{7.4}$$

$X = (X_i)_{i=1\ldots J}$ étant les nombres de sujets infectés observés aux temps $(t_1, \ldots, t_J)$. On supposera ici que les observations x_j sont suffisamment espacées pour être quasi-indépendantes. Il devient clair maintenant que les paramètres α et β ne sont pas identifiables car sont représentés uniquement par leur rapport. Ce rapport est un taux de reproduction R_0 et ainsi le modèle (M) peut s'écrit

$$X_i \sim \mathcal{N}\left(N(1 - \frac{1}{R_0}), N\frac{1}{R_0}\right) \tag{7.5}$$

et le modèle (S)

$$X_i \sim \mathcal{N}\left(N(1 - \frac{1}{R_0}), \sigma^2\right) \tag{7.6}$$

Si les observations proviennent de l'état stationnaire, l'estimation du paramètre i_0 n'a plus de sens.

On va démontrer que pour le modèle (M^s), l'estimateur du maximum de vraisemblance de R_0 est de la forme

$$\widehat{R_0} = NJ\frac{1+\sqrt{1+4\frac{1}{J}\sum_{i=1}^{J}(X_i-N)^2}}{2\sum_{i=1}^{J}(X_i-N)^2} \tag{7.7}$$

et qu'il est asymptotiquement sans biais. Il est aussi asymptotiquement efficace avec une variance

$$var(\widehat{R_0}) \overset{as}{\approx} \frac{2R_0^3}{J(R_0+2N)} \tag{7.8}$$

égale à la borne de Cramer-Rao.

On va démontrer ensuite que sous le modèle (S^s), l'estimateur du maximum de vraisemblance de R_0 est de la forme

$$\widetilde{R_0} = \frac{N}{N-\bar{X}} \tag{7.9}$$

où $\bar{x} = \sum_{i=1}^{J} X_i$ avec une variance asymptotique

$$var(\widetilde{R_0}) \overset{as}{\approx} \frac{R_0^4\sigma^2}{JN^2} \tag{7.10}$$

L'estimateur du maximum de vraisemblance de σ^2 est

$$\widetilde{\sigma^2} = \frac{1}{J}\sum_{i=1}^{J} X_i^2 - \bar{X}^2 \tag{7.11}$$

Il n'est cependant pas sans biais car l'espérance $E(\widetilde{\sigma^2}) = \frac{J-1}{J}\sigma^2$. L'estimateur sans biais $\widetilde{\delta^2}$ déduit de $\widetilde{\sigma^2}$: $\widetilde{\delta^2} = \frac{J}{J-1}\widetilde{\sigma^2}$ n'est pas efficace. Comme illustré sur les graphiques (7.1), (7.2) et (7.3) dans la section (7.1), la solution de la variance obtenue via le développement (6.66) de l'équation maîtresse $N\langle\xi^2\rangle$ correspond de manière très adéquate à la variance des 1000 trajectoires du processus de naissance et de mort surtout dans sa phase stationnaire. Intuitivement les modèles (M^s) et (S^s) devraient être identiques pour l'état stationnaire. Les moyennes des deux modèles sont identiques. Les variances seront aussi les mêmes si $\sigma^2 = \frac{N}{R_0}$. Dans ce cas, la variance (7.10) de l'estimateur $\widetilde{R_0}$ devient

$$Var(\widetilde{R_0}) = \frac{R_0^3}{JN} > \frac{R_0^3}{\frac{J}{2}R_0 + JN} \tag{7.12}$$

qui est plus grande que la variance (7.8) de l'estimateur $\widehat{R_0}$. Si on suppose que la variance stationnaire σ^2 est égale à $\frac{N}{R_0}$, on trouve que l'estimateur du maximum de vraisemblance

sous le modèle (M^s) est asymptotiquement plus efficace que l'estimateur du maximum de vraisemblance sous le modèle (S^s).

Nåsell (2002) via l'approximation de la diffusion (Kurtz, 1971; Barbour, 1976) arrive au même résultat (7.5) comme solution stationnaire du processus d'Ornstein-Uhlenbeck. La méthode de développement est valide pour $\alpha > \beta$ ce qui correspond au seuil déterministe $R_0 > 1$. Nåsell (1999) donne un argument pratique pour la validité de l'approximation normale qui relie la valeur R_0 avec la taille de la population N. La densité stationnaire présente trois comportements différents selon la valeur de R_0 : $R_0 > 1$, $R_0 < 1$ et $R_0 \approx 1$. Nåsell (1996) propose de reparamétrer R_0 par $R_0 = 1+\rho/\sqrt{N}$ pour mieux saisir ces trois régions. D'un point de vue pratique, Nåsell (1999) suggère d'identifier la région transitive là où R_0 est proche de 1 par $|\rho| = |N^{1/2}(R_0 - 1)| < 3$. Il justifie le choix de cette relation via une certaine particularité de la loi normale : une variable normale se situe avec la plus grand probabilité dans l'intervalle centré sur sa moyenne ± 3 fois son écart-type. Pour les valeurs R_0 nettement plus grandes que 1, la relation $|\rho| = |N^{1/2}(R_0 - 1)| > 3$ indique donc une bonne approximation normale de la loi stationnaire.

Les estimateurs sont appliqués aux données simulées décrites dans la section (7.2). La démonstration algébrique des propriétés des estimateurs de R_0 se trouve en annexe (IV).

7.4.2 Application aux données simulées

L'estimateur $\widehat{R_0}$ (7.7) avec la variance (7.8) et l'estimateur $\widetilde{R_0}$ (7.9) avec la variance (7.10) et avec la variance stationnaire $\widetilde{\sigma^2}$ (7.11) du modèle (S^s) ont été appliqués aux données simulées dans la section (7.2) provenant de la phase stationnaire. Les résultats des estimations sont présentés dans la table (7.6) et sur les graphiques (7.13) et (7.14). En analysant les résultats de la table (7.6) et des box-plots des estimations (7.13) obtenues, on constate que pour les deux modèles les estimations sont très comparables en termes de moyennes, biais, erreurs quadratiques moyennes et efficacités relatives. Cette dernière caractéristique étant systématiquement égale à 1. Cependant en examinant les écarts-types des estimations et les moyennes des écarts-types asymptotiques, on les trouve légèrement plus proches sous le modèle (M^s). Les box-plots (7.14) montrent que les écarts-types asymptotiques sont plus petits sous le

modèle (M^s). Dans les deux modèles, les écarts-types asymptotiques restent inférieurs aux écarts-types des estimations, surtout dans les paramétrisations avec $s = 25$ pour lesquelles la vraie valeur de R_0 est proche de 1 (1.33). Le modèle (M^s) semble être meilleur en terme d'estimations de l'écart-type asymptotique. (Les écarts-types ont été utilisés pour améliorer la lecture des box-plots). Les PP-plots (7.15) sont très comparables sous les deux modèles avec une très petite amélioration des propriétés asymptotiques sous le modèle (M^s). On peut observer une amélioration des propriétés asymptotiques quand la valeur de R_0 augmentent. En effet, dans le graphique (7.15), les PP-plots dans la première et la quatrième colonne correspondent aux paramétrisations pour lesquelles $R_0 = 1.33$, ceux dans la deuxième et la cinquième colonne aux paramétrisations avec $R_0 = 2$, et les PP-plots de la troisième et la sixième colonne correspondent aux paramétrisations avec $R_0 = 4$. Dans les paramétrisations avec $R_0 = 4$, les probabilités estimées et théoriques sont pratiquement alignées sur la diagonale. Dans ces cas, l'approximation normale est justifiée selon le critère de Nåsell (1999) mentionné ci-dessus.

TABLE 7.6: Résultats d'estimation par maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after » - nouveaux estimateurs de R_0 pour les observations dans l'état stationnaire

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
mean	1.30	1.29	1.97	1.96	3.98	3.97	1.29	1.29	1.97	1.97	3.99	3.97
bias	-0.04	-0.04	-0.03	-0.04	-0.02	-0.03	-0.04	-0.05	-0.03	-0.03	-0.01	-0.03
rel. bias	-0.03	-0.03	-0.02	-0.02	0.00	-0.01	-0.03	-0.03	-0.01	-0.02	0.00	-0.01
|bias|	0.07	0.07	0.12	0.12	0.30	0.30	0.08	0.08	0.12	0.12	0.31	0.31
sd	0.08	0.08	0.15	0.14	0.38	0.37	0.09	0.09	0.15	0.15	0.39	0.39
MSE	0.01	0.01	0.02	0.02	0.14	0.14	0.01	0.01	0.02	0.02	0.15	0.15
RE	1.00	0.98	1.00	1.00	1.00	1.04	1.00	0.97	1.00	1.00	1.00	1.01
Mean.as.sd	0.06	0.05	0.11	0.10	0.32	0.27	0.06	0.05	0.11	0.09	0.32	0.27

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
mean	1.29	1.29	1.97	1.97	3.99	3.97	1.29	1.28	1.97	1.97	4.02	4.00
bias	-0.04	-0.04	-0.03	-0.03	-0.01	-0.03	-0.05	-0.05	-0.03	-0.03	0.02	0.00
rel. bias	-0.03	-0.03	-0.01	-0.02	0.00	-0.01	-0.03	-0.04	-0.01	-0.02	0.00	0.00
|bias|	0.07	0.07	0.11	0.11	0.30	0.29	0.08	0.08	0.12	0.12	0.32	0.32
sd	0.08	0.08	0.14	0.14	0.38	0.37	0.09	0.09	0.15	0.15	0.41	0.40
MSE	0.01	0.01	0.02	0.02	0.14	0.14	0.01	0.01	0.02	0.02	0.17	0.16
RE	1.00	0.97	1.00	1.00	1.00	1.03	1.00	0.97	1.00	1.00	1.00	1.03
Mean.as.sd	0.06	0.05	0.11	0.10	0.32	0.27	0.06	0.04	0.11	0.09	0.33	0.27

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
mean	1.29	1.29	1.97	1.96	4.01	3.99	1.29	1.29	1.98	1.97	4.01	3.99
bias	-0.04	-0.05	-0.03	-0.04	0.01	-0.01	-0.04	-0.05	-0.02	-0.03	0.01	-0.01
rel. bias	-0.03	-0.04	-0.02	-0.02	0.00	0.00	-0.03	-0.03	-0.01	-0.01	0.00	0.00
|bias|	0.07	0.07	0.11	0.11	0.31	0.31	0.08	0.08	0.12	0.12	0.32	0.31
sd	0.08	0.08	0.14	0.14	0.39	0.39	0.09	0.09	0.15	0.15	0.40	0.39
MSE	0.01	0.01	0.02	0.02	0.15	0.15	0.01	0.01	0.02	0.02	0.16	0.15
RE	1.00	0.97	1.00	1.00	1.00	1.03	1.00	0.98	1.00	1.00	1.00	1.03
Mean.as.sd	0.06	0.05	0.11	0.09	0.33	0.28	0.06	0.04	0.11	0.09	0.33	0.27

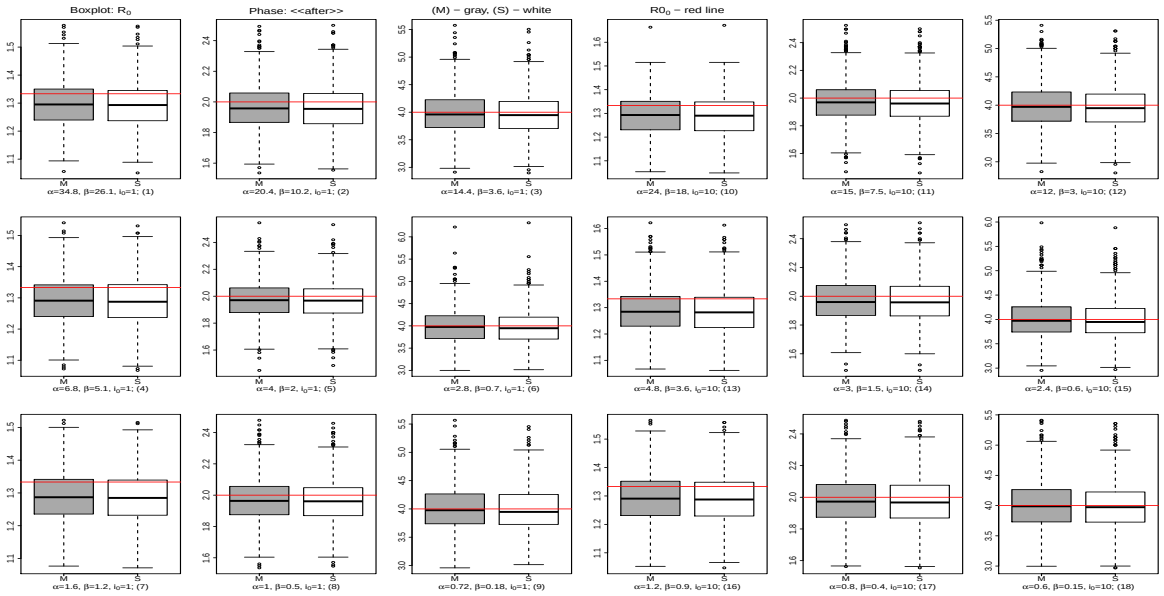

FIGURE 7.13: Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0 ; (M) en rouge, (S) en bleu

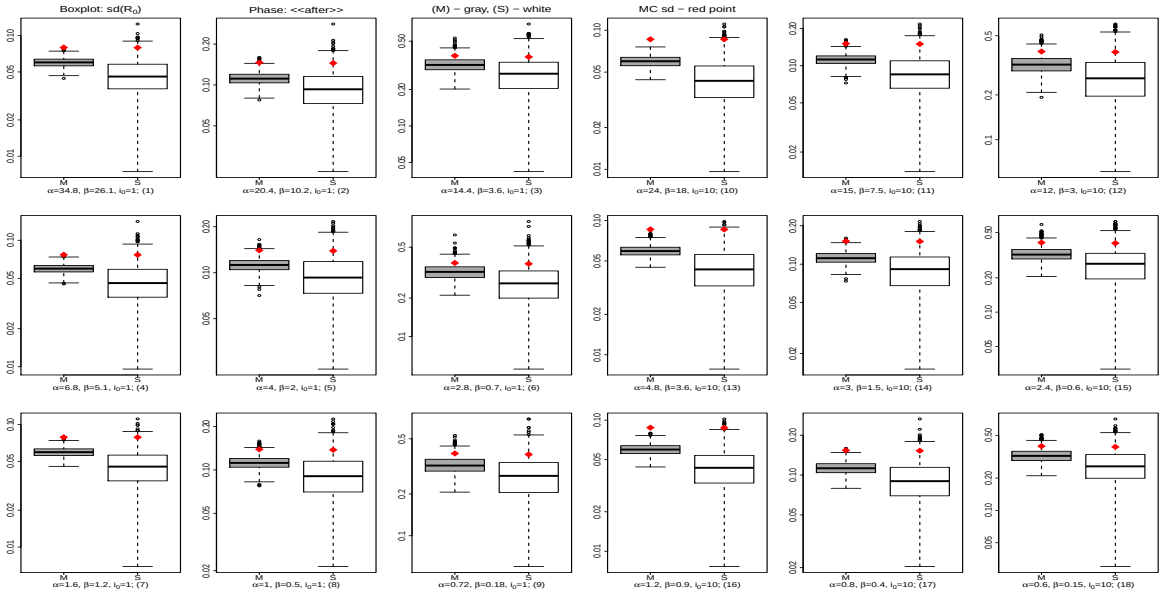

FIGURE 7.14: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0

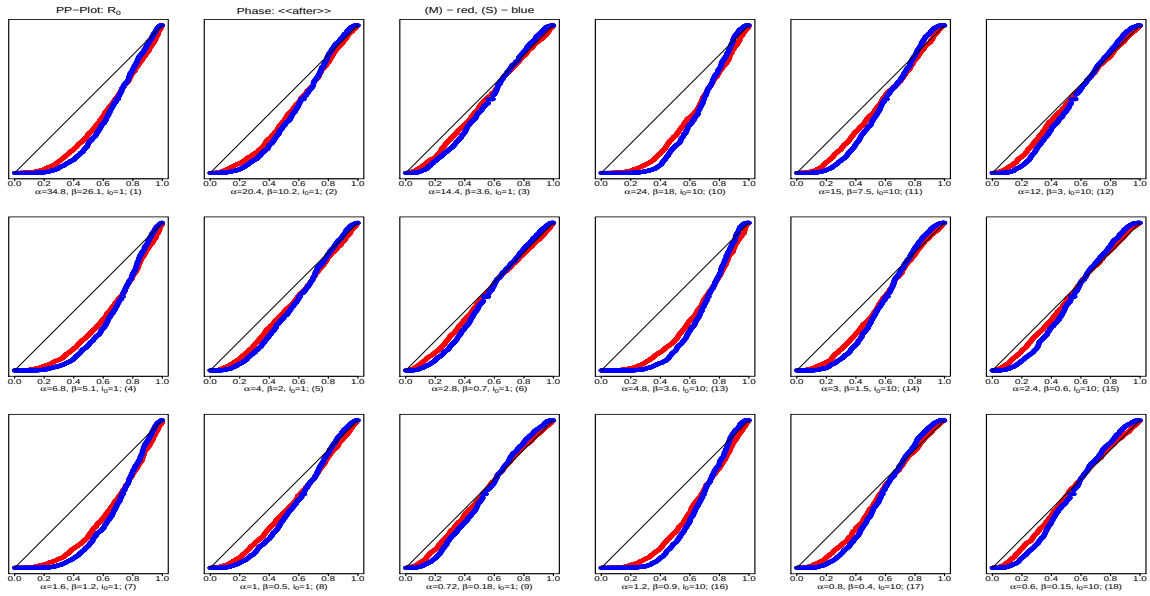

FIGURE 7.15: PP-plots de R_0 pour les observations de la phase stationnaire : « **after** » - nouveaux estimateurs de R_0

7.5 Résultats des estimations bayésiennes

L'estimation bayésienne a été effectuée seulement pour une sélection de paramétrisations. Selon le tableau (7.1), il s'agit de paramétrisations avec $t = 5$ (#4), (#5), (#6) et (#13), (#14), (#15) dans le scénario où les observations sont dans la phase évolutive « during ». Les vraisemblances des modèles (M) et (S) sont décrites dans la section (6.6).

7.5.1 Choix des lois a priori

Les lois a priori choisies pour ces modèles sont présentées dans le tableau (7.16).

Ces lois a priori ont été choisies informatives. La loi du paramètre β est une gaussienne avec une moyenne égale à la vraie valeur β_0 et avec une variance de 25 et tronquée sur l'intervalle $\left[10^{-3}, N\right]$. Dans la construction de la loi a priori sur α, il a été décidé que son espérance est le double de la vraie valeur $2\alpha_0$ mais avec une precision suffisamment grande pour que la loi a priori puisse couvrir la vraie valeur α_0. L'information apportée par la loi a priori suppose donc que le taux de contact est en moyenne plus grand que sa vraie valeur. En

FIGURE 7.16: Lois a priori pour les modèles (M) et (S)

$\alpha = \beta + \gamma$ pour garantir que $\alpha > \beta$
$\gamma \sim \mathcal{G}amma(\nu_1, \nu_1)$
$\beta \sim \mathcal{N}(\beta_0, \sigma_\beta^2) I(10^{-3}, N)$
$i_0 \sim \mathcal{P}oisson(\lambda) I(1, N)$
$\sigma^2 \sim \mathcal{IG}amma(\tau_1, \tau_2)$ pour le modèle (S)

FIGURE 7.17: Les paramètres de la loi a priori inverse gamma sur σ^2

	(#4) et (#13)	(#4) et (#14)	(#6) et (#15)
τ_1	4	3	2.5
τ_2	250	115	45

faisant l'inférence bayésienne, on s'attend à ce que l'information apportée par les données x permette de retrouver la vraie valeur moyenne de α par l'espérance a posteriori.

Un nouveau paramètre $\gamma = \alpha - \beta$ positif avec une loi a priori $\mathcal{G}amma(\nu_1, \nu_1)$ a été introduit dans les modèles pour garantir que $\alpha > \beta$. Les paramètres de la loi a priori de γ, ν_1 et ν_2, ont été choisis de manière à ce que l'espérance et la variance de γ sous cette loi soient $2\alpha_0 - \beta_0$ et 100 respectivement.

La loi a priori du paramètre i_0 a pour espérance la vraie valeur de i_0 augmentée de 4 dans les paramétrisations avec $i0 = 1$ (#4)-(#6) et de 5 dans les paramétrisations avec $i_0 = 10$ (#13)-(#15).

Pour le modèle (S), la loi a priori de la variance résiduelle σ^2 a été choisie cohérente avec les résultats de la modélisation de la section (7.1). La variance stationnaire du PNM du nombre d'infectés semble correspondre à la variance obtenue par la méthode de développement $N\beta/\alpha$. Cependant la variance dans la phase évolutive du processus est beaucoup plus grande. La loi a priori sur σ^2 a donc été choisie comme une loi inverse gamma dont la moyenne est proche de $N\beta_0/\alpha_0$ mais avec une queue de distribution couvrant les grandes valeurs. Les paramètres τ_1 et τ_2 de cette inverse gamma sont présentés dans le tableau (7.17).

7.5.2 Estimation des paramètres - utilisation de la stratégie de mélange

Les estimations bayésiennes ont été obtenues pour la totalité de 1000 jeux de données simulées. Il s'agit des mêmes données utilisées précédemment lors de l'inférence fréquentiste.

Les inférences obtenues classiquement par les méthodes MCMC ont été obtenues avec le logiciel WinBUGS en y ajoutant un module supplémentaire WBdiff qui traite les équations différentielles. WinBUGS a été lancé (en mode *batch*) sous R via le package R2WinBUGS. Pour chaque chaîne, 145000 itérations ont été faites avec un temps de chauffe fixé à 5000 (convergence atteinte). L'éclaircissement (*thin*) a été fixé a 20 pour réduire les dépendances entre les itérations. Ainsi, 7000 réalisations de la loi a posteriori sont retenues quasi-indépendantes. Pour chaque jeu de données x, le temps nécessaire pour obtenir une réalisation de la loi a posteriori est long. En effet, dans le cas de ces modèles épidémiques, l'évaluation de la vraisemblance est couteuse en temps car requiert l'évaluation de la solution des équations différentielles. Les améliorations algorithmiques présentées lors de la première partie ont donc leur rôle à jouer ici. Par contre, parmi les méthodes proposées, il est essentiel de choisir une méthode qui nécessite le moins possible le calcul de la vraisemblance. Nous avons donc décidé de retenir la stratégie de mélange (voir équation 2.45) où la fonction d'importance construite comme un mélange de densités est la même pour tous les jeux de données où l'Importance Sampling est utilisé.

Les algorithmes MCMC ont donc été employés seulement pour les 50 premiers jeux de données. Pour contrôler la convergence, il a été vérifié pour chaque paramètre que l'erreur de Monte Carlo de la moyenne a posteriori (MCerror) soit inférieure à 5% de l'écart-type des moyennes a posteriori. Des outils graphiques (history, autocorrélation) ont été utilisés pour voir si il y avait une bonne mélangeance des chaînes. Les estimations des paramètres sont les moyennes a posteriori, obtenues comme moyennes empiriques des 7000 itération avec une variance étant le carré de MCerror.

Pour effectuer les estimations sur les 950 autres jeux de données, un estimateur d'échantillonnage préférentiel a été utilisé avec la stratégie de mélange. La densité d'échantillonnage (ou la fonction d'importance) a été choisie comme le mélange des 50 densités a posteriori obtenues pour les 50 premiers jeux de données. Les constantes de normalisation ont été calculées par

la méthode de *reciprocal importance sampling* comme présentée dans le paragraphe (2.2.5.2) de la première partie. Pour calculer chaque constante, la densité instrumentale a été choisie comme un produit de lois normales de moyenne et de variance correspondant à la moyenne et la variance obtenues par MCMC pour chaque paramètre.

7.5.3 Analyse des résultats

Les résultats des estimations sont présentées sur les graphiques (77)-(84) et dans les tableaux (35)-(38) en annexe. Pour tous les paramètres, les estimations sous les deux modèles dans la plupart de cas indiquent le même comportement. Le paramètre α semble être mieux estimé en termes de biais et d'efficacité relative par le modèle (S) tandis que le paramètre β est mieux estimé sous le modèle (S) mais seulement pour les estimations avec l'échantillonnage préférentiel (IS). Les estimations avec MCMC sont meilleures sous le modèle (S) seulement pour la moitié des paramétrisations. Les estimations de i_0 sont sur-estimées par rapport aux vraie valeurs. Enfin R_0 est généralement mieux estimé sous le modèle (S). Cependant il y a certaines paramétrisations pour lesquelles les estimations avec IS sont très grandes sous le modèle (M). Pour ces mêmes paramétrisations, le paramètre β est souvent sous-estimé sous ce modèle. Les écart-types estimés sous-estiment l'écart-type des estimations sous les deux modèles et pour les deux types d'estimation. Les propriétés asymptotiques sont donc médiocres.

La comparaison entre les résultats des estimation avec MCMC et avec IS montre une certaine correspondance principalement sous le modèle (S).

Le graphique (85) en annexe illustre les estimations de α en fonction de β. On obtient la même relation entre les paramètres que celle observée lors de l'inférence fréquentiste. Les estimations sont partiellement alignées sur la droite $\alpha_0 = R_0/\beta_0$ et s'écartent de la droite pour les petites valeurs de β. On retrouve ainsi le problème d'identification du paramètre β

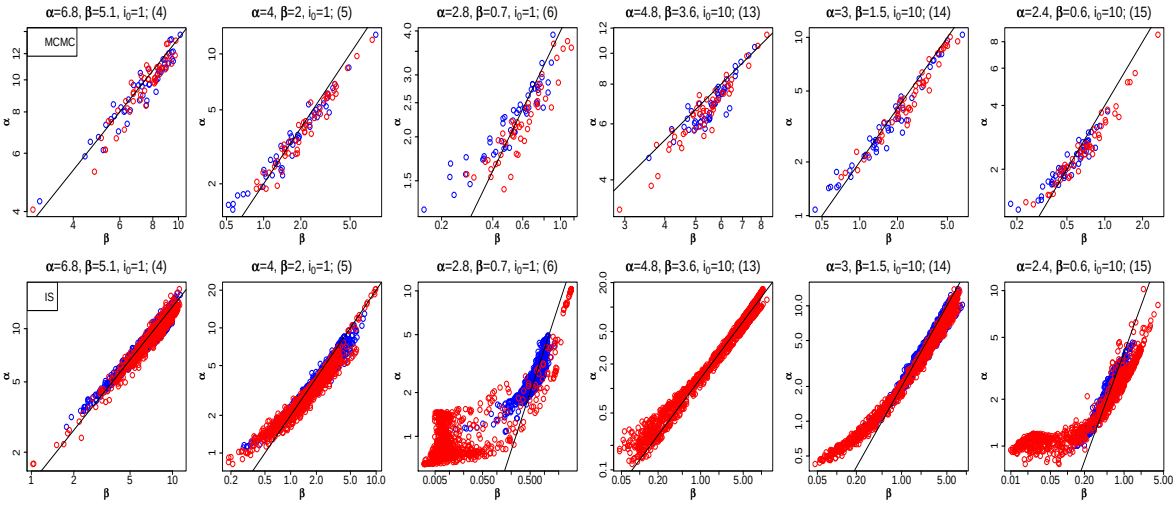

FIGURE 7.18: Les estimations bayésiennes de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during ». Résultats MCMC en première ligne, IS en deuxième ligne.

déjà évoqué précedemment.

8 Discussion

L'objectif de ce travail était de construire un modèle de transmission d'une maladie dans une population pouvant capturer l'hétérogénéité à partir des effectifs observés. Le modèle épidémique peut être représenté comme un modèle compartimental, les taux de transition entre les compartiments en étant les paramètres. Classiquement, un tel modèle s'exprime comme un système d'équations différentielles. Mais, afin de prendre en compte le caractère aléatoire du processus épidémique sous-jacent, on s'est dirigé vers un modèle stochastique. Il existe différentes façons d'incorporer une partie aléatoire dans le modèle déterministe : via une modélisation par chaîne de Markov, via une modélisation par équations différentielles stochastiques en y ajoutant d'un processus de Wiener ou via une modélisation bayésienne par

attribution à certains paramètres du modèle d'une distribution de probabilité.

Dans ce travail de thèse, seulement la première et la troisième possibilité ont été étudiées. Si le modèle compartimental est décrit comme une chaîne de Markov, la probabilité de transition depuis et vers chaque état est donnée par une équation maîtresse étant elle même une équation différentielle. La solution de cette équation n'est possible que si les taux de transition sont des fonctions linéaires des effectifs. Si les taux de transition sont non-linéaires (situations réalistes relativement fréquentes), une solution approchée est alors indispensable. La solution étudiée ici est obtenue via la méthode de développement de l'équation maîtresse de Van Kampen (2007).

Initialement, il était prévu d'étudier les modèles infectieux muti-états. Cependant la complexité de la procédure de développement nous a orienté vers un modèle épidémique plus simple. Il s'agit d'un modèle SIS (susceptible-infecté-susceptible) qui se caractérise seulement par les deux états et s'écrit avec deux paramètres de transition : α - taux de contacts infectieux et β - taux de guérison.

Pour le modèle SIS représenté comme une chaîne de Markov, le développement de l'équation maîtresse associée donne une solution gaussienne. La moyenne et la variance de cette gaussienne sont décrites par un système d'equations différentielles. L'équation sur la moyenne est équivalente à l'équation du nombre de sujets infectés du modèle déterministe.

La méthode de développement est particulièrement bien adaptée pour le processus avec le paramètre α plus grand que β et quand le nombre initial d'infectés est suffisamment grand. Une série de simulations de chaîne de Markov du processus SIS a montré que leur moyenne et leur variance sont représentées de manière très fidèle par les solutions obtenues par le développement (ceci étant vrai également pour la solution de la moyenne du modèle déterministe).

Pour étudier les performances des estimations des paramètres du modèle gaussien avec moyenne et variance obtenues par le développement de l'équation maîtresse, une étude de simulation à été mise en place. Puisque la moyenne de ce modèle noté (M) correspond à la moyenne du modèle déterministe (deuxième modèle noté (S)), il a été construit avec la même moyenne et une variance résiduelle constante.

Différents scénarios sous les différentes paramétrisations ont été envisagés.

Pour placer les analyses dans un contexte réaliste, des jeux de données de petite taille ont été construits à partir des réalisations de chaînes de Markov imitant l'épidémie. Chaque jeu de données est constitué de 6 couples (date/nombre de sujets infectés) observés à des instants équidistants. Les données correspondent donc à des mesures répétées dans la population lors d'une enquête. Un total de 1000 jeux de données a été simulé.

Une analyse fréquentiste a été réalisée montrant que le modèle plus simple (S) est souvent plus performant que le modèle (M) pour l'estimation de R_0. C'est l'inverse pour les estimations de α et β.

Lors des inférences par maximum de vraisemblance, de nombreuses estimations n'ont pas pu être qualifiées comme les extrêma recherchés suite à des problèmes sans doute numériques. Un problème d'identifiabilité pour les paramètres α et β a été détecté. De plus, une partie des estimations de β est sous-estimée, suggérant que les modèles sont mal conditionnés pour ce paramètre. Dans un deuxième temps, les modèles avec le paramètres β fixé ont été étudiés montrant que le modèle (S) est toujours plus performant dans l'estimation de R_0.

L'écart-type asymptotique est souvent mieux estimé sous le modèle (M) mais les résultats en terme de biais et d'erreurs quadratiques moyennes peuvent être plus défavorables que sous le modèle (S).

Le modèle (S) est surtout meilleur dans les scénarios quand les observations proviennent de la phase initiale ou évolutive. Quand les observation proviennent de la phase (quasi-)stationnaire, les deux modèles sont dans l'ensemble très comparables.

Pour cette raison, les modèle (M) et (S) ont été étudiés dans l'état stationnaire. On a procédé de manière différente en prenant les solutions stationnaires des équations différentielles de la moyenne et la variance. Elles ont alors une forme explicite et dans l'écriture de la vraisemblance, les paramètres α et β existent seulement à travers le rapport de α/β qui vaut R_0. Les estimateurs du maximum de vraisemblance de R_0 s'obtiennent directement. Les estimations sont à nouveau très comparables et proches des vraies valeurs fixées dans les simulations mais les écarts-types des estimateurs sont sous-estimés. Cet exemple montre clairement que les paramètres α et β ne peuvent pas être estimés ensembles sous le modèle SIS.

L'inférence bayésienne a été effectuée pour les modèles avec les deux paramètre α et β et seulement pour une partie des paramétrisations. L'intégralité des estimations est obtenue via les méthodes MCMC et IS combiné avec MCMC développées dans la première partie de la thèse. Dans les modèles étudiés, les méthodes MCMC sont très longues et l'approche combinée a permis d'accélérer les calculs. Les résultats de l'estimation bayésienne sont compatibles avec ceux de l'estimation par maximum de vraisemblance. Le modèle (S) semble être meilleur en terme de biais et d'efficacité relative. Les écarts-types sont sous-estimés et les propriétés asymptotiques sont médiocres pour les deux modèles. Les conclusions obtenues dans l'inférence avec MCMC sont comparables avec les conclusions apportées par la stratégie combinée.

L'analyse bayésienne effectuée dans cette partie a attiré notre attention sur la construction des stratégies combinées. Les deux modèles ont une vraisemblance complexe à évaluer à cause des équations différentielles même si la structure hiérarchique est très simple. La plus grande difficulté qui se traduit par une durée importante des algorithmes MCMC, est l'évaluation de la solution d'équations différentielles lors de chaque itération de la chaîne. Cette évaluation de la vraisemblance est ici coûteuse en temps. Dans ce cas particulier, les stratégies combinées avec l'échantillonnage qui nécessitent le calcul successif du rapport des vraisemblances peuvent s'avérer longues aussi et donc un mauvais choix. Parmi les stratégies mises en place, celle du mélange des densités a posteriori est la plus adaptée. L'éclaircissement (subsampling) a permis de réduire également le nombre d'itérations tout en gardant une bonne représentation du mélange des densités.

Une analyse supplémentaire avec le paramètre β fixe doit être encore effectuée dans le cadre bayésien pour comparer les deux modèles en terme d'amélioration des estimations par rapport aux vraies valeurs mais aussi pour comparer les performances des méthodes MCMC par rapport aux méthodes combinées.

Le paramètre le plus intéressant est le taux de reproduction R_0 et la sous-estimation du paramètre β affecte de manière très importante son estimation. Il semble dont prudent d'utiliser d'autres procédures pour évaluer le paramètre β avant de l'introduire dans le modèle SIS. Le paramètre i_0 est un paramètre de nuisance et n'a pas d'intérêt épidémique mais lors du développement de l'équation maîtresse, de petites valeurs i_0 peut augmenter la variance

du processus autrement bien représentée.

Dans le cas d'estimation fréquentiste, il pourrait être intéressant d'utiliser les techniques numériques d'estimations plus efficaces basées sur les vraisemblances pénalisées (Ramsay et al., 2007).

Dans le cadre d'une recherche plus approfondie, les comparaisons avec d'autres méthodes comme le *moment closure* (Gillespie, 2009) et les méthodes avec les équations différentielles stochastiques seront très intéressantes aussi bien dans la modélisation que pour un ajustement du modèle afin de pouvoir finalement élargir les conclusions aux modèles plus complexes.

Une dernière piste de recherche est la modélisation des extinctions précoces du modèle stochastique.

Conclusion générale

La première partie de ce travail sur l'optimisation algorithmique a permis d'établir de bons résultats (en terme de réduction de temps calcul et en terme de précision des estimations) des approches qui combinent les algorithmes MCMC et les techniques d'Importance Sampling. Différentes méthodes combinées ont été présentées avec leurs performances respectives sur des modèles simples de type poissonnien. De manière générale, principalement deux méthodes nous semblent les plus performantes : celle dite de la "référence choisie" où pour chaque jeu de données une fonction d'importance est choisie comme la plus adéquate ou celle dite de "mélange" où la fonction d'importance est la même pour tous les jeux de données et est construite comme un mélange de lois a posteriori. Cette approche par mélange nécessiterait des améliorations. Tout d'abord, elle requiert le calcul de constantes de normalisation qui peut être difficile à faire et couteux en temps. D'autres techniques que celles présentées dans ce mémoire sont envisageables. Ensuite, les poids des composantes du mélange ont été ici choisis comme égaux. On peut étendre ce travail en introduisant la possibilité d'avoir des poids inégaux en les choisissant de manière à optimiser les estimations. Néanmoins, ces extensions doivent être faites en gardant le soucis de ne pas alourdir trop les temps de calculs.

La seconde partie a permis l'étude des modèles simples de type SIS sous leur forme stochastique. Malgré la forme simple de tels modèles, les solutions analytiques ne sont pas accessibles quand les taux de transition s'expriment sous une forme non linéaire des états. Une solution approchée est alors indispensable. La solution étudiée ici est obtenue via la méthode de développement de l'équation maîtresse. Cette procédure s'est avérée parfois complexe même dans le cadre de modèles SIS. Nous avons montré que la qualité des estimations des paramètres dépend de certaines conditions comme la taille du nombre initial d'infectés. Pour étudier les performances des estimations des paramètres, une étude par simulations à été mise en place où, pour chaque situation, un total de 1000 jeux de données a été simulé. Les qualités des estimateurs ont été étudiées dans un cadre fréquentiste et Bayésien. Il a été particulièrement intéressant dans ce contexte d'utiliser les accélérations algorithmiques proposées lors de la première partie. En effet, les méthodes MCMC se sont avérées très longues et l'approche combinée (MCMC et Importance Sampling) a permis d'accélérer les calculs. Dans le cadre de ces modèles, l'évaluation de la vraisemblance est coûteuse en temps. Ce point est intéressant car

distingue cette application des applications faites lors de la première partie où les modèles poissonniens considérés ne présentaient aucun problème de calcul de la vraisemblance. La lourdeur du calcul de la vraisemblance guide le choix entre les différentes méthodes proposées dans la première partie. En particulier, la méthode dite de "référence choisie" est à écarter car nécessite trop de calculs de la vraisemblance. Nous avons donc retenue ici la méthode dite de "mélange" qui est la plus adaptée car une seule fonction d'importance est utilisée pour l'ensemble des jeux de données. Les résultats de l'estimation bayésienne ont été en général compatibles avec ceux de l'estimation par maximum de vraisemblance. Dans le cas des estimations fréquentistes, il sera intéressant d'utiliser des techniques numériques d'estimations plus efficaces que celles explorées dans ce travail.

Bibliographie

Allen, L. J. and Burgin, A. M. (2000). Comparison of deterministic and stochastic sis and sir models in discrete time. *Mathematical Biosciences*, 163 :1–33.

Alonso, D. and McKane, A. (2002). Extinction dynamics in mainland-island metapopulations : An n-patch stochastic model. *Bulletin of Mathematical Biology*, 64(5) :913 – 958.

Alonso, D., McKane, A. J., and Pascual, M. (2007). Stochastic amplification in epidemics. *Journal of The Royal Society Interface*, 4(14) :575–582.

Anderson, R. M. and May, R. M. (1992). *Infectious Diseases of Humans. Dynamics and Control*. Oxford University Press, USA.

Asmussen, S., Kroese, D. P., and Rubinstein, R. Y. (2005). Heavy tails, importance sampling and cross-entropy. *Stochastic Models*, 21(1) :57–76.

Barbour, A. D. (1976). Quasi-stationary distributions in markov population processes. *Advances in Applied Probability*, 8(2) :296–314.

Breslow, N. E. (1984). Extra-poisson variation in log-linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(1) :38–44.

Brooks, S. P. and Roberts, G. O. (1998). Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing*, 8 :319–335.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(6) :1190–1208.

Cappé, O., Douc, R., Guillin, A., Marin, J. M., and Robert, C. P. (2008). Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18(4) :447–459.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Press.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3) :167–174.

Chen, M.-H. (1994). Importance-weighted marginal bayesian posterior density estimation. *Journal of the American Statistical Association*, 89(427) :818–824.

Chen, M.-H. and Shao, Q.-M. (1997). On monte carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics*, 25(4) :1563–1594.

Chen, W.-Y. and Bokka, S. (2005). Stochastic modeling of nonlinear epidemiology. *Journal of Theoretical Biology*, 234(4) :455 – 470.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics : A comparative review. *Journal of the American Statistical Association*, 91(434) :883–904.

de Valpine, P. (2008). Improved estimation of normalizing constants from markov chain monte carlo output. *Journal of Computational and Graphical Statistics*, 17(2) :333–351.

Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag.

DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997). Computing bayes factors by combining simulation and asymptotic approximations. *The Annals of Statistics*, 25(4) :1563–1594.

Diekmann, O., Heesterbeek, J. A., and Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4) :365–382.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research.*, 2(1) :23–41.

Doss, H. (1994). Discussion of the paper "markov chains for exploring posterior distributions" by luke tierney. *The Annals of Statistics*, 22(4) :1728–1734.

Doss, H. (2010). Estimation of large families of bayes factors from markov chain output. *Statistica Sinica*, 20 :537–560.

Douc, R., Guillin, A., Marin, J. M., and Robert, C. P. (2007). Minimum variance importance sampling via population monte carlo. *ESAIM : P&S 11 (2007)*, 11 :427–447.

Gajda, D., Guihenneuc-Jouyaux, C., and Rousseau, J. (2011). Melange de distributions comme fonction d'importance dans l'echantillonnagee preferentiel combine avec l'algorithme de monte carlo par chaine de markov. *Revue des Nouvelles Technologies de l'Information (RNTI)*, S-1, le numéro special JDS2009. :59–71.

Gajda, D., Guihenneuc-Jouyaux, C., Rousseau, J., Mengersen, K., and Nur, D. (2010). Use in practice of importance sampling for repeated mcmc for poisson models. *Electronic Journal of Statistics*, 4 :361–383.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice : Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3) :501–514.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412) :972–985.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410) :398–409.

Gelman, A. and Meng, X. L. (1994). Path sampling for computing normalizing constants : identities and theory. Technical Report 376, Dept. Statistics, Univ. Chicago.

Gelman, A. and Meng, X. L. (1998). Simulating normalizing constants : From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2) :163–185.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 :721–740.

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6) :1317–1339.

Geyer, C. J. (1993). Estimating normalizing constants and reweighting mixtures in markov chain monte carlo. Technical report, School of Statistics, University of Minnesota.

Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. D. and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 169–194. Clarendon Press, Oxford, UK.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995a). Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4) :455–472.

Gilks, W. R., Richardson, S., and Spiegelhalter (1995b). *Markov Chain Monte Carlo in Practice : Interdisciplinary Statistics*. Chapman and Hall/CRC, 1 edition.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2) :337–348.

Gillespie, C. S. (2009). Moment-closure approximations for mass-action models. *IET Systems Biology*, 3(1) :52–58.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25) :2340–2361.

Gómez-Uribe, C. A. and Verghese, G. C. (2007). Mass fluctuation kinetics : Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *The Journal of Chemical Physics*, 126(2) :024109.

Goodman, L. A. (1953). Population growth of the sexes. *Biometrics*, 9(2) :212–225.

Goutsias, J. (2007). Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophysical Journal*, 92(7) :2350–2365.

Gustafson, P. and Wasserman, L. (1995). Local sensitivity diagnostics for Bayesian inference. *The Annals of Statistics*, 23(6) :2153–2167.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1) :97–109.

Jensen, S. T., Johansen, S., and Lauritzen, S. L. (1991). Globally convergent algorithms for maximizing a likelihood function. *Biometrika*, 78(4) :867–877.

Johnson, N. l., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume Vol. 1-2. John Wiley & Sons, 2nd edition.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding groups in data : An introduction to cluster analysis*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics.

Keeling, M. J. (2000). Multiplicative moments and measures of persistence in ecology. *Journal of Theoretical Biology*, 205(2) :269–281.

Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological),*, 11(2) :230–282.

Krishnarajah, I., Cook, A., Marion, G., and Gibson, G. (2005). Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology*, 67(4) :855–873.

Kryscio, R. J. and Lefévre, C. (1989). On the extinction of the s-i-s stochastic logistic epidemic. *Journal of Applied Probability*, 26(4) :685–694.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7(1) :49–58.

Kurtz, T. G. (1971). Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2) :344–356.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs - a bayesian modelling framework : Concepts, structure, and extensibility. *Statistics and Computing*, 10 :325–337.

Maechler, M., Rousseeuw, P., Struyf, A., and Hubert, M. (2005). Cluster analysis basics and extensions. Rousseeuw et al provided the S original which has been ported to R by Kurt Hornik and has since been enhanced by Martin Maechler : speed improvements, silhouette() functionality, bug fixes, etc. See the 'Changelog' file (in the package source).

Matis, J. H. and Kiffe, T. R. (1996). On approximating the moments of the equilibrium distributions of a stochastic logistic model. *Biometrics*, 52 :980–991.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity : A theoretical exploration. *Statistica Sinica*, 6 :831–860.

Mengersen, K. L., Robert, C. P., and Guihenneuc-Jouyaux, C. (1999). Mcmc convergence diagnostics : a reviewww. In J. M. Bernardo, J. O. Berger, A. P. D. and Smith, A. F. M., editors, *Bayesian Statistics 6*, pages 415–440. Oxford Press, Oxford, UK.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6) :1087–1092.

Michael, J. R. (1983). The stabilized probability plot. *Biometrika*, 70(1) :11–17.

Nakache, J. P. and Confais, J. (2005). *Approche pragmatique de la classification.* Technip.

Neal, R. M. (1997). Markov chain monte carlo methods based on 'slicing' the density function. Technical report, Department of Statistics, University of Toronto.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3) :705–767.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4) :308–313.

Ng, R. and Han, J. (1994). Efficient and effective clustering methods for spatial data mining. In *Proceedings of the 20th Conference on VLDB, Santiago, Chili*, pages 144–155.

Nåsell, I. (1996). The quasi-stationary distribution of the closed endemic sis model. *Advances in Applied Probability*, 28(3) :895–932.

Nåsell, I. (1999). On the time to extinction in recurrent epidemics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(2) :309–330.

Nåsell, I. (2002). Stochastic models of some endemic infections. *Mathematical Biosciences*, 179(1) :1–19.

Nåsell, I. (2003). Moment closure and the stochastic logistic model. *Theoretical Population Biology*, 63(2) :159–168.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Wiley series in computational statistics. Wiley.

Picard, P. (1965). Sur les modèles stochastiques logistiques en démographie. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 2(2) :151–172.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations : a generalized smoothing approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69(5) :741–796.

Ripley, B. (1987). *Stochastic simulation*. Wiley Series in Probability and Statistics. J. Wiley.

Robert, C. P. (2007). *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer, 2 edition.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2nd. edition.

Robert, C. P. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Use R ! Springer.

Rubinstein, R. Y. and Kroese, D. P. (2004a). *The cross-entropy method : A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning.* Information Science and Statistics. Springer-Verlag, New-York.

Rubinstein, R. Y. and Kroese, D. P. (2004b). *The cross-entropy method : A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning.* Information Science and Statistics. Springer-Verlag, New York.

Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo method.* Wiley series in probability and mathematical statistics. Probability and mathematical statistics. John Wiley & Sons.

Singh, A. and Hespanha, J. (2007). A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of Mathematical Biology*, 69 :1909–1925.

Stollenwerk, N. and Briggs, K. M. (2000). Master equation solution of a plant disease model. *Physics Letters A*, 274 :84–91.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22 :1701–1762.

Van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry.* North Holland, 1st edition.

Van Kampen, N. G. (2007). *Stochastic processes in physics and chemistry.* North Holland, 3rd edition.

Whittle, P. (1957). On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 19(2) :268–281.

Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1) :1–17.

Woo, K. G., Lee, J. H., Kim, M. H., and Lee, Y. I. (2004). Findit : a fast and intelligent subspace clustering algorithm using dimension voting. *Informations & Software Technology*, 46(4) :255–271.

Annexe de la modélisation bayésienne

Démonstration de la convergence vers la loi normale et du terme de variance de l'estimateur IS "normalisé"

Convergence vers la loi normale

On montre simplement que

$$\bar{g}_m^{SNIS} - E_f\left[g(\theta)\right] = \frac{\frac{1}{m}\sum_{j=1}^{m}\left(g(\theta_j)\tilde{w}(\theta_j) - E_f\left[g(\theta)\right]\tilde{w}(\theta_j)\right)}{1} \times \frac{1}{\frac{1}{m}\sum_{j=1}^{m}\tilde{w}(\theta_j)} \quad (1)$$

La loi forte des grands nombres établit la convergence presque sûre des moyennes suivantes

$$\frac{1}{m}\sum_{j=1}^{m} g(\theta_j)\tilde{w}(\theta_j) \xrightarrow{p.s.} cdE_f\left[g(\theta)\right] \quad \text{et} \quad \frac{1}{m}\sum_{j=1}^{m}\tilde{w}(\theta_j) \xrightarrow{p.s.} cd \quad (2)$$

D'après le théorème centrale limite, le premier facteur de (1) converge en loi vers une variable normale de moyenne 0 avec une certaine variance. Le deuxième facteur converge presque sûrement vers $\frac{1}{cd}$ (convergence en probabilité). Ainsi, le produit converge en loi vers une variable normale de moyenne nulle et de variance : $var_h\left[\frac{\frac{1}{m}\sum_{j=1}^{m}\left(g(\theta_j)\tilde{w}(\theta_j) - E_f[g(\theta)]\tilde{w}(\theta_j)\right)}{cd}\right]$. On montre que $var_h\left[\frac{\frac{1}{m}\sum_{j=1}^{m}\left(g(\theta_j)\tilde{w}(\theta_j) - E_f[g(\theta)]\tilde{w}(\theta_j)\right)}{cd}\right] = \frac{1}{m}\frac{1}{c^2d^2}\int\left(g(\theta) - E_f\left[g(\theta)\right]\right)^2\tilde{w}^2(\theta)h(\theta)d\theta$.

Terme de variance

$$
\begin{aligned}
& var_h\left[\frac{\frac{1}{m}\sum_{j=1}^{m}\left(g(\theta_j)\tilde{w}(\theta_j)-E_f\left[g(\theta)\right]\tilde{w}(\theta_j)\right)}{cd}\right] \\
&= \frac{1}{m^2}\frac{m}{c^2d^2}var_h\left[g(\theta)\tilde{w}(\theta)-E_f\left[g(\theta)\right]\tilde{w}(\theta)\right] \\
&= \frac{1}{m}\frac{1}{c^2d^2}\Big\{var_h\left[g(\theta)\tilde{w}(\theta)\right]+E_f\left[g(\theta)\right]^2 var_h\left[\tilde{w}(\theta)\right] \\
&-2E_f\left[g(\theta)\right]cov_h\left[g(\theta)\tilde{w}(\theta);\tilde{w}(\theta)\right]\Big\} \qquad (3)
\end{aligned}
$$

. *On remarque que* $E_h\left[\tilde{w}(\theta)\right]=cd$ *et* $E_h\left[g(\theta)\tilde{w}(\theta)\right]=cdE_f\left[g(\theta)\right]$ *puis on développe (4).*

$$
\begin{aligned}
& \frac{1}{m}\frac{1}{c^2d^2}\Big\{var_h\left[g(\theta)\tilde{w}(\theta)\right]+E_f\left[g(\theta)\right]^2 var_h\left[\tilde{w}(\theta)\right]-2E_f\left[g(\theta)\right]cov_h\left[g(\theta)\tilde{w}(\theta);\tilde{w}(\theta)\right]\Big\} \\
&= \frac{1}{m}\frac{1}{c^2d^2}\Big\{E_h\left[g^2(\theta)\tilde{w}(\theta)\right]-E_h^2\left[g(\theta)\tilde{w}(\theta)\right] \\
&\quad +E_f^2\left[g(\theta)\right]E_h\left[\tilde{w}^2(\theta)\right]-E_f^2\left[g(\theta)\right]E_h^2\left[\tilde{w}(\theta)\right] \\
&\quad -2E_f\left[g(\theta)\right]E_h\left[g(\theta)\tilde{w}^2(\theta)\right]+2E_f\left[g(\theta)\right]E_h\left[g(\theta)\tilde{w}(\theta)E_h\left[\tilde{w}(\theta)\right]\right]\Big\} \\
&= \frac{1}{m}\frac{1}{c^2d^2}\Big\{E_h\left[g^2(\theta)\tilde{w}^2(\theta)\right]-c^2d^2E_f^2\left[g(\theta)\right]+E_f^2\left[g(\theta)\right]E_h\left[\tilde{w}^2(\theta)\right] \\
&\quad -E_h^2\left[g(\theta)\right]c^2d^2-2E_f\left[g(\theta)\right]E_h\left[g(\theta)\tilde{w}^2(\theta)\right]+2E_f^2\left[g(\theta)\right]c^2d^2\Big\} \\
&= \frac{1}{m}\frac{1}{c^2d^2}\Big\{E_h\left[g^2(\theta)\tilde{w}^2(\theta)\right]+E_f^2\left[g(\theta)\right]E_h\left[g(\theta)\tilde{w}^2(\theta)\right]-2E_f\left[g(\theta)\right]E_h\left[g(\theta)\tilde{w}^2(\theta)\right]\Big\} \\
&= \frac{1}{m}\frac{1}{c^2d^2}E_h\left[g^2(\theta)\tilde{w}^2(\theta)+E_f^2\left[g(\theta)\right]\tilde{w}^2(\theta)-2E_f\left[g(\theta)\right]g(\theta)\tilde{w}^2(\theta)\right] \\
&= \frac{1}{m}\frac{1}{c^2d^2}E_h\left[\Big(g^2(\theta)+E_f^2\left[g(\theta)\right]-2E_f\left[g(\theta)\right]g(\theta)\Big)\tilde{w}^2(\theta)\right] \\
&= \frac{1}{m}\frac{1}{c^2d^2}E_h\left[\left(g(\theta)-E_f\left[g(\theta)\right]\right)^2\tilde{w}^2(\theta)\right]=\frac{1}{m}\frac{1}{c^2d^2}\int\left(g(\theta)-E_f\left[g(\theta)\right]\right)^2\tilde{w}^2(\theta)h(\theta)d\theta
\end{aligned}
$$

Démonstration de la convergence de l'estimateur IS autonormalisé

On peut écrire le rapport intervenant dans l'approximation (2.37) comme

$$
\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})}\cdot\frac{\pi(x^{(1)})}{\pi(x^{(2)})}=\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})}\cdot\frac{1}{C} \qquad (4)
$$

où $C=\frac{\pi(x^{(2)})}{\pi(x^{(1)})}$ peut être considerée comme une constante car elle ne depend pas de θ.

D'après cette formule (4), l'approximation (2.37) de $E_{[\theta|X^{(2)}]}[g(\theta)]$ peut alors s'écrire

$$
\frac{1}{M}\sum_{i=1}^{M}g(\theta_i^{(1)})\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})}\cdot\frac{\pi(x^{(1)})}{\pi(x^{(2)})}=\frac{1}{CM}\cdot\sum_{i=1}^{M}g(\theta_i^{(1)})\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})} \qquad (5)
$$

Comme $\pi(\theta|x^{(2)})$ est une densité de probabilité, alors

$$\int \pi(\theta|x^{(2)})d\theta = 1 \Leftrightarrow \int \frac{\pi(\theta|x^{(2)})}{\pi(\theta|x^{(1)})}\pi(\theta|x^{(1)})d\theta = 1$$

Ainsi, d'après l'équation (4),

$$\int \pi(\theta|x^{(2)})d\theta = \int \frac{\pi(\theta|X^{(2)})}{\pi(\theta|X^{(1)})}\pi(\theta|X^{(1)})d\theta = \frac{1}{C}\cdot\int \frac{\pi(x^{(2)}|\theta)}{\pi(x^{(1)}|\theta)}\pi(\theta|x^{(1)})d\theta = 1$$

et donc

$$C = \int \frac{\pi(x^{(2)}|\theta)}{\pi(x^{(1)}|\theta)}\pi(\theta|x^{(1)})d\theta$$

Par approximation de Monte carlo, si $\theta_1^{(1)},\ldots,\theta_M^{(1)}$ est une réalisation de $\pi(\theta|X^{(1)})$:

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})} \to C \tag{6}$$

D'après les approximations (5) et (6), avec les propriétés de convergence presque sûre, on obtient

$$\frac{\sum_{i=1}^{M} g(\theta_i^{(1)})\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})}}{\sum_{i=1}^{M}\frac{\pi(x^{(2)}|\theta_i^{(1)})}{\pi(x^{(1)}|\theta_i^{(1)})}} = \sum_{i=1}^{M} g(\theta_i^{(1)})\frac{w_i}{\sum_i^M w_i} \to E_{[\theta|X^{(2)}]}[g(\theta)] \tag{7}$$

TABLE 1: Model 1 : Estimation of $E_{[\lambda|X]}(g(\lambda))$ for $\lambda = 1$

$\lambda = 1$	Exact estimation	Approximation via MCMC	Approximation via IS
$\hat{E}_{[\lambda\|X]}(\lambda)$ [1]	1.017	1.016	1.016
CI [4]	[0.650 ;1.400]	[0.650 ;1.398]	[0.652 ;1.402]
$\hat{sd}_{[\lambda\|X]}(\lambda)$ [2]	0.224	0.225	0.225
CI	[0.180 ;0.264]	[0.182 ;0.266]	[0.177 ;0.268]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.975})$ [3]	0.975	0.975	0.977
CI		[0.974 ;0.975]	[0.967 ;1.000]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.95})$ [3]	0.95	0.948	0.950
CI		[0.947 ;0.950]	[0.939 ;0.956]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.75})$ [3]	0.75	0.747	0.749
CI		[0.746 ;0.748]	[0.741 ;0.768]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.50})$ [3]	0.5	0.498	0.498
CI		[0.497 ;0.499]	[0.493 ;0.508]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.25})$ [3]	0.25	0.250	0.249
CI		[0.250,0.251]	[0.246 ;0.255]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.05})$ [3]	0.05	0.050	0.050
CI		[0.050 ;0.051]	[0.049 ;0.053]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.025})$ [3]	0.025	0.026	0.025
CI		[0.025 ;0.026]	[0.024 ;0.026]

[1] Moyenne des 100 approximations de l'espérance a posteriori de λ
[2] Moyenne des 100 approximations de l'écart-type a posteriori de θ
[3] Moyenne des 100 approximations des proportions a posteriori de θ où θ_q est le quantile d'ordre q de $\pi(\theta|X)$
[4] Intervalle contenant 95% des approximations

TABLE 2: Modèle 1 : Estimation de $E_{[\lambda|X]}(g(\lambda))$ pour $\lambda = 20$

$\lambda = 20$	Vraie estimation	Estimation via MCMC	Estimation via IS
$\hat{E}_{[\lambda\|X]}(\lambda)$ [1]	19.887	19.878	19.917
CI [4]	[17.142 ;22.339]	[17.140 ;22.330]	[17.153 ;22.655]
$\hat{sd}_{[\lambda\|X]}(\lambda)$ [2]	0.996	0.998	1.018
CI	[0.926 ;1.057]	[0.927 ;1.058]	[0.905 ;1.169]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.975})$ [3]	0.975	0.974	0.975
CI		[0.973 ;0.974]	[0.973 ;0.982]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.95})$ [3]	0.95	0.950	0.951
CI		[0.950 ;0.950]	[0.947 ;0.962]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.75})$ [3]	0.75	0.748	0.753
CI		[0.748 ;0.748]	[0.746 ;0.800]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.50})$ [3]	0.5	0.496	0.507
CI		[0.496 ;0.496]	[0.495 ;0.621]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.25})$ [3]	0.25	0.250	0.262
CI		[0.250,0.250]	[0.240 ;0.420]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.05})$ [3]	0.05	0.050	0.057
CI		[0.050 ;0.050]	[0.000 ;0.138]
$\hat{P}_{[\lambda\|X]}(\lambda < \lambda_{.025})$ [3]	0.025	0.025	0.030
CI		[0.025 ;0.025]	[0.000 ;0.095]

[1] Moyenne des 100 approximations de l'espérance a posteriori de λ
[2] Moyenne des 100 approximations de l'écart-type a posteriori de λ
[3] Moyenne des 100 approximations des proportions a posteriori de λ où λ_q est le quantile d'ordre q de $\pi(\lambda|X)$
[4] Intervalle contenant 95% des approximations

TABLE 3: Modèle 2 : Estimation de $E_{[\theta|X]}(g(\theta))$ for $\theta = (a = 0, b = 0.5)$

$a = 0, b = 0.5$	MCMC $g(\theta) = a$	IS $g(\theta) = a$	MCMC $g(\theta) = b$	IS $g(\theta) = b$
$\hat{E}_{[\theta\|X]}(\theta)$ [1]	−0.096	−0.079	0.503	0.511
CI [4]	[-0.524 ;0.273]	[-0.474 ;0.271]	[0.083 ;0.897]	[0.157 ;0.887]
$\hat{sd}_{[\theta\|X]}(\theta)$ [2]	0.252	0.242	0.246	0.235
CI	[0.205 ;0.323]	[0.198 ;0.287]	[0.157 ;0.348]	[0.157 ;0.328]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.975})$ [3]	0.975	0.982	0.975	0.982
CI		[0.973 ;1]		[0.969 ;1]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.95})$ [3]	0.950	0.961	0.950	0.959
CI		[0.947 ;1]		[0.937 ;1]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.75})$ [3]	0.750	0.771	0.750	0.760
CI		[0.746 ;0.865]		[0.719 ;0.936]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.50})$ [3]	0.500	0.519	0.500	0.508
CI		[0.483 ;0.605]		[0.471 ;0.609]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.25})$ [3]	0.250	0.267	0.250	0.253
CI		[0.236 ;0.334]		[0.230 ;0.307]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.05})$ [3]	0.050	0.054	0.050	0.049
CI		[0.017 ;0.093]		[0.034 ;0.061]
$\hat{P}_{[\theta\|X]}(\theta < \theta_{.025})$ [3]	0.025	0.028	0.025	0.024
CI		[0.009 ;0.050]		[0.012 ;0.031]

[1]Moyenne des 100 approximations de l'espérance a posteriori de θ
[2]Moyenne des 100 approximations de l'écart-type a posteriori de θ
[3]Moyenne des 100 approximations des proportions a posteriori de θ où θ_q est le quantile d'ordre q de $\pi(\theta|X)$
[4]Intervalle contenant 95% des approximations

TABLE 4: Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=$(a = 0\ ,\ b = 1/2\ ,\ \sigma^2$=$1/8)$

a=0, b=1/2, σ^2=1/8	MCMC $g(\theta) = a$	IS $g(\theta) = a$	MCMC $g(\theta) = b$	IS $g(\theta) = b$
$\hat{E}_{[\theta\|X]}(g(\theta))$ [1]	−0.199	−0.140	0.534	0.510
CI [4]	[-0.810 ;0.274]	[-0.542 ;0.267]	[-0.043 ;0.988]	[-0.112 ;0.929]
$\hat{sd}_{[\theta\|X]}(g(\theta))$ [2]	0.317	0.255	0.284	0.242
CI	[0.218 ;0.560]	[0.169 ;0.340]	[0.172 ;0.479]	[0.159 ;0.355]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.975}))$ [3]	0.975	0.988	0.975	0.979
CI		[0.920 ;1.000]		[0.958 ;0.997]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.95}))$ [3]	0.950	0.976	0.950	0.955
CI		[0.912 ;1.000]		[0.915 ;0.992]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.75}))$ [3]	0.750	0.801	0.750	0.742
CI		[0.534 ;0.996]		[0.637 ;0.865]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.50}))$ [3]	0.500	0.549	0.500	0.484
CI		[0.316 ;0.831]		[0.268 ;0.680]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.25}))$ [3]	0.250	0.279	0.250	0.221
CI		[0.101 ;0.466]		[0.012 ;0.385]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.05}))$ [3]	0.050	0.056	0.050	0.031
CI		[0.007 ;0.150]		[0.000 ;0.084]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.025}))$ [3]	0.025	0.029	0.025	0.006
CI		[0.001 ;0.094]		[0.000 ;0.056]

[1]Moyenne des 100 approximations de l'espérance a posteriori de θ
[2]Moyenne des 100 approximations de l'écart-type a posteriori de θ
[3]Moyenne des 100 approximations des proportions a posteriori de θ où θ_q est le quantile d'ordre q de $\pi(\theta|X)$
[4]Intervalle contenant 95% des approximations

TABLE 5: Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=1/4)

a=0, b=1/2, σ^2=1/4	MCMC $g(\theta) = a$	IS $g(\theta) = a$	MCMC $g(\theta) = b$	IS $g(\theta) = b$
$\hat{E}_{[\theta\|X]}(g(\theta))$ [1]	−0.191	−0.136	0.534	0.436
CI [4]	[-0.793 ;0.277]	[-0.604 ;0.203]	[0.040 ;1.069]	[0.024 ;0.827]
$\hat{sd}_{[\theta\|X]}(g(\theta))$ [2]	0.323	0.252	0.287	0.219
CI	[0.222 ;0.545]	[0.174 ;0.364]	[0.186 ;0.456]	[0.125 ;0.308]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.975}))$ [3]	0.975	0.988	0.975	0.961
CI		[0.949 ;1.000]		[0.848 ;0.998]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.95}))$ [3]	0.950	0.972	0.950	0.927
CI		[0.878 ;1.000]		[0.743 ;0.992]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.75}))$ [3]	0.750	0.796	0.750	0.647
CI		[0.484 ;1.000]		[0.186 ;0.857]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.50}))$ [3]	0.500	0.561	0.500	0.360
CI		[0.276 ;0.948]		[0.019 ;0.678]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.25}))$ [3]	0.250	0.273	0.250	0.140
CI		[0.041 ;0.682]		[0.000 ;0.450]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.05}))$ [3]	0.050	0.042	0.050	0.022
CI		[0.000 ;0.134]		[0.000 ;0.129]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.025}))$ [3]	0.025	0.016	0.025	0.006
CI		[0.000 ;0.060]		[0.000 ;0.052]

[1] Moyenne des 100 approximations de l'espérance a posteriori de θ
[2] Moyenne des 100 approximations de l'écart-type a posteriori de θ
[3] Moyenne des 100 approximations des proportions a posteriori de θ où θ_q est le quantile d'ordre q de $\pi(\theta|X)$
[4] Intervalle contenant 95% des approximations

TABLE 6: Modèle 3 : Estimation de $E_{[\theta|X]}(g(\theta))$ pour θ=($a = 0$, $b = 1/2$, σ^2=1/2)

a=0, b=1/2, σ^2=1/2	MCMC $g(\theta) = a$	IS $g(\theta) = a$	MCMC $g(\theta) = b$	IS $g(\theta) = b$
$\hat{E}_{[\theta\|X]}(g(\theta))$ [1]	−0.149	−0.046	0.548	0.515
CI [4]	[-0.803 ;0.350]	[-0.429 ;0.402]	[0.008 ;1.040]	[-0.013 ;0.967]
$\hat{sd}_{[\theta\|X]}(g(\theta))$ [2]	0.340	0.253	0.301	0.232
CI	[0.229 ;0.619]	[0.172 ;0.322]	[0.183 ;0.544]	[0.122 ;0.347]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.975}))$ [3]	0.975	0.994	0.975	0.977
CI		[0.973 ;1.000]		[0.928 ;1.000]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.95}))$ [3]	0.950	0.983	0.950	0.956
CI		[0.942 ;1.000]		[0.862 ;1.000]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.75}))$ [3]	0.750	0.837	0.750	0.731
CI		[0.671 ;0.999]		[0.356 ;0.975]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.50}))$ [3]	0.500	0.606	0.500	0.468
CI		[0.367 ;0.932]		[0.118 ;0.835]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.25}))$ [3]	0.250	0.318	0.250	0.209
CI		[0.089 ;0.719]		[0.018 ;0.380]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.05}))$ [3]	0.050	0.070	0.050	0.025
CI		[0.006 ;0.233]		[0.000 ;0.083]
$\hat{P}_{[\theta\|X]}(g(\theta) < g(\theta_{.025}))$ [3]	0.025	0.041	0.025	0.009
CI		[0.001 ;0.207]		[0.000 ;0.035]

[1] Moyenne des 100 approximations de l'espérance a posteriori de θ
[2] Moyenne des 100 approximations de l'écart-type a posteriori de θ
[3] Moyenne des 100 approximations des proportions a posteriori de θ où θ_q est le quantile d'ordre q de $\pi(\theta|X)$
[4] Intervalle contenant 95% des approximations

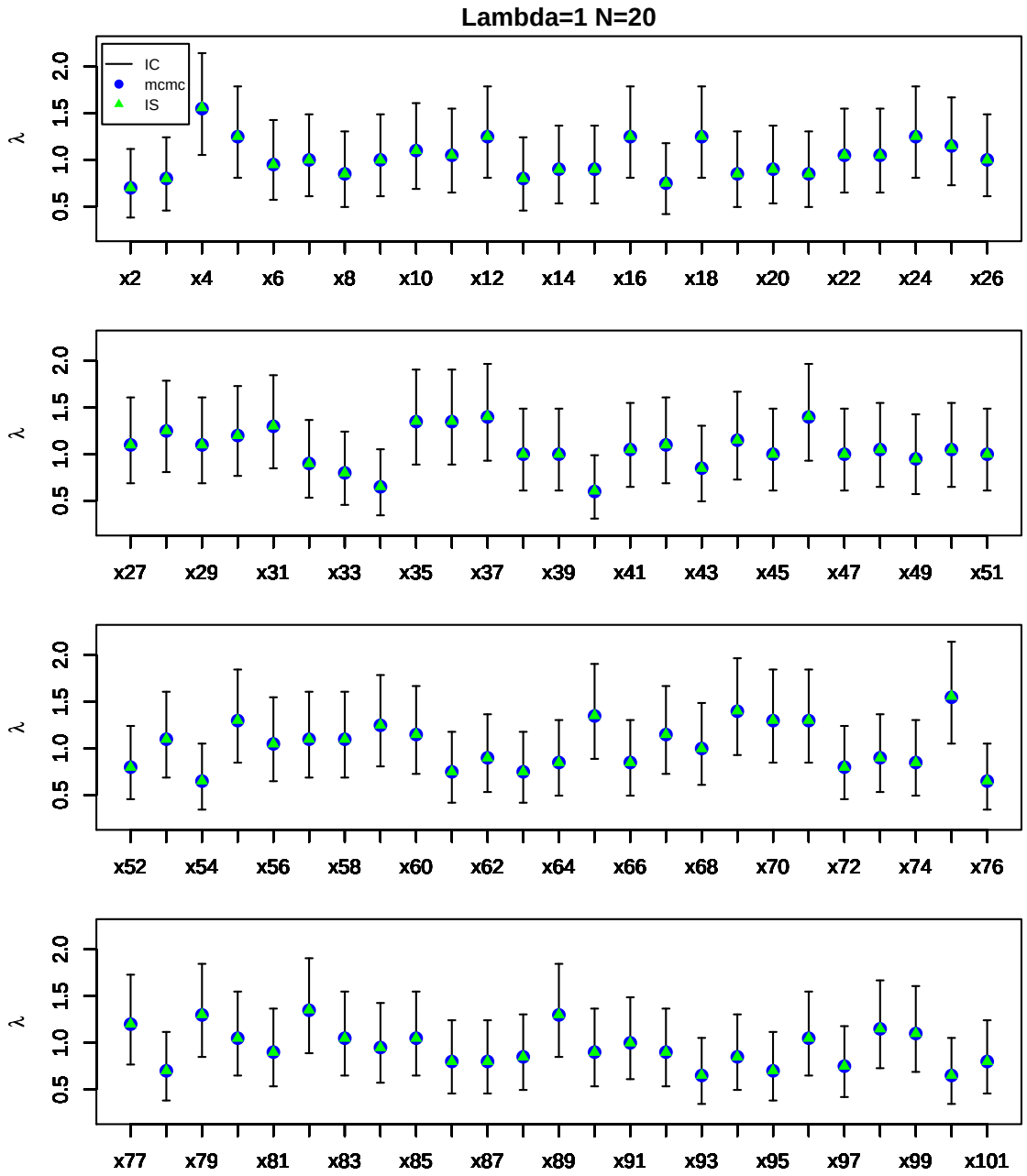

FIGURE 1: Performance d'inférence pour $g(\theta) = \lambda$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 1 avec $\lambda = 1$

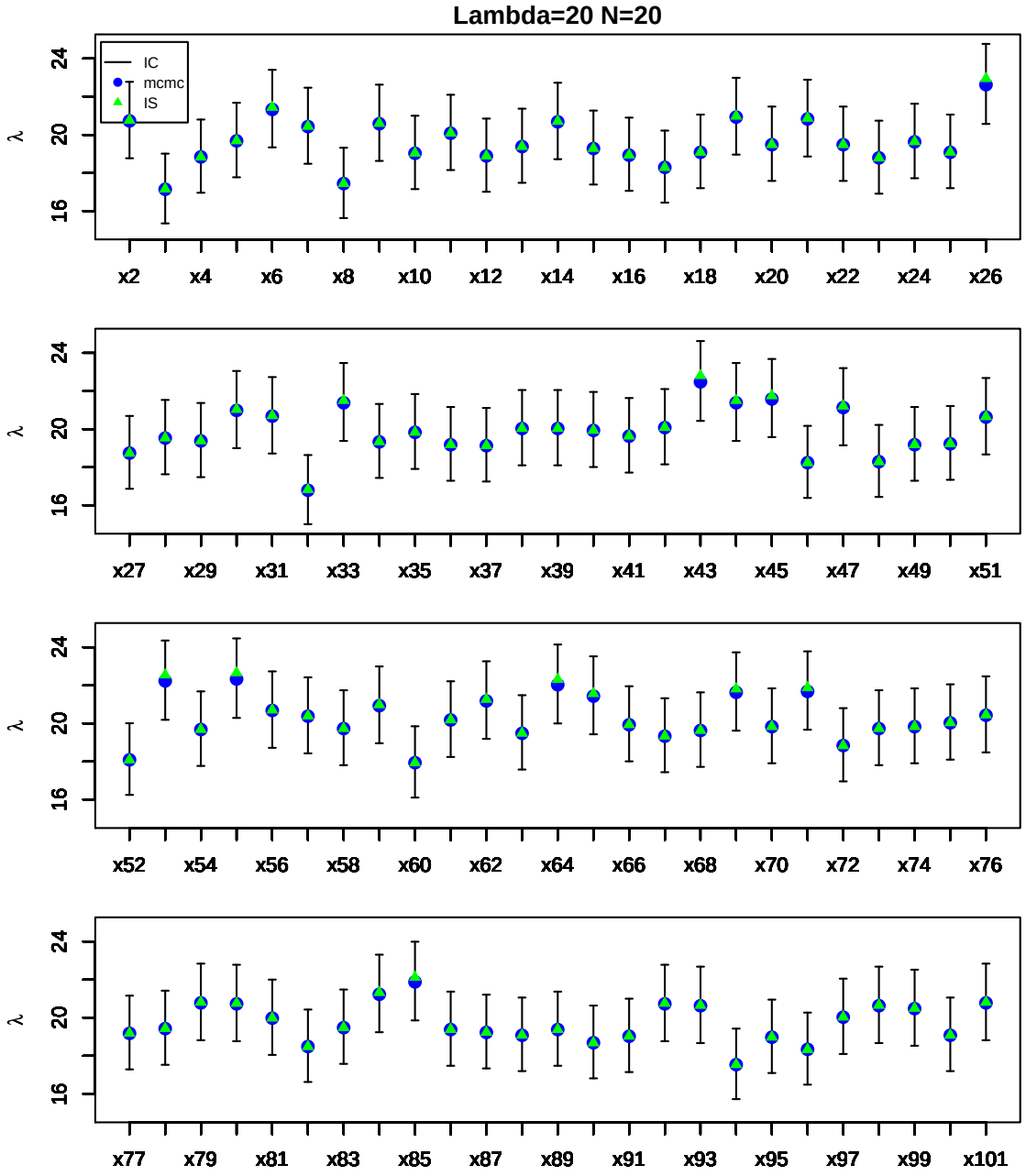

FIGURE 2: Performance d'inférence pour $g(\theta) = \lambda$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 1 avec $\lambda = 20$

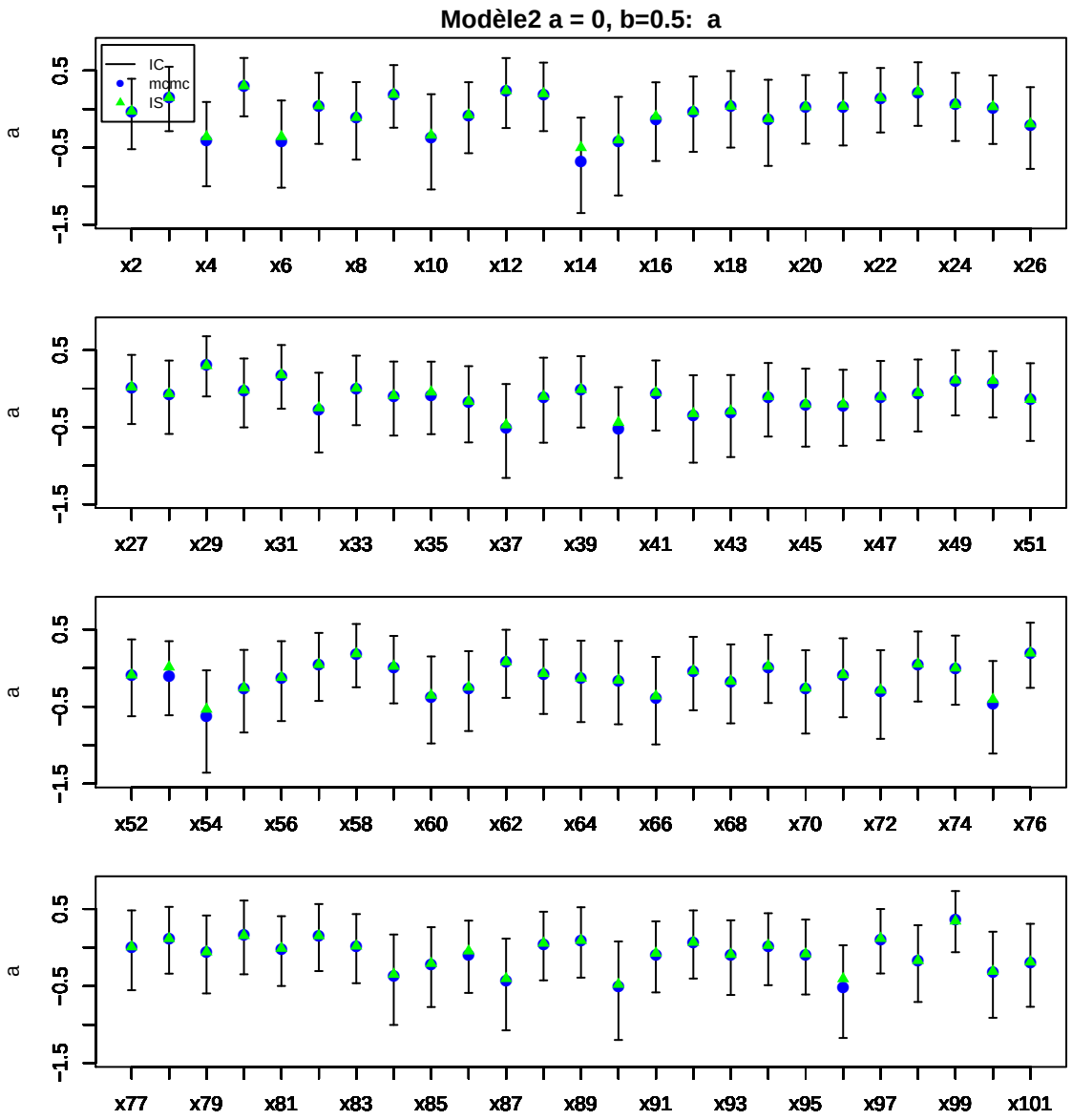

FIGURE 3: Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 2 avec $a = 0$ et $b = 0.5$

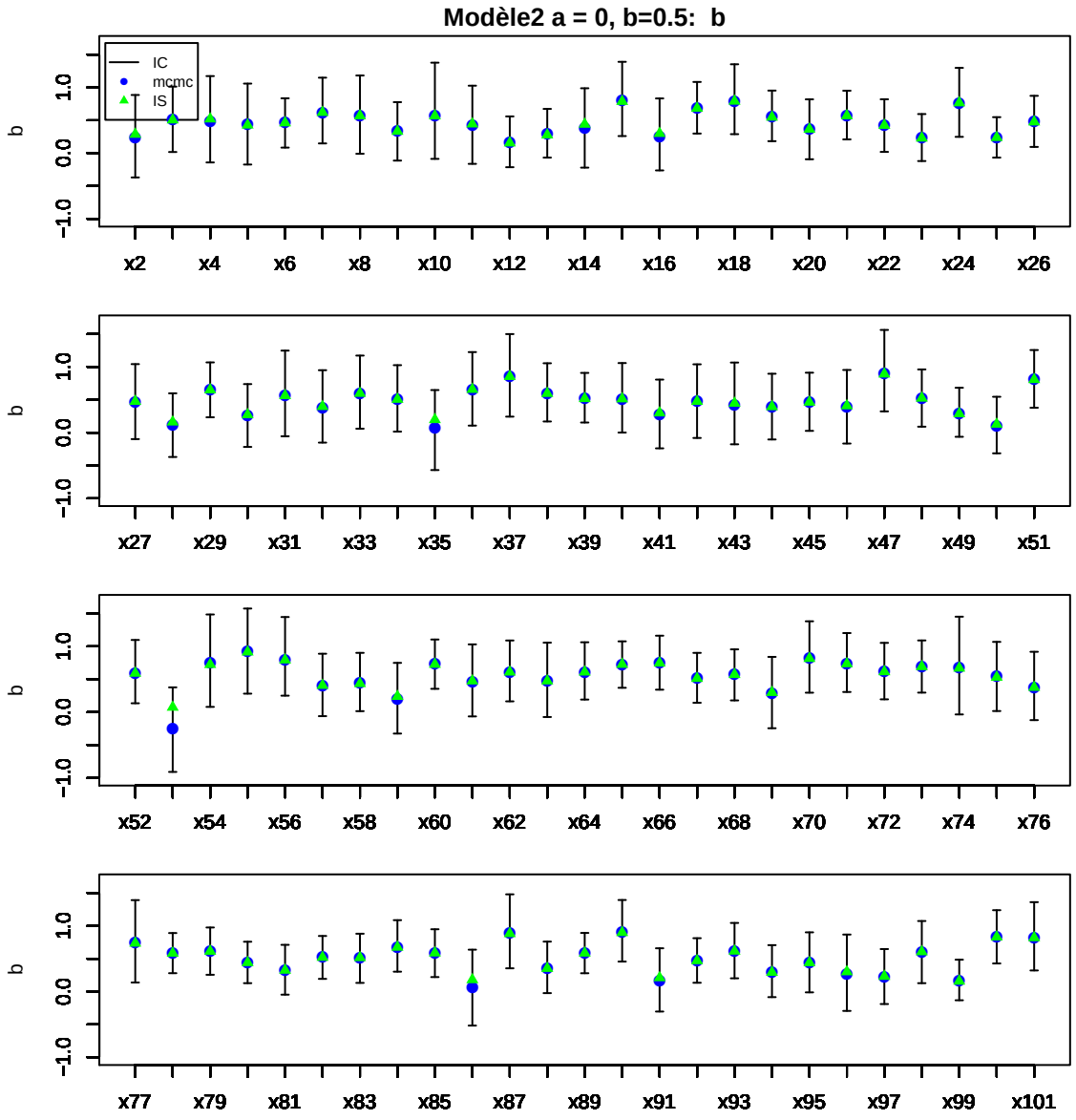

FIGURE 4: Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 2 avec $a = 0$ et $b = 0.5$

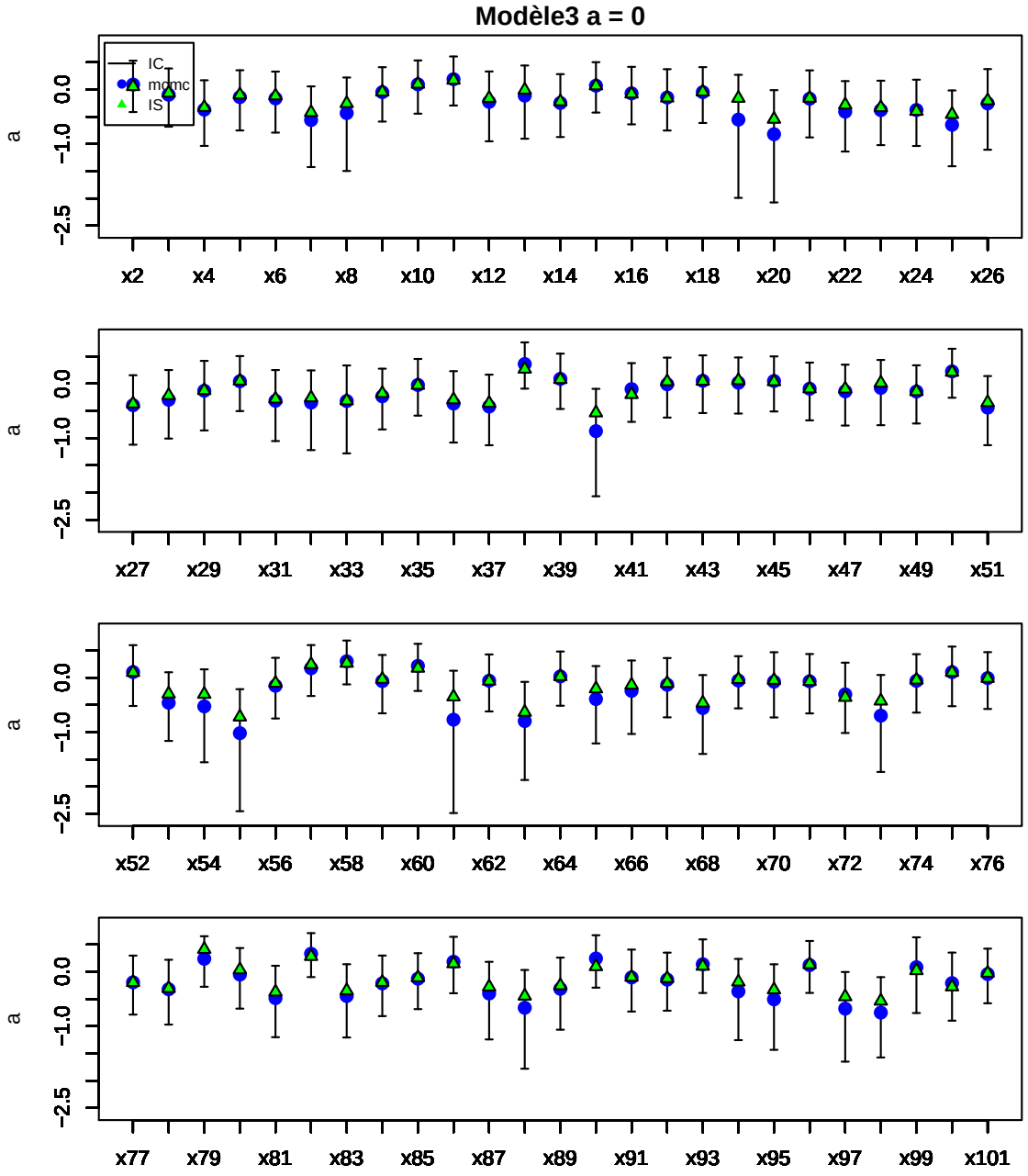

FIGURE 5: Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$

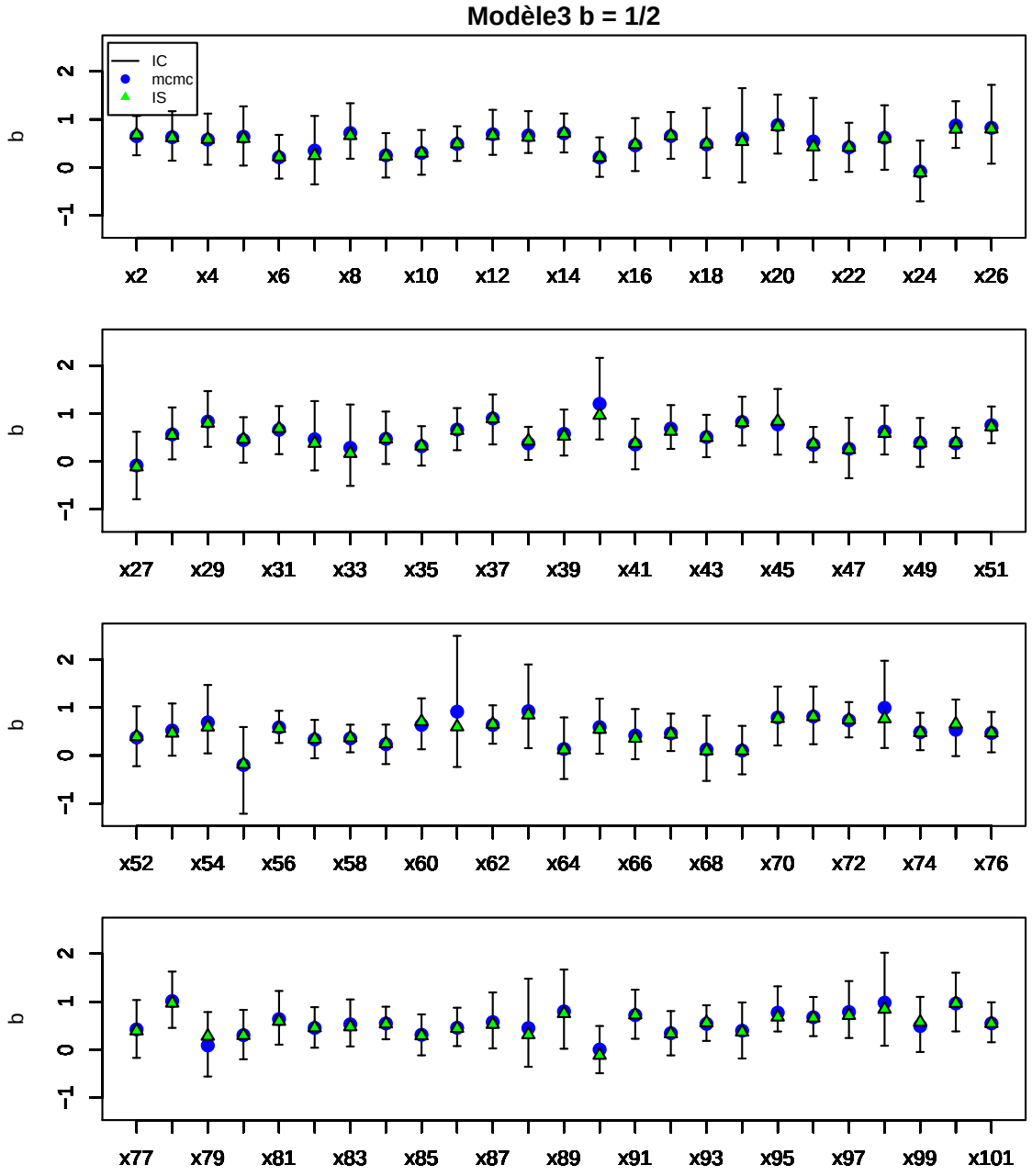

FIGURE 6: Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$

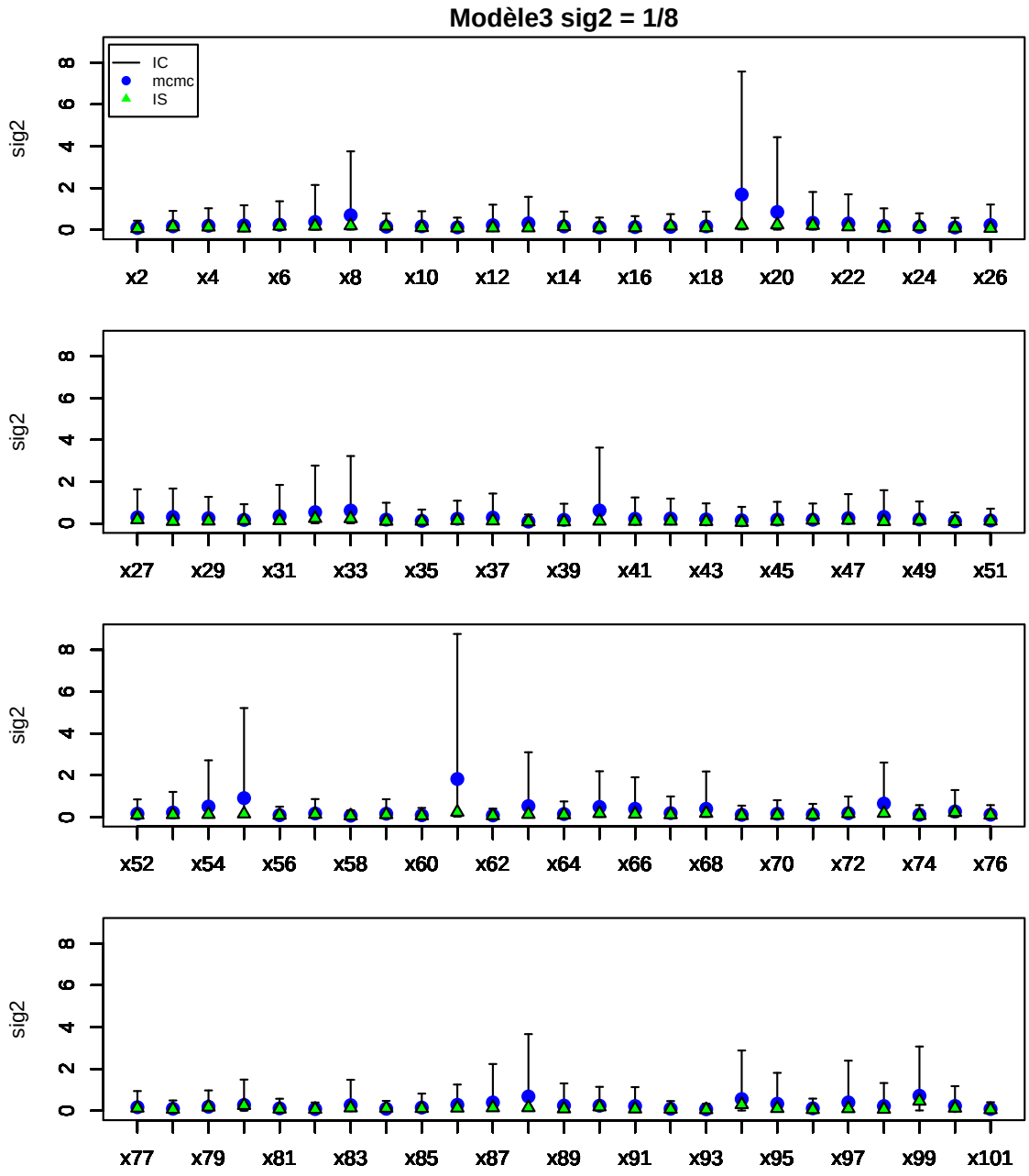

FIGURE 7: Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/8$

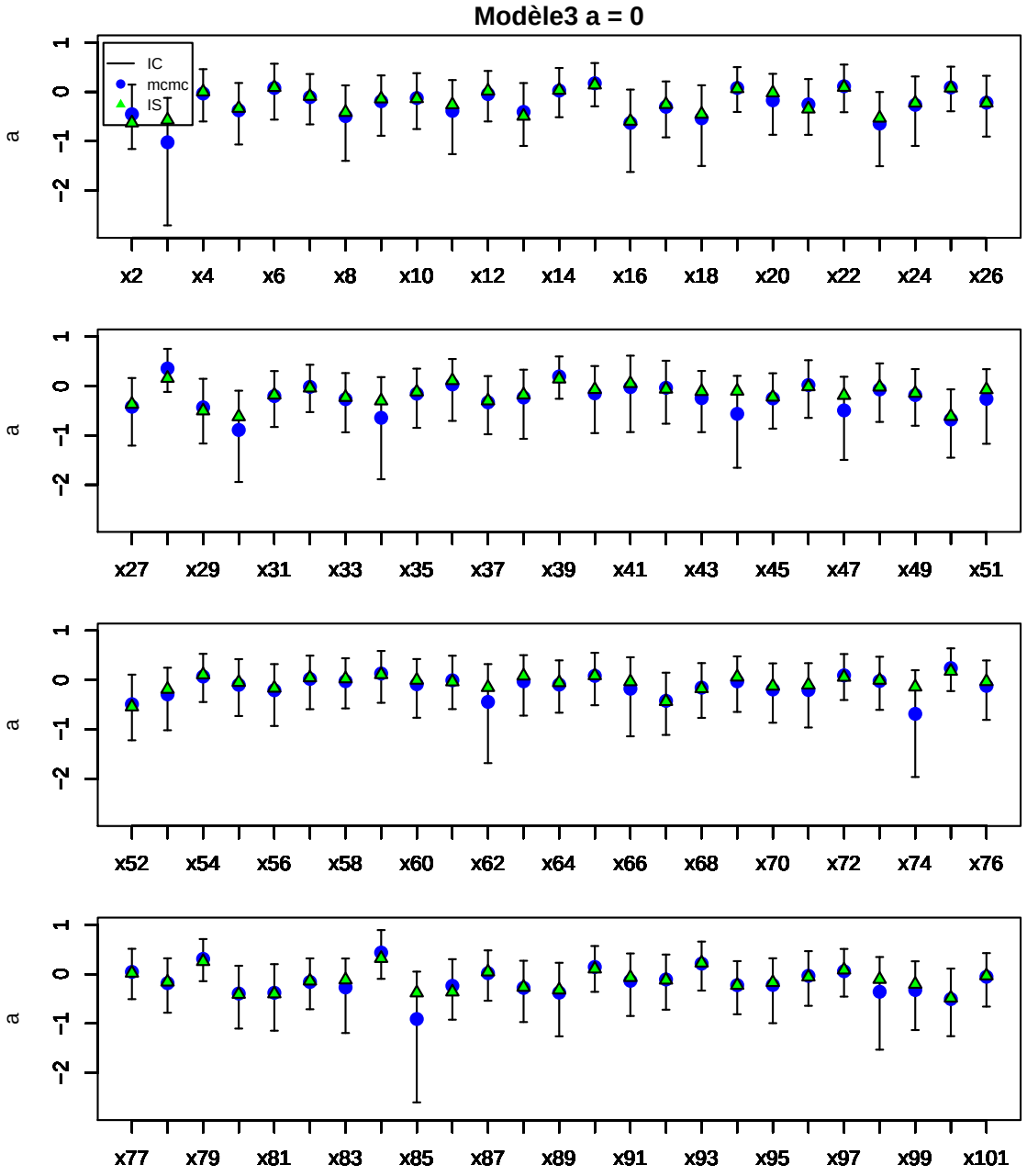

FIGURE 8: Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$

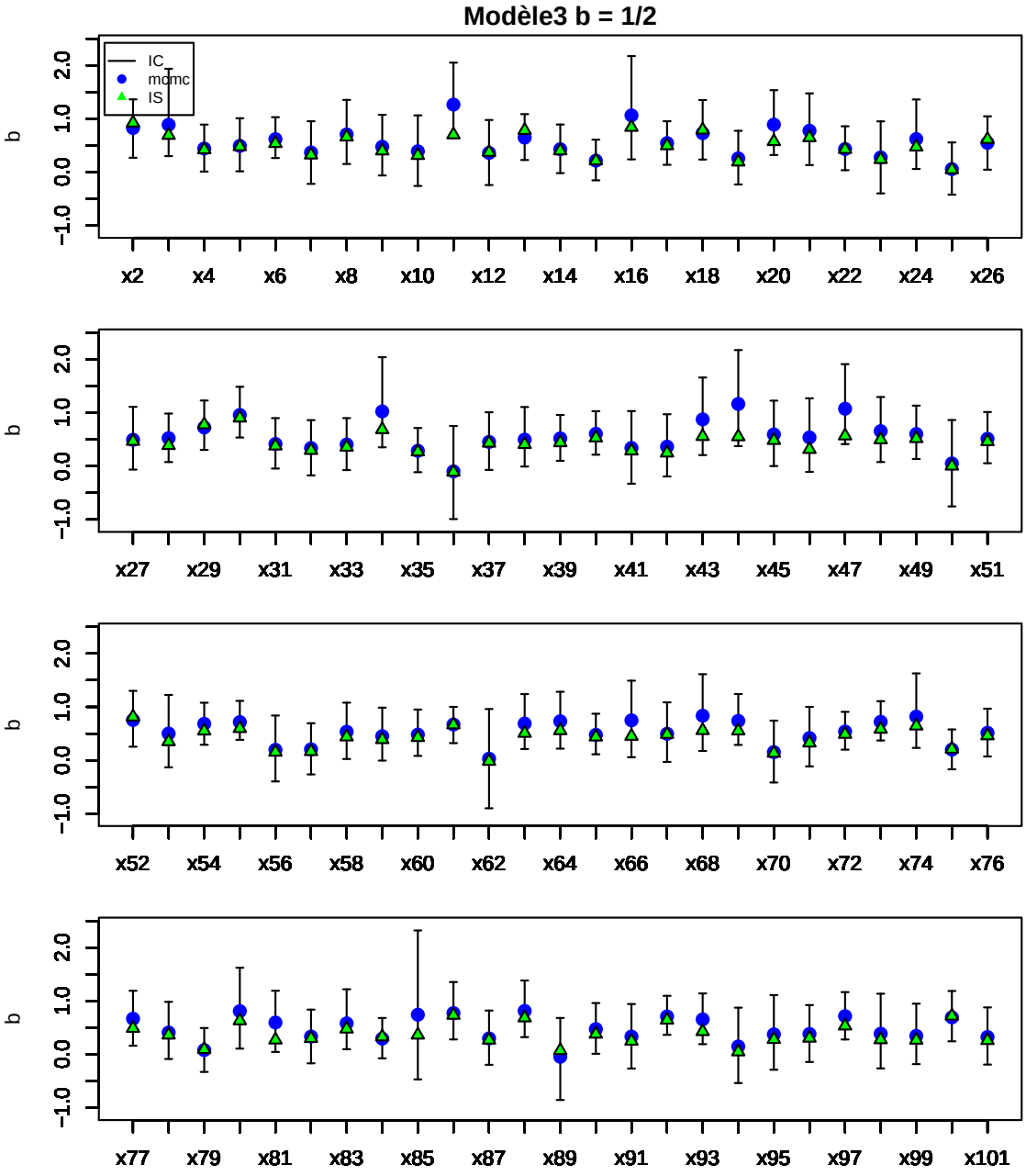

FIGURE 9: Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$

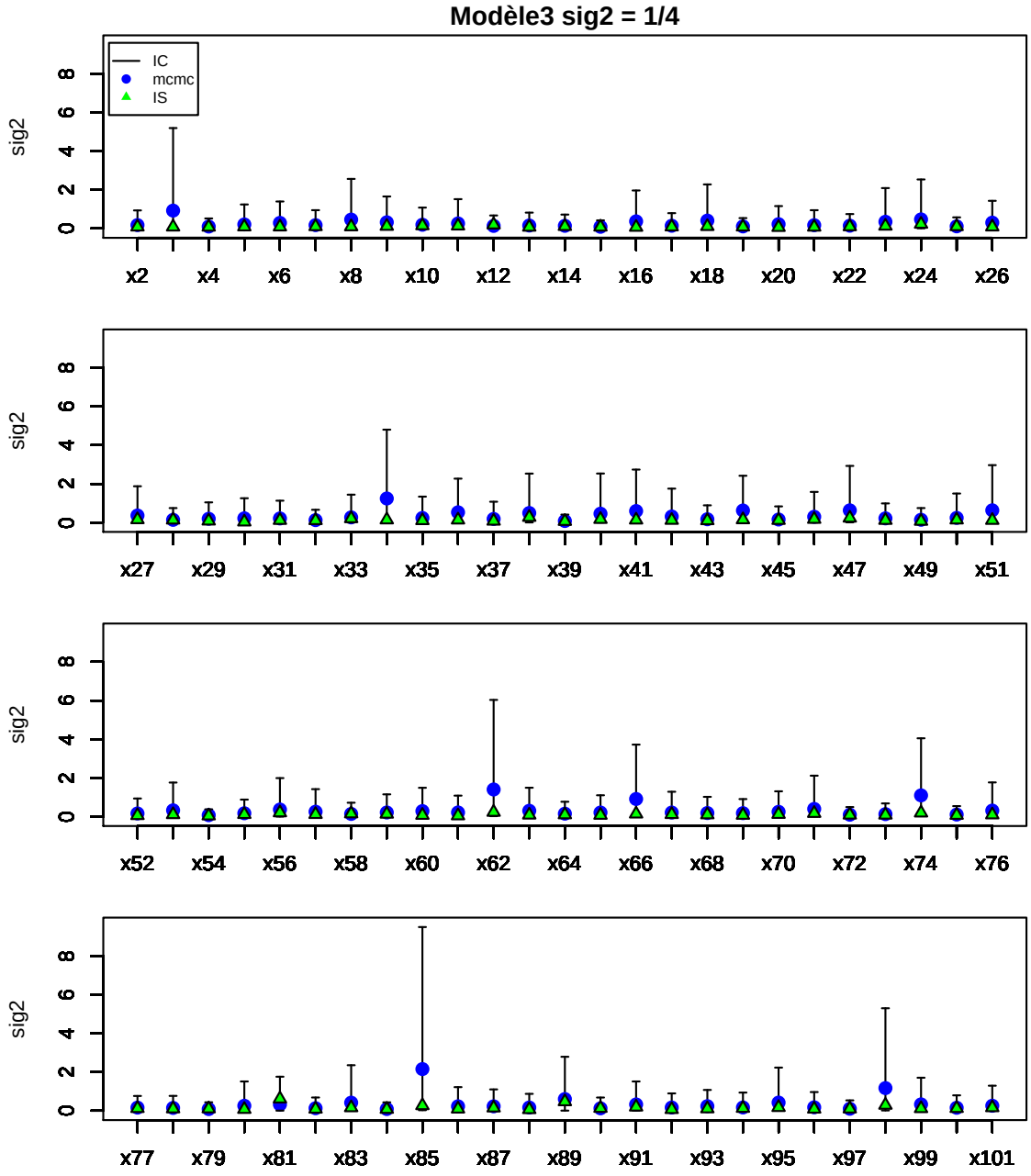

FIGURE 10: Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/4$

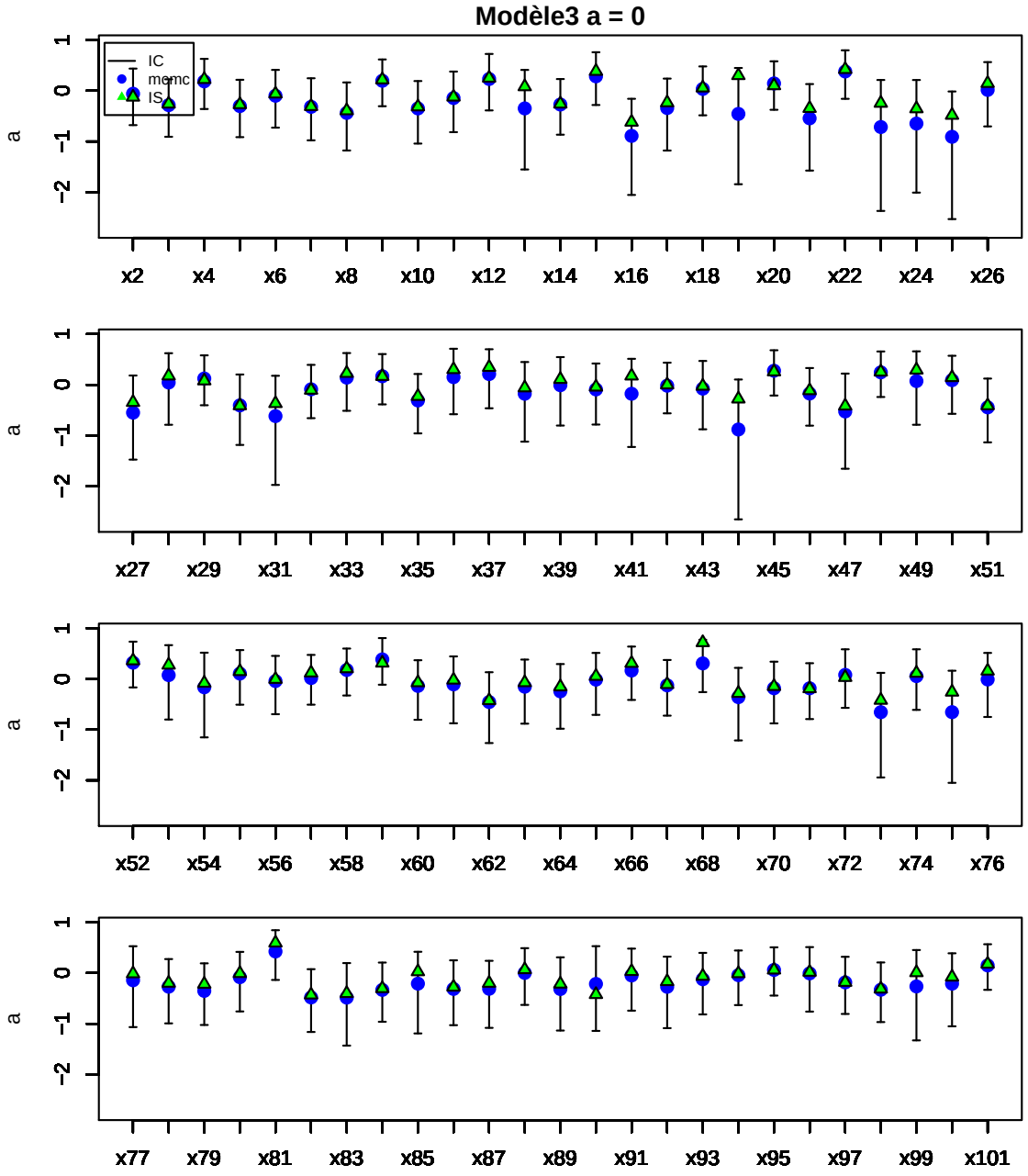

FIGURE 11: Performance d'inférence pour $g(\theta) = a$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$

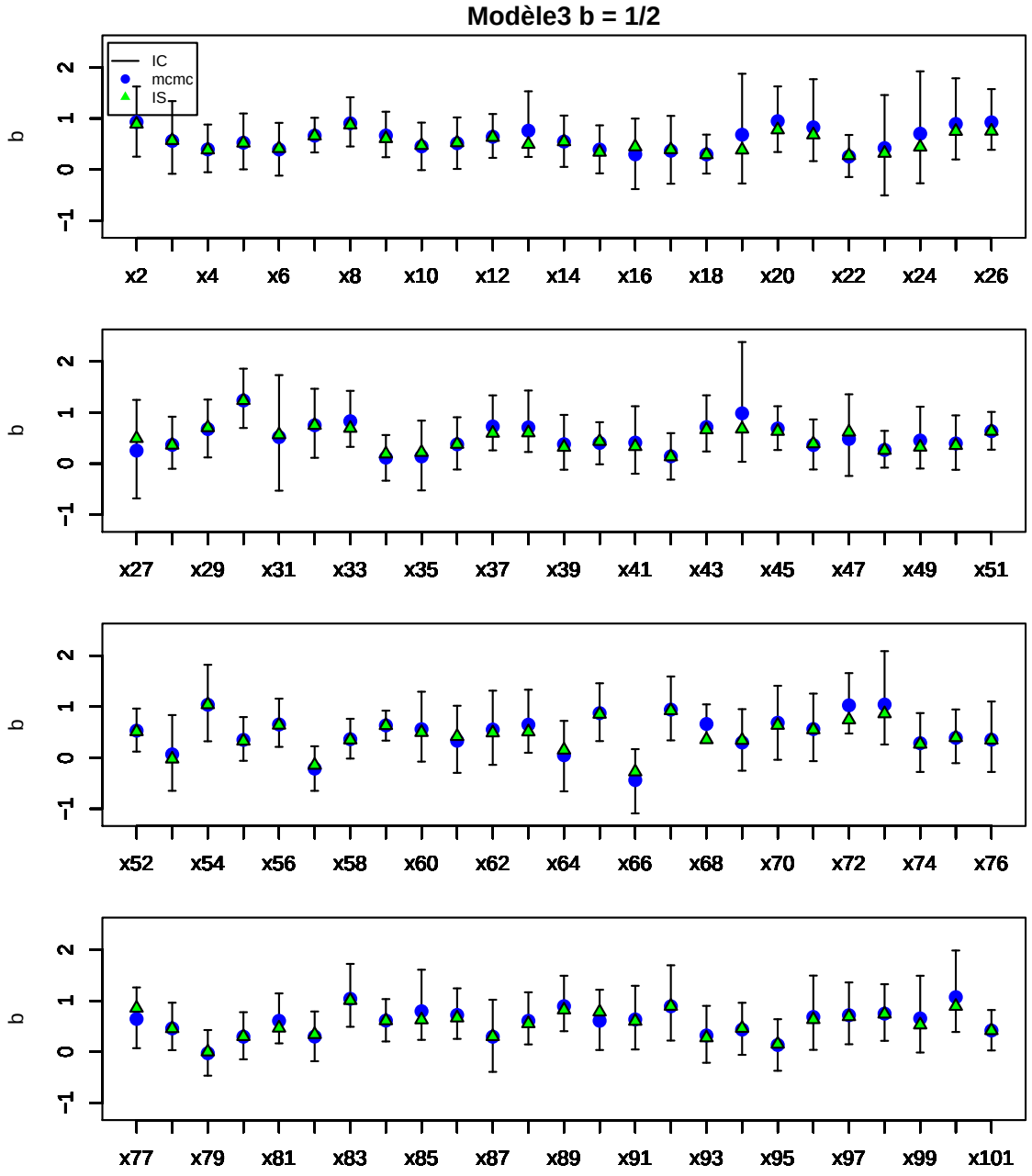

FIGURE 12: Performance d'inférence pour $g(\theta) = b$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$

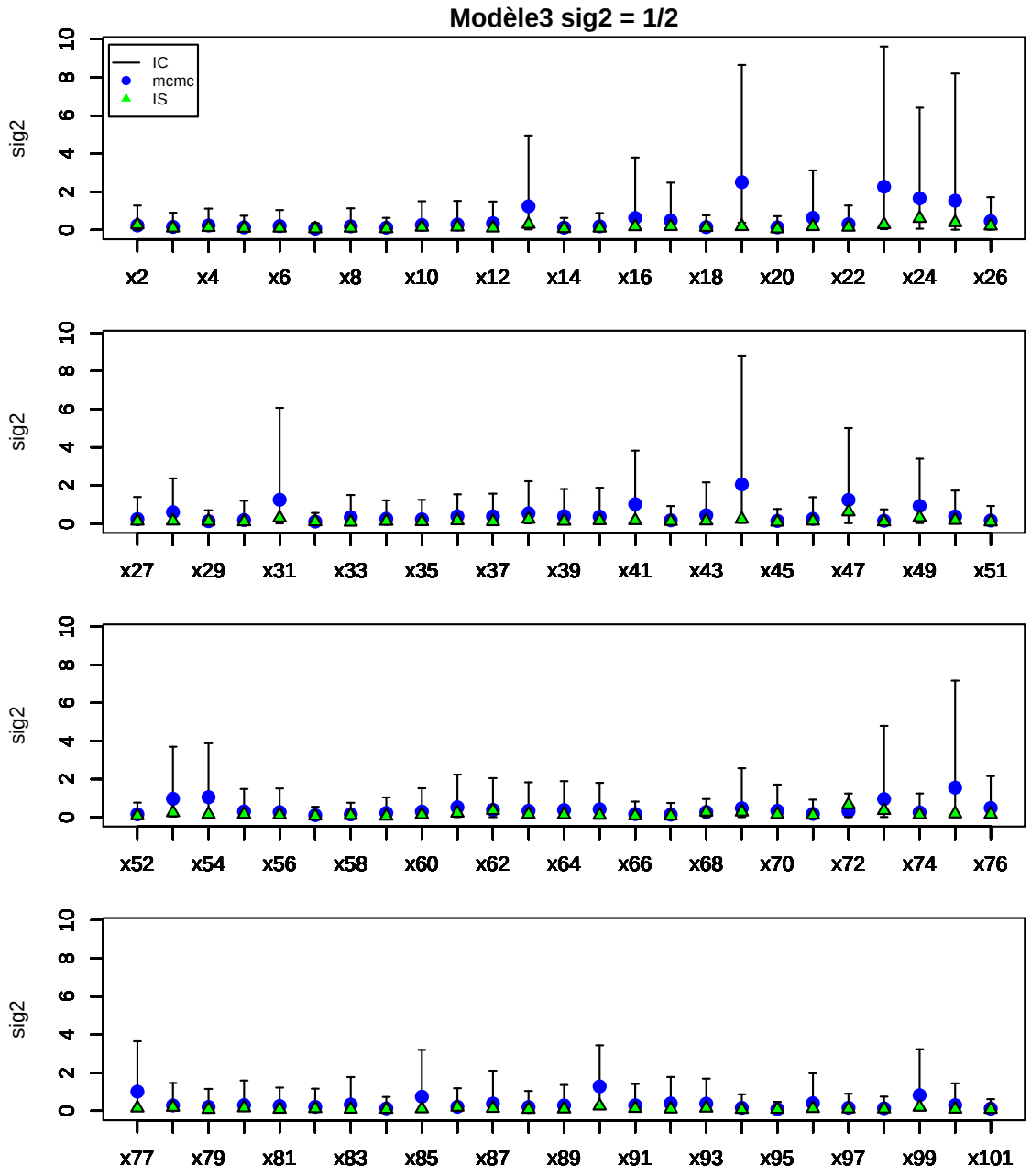

FIGURE 13: Performance d'inférence pour $g(\theta) = \sigma^2$ avec le premier jeu de données $X^{(1)}$ comme la référence fixe Modèle 3 avec $a = 0$, $b = 0.5$ et $\sigma^2 = 1/2$

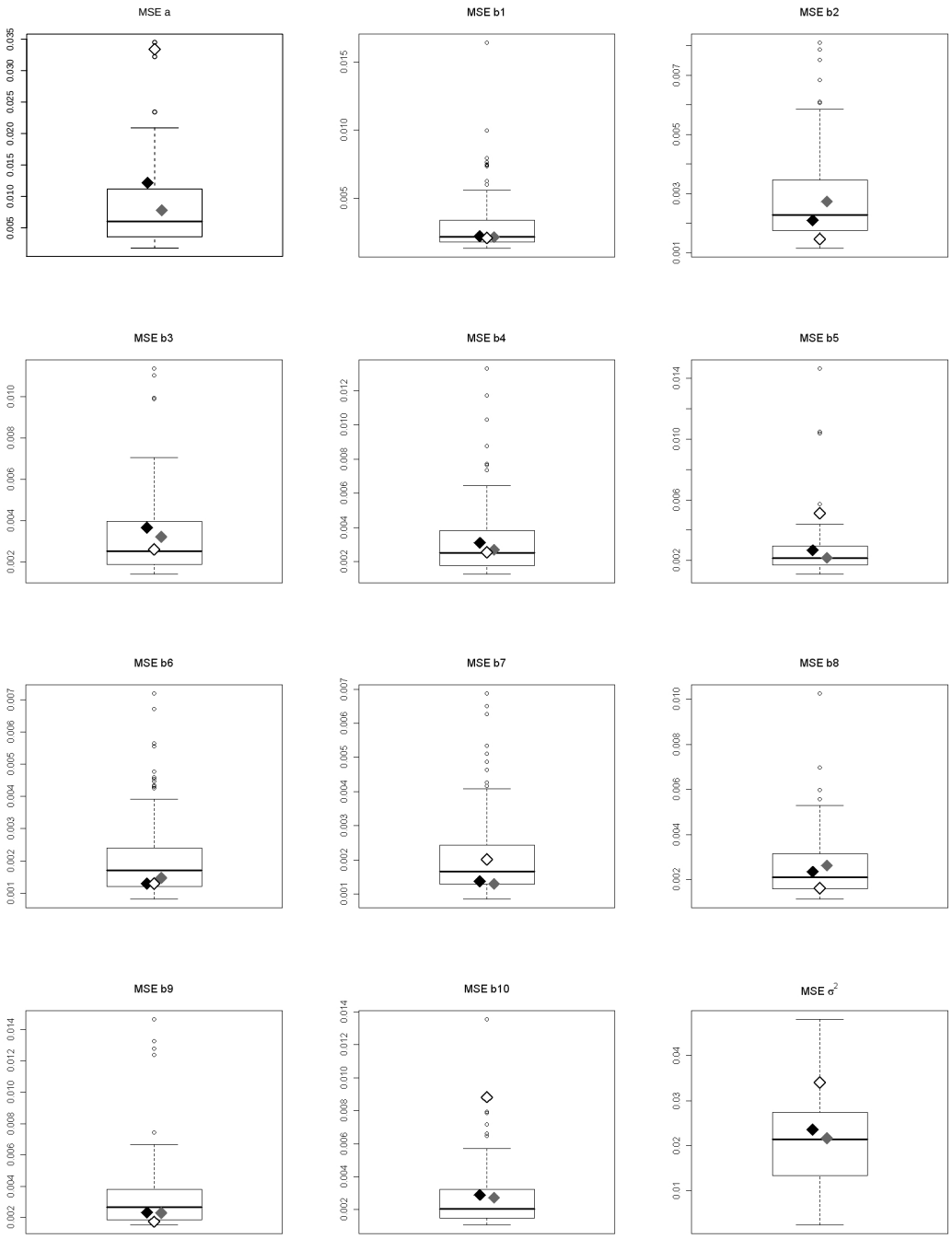

FIGURE 14: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0, b_1, \ldots, b_{10} = 0.05$ (normal) et $\sigma^2 = 1/2$, pour $g(\theta) = a, g(\theta) = b_1, \ldots, b_{10}$ et $g(\theta) = \sigma^2$; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

TABLE 7: Erreurs quadratique moyennes par rapport à MCMC dans le Modèle 3 avec 10 covaribles normales et avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $g(\theta) = \sigma^2$ (extension)

strategy	a		b_1		b_2	
$\sigma^2 = 1/2$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	1.766	0.066	1.291	0.087	1.150	0.045
La pire ref. "fixe" [4]	34.501	0.743	16.471	0.435	8.107	0.198
Stratégie de réf. "choisie" avec le critère 1 [5]	12.163	0.482	2.180	0.165	2.097	0.107
Stratégie de réf. "choisie" avec le critère 2 [5]	33.395	0.819	2.058	0.094	1.458	0.069
Stratégie de réf. "choisie" avec le critère 3 [5]	7.790	3.736	2.118	0.120	2.733	0.110
	b_3		b_4		b_5	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	1.425	0.052	1.256	0.068	1.090	0.054
La pire ref. "fixe" [4]	11.367	0.294	13.330	0.326	14.674	0.319
Stratégie de réf. "choisie" avec le critère 1 [5]	3.672	0.159	3.102	0.198	2.676	0.169
Stratégie de réf. "choisie" avec le critère 2 [5]	2.617	0.144	2.537	0.119	5.118	0.224
Stratégie de réf. "choisie" avec le critère 3 [5]	3.234	0.175	2.692	0.165	2.180	0.098
	b_6		b_7		b_8	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	0.815	0.040	0.845	0.032	1.126	0.050
La pire ref. "fixe" [4]	7.200	0.298	6.885	0.201	10.272	0.246
Stratégie de réf. "choisie" avec le critère 1 [5]	1.297	0.069	1.374	0.0577	2.347	0.110
Stratégie de réf. "choisie" avec le critère 2 [5]	1.300	0.065	2.007	0.099	1.611	0.081
Stratégie de réf. "choisie" avec le critère 3 [5]	1.473	0.065	1.292	0.055	2.623	0.148
	b_9		b_{10}		σ^2	
$\sigma^2 = 1/2$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	1.526	0.080	1.063	0.051	2.347	0.134
La pire ref. "fixe" [4]	14.668	0.376	13.562	0.294	48.125	1.014
Stratégie de réf. "choisie" avec le critère 1 [5]	2.326	0.094	2.900	0.117	23.535	0.177
Stratégie de réf. "choisie" avec le critère 2 [5]	1.736	0.070	8.824	0.296	34.034	0.918
Stratégie de réf. "choisie" avec le critère 3 [5]	2.298	0.097	2.725	0.136	21.677	0.837

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

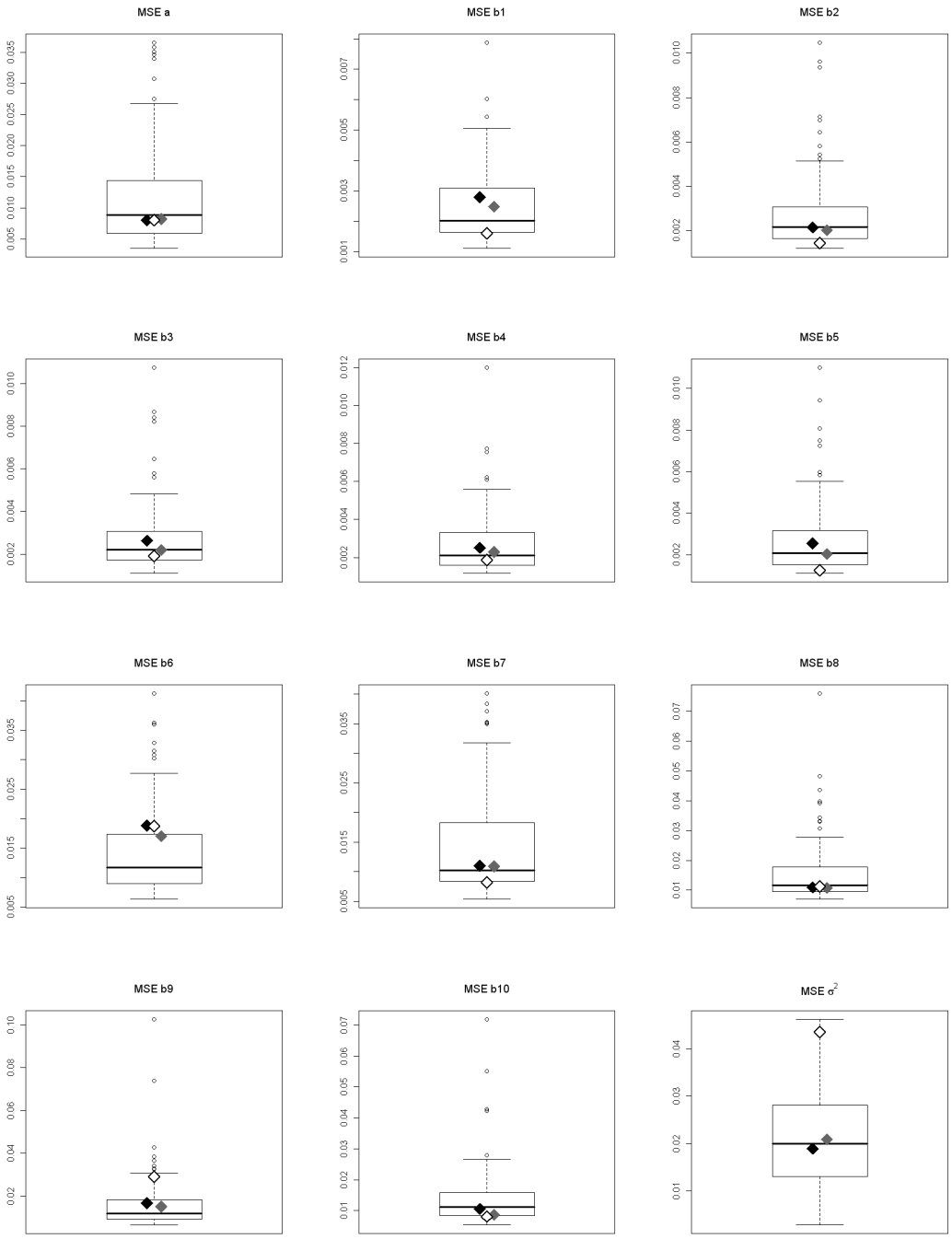

FIGURE 15: Box-plot de $EQ_{(fixe)}$ pour toutes les valeurs possibles m avec la stratégie "référence fixe" et de $EQ_{(choisie)}$ avec la stratégie "référence choisie" selon les trois critères dans le Modèle 3 avec $a = 0$, $b_1, \ldots, b_{10} = 0.05$ (normal + bernoulli) et $\sigma^2 = 1/2$, pour $g(\theta) = a$, $g(\theta) = b_1, \ldots, b_{10}$ et $g(\theta) = \sigma^2$; losange noir pour le premier critère, losange blanc pour le second et losange gris pour le troisième.

TABLE 8: Erreurs quadratique moyennes par rapport à MCMC dans le Modèle 3 avec 5 covariables normales et 5 covariables de Bernoulli avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$, $\sigma^2 = 1/2$ et $g(\theta) = \sigma^2$ (extension)

strategy	a		b_1		b_2	
$\sigma^2 = 1/2$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	3.479	0.163	1.108	0.044	1.202	0.058
La pire ref. "fixe" [4]	36.598	0.851	7.887	0.223	10.483	0.291
Stratégie de réf. "choisie" avec le critère 1 [5]	8.017	0.404	2.795	0.146	2.148	0.089
Stratégie de réf. "choisie" avec le critère 2 [5]	8.002	0.383	1.597	0.097	1.439	0.066
Stratégie de réf. "choisie" avec le critère 3 [5]	8.235	0.562	2.484	0.135	2.026	0.116
	b_3		b_4		b_5	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	1.096	0.048	1.165	0.061	1.097	0.044
La pire ref. "fixe" [4]	10.755	0.244	12.008	0.314	11.020	0.261
Stratégie de réf. "choisie" avec le critère 1 [5]	2.626	0.132	2.515	0.135	2.547	0.149
Stratégie de réf. "choisie" avec le critère 2 [5]	1.900	0.104	1.859	0.066	1.252	0.062
Stratégie de réf. "choisie" avec le critère 3 [5]	2.198	0.097	2.295	0.139	2.032	0.115
	b_6		b_7		b_8	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	6.331	0.296	5.327	0.224	6.955	0.343
La pire ref. "fixe" [4]	41.283	1.494	40.143	1.064	76.085	1.881
Stratégie de réf. "choisie" avec le critère 1 [5]	18.825	0.921	11.001	0.573	10.941	0.524
Stratégie de réf. "choisie" avec le critère 2 [5]	18.670	0.641	8.172	0.358	11.182	0.548
Stratégie de réf. "choisie" avec le critère 3 [5]	17.039	0.653	10.857	0.576	10.771	0.594
	b_9		b_{10}		σ^2	
$\sigma^2 = 1/2$	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
La meilleure ref. "fixe" [3]	6.406	0.416	5.355	0.302	2.811	0.108
La pire ref. "fixe" [4]	102.678	2.150	71.962	1.636	46.223	0.901
Stratégie de réf. "choisie" avec le critère 1 [5]	16.668	0.796	10.560	0.586	18.915	0.725
Stratégie de réf. "choisie" avec le critère 2 [5]	28.939	1.280	8.092	0.403	43.450	0.995
Stratégie de réf. "choisie" avec le critère 3 [5]	15.043	0.711	8.634	0.418	20.845	0.637

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

TABLE 9: Erreurs quadratiques moyennes par rapport à MCMC dans le Modèle 2 avec 10 covaribles normales avec $a = 0$, $b_j = 0.05$ pour $j = 1, \ldots, 10$ et $n = 1,000$ pour $g(\theta) = a$, $g(\theta) = b_j$, $j = 1, \ldots, 10$

strategy	a		b_1		b_2	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
Stratégie de réf. "choisie" avec le critère 1 [5]	0.071	0.005	0.078	0.005	0.062	0.004
Stratégie de réf. "choisie" avec le critère 2 [5]	0.249	0.016	0.245	0.017	0.251	0.020
Stratégie de réf. "choisie" avec le critère 3 [5]	0.110	0.010	0.160	0.013	0.126	0.008
Stratégie mélange	**0.029**	**0.002**	**0.032**	**0.003**	**0.044**	**0.003**
	b_3		b_4		b_5	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
Stratégie de réf. "choisie" avec le critère 1 [5]	0.091	0.005	0.063	0.004	0.070	0.005
Stratégie de réf. "choisie" avec le critère 2 [5]	0.374	0.033	0.203	0.013	0.182	0.086
Stratégie de réf. "choisie" avec le critère 3 [5]	0.119	0.010	0.116	0.010	0.069	0.006
Stratégie mélange	**0.043**	**0.002**	**0.032**	**0.002**	**0.032**	**0.003**
	b_6		b_7		b_8	
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]
Stratégie de réf. "choisie" avec le critère 1 [5]	0.085	0.008	0.075	0.005	0.063	0.003
Stratégie de réf. "choisie" avec le critère 2 [5]	0.268	0.014	0.287	0.017	0.275	0.014
Stratégie de réf. "choisie" avec le critère 3 [5]	0.179	0.013	0.115	0.009	0.136	0.013
Stratégie mélange	**0.030**	**0.004**	**0.052**	**0.005**	**0.031**	**0.003**
	b_9		b_{10}			
	EQ [1]	$q_{97.5\%}$ [2]	EQ [1]	$q_{97.5\%}$ [2]		
Stratégie de réf. "choisie" avec le critère 1 [5]	0.052	0.002	0.071	0.006		
Stratégie de réf. "choisie" avec le critère 2 [5]	0.261	0.020	0.247	0.014		
Stratégie de réf. "choisie" avec le critère 3 [5]	0.118	0.008	0.096	0.006		
Stratégie mélange	**0.033**	**0.002**	**0.041**	**0.003**		

[1] Erreur quadratique moyenne par rapport à MCMC sur (K-M) jeux de données ($\times 10^3$)
[2] Terme de rang 97,5% intervenant dans l'erreur quadratique moyenne ($\times 10^3$)
[3] IS avec la stratégie de réf. "fixe" - le meilleur cas ($\times 10^3$)
[4] IS avec la stratégie de réf. "fixe" - le pire cas ($\times 10^3$)
[5] IS avec la stratégie de réf. "choisie" ($\times 10^3$)

TABLE 10: Les vraisemblances, la densités a priori et a posteriori, les espérances a posteriori et les rapport des constantes de normalisation des modèles : Normal-Normal, Poisson-Gamma, Binomial-Beta, Gamma-Gamma et Négative Binomial Beta.

Modèle Normal Normal	
Modèle	$\mathcal{N}(\theta, \sigma^2)$
Vraisemblance	$V(x\|\theta) = (2\pi\sigma^2)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(\sum x_i^2 - 2\theta \sum x_i + n\theta^2)\right\}$
Prior	$\mathcal{N}(\mu, \tau^2)$ fdd. $\pi(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$
Posterior	$\mathcal{N}(\rho(\sigma^2\mu + \tau^2 \sum x_i), \rho\sigma^2\tau^2)$, $\rho^{-1} = \sigma^2 + n\tau^2$
Espérance a posteriori	$E_{\theta\|x}(\theta) = (\sigma^2\mu + \tau^2 \sum x_i)/(\sigma^2 + n\tau^2)$
Rapport des constantes de normalisation	$\frac{c_m}{c_p} = \exp\left\{-\frac{1}{2}\left[\frac{\sum (x_i^{(m)2} - x_i^{(p)2})}{\sigma^2} - \left(\frac{\sum (x_i^{(m)} - x_i^{(p)})}{\sigma^2} + \frac{2\mu}{\tau^2}\right)\frac{\tau^2 \sum (x_i^{(m)} - x_i^{(p)})}{n\tau^2 + \sigma^2}\right]\right\}$

Modèle Poisson Gamma	
Modèle	$\mathcal{P}(\theta)$
Vraisemblance	$V(x\|\theta) = e^{-n\theta}\frac{\theta^{\sum x_i}}{\prod x_i!}$
Prior	$\mathcal{G}(\alpha, \beta)$ fdd. $\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \quad \theta \geq 0$
Posterior	$\mathcal{G}(\alpha + \sum x_i, \beta + n)$
Espérance a posteriori	$E_{\theta\|x}(\theta) = (\alpha + \sum x_i)/(\beta + n)$
Rapport des constantes de normalisation	$\frac{c_m}{c_p} = (n+\beta)^{\sum (x_i^{(p)} - x_i^{(m)})} \frac{\Gamma(\alpha + \sum x_i^{(m)})}{\Gamma(\alpha + \sum x_i^{(p)})} \frac{\Pi x_i^{(p)}!}{\Pi x_i^{(m)}!}$

Modèle Binomial Beta	
Modèle	$\mathcal{B}in(N, \theta)$
Vraisemblance	$V(x\|\theta) = \prod \binom{N}{x_i} \theta^{\sum x_i}(1-\theta)^{nN - \sum x_i}, \quad x_i \in \mathbb{N}$
Prior	$\mathcal{G}(\alpha, \beta)$ fdd. $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, \quad \theta \in [0,1]$
Posterior	$\mathcal{B}(\alpha + \sum x_i, \beta + nN - \sum x_i)$
Espérance a posteriori	$E_{\theta\|x}(\theta) = (\alpha + \sum x_i)/(\alpha + \beta + nN)$
Rapport des constantes de normalisation	$\frac{c_m}{c_p} = \frac{\Pi \binom{n}{x_i^{(m)}}}{\Pi \binom{n}{x_i^{(p)}}} \frac{B(\sum x_i^{(m)} + \alpha, nN - \sum x_i^{(m)})}{B(\sum x_i^{(p)} + \alpha, nN - \sum x_i^{(p)})}$

Modèle Gamma Gamma	
Modèle	$\mathcal{G}(\nu, \theta)$
Vraisemblance	$V(x\|\theta) = \frac{\theta^{n\nu}}{\Gamma(\nu)^n}(\prod x_i)^{\nu-1} e^{-\theta \sum x_i}, \quad x_i \geq 0$
Prior	$\mathcal{G}(\alpha, \beta)$ fdd. $\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \quad \theta \geq 0$
Posterior	$\mathcal{G}(\alpha + n\nu, \beta + \sum x_i)$
Espérance a posteriori	$E_{\theta\|x}(\theta) = (\alpha + n\nu)/(\beta + \sum x_i)$
Rapport des constantes de normalisation	$\frac{c_m}{c_p} = \left(\frac{\prod x_i^{(m)}}{\prod x_i^{(p)}}\right)^{\nu-1} \cdot \left(\frac{\sum x_i^{(p)} + \beta}{\sum x_i^{(m)} + \beta}\right)^{n\nu+\alpha}$

Modèle Négative Binomial Beta	
Modèle	$\mathcal{N}eg(N, \theta)$
Vraisemblance	$V(x\|\theta) = \prod \binom{n+x_i-1}{x_i} \theta^{nN}(1-\theta)^{\sum x_i}, \quad x_i \in \mathbb{N}$
Prior	$\mathcal{G}(\alpha, \beta)$ fdd. $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, \quad \theta \in [0,1]$
Posterior	$\mathcal{B}(\alpha + nN, \beta + \sum x_i)$
Espérance a posteriori	$E_{\theta\|x}(\theta) = (\alpha + nN)/(\alpha + nN + \beta + \sum x_i)$
Rapport des constantes de normalisation	$\frac{c_m}{c_p} = \frac{\Pi \binom{n+x_i^{(m)}-1}{x_i^{(m)}}}{\Pi \binom{n+x_i^{(p)}-1}{x_i^{(p)}}} \frac{B(nN+\alpha, \beta + \sum x_i^{(m)})}{B(nN+\alpha, \beta + \sum x_i^{(p)})}$

TABLE 11: Estimation des rapports des constantes de normalisation via trois méthodes : importance sampling ($\hat{r}_{is}$), reciprocal importance sampling ($\hat{r}_{recip}$) et ratio importance sampling ($\hat{r}_{is}$). Les vraies valeurs sont rapportées dans la colonne dénotée par r_0.

Méthode :	r_0	$\hat{r}_{is}$	$\hat{r}_{recip}$	$\hat{r}_{ris}$
	Modèle Normal Normal ($p = 8$)			
r_1	2.210e-02	2.203e-02	2.210e-02	2.209e-02
r_2	5.520e-01	5.566e-01	5.520e-01	5.581e-01
r_3	3.491e+00	3.477e+00	3.491e+00	3.507e+00
r_4	2.187e+00	2.200e+00	2.187e+00	2.203e+00
r_5	2.011e+01	2.003e+01	2.011e+01	2.015e+01
r_6	3.524e-04	3.559e-04	3.524e+04	3.571e-04
r_7	2.967e-01	2.958e-01	2.967e-01	2.965e-01
r_9	7.053e-03	7.133e-03	7.053e-03	7.161e-03
r_{10}	7.051e+00	7.022e+00	7.051e+00	7.070e+00
	Modèle Poisson Gamma ($p = 3$)			
r_1	1.861 002e-01	1.841e-01	1.859e-01	1.872e-01
r_2	4.225 230e-02	4.077e-02	4.220e-02	4.155e-02
r_4	1.966 667e-05	1.963e-05	1.964e-05	1.952e-05
r_5	4.279 676e-03	4.276e-03	4.263e-03	4.253e-03
r_6	1.568 525e-04	1.451e-04	1.566e-04	1.547e-04
r_7	8.593 200e-02	7.785e-02	8.582e-02	8.696e-02
r_8	1.603 497e-03	1.557e-03	1.600e-03	1.617e-03
r_9	3.878 966e-01	3.677e-01	3.882e-01	3.811e-01
r_{10}	5.652 794e-03	5.653e-03	5.654e-03	5.626e-03
	Modèle Binomial Beta ($p = 1$)			
r_2	14.688	14.708	14.655	14.700
r_3	130.692	129.186	130.503	129.936
r_4	0.500	0.502	0.500	0.503
r_5	81.047	81.289	81.029	81.594
r_6	16.447	16.470	16.410	16.461
r_7	0.940	0.938	0.939	0.939
r_8	1.859	1.863	1.858	1.864
r_9	0.2336	0.234	0.233	0.236
r_{10}	19.433	19.481	19.422	19.540
	Modèle Gamma Gamma ($p = 9$)			
r_1	1.899e-02	1.895e-02	1.899e-02	2.040e-02
r_2	3.959e-02	3.950e-02	3.959e-02	4.188e-02
r_3	1.651e-01	1.649e-01	1.651e-01	1.709e-01
r_4	3.046e+01	2.794e+01	3.046e+01	2.442e+01
r_5	8.788e-04	9.123e-04	8.788e-04	1.017e-03
r_6	1.239e-02	1.238e-02	1.239e-02	1.347e-02
r_7	1.476e-04	1.608e-04	1.476e-04	1.756e-04
r_8	7.981e-04	8.678e-04	7.981e-04	9.484e-04
r_{10}	8.417e-02	8.400e-02	8.417e-02	9.053e-02
	Modèle Négative Binomial ($p = 8$)			
r_1	8.976e+01	9.240e+01	8.982e+01	9.140e+01
r_2	5.211e-02	5.287e-02	5.210e-02	5.266e-02
r_3	2.140e-01	2.183e-01	2.140e-01	2.169e-01
r_4	8.097e-01	8.368e-01	8.090e-01	8.262e-01
r_5	9.089e+02	8.994e+02	9.092e+02	8.917e+02
r_6	6.273e+03	5.854e+03	6.276e+03	6.281e+03
r_7	5.224e-02	5.265e-02	5.228e-02	5.393e-02
r_9	9.451e-01	9.366e-01	9.411e-01	9.310e-01
r_{10}	1.701e+01	1.703e+01	1.702e+01	1.703e+01

Annexe de la modélisation épidémique

Développement de Taylor au voisinage de ϕ

$$
\begin{aligned}
\frac{\partial \Pi}{\partial t} &- \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} \\
&= f(\Omega)(\mathbb{E}_\xi - 1)\left[\rho_0(\phi) + \Omega^{-1/2}\rho_0'(\phi)\xi + \Omega^{-1}\rho_0''(\phi)\xi^2 + \ldots\right]\Pi(\xi,t) \\
&+ \Omega^{-1}f(\Omega)(\mathbb{E}_\xi - 1)\left[\rho_1(\phi) + \Omega^{-1/2}\rho_1'(\phi)\xi + \Omega^{-1}\rho_1''(\phi)\xi^2 + \ldots\right]\Pi(\xi,t) \\
&+ \ldots \\
&+ f(\Omega)(\mathbb{E}_\xi^{-1} - 1)\left[\gamma_0(\phi) + \Omega^{-1/2}\gamma_0'(\phi)\xi + \Omega^{-1}\gamma_0''(\phi)\xi^2 + \ldots\right]\Pi(\xi,t) \\
&+ \Omega^{-1}f(\Omega)(\mathbb{E}_\xi - 1)\left[\gamma_1(\phi) + \Omega^{-1/2}\gamma_1'(\phi)\xi + \Omega^{-1}\gamma_1''(\phi)\xi^2 + \ldots\right]\Pi(\xi,t) \\
&+ \ldots
\end{aligned}
\tag{8}
$$

En appliquant les opérateurs $\mathbb{E}_\xi$ et $\mathbb{E}_\xi^{-1}$ on obtient donc

$$
\begin{aligned}
\frac{\partial \Pi}{\partial t} &- \Omega^{1/2}\frac{d\phi}{dt}\frac{\partial \Pi}{\partial \xi} \\
&= f(\Omega)(\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} + \ldots)\left[\rho_0(\phi) + \Omega^{-1/2}\rho_0'(\phi)\xi + \Omega^{-1}\rho_0''(\phi)\xi^2 + \ldots\right]\Pi \\
&+ \Omega^{-1}f(\Omega)(\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} + \ldots)\left[\rho_1(\phi) + \Omega^{-1/2}\rho_1'(\phi)\xi + \Omega^{-1}\rho_1''(\phi)\xi^2 + \ldots\right]\Pi \\
&+ \ldots \\
&+ f(\Omega)(-\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} - \ldots)\left[\gamma_0(\phi) + \Omega^{-1/2}\gamma_0'(\phi)\xi + \Omega^{-1}\gamma_0''(\phi)\xi^2 + \ldots\right]\Pi \\
&+ \Omega^{-1}f(\Omega)(-\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} - \ldots)\left[\gamma_1(\phi) + \Omega^{-1/2}\gamma_1'(\phi)\xi + \Omega^{-1}\gamma_1''(\phi)\xi^2 + \ldots\right]\Pi \\
&+ \ldots
\end{aligned}
\tag{9}
$$

Il est possible de changer l'échelle du temps en posant $\Omega^{-1}f(\Omega)t = \tau$ et ainsi la dérivé d'une fonction h devient $\frac{dh}{dt} = \frac{dh}{d\tau}\Omega^{-1}f(\Omega)$. En mettant en facteur $\Omega^{-1}f(\Omega)$ dans (9) on obtient

$$
\begin{aligned}
\frac{\partial \Pi}{\partial \tau} &- \Omega^{1/2}\frac{d\phi}{d\tau}\frac{\partial \Pi}{\partial \xi} \qquad (10)\\
&= (\Omega^{1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{0}\frac{\partial^2}{\partial \xi^2} + \ldots)\left[\rho_0(\phi) + \Omega^{-1/2}\rho_0'(\phi)\xi + \Omega^{-1}\rho_0''(\phi)\xi^2 + \ldots\right]\Pi\\
&+ (\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} + \ldots)\left[\rho_1(\phi) + \Omega^{-1/2}\rho_1'(\phi)\xi + \Omega^{-1}\rho_1''(\phi)\xi^2 + \ldots\right]\Pi\\
&+ \ldots\\
&+ (-\Omega^{1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{0}\frac{\partial^2}{\partial \xi^2} - \ldots)\left[\gamma_0(\phi) + \Omega^{-1/2}\gamma_0'(\phi)\xi + \Omega^{-1}\gamma_0''(\phi)\xi^2 + \ldots\right]\Pi\\
&+ (-\Omega^{-1/2}\frac{\partial}{\partial \xi} + \frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial \xi^2} - \ldots)\left[\gamma_1(\phi) + \Omega^{-1/2}\gamma_1'(\phi)\xi + \Omega^{-1}\gamma_1''(\phi)\xi^2 + \ldots\right]\Pi\\
&+ \ldots
\end{aligned}
$$

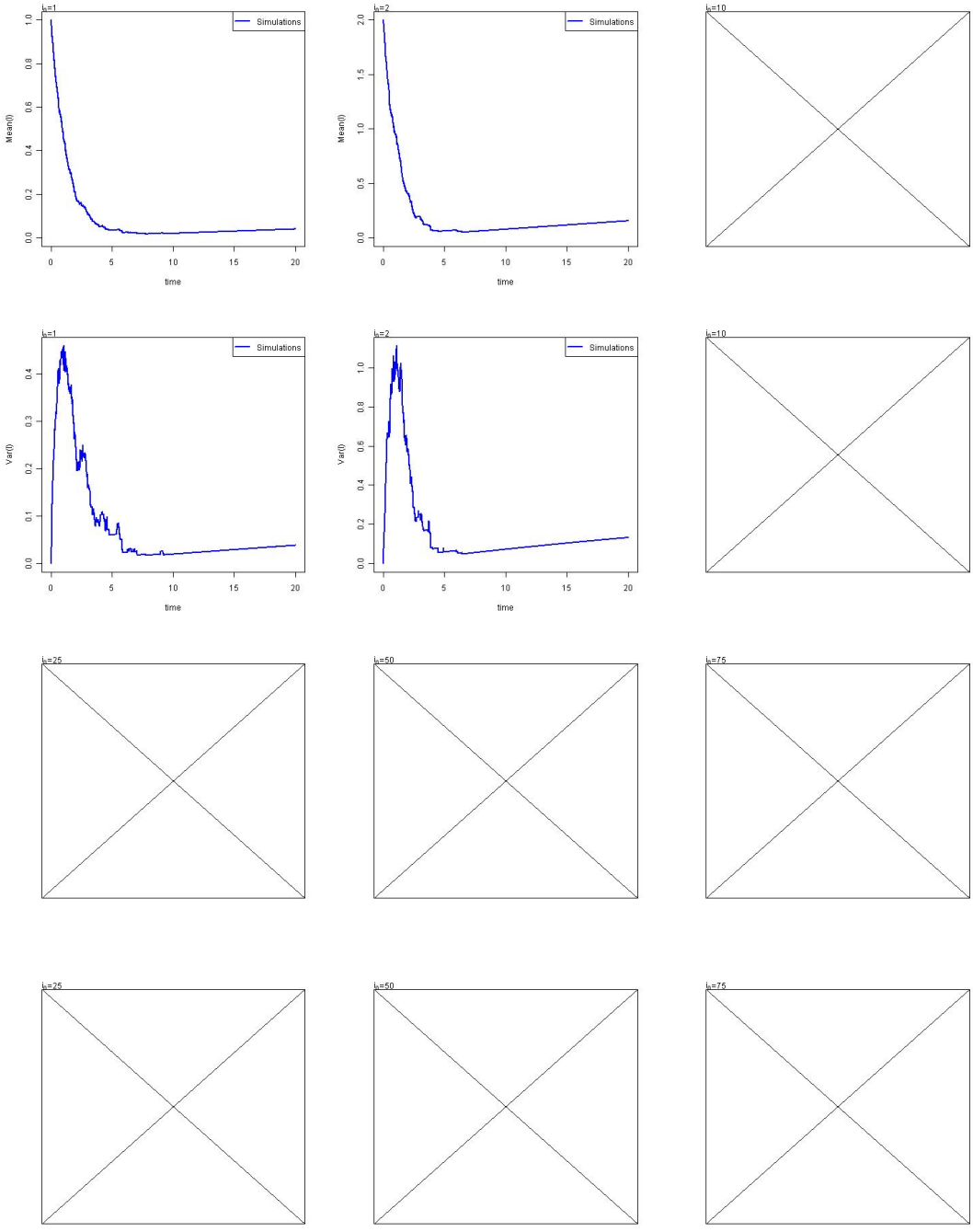

FIGURE 16: $\alpha = 1.2$, $\beta = 0.3$, extinction

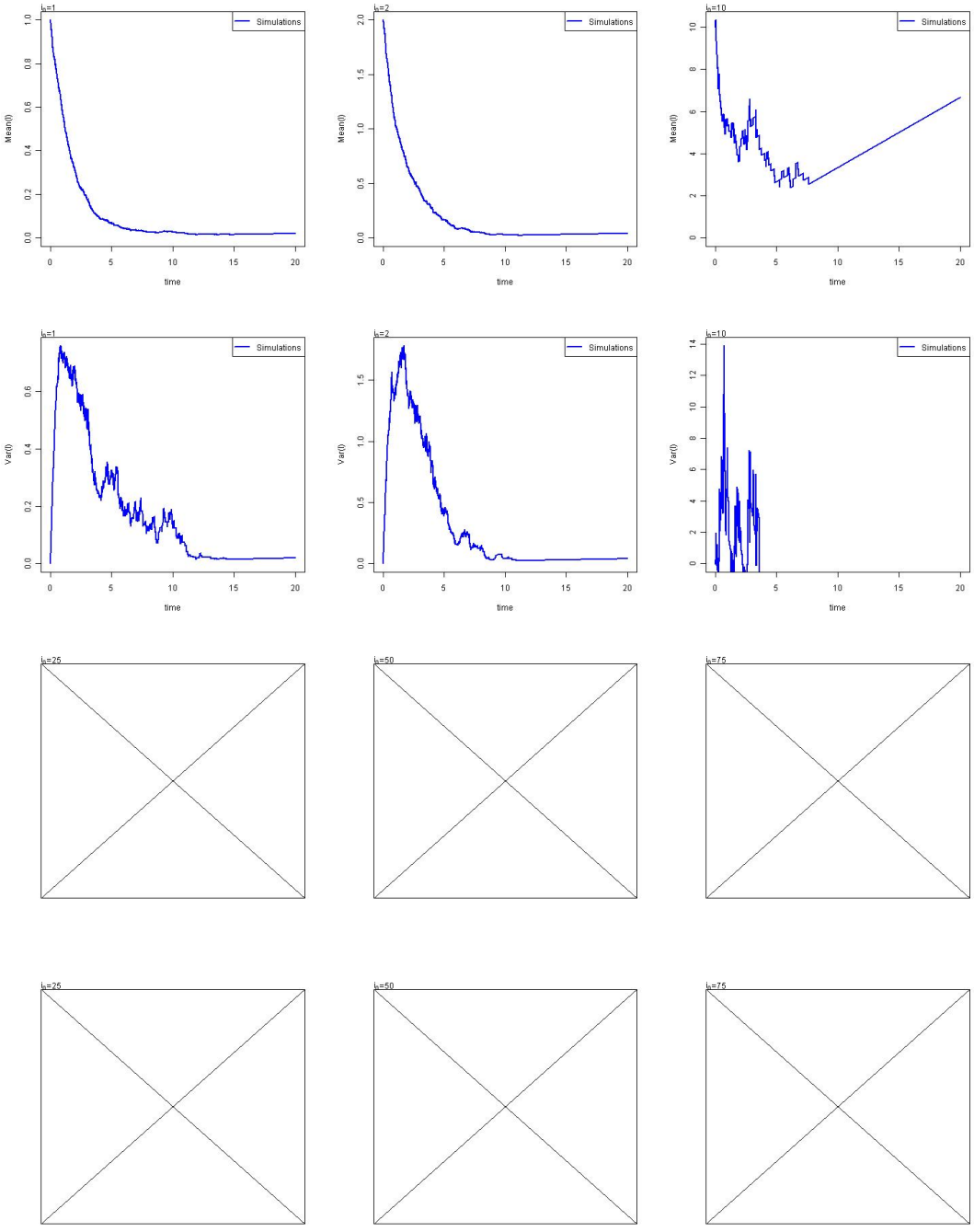

FIGURE 17: $\alpha = 1.2$, $\beta = 0.6$, extinction

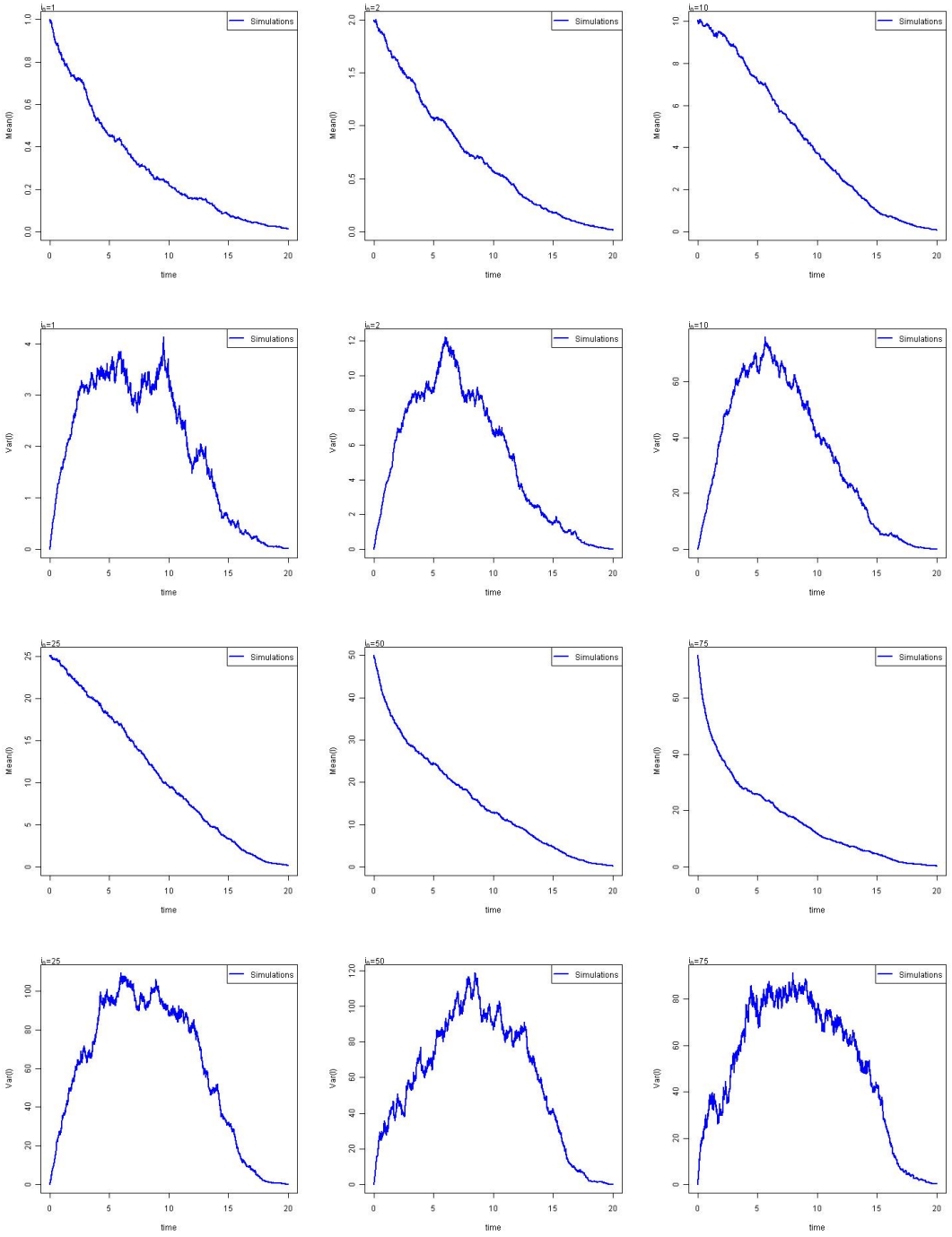

FIGURE 18: $\alpha = 1.2$, $\beta = 0.9$, extinction

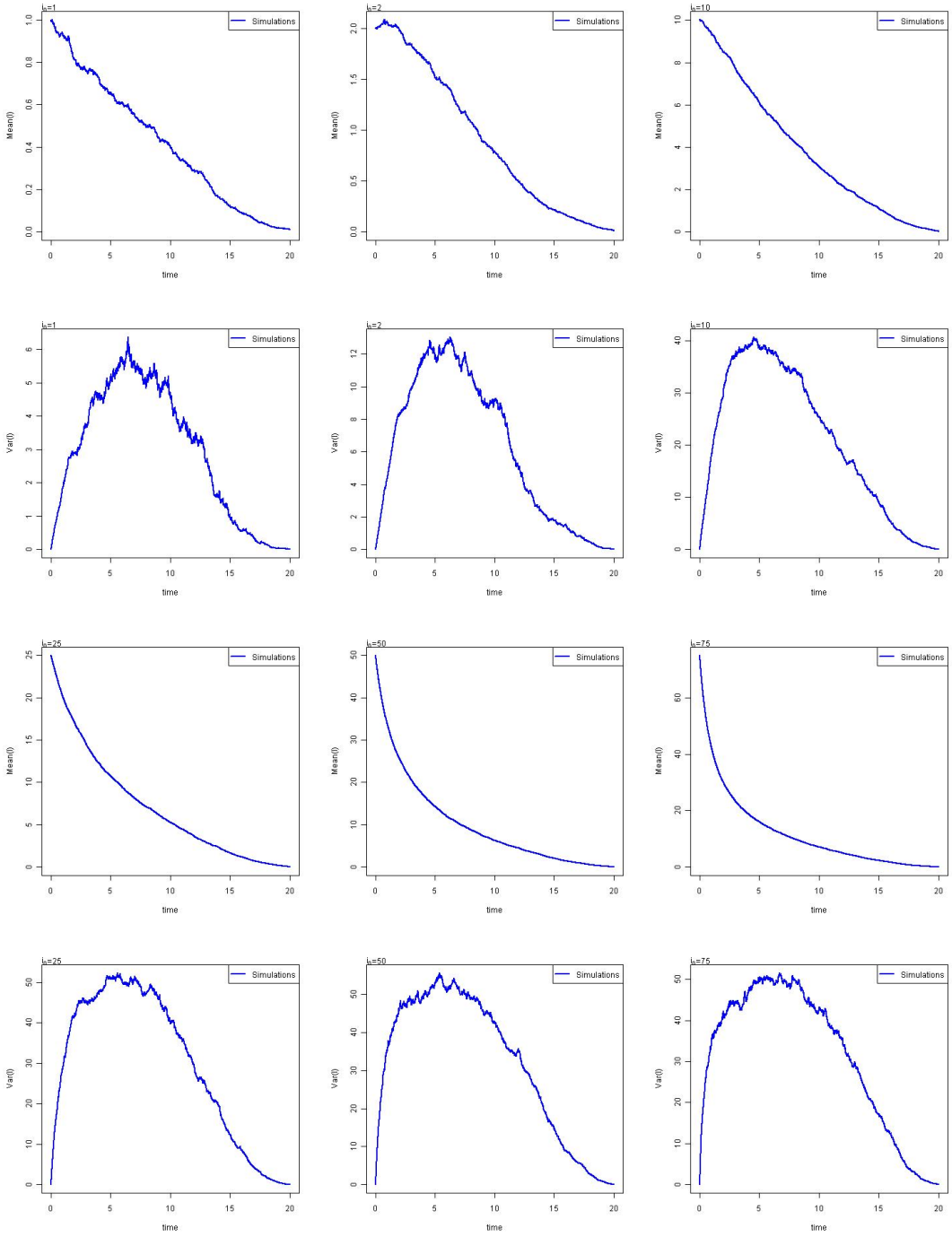

FIGURE 19: $\alpha = 1.2$, $\beta = 1.1$, extinction

TABLE 12: Le nombre des problèmes selon la catégorie

	Conv [1]		Alg [2]		VarNeg [3]		Borne [4]	
« early »	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
1	0	0	1	0	178	319	352	604
2	0	0	2	0	293	102	502	199
3	0	0	0	0	516	86	727	240
4	1	1	41	0	107	208	225	233
5	0	0	4	0	239	106	441	174
6	0	0	0	0	513	101	741	271
7	0	0	34	0	92	190	239	91
8	0	0	3	0	236	132	456	160
9	0	0	8	0	497	70	715	244
10	0	0	0	0	91	356	309	643
11	0	0	0	0	187	91	403	149
12	0	0	0	0	288	41	477	72
13	0	0	5	0	77	199	252	254
14	0	0	0	0	173	131	369	92
15	0	0	0	0	219	81	442	76
16	0	0	4	0	84	261	225	140
17	0	0	3	0	145	125	315	95
18	0	0	0	0	221	18	450	71
« during »	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
1	1	1	82	0	172	343	259	494
2	0	0	88	0	111	64	57	44
3	0	0	35	0	98	36	27	6
4	2	2	125	0	116	170	96	74
5	0	0	49	0	103	96	62	22
6	0	0	25	0	26	19	49	10
7	2	2	105	0	141	259	72	73
8	1	1	48	0	44	79	37	26
9	0	0	11	0	19	51	28	25
10	0	0	20	0	200	326	353	567
11	0	0	3	0	119	117	83	72
12	0	0	0	0	96	68	17	4
13	0	0	31	0	130	186	169	160
14	0	0	0	0	116	155	60	22
15	0	0	0	0	48	56	5	1
16	1	1	15	5	211	502	147	414
17	1	1	4	0	98	175	37	26
18	0	0	0	0	20	24	6	7
« after »	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
1	0	0	66	4	519	398	613	691
2	1	1	45	0	487	390	91	491
3	0	0	10	3	554	456	56	252
4	4	4	78	2	434	403	140	327
5	2	2	11	3	550	445	54	222
6	2	2	10	3	566	541	24	57
7	1	1	42	9	482	500	108	205
8	2	2	45	17	568	512	16	178
9	2	2	14	23	594	576	11	43
10	0	0	44	3	455	379	615	569
11	0	0	4	1	506	488	108	415
12	0	0	4	0	520	486	71	42
13	0	0	21	1	427	446	188	238
14	2	2	6	4	583	551	79	85
15	0	0	9	1	551	559	41	8
16	0	0	31	39	521	607	142	346
17	1	1	58	5	577	579	41	94
18	3	3	38	10	587	667	20	7

[1] Conv - Problème de Convergence d'algorithme de Nealder-Mead dans ConstrOptim

[2] Alg - Échec d'estimations malgré la convergence

[3] NegVar - Estimation négative de la variance asymptotique au moins d'un paramètre

[4] Borne - Les estimations réalisant une de contrainte fixée dans ConstrOptim ou $\alpha < \beta$

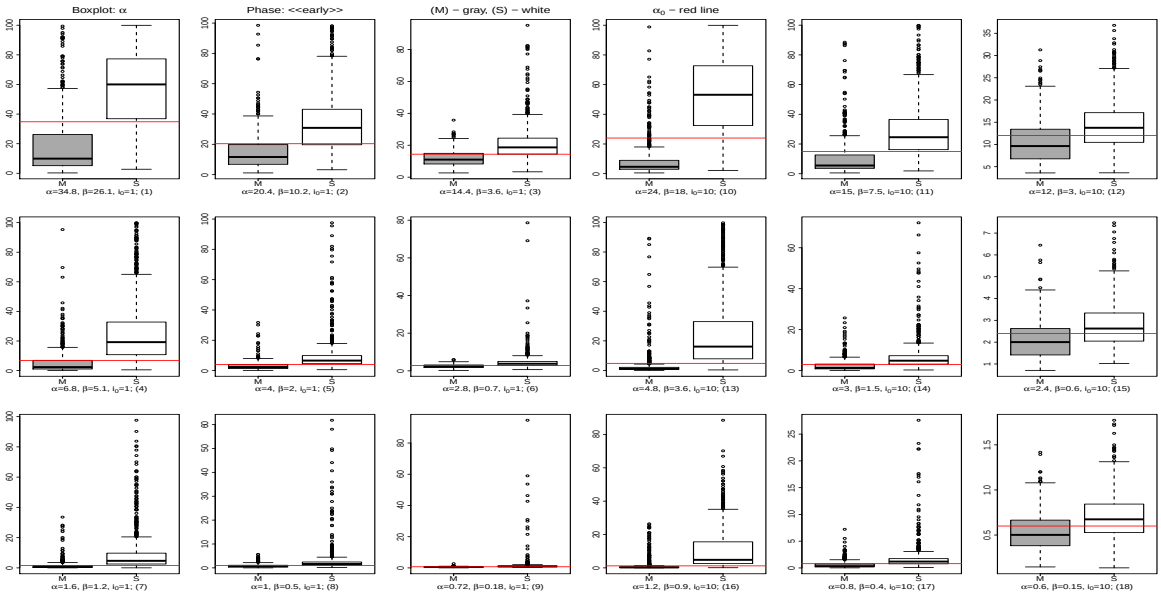

FIGURE 20: Boxplots des estimations de α pour les observations de la phase initale : « early »

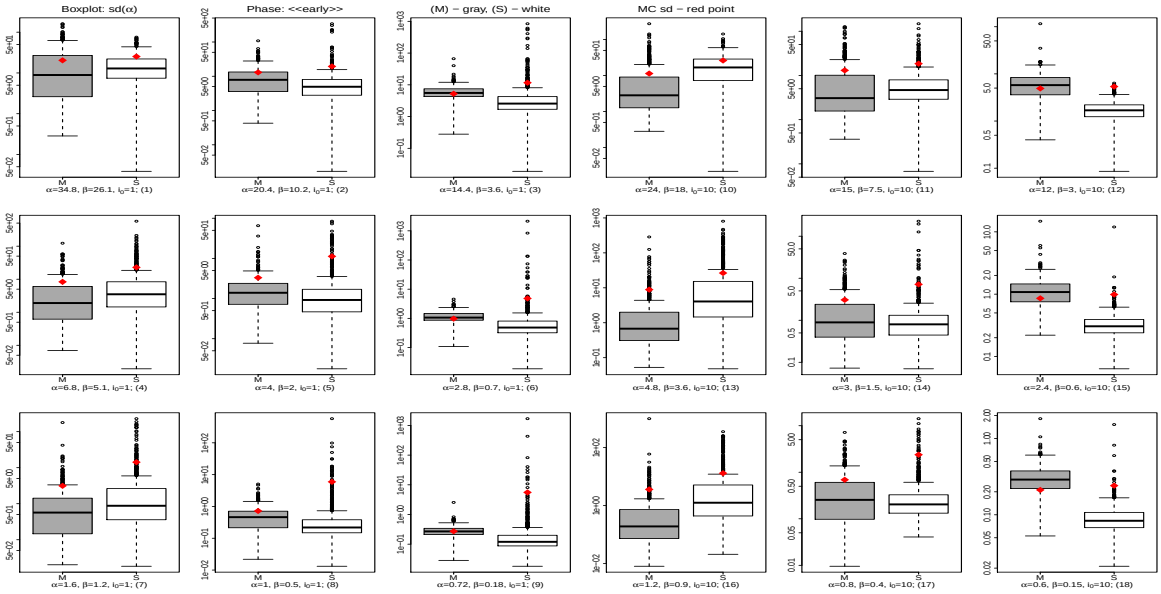

FIGURE 21: Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase initale : « early »

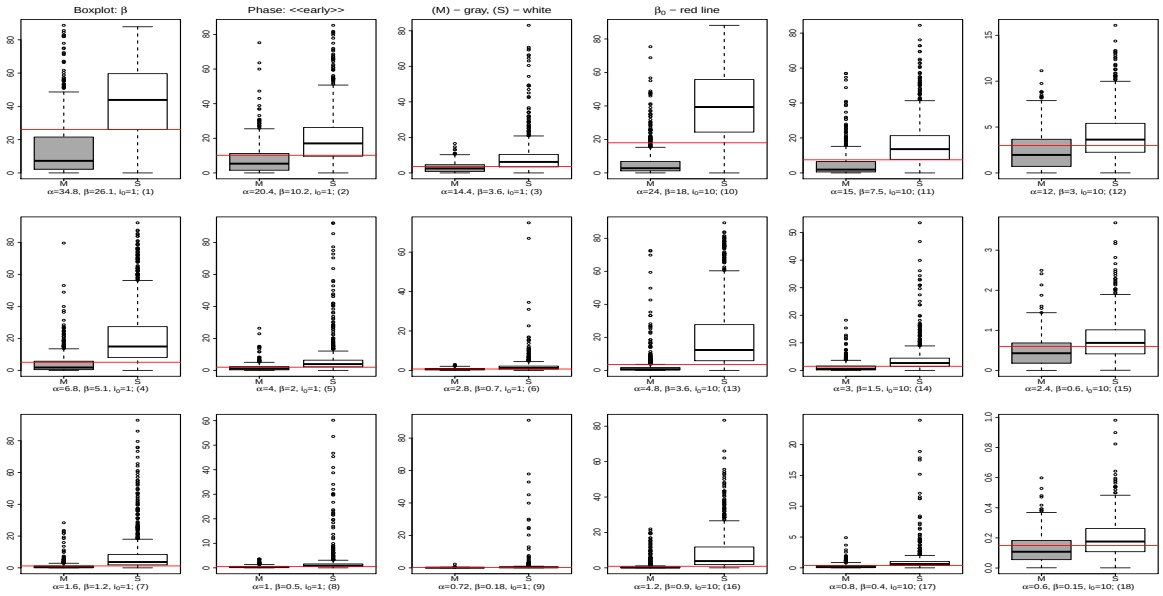

FIGURE 22: Boxplots des estimations de β pour les observations de la phase initiale : « early »

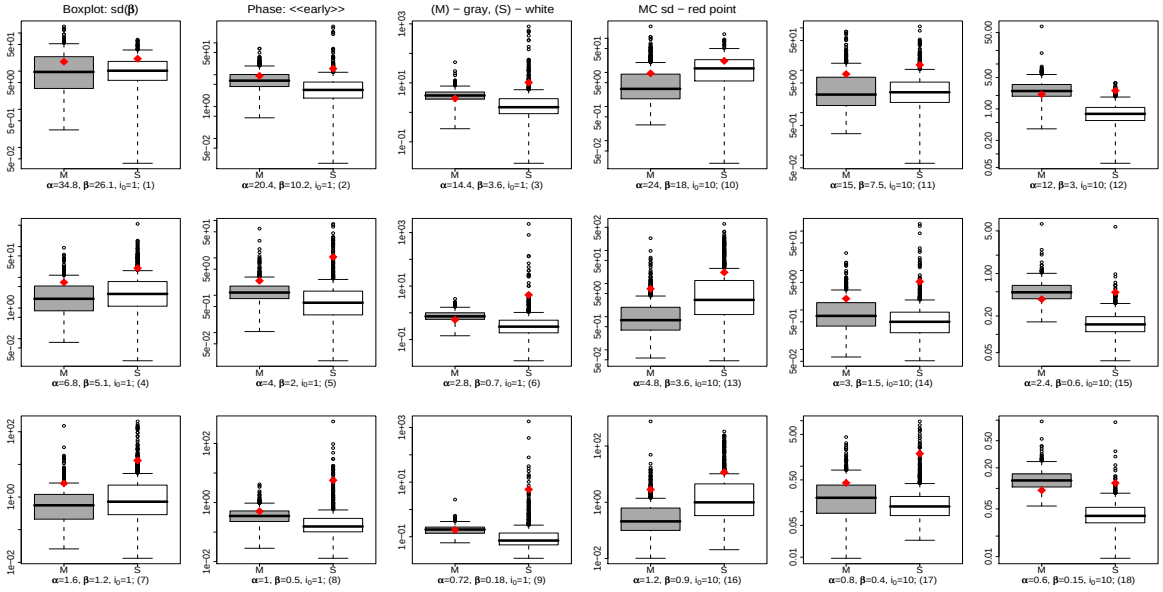

FIGURE 23: Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase initiale : « early »

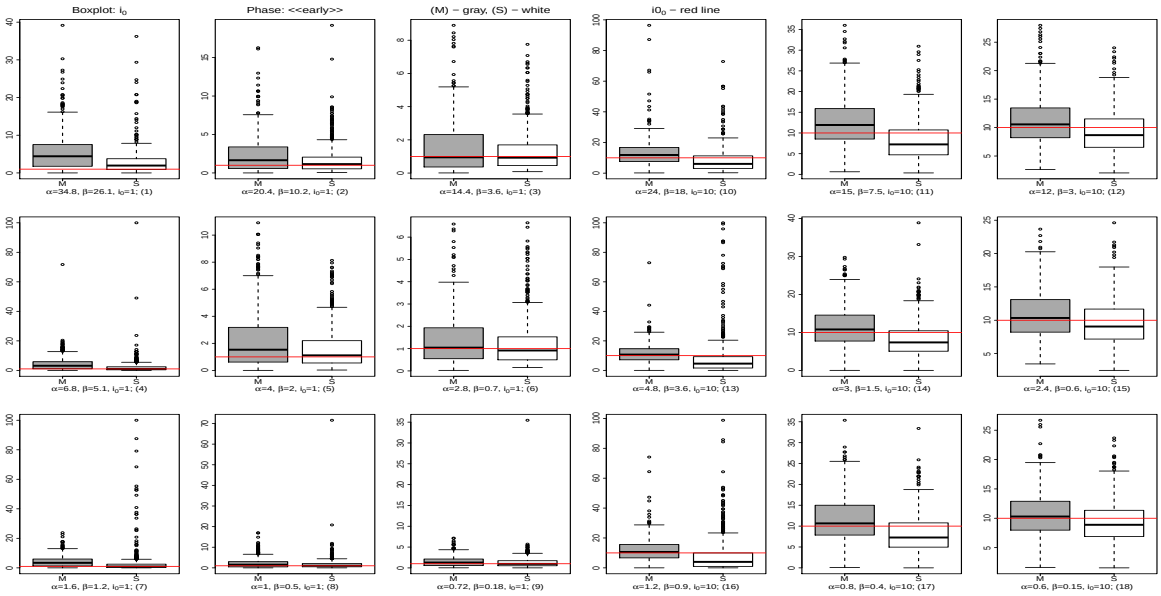

FIGURE 24: Boxplots des estimations de i_0 pour les observations de la phase initiale : « early »

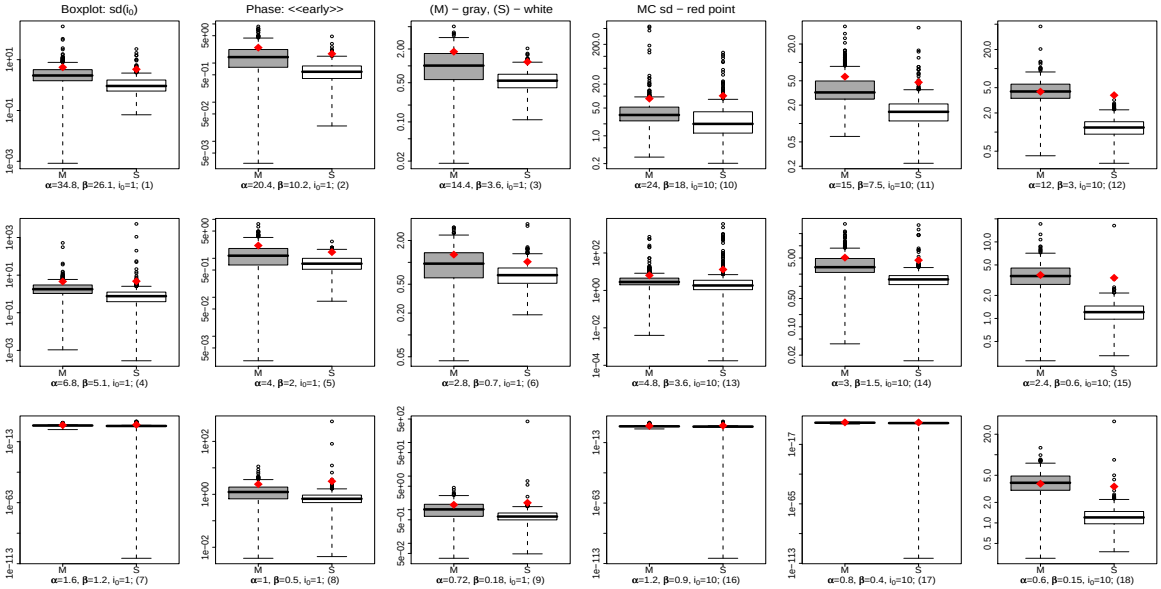

FIGURE 25: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase initiale : « early »

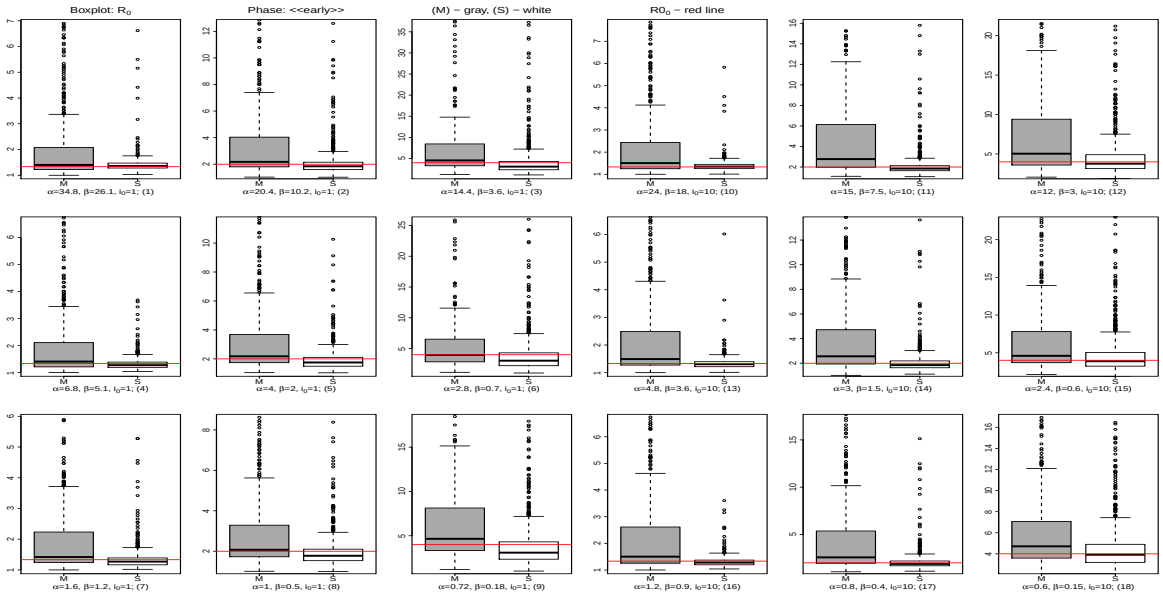

FIGURE 26: Boxplots des estimations de R_0 pour les observations de la phase initiale : « early »

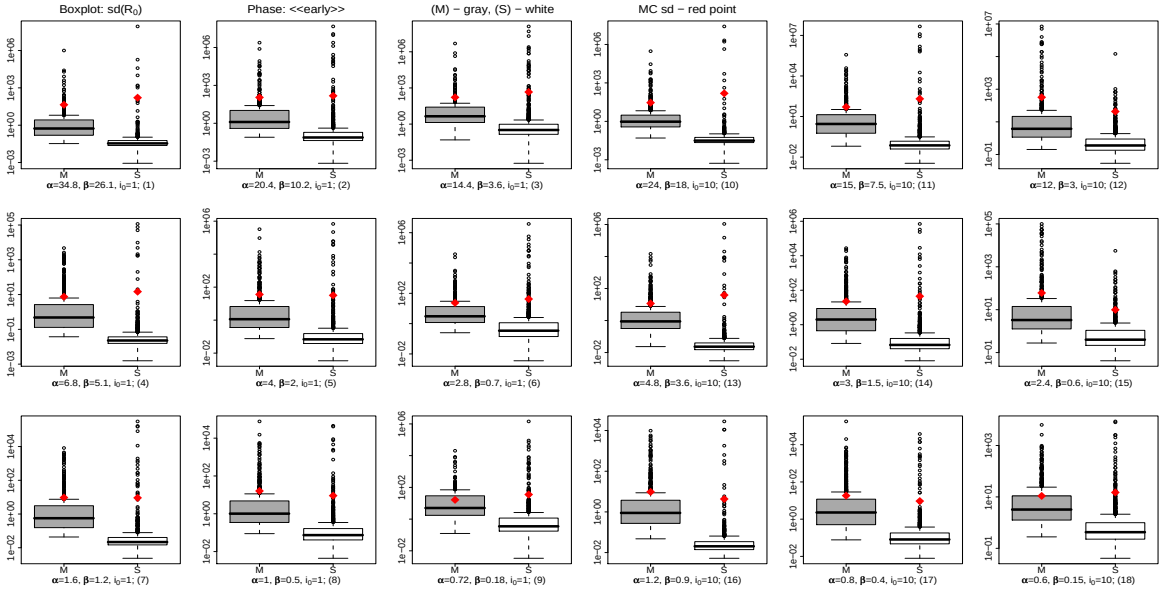

FIGURE 27: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase initiale : « early »

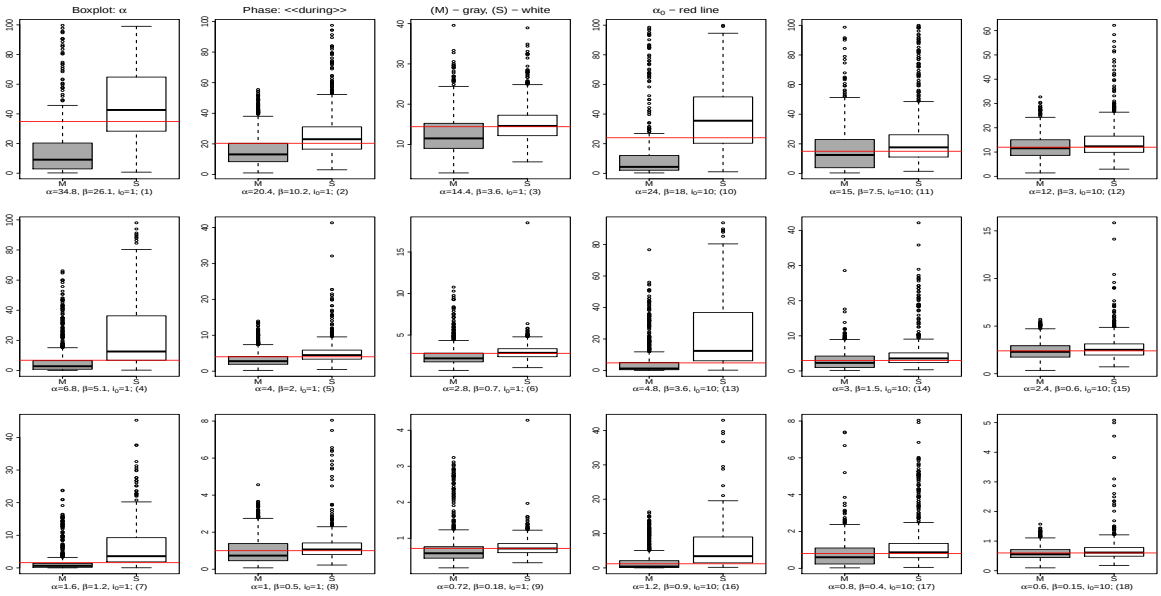

FIGURE 28: Boxplots des estimations de α pour les observations de la phase évolutive : « during »

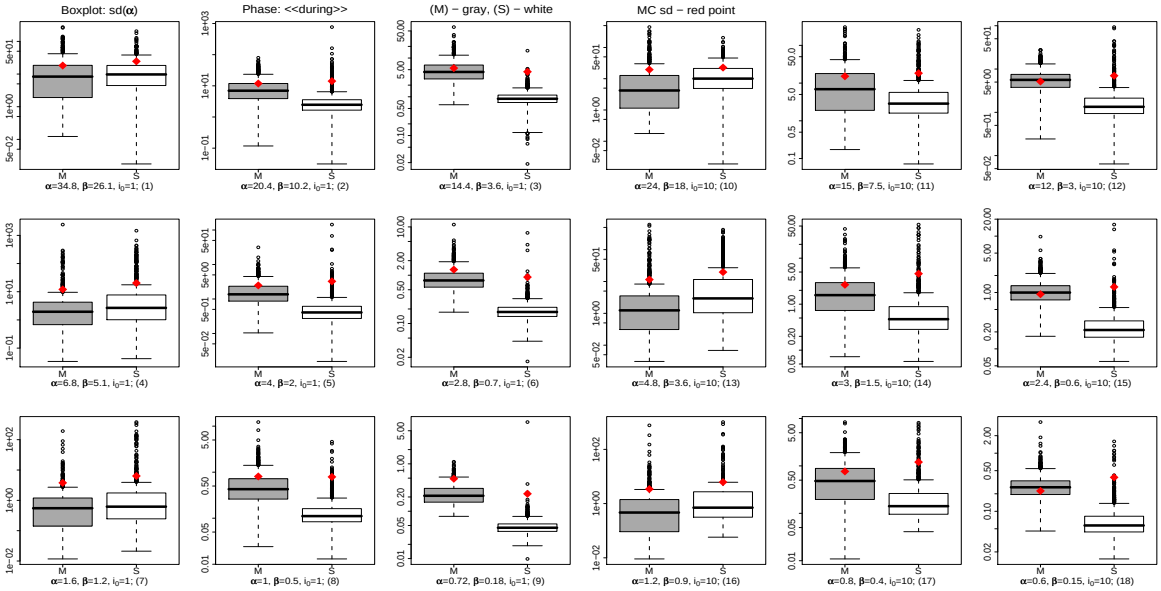

FIGURE 29: Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase évolutive : « during »

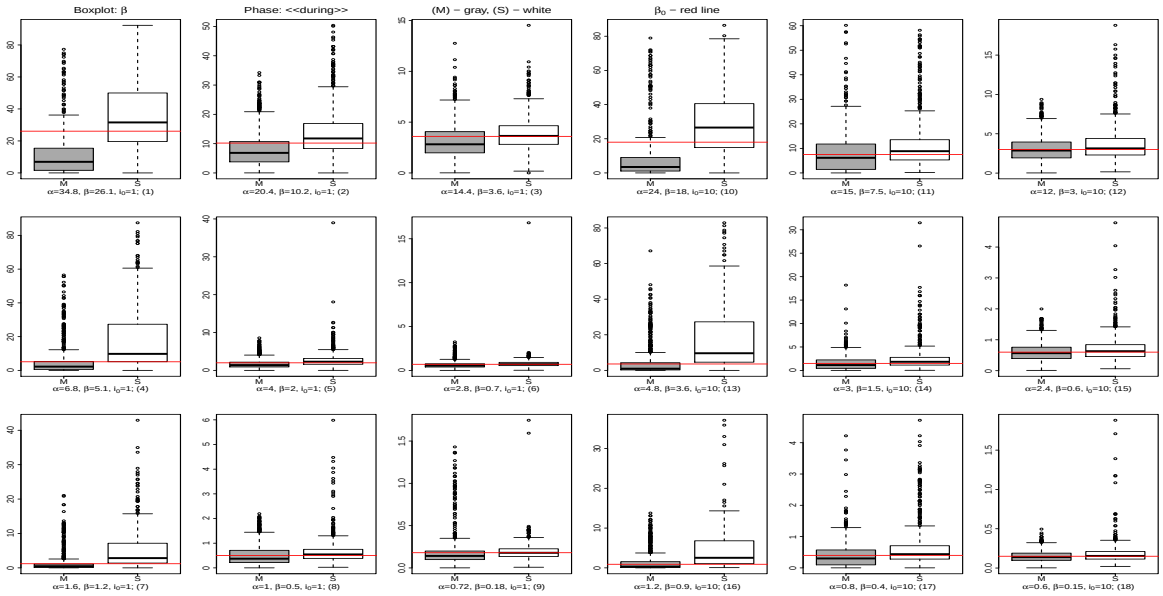

FIGURE 30: Boxplots des estimations de β pour les observations de la phase évolutive : « during »

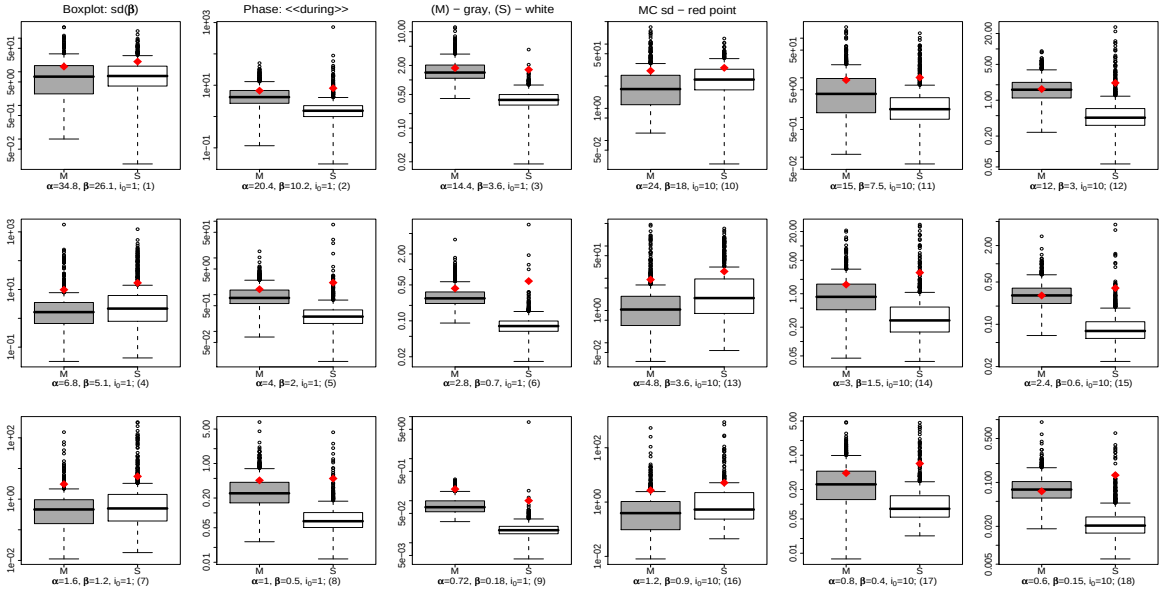

FIGURE 31: Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase évolutive : « during »

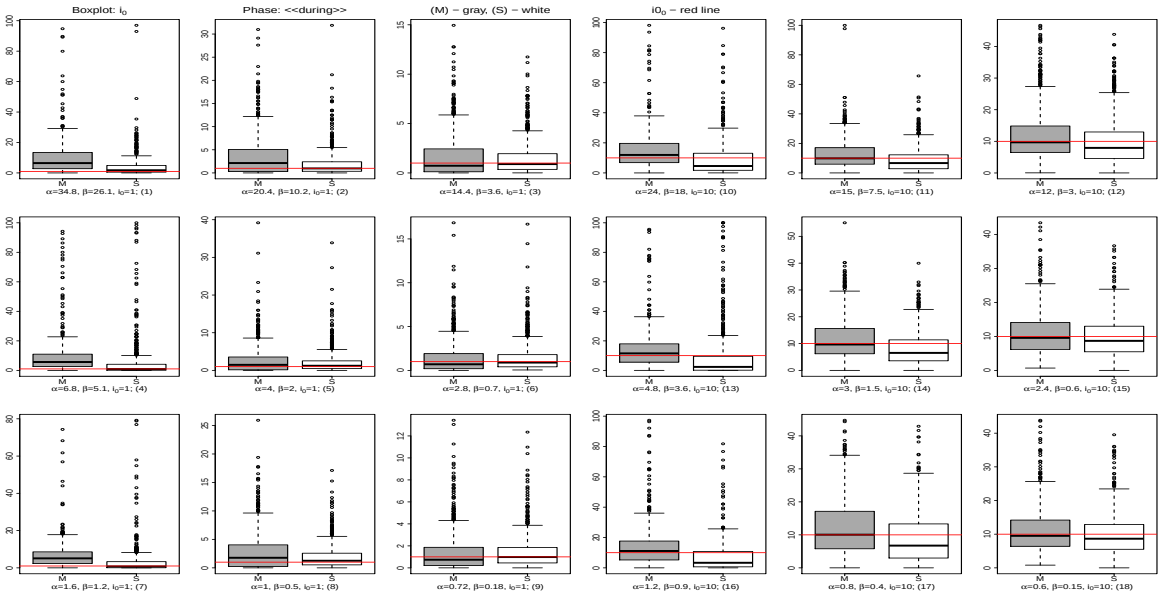

FIGURE 32: Boxplots des estimations de i_0 pour les observations de la phase évolutive : « during »

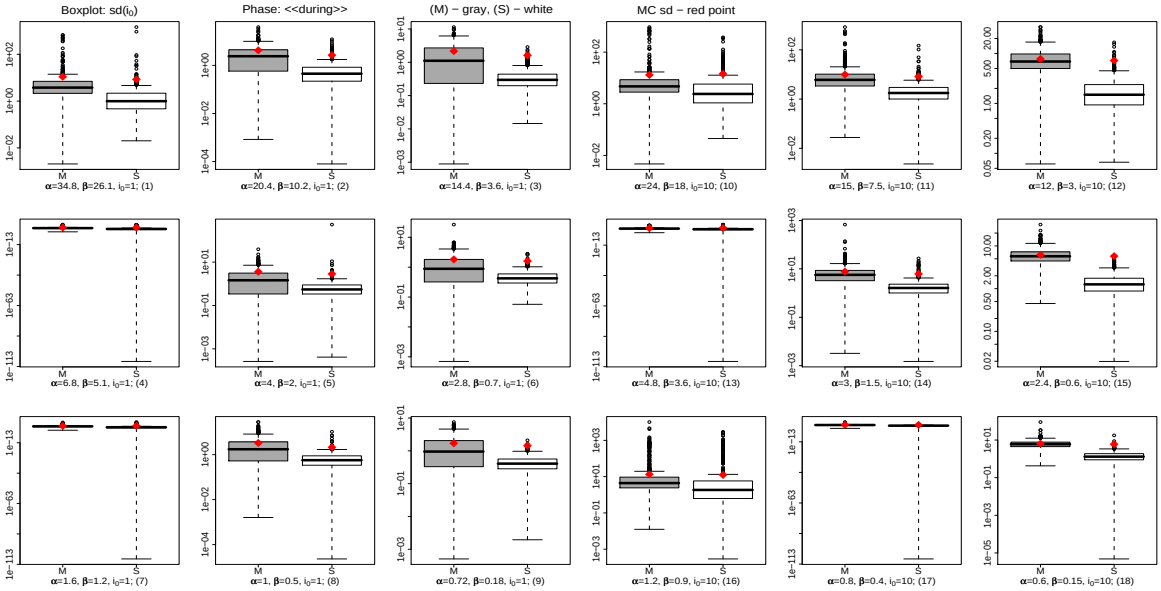

FIGURE 33: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase évolutive : « during »

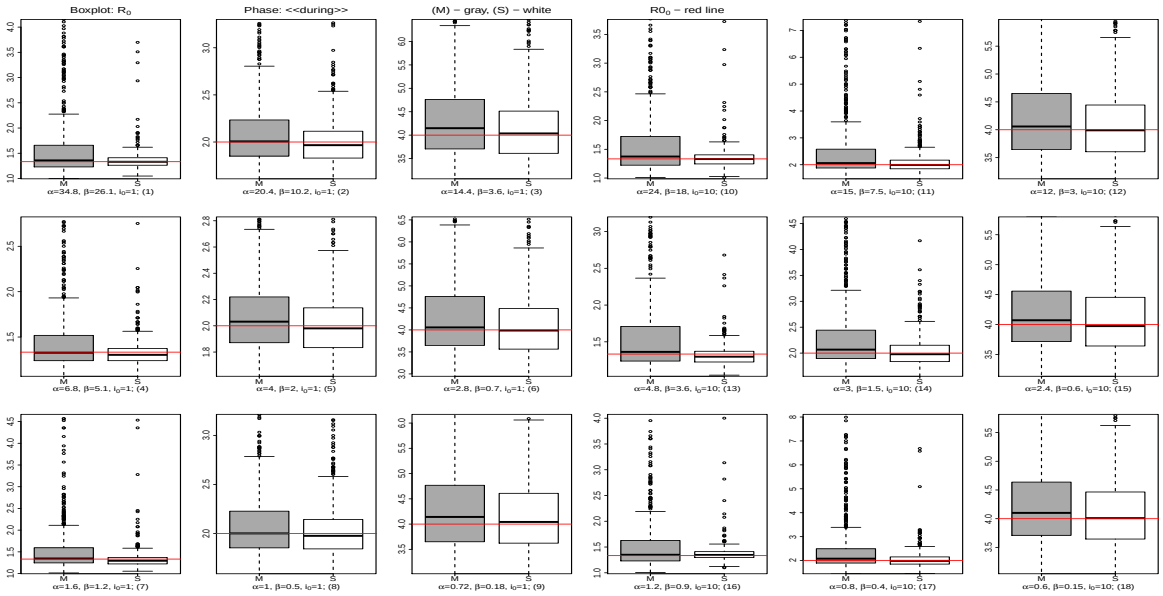

FIGURE 34: Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during »

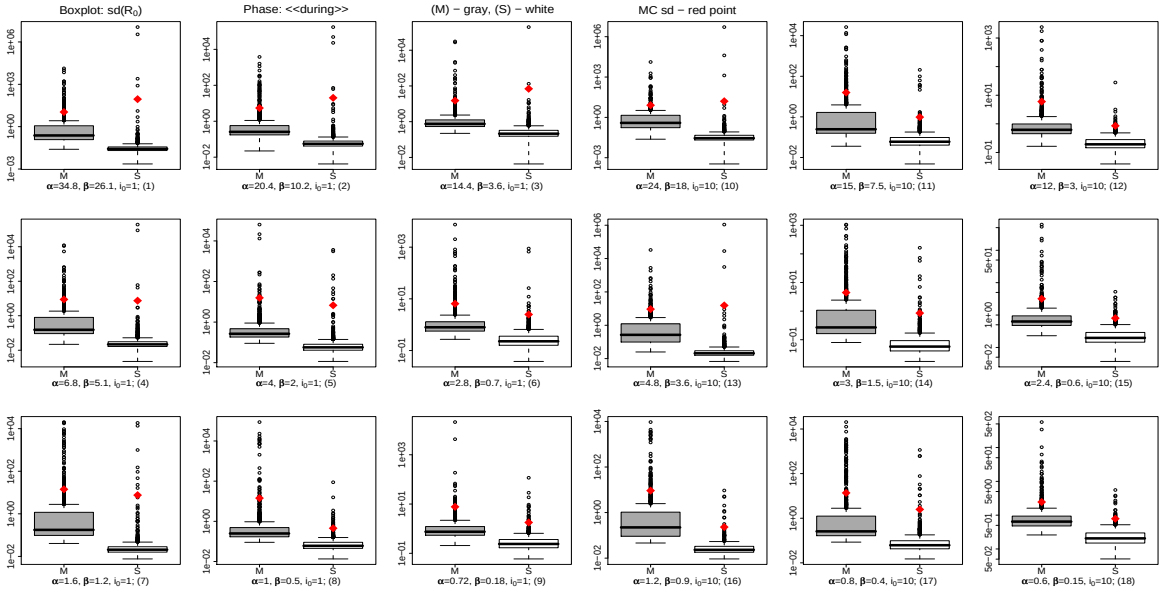

FIGURE 35: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during »

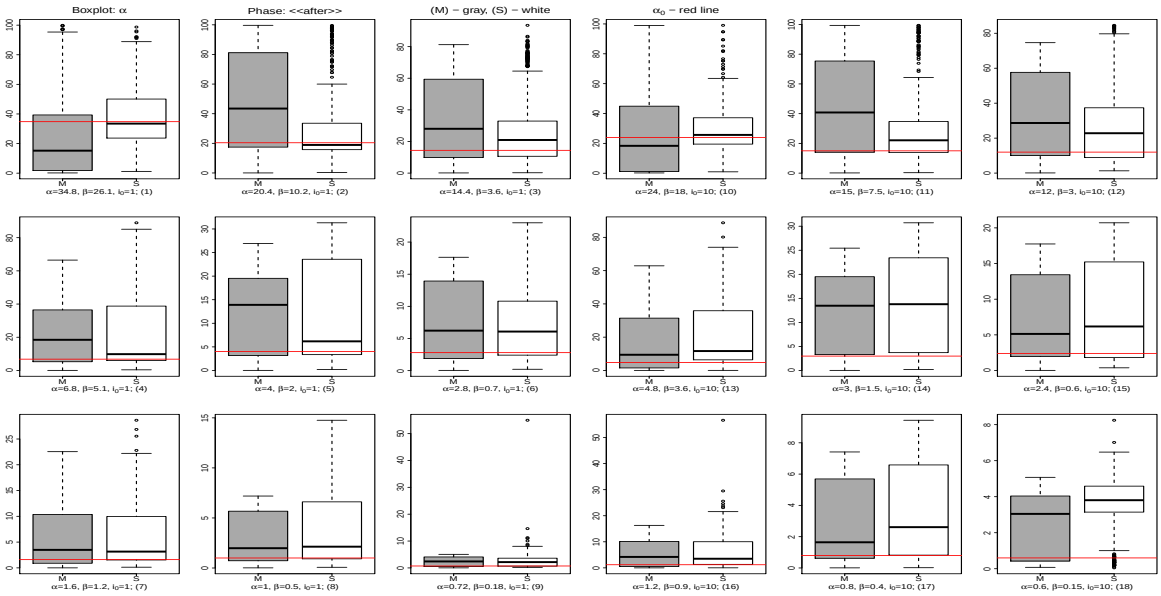

FIGURE 36: Boxplots des estimations de α pour les observations de la phase stationnaire : « after »

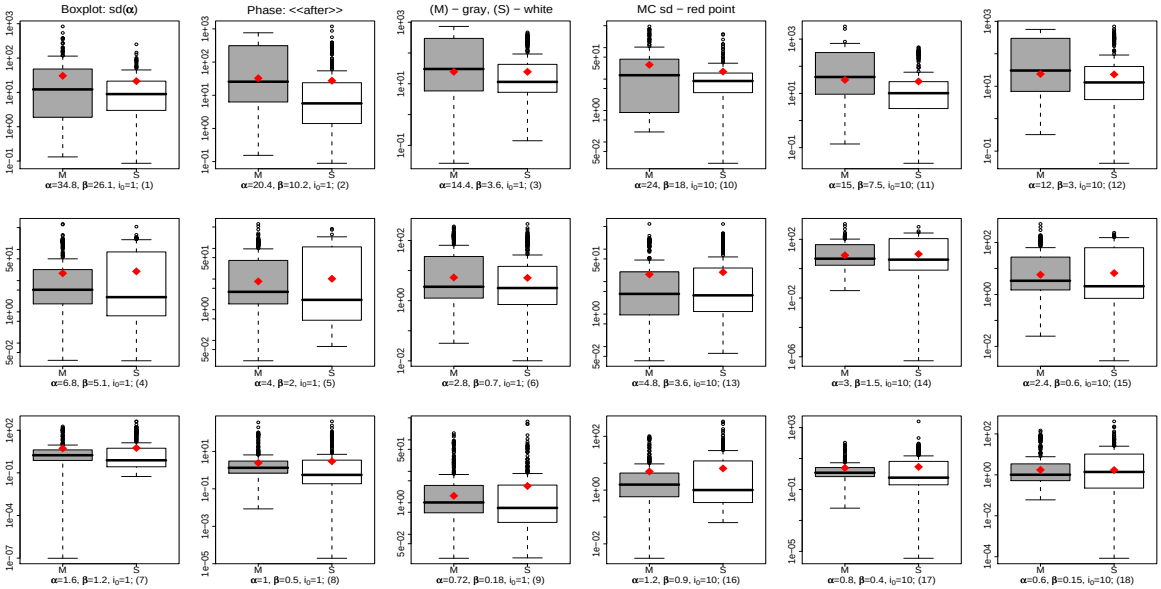

FIGURE 37: Boxplots des estimations de l'écart-type asymptotique de α pour les observations de la phase stationnaire : « after »

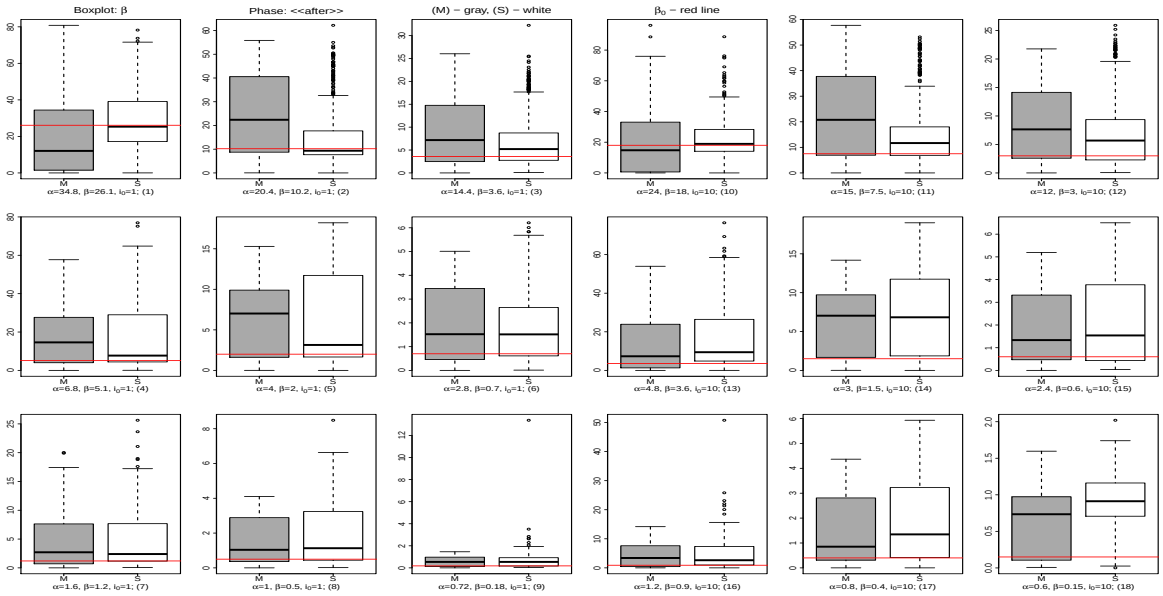

FIGURE 38: Boxplots des estimations de β pour les observations de la phase stationnaire : « after »

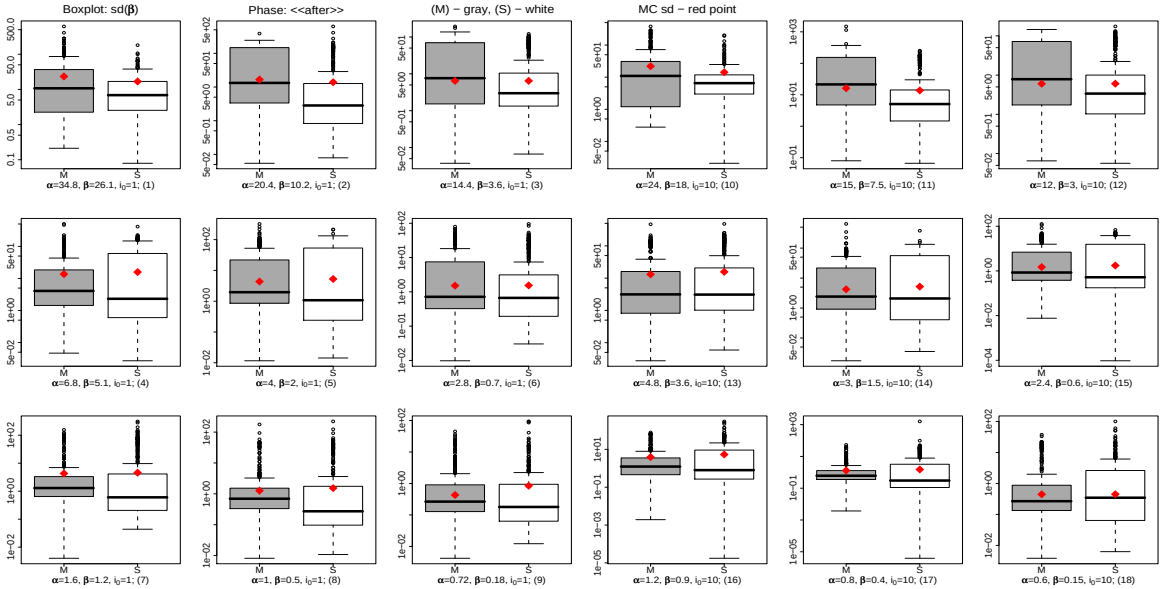

FIGURE 39: Boxplots des estimations de l'écart-type asymptotique de β pour les observations de la phase stationnaire : « after »

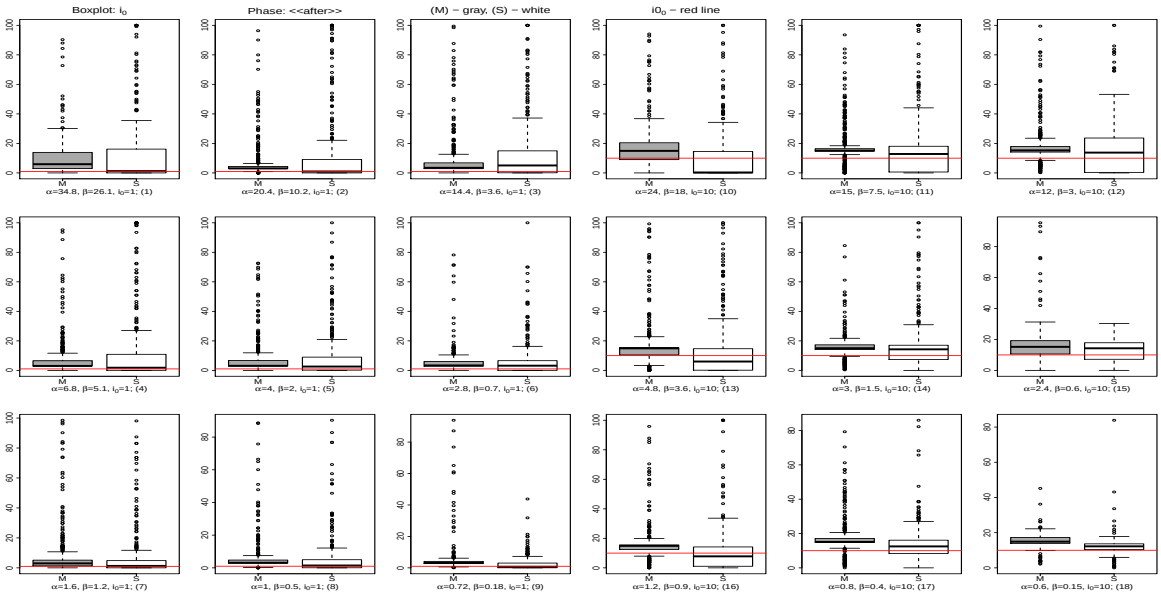

FIGURE 40: Boxplots des estimations de i_0 pour les observations de la phase stationnaire : « after »

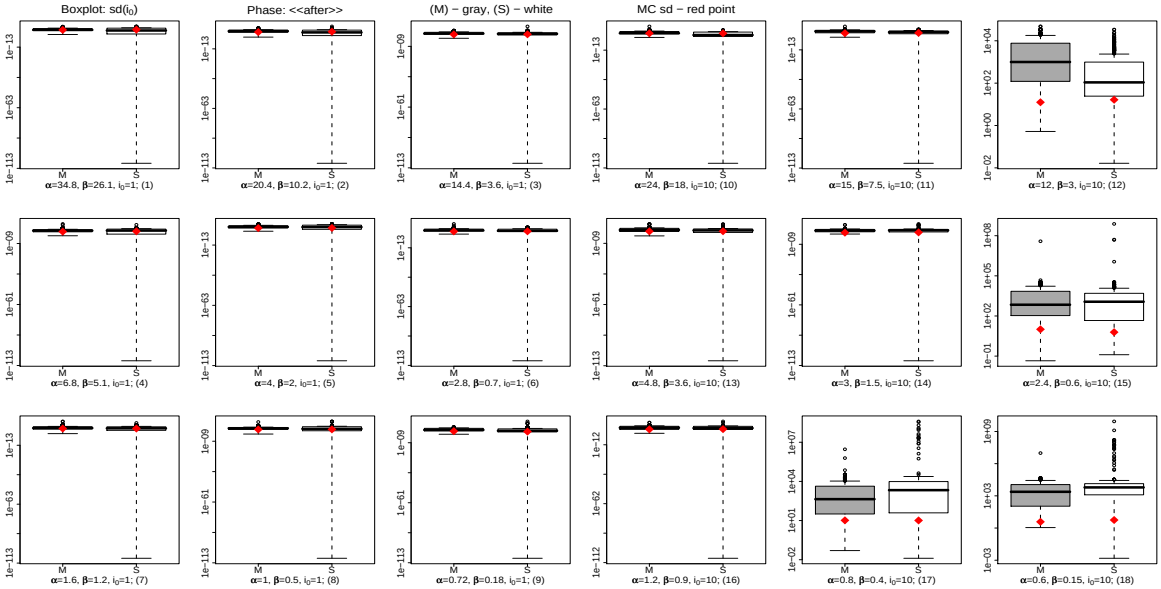

FIGURE 41: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase stationnaire : « after »

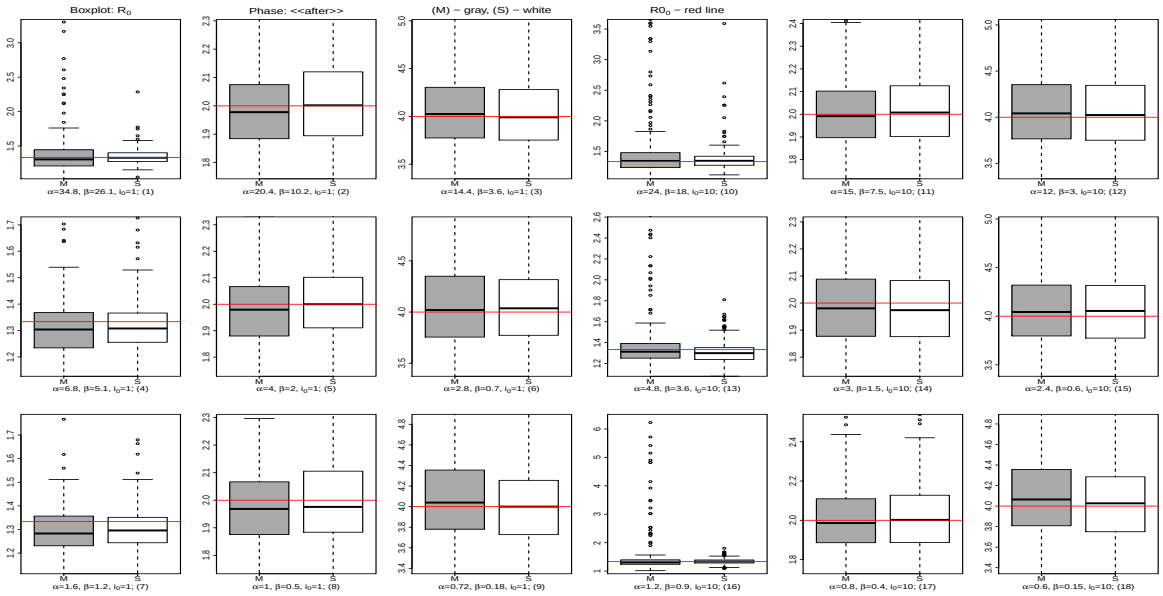

FIGURE 42: Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « after »

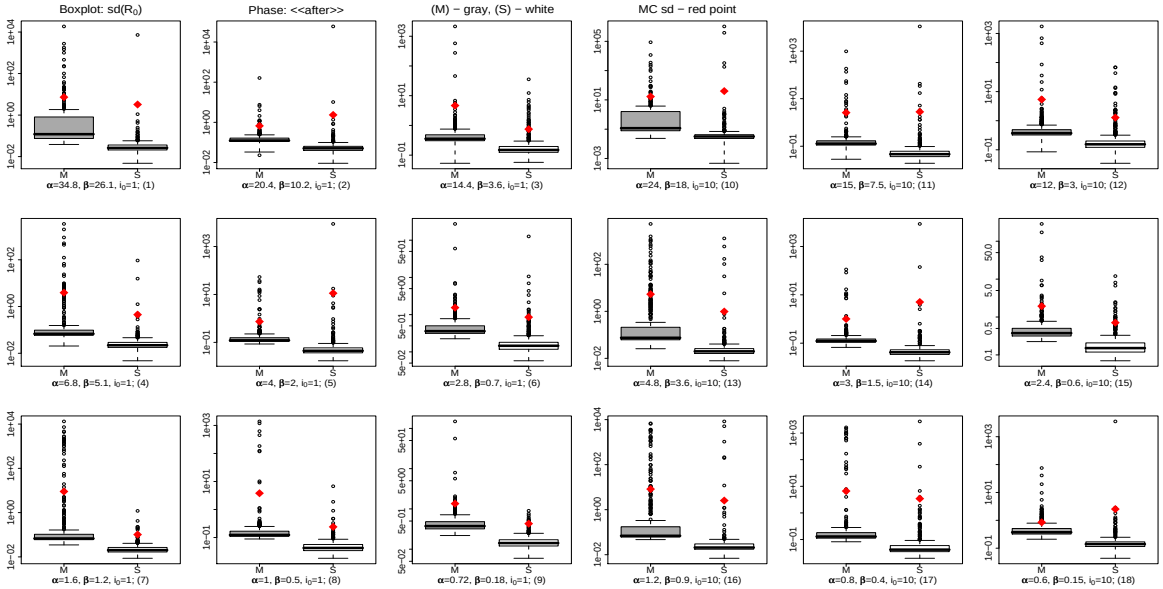

FIGURE 43: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « after »

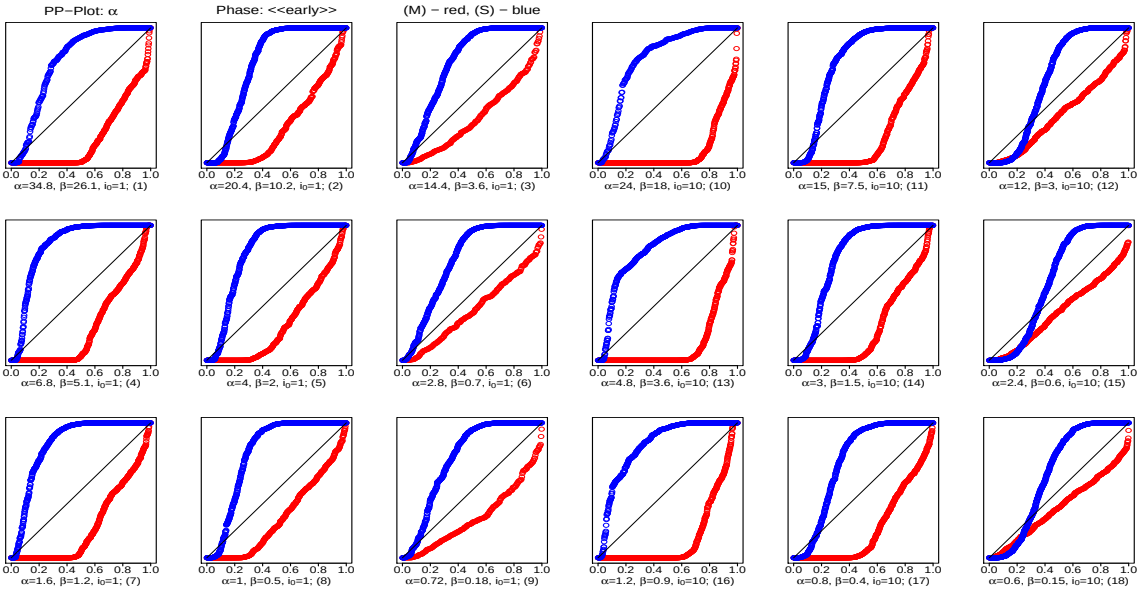

FIGURE 44: PP-plots de $\theta = \alpha$ pour les observations de la phase initiale : « early »

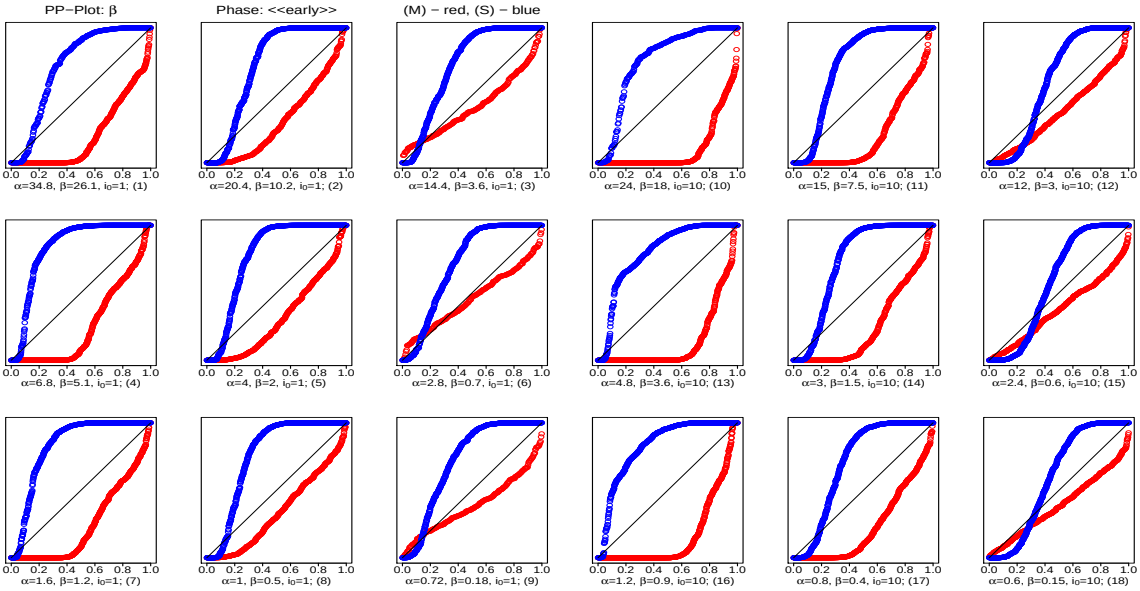

FIGURE 45: PP-plots de $\theta = \beta$ pour les observations de la phase initiale : « early »

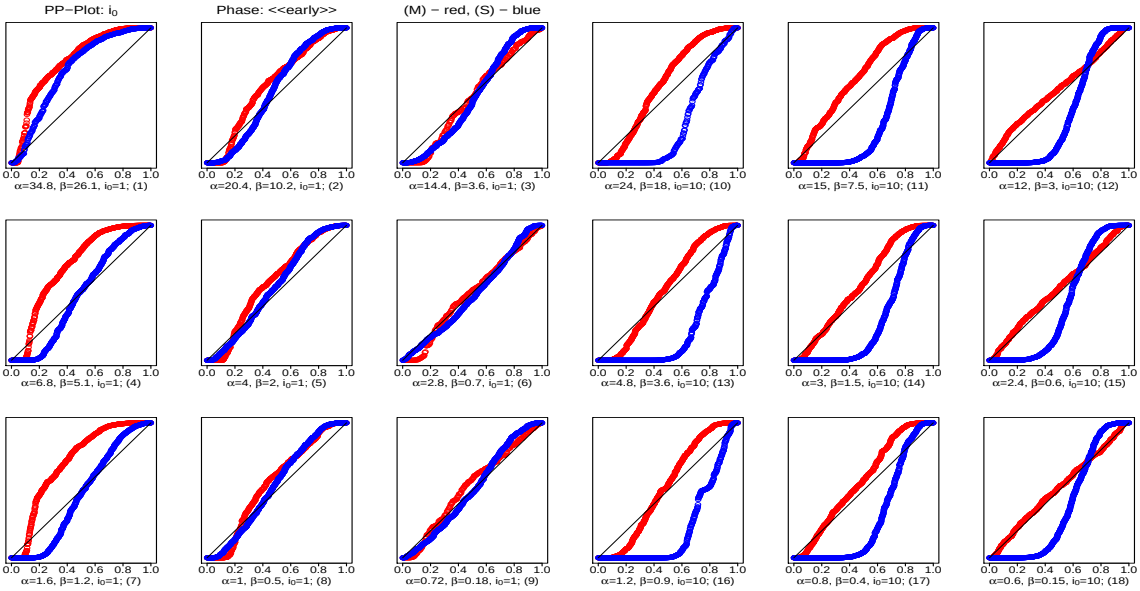

FIGURE 46: PP-plots de $\theta = i_0$ pour les observations de la phase initiale : « early »

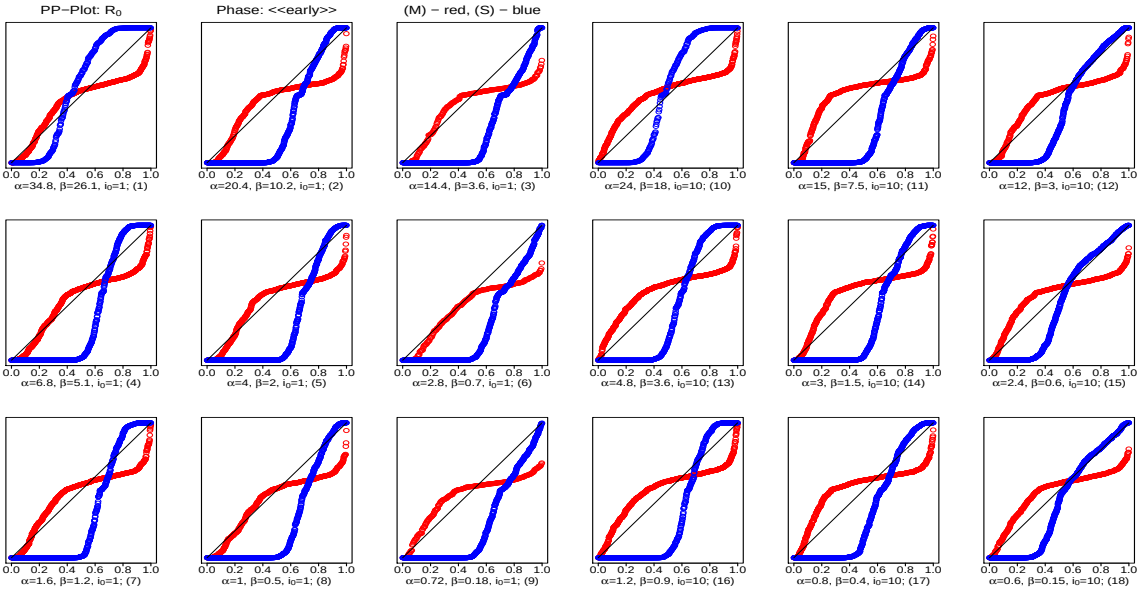

FIGURE 47: PP-plots de R_0 pour les observations de la phase initiale : « early »

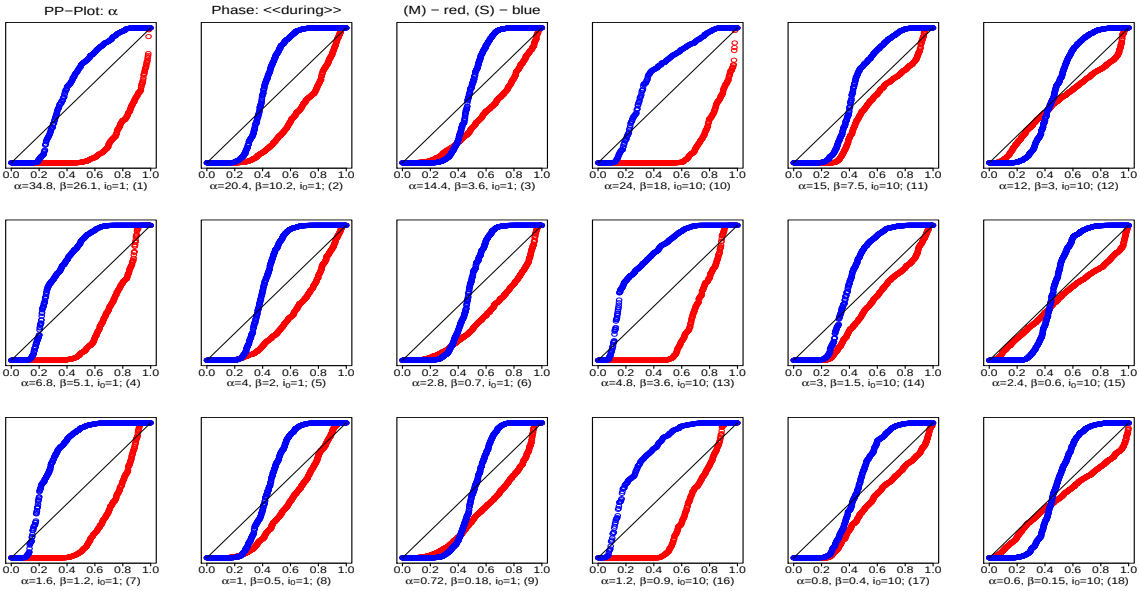

FIGURE 48: PP-plots de $\theta = \alpha$ pour les observations de la phase évolutive : « during »

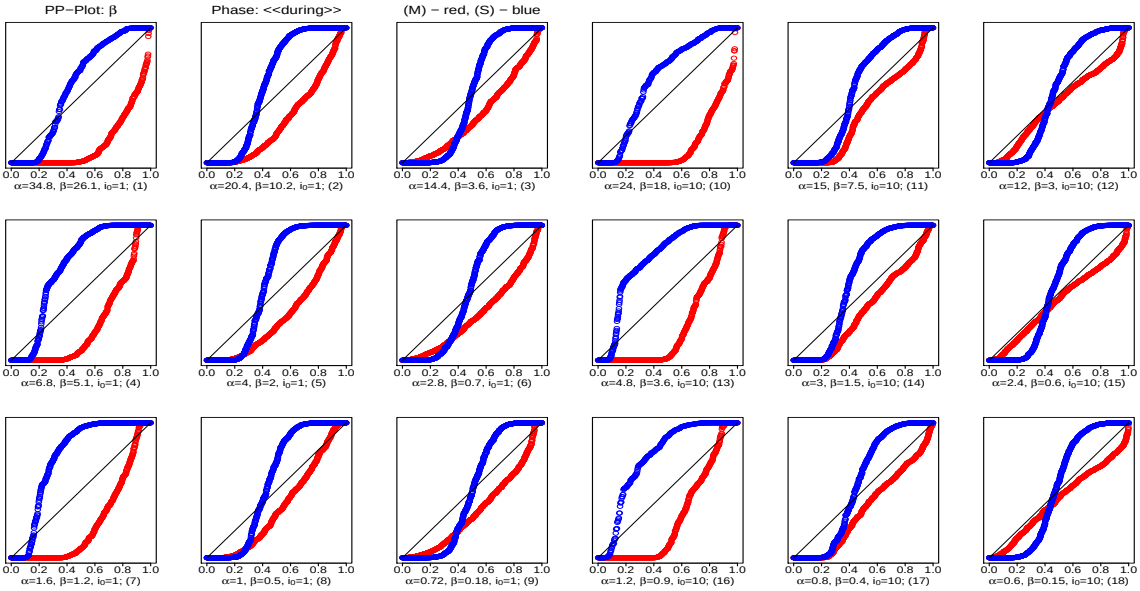

FIGURE 49: PP-plots de $\theta = \beta$ pour les observations de la phase evolutive : « during »

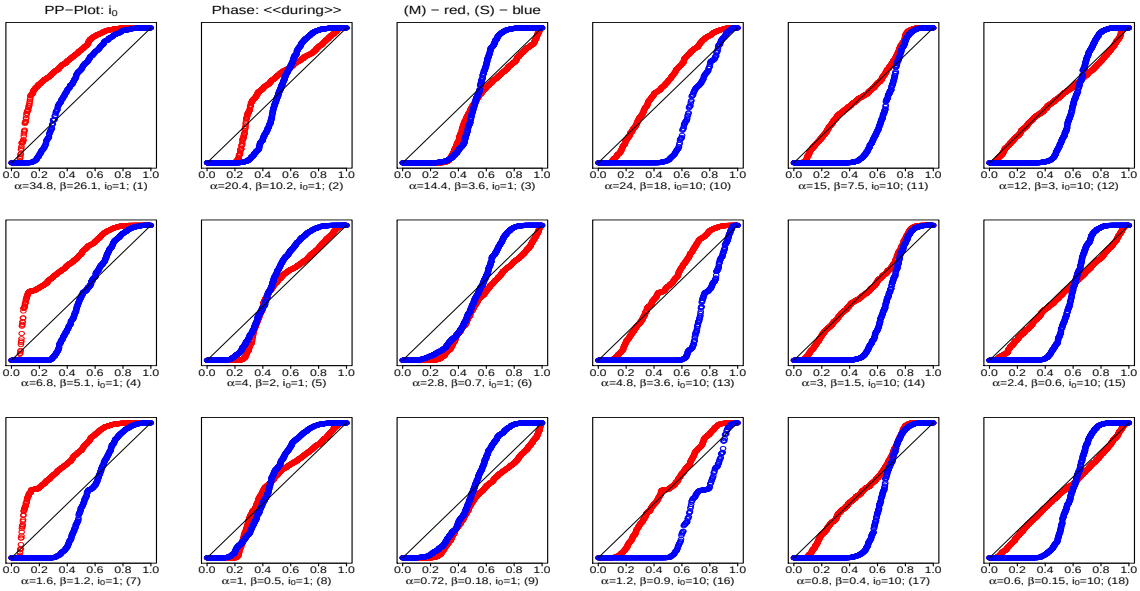

FIGURE 50: PP-plots de $\theta = i_0$ pour les observations de la phase evolutive : « during »

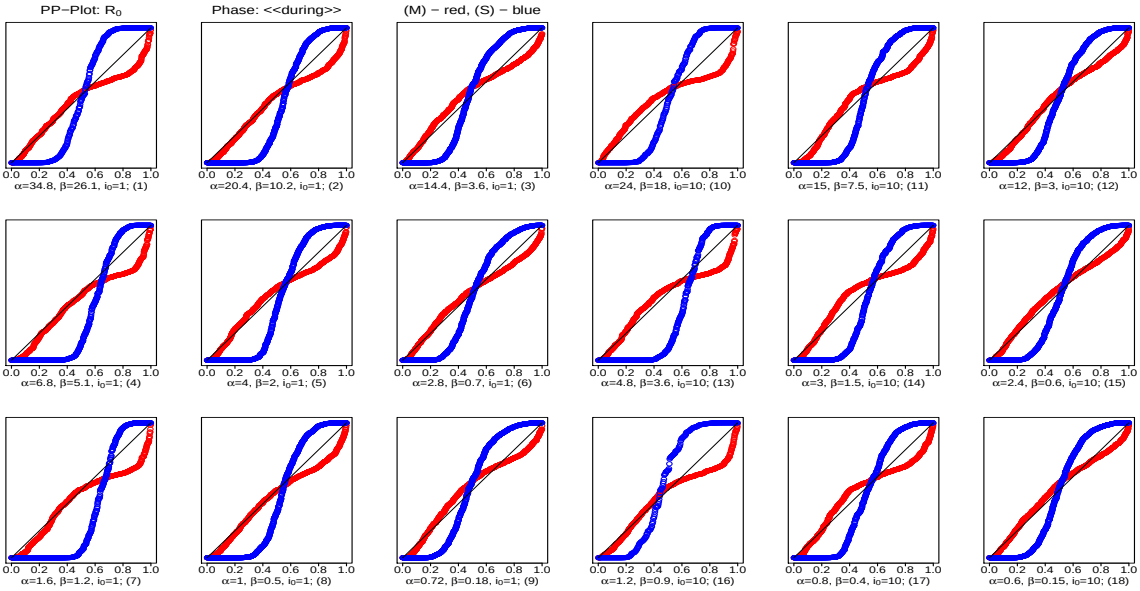

FIGURE 51: PP-plots de R_0 pour les observations de la phase evolutive : « during »

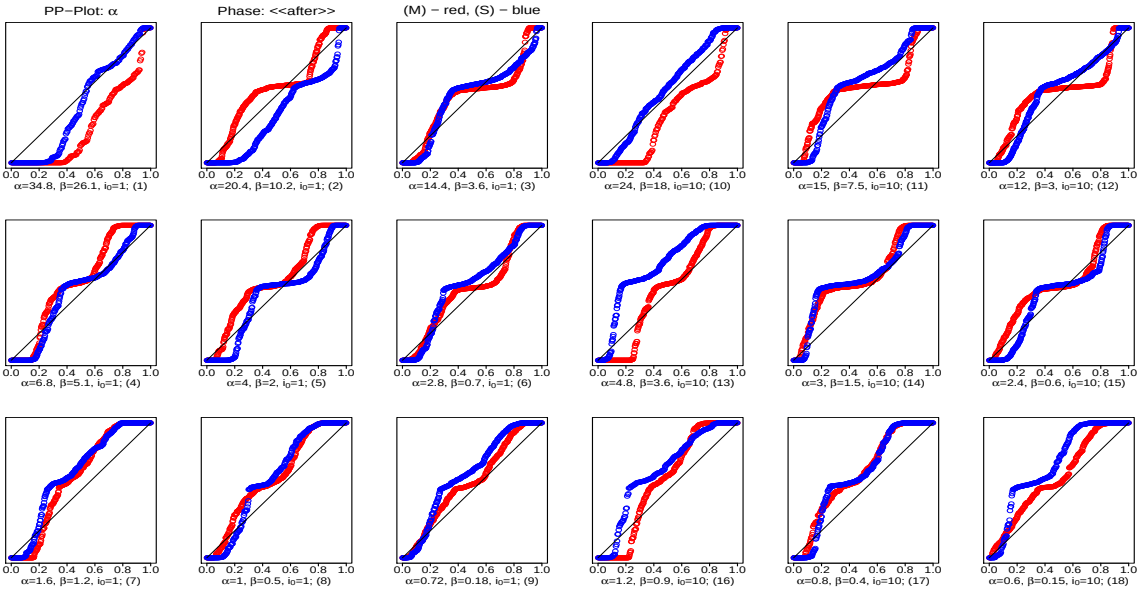

FIGURE 52: PP-plots de $\theta = \alpha$ pour les observations de la phase stationnaire : « after »

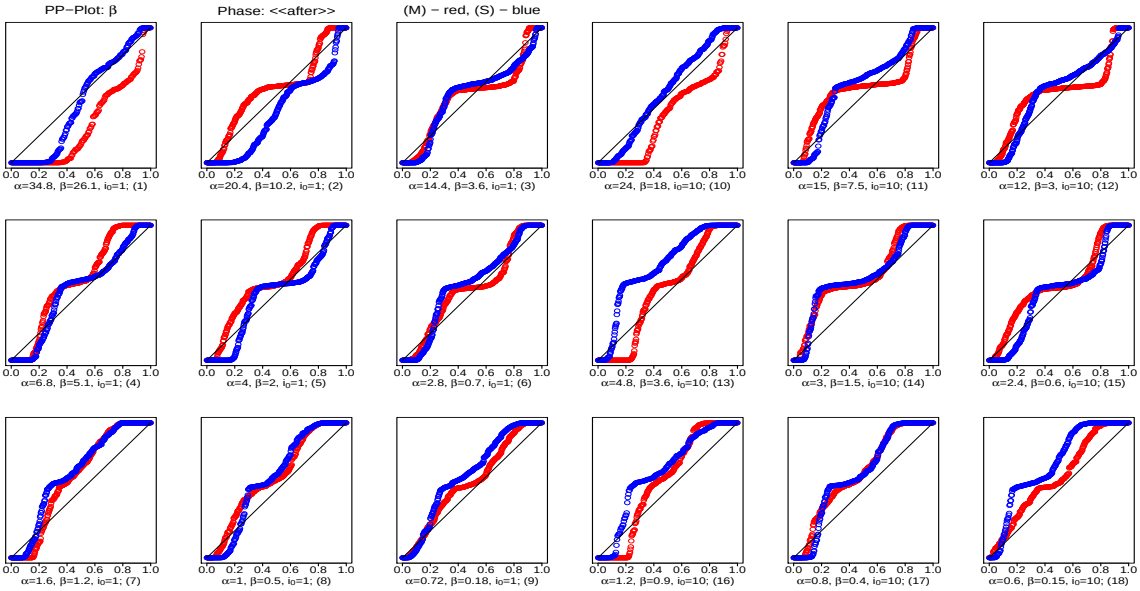

FIGURE 53: PP-plots de $\theta = \beta$ pour les observations de la phase stationnaire : « after »

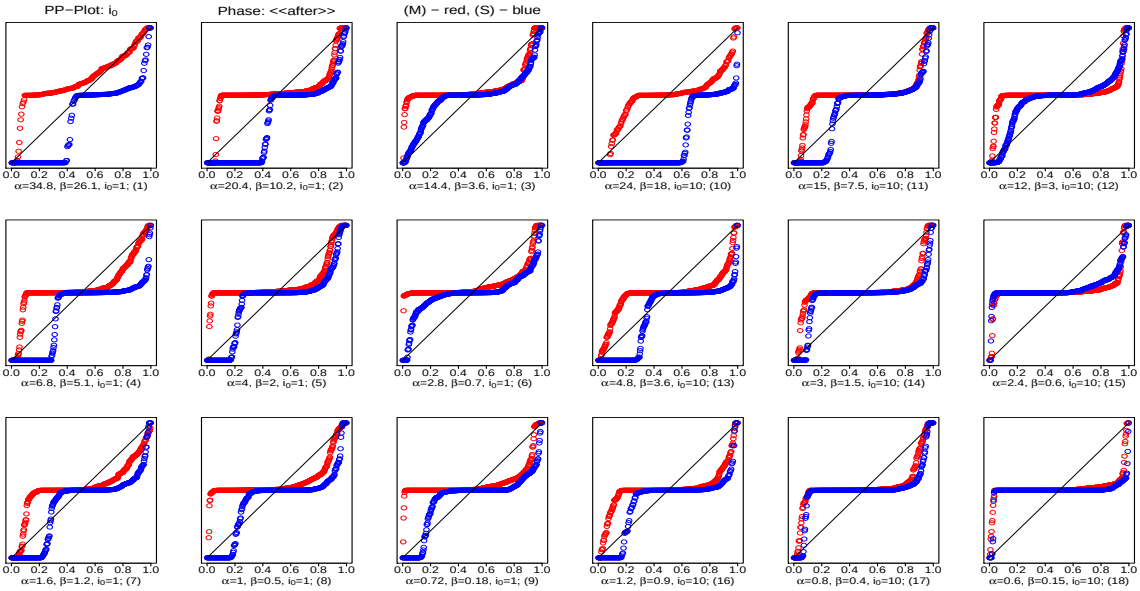

FIGURE 54: PP-plots de $\theta = i_0$ pour les observations de la phase stationnaire : « after »

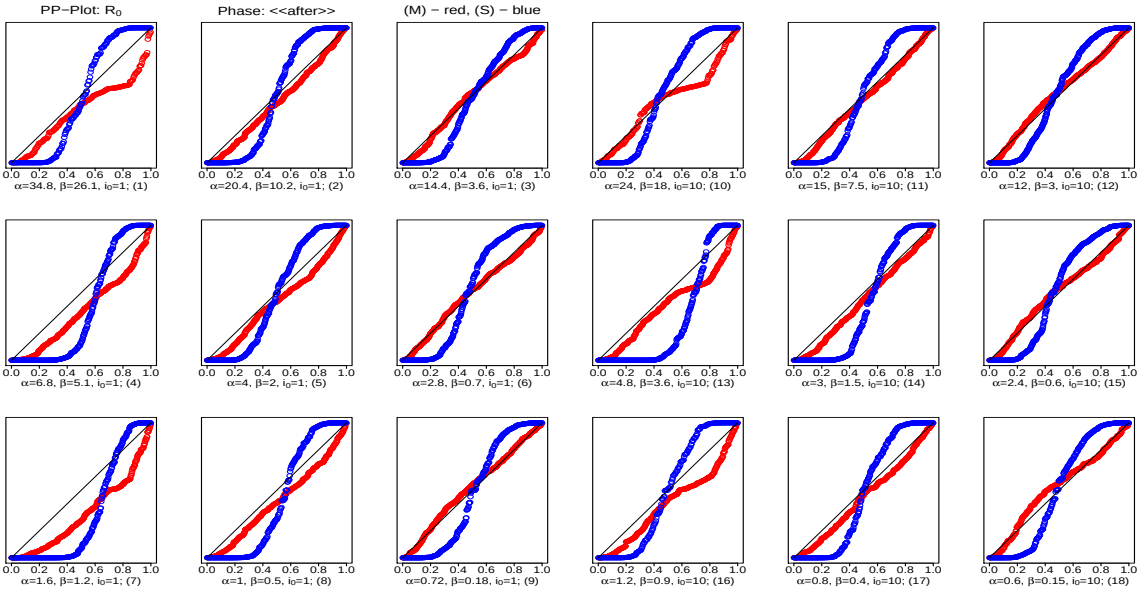

FIGURE 55: PP-plots de R_0 pour les observations de la phase stationnaire : « after »

TABLE 13: Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase initiale du processus « early »

	#1		#2		#3		#10		#11		#12	
α_0	34.8		20.4		14.4		24		15		12	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	617	389	471	774	260	724	664	340	569	812	501	891
mean	19.14	57.14	15.55	34.39	12.24	21.31	8.77	53.13	10.72	28.60	10.69	14.38
bias	-15.66	22.34	-4.85	13.99	-2.16	6.91	-15.23	29.13	-4.28	13.60	-1.31	2.38
rel. bias	-0.45	0.64	-0.24	0.69	-0.15	0.48	-0.63	1.21	-0.29	0.91	-0.11	0.20
|bias|	23.30	28.10	11.21	17.27	4.74	8.52	17.91	32.60	10.65	16.05	4.23	4.33
sd	20.39	25.28	12.99	19.37	5.24	11.65	11.38	26.20	13.04	18.47	4.91	5.38
MSE	660.56	1136.63	191.98	570.51	32.06	183.30	361.16	1533.41	188.01	525.68	25.78	34.63
RE1	1.00	0.58	1.00	0.34	1.00	0.17	1.00	0.24	1.00	0.36	1.00	0.74
Mean.as.sd	18.53	16.33	10.40	8.60	6.44	9.80	9.27	21.14	8.93	7.19	6.57	1.74

	#4		#5		#6		#13		#14		#15	
α_0	6.8		4		2.8		4.8		3		2.4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	682	708	534	770	248	695	705	633	590	798	542	858
mean	5.59	26.48	3.22	9.56	2.49	4.61	3.40	26.64	2.52	6.68	2.12	2.78
bias	-1.21	19.68	-0.78	5.56	-0.31	1.81	-1.40	21.84	-0.48	3.68	-0.28	0.38
rel. bias	-0.18	2.89	-0.19	1.39	-0.11	0.64	-0.29	4.55	-0.16	1.23	-0.12	0.16
|bias|	5.58	20.32	2.27	6.04	0.84	2.08	4.80	22.26	2.17	4.13	0.74	0.78
sd	8.33	22.84	3.30	11.24	0.98	4.83	8.86	26.47	3.06	7.16	0.86	0.99
MSE	70.69	908.28	11.50	157.03	1.04	26.52	80.29	1176.33	9.57	64.68	0.81	1.12
RE1	1.00	0.08	1.00	0.07	1.00	0.04	1.00	0.07	1.00	0.15	1.00	0.73
Mean.as.sd	4.86	10.72	2.07	2.51	1.22	5.94	3.33	26.11	2.05	2.19	1.19	0.35

	#7		#8		#9		#16		#17		#18	
α_0	1.6		1		0.72		1.2		0.8		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	699	728	523	756	268	729	717	630	658	801	549	916
mean	1.47	9.93	0.84	2.95	0.59	1.72	1.31	10.84	0.60	1.70	0.54	0.70
bias	-0.13	8.33	-0.16	1.95	-0.13	1.00	0.11	9.64	-0.20	0.90	-0.06	0.10
rel. bias	-0.08	5.21	-0.16	1.95	-0.19	1.38	0.10	8.03	-0.25	1.13	-0.10	0.17
|bias|	1.49	8.53	0.54	2.09	0.24	1.07	1.62	9.73	0.54	1.04	0.18	0.20
sd	3.13	14.20	0.72	5.93	0.27	5.55	3.61	13.08	0.69	2.38	0.21	0.24
MSE	9.79	270.66	0.54	38.96	0.09	31.75	13.04	263.56	0.52	6.47	0.05	0.07
RE1	1.00	0.04	1.00	0.01	1.00	0.00	1.00	0.05	1.00	0.08	1.00	0.70
Mean.as.sd	1.54	5.54	0.55	1.81	0.29	3.51	2.40	12.71	0.48	0.43	0.31	0.09

TABLE 14: Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase initiale du processus « early »

	#1		#2		#3		#10		#11		#12	
β_0	26.1		10.2		3.6		18		7.5		3	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	617	389	471	774	260	724	664	340	569	812	501	891
mean	14.70	42.89	8.00	20.39	3.22	8.87	6.29	40.47	5.11	16.72	2.48	4.07
bias	-11.40	16.79	-2.20	10.19	-0.38	5.27	-11.71	22.47	-2.39	9.22	-0.52	1.07
rel. bias	-0.44	0.64	-0.22	1.00	-0.11	1.46	-0.65	1.25	-0.32	1.23	-0.17	0.36
|bias|	18.70	22.25	7.28	12.71	2.36	6.23	14.02	25.71	6.42	10.93	1.83	2.02
sd	17.66	21.17	9.25	15.54	2.96	10.29	9.40	21.79	8.13	13.34	2.11	2.56
MSE	441.36	729.18	90.31	345.12	8.86	133.53	225.29	978.13	71.63	262.59	4.73	7.67
RE	1.00	0.61	1.00	0.26	1.00	0.07	1.00	0.23	1.00	0.27	1.00	0.62
Mean.as.sd	16.93	13.48	8.40	6.97	4.41	8.58	8.36	17.13	6.01	4.92	2.97	0.85

	#4		#5		#6		#13		#14		#15	
β_0	5.1		2		0.7		3.6		1.5		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	682	708	534	770	248	695	705	633	590	798	542	858
mean	4.34	21.65	1.72	6.66	0.71	2.15	2.50	21.31	1.22	4.02	0.49	0.77
bias	-0.76	16.55	-0.28	4.66	0.01	1.45	-1.10	17.71	-0.28	2.52	-0.11	0.17
rel.bias	-0.15	3.25	-0.14	2.33	0.02	2.07	-0.31	4.92	-0.18	1.68	-0.19	0.28
|bias|	4.56	17.16	1.50	5.04	0.44	1.63	3.76	18.11	1.32	2.84	0.32	0.37
sd	6.87	19.88	2.45	10.61	0.55	4.68	6.94	21.77	1.93	5.21	0.38	0.49
MSE	47.66	668.41	6.05	134.16	0.31	23.98	49.38	786.81	3.81	33.44	0.16	0.27
RE	1.00	0.07	1.00	0.05	1.00	0.01	1.00	0.06	1.00	0.11	1.00	0.58
Mean.as.sd	4.36	9.24	1.73	2.15	0.86	5.64	2.81	21.13	1.35	1.51	0.55	0.17

	#7		#8		#9		#16		#17		#18	
β_0	1.2		0.5		0.18		0.9		0.4		0.15	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	699	728	523	756	268	729	717	630	658	801	549	916
mean	1.15	8.54	0.45	2.21	0.15	1.08	0.99	9.03	0.28	1.04	0.13	0.19
bias	-0.05	7.34	-0.05	1.71	-0.03	0.90	0.09	8.13	-0.12	0.64	-0.02	0.04
rel. bias	-0.04	6.12	-0.11	3.43	-0.17	5.00	0.10	9.03	-0.30	1.60	-0.15	0.30
|bias|	1.23	7.53	0.36	1.82	0.11	0.95	1.27	8.22	0.33	0.73	0.08	0.09
sd	2.62	13.14	0.51	5.68	0.18	5.43	2.84	11.66	0.43	1.89	0.09	0.12
MSE	6.86	226.41	0.26	35.14	0.03	30.21	8.09	201.79	0.20	3.97	0.01	0.02
RE	1.00	0.03	1.00	0.01	1.00	0.00	1.00	0.04	1.00	0.05	1.00	0.57
Mean.as.sd	1.36	4.99	0.44	1.69	0.21	3.39	1.89	10.77	0.32	0.30	0.14	0.05

TABLE 15: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase initiale du processus « early »

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	617	389	471	774	260	724	664	340	569	812	501	891
mean	5.46	3.20	2.40	1.62	1.67	1.29	13.08	9.27	12.63	8.10	11.24	9.32
bias	4.46	2.20	1.40	0.62	0.67	0.29	3.08	-0.73	2.63	-1.90	1.24	-0.68
rel. bias	4.46	2.20	1.40	0.62	0.67	0.29	0.31	-0.07	0.26	-0.19	0.12	-0.07
|bias|	4.67	2.42	1.84	1.07	1.26	0.81	5.89	6.96	4.85	4.21	3.34	3.16
sd	5.01	4.18	2.48	1.72	1.79	1.18	8.63	10.15	5.84	4.72	4.33	3.80
MSE	44.97	22.29	8.10	3.36	3.64	1.47	83.92	103.24	41.02	25.89	20.24	14.91
RE	1.00	2.02	1.00	2.41	1.00	2.47	1.00	0.81	1.00	1.58	1.00	1.36
Mean.as.sd	3.67	1.47	1.65	0.68	1.21	0.58	7.22	5.15	4.35	1.80	4.75	1.22

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	682	708	534	770	248	695	705	633	590	798	542	858
mean	4.14	2.07	2.18	1.58	1.42	1.22	11.41	8.16	11.42	8.16	10.84	9.61
bias	3.14	1.07	1.18	0.58	0.42	0.22	1.41	-1.84	1.42	-1.84	0.84	-0.39
rel.bias	3.14	1.07	1.18	0.58	0.42	0.22	0.14	-0.18	0.14	-0.18	0.08	-0.04
|bias|	3.45	1.66	1.60	1.01	0.90	0.69	4.69	8.09	4.05	3.89	2.94	2.77
sd	4.59	4.76	2.13	1.44	1.28	1.02	6.36	13.11	5.11	4.41	3.70	3.39
MSE	30.91	23.73	5.89	2.41	1.80	1.09	42.43	174.89	28.09	22.81	14.34	11.63
RE	1.00	1.30	1.00	2.44	1.00	1.66	1.00	0.24	1.00	1.23	1.00	1.23
Mean.as.sd	3.60	10.48	1.34	0.78	1.03	0.70	7.41	28.35	3.99	1.70	3.73	1.25

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	699	728	523	756	268	729	717	630	658	801	549	916
mean	4.08	2.91	2.18	1.73	1.59	1.31	11.66	7.89	11.50	8.10	10.62	9.37
bias	3.08	1.91	1.18	0.73	0.59	0.31	1.66	-2.11	1.50	-1.90	0.62	-0.63
rel. bias	3.08	1.91	1.18	0.73	0.59	0.31	0.17	-0.21	0.15	-0.19	0.06	-0.06
|bias|	3.37	2.58	1.66	1.18	1.03	0.77	5.23	8.28	4.18	4.07	2.96	2.81
sd	3.60	8.44	2.41	3.11	1.40	1.60	7.17	11.40	5.32	4.54	3.76	3.42
MSE	22.41	74.86	7.18	10.19	2.31	2.67	54.15	134.28	30.47	24.15	14.49	12.07
RE	1.00	0.30	1.00	0.70	1.00	0.87	1.00	0.40	1.00	1.26	1.00	1.20
Mean.as.sd	4.64	28.55	1.40	1.62	1.11	1.26	19.95	112.62	3.93	1.71	3.99	1.27

TABLE 16: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early »

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	617	389	471	774	260	724	664	340	569	812	501	891
mean	5.25	10.28	15.53	15.27	29.17	50.21	4.99	15.35	11.32	18.89	53.57	6.58
bias	3.91	8.95	13.53	13.27	25.17	46.21	3.66	14.01	9.32	16.89	49.57	2.58
rel. bias	2.93	6.71	6.76	6.63	6.29	11.55	2.75	10.51	4.66	8.45	12.39	0.65
|bias|	4.03	9.02	13.77	13.72	26.01	48.15	3.77	14.11	9.47	17.30	50.06	3.51
sd	43.76	161.76	111.24	149.35	165.63	458.95	28.76	153.07	50.84	203.92	319.48	43.01
MSE$\times 10^2$	19.27	261.80	125.30	224.52	279.61	2124.75	8.39	235.58	26.67	418.17	1043.21	18.54
RE	1.00	0.07	1.00	0.56	1.00	0.13	1.00	0.04	1.00	0.06	1.00	56.26
Mean.as.sd $\times 10^2$	17.39	2230.38	74.07	932.12	238.52	2859.90	5.50	873.34	9.93	870.65	353.52	1.71

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	682	708	534	770	248	695	705	633	590	798	542	858
mean	3.38	2.59	7.80	4.65	9.44	9.59	3.60	4.17	7.94	5.36	18.20	5.50
bias	2.04	1.25	5.80	2.65	5.44	5.59	2.27	2.83	5.94	3.36	14.20	1.50
rel. bias	1.53	0.94	2.90	1.33	1.36	1.40	1.70	2.13	2.97	1.68	3.55	0.37
|bias|	2.17	1.41	6.10	3.24	6.59	7.61	2.37	2.96	6.09	3.77	14.62	2.34
sd	7.51	14.71	36.13	31.89	23.60	42.04	11.64	39.19	22.58	44.73	60.49	9.90
MSE	60.51	217.67	1336.81	1022.67	584.23	1796.32	140.44	1541.42	544.24	2009.72	3853.62	100.06
RE	1.00	0.28	1.00	1.31	1.00	0.33	1.00	0.09	1.00	0.27	1.00	38.51
Mean.as.sd	43.11	369.50	921.94	1448.69	320.79	7607.04	66.48	2044.66	227.09	1665.12	1104.76	11.90

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	699	728	523	756	268	729	717	630	658	801	549	916
MC mean	3.71	1.98	5.69	2.80	9.06	8.55	3.90	1.69	8.19	3.15	7.81	6.11
MC bias	2.38	0.65	3.69	0.80	5.06	4.55	2.57	0.36	6.19	1.15	3.81	2.11
MC rel. bias	1.79	0.49	1.85	0.40	1.26	1.14	1.93	0.27	3.09	0.58	0.95	0.53
MC |bias|	2.49	0.83	4.00	1.37	5.78	6.51	2.67	0.51	6.32	1.52	4.26	2.96
MC sd	9.43	9.07	16.08	9.07	16.37	35.40	9.65	4.31	18.29	9.36	10.94	14.94
MC MSE	94.51	82.63	271.68	82.86	292.59	1272.25	99.62	18.71	372.49	88.88	133.93	227.37
MC RE1	1.00	1.14	1.00	3.28	1.00	0.23	1.00	5.32	1.00	4.19	1.00	0.59
Mean.as.sd	71.55	995.77	264.38	150.68	211.17	2395.12	69.54	84.63	424.34	134.32	47.09	32.24

TABLE 17: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early » (95% des estimations)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	585.00	369.00	447.00	734.00	246.00	686.00	630.00	322.00	539.00	770.00	475.00	845.00
mean [1]	2.19	1.42	4.88	2.10	9.63	4.79	2.41	1.40	5.80	2.04	12.05	4.39
bias	0.86	0.09	2.88	0.10	5.63	0.79	1.08	0.06	3.80	0.04	8.05	0.39
rel. bias	0.65	0.07	1.44	0.05	1.41	0.20	0.81	0.05	1.90	0.02	2.01	0.10
|bias|	0.97	0.15	3.10	0.52	6.39	2.69	1.17	0.15	3.92	0.43	8.49	1.27
sd	2.11	0.35	9.35	1.13	15.29	7.21	2.57	0.32	8.05	0.85	29.72	2.07
MSE	5.21	0.13	95.59	1.27	264.49	52.55	7.77	0.11	79.14	0.73	946.34	4.44
RE	1.00	40.15	1.00	74.98	1.00	5.03	1.00	73.74	1.00	108.23	1.00	212.96
Mean.as.sd	5.58	0.60	80.86	1.62	152.48	463.48	6.31	0.20	37.20	0.68	282.24	1.10

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	646.00	672.00	506.00	730.00	234.00	659.00	669.00	601.00	560.00	758.00	514.00	814.00
mean	2.37	1.31	4.14	1.91	6.45	4.29	2.33	1.32	4.95	1.99	9.51	4.54
bias	1.04	-0.03	2.14	-0.09	2.45	0.29	0.99	-0.01	2.95	-0.01	5.51	0.54
relative bias	0.78	-0.02	1.07	-0.04	0.61	0.07	0.75	-0.01	1.48	-0.01	1.38	0.13
|bias|	1.16	0.12	2.40	0.49	3.52	2.26	1.08	0.11	3.08	0.38	5.88	1.33
sd	2.79	0.16	6.14	0.76	6.77	4.63	2.08	0.14	6.76	0.60	16.27	2.21
MSE	8.87	0.03	42.14	0.58	51.69	21.50	5.30	0.02	54.41	0.36	294.59	5.17
RE	1.00	338.21	1.00	72.05	1.00	2.40	1.00	253.74	1.00	149.62	1.00	57.01
Mean of as.sd	9.92	0.05	34.20	0.58	33.67	6.95	5.11	0.04	25.85	0.35	94.67	1.38

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
true value	1.33	1.33	2.00	2.00	4.00	4.00	1.33	1.33	2.00	2.00	4.00	4.00
# sims	663.00	690.00	495.00	718.00	254.00	691.00	681.00	598.00	624.00	759.00	521.00	870.00
mean	2.36	1.32	3.54	1.93	6.94	4.19	2.62	1.30	5.88	2.03	6.58	4.53
bias	1.03	-0.02	1.54	-0.07	2.94	0.19	1.28	-0.04	3.88	0.03	2.58	0.53
relative bias	0.77	-0.01	0.77	-0.03	0.74	0.05	0.96	-0.03	1.94	0.01	0.64	0.13
|bias|	1.13	0.16	1.81	0.48	3.57	2.10	1.38	0.11	3.98	0.37	2.98	1.33
sd	2.82	0.25	4.32	0.73	6.03	4.12	3.18	0.14	9.33	0.66	5.14	2.36
MSE	9.00	0.06	20.97	0.54	44.84	16.98	11.75	0.02	102.01	0.43	33.03	5.83
RE	1.00	148.64	1.00	38.94	1.00	2.64	1.00	578.43	1.00	237.02	1.00	5.67
Mean of as.sd	8.34	0.34	18.30	0.53	32.63	62.73	8.99	0.04	43.39	0.31	17.30	1.41

[1] Les caractéristiques présentées dans le tableau sont évaluées pour 95% des estimations.
5% des estimations extrêmes ont été écartées

TABLE 18: Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase évolutive du processus « during »

	#1		#2		#3		#10		#11		#12	
α_0	34.8		20.4		14.4		24		15		12	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	621	478	767	897	850	961	577	405	819	802	892	912
mean	15.20	46.23	16.70	25.88	12.81	15.19	12.06	38.47	16.19	22.43	12.16	14.07
bias	-19.60	11.43	-3.70	5.48	-1.59	0.79	-11.94	14.47	1.19	7.43	0.16	2.07
rel. bias	-0.56	0.33	-0.18	0.27	-0.11	0.05	-0.50	0.60	0.08	0.50	0.01	0.17
|bias|	24.48	21.30	10.52	9.99	4.79	3.26	19.93	20.57	11.26	11.79	3.99	4.56
sd	18.21	24.54	11.90	14.16	5.50	4.43	19.53	22.74	15.16	18.28	5.16	7.01
MSE	715.26	731.57	155.17	230.44	32.74	20.21	523.30	725.14	231.04	388.79	26.66	53.35
RE	1.00	0.98	1.00	0.67	1.00	1.62	1.00	0.72	1.00	0.59	1.00	0.50
Mean.as.sd	15.12	15.69	9.82	5.01	5.41	0.98	15.49	17.22	15.51	8.04	6.16	2.58

	#4		#5		#6		#13		#14		#15	
α_0	6.8		4		2.8		4.8		3		2.4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	675	731	813	880	912	974	686	658	842	779	948	941
mean	7.56	21.81	3.51	5.10	2.51	2.94	6.64	21.51	3.00	4.77	2.41	2.71
bias	0.76	15.01	-0.49	1.10	-0.29	0.14	1.84	16.71	0.00	1.77	0.01	0.31
rel. bias	0.11	2.21	-0.12	0.27	-0.10	0.05	0.38	3.48	0.00	0.59	0.00	0.13
|bias|	7.70	16.37	2.03	1.96	0.97	0.61	7.15	17.33	1.96	2.54	0.73	0.80
sd	12.09	20.38	2.48	3.25	1.31	0.91	11.49	19.72	2.64	4.60	0.93	1.26
MSE	146.64	640.03	6.37	11.74	1.79	0.85	135.30	667.30	6.98	24.24	0.87	1.67
RE	1.00	0.23	1.00	0.54	1.00	2.09	1.00	0.20	1.00	0.29	1.00	0.52
Mean.as.sd	11.91	19.44	1.93	0.91	0.95	0.20	10.37	20.90	2.41	1.42	1.12	0.33

	#7		#8		#9		#16		#17		#18	
α_0	1.6		1		0.72		1.2		0.8		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	691	675	888	887	956	932	640	358	874	744	974	963
mean	1.99	6.20	1.03	1.23	0.71	0.74	2.05	5.81	0.80	1.25	0.59	0.68
bias	0.39	4.60	0.03	0.23	-0.01	0.02	0.85	4.61	0.00	0.45	-0.01	0.08
relative bias	0.24	2.87	0.03	0.23	-0.01	0.03	0.71	3.85	-0.01	0.56	-0.01	0.14
|bias|	2.08	4.86	0.63	0.46	0.30	0.16	2.07	4.77	0.55	0.70	0.17	0.20
sd	3.78	6.35	0.81	0.79	0.49	0.23	3.43	6.20	0.77	1.20	0.22	0.38
MSE	14.40	61.45	0.65	0.67	0.24	0.06	12.49	59.60	0.58	1.64	0.05	0.15
RE	1.00	0.23	1.00	0.97	1.00	4.28	1.00	0.21	1.00	0.36	1.00	0.33
Mean.as.sd	1.77	6.46	0.59	0.17	0.25	0.06	3.72	12.87	0.68	0.34	0.29	0.08

TABLE 19: Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase évolutive du processus « during »

	#1		#2		#3		#10		#11		#12	
β_0	26.1		10.2		3.6		18		7.5		3	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	621	478	767	897	850	961	577	405	819	802	892	912
mean	11.63	35.34	8.53	13.49	3.18	3.87	9.06	29.42	8.35	11.65	3.09	3.65
bias	-14.47	9.24	-1.67	3.29	-0.42	0.27	-8.94	11.42	0.85	4.15	0.09	0.65
rel. bias	-0.55	0.35	-0.16	0.32	-0.12	0.07	-0.50	0.63	0.11	0.55	0.03	0.22
|bias|	18.56	17.20	5.64	5.79	1.48	1.20	15.04	16.23	6.29	6.58	1.27	1.44
sd	14.33	20.09	6.70	8.02	1.77	1.63	14.76	18.23	8.80	10.17	1.65	2.18
MSE	414.41	488.06	47.57	75.13	3.29	2.74	297.45	461.73	78.08	120.55	2.73	5.14
RE	1.00	0.85	1.00	0.63	1.00	1.20	1.00	0.64	1.00	0.65	1.00	0.53
Mean.as.sd	12.61	12.49	5.71	3.25	1.74	0.42	12.39	13.49	8.67	4.45	1.85	0.81

	#4		#5		#6		#13		#14		#15	
β_0	5.1		2		0.7		3.6		1.5		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	675	731	813	880	912	974	686	658	842	779	948	941
mean	5.91	17.29	1.79	2.68	0.63	0.77	5.10	17.02	1.53	2.53	0.61	0.70
bias	0.81	12.19	-0.21	0.68	-0.07	0.07	1.50	13.42	0.03	1.03	0.01	0.10
rel. bias	0.16	2.39	-0.11	0.34	-0.09	0.10	0.42	3.73	0.02	0.69	0.01	0.17
|bias|	6.01	13.28	1.10	1.18	0.30	0.24	5.50	13.92	1.09	1.47	0.23	0.26
sd	9.81	16.95	1.40	2.12	0.43	0.59	9.09	16.31	1.55	2.76	0.30	0.40
MSE	96.85	435.62	2.01	4.97	0.19	0.36	84.77	445.58	2.41	8.68	0.09	0.17
RE	1.00	0.22	1.00	0.40	1.00	0.53	1.00	0.19	1.00	0.28	1.00	0.53
Mean.as.sd	9.40	15.72	1.12	0.54	0.33	0.10	8.12	16.74	1.35	0.80	0.34	0.11

	#7		#8		#9		#16		#17		#18	
β_0	1.2		0.5		0.18		0.9		0.4		0.15	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	691	675	888	887	956	932	640	358	874	744	974	963
mean	1.57	5.00	0.53	0.65	0.19	0.19	1.59	4.51	0.41	0.65	0.15	0.18
bias	0.37	3.80	0.03	0.15	0.01	0.01	0.69	3.61	0.01	0.25	0.00	0.03
relative bias	0.31	3.17	0.06	0.29	0.05	0.06	0.77	4.01	0.02	0.63	0.00	0.18
|bias|	1.64	4.02	0.34	0.28	0.10	0.06	1.61	3.74	0.30	0.39	0.06	0.07
sd	3.08	5.52	0.46	0.50	0.19	0.10	2.71	5.16	0.43	0.68	0.07	0.13
MSE	9.63	44.87	0.21	0.27	0.04	0.01	7.83	39.55	0.19	0.52	0.01	0.02
RE	1.00	0.21	1.00	0.78	1.00	3.51	1.00	0.20	1.00	0.36	1.00	0.29
Mean.as.sd	1.45	5.41	0.34	0.11	0.09	0.03	2.80	10.37	0.37	0.19	0.09	0.03

TABLE 20: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase évolutive du processus « during »

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	621	478	767	897	850	961	577	405	819	802	892	912
mean	9.56	4.49	3.44	1.88	1.64	1.45	14.82	10.50	12.77	8.70	11.76	9.74
bias	8.56	3.49	2.44	0.88	0.64	0.45	4.82	0.50	2.77	-1.30	1.76	-0.26
rel. bias	8.56	3.49	2.44	0.88	0.64	0.45	0.48	0.05	0.28	-0.13	0.18	-0.03
|bias|	8.76	3.95	2.98	1.49	1.47	1.06	8.53	9.67	6.95	6.49	5.52	5.56
sd	11.02	8.47	4.22	2.65	2.17	1.62	13.09	14.53	9.92	8.26	7.79	7.25
MSE	194.55	83.85	23.74	7.81	5.10	2.81	194.16	210.92	106.00	69.77	63.72	52.55
RE	1.00	2.32	1.00	3.04	1.00	1.81	1.00	0.92	1.00	1.52	1.00	1.21
Mean.as.sd	12.82	7.48	3.14	0.68	1.63	0.37	21.67	10.09	12.74	2.86	7.80	1.97

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims 675	731	813	880	912	974	686	658	842	779	948	941	
mean	9.03	4.98	2.58	2.08	1.37	1.41	13.91	8.54	11.71	8.10	10.78	9.75
bias	8.03	3.98	1.58	1.08	0.37	0.41	3.91	-1.46	1.71	-1.90	0.78	-0.25
rel. bias	8.03	3.98	1.58	1.08	0.37	0.41	0.39	-0.15	0.17	-0.19	0.08	-0.03
|bias|	8.20	4.77	2.25	1.58	1.16	0.97	8.13	10.45	5.79	5.35	4.66	4.50
sd	12.44	12.73	3.59	2.85	1.82	1.60	12.60	16.07	7.71	6.18	6.20	5.76
MSE	219.08	177.65	15.33	9.29	3.43	2.72	173.90	260.05	62.27	41.80	38.97	33.23
RE	1.00	1.23	1.00	1.65	1.00	1.26	1.00	0.67	1.00	1.49	1.00	1.17
Mean.as.sd	40.45	30.14	2.21	1.32	1.27	0.47	51.92	69.87	7.91	2.02	5.98	1.43

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	691	675	888	887	956	932	640	358	874	744	974	963
mean	6.52	3.26	2.82	1.93	1.36	1.41	13.93	8.07	12.33	8.87	10.94	9.85
bias	5.52	2.26	1.82	0.93	0.36	0.41	3.93	-1.93	2.33	-1.13	0.94	-0.15
rel. bias	5.52	2.26	1.82	0.93	0.36	0.41	0.39	-0.19	0.23	-0.11	0.09	-0.02
|bias|	5.71	3.10	2.41	1.41	1.10	0.93	8.58	8.95	6.55	6.30	4.72	4.59
sd	7.10	7.99	3.33	2.12	1.70	1.44	13.48	12.44	8.47	7.60	6.44	5.99
MSE	80.92	68.79	14.39	5.37	3.01	2.23	196.92	158.03	77.12	58.97	42.26	35.83
RE	1.00	1.18	1.00	2.68	1.00	1.35	1.00	1.25	1.00	1.31	1.00	1.18
Mean.as.sd	28.60	39.86	2.62	0.70	1.33	0.45	81.67	84.18	9.63	2.09	6.28	1.52

TABLE 21: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase évolutive du processus « during »

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	621	478	767	897	850	961	577	405	819	802	892	912
mean	3.15	6.78	2.87	3.22	5.57	6.48	2.54	2.33	4.65	2.12	4.85	4.12
bias	1.82	5.44	0.87	1.22	1.57	2.48	1.21	0.99	2.65	0.12	0.85	0.12
rel. bias	1.36	4.08	0.44	0.61	0.39	0.62	0.91	0.75	1.33	0.06	0.21	0.03
|bias|	1.94	5.52	1.03	1.41	1.92	2.91	1.33	1.09	2.79	0.29	1.23	0.53
var	118.51	7850.17	29.53	400.52	228.70	4814.99	64.45	251.90	249.55	0.94	36.01	0.73
sd	10.89	88.60	5.43	20.01	15.12	69.39	8.03	15.87	15.80	0.97	6.00	0.85
MSE	121.62	7863.37	30.25	401.56	230.90	4816.15	65.81	252.27	256.28	0.96	36.70	0.74
RE	1.00	0.02	1.00	0.08	1.00	0.05	1.00	0.26	1.00	268.14	1.00	49.50
Mean.as.sd	65.14	30829.73	13.44	273.82	78.09	205.68	34.78	12925.87	85.07	0.72	8.90	0.28

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	675	731	813	880	912	974	686	658	842	779	948	941
mean	2.55	1.69	3.18	2.40	5.03	4.25	2.52	2.19	3.07	2.09	4.54	4.12
bias	1.22	0.36	1.18	0.40	1.03	0.25	1.19	0.86	1.07	0.09	0.54	0.12
rel. bias	0.91	0.27	0.59	0.20	0.26	0.06	0.89	0.64	0.53	0.04	0.13	0.03
|bias|	1.33	0.47	1.33	0.60	1.44	0.74	1.29	0.97	1.20	0.28	0.88	0.54
sd	8.81	7.45	15.67	6.72	6.48	2.49	9.27	15.46	4.45	0.85	3.19	0.80
MSE	79.04	55.61	246.60	45.34	43.02	6.26	87.22	239.50	20.89	0.73	10.42	0.66
RE	1.00	1.42	1.00	5.44	1.00	6.87	1.00	0.36	1.00	28.44	1.00	15.86
Mean.as.sd	48.97	385.52	123.28	8.25	16.59	1.98	60.37	1707.24	8.51	0.53	3.08	0.27

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	691	675	888	887	956	932	640	358	874	744	974	963
mean	4.29	1.78	3.96	2.04	4.76	4.30	3.16	1.37	5.57	2.23	4.44	4.12
bias	2.96	0.44	1.96	0.04	0.76	0.30	1.83	0.04	3.57	0.23	0.44	0.12
rel. bias	2.22	0.33	0.98	0.02	0.19	0.08	1.37	0.03	1.78	0.11	0.11	0.03
|bias|	3.06	0.57	2.12	0.23	1.18	0.73	1.95	0.11	3.70	0.39	0.80	0.54
sd	14.22	7.55	14.29	0.45	7.70	1.79	9.19	0.23	13.64	2.49	2.45	0.78
MSE	210.65	57.13	207.83	0.21	59.85	3.29	87.75	0.05	198.54	6.22	6.18	0.63
RE	1.00	3.69	1.00	1009.91	1.00	18.22	1.00	1604.71	1.00	31.90	1.00	9.82
Mean.as.sd	180.37	49.93	230.19	0.22	28.37	0.54	63.06	0.10	105.85	3.54	2.26	0.27

TABLE 22: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « during » (95% des estimations)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	589.00	454.00	727.00	851.00	806.00	911.00	547.00	383.00	777.00	760.00	846.00	866.00
mean [1]	1.87	1.35	2.25	1.99	4.38	4.13	1.75	1.34	2.99	2.03	4.26	4.07
bias	0.54	0.02	0.25	-0.01	0.38	0.13	0.42	0.01	0.99	0.03	0.26	0.07
relative bias	0.40	0.01	0.12	-0.01	0.09	0.03	0.32	0.00	0.50	0.02	0.06	0.02
|bias|	0.65	0.09	0.40	0.16	0.69	0.51	0.53	0.09	1.11	0.19	0.61	0.45
sd	1.56	0.12	0.93	0.21	1.00	0.66	1.12	0.12	2.63	0.25	0.86	0.57
MSE	2.70	0.01	0.93	0.04	1.15	0.46	1.42	0.02	7.91	0.06	0.81	0.33
RE	1.00	200.60	1.00	21.58	1.00	2.52	1.00	92.95	1.00	128.90	1.00	2.43
Mean.as.sd	3.05	0.04	1.26	0.07	1.34	0.26	1.79	0.04	4.51	0.09	1.00	0.23

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	641.00	693.00	771.00	836.00	866.00	924.00	650.00	624.00	798.00	739.00	900.00	893.00
mean	1.66	1.31	2.12	2.00	4.36	4.08	1.69	1.30	2.51	2.01	4.22	4.08
bias	0.33	-0.02	0.12	0.00	0.36	0.08	0.35	-0.03	0.51	0.01	0.22	0.08
relative bias	0.25	-0.02	0.06	0.00	0.09	0.02	0.27	-0.02	0.25	0.01	0.05	0.02
|bias|	0.43	0.08	0.25	0.17	0.73	0.52	0.45	0.08	0.62	0.18	0.52	0.45
sd	1.08	0.10	0.43	0.21	1.15	0.67	0.92	0.10	1.38	0.24	0.70	0.57
MSE	1.27	0.01	0.20	0.04	1.45	0.45	0.98	0.01	2.16	0.06	0.54	0.33
RE	1.00	123.22	1.00	4.38	1.00	3.20	1.00	93.77	1.00	38.00	1.00	1.61
Mean.as.sd	1.69	0.03	0.59	0.07	1.55	0.30	1.50	0.03	1.84	0.09	0.86	0.24

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
true value	1.33	1.33	2.00	2.00	4.00	4.00	1.33	1.33	2.00	2.00	4.00	4.00
# sims	655.00	641.00	842.00	841.00	908.00	884.00	608.00	340.00	830.00	706.00	924.00	913.00
mean	2.22	1.31	2.19	2.01	4.29	4.16	1.84	1.35	3.74	2.02	4.23	4.08
bias	0.88	-0.03	0.19	0.01	0.29	0.16	0.50	0.02	1.74	0.02	0.23	0.08
relative bias	0.66	-0.02	0.10	0.00	0.07	0.04	0.38	0.01	0.87	0.01	0.06	0.02
|bias|	0.98	0.09	0.34	0.18	0.65	0.54	0.62	0.08	1.85	0.17	0.55	0.45
sd	3.54	0.11	0.74	0.23	0.84	0.68	1.70	0.10	6.57	0.23	0.73	0.57
MSE	13.31	0.01	0.59	0.05	0.79	0.49	3.14	0.01	46.13	0.05	0.58	0.33
RE	1.00	1001.88	1.00	11.30	1.00	1.61	1.00	304.76	1.00	878.10	1.00	1.77
Mean.as.sd	11.23	0.04	0.92	0.08	1.08	0.29	2.68	0.03	17.30	0.10	0.86	0.24

[1] Les caractéristiques présentées dans le tableau sont évaluées pour 95% des estimations.
5% des estimations extrêmes ont été écartées

TABLE 23: Résultats d'estimation par le maximum de vraisemblance du paramètre α avec les données provenant de la phase stationnaire du processus « after »

	#1		#2		#3		#10		#11		#12	
α_0	34.8		20.4		14.4		24		15		12	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	205	223	408	359	410	420	243	331	415	320	448	496
mean	27.30	38.43	48.33	31.16	34.51	28.60	26.72	30.21	44.75	31.85	33.48	28.06
bias	-7.50	3.63	27.93	10.76	20.11	14.20	2.72	6.21	29.75	16.85	21.48	16.06
rel. bias	-0.22	0.10	1.37	0.53	1.40	0.99	0.11	0.26	1.98	1.12	1.79	1.34
|bias|	27.68	16.35	33.17	17.31	24.61	18.54	22.64	12.37	33.21	20.43	24.59	19.67
sd	30.46	21.29	32.59	27.51	24.54	24.50	28.48	17.33	31.33	27.44	23.83	23.08
MSE	979.61	464.20	1839.62	870.16	1005.21	800.37	815.13	337.94	1864.00	1034.55	1027.90	789.36
RE1	1.00	2.11	1.00	2.11	1.00	1.26	1.00	2.41	1.00	1.80	1.00	1.30
Mean.as.sd	48.62	17.95	114.96	55.56	122.49	70.23	46.05	14.76	151.31	56.52	126.19	57.93

	#4		#5		#6		#13		#14		#15	
α_0	6.8		4		2.8		4.8		3		2.4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	374	363	397	384	412	422	406	405	361	398	423	436
mean	21.06	20.39	11.94	12.15	7.71	7.40	16.81	20.95	11.67	13.69	7.45	8.45
bias	14.26	13.59	7.94	8.15	4.91	4.60	12.01	16.15	8.67	10.69	5.05	6.05
rel. bias	2.10	2.00	1.99	2.04	1.75	1.64	2.50	3.36	2.89	3.56	2.10	2.52
|bias|	17.04	15.01	9.04	9.20	5.82	5.28	14.74	16.73	9.24	11.11	5.69	6.75
sd	17.18	19.77	8.38	10.23	5.90	5.72	16.60	19.45	8.17	9.88	5.75	6.65
MSE	497.61	574.49	133.06	170.69	58.84	53.86	418.96	638.02	141.71	211.65	58.47	80.70
RE	1.00	0.87	1.00	0.78	1.00	1.09	1.00	0.66	1.00	0.67	1.00	0.72
Mean.as.sd	40.43	47.72	40.73	47.96	34.21	22.39	32.89	43.94	43.18	51.21	31.03	32.38

	#7		#8		#9		#16		#17		#18	
α_0	1.6		1		0.72		1.2		0.8		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	387	355	381	318	388	367	334	246	346	357	365	319
MC mean	5.61	5.84	3.04	3.51	2.21	2.60	5.32	6.06	2.97	3.63	2.34	3.39
MC bias	4.01	4.24	2.04	2.51	1.49	1.88	4.12	4.86	2.17	2.83	1.74	2.79
rel. bias	2.51	2.65	2.04	2.51	2.07	2.61	3.43	4.05	2.71	3.54	2.90	4.66
MC |bias|	4.74	4.50	2.35	2.73	1.71	2.00	4.77	5.10	2.41	3.02	1.92	2.88
MC sd	5.28	5.68	2.43	2.96	1.67	3.41	4.92	6.40	2.50	2.91	1.75	1.72
MC MSE	43.89	50.14	10.06	15.03	5.00	15.13	41.16	64.45	10.93	16.42	6.08	10.75
MC RE	1.00	0.88	1.00	0.67	1.00	0.33	1.00	0.64	1.00	0.67	1.00	0.57
Mean.as.sd	8.62	17.35	7.70	13.60	8.36	8.62	6.98	17.83	4.21	18.77	8.91	12.75

TABLE 24: Résultats d'estimation par le maximum de vraisemblance du paramètre β avec les données provenant de la phase stationnaire du processus « after »

	#1		#2		#3		#10		#11		#12	
β_0	26.1		10.2		3.6		18		7.5		3	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	205	223	408	359	410	420	243	331	415	320	448	496
mean	21.21	29.21	24.46	15.65	8.76	7.30	20.39	22.94	22.65	15.94	8.32	7.09
bias	-4.89	3.11	14.26	5.45	5.16	3.70	2.39	4.94	15.15	8.44	5.32	4.09
rel. bias	-0.19	0.12	1.40	0.53	1.43	1.03	0.13	0.27	2.02	1.13	1.77	1.36
|bias|	20.90	13.15	16.83	8.87	6.33	4.82	17.23	10.10	16.91	10.33	6.12	5.01
sd	23.33	16.78	16.68	13.89	6.42	6.45	21.99	14.25	16.15	13.74	6.02	6.05
MSE	565.54	290.14	480.78	222.04	67.68	55.13	487.30	226.80	489.62	259.48	64.49	53.22
RE	1.00	1.95	1.00	2.17	1.00	1.23	1.00	2.15	1.00	1.89	1.00	1.21
Mean.as.sd	37.64	13.91	57.88	28.02	30.98	17.97	35.20	11.46	76.51	28.18	31.29	14.74

	#4		#5		#6		#13		#14		#15	
β_0	5.1		2		0.7		3.6		1.5		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	374	363	397	384	412	422	406	405	361	398	423	436
mean	16.40	15.85	6.09	6.15	1.94	1.89	12.99	16.40	5.94	7.01	1.88	2.14
bias	11.30	10.75	4.09	4.15	1.24	1.19	9.39	12.80	4.44	5.51	1.28	1.54
rel. bias	2.22	2.11	2.05	2.08	1.77	1.71	2.61	3.56	2.96	3.68	2.13	2.57
|bias|	13.33	11.86	4.64	4.69	1.48	1.37	11.42	13.27	4.74	5.73	1.44	1.73
sd	13.66	15.77	4.35	5.26	1.51	1.54	13.12	15.67	4.23	5.18	1.49	1.74
MSE	313.73	363.61	35.60	44.89	3.82	3.79	259.95	408.63	37.56	57.19	3.84	5.41
RE	1.00	0.86	1.00	0.79	1.00	1.01	1.00	0.64	1.00	0.66	1.00	0.71
Mean.as.sd	31.27	37.12	20.88	24.26	8.67	5.90	25.35	34.34	21.92	26.11	7.76	8.31

	#7		#8		#9		#16		#17		#18	
β_0	1.2		0.5		0.18		0.9		0.4		0.15	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	387	355	381	318	388	367	334	246	346	357	365	319.00
mean	4.43	4.56	1.56	1.79	0.55	0.66	4.12	4.70	1.51	1.84	0.58	0.85
bias	3.23	3.36	1.06	1.29	0.37	0.48	3.22	3.80	1.11	1.44	0.43	0.70
rel. bias	2.69	2.80	2.12	2.58	2.08	2.65	3.57	4.23	2.79	3.60	2.89	4.69
|bias|	3.77	3.56	1.21	1.40	0.43	0.51	3.70	3.99	1.24	1.54	0.48	0.73
sd	4.30	4.56	1.26	1.53	0.43	0.84	3.88	5.36	1.30	1.51	0.44	0.45
MSE	28.84	32.09	2.71	4.02	0.32	0.94	25.34	43.12	2.92	4.34	0.39	0.70
RE	1.00	0.90	1.00	0.68	1.00	0.35	1.00	0.59	1.00	0.67	1.00	0.55
Mean.as.sd	6.75	13.47	3.94	6.92	2.11	2.24	5.33	13.67	2.14	10.37	2.22	3.23

TABLE 25: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase stationnaire du processus « after »

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	205	223	408	359	410	420	243	331	415	320	448	496
mean	11.73	14.83	7.21	12.92	8.65	12.47	19.02	9.77	17.67	16.45	17.76	14.93
bias	10.73	13.83	6.21	11.92	7.65	11.47	9.02	-0.23	7.67	6.45	7.76	4.93
rel. bias	10.73	13.83	6.21	11.92	7.65	11.47	0.90	-0.02	0.77	0.64	0.78	0.49
|bias|	10.87	14.70	6.38	12.85	7.80	12.09	12.18	12.82	9.77	13.30	9.02	12.62
sd	15.37	25.85	12.84	26.01	15.23	20.08	18.53	18.85	13.12	21.23	12.51	16.44
MSE	350.32	856.42	203.06	817.00	290.09	533.79	423.53	354.33	230.62	491.04	216.33	294.12
RE	1.00	0.41	1.00	0.25	1.00	0.54	1.00	1.20	1.00	0.47	1.00	0.74
Mean.as.sd $\times 10^{-2}$	3.29	6.90	12.40	31.84	21.62	2686.56	561.81	3.80	199.99	13.56	54.15	16.29

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	374	363	397	384	412	422	406	405	361	398	423	436
mean	7.59	12.29	7.86	8.67	5.60	5.84	16.70	12.31	15.79	15.54	16.18	12.77
bias	6.59	11.29	6.86	7.67	4.60	4.84	6.70	2.31	5.79	5.54	6.18	2.77
rel. bias	6.59	11.29	6.86	7.67	4.60	4.84	0.67	0.23	0.58	0.55	0.62	0.28
|bias|	6.77	12.02	6.93	8.26	4.71	5.58	9.35	11.75	7.43	9.85	7.17	5.85
sd	13.14	23.55	12.04	16.03	8.46	10.97	15.76	18.94	9.16	15.54	10.07	6.15
MSE	215.58	680.71	191.58	315.31	92.61	143.42	292.67	363.14	117.25	271.52	139.37	45.34
RE	1.00	0.32	1.00	0.61	1.00	0.65	1.00	0.81	1.00	0.43	1.00	3.07
Mean.as.sd $\times 10^{-3}$	156.12	26.76	1.72	2.80	53.79	1.56	238.69	106.77	242.420	1141.41	96.06	1982.82

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	387	355	381	318	388	367	334	246	346	357	365	319.00
mean	8.03	6.90	6.34	6.14	5.35	2.34	16.40	12.70	17.33	13.54	15.36	12.23
bias	7.03	5.90	5.34	5.14	4.35	1.34	6.40	2.70	7.33	3.54	5.36	2.23
rel. bias	7.03	5.90	5.34	5.14	4.35	1.34	0.64	0.27	0.73	0.35	0.54	0.22
|bias|	7.39	6.73	5.43	5.86	4.37	2.21	8.16	10.66	8.63	6.72	5.75	3.36
sd	16.11	15.69	11.23	13.10	10.42	4.29	12.88	18.88	10.36	10.19	3.79	5.60
MSE	308.30	280.21	154.34	197.65	127.12	20.14	206.36	362.31	160.75	116.17	43.09	36.19
RE	1.00	1.10	1.00	0.78	1.00	6.31	1.00	0.57	1.00	1.38	1.00	1.19
Mean of as.sd $\times 10^{-3}$	75.18	3.95	73.75	1100.31	2.73	8881.57	43.37	293.36	14.11	4084.11	33.50	35488.51

TABLE 26: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after »

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	205	223	408	359	410	420	243	331	415	320	448	496
mean	2.83	1.56	2.01	2.15	4.46	4.10	3.97	4.59	2.25	2.28	4.55	4.20
bias	1.49	0.23	0.01	0.15	0.46	0.10	2.63	3.26	0.25	0.28	0.55	0.20
rel. bias	1.12	0.17	0.01	0.08	0.11	0.03	1.98	2.44	0.13	0.14	0.14	0.05
|bias|	1.64	0.30	0.16	0.27	0.74	0.40	2.75	3.32	0.39	0.39	0.84	0.49
sd	7.29	3.23	0.67	2.34	4.66	0.75	16.87	39.92	2.59	2.80	5.38	1.28
MSE	55.12	10.47	0.44	5.48	21.84	0.57	290.32	1599.13	6.74	7.87	29.19	1.68
RE	1.00	5.26	1.00	0.08	1.00	38.33	1.00	0.18	1.00	0.86	1.00	17.38
Mean.as.sd	132.08	32.48	0.60	166.45	7.90	0.42	451.76	4317.80	3.70	34.70	7.28	0.72

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	374	363	397	384	412	422	406	405	361	398	423	436
mean	1.83	1.35	2.07	2.61	4.19	4.12	2.35	1.40	2.10	2.28	4.23	4.13
bias	0.50	0.02	0.07	0.61	0.19	0.12	1.01	0.07	0.10	0.28	0.23	0.13
rel. bias	0.38	0.01	0.03	0.31	0.05	0.03	0.76	0.05	0.05	0.14	0.06	0.03
|bias|	0.62	0.11	0.21	0.72	0.46	0.40	1.12	0.18	0.24	0.42	0.50	0.40
sd	3.92	0.45	0.74	11.26	1.54	0.86	5.35	0.98	1.01	5.05	1.93	0.71
MSE	15.55	0.20	0.55	126.77	2.41	0.75	29.54	0.96	1.02	25.47	3.78	0.52
RE	1.00	77.07	1.00	0.00	1.00	3.23	1.00	30.80	1.00	0.04	1.00	7.29
Mean.as.sd	20.66	0.33	0.65	23.05	1.34	0.55	33.35	5.33	0.89	22.04	1.79	0.29

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	387	355	381	318	388	367	334	246	346	357	365	319
mean	3.41	1.30	2.46	2.01	4.18	4.02	3.48	1.66	3.35	2.26	4.16	4.18
bias	2.08	-0.03	0.46	0.01	0.18	0.02	2.15	0.33	1.35	0.26	0.16	0.18
rel. bias	1.56	-0.02	0.23	0.01	0.05	0.00	1.61	0.25	0.67	0.13	0.04	0.04
|bias|	2.21	0.07	0.60	0.14	0.44	0.32	2.27	0.41	1.49	0.39	0.41	0.48
sd	8.99	0.10	3.81	0.24	1.35	0.43	8.22	2.55	6.64	3.43	0.85	2.53
MSE	84.99	0.01	14.67	0.06	1.85	0.18	71.98	6.58	45.72	11.82	0.74	6.41
RE	1.00	7622.92	1.00	254.20	1.00	10.04	1.00	10.94	1.00	3.87	1.00	0.12
Mean.as.sd	142.58	0.03	10.90	0.09	1.00	0.16	127.59	60.64	31.97	9.04	0.92	11.28

TABLE 27: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « after » (95% des estimations)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	193.00	211.00	386.00	341.00	388.00	398.00	229.00	313.00	393.00	304.00	424.00	470.00
mean [1]	1.71	1.34	1.98	2.02	4.05	4.05	1.83	1.36	2.01	2.03	4.08	4.08
bias	0.37	0.01	-0.02	0.02	0.05	0.05	0.50	0.03	0.01	0.03	0.08	0.08
rel. bias	0.28	0.01	-0.01	0.01	0.01	0.01	0.37	0.02	0.00	0.01	0.02	0.02
|bias|	0.51	0.07	0.11	0.12	0.30	0.31	0.60	0.08	0.12	0.13	0.31	0.34
sd	1.53	0.09	0.13	0.15	0.37	0.41	1.83	0.11	0.17	0.17	0.39	0.43
MSE	2.47	0.01	0.02	0.02	0.14	0.17	3.60	0.01	0.03	0.03	0.16	0.19
RE	1.00	309.63	1.00	0.74	1.00	0.85	1.00	275.02	1.00	0.97	1.00	0.82
Mean.as.sd	5.10	0.03	0.15	0.06	0.47	0.23	5.78	0.06	0.18	0.08	0.45	0.23

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	354.00	343.00	377.00	364.00	390.00	400.00	384.00	383.00	341.00	378.00	401.00	414.00
mean	1.36	1.32	1.98	2.01	4.07	4.06	1.61	1.30	1.99	1.99	4.08	4.08
bias	0.03	-0.02	-0.02	0.01	0.07	0.06	0.28	-0.03	-0.01	-0.01	0.08	0.08
rel. bias	0.02	-0.01	-0.01	0.01	0.02	0.02	0.21	-0.02	0.00	-0.01	0.02	0.02
|bias|	0.14	0.07	0.11	0.11	0.31	0.31	0.38	0.07	0.11	0.11	0.31	0.32
sd	0.30	0.08	0.14	0.14	0.38	0.38	1.32	0.09	0.14	0.14	0.39	0.40
MSE	0.09	0.01	0.02	0.02	0.15	0.15	1.82	0.01	0.02	0.02	0.15	0.16
RE	1.00	13.76	1.00	1.12	1.00	1.02	1.00	211.78	1.00	1.04	1.00	0.94
Mean.as.sd	0.49	0.03	0.15	0.06	0.48	0.19	3.06	0.02	0.14	0.05	0.55	0.21

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
true value	1.33	1.33	2.00	2.00	4.00	4.00	1.33	1.33	2.00	2.00	4.00	4.00
# sims	367.00	337.00	361.00	302.00	368.00	347.00	316.00	232.00	328.00	339.00	345.00	303.00
mean	2.21	1.30	1.98	2.00	4.08	4.00	2.40	1.33	2.37	2.02	4.09	4.03
bias	0.88	-0.03	-0.02	0.00	0.08	0.00	1.07	0.00	0.37	0.02	0.09	0.03
rel. bias	0.66	-0.02	-0.01	0.00	0.02	0.00	0.80	0.00	0.19	0.01	0.02	0.01
|bias|	1.01	0.06	0.11	0.12	0.31	0.28	1.18	0.07	0.50	0.13	0.30	0.30
sd	4.81	0.07	0.13	0.15	0.39	0.34	4.77	0.10	2.52	0.17	0.37	0.37
MSE	23.89	0.01	0.02	0.02	0.16	0.11	23.80	0.01	6.46	0.03	0.14	0.14
RE	1.00	4176.77	1.00	0.84	1.00	1.38	1.00	2465.53	1.00	218.56	1.00	1.05
Mean.as.sd	36.43	0.03	0.15	0.05	0.43	0.15	35.37	0.04	5.48	0.07	0.51	0.17

[1] Les caractéristiques présentées dans le tableau sont évaluées pour 95% des estimations.
5% des estimations extrêmes ont été écartées

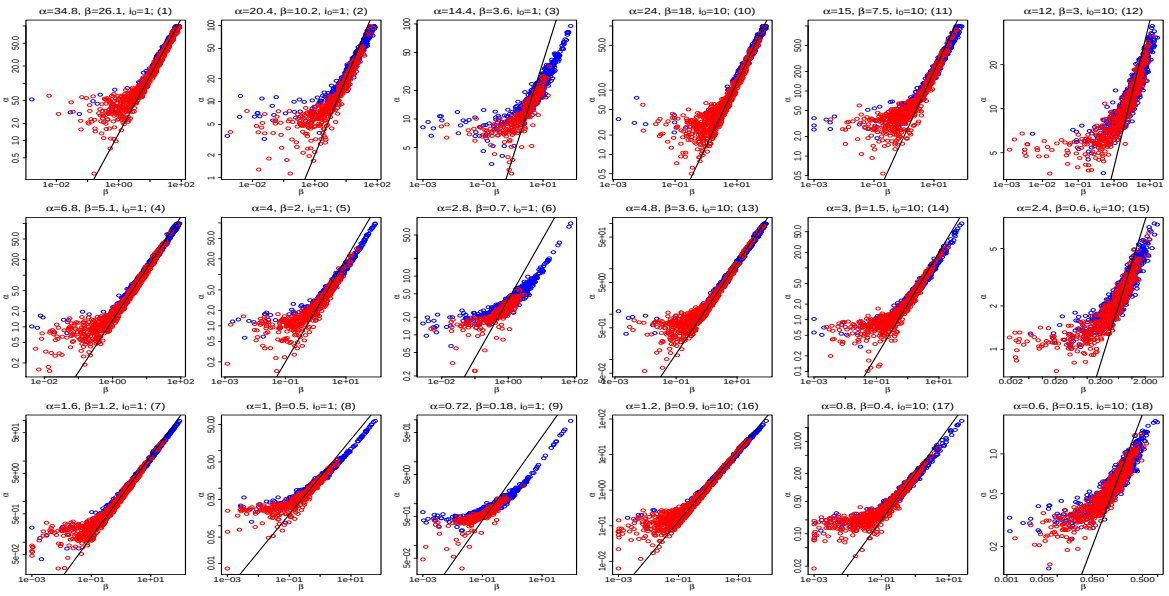

FIGURE 56: Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase initiale : « early » ; modèle (M) - points rouges, modèle (S) - points bleus

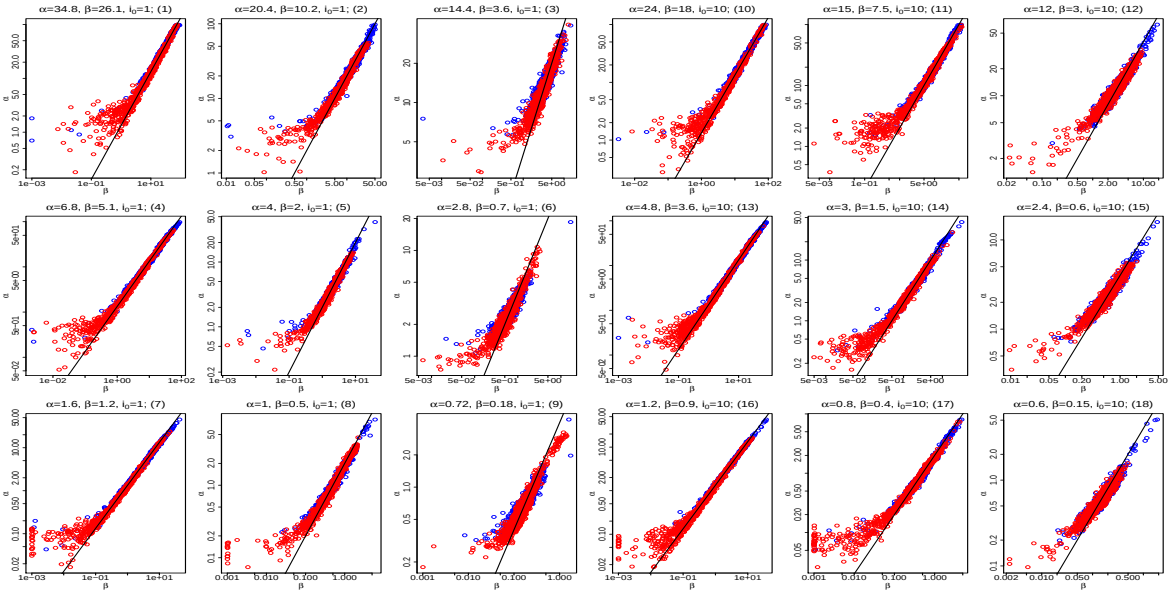

FIGURE 57: Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus

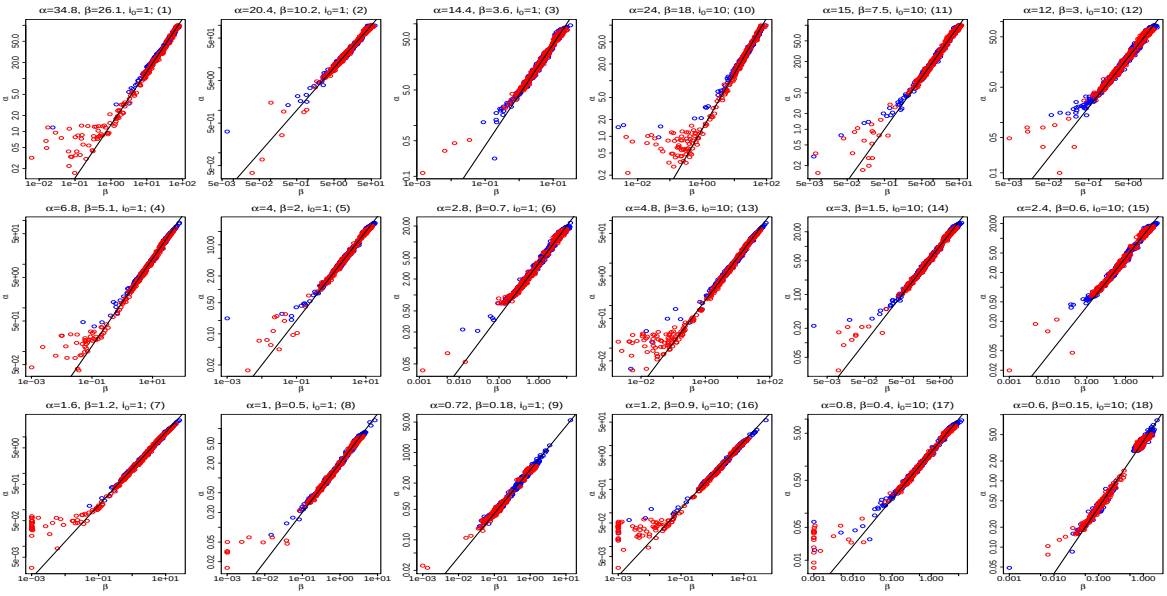

FIGURE 58: Les estimations de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase stationnaire : « after » ; modèle (M) - points rouges, modèle (S) - points bleus

TABLE 28: Le nombre des estimations « utilisables » (β fixe)

	Phase :	« early »			« during »			« after »		
#	Modèle :	(M)	(S)	(M)∩(S)	(M)	(S)	(M)∩(S)	(M)	(S)	(M)∩(S)
1		1000	1000	1000	873	888	804	629	555	372
2		942	976	938	985	998	984	584	599	325
3		977	992	975	995	998	993	603	664	366
4		866	959	856	950	986	947	368	637	181
5		934	979	930	1000	999	999	561	624	304
6		977	989	976	994	999	993	569	692	363
7		830	940	818	951	965	934	320	679	163
8		937	974	928	991	989	981	576	624	336
9		985	992	983	998	969	967	588	566	275
10		771	899	758	859	918	855	268	815	192
11		979	996	979	991	996	987	436	801	382
12		993	1000	993	995	1000	995	512	647	311
13		817	920	811	925	969	923	256	798	205
14		987	998	987	991	996	987	442	700	301
15		999	997	996	994	1000	994	550	626	303
16		823	943	817	932	960	917	277	708	174
17		985	997	984	989	991	981	469	650	284
18		996	1000	996	994	997	991	532	633	312

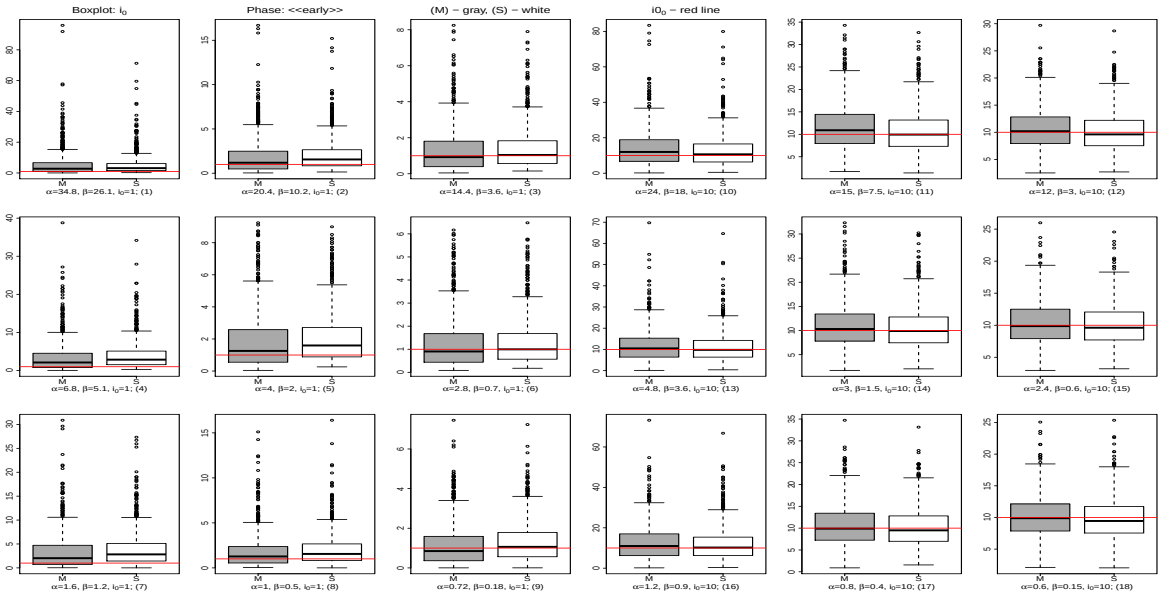

FIGURE 59: Boxplots des estimations de i_0 pour les observations de la phase initiale : « early » (β-fixe)

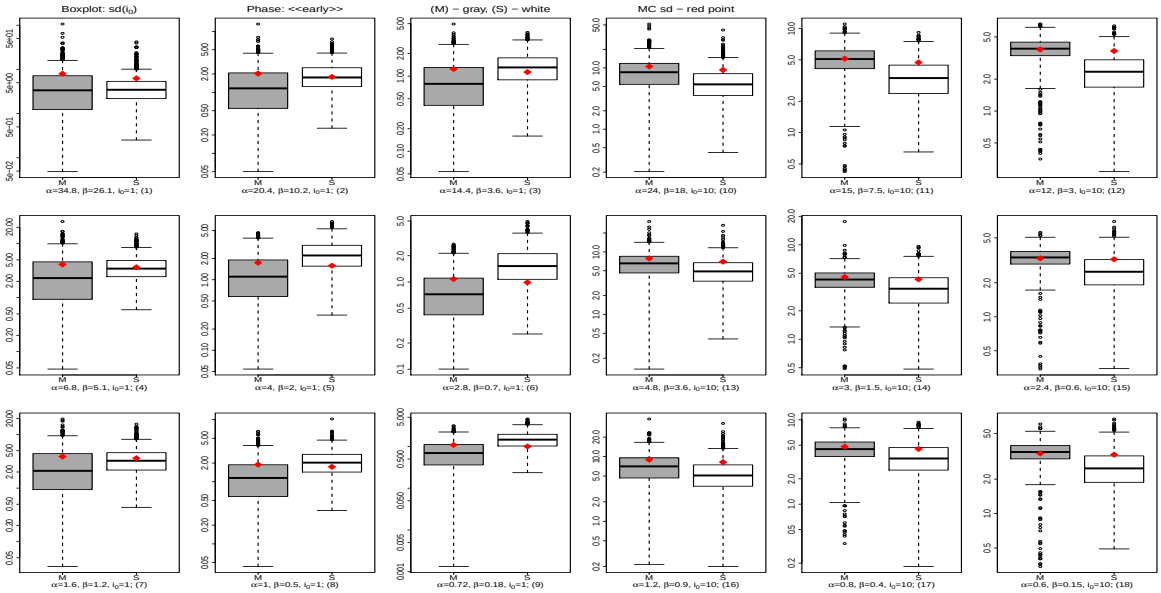

FIGURE 60: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase initiale : « early » (β-fixe)

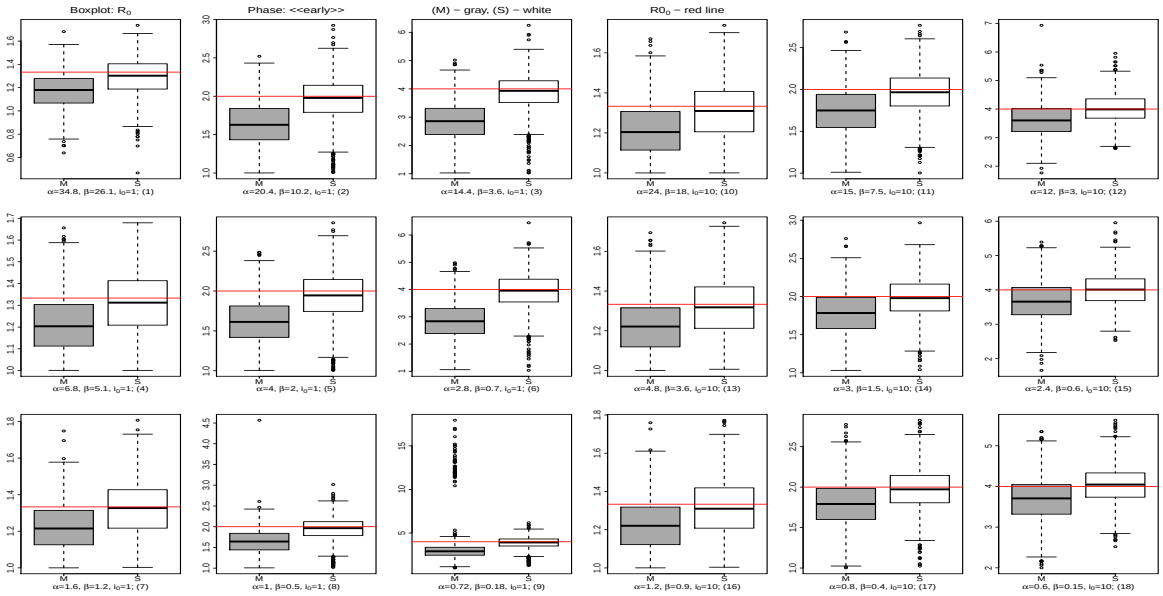

FIGURE 61: Boxplots des estimations de R_0 pour les observations de la phase initiale : « early »

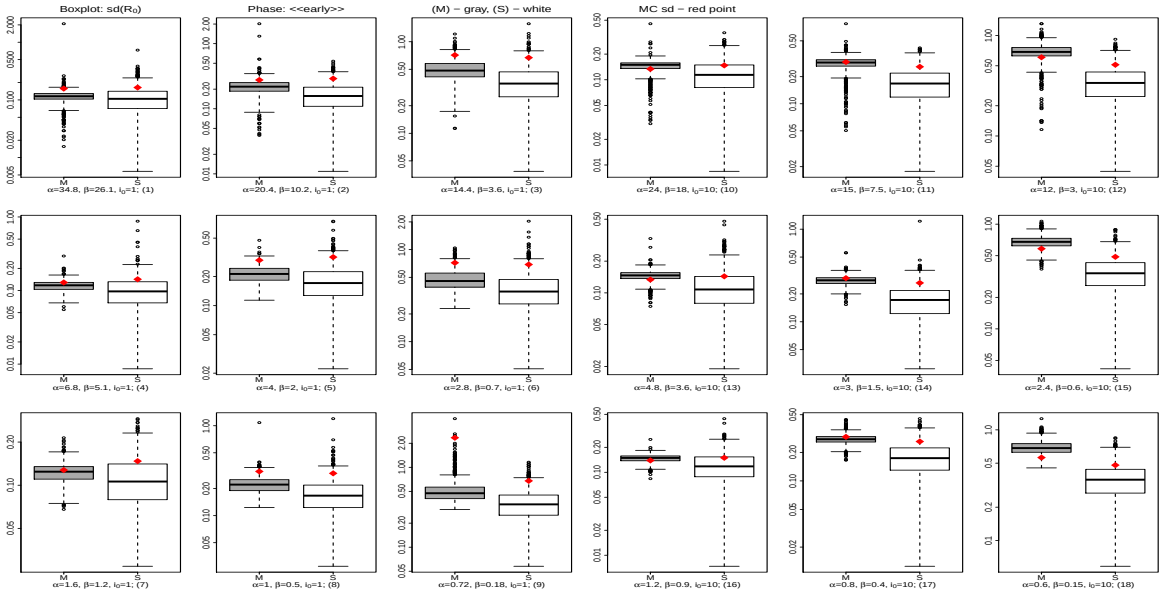

FIGURE 62: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase initiale : « early » (β-fixe)

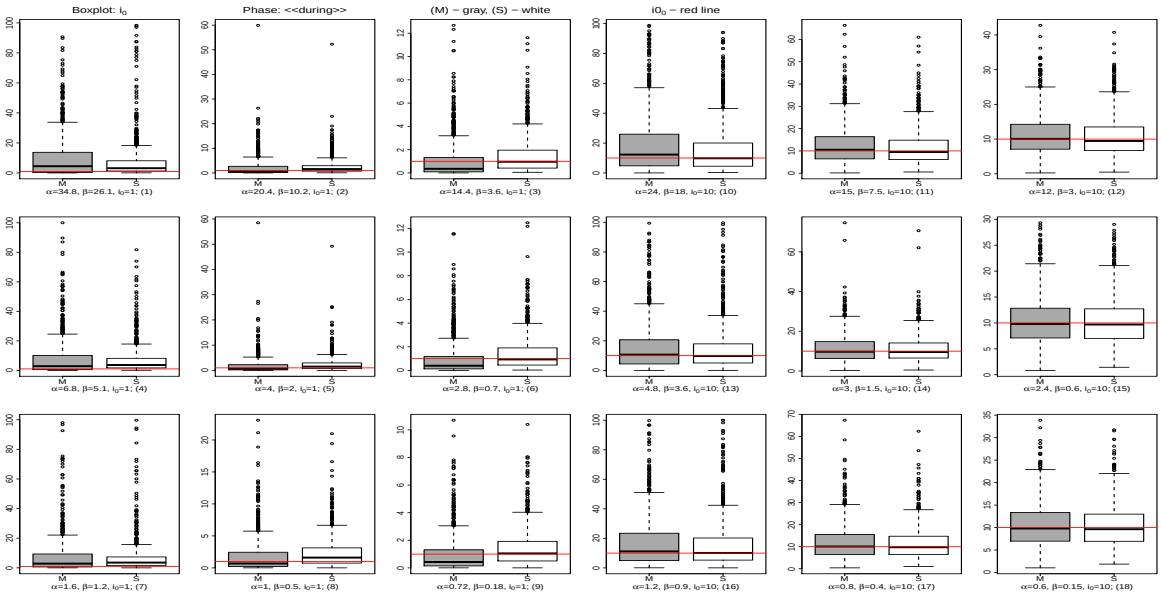

FIGURE 63: Boxplots des estimations de i_0 pour les observations de la phase évolutive : « during » (β-fixe)

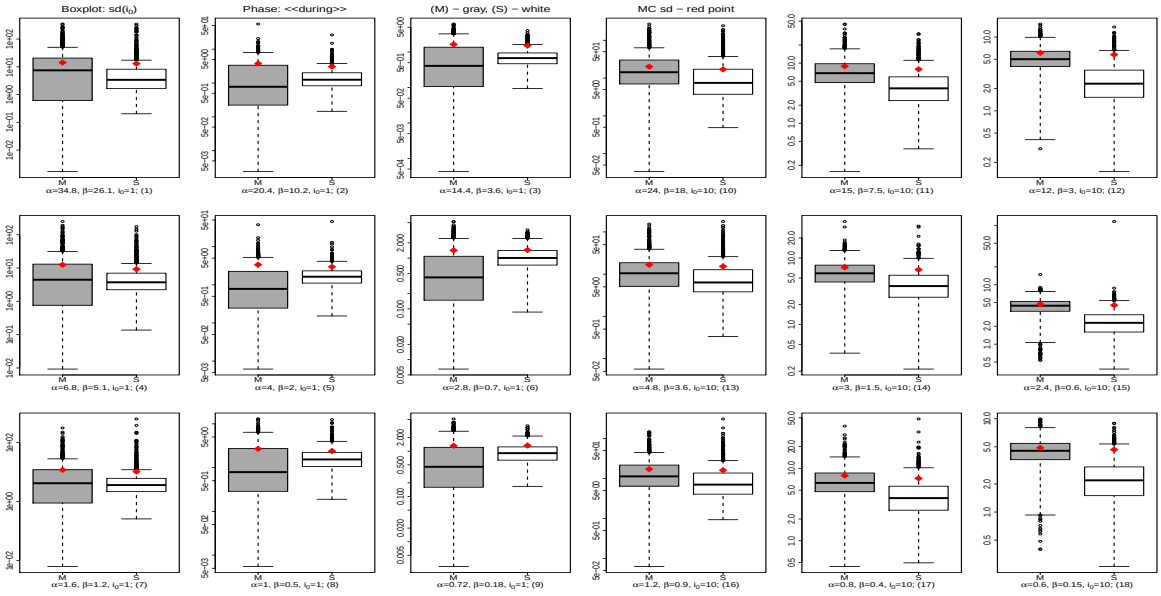

FIGURE 64: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase évolutive : « during » (β-fixe)

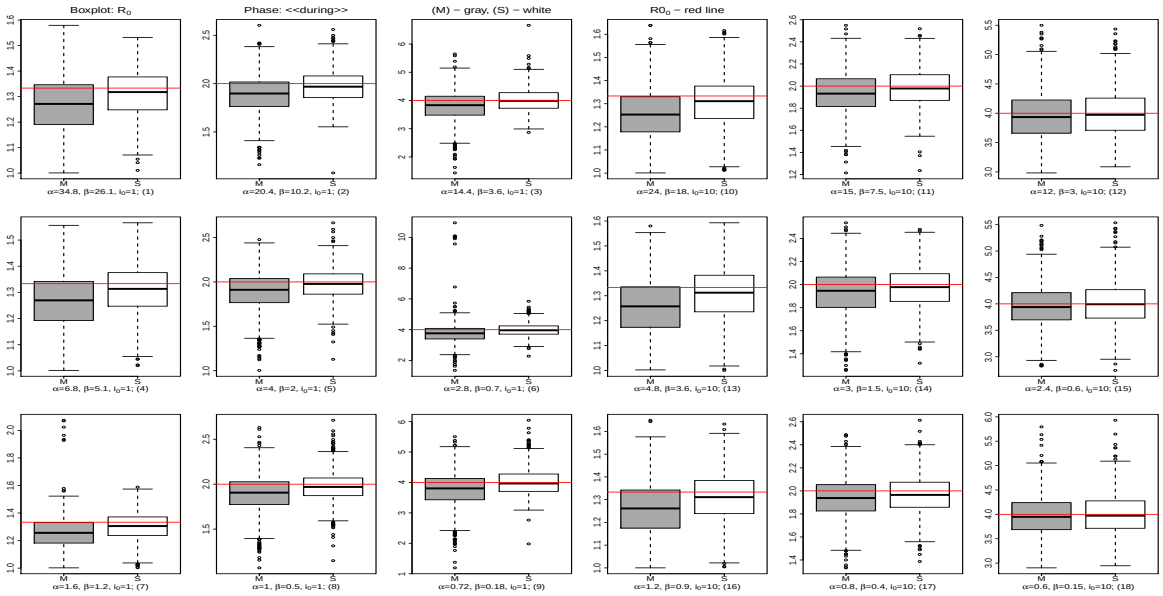

FIGURE 65: Boxplots des estimations de R_0 pour les observations de la phase évolutive : « during » (β-fixe)

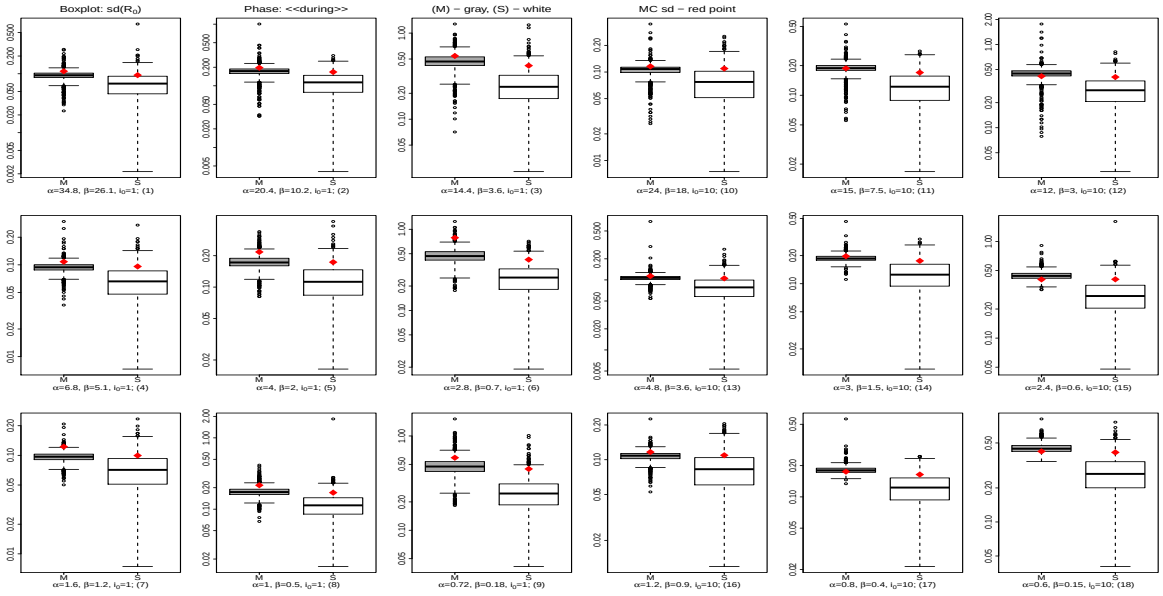

FIGURE 66: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase évolutive : « during » (β-fixe)

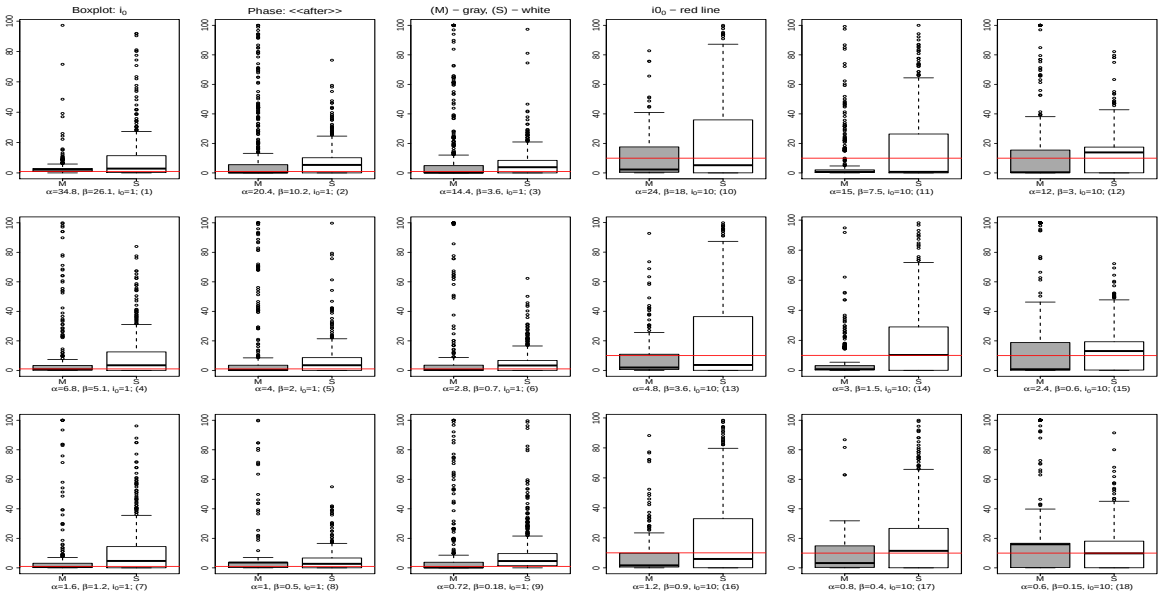

FIGURE 67: Boxplots des estimations de i_0 pour les observations de la phase stationnaire : « after » (β-fixe)

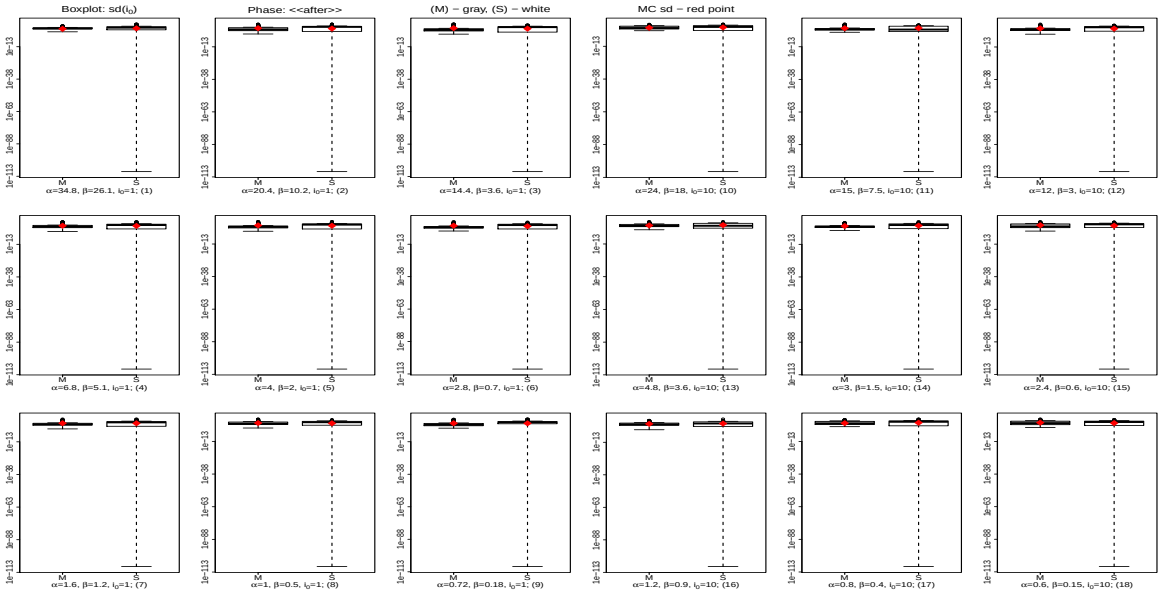

FIGURE 68: Boxplots des estimations de l'écart-type asymptotique de i_0 pour les observations de la phase stationnaire : « after » (β-fixe)

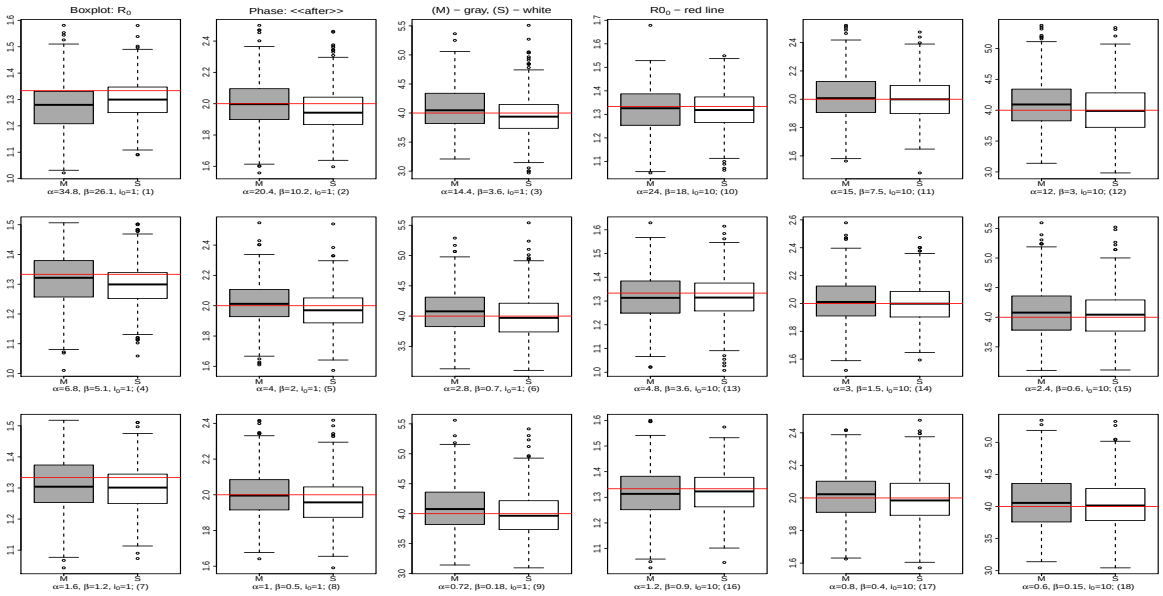

FIGURE 69: Boxplots des estimations de R_0 pour les observations de la phase stationnaire : « after » (β-fixe)

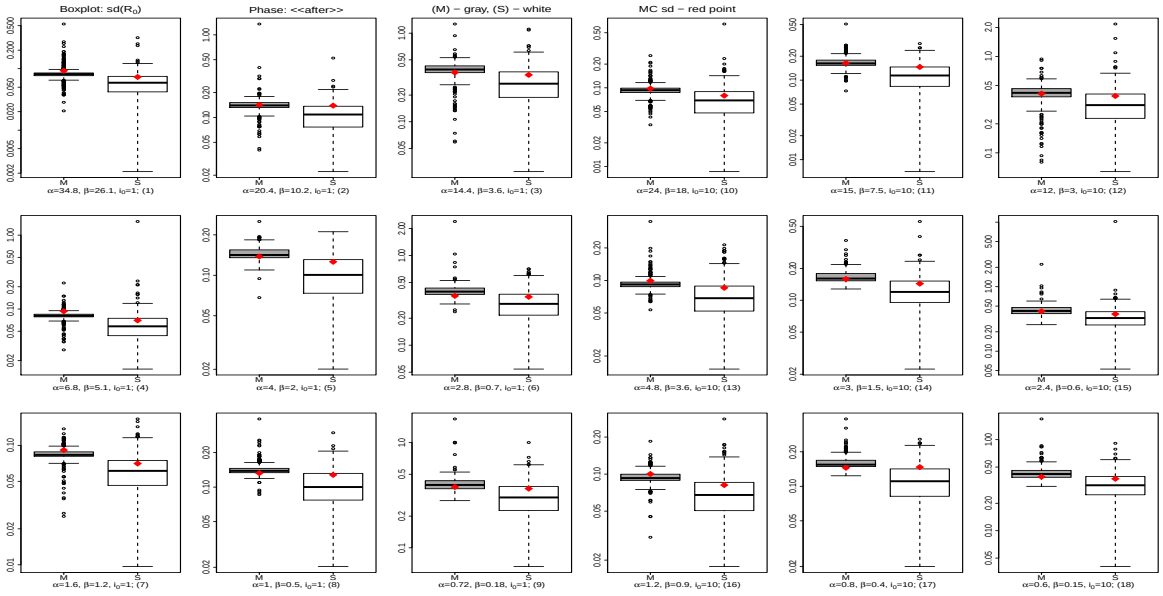

FIGURE 70: Boxplots des estimations de l'écart-type asymptotique de R_0 pour les observations de la phase stationnaire : « after » (β-fixe)

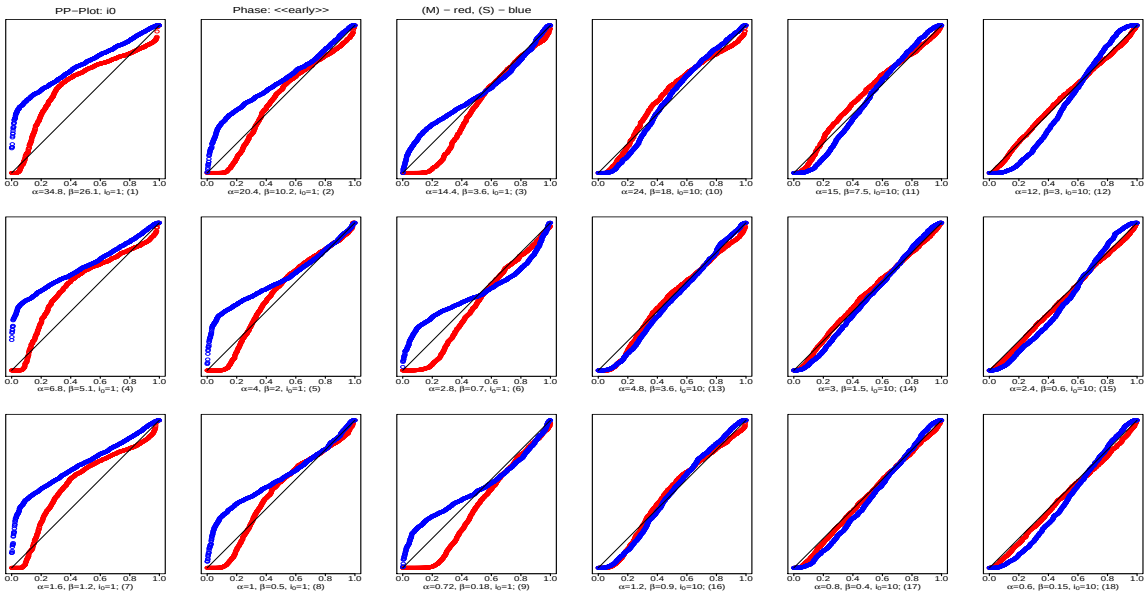

FIGURE 71: PP-plots de $\theta = i_0$ pour les observations de la phase initale : « early » (β-fixe)

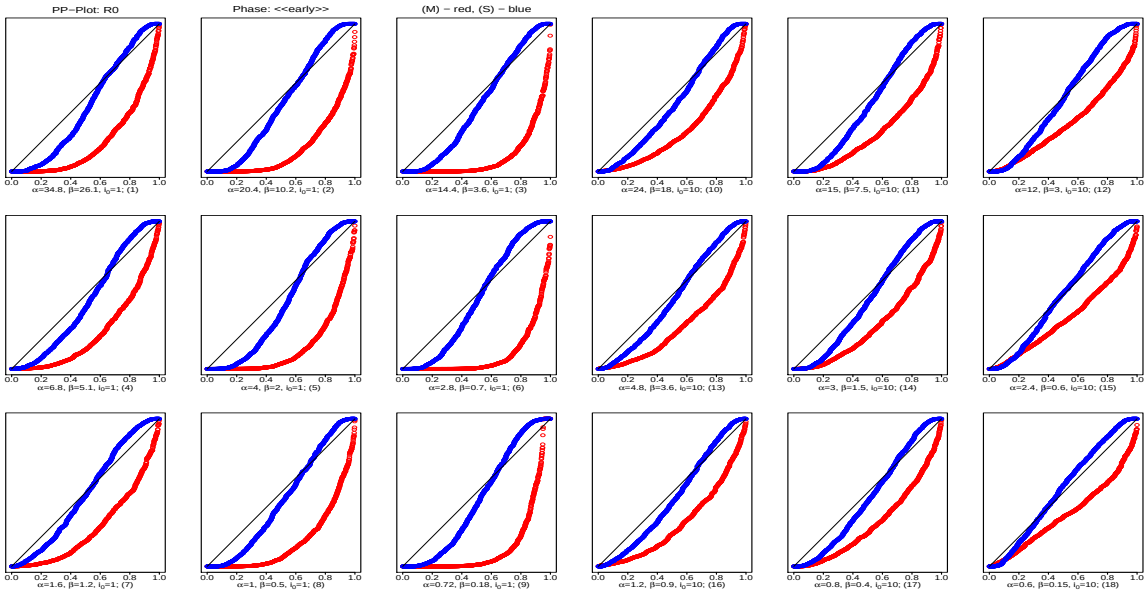

FIGURE 72: PP-plots de R_0 pour les observations de la phase initale : « early » (β-fixe)

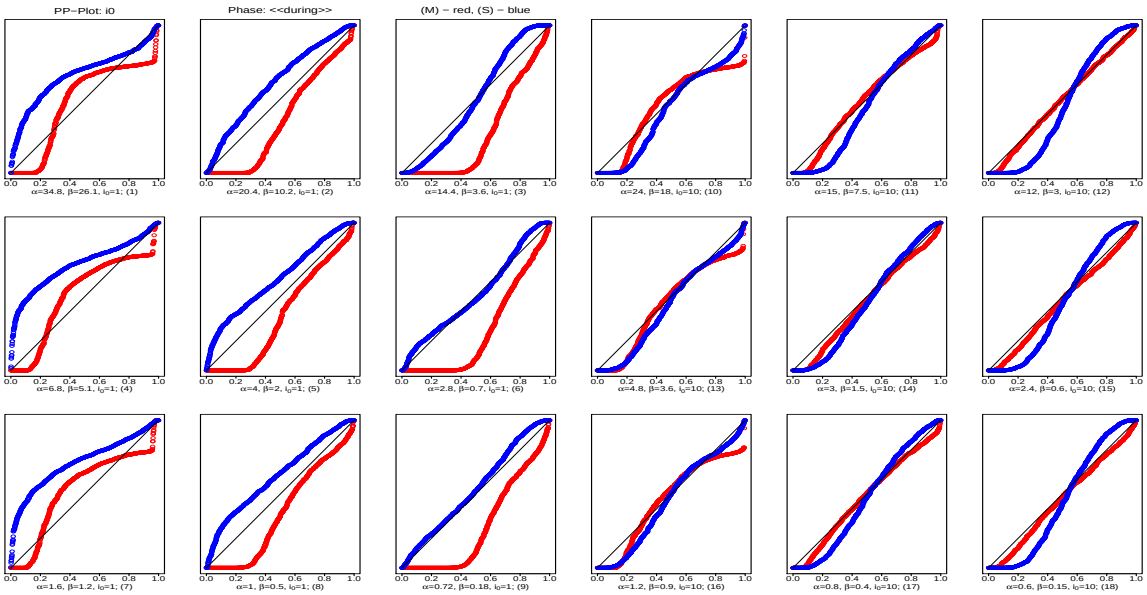

FIGURE 73: PP-plots de $\theta = i_0$ pour les observations de la phase évolutive : « during » (β-fixe)

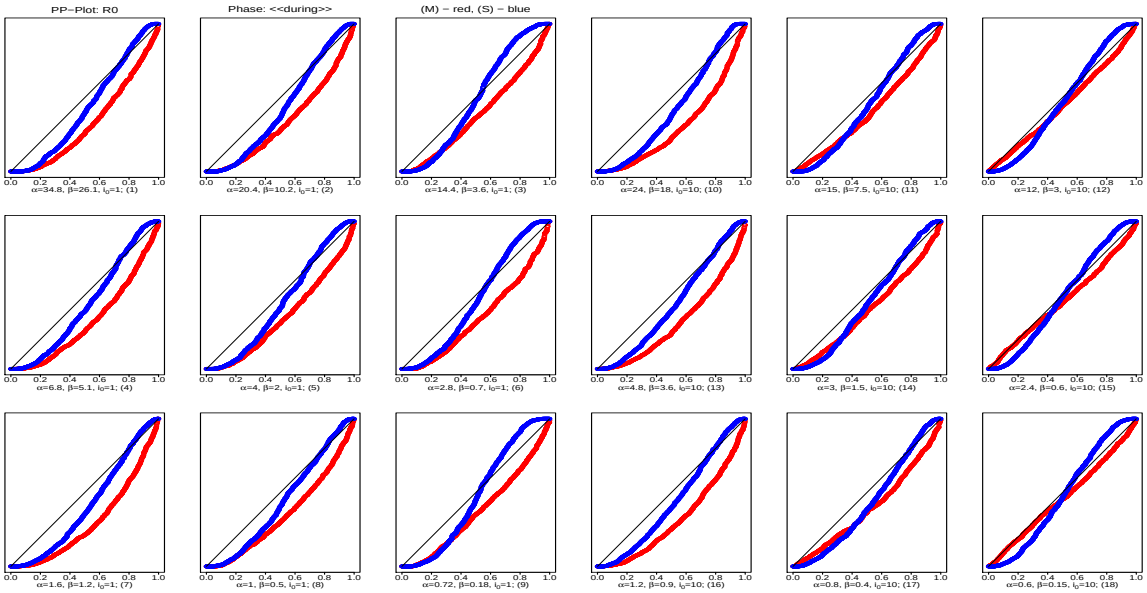

FIGURE 74: PP-plots de R_0 pour les observations de la phase évolutive : « during » (β-fixe)

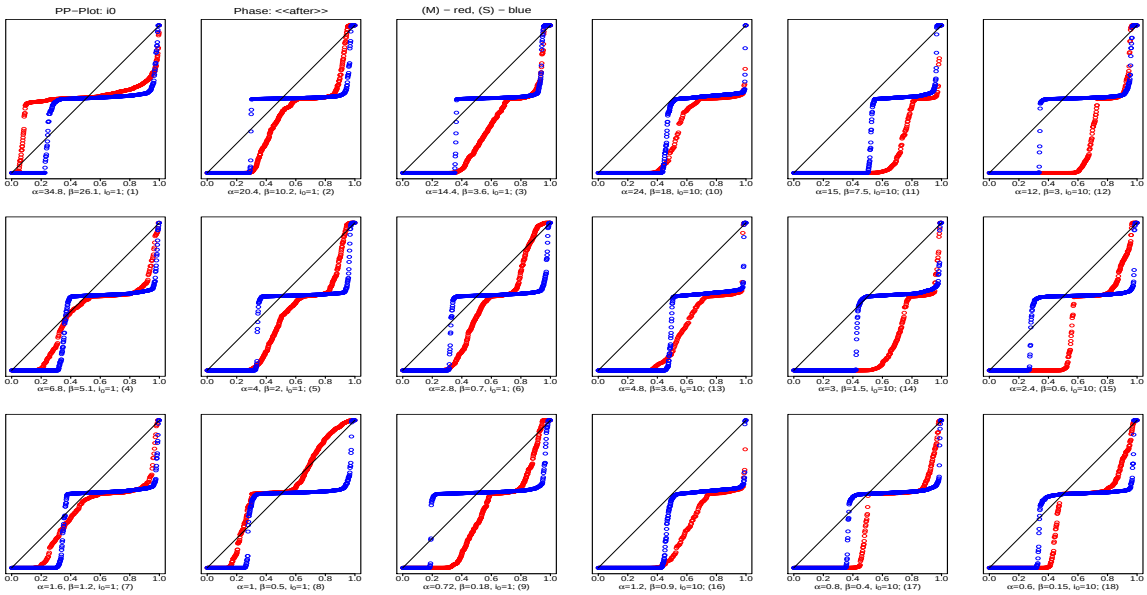

FIGURE 75: PP-plots de $\theta = i_0$ pour les observations de la phase stationnaire : « after » (β-fixe)

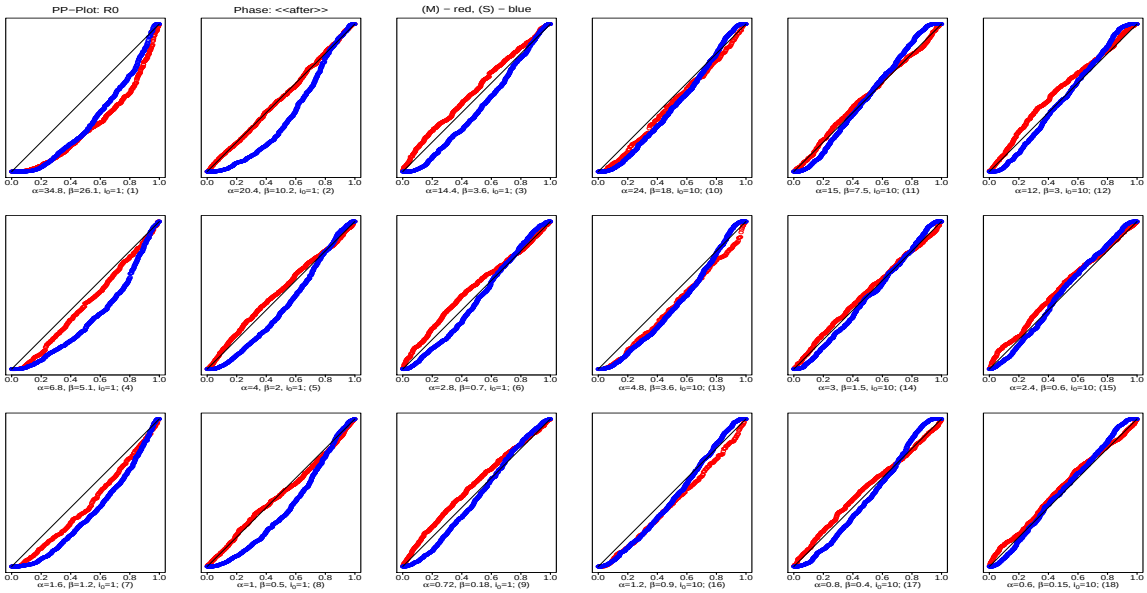

FIGURE 76: PP-plots de R_0 pour les observations de la phase stationnaire : « after » (β-fixe)

TABLE 29: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase initiale du processus « early » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	835.00	928.00	942.00	976.00	977.00	992.00	771.00	899.00	979.00	996.00	993.00	1000.00
mean	4.81	4.79	1.86	2.06	1.32	1.37	14.10	12.76	11.50	10.62	10.67	10.13
bias	3.81	3.79	0.86	1.06	0.32	0.37	4.10	2.76	1.50	0.62	0.67	0.13
rel. bias	3.81	3.79	0.86	1.06	0.32	0.37	0.41	0.28	0.15	0.06	0.07	0.01
|bias|	4.10	3.86	1.34	1.30	0.89	0.78	7.60	6.42	4.05	3.64	2.96	2.87
sd	7.08	5.47	2.03	1.79	1.25	1.13	10.55	9.26	5.13	4.72	3.81	3.69
MSE	64.59	44.30	4.86	4.33	1.65	1.42	127.89	93.21	28.52	22.61	14.95	13.59
RE	1.00	1.46	1.00	1.12	1.00	1.16	1.00	1.37	1.00	1.26	1.00	1.10
Mean.as.sd	3.11	4.21	2.27	1.96	1.82	1.38	2.59	6.38	2.10	3.55	2.06	2.43

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	866.00	959.00	934.00	979.00	977.00	989.00	817.00	920.00	987.00	998.00	999.00	997.00
mean	3.42	3.90	1.83	2.04	1.24	1.30	11.83	11.14	10.95	10.49	10.33	10.08
bias	2.42	2.90	0.83	1.04	0.24	0.30	1.83	1.14	0.95	0.49	0.33	0.08
rel. bias	2.42	2.90	0.83	1.04	0.24	0.30	0.18	0.11	0.10	0.05	0.03	0.01
|bias|	2.73	2.99	1.27	1.26	0.76	0.68	5.66	4.99	3.47	3.27	2.61	2.58
sd	4.11	3.65	1.74	1.58	1.08	0.99	7.83	6.97	4.56	4.31	3.30	3.23
MSE	22.69	21.71	3.72	3.58	1.22	1.06	64.64	49.86	21.67	18.83	10.97	10.44
RE	1.00	1.05	1.00	1.04	1.00	1.15	1.00	1.30	1.00	1.15	1.00	1.05
Mean.as.sd	0.60	3.88	0.43	2.37	0.34	1.65	0.52	5.43	0.42	3.60	0.41	2.60

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	830.00	940.00	937.00	974.00	985.00	992.00	823.00	943.00	985.00	997.00	996.00	1000.00
mean	3.33	3.86	1.81	1.99	1.17	1.33	12.78	12.07	10.69	10.23	10.19	9.89
bias	2.33	2.86	0.81	0.99	0.17	0.33	2.78	2.07	0.69	0.23	0.19	-0.11
rel. bias	2.33	2.86	0.81	0.99	0.17	0.33	0.28	0.21	0.07	0.02	0.02	-0.01
|bias|	2.66	2.98	1.25	1.27	0.79	0.73	6.62	5.83	3.73	3.51	2.63	2.59
sd	3.89	3.61	1.89	1.74	1.09	1.00	8.96	8.23	4.80	4.51	3.36	3.27
MSE	20.52	21.24	4.22	4.00	1.21	1.10	87.99	71.88	23.53	20.39	11.29	10.69
RE	1.00	0.97	1.00	1.05	1.00	1.10	1.00	1.22	1.00	1.15	1.00	1.06
Mean.as.sd	0.15	3.69	0.11	2.18	0.09	1.55	0.13	5.84	0.11	3.69	0.10	2.56

TABLE 30: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase initiale du processus « early » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	835.00	928.00	942.00	976.00	977.00	992.00	771.00	899.00	979.00	996.00	993.00	1000.00
mean	1.21	1.31	1.63	1.95	2.86	3.86	1.22	1.31	1.74	1.96	3.62	4.01
bias	-0.12	-0.02	-0.37	-0.05	-1.14	-0.14	-0.12	-0.03	-0.26	-0.04	-0.38	0.01
rel. bias	-0.09	-0.02	-0.18	-0.03	-0.29	-0.03	-0.09	-0.02	-0.13	-0.02	-0.10	0.00
|bias|	0.15	0.11	0.39	0.22	1.18	0.50	0.15	0.12	0.31	0.20	0.58	0.40
sd	0.12	0.14	0.28	0.29	0.71	0.67	0.13	0.15	0.29	0.26	0.61	0.51
MSE	0.03	0.02	0.21	0.09	1.81	0.47	0.03	0.02	0.15	0.07	0.51	0.26
RE	1.00	1.55	1.00	2.44	1.00	3.88	1.00	1.40	1.00	2.22	1.00	1.96
Mean.as.sd $\times 10^2$	0.12	0.10	0.22	0.16	0.51	0.37	0.14	0.12	0.28	0.17	0.69	0.35

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	866.00	959.00	934.00	979.00	977.00	989.00	817.00	920.00	987.00	998.00	999.00	997.00
mean	1.22	1.31	1.62	1.93	2.85	3.91	1.22	1.32	1.78	1.98	3.67	4.02
bias	-0.12	-0.02	-0.38	-0.07	-1.15	-0.09	-0.11	-0.02	-0.22	-0.02	-0.33	0.02
rel. bias	-0.09	-0.02	-0.19	-0.04	-0.29	-0.02	-0.08	-0.01	-0.11	-0.01	-0.08	0.00
|bias|	0.15	0.12	0.41	0.25	1.19	0.53	0.14	0.12	0.30	0.21	0.54	0.38
sd	0.13	0.14	0.30	0.32	0.72	0.69	0.13	0.14	0.30	0.26	0.58	0.49
MSE	0.03	0.02	0.23	0.11	1.85	0.48	0.03	0.02	0.14	0.07	0.45	0.24
RE	1.00	1.46	1.00	2.16	1.00	3.84	1.00	1.42	1.00	1.94	1.00	1.89
Mean.as.sd	0.12	0.10	0.22	0.18	0.48	0.38	0.15	0.12	0.28	0.18	0.68	0.35

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	830.00	940.00	937.00	974.00	985.00	992.00	823.00	943.00	985.00	997.00	996.00	1000.00
mean	1.23	1.32	1.65	1.94	3.37	3.88	1.23	1.31	1.79	1.97	3.69	4.05
bias	-0.11	-0.01	-0.35	-0.06	-0.63	-0.12	-0.10	-0.02	-0.21	-0.03	-0.31	0.05
rel. bias	-0.08	-0.01	-0.18	-0.03	-0.16	-0.03	-0.08	-0.01	-0.11	-0.01	-0.08	0.01
|bias|	0.14	0.12	0.39	0.22	1.57	0.51	0.15	0.12	0.29	0.20	0.51	0.38
sd	0.13	0.15	0.31	0.30	2.35	0.68	0.14	0.15	0.29	0.26	0.56	0.48
MSE	0.03	0.02	0.22	0.09	5.90	0.47	0.03	0.02	0.13	0.07	0.41	0.23
RE	1.00	1.28	1.00	2.43	1.00	12.53	1.00	1.33	1.00	1.89	1.00	1.78
Mean.as.sd	0.12	0.11	0.22	0.18	0.53	0.37	0.15	0.12	0.28	0.18	0.69	0.36

TABLE 31: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase évolutive du processus « during » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	873.00	888.00	985.00	998.00	995.00	998.00	859.00	918.00	991.00	996.00	995.00	1000.00
MC mean	9.83	7.53	2.10	2.38	1.06	1.46	19.00	15.80	12.51	11.58	11.37	10.81
bias	8.83	6.53	1.10	1.38	0.06	0.46	9.00	5.80	2.51	1.58	1.37	0.81
rel. bias	8.83	6.53	1.10	1.38	0.06	0.46	0.90	0.58	0.25	0.16	0.14	0.08
|bias|	9.36	6.71	1.92	1.70	1.10	1.01	14.00	10.97	6.42	5.63	4.51	4.28
sd	14.20	13.04	3.76	3.05	1.65	1.52	20.02	16.96	8.87	7.91	6.11	5.78
MSE	279.39	212.51	15.29	11.20	2.72	2.53	481.13	321.10	84.80	65.01	39.22	33.98
RE	1.00	1.31	1.00	1.37	1.00	1.08	1.00	1.50	1.00	1.30	1.00	1.15
Mean.as.sd	2.46	11.43	1.77	1.65	1.70	0.78	1.89	16.53	1.42	4.90	1.35	2.80

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950.00	986.00	1000.00	999.00	994.00	999.00	925.00	969.00	991.00	996.00	994.00	1000.00
mean	7.94	6.82	1.85	2.35	0.97	1.41	16.01	14.62	11.20	11.00	10.45	10.36
bias	6.94	5.82	0.85	1.35	-0.03	0.41	6.01	4.62	1.20	1.00	0.45	0.36
rel. bias	6.94	5.82	0.85	1.35	-0.03	0.41	0.60	0.46	0.12	0.10	0.05	0.04
|bias|	7.39	5.92	1.63	1.67	0.96	0.94	11.14	9.47	5.20	4.75	3.59	3.48
sd	12.45	9.25	3.36	2.96	1.43	1.45	16.68	15.25	7.22	6.65	4.70	4.53
MSE	203.10	119.22	11.99	10.55	2.03	2.26	314.11	253.61	53.48	45.19	22.27	20.65
RE	1.00	1.70	1.00	1.14	1.00	0.90	1.00	1.24	1.00	1.18	1.00	1.08
Mean.as.sd	0.48	7.35	0.35	1.90	0.34	1.11	0.39	11.43	0.28	4.39	0.27	2.60

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	951.00	965.00	991.00	989.00	998.00	969.00	932.00	960.00	989.00	991.00	994.00	997.00
mean	7.32	6.65	1.82	2.29	0.96	1.43	16.94	15.53	11.83	11.46	10.43	10.33
bias	6.32	5.65	0.82	1.29	-0.04	0.43	6.94	5.53	1.83	1.46	0.43	0.33
rel. bias	6.32	5.65	0.82	1.29	-0.04	0.43	0.69	0.55	0.18	0.15	0.04	0.03
|bias|	6.72	5.78	1.59	1.62	0.92	0.92	11.87	10.24	5.70	5.20	3.75	3.59
sd	11.86	10.40	2.69	2.38	1.32	1.33	17.08	15.99	7.95	7.28	4.87	4.66
MSE	180.48	139.86	7.90	7.34	1.74	1.96	339.44	286.05	66.41	55.03	23.90	21.80
RE	1.00	1.29	1.00	1.08	1.00	0.88	1.00	1.19	1.00	1.21	1.00	1.10
Mean.as.sd	0.12	8.18	0.09	1.79	0.09	1.00	0.10	12.65	0.07	4.57	0.07	2.45

TABLE 32: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase évolutive du processus « during » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	873.00	888.00	985.00	998.00	995.00	998.00	859.00	918.00	991.00	996.00	995.00	1000.00
mean	1.27	1.31	1.89	1.97	3.80	4.02	1.26	1.31	1.94	1.98	3.96	4.00
bias	-0.06	-0.02	-0.11	-0.03	-0.20	0.02	-0.08	-0.03	-0.06	-0.02	-0.04	0.00
rel. bias	-0.05	-0.02	-0.06	-0.01	-0.05	0.00	-0.06	-0.02	-0.03	-0.01	-0.01	0.00
|bias|	0.10	0.08	0.18	0.13	0.44	0.33	0.11	0.09	0.16	0.14	0.34	0.32
sd	0.11	0.09	0.20	0.17	0.54	0.42	0.12	0.11	0.19	0.17	0.42	0.41
MSE	0.02	0.01	0.05	0.03	0.33	0.18	0.02	0.01	0.04	0.03	0.18	0.17
RE	1.00	1.73	1.00	1.76	1.00	1.90	1.00	1.53	1.00	1.33	1.00	1.06
Mean.as.sd	0.09	0.07	0.17	0.12	0.47	0.26	0.11	0.08	0.19	0.12	0.45	0.29

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950.00	986.00	1000.00	999.00	994.00	999.00	925.00	969.00	991.00	996.00	994.00	1000.00
mean	1.27	1.31	1.89	1.98	3.76	3.99	1.26	1.31	1.93	1.97	3.97	4.01
bias	-0.07	-0.02	-0.11	-0.02	-0.24	-0.01	-0.08	-0.03	-0.07	-0.03	-0.03	0.01
rel. bias	-0.05	-0.02	-0.06	-0.01	-0.06	0.00	-0.06	-0.02	-0.03	-0.01	-0.01	0.00
|bias|	0.10	0.08	0.18	0.14	0.53	0.33	0.11	0.09	0.16	0.14	0.32	0.32
sd	0.11	0.10	0.22	0.17	0.79	0.42	0.11	0.11	0.20	0.18	0.40	0.41
MSE	0.02	0.01	0.06	0.03	0.68	0.18	0.02	0.01	0.04	0.03	0.16	0.16
RE	1.00	1.64	1.00	1.94	1.00	3.79	1.00	1.59	1.00	1.38	1.00	1.00
Mean.as.sd	0.09	0.07	0.18	0.12	0.48	0.26	0.11	0.08	0.19	0.13	0.45	0.28

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	951.00	965.00	991.00	989.00	998.00	969.00	932.00	960.00	989.00	991.00	994.00	997.00
mean	1.26	1.30	1.89	1.98	3.76	4.01	1.26	1.31	1.94	1.97	3.97	4.00
bias	-0.07	-0.03	-0.11	-0.02	-0.24	0.01	-0.08	-0.02	-0.06	-0.03	-0.03	0.00
rel. bias	-0.05	-0.02	-0.05	-0.01	-0.06	0.00	-0.06	-0.02	-0.03	-0.02	-0.01	0.00
|bias|	0.11	0.08	0.18	0.13	0.48	0.35	0.11	0.09	0.15	0.13	0.33	0.33
sd	0.12	0.10	0.22	0.17	0.59	0.45	0.12	0.11	0.18	0.16	0.42	0.41
MSE	0.02	0.01	0.06	0.03	0.40	0.20	0.02	0.01	0.03	0.03	0.18	0.17
RE	1.00	1.87	1.00	1.96	1.00	2.01	1.00	1.53	1.00	1.24	1.00	1.05
Mean.as.sd	0.10	0.07	0.18	0.12	0.49	0.26	0.11	0.08	0.18	0.12	0.45	0.28

TABLE 33: Résultats d'estimation par le maximum de vraisemblance du paramètre i_0 avec les données provenant de la phase stationnaire du processus « after » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	629.00	555.00	584.00	599.00	603.00	664.00	268.00	815.00	436.00	801.00	512.00	647.00
mean	2.99	8.57	9.60	7.92	9.05	6.17	8.69	19.88	6.35	14.75	10.03	13.05
bias	1.99	7.57	8.60	6.92	8.05	5.17	-1.31	9.88	-3.65	4.75	0.03	3.05
rel. bias	1.99	7.57	8.60	6.92	8.05	5.17	-0.13	0.99	-0.36	0.47	0.00	0.31
|bias|	2.22	8.12	9.63	7.51	9.32	5.89	9.99	19.58	11.44	15.30	14.00	10.01
sd	6.00	14.15	20.83	10.01	21.94	9.00	12.98	24.96	14.95	19.92	21.51	13.02
MSE	39.90	257.03	506.85	148.04	545.22	107.59	169.45	719.89	236.42	418.67	461.71	178.58
RE	1.00	0.16	1.00	3.42	1.00	5.07	1.00	0.24	1.00	0.56	1.00	2.59
Mean.as.sd	2.20	605.80	1.47	744.32	1.40	260.04	1.70	325.58	1.27	283.70	1.27	324.79

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	368.00	637.00	561.00	624.00	569.00	692.00	256.00	798.00	442.00	700.00	550.00	626.00
mean	7.11	8.74	7.38	6.44	6.48	5.34	8.24	21.01	5.43	16.97	18.60	13.86
bias	6.11	7.74	6.38	5.44	5.48	4.34	-1.76	11.01	-4.57	6.97	8.60	3.86
rel. bias	6.11	7.74	6.38	5.44	5.48	4.34	-0.18	1.10	-0.46	0.70	0.86	0.39
|bias|	6.92	8.48	7.46	6.14	6.48	5.01	10.32	20.86	9.69	15.92	19.62	10.06
sd	18.09	13.20	19.98	10.39	19.26	7.51	13.92	26.25	10.86	20.79	30.96	13.08
MSE	363.81	233.87	439.22	137.35	400.36	75.13	196.17	809.35	138.42	480.22	1030.95	185.67
RE	1.00	1.56	1.00	3.20	1.00	5.33	1.00	0.24	1.00	0.29	1.00	5.55
Mean.as.sd	0.42	374.34	0.29	383.75	0.29	185.04	0.34	334.19	0.25	323.87	0.26	373.56

	#7		#8		#9		#16		#17		#18	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	320.00	679.00	576.00	624.00	588.00	566.00	277.00	708.00	469.00	650.00	532.00	633.00
mean	6.13	10.30	4.59	5.00	6.40	9.25	8.01	18.96	8.34	17.64	18.15	11.96
bias	5.13	9.30	3.59	4.00	5.40	8.25	-1.99	8.96	-1.66	7.64	8.15	1.96
rel. bias	5.13	9.30	3.59	4.00	5.40	8.25	-0.20	0.90	-0.17	0.76	0.82	0.20
|bias|	5.97	10.01	4.11	4.61	6.45	8.66	10.67	18.63	7.89	15.49	17.20	9.65
sd	17.78	15.12	11.38	7.12	16.67	14.15	14.47	24.68	9.71	21.44	28.49	12.96
MSE	341.57	314.85	142.23	66.58	306.49	268.07	212.70	688.74	96.89	517.25	876.50	171.55
RE	1.00	1.08	1.00	2.14	1.00	1.14	1.00	0.31	1.00	0.19	1.00	5.11
Mean.as.sd	0.10	409.28	0.07	162.89	0.07	756.48	0.09	435.24	0.06	455.75	0.07	298.62

TABLE 34: Résultats d'estimation par le maximum de vraisemblance du paramètre R_0 avec les données provenant de la phase stationnaire du processus « after » (β-fixe)

	#1		#2		#3		#10		#11		#12	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	629.00	555.00	584.00	599.00	603.00	664.00	268.00	815.00	436.00	801.00	512.00	647.00
mean	1.27	1.30	2.00	1.96	4.09	3.95	1.32	1.32	2.01	2.00	4.10	4.02
bias	-0.06	-0.03	0.00	-0.04	0.09	-0.05	-0.01	-0.02	0.01	0.00	0.10	0.02
rel. bias	-0.05	-0.03	0.00	-0.02	0.02	-0.01	-0.01	-0.01	0.01	0.00	0.03	0.00
|bias|	0.09	0.06	0.11	0.12	0.30	0.26	0.08	0.07	0.13	0.12	0.33	0.31
sd	0.09	0.07	0.14	0.14	0.36	0.34	0.10	0.08	0.16	0.15	0.41	0.39
MSE	0.01	0.01	0.02	0.02	0.14	0.12	0.01	0.01	0.03	0.02	0.18	0.15
RE	1.00	1.89	1.00	0.97	1.00	1.17	1.00	1.48	1.00	1.26	1.00	1.19
Mean.as.sd	0.08	0.06	0.14	0.11	0.39	0.28	0.09	0.07	0.17	0.12	0.42	0.33

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	368.00	637.00	561.00	624.00	569.00	692.00	256.00	798.00	442.00	700.00	550.00	626.00
mean	1.31	1.30	2.01	1.97	4.09	3.98	1.31	1.31	2.02	2.00	4.11	4.04
bias	-0.02	-0.04	0.01	-0.03	0.09	-0.02	-0.02	-0.02	0.02	0.00	0.11	0.04
rel. bias	-0.02	-0.03	0.01	-0.01	0.02	0.00	-0.02	-0.01	0.01	0.00	0.03	0.01
|bias|	0.07	0.06	0.11	0.10	0.29	0.28	0.08	0.07	0.13	0.11	0.34	0.30
sd	0.09	0.07	0.14	0.13	0.35	0.35	0.10	0.09	0.16	0.14	0.41	0.38
MSE	0.01	0.01	0.02	0.02	0.13	0.12	0.01	0.01	0.03	0.02	0.18	0.14
RE	1.00	1.48	1.00	1.15	1.00	1.11	1.00	1.35	1.00	1.23	1.00	1.28
Mean.as.sd	0.08	0.06	0.15	0.10	0.41	0.30	0.10	0.07	0.17	0.13	0.44	0.35

	#7		#8		#9		#16		#17		#18	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	320.00	679.00	576.00	624.00	588.00	566.00	277.00	708.00	469.00	650.00	532.00	633.00
mean	1.31	1.30	2.00	1.96	4.10	3.99	1.31	1.32	2.01	2.00	4.08	4.04
bias	-0.03	-0.04	0.00	-0.04	0.10	-0.01	-0.02	-0.01	0.01	0.00	0.08	0.04
rel. bias	-0.02	-0.03	0.00	-0.02	0.02	0.00	-0.02	-0.01	0.01	0.00	0.02	0.01
|bias|	0.08	0.06	0.10	0.11	0.31	0.29	0.08	0.07	0.12	0.12	0.32	0.29
sd	0.09	0.07	0.13	0.13	0.38	0.37	0.10	0.08	0.15	0.15	0.39	0.37
MSE	0.01	0.01	0.02	0.02	0.15	0.13	0.01	0.01	0.02	0.02	0.16	0.14
RE	1.00	1.45	1.00	1.00	1.00	1.15	1.00	1.54	1.00	0.98	1.00	1.14
Mean.as.sd	0.08	0.06	0.14	0.10	0.41	0.31	0.09	0.07	0.16	0.11	0.43	0.33

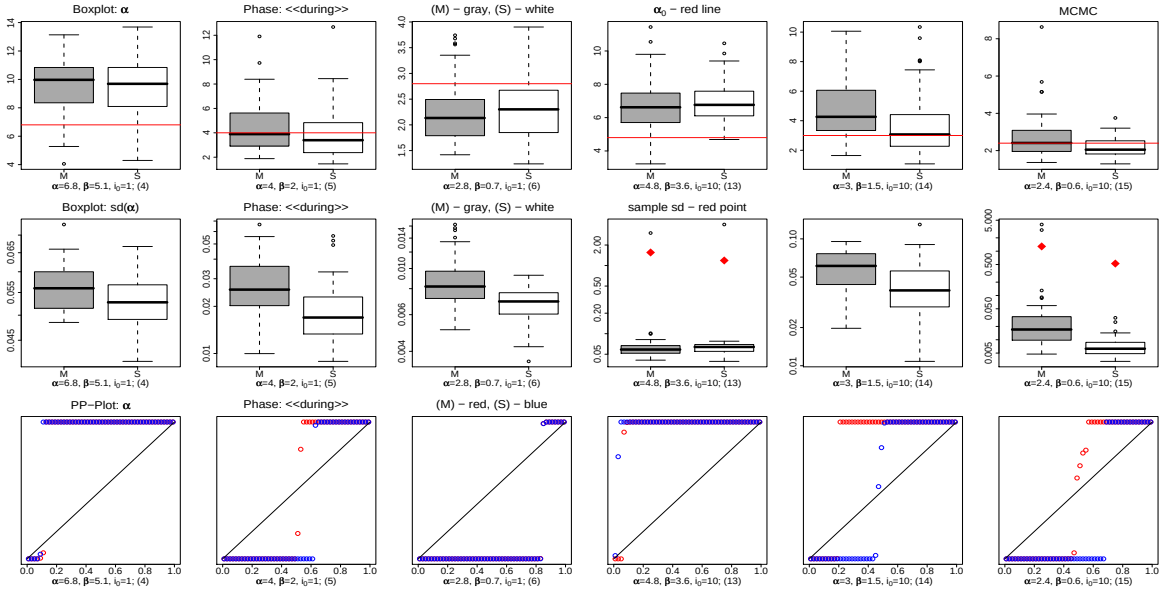

FIGURE 77: Résultats d'estimations bayésiennes par MCMC de α pour les observations de la phase évolutive : « during »

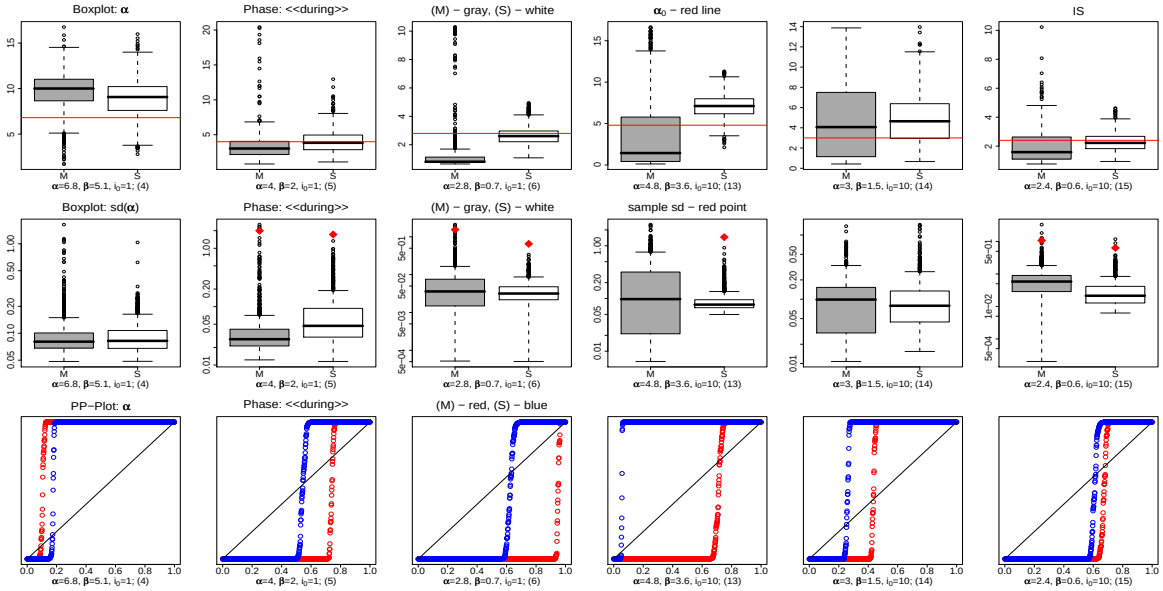

FIGURE 78: Résultats d'estimations bayésiennes IS de α pour les observations de la phase évolutive : « during »

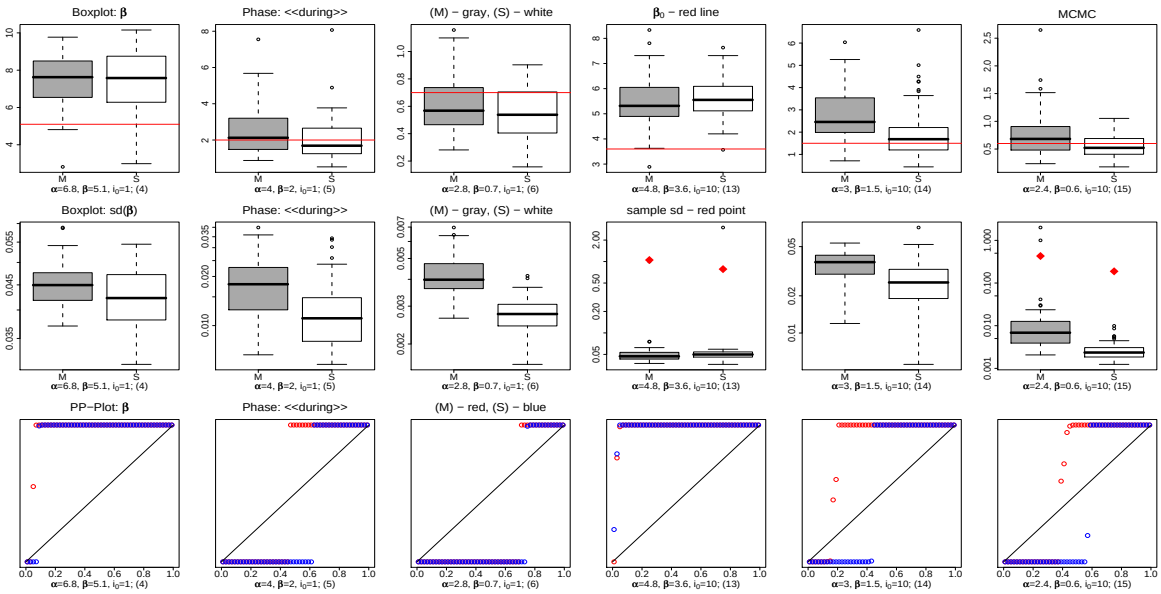

FIGURE 79: Résultats d'estimations bayésiennes par MCMC de β pour les observations de la phase évolutive : « during »

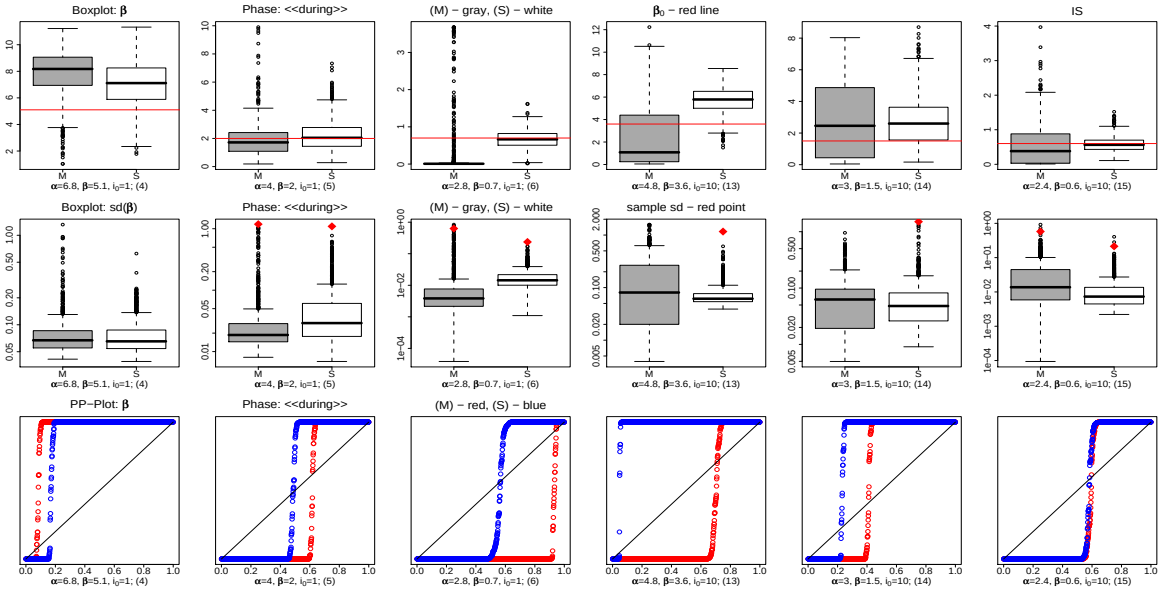

FIGURE 80: Résultats d'estimations bayésiennes IS de β pour les observations de la phase évolutive : « during »

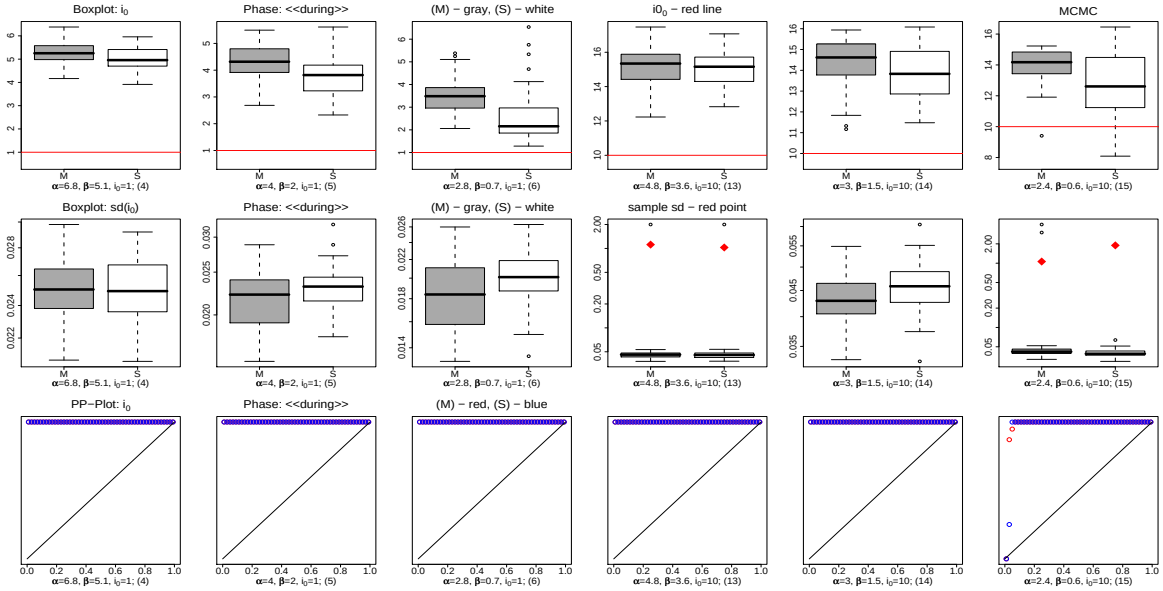

FIGURE 81: Résultats d'estimations bayésiennes par MCMC de i_0 pour les observations de la phase évolutive : « during »

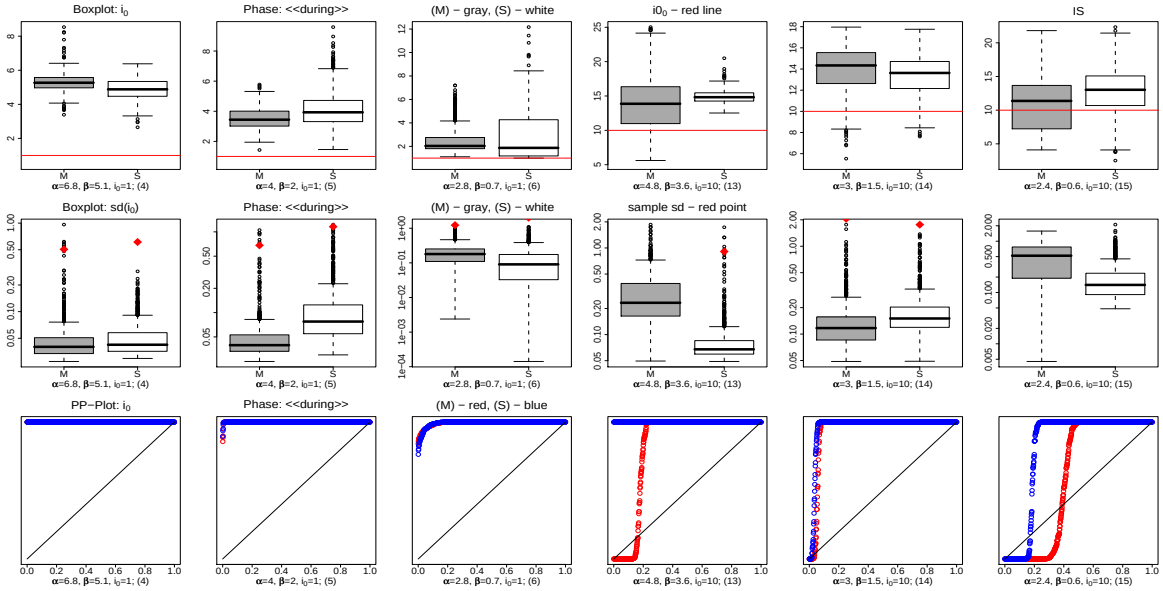

FIGURE 82: Résultats d'estimations bayésiennes IS de i_0 pour les observations de la phase évolutive : « during »

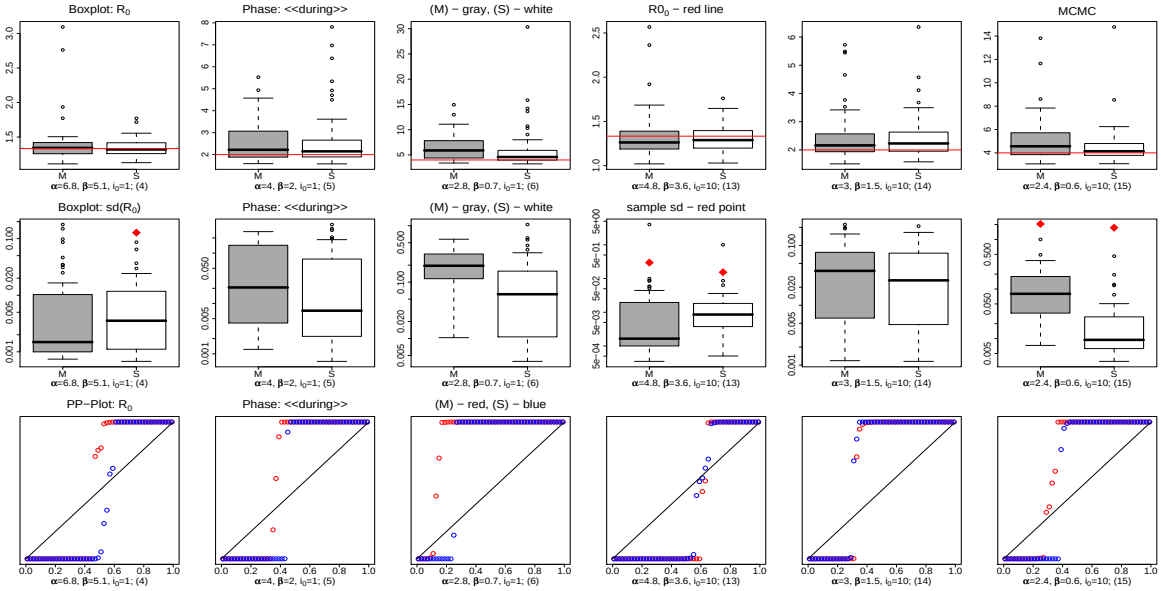

FIGURE 83: Résultats d'estimations bayésiennes par MCMC de R_0 pour les observations de la phase évolutive : « during »

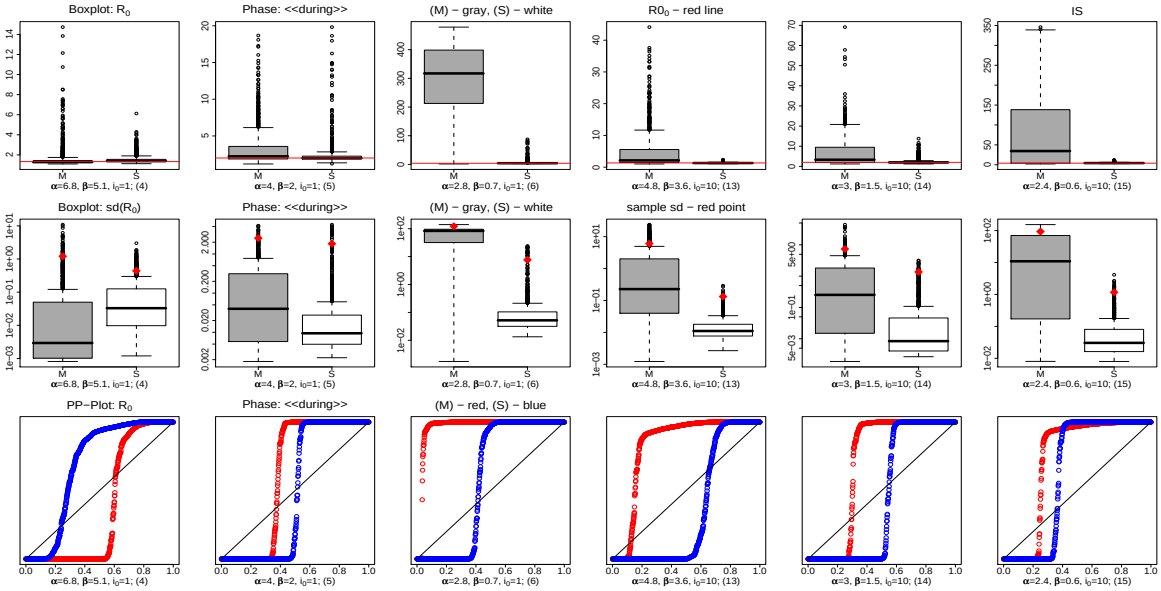

FIGURE 84: Résultats d'estimations bayésiennes IS de R_0 pour les observations de la phase évolutive : « during »

TABLE 35: Résultats d'estimation bayésienne du paramètre α avec les données provenant de la phase évolutive du processus « during »

MCMC

	#4		#5		#6		#13		#14		#15	
α_0	6.8		4		2.8		4.8		3		2.4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
mean	9.54	9.56	4.31	3.81	2.25	2.27	6.71	6.95	4.96	3.75	2.72	2.18
bias	2.74	2.76	0.31	-0.19	-0.55	-0.53	1.91	2.15	1.96	0.75	0.32	-0.22
rel. bias	0.40	0.41	0.08	-0.05	-0.20	-0.19	0.40	0.45	0.65	0.25	0.13	-0.09
|bias|	2.97	2.94	1.50	1.50	0.71	0.62	2.04	2.15	2.25	1.57	0.79	0.47
sd	1.97	2.03	2.05	2.02	0.59	0.54	1.56	1.19	2.32	2.25	1.28	0.53
MSE	11.29	11.64	4.23	4.02	0.64	0.56	6.03	5.99	9.10	5.55	1.71	0.32
RE	1.00	0.97	1.00	1.05	1.00	1.13	1.00	1.01	1.00	1.64	1.00	5.36
Mean.as.sd	0.06	0.05	0.03	0.02	0.01	0.01	0.12	0.14	0.06	0.04	0.16	0.01

IS

	#4		#5		#6		#13		#14		#15	
α_0	6.8		4		2.8		4.8		3		2.4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00
mean	9.67	8.86	3.33	4.06	1.26	2.66	3.54	7.12	4.65	4.87	1.96	2.28
bias	2.87	2.06	-0.67	0.06	-1.54	-0.14	-1.26	2.32	1.65	1.87	-0.44	-0.12
relative bias	0.42	0.30	-0.17	0.02	-0.55	-0.05	-0.26	0.48	0.55	0.62	-0.18	-0.05
|bias|	3.17	2.52	1.46	1.33	1.98	0.52	3.83	2.41	3.16	2.38	0.97	0.53
sd	2.09	2.14	1.99	1.72	1.55	0.65	4.16	1.46	3.54	2.48	1.08	0.64
MSE	12.59	8.80	4.42	2.96	4.76	0.44	18.86	7.51	15.24	9.62	1.35	0.42
RE	1.00	1.43	1.00	1.49	1.00	10.70	1.00	2.51	1.00	1.59	1.00	3.24
Mean.as.sd	0.10	0.10	0.08	0.10	0.09	0.04	0.24	0.10	0.12	0.12	0.09	0.04

TABLE 36: Résultats d'estimation bayésienne du paramètre β avec les données provenant de la phase évolutive du processus « during »

MCMC

	#4		#5		#6		#13		#14		#15	
α_0	5.1		2		0.7		3.6		1.5		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
mean	7.43	7.45	2.40	2.03	0.62	0.55	5.41	5.60	2.75	2.01	0.77	0.55
bias	2.33	2.35	0.40	0.03	-0.08	-0.15	1.81	2.00	1.25	0.51	0.17	-0.05
rel. bias	0.46	0.46	0.20	0.01	-0.12	-0.22	0.50	0.56	0.83	0.34	0.28	-0.08
|bias|	2.43	2.48	0.92	0.92	0.18	0.20	1.83	2.00	1.39	0.92	0.29	0.16
sd	1.46	1.61	1.29	1.30	0.20	0.19	1.05	0.78	1.29	1.30	0.43	0.19
MSE	7.51	8.04	1.79	1.67	0.05	0.06	4.34	4.60	3.19	1.91	0.21	0.04
RE	1.00	0.93	1.00	1.07	1.00	0.79	1.00	0.94	1.00	1.67	1.00	5.62
Mean.as.sd	0.05	0.04	0.02	0.01	>0.01	>0.01	0.07	0.11	0.04	0.03	0.07	>0.01

IS

	#4		#5		#6		#13		#14		#15	
α_0	5.1		2		0.7		3.6		1.5		0.6	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00
mean	7.80	6.91	1.88	2.21	0.18	0.67	2.61	5.72	2.78	2.74	0.54	0.58
bias	2.70	1.81	-0.12	0.21	-0.52	-0.03	-0.99	2.12	1.28	1.24	-0.06	-0.02
relative bias	0.53	0.35	-0.06	0.10	-0.75	-0.05	-0.28	0.59	0.85	0.83	-0.10	-0.04
|bias|	2.91	2.20	0.85	0.84	0.73	0.19	2.85	2.18	2.12	1.50	0.48	0.17
sd	1.74	1.78	1.19	1.09	0.63	0.24	3.00	1.17	2.29	1.51	0.58	0.21
MSE	10.34	6.43	1.42	1.23	0.67	0.06	9.99	5.84	6.86	3.80	0.34	0.05
RE	1.00	1.61	1.00	1.15	1.00	11.74	1.00	1.71	1.00	1.81	1.00	7.29
Mean.as.sd	0.09	0.08	0.05	0.06	0.03	0.02	0.18	0.08	0.08	0.08	0.04	0.01

TABLE 37: Résultats d'estimation bayésienne du paramètre i_0 avec les données provenant de la phase évolutive du processus « during »

MCMC

	#4		#5		#6		#13		#14		#15	
$i0_0$	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
mean	5.25	4.99	4.25	3.77	3.56	2.54	15.18	15.06	14.30	13.82	13.99	12.86
bias	4.25	3.99	3.25	2.77	2.56	1.54	5.18	5.06	4.30	3.82	3.99	2.86
rel. bias	4.25	3.99	3.25	2.77	2.56	1.54	0.52	0.51	0.43	0.38	0.40	0.29
|bias|	4.25	3.99	3.25	2.77	2.56	1.54	5.18	5.06	4.30	3.82	4.02	2.94
sd	0.40	0.50	0.68	0.75	0.78	1.14	1.12	1.03	1.23	1.30	1.06	1.89
MSE	18.24	16.21	11.04	8.22	7.12	3.62	28.04	26.68	19.98	16.28	17.05	11.69
RE	1.00	1.13	1.00	1.34	1.00	1.97	1.00	1.05	1.00	1.23	1.00	1.46
Mean.as.sd	0.03	0.03	0.02	0.02	0.02	0.02	0.08	0.08	0.04	0.05	0.18	0.04

IS

	#4		#5		#6		#13		#14		#15	
ß0_0	1		1		1		10		10		10	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950	950	950	950	950	950	950	950	950	950	950	950
mean	5.26	4.88	3.54	4.12	2.52	2.83	14.08	14.89	13.88	13.41	10.92	12.86
bias	4.26	3.88	2.54	3.12	1.52	1.83	4.08	4.89	3.88	3.41	0.92	2.86
rel. bias	4.26	3.88	2.54	3.12	1.52	1.83	0.41	0.49	0.39	0.34	0.09	0.29
|bias|	4.26	3.88	2.54	3.12	1.52	1.83	4.74	4.89	4.00	3.48	3.39	3.60
sd	0.51	0.61	0.68	1.16	1.23	2.01	4.34	0.91	2.06	1.76	3.87	3.27
MSE	18.37	15.46	6.89	11.09	3.82	7.39	35.48	24.69	19.32	14.76	15.83	18.89
RE	1.00	1.19	1.00	0.62	1.00	0.52	1.00	1.44	1.00	1.31	1.00	0.84
Mean.as.sd	0.05	0.05	0.06	0.13	0.22	0.13	0.31	0.09	0.15	0.19	0.51	0.20

TABLE 38: Résultats d'estimation bayésienne du paramètre R_0 avec les données provenant de la phase évolutive du processus « during »

MCMC

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
mean	1.42	1.35	2.61	2.67	6.53	6.13	1.35	1.30	2.49	2.45	5.20	4.59
bias	0.08	0.01	0.61	0.67	2.53	2.13	0.02	-0.03	0.49	0.45	1.20	0.59
rel. bias	0.06	0.01	0.31	0.33	0.63	0.53	0.01	-0.02	0.24	0.23	0.30	0.15
|bias|	0.16	0.10	0.73	0.81	2.60	2.28	0.19	0.12	0.60	0.54	1.41	0.84
sd	0.34	0.13	0.97	1.41	2.65	4.48	0.30	0.15	0.98	0.84	2.05	1.73
MSE	0.12	0.02	1.29	2.39	13.30	24.18	0.09	0.02	1.17	0.90	5.56	3.27
RE	1.00	7.54	1.00	0.54	1.00	0.55	1.00	3.61	1.00	1.30	1.00	1.70
Mean.as.sd	0.02	0.01	0.07	0.06	0.22	0.13	0.09	0.03	0.05	0.04	0.17	0.03

IS

	#4		#5		#6		#13		#14		#15	
$R0_0$	1.33		2		4		1.33		2		4	
modèle	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)	(M)	(S)
# sims	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00	950.00
mean	1.64	1.54	3.29	2.43	288.46	5.56	4.88	1.31	6.79	2.35	80.16	4.51
bias	0.31	0.21	1.29	0.43	284.46	1.56	3.55	-0.02	4.79	0.35	76.16	0.51
rel. bias	0.23	0.16	0.65	0.22	71.12	0.39	2.66	-0.02	2.39	0.17	19.04	0.13
|bias|	0.42	0.24	1.46	0.61	284.55	1.95	3.59	0.10	4.96	0.59	76.69	0.76
sd	1.21	0.44	2.53	1.82	122.55	7.67	6.08	0.13	7.36	1.36	93.22	1.17
MSE	1.56	0.24	8.05	3.51	95919.13	61.14	49.57	0.02	77.06	1.97	14481.47	1.63
RE	1.00	6.59	1.00	2.30	1.00	1568.81	1.00	2737.68	1.00	39.09	1.00	8887.21
Mean.as.sd	0.21	0.13	0.30	0.13	67.81	0.56	1.96	0.02	1.52	0.12	35.45	0.12

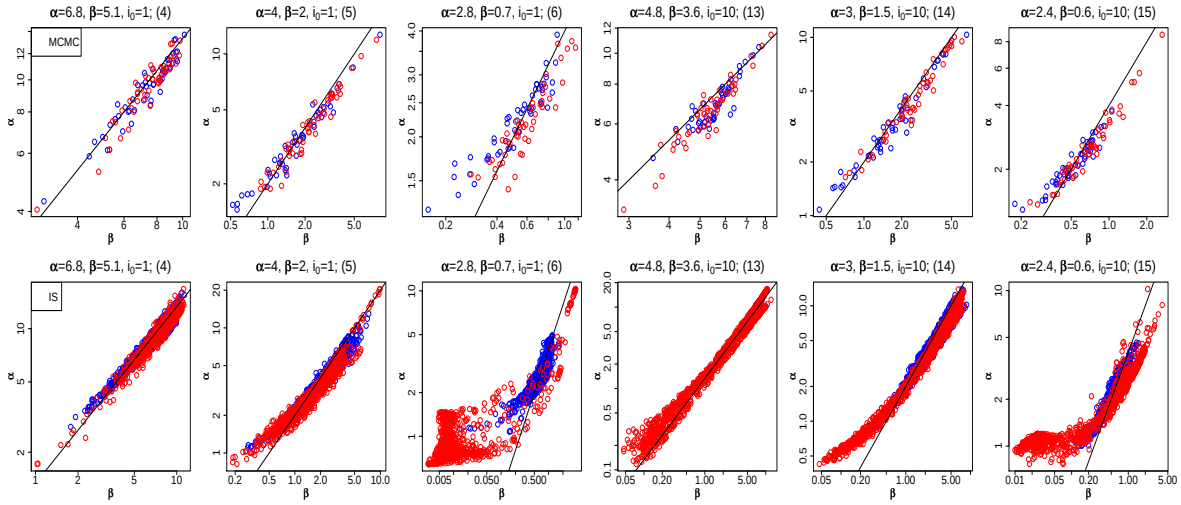

FIGURE 85: Les estimations bayésiennes de α en fonction des estimations de β par rapport à la vraie droite $\alpha = R_0\beta$ pour les observations de la phase évolutive : « during » ; modèle (M) - points rouges, modèle (S) - points bleus

Démonstration algébrique des propriétés des estimateurs de R_0

La vraisemblance du modèle (M^s) est

$$\begin{aligned} V_{R_0}(X) &= \frac{1}{\sqrt{2\pi N/R_0}} \exp\{-\frac{1}{2N/R_0}\sum_{i=1}^{J}(x_i - N(1-\frac{1}{R_0}))^2\} \quad (11)\\ &= (2\pi N/R_0)^{-\frac{J}{2}} \exp\{-\frac{R_0}{2N}\left(\sum_{i=1}^{J}x_i^2 - 2N\frac{R_0-1}{R_0}\sum_{i=1}^{J}x_i + JN^2(\frac{R_0-1}{R_0})^2)\right)\}\\ &= (2\pi N/R_0)^{-\frac{J}{2}} \exp\{-\frac{R_0}{2N}\sum_{i=1}^{J}x_i^2 + (R_0-1)\sum_{i=1}^{J}x_i - \frac{JN(R_0^2-2R_0+1)}{2R_0}\} \end{aligned}$$

La log-vraisemblance est donc

$$l = \ln V = -\frac{J}{2}\ln 2\pi N + \frac{J}{2}\ln R_0 - \frac{R_0}{2N}\sum_{i=1}^{J}x_i^2 + (R_0-1)\sum_{i=1}^{J}x_i - \frac{JN}{2}(R_0 - 2 + \frac{1}{R_0}) \quad (12)$$

La dérivée de la log-vraisemblance est

$$\frac{\partial l}{\partial R_0} = \frac{J}{2}\frac{1}{R_0} - \frac{1}{2N}\sum_{i=1}^{J}x_i^2 + \sum_{i=1}^{J}x_i - \frac{JN}{2} + \frac{1}{R_0^2}\frac{JN}{2} \quad (13)$$

En annulant la dérivée de log-vraisemblance, $\frac{\partial l}{\partial R_0} = 0$, on obtient

$$\begin{aligned} JR_0 - \frac{R_0^2}{N}\sum_{i=1}^{J}x_i^2 + 2R_0^2\sum_{i=1}^{J}x_i - JNR_0^2 + JN &= 0 \quad (14)\\ -R_0^2\left(\frac{1}{N}\sum_{i=1}^{J}x_i^2 - 2\sum_{i=1}^{J}x_i + JN\right) + JR_0 + JN &= 0\\ -R_0^2\frac{1}{N}\left(\sum_{i=1}^{J}x_i^2 - \sum_{i=1}^{J}2Nx_i + \sum_{i=1}^{J}N^2\right) + JR_0 + JN &= 0\\ -R_0^2\frac{1}{N}\sum_{i=1}^{J}(x_i - N)^2 + JR_0 + JN &= 0 \quad (15) \end{aligned}$$

qui est une équation du second degré par rapport à R_0 avec un coefficient d'ordre 2 négatif $(-\frac{1}{N}\sum_{i=1}^{J}(x_i-N)^2 < 0)$. Le discriminant de cette équation est $\Delta = J^2\left(1 + 4\frac{1}{J}\sum_{i=1}^{J}(x_i-N)^2\right)$. L'équation (15) admet deux racines réelles mais dont une seule est toujours positive :

$$R_0 = NJ\frac{1 + \sqrt{1 + 4\frac{1}{J}\sum_{i=1}^{J}(X_i-N)^2}}{2\sum_{i=1}^{J}(X_i-N)^2} \quad (16)$$

La dérivée seconde de la log-vraisemblance est

$$\frac{\partial^2 l}{\partial R_0^2} = -\frac{J}{2R_0^2} - \frac{JN}{R_0^3} \quad (17)$$

et l'information de Fisher est donc

$$I(R_0) = -E(\frac{\partial^2 l}{\partial R_0^2}|R_0) = \frac{J}{2R_0^2} + \frac{JN}{R_0^3} \tag{18}$$

La borne inférieure de Cramer-Rao est ainsi

$$B(R_0) = \frac{1}{I(R_0)} = \frac{2R_0^3}{J(R_0 + 2N)} \tag{19}$$

On va montrer que l'estimateur $\widehat{R_0}$ (7.7) est asymptotiquement sans biais et que sa variance asymptotique est égale à la borne de Cramer-Rao.

Pour fixer les notations, on rappelle que la somme de carrées des ν variables normales indépendantes $Z_i \sim \mathcal{N}(\omega_i, 1)$, $i = 1, \ldots, \nu$ suit une loi khi-2 non-centrée à ν degrés de liberté et avec un paramètre de non-centralité $\lambda = \sum_{i=1}^{\nu} \omega_i^2$ (Johnson et al., 1995). On notera donc que $\sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2(\lambda)$. L'espérance et la variance de la loi khi-2 non-centrée sont $E\left[\chi_\nu^2(\lambda)\right] = \nu + \lambda$ et $Var\left[\chi_\nu^2(\lambda)\right] = 2(\nu + 2\lambda)$.

Pour démontrer les propriétés asymptotiques de l'estimateur, on utilisera la delta méthode (Casella and Berger, 2002) en déterminant les moments d'ordre un et deux de $\sum_{i=1}^{J} \frac{(X_i - N)^2}{N/R_0}$. Puisque la loi de X_i est donnée par (7.5) et grâce aux propriétés de la loi normale, on peut effectuer les changements de variables pour tout $i = 1, \ldots, n$

$$X_i \sim \mathcal{N}(N - \frac{N}{R_0}, \frac{N}{R_0}) \quad \Rightarrow \quad X_i - N \sim \mathcal{N}(-\frac{N}{R_0}, \frac{N}{R_0}) \quad \Rightarrow \quad \frac{X_i - N}{\sqrt{\frac{N}{R_0}}} \sim \mathcal{N}(-\frac{\frac{N}{R_0}}{\sqrt{\frac{N}{R_0}}}, 1) \tag{20}$$

Posons $Y = \sum_{i=1}^{J} \frac{(X_i - N)^2}{N/R_0}$. La nouvelle variable Y suit donc une loi de khi-2 non-centrée à J degrés de liberté avec un paramètre de non-centralité $\lambda = \frac{JN}{R_0}$ (car $\sum_{i=1}^{J} \left(-\frac{N}{R_0}/\sqrt{\frac{N}{R_0}}\right)^2 = \sum_{i=1}^{J} \frac{N}{R_0} = \frac{JN}{R_0}$) et donc une espérance $E(Y) = J(1 + \frac{N}{R_0})$ et une variance $Var(Y) = 2J(1 + 2\frac{N}{R_0})$. L'estimateur $\widehat{R_0}$ devient, en fonction de Y :

$$\widehat{R_0} = NJ \frac{1 + \sqrt{1 + \frac{4}{J} Y \frac{N}{R_0}}}{2\frac{N}{R_0} Y} \tag{21}$$

Posons que $\widehat{R_0}$ correspond à une fonction $g(Y)$ différentiable et que l'espérance de Y définie ci-dessus est $\theta = J(1 + \frac{N}{R_0})$. Dans ce cas, on peut écrire que

$$E(g(Y)) \approx g(\theta) + g'(\theta) E(Y - \theta) = g(\theta) \tag{22}$$

$$Var(g(Y)) \approx [g'(\theta)]^2 Var(Y) \tag{23}$$

où $g'(\theta) = \frac{\partial}{\partial y} g(y)|_{y=\theta}$. On obtient $g(\theta)$ en remplaçant dans (21) Y par $\theta = J(1 + \frac{N}{R_0})$.

$$\begin{aligned} g(\theta) &= NJ \frac{1 + \sqrt{1 + \frac{4N}{JR_0} J(1 + \frac{N}{R_0})}}{2\frac{N}{R_0} J(1 + \frac{N}{R_0})} = \frac{1 + \frac{1}{R_0}\sqrt{R_0^2 + 4N(R_0 + N)}}{\frac{2}{R_0^2}(R_0 + N)} = \\ &= \frac{R_0 + \sqrt{(R_0 + 2N)^2}}{\frac{2}{R_0}(R_0 + N)} = \frac{R_0 + R_0 + 2N}{\frac{2}{R_0}(R_0 + N)} = \frac{2(R_0 + N)}{\frac{2}{R_0}(R_0 + N)} = R_0 \end{aligned} \tag{24}$$

On obtient donc que $E(g(Y)) = E(\widehat{R_0}) \overset{as}{\approx} R_0$, $\widehat{R_0}$ est asymptotiquement sans biais.

La dérivée de $g(Y)$ est

$$g'(Y) = \frac{JR_0}{2Y^2} \left(\frac{\frac{2N}{JR_0} Y}{\sqrt{1 + \frac{4N}{JR_0} Y}} - \sqrt{1 + \frac{4N}{JR_0} Y} - 1 \right) \tag{25}$$

En remplaçant Y par θ $g'(\theta)$:

$$\begin{aligned} g'(\theta) &= \frac{JR_0}{2J^2(1 + \frac{N}{R_0})^2} \left(\frac{\frac{2N}{JR_0} J(1 + \frac{N}{R_0})}{\sqrt{1 + \frac{4N}{JR_0} J(1 + \frac{N}{R_0})}} - \sqrt{1 + \frac{4N}{JR_0} J(1 + \frac{N}{R_0})} - 1 \right) = \\ &= \frac{R_0}{2J(R_0 + N)^2 \frac{1}{R_0^2}} \left(\frac{\frac{2N}{R_0^2}(R_0 + N)}{\frac{1}{R_0}\sqrt{R_0^2 + 4N(R_0 + N)}} - \frac{1}{R_0}\sqrt{R_0^2 + 4N(R_0 + N)} - 1 \right) \\ &= \frac{R_0^3}{2J(R_0 + N)^2} \left(\frac{\frac{2N}{R_0}(R_0 + N)}{\sqrt{(R_0 + 2N)^2}} - \frac{1}{R_0}\sqrt{(R_0 + 2N)^2} - 1 \right) \\ &= \frac{R_0^2}{2J(R_0 + N)^2} \left(\frac{2N(R_0 + N)}{R_0 + 2N} - (R_0 + 2N) - R_0 \right) \\ &= \frac{R_0^2}{2J(R_0 + N)^2} \left(\frac{2N(R_0 + N) - (R_0 + 2N)^2 - (R_0 + 2N)R_0}{R_0 + 2N} \right) \\ &= \frac{R_0^2}{2J(R_0 + N)^2} \left(\frac{-2R_0^2 - 4NR_0 - 2N^2}{R_0 + 2N} \right) \\ &= \frac{R_0^2}{2J(R_0 + N)^2} \frac{(-2)(R_0 + N)^2}{R_0 + 2N} = \frac{-R_0^2}{J(R_0 + 2N)} \end{aligned} \tag{26, 27}$$

D'où le résultat de la variance asymptotique selon la formule (23)

$$Var(g(Y)) = Var(\widehat{R_0}) \overset{as}{\approx} \frac{R_0^4}{J^2(R_0 + 2N)^2} 2J(1 + 2\frac{N}{R_0}) = \frac{2R_0^3}{J(R_0 + 2N)} \tag{28}$$

Cette variance correspond à la borne inférieure de Cramer-Rao (19), donc l'estimateur $\widehat{R_0}$ est asymptotiquement efficace.

On va maintenant démontrer que l'estimateur du maximum de vraisemblance de R_0 sous le modèle (S^s) (7.6) a la forme (7.9) avec la variance (7.10) et que l'estimateur de σ^2 est (7.11).

La vraisemblance du modèle (S^s) est

$$\begin{aligned} V_{R_0}(X) &= (2\pi\sigma^2)^{-\frac{J}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{J}(x_i - N(1-\frac{1}{R_0}))^2\} \\ &= (2\pi\sigma^2)^{-\frac{J}{2}} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{J}x_i^2 - 2N(1-\frac{1}{R_0})\sum_{i=1}^{J}x_i + JN^2(1-\frac{1}{R_0})^2)\} \\ &= (2\pi\sigma^2)^{-\frac{J}{2}} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{J}x_i^2 - 2N(1-\frac{1}{R_0})\sum_{i=1}^{J}x_i + JN^2(1-2\frac{1}{R_0}+\frac{1}{R_0^2}))\} \end{aligned} \tag{29}$$

D'où la log-vraisemblance :

$$l = \ln V = -\frac{J}{2}\ln 2\pi - \frac{J}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{J}x_i^2 - 2\sum_{i=1}^{J}x_i + \frac{2N}{R_0}\sum_{i=1}^{J}x_i + JN^2 - \frac{2JN^2}{R_0} + \frac{JN^2}{R_0^2}\right) \tag{30}$$

La dérivée de log-vraisemblance par rapport à R_0 est

$$\frac{\partial l}{\partial R_0} = -\frac{1}{2\sigma^2}\left(-\frac{2N}{R_0^2}\sum_{i=1}^{J}x_i + \frac{2JN^2}{R_0^2} - 2\frac{JN^2}{R_0^3}\right) \tag{31}$$

En comparant la dérivée de log-vraisemblance à zéro $\frac{\partial l}{\partial R_0} = 0$, on obtient les équations équivalentes

$$\begin{aligned} -\frac{2N}{R_0^2}\left(\sum_{i=1}^{J}x_i - JN + \frac{JN}{R_0}\right) &= 0 \\ \sum_{i=1}^{J}x_i - JN + \frac{JN}{R_0} &= 0 \\ -\sum_{i=1}^{J}x_i + JN &= \frac{JN}{R_0} \end{aligned} \tag{32}$$

D'où l'estimateur du maximum de vraisemblance suivant

$$R_0 = \frac{JN}{JN - \sum_{i=i}^{J}x_i} = \frac{N}{N-\bar{X}} \tag{33}$$

La dérivée de la log-vraisemblance par rapport à σ^2 est

$$\frac{\partial l}{\partial \sigma^2} = -\frac{J}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\sum_{i=1}^{J}x_i^2 - 2N(1-\frac{1}{R_0})\sum_{i=1}^{J}x_i + JN^2(1-\frac{1}{R_0})^2\right) \tag{34}$$

La dérivée de la log-vraisemblance est ensuite comparée à zéro et en multipliant par σ^4 on obtient

$$-n\sigma^2 + \left(\sum_{i=1}^{J} x_i^2 - 2N(1 - \frac{1}{R_0}\sum_{i=1}^{J} x_i + JN^2(1 - \frac{1}{R_0})^2\right) = 0 \tag{35}$$

$$\sigma^2 = \frac{1}{J}\left(\sum_{i=1}^{J} x_i^2 - 2N\frac{1-R_0}{R0}\sum_{i=1}^{J} x_i + JN^2(\frac{1-R_0}{R_0})^2\right) \tag{36}$$

En remplaçant R_0 par son estimateur dans l'expression ci-dessus, on obtient

$$\sigma^2 = \frac{1}{J}\left(\sum_{i=1}^{J} x_i^2 - 2\bar{X}\sum_{i=1}^{J} x_i + J\bar{X}^2\right) = \frac{1}{J}\left(\sum_{i=1}^{J} x_i^2 - J\bar{X}^2\right) \tag{37}$$

d'où l'estimateur du maximum de vraisemblance :

$$\sigma^2 = \frac{1}{J}\sum_{i=1}^{J} x_i^2 - \bar{X}^2 \tag{38}$$

Les dérivées secondes de la log-vraisemblance pas rapport à R_0 et σ^2 sont les suivantes

$$\frac{\partial^2 l}{\partial R_0^2} = -\frac{1}{\sigma^2}\left(\frac{2N}{R_0^3}\sum_{i=1}^{J} x_i - \frac{2JN^2}{R_0^3} + 3\frac{JN^2}{R_0^4}\right) \tag{39}$$

$$\frac{\partial^2 l}{\partial R_0 \partial \sigma^2} = -\frac{1}{\sigma^4}\left(\frac{N}{R_0^2}\sum_{i=1}^{J} x_i - \frac{JN^2}{R_0^2} + \frac{JN^2}{R_0^3}\right) \tag{40}$$

$$\frac{\partial^2 l}{\partial \sigma^4} = \frac{J}{2\sigma^4} - \frac{1}{\sigma^6}\left(\sum_{i=1}^{J} x_i^2 - 2N(1 - \frac{1}{R_0})\sum_{i=1}^{J} x_i + JN^2(1 - \frac{1}{R_0})^2\right) \tag{41}$$

Pour calculer la matrice d'information de Fisher, on calcule d'abord l'espérance de $E(\sum_{i=1}^{J} X_i)$ et $E(\sum_{i=1}^{J} X_i^2)$. Puisque les variables X_i sont supposées indépendantes et suivant la même loi normale (7.6) alors on obtient

$$E(\sum_{i=1}^{J} X_i) = \sum_{i=1}^{J} E(X_i) = JN(1 - \frac{1}{R_0}) \tag{42}$$

$$E(\sum_{i=1}^{J} X_i^2) = \sum_{i=1}^{J} E(X_i^2) = \sum_{i=1}^{J} Var(X_i) + [E(X_i)]^2 = J\sigma^2 + JN^2(1 - \frac{1}{R_0})^2 \tag{43}$$

d'où le calcul

$$
\begin{aligned}
-E\left(\frac{\partial^2 l}{\partial R_0^2}\right) &= \frac{1}{\sigma^2}\left(\frac{2N}{R_0^3}JN(1-\frac{1}{R_0}) - \frac{JN^2}{R_0^3} + 3\frac{JN^2}{R_0^4}\right) = \frac{JN^2}{\sigma^2 R_0^4} \qquad (44)\\
-E\left(\frac{\partial^2 l}{\partial R_0 \partial \sigma^2}\right) &= \frac{1}{\sigma^4}\left(\frac{N}{R_0^2}JN(1-\frac{1}{R_0}) - \frac{JN^2}{R_0^2} + \frac{JN^2}{R_0^3}\right) = 0 \qquad (45)\\
-E\left(\frac{\partial^2 l}{\partial \sigma^4}\right) &= -\left(\frac{J}{2\sigma^4} - \frac{1}{\sigma^6}\left(J\sigma^2 + JN^2(1-\frac{1}{R_0})^2 - 2JN^2(1-\frac{1}{R_0})^2 + JN^2(1-\frac{1}{R_0})^2\right)\right)\\
&= \frac{J}{2\sigma^4} \qquad (46)
\end{aligned}
$$

Alors la matrice d'information de Fisher et son inverse sont

$$
I(R_0,\sigma^2) = \begin{pmatrix} \frac{JN^2}{\sigma^2 R_0^4} & 0 \\ 0 & \frac{J}{2\sigma^2} \end{pmatrix}, \qquad I^{-1}(R_0,\sigma^2) = \begin{pmatrix} \frac{\sigma^2 R_0^4}{JN^2} & 0 \\ 0 & \frac{2\sigma^2}{J} \end{pmatrix}
$$

Les bornes de Cramer-Rao pour les variances asymptotiques de $\widetilde{R_0}$ et $\widetilde{\sigma^2}$ sont

$$
Var(\widetilde{R_0}) \geq \frac{\sigma^2 R_0^4}{JN^2} \qquad (47)
$$

$$
Var(\widetilde{\sigma^2}) \geq \frac{2\sigma^4}{J} \qquad (48)
$$

Pour démontrer que l'estimateur $\widetilde{R_0}$ (7.9) est sans biais et efficace, la delta méthode est utilisée comme dans les équations (22) et (23). Posons cette fois-ci que $Y = \bar{X}$ dans l'expression de $\widetilde{R_0}$. La fonction $g(Y)$ est donc $g(Y) = \frac{N}{N-Y}$. L'espérance et la variance de Y sont $E(Y) = N(1-\frac{1}{R_0})$ et $Var(Y) = \frac{\sigma^2}{J}$, donc $\theta = N(1-\frac{1}{R_0})$. $g(\theta)$ est donc

$$
g(\theta) = \frac{N}{N - N(1-\frac{1}{R_0})} = R_0 \qquad (49)
$$

$g'(Y)$ et $g'(\theta)$ sont

$$
g'(Y) = \frac{N}{(N-Y)^2}, \quad g'(\theta) = \frac{R_0^2}{N} \qquad (50)
$$

On obtient que $E(g(Y)) = E(g(\theta)) = g(\theta)$. L'estimateur $\widetilde{R_0}$ est sans biais car $E(\widetilde{R_0}) = R_0$. La variance asymptotique est déduite de l'équation (23)

$$
Var(\widetilde{R_0}) = \frac{R_0^4 \sigma^2}{JN^2} \qquad (51)
$$

qui correspond à la borne dans (47), $\widetilde{R_0}$ est donc asymptotiquement efficace.

La variance de $\widetilde{\sigma^2}$ se calcule directement car $\widetilde{\sigma^2} = \frac{1}{J}\sum_{i=1}^{J} X_i^2 - \bar{X}^2 = \frac{1}{J}\sum_{i=1}^{J}(X_i - \bar{X})^2$. En effet, effectuons les transformations suivantes sur $\widetilde{\sigma^2}$

$$\begin{aligned}\widetilde{\sigma^2} &= \frac{1}{J}\sum_{i=1}^{J}(X_i - \bar{X})^2 = \frac{1}{J}\sum_{i=1}^{J}\left(X_i - N(1-\frac{1}{R_0}) - \bar{X} + N(1-\frac{1}{R_0})\right)^2 \\ &= \sigma^2\frac{1}{J}\sum_{i=1}^{J}\left(\frac{X_i - N(1-\frac{1}{R_0})}{\sigma} - \frac{\bar{X} - N(1-\frac{1}{R_0})}{\sigma}\right)^2 = \sigma^2\frac{1}{J}\sum_{i=1}^{J}(Z_i - \bar{Z}) \qquad (52)\end{aligned}$$

où $Z_i = \frac{X_i - N(1-\frac{1}{R_0})}{\sigma}$ est une nouvelle variable qui suit une loi normale centrée réduite. La somme $\sum_{i=1}^{J}(Z_i - \bar{Z})^2$ suit donc une loi du chi-2 (centrée) à $n-1$ degrés de liberté avec une espérance $J-1$ et une variance $2(J-1)$. On trouve donc aisément que l'espérance et la variance de $\widetilde{\sigma^2}$ valent

$$E(\widetilde{\sigma^2}) = \frac{J-1}{J}\sigma^2, \quad \text{et} \quad Var(\widetilde{\sigma^2}) = \frac{2\sigma^4(J-1)}{J^2} \qquad (53)$$

On constate donc que l'estimateur $\widetilde{\sigma^2}$ a un biais, cependant nous pouvons en déduire un estimateur de σ^2 sans biais qui est le suivant $\widetilde{\delta^2} = \frac{J}{J-1}\widetilde{\sigma^2}$. Sa variance est $Var(\widetilde{\delta^2}) = \frac{2\sigma^4}{J-1}$ alors $\widetilde{\delta^2}$ n'est plus efficace car $\frac{2\sigma^4}{J-1} > \frac{2\sigma^4}{J}$.

Articles