

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATION DE PARAMÈTRES DANS DES MODÈLES D'ÉPIDÉMIES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

HUGUES LEDUC

AVRIL 2011

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je tiens d'abord à remercier ma directrice de recherche, Sorana Froda, pour sa passion, sa disponibilité et la confiance qu'elle m'a témoignée tout au long de mes études à la maîtrise en statistique à l'UQAM. Je remercie également Jean-François Boudreau, chercheur à l'Institut de recherche d'Hydro-Québec, avec qui j'ai fait un stage pendant ma maîtrise.

Je souhaite ensuite remercier le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour le soutien financier qu'il m'a accordé et grâce auquel j'ai pu me consacrer à temps plein à mes études pendant ma maîtrise. Je remercie également la Fondation de l'UQAM et Yves Bélair pour le soutien financier offert aux étudiants en situation de handicap.

J'exprime ma gratitude envers les professeurs qui m'ont enseigné, tant au baccalauréat qu'à la maîtrise à l'UQAM. En particulier, je remercie Fabrice Larribe, qui m'a donné ma première opportunité d'être démonstrateur et François Watier, avec qui j'ai fait un stage d'initiation à la recherche pendant mon baccalauréat. Je remercie Gisèle Legault, responsable des laboratoires informatiques, qui a rendu mon travail sur ordinateur plus facile en me fournissant un plus grand écran. Je remercie également Manon Gauthier, assistante à la gestion des programmes de cycles supérieurs en mathématiques, pour son aide au niveau administratif. Je remercie Gilles Ouellet, conseiller au Service d'intégration des étudiants en situation de handicap de l'UQAM pour son aide, son support et pour les Hiboux !

Je tiens à remercier Malorie Chabot-Blanchet et Marie Forest, mes collègues et amies, avec qui j'ai étudié au baccalauréat et à la maîtrise en statistique. Je remercie également

mon père, ma mère et ma soeur pour leur soutien et leurs encouragements.

En terminant, je remercie à nouveau Sorana et Fabrice qui, grâce à leur intervention en mars 2009, m'ont amené à poursuivre mes études malgré les difficultés.

Merci Fabrice!

Merci Sorana!

## TABLE DES MATIÈRES

LISTE DES FIGURES . . . . .	ix
LISTE DES TABLEAUX . . . . .	xi
RÉSUMÉ . . . . .	xiii
INTRODUCTION . . . . .	1
CHAPITRE I	
NOTIONS DE BASE SUR LES PROCESSUS DE NAISSANCE ET DE MORT	5
1.1 Processus stochastiques : définitions de base . . . . .	5
1.2 Loi de Poisson et loi binomiale . . . . .	6
1.3 Processus de Poisson . . . . .	9
1.3.1 Processus de Poisson homogène . . . . .	9
1.3.2 Processus de Poisson non homogène . . . . .	9
1.3.3 Processus de Poisson composé . . . . .	10
1.4 Processus de naissance et de mort . . . . .	13
1.4.1 Chaînes de Markov à temps continu . . . . .	13
1.4.2 Processus de naissance et de mort homogène . . . . .	14
1.4.3 Processus de naissance et de mort non homogène . . . . .	15
CHAPITRE II	
MODÈLES DE PROPAGATION DES ÉPIDÉMIES DANS LE TEMPS . . . . .	19
2.1 Terminologie . . . . .	19
2.2 Modèles de la littérature . . . . .	20
2.2.1 Modèle déterministe : modèle SIR . . . . .	20
2.2.2 Modèle stochastique à partir d'une chaîne de Markov à temps continu à trois dimensions . . . . .	27
2.2.3 Modèle stochastique à partir de processus de naissance ou de mort purs non homogènes . . . . .	28
2.3 Modèles stochastiques proposés . . . . .	30
2.3.1 Premier modèle proposé . . . . .	30

2.3.2	Deuxième modèle proposé . . . . .	40
CHAPITRE III		
ESTIMATION DE $R_0$ . . . . .		
3.1	Suppositions au niveau des observations . . . . .	45
3.2	Procédure d'estimation de $R_0$ . . . . .	46
3.3	Méthode d'estimation de $\rho$ proposée . . . . .	46
3.3.1	Estimation de $\rho$ avec $t_z$ connu . . . . .	47
3.3.2	Estimation de $\rho$ avec $t_z$ inconnu . . . . .	50
3.4	Intervalles de confiance asymptotiques pour $\rho$ et $R_0$ . . . . .	62
3.4.1	Intervalles de confiance asymptotiques pour $\rho$ . . . . .	63
3.4.2	Intervalles de confiance asymptotiques pour $R_0$ . . . . .	65
3.4.3	Estimation de la variance des estimateurs de $\rho$ . . . . .	69
3.5	Calculs supplémentaires et démonstrations . . . . .	70
3.5.1	Calcul de la fonction de masse de $T_z$ . . . . .	70
3.5.2	Calcul de la fonction de masse de $\tilde{T}_z$ . . . . .	73
3.5.3	Démonstration de la convergence presque sûre de $\tilde{T}_z$ vers $t_z$ . . . . .	74
3.5.4	Démonstration de la convergence presque sûre de $\bar{\rho}(\tilde{T}_z)$ vers $\rho$ . . . . .	75
3.5.5	Démonstration de la convergence en loi de $\bar{\rho}(\tilde{T}_z)$ vers une loi normale . . . . .	77
CHAPITRE IV		
APPLICATIONS NUMÉRIQUES . . . . .		
4.1	Valeurs de départ des paramètres . . . . .	82
4.2	Reparamétrisation en $\beta$ et $\gamma$ . . . . .	83
4.3	Solutions du modèle déterministe . . . . .	84
4.4	Simulation des données à partir des modèles stochastiques proposés . . . . .	85
4.4.1	Modèle avec la loi binomiale . . . . .	85
4.4.2	Modèle avec la loi de Poisson . . . . .	88
4.4.3	Résultats obtenus en simulation . . . . .	89
4.5	Approximation de la fonction de masse de $T_z$ . . . . .	89
4.6	Approximation de la fonction de masse de $\tilde{T}_z$ . . . . .	93
4.7	Propriétés des estimateurs de $\rho$ . . . . .	96

4.8 Propriétés des estimateurs de $R_0$ . . . . .	102
4.9 Intervalles de confiance asymptotiques pour $\rho$ et $R_0$ . . . . .	106
4.10 Synthèse des résultats . . . . .	115
CONCLUSION . . . . .	117
APPENDICE A	
RÉSULTATS POUR LE MODÈLE AVEC LA LOI DE POISSON . . . . .	119
RÉFÉRENCES . . . . .	125



## LISTE DES FIGURES

Figure	Page
2.1 Solutions du modèle déterministe SIR . . . . .	21
3.1 Graphique de $d'(t)$ . . . . .	52
4.1 Solutions du modèle SIR : $x(t)$ , $y(t)$ et $z(t)$ . . . . .	84
4.2 Solutions du modèle SIR : $-x'(t)$ et $z'(t)$ . . . . .	85
4.3 Graphique de $\ell_D(\tau)$ . . . . .	90
4.4 Approximation de la fonction de masse de $T_Z$ . . . . .	92
4.5 Approximation de la fonction de masse de $\tilde{T}_Z$ avec $k = 10$ . . . . .	94
4.6 Biais relatif, variance et erreur quadratique moyenne approximés de $\tilde{T}_Z$ en fonction de $k$ . . . . .	96
4.7 Biais relatif de $\bar{\rho}(t_Z)$ , $\bar{\rho}(\mathbf{T}_Z)$ et $\bar{\rho}(\tilde{T}_Z)$ en fonction de $k$ . . . . .	98
4.8 Valeur absolue du biais relatif de $\bar{\rho}(\mathbf{T}_Z)$ et $\bar{\rho}(\tilde{T}_Z)$ en fonction de $k$ . . . . .	98
4.9 Variance de $\bar{\rho}(t_Z)$ , $\bar{\rho}(\mathbf{T}_Z)$ et $\bar{\rho}(\tilde{T}_Z)$ en fonction de $k$ , modèle avec la loi binomiale . . . . .	99
4.10 Coefficient de variation de $\bar{\rho}(t_Z)$ , $\bar{\rho}(\mathbf{T}_Z)$ et $\bar{\rho}(\tilde{T}_Z)$ en fonction de $k$ , modèle avec la loi binomiale . . . . .	101
4.11 Erreur quadratique moyenne de $\bar{\rho}(t_Z)$ , $\bar{\rho}(\mathbf{T}_Z)$ et $\bar{\rho}(\tilde{T}_Z)$ en fonction de $k$ , modèle avec la loi binomiale . . . . .	102

4.12	Biais relatif de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ . . . . .	104
4.13	Valeur absolue du biais relatif de $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ . . .	104
4.14	Variance de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi binomiale . . . . .	105
4.15	Erreur quadratique moyenne de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi binomiale . . . . .	106
4.16	Biais relatif de $S(\tilde{T}_z)$ . . . . .	108
A.1	Variance de $\bar{\rho}(t_z)$ , $\bar{\rho}(\mathbf{T}_z)$ et $\bar{\rho}(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi de Poisson . . . . .	119
A.2	Coefficient de variation de $\bar{\rho}(t_z)$ , $\bar{\rho}(\mathbf{T}_z)$ et $\bar{\rho}(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi de Poisson . . . . .	120
A.3	Erreur quadratique moyenne de $\bar{\rho}(t_z)$ , $\bar{\rho}(\mathbf{T}_z)$ et $\bar{\rho}(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi de Poisson . . . . .	120
A.4	Variance de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi de Poisson . . . . .	121
A.5	Erreur quadratique moyenne de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ en fonction de $k$ , modèle avec la loi de Poisson . . . . .	121

## LISTE DES TABLEAUX

Tableau	Page
4.1 Proportion des intervalles de confiance asymptotiques pour $\rho$ construits à partir de $\bar{\rho}(t_z)$ , $\bar{\rho}(\mathbf{T}_z)$ et $\bar{\rho}(\tilde{T}_z)$ qui contiennent $\rho$ , modèle avec la loi binomiale . . . . .	110
4.2 Proportion des intervalles de confiance asymptotiques pour $R_0$ construits à partir de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ qui contiennent $R_0$ , modèle avec la loi binomiale . . . . .	113
4.3 Précision moyenne de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ avec la variance estimée et le quantile de la loi de Student, modèle avec la loi binomiale . . . . .	114
A.1 Proportion des intervalles de confiance asymptotiques pour $\rho$ construits à partir de $\bar{\rho}(t_z)$ , $\bar{\rho}(\mathbf{T}_z)$ et $\bar{\rho}(\tilde{T}_z)$ qui contiennent $\rho$ , modèle avec la loi de Poisson . . . . .	122
A.2 Proportion des intervalles de confiance asymptotiques pour $R_0$ construits à partir de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ qui contiennent $R_0$ , modèle avec la loi de Poisson . . . . .	123
A.3 Précision moyenne de $\tilde{R}_0(t_z)$ , $\tilde{R}_0(\mathbf{T}_z)$ et $\tilde{R}_0(\tilde{T}_z)$ avec la variance estimée et le quantile de la loi de Student, modèle avec la loi de Poisson . . . . .	124

## RÉSUMÉ

Ce mémoire porte sur l'estimation de paramètres dans des modèles de propagation des épidémies dans le temps. On considère le cas d'une population fermée dans laquelle chacun des individus est soit susceptible, soit infecté, soit retiré (S-I-R). Un individu est dit infecté lorsqu'il est atteint d'une maladie infectieuse et qu'il est contagieux. Un individu qui n'a pas encore été infecté est dit susceptible, alors qu'un individu qui a été atteint par la maladie et qui n'est plus infecté est dit retiré (immunisé ou décédé). Un paramètre important dans ce type de modèles est le taux de reproduction  $R_0$ , qui s'interprète comme le nombre moyen d'individus à qui un individu infecté transmet la maladie, au début de l'épidémie. Plus la valeur de  $R_0 > 1$  est grande, plus l'épidémie est importante. On propose d'abord deux modèles stochastiques pour l'évolution d'une épidémie en se basant sur un modèle déterministe classique, le modèle SIR de Kermack et McKendrick (1927). Les modèles tiennent compte du type de données disponibles en pratique. Par la suite, on étudie une nouvelle méthode d'estimation de  $R_0$  et on construit un intervalle de confiance asymptotique pour  $R_0$ . Finalement, on présente des résultats obtenus en appliquant la méthode d'estimation de  $R_0$  sur des données simulées à l'aide des modèles proposés.

**Mots clés :** modèle SIR de propagation d'une épidémie, taux de reproduction  $R_0$ , estimation, processus de naissance et de mort non homogène, processus de Poisson non homogène.

—

—

—

---

## INTRODUCTION

Dans ce mémoire, on traite de la modélisation de la propagation des épidémies dans le temps et de l'estimation des paramètres de ces modèles. Un des deux objectifs principaux est de proposer de nouveaux modèles de propagation des épidémies dans le temps, modèles qui reflètent bien le type de données disponibles en pratique. L'autre objectif principal est de proposer une nouvelle méthode d'estimation à partir de telles données.

On considère une population fermée, comme c'est le cas pour le modèle déterministe SIR (Kermack et McKendrick, 1927) et son équivalent stochastique, présenté entre autres par Kendall (1956). Dans ces modèles, à chaque instant  $t \geq 0$ , chacun des individus est soit susceptible, soit infecté, soit retiré (S-I-R). Un individu est dit infecté lorsqu'il est atteint d'une maladie infectieuse et qu'il est contagieux. Un individu qui n'a pas encore été infecté est dit susceptible, alors qu'un individu qui a été atteint par la maladie et qui n'est plus infecté est dit retiré (immunisé ou décédé).

Un paramètre d'intérêt de ces modèles est le taux de reproduction  $R_0$ , qui s'interprète comme le nombre moyen d'individus à qui un individu infecté transmet la maladie, au début de l'épidémie. Ce paramètre est très important dans l'étude de la propagation des épidémies, étant donné qu'il permet de quantifier la virulence d'une maladie infectieuse et, donc, de prévoir l'évolution de l'épidémie dans le temps. De plus, en pratique, on s'intéresse souvent à la valeur de  $R_0$  pour juger de l'efficacité de certaines mesures visant à ralentir la propagation d'une épidémie, comme la vaccination. En effet, on cherche à savoir de quelle façon la valeur de  $R_0$  a été modifiée suite à ces mesures. Enfin, mentionnons qu'une épidémie est importante si  $R_0 > 1$ . La valeur 1 représente donc un seuil pour le paramètre  $R_0$ .

Les nouveaux modèles de propagation des épidémies dans le temps proposés dans ce mémoire permettent d'estimer  $R_0$  à partir du type de données disponibles en pratique. En général, pour certains temps  $t_1, t_2, \dots, t_n$ , on a des données sur le nombre total d'individus qui ont été infectés par la maladie jusqu'au temps  $t_i$  et sur le nombre total d'individus qui en sont décédés jusqu'au temps  $t_i$ ,  $i = 1, 2, \dots, n$ . Par exemple, pour une épidémie de grippe saisonnière, ces données peuvent être rapportées à chaque semaine. Donc, d'une part, on ne dispose ni du nombre d'individus retirés, ni du nombre d'individus encore infectés au temps  $t_i$ ,  $i = 1, 2, \dots, n$ . D'autre part, même si, conceptuellement, l'évolution de l'épidémie a lieu en temps continu, en pratique, les données sont recueillies à des temps discrets. Un autre objectif de la modélisation est que les espérances du nombre d'individus susceptibles et retirés dans les modèles stochastiques proposés soient égales aux solutions du modèle déterministe SIR. Ce modèle déterministe décrit de façon simple et adéquate le déroulement moyen de l'épidémie dans une grande population. Notons qu'on y fait encore appel dans des études épidémiologiques sur la grippe (voir Cruz-Pacheco *et al.*, 2009).

---

On propose également une nouvelle méthode d'estimation de  $R_0$  qui se base sur les données disponibles en pratique. En fait, on estime  $R_0$  à partir du nombre d'individus susceptibles et du nombre d'individus décédés, en tenant compte de la discrétisation du temps d'observation. On cherche à proposer une méthode d'estimation pour laquelle on n'a pas à faire de suppositions quant aux valeurs de certains autres paramètres du modèle, tel que rencontré dans certaines études épidémiologiques (voir Cruz-Pacheco *et al.*, 2009).

La structure du mémoire est la suivante. Au chapitre 1, on introduit certaines notions de base sur les processus de naissance et de mort qui sont nécessaires au développement des modèles. Par la suite, au chapitre 2, on présente sommairement certains des modèles de propagation des épidémies dans le temps de la littérature, soit le modèle déterministe SIR et deux modèles stochastiques, en plus de présenter de façon plus détaillée les deux nouveaux modèles que l'on propose. Au chapitre 3, on explique la nouvelle méthode

d'estimation de  $R_0$ , en plus de développer certaines propriétés des estimateurs et de construire des intervalles de confiance pour  $R_0$  à partir de ces estimateurs. Enfin, au chapitre 4, on présente et on commente les résultats obtenus numériquement et en simulation lorsqu'on applique la méthode d'estimation de  $R_0$  à des données simulées à partir des modèles proposés.



# CHAPITRE I

## NOTIONS DE BASE SUR LES PROCESSUS DE NAISSANCE ET DE MORT

Dans ce chapitre, on présente certains concepts de base sur les processus de naissance et de mort. Nous suivons l'approche de Parzen (1962) dans les chapitres 1, 4, 6 et 7. Les notions et les résultats présentés dans ce chapitre sont nécessaires dans les modèles introduits au chapitre 2.

### 1.1 Processus stochastiques : définitions de base

**Définition 1.1.1** (Processus stochastique). *Un processus stochastique  $\{X(t), t \in \mathbb{T}\}$  est une collection de variables aléatoires, c'est-à-dire que, pour tout  $t \in \mathbb{T}$ ,  $X(t)$  est une variable aléatoire.*

L'indice  $t$  est souvent interprété comme étant le temps, tandis que  $X(t)$  est appelé l'état du processus au temps  $t$ . L'ensemble des valeurs que peut prendre  $X(t)$ , noté  $\mathbb{S}$ , est appelé l'espace des états du processus et l'ensemble  $\mathbb{T}$  est appelé l'ensemble des indices du processus. Lorsque  $t$  représente le temps et que  $\mathbb{T}$  est dénombrable, on dit que le processus est à temps discret, alors qu'il est à temps continu si  $\mathbb{T}$  est un intervalle dans  $\mathbb{R}$ .

**Définition 1.1.2** (Processus de comptage). *Un processus stochastique à temps continu  $\{N(t), t \geq 0\}$  est un processus de comptage si  $N(t)$  représente le nombre total d'événements s'étant produits jusqu'au temps  $t$ .*

Un processus de comptage possède donc les propriétés suivantes :

- (i)  $N(t) \geq 0$  ;
- (ii)  $N(t) \in \mathbb{N}$  ;
- (iii) si  $0 \leq s < t$ , alors  $N(s) \leq N(t)$  ;
- (iv) pour  $0 \leq s < t$ , l'accroissement  $N(t) - N(s)$  correspond au nombre d'événements se produisant dans l'intervalle  $(s, t]$ .

**Définition 1.1.3** (Accroissements indépendants). *On dit qu'un processus de comptage a des accroissements indépendants si les variables aléatoires qui comptent le nombre d'événements se produisant dans des intervalles de temps disjoints sont indépendantes.*

**Définition 1.1.4** (Accroissements stationnaires). *On dit qu'un processus de comptage a des accroissements stationnaires si la loi que suit la variable aléatoire comptant le nombre d'événements se produisant entre les temps  $t$  et  $t + h$  ne dépend que de  $h$ , pour tout  $t \geq 0$ .*

Dans la section suivante, on introduit la loi de Poisson et la loi binomiale. On présente également deux propositions sur lesquelles se basent certains des résultats présentés dans la suite du chapitre 1 et au chapitre 2.

## 1.2 Loi de Poisson et loi binomiale

**Définition 1.2.1** (Loi de Poisson). *Une variable aléatoire discrète est distribuée selon une loi de Poisson de paramètre  $\lambda$ , notée  $\mathcal{P}(\lambda)$ , si sa fonction de masse est*

$$P[X = k] = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!}, & k = 0, 1, \dots, \\ 0, & \text{sinon.} \end{cases} \quad (1.1)$$

On a que  $\mathbb{E}[X] = \text{Var}[X] = \lambda$ .

**Définition 1.2.2** (Loi binomiale). Une variable aléatoire discrète est distribuée selon une loi binomiale de paramètres  $n$  et  $p$ , notée  $\mathcal{B}(n, p)$ , si sa fonction de masse est

$$P[X = k] = \begin{cases} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, & k = 0, 1, \dots, n, \\ 0, & \text{sinon.} \end{cases} \quad (1.2)$$

On a que  $\mathbb{E}[X] = np$  et  $\text{Var}[X] = np(1-p)$ .

**Proposition 1.2.1.** Si  $X$  est de loi binomiale  $\mathcal{B}(n, p_1)$  et si la loi conditionnelle de  $Y$  sachant  $X$  est une loi binomiale  $\mathcal{B}(X, p_2)$ , alors  $Y$  est de loi binomiale  $\mathcal{B}(n, p_1 p_2)$ .

*Démonstration.* Par hypothèse, pour  $k = 0, 1, \dots, n$ , on a que

$$P[Y = k | X = j] = \begin{cases} \frac{j!}{k!(j-k)!} p_2^k (1-p_2)^{j-k}, & k \leq j, \\ 0, & \text{sinon.} \end{cases}$$

Alors, pour  $k = 0, 1, \dots, n$ , on a que

$$\begin{aligned} P[Y = k] &= \\ &= \sum_{j=0}^n P[Y = k | X = j] P[X = j] \\ &= \sum_{j=k}^n \frac{j!}{k!(j-k)!} p_2^k (1-p_2)^{j-k} \frac{n!}{j!(n-j)!} p_1^j (1-p_1)^{n-j} \\ &= \sum_{m=0}^{n-k} \frac{1}{k!m!} p_2^k (1-p_2)^m \frac{n!}{[(n-k)-m]!} p_1^{m+k} (1-p_1)^{(n-k)-m} \\ &= \frac{n!}{k!(n-k)!} (p_1 p_2)^k (1-p_1 p_2)^{n-k} \\ &\quad \sum_{m=0}^{n-k} \left( \frac{(n-k)!}{m![(n-k)-m]!} \left( \frac{p_1(1-p_2)}{1-p_1 p_2} \right)^m \left( \frac{1-p_1}{1-p_1 p_2} \right)^{(n-k)-m} \right). \end{aligned} \quad (1.3)$$

Comme

$$\begin{aligned} 1 - \frac{p_1(1-p_2)}{1-p_1 p_2} &= \frac{1-p_1 p_2 - p_1 + p_1 p_2}{1-p_1 p_2} \\ &= \frac{1-p_1}{1-p_1 p_2}, \end{aligned}$$

on a que

$$\begin{aligned}
& \sum_{m=0}^{n-k} \left( \frac{(n-k)!}{m![(n-k)-m]!} \left( \frac{p_1(1-p_2)}{1-p_1p_2} \right)^m \left( \frac{1-p_1}{1-p_1p_2} \right)^{(n-k)-m} \right) \\
&= \sum_{m=0}^{n-k} \left( \frac{(n-k)!}{m![(n-k)-m]!} \left( \frac{p_1(1-p_2)}{1-p_1p_2} \right)^m \left( 1 - \frac{p_1(1-p_2)}{1-p_1p_2} \right)^{(n-k)-m} \right) \\
&= 1,
\end{aligned} \tag{1.4}$$

car c'est la somme de la fonction de masse d'une variable aléatoire de loi binomiale  $\mathcal{B}\left(n-k, \frac{p_1(1-p_2)}{1-p_1p_2}\right)$  sur toutes ses valeurs possibles. À partir des équations (1.3) et (1.4), on a donc que

$$P[Y = k] = \frac{n!}{k!(n-k)!} (p_1p_2)^k (1-p_1p_2)^{n-k},$$

$k = 0, 1, \dots, n$ . Donc,  $Y$  est de loi binomiale  $\mathcal{B}(n, p_1p_2)$ .  $\square$

**Proposition 1.2.2.** *Si  $X$  est de loi de Poisson  $\mathcal{P}(\lambda)$  et si la loi conditionnelle de  $Y$  sachant  $X$  est une loi binomiale  $\mathcal{B}(X, p)$ , alors  $Y$  est de loi de Poisson  $\mathcal{P}(p\lambda)$ .*

*Démonstration.* Par hypothèse, pour  $k = 0, 1, \dots$ , on a que

$$P[Y = k | X = j] = \begin{cases} \frac{j!}{k!(j-k)!} p^k (1-p)^{j-k}, & k \leq j, \\ 0, & \text{sinon.} \end{cases}$$

Alors, pour  $k = 0, 1, \dots$ , on a que

$$\begin{aligned}
P[Y = k] &= \sum_{j=0}^{\infty} P[Y = k | X = j] P[X = j] \\
&= \sum_{j=k}^{\infty} \frac{j!}{k!(j-k)!} p^k (1-p)^{j-k} \frac{e^{-\lambda} \lambda^j}{j!} \\
&= \sum_{m=0}^{\infty} \frac{1}{k!m!} p^k (1-p)^m e^{-\lambda} \lambda^m \lambda^k \\
&= \frac{e^{-\lambda} (p\lambda)^k}{k!} \sum_{m=0}^{\infty} \frac{[\lambda(1-p)]^m}{m!} \\
&= \frac{e^{-\lambda} (p\lambda)^k}{k!} e^{\lambda(1-p)} \\
&= \frac{e^{-p\lambda} (p\lambda)^k}{k!}.
\end{aligned}$$

$k = 0, 1, \dots$ . Donc,  $Y$  est de loi de Poisson  $\mathcal{P}(p\lambda)$ . □

### 1.3 Processus de Poisson

#### 1.3.1 Processus de Poisson homogène

**Définition 1.3.1** (Processus de Poisson homogène). *Un processus de comptage  $\{N(t), t \geq 0\}$  est un processus de Poisson homogène d'intensité  $\lambda$ ,  $\lambda > 0$ , si :*

- (i)  $N(0) = 0$ ;
- (ii) il a des accroissements indépendants ;
- (iii) il a des accroissements stationnaires ;
- (iv) si  $0 \leq s < t$ , alors  $N(t) - N(s)$  est de loi de Poisson  $P(\lambda(t - s))$ .

Comme  $N(t)$  est de loi de Poisson  $\mathcal{P}(\lambda t)$ , on a que  $\mathbb{E}[N(t)] = \text{Var}[N(t)] = \lambda t$ .

#### 1.3.2 Processus de Poisson non homogène

**Définition 1.3.2** (Processus de Poisson non homogène). *Un processus de comptage  $\{N(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $\lambda(t)$ ,  $\lambda(t) > 0$ ,  $t \geq 0$ , si :*

- (i)  $N(0) = 0$ ;
- (ii) il a des accroissements indépendants ;
- (iii) si  $0 \leq s < t$ , alors  $N(t) - N(s)$  est de loi de Poisson de  $P(m(t) - m(s))$ , où
 
$$m(t) = \int_0^t \lambda(\tau) d\tau.$$

Comme  $N(t)$  est de loi de Poisson  $\mathcal{P}(m(t))$ , on a que  $\mathbb{E}[N(t)] = \text{Var}[N(t)] = m(t)$ . On appelle d'ailleurs  $m(t)$  la fonction moyenne du processus de Poisson non homogène d'intensité  $\lambda(t)$ . Il est à noter que les accroissements d'un processus de Poisson non homogène ne sont généralement pas stationnaires et que le processus de Poisson homogène est un cas particulier du processus de Poisson non homogène pour lequel la fonction  $\lambda(t) = \lambda$  pour tout  $t \geq 0$ , où  $\lambda > 0$ .

**Proposition 1.3.1.** *Soit  $\{N(t), t \geq 0\}$  un processus de Poisson d'intensité  $\lambda(t)$  et soit  $0 \leq s < t$ . Alors, on a que*

$$\text{Cov}[N(s), N(t)] = \text{Var}[N(s)].$$

*Démonstration.* Soit  $0 \leq s < t$ . Comme  $\{N(t), t \geq 0\}$  est un processus de Poisson, on a que les variables  $N(s) - N(0)$  et  $N(t) - N(s)$  sont indépendantes. Comme  $N(0) = 0$ ,  $N(s)$  et  $N(t) - N(s)$  sont indépendantes. On a donc que  $\text{Cov}[N(s), N(t) - N(s)] = 0$ , d'où

$$\begin{aligned} \text{Cov}[N(s), N(t)] &= \text{Cov}[N(s), \{N(t) - N(s)\} + N(s)] \\ &= \text{Cov}[N(s), N(t) - N(s)] + \text{Cov}[N(s), N(s)] \\ &= \text{Cov}[N(s), N(s)] \\ &= \text{Var}[N(s)]. \end{aligned}$$

□

### 1.3.3 Processus de Poisson composé

**Définition 1.3.3** (Processus de Poisson composé). *Un processus stochastique  $\{M(t), t \geq 0\}$  est un processus de Poisson composé si  $M(t)$  peut être exprimé de la façon suivante :*

$$M(t) = \begin{cases} \sum_{j=1}^{N(t)} Y_j, & N(t) > 0, \\ 0, & \text{sinon,} \end{cases}$$

*où  $\{N(t), t \geq 0\}$  est un processus de Poisson et  $\{Y_j, j = 1, 2, \dots\}$  est une collection de variables aléatoires indépendantes et identiquement distribuées. On suppose que  $\{N(t), t \geq 0\}$  et  $\{Y_j, j = 1, 2, \dots\}$  sont indépendants.*

Habituellement,  $\{N(t), t \geq 0\}$  est un processus de Poisson homogène. C'est d'ailleurs le cas dans Parzen (1962). Dans nos applications, on considère que  $\{N(t), t \geq 0\}$  est un processus de Poisson non homogène. On prouve le résultat suivant sur lequel se basent certains des calculs et des simulations des chapitres 2, 3 et 4.

**Proposition 1.3.2.** Soit  $\{M(t), t \geq 0\}$  un processus de Poisson composé tel que défini précédemment. Supposons que  $\{N(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $\lambda(t)$ . Si  $Y_j$  est de loi binomiale  $\mathcal{B}(1, p)$ , alors  $\{M(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\lambda(t)$ .

*Démonstration.* Afin de démontrer cette proposition, on vérifie que  $\{M(t), t \geq 0\}$  satisfait la définition 1.3.2.

(i) Comme  $\{N(t), t \geq 0\}$  est un processus de Poisson,  $N(0) = 0$ . Cela entraîne que  $M(0) = 0$ .

(ii) Soit  $0 \leq s < t \leq u < v$  et soit

$$M_1 = M(t) - M(s),$$

$$M_2 = M(v) - M(u),$$

$$N_1 = N(t) - N(s),$$

$$N_2 = N(v) - N(u).$$

La loi conditionnelle de  $M_1$  sachant  $N_1$  et  $N_2$  est la même que la loi conditionnelle de  $M_1$  sachant  $N_1$  car

$$M(t) - M(s) = \sum_{j=N(s)+1}^{N(t)} Y_j.$$

De même, la loi conditionnelle de  $M_2$  sachant  $N_1$  et  $N_2$  est la même que la loi conditionnelle de  $M_2$  sachant  $N_2$  car

$$M(v) - M(u) = \sum_{j=N(u)+1}^{N(v)} Y_j.$$

De plus, comme  $N_1$  et  $N_2$  sont indépendants et comme les variables  $Y_j$  sont

indépendantes,  $M_1$  et  $M_2$  sont indépendantes conditionnellement à  $N_1$  et  $N_2$ . Alors

$$\begin{aligned}
& P[M_1 = k_1, M_2 = k_2] \\
&= \sum_{n_1=k_1}^{\infty} \sum_{n_2=k_2}^{\infty} P[M_1 = k_1, M_2 = k_2 | N_1 = n_1, N_2 = n_2] P[N_1 = n_1, N_2 = n_2] \\
&= \sum_{n_1=k_1}^{\infty} \sum_{n_2=k_2}^{\infty} \left( P[M_1 = k_1 | N_1 = n_1, N_2 = n_2] P[M_2 = k_2 | N_1 = n_1, N_2 = n_2] \right. \\
&\quad \left. P[N_1 = n_1] P[N_2 = n_2] \right) \\
&= \left( \sum_{n_1=k_1}^{\infty} P[M_1 = k_1 | N_1 = n_1] P[N_1 = n_1] \right) \cdot \\
&\quad \left( \sum_{n_2=k_2}^{\infty} P[M_2 = k_2 | N_2 = n_2] P[N_2 = n_2] \right) \\
&= P[M_1 = k_1] P[M_2 = k_2].
\end{aligned}$$

Cela entraîne que  $M_1$  et  $M_2$  sont indépendants. De la même façon, on pourrait prouver que les accroissements sur  $n$  intervalles disjoints sont indépendants.  $\{M(t), t \geq 0\}$  a donc des accroissements indépendants.

- (iii) Soit  $0 \leq s < t$ . Comme  $\{N(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $\lambda(t)$ ,  $N(t) - N(s)$  est de loi de Poisson  $\mathcal{P}(m(t) - m(s))$ , où

$$m(t) = \int_0^t \lambda(\tau) d\tau.$$

De plus, la loi conditionnelle de  $M(t) - M(s)$  sachant  $N(t) - N(s)$  est une loi binomiale  $\mathcal{B}(N(t) - N(s), p)$ . Par la proposition 1.2.2, on a que  $M(t) - M(s)$  est de loi de Poisson  $\mathcal{P}[p(m(t) - m(s))]$ . Soit

$$m^*(t) = \int_0^t p\lambda(\tau) d\tau.$$

Alors,  $M(t) - M(s)$  est de loi de Poisson  $\mathcal{P}(m^*(t) - m^*(s))$ .

Donc,  $\{M(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\lambda(t)$ .  $\square$

## 1.4 Processus de naissance et de mort

### 1.4.1 Chaînes de Markov à temps continu

**Définition 1.4.1** (Chaîne de Markov à temps continu). *Un processus stochastique à temps continu  $\{N(t), t \geq 0\}$  est une chaîne de Markov à temps continu si :*

- (i) *pour n'importe quel ensemble de  $n$  temps  $t_1 < t_2 < \dots < t_n$ , la loi conditionnelle de  $N(t_n)$  sachant  $N(t_1), N(t_2), \dots, N(t_{n-1})$  ne dépend que de  $N(t_{n-1})$  ;*
- (ii)  *$\mathbb{S}$ , l'espace des états du processus, est dénombrable.*

**Définition 1.4.2** (Fonction de transition). *Soit  $j$  et  $k$  deux états de la chaîne de Markov à temps continu  $\{N(t), t \geq 0\}$  et soit  $0 \leq s \leq t$ . La fonction de transition de  $\{N(t), t \geq 0\}$  est donnée par :*

$$p_{j,k}(s, t) = P[X(t) = k | X(s) = j]. \quad (1.5)$$

Une chaîne de Markov à temps continu est dite homogène si sa fonction de transition dépend uniquement de la différence  $(t - s)$ . Dans ce cas, sa fonction de transition est notée

$$p_{j,k}(t) = P[X(t + u) = k | X(u) = j], \quad (1.6)$$

pour tout  $u \geq 0$ . Une chaîne de Markov à temps continu dont la fonction de transition ne dépend pas uniquement de la différence  $(t - s)$  est dite non homogène.

**Définition 1.4.3** (Taux de transition instantané). *Soit  $\{N(t), t \geq 0\}$  une chaîne de Markov à temps continu. Si  $\{N(t), t \geq 0\}$  est homogène, le taux de transition instantané à partir de l'état  $k$  est donné par*

$$q_k = \lim_{u \rightarrow 0} \frac{1 - p_{k,k}(u)}{u}$$

*et le taux de transition instantané de l'état  $j$  vers l'état  $k$  est donné par*

$$q_{j,k} = \lim_{u \rightarrow 0} \frac{p_{j,k}(u)}{u}.$$

Si  $\{N(t), t \geq 0\}$  est non homogène, le taux de transition instantané à partir de l'état  $k$  au temps  $t$  est donné par

$$q_k(t) = \lim_{u \rightarrow 0} \frac{1 - p_{k,k}(t, t+u)}{u}$$

et le taux de transition instantané de l'état  $j$  à l'état  $k$  au temps  $t$  est donné par

$$q_{j,k}(t) = \lim_{u \rightarrow 0} \frac{p_{j,k}(t, t+u)}{u}.$$

Pour alléger le texte, on utilisera l'expression *taux de transition* pour désigner le *taux de transition instantané*.

#### 1.4.2 Processus de naissance et de mort homogène

**Définition 1.4.4** (Processus de naissance et de mort homogène). *Une chaîne de Markov homogène à temps continu  $\{N(t), t \geq 0\}$  dont l'espace des états est  $\mathbb{N}$  est un processus de naissance et de mort homogène si, pour tous les couples d'états  $(j, k)$  tels que  $|j - k| > 1$ , on a que*

$$q_{j,k} = 0.$$

Pour le processus de naissance et de mort homogène, on introduit la notation suivante :

$$\begin{aligned} \lambda_j &= q_{j,j+1} = \lim_{u \rightarrow 0} \frac{p_{j,j+1}(u)}{u}, \quad j = 0, 1, \dots, \\ \mu_j &= q_{j,j-1} = \lim_{u \rightarrow 0} \frac{p_{j,j-1}(u)}{u}, \quad j = 1, 2, \dots, \end{aligned}$$

et  $\mu_0 = 0$ . Il est à noter que les limites sont uniformes en  $j$ . On appelle  $\lambda_j$  le taux de naissance à l'état  $j$  et  $\mu_j$  le taux de mort à l'état  $j$ . On a que

$$\lambda_j + \mu_j = q_j,$$

$j = 0, 1, \dots$ . Soit  $\{N(t), t \geq 0\}$  un processus de naissance et de mort homogène. Dans un intervalle de temps très petit, la valeur de  $N(t)$  peut soit augmenter de 1, soit diminuer de 1, soit rester la même. Si  $\mu_j = 0$  pour tout  $j \in \mathbb{N}$ , alors on dit que  $\{N(t), t \geq 0\}$  est un processus de naissance pur homogène. On peut montrer qu'un processus de naissance

pur homogène pour lequel  $\lambda_j = \lambda$  pour tout  $j \in \mathbb{N}$  est équivalent à un processus de Poisson homogène. Si  $\lambda_j = 0$  pour tout  $j \in \mathbb{N}$ , alors on dit que  $\{N(t), t \geq 0\}$  est un processus de mort pur homogène.

### 1.4.3 Processus de naissance et de mort non homogène

**Définition 1.4.5** (Processus de naissance et de mort non homogène). *Une chaîne de Markov non homogène à temps continu  $\{N(t), t \geq 0\}$  dont l'espace des états est  $\mathbb{N}$  est un processus de naissance et de mort non homogène si, pour tous les couples d'états  $(j, k)$  tels que  $|j - k| > 1$ , on a que*

$$q_{j,k}(t) = 0,$$

pour tout  $t \geq 0$ .

Pour le processus de naissance et de mort non homogène, on introduit la notation suivante :

$$\begin{aligned} \lambda_j(t) &= q_{j,j+1}(t) = \lim_{u \rightarrow 0} \frac{p_{j,j+1}(t, t+u)}{u}, \quad j = 0, 1, \dots, \\ \mu_j(t) &= q_{j,j-1}(t) = \lim_{u \rightarrow 0} \frac{p_{j,j-1}(t, t+u)}{u}, \quad j = 1, 2, \dots, \end{aligned}$$

et  $\mu_0(t) = 0, t \geq 0$ . Il est à noter que les limites sont uniformes en  $j$ . On appelle  $\lambda_j(t)$  le taux de naissance à l'état  $j$  au temps  $t$  et  $\mu_j(t)$  le taux de mort à l'état  $j$  au temps  $t$ . On a que

$$\lambda_j(t) + \mu_j(t) = q_j(t),$$

$j = 0, 1, \dots$  et  $t \geq 0$ . Soit  $\{N(t), t \geq 0\}$  un processus de naissance et de mort non homogène. Si  $\mu_j(t) = 0$  pour tout  $j \in \mathbb{N}$  et pour tout  $t \geq 0$ , alors on dit que  $\{N(t), t \geq 0\}$  est un processus de naissance pur non homogène. On peut montrer qu'un processus de naissance pur non homogène pour lequel  $\lambda_j(t) = \lambda(t)$  pour tout  $j \in \mathbb{N}$  est équivalent à un processus de Poisson non homogène. Si  $\lambda_j(t) = 0$  pour tout  $j \in \mathbb{N}$  et pour tout  $t \geq 0$ , alors on dit que  $\{N(t), t \geq 0\}$  est un processus de mort pur non homogène. Dans les modèles présentés au chapitre 2, on s'intéresse à un cas particulier de processus de

mort purs, c'est-à-dire le cas où  $\mu_j(t) = j\mu(t)$ , pour tout  $j \in \mathbb{N}$ . Afin de déterminer la fonction de transition d'un processus de mort pur non homogène avec  $\mu_j(t) = j\mu(t)$  pour tout  $j \in \mathbb{N}$ , on doit introduire la notion de fonction génératrice des probabilités de transition et énoncer deux théorèmes. La preuve de ces théorèmes n'est pas présentée, étant donné qu'elle dépasse le niveau de ce mémoire.

**Définition 1.4.6** (Fonction génératrice des probabilités de transition). *La fonction génératrice des probabilités de transition d'une chaîne de Markov à temps continu est donnée par*

$$\psi_{j,s}(z,t) = \sum_{k=0}^{\infty} z^k p_{j,k}(s,t). \quad (1.7)$$

Cette définition est tirée de Parzen (1962), chapitre 7.

**Théorème 1.4.1** (Théorème 5A, chapitre 7, Parzen 1962). *Soit  $\{N(t), t \geq 0\}$  un processus de naissance et de mort non homogène avec  $\lambda_j(t) = j\lambda(t)$  et  $\mu_j(t) = j\mu(t)$ , pour tout  $j \in \mathbb{N}$  et pour tout  $t \geq 0$ . Soit  $0 \leq s < t$  et  $|z| < 1$ . Alors,*

$$\frac{\partial}{\partial t} \psi_{j,s}(z,t) = \frac{\partial}{\partial z} \psi_{j,s}(z,t) \left[ (z-1)(z\lambda(t) - \mu(t)) \right], \quad (1.8)$$

avec la condition limite suivante :

$$\psi_{j,s}(z,s) = z^j \text{ si } P[N(s) = j] = 1, \quad (1.9)$$

$j \in \mathbb{N}$ .

**Théorème 1.4.2** (Théorème 5B, chapitre 7, Parzen 1962). *Supposons que la fonction génératrice des probabilités de transition satisfait*

$$\frac{\partial}{\partial t} \psi_{j,s}(z,t) = \alpha(z,t) \frac{\partial}{\partial z} \psi_{j,s}(z,t) \quad (1.10)$$

avec la condition limite donnée par l'équation (1.9). Soit  $u(\cdot, \cdot)$  une fonction qui est telle que la solution  $z(t)$  de l'équation

$$\frac{dz}{dt} + \alpha(z,t) = 0 \quad (1.11)$$

satisfait

$$u(z,t) = c, \quad (1.12)$$

où  $c$  est une constante. Soit

$$g_t(z) = u(z, t).$$

Alors

$$\psi_{j,s}(z, t) = \{g_s^{-1}(u(z, t))\}^j. \quad (1.13)$$

En se basant sur le théorème 1.4.1 et le théorème 1.4.2, Parzen (1962) présente le calcul de la fonction de transition pour un processus de naissance et de mort non homogène avec  $\lambda_j(t) = j\lambda(t)$  et  $\mu_j(t) = j\mu(t)$ . D'une façon analogue à celle de Parzen, on présente le calcul de la fonction de transition pour un processus de mort pur non homogène avec  $\mu_j(t) = j\mu(t)$ .

**Proposition 1.4.1.** *Soit  $\{N(t), t \geq 0\}$  un processus de mort pur non homogène avec  $\mu_j(t) = j\mu(t)$ , pour tout  $j \in \mathbb{N}$  et pour tout  $t \geq 0$ , et soit  $0 \leq s < t$ . Alors, la loi conditionnelle de  $N(t)$  sachant  $N(s)$  est une loi binomiale  $\mathcal{B}(N(s), \exp\{\eta(s) - \eta(t)\})$ , où  $\eta(t) = \int_0^t \mu(\tau) d\tau$ .*

*Démonstration.* De l'équation (1.8), on a

$$\frac{\partial}{\partial t} \psi_{j,s}(z, t) = -(z-1)\mu(t) \frac{\partial}{\partial z} \psi_{j,s}(z, t).$$

En identifiant les termes dans l'équation (1.10), on a que

$$\alpha(z, t) = -(z-1)\mu(t).$$

L'équation (1.11) peut donc s'écrire comme suit :

$$\begin{aligned} \frac{dz}{dt} - (z-1)\mu(t) = 0 &\Leftrightarrow \frac{dz}{dt} \frac{1}{z-1} - \mu(t) dt = 0 \\ &\Leftrightarrow \log(z-1) - \int_0^t \mu(\tau) d\tau = 0. \end{aligned}$$

Soit

$$\begin{aligned} \eta(t) &= \int_0^t \mu(\tau) d\tau, \\ u(z, t) &= \log(z-1) - \eta(t) \end{aligned}$$

et  $g_t(z) = u(z, t)$ . Pour appliquer le théorème 1.4.2, on doit trouver l'expression de  $g_s^{-1}(\cdot)$ . Soit  $x = u(z, s)$ . On a que

$$\begin{aligned} x = \log(z - 1) - \eta(s) &\Leftrightarrow z - 1 = \exp\{x + \eta(s)\} \\ &\Leftrightarrow z = \exp\{x + \eta(s)\} + 1. \end{aligned} \quad (1.14)$$

En posant

$$z = g_s^{-1}(x) = g_s^{-1}(u(z, t))$$

et

$$x = u(z, t) = \log(z - 1) - \eta(t)$$

dans l'équation (1.14), on obtient

$$\begin{aligned} g_s^{-1}(u(z, t)) &= 1 + \exp\{\log(z - 1) - \eta(t) + \eta(s)\} \\ &= 1 + (z - 1) \exp\{\eta(s) - \eta(t)\}. \end{aligned} \quad (1.15)$$

On obtient ensuite la fonction génératrice des probabilités de transition à partir des équations (1.13) et (1.15) :

$$\begin{aligned} \psi_{j,s}(z, t) &= \left[ g_s^{-1}(u(z, t)) \right]^j \\ &= \left[ 1 + (z - 1) \exp\{\eta(s) - \eta(t)\} \right]^j \\ &= \left[ 1 - \exp\{\eta(s) - \eta(t)\} + z \exp\{\eta(s) - \eta(t)\} \right]^j, \end{aligned}$$

qui est la fonction génératrice des probabilités d'une variable aléatoire de loi binomiale  $\mathcal{B}(j, \exp\{\eta(s) - \eta(t)\})$ . On a donc que la loi conditionnelle de  $N(t)$  sachant  $N(s)$  est une loi binomiale  $\mathcal{B}(N(s), \exp\{\eta(s) - \eta(t)\})$ , où  $\eta(t) = \int_0^t \mu(\tau) d\tau$ .  $\square$

En particulier, on a que la loi conditionnelle de  $N(t) - N(0)$  sachant  $N(0)$  est une loi binomiale  $\mathcal{B}(N(0), \exp\{\eta(t)\})$ . Comme la loi conditionnelle de  $N(t)$  sachant  $N(s)$  est une loi binomiale  $\mathcal{B}(N(s), \exp\{\eta(s) - \eta(t)\})$ , on a que la loi conditionnelle de  $N(t) - N(s)$  sachant  $N(s)$  est une loi binomiale  $\mathcal{B}(N(s), 1 - \exp\{\eta(s) - \eta(t)\})$ . Le nombre de morts dans l'intervalle  $(s, t]$  correspond donc au nombre de succès parmi  $N(s)$  essais de Bernoulli indépendants, pour lesquels la probabilité de succès est égale à  $1 - \exp\{\eta(s) - \eta(t)\}$ .

## CHAPITRE II

### MODÈLES DE PROPAGATION DES ÉPIDÉMIES DANS LE TEMPS

Dans ce chapitre, on introduit d'abord la terminologie relative à l'étude des épidémies. Par la suite, on présente trois modèles de propagation d'une épidémie dans le temps, dont deux classiques, soit un modèle déterministe et deux modèles stochastiques. Enfin, on propose deux modèles stochastiques qui sont utilisés dans les chapitres 3 et 4.

#### 2.1 Terminologie

Dans cette section, on définit certains termes liés à l'étude des épidémies. Une **maladie infectieuse** est une maladie due à un agent infectieux, comme une bactérie, un virus ou un parasite, qui s'introduit dans le corps par l'alimentation, la respiration ou les fluides corporels (sang, sueur, etc.). Il s'y reproduit en grand nombre, puis va infecter d'autres individus, ce qu'on appelle la **contagion**. Une **épidémie** est la propagation rapide d'une maladie infectieuse à un grand nombre d'individus par contagion.

Dans ce travail, un individu est dit **infecté** lorsqu'il est atteint d'une maladie infectieuse. On fait la supposition que l'individu infecté est contagieux, c'est-à-dire qu'il peut transmettre la maladie à d'autres individus, pendant toute la durée des symptômes de la maladie. Un individu qui n'a pas encore été infecté est dit **susceptible**, alors qu'un individu qui a été atteint par la maladie et qui n'est plus infecté est dit **retiré**. Ce retrait peut correspondre à une guérison ou à un décès. On considère seulement les maladies

infectieuses pour lesquelles un individu retiré qui n'est pas décédé est immunisé contre la maladie, c'est-à-dire qu'il ne peut être infecté à nouveau.

## 2.2 Modèles de la littérature

Dans cette section, on présente trois modèles de propagation d'une épidémie dans le temps, dont deux classiques, soit un modèle déterministe et deux modèles stochastiques. Ces modèles sont directement reliés aux modèles proposés dans ce mémoire, qui sont présentés à la section 2.3.

### 2.2.1 Modèle déterministe : modèle SIR

Dans le modèle déterministe SIR (Kermack et McKendrick, 1927), une population de taille  $N$  est divisée en trois catégories, les individus étant soit susceptibles, soit infectés, soit retirés. Soit  $x(t)$  le nombre d'individus susceptibles,  $y(t)$  le nombre d'individus infectés et  $z(t)$  le nombre d'individus retirés, chacun au temps  $t \geq 0$ . Soit  $x'(t) = \frac{d}{dt}x(t)$ ,  $y'(t) = \frac{d}{dt}y(t)$  et  $z'(t) = \frac{d}{dt}z(t)$ . Les valeurs de  $x(t)$ ,  $y(t)$  et  $z(t)$  correspondent aux solutions du système de trois équations différentielles suivant :

$$\begin{cases} x'(t) = -\beta x(t)y(t), \\ y'(t) = \beta x(t)y(t) - \gamma y(t), \\ z'(t) = \gamma y(t), \end{cases} \quad (2.1)$$

de conditions initiales  $z(0) = 0$ ,  $x(0) > 0$  et  $y(0) > 0$ , avec  $x(0) + y(0) = N$ . Les paramètres  $\beta$  et  $\gamma$  prennent des valeurs strictement positives. En faisant la somme des trois équations de (2.1), on obtient :

$$\frac{d}{dt} [x(t) + y(t) + z(t)] = 0,$$

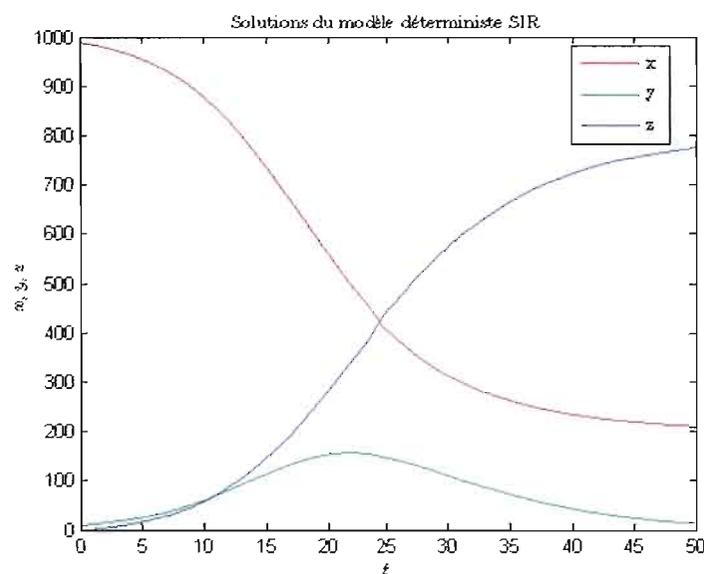
ce qui signifie que la taille de la population est supposée constante dans le temps. On note par  $\rho$  le rapport de  $\gamma$  sur  $\beta$ , c'est-à-dire

$$\rho = \frac{\gamma}{\beta}.$$

Soit  $x_0$  le nombre initial d'individus susceptibles. On note par  $R_0$  le rapport du nombre initial d'individus susceptibles sur  $\rho$ , c'est-à-dire

$$R_0 = \frac{x_0}{\rho}.$$

Les solutions du système (2.1) sont représentées graphiquement à la figure 2.1, avec  $N = 1000$ ,  $x_0 = 990$ ,  $\beta = 0,0004$  et  $\gamma = 0,2$ . Le paramètre  $\beta$  représente le taux



**Figure 2.1** Solutions du modèle déterministe SIR

d'infection. Dans la première équation de (2.1), on remarque que plus la valeur de  $\beta$  est grande, plus la décroissance du nombre d'individus susceptibles est rapide, pour des valeurs fixes de  $x(t)$  et  $y(t)$ . Le paramètre  $\beta$  représente en quelque sorte la facilité avec laquelle la maladie se transmet d'un individu infecté à un individu susceptible. Plus la valeur de  $\beta$  est grande, plus la maladie se transmet facilement. Le paramètre  $\gamma$  représente quant à lui le taux de retrait. L'inverse du paramètre  $\gamma$  est la période de temps moyenne pendant laquelle un individu est infecté. Dans la troisième équation de (2.1), on remarque que plus la valeur de  $\gamma$  est grande, plus la croissance du nombre d'individus retirés est rapide, pour une valeur fixe de  $y(t)$ , étant donné que le temps

moyen d'infection est petit. Enfin, le paramètre  $R_0$  représente en quelque sorte le nombre moyen d'individus susceptibles à qui un individu infecté transmet la maladie au début de l'épidémie. Pour mieux comprendre l'effet de  $R_0$  sur l'évolution de l'épidémie, on réécrit la deuxième équation de (2.1) en fonction de  $R_0$  :

$$\begin{aligned}
 y'(t) = \beta x(t)y(t) - \gamma y(t) &\Leftrightarrow y'(t) = (\beta x(t) - \gamma)y(t) \\
 &\Leftrightarrow y'(t) = \left(\frac{\beta}{\gamma}x(t) - 1\right)\gamma y(t) \\
 &\Leftrightarrow y'(t) = \left(\frac{x(t)}{\rho} - 1\right)\gamma y(t) \\
 &\Leftrightarrow y'(t) = \left(\frac{x(t)}{\frac{x_0}{R_0}} - 1\right)\gamma y(t) \\
 &\Leftrightarrow y'(t) = \left(R_0 \frac{x(t)}{x_0} - 1\right)\gamma y(t). \quad (2.2)
 \end{aligned}$$

En posant  $t = 0$  dans l'équation (2.2), on obtient

$$y'(0) = \left(R_0 \frac{x_0}{x_0} - 1\right)\gamma y(0) \Leftrightarrow y'(0) = (R_0 - 1)\gamma y(0). \quad (2.3)$$

Comme  $y(0)$  et  $\gamma$  sont supérieurs à 0, si  $R_0$  est supérieur à 1, alors  $y'(0)$  est supérieur à 0. Cela signifie que le nombre d'individus infectés croît dès le début et donc que l'épidémie sera importante. Par contre, si  $R_0$  est inférieur à 1,  $y'(0)$  est inférieur à 0. Le nombre d'individus infectés décroît donc dès le début jusqu'à ce qu'il atteigne 0, ce qui correspond à une épidémie de faible importance. La valeur 1 représente donc un seuil pour le paramètre  $R_0$ . De plus, la valeur de  $R_0$ , lorsqu'elle est supérieure à 1, permet de quantifier le niveau de sévérité d'une épidémie. Le paramètre  $R_0$  est très important dans l'étude de la propagation des épidémies.

Pour certaines valeurs de  $t$ , il est possible d'exprimer  $\rho$  et  $R_0$  en fonction de  $x(t)$ ,  $y(t)$  et  $z(t)$ . La première valeur d'intérêt est celle pour laquelle  $y(t)$  atteint son maximum. Par la troisième équation de (2.1), on a que la valeur de  $t$  pour laquelle  $y(t)$  atteint son maximum est également celle pour laquelle  $z'(t)$  atteint son maximum. Étant donné que le nombre d'individus infectés à un moment précis n'est pas observé en pratique,

on s'intéresse davantage à  $z'(t)$  qu'à  $y(t)$ . On note par  $t_z$  la valeur de  $t$  à laquelle  $z'(t)$  atteint son maximum. On cherche à exprimer  $\rho$  en fonction de  $x(t_z)$ ,  $y(t_z)$  et  $z(t_z)$ . À partir de la troisième équation de (2.1), on obtient :

$$\frac{d}{dt}z'(t)|_{t_z} = 0 \Leftrightarrow \frac{d}{dt}y(t)|_{t_z} = 0.$$

Ensuite, à partir de la deuxième équation de (2.1), on obtient :

$$\begin{aligned} \frac{d}{dt}y(t)|_{t_z} = 0 &\Leftrightarrow \beta x(t_z)y(t_z) - \gamma y(t_z) = 0 \\ &\Leftrightarrow \beta x(t_z)y(t_z) = \gamma y(t_z) \\ &\Leftrightarrow \frac{\gamma}{\beta} = x(t_z) \Leftrightarrow \rho = x(t_z), \end{aligned} \tag{2.4}$$

car  $y(t_z) > 0$ . La valeur du nombre d'individus susceptibles au moment où  $z'(t)$  atteint son maximum est donc égale à  $\rho$ . L'équation (2.4) est à la base de la méthode d'estimation du paramètre  $\rho$  présentée au chapitre 3. De même, on peut exprimer  $R_0$  en fonction de  $x(t_z)$  de la façon suivante :

$$R_0 = \frac{x_0}{x(t_z)}.$$

La deuxième valeur d'intérêt est celle pour laquelle  $-x'(t)$  atteint son maximum. Soit  $t_x$  la valeur de  $t$  à laquelle  $-x'(t)$  atteint son maximum. On cherche à exprimer  $\rho$  en fonction de  $x(t_x)$ ,  $y(t_x)$  et  $z(t_x)$ . À partir des deux premières équations de (2.1), on obtient :

$$\begin{aligned} \frac{d}{dt}x'(t)|_{t_x} = 0 &\Leftrightarrow \frac{d}{dt} \left[ \beta x(t)y(t) \right] |_{t_x} = 0 \\ &\Leftrightarrow x(t_x)y'(t_x) + x'(t_x)y(t_x) = 0 \\ &\Leftrightarrow x(t_x)y'(t_x) = -x'(t_x)y(t_x) \\ &\Leftrightarrow x(t_x) \left[ \beta x(t_x)y(t_x) - \gamma y(t_x) \right] = \left[ \beta x(t_x)y(t_x) \right] y(t_x) \\ &\Leftrightarrow \beta x^2(t_x)y(t_x) - \gamma x(t_x)y(t_x) = \beta x(t_x)y^2(t_x). \end{aligned}$$

Comme  $x(t_x) > 0$  et  $y(t_x) > 0$ , on peut simplifier et on obtient

$$\begin{aligned}
 \beta x^2(t_x)y(t_x) - \gamma x(t_x)y(t_x) &= \beta x(t_x)y^2(t_x) \Leftrightarrow \beta x(t_x) - \gamma = \beta y(t_x) \\
 &\Leftrightarrow \beta [x(t_x) - y(t_x)] = \gamma \\
 &\Leftrightarrow \frac{\gamma}{\beta} = x(t_x) - y(t_x) \\
 &\Leftrightarrow \rho = x(t_x) - y(t_x). \quad (2.5)
 \end{aligned}$$

La différence entre le nombre d'individus susceptibles et le nombre d'individus infectés au moment où  $-x'(t)$  atteint son maximum est donc égale à  $\rho$ . Il serait également possible de proposer une méthode d'estimation à partir de l'équation (2.5), comme ce qui est fait au chapitre 3 à partir de l'équation (2.4).

Intuitivement, on s'attend à ce que  $t_x < t_z$ . En effet, c'est à l'instant  $t_x$  que la variation du nombre d'individus susceptibles est maximale en valeur absolue. Cela entraîne qu'à l'instant  $t_x$ , un grand nombre d'individus deviennent infectés. Après un certain temps, ces individus deviennent retirés. Il y a donc un grand nombre d'individus qui deviennent retirés à ce moment, ce qui entraîne que la variation du nombre d'individus retirés y est maximale. Comme ce maximum est atteint à l'instant  $t_z$ , on a que  $t_x < t_z$ . Ce résultat est démontré formellement dans la proposition suivante.

**Proposition 2.2.1.** *Soit le système d'équations (2.1). Alors, on a que  $t_x < t_z$ .*

*Démonstration.* À partir des équations (2.4) et (2.5), on a que

$$x(t_z) = \rho \Leftrightarrow x(t_z) = x(t_x) - y(t_x).$$

Comme  $y(t_x) > 0$ , on a que  $x(t_x) > x(t_z)$ . Cela implique que  $t_x < t_z$ , car  $x(t)$  est une fonction strictement décroissante.  $\square$

Au chapitre 3, pour l'estimation du paramètre  $\rho$ , on dispose de données observées à des valeurs entières de temps. Il serait donc convenable que  $t_z$  prenne une valeur entière. Or,  $t_z$  n'est pas nécessairement un nombre entier. Par contre, on peut montrer qu'à

travers une reparamétrisation qui n'affecte pas les valeurs de  $\rho$  et  $R_0$ , on peut supposer que  $t_z$  est un entier. On a, en effet, la propriété suivante.

**Proposition 2.2.2.** *Soit le système d'équations (2.1),  $t_z$  la valeur de  $t$  pour laquelle  $z'(t)$  atteint son maximum et  $\lambda > 0$  tel que  $\frac{t_z}{\lambda}$  est un nombre entier. Alors, on peut trouver une reparamétrisation de (2.1) qui fait en sorte que*

- (i) *la valeur de  $t$  pour laquelle la dérivée du nombre d'individus retirés est maximale est un nombre entier ;*
- (ii) *le paramètre  $\rho$  est le même que celui du système avec la paramétrisation d'origine ;*
- (iii) *les valeurs initiales du nombre d'individus susceptibles, infectés et retirés sont les mêmes que celles du système avec la paramétrisation d'origine.*

*Démonstration.* (i) Soit

$$\begin{aligned} u(t) &= x(\lambda t), \\ v(t) &= y(\lambda t), \\ w(t) &= z(\lambda t). \end{aligned}$$

À partir des trois équations de (2.1), on a que

$$\begin{aligned} \frac{d}{dt}u(t) &= \frac{d}{dt}x(\lambda t) \\ &= \lambda x'(\lambda t) \\ &= \lambda \left[ -\beta x(\lambda t)y(\lambda t) \right] \\ &= -(\lambda\beta)u(t)v(t); \end{aligned} \tag{2.6}$$

$$\begin{aligned} \frac{d}{dt}v(t) &= \frac{d}{dt}y(\lambda t) \\ &= \lambda y'(\lambda t) \\ &= \lambda \left[ \beta x(\lambda t)y(\lambda t) - \gamma y(\lambda t) \right] \\ &= (\lambda\beta)u(t)v(t) - (\lambda\gamma)v(t); \end{aligned} \tag{2.7}$$

$$\begin{aligned}
\frac{d}{dt}w(t) &= \frac{d}{dt}z(\lambda t) \\
&= \lambda z'(\lambda t) \\
&= \lambda [\gamma y(\lambda t)] \\
&= (\lambda \gamma)w(t).
\end{aligned} \tag{2.8}$$

Les valeurs de  $u(t)$ ,  $v(t)$  et  $w(t)$  correspondent donc aux solutions de (2.1) avec les paramètres  $\lambda\beta$  et  $\lambda\gamma$ . Soit  $t_z^*$  la valeur de  $t$  pour laquelle  $w'(t)$  atteint son maximum. Étant donné que  $t_z$  est la valeur de  $t$  pour laquelle  $z'(t)$  atteint son maximum, on a que :

$$\begin{aligned}
w'(t_z^*) = 0 &\Leftrightarrow z'(\lambda t_z^*) = 0 \\
&\Leftrightarrow \lambda t_z^* = t_z \\
&\Leftrightarrow t_z^* = \frac{t_z}{\lambda}.
\end{aligned}$$

Comme  $\lambda$  est tel que  $\frac{t_z}{\lambda}$  est un nombre entier,  $t_z^*$  est un nombre entier.

(ii) Par les équations (2.6), (2.7) et (2.8), on a que les valeurs de  $u(t)$ ,  $v(t)$  et  $w(t)$  sont les solutions du système d'équations différentielles suivant :

$$\begin{cases} u'(t^*) = -(\lambda\beta)u(t)v(t), \\ v'(t^*) = (\lambda\beta)u(t)v(t) - (\lambda\gamma)v(t), \\ w'(t^*) = (\lambda\gamma)v(t), \end{cases} \tag{2.9}$$

Soit  $\rho^*$  la valeur du paramètre  $\rho$  pour le système (2.9). On a que

$$\rho^* = \frac{\lambda\gamma}{\lambda\beta} = \frac{\gamma}{\beta} = \rho.$$

La reparamétrisation n'a donc aucun effet sur la valeur de  $\rho$ .

(iii) Les valeurs initiales de (2.9) sont :

$$\begin{aligned}
u(0) &= x(\lambda \cdot 0) = x(0); \\
v(0) &= y(\lambda \cdot 0) = y(0); \\
w(0) &= z(\lambda \cdot 0) = z(0).
\end{aligned}$$

Les valeurs initiales de (2.9) sont donc les mêmes que celles de (2.1). La reparamétrisation n'a donc pas d'effet sur la valeur de  $R_0$ , étant donné qu'elle n'en a ni sur  $\rho$ , ni sur  $x_0$ .

□

### 2.2.2 Modèle stochastique à partir d'une chaîne de Markov à temps continu à trois dimensions

Le modèle stochastique présenté dans cette section a d'abord été proposé par McKendrick (1926). Certaines propriétés de ce modèle ont par la suite été développées par Bartlett (1949,1955) et Kendall (1956). Kendall qualifie ce modèle d'équivalent stochastique naturel du modèle déterministe SIR, que l'on a présenté à la section 2.2.1.

Soit  $X(t)$  le nombre d'individus susceptibles,  $Y(t)$  le nombre d'individus infectés et  $Z(t)$  le nombre d'individus retirés, chacun au temps  $t \geq 0$ . Soit  $N$  la taille de la population. Dans ce modèle,  $\{(X(t), Y(t), Z(t)), t \geq 0\}$  est une chaîne de Markov à temps continu de dimension 3 telle que :

- (i) l'espace des états est  $\mathbb{N}^3$  ;
- (ii)  $(X(0), Y(0), Z(0)) = (x_0, N - x_0, 0)$  ;
- (iii) le taux de transition de l'état  $(X(t), Y(t), Z(t))$  à l'état  $(X(t) - 1, Y(t) + 1, Z(t))$  est  $\beta X(t)Y(t)$  (*une infection*) ;
- (iv) le taux de transition de l'état  $(X(t), Y(t), Z(t))$  à l'état  $(X(t), Y(t) - 1, Z(t) + 1)$  est  $\gamma Y(t)$  (*un retrait*) ;
- (v) le taux de transition de l'état  $(X(t), Y(t), Z(t))$  à l'état  $(X(t), Y(t), Z(t))$  est

$$1 - [\beta X(t) + \gamma] Y(t)$$

(*pas de changement*) ;

- (vi) le taux de transition de l'état  $(X(t), Y(t), Z(t))$  vers n'importe quel autre état est égal à 0.

On voit que, dans ce modèle, les valeurs de  $X(t)$ ,  $Y(t)$  et  $Z(t)$  sont telles que  $X(t) + Y(t) + Z(t) = N$ , pour tout  $t \geq 0$ , comme dans le modèle déterministe SIR.

### 2.2.3 Modèle stochastique à partir de processus de naissance ou de mort purs non homogènes

Van Den Broek et Heesterbeek (2007) présentent deux modèles stochastiques, soit un modèle pour le nombre d'individus susceptibles et un modèle pour le nombre d'individus infectés. Ces modèles se basent sur la première équation du système (2.1), dans laquelle  $\beta$  est remplacé par  $\beta(t) > 0$ ,  $t \geq 0$ . Dans ce cas, la première équation de (2.1) devient :

$$x'(t) = -\beta(t)x(t)y(t). \quad (2.10)$$

On permet donc au taux d'infection de changer dans le temps. Il est à noter que ces auteurs définissent  $x(t)$ ,  $y(t)$  et  $z(t)$  comme la **proportion** des individus qui sont susceptibles, infectés et retirés dans la population. Étant donné qu'on a défini ces fonctions comme étant le **nombre** d'individus susceptibles, infectés et retirés, les modèles présentés par ces auteurs ont été adaptés en conséquence.

On présente tout d'abord le modèle pour le nombre d'individus susceptibles proposé par Van Den Broek et Heesterbeek. En notant

$$\mu(t) = \beta(t)y(t),$$

on peut réécrire l'équation (2.10) de la façon suivante :

$$x'(t) = -\mu(t)x(t). \quad (2.11)$$

Soit  $X(t)$  le nombre d'individus susceptibles au temps  $t$ . En se basant sur l'équation (2.11), il est proposé que  $\{X(t), t \geq 0\}$  soit un processus de mort pur non homogène tel que

- (i) l'espace des états est  $\mathbb{N}$  ;
- (ii)  $X(0) = x_0$  ;

(iii) le taux de mort à l'état  $X(t)$  et au temps  $t$  est égal à  $\mu(t)X(t)$ .

Dans ce modèle, on a que  $\mathbb{E}[X(t)] = x(t)$ .

On présente ensuite le modèle pour le nombre d'individus infectés proposé par Van Den Broek et Heesterbeek. Dans ce modèle, on suppose qu'on est au début de l'épidémie et que, par conséquent, le nombre d'individus retirés est égal à 0. On a donc

$$x(t) + y(t) = N, \quad (2.12)$$

pour les valeurs de  $t \geq 0$  à l'étude. À partir de l'équation (2.12), on peut réécrire l'équation (2.10) comme suit :

$$\begin{aligned} \frac{d}{dt}x(t) = -\beta(t)x(t)y(t) &\Leftrightarrow \frac{d}{dt}[N - y(t)] = -\beta(t)[N - y(t)]y(t) \\ &\Leftrightarrow y'(t) = \lambda(t)y(t), \end{aligned} \quad (2.13)$$

où

$$\lambda(t) = \beta(t)[N - y(t)].$$

Soit  $Y(t)$  le nombre d'individus infectés au temps  $t$ . En se basant sur l'équation (2.13), il est proposé que  $\{Y(t), t \geq 0\}$  soit un processus de naissance pur non homogène tel que

(i) l'espace des états est  $\mathbb{N}$  ;

(ii)  $Y(0) = y_0$  ;

(iii) le taux de naissance à l'état  $Y(t)$  et au temps  $t$  est égal à  $\lambda(t)Y(t)$ .

Dans ce modèle, on a que  $\mathbb{E}[Y(t)] = y(t)$ . On peut montrer que les fonctions  $\mu(t)$  et  $\lambda(t)$  sont reliées de la façon suivante :

$$\mu(t) = \frac{y(t)}{N - y(t)}\lambda(t).$$

On remarque que dans le modèle pour le nombre d'individus infectés,  $Y(t)$  peut être supérieur à  $N$ , car  $\{Y(t), t \geq 0\}$  est un processus de naissance pur. Contrairement au modèle présenté à la section 2.2.2, on n'a donc pas que  $X(t) + Y(t) + Z(t) = N$ . Cela ne pose toutefois pas de problème à ces auteurs, étant donné qu'ils n'utilisent pas cette forme générale de leur modèle pour l'estimation des paramètres.

## 2.3 Modèles stochastiques proposés

Les deux modèles proposés dans ce mémoire tiennent compte du fait qu'en pratique, on a des données sur le nombre d'individus susceptibles et décédés, mais pas sur le nombre d'individus infectés à des instants précis. Le nombre d'individus infectés ne peut donc pas être utilisé pour l'estimation des paramètres. Un autre but de cette modélisation est d'obtenir des processus dont l'espérance est égale à une des solutions du modèle déterministe SIR, qui sont données par (2.1). Ainsi, soit  $X(t)$  le nombre d'individus susceptibles,  $Z(t)$  le nombre d'individus retirés et  $D(t)$  le nombre d'individus décédés, chacun au temps  $t \geq 0$ . Les modèles proposés s'expriment en fonction de  $X(t)$ ,  $Z(t)$  et  $D(t)$ .

### 2.3.1 Premier modèle proposé

Le premier modèle proposé s'inspire du modèle de Kendall (1956), présenté à la section 2.2.2. On le réduit cependant à deux processus stochastiques, car  $Y(t)$  n'apparaît pas explicitement. L'idée est la suivante : chaque fois que  $X(t)$  décroît d'une unité,  $Y(t)$  croît d'une unité, et, de façon similaire, chaque fois que  $Z(t)$  croît d'une unité,  $Y(t)$  décroît d'une unité. En vertu de ces relations, on propose le modèle suivant pour le processus  $\{(X(t), Z(t)), t \geq 0\}$  :

**Modèle 2.3.1.**  $\{(X(t), Z(t)), t \geq 0\}$  est une chaîne de Markov à temps continu non homogène de dimension 2 telle que

- (i) l'espace des états est  $\mathbb{N}^2$  ;
- (ii)  $(X(0), Z(0)) = (x_0, 0)$  ;
- (iii) le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t) - 1, Z(t))$  est  $\beta y(t)X(t)$  (une infection) ;
- (iv) le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t), Z(t) + 1)$  est  $\gamma y(t)$  (un retrait) ;

(v) le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t), Z(t))$  est

$$1 - [\beta X(t) + \gamma]y(t)$$

(pas de changement);

(vi) le taux de transition de l'état  $(X(t), Z(t))$  vers n'importe quel autre état est égal à 0;

(vii) les limites pour le calcul des taux de transition à partir de l'état  $(j, k)$  sont uniformes en  $(j, k)$ .

Dans ce modèle, on voit que les taux de transition du processus  $\{X(t), t \geq 0\}$  ne dépendent pas de  $Z(t)$  et ceux de  $\{Z(t), t \geq 0\}$  ne dépendent pas de  $X(t)$ . Les processus  $\{X(t), t \geq 0\}$  et  $\{Z(t), t \geq 0\}$  sont décrits en détail dans ce qui suit.

**Proposition 2.3.1.** *Soit le modèle 2.3.1. Alors,  $\{X(t), t \geq 0\}$  est un processus de mort pur de taux  $\mu_X(t) = \beta y(t)X(t)$ .*

*Démonstration.* Soit  $k \in \mathbb{N}$  tel que  $0 < k \leq x_0$ . Afin de montrer que  $\{X(t), t \geq 0\}$  est un processus de mort pur de taux  $\mu_X(t) = \beta y(t)X(t)$ , on doit montrer que

$$\lim_{h \rightarrow 0} \frac{P[X(t+h) = k-1 | X(t) = k]}{h} = \beta y(t)k,$$

$$\lim_{h \rightarrow 0} \frac{P[X(t+h) = k | X(t) = k]}{h} = 1 - \beta y(t)k$$

et

$$\lim_{h \rightarrow 0} \frac{P[X(t+h) = j | X(t) = k]}{h} = 0,$$

où  $j \notin \{k-1, k\}$ . On évalue d'abord

$$P[X(t+h) = k-1 | X(t) = k] = \frac{P[X(t+h) = k-1, X(t) = k]}{P[X(t) = k]}. \quad (2.14)$$

Étant donné que le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t) - 1, Z(t))$  est égal à  $\beta y(t)X(t)$  et que le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t) - 1, Z(t) + m)$  est nul pour tout  $m \neq 0$ , on écrit le numérateur de l'équation (2.14) comme suit :

$$\begin{aligned} & P[X(t+h) = k-1, X(t) = k] \\ &= \sum_{\substack{j \in \mathbb{N} \\ j'=j}} P[X(t+h) = k-1, X(t) = k, Z(t+h) = j', Z(t) = j] \\ &+ \sum_{\substack{j \in \mathbb{N} \\ j' \neq j}} P[X(t+h) = k-1, X(t) = k, Z(t+h) = j', Z(t) = j] \\ &= \sum_{j \in \mathbb{N}} P[X(t+h) = k-1, Z(t+h) = j | X(t) = k, Z(t) = j] \cdot \\ &\quad P[X(t) = k, Z(t) = j] \\ &+ \sum_{\substack{j \in \mathbb{N} \\ j' \neq j}} P[X(t+h) = \bar{k}-1, \bar{Z}(t+h) = j' | X(t) = \bar{k}, \bar{Z}(t) = j] \cdot \\ &\quad P[X(t) = \bar{k}, \bar{Z}(t) = j]. \end{aligned}$$

On pose

$$\begin{aligned} & P[X(t+h) = k-1, Z(t+h) = j | X(t) = k, Z(t) = j] \\ &\approx P[X(t+h) = k-1, Z(t+h) = 1 | X(t) = k, Z(t) = 1], \quad (2.15) \end{aligned}$$

pour tout  $j \in \mathbb{N}$ , car lorsque  $h$  tend vers 0, les deux membres de l'équation (2.15) tendent vers  $\beta y(t)k$  et les limites sont uniformes en  $j$ . De même, on pose

$$\begin{aligned} & P[X(t+h) = k-1, Z(t+h) = j' | X(t) = k, Z(t) = j] \\ &\approx P[X(t+h) = k-1, Z(t+h) = 2 | X(t) = k, Z(t) = 1]. \end{aligned}$$

pour tout  $j \in \mathbb{N}$ ,  $j' \neq j$ . On a donc que

$$\begin{aligned}
& P[X(t+h) = k-1, X(t) = k] \\
& \approx P[X(t+h) = k-1, Z(t+h) = 1 | X(t) = k, Z(t) = 1] \cdot \\
& \quad \sum_{j \in \mathbb{N}} P[X(t) = k, Z(t) = j] \\
& + P[X(t+h) = k-1, Z(t+h) = 2 | X(t) = k, Z(t) = 1] \cdot \\
& \quad \sum_{j \in \mathbb{N}} P[X(t) = k, Z(t) = j] \\
& = P[X(t+h) = k-1, Z(t+h) = 1 | X(t) = k, Z(t) = 1] P[X(t) = k] \\
& + P[X(t+h) = k-1, Z(t+h) = 2 | X(t) = k, Z(t) = 1] P[X(t) = k]. \quad (2.16)
\end{aligned}$$

À partir de l'équation (2.16), on peut réécrire l'équation (2.14) comme suit :

$$\begin{aligned}
& P[X(t+h) = k-1 | X(t) = k] \\
& = \frac{P[X(t+h) = k-1, X(t) = k]}{P[X(t) = k]} \\
& = P[X(t+h) = k-1, Z(t+h) = 1 | X(t) = k, Z(t) = 1] \\
& + P[X(t+h) = k-1, Z(t+h) = 2 | X(t) = k, Z(t) = 1],
\end{aligned}$$

ce qui implique que

$$\begin{aligned}
& \lim_{h \rightarrow 0} \frac{P[X(t+h) = k-1 | X(t) = k]}{h} \\
& = \lim_{h \rightarrow 0} \frac{P[X(t+h) = k-1, Z(t+h) = 1 | X(t) = k, Z(t) = 1]}{h} \\
& + \lim_{h \rightarrow 0} \frac{P[X(t+h) = k-1, Z(t+h) = 2 | X(t) = k, Z(t) = 1]}{h} \\
& = \beta y(t)k + 0 \\
& = \beta y(t)k.
\end{aligned}$$

Ensuite, on évalue

$$P[X(t+h) = k | X(t) = k] = \frac{P[X(t+h) = k, X(t) = k]}{P[X(t) = k]}. \quad (2.17)$$

Étant donné que le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t), Z(t) + 1)$  est égal à  $\gamma y(t)$ , que le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t), Z(t))$  est

égal à  $1 - \beta y(t)X(t) - \gamma y(t)$  et que le taux de transition de l'état  $(X(t), Z(t))$  à l'état  $(X(t), Z(t) + m)$  est nul pour tout  $m \notin \{0, 1\}$ , on écrit le numérateur de l'équation (2.16) comme suit :

$$\begin{aligned}
& P[X(t+h) = k, X(t) = k] \\
&= \sum_{\substack{j \in \mathbb{N} \\ j'=j}} P[X(t+h) = k, X(t) = k, Z(t+h) = j', Z(t) = j] \\
&\quad + \sum_{\substack{j \in \mathbb{N} \\ j'=j+1}} P[X(t+h) = k, X(t) = k, Z(t+h) = j', Z(t) = j] \\
&\quad + \sum_{\substack{j \in \mathbb{N} \\ j' \notin \{j, j+1\}}} P[X(t+h) = k, X(t) = k, Z(t+h) = j', Z(t) = j] \\
&= \sum_{j \in \mathbb{N}} P[X(t+h) = k, Z(t+h) = j | X(t) = k, Z(t) = j] \cdot \\
&\quad P[X(t) = k, Z(t) = j] \\
&\quad + \sum_{\substack{j \in \mathbb{N} \\ j'=j+1}} P[X(t+h) = k, Z(t+h) = j+1 | X(t) = k, Z(t) = j] \cdot \\
&\quad P[X(t) = k, Z(t) = j] \\
&\quad + \sum_{\substack{j \in \mathbb{N} \\ j' \notin \{j, j+1\}}} P[X(t+h) = k, Z(t+h) = j' | X(t) = k, Z(t) = j] \cdot \\
&\quad P[X(t) = k, Z(t) = j].
\end{aligned}$$

Comme précédemment, on isole

$$\sum_{j \in \mathbb{N}} P[X(t) = k, Z(t) = j]$$

qui est égal à

$$P[X(t) = k]$$

et on obtient

$$\begin{aligned}
& P[X(t+h) = k | X(t) = k] \\
&= \frac{P[X(t+h) = k, X(t) = k]}{P[X(t) = k]} \\
&\approx P[X(t+h) = k, Z(t+h) = 1 | X(t) = k, Z(t) = 1] \\
&\quad + P[X(t+h) = k, Z(t+h) = 2 | X(t) = k, Z(t) = 1] \\
&\quad + P[X(t+h) = k, Z(t+h) = 3 | X(t) = k, Z(t) = 1]. \tag{2.18}
\end{aligned}$$

L'équation (2.18) implique que

$$\begin{aligned}
& \lim_{h \rightarrow 0} \frac{P[X(t+h) = k | X(t) = k]}{h} \\
&= \lim_{h \rightarrow 0} \frac{P[X(t+h) = k, Z(t+h) = 1 | X(t) = k, Z(t) = 1]}{h} \\
&\quad + \lim_{h \rightarrow 0} \frac{P[X(t+h) = k, Z(t+h) = 2 | X(t) = k, Z(t) = 1]}{h} \\
&\quad + \lim_{h \rightarrow 0} \frac{P[X(t+h) = k, Z(t+h) = 3 | X(t) = k, Z(t) = 1]}{h} \\
&= [1 - \beta y(t)k - \gamma y(t)] + \gamma y(t) + 0 \\
&= 1 - \beta y(t)k.
\end{aligned}$$

Enfin, étant donné que

$$\begin{aligned}
& \sum_{\substack{m=1 \\ m \notin \{k-1, k\}}}^{x_0} P[X(t+h) = m, X(t) = k] \\
&= 1 - P[X(t+h) = k-1, X(t) = k] - P[X(t+h) = k, X(t) = k],
\end{aligned}$$

on a que

$$\begin{aligned}
& \lim_{h \rightarrow 0} \sum_{\substack{m=1 \\ m \notin \{k-1, k\}}}^{x_0} \frac{P[X(t+h) = m | X(t) = k]}{h} \\
&= 1 - \lim_{h \rightarrow 0} \frac{P[X(t+h) = k-1 | X(t) = k]}{h} - \lim_{h \rightarrow 0} \frac{P[X(t+h) = k | X(t) = k]}{h} \\
&= 1 - \beta k y(t) - (1 - \beta k y(t)) \\
&= 0.
\end{aligned}$$

Cela donne

$$\lim_{h \rightarrow 0} \frac{P[X(t+h) = m | X(t) = k]}{h} = 0,$$

pour  $m \notin \{k-1, k\}$ .

Ainsi, on conclut que  $\{X(t), t \geq 0\}$  est un processus de mort pur de taux  $\mu_X(t) = \beta y(t)X(t)$ .  $\square$

Dans ce qui suit, on étudie le processus  $\{X(t), t \geq 0\}$  et on prouve, entre autres, que son espérance est égale à  $x(t)$ . Les propriétés du processus présentées à la proposition 2.3.2 sont utilisées pour l'estimation des paramètres.

**Proposition 2.3.2.** *Soit  $\{X(t), t \geq 0\}$  un processus de mort pur avec  $\mu_X(t) = \beta y(t)X(t)$ ,  $X(0) = x_0$  et soit  $0 \leq s < t$ . Alors*

(i) *la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est une loi binomiale  $\mathcal{B}\left(X(s), \frac{x(t)}{x(s)}\right)$ ;*

(ii)  *$X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(t)}{x_0}\right)$ ;*

(iii)  $\mathbb{E}[X(t)] = x(t)$ ;

(iv)  $\text{Var}[X(t)] = \frac{x(t)}{x_0} \left(x_0 - x(t)\right)$ ;

(v)  $X(s) - X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(s) - x(t)}{x_0}\right)$ ;

(vi)  $\text{Cov}[X(s), X(t)] = \frac{x(t)}{x(s)} \text{Var}[X(s)]$ .

*Démonstration.* (i) De la proposition 1.4.1, la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est une loi binomiale  $\mathcal{B}(X(s), \exp\{\eta(s) - \eta(t)\})$ , où  $\eta(t) = \int_0^t \beta y(\tau) d\tau$ . De la première équation de (2.1), on a que

$$\begin{aligned} x'(\tau) = -\beta x(\tau)y(\tau) &\Leftrightarrow \beta y(\tau) = -\frac{x'(\tau)}{x(\tau)} \\ &\Leftrightarrow \int_0^t \beta y(\tau) d\tau = \log\{x_0\} - \log\{x(t)\}. \end{aligned}$$

On a donc que

$$\begin{aligned} \exp\{\eta(s) - \eta(t)\} &= \exp\{\log\{x_0\} - \log\{x(s)\} - \log\{x_0\} + \log\{x(t)\}\} \\ &= \exp\left[\log\left(\frac{x(t)}{x(s)}\right)\right] \\ &= \frac{x(t)}{x(s)}. \end{aligned}$$

Donc, la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est une loi binomiale  $\mathcal{B}\left(X(s), \frac{x(t)}{x(s)}\right)$ .

(ii) Comme  $X(0) = x_0$  avec probabilité 1, on a que

$$\begin{aligned} P[X(t) = k] &= \sum_{m=0}^{x_0} P[X(t) = k | X(0) = m] P[X(0) = m] \\ &= P[X(t) = k | X(0) = x_0] \\ &= \frac{x_0!}{k!(x_0 - k)!} \left(\frac{x(t)}{x_0}\right)^k \left(1 - \frac{x(t)}{x_0}\right)^{x_0 - k}. \end{aligned}$$

Donc,  $X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(t)}{x_0}\right)$ .

(iii) Comme  $X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(t)}{x_0}\right)$ , on a que

$$\mathbb{E}[X(t)] = x_0 \frac{x(t)}{x_0} = x(t).$$

(iv) Comme  $X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(t)}{x_0}\right)$ , on a que

$$\begin{aligned} \text{Var}[X(t)] &= x_0 \frac{x(t)}{x_0} \left(1 - \frac{x(t)}{x_0}\right) \\ &= \frac{x(t)}{x_0} (x_0 - x(t)). \end{aligned}$$

(v) Comme la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est binomiale  $\mathcal{B}\left(X(s), \frac{x(t)}{x(s)}\right)$ , on a que la loi conditionnelle de  $X(s) - X(t)$  sachant  $X(s)$  est une loi binomiale  $\mathcal{B}\left(X(s), 1 - \frac{x(t)}{x(s)}\right)$ . De plus,  $X(s)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(s)}{x_0}\right)$ . Par la proposition 1.2.1, on a que  $X(s) - X(t)$  est de loi binomiale  $\mathcal{B}(x_0, \tilde{p})$  où

$$\begin{aligned} \tilde{p} &= \left(\frac{x(s)}{x_0}\right) \left(1 - \frac{x(t)}{x(s)}\right) \\ &= \left(\frac{x(s)}{x_0}\right) \left(\frac{x(s) - x(t)}{x(s)}\right) \\ &= \frac{x(s) - x(t)}{x_0}. \end{aligned}$$

Donc  $X(s) - X(t)$  est de loi binomiale  $\mathcal{B}\left(x_0, \frac{x(s) - x(t)}{x_0}\right)$ .

(vi) Si  $0 \leq s < t$ , la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est une loi binomiale  $\mathcal{B}\left(X(s), \frac{x(t)}{x(s)}\right)$ . On a donc que

$$\begin{aligned}
\text{Cov}[X(s), X(t)] &= \mathbb{E}[X(s)X(t)] - \mathbb{E}[X(s)]\mathbb{E}[X(t)] \\
&= \mathbb{E}\left(\mathbb{E}[X(s)X(t)|X(s)]\right) - \mathbb{E}[X(s)]\mathbb{E}\left(\mathbb{E}[X(t)|X(s)]\right) \\
&= \mathbb{E}\left(X(s)\mathbb{E}[X(t)|X(s)]\right) - \mathbb{E}[X(s)]\mathbb{E}\left(X(s)\frac{x(t)}{x(s)}\right) \\
&= \mathbb{E}\left(\frac{x(t)}{x(s)}X^2(s)\right) - \frac{x(t)}{x(s)}\left(\mathbb{E}[X(s)]\right)^2 \\
&= \frac{x(t)}{x(s)}\left[\mathbb{E}[X^2(s)] - \left(\mathbb{E}[X(s)]\right)^2\right] \\
&= \frac{x(t)}{x(s)}\text{Var}[X(s)].
\end{aligned}$$

□

Pour le modèle 2.3.1, on peut montrer que  $\{Z(t), t \geq 0\}$  est un processus de Poisson non-homogène d'intensité  $\gamma y(t)$ . Étant donné que la preuve est semblable à celle de la proposition 2.3.1, on ne la présente pas. Soit  $p$  la probabilité qu'un individu retiré soit en fait décédé et soit  $V_j$  une variable aléatoire qui prend la valeur 1 si le  $j^{\text{ième}}$  individu retiré est décédé et qui prend la valeur 0 sinon.  $V_j$  est donc de loi binomiale  $\mathcal{B}(1, p)$ . On suppose que les  $V_j$  sont des variables indépendantes. On propose que  $\{D(t), t \geq 0\}$  soit un processus de Poisson composé tel que

$$D(t) = \begin{cases} \sum_{j=1}^{Z(t)} V_j, & Z(t) > 0 \\ 0, & \text{sinon.} \end{cases}$$

Par la proposition 1.3.2,  $\{D(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\gamma y(t)$ .

**Proposition 2.3.3.** *Soit  $\{D(t), t \geq 0\}$  un processus de Poisson non homogène d'intensité  $p\gamma y(t)$  et soit  $0 \leq s < t \leq u < v$ . Alors*

(i)  $D(t) - D(s)$  et  $D(v) - D(u)$  sont indépendants ;

- (ii)  $D(t) - D(s)$  est de loi de Poisson  $\mathcal{P}(p\{z(t) - z(s)\})$  ;
- (iii)  $D(t)$  est de loi de Poisson  $\mathcal{P}(pz(t))$  ;
- (iv)  $\mathbb{E}[D(t)] = \text{Var}[D(t)] = pz(t)$  ;
- (v)  $\text{Cov}[D(s), D(t)] = \text{Var}[D(s)]$ .

*Démonstration.* (i) Comme  $\{D(t), t \geq 0\}$  est un processus de Poisson,  $D(t) - D(s)$  et  $D(v) - D(u)$  sont indépendants ;

- (ii) Comme  $\{D(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\gamma y(t)$ ,  $D(t) - D(s)$  est de loi de Poisson  $\mathcal{P}(m(t) - m(s))$ , où  $m(t) = \int_0^t p\gamma y(\tau) d\tau$ . À partir de la troisième équation de (2.1), on a que

$$\begin{aligned}
 m(t) &= \int_0^t p\gamma y(\tau) d\tau \\
 &= p \int_0^t z'(\tau) d\tau \\
 &= p(z(t) - z(0)) \\
 &= pz(t),
 \end{aligned}$$

car  $z(0) = 0$ . Donc,  $D(t) - D(s)$  est de loi de Poisson  $\mathcal{P}(p\{z(t) - z(s)\})$ .

- (iii) Comme  $\{D(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\gamma y(t)$ ,  $D(t)$  est de loi de Poisson  $\mathcal{P}(m(t))$ , où  $m(t) = pz(t)$ . Donc,  $D(t)$  est de loi de Poisson  $\mathcal{P}(pz(t))$ .
- (iv) Comme  $D(t)$  est de loi de Poisson  $\mathcal{P}(pz(t))$ , on a que  $\mathbb{E}[D(t)] = \text{Var}[D(t)] = pz(t)$  ;
- (v) Comme  $\{D(t), t \geq 0\}$  est un processus de Poisson, par la proposition 1.3.1, on a que

$$\text{Cov}[D(s), D(t)] = \text{Var}[D(s)].$$

□

Pour l'estimation du paramètre  $\rho$  et pour les simulations, on a plutôt considéré qu'on disposait de deux processus indépendants,  $\{X(t), t \geq 0\}$  et  $\{Z(t), t \geq 0\}$ , tels que

décrits plus haut. Ainsi, on permettrait qu'il y ait simultanément une infection et un retrait avec une probabilité non nulle. Comme les taux de transition ne dépendent plus du nombre d'individus infectés au temps  $t$ , noté  $Y(t)$ , mais seulement du nombre moyen d'individus infectés,  $y(t)$ , cette supposition pourrait être raisonnable dans notre cas. De plus, il est assez réaliste de permettre la simultanéité de ces événements. Par contre, avec ce modèle, si on définit  $Y(t) = N - X(t) - Z(t)$ , alors il n'est pas exclu que  $Y(t) < 0$ . Cependant, il est raisonnable de croire que cela se produit à la fin de l'épidémie, c'est-à-dire pour de grandes valeurs de  $t$ .

### 2.3.2 Deuxième modèle proposé

Le deuxième modèle proposé est une version simplifiée du modèle présenté à la section 2.3.1. En fait, on se réduit à deux processus de Poisson non homogènes, que l'on suppose indépendants pour la simplicité. On considère toujours les couples  $(X(t), Z(t))$ , où  $\{Z(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $\gamma y(t)$ , comme pour le modèle 2.3.1. De la même façon que pour le modèle 2.3.1, on considère que  $\{D(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $p\gamma y(t)$ .

Pour modéliser le nombre d'individus susceptibles, on considère d'abord  $\{N_X(t), t \geq 0\}$ , un processus de Poisson non homogène d'intensité  $\beta x(t)y(t)$  qui compte le nombre total d'individus infectés par la maladie jusqu'au temps  $t$ . On a donc que le nombre d'individus susceptibles au temps  $t$  est donné par

$$X(t) = x_0 - N_X(t), \quad (2.19)$$

$t \geq 0$ . Dans ce qui suit, on présente certaines propriétés du processus  $\{X(t), t \geq 0\}$ . Les propriétés du processus présentées à la proposition 2.3.4 sont utilisées pour l'estimation des paramètres.

**Proposition 2.3.4.** *Soit  $\{N_X(t), t \geq 0\}$  un processus de Poisson non homogène d'intensité  $\beta x(t)y(t)$ , soit  $X(t) = x_0 - N_X(t)$  et soit  $0 \leq s < t \leq u < v$ . Alors*

- (i)  $X(s) - X(t)$  et  $X(u) - X(v)$  sont indépendants ;

- (ii)  $X(s) - X(t)$  est de loi de Poisson  $\mathcal{P}(x(s) - x(t))$ ;
- (iii)  $\mathbb{E}[X(t)] = x(t)$ ;
- (iv)  $\text{Var}[X(t)] = x_0 - x(t)$ ;
- (v)  $\text{Cov}[X(s), X(t)] = \text{Var}[X(s)]$ .

*Démonstration.* (i) Étant donné que  $X(t) = x_0 - N_X(t)$ , on a que

$$\begin{aligned} N_X(t) - N_X(s) &= x_0 - X(t) - [x_0 - X(s)] \\ &= X(s) - X(t). \end{aligned}$$

De même, on a que

$$N_X(v) - N_X(u) = X(u) - X(v).$$

Comme  $\{N_X(t), t \geq 0\}$  est un processus de Poisson non homogène,  $N_X(t) - N_X(s)$  et  $N_X(v) - N_X(u)$  sont indépendants. On a donc que  $X(s) - X(t)$  et  $X(u) - X(v)$  sont indépendants.

- (ii) Comme  $\{N_X(t), t \geq 0\}$  est un processus de Poisson non homogène d'intensité  $\beta x(t)y(t)$ ,  $N_X(t) - N_X(s)$  est de loi de Poisson  $\mathcal{P}(m(t) - m(s))$ , où  $m(t) = \int_0^t \beta x(\tau)y(\tau)d\tau$ . À partir de la première équation de (2.1), on a que

$$\begin{aligned} m(t) &= \int_0^t \beta x(\tau)y(\tau)d\tau \\ &= - \int_0^t x'(\tau)d\tau \\ &= -x(t). \end{aligned}$$

Donc,  $N_X(t) - N_X(s)$  est de loi de Poisson  $\mathcal{P}(x(s) - x(t))$ . Comme

$$X(s) - X(t) = N_X(t) - N_X(s),$$

on a que  $X(s) - X(t)$  est de loi de Poisson  $\mathcal{P}(x(s) - x(t))$ .

- (iii) Comme  $X(0) - X(t)$  est de loi de Poisson  $\mathcal{P}(x(0) - x(t))$  et comme  $X(0) = x_0$  avec probabilité 1, on a que

$$\begin{aligned} \mathbb{E}[X(0) - X(t)] = x(0) - x(t) &\Leftrightarrow x_0 - \mathbb{E}[X(t)] = x_0 - x(t) \\ &\Leftrightarrow \mathbb{E}[X(t)] = x(t). \end{aligned}$$

(iv) Comme  $X(0) - X(t)$  est de loi de Poisson  $\mathcal{P}(x(0) - x(t))$  et comme  $X(0) = x_0$  avec probabilité 1, on a que

$$\begin{aligned}\text{Var}[X(0) - X(t)] = x_0 - x(t) &\Leftrightarrow \text{Var}[x_0 - X(t)] = x_0 - x(t) \\ &\Leftrightarrow \text{Var}[X(t)] = x_0 - x(t),\end{aligned}$$

(v) On sait que  $X(0) - X(s)$  et  $X(s) - X(t)$  sont indépendants. Cela entraîne que

$$\begin{aligned}\text{Cov}[X(s), X(t)] &= \text{Cov}[X(s), X(s) + \{X(t) - X(s)\}] \\ &= \text{Cov}[X(s), X(s)] \\ &\quad + \text{Cov}[X(0) - (X(0) - X(s)), -(X(s) - X(t))] \\ &= \text{Cov}[X(s), X(s)] - \text{Cov}[X(0), X(s) - X(t)] \\ &\quad + \text{Cov}[X(0) - X(s), X(s) - X(t)] \\ &= \text{Var}[X(s)],\end{aligned}$$

car  $X(0) = x_0$  avec probabilité 1.

□

À la proposition 2.3.4, on a montré que l'espérance de  $X(t)$  dans le deuxième modèle proposé est égale à  $x(t)$ , comme dans le premier modèle proposé. Il reste que, cette fois-ci, la probabilité que  $X(t) < 0$  est non nulle. Par contre, cette probabilité est plutôt faible lorsque  $x(t)$  est suffisamment grand, c'est-à-dire avant d'arriver vers la fin de l'épidémie. L'avantage de cette approche vient du fait que les deux processus,  $\{N_X(t), t \geq 0\}$  et  $\{Z(t), t \geq 0\}$ , sont du même type, c'est-à-dire des processus de Poisson.

À la proposition 2.3.2, on a montré que sur un intervalle de temps, la variable comptant le nombre de nouveaux individus infectés est de loi binomiale dans le premier modèle proposé, alors qu'à la proposition 2.3.4, on a montré qu'elle est de loi de Poisson dans le second modèle. Dans la suite, on désigne par **modèle avec la loi binomiale** le premier modèle proposé, alors qu'on désigne par **modèle avec la loi de Poisson** le deuxième

modèle proposé. Rappelons que pour les deux modèles, la variable comptant le nombre de nouveaux individus décédés dans un intervalle de temps est de loi de Poisson.

—

---

---

—

## CHAPITRE III

### ESTIMATION DE $R_0$

Ce chapitre traite de l'estimation du paramètre  $R_0$ . On commence par présenter les suppositions sur les données observées. Dans un second temps, on introduit la procédure d'estimation de  $R_0$  qui se base sur l'estimation du paramètre  $\rho$ . Par la suite, on propose une méthode d'estimation de  $\rho$ . Enfin, à partir des estimateurs de  $\rho$  et  $R_0$ , on construit des intervalles de confiance asymptotiques pour  $\rho$  et  $R_0$ .

#### 3.1 Suppositions au niveau des observations

On suppose tout d'abord que l'on a des observations pour  $k$  épidémies, où  $k > 1$ , et que les données de deux épidémies différentes sont indépendantes. Par exemple, chacune des  $k$  épidémies peut correspondre à une épidémie de grippe pour une année donnée. Pour chacune de ces épidémies, on suppose que l'on a des observations à  $n + 1$  instants, les  $n$  intervalles de temps entre les observations étant de même longueur. Pour simplifier la notation, l'unité de temps utilisée est la longueur de l'intervalle de temps entre deux observations. On a donc des observations aux instants  $0, 1, \dots, n$ . Par exemple,  $x(1)$  représente le nombre d'individus susceptibles dans le modèle déterministe au deuxième instant pour lequel on a des observations, alors que  $x(1, 5)$  représente le nombre d'individus susceptibles dans le modèle déterministe au milieu de l'intervalle de temps entre le deuxième et le troisième instant pour lequel on a des observations.

Soit  $X_i(t)$  le nombre d'individus susceptibles et  $D_i(t)$  le nombre d'individus décédés

pour la  $i^{\text{ième}}$  épidémie au temps  $t > 0$ ,  $i = 1, 2, \dots, k$ . Pour chacune des épidémies, on suppose que l'on connaît le nombre initial d'individus susceptibles, soit  $X_i(0)$ , et que  $X_i(0) = x_0$ ,  $i = 1, 2, \dots, k$ . De plus, on suppose que le nombre initial d'individus décédés est nul pour chacune des épidémies, c'est-à-dire que  $D_i(0) = 0$ ,  $i = 1, 2, \dots, k$ . Enfin, pour les instants 1 à  $n$ , on suppose que l'on connaît le nombre total d'individus infectés jusqu'à l'instant  $\tau$ , soit  $X_i(0) - X_i(\tau)$ , et le nombre d'individus décédés jusqu'à l'instant  $\tau$ , soit  $D_i(\tau)$ ,  $\tau = 1, 2, \dots, n$ ,  $i = 1, 2, \dots, k$ . Ces suppositions correspondent aux données disponibles en pratique.

### 3.2 Procédure d'estimation de $R_0$

La procédure d'estimation de  $R_0$  proposée comporte trois étapes. Tout d'abord, un estimateur de  $\rho$  est obtenu pour chacune des  $k$  épidémies. Ces estimateurs sont notés  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_k$ . Ensuite, un nouvel estimateur de  $\rho$ , noté  $\bar{\rho}$ , est obtenu en calculant la moyenne des  $k$  estimateurs obtenus précédemment. Enfin, on obtient l'estimateur de  $R_0$ , noté  $\tilde{R}_0$ , de la façon suivante :

$$\tilde{R}_0 = \frac{x_0}{\bar{\rho}}. \quad (3.1)$$

### 3.3 Méthode d'estimation de $\rho$ proposée

La méthode d'estimation de  $\rho$  proposée se base sur l'équation (2.4), présentée à la section 2.2.1. Rappelons que  $t_z$  représente l'instant auquel  $z'(t)$  atteint son maximum et que l'on avait obtenu la relation suivante :

$$\rho = x(t_z).$$

L'idée de cette méthode d'estimation de  $\rho$  est d'estimer  $\rho$  par le nombre d'individus susceptibles à l'instant  $t_z$ . On propose trois estimateurs de  $\rho$  basés sur cette idée.

Soit  $X(t)$  le nombre d'individus susceptibles au temps  $t$  dans le modèle stochastique. Rappelons que deux modèles stochastiques ont été proposés à la section 2.3, soit le

modèle avec la loi binomiale et le modèle avec la loi de Poisson. L'espérance de  $X(t)$  est la même dans les deux modèles, alors que la variance de  $X(t)$  est différente. Comme l'expression des estimateurs de  $\rho$  proposés dépend du nombre d'individus susceptibles, on a une seule expression pour l'espérance de l'estimateur, alors que l'on a une expression différente pour la variance des estimateurs pour chaque modèle. Il est à noter que pour chacun des estimateurs de  $\rho$  présentés dans cette section, le développement menant à l'expression de la variance de l'estimateur de  $\rho$  est présenté uniquement pour le modèle avec la loi de Poisson, étant donné que le développement pour le modèle avec la loi binomiale est similaire, mais un peu plus long.

Tel que mentionné à la section 3.1, les seuls instants pour lesquels on a des observations sont les nombres entiers de 0 à  $n$  inclusivement. À la proposition 2.2.2, on a montré qu'on peut choisir une paramétrisation en  $\beta$  et  $\gamma$  qui est telle que  $t_z$  soit un nombre entier. À partir d'ici, on suppose que l'on a une telle paramétrisation et, donc, que  $t_z$  prend une valeur entière.

### 3.3.1 Estimation de $\rho$ avec $t_z$ connu

Supposons d'abord que l'on connaisse  $t_z$ . Soit  $\hat{\rho}_i(t_z)$  l'estimateur de  $\rho$  pour la  $i^{\text{ième}}$  épidémie basé sur  $t_z$ ,  $i = 1, 2, \dots, k$ . On estime  $\rho$  pour la  $i^{\text{ième}}$  épidémie en prenant la valeur du nombre d'individus susceptibles observés à l'instant  $t_z$ , c'est-à-dire

$$\hat{\rho}_i(t_z) = X_i(t_z), \quad (3.2)$$

$i = 1, 2, \dots, k$ .

**Proposition 3.3.1.** *On a les propriétés suivantes pour  $\hat{\rho}_i(t_z)$ ,  $i = 1, 2, \dots, k$  :*

(i) *Pour les deux modèles, l'espérance de  $\hat{\rho}_i(t_z)$  est donnée par*

$$\mathbb{E}[\hat{\rho}_i(t_z)] = \rho; \quad (3.3)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\hat{\rho}_i(t_z)$  est donnée par*

$$\text{Var}[\hat{\rho}_i(t_z)] = x_0 - \rho; \quad (3.4)$$

(iii) Pour le modèle avec la loi binomiale, la variance de  $\hat{\rho}_i(t_Z)$  est donnée par

$$\text{Var}[\hat{\rho}_i(t_Z)] = \frac{\rho}{x_0} \left( x_0 - \rho \right). \quad (3.5)$$

*Démonstration.* (i) Étant donné que

$$\mathbb{E}[X_i(t)] = x(t),$$

on a que

$$\begin{aligned} \mathbb{E}[\hat{\rho}_i(t_Z)] &= \mathbb{E}[X_i(t_Z)] \\ &= x(t_Z) \\ &= \rho. \end{aligned}$$

(ii) Étant donné que

$$\text{Var}[X(t)] = x_0 - x(t)$$

dans le modèle avec la loi de Poisson, on a que

$$\begin{aligned} \text{Var}[\hat{\rho}_i(t_Z)] &= \text{Var}[X_i(t_Z)] \\ &= x_0 - x(t_Z) \\ &= x_0 - \rho. \end{aligned}$$

(iii) La preuve de (3.5) étant similaire à celle de (3.4), on ne la présente pas. Rappelons toutefois que

$$\text{Var}[X(t)] = \frac{x(t)}{x_0} \left( x_0 - x(t) \right)$$

dans le modèle avec la loi binomiale.

□

Comme l'espérance de  $\hat{\rho}_i(t_Z)$  est égale à  $\rho$ ,  $\hat{\rho}_i(t_Z)$  est un estimateur sans biais pour  $\rho$ . Soit  $\bar{\rho}(t_Z)$  l'estimateur de  $\rho$  obtenu en prenant la moyenne des  $\hat{\rho}_i(t_Z)$ , c'est-à-dire

$$\bar{\rho}(t_Z) = \frac{\sum_{i=1}^k \hat{\rho}_i(t_Z)}{k}. \quad (3.6)$$

**Proposition 3.3.2.** *On a les propriétés suivantes pour  $\bar{\rho}(t_z)$  :*

(i) *Pour les deux modèles, l'espérance de  $\bar{\rho}(t_z)$  est donnée par*

$$\mathbb{E}[\bar{\rho}(t_z)] = \rho; \quad (3.7)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\bar{\rho}(t_z)$  est donnée par*

$$\text{Var}[\bar{\rho}(t_z)] = \frac{1}{k} \left( x_0 - \rho \right); \quad (3.8)$$

(iii) *Pour le modèle avec la loi binomiale, la variance de  $\bar{\rho}(t_z)$  est donnée par*

$$\text{Var}[\bar{\rho}(t_z)] = \frac{\rho}{kx_0} \left( x_0 - \rho \right). \quad (3.9)$$

*Démonstration.* (i) De l'équation (3.3), on a que

$$\begin{aligned} \mathbb{E}[\bar{\rho}(t_z)] &= \mathbb{E} \left[ \frac{\sum_{i=1}^k \hat{\rho}_i(t_z)}{k} \right] \\ &= \mathbb{E}[\hat{\rho}_1(t_z)] \\ &= \rho. \end{aligned}$$

(ii) De l'équation (3.4), on a que

$$\begin{aligned} \text{Var}[\bar{\rho}(t_z)] &= \text{Var} \left[ \frac{\sum_{i=1}^k \hat{\rho}_i(t_z)}{k} \right] \\ &= \frac{1}{k} \text{Var}[\hat{\rho}_1(t_z)] \\ &= \frac{1}{k} \left( x_0 - \rho \right), \end{aligned}$$

car les variables aléatoires  $\hat{\rho}_i(t_z)$  sont indépendantes.

(iii) La preuve de (3.9) étant similaire à celle de (3.8), on ne la présente pas.

□

Comme  $\hat{\rho}_i(t_z)$ ,  $\bar{\rho}(t_z)$  est un estimateur sans biais pour  $\rho$ . Donc, si  $t_z$  était connu, la méthode d'estimation proposée permettrait d'obtenir un estimateur non biaisé pour  $\rho$ .

### 3.3.2 Estimation de $\rho$ avec $t_Z$ inconnu

En pratique, on ne connaît pas  $t_Z$  et on doit l'estimer. Rappelons que l'on a supposé à la section 3.1 que l'on possède des données sur  $k$  épidémies. On propose deux estimateurs de  $t_Z$ . Le premier est obtenu à partir des données d'une seule des épidémies à la fois, alors que le second est obtenu à partir des données de toutes les épidémies. Chacun des deux estimateurs de  $t_Z$  permet de construire un estimateur de  $\rho$  pour chacune des  $k$  épidémies. Tel que mentionné à la section 3.2, un nouvel estimateur de  $\rho$ , noté  $\bar{\rho}$ , est ensuite obtenu en calculant la moyenne des  $k$  estimateurs de  $\rho$  obtenus au préalable. Pour le moment, on présente une justification intuitive des estimateurs de  $t_Z$  et  $\rho$  proposés. Au chapitre 4, on verra que les suppositions faites dans cette section sont vérifiées lorsqu'on obtient les solutions numériques du système (2.1) pour un ensemble de paramètres donné.

#### Estimation de $t_Z$ à partir de chaque épidémie séparément

Comme  $t_Z$  est l'instant auquel  $z'(t)$  atteint sa valeur maximale, il serait intéressant de pouvoir l'estimer à partir des accroissements du nombre d'individus retirés entre les instants auxquels on a des observations. Cependant, on n'observe pas le nombre d'individus retirés, mais bien le nombre d'individus décédés. Rappelons que l'on a supposé que la probabilité qu'un individu retiré soit en fait décédé, notée  $p$ , est constante dans le temps pour les deux modèles stochastiques proposés à la section 2.3. Cette supposition permet d'estimer  $t_Z$  à partir des accroissements du nombre d'individus décédés entre les instants auxquels on a des observations, plutôt qu'à partir du nombre d'individus retirés.

Supposons que la proportion des individus retirés qui sont en fait décédés dans le modèle déterministe est constante dans le temps et est égale à  $p$ . Soit  $d(t)$  le nombre d'individus décédés dans le modèle déterministe au temps  $t$ , c'est-à-dire

$$d(t) = pz(t),$$

$t \geq 0$ . Il est clair que  $t_Z$  est également l'instant auquel  $d'(t)$  atteint son maximum. Soit  $\ell_D(\tau)$  le nombre d'individus décédés dans le modèle déterministe entre les instants  $\tau - 1$

et  $\tau$ , c'est-à-dire

$$\ell_D(\tau) = d(\tau) - d(\tau - 1)$$

et soit  $L_D(\tau)$  le nombre d'individus décédés dans le modèle stochastique entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$L_D(\tau) = D(\tau) - D(\tau - 1),$$

$\tau = 1, 2, \dots, n$ . Étant donné que  $L_D(1), L_D(2), \dots, L_D(n)$  sont des variables aléatoires, la valeur maximale des  $L_D(\tau)$  peut correspondre à plus d'une valeur de  $\tau$ . Par conséquent, l'estimateur de  $t_z$ , noté  $T_z$ , est obtenu en calculant la moyenne des valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximal,  $\tau = 1, 2, \dots, n$ . Si la valeur obtenue n'est pas entière, alors on l'arrondit afin que  $T_z$  soit un nombre entier entre 1 et  $n$  inclusivement. Il est nécessaire que  $T_z$  prenne une valeur entière, étant donné qu'on utilisera entre autres le nombre d'individus susceptibles observés à l'instant  $T_z$  pour estimer  $\rho$  et que l'on a seulement des observations aux instants  $0, 1, \dots, n$ . Pour ce faire, on peut soit arrondir la valeur obtenue à l'entier le plus près, soit prendre sa partie entière, soit prendre sa partie entière à laquelle on ajoute 1. On propose de choisir la façon d'arrondir pour laquelle le biais de  $T_z$  est le plus petit. Plus de détails à ce sujet sont présentés au chapitre 4.

Afin de construire un estimateur de  $\rho$  à partir de  $T_z$ , on propose de se baser sur l'espérance de  $T_z$ . Étant donné que  $\{D(t), t \geq 0\}$  a été modélisé par un processus de Poisson non homogène d'intensité  $pz'(t)$  dans les deux modèles proposés à la section 2.3 et que  $L_D(\tau)$  est l'accroissement de ce processus entre les instants  $\tau - 1$  et  $\tau$ , les variables aléatoires  $L_D(\tau)$  sont indépendantes et distribuées selon une loi de Poisson. Le paramètre de chacune des ces variables est donné par

$$\begin{aligned} \int_{\tau-1}^{\tau} pz'(t)dt &= \int_{\tau-1}^{\tau} d'(t)dt \\ &= d(\tau) - d(\tau - 1) \\ &= \ell_D(\tau), \end{aligned} \tag{3.10}$$

$\tau = 1, 2, \dots, n$ . On a donc que  $L_D(\tau)$  est de loi de Poisson  $\mathcal{P}(\ell_D(\tau))$ ,  $\tau = 1, 2, \dots, n$ . L'équation (3.10) met en évidence le fait que  $\ell_D(\tau)$  correspond à l'aire sous la courbe  $d'(t)$  entre  $t = \tau - 1$  et  $t = \tau$ . À la figure 3.1, on remarque que les deux plus grandes aires sont situées de part et d'autre de  $t_Z$ . Ces valeurs correspondent à  $\ell_D(t_Z)$  et  $\ell_D(t_Z + 1)$ .

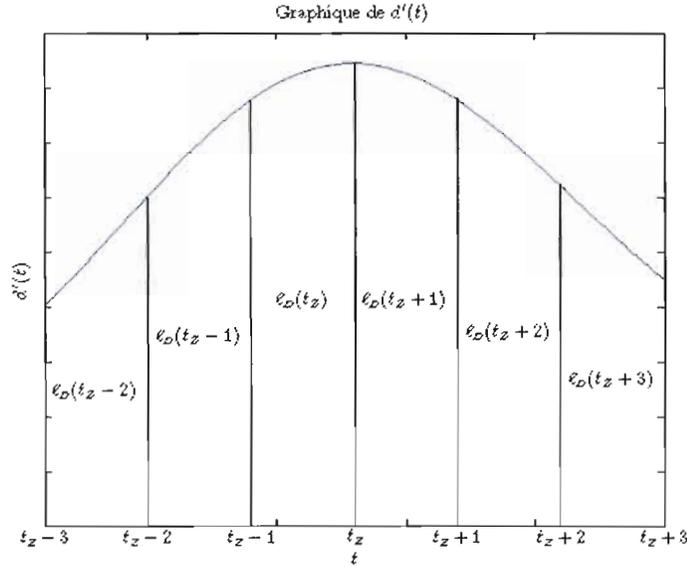


Figure 3.1 Graphique de  $d'(t)$

Par conséquent, les variables aléatoires  $L_D(t_Z)$  et  $L_D(t_Z + 1)$  ont une probabilité plus grande que les autres de correspondre au maximum des  $L_D(\tau)$ ,  $\tau = 1, 2, \dots, n$ . On s'attend donc à ce que

$$P[T_Z = t_Z] > P[T_Z = \tau], \tau \notin \{t_Z, t_Z + 1\}$$

et

$$P[T_Z = t_Z + 1] > P[T_Z = \tau], \tau \notin \{t_Z, t_Z + 1\}.$$

Supposons que la valeur maximale des  $L_D(\tau)$  ne puisse être atteinte que pour  $\tau = t_Z$  et  $\tau = t_Z + 1$ . Avec cette supposition, les seules valeurs possibles de  $T_Z$  sont  $t_Z$  et  $t_Z + 1$ . Étant donné que  $d'(t)$  est à peu près symétrique par rapport à  $t_Z$  dans l'intervalle  $(t_Z - 1, t_Z + 1)$ , les valeurs de  $\ell_D(t_Z)$  et  $\ell_D(t_Z + 1)$  sont à peu près égales. Cela implique

que

$$P[T_Z = t_Z] \approx P[T_Z = t_Z + 1].$$

Intuitivement, on s'attend donc à avoir que

$$t_Z < \mathbb{E}[T_Z] < t_Z + 1. \quad (3.11)$$

En supposant que l'équation (3.11) est vérifiée, on va maintenant construire un estimateur de  $\rho$  à partir de  $T_Z$ . Pour ce faire, réécrivons (3.11) de la façon suivante :

$$\begin{aligned} t_Z < \mathbb{E}[T_Z] < t_Z + 1 &\Leftrightarrow 0 < \mathbb{E}[T_Z] - t_Z < 1 \\ &\Leftrightarrow -\mathbb{E}[T_Z] < -t_Z < -\mathbb{E}[T_Z] + 1 \\ &\Leftrightarrow \mathbb{E}[T_Z] - 1 < t_Z < \mathbb{E}[T_Z]. \end{aligned}$$

Soit  $T_Z^{(i)}$  l'estimateur de  $t_Z$  pour la  $i^{\text{ième}}$  épidémie et soit  $\hat{\rho}_i(T_Z^{(i)})$  l'estimateur de  $\rho$  basé sur  $T_Z^{(i)}$  pour la  $i^{\text{ième}}$  épidémie,  $i = 1, 2, \dots, k$ . Comme  $t_Z$  est inclus dans l'intervalle  $(\mathbb{E}[T_Z^{(i)}] - 1, \mathbb{E}[T_Z^{(i)}])$ , on propose d'estimer  $\rho$  en calculant la moyenne du nombre d'individus susceptibles observés aux instants  $T_Z^{(i)} - 1$  et  $T_Z^{(i)}$ , c'est-à-dire :

$$\hat{\rho}_i(T_Z^{(i)}) = \frac{X_i(T_Z^{(i)} - 1) + X_i(T_Z^{(i)})}{2}. \quad (3.12)$$

On va maintenant développer les propriétés de  $T_Z^{(i)}$  et  $\hat{\rho}_i(T_Z^{(i)})$ ,  $i = 1, 2, \dots, k$ . Afin de calculer l'espérance et la variance de  $T_Z^{(i)}$ , on doit connaître sa fonction de masse. Le calcul de la fonction de masse de  $T_Z^{(i)}$  est présenté à la section 3.5.1 et son approximation numérique est présentée à la section 4.5. À partir de la fonction de masse de  $T_Z^{(i)}$ , on peut calculer son espérance et sa variance comme suit :

$$\begin{aligned} \mathbb{E}[T_Z^{(i)}] &= \sum_{\tau=1}^n \tau P[T_Z^{(i)} = \tau], \\ \text{Var}[T_Z^{(i)}] &= \sum_{\tau=1}^n (\tau - \mathbb{E}[T_Z^{(i)}])^2 P[T_Z^{(i)} = \tau]. \end{aligned}$$

Comme on s'attend à ce que l'espérance de  $T_Z^{(i)}$  soit dans l'intervalle  $(t_Z, t_Z + 1)$ ,  $T_Z^{(i)}$  semble être un estimateur biaisé pour  $t_Z$ .

**Proposition 3.3.3.** *On a les propriétés suivantes pour  $\hat{\rho}_i(T_Z^{(i)})$ ,  $i = 1, 2, \dots, k$  :*

(i) *Pour les deux modèles, l'espérance de  $\hat{\rho}_i(T_Z^{(i)})$  est donnée par*

$$\mathbb{E}[\hat{\rho}_i(T_Z^{(i)})] = \frac{\mathbb{E}[x(T_Z^{(i)} - 1)] + \mathbb{E}[x(T_Z^{(i)})]}{2}; \quad (3.13)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\hat{\rho}_i(T_Z^{(i)})$  est donnée par*

$$\begin{aligned} \text{Var}[\hat{\rho}_i(T_Z^{(i)})] &= x_0 - \frac{3}{4}\mathbb{E}[x(T_Z^{(i)} - 1)] - \frac{1}{4}\mathbb{E}[x(T_Z^{(i)})] + \frac{1}{4}\text{Var}[x(T_Z^{(i)} - 1)] \\ &\quad + \frac{1}{4}\text{Var}[x(T_Z^{(i)})] + \frac{1}{2}\text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})]; \end{aligned} \quad (3.14)$$

(iii) *Pour le modèle avec la loi binomiale, la variance de  $\hat{\rho}_i(T_Z^{(i)})$  est donnée par*

$$\begin{aligned} \text{Var}[\hat{\rho}_i(T_Z^{(i)})] &= \frac{1}{4}\mathbb{E}[x(T_Z^{(i)} - 1)] - \frac{1}{4x_0}\mathbb{E}[x^2(T_Z^{(i)} - 1)] - \frac{1}{2x_0}\mathbb{E}[x(T_Z^{(i)} - 1)x(T_Z^{(i)})] \\ &\quad + \frac{3}{4}\mathbb{E}[x(T_Z^{(i)})] - \frac{1}{4x_0}\mathbb{E}[x^2(T_Z^{(i)})] + \frac{1}{4}\text{Var}[x(T_Z^{(i)} - 1)] \\ &\quad + \frac{1}{4}\text{Var}[x(T_Z^{(i)})] + \frac{1}{2}\text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})]. \end{aligned} \quad (3.15)$$

*Démonstration.* (i) Étant donné que, pour  $t$  fixe,

$$\mathbb{E}[X(t)] = x(t),$$

on a que

$$\begin{aligned} \mathbb{E}[\hat{\rho}_i(T_Z^{(i)})] &= \mathbb{E}(\mathbb{E}[\hat{\rho}_i(T_Z^{(i)})|T_Z^{(i)}]) \\ &= \mathbb{E}\left(\mathbb{E}\left[\frac{X_i(T_Z^{(i)} - 1) + X_i(T_Z^{(i)})}{2} \middle| T_Z^{(i)}\right]\right) \\ &= \mathbb{E}\left(\frac{\mathbb{E}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}] + \mathbb{E}[X_i(T_Z^{(i)})|T_Z^{(i)}]}{2}\right) \\ &= \mathbb{E}\left(\frac{x(T_Z^{(i)} - 1) + x(T_Z^{(i)})}{2}\right) \\ &= \frac{\mathbb{E}[x(T_Z^{(i)} - 1)] + \mathbb{E}[x(T_Z^{(i)})]}{2}. \end{aligned}$$

(ii) Pour le modèle avec la loi de Poisson, par la proposition 2.3.4, on a que

$$\text{Cov}[X_i(T_Z^{(i)} - 1), X_i(T_Z^{(i)})|T_Z^{(i)}] = \text{Var}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}].$$

De plus, on a que, pour  $t$  fixe,

$$\text{Var}[X(t)] = x_0 - x(t)$$

dans le modèle avec la loi de Poisson. Donc, on a

$$\begin{aligned}
& \text{Var}[\hat{\rho}_i(T_Z^{(i)})] \\
&= \mathbb{E}(\text{Var}[\hat{\rho}_i(T_Z^{(i)})|T_Z^{(i)}]) + \text{Var}(\mathbb{E}[\hat{\rho}_i(T_Z^{(i)})|T_Z^{(i)}]) \\
&= \mathbb{E}\left(\text{Var}\left[\frac{X_i(T_Z^{(i)} - 1) + X(T_Z^{(i)})}{2} \middle| T_Z^{(i)}\right]\right) \\
&\quad + \text{Var}\left(\mathbb{E}\left[\frac{X_i(T_Z^{(i)} - 1) + X_i(T_Z^{(i)})}{2} \middle| T_Z^{(i)}\right]\right) \\
&= \mathbb{E}\left(\frac{\text{Var}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}] + \text{Var}[X_i(T_Z^{(i)})|T_Z^{(i)}]}{4}\right) \\
&\quad + \mathbb{E}\left(\frac{\text{Cov}[X_i(T_Z^{(i)} - 1), X_i(T_Z^{(i)})|T_Z^{(i)}]}{2}\right) \\
&\quad + \text{Var}\left(\frac{\mathbb{E}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}] + \mathbb{E}[X_i(T_Z^{(i)})|T_Z^{(i)}]}{2}\right) \\
&= \mathbb{E}\left(\frac{\text{Var}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}] + \text{Var}[X_i(T_Z^{(i)})|T_Z^{(i)}] + 2\text{Var}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}]}{4}\right) \\
&\quad + \text{Var}\left(\frac{x(T_Z^{(i)} - 1) + x(T_Z^{(i)})}{2}\right) \\
&= \mathbb{E}\left(\frac{3\text{Var}[X_i(T_Z^{(i)} - 1)|T_Z] + \text{Var}[X_i(T_Z^{(i)})|T_Z^{(i)}]}{4}\right) \\
&\quad + \text{Var}\left(\frac{x(T_Z^{(i)} - 1) + x(T_Z^{(i)})}{2}\right) \\
&= \mathbb{E}\left(\frac{4x_0 - 3x(T_Z^{(i)} - 1) - x(T_Z^{(i)})}{4}\right) + \text{Var}\left(\frac{x(T_Z^{(i)} - 1) + x(T_Z^{(i)})}{2}\right) \\
&= \frac{\mathbb{E}[4x_0 - 3x(T_Z^{(i)} - 1) - x(T_Z^{(i)})]}{4} + \frac{\text{Var}[x(T_Z^{(i)} - 1) + x(T_Z^{(i)})]}{4} \\
&= \frac{4x_0 - 3\mathbb{E}[x(T_Z^{(i)} - 1)] - \mathbb{E}[x(T_Z^{(i)})]}{4} \\
&\quad + \frac{\text{Var}[x(T_Z^{(i)} - 1)] + \text{Var}[x(T_Z^{(i)})] + 2\text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})]}{4} \\
&= x_0 - \frac{3}{4}\mathbb{E}[x(T_Z^{(i)} - 1)] - \frac{1}{4}\mathbb{E}[x(T_Z^{(i)})] + \frac{1}{4}\text{Var}[x(T_Z^{(i)} - 1)] + \frac{1}{4}\text{Var}[x(T_Z^{(i)})] \\
&\quad + \frac{1}{2}\text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})].
\end{aligned}$$

(iii) La preuve de (3.15) étant similaire à celle de (3.14), on ne la présente pas. Rappelons toutefois que, dans le modèle avec la loi binomiale, par la proposition 2.3.2, on a que

$$\text{Cov}[X_i(T_Z^{(i)} - 1), X_i(T_Z^{(i)})|T_Z^{(i)}] = \frac{x(T_Z^{(i)})}{x(T_Z^{(i)} - 1)} \text{Var}[X_i(T_Z^{(i)} - 1)|T_Z^{(i)}].$$

□

Les valeurs de  $\mathbb{E}[x(T_Z^{(i)} - 1)]$ ,  $\mathbb{E}[x(T_Z^{(i)})]$ ,  $\mathbb{E}[x^2(T_Z^{(i)} - 1)]$ ,  $\mathbb{E}[x^2(T_Z^{(i)})]$ ,  $\mathbb{E}[x(T_Z^{(i)} - 1)x(T_Z^{(i)})]$ ,  $\text{Var}[x(T_Z^{(i)} - 1)]$ ,  $\text{Var}[x(T_Z^{(i)})]$  et  $\text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})]$  dans les équations (3.13), (3.14) et (3.15) sont calculées à partir de la fonction de masse de  $T_Z^{(i)}$  comme suit :

$$\begin{aligned} \mathbb{E}[x(T_Z^{(i)} - 1)] &= \sum_{\tau=1}^n x(\tau - 1)P[T_Z^{(i)} = \tau], \\ \mathbb{E}[x(T_Z^{(i)})] &= \sum_{\tau=1}^n x(\tau)P[T_Z^{(i)} = \tau], \\ \mathbb{E}[x^2(T_Z^{(i)} - 1)] &= \sum_{\tau=1}^n x^2(\tau - 1)P[T_Z^{(i)} = \tau], \\ \mathbb{E}[x^2(T_Z^{(i)})] &= \sum_{\tau=1}^n x^2(\tau)P[T_Z^{(i)} = \tau], \\ \mathbb{E}[x(T_Z^{(i)} - 1)x(T_Z^{(i)})] &= \sum_{\tau=1}^n x(\tau - 1)x(\tau)P[T_Z^{(i)} = \tau], \\ \text{Var}[x(T_Z^{(i)} - 1)] &= \sum_{\tau=1}^n \left( x(\tau - 1) - \mathbb{E}[x(T_Z^{(i)} - 1)] \right)^2 P[T_Z^{(i)} = \tau], \\ \text{Var}[x(T_Z^{(i)})] &= \sum_{\tau=1}^n \left( x(\tau) - \mathbb{E}[x(T_Z^{(i)})] \right)^2 P[T_Z^{(i)} = \tau], \\ \text{Cov}[x(T_Z^{(i)} - 1), x(T_Z^{(i)})] &= \sum_{\tau=1}^n \left( x(\tau - 1) - \mathbb{E}[x(T_Z^{(i)} - 1)] \right) \left( x(\tau) - \mathbb{E}[x(T_Z^{(i)})] \right) \\ &\quad P[T_Z^{(i)} = \tau]. \end{aligned}$$

Étant donné que  $\rho = x(t_Z)$  et qu'il est raisonnable de croire que

$$\frac{\mathbb{E}[x(T_Z^{(i)} - 1)] + \mathbb{E}[x(T_Z^{(i)})]}{2} \neq x(t_Z),$$

on s'attend à ce que  $\hat{\rho}_i(T_Z^{(i)})$  soit un estimateur biaisé pour  $\rho$ ,  $i = 1, 2, \dots, k$ .

Soit  $\mathbf{T}_Z = (T_Z^{(1)}, T_Z^{(2)}, \dots, T_Z^{(k)})$  et soit  $\bar{\rho}(\mathbf{T}_Z)$  l'estimateur de  $\rho$  obtenu en prenant la moyenne des  $\hat{\rho}_i(T_Z^{(i)})$ , c'est-à-dire

$$\bar{\rho}(\mathbf{T}_Z) = \frac{\sum_{i=1}^k \hat{\rho}_i(T_Z^{(i)})}{k}. \quad (3.16)$$

**Proposition 3.3.4.** *On a les propriétés suivantes pour  $\bar{\rho}(\mathbf{T}_Z)$  :*

(i) *Pour les deux modèles, l'espérance de  $\bar{\rho}(\mathbf{T}_Z)$  est donnée par :*

$$\mathbb{E}[\bar{\rho}(\mathbf{T}_Z)] = \frac{\mathbb{E}[x(T_Z^{(1)} - 1)] + \mathbb{E}[x(T_Z^{(1)})]}{2}; \quad (3.17)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\bar{\rho}(\mathbf{T}_Z)$  est donnée par :*

$$\begin{aligned} \text{Var}[\bar{\rho}(\mathbf{T}_Z)] &= \frac{1}{k} \left[ x_0 - \frac{3}{4} \mathbb{E}[x(T_Z^{(1)} - 1)] - \frac{1}{4} \mathbb{E}[x(T_Z^{(1)})] + \frac{1}{4} \text{Var}[x(T_Z^{(1)} - 1)] \right. \\ &\quad \left. + \frac{1}{4} \text{Var}[x(T_Z^{(1)})] + \frac{1}{2} \text{Cov}[x(T_Z^{(1)} - 1), x(T_Z^{(1)})] \right]; \end{aligned} \quad (3.18)$$

(iii) *Pour le modèle avec la loi binomiale, la variance de  $\bar{\rho}(\mathbf{T}_Z)$  est donnée par :*

$$\begin{aligned} \text{Var}[\bar{\rho}(\mathbf{T}_Z)] &= \frac{1}{k} \left[ \frac{1}{4} \mathbb{E}[x(T_Z^{(1)} - 1)] - \frac{1}{4x_0} \mathbb{E}[x^2(T_Z^{(1)} - 1)] \right. \\ &\quad - \frac{1}{2x_0} \mathbb{E}[x(T_Z^{(1)} - 1)x(T_Z^{(1)})] + \frac{3}{4} \mathbb{E}[x(T_Z^{(1)})] - \frac{1}{4x_0} \mathbb{E}[x^2(T_Z^{(1)})] \\ &\quad + \frac{1}{4} \text{Var}[x(T_Z^{(1)} - 1)] + \frac{1}{4} \text{Var}[x(T_Z^{(1)})] \\ &\quad \left. + \frac{1}{2} \text{Cov}[x(T_Z^{(1)} - 1), x(T_Z^{(1)})] \right]. \end{aligned} \quad (3.19)$$

*Démonstration.* (i) De l'équation (3.13), on a que

$$\begin{aligned} \mathbb{E}[\bar{\rho}(\mathbf{T}_Z)] &= \mathbb{E} \left[ \frac{\sum_{i=1}^k \hat{\rho}_i(T_Z^{(i)})}{k} \right] \\ &= \mathbb{E}[\hat{\rho}_1(T_Z^{(1)})] \\ &= \frac{\mathbb{E}[x(T_Z^{(1)} - 1)] + \mathbb{E}[x(T_Z^{(1)})]}{2}. \end{aligned}$$

(ii) De l'équation (3.14), on a que

$$\begin{aligned}
\text{Var}[\bar{\rho}(\mathbf{T}_Z)] &= \text{Var}\left[\frac{\sum_{i=1}^k \hat{\rho}_i(T_Z^{(i)})}{k}\right] \\
&= \frac{1}{k} \text{Var}[\hat{\rho}_1(T_Z^{(1)})] \\
&= \frac{1}{k} \left[ x_0 - \frac{3}{4} \mathbb{E}[x(T_Z^{(1)}) - 1] - \frac{1}{4} \mathbb{E}[x(T_Z^{(1)})] + \frac{1}{4} \text{Var}[x(T_Z^{(1)}) - 1] \right. \\
&\quad \left. + \frac{1}{4} \text{Var}[x(T_Z^{(1)})] + \frac{1}{2} \text{Cov}[x(T_Z^{(1)}) - 1, x(T_Z^{(1)})] \right],
\end{aligned}$$

car les variables aléatoires  $\hat{\rho}_i(T_Z^{(i)})$  sont indépendantes.

(iii) La preuve de (3.19) étant similaire à celle de (3.18), on ne la présente pas.

□

Comme l'espérance de  $\bar{\rho}(\mathbf{T}_Z)$  est égale à l'espérance de  $\hat{\rho}_i(T_Z^{(i)})$ , on s'attend à ce que  $\bar{\rho}(\mathbf{T}_Z)$  soit lui aussi un estimateur biaisé pour  $\rho$ . De plus, son biais ne décroît pas lorsque  $k$  augmente, car la fonction de masse de  $T_Z^{(i)}$  ne fait pas intervenir  $k$ . Cependant, la variance de  $\bar{\rho}(\mathbf{T}_Z)$  décroît lorsque  $k$  augmente.

### Estimation de $t_Z$ à partir de toutes les épidémies

On a vu que les estimateurs  $T_Z^{(i)}$ ,  $i = 1, 2, \dots, k$  et  $\bar{\rho}(\mathbf{T}_Z)$  sont biaisés. De plus, la valeur de leur biais ne décroît pas lorsque  $k$ , le nombre d'épidémies considérées pour calculer  $\bar{\rho}(\mathbf{T}_Z)$ , augmente. Il serait intéressant d'avoir des estimateurs de  $t_Z$  et  $\rho$  qui soient asymptotiquement non biaisés, c'est-à-dire que leurs biais respectifs tendent vers 0 lorsque  $k$  tend vers l'infini. C'est la raison pour laquelle un deuxième estimateur de  $t_Z$  est proposé.

Rappelons que les données pour deux épidémies différentes sont supposées indépendantes. Rappelons également que  $T_Z^{(i)}$  est l'estimateur de  $t_Z$  obtenu à partir des données de la  $i^{\text{ème}}$  épidémie,  $i = 1, 2, \dots, k$ . Soit  $\bar{T}_Z$  la moyenne des  $T_Z^{(i)}$ , c'est-à-dire

$$\bar{T}_Z = \frac{\sum_{i=1}^k T_Z^{(i)}}{k}. \quad (3.20)$$

Soit  $\tilde{T}_z$  la partie entière de  $\bar{T}_z$ . On peut montrer que  $\tilde{T}_z$  est un estimateur fortement convergent pour  $t_z$ . La preuve est présentée à la section 3.5.3. Étant donné que  $\tilde{T}_z$  est un estimateur fortement convergent pour  $t_z$ , on propose d'estimer  $\rho$  à partir de  $\tilde{T}_z$  pour la  $i^{\text{ème}}$  épidémie de la façon suivante :

$$\hat{\rho}_i(\tilde{T}_z) = X_i(\tilde{T}_z), \quad (3.21)$$

$i = 1, 2, \dots, k$ . On estime ensuite  $\rho$  à partir des  $\hat{\rho}_i(\tilde{T}_z)$  de la façon suivante :

$$\bar{\rho}(\tilde{T}_z) = \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_z)}{k}. \quad (3.22)$$

On peut montrer que  $\bar{\rho}(\tilde{T}_z)$  est un estimateur fortement convergent pour  $\rho$ . La preuve est présentée à la section 3.5.4.

On va maintenant développer les propriétés de  $\tilde{T}_z$ ,  $\hat{\rho}_i(\tilde{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  pour  $k$  fixe. Afin de calculer l'espérance et la variance de  $\tilde{T}_z$ , on doit connaître sa fonction de masse. Le calcul de la fonction de masse de  $\tilde{T}_z$  est présenté à la section 3.5.2 et son approximation numérique est présentée à la section 4.6. À partir de la fonction de masse de  $\tilde{T}_z$ , on peut calculer son espérance et sa variance comme suit :

$$\begin{aligned} \mathbb{E}[\tilde{T}_z] &= \sum_{\tau=1}^n \tau P[\tilde{T}_z = \tau], \\ \text{Var}[\tilde{T}_z] &= \sum_{\tau=1}^n (\tau - \mathbb{E}[\tilde{T}_z])^2 P[\tilde{T}_z = \tau]. \end{aligned}$$

**Proposition 3.3.5.** *On a les propriétés suivantes pour  $\hat{\rho}_i(\tilde{T}_z)$ ,  $i = 1, 2, \dots, k$  :*

(i) *Pour les deux modèles, l'espérance de  $\hat{\rho}_i(\tilde{T}_z)$  est donnée par*

$$\mathbb{E}[\hat{\rho}_i(\tilde{T}_z)] = \mathbb{E}[x(\tilde{T}_z)]; \quad (3.23)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\hat{\rho}_i(\tilde{T}_z)$  est donnée par*

$$\text{Var}[\hat{\rho}_i(\tilde{T}_z)] = x_0 - \mathbb{E}[x(\tilde{T}_z)] + \text{Var}[x(\tilde{T}_z)]; \quad (3.24)$$

(iii) *Pour le modèle avec la loi binomiale, la variance de  $\hat{\rho}_i(\tilde{T}_z)$  est donnée par*

$$\text{Var}[\hat{\rho}_i(\tilde{T}_z)] = \mathbb{E}[x(\tilde{T}_z)] - \frac{\mathbb{E}[x^2(\tilde{T}_z)]}{x_0} + \text{Var}[x(\tilde{T}_z)]. \quad (3.25)$$

*Démonstration.* (i) Étant donné que, pour  $t$  fixe,

$$\mathbb{E}[X(t)] = x(t),$$

pour les deux modèles, on a que

$$\begin{aligned} \mathbb{E}[\hat{\rho}_i(\tilde{T}_Z)] &= \mathbb{E}(\mathbb{E}[\hat{\rho}_i(\tilde{T}_Z)|\tilde{T}_Z]) \\ &= \mathbb{E}(\mathbb{E}[X_i(\tilde{T}_Z)|\tilde{T}_Z]) \\ &= \mathbb{E}[x(\tilde{T}_Z)]. \end{aligned}$$

(ii) Étant donné que, pour  $t$  fixe,

$$\text{Var}[X(t)] = x_0 - x(t)$$

pour le modèle avec la loi de Poisson, on a que

$$\begin{aligned} \text{Var}[\hat{\rho}_i(\tilde{T}_Z)] &= \mathbb{E}(\text{Var}[\hat{\rho}_i(\tilde{T}_Z)|\tilde{T}_Z]) + \text{Var}(\mathbb{E}[\hat{\rho}_i(\tilde{T}_Z)|\tilde{T}_Z]) \\ &= \mathbb{E}(\text{Var}[X_i(\tilde{T}_Z)|\tilde{T}_Z]) + \text{Var}(\mathbb{E}[X_i(\tilde{T}_Z)|\tilde{T}_Z]) \\ &= \mathbb{E}[x_0 - x(\tilde{T}_Z)] + \text{Var}[x(\tilde{T}_Z)] \\ &= x_0 - \mathbb{E}[x(\tilde{T}_Z)] + \text{Var}[x(\tilde{T}_Z)]. \end{aligned}$$

(iii) La preuve de (3.25) étant similaire à celle de (3.24), on ne la présente pas.

□

Les valeurs de  $\mathbb{E}[x(\tilde{T}_Z)]$ ,  $\mathbb{E}[x^2(\tilde{T}_Z)]$  et  $\text{Var}[x(\tilde{T}_Z)]$  dans les équations (3.23), (3.24) et (3.25) sont calculées à partir de la fonction de masse de  $\tilde{T}_Z$  comme suit :

$$\begin{aligned} \mathbb{E}[x(\tilde{T}_Z)] &= \sum_{\tau=1}^n x(\tau)P[\tilde{T}_Z = \tau], \\ \mathbb{E}[x^2(\tilde{T}_Z)] &= \sum_{\tau=1}^n x^2(\tau)P[\tilde{T}_Z = \tau]. \\ \text{Var}[x(\tilde{T}_Z)] &= \sum_{\tau=1}^n \left(x(\tau) - \mathbb{E}[x(\tilde{T}_Z)]\right)^2 P[\tilde{T}_Z = \tau]. \end{aligned}$$

**Proposition 3.3.6.** *On a les propriétés suivantes pour  $\bar{\rho}(\tilde{T}_z)$  :*

(i) *Pour les deux modèles, l'espérance de  $\bar{\rho}(\tilde{T}_z)$  est donnée par*

$$\mathbb{E}[\bar{\rho}(\tilde{T}_z)] = \mathbb{E}[x(\tilde{T}_z)]. \quad (3.26)$$

(ii) *Pour le modèle avec la loi de Poisson, la variance de  $\bar{\rho}(\tilde{T}_z)$  est donnée par*

$$\text{Var}[\bar{\rho}(\tilde{T}_z)] = \frac{1}{k} \left[ x_0 - \mathbb{E}[x(\tilde{T}_z)] \right] + \text{Var}[x(\tilde{T}_z)]; \quad (3.27)$$

(iii) *Pour le modèle avec la loi binomiale, la variance de  $\bar{\rho}(\tilde{T}_z)$  est donnée par*

$$\text{Var}[\bar{\rho}(\tilde{T}_z)] = \frac{1}{k} \left[ \mathbb{E}[x(\tilde{T}_z)] - \frac{\mathbb{E}[x^2(\tilde{T}_z)]}{x_0} \right] + \text{Var}[x(\tilde{T}_z)]; \quad (3.28)$$

*Démonstration.* (i) De l'équation (3.23), on a que

$$\begin{aligned} \mathbb{E}[\bar{\rho}(\tilde{T}_z)] &= \mathbb{E} \left[ \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_z)}{k} \right] \\ &= \mathbb{E}[\hat{\rho}_1(\tilde{T}_z)] \\ &= \mathbb{E}[x(\tilde{T}_z)]. \end{aligned}$$

(ii) Tout d'abord, étant donné que les  $\hat{\rho}_i(\tilde{T}_z)$  sont des fonctions de  $\tilde{T}_z$ , ce sont des variables aléatoires dépendantes. Par conséquent, on a que

$$\begin{aligned} \text{Var}[\bar{\rho}(\tilde{T}_z)] &= \text{Var} \left[ \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_z)}{k} \right] \\ &= \frac{1}{k^2} \left[ k \text{Var}[\hat{\rho}_1(\tilde{T}_z)] + k(k-1) \text{Cov}[\hat{\rho}_1(\tilde{T}_z), \hat{\rho}_2(\tilde{T}_z)] \right] \\ &= \frac{\text{Var}[\hat{\rho}_1(\tilde{T}_z)]}{k} + \frac{k-1}{k} \text{Cov}[\hat{\rho}_1(\tilde{T}_z), \hat{\rho}_2(\tilde{T}_z)]. \end{aligned} \quad (3.29)$$

Soit  $1 \leq i < j \leq k$ . La covariance entre les estimateurs  $\hat{\rho}_i(\tilde{T}_z)$  et  $\hat{\rho}_j(\tilde{T}_z)$  est donnée

par

$$\begin{aligned}
\text{Cov}[\hat{\rho}_i(\tilde{T}_z), \hat{\rho}_j(\tilde{T}_z)] &= \mathbb{E}[\hat{\rho}_i(\tilde{T}_z)\hat{\rho}_j(\tilde{T}_z)] - \mathbb{E}[\hat{\rho}_i(\tilde{T}_z)] \cdot \mathbb{E}[\hat{\rho}_j(\tilde{T}_z)] \\
&= \mathbb{E}[X_i(\tilde{T}_z)X_j(\tilde{T}_z)] - \mathbb{E}[X_i(\tilde{T}_z)] \cdot \mathbb{E}[X_j(\tilde{T}_z)] \\
&= \mathbb{E}(\mathbb{E}[X_i(\tilde{T}_z)X_j(\tilde{T}_z)|\tilde{T}_z]) \\
&\quad - \mathbb{E}(\mathbb{E}[X_i(\tilde{T}_z)|\tilde{T}_z]) \cdot \mathbb{E}(\mathbb{E}[X_j(\tilde{T}_z)|\tilde{T}_z]) \\
&= \mathbb{E}(\mathbb{E}[X_i(\tilde{T}_z)|\tilde{T}_z] \cdot \mathbb{E}[X_j(\tilde{T}_z)|\tilde{T}_z]) - \mathbb{E}[x(\tilde{T}_z)]^2 \\
&= \mathbb{E}[x^2(\tilde{T}_z)] - \mathbb{E}[x(\tilde{T}_z)]^2 \\
&= \text{Var}[x(\tilde{T}_z)], \tag{3.30}
\end{aligned}$$

car  $X_i(\tilde{T}_z)$  et  $X_j(\tilde{T}_z)$  sont indépendantes, conditionnellement à  $\tilde{T}_z$ , lorsque  $i \neq j$ . À partir des équations (3.24) et (3.30), on peut exprimer l'équation (3.29) de la façon suivante :

$$\begin{aligned}
\text{Var}[\bar{\rho}(\tilde{T}_z)] &= \frac{\text{Var}[\hat{\rho}_1(\tilde{T}_z)]}{k} + \frac{k-1}{k} \text{Cov}[\hat{\rho}_1(\tilde{T}_z), \hat{\rho}_2(\tilde{T}_z)] \\
&= \frac{1}{k} \left[ x_0 - \mathbb{E}[x(\tilde{T}_z)] + \text{Var}[x(\tilde{T}_z)] \right] + \frac{k-1}{k} \text{Var}[x(\tilde{T}_z)] \\
&= \frac{1}{k} \left[ x_0 - \mathbb{E}[x(\tilde{T}_z)] \right] + \text{Var}[x(\tilde{T}_z)].
\end{aligned}$$

□

Étant donné que  $\tilde{T}_z$  est un estimateur fortement convergent pour  $t_z$ , la variance de  $x(\tilde{T}_z)$  tend vers 0 lorsque  $k$  tend vers l'infini. Donc, on a bien que la variance de  $\bar{\rho}(\tilde{T}_z)$  tend vers 0 lorsque  $k$  tend vers l'infini.

### 3.4 Intervalles de confiance asymptotiques pour $\rho$ et $R_0$

Dans cette section, on présente des intervalles de confiance asymptotiques pour les paramètres  $\rho$  et  $R_0$ . Ces intervalles de confiance sont obtenus à partir des estimateurs de  $\rho$  présentés à la section 3.3.

### 3.4.1 Intervalles de confiance asymptotiques pour $\rho$

À la section 3.3, trois estimateurs de  $\rho$  ont été proposés, soit  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$ . À partir de chacun de ces estimateurs, on peut construire un intervalle de confiance asymptotique pour  $\rho$ .

Tout d'abord, on construit un intervalle de confiance asymptotique pour  $\rho$  à partir de  $\bar{\rho}(t_z)$ . Rappelons que

$$\bar{\rho}(t_z) = \frac{\sum_{i=1}^k \hat{\rho}_i(t_z)}{k},$$

où  $\hat{\rho}_i(t_z)$  est l'estimateur de  $\rho$  pour la  $i^{\text{ième}}$  épidémie lorsqu'on considère  $t_z$  connu,  $i = 1, 2, \dots, k$ . À la section 3.3.1, on a montré que  $\mathbb{E}[\bar{\rho}(t_z)] = \rho$ . Comme les  $\hat{\rho}_i(t_z)$  sont indépendants, on a que  $\bar{\rho}(t_z)$  est un moyenne de variables indépendantes et identiquement distribuées. Par le théorème de limite centrale, on a que

$$\sqrt{k} [\bar{\rho}(t_z) - \rho]$$

converge en loi vers une loi normale

$$\mathcal{N}(0, \text{Var}[\hat{\rho}_1(t_z)]),$$

où  $\text{Var}[\hat{\rho}_1(t_z)]$  est donnée par les équations (3.4) ou (3.5) selon le modèle considéré. L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $\rho$  construit à partir de  $\bar{\rho}(t_z)$  est donc donné par :

$$\left( \bar{\rho}(t_z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}}, \bar{\rho}(t_z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}} \right),$$

où  $z_{\alpha/2}$  est tel que

$$P[Z > z_{\alpha/2}] = \frac{\alpha}{2},$$

si  $Z$  est de loi normale  $\mathcal{N}(0, 1)$ .

On construit ensuite un intervalle de confiance asymptotique pour  $\rho$  à partir de  $\bar{\rho}(\mathbf{T}_z)$ .

Rappelons que

$$\bar{\rho}(\mathbf{T}_z) = \frac{\sum_{i=1}^k \hat{\rho}_i(T_z^{(i)})}{k},$$

où  $\hat{\rho}_i(T_Z^{(i)})$  est l'estimateur de  $\rho$  pour la  $i^{\text{ième}}$  épidémie lorsqu'on considère  $t_Z$  inconnu et qu'on l'estime à partir des données de la  $i^{\text{ième}}$  épidémie,  $i = 1, 2, \dots, k$ . Comme les  $\hat{\rho}_i(T_Z^{(i)})$  sont indépendants, on a que  $\bar{\rho}(\mathbf{T}_Z)$  est une moyenne de variables indépendantes et identiquement distribuées. Par le théorème de limite centrale, on a que la loi asymptotique de

$$\sqrt{k} \left[ \bar{\rho}(\mathbf{T}_Z) - \mathbb{E}[\bar{\rho}(\mathbf{T}_Z)] \right]$$

est une loi normale

$$\mathcal{N}(0, \text{Var}[\hat{\rho}_1(T_Z^{(1)})]),$$

où  $\mathbb{E}[\bar{\rho}(\mathbf{T}_Z)]$  est donnée par l'équation (3.17) et  $\text{Var}[\hat{\rho}_1(T_Z^{(1)})]$  est donnée par les équations (3.14) ou (3.15) selon le modèle considéré. L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $\rho$  construit à partir de  $\bar{\rho}(\mathbf{T}_Z)$  est donc donné par :

$$\left( \bar{\rho}(\mathbf{T}_Z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_Z^{(1)})]}{k}}, \bar{\rho}(\mathbf{T}_Z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_Z^{(1)})]}{k}} \right).$$

Pour de grandes valeurs de  $k$ , la probabilité que cet intervalle de confiance recouvre la valeur de  $\mathbb{E}[\bar{\rho}(\mathbf{T}_Z)]$  devrait être environ égale à  $1 - \alpha$ . Cependant, on s'attend à ce que l'espérance de  $\bar{\rho}(\mathbf{T}_Z)$  soit différente de  $\rho$ . Par conséquent, on s'attend à ce que la probabilité que l'intervalle de confiance recouvre la valeur de  $\rho$  soit inférieure à  $1 - \alpha$  pour de grandes valeurs de  $k$ .

Enfin, on construit un intervalle de confiance asymptotique pour  $\rho$  à partir de  $\bar{\rho}(\tilde{T}_Z)$ .

Rappelons que

$$\bar{\rho}(\tilde{T}_Z) = \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_Z)}{k},$$

où  $\hat{\rho}_i(\tilde{T}_Z)$  est l'estimateur de  $\rho$  pour la  $i^{\text{ième}}$  épidémie lorsqu'on considère  $t_Z$  inconnu et qu'on l'estime à partir des données de toutes les épidémies,  $i = 1, 2, \dots, k$ . Étant donné que chacun des  $\hat{\rho}_i(\tilde{T}_Z)$  dépend de  $\tilde{T}_Z$ , les  $\hat{\rho}_i(\tilde{T}_Z)$  sont dépendants. Contrairement à  $\bar{\rho}(t_Z)$  et  $\bar{\rho}(\mathbf{T}_Z)$ ,  $\bar{\rho}(\tilde{T}_Z)$  n'est donc pas une moyenne de variables indépendantes, mais bien une moyenne de variables dépendantes et identiquement distribuées. Cependant, on peut montrer que le théorème de limite centrale s'applique quand même à  $\bar{\rho}(\tilde{T}_Z)$ . La preuve

détaillée est présentée à la section 3.5.5. L'idée est qu'au-delà d'une certaine valeur de  $k$ , notée  $k_0$ ,  $\tilde{T}_Z = t_Z$  avec probabilité 1. On a alors que  $\bar{\rho}(\tilde{T}_Z) = \bar{\rho}(t_Z)$  avec probabilité 1. On a donc que

$$\sqrt{k}[\bar{\rho}(\tilde{T}_Z) - \rho]$$

converge en loi vers une loi normale

$$\mathcal{N}(0, \text{Var}[\hat{\rho}_1(t_Z)]),$$

où  $\text{Var}[\hat{\rho}_1(t_Z)]$  est donnée par les équations (3.4) ou (3.5) selon le modèle considéré. L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $\rho$  construit à partir de  $\bar{\rho}(\tilde{T}_Z)$  est donc donné par :

$$\left( \bar{\rho}(\tilde{T}_Z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}}, \bar{\rho}(\tilde{T}_Z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}} \right).$$

### 3.4.2 Intervalles de confiance asymptotiques pour $R_0$

Les intervalles de confiance asymptotiques pour  $R_0$  sont construits en se basant sur les deux équations suivantes :

$$R_0 = \frac{x_0}{\rho},$$

$$\tilde{R}_0 = \frac{x_0}{\bar{\rho}},$$

où  $\tilde{R}_0$  et  $\bar{\rho}$  sont les estimateurs de  $R_0$  et  $\rho$  respectivement.

On démontre d'abord le théorème qu'on va utiliser pour construire les intervalles de confiance asymptotiques pour  $R_0$ . Pour ce faire, on a besoin du théorème de Slutsky, qui est énoncé sans démonstration.

**Lemme 3.4.1** (Bickel et Doksum, 2001, Appendice A). *Si  $X_n$  converge en probabilité vers  $c$  et si  $Y_n$  converge en loi vers  $Y$ , alors  $X_n Y_n$  converge en loi vers  $cY$ .*

**Théorème 3.4.1.** *Soit  $T_n$  une statistique qui converge en probabilité vers  $\mu_t$  et qui est telle que l'on a la convergence*

$$\sqrt{n}(T_n - \mu_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2).$$

Soit  $h(\cdot)$  une fonction telle que  $h'(\cdot)$  est continue et  $h'(\mu_t) \neq 0$ . Alors, on a la convergence

$$\sqrt{n} \left[ h(T_n) - h(\mu_t) \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2 [h'(\mu_t)]^2).$$

*Démonstration.* Soit  $T_n^*$  tel que

$$h(T_n) = h(\mu_t) + h'(T_n^*)(T_n - \mu_t), \quad (3.31)$$

où  $\mu_t < T_n^* < T_n$  ou  $T_n < T_n^* < \mu_t$ . Comme  $T_n$  converge en probabilité vers  $\mu_t$ , on a que  $T_n^*$  converge en probabilité vers  $\mu_t$ . Comme  $h'(\cdot)$  est continue, on a que  $h'(T_n^*)$  converge en probabilité vers  $h'(\mu_t)$ . Par le lemme 3.4.1, étant donné que  $h'(T_n^*)$  converge en probabilité vers  $h'(\mu_t)$  et que l'on a la convergence

$$\sqrt{n}(T_n - \mu_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2),$$

on a la convergence

$$h'(T_n^*) \left[ \sqrt{n}(T_n - \mu_t) \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2 [h'(\mu_t)]^2).$$

Or, de l'équation (3.31), on a que

$$\sqrt{n} \left[ h(T_n) - h(\mu_t) \right] = \sqrt{n} h'(T_n^*)(T_n - \mu_t),$$

ce qui entraîne que l'on a la convergence

$$\sqrt{n} \left[ h(T_n) - h(\mu_t) \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_t^2 [h'(\mu_t)]^2).$$

□

Soit

$$\begin{aligned} \tilde{R}_0(t_z) &= \frac{x_0}{\tilde{\rho}(t_z)}, \\ \tilde{R}_0(\mathbf{T}_z) &= \frac{x_0}{\tilde{\rho}(\mathbf{T}_z)}, \\ \tilde{R}_0(\tilde{T}_z) &= \frac{x_0}{\tilde{\rho}(\tilde{T}_z)}, \end{aligned}$$

les estimateurs de  $R_0$  obtenus à partir des différents estimateurs de  $\rho$ . Afin d'appliquer le théorème 3.4.1 pour construire les intervalles de confiance asymptotiques pour  $R_0$ , on définit la fonction  $h(\cdot)$  de la façon suivante :

$$h(u) = \frac{x_0}{u},$$

$u > 0$ . La fonction  $h(\cdot)$  est telle que  $R_0 = h(\rho)$  et  $\tilde{R}_0 = h(\bar{\rho})$ , où  $\tilde{R}_0$  et  $\bar{\rho}$  sont les estimateurs de  $R_0$  et  $\rho$  respectivement. On a que

$$h'(u) = -\frac{x_0}{u^2},$$

$u > 0$ .

Tout d'abord, on construit un intervalle de confiance asymptotique pour  $R_0$  à partir de  $\tilde{R}_0(t_z)$ . Étant donné que  $\mathbb{E}[\bar{\rho}(t_z)] = \rho$  et que  $\text{Var}[\bar{\rho}(t_z)]$  tend vers 0 lorsque  $k$  tend vers l'infini, on a que  $\bar{\rho}(t_z)$  converge en probabilité vers  $\rho$ . De plus, on a vu à la section 3.4.1 que l'on a la convergence

$$\sqrt{k}[\bar{\rho}(t_z) - \rho] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}[\hat{\rho}_1(t_z)]).$$

Par le théorème 3.4.1, on a la convergence

$$\sqrt{k}[\tilde{R}_0(t_z) - R_0] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{x_0^2}{\rho^4} \text{Var}[\hat{\rho}_1(t_z)]\right).$$

L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $R_0$  construit à partir de  $\tilde{R}_0(t_z)$  est donc donné par :

$$\left( \tilde{R}_0(t_z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}}, \tilde{R}_0(t_z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}} \right).$$

On construit ensuite un intervalle de confiance asymptotique pour  $R_0$  à partir de  $\tilde{R}_0(\mathbf{T}_z)$ . Rappelons que l'espérance de  $\bar{\rho}(\mathbf{T}_z)$  n'est pas égale à  $\rho$  et que son expression ne dépend pas de  $k$ . Comme la variance de  $\bar{\rho}(\mathbf{T}_z)$  tend vers 0 lorsque  $k$  tend vers l'infini, on a que  $\bar{\rho}(\mathbf{T}_z)$  converge en probabilité vers  $\mathbb{E}[\bar{\rho}(\mathbf{T}_z)] \neq \rho$ . De plus, on a vu à la section 3.4.1 que l'on a la convergence

$$\sqrt{k}[\bar{\rho}(\mathbf{T}_z) - \mathbb{E}[\bar{\rho}(\mathbf{T}_z)]] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}[\hat{\rho}_1(T_z^{(1)})]).$$

Par le théorème 3.4.1, on a la convergence

$$\sqrt{k} \left[ \tilde{R}_0(\mathbf{T}_z) - \frac{x_0}{\mathbb{E}[\tilde{\rho}(\mathbf{T}_z)]} \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{x_0^2}{\mathbb{E}[\tilde{\rho}(\mathbf{T}_z)]^4} \text{Var}[\hat{\rho}_1(T_z^{(1)})] \right).$$

Au chapitre 4, on verra empiriquement que  $\mathbb{E}[\tilde{\rho}(\mathbf{T}_z)] \approx \rho$ . En supposant que  $\mathbb{E}[\tilde{\rho}(\mathbf{T}_z)] = \rho$ , on a la convergence

$$\sqrt{k} \left[ \tilde{R}_0(\mathbf{T}_z) - R_0 \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{x_0^2}{\rho^4} \text{Var}[\hat{\rho}_1(T_z^{(1)})] \right).$$

L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $R_0$  construit à partir de  $\tilde{R}_0(\mathbf{T}_z)$  est donc donné par :

$$\left( \tilde{R}_0(\mathbf{T}_z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_z^{(1)})]}{k}}, \tilde{R}_0(\mathbf{T}_z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_z^{(1)})]}{k}} \right).$$

Comme pour l'intervalle de confiance pour  $\rho$  construit à partir de  $\tilde{\rho}(\mathbf{T}_z)$ , on s'attend à ce que la probabilité que l'intervalle de confiance pour  $R_0$  construit à partir de  $\tilde{R}_0(\mathbf{T}_z)$  contienne la valeur de  $R_0$  soit différente de  $1 - \alpha$ .

Enfin, on construit un intervalle de confiance asymptotique pour  $R_0$  à partir de  $\tilde{R}_0(\tilde{T}_z)$ . On a mentionné à la section 3.3.2 que  $\tilde{\rho}(\tilde{T}_z)$  converge presque sûrement vers  $\rho$ . La preuve est d'ailleurs présentée à la section 3.5.3. De plus, on a mentionné à la section 3.4.1 que l'on a la convergence

$$\sqrt{k} \left[ \tilde{\rho}(\tilde{T}_z) - \rho \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}[\hat{\rho}_1(t_z)]).$$

La preuve est présentée à la section 3.5.5. Par le théorème 3.4.1, on a la convergence

$$\sqrt{k} \left[ \tilde{R}_0(\tilde{T}_z) - R_0 \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{x_0^2}{\rho^4} \text{Var}[\hat{\rho}_1(t_z)] \right).$$

L'intervalle de confiance asymptotique de niveau  $(1 - \alpha) \cdot 100\%$  pour  $R_0$  construit à partir de  $\tilde{R}_0(\tilde{T}_z)$  est donc donné par :

$$\left( \tilde{R}_0(\tilde{T}_z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}}, \tilde{R}_0(\tilde{T}_z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_z)]}{k}} \right).$$

### 3.4.3 Estimation de la variance des estimateurs de $\rho$

En pratique, on ne connaît pas les variances de  $\hat{\rho}_1(t_z)$  et  $\hat{\rho}_1(T_z^{(1)})$ . On doit donc les estimer afin de construire les intervalles de confiance asymptotiques pour  $\rho$  et  $R_0$  présentés aux sections 3.4.1 et 3.4.2.

On utilise la variance de  $\hat{\rho}_1(t_z)$  dans les intervalles de confiance asymptotiques pour  $\rho$  construits à partir de  $\bar{\rho}(t_z)$  et dans ceux construits à partir de  $\bar{\rho}(\tilde{T}_z)$ . On propose d'estimer la variance de  $\hat{\rho}_1(t_z)$  à partir des  $\hat{\rho}_i(t_z)$  de la façon suivante :

$$S(t_z) = \sum_{i=1}^k \frac{[\hat{\rho}_i(t_z) - \bar{\rho}(t_z)]^2}{k-1},$$

alors qu'on propose d'estimer la variance de  $\hat{\rho}_1(t_z)$  à partir des  $\hat{\rho}_i(\tilde{T}_z)$  de la façon suivante :

$$S(\tilde{T}_z) = \sum_{i=1}^k \frac{[\hat{\rho}_i(\tilde{T}_z) - \bar{\rho}(\tilde{T}_z)]^2}{k-1}.$$

Par contre, on utilise la variance de  $\hat{\rho}_1(T_z^{(1)})$  dans l'intervalle de confiance asymptotique pour  $\rho$  construit à partir de  $\bar{\rho}(\mathbf{T}_z)$ . On propose d'estimer la variance de  $\hat{\rho}_1(T_z^{(1)})$  à partir des  $\hat{\rho}_i(T_z^{(i)})$  de la façon suivante :

$$S(\mathbf{T}_z) = \sum_{i=1}^k \frac{[\hat{\rho}_i(T_z^{(i)}) - \bar{\rho}(\mathbf{T}_z)]^2}{k-1}.$$

Il est clair que les estimateurs  $S(t_z)$  et  $S(\mathbf{T}_z)$  sont sans biais pour  $\text{Var}[\hat{\rho}_1(t_z)]$  et  $\text{Var}[\hat{\rho}_1(T_z^{(1)})]$  respectivement. L'espérance de  $S(\tilde{T}_z)$  est, quant à elle, donnée par :

$$\begin{aligned} \mathbb{E}[S(\tilde{T}_z)] &= \mathbb{E}(\mathbb{E}[S(\tilde{T}_z)|\tilde{T}_z]) \\ &= \mathbb{E} \left( \mathbb{E} \left[ \sum_{i=1}^k \frac{[\hat{\rho}_i(\tilde{T}_z) - \bar{\rho}(\tilde{T}_z)]^2}{k-1} \middle| \tilde{T}_z \right] \right) \\ &= \mathbb{E}(\text{Var}[\hat{\rho}_1(\tilde{T}_z)|\tilde{T}_z]), \end{aligned}$$

car

$$\mathbb{E} \left[ \sum_{i=1}^k \frac{[\hat{\rho}_i(\tilde{T}_z) - \bar{\rho}(\tilde{T}_z)]^2}{k-1} \middle| \tilde{T}_z \right] = \text{Var}[\hat{\rho}_1(\tilde{T}_z)|\tilde{T}_z].$$

Au chapitre 4, on verra empiriquement que  $\mathbb{E}(\text{Var}[\hat{\rho}_1(\tilde{T}_z)|\tilde{T}_z]) \approx \text{Var}[\hat{\rho}_1(t_z)]$ . On a donc que  $S(\tilde{T}_z)$  est approximativement sans biais pour  $\text{Var}[\hat{\rho}_1(t_z)]$ .

On a donc les intervalles de confiance asymptotiques de niveau  $(1 - \alpha) \cdot 100\%$  suivants pour  $\rho$  construits à partir des différents estimateurs de  $\rho$  et avec la variance estimée :

$$\begin{aligned} & \left( \bar{\rho}(t_z) - z_{\alpha/2} \sqrt{\frac{S(t_z)}{k}}, \bar{\rho}(t_z) + z_{\alpha/2} \sqrt{\frac{S(t_z)}{k}} \right), \\ & \left( \bar{\rho}(\mathbf{T}_z) - z_{\alpha/2} \sqrt{\frac{S(\mathbf{T}_z)}{k}}, \bar{\rho}(\mathbf{T}_z) + z_{\alpha/2} \sqrt{\frac{S(\mathbf{T}_z)}{k}} \right), \\ & \left( \bar{\rho}(\tilde{T}_z) - z_{\alpha/2} \sqrt{\frac{S(\tilde{T}_z)}{k}}, \bar{\rho}(\tilde{T}_z) + z_{\alpha/2} \sqrt{\frac{S(\tilde{T}_z)}{k}} \right). \end{aligned}$$

Pour les intervalles de confiance asymptotiques pour  $R_0$ , en plus d'estimer la variance de l'estimateur de  $\rho$ , on doit également estimer  $\rho$ , car il apparaît dans la formule de ces intervalles de confiance. Naturellement, on propose de l'estimer par l'estimateur de  $\rho$  à partir duquel l'intervalle de confiance a été construit. On a donc les intervalles de confiance asymptotiques suivants pour  $R_0$  construits à partir des différents estimateurs de  $R_0$  et avec la variance estimée :

$$\begin{aligned} & \left( \tilde{R}_0(t_z) - z_{\alpha/2} \frac{x_0}{\bar{\rho}(t_z)^2} \sqrt{\frac{S(t_z)}{k}}, \tilde{R}_0(t_z) + z_{\alpha/2} \frac{x_0}{\bar{\rho}(t_z)^2} \sqrt{\frac{S(t_z)}{k}} \right), \\ & \left( \tilde{R}_0(\mathbf{T}_z) - z_{\alpha/2} \frac{x_0}{\bar{\rho}(\mathbf{T}_z)^2} \sqrt{\frac{S(\mathbf{T}_z)}{k}}, \tilde{R}_0(\mathbf{T}_z) + z_{\alpha/2} \frac{x_0}{\bar{\rho}(\mathbf{T}_z)^2} \sqrt{\frac{S(\mathbf{T}_z)}{k}} \right), \\ & \left( \tilde{R}_0(\tilde{T}_z) - z_{\alpha/2} \frac{x_0}{\bar{\rho}(\tilde{T}_z)^2} \sqrt{\frac{S(\tilde{T}_z)}{k}}, \tilde{R}_0(\tilde{T}_z) + z_{\alpha/2} \frac{x_0}{\bar{\rho}(\tilde{T}_z)^2} \sqrt{\frac{S(\tilde{T}_z)}{k}} \right), \end{aligned}$$

### 3.5 Calculs supplémentaires et démonstrations

#### 3.5.1 Calcul de la fonction de masse de $T_z$

Rappelons d'abord que  $T_z$  est l'estimateur de  $t_z$  obtenu à partir des données d'une seule épidémie. Ces données sont  $L_D(1), L_D(2), \dots, L_D(n)$ . Soit

$$\mathbf{L}_D = \{L_D(\tau), \tau = 1, 2, \dots, n\}.$$

À la section 2.3.1, on a montré que  $\mathbf{L}_D$  est un ensemble de  $n$  variables aléatoires indépendantes de loi de Poisson. Rappelons ensuite que  $T_Z$  est calculé en arrondissant la moyenne des valeurs de  $\tau$  pour lesquelles

$$L_D(\tau) = \max(\mathbf{L}_D). \quad (3.32)$$

$T_Z$  peut donc prendre des valeurs entières entre 1 et  $n$  inclusivement. Supposons que l'on arrondisse en prenant la partie entière. Soit  $M$  le nombre de valeurs de  $\tau$  qui satisfont l'équation (3.32).  $M$  peut prendre des valeurs entières entre 1 et  $n$  inclusivement. Toutefois, on verra empiriquement au chapitre 4 que la probabilité que le maximum soit atteint pour plus de 5 valeurs de  $\tau$  est très faible.

Soit  $k \in \mathbb{N}$  tel que  $1 \leq k \leq n$ . Afin de calculer la probabilité que  $T_Z$  prenne la valeur  $k$ , on doit considérer tous les événements qui font en sorte que  $T_Z$  soit égal à  $k$  et que la variable aléatoire  $M$  soit égale à  $m$ , pour  $m = 1, 2, \dots, n$ . Tout d'abord, on considère le cas où il n'y a qu'une seule valeur de  $\tau$  pour laquelle le maximum est atteint et qui satisfait donc l'équation (3.32). Cela correspond au cas où  $M = 1$ . Dans ce cas, le seul événement pour lequel  $T_Z$  est égal à  $k$  est celui pour lequel  $L_D(k)$  est supérieure à toutes les autres variables aléatoires de l'ensemble  $\mathbf{L}_D$ . La probabilité que  $T_Z$  soit égal à  $k$  et que la variable aléatoire  $M$  soit égale à 1 se calcule donc comme suit :

$$P[T_Z = k, M = 1] = \prod_{\substack{j=1 \\ j \neq k}}^n P[L_D(k) > L_D(j)].$$

Ensuite, on considère les cas où  $T_Z$  prend la valeur  $k$  et  $M$  prend la valeur 2. En général, plusieurs événements correspondent à ce cas. En voici d'ailleurs quelques exemples. Pour chacun, on donne les valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximale, ainsi que leur moyenne :

- (i) pour  $\tau = k - 1$  et  $\tau = k + 1$ , la moyenne est  $k$  ;
- (ii) pour  $\tau = k$  et  $\tau = k + 1$ , la moyenne est

$$\frac{k + k + 1}{2} = k + \frac{1}{2},$$

dont la partie entière est  $k$  ;

(iii) pour  $\tau = k - 2$  et  $\tau = k + 2$ , la moyenne est  $k$  ;

(iv) pour  $\tau = k - 1$  et  $\tau = k + 2$ , la moyenne est

$$\frac{k - 1 + k + 2}{2} = k + \frac{1}{2},$$

dont la partie entière est  $k$ .

On peut faire la liste de tous les événements pour lesquels  $T_Z = k$  et  $M = 2$  et calculer leur probabilité respective. Par exemple, considérons un de ces événements, qui est le cas où les valeurs de  $L_D(k - 1)$  et  $L_D(k + 1)$  sont égales et où les valeurs de  $L_D(\tau)$  pour toutes les autres valeurs de  $\tau$  leur sont inférieures. En notant cet événement par  $A(k - 1, k + 1)$ , on peut calculer sa probabilité comme suit :

$$P[A(k - 1, k + 1)] = \sum_{i=1}^{\infty} \left( P[L_D(k - 1) = i] P[L_D(k + 1) = i] \prod_{\substack{j=1 \\ j \notin \{k-1, k+1\}}}^n P[L_D(j) < i] \right).$$

On calcule  $P[T_Z = k, M = 2]$  en faisant la somme des probabilités de tous les événements pour lesquels  $T_Z$  prend la valeur  $k$  et  $M$  prend la valeur 2. Par la suite, on considère successivement les cas où  $M = 3$ ,  $M = 4$ , jusqu'à  $M = n$ . À partir des probabilités calculées pour chaque valeur de  $M$ , la fonction de masse de  $T_Z$  se calcule comme suit :

$$P[T_Z = k] = \sum_{m=1}^n P[T_Z = k, M = m], \quad (3.33)$$

$k = 1, 2, \dots, n$ .

Afin d'illustrer le calcul de la fonction de masse de  $T_Z$ , le cas où  $n = 3$  est présenté au complet. Dans cet exemple, lorsque la moyenne des valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximale n'est pas égale à un nombre entier, on prend sa partie entière. Soit  $L_D(1)$ ,  $L_D(2)$  et  $L_D(3)$  les trois observations du nombre de nouveaux individus décédés. On veut calculer la fonction de masse de  $T_Z$  dans ce cas. On commence par calculer la

probabilité que  $T_Z$  soit égal à 1 :

$$\begin{aligned}
P[T_Z = 1] &= \sum_{m=1}^3 P[T_Z = 1, M = m] \\
&= P[T_Z = 1, M = 1] + P[T_Z = 1, M = 2] + P[T_Z = 1, M = 3] \\
&= P\left[\left(L_D(1) > L_D(2)\right), \left(L_D(1) > L_D(3)\right)\right] \\
&\quad + P\left[\left(L_D(1) = L_D(2)\right), \left(L_D(1) > L_D(3)\right)\right] \\
&\quad + 0,
\end{aligned}$$

car la moyenne de 1 et 2 est égale à 1,5 et la partie entière de 1,5 est 1. Ensuite, on calcule la probabilité que  $T_Z$  soit égal à 2 :

$$\begin{aligned}
P[T_Z = 2] &= \sum_{m=1}^3 P[T_Z = 2, M = m] \\
&= P[T_Z = 2, M = 1] + P[T_Z = 2, M = 2] + P[T_Z = 2, M = 3] \\
&= P\left[\left(L_D(2) > L_D(1)\right), \left(L_D(2) > L_D(3)\right)\right] \\
&\quad + P\left[\left(L_D(2) = L_D(3)\right), \left(L_D(2) > L_D(1)\right)\right] \\
&\quad + P\left[L_D(1) = L_D(2), L_D(2) = L_D(3)\right],
\end{aligned}$$

car la moyenne de 2 et 3 est égale à 2,5 et la partie entière de 2,5 est 2. Enfin, on calcule la probabilité que  $T_Z$  soit égal à 3 :

$$\begin{aligned}
P[T_Z = 3] &= \sum_{m=1}^3 P[T_Z = 3, M = m] \\
&= P[T_Z = 3, M = 1] + P[T_Z = 3, M = 2] + P[T_Z = 3, M = 3] \\
&= P\left[\left(L_D(3) > L_D(1)\right), \left(L_D(3) > L_D(2)\right)\right] \\
&\quad + 0 + 0.
\end{aligned}$$

### 3.5.2 Calcul de la fonction de masse de $\tilde{T}_Z$

Rappelons d'abord que  $\tilde{T}_Z$  est l'estimateur de  $t_Z$  obtenu à partir des données de  $k$  épidémies.  $\tilde{T}_Z$  est calculé en prenant la partie entière de la moyenne de  $T_Z^{(1)}, T_Z^{(2)}, \dots,$

$T_Z^{(k)}$ , où  $T_Z^{(i)}$  est l'estimateur de  $t_Z$  obtenu à partir des données de la  $i^{\text{ième}}$  épidémie.  $\tilde{T}_Z$  peut donc prendre des valeurs entières entre 1 et  $n$  inclusivement. Afin de calculer la probabilité que  $\tilde{T}_Z$  soit égal à  $k$ , on fait la somme des probabilités de tous les événements qui font en sorte que  $\tilde{T}_Z$  soit égal à  $k$ .

Au chapitre 4, on explique la façon de calculer numériquement la fonction de masse de  $T_Z$  et de  $\tilde{T}_Z$ .

### 3.5.3 Démonstration de la convergence presque sûre de $\tilde{T}_Z$ vers $t_Z$

Pour la présentation de la preuve que  $\tilde{T}_Z$  est un estimateur fortement convergent pour  $t_Z$ , c'est-à-dire que l'on a la convergence presque sûre de  $\tilde{T}_Z$  vers  $t_Z$ , on introduit une notation quelque peu différente de celle utilisée à la section 3.3.2. Rappelons que  $T_Z^{(i)}$  est l'estimateur de  $t_Z$  obtenu à partir des données de la  $i^{\text{ième}}$  épidémie. On note par  $\tau^*$  l'espérance de  $T_Z^{(i)}$ ,  $i = 1, 2, \dots$ . Soit

$$\bar{T}_Z^k = \frac{\sum_{i=1}^k T_Z^{(i)}}{k}$$

et soit  $\tilde{T}_Z^k$  la partie entière de  $\bar{T}_Z^k$ . Formellement, la convergence presque sûre de  $\tilde{T}_Z^k$  vers  $t_Z$  signifie que

$$P\left[\lim_{k \rightarrow \infty} \tilde{T}_Z^k = t_Z\right] = 1. \quad (3.34)$$

Dans notre cas, on a ce résultat plus fort : pour tout  $\delta > 0$ , il existe  $k^*$  tel que, pour  $k \geq k^*$ , on a

$$P\left[\tilde{T}_Z^m = t_Z, \forall m \geq k\right] > 1 - \delta. \quad (3.35)$$

On va montrer (3.35) sous la supposition que

$$t_Z < \tau^* < t_Z + 1. \quad (3.36)$$

Comme les variables  $T_Z^{(i)}$ ,  $i = 1, 2, \dots$ , sont indépendantes, identiquement distribuées et de moyenne  $\tau^*$ , on a la convergence presque sûre de  $\bar{T}_Z^k$  vers  $\tau^*$ , c'est-à-dire

$$P\left[\lim_{k \rightarrow \infty} \bar{T}_Z^k = \tau^*\right] = 1. \quad (3.37)$$

Une définition équivalente à (3.37) est : pour tout  $\epsilon > 0$ , on a que

$$\lim_{k \rightarrow \infty} P\left[|\bar{T}_Z^m - \tau^*| < \epsilon, \forall m \geq k\right] = 1.$$

Soit  $\epsilon > 0$  fixé. Alors, pour tout  $\delta > 0$ , il existe  $k^{**}$  tel que, pour  $k \geq k^{**}$ , on a que

$$P\left[|\bar{T}_Z^m - \tau^*| < \epsilon, \forall m \geq k\right] > 1 - \delta.$$

Fixons

$$\epsilon_0 = \min\left(\frac{\tau^* - t_Z}{2}, \frac{t_Z + 1 - \tau^*}{2}\right)$$

et soit  $\delta > 0$ . Alors, il existe  $k_0$  tel que, pour  $k \geq k_0$ , on a que

$$\begin{aligned} P\left[|\bar{T}_Z^m - \tau^*| < \epsilon_0, \forall m \geq k\right] &> 1 - \delta \\ \Leftrightarrow P\left[\tau^* - \epsilon_0 < \bar{T}_Z^m < \tau^* + \epsilon_0, \forall m \geq k\right] &> 1 - \delta \\ \Leftrightarrow P\left[t_Z < \tau^* - \epsilon_0 < \bar{T}_Z^m < \tau^* + \epsilon_0 < t_Z + 1, \forall m \geq k\right] &> 1 - \delta. \end{aligned} \quad (3.38)$$

Forcément, en prenant en (3.38) la partie entière de  $\bar{T}_Z^m$ , il ne reste qu'un choix pour  $\tilde{T}_Z^m$ , c'est-à-dire  $t_Z$ . Donc, pour  $k \geq k_0$ , on a que

$$P[\tilde{T}_Z^m = t_Z, \forall m \geq k] > 1 - \delta \quad (3.39)$$

car

$$\{\tau^* - \epsilon_0 < \bar{T}_Z^m < \tau^* + \epsilon_0\} \subset \{\tilde{T}_Z^m = t_Z\}.$$

En répétant pour tout  $\delta$ , l'inégalité (3.39) donne que

$$P\left[\lim_{k \rightarrow \infty} \tilde{T}_Z^k = t_Z\right] = 1.$$

$\tilde{T}_Z$  est donc un estimateur fortement convergent pour  $t_Z$ . En fait, on a même l'équation (3.35).

### 3.5.4 Démonstration de la convergence presque sûre de $\bar{\rho}(\tilde{T}_Z)$ vers $\rho$

Pour la présentation de la preuve que  $\bar{\rho}(\tilde{T}_Z)$  est un estimateur fortement convergent pour  $\rho$ , c'est-à-dire que l'on a la convergence presque sûre de  $\bar{\rho}(\tilde{T}_Z)$  vers  $\rho$ , on introduit

une notation quelque peu différente de celle utilisée à la section 3.3.2. Rappelons qu'à la section 3.5.3, on a noté par  $\tilde{T}_z^k$  l'estimateur de  $t_z$  obtenu à partir des données des épidémies 1 à  $k$ . Soit

$$\bar{\rho}_k(t_z) = \frac{\sum_{i=1}^k \hat{\rho}_i(t_z)}{k}$$

et

$$\bar{\rho}(\tilde{T}_z^k) = \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_z^k)}{k},$$

où  $\hat{\rho}_i(t_z)$  et  $\hat{\rho}_i(\tilde{T}_z^k)$  sont les estimateurs de  $\rho$  obtenus à partir de  $t_z$  et  $\tilde{T}_z^k$  respectivement pour la  $i^{\text{ième}}$  épidémie. Formellement, la convergence presque sûre de  $\bar{\rho}(\tilde{T}_z^k)$  vers  $\rho$  signifie que

$$P\left[\lim_{k \rightarrow \infty} \bar{\rho}(\tilde{T}_z^k) = \rho\right] = 1. \quad (3.40)$$

On va montrer (3.40) sous la supposition que  $\tilde{T}_z^k$  satisfait l'équation (3.35). Soit  $\epsilon > 0$  et  $\delta > 0$ . Alors, il existe  $k^*$  tel que, pour  $k \geq k^*$ , on a que

$$P\left[\tilde{T}_z^m = t_z, \forall m \geq k\right] > 1 - \frac{\delta}{2}. \quad (3.41)$$

Comme les variables aléatoires  $\hat{\rho}_i(t_z)$ ,  $i = 1, 2, \dots, k$ , sont indépendantes, identiquement distribuées et de moyenne  $x(t_z) = \rho$ , il existe  $k^{**}$  tel que pour  $k \geq k^{**}$ , on a que

$$P\left[|\bar{\rho}_m(t_z) - \rho| < \epsilon, \forall m \geq k\right] > 1 - \frac{\delta}{2}. \quad (3.42)$$

Considérons l'intersection des événements en (3.41) et (3.42) pour  $k \geq \max(k^*, k^{**})$ . Étant donné que sur l'intersection,  $\tilde{T}_z^m$  est égal à  $t_z$ , l'intersection de ces événements est équivalente à l'événement suivant :

$$|\bar{\rho}(\tilde{T}_z^m) - \rho| < \epsilon, \forall m \geq k.$$

En utilisant l'inégalité

$$P[A \cap B] \geq P[A] + P[B] - 1, \quad (3.43)$$

pour les événements en (3.41) et (3.42), on a que, pour  $k \geq \max(k^*, k^{**})$ ,

$$\begin{aligned} & P\left[\left(\tilde{T}_z^m = t_z, \forall m \geq k\right), \left(|\bar{\rho}_m(t_z) - \rho| < \epsilon, \forall m \geq k\right)\right] \\ & \geq P\left[\tilde{T}_z^m = t_z, \forall m \geq k\right] + P\left[|\bar{\rho}_m(t_z) - \rho| < \epsilon, \forall m \geq k\right] - 1, \end{aligned}$$

ce qui donne, en vertu de (3.41) et (3.42),

$$\begin{aligned} P\left[|\bar{\rho}(\tilde{T}_Z^m) - \rho| < \epsilon, \forall m \geq k\right] &> \left(1 - \frac{\delta}{2}\right) + \left(1 - \frac{\delta}{2}\right) - 1 \\ \Leftrightarrow P\left[|\bar{\rho}(\tilde{T}_Z^m) - \rho| < \epsilon, \forall m \geq k\right] &> 1 - \delta. \end{aligned} \quad (3.44)$$

En répétant pour tout  $\delta$ , l'inégalité (3.44) donne que

$$\lim_{k \rightarrow \infty} P\left[|\bar{\rho}(\tilde{T}_Z^m) - \rho| < \epsilon, \forall m \geq k\right] = 1. \quad (3.45)$$

Enfin, comme (3.45) est vraie pour tout  $\epsilon > 0$ , on a que

$$P\left[\lim_{k \rightarrow \infty} \bar{\rho}(\tilde{T}_Z^k) = \rho\right] = 1.$$

$\bar{\rho}(\tilde{T}_Z^k)$  est donc un estimateur fortement convergent pour  $\rho$ .

### 3.5.5 Démonstration de la convergence en loi de $\bar{\rho}(\tilde{T}_Z)$ vers une loi normale

Dans cette section, on montre que l'on a la convergence

$$\sqrt{k}\left[\bar{\rho}(\tilde{T}_Z) - \rho\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}[\hat{\rho}_1(t_Z)]).$$

Rappelons qu'à la section 3.5.3, on a noté par  $\tilde{T}_Z^k$  l'estimateur de  $t_Z$  obtenu à partir des données des épidémies 1 à  $k$ . À la section 3.5.4, on a défini

$$\bar{\rho}_k(t_Z) = \frac{\sum_{i=1}^k \hat{\rho}_i(t_Z)}{k},$$

et

$$\bar{\rho}(\tilde{T}_Z^k) = \frac{\sum_{i=1}^k \hat{\rho}_i(\tilde{T}_Z^k)}{k},$$

où  $\hat{\rho}_i(t_Z)$  et  $\hat{\rho}_i(\tilde{T}_Z^k)$  sont les estimateurs de  $\rho$  obtenus à partir de  $t_Z$  et  $\tilde{T}_Z^k$  respectivement pour la  $i^{\text{ième}}$  épidémie,  $i = 1, 2, \dots, k$ .

On définit, pour  $k = 1, 2, \dots$  et  $u \in \mathbb{R}$  fixé :

$$\begin{aligned}\tilde{U}_k &= \frac{\sqrt{k}[\bar{\rho}(\tilde{T}_Z^k) - \rho]}{\sqrt{\text{Var}[\hat{\rho}_1(t_Z)]}}, \\ \tilde{V}_k &= \frac{\sqrt{k}[\bar{\rho}_k(t_Z) - \rho]}{\sqrt{\text{Var}(\hat{\rho}_1(t_Z))}}, \\ A_k &= \{\tilde{U}_k \leq u\}, \\ B_k &= \{\tilde{V}_k \leq u\}, \\ C_k &= \{\tilde{T}_Z^m = t_Z, \forall m \geq k\}.\end{aligned}$$

À la section 3.5.3, on a montré que, si  $k \rightarrow \infty$ ,

$$P[C_k] \rightarrow 1,$$

d'où

$$P[\bar{C}_k] \rightarrow 0,$$

où  $\bar{C}_k$  est le complément de  $C_k$ . De plus, comme  $\bar{\rho}_k(t_Z)$  est une moyenne de  $k$  variables aléatoires indépendantes et identiquement distribuées d'espérance  $\rho$  et de variance  $\text{Var}[\hat{\rho}_1(t_Z)]$ , on a que, si  $k \rightarrow \infty$ ,

$$P[B_k] \rightarrow \Phi(u)$$

où  $\Phi(u)$  est la fonction de répartition de la loi  $\mathcal{N}(0, 1)$ . On veut montrer que, si  $k \rightarrow \infty$ ,

$$P[A_k] \rightarrow \Phi(u).$$

De l'équation (3.43), on a que

$$P[B_k] + P[C_k] - 1 \leq P[B_k \cap C_k] \leq P[B_k].$$

Cela implique que, si  $k \rightarrow \infty$ ,

$$P[B_k \cap C_k] \rightarrow \Phi(u). \quad (3.46)$$

D'autre part, pour tout  $k$ ,

$$A_k \cap C_k = B_k \cap C_k,$$

ce qui implique que

$$P[A_k \cap C_k] = P[B_k \cap C_k].$$

Donc, si  $k \rightarrow \infty$ ,

$$P[A_k \cap C_k] \rightarrow \Phi(u). \quad (3.47)$$

Finalement, comme

$$P[A_k \cap \bar{C}_k] \leq P[\bar{C}_k],$$

on a que

$$\begin{aligned} \lim_{k \rightarrow \infty} P[A_k] &= \lim_{k \rightarrow \infty} P[A_k \cap \bar{C}_k] + \lim_{k \rightarrow \infty} P[A_k \cap C_k] \\ &= 0 + \Phi(u) \\ &= \Phi(u), \end{aligned}$$

tel que désiré.

- - - - -

## CHAPITRE IV

### APPLICATIONS NUMÉRIQUES

Dans ce chapitre, on présente les résultats de l'application de la méthode d'estimation de  $R_0$  du chapitre 3 sur des données simulées à partir des modèles présentés à la section 2.3.

Rappelons que la procédure d'estimation de  $R_0$  proposée passe par l'estimation de  $\rho$ . La méthode d'estimation de  $\rho$  se base sur l'équation (2.4), présentée à la section 2.2.1 :

$$\rho = x(t_z),$$

où  $t_z$  représente l'instant auquel  $z'(t)$  atteint son maximum. Trois estimateurs de  $\rho$  ont été proposés, chacun étant obtenu à partir des données de  $k$  épidémies. Pour le premier estimateur, on suppose que  $t_z$  est connu, alors qu'on le suppose inconnu pour les deux autres estimateurs. Le premier estimateur est donc obtenu à partir de  $t_z$  et est noté par  $\bar{\rho}(t_z)$ . Le second estimateur de  $\rho$  est obtenu à partir de  $\mathbf{T}_z = (T_z^{(1)}, T_z^{(2)}, \dots, T_z^{(k)})$ , où  $T_z^{(i)}$  est l'estimateur de  $t_z$  obtenu à partir des données de la  $i^{\text{ième}}$  épidémie. Cet estimateur est noté par  $\bar{\rho}(\mathbf{T}_z)$ . Enfin, le troisième estimateur de  $\rho$  est obtenu à partir de  $\tilde{T}_z$ , l'estimateur de  $t_z$  obtenu à partir des données de toutes les épidémies et est noté par  $\bar{\rho}(\tilde{T}_z)$ .

Dans ce chapitre, on explique d'abord le choix des valeurs des paramètres utilisées pour faire les calculs et les simulations, puis on fait une reparamétrisation en  $\beta$  et  $\gamma$  qui fait en sorte que  $t_z$  prenne une valeur entière. Dans un second temps, on présente les solutions du modèle déterministe correspondant aux paramètres obtenus. Par la suite, on explique la façon dont les valeurs du nombre d'individus susceptibles et décédés ont été générées

à partir des deux modèles et on fait quelques commentaires sur les données générées. On présente ensuite l'approximation de la fonction de masse de  $T_Z$  et de  $\tilde{T}_Z$ . Par la suite, on présente les propriétés des trois estimateurs de  $\rho$  obtenues numériquement, les propriétés des trois estimateurs de  $R_0$  obtenues en simulation et les résultats obtenus en simulation pour la proportion des intervalles de confiance asymptotiques pour  $\rho$  et  $R_0$  qui contiennent  $\rho$  et  $R_0$  respectivement. On termine avec une synthèse de tous les résultats.

#### 4.1 Valeurs de départ des paramètres

On considère une population de 100 000 individus, dont 10 sont initialement infectés par la maladie. Le choix de cette taille de population est justifiée par le fait que les données sont souvent présentées comme le nombre de cas parmi 100 000 habitants. On suppose qu'on observe des données sur la grippe saisonnière dans cette population. La valeur du paramètre  $R_0$  pour l'épidémie considérée est 1,5 et la probabilité qu'un individu retiré soit en fait décédé, notée  $p$ , est égale à 0,001. Ces deux valeurs sont raisonnables pour la grippe saisonnière. On observe des données pour une année à raison d'une observation par semaine. On a donc que  $n = 52$ . On considère que, en moyenne, un individu est infecté pendant 5 jours. Étant donné que l'inverse du paramètre  $\gamma$  représente la période de temps moyenne pendant laquelle un individu est infecté et que l'unité de temps considérée est la semaine, on a que  $\gamma = \frac{7}{5}$ . En résumé, les valeurs des paramètres sont :

$$N = 100\,000;$$

$$y_0 = 10;$$

$$R_0 = 1,5;$$

$$p = 0,001;$$

$$n = 52;$$

$$\gamma = \frac{7}{5}.$$

À partir de ces valeurs, on peut déduire celles des autres paramètres importants. Tout d'abord, comme

$$x_0 = N - y_0,$$

on a que  $x_0 = 999\,990$ . Ensuite, comme

$$\rho = \frac{x_0}{R_0},$$

on a que  $\rho = 66\,660$ . Enfin, comme

$$\beta = \frac{\gamma}{\rho},$$

on a que  $\beta = \frac{7}{333\,300}$ . Pour les calculs et les simulations, certaines des valeurs des paramètres ont été légèrement modifiées afin que la méthode d'estimation puisse être appliquée. Les détails à ce sujet sont d'ailleurs présentés à la section suivante.

## 4.2 Reparamétrisation en $\beta$ et $\gamma$

Tel que mentionné à la section 3.3, on doit avoir une paramétrisation en  $\beta$  et  $\gamma$  qui est telle que  $t_z$  prenne une valeur entière afin d'appliquer la première méthode d'estimation. Avec les valeurs initiales de  $\beta$  et  $\gamma$  données à la section 4.1, c'est-à-dire

$$\beta = \frac{7}{333\,300};$$

$$\gamma = \frac{7}{5},$$

on a que  $t_z = 11,47$ . La paramétrisation initiale en  $\beta$  et  $\gamma$  n'est donc pas telle que  $t_z$  prenne une valeur entière. On cherche une paramétrisation qui fait en sorte que  $t_z = 12$ . À la proposition 2.2.2, on a montré qu'en multipliant  $\beta$  et  $\gamma$  par

$$\lambda = \frac{11,47}{12} = 0,956,$$

on obtient une telle paramétrisation. Les valeurs de  $\beta$  et  $\gamma$  correspondantes sont

$$\beta = 2,007 \times 10^{-5},$$

$$\gamma = 1,338.$$

Ces valeurs sont légèrement différentes de celles de départ, mais leur interprétation reste essentiellement la même.

### 4.3 Solutions du modèle déterministe

Une fois qu'on a obtenu une paramétrisation en  $\beta$  et  $\gamma$  qui est telle que  $t_z$  prenne une valeur entière, on résout (2.1) afin d'obtenir les solutions  $x(t)$ ,  $y(t)$  et  $z(t)$ . Rappelons que  $x(t)$ ,  $y(t)$  et  $z(t)$  représentent respectivement le nombre d'individus susceptibles, infectés et retirés, chacun au temps  $t \geq 0$ . Ces solutions sont utilisées pour générer des réalisations des modèles stochastiques, pour calculer la fonction de masse de  $T_z$  et de  $\tilde{T}_z$  et pour obtenir les propriétés des estimateurs de  $\rho$ .

Le système d'équations différentielles (2.1) ne possède pas de solution explicite. Pour approximer numériquement les solutions de (2.1), on utilise la fonction « ode45 » du programme MATLAB. Pour ce faire, il faut spécifier un intervalle de temps sur lequel les solutions sont calculées. L'intervalle de temps choisi est  $[0, n]$ , car on veut générer des réalisations des modèles stochastiques jusqu'au temps  $n$ . Rappelons qu'on a fixé  $n = 52$ .

Les fonctions  $x(t)$ ,  $y(t)$ ,  $z(t)$  sont représentées graphiquement à la figure 4.1. Les fonctions  $-x'(t)$  et  $z'(t)$  sont quant à elles représentées graphiquement à la figure 4.2. On présente les fonctions jusqu'à  $t = 25$ , étant donné qu'elles sont pratiquement constantes pour des valeurs de  $t$  supérieures à 25.

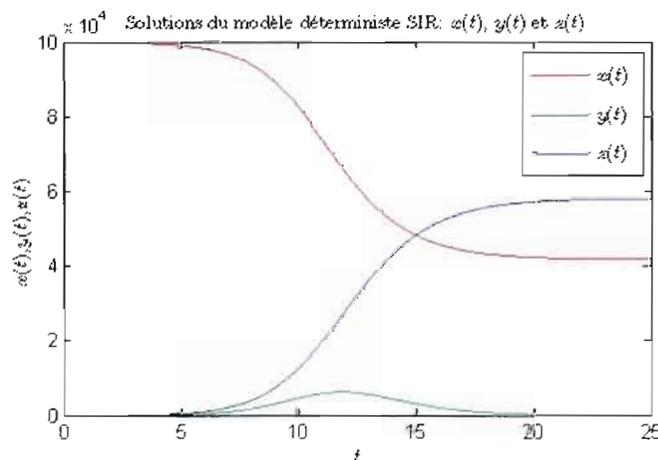


Figure 4.1 Solutions du modèle SIR :  $x(t)$ ,  $y(t)$  et  $z(t)$

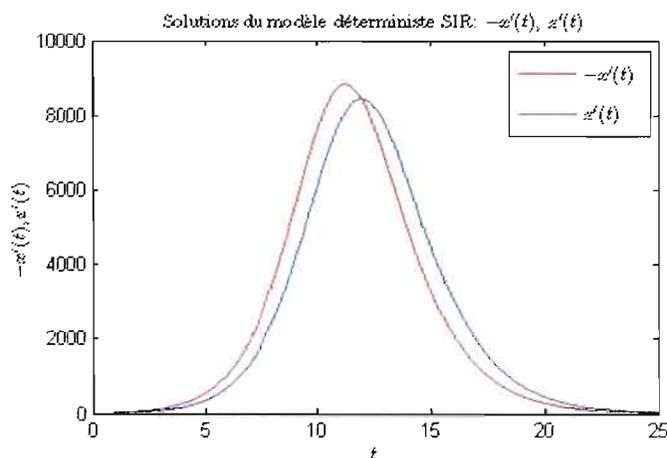


Figure 4.2 Solutions du modèle SIR :  $-x'(t)$  et  $z'(t)$

#### 4.4 Simulation des données à partir des modèles stochastiques proposés

Dans cette section, on présente la façon de simuler des données à partir des deux modèles stochastiques proposés à la section 2.3. Rappelons que dans ces modèles, on modélise le nombre d'individus susceptibles, retirés et décédés, notés respectivement par  $X(t)$ ,  $Z(t)$  et  $D(t)$ , où  $t \geq 0$ . De plus, ces modèles font intervenir certaines des solutions de (2.1), le système d'équations différentielles permettant d'obtenir le nombre d'individus susceptibles, infectés et retirés dans le modèle déterministe SIR.

##### 4.4.1 Modèle avec la loi binomiale

À la section 3.1, on a supposé qu'on a des observations du nombre d'individus susceptibles et décédés aux instants  $\tau = 0, 1, \dots, n$ . On a également supposé que le nombre initial d'individus susceptibles est égal à  $x_0$ , tandis que le nombre initial d'individus décédés est égal à 0. Par conséquent, on doit générer des valeurs du nombre d'individus susceptibles et décédés aux instants  $\tau = 1, 2, \dots, n$ .

Rappelons que pour le modèle avec la loi binomiale,  $\{X(t), t \geq 0\}$  est modélisé par

un processus de mort pur avec  $\mu_X(t) = \beta y(t)X(t)$ . À la proposition 2.3.2, on a montré que la loi conditionnelle de  $X(t)$  sachant  $X(s)$  est une loi binomiale

$$\mathcal{B}\left(X(s), \frac{x(t)}{x(s)}\right).$$

On propose l'algorithme suivant pour générer des réalisations du nombre d'individus susceptibles aux instants  $\tau = 1, 2, \dots, n$  :

- (i) on pose  $\tau = 1$  ;
- (ii) on génère  $X(\tau)$ , qui, conditionnellement à  $X(\tau - 1)$ , est de loi binomiale

$$\mathcal{B}\left(X(\tau - 1), \frac{x(\tau)}{x(\tau - 1)}\right) ;$$

- (iii) on pose  $\tau = \tau + 1$  et on retourne à l'étape (ii) si  $\tau \leq n$ .

Pour ce qui est du nombre d'individus retirés et décédés, rappelons que pour le modèle avec la loi binomiale,  $\{Z(t), t \geq 0\}$  est modélisé par un processus de Poisson non homogène d'intensité  $\gamma y(t)$  et que

$$D(t) = \begin{cases} \sum_{j=1}^{Z(t)} V_j, & Z(t) > 0 \\ 0, & \text{sinon,} \end{cases}$$

où les  $V_j$  sont des variables aléatoires indépendantes de loi binomiale  $\mathcal{B}(1, p)$  et  $t \geq 0$ . On a donc que pour  $0 \leq s < t \leq u < v$ ,  $Z(t) - Z(s)$  et  $Z(v) - Z(u)$  sont indépendants et  $Z(t) - Z(s)$  est de loi de Poisson

$$\mathcal{P}\left[z(t) - z(s)\right].$$

Soit  $L_Z(\tau)$  le nombre de nouveaux individus retirés entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$L_Z(\tau) = Z(\tau) - Z(\tau - 1),$$

$\tau = 1, 2, \dots, n$ . On a que les  $L_Z(\tau)$  sont indépendants. Soit  $L_D(\tau)$  le nombre de nouveaux individus décédés entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$L_D(\tau) = D(\tau) - D(\tau - 1),$$

$\tau = 1, 2, \dots, n$ . Étant donné que

$$\begin{aligned} L_D(\tau) &= D(\tau) - D(\tau - 1) \\ &= \sum_{j=1}^{Z(\tau)} V_j - \sum_{j=1}^{Z(\tau-1)} V_j \\ &= \sum_{j=Z(\tau-1)+1}^{Z(\tau)} V_j, \end{aligned}$$

$L_D(\tau)$  est la somme de  $L_Z(\tau)$  variables aléatoires indépendantes de loi binomiale  $B(1, p)$ .

On a donc que la loi conditionnelle de  $L_D(\tau)$  sachant  $L_Z(\tau)$  est une loi binomiale

$$\mathcal{B}\left[L_Z(\tau), p\right].$$

On propose l'algorithme suivant pour générer des réalisations du nombre d'individus retirés et décédés aux instants  $\tau = 1, 2, \dots, n$  :

- (i) on pose  $\tau = 1$  ;
- (ii) on génère  $L_Z(\tau)$ , qui est de loi de Poisson

$$\mathcal{P}\left[z(\tau) - z(\tau - 1)\right]$$

et qui est indépendant de  $Z(\tau - 1)$  ;

- (iii) on pose  $Z(\tau) = Z(\tau - 1) + L_Z(\tau)$  ;
- (iv) on génère  $L_D(\tau)$ , qui, conditionnellement à  $L_Z(\tau)$ , est de loi binomiale

$$\mathcal{B}\left[L_Z(\tau), p\right]$$

et est indépendant de  $D(\tau - 1)$  ;

- (v) on pose  $D(\tau) = D(\tau - 1) + L_D(\tau)$  ;
- (vi) on pose  $\tau = \tau + 1$  et on retourne à l'étape (ii) si  $\tau \leq n$ .

Mentionnons que pour chaque valeur de  $\tau$  entre 1 et  $n$ , les valeurs de  $X(\tau)$  et  $Z(\tau)$  sont générées de façon indépendante, alors qu'il existe une dépendance entre ces deux variables dans le modèle avec la loi binomiale. Cette dépendance vient du fait qu'une infection et un retrait ne puissent pas se produire simultanément. Étant donné qu'on génère le nombre de nouveaux individus infectés et retirés sur un intervalle de temps, on s'est permis de les générer de façon indépendante.

#### 4.4.2 Modèle avec la loi de Poisson

Soit  $N_X(t)$  le nombre total d'individus infectés au temps  $t$ . Rappelons que pour le modèle avec la loi de Poisson,  $\{N_X(t), t \geq 0\}$  est modélisé par un processus de Poisson non homogène d'intensité  $\beta x(t)y(t)$  et qu'on a posé  $X(t) = x_0 - N_X(t)$ . À la proposition 2.3.4, on a montré que pour  $0 \leq s < t \leq u < v$ ,  $X(s) - X(t)$  et  $X(u) - X(v)$  sont indépendants et  $X(s) - X(t)$  est de loi de Poisson

$$\mathcal{P}[x(s) - x(t)].$$

Soit  $L_X(\tau)$  le nombre de nouveaux individus infectés entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$L_X(\tau) = X(\tau - 1) - X(\tau),$$

$\tau = 1, 2, \dots, n$ . On a que les  $L_X(\tau)$  sont indépendants. On propose donc l'algorithme suivant pour générer des réalisations du nombre d'individus susceptibles aux instants  $\tau = 1, 2, \dots, n$  :

- (i) on pose  $\tau = 1$ ;
- (ii) on génère  $L_X(\tau)$ , qui est de loi de Poisson

$$\mathcal{P}[x(\tau - 1) - x(\tau)]$$

et qui est indépendant de  $X(\tau - 1)$ ;

- (iii) on pose  $X(\tau) = X(\tau - 1) - L_X(\tau)$ ;
- (iv) on pose  $\tau = \tau + 1$  et on retourne à l'étape (ii) si  $\tau \leq n$ .

Pour ce qui est du nombre d'individus retirés et décédés, on en génère des réalisations à l'aide de l'algorithme présenté à la section 4.4.1 pour le nombre d'individus retirés et décédés, étant donné qu'ils sont modélisés de la même façon dans les deux modèles. Pour chaque valeur de  $\tau$  entre 1 et  $n$ , les valeurs de  $X(\tau)$  et  $Z(\tau)$  ont été générées de façon indépendante, étant donné qu'on a supposé l'indépendance entre ces deux variables dans le modèle avec la loi de Poisson.

### 4.4.3 Résultats obtenus en simulation

Un des objectifs des modèles proposés est que les espérances de  $X(t)$  et  $Z(t)$  soient respectivement égales à  $x(t)$  et à  $z(t)$ , les solutions du modèle déterministe. Pour les deux modèles, on a simulé 5000 fois les valeurs de  $X(\tau)$  et  $Z(\tau)$  pour  $\tau = 1, 2, \dots, n$  à l'aide des algorithmes présentés aux sections 4.4.1 et 4.4.2. Dans les deux cas, on obtient que la moyenne des valeurs générées est pratiquement égale aux solutions du système déterministe SIR obtenues numériquement, tel que désiré.

Soit  $Y(t)$  le nombre d'individus infectés au temps  $t$ . À la section 2.3, on a mentionné que si on posait  $Y(t) = N - X(t) - Z(t)$ , il se pourrait que  $Y(t) < 0$  pour les deux modèles. De plus, il se pourrait que  $X(t) < 0$  pour le modèle avec la loi de Poisson. En simulation, pour les deux modèles, ces problèmes surviennent à partir de  $t = 20$ . Cela s'explique par le fait que pour des valeurs de  $t$  supérieures ou égales à 20,  $y(t) \approx 0$  dans le système déterministe. Cependant, pour l'estimation de  $R_0$ , on s'intéresse à des valeurs de  $\tau$  autour de  $t_z = 12$ . Les cas où  $X(t) < 0$  ou  $Y(t) < 0$  n'influencent donc pas les résultats obtenus en simulation pour l'estimation de  $R_0$ .

### 4.5 Approximation de la fonction de masse de $T_z$

L'étape suivante consiste à calculer la fonction de masse de  $T_z$ , l'estimateur de  $t_z$  obtenu à partir des données d'une seule épidémie. À la section 3.3, on a défini  $\ell_D(\tau)$  comme le nombre d'individus décédés dans le modèle déterministe entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$\ell_D(\tau) = p \left[ z(\tau) - z(\tau - 1) \right]$$

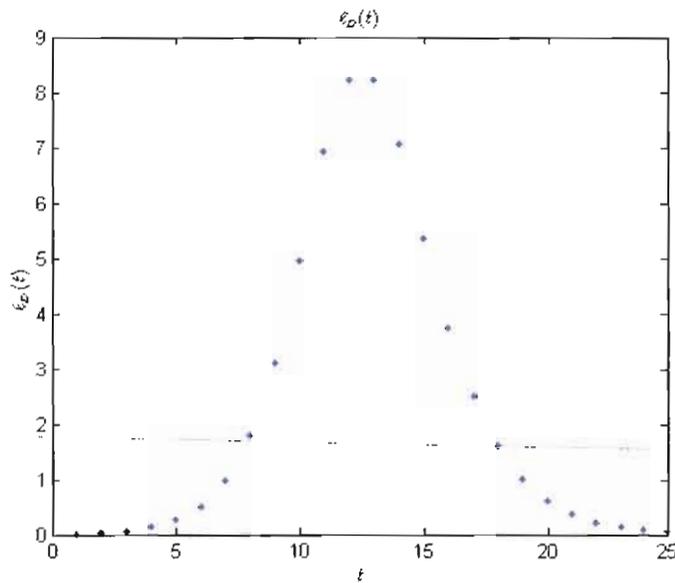
et  $L_D(\tau)$  comme le nombre d'individus décédés dans le modèle stochastique entre les instants  $\tau - 1$  et  $\tau$ , c'est-à-dire

$$L_D(\tau) = D(\tau) - D(\tau - 1),$$

$\tau = 1, 2, \dots, n$ . À la section 2.3.1, on a montré que  $L_D(\tau)$  est de loi de Poisson

$$\mathcal{P}(\ell_D(\tau)),$$

$\tau = 1, 2, \dots, n$ . Les valeurs de  $\ell_D(\tau)$  sont représentées graphiquement à la figure 4.3. On présente les valeurs de  $\ell_D(\tau)$  jusqu'à  $\tau = 25$ , étant donné que  $\ell_D(\tau) \approx 0$  pour des valeurs de  $\tau$  supérieures à 25. À la section 3.3.2, on a mentionné que  $T_Z$  est obtenu en calculant



**Figure 4.3** Graphique de  $\ell_D(\tau)$

la moyenne des valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximal,  $\tau = 1, 2, \dots, n$ . Les valeurs de  $\ell_D(\tau)$  sont donc les paramètres des  $n$  variables aléatoires de loi de Poisson à partir desquelles on calcule  $T_Z$ . Si la valeur obtenue n'est pas entière, alors on l'arrondit afin que  $T_Z$  soit un nombre entier entre 1 et  $n$  inclusivement. Pour ce faire, on arrondit en prenant la partie entière, car c'est la méthode qui fait en sorte que le biais empirique de  $T_Z$  est le plus petit. À la section 3.5.1, on a présenté le calcul de la fonction de masse de  $T_Z$ , qui peut se calculer comme suit :

$$P[T_Z = k] = \sum_{m=1}^n P[T_Z = k, M = m], \quad (4.1)$$

où  $M$  est le nombre de valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximal. Il n'est cependant pas possible d'obtenir la fonction de masse exacte de  $T_Z$  à partir de l'équation (4.1) et des valeurs de  $\ell_D(\tau)$  obtenues numériquement. Premièrement, étant donné que

$$P[L_D(1) > L_D(2)] = \sum_{i=1}^{\infty} P[L_D(1) = i]P[L_D(2) < i]$$

et qu'on doit arrêter la somme à une valeur finie en pratique, on ne peut qu'obtenir une approximation de la probabilité que  $L_D(1)$  soit supérieur à  $L_D(2)$ . On note cette valeur par  $a$ . De plus, étant donné que la valeur de  $n$  est grande et que  $M$  peut prendre des valeurs entières entre 1 et  $n$ , le temps de calcul de la fonction de masse de  $T_Z$  à partir de l'équation (4.1) est extrêmement long. On fixe donc une valeur maximale pour  $M$ , le nombre de valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximal, que l'on note par  $b$ . On choisit une valeur de  $b$  qui fait en sorte que  $P[M > b]$  soit très petite. Le choix d'une valeur de  $b$  inférieure à  $n$  fait également en sorte que la fonction de masse est approximée. Comme on approxime la fonction de masse de  $T_Z$ , la somme des probabilités n'est pas égale à 1. On propose de normaliser la fonction de masse en divisant chacune des valeurs par la somme des probabilités de la fonction de masse approximée. De cette façon, la somme des probabilités de la fonction de masse normalisée est égale à 1.

Soit

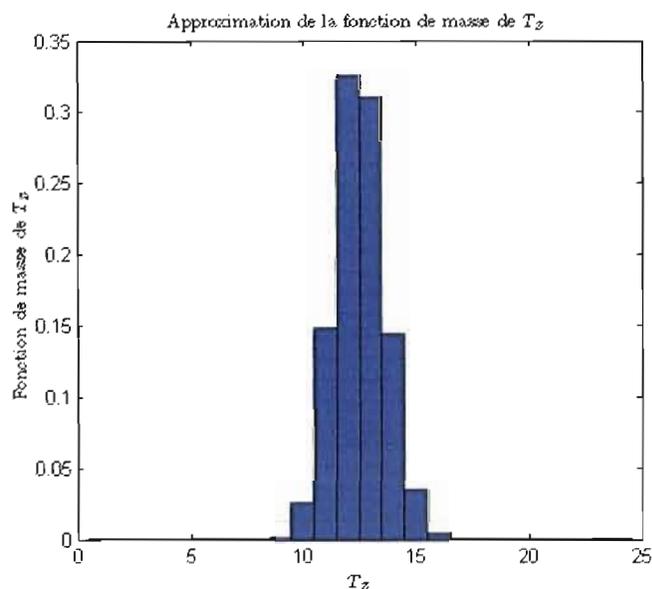
$$c = \max\left(\{\ell_D(\tau), \tau = 1, 2, \dots, n\}\right).$$

Pour le calcul de la fonction de masse de  $T_Z$ , on a utilisé

$$a = c + 3\sqrt{c},$$

car la probabilité qu'une variable aléatoire de loi normale  $\mathcal{N}(c, c)$  prenne une valeur supérieure à  $c + 3\sqrt{c}$  est très petite et qu'une variable aléatoire de loi de Poisson  $\mathcal{P}(c)$  est approximativement de loi normale  $\mathcal{N}(c, c)$  lorsque  $c$  est relativement grand. De même, la probabilité qu'une variable aléatoire de loi de Poisson  $\mathcal{P}(\lambda)$  pour laquelle  $\lambda < c$  prenne une valeur supérieure à  $c + 3\sqrt{c}$  est très petite. Pour ce qui est de la valeur maximale du nombre de valeurs de  $\tau$  pour lesquelles  $L_D(\tau)$  est maximal, on a considéré le cas où

$b = 5$ . Après normalisation, on obtient la fonction de masse approximée de  $T_Z$  qui est présentée à la figure 4.4. On présente les valeurs de la fonction de masse jusqu'à  $\tau = 24$ , étant donné qu'elles sont pratiquement égales à 0 pour des valeurs de  $\tau$  supérieures à 24. Il est à noter que la somme des probabilités avant normalisation est égale à 0,995.



**Figure 4.4** Approximation de la fonction de masse de  $T_Z$

Comme cette valeur est presque égale à 1, on en conclut que les événements qui n'ont pas été considérés dans le calcul de la fonction de masse de  $T_Z$  ont une très faible probabilité de se produire, ce qui est conforme à qu'on avait prévu. À partir de l'approximation de la fonction de masse de  $T_Z$ , on peut obtenir une approximation de l'espérance et de la variance de  $T_Z$ . Les valeurs obtenues sont :

$$\mathbb{E}[T_Z] = 12,521 ;$$

$$\text{Var}[T_Z] = 1,305 .$$

On peut ensuite calculer le biais, le biais relatif et l'erreur quadratique moyenne de  $T_z$ , notés respectivement par  $B[T_z]$ ,  $BR[T_z]$  et  $EQM[T_z]$ , de la façon suivante :

$$\begin{aligned} B[T_z] &= \mathbb{E}[T_z] - t_z, \\ BR[T_z] &= \frac{B[T_z]}{t_z}, \\ EQM[T_z] &= \text{Var}[T_z] + \left(B[T_z]\right)^2, \end{aligned}$$

Les valeurs obtenues sont :

$$\begin{aligned} B[T_z] &= 0,521; \\ BR[T_z] &= 0,043; \\ EQM[T_z] &= 1,576. \end{aligned}$$

Étant donné que  $t_z = 12$  et que  $\mathbb{E}[T_z] = 12,521$ , on a que

$$\mathbb{E}[T_z] - 1 < t_z < \mathbb{E}[T_z] \tag{4.2}$$

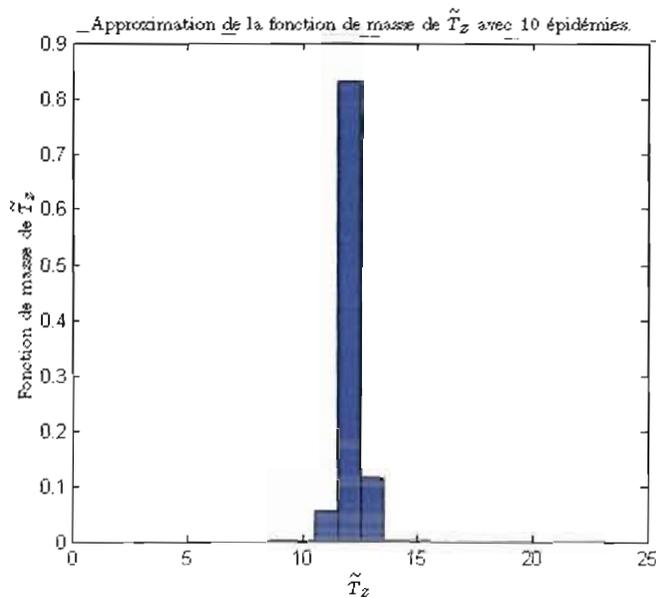
et que  $T_z$  est un estimateur biaisé pour  $T_z$ , ce à quoi on s'attendait intuitivement. À la section 3.3.2, on s'était basé sur l'équation (4.2) pour justifier la formule de  $\hat{\rho}(T_z)$ , l'estimateur de  $\rho$  obtenu à partir de  $T_z$  et des données d'une seule épidémie. Cette justification est donc correcte avec les valeurs des paramètres utilisées dans ce chapitre.

#### 4.6 Approximation de la fonction de masse de $\tilde{T}_z$

À partir de la fonction de masse de  $T_z$ , on peut calculer la fonction de masse de  $\tilde{T}_z$ , l'estimateur de  $t_z$  obtenu à partir des données de toutes les épidémies. Rappelons que  $\tilde{T}_z$  est calculé en prenant la partie entière de la moyenne de  $T_z^{(1)}, T_z^{(2)}, \dots, T_z^{(k)}$ , où  $T_z^{(i)}$  est l'estimateur de  $t_z$  obtenu à partir des données de la  $i^{\text{ème}}$  épidémie.  $\tilde{T}_z$  peut donc prendre des valeurs entières entre 1 et  $n$  inclusivement. Afin de calculer la probabilité que  $\tilde{T}_z$  soit égal à  $k$ , on fait la somme des probabilités de tous les événements qui font en sorte que  $\tilde{T}_z$  soit égal à  $k$ . Comme dans le cas du calcul de la fonction de masse de  $T_z$ , on ne peut obtenir qu'une approximation de la fonction de masse de  $\tilde{T}_z$ , car elle est

obtenue à partir de l'approximation de la fonction de masse de  $T_Z$ . De plus, le fait que  $n$  soit grand implique que le temps de calcul de la fonction de masse est énorme si on considère tous les événements possibles. Pour réduire le temps de calcul, on ne considère que les valeurs de  $T_Z$  dont la probabilité est supérieure à 0,001 pour calculer la fonction de masse de  $\tilde{T}_Z$ , étant donné que la probabilité des événements faisant intervenir les autres valeurs possibles de  $T_Z$  est très faible. Comme à la section 4.5, on propose de normaliser la fonction de masse en divisant chacune des valeurs par la somme des probabilités de la fonction de masse approximée. De cette façon, la somme des probabilités de la fonction de masse normalisée est égale à 1.

Soit  $k$  le nombre d'épidémies considérées. Pour  $k = 10$ , on obtient la fonction de masse approximée de  $\tilde{T}_Z$  qui est présentée à la figure 4.5. On présente les valeurs de la fonction de masse jusqu'à  $\tau = 24$ , étant donné qu'elles sont pratiquement égales à 0 pour des valeurs de  $\tau$  supérieures à 24.



**Figure 4.5** Approximation de la fonction de masse de  $\tilde{T}_Z$  avec  $k = 10$

Il est à noter que la somme des probabilités avant normalisation est égale à 0,994.

Comme cette valeur est presque égale à 1, on en conclut que les événements qui n'ont pas été considérés dans le calcul de la fonction de masse de  $\tilde{T}_z$  ont une très faible probabilité de se produire, ce qui est conforme à qu'on avait prévu.

À partir de l'approximation de la fonction de masse de  $\tilde{T}_z$ , on peut obtenir une approximation de l'espérance, de la variance, du biais, du biais relatif et de l'erreur quadratique moyenne de  $\tilde{T}_z$ , de la même façon que pour  $T_z$ . Les valeurs obtenues sont :

$$E[\tilde{T}_z] = 12,060 ;$$

$$\text{Var}[\tilde{T}_z] = 0,167 ;$$

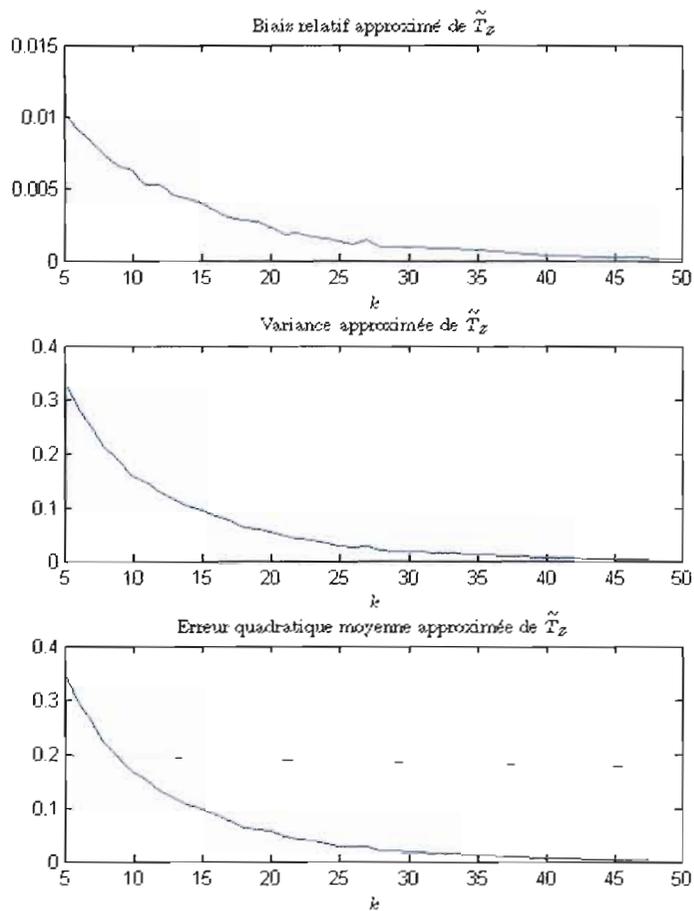
$$B[\tilde{T}_z] = 0,060 ;$$

$$\text{BR}[\tilde{T}_z] = 0,005 ;$$

$$\text{EQM}[\tilde{T}_z] = 0,171 .$$

Tel qu'attendu, le biais et la variance de  $\tilde{T}_z$  sont inférieurs à ceux de  $T_z$ . Au-delà de 10 épidémies, l'approximation de la fonction de masse est très longue à calculer de cette façon. On peut cependant obtenir la distribution de  $\tilde{T}_z$  en simulation, qui constitue une autre approximation de la fonction de masse de  $\tilde{T}_z$ . Pour ce faire, on génère 5000 fois  $k$  réalisations de  $L_D(1), L_D(2), \dots, L_D(n)$ . Pour chacune des 5000 simulations, on calcule  $\tilde{T}_z$  à partir des valeurs générées. Par la suite, on calcule la proportion des simulations qui ont mené à  $\tilde{T}_z = \tau$ ,  $\tau = 1, 2, \dots, n$ . Ces proportions constituent l'approximation de la fonction de masse de  $\tilde{T}_z$ .

Cette deuxième procédure d'approximation de la fonction de masse de  $\tilde{T}_z$  a été appliquée pour différentes valeurs de  $k$ . On a représenté graphiquement le biais relatif, la variance et l'erreur quadratique moyenne de  $\tilde{T}_z$  en fonction de  $k$  à la figure 4.6. On remarque que le biais relatif, la variance et l'erreur quadratique moyenne de  $\tilde{T}_z$  se rapprochent de 0 lorsque  $k$  augmente. Ce comportement est conforme à ce qui était attendu, car on a montré à la section 3.5.3 que  $\tilde{T}_z$  est un estimateur fortement convergent pour  $t_z$ .



**Figure 4.6** Biais relatif, variance et erreur quadratique moyenne approximés de  $\tilde{T}_Z$  en fonction de  $k$

#### 4.7 Propriétés des estimateurs de $\rho$

On présente maintenant les résultats obtenus numériquement pour les propriétés de  $\bar{\rho}(t_Z)$ ,  $\bar{\rho}(\mathbf{T}_Z)$  et  $\bar{\rho}(\tilde{T}_Z)$ , les trois estimateurs de  $\rho$  proposés au chapitre 3. Mentionnons tout d'abord que la plus petite valeur de  $k$ , le nombre d'épidémies, qui a été considérée est 2, étant donné que si  $k = 1$ , on ne peut pas estimer la variance des estimateurs. Les expressions permettant de calculer l'espérance et la variance de  $\bar{\rho}(t_Z)$ ,  $\bar{\rho}(\mathbf{T}_Z)$  et

$\bar{\rho}(\tilde{T}_Z)$  sont données à la proposition 3.3.1, à la proposition 3.3.3 et à la proposition 3.3.5 respectivement. Rappelons qu'à ces propositions, on n'a qu'une seule expression pour l'espérance de l'estimateur, alors qu'on a deux expressions pour sa variance, soit une pour chaque modèle.

Cependant, les résultats obtenus numériquement et en simulation sont pratiquement les mêmes pour les deux modèles. Pour cette raison, on présente et on commente uniquement les résultats obtenus à partir du modèle avec la loi binomiale dans cette section, ainsi qu'aux sections 4.8 et 4.9. Les résultats obtenus à partir du modèle avec la loi de Poisson sont présentés à l'appendice A.

Tout d'abord, on a représenté graphiquement le biais relatif des trois estimateurs de  $\rho$  en fonction de  $k$ , le nombre d'épidémies, à la figure 4.7. Au chapitre 3, on a montré que le biais relatif de  $\bar{\rho}(t_Z)$  est égal à 0. Pour ce qui est de  $\bar{\rho}(\mathbf{T}_Z)$ , on s'attendait à ce qu'il soit biaisé et que son biais ne change pas avec  $k$ . On remarque en effet que le biais relatif de  $\bar{\rho}(\mathbf{T}_Z)$  est égal à 0,78%, ce qui est toutefois très faible. Enfin, on a montré au chapitre 3 que  $\bar{\rho}(\tilde{T}_Z)$  est un estimateur convergent pour  $\rho$ . Graphiquement, on voit bien que le biais relatif de  $\bar{\rho}(\tilde{T}_Z)$  se rapproche de 0 lorsque  $k$  augmente. Cependant, il est plutôt élevé en valeur absolue pour de petites valeurs de  $k$ .

Afin de pouvoir comparer plus facilement le biais relatif de  $\bar{\rho}(\mathbf{T}_Z)$  et  $\bar{\rho}(\tilde{T}_Z)$  pour différentes valeurs de  $k$ , on a représenté graphiquement la valeur absolue de ces biais relatifs à la figure 4.8. On remarque que le biais relatif de  $\bar{\rho}(\mathbf{T}_Z)$  est inférieur en valeur absolue à celui de  $\bar{\rho}(\tilde{T}_Z)$  pour des valeurs de  $k \leq 10$ , alors qu'il lui est supérieur pour des valeurs de  $k \geq 11$ .

On a représenté graphiquement la variance des trois estimateurs de  $\rho$  en fonction de  $k$  pour le modèle avec la loi binomiale à la figure 4.9. Il faut noter que le paramètre  $\rho$  est égal à 66 660, d'où les grandes valeurs de la variance des estimateurs.

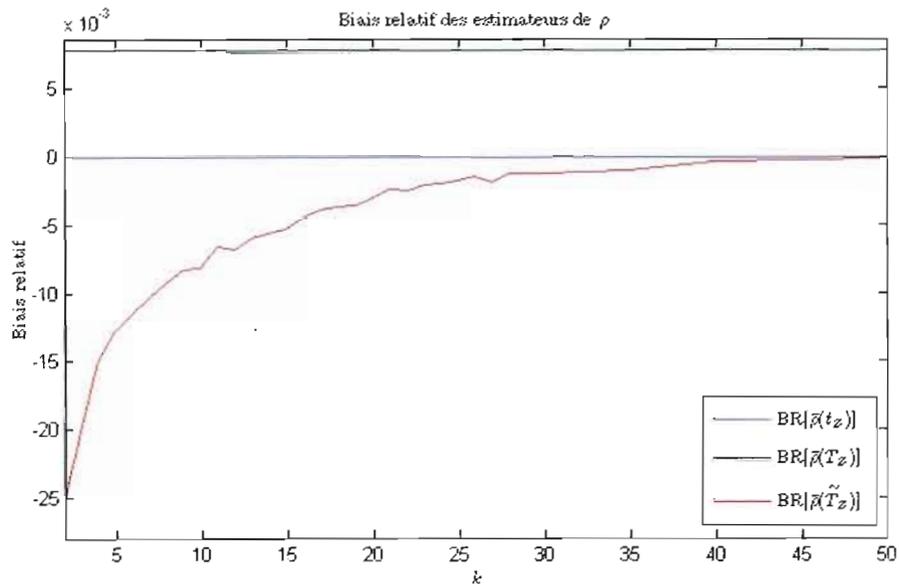


Figure 4.7 Biais relatif de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{\mathbf{T}}_z)$  en fonction de  $k$

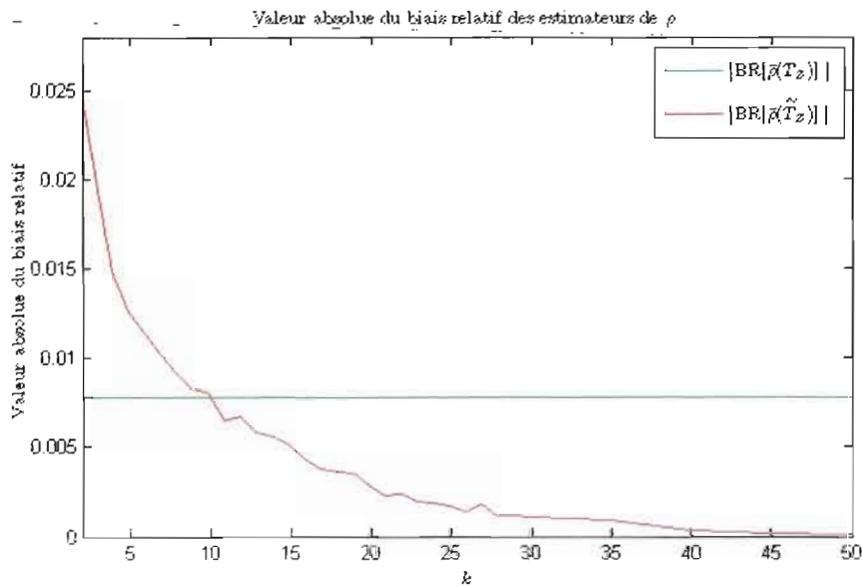
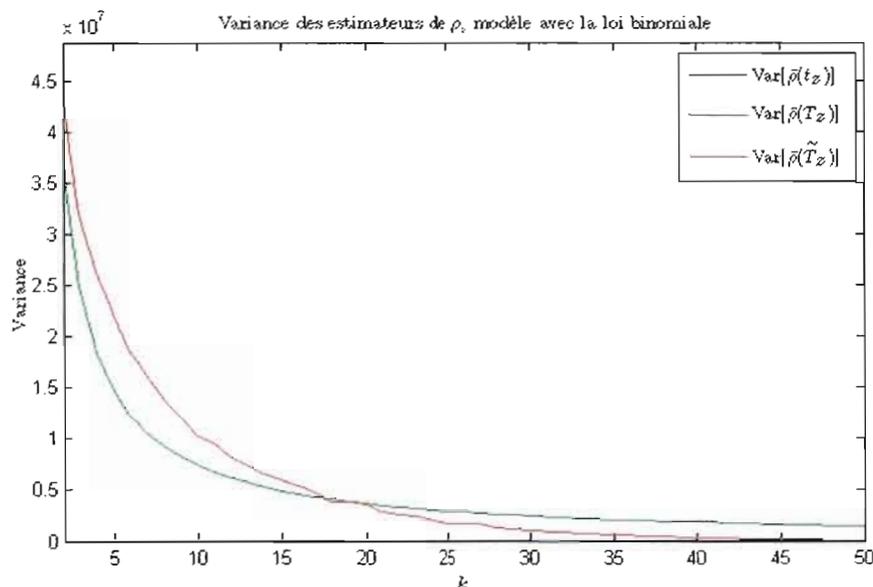


Figure 4.8 Valeur absolue du biais relatif de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{\mathbf{T}}_z)$  en fonction de  $k$



**Figure 4.9** Variance de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi binomiale

On remarque que la variance de  $\bar{\rho}(t_z)$  est beaucoup plus petite que celle des deux autres estimateurs de  $\rho$ . Le fait d'estimer  $t_z$ , que ce soit par  $T_z$  ou  $\tilde{T}_z$ , entraîne une augmentation considérable de la variance de l'estimateur de  $\rho$  obtenu à partir de cet estimateur de  $t_z$ . Au chapitre 3, on a calculé la variance de  $\hat{\rho}_1(T_z^{(1)})$  à partir de l'équation suivante :

$$\text{Var}[\hat{\rho}_1(T_z^{(1)})] = \mathbb{E}(\text{Var}[\hat{\rho}_1(T_z^{(1)})|T_z^{(1)}]) + \text{Var}(\mathbb{E}[\hat{\rho}_1(T_z^{(1)})|T_z^{(1)}])$$

et la variance de  $\bar{\rho}(\mathbf{T}_z)$  à partir de l'équation suivante :

$$\text{Var}[\bar{\rho}(\mathbf{T}_z)] = \frac{1}{k} \text{Var}[\hat{\rho}_1(T_z^{(1)})].$$

Environ 99,9% de la variabilité de  $\bar{\rho}(\mathbf{T}_z)$  est due à la variabilité des  $T_z^{(i)}$ , c'est-à-dire au terme

$$\text{Var}(\mathbb{E}[\hat{\rho}_1(T_z^{(1)})|T_z^{(1)}]).$$

Cette proportion est la même peu importe la valeur de  $k$ . Pour ce qui est de  $\bar{\rho}(\tilde{T}_z)$ , on

a calculé sa variance à partir de l'équation suivante :

$$\text{Var}[\bar{\rho}(\tilde{T}_z)] = \mathbb{E}(\text{Var}[\bar{\rho}(\tilde{T}_z)]|\tilde{T}_z) + \text{Var}(\mathbb{E}[\bar{\rho}(\tilde{T}_z)]|\tilde{T}_z) + \frac{k-1}{k} \text{Cov}[\hat{\rho}_1(\tilde{T}_z), \hat{\rho}_2(\tilde{T}_z)].$$

Pour  $k = 10$ , environ 10% de la variabilité de  $\bar{\rho}(\tilde{T}_z)$  est due à la variabilité de  $\tilde{T}_z$ , c'est-à-dire au terme

$$\text{Var}(\mathbb{E}[\bar{\rho}(\tilde{T}_z)]|\tilde{T}_z).$$

La plus grande partie de sa variabilité, soit environ 89,9%, est due à la dépendance entre les  $\hat{\rho}_i(\tilde{T}_z)$ , c'est-à-dire au terme

$$\frac{k-1}{k} \text{Cov}[\hat{\rho}_1(\tilde{T}_z), \hat{\rho}_2(\tilde{T}_z)].$$

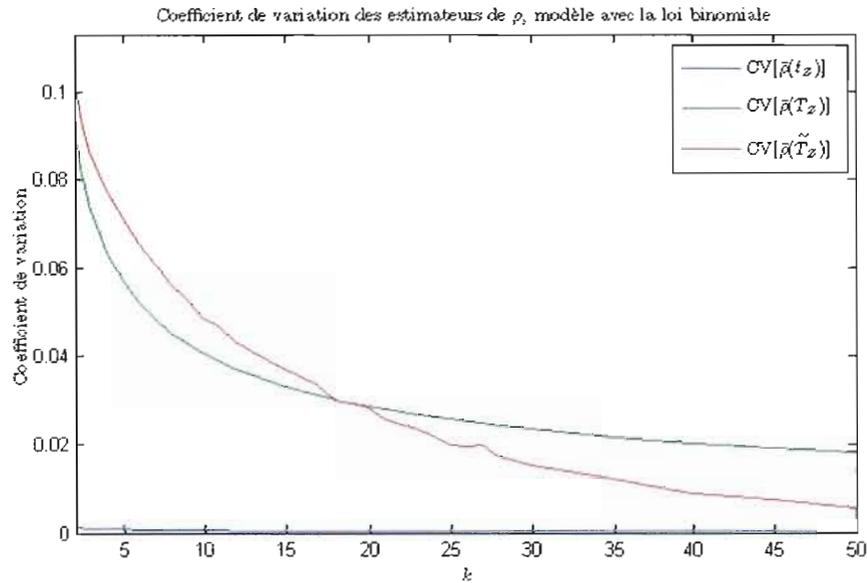
Le reste de la variabilité de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$ , qui est inférieure à 0,1% de leur variabilité totale, est due à la variabilité du nombre d'individus susceptibles aux instants  $\tau = 1, 2, \dots, n$ . Cette source de variabilité est d'ailleurs la seule qui contribue à la variance de  $\bar{\rho}(t_z)$ , d'où la petite variance de  $\bar{\rho}(t_z)$ .

Malgré le fait que leur variance soit beaucoup plus élevée que celle de  $\bar{\rho}(t_z)$ , les estimateurs  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  sont plus réalistes, car en pratique, on ne connaît pas  $t_z$  et on doit l'estimer. On peut comparer la variance de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  en fonction de  $k$ . On remarque que la variance de  $\bar{\rho}(\mathbf{T}_z)$  est inférieure à la variance de  $\bar{\rho}(\tilde{T}_z)$  pour des valeurs de  $k \leq 17$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 18$ .

Étant donné que la valeur de  $\rho$  est égale à 66 660 et que, par conséquent, la variance de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  est très grande, il peut être intéressant de calculer le coefficient de variation de ces estimateurs. En effet, le coefficient de variation mesure la dispersion relative des estimateurs par rapport à leur moyenne. Rappelons que la moyenne de  $\bar{\rho}(t_z)$  est égale à  $\rho$ , alors que la moyenne de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  est presque égale à  $\rho$ . De façon générale, le coefficient de variation d'une variable aléatoire  $X$  de moyenne  $\mu$  et de variance  $\sigma^2$  est donné par :

$$\text{CV}(X) = \frac{\sigma}{\mu}.$$

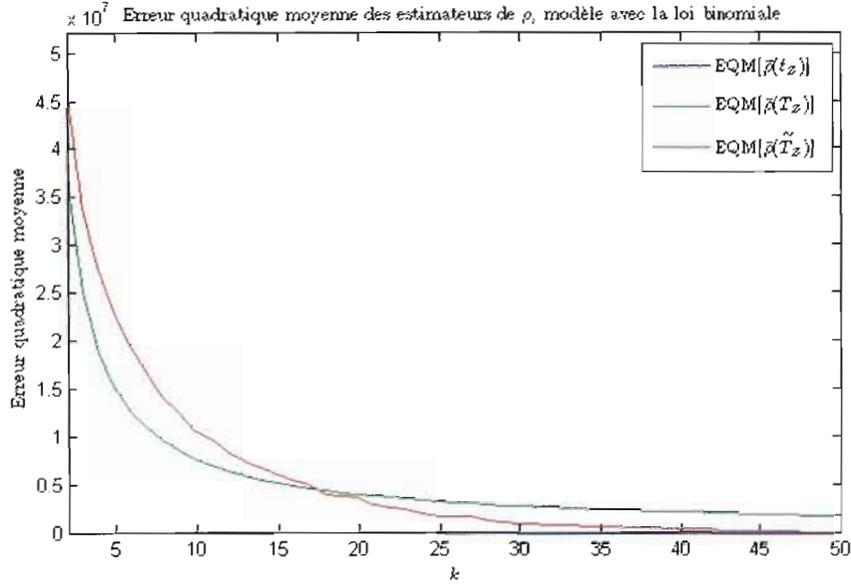
On a représenté graphiquement le coefficient de variation des trois estimateurs de  $\rho$  en fonction de  $k$  pour le modèle avec la loi binomiale à la figure 4.10.



**Figure 4.10** Coefficient de variation de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi binomiale

On remarque que le coefficient de variation de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  n'est pas trop élevé. En effet, il est inférieur à 10%, même pour de petites valeurs de  $k$ .

Finalement, on a représenté graphiquement l'erreur quadratique moyenne des trois estimateurs de  $\rho$  en fonction de  $k$  pour le modèle avec la loi binomiale à la figure 4.11. Les observations sur l'erreur quadratique moyenne des estimateurs de  $\rho$  sont pratiquement les mêmes que celles sur leur variance, étant donné qu'une grande partie de l'erreur quadratique moyenne des estimateurs est due à leur variance. On remarque en effet que, comme dans le cas de la variance, l'erreur quadratique moyenne de  $\bar{\rho}(\mathbf{T}_z)$  est inférieure à l'erreur quadratique moyenne de  $\bar{\rho}(\tilde{T}_z)$  pour des valeurs de  $k \leq 17$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 18$ .



**Figure 4.11** Erreur quadratique moyenne de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi binomiale

#### 4.8 Propriétés des estimateurs de $R_0$

On présente maintenant les résultats obtenus pour les propriétés des trois estimateurs de  $R_0$  obtenus à partir des trois estimateurs de  $\rho$ , soit  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$ . Étant donné qu'on n'a pas d'expression permettant de calculer l'espérance et la variance de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$ , on présente les espérances et les variances obtenues en simulation. Pour obtenir ces valeurs, on génère d'abord des données du nombre d'individus susceptibles et décédés aux instants  $\tau = 1, 2, \dots, n$  pour  $k$  épidémies. Ces données sont générées avec chacun des deux modèles proposés en suivant la procédure décrite à la section 4.4. Par la suite, on calcule les estimateurs  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  à partir des données générées. On calcule ensuite les estimateurs  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  à partir de la formule suivante :

$$\tilde{R}_0 = \frac{x_0}{\bar{\rho}}.$$

On fait 5000 simulations et on calcule la valeur moyenne et la variance des 5000 réalisations de chacun des estimateurs de  $R_0$ . À partir de ces valeurs, on peut calculer le biais relatif et l'erreur quadratique moyenne des estimateurs de  $R_0$ . Cette procédure est répétée pour différentes valeurs de  $k$ , le nombre d'épidémies. La plus petite valeur de  $k$  qui a été considérée est 2, étant donné que si  $k = 1$ , on ne peut pas estimer la variance des estimateurs.

Comme à la sections 4.7, on ne présente que les résultats pour le modèle avec la loi binomiale dans cette section. Les résultats pour le modèle avec la loi Poisson sont présentés à l'appendice A.

On a représenté graphiquement le biais relatif des trois estimateurs de  $R_0$  en fonction de  $k$  à la figure 4.12. On remarque d'abord que le biais relatif de  $\tilde{R}_0(t_z)$  obtenu en simulation est approximativement égal à 0 et ce, peu importe la valeur de  $k$ . Ce résultat est similaire à celui obtenu pour  $\bar{\rho}(t_z)$ , qui est un estimateur sans biais de  $\rho$ . Pour ce qui est de  $\tilde{R}_0(\mathbf{T}_z)$ , on remarque d'abord que son biais relatif obtenu en simulation est pratiquement nul pour 2 et 3 épidémies. Pour des valeurs de  $k > 3$ , le biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  varie entre  $-0,3\%$  et  $-0,7\%$  environ, selon la valeur de  $k$ . Rappelons qu'il est calculé à partir de  $\bar{\rho}(\mathbf{T}_z)$ , qui est un estimateur biaisé pour  $\rho$ . Enfin, on remarque que le biais relatif de  $\tilde{R}_0(\tilde{T}_z)$  se rapproche de 0 lorsque  $k$  augmente. Ce résultat est similaire à celui obtenu pour  $\bar{\rho}(\tilde{T}_z)$ , qui est un estimateur fortement convergent pour  $\rho$ . Cependant, il est plutôt élevé en valeur absolue pour de petites valeurs de  $k$ , comme c'est le cas pour le biais relatif de  $\bar{\rho}(\tilde{T}_z)$ .

Comme on l'a fait pour  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$ , on a représenté graphiquement la valeur absolue du biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction de  $k$  à la figure 4.13 afin de pouvoir les comparer plus facilement. On remarque que le biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieur en valeur absolue à celui de  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k \leq 13$ , alors qu'il lui est supérieur pour des valeurs de  $k \geq 14$ . Ce résultat est semblable à celui obtenu pour  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$ . Rappelons que la valeur de  $k$  à partir de laquelle le biais relatif de  $\bar{\rho}(\tilde{T}_z)$  devient

inférieur à celui de  $\bar{\rho}(\mathbf{T}_Z)$  est 11.

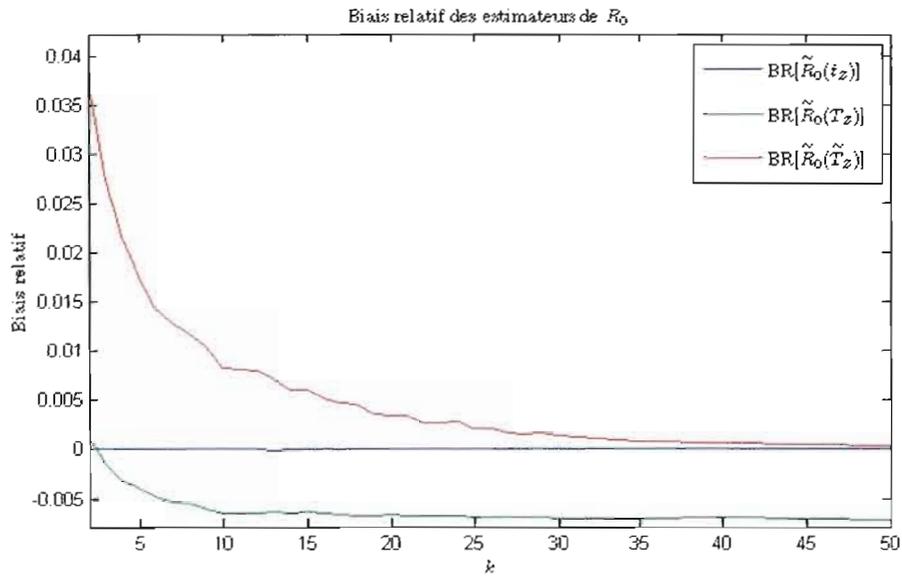


Figure 4.12 Biais relatif de  $\tilde{R}_0(t_Z)$ ,  $\tilde{R}_0(\mathbf{T}_Z)$  et  $\tilde{R}_0(\tilde{T}_Z)$  en fonction de  $k$

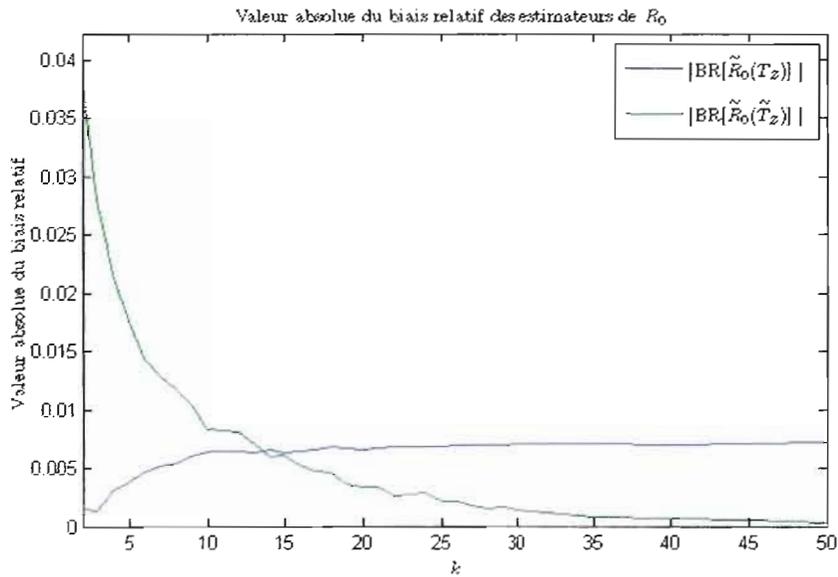
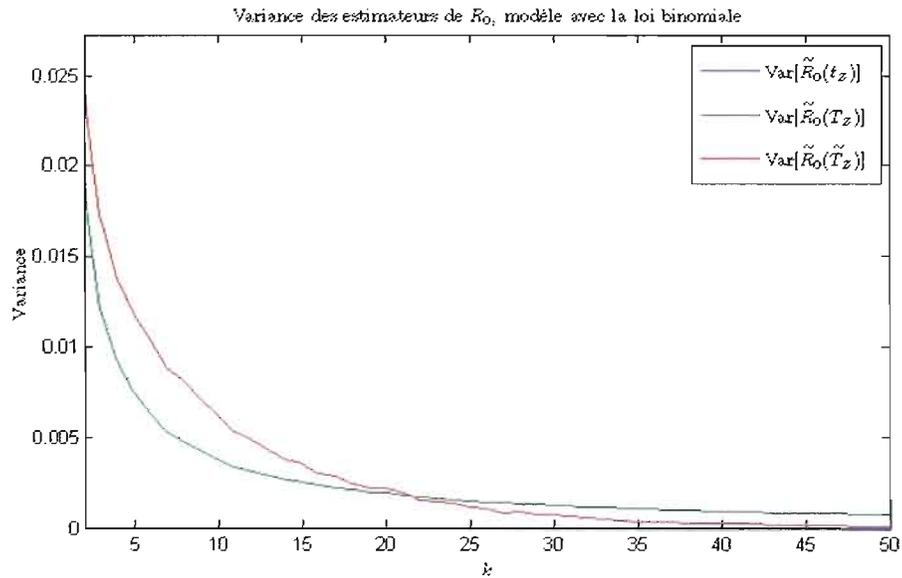


Figure 4.13 Valeur absolue du biais relatif de  $\tilde{R}_0(\mathbf{T}_Z)$  et  $\tilde{R}_0(\tilde{T}_Z)$  en fonction de  $k$

On a représenté graphiquement la variance des trois estimateurs de  $R_0$  en fonction de  $k$  pour le modèle avec la loi binomiale à la figure 4.14.

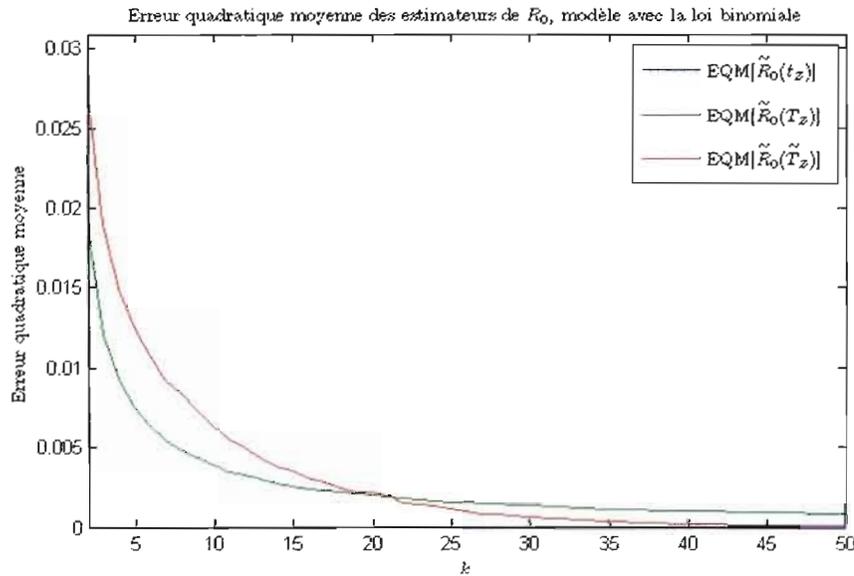


**Figure 4.14** Variance de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi binomiale

Tel qu'on l'avait observé pour les estimateurs de  $\rho$ , la variance de l'estimateur de  $R_0$  obtenu à partir de  $t_z$ ,  $\tilde{R}_0(t_z)$ , est beaucoup plus petite que celle de  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$ . Cette différence entre les variances s'explique d'ailleurs par la différence entre les variances des estimateurs de  $\rho$ .

On remarque que la variance de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieure à celle de  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k \leq 21$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 22$ . Ces résultats sont semblables à ceux obtenus pour les estimateurs de  $\rho$ . Rappelons que la variance de  $\tilde{\rho}(\mathbf{T}_z)$  est inférieure à la variance de  $\tilde{\rho}(\tilde{T}_z)$  pour des valeurs de  $k \leq 17$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 18$ . Mentionnons également que la variance des trois estimateurs de  $R_0$  décroît lorsque la valeur de  $k$  augmente.

On a représenté graphiquement l'erreur quadratique moyenne des trois estimateurs de  $R_0$  en fonction de  $k$  pour le modèle avec la loi binomiale à la figure 4.15.



**Figure 4.15** Erreur quadratique moyenne de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi binomiale

On remarque que l'erreur quadratique moyenne de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieure à celle de  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k \leq 21$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 22$ . Ces résultats sont semblables à ceux obtenus pour les estimateurs de  $\rho$ . Rappelons que l'erreur quadratique moyenne de  $\bar{\rho}(\mathbf{T}_z)$  est inférieure à l'erreur quadratique moyenne de  $\bar{\rho}(\tilde{T}_z)$  pour des valeurs de  $k \leq 17$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 18$ .

#### 4.9 Intervalles de confiance asymptotiques pour $\rho$ et $R_0$

À la section 3.4, on a présenté des formules pour construire des intervalles de confiance asymptotiques de niveau  $(1 - \alpha) \cdot 100\%$  pour les paramètres  $\rho$  et  $R_0$  à partir de  $\bar{\rho}(t_z)$ ,

$\bar{\rho}(\mathbf{T}_Z)$ ,  $\bar{\rho}(\tilde{T}_Z)$ ,  $\tilde{R}_0(t_Z)$ ,  $\tilde{R}_0(\mathbf{T}_Z)$  et  $\tilde{R}_0(\tilde{T}_Z)$ . Les formules sont les suivantes :

$$\begin{aligned} & \left( \bar{\rho}(t_Z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}}, \bar{\rho}(\mathbf{T}_Z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}} \right), \\ & \left( \bar{\rho}(\mathbf{T}_Z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_Z^{(1)})]}{k}}, \bar{\rho}(\mathbf{T}_Z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(T_Z^{(1)})]}{k}} \right), \\ & \left( \bar{\rho}(\tilde{T}_Z) - z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}}, \bar{\rho}(\tilde{T}_Z) + z_{\alpha/2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}} \right), \end{aligned}$$

lorsqu'on utilise la variance théorique pour construire les intervalles de confiance pour  $\rho$ .

Pour les intervalles de confiance pour  $R_0$ , on a les formules suivantes :

$$\begin{aligned} & \left( \tilde{R}_0(t_Z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}}, \tilde{R}_0(t_Z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}} \right), \\ & \left( \tilde{R}_0(\mathbf{T}_Z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(\mathbf{T}_Z^{(1)})]}{k}}, \tilde{R}_0(\mathbf{T}_Z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(\mathbf{T}_Z^{(1)})]}{k}} \right), \\ & \left( \tilde{R}_0(\tilde{T}_Z) - z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}}, \tilde{R}_0(\tilde{T}_Z) + z_{\alpha/2} \frac{x_0}{\rho^2} \sqrt{\frac{\text{Var}[\hat{\rho}_1(t_Z)]}{k}} \right). \end{aligned}$$

En pratique, on doit estimer  $\text{Var}[\hat{\rho}_1(t_Z)]$  et  $\text{Var}[\hat{\rho}_1(T_Z^{(1)})]$  pour construire les intervalles de confiance pour  $\rho$  et  $R_0$ . À la section 3.4.3, on a proposé un estimateur sans biais pour  $\text{Var}[\hat{\rho}_1(t_Z)]$  obtenu à partir des  $\hat{\rho}_i(t_Z)$ , que l'on a noté  $S(t_Z)$ . De même, on a proposé un estimateur sans biais pour  $\text{Var}[\hat{\rho}_1(T_Z^{(1)})]$  obtenu à partir des  $\hat{\rho}_i(T_Z^{(i)})$ , que l'on a noté  $S(\mathbf{T}_Z)$ . Pour ce qui est de l'estimateur proposé pour  $V[\hat{\rho}_i(t_Z)]$  obtenu à partir des  $\hat{\rho}_i(\tilde{T}_Z)$ , que l'on a noté  $S(\tilde{T}_Z)$ , son espérance n'est pas égale à  $\text{Var}[\hat{\rho}_1(t_Z)]$ . Elle est en fait égale à  $\mathbb{E}(\text{Var}[\hat{\rho}_1(\tilde{T}_Z)|\tilde{T}_Z])$ . Cette valeur dépend de  $k$ , étant donné que la fonction de masse de  $\tilde{T}_Z$  dépend de  $k$ . Afin de voir l'ordre de grandeur du biais de  $S(\tilde{T}_Z)$  et son comportement en fonction de  $k$ , on l'a représenté graphiquement en fonction de  $k$  à la figure 4.16. On remarque que le biais relatif de  $S(\tilde{T}_Z)$  est inférieur à 0,5% pour de petites valeurs de  $k$  et qu'il diminue lorsque  $k$  augmente. Les intervalles de confiance construits à partir de  $S(\tilde{T}_Z)$  devraient donc bien se comporter, étant donné que le biais relatif de  $S(\tilde{T}_Z)$  est très petit.

En théorie, la proportion des intervalles de confiance à  $(1 - \alpha) \cdot 100\%$  obtenus en simulation qui contiennent le paramètre pour lequel les intervalles ont été construits

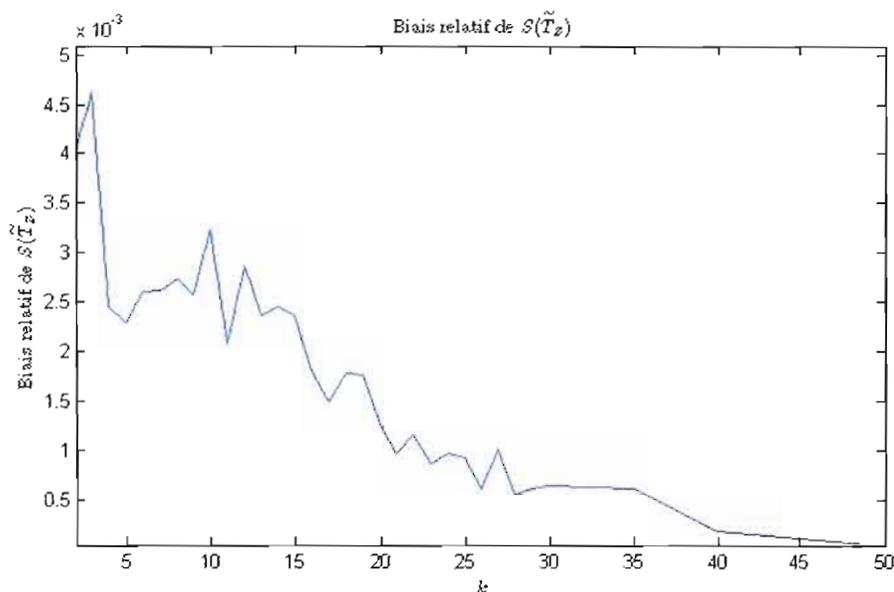


Figure 4.16 Biais relatif de  $S(\tilde{T}_Z)$

devrait être approximativement égale à  $(1 - \alpha) \cdot 100\%$ . Cependant, étant donné que les intervalles de confiance pour  $\rho$  et  $R_0$  sont asymptotiques, cette proportion n'est pas nécessairement approximativement égale à  $(1 - \alpha) \cdot 100\%$  pour de petites valeurs de  $k$ , le nombre d'épidémies considérées pour estimer  $\rho$ . On procède donc à une étude en simulation pour obtenir la proportion des intervalles de confiance pour  $\rho$  et  $R_0$  qui contiennent  $\rho$  ou  $R_0$ , selon le cas, pour différentes valeurs de  $k$ . Pour cette étude, on fixe le niveau de confiance des intervalles de confiance à 95%, ce qui correspond à  $\alpha = 0,05$ .

On commence par générer des données du nombre d'individus susceptibles et décédés aux instants  $\tau = 1, 2, \dots, n$  pour  $k$  épidémies. Ces données sont générées avec chacun des deux modèles proposés en suivant la procédure décrite à la section 4.4. Par la suite, on obtient les estimateurs  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$ ,  $\bar{\rho}(\tilde{T}_z)$ ,  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  à partir des données générées. On construit ensuite les intervalles de confiance asymptotiques à 95% pour  $\rho$  et  $R_0$  à partir des estimateurs de  $\rho$  et de  $R_0$ . Les valeurs de  $\rho$  et  $R_0$  sont connues, étant donné que les données ont été générées à partir des solutions de (2.1) obtenues

avec les paramètres présentés à la section 4.1. Rappelons qu'on a  $\rho = 66\,660$  et  $R_0 = 1, 5$ . On fait 5000 simulations et on compte le nombre de simulations pour lesquelles les intervalles de confiance asymptotiques pour  $\rho$  et  $R_0$  contiennent la vraie valeur de  $\rho$  et  $R_0$  respectivement. Cette procédure est répétée pour différentes valeurs de  $k$ .

Pour chacune des simulations, trois intervalles de confiance pour  $\rho$  et  $R_0$  ont été construits à partir de chacun des trois estimateurs, pour un total de neuf intervalles de confiance pour chacun des paramètres. Tout d'abord, un intervalle de confiance a été construit en utilisant la variance théorique de l'estimateur de  $\rho$  et  $z_{\alpha/2}$ , le quantile de la loi normale. Par la suite, deux intervalles ont été construits à partir de la variance estimée. Pour le premier intervalle construit à partir de la variance estimée, on a utilisé  $z_{\alpha/2}$ , alors que pour le deuxième, étant donné qu'on utilise la variance estimée et que certaines des valeurs de  $k$  considérées sont petites, on a utilisé  $t_{\alpha/2, k-1}$ , le quantile de la loi de Student à  $k - 1$  degrés de liberté. Rappelons que  $t_{\alpha/2, k-1}$  est tel que

$$P[T_{k-1} > t_{\alpha/2, k-1}] = \frac{\alpha}{2},$$

si  $T_{k-1}$  est de loi de Student à  $k - 1$  degrés de liberté.

Pour la présentation des résultats dans les tableaux, la lettre  $V$  signifie que les intervalles de confiance ont été construits à partir de la variance théorique, alors que la lettre  $S$  signifie qu'ils ont été construits à partir de la variance estimée. Pour ce qui est des quantiles, la lettre  $z$  signifie que les intervalles de confiance ont été construits en utilisant  $z_{\alpha/2}$ , alors que la lettre  $t$  signifie qu'ils ont été construits en utilisant  $t_{\alpha/2, k-1}$ . Rappelons qu'étant donné que les intervalles de confiance pour  $R_0$  font intervenir  $\rho$ , on remplace  $\rho$  par la valeur estimée de  $\rho$  lorsqu'on construit l'intervalle de confiance à partir de la variance estimée de l'estimateur de  $\rho$ .

Comme aux sections 4.7 et 4.8, on ne présente que les résultats pour le modèle avec la loi binomiale dans cette section. Les résultats pour le modèle avec la loi Poisson sont présentés à l'appendice A. Le tableau 4.1 présente la proportion des intervalles de

confiance asymptotiques à 95% pour  $\rho$  qui contiennent  $\rho$  pour les données générées par le modèle avec la loi binomiale.

**Tableau 4.1** Proportion des intervalles de confiance asymptotiques pour  $\rho$  construits à partir de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  qui contiennent  $\rho$ , modèle avec la loi binomiale

k	$\bar{\rho}(t_z)$			$\bar{\rho}(\mathbf{T}_z)$			$\bar{\rho}(\tilde{T}_z)$		
	V, z	S, z	S, t	V, z	S, z	S, t	V, z	S, z	S, t
2	0,951	0,703	0,948	0,932	0,555	<b>0,762</b>	0,430	0,316	0,425
5	0,947	0,882	0,952	0,943	0,905	<b>0,943</b>	0,632	0,589	0,634
10	0,954	0,924	0,951	0,945	0,908	<b>0,937</b>	0,784	0,760	0,782
15	0,954	0,934	0,950	0,942	0,926	<b>0,944</b>	0,860	0,841	0,857
20	0,952	0,935	0,951	0,942	0,921	<b>0,940</b>	0,894	0,877	0,892
25	0,949	0,940	0,952	0,941	0,925	<b>0,938</b>	0,919	0,909	0,920
30	0,952	0,946	0,953	0,931	0,923	<b>0,935</b>	0,933	0,928	0,934
35	0,955	0,945	0,951	0,935	0,928	0,936	0,944	0,934	<b>0,941</b>
40	0,952	0,942	0,950	0,934	0,927	0,936	0,945	0,936	<b>0,944</b>
45	0,955	0,944	0,949	0,928	0,924	0,931	0,951	0,940	<b>0,944</b>
50	0,949	0,945	0,949	0,933	0,926	0,934	0,946	0,942	<b>0,946</b>
100	0,954	0,950	0,952	0,912	0,911	0,913	0,954	<b>0,949</b>	0,952

Tout d'abord, on commente les résultats obtenus pour les intervalles de confiance asymptotiques construits à partir de la variance théorique, c'est-à-dire ceux dans les colonnes nommées V, z. Pour les intervalles de confiance construits à partir de  $\bar{\rho}(t_z)$ , la proportion des intervalles qui contiennent  $\rho$  est approximativement égale à 95% et ce, peu importe la valeur de k. On en conclut que la distribution de  $\bar{\rho}(t_z)$  s'approche bien par une loi normale même pour de petites valeurs de k.

À la section 3.4, on a mentionné qu'étant donné que  $\bar{\rho}(\mathbf{T}_z)$  est un estimateur biaisé pour  $\rho$ , on s'attend à ce que la probabilité que l'intervalle de confiance pour  $\rho$  construit

à partir de  $\bar{\rho}(\mathbf{T}_z)$  recouvre la valeur de  $\rho$  soit différente de  $1 - \alpha$ . Pour cet intervalle de confiance, on remarque en effet que la proportion des intervalles qui contiennent  $\rho$  diminue au fur et à mesure que  $k$  augmente. Cela s'explique par le fait que plus  $k$  augmente, plus les intervalles de confiance sont étroits et concentrés autour de l'espérance de  $\bar{\rho}(\mathbf{T}_z)$ , qui n'est pas égale à  $\rho$ . Donc, les intervalles de confiance contiennent  $\mathbb{E}[\bar{\rho}(\mathbf{T}_z)]$  environ 95% du temps, mais ils contiennent  $\rho$  moins souvent.

Enfin, pour de grandes valeurs de  $k$ , la proportion des intervalles de confiance construits à partir de  $\bar{\rho}(\tilde{T}_z)$  qui contiennent  $\rho$  est approximativement égale à 95%. Cependant, pour de petites valeurs de  $k$ , cette proportion est très petite. Cela s'explique par le fait qu'on utilise  $\text{Var}[\hat{\rho}_1(t_z)]$  plutôt que  $\text{Var}[\hat{\rho}_1(\tilde{T}_z)]$  pour construire les intervalles de confiance. À la figure 4.9, on remarque que la variance de  $\bar{\rho}(\tilde{T}_z)$  est beaucoup plus grande que celle de  $\bar{\rho}(t_z)$  pour de petites valeurs de  $k$ , ce qui est également le cas pour la variance de  $\hat{\rho}_1(\tilde{T}_z)$  par rapport à la variance de  $\hat{\rho}_1(t_z)$ . Étant donné que  $\bar{\rho}(\tilde{T}_z)$  est trop variable par rapport à la largeur des intervalles de confiance pour de petites valeurs de  $k$ , peu d'intervalles de confiance contiennent  $\rho$ .

Comme on l'a fait pour le biais relatif de  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$ , on s'intéresse à la valeur de  $k$  à partir de laquelle la proportion des intervalles de confiance qui contiennent  $\rho$  est plus près de 95% avec  $\bar{\rho}(\tilde{T}_z)$  qu'avec  $\bar{\rho}(\mathbf{T}_z)$ . On remarque que cette valeur se situe entre 25 et 30 épidémies.

En pratique, on ne connaît pas la variance des estimateurs de  $\rho$  et on doit l'estimer pour construire les intervalles de confiance pour  $\rho$ . Les résultats pour les intervalles de confiance construits à partir de la variance estimée sont ceux des colonnes nommées  $S, z$  et  $S, t$ . Les commentaires relatifs aux intervalles de confiance construits à partir de la variance théorique s'appliquent également pour ceux construits à partir de la variance estimée. On remarque toutefois que peu importe l'estimateur de  $\rho$  utilisé pour construire l'intervalle de confiance, la proportion des intervalles de confiance construits avec le quantile de la loi de Student (colonne  $S, t$ ) qui contiennent  $\rho$  est plus près de

95% que celle obtenue avec le quantile de la loi normale  $(S, z)$ . Il est donc préférable d'utiliser  $t_{\alpha/2, k-1}$  plutôt que  $z_{\alpha/2}$  pour construire les intervalles de confiance à partir de la variance estimée.

Comme on l'a fait pour les intervalles de confiance construits à partir de la variance théorique, on s'intéresse à la valeur de  $k$  à partir de laquelle la proportion des intervalles de confiance qui contiennent  $\rho$  est plus près de 95% avec  $\bar{\rho}(\tilde{T}_z)$  qu'avec  $\bar{\rho}(\mathbf{T}_z)$ , dans le cas où on utilise la variance estimée et  $t_{\alpha/2, k-1}$ . On remarque que cette valeur est entre 30 et 35 épidémies.

Le tableau 4.2 présente la proportion des intervalles de confiance asymptotiques à 95% pour  $R_0$  qui contiennent  $R_0$  pour les données générées par le modèle avec la loi binomiale. Les commentaires formulés sur les intervalles de confiance pour  $\rho$  s'appliquent également dans le cas des intervalles de confiance pour  $R_0$ . Notons que dans ce cas-ci, la valeur de  $k$  à partir de laquelle la proportion des intervalles de confiance qui contiennent  $R_0$  est plus près de 95% avec  $\tilde{R}_0(\tilde{T}_z)$  qu'avec  $\tilde{R}_0(\mathbf{T}_z)$ , est entre 25 et 30 épidémies, lorsqu'on utilise la variance estimée et  $t_{\alpha/2, k-1}$ .

Pour de petites valeurs de  $k$ , on peut s'attendre à ce que la précision des estimateurs de  $R_0$ , c'est-à-dire la moitié de la largeur de l'intervalle de confiance, soit grande par rapport à la valeur de  $R_0$ , étant donné que la variance des estimateurs de  $\rho$  est plutôt grande. On a calculé la précision moyenne des estimateurs de  $R_0$  en simulation, lorsqu'on estime la variance de l'estimateur de  $\rho$  et qu'on utilise  $t_{\alpha/2, k-1}$ . Le tableau 4.3 présente les résultats obtenus pour  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  avec des valeurs de  $k \leq 10$ . On remarque que la précision moyenne de  $\tilde{R}_0(t_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  est toujours inférieure à 0,03, ce qui est très petit par rapport à la valeur de  $R_0$  (1,5). Par contre, pour ce qui est de la précision moyenne de  $\tilde{R}_0(\mathbf{T}_z)$ , elle est égale à 1,359 lorsque  $k = 2$ , ce qui est très élevé, et égale à 0,425 pour  $k = 3$ . Elle est inférieure à 0,3 pour des valeurs de  $k \geq 4$ .

**Tableau 4.2** Proportion des intervalles de confiance asymptotiques pour  $R_0$  construits à partir de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  qui contiennent  $R_0$ , modèle avec la loi binomiale

k	$\tilde{R}_0(t_z)$			$\tilde{R}_0(\mathbf{T}_z)$			$\tilde{R}_0(\tilde{T}_z)$		
	V,z	S,z	S,t	V,z	S,z	S,t	V,z	S,z	S,t
2	0,951	0,703	0,948	0,952	0,625	<b>0,761</b>	0,430	0,317	0,425
5	0,947	0,882	0,952	0,957	0,894	<b>0,946</b>	0,632	0,589	0,634
10	0,954	0,925	0,951	0,945	0,906	<b>0,937</b>	0,784	0,760	0,782
15	0,955	0,933	0,950	0,950	0,922	<b>0,942</b>	0,860	0,841	0,857
20	0,952	0,935	0,951	0,943	0,918	<b>0,934</b>	0,894	0,878	0,892
25	0,949	0,940	0,951	0,946	0,921	<b>0,932</b>	0,919	0,909	0,920
30	0,952	0,946	0,953	0,940	0,917	0,927	0,933	0,928	<b>0,934</b>
35	0,955	0,945	0,951	0,942	0,922	0,930	0,944	0,934	<b>0,941</b>
40	0,952	0,942	0,950	0,940	0,922	0,930	0,946	0,936	<b>0,944</b>
45	0,955	0,944	0,949	0,933	0,916	0,924	0,951	0,940	<b>0,944</b>
50	0,949	0,945	0,949	0,939	0,918	0,924	0,945	0,942	<b>0,946</b>
100	0,954	0,949	0,952	0,917	0,905	0,908	0,954	<b>0,949</b>	0,952

**Tableau 4.3** Précision moyenne de  $\tilde{R}_0(t_Z)$ ,  $\tilde{R}_0(\mathbf{T}_Z)$  et  $\tilde{R}_0(\tilde{T}_Z)$  avec la variance estimée et le quantile de la loi de Student, modèle avec la loi binomiale

k	$\tilde{R}_0(t_Z)$	$\tilde{R}_0(\mathbf{T}_Z)$	$\tilde{R}_0(\tilde{T}_Z)$
2	0,024	1,359	0,027
3	0,007	0,425	0,008
4	0,005	0,283	0,005
5	0,004	0,226	0,004
6	0,003	0,192	0,004
7	0,003	0,172	0,003
8	0,003	0,156	0,003
9	0,003	0,144	0,003
10	0,002	0,134	0,003

#### 4.10 Synthèse des résultats

Rappelons que l'objectif était d'estimer  $R_0$  quand on dispose des données de  $k$  épidémies. Pour ce faire, trois estimateurs de  $R_0$  ont été proposés, soit  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$ . Aux sections 4.7, 4.8 et 4.9, on a comparé ces estimateurs en se basant sur les critères de performance suivants :

- (i) le biais relatif de l'estimateur ;
- (ii) l'erreur quadratique moyenne de l'estimateur ;
- (iii) la probabilité de couverture des intervalles de confiance construits à partir de l'estimateur ;
- (iv) la précision moyenne de l'estimateur.

Dans ce qui suit, on fait une synthèse de ces comparaisons. L'estimateur de  $R_0$  pour lequel les résultats sont les plus intéressants en simulation pour chacun des critères est  $\tilde{R}_0(t_z)$ . Cependant, comme on l'a déjà mentionné, cet estimateur n'est pas intéressant en pratique, étant donné qu'on ne connaît pas la valeur de  $t_z$ .

On s'attarde donc à la performance de  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction du nombre d'épidémies ( $k$ ). On a vu à la section 4.8 que le biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieur en valeur absolue à celui de  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k \leq 13$ , alors qu'il lui est supérieur pour des valeurs de  $k \geq 14$ . On a également vu que l'erreur quadratique moyenne de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieure à celle de  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k \leq 21$ , alors qu'elle lui est supérieure pour des valeurs de  $k \geq 22$ . Enfin, à la section 4.9, on a vu que la valeur de  $k$  à partir de laquelle la proportion des intervalles de confiance qui contiennent  $R_0$  est plus près de 95% avec  $\tilde{R}_0(\tilde{T}_z)$  qu'avec  $\tilde{R}_0(\mathbf{T}_z)$ , est entre 25 et 30 épidémies, lorsqu'on utilise la variance estimée et  $t_{\alpha/2, k-1}$ .

À la lumière de ces résultats, on suggère d'estimer  $R_0$  par  $\tilde{R}_0(\mathbf{T}_z)$  si le nombre d'épidémies ( $k$ ) pour lesquelles des données sont disponibles est inférieur à 20 et par  $\tilde{R}_0(\tilde{T}_z)$  si  $k \geq 20$ . Cette valeur de  $k$  représente en quelque sorte un compromis entre le biais

relatif et la probabilité de couverture des intervalles de confiance. En effet, le biais relatif de  $\tilde{R}_0(\tilde{T}_z)$  est inférieur à celui de  $\tilde{R}_0(\mathbf{T}_z)$  pour des valeurs de  $k \geq 14$ , alors que la proportion des intervalles de confiance à 95% qui contiennent  $R_0$  est plus grande pour les intervalles construits à partir de  $\tilde{R}_0(\tilde{T}_z)$  que pour ceux construits à partir de  $\tilde{R}_0(\mathbf{T}_z)$  pour des valeurs de  $k \geq 30$ . Les valeurs de  $k$  qu'on aurait pu choisir comme seuil à partir duquel il est préférable d'utiliser  $\tilde{R}_0(\tilde{T}_z)$  sont entre 14 et 30 épidémies. On a choisi  $k = 20$ , car pour des valeurs de  $k > 20$ , le biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  est plus du double de celui de  $\tilde{R}_0(\tilde{T}_z)$ .

Lorsque le nombre d'épidémies est égal à 2 ou 3, le biais relatif de  $\tilde{R}_0(\mathbf{T}_z)$  est inférieur à 0,2% en valeur absolue, ce qui est inférieur aux autres biais obtenus avec  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  pour des valeurs de  $k < 25$ . Cependant, la variance de  $\tilde{R}_0(\mathbf{T}_z)$  est très grande pour ces valeurs de  $k$ . Cela fait en sorte que la précision des estimateurs est, en moyenne, égale à 1,359 pour  $k = 2$  et à 0,425 pour  $k = 3$  en simulation. Comme  $R_0 = 1,5$ , ces valeurs sont beaucoup trop élevées. De plus, la proportion des intervalles de confiance à 95% qui contiennent  $R_0$  en simulation est d'environ 75% pour ces valeurs de  $k$ , ce qui est très faible. On suggère donc d'utiliser la méthode d'estimation proposée dans ce mémoire quand on dispose d'au moins 4 épidémies pour estimer  $R_0$ , étant donné que la précision des estimateurs est inférieure à 0,3 si  $k \geq 4$ .

## CONCLUSION

Dans ce mémoire, on a proposé deux nouveaux modèles pour la propagation des épidémies et une nouvelle méthode d'estimation du paramètre  $R_0$ , qui est d'une importance cruciale pour évaluer la virulence d'une épidémie.

Mentionnons qu'en simulation, les modèles se comportent bien pour les valeurs de  $t$  utiles pour l'estimation. L'espérance du nombre d'individus susceptibles et retirés est bien égale à la solution du système déterministe SIR, tel que désiré.

Par rapport à d'autres approches, la méthode d'estimation de  $R_0$  proposée a l'avantage d'être basée sur des données disponibles en pratique. De plus, on n'a pas à faire de suppositions quant aux valeurs de certains autres paramètres des modèles. Un désavantage de la méthode est qu'on a besoin de données sur plusieurs épidémies. Dans le cas où on considère le nombre d'épidémies ( $k$ ) comme étant le nombre d'années pour lesquelles on a des données sur la grippe saisonnière, par exemple, la valeur de  $k$  n'est pas très élevée. Cependant, si on considère  $k$  comme étant le nombre d'emplacements géographiques pour lesquels on a des données sur la grippe saisonnière pour une année,  $k$  peut prendre des valeurs plus grandes.

La procédure d'estimation de  $R_0$  proposée passe par l'estimation de  $\rho$ , un autre paramètre des modèles. On a présenté trois estimateurs pour  $\rho$ , chacun menant à un estimateur de  $R_0$ . Tout d'abord,  $\tilde{R}_0(t_z)$  a été présenté pour mettre en évidence le fait que l'estimateur de  $\rho$  associé,  $\tilde{\rho}(t_z)$ , est sans biais pour  $\rho$ , lorsqu'on suppose qu'on connaît la valeur de  $t_z$ . Étant donné qu'on ne connaît pas cette valeur, on a proposé deux estimateurs de  $R_0$  pour lesquels  $t_z$  est estimé, c'est-à-dire  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$ , qui peuvent être calculés en pratique.  $\tilde{R}_0(\tilde{T}_z)$  a l'avantage d'être obtenu à partir de  $\tilde{\rho}(\tilde{T}_z)$ ,

qui est un estimateur fortement convergent pour  $\rho$ . Quant à lui,  $\tilde{R}_0(\mathbf{T}_z)$  est obtenu à partir de  $\bar{\rho}(\mathbf{T}_z)$ , qui est un estimateur biaisé pour  $\rho$ . Le biais de  $\bar{\rho}(\mathbf{T}_z)$  ne diminue d'ailleurs pas lorsque  $k$  augmente.

On suggère d'utiliser  $\tilde{R}_0(\mathbf{T}_z)$  si  $k < 20$  et  $\tilde{R}_0(\tilde{T}_z)$  si  $k \geq 20$ . En suivant ces recommandations, le biais relatif de l'estimateur de  $R_0$  est inférieur à 0,7% peu importe la valeur de  $k$ . Pour des valeurs de  $k < 4$ , la précision des estimateurs est toutefois moins bonne. On suggère donc d'utiliser la méthode d'estimation proposée dans ce mémoire quand on dispose d'au moins 4 épidémies. Ces valeurs de  $k$  ont été obtenues en simulation, à partir des valeurs qu'on a choisies pour les paramètres. Elles peuvent donc varier selon le choix des valeurs des paramètres. Une recommandation générale serait donc d'utiliser  $\tilde{R}_0(\mathbf{T}_z)$  lorsque le nombre d'épidémies est petit et d'utiliser  $\tilde{R}_0(\tilde{T}_z)$  lorsque le nombre d'épidémies est grand.

Rappelons que la méthode d'estimation proposée se base sur l'instant auquel le nombre de nouveaux individus décédés est maximal. Une autre approche consisterait à se baser sur l'instant auquel le nombre de nouveaux individus infectés est maximal. Cette approche a d'ailleurs été explorée. Elle a l'avantage de produire des estimateurs dont la variance est beaucoup plus petite que ceux obtenus avec la méthode présentée dans ce mémoire. Cependant, pour calculer ces estimateurs, on doit avoir des données sur le nombre d'individus retirés, ce que l'on n'a pas en pratique. On peut obtenir une approximation de ces valeurs à partir du nombre d'individus décédés, mais cela suppose qu'on connaît la proportion des individus retirés qui sont en fait décédés. On pourrait estimer cette proportion, mais à ce moment-là, la variabilité des estimateurs augmenterait. En somme, cette façon de faire, bien que prometteuse à première vue, semble moins intéressante en pratique que celle qu'on a présentée dans ce mémoire.

## APPENDICE A

### RÉSULTATS POUR LE MODÈLE AVEC LA LOI DE POISSON

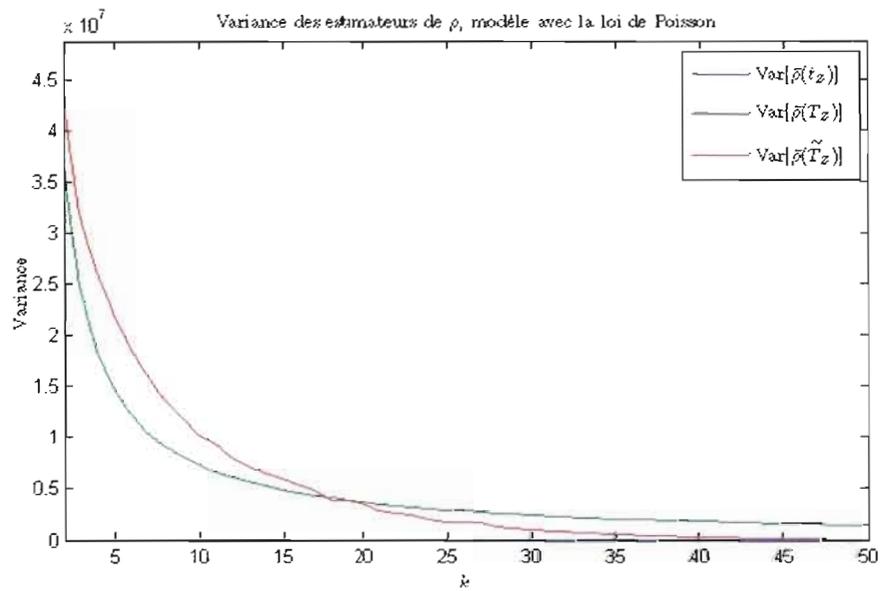
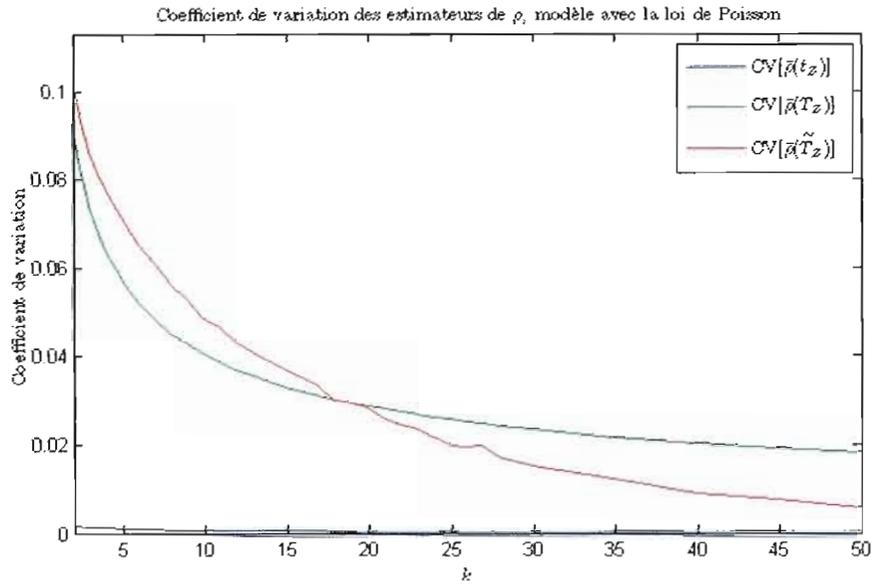
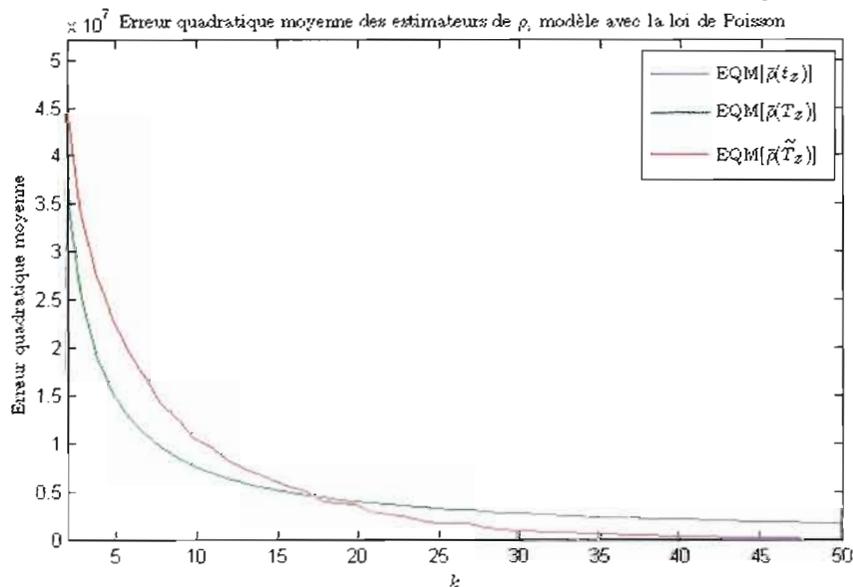


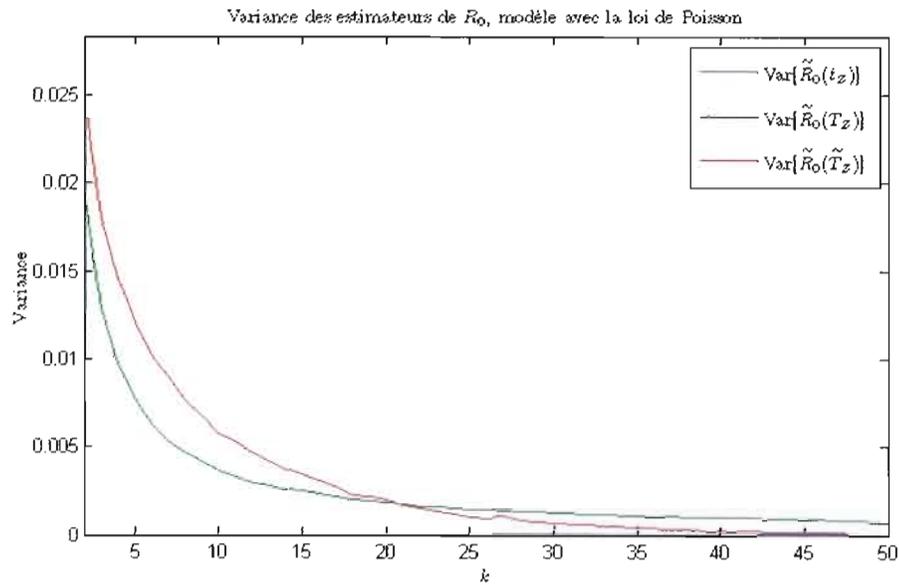
Figure A.1 Variance de  $\hat{\rho}(t_z)$ ,  $\hat{\rho}(T_z)$  et  $\hat{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi de Poisson



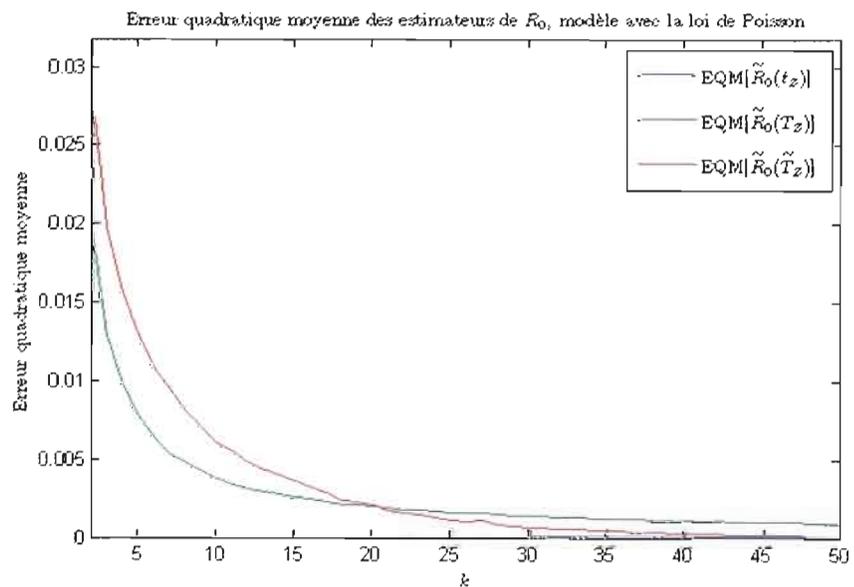
**Figure A.2** Coefficient de variation de  $\hat{\rho}(t_z)$ ,  $\hat{\rho}(T_z)$  et  $\hat{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi de Poisson



**Figure A.3** Erreur quadratique moyenne de  $\hat{\rho}(t_z)$ ,  $\hat{\rho}(T_z)$  et  $\hat{\rho}(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi de Poisson



**Figure A.4** Variance de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi de Poisson



**Figure A.5** Erreur quadratique moyenne de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  en fonction de  $k$ , modèle avec la loi de Poisson

**Tableau A.1** Proportion des intervalles de confiance asymptotiques pour  $\rho$  construits à partir de  $\bar{\rho}(t_z)$ ,  $\bar{\rho}(\mathbf{T}_z)$  et  $\bar{\rho}(\tilde{T}_z)$  qui contiennent  $\rho$ , modèle avec la loi de Poisson

k	$\bar{\rho}(t_z)$			$\bar{\rho}(\mathbf{T}_z)$			$\bar{\rho}(\tilde{T}_z)$		
	V, z	S, z	S, t	V, z	S, z	S, t	V, z	S, z	S, t
2	0,951	0,685	0,947	0,927	0,552	<b>0,757</b>	0,435	0,313	0,434
5	0,952	0,875	0,946	0,939	0,898	<b>0,937</b>	0,627	0,578	0,623
10	0,951	0,919	0,951	0,957	0,923	<b>0,951</b>	0,795	0,768	0,795
15	0,952	0,933	0,953	0,948	0,928	<b>0,948</b>	0,858	0,843	0,859
20	0,952	0,938	0,951	0,945	0,928	<b>0,947</b>	0,898	0,885	0,898
25	0,952	0,937	0,949	0,943	0,934	<b>0,945</b>	0,925	0,911	0,923
30	0,951	0,945	0,953	0,936	0,926	<b>0,939</b>	0,935	0,929	0,937
35	0,952	0,946	0,955	0,933	0,925	0,935	0,941	0,936	<b>0,944</b>
40	0,948	0,944	0,950	0,934	0,933	0,941	0,942	0,938	<b>0,944</b>
45	0,951	0,945	0,951	0,930	0,926	0,931	0,946	0,941	<b>0,947</b>
50	0,947	0,946	0,952	0,928	0,924	0,935	0,945	0,944	<b>0,949</b>
100	0,950	0,948	0,950	0,906	0,906	0,909	0,950	0,948	<b>0,950</b>

**Tableau A.2** Proportion des intervalles de confiance asymptotiques pour  $R_0$  construits à partir de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  qui contiennent  $R_0$ , modèle avec la loi de Poisson

k	$\tilde{R}_0(t_z)$			$\tilde{R}_0(\mathbf{T}_z)$			$\tilde{R}_0(\tilde{T}_z)$		
	V, z	S, z	S, t	V, z	S, z	S, t	V, z	S, z	S, t
2	0,952	0,686	0,947	0,943	0,604	<b>0,758</b>	0,435	0,314	0,434
5	0,952	0,874	0,946	0,953	0,887	<b>0,938</b>	0,627	0,578	0,623
10	0,951	0,919	0,950	0,957	0,923	<b>0,949</b>	0,795	0,768	0,795
15	0,952	0,933	0,953	0,955	0,924	<b>0,947</b>	0,858	0,843	0,860
20	0,952	0,938	0,951	0,950	0,926	<b>0,939</b>	0,898	0,885	0,897
25	0,952	0,936	0,949	0,950	0,925	<b>0,940</b>	0,925	0,911	0,923
30	0,951	0,945	0,952	0,944	0,920	0,932	0,936	0,929	<b>0,936</b>
35	0,952	0,946	0,955	0,937	0,919	0,932	0,941	0,936	<b>0,944</b>
40	0,948	0,944	0,950	0,943	0,925	0,934	0,942	0,938	<b>0,944</b>
45	0,951	0,945	0,951	0,936	0,920	0,927	0,946	0,941	<b>0,947</b>
50	0,947	0,946	0,952	0,935	0,916	0,925	0,945	0,944	<b>0,949</b>
100	0,951	0,948	0,950	0,913	0,899	0,904	0,951	0,948	<b>0,950</b>

**Tableau A.3** Précision moyenne de  $\tilde{R}_0(t_z)$ ,  $\tilde{R}_0(\mathbf{T}_z)$  et  $\tilde{R}_0(\tilde{T}_z)$  avec la variance estimée et le quantile de la loi de Student, modèle avec la loi de Poisson

k	$\tilde{R}_0(t_z)$	$\tilde{R}_0(\mathbf{T}_z)$	$\tilde{R}_0(\tilde{T}_z)$
2	0,029	1,354	0,033
3	0,009	0,427	0,010
4	0,006	0,285	0,006
5	0,005	0,228	0,005
6	0,004	0,194	0,004
7	0,004	0,173	0,004
8	0,003	0,157	0,004
9	0,003	0,144	0,003
10	0,003	0,135	0,003

## RÉFÉRENCES

- Bartlett M. S., 1949. « Some evolutionary stochastic processes ». *Journal of the Royal Statistical Society*, ser. B, vol. 11, p. 211–229.
- Bartlett M. S., 1966. *An Introduction to Stochastic Processes : with special reference to methods and applications*, Cambridge (MA) : Cambridge University Press, 362 p.
- Bickel P.J., Doksum K.A., 2001. *Mathematical Statistics : Basic Ideas and Selected Topics, Volume I*, 2<sup>e</sup> éd. rev. et augm., Upper Saddle River (N.J.) : Prentice-Hall, 556 p.
- Cruz-Pacheco G., Duran L., Esteva L., Minzoni A., Lopez-Cervantes M., Panayotaros P., Ahued Ortega A., Villasenor Ruiz I., 2009. « Modelling of the influenza A(H1N1)v outbreak in Mexico City, April-May 2009, with control sanitary measures ». *Eurosurveillance*, vol. 14, no 26, p. 19254.
- Kendall D. G., 1956. « Deterministic and stochastic epidemics in closed populations ». *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, p. 149–165.
- Kermack W. O., McKendrick A. G., 1927. « A contribution to the mathematical theory of epidemics ». *Proceedings of the Royal Society*, ser A, vol. 115, p. 700–721.
- McKendrick A. G., 1926. « Applications of mathematics to medical problems ». *Proceedings of the Edinburgh Mathematical Society*, vol. 44, p. 1–34.
- Parzen E., 1962. *Stochastic Processes*, Oakland (CA) : Holden-Day, 324 p.
- Van Den Broek J., Heesterbeek H., 2007. « Nonhomogeneous birth and death models for epidemic outbreak data ». *Biostatistics*, vol. 8, p. 453–467.