



**HAL**  
open science

# Micro-macro analysis of heterogenous age-structured populations dynamics. Application to self-exciting processes and demography

Alexandre Boumezoued

► **To cite this version:**

Alexandre Boumezoued. Micro-macro analysis of heterogenous age-structured populations dynamics. Application to self-exciting processes and demography. Modeling and Simulation. Université Pierre et Marie Curie - Paris VI, 2016. English. NNT: 2016PA066085 . tel-01508922v2

**HAL Id: tel-01508922**

**<https://theses.hal.science/tel-01508922v2>**

Submitted on 15 Apr 2017

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

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



**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité

**Mathématiques Appliquées**

École doctorale Sciences Mathématiques de Paris Centre

Présentée par

**Alexandre BOUMEZOUED**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

**Approches micro-macro des dynamiques de  
populations hétérogènes structurées par âge  
Application aux processus auto-excitants et à la démographie**

*Micro-macro analysis of heterogenous age-structured populations  
dynamics*

*Application to self-exciting processes and demography*

Directrice de thèse : Mme Nicole EL KAROUI

Co-directeur de thèse : M. Stéphane LOISEL

---

# Remerciements

# Table des matières

<b>1</b>	<b>Présentation des motivations et travaux</b>	<b>10</b>
1.1	Motivations . . . . .	10
1.2	Modélisation dynamique d'une population . . . . .	36
1.3	La pyramide des âges des processus de Hawkes . . . . .	67
1.4	Les causes de décès dans la dynamique de population . . . . .	87
1.5	Estimation de l'hétérogénéité de la mortalité en présence de censure par intervalle . . . . .	104
1.6	Comportement macroscopique de trajectoires de vie accélérées . . . . .	143
<b>2</b>	<b>Bridging the gap from microsimulation practice to population mo- dels: a survey<sup>†</sup></b>	<b>157</b>
2.1	Introduction . . . . .	158
2.2	Complexity and heterogeneity in demographic transition analysis . . . . .	159
2.3	Point and population processes . . . . .	169
2.4	Microsimulation . . . . .	185
2.5	Macroscopic behavior of linear and non-linear dynamics . . . . .	196
<b>3</b>	<b>Population viewpoint on Hawkes processes<sup>†</sup></b>	<b>209</b>
3.1	Introduction . . . . .	209
3.2	Population point of view . . . . .	212
3.3	The exponential case generalized . . . . .	215
3.4	Pathwise representation of Hawkes population . . . . .	221
3.5	Towards more general Hawkes processes . . . . .	225
<b>4</b>	<b>Cause-of-death mortality: what can be learned from population dynamics?<sup>†</sup></b>	<b>237</b>
4.1	Introduction . . . . .	237
4.2	The Model . . . . .	240
4.3	Data . . . . .	250
4.4	Results . . . . .	253
4.5	Concluding Remarks . . . . .	262

<b>5</b>	<b>Measuring mortality heterogeneity dynamics with interval-censored data<sup>†</sup></b>	<b>267</b>
5.1	Introduction . . . . .	268
5.2	Longitudinal data . . . . .	271
5.3	Parametric maximum likelihood method for interval-censored data . .	274
5.4	Results . . . . .	281
5.5	Aggregate mortality and the impact of heterogeneity on life insurance products . . . . .	286
<b>6</b>	<b>Macroscopic behavior of heterogenous populations with fast random life histories<sup>†</sup></b>	<b>298</b>
6.1	Introduction . . . . .	299
6.2	Setting and main results . . . . .	300
6.3	Proofs . . . . .	307
6.4	Examples and numerical illustration . . . . .	319
6.5	Conclusion . . . . .	323

## Résumé

Cette thèse porte sur la modélisation de la dynamique des populations et de ses applications, à la démographie et l'actuariat d'une part, et à l'étude des processus de Hawkes d'autre part. Ces travaux de thèse proposent d'explorer à travers différents points de vue comment se déforme la structure d'une population, tant concernant la répartition des âges que sa composition en terme de caractéristiques. À travers cinq chapitres, nous déclinons une même philosophie qui, pour comprendre comment évoluent des quantités agrégées, propose d'étudier la dynamique de la population à une échelle plus fine, celle de l'individu.

Après un premier chapitre introductif en langue française, détaillant les motivations et les principales contributions, nous proposons d'abord dans le Chapitre 2 la description du cadre général de la modélisation dynamique aléatoire de populations structurées en caractéristiques et en âges, sur la base de Bensusan et al. (2010–2015), ainsi que plusieurs exemples motivés par les applications démographiques et actuarielles. Nous détaillons la construction mathématique de tels processus ainsi que le lien avec les équations déterministes classiques en démographie. Nous discutons également l'impact de l'hétérogénéité sur l'exemple d'un *effet cohorte*, ainsi que le rôle de l'environnement aléatoire.

Les deux chapitres suivants mettent en avant l'importance de la pyramide des âges. Le modèle de population général issu du Chapitre 2 est décliné dans le Chapitre 3 pour étudier des processus de Hawkes avec immigrants généraux, pour lesquels nous exploitons le concept de pyramide des âges. Dans cette étude théorique, basée sur Boumezoued (2015b), nous établissons de nouveaux résultats sur leur distribution pour une classe de fonctions qui généralisent le cas exponentiel étudié jusqu'ici. Dans le Chapitre 4, qui reprend Arnold et al. (2015), nous analysons l'impact de changements dans la mortalité par causes de décès sur la dynamique de la pyramide des âges, et en particulier sur le ratio de dépendance qui est un indicateur crucial du vieillissement de la population. En incluant le jeu des naissances dans la dynamique, ce travail de simulations, basé sur les données de l'OMS, permet de compléter la littérature existante sur les causes de décès qui se focalise traditionnellement sur des indicateurs de mortalité.

Les deux derniers chapitres étudient plus particulièrement l'hétérogénéité des populations. Le Chapitre 5, basé sur Boumezoued et al. (2015), propose de mesurer l'hétérogénéité de la mortalité dans les données de l'Échantillon Démographique Permanent de l'INSEE. Dans le cadre de cette contribution d'adaptation de méthodes statistiques et de sa mise en oeuvre sur données réelles, nous proposons une méthode d'estimation paramétrique par maximum de vraisemblance pour les modèles multi-états qui prend en compte à la fois la censure par intervalle, caractéristique des données longitudinales issues du recensement, et également le retour dans les états

intermédiaires. Enfin, le Chapitre 6, tiré de Boumezoued (2015a), reprend le modèle général du Chapitre 2 dans lequel les individus peuvent donner naissance, changer de caractéristiques et décéder. La contribution de cette partie théorique est d'étudier le comportement de la population lorsque les caractéristiques individuelles changent fréquemment. Nous établissons un théorème limite en grande population pour le processus de pyramide des âges, dont le comportement est alors décrit par des taux de naissance et mort agrégés sur la structure stable en terme de caractéristiques.

**Mots clés :** Dynamique des populations, démographie mathématique, science actuarielle, microsimulation, longévité, mortalité, pyramide des âges, hétérogénéité, processus de naissance et mort, processus ponctuels, processus de Hawkes, mortalité par causes de décès, modèles multi-états, censure par intervalle, maximum de vraisemblance paramétrique, théorèmes limites pour les processus, trajectoires de vie accélérées.

## Abstract

This thesis focuses on population dynamics models and their applications, on one hand to demography and actuarial science, and on the other hand to Hawkes processes. This work explores through several viewpoints how population structures evolve over time, both in terms of ages and characteristics. In five chapters, we develop a common philosophy which studies the population at the scale of the individual in order to better understand the behavior of aggregate quantities.

The first chapter introduces the motivations and details the main contributions in French. In Chapter 2, based on Bensusan et al. (2010–2015), we survey the modeling of characteristic and age-structured populations and their dynamics, as well as several examples motivated by demographic issues. We detail the mathematical construction of such population processes, as well as their link with well known deterministic equations in demography. We illustrate the simulation algorithm on an example of *cohort effect*, and we also discuss the role of the random environment. The two following chapters emphasize on the importance of the age pyramid. Chapter 3 uses a particular form of the general model introduced in Chapter 2 in order to study Hawkes processes with general immigrants. In this theoretical part based on Boumezoued (2015b) we use the concept of age pyramid to derive new distribution properties for a class of fertility functions which generalize the popular exponential case. Chapter 4 is based on Arnold et al. (2015) and analyses the impact of cause-of-death mortality changes on the population age pyramid, and in particular on the dependency ratio which is crucial to measure population ageing. By including birth patterns, this numerical work based on WHO data gives additional insights compared to the existing literature on causes of death focusing only on mortality indicators.

The last two chapters focus on population heterogeneity. The aim of Chapter 5, based on Boumezoued et al. (2015), is to measure mortality heterogeneity on French longitudinal data called *Échantillon Démographique Permanent*. In this work, inspired by recent advances in the statistical literature, we develop a parametric maximum likelihood method for multi-state models which takes into account both interval censoring and reversible transitions. Finally, Chapter 6, based on Boumezoued (2015a), considers the general model introduced in Chapter 2 in which individuals can give birth, change their characteristics and die. The contribution of this theoretical work is the analysis of the population behavior when individual characteristics change very often. We establish a large population limit theorem for the age pyramid process, whose dynamics is described at the limit by birth and death rates which are averaged over the stable population composition.

**Keywords:** Population dynamics, mathematical demography, actuarial science, microsimulation, longevity, mortality, age pyramid, heterogeneity, birth-death pro-



cesses, point processes, Hawkes processes, causes-of-death mortality, multi-state models, interval censoring, parametric maximum likelihood, limit theorems for stochastic processes, fast life trajectories.

## Contributions issues de la thèse

Nous mentionnons ici les contributions issues de la thèse, qui seront introduites dans les Sections 1.2 à 1.6 et détaillées dans leur intégralité dans les Chapitres 2 à 6.

- Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Bridging the gap from microsimulation practice to population models: a survey. *Work in progress*
- Boumezoued, A. 2015. Population viewpoint on Hawkes processes. HAL preprint Id: hal-01149752, To appear in *Advances in Applied Probability* 48.2 (June 2016)
- Arnold, S., A. Boumezoued, H. Labit Hardy, N. El Karoui. 2015. Cause-of-death mortality: What can be learned from population dynamics? HAL preprint Id: hal-01157900
- Boumezoued, A., N. El Karoui, S. Loisel 2015, Measuring mortality heterogeneity dynamics with interval-censored data. *Working paper*
- Boumezoued, A. 2015, Macroscopic behavior of heterogenous populations with fast random life histories. *Working paper*

# Chapitre 1

## Présentation des motivations et travaux

Cette thèse se divise en six chapitres. Nous décrivons dans ce chapitre introductif les motivations qui sous-tendent les différents travaux, et nous détaillons pour chaque thème le cadre et les principales contributions. Les cinq chapitres suivants peuvent être appréhendés indépendamment.

### 1.1 Motivations

Dans la plupart des pays développés, on assiste à un accroissement de la durée de vie, matérialisé par l'indicateur d'espérance de vie. Ces améliorations de longévité mettent en difficulté les systèmes existants de mutualisation et de transfert des risques, en particulier les systèmes publics de retraite et de santé, les fonds de pension et les sociétés d'assurance. Pour les produits d'assurance qui concernent le versement d'une rente jusqu'au décès, le risque de longévité consiste en l'allongement imprévu de la durée de vie. Pour faire face à ce risque, les sociétés d'assurance doivent prévoir des montants de capitaux toujours plus importants pour faire face à un scénario de longévité extrême, ce qui est requis par la réglementation européenne Solvabilité 2.

La mesure de longévité à l'échelle d'un pays passe par des tables de mortalité qui donnent, pour les hommes et les femmes, une estimation de taux de mortalité pour une classe d'âge et une année données. Ces indicateurs sont cruciaux d'un point de vue démographique pur, mais également car ils servent de référence pour des secteurs comme l'assurance et l'économie. Pour les calculer, il est nécessaire de mesurer le nombre d'individus en vie et décédés par classe d'âge, ce qui requiert une bonne compréhension de la pyramide des âges. L'allongement des durées de vie s'ajoute au jeu de la fertilité des populations, ce qui a pour conséquence une modification des équilibres dans cette pyramide des âges. Celle-ci permet de calculer les indicateurs

de mortalité agrégés, et elle est aussi au coeur des systèmes de retraite par répartition qui doivent prévoir au mieux les transferts inter-générationnels.

L'évolution dans le temps d'un indicateur de mortalité agrégé, c'est-à-dire calculé de manière globale sur une population d'intérêt, est complexe. Des modifications de l'environnement des individus, par exemple en terme d'hygiène et d'innovation médicale, sont des facteurs communs à l'ensemble de la population qui amènent la longévité à s'améliorer. En outre, à un niveau individuel, de nombreux facteurs semblent influencer sur la durée de vie. En particulier, selon leurs caractéristiques propres, les individus présentent des niveaux de longévité différents. On citera par exemple les grandes différences qui persistent entre les individus selon leur catégorie sociale ou leur statut matrimonial, et qui seront d'intérêt dans ces travaux de thèse. La population nationale, faite de différentes sous-populations aux caractéristiques spécifiques, évolue ainsi de manière complexe selon les niveaux de mortalité et de fertilité de chacune. Les assurances et les fonds de pension font également face à cette hétérogénéité. En particulier, il y a un biais entre la mortalité d'un portefeuille d'assurés et la mortalité nationale de référence : cela s'appelle le risque de base.

Dans ce cadre, nous proposons un point de vue microscopique sur la quantification de la longévité. Ce point de vue est complémentaire car il permet de répondre à un certain nombre de questions au-delà de celles traitées habituellement. En particulier, il met l'accent sur la dynamique de la population à l'échelle individuelle et propose une description détaillée de la dynamique de la pyramide des âges et de l'hétérogénéité des populations. Dans le même temps, ce point de vue permet d'affiner la compréhension des quantités d'intérêt traditionnelles en se donnant davantage d'information sur les phénomènes en jeu.

### 1.1.1 Quantifier la mortalité agrégée

Commençons par décrire un formalisme simple. Soit  $\tau$  la *durée de vie* d'un individu représentatif. La *force de mortalité à l'âge  $a$* , notée  $\mu(a)$ , caractérise la loi de cette durée de vie à travers la *probabilité de survie* jusqu'à l'âge  $a$  :

$$S(a) = \mathbb{P}(\tau > a) = \exp\left(-\int_0^a \mu(s)ds\right).$$

Comme  $\mu(a)da = \mathbb{P}(\tau \in [a, a + da) \mid \tau > a)$ , on appelle aussi  $\mu(a)$  la probabilité instantanée de décès à l'âge  $a$ . Ainsi, pour comprendre comment évolue la mortalité avec l'âge, il est intéressant d'étudier la forme de la force de mortalité  $\mu(a)$  en fonction de  $a$ . La loi de Gompertz (voir Gompertz (1825)) stipule que pour les populations humaines, la force de mortalité est exponentielle en l'âge pour des âges suffisamment élevés, c'est-à-dire qu'il existe deux constantes positives  $\alpha$  et  $\beta$  telles que pour de tels âges  $a$ , la relation  $\mu(a) = \alpha \exp(\beta a)$  est vérifiée. Pour s'en

convaincre, on peut tracer le logarithme de la force de mortalité en fonction de l'âge et vérifier pour quelle plage d'âges elle est bien linéaire. Les données pour la France peuvent par exemple être trouvées dans la *Human Mortality Database* (voir HMD (2014)). Nous reviendrons dans la suite sur le calcul de ces indicateurs nationaux. Pour l'instant, regardons pour l'année 2008 par exemple la force de mortalité et son logarithme, représentés en Figure 1.1. Décrivons la structure de la mortalité en fonction de l'âge sur le logarithme de la force de mortalité, qui a l'avantage de pouvoir mieux comparer les différents âges. On note d'abord la mortalité infantile, très présente jusqu'aux âges autour de 2-3 ans, le creux de mortalité pour les âges autour de 10 ans, la bosse à accident des personnes d'âges autour de 20 ans, puis à partir de 40 ans une croissance à *peu près* linéaire de la force de mortalité jusqu'aux grands âges. À ce stade, on dispose d'une estimation de la "mortalité" nationale. En pratique, on sait que celle-ci diffère suivant le sexe, ce qui est en fait catégorisé dans les données agrégées dans HMD (2014). Sur la Figure 1.2, on représente la force de mortalité et son logarithme pour les hommes d'une part et les femmes d'autre part. Plusieurs observations majeures apparaissent : d'abord, les femmes présentent un niveau de mortalité inférieur comparé aux hommes. Ensuite, cette différence se fait surtout entre les âges 15 et 100 ans, tandis que la mortalité infantile et celle aux âges très élevés semble du même ordre de grandeur entre les deux sexes. Enfin, la structure en âge de la mortalité entre les deux sexes est différente : l'observation la plus frappante concerne la bosse à accidents très accentuée chez les hommes et moins présente chez les femmes.

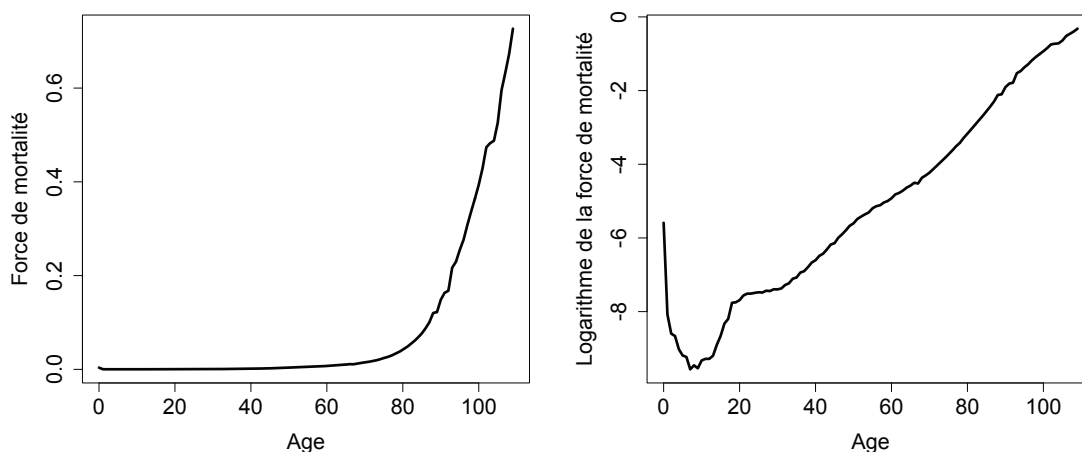


FIGURE 1.1 – Force de mortalité et son logarithme pour la France en 2008

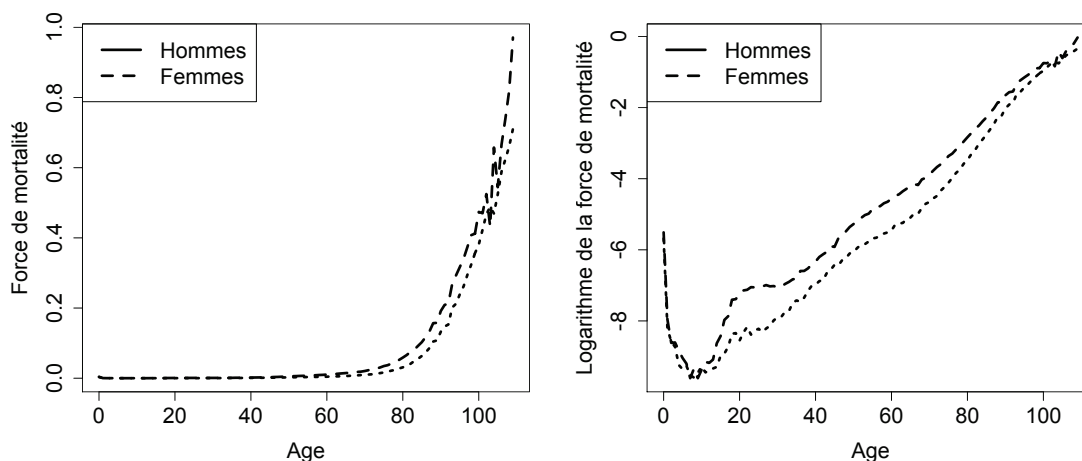


FIGURE 1.2 – Force de mortalité et son logarithme pour la France en 2008 et par sexe

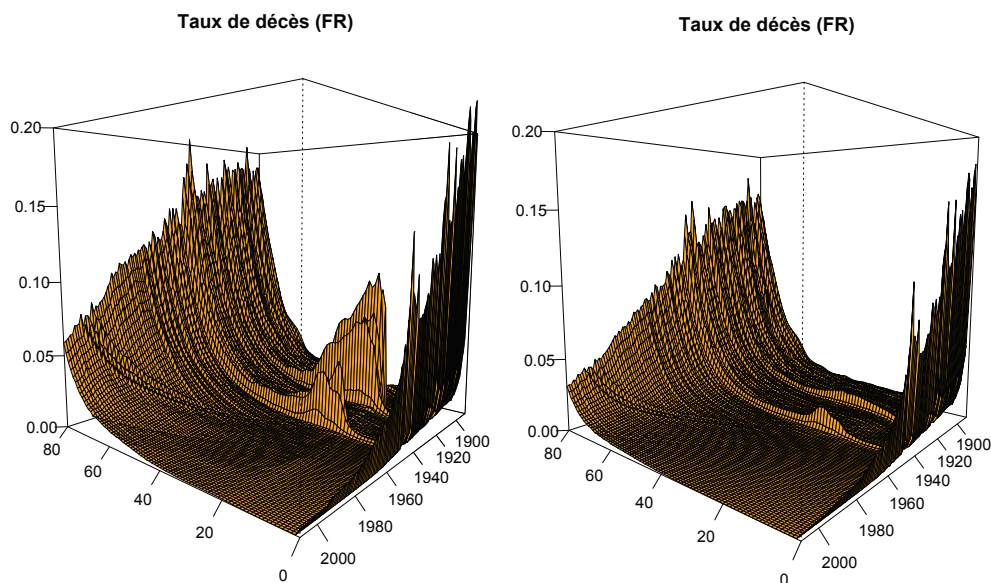


FIGURE 1.3 – Force de mortalité pour la France pour les hommes (gauche) et les femmes (droite)

Ces analyses statiques sur la structure en âge de la mortalité donnent de premières intuitions. En pratique, on sait qu'au global, la mortalité diminue au cours du temps. Pour s'en convaincre, on peut observer la force de mortalité en fonction de l'âge et pour différentes années (ce qui sera rigoureusement défini dans la suite), ce qui est représenté sur la Figure 1.3 pour les hommes et les femmes. Sur ces graphiques, on distingue la diminution de la mortalité, en particulier de la mortalité infantile et de celle aux âges élevés. On remarque également grâce à la dimension temporelle l'impact des événements catastrophiques : l'effet conjoint de la Première Guerre Mondiale et de la grippe espagnole de 1918, et celui de la Seconde Guerre

Mondiale. Si les niveaux entre hommes et femmes sont différents, l'impact des événements catastrophiques est également modulé selon la catégorie (homme ou femme) que l'on regarde. Par ces observations, on a envie de définir une force de mortalité  $\mu(a, t)$  qui serait la probabilité, sachant que l'on a l'âge  $a$  au temps  $t$  de décéder instantanément. Pour définir cette quantité, il faut en fait augmenter la dimension du problème, par exemple en ajoutant la date de naissance dans les variables d'intérêt. Notons  $\nu$  la date de naissance (aléatoire) de l'individu de durée de vie  $\tau$ . On peut définir la probabilité de décès instantanée pour cet individu en remarquant que s'il est né au temps  $\nu$ , alors il a l'âge  $a$  au temps  $a + \nu$  et ainsi

$$\mathbb{P}(\tau \in [a, a + da) \mid \tau > a, \nu) = \mu(a, \nu + a)da.$$

En d'autres termes, la probabilité de survie dépend de la date de naissance de l'individu et s'exprime comme

$$S(a \mid \nu) = \exp \left( - \int_0^a \mu(s, \nu + s) ds \right).$$

La lecture par âge et temps de ces forces de mortalité peut se faire dans ce que l'on appelle un diagramme de Lexis, représenté sous une forme simple en Figure 1.4. Ce diagramme se lit de la façon suivante : en abscisse se trouvent les années, et pour une année donnée, en traçant une droite verticale, on fait une observation de tous les âges dans la population au temps  $t$ . Ceci donne une photographie de la *pyramide des âges* au temps  $t$ . En ordonnée l'axe des âges permet, en traçant une droite horizontale, de regarder à travers le temps toutes les personnes d'un même âge. Ceci est utile en particulier pour comparer la mortalités et son amélioration, en les étudiant classe d'âge par classe d'âge. Faire cette observation revient à regarder l'évolution dans le temps de la pyramide des âges, ce qui est en fait un phénomène complexe. Enfin, il reste à parcourir le diagramme en diagonale, c'est à dire en choisissant tous les âges  $a$  et temps  $t$  tel que  $t - a$  reste constant. Comme  $t - a$  est la date de naissance, on ne fait rien d'autre que d'observer les personnes nées au même moment, ce que l'on appelle une *cohorte* (voir Figure 1.4). En lisant en diagonale, on suit donc le même groupe de personnes qui vieillit au cours du temps. Si l'on connaît le nombre de personnes nées dans cette cohorte, ainsi que la force de mortalité, on peut en déduire la trajectoire moyenne de la cohorte qui vieillit au cours du temps. Si on note  $g(0, \nu)$  la quantité d'individus d'âge 0 au temps  $\nu$ , c'est-à-dire le nombre de personnes nées au temps  $\nu$ , alors la quantité *moyenne*  $g(a, \nu + a)$  d'individus d'âge  $a$  au temps  $\nu + a$  est donnée comme la fraction d'individus ayant survécu par

$$g(a, \nu + a) = g(0, \nu)S(a \mid \nu) = g(0, \nu) \exp \left( - \int_0^a \mu(s, \nu + s) ds \right).$$

Si l'on dérive par rapport à l'âge  $a$ , on obtient la dynamique

$$\frac{dg(a, \nu + a)}{da} = -\mu(a, \nu + a)g(a, \nu + a) \text{ ou encore si l'on dérive conjointement par rapport}$$

à l'âge et au temps,

$$(\partial_a + \partial_t)g(a, t) = -\mu(a, t)g(a, t). \quad (1.1)$$

Avec cette équation aux dérivées partielles, nous venons d'écrire la dynamique déterministe d'une cohorte qui sera une composante essentielle des modèles de population. Notons que l'on a une forme très simple d'équation dite de *transport*, qui traduit le fait qu'au cours du temps, le nombre d'individus décroît du fait des morts, mais aussi que l'âge se transporte (terme en  $\partial_a$ ) le long de l'axe du temps : c'est la formalisation du vieillissement.

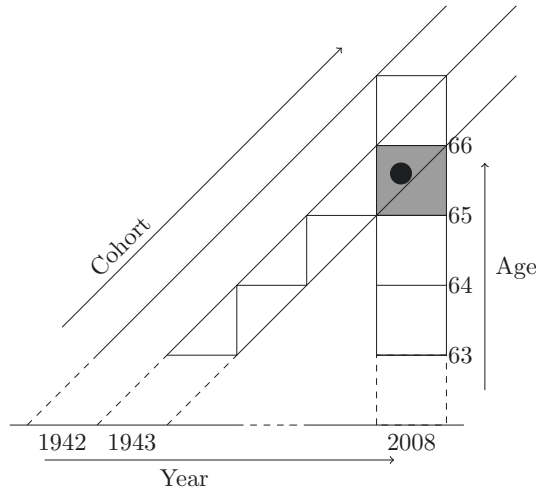


FIGURE 1.4 – Diagramme de Lexis simplifié  
Figure 1.1: Lexis diagram: Age-year-cohort diagram.

Revenons dans ce cadre dynamique à la forme de  $\mu(a, t)$  en fonction de l'âge et en particulier à la loi de Gompertz. On peut s'intéresser à cette loi à une date donnée, en affirmant qu'à chaque année  $t$  et pour des âges raisonnables, on a la relation suivante

$$\mu(a, t) \approx \exp(\beta_0 a) \text{ ou encore } \ln(\mu(a, t)) = \ln(\alpha_0) + \beta_0 a.$$

On retrouve alors la structure du modèle de Cairns et al. (2006), utilisé dans la pratique sous des forme plus générales pour étudier la mortalité aux âges élevés. En fait, le modèle de Cairns et al. (2006) s'écrit plutôt à temps discret (i.e. pour chaque année) sur les probabilités annuelles de décès  $q(a, t)$ . Celles-ci donnent la probabilité de décéder dans l'année  $[t, t + 1)$  ( $t$  entier) sachant que l'individu a l'âge (révolu)  $a$  au temps  $t$ . Notons que s'il est né au temps  $\nu$ , son âge au temps  $t$ , s'il est atteint, est  $t - \nu$ , ainsi nous pouvons écrire

$$q(a, t) = \mathbb{P}(\tau < t - \nu + 1 \mid \tau \geq t - \nu, t - \nu \in [a, a + 1)).$$

Nous donnons ici cette définition mais il faut garder à l'esprit que celle-ci se rapporte à un découpage particulier du diagramme de Lexis et que celui-ci peut varier d'un



pays à l'autre pour le calcul des données nationales. Le modèle de Cairns et al. (2006), que l'on peut voir comme un modèle de Gompertz à chaque année, s'écrit alors pour des âge  $a$  et des temps  $t$  entiers (en années) comme<sup>†</sup>

$$\ln(q(a, t)) = \gamma_t + \beta_t a. \quad (1.2)$$

Cette formulation est intéressante car elle donne un cadre formel léger pour étudier la déformation de la mortalité dans le temps. Pour calibrer les coefficients  $\gamma_t$  et  $\beta_t$ , il suffit pour chaque année d'effectuer la régression linéaire de  $\ln(q(a, t))$  en fonction de  $a$ . Les estimations sont représentées sur la Figure 1.5, avec à gauche (resp. à droite) la valeur de  $\gamma_t$  (resp.  $\beta_t$ ) estimée en fonction de l'année  $t$ . La première chose qui peut venir à l'esprit en observant ce graphique concerne les variations d'un des deux coefficients par rapport à l'autre : celles-ci sont inversées. Lorsque  $\gamma$  augmente d'une année sur l'autre,  $\beta$  semble lui diminuer. En particulier, on a le sentiment que beaucoup d'information sur l'un est donnée par l'autre. Quelle pourrait être la contrainte qui lie ces deux processus ? Nous savons que la relation (1.2) ne peut pas être satisfaite pour des âges trop élevés. En effet, le terme de gauche est contraint : la probabilité  $q(a, t)$  vaut au maximum 1, donc son logarithme vaut au plus 0. Notons  $\bar{a}_t$  l'âge pour lequel la probabilité de décéder est 1, c'est à-dire tel que  $q(\bar{a}_t, t) = 1$ . On peut appeler  $\bar{a}_t$  *l'âge limite pour l'année  $t$  dans le modèle de Gompertz*, car pour cette année on ne pourra pas observer d'âge plus grand dans la population si l'on en croit ce modèle. Grâce à l'Équation (1.2) nous avons la relation  $\gamma_t + \beta_t \bar{a}_t = 0$  soit formellement

$$\bar{a}_t = -\frac{\gamma_t}{\beta_t}. \quad (1.3)$$

La dynamique de cet âge limite est représentée sur la Figure 1.6. Notons en première observation que c'est bien l'ordre de grandeur que l'on a en tête pour un âge qui serait maximal. Notons ensuite que celui-ci a augmenté au cours des 60 dernières années, ce qui est conforme à l'intuition. Également, son augmentation est relativement modérée puisque cette quantité a augmenté d'environ 8% en 60 ans. Enfin, observons que sur des plages de temps réduites, cet âge limite est *relativement* constant par morceaux, en particulier sur les plages d'années 1955-1965, 1965-1985 et 1995-2005. Quelle est la conséquence élémentaire de cette stabilité ? Pour une collection d'années  $t$  pour laquelle l'âge limite reste constant à une valeur  $\bar{a}_t = a^*$ , alors d'après l'Équation (1.3) nous obtenons la contrainte  $\gamma_t = -a^* \beta_t$ . Ainsi, pour une période pendant laquelle l'âge limite reste stable, on peut déduire le deuxième paramètres à partir du premier, et dans ce cas les variations sont parfaitement inversées.

---

<sup>†</sup>Le modèle de Cairns et al. (2006) s'écrit plutôt avec la fonction lien  $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ . Nous utilisons ici la fonction lien  $\text{log}$  qui permet une discussion simple sur l'âge limite.

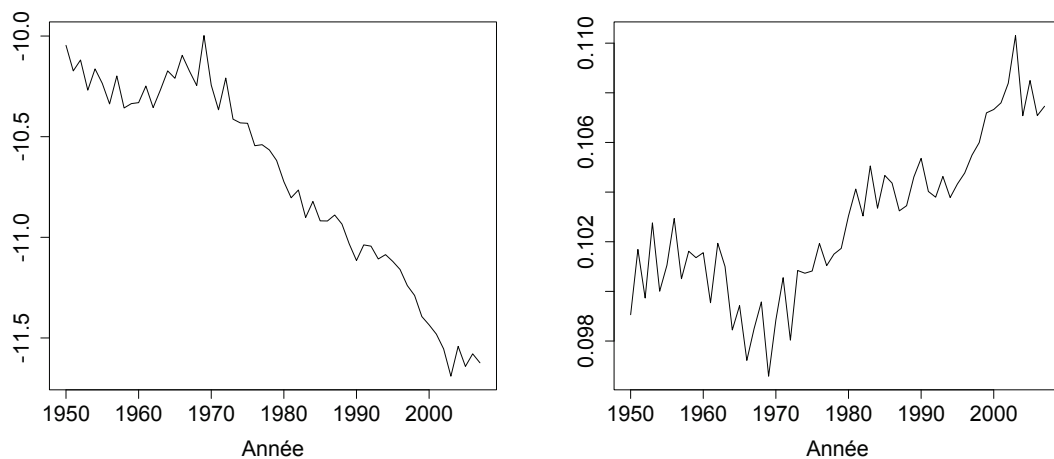
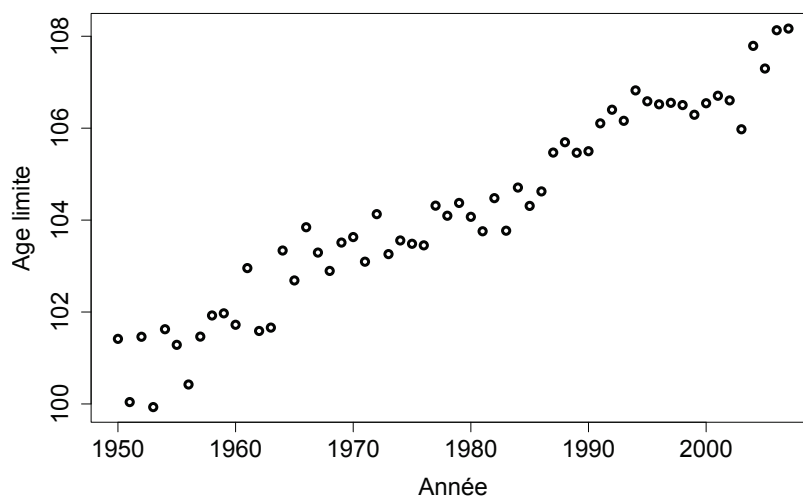
FIGURE 1.5 – Coefficients  $\gamma_t$  et  $\beta_t$  en fonction de l'année  $t$ 

FIGURE 1.6 – Dynamique de l'âge limite calculé dans le modèle de Gompertz

Nous venons de décrire sur un modèle très simple un phénomène bien connu des démographes, que l'on retrouve sous différentes formes : loi de compensation, rectangularisation de la courbe de survie, compression de la distribution des âges au décès, phénomène de rattrapage. Ceux-ci apparaissent si l'on compare une "même" population à différentes dates, comme nous venons de le faire, ou également si l'on compare plusieurs sous-populations au même moment (voir sur ce point la Partie 1.5 et le Chapitre 5 correspondant). Tous ces effets traduisent le fait que si l'on améliore la mortalité à des âges intermédiaires, alors la contrainte de l'existence d'un âge limite commun fait augmenter l'intensité de décès aux grands âges. La loi de compensation de la mortalité et ses effets associés a été étudié à travers différents

points de vue dans la littérature ; sur ce thème, nous renvoyons le lecteur par exemple à Strehler and Mildvan (1960), Fries (1980), Gavrilov and Gavrilova (1991), Wilmoth and Horiuchi (1999) et Strulik and Vollmer (2013).

### 1.1.2 L'importance de la pyramide des âges

Nous avons vu que les données de mortalité nationales nous donnent de nombreuses intuitions sur les phénomènes en jeu. Mais comment ces données nationales sont-elles calculées ? Ce que l'on veut estimer, ce sont des forces de mortalité considérées annuellement, et que l'on suppose constantes par morceaux. En fait, ces morceaux sont des triangles dans le diagramme de Lexis (voir Figure 1.4). Pourquoi une telle hypothèse ? Lorsque l'on regarde un problème à plusieurs dimensions, ici âge  $\times$  temps, l'estimation statistique (non-paramétrique classique) ne peut pas se faire en âge et temps qui varient continûment (voir par exemple Keiding (1990)). C'est pourquoi il faut faire une hypothèse sur la structure des forces de mortalité dans le diagramme de Lexis, en d'autres termes, il faut indiquer quelle sous-population participe à l'estimation de  $\mu(a, t)$ . Sur une telle sous population qui participe à l'estimation de  $\mu(a, t)$ , on va mesurer deux quantités : le nombre de décès dans l'année, noté  $D(a, t)$ , et l'*exposition au risque* notée  $E(a, t)$ . L'exposition au risque correspond à la durée totale vécue dans l'année par les individus de la sous-population, autrement dit la somme des durées de vie dans l'année de chaque individu. Un estimateur de la force de mortalité annuelle est alors (voir par exemple Delwarde and Denuit (2006))

$$\hat{\mu}(a, t) = \frac{D(a, t)}{E(a, t)}.$$

En pratique, les organismes nationaux ne disposent malheureusement pas de ces informations de manière fiable. En particulier, mesurer la durée de vie totale vécue dans l'année reste un problème, et on préfère souvent l'approcher par le nombre d'individus en vie au milieu de l'année. Comme nous l'avons déjà évoqué, le découpage du diagramme de Lexis est différent d'un pays à l'autre. Citons l'exemple du Royaume-Uni, dont l'institution en charge de ces questions s'appelle *Office for National Statistics* (ONS), et pour lequel  $D(a, t)$  compte le nombre de décès dans l'année  $t$  d'individus d'âge  $a$  au moment du décès et l'exposition au risque est approchée par le nombre d'individus d'âge  $a$  au milieu de l'année  $t$  (voir Cairns et al. (2014) et la discussion ci-dessous).

**Les fantômes ne meurent jamais** Le vrai problème avec les fantômes est qu'ils ne meurent pas, et que cela a des conséquences sur le calcul des indicateurs démographiques de référence. C'est sur la base des travaux de Cairns et al. (2014), dont vient aussi le titre de cette partie, que nous allons argumenter en ce sens. Ces

considérations constituent des motivations importantes au recours à des modèles de population. Comme nous l'avons vu, le calcul de l'exposition au risque est crucial pour estimer la force de mortalité annuelle. En pratique néanmoins, de nombreuses erreurs peuvent se glisser, en particulier :

- (i) l'approximation de l'exposition au risque par le nombre d'individus en milieu d'année peut être inadaptée,
- (ii) le calcul de cette population en milieu d'année peut lui-même souffrir d'un certain nombre de biais.

Notons que ces sources d'erreur correspondent aux deux principaux points discutés par Cairns et al. (2014). Concentrons nous sur le premier point, et notons  $P(a, s)$  le nombre d'individus d'âge dans  $[a, a + 1)$  à la date exacte  $s$ . Le point (i) revient à approcher l'exposition au risque  $E(a, t)$ , la durée totale vécue par les individus d'âge (entier)  $a$  au cours de l'année  $[t, t + 1)$ , par  $P(a, t + \frac{1}{2})$ . Or l'exposition au risque et la quantité  $P(a, s)$  s'écrivent respectivement

$$E(a, t) = \int_t^{t+1} P(a, s) ds \text{ et } P(a, s) = \int_a^{a+1} g(u, s) du,$$

où nous rappelons que  $g(u, s)$  désigne le nombre d'individus d'âge (exact)  $u$  au temps (exact)  $s$ . Ainsi, l'approximation s'écrit

$$\int_t^{t+1} \int_a^{a+1} g(u, s) dud s \approx \int_a^{a+1} g(u, t + 1/2) du,$$

et la qualité de cette approximation va dépendre de la forme de la fonction  $s \mapsto \int_a^{a+1} g(u, s) du$  sur l'intervalle de temps  $[t, t + 1)$ . Pour comprendre comment cette quantité évolue au cours du temps, il faut pouvoir capter la *dynamique* de la pyramide des âges au cours du temps, qui va dépendre fondamentalement du jeu des naissances et des morts (ainsi que des flux migratoires). Quant au point (ii), celui-ci est abordé de manière très spécifique par l'ONS et là aussi, la pyramide des âges s'avère cruciale. L'objectif de l'ONS est d'estimer  $P(a, t + \frac{1}{2})$ , le nombre d'individus d'âge  $a$  (entier) à leur dernier anniversaire avant le milieu de l'année. Malheureusement, il n'est pas toujours possible de faire cette observation au milieu de l'année, mais plutôt par exemple dans la première moitié de l'année, disons à un temps  $s \in (t, t + \frac{1}{2})$ . Enfin, pour compliquer encore les choses, on ne dispose pas de l'âge exact des individus au temps  $s$ , mais seulement de leur âge entier (donc au dernier anniversaire). En d'autres termes, on dispose d'une estimation de  $P(a, s)$ , mais aussi de  $P(a - 1, s)$  par exemple. Comment estimer  $P(a, t + \frac{1}{2})$  à partir des deux quantités précédentes ? En fait, la population d'individus d'âge  $a$  (entier) à la date  $t + \frac{1}{2}$  sera constituée d'une part des individus d'âge  $a$  au temps  $s$  qui n'ont pas eu d'anniversaire entre  $s$  et  $t + \frac{1}{2}$  (et qui ne sont pas décédés), et d'autre part des individus d'âge  $a - 1$  au temps  $s$  qui ont eu leur fête d'anniversaire dans  $s$  et

$t + \frac{1}{2}$  (et qui ne sont pas décédés), le tout corrigé des flux migratoires. En dehors de la correction des décès et de la migration, il s'agit de déterminer les proportions de chaque population. Pour cela, l'ONS fait l'hypothèse d'une répartition uniforme des naissances dans l'année ; cette hypothèse forte ne tient pas compte de la saisonnalité des naissances et s'avère surtout problématique pour les années présentant des chocs de taux de naissance *dans* l'année (voir encore Cairns et al. (2014) et la discussion ci-dessous).

Ces enjeux de calcul d'indicateurs ne sont pas anecdotiques : en 2012, l'ONS a publié une révision à la hausse des taux de mortalité à tous les âges, très importante pour les âges supérieurs à 88 ans, pour la cohorte des personnes nées en 1919. Cette sous-estimation de la mortalité semble être due à une sur-estimation des expositions au risque : une surpopulation de fantômes dans la cohorte de 1919 qui ne s'éteint jamais.

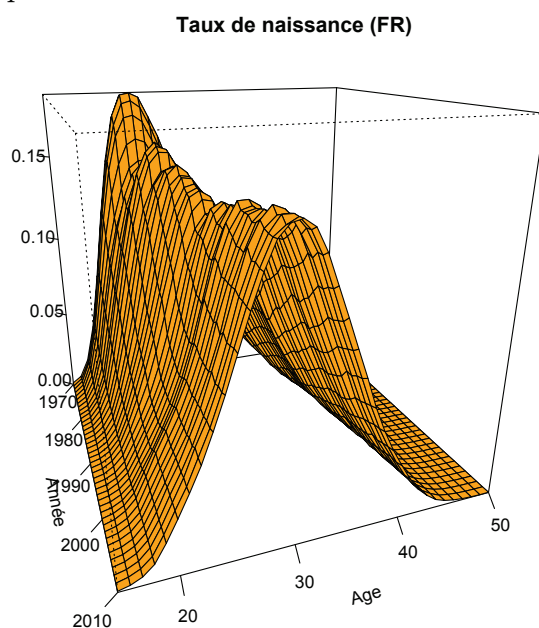


FIGURE 1.7 – Taux de naissance par femme en France

Avec ces deux exemples, il s'avère que comprendre le jeu des naissances dans la dynamique est crucial, en particulier à une échelle plus fine que l'année. De manière analogue à la force de mortalité  $\mu(a, t)$ , aussi appelée *taux de décès*, on peut définir un *taux de naissance*  $b(a, t)$  qui caractérise la fréquence avec laquelle chaque individu donne naissance selon la date et son âge à cette date. En revanche, contrairement au décès, on peut avoir plusieurs événements de naissance, ceux-ci se produisant jusqu'au décès de l'individu. On peut trouver les taux de naissance par âge depuis 1901 pour la France sur le site de l'INSEE, ce que l'on représente sur la Figure 1.7. À noter que ce sont des taux de naissance par âge de la mère, l'âge du père à la naissance étant une donnée moins fiable. Ce choix n'est pas anodin en terme de modélisation puisqu'à nombre de femmes fixé, le nombre de naissances sera insensible au nombre d'hommes dans la population (voir la discussion dans le Chapitre 4). Néanmoins,

l'observation de ces taux de naissance donne de premières intuitions sur sa structure en âge et sa déformation dans le temps. Pour chaque année, le taux de naissance a une forme en cloche centrée en un âge qui a augmenté ces dernières années jusqu'à environ 30 ans en 2010 (voir Figure 1.7).

Avec l'Équation (1.1), nous avons la dynamique (déterministe) d'une cohorte : étant donné le nombre de nouveau nés dans cette cohorte  $g(0, \nu)$ , le nombre d'individus encore vivants à l'âge  $a$  est donné par

$$g(a, \nu + a) = g(0, \nu) \exp \left( - \int_0^a \mu(s, \nu + s) ds \right). \quad (1.4)$$

Pour comprendre la dynamique complète, il reste à traduire le fait que le nombre  $g(0, \nu)$  de personnes nées au temps  $\nu$  est la somme des naissances de tous les parents possibles dans la population au temps  $\nu$ , soit

$$g(0, \nu) = \int_0^\infty g(a, \nu) b(a, \nu) da. \quad (1.5)$$

En rassemblant les Équations (1.4) et (1.5), nous obtenons une dynamique de population (déterministe) connue sous le nom de McKendrick-Von Foerster (voir McKendrick (1926) et Von Foerster (1959)). Nous avons déjà interprété le terme de *transport* de l'Équation (1.1) auquel s'ajoute maintenant le terme (1.5) dit *de renouvellement* qui, comme son nom l'indique, décrit comment la population se renouvelle au cours du temps. Ainsi, si l'on a accès à des indicateurs de naissance à une échelle plus fine que l'année, on pourra apporter des informations supplémentaires et corriger (au moins en partie) le calcul des indicateurs agrégés (voir encore Cairns et al. (2014) sur ce sujet et le cas du Royaume-Uni). On peut trouver dans la *Human Fertility Database* (voir HFD (2014)), la petite sœur de la *Human Mortality Database*, des informations complémentaires sur les taux de naissance pour de nombreux pays. Pour la France par exemple, on trouve le nombre de naissances par mois depuis l'année 1861, ce qui constitue une profondeur d'historique remarquable. Néanmoins, le périmètre géographique de ce calcul a varié (avec ou sans l'Alsace-Lorraine, la Corse,...) et n'est stable que depuis 1946. Nous représentons sur la Figure 1.8 ce nombre de naissances par mois depuis 1861 et jusqu'en 2012. Le nombre de naissances pour le mois  $[\nu, \nu + 1/12)$  correspond, dans le modèle de population que nous avons discuté, à  $\int_\nu^{\nu+1/12} g(0, \nu) d\nu$ . Sur la Figure 1.8, nous distinguons d'abord clairement l'impact des deux guerres mondiales, et également les changements de périmètre géographique, en particulier celui de 1946. De manière commune, ce nombre de naissance par mois présente une forte saisonnalité dans l'année comme nous en avons l'intuition. Nous l'illustrons plus en détail sur les dernières années en Figure 1.9. Enfin, dans la période de périmètre géographique stable (après 1946), on observe des chocs qui peuvent avoir un impact sur les points (i) et (ii) des calculs des indicateurs agrégés.

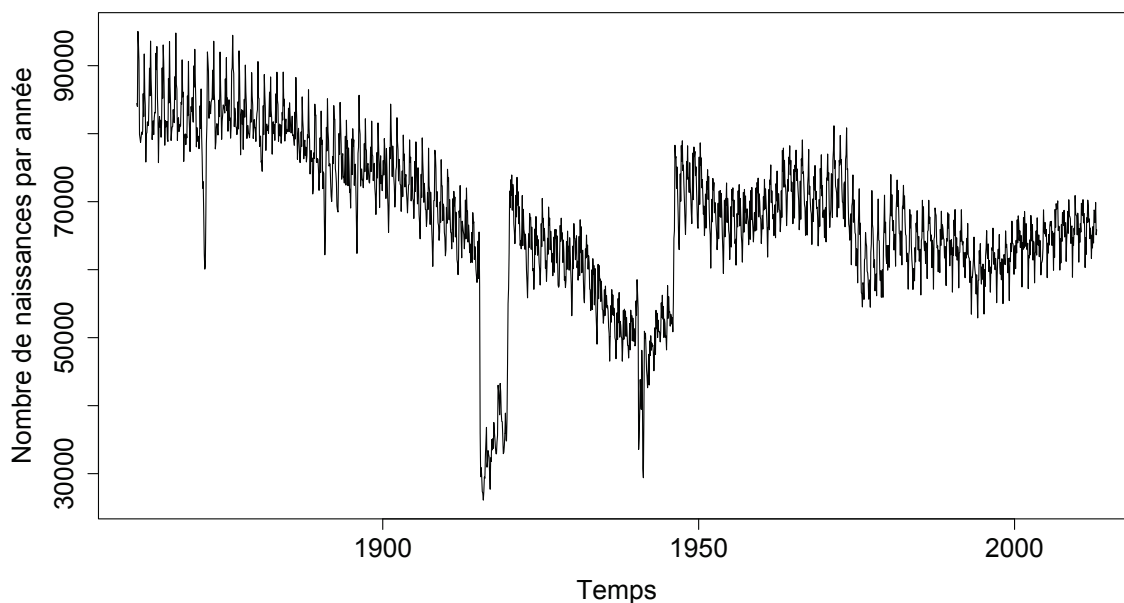


FIGURE 1.8 – Nombre de naissances par mois de 1861 à 2012 pour la France

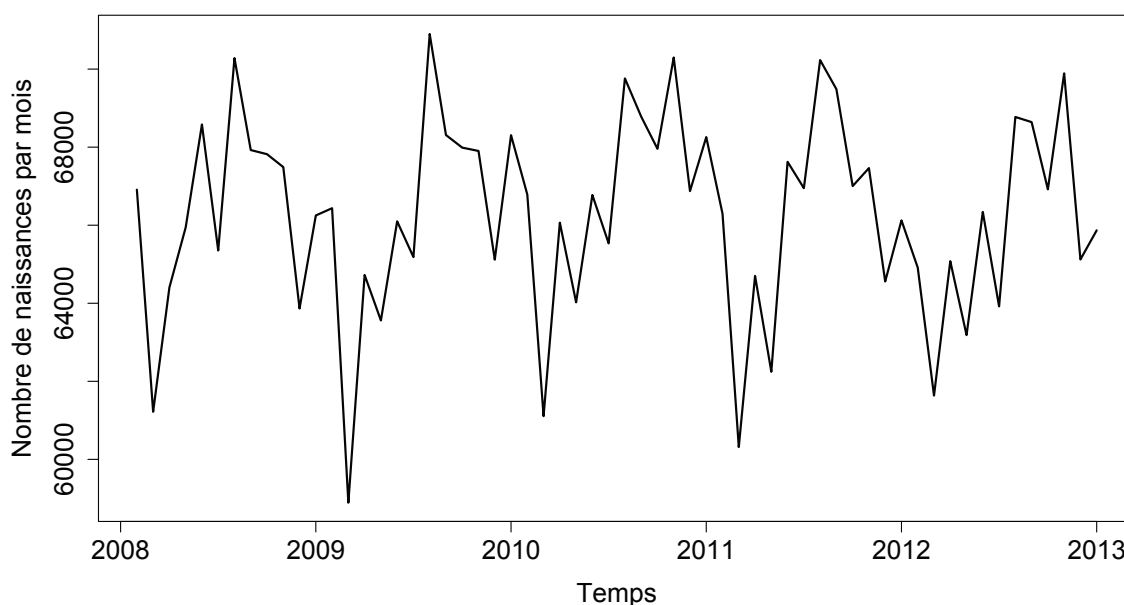


FIGURE 1.9 – Nombre de naissances par mois de 2008 à 2012 pour la France

Nous avons argumenté sur la nécessité de comprendre la dynamique de la pyramide des âges pour le calcul des indicateurs de mortalité nationaux, qui constituent des références en particulier pour les acteurs de l'assurance. Nous souhaitons brièvement décrire une autre information cruciale donnée par cette pyramide des âges :

l'équilibre entre les générations.

**L'équilibre entre les générations** Au-delà des calculs des indicateurs de mortalité de référence, la pyramide des âges est cruciale pour traduire les proportions des différentes classes d'âge. Par exemple, un indicateur crucial pour la question des retraites est celui du ratio de dépendance, qui traduit la proportion d'individus plus âgés que 65 sur les personnes d'âge entre 15 et 65 ans. Pour mesurer le nombre d'individus dans une classe d'âge  $[a_1, a_2]$  au temps  $t$ , on écrit  $\int_{a_1}^{a_2} g(a, t) da$ . Formellement, le ratio de dépendance  $r_t$  peut alors être calculé sur la pyramide des âges déterministe de la façon suivante :

$$r_t = \frac{\int_{65}^{\infty} g(a, t) da}{\int_{15}^{65} g(a, t) da}.$$

Pour avoir une intuition de la forme de la pyramide des âges, on peut se placer dans le cadre de *population stable* largement répandu en démographie. Il s'agit d'oublier le fait que les taux de naissance et mort dépendent du temps ; notons les  $b(a)$  et  $\mu(a)$ . Dans ce cadre, la pyramide des âges se stabilise en temps long, ainsi les proportions d'individus dans chaque classe restent constantes. On dispose même de formules semi-explicites pour ces proportions, que nous donnons dans la suite. Introduisons d'abord un indicateur clef : le nombre moyen d'enfants par individu, noté  $R_0$ , qui s'écrit

$$R_0 = \int_0^{\infty} b(a) e^{-\int_0^a \mu(s) ds} da.$$

Si l'on prend les données nationales pour la France concernant les taux de naissance et mort, on obtient par exemple<sup>†</sup>  $R_0 = 0.964$  (ce qui ne tient pas compte des migrations). Dans ce type de modèle, après un certain temps, la population décroît si  $R_0 < 1$ , est stable si  $R_0 = 1$  et croît si  $R_0 > 1$ . On a également une idée de la vitesse à laquelle ces variations ont lieu. La taille de la population varie exponentiellement vite avec pour paramètre  $\lambda^*$ , le taux de croissance de la population, solution de l'équation

$$\int_0^{\infty} b(a) e^{-\lambda^* a - \int_0^a \mu(s) ds} da = 1.$$

Grâce à ce taux de croissance, on obtient enfin la forme de la pyramide des âges en temps long, résultat connu sous le nom de théorème de Sharpe-Lotka. En temps grand, à une constante multiplicative près,

$$e^{-\lambda^* t} g(a, t) \sim e^{-\lambda^* a - \int_0^a \mu(s) ds}. \quad (1.6)$$

Ce résultat est intéressant car il donne beaucoup d'intuitions sur l'impact des naissances sur l'équilibre de la pyramide des âges. Commençons par remarquer que si la

---

<sup>†</sup>Voir le Chapitre 4.



population est stable, c'est-à-dire que chaque individu a en moyenne exactement un enfant et ainsi  $\lambda^* = 0$ , alors la forme de la pyramide des âges est la même que celle de la fonction de survie  $S(a) = e^{-\int_0^a \mu(s) ds}$ . Si on tirait un individu au hasard dans la population, la probabilité de tomber sur un individu jeune serait naturellement plus importante, et cette probabilité est mesurée par la fonction de survie elle-même. Sinon, il y a un biais dû à l'asymétrie entre les générations successives. Par exemple, si  $\lambda^* > 0$ , cela signifie qu'il y a plus de personnes dans la génération d'après, et donc comparé au cas précédent ( $\lambda^* = 0$ ), la probabilité de tomber sur un jeune en tirant un individu au hasard est encore plus importante. Pour traduire ce biais, on pénalise davantage les âges élevés par le facteur  $-\lambda^* a$  dans l'Équation (1.6). Inversement, si  $\lambda^* < 0$ , on favorise les âges jeunes dans la répartition de la pyramide des âges. Ces proportions, dont le ratio de dépendance est un indicateur, nous permettent en retour de mieux comprendre les modifications des conditions démographiques, par exemple sur la façon dont se propage une modification de la mortalité. Ceci sera l'objet du Chapitre 4 qui se focalise sur l'impact au niveau de la pyramide des âges de suppressions ou réductions de causes de décès.

**Bruit d'environnement et aléa démographique** Revenons à un cadre de taux de naissance et de mort dépendant du temps. Dans ce cadre, nous avons observé les variations des taux de mortalité et de naissance d'une année sur l'autre. L'incertitude qui persiste concernant leur développement futur est appelé *risque systématique* : c'est le risque que le niveau général de la mortalité de la population de référence (par exemple la population nationale) varie de manière inattendue. En terme de modélisation, cela traduit le fait que l'évolution des taux démographiques de la population de référence est aléatoire. Pour les taux de mortalité en particulier, il est souvent fait référence d'une part à des chocs de court terme qui augmentent la mortalité, comme des catastrophes et des pandémies, et qui sont absorbés au cours du temps, et d'autre part à des changements structurels persistants sur le long terme de diminution de la mortalité, dus par exemple à des innovations médicales (voir Figure 1.3). Ces deux types de changements au niveau de la mortalité agrégée sont respectivement caractéristiques des risques de mortalité d'une part et de longévité d'autre part. Le premier risque concerne davantage les bénéficiaires au décès (comme une assurance qui verse une somme au décès de l'assuré), tandis que le second porte plutôt sur les rentes, par exemple les retraites. Pour capter ce risque systématique, ce qui est fait en pratique est de calibrer un modèle de série temporelle sur les observations passées des *composantes temporelles*, ce qui correspond dans nos discussions précédentes aux facteurs  $\gamma_t$  et  $\beta_t$  (voir Figure 1.5). Cette série temporelle est ensuite projetée à l'aide d'un modèle dynamique pour obtenir des réalisations de taux de mortalité dans le futur. Dans la suite, nous désignerons par *bruit d'environnement*

la stochasticité des taux démographiques.

Les variations de populations que nous avons mises en équation sont, étant donnés les taux démographiques, déterministes : il s'agit de la dynamique du *nombre moyen*  $g(a, t)$  d'individus d'âge  $a$  au temps  $t$ . En réalité, même si les taux de mortalité et de naissance étaient anticipés, i.e. formellement si les taux démographiques dans le futur étaient déterministes, il subsisterait un aléa dû à la taille finie de la population dans laquelle les décès et les naissances ont lieu aléatoirement : plusieurs scénarios de réalisation des événements démographiques peuvent se produire. On parle dans ce cas de *risque idiosyncratique* ou *risque démographique*. Celui-ci est d'autant plus important que la population est petite. C'est surtout le cas par exemple pour des portefeuilles d'assurés, mais même pour une population nationale, on peut avoir un nombre réduit de personnes par classe d'âge et donc un risque démographique non-négligeable. Ce risque idiosyncratique, aussi appelé *non-systématique*, peut être réduit du point de vue d'un assureur par exemple en augmentant la taille de la population assurée. Cependant, en pratique on fait souvent face à des populations de taille réduite, comme par exemple certains fonds de pension au Royaume-Uni (voir par exemple Cairns et al. (2008)). Ces effets sont également un enjeu pour certaines institutions qui ont recours à la simulation dans le temps d'échantillons représentatifs. Par exemple, l'INSEE utilise le modèle DESTINIE sur une telle sous-population et note dans les trajectoires obtenues des fluctuations liées à ce risque idiosyncratique (voir Blanchet et al. (2009)). Dans toutes ces applications, on cherche souvent à avoir un ordre de grandeur de ces deux grands risques, c'est-à-dire à savoir si les fluctuations démographiques ont plus d'impact que l'aléa des taux eux-mêmes, que l'on a appelé *risque d'environnement* car commun à tous les individus. Cette question est discutée au-delà du périmètre de la science actuarielle. En écologie par exemple, la question est de savoir si l'extinction d'une espèce est due à un scénario défavorable du bruit d'environnement (i.e. des taux démographiques), ou bien dû à une réalisation adverse des événements démographiques eux-mêmes dans une population de taille finie (voir par exemple Engen et al. (1998) sur ce thème). Ces considérations suggèrent qu'il est nécessaire d'utiliser un cadre stochastique pour la dynamique de population. La construction d'un tel cadre stochastique dans une optique *trajectorielle* sera présentée dans la Partie 1.2 et détaillée dans le Chapitre 2. Celui-ci permet de décrire l'état de la population et son évolution, en présence de l'aléa démographique et du bruit d'environnement. En particulier, nous donnerons un exemple de décomposition entre ces deux sources de bruit.

Ce cadre stochastique a l'avantage d'être lié aux équations déterministes à deux niveaux, que nous discuterons en particulier dans la Partie 1.2 et le Chapitre 2 : la dynamique déterministe représente à la fois le scénario moyen, et également la limite en grande population au sens de la convergence des processus. Ce deuxième

aspect est intéressant car il donne un cadre asymptotique pour étudier la *dynamique macroscopique* de la population, afin d'en dégager les grandes tendances. Un cadre particulier de limite macroscopique, lorsque les individus ont des trajectoires de vie accélérées, sera étudié dans le Chapitre 6.

**La pyramide des âges des processus de Hawkes** La philosophie qui ressort de toutes les questions précédentes liées à la longévité est que la pyramide des âges donne de l'information sur les phénomènes en jeu, de manière générale sur toutes les quantités agrégées que l'on calcule sur celle-ci. Nous allons décliner ce point de vue de manière théorique dans une contribution à l'étude des processus de Hawkes que nous décrirons dans la suite. Soit  $(N_t)$  un *processus de comptage*, c'est-à-dire un processus qui part de  $N_0 = 0$  et dont la valeur  $N_t$  compte le nombre d'événements qui ont eu lieu jusqu'au temps  $t$ , de un en un. Ces événements se produisent à des temps (aléatoires) que l'on note  $T_n$ , pour  $n \geq 1$ . On peut voir  $N_t$  comme la taille d'une population qui croît de un en un au cours du temps, et interpréter  $T_n$  comme le temps de naissance de l'individu arrivé en position  $n$ . À quelle fréquence se produisent ces naissances ? Considérons le modèle dans lequel un individu d'âge  $a$  dans la population donne naissance au taux  $\phi(a)$ . Nous conserverons cette notation du taux de naissance pour différencier cette étude par rapport au cadre démographique. Alors, si les individus donnent naissance indépendamment, la fréquence totale d'apparition d'un nouvel individu est la somme des taux de naissance de tous les individus dans la population. Or si un individu est né au temps  $T_n$ , son âge au temps  $t$  est  $t - T_n$ . Ainsi le taux de naissance global au temps  $t$  est donné par  $\sum_{T_n < t} \phi(t - T_n)$ . Pour *lancer* la dynamique, on peut ajouter une constante  $\mu$  à cette fréquence globale, que l'on peut voir comme un taux d'immigration, si bien que l'on obtient l'intensité  $\lambda_t$  du processus  $N_t$  que l'on appelle processus de Hawkes, avec

$$\lambda_t = \mu + \sum_{T_n < t} \phi(t - T_n).$$

Ce processus a été introduit par Hawkes (1971), d'abord vu comme processus de comptage, puis interprété en terme de population (dans un cadre stationnaire) par Hawkes and Oakes (1974).

Pour étudier le processus  $(N_t)$ , la notion d'information est cruciale. Par exemple, si le taux de naissance ne dépend pas de l'âge, c'est-à-dire si  $\phi(a) = \bar{\phi}$ , alors on a un processus très simple de naissance à taux constants dont l'intensité vaut  $\lambda_t = \mu + \bar{\phi}N_{t-}$ . Dans ce cas, il suffit d'observer le présent de  $(N_t)$  pour en déduire son futur immédiat et ainsi le processus  $(N_t)$  possède la propriété de Markov. En revanche, si  $\phi$  dépend de l'âge, le processus de Hawkes n'est plus un processus de Markov. On connaît pourtant un cas particulier (voir par exemple Oakes (1975)) qui est que lorsque  $\phi$  est exponentiel, disons  $\phi(a) = \alpha e^{\beta a}$ , c'est l'intensité  $(\lambda_t)$  qui

est Markovienne. On en déduit que si le taux de naissance s'écrit  $\phi(a) = \bar{\phi} + \alpha e^{\beta a}$ , le couple  $(N_t, \lambda_t)$  est Markovien. Lorsque  $\phi$  dépend de l'âge de manière plus complexe, ces considérations ne sont plus valables. Dans la description de la dynamique des naissances que nous avons abordée, il était clair que connaître l'état de la population à un instant donné nous permettait de construire la suite de la dynamique. En d'autres termes, connaître la *pyramide des âges* sous-jacente permet de comprendre la dynamique du processus  $(N_t)$ . En fait, quel que soit le taux de naissance, la pyramide des âges est toujours Markovienne. L'exploiter nous permettra d'affiner l'étude des processus de Hawkes pour des taux de naissance qui généralisent le cas exponentiel. Ce sera l'objet de la Partie 1.3 de ce chapitre introductif et, sous une forme plus détaillée, du Chapitre 3.

### 1.1.3 L'impact de l'hétérogénéité

Revenons à des considérations liées à la longévité. Comme nous l'avons vu, les niveaux de mortalité agrégés, distingués selon le sexe, constituent des indicateurs de référence pour les politiques publiques ou la gestion des risques dans le secteur de l'assurance. Pour comprendre l'évolution des taux de mortalité agrégés et les fluctuations démographiques, il est nécessaire d'étudier la population à un niveau plus fin, *microscopique*. À cette échelle, on distingue les différents individus qui constituent la population, avec leur âge exact et un certain nombre de caractéristiques propres. Ces variables définissent la façon dont chaque individu se comporte dans la population. En particulier, les niveaux de mortalité diffèrent selon le pays, le sexe, mais aussi un certain nombre de caractéristiques individuelles, comme le niveau de revenu, le statut marital ou le niveau de santé par exemple. Également, les taux de naissance peuvent être très différents selon ces différentes catégories, ce qui rend la dynamique de la population complexe.

**L'hétérogénéité de la population** Nous restons dans le cadre d'une dynamique de population *en moyenne*, autrement dit déterministe étant donnés les taux de naissance et de mort. La construction du processus de population aléatoire sera abordée dans la Partie 1.2 et sous une forme plus détaillée dans le Chapitre 2. Concentrons nous d'abord sur la dynamique d'une cohorte. On se donne maintenant  $\mathcal{X}$  l'ensemble de toutes les caractéristiques individuelles possibles. Nous avons noté  $g(0, \nu)$  le nombre d'individus nés dans la cohorte de date de naissance  $\nu$ . Celle-ci est en réalité composée de différentes sous-populations de caractéristique  $x \in \mathcal{X}$ , notée  $g(x, 0, \nu)$ , de sorte que

$$g(0, \nu) = \sum_{x \in \mathcal{X}} g(x, 0, \nu).$$

Le taux de mortalité d'un individu va maintenant dépendre de sa caractéristique : si un individu a la caractéristique  $x$ , et l'âge  $a$  au temps  $t$ , sa force de mortalité est  $\mu(x, a, t)$ . Le taux de mortalité agrégé de chaque cohorte dépend des taux de mortalité de chaque classe et de leurs proportions respectives, de sorte que

$$\mu(a, \nu + a) = \frac{\sum_{x \in \mathcal{X}} g(x, a, \nu + a) \mu(x, a, \nu + a)}{\sum_{x \in \mathcal{X}} g(x, a, \nu + a)}.$$

La proportion de survivants à l'âge  $a$  dans chaque groupe étant donnée par

$$g(x, a, \nu + a) = g(x, 0, \nu) e^{-\int_0^a \mu(x, s, \nu + s) ds}, \quad (1.7)$$

et on obtient finalement :

$$\mu(a, \nu + a) = \frac{\sum_{x \in \mathcal{X}} g(x, 0, \nu) e^{-\int_0^a \mu(x, s, \nu + s) ds} \mu(x, a, \nu + a)}{\sum_{x \in \mathcal{X}} g(x, 0, \nu) e^{-\int_0^a \mu(x, s, \nu + s) ds}}. \quad (1.8)$$

À ce stade, nous pouvons déjà discuter plusieurs effets. Le premier effet est celui du *ratissage* : le taux de mortalité agrégé a tendance à rattraper celui du groupe qui a le taux le plus faible, étant donné que ceux-ci vivent plus longtemps et sont de plus en plus représentés quand l'âge augmente. Ceci est effectivement observé lorsque les caractéristiques sont stables au cours de la vie, ce qui est le cas dans le modèle que nous venons d'écrire. Si celles-ci peuvent évoluer, la structure en âge du taux de mortalité agrégée doit s'analyser de manière plus fine, ce qui sera discuté dans les Parties 1.5 et 1.6 et détaillé dans les Chapitres 5 et 6 correspondants. Un deuxième effet que l'on peut discuter est le paradoxe de l'hypothèse de loi de Gompertz : supposons que chaque classe satisfait cette loi sur la génération, soit  $\mu(x, a, \nu + a) = \alpha_\nu^x \exp(\beta_\nu^x a)$ , alors le taux de mortalité agrégé donné dans l'Équation (1.8) ne satisfait pas une loi de Gompertz. Au-delà, il faut mentionner ici que dans le domaine de l'assurance, le taux de mortalité national  $\mu(a, t)$  sert souvent de référence pour comprendre le taux de mortalité propre à un groupe d'assurés, tant par sa structure en âge que son évolution en temps. Cet indicateur national sert de plus de base aux produits financiers de transfert du risque de longévité dits standardisés. Une façon de lier la mortalité d'un groupe d'assurés et celle de la population nationale est de calibrer ce que l'on appelle un *modèle relationnel*, qui exprime la première mortalité comme une fonction de la seconde, par exemple linéaire.

En résumé, le taux de mortalité agrégé dépend à travers l'Équation (1.8) des taux de mortalité par groupe d'individus  $\mu(x, a, \nu + a)$  d'une part, et de la répartition initiale de la cohorte composée des  $g(x, 0, \nu)$  d'autre part. Dans la suite, nous présentons comment ces répartitions initiales peuvent être modélisées, en suivant en parallèle un certain nombre d'intuitions sur l'effet cohorte.

**Point de vue sur l'effet cohorte** Dans un modèle de population hétérogène, les individus possèdent des taux démographiques spécifiques à leur caractéristique. En particulier, nous pouvons définir un taux de naissance  $b(x', a, t)$  pour les individus de caractéristique  $x'$  et d'âge  $a$  au temps  $t$ . Nous souhaitons maintenant traduire le fait que les individus de caractéristique  $x$  nés au temps  $t$ , dont le nombre initial est  $g(x, 0, t)$ , sont issus des parents de la population au temps  $t$ , avec leur propre âge et caractéristique. Un parent d'âge  $a$  et de caractéristique  $x'$  au temps  $t$  (ce que l'on note  $(x', a, t)$ ) donne naissance au taux  $b(x', a, t)$  à des individus dont on doit déterminer le type. On pourrait traduire une hérédité parfaite, c'est-à-dire faire conserver la caractéristique aux enfants, ou proposer une distribution équi-répartie. De manière générale, on peut proposer une distribution générique  $k_b(x', a, t, \cdot)$ , où pour tout  $x \in \mathcal{X}$ ,  $k_b(x', a, t, x)$  donne la probabilité que, si le parent  $(x', a, t)$  donne naissance, le nouvel individu ait la caractéristique  $x$ . C'est ce que l'on appelle un *noyau de mutation* en référence aux modèles motivés par des questions écologiques. Ainsi le nombre d'individus de caractéristique  $x$  dans la cohorte de date de naissance  $\nu$  est obtenu en sommant sur tous les parents possibles, soit

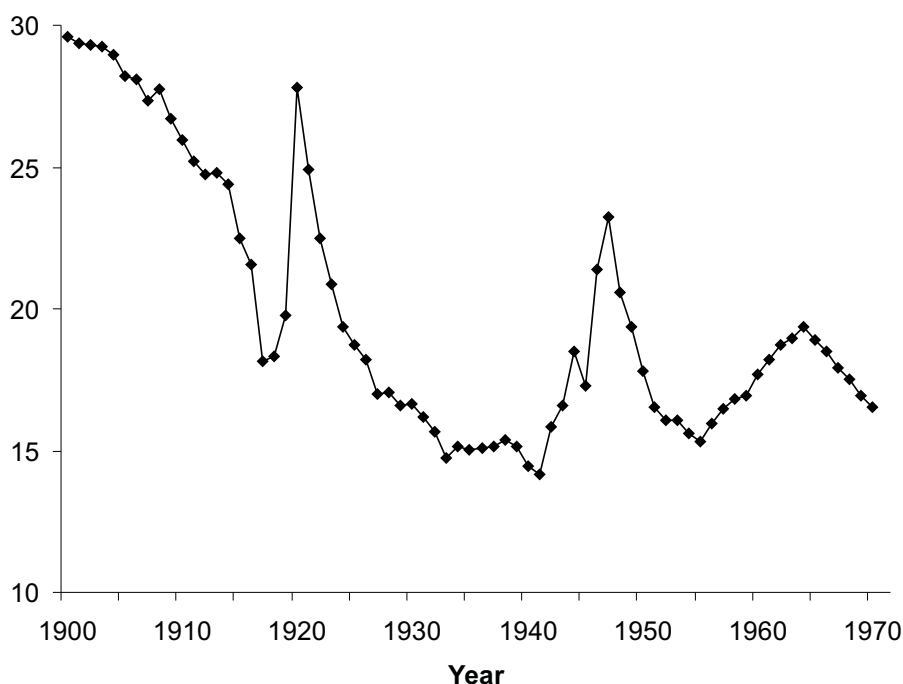
$$g(x, 0, \nu) = \sum_{x' \in \mathcal{X}} \int_0^\infty g(x', a, \nu) b(x', a, \nu) k_b(x', a, \nu, x) da. \quad (1.9)$$

Avec le terme (1.7), analogue au terme de transport, et l'équation de renouvellement (1.9), nous venons de donner le comportement déterministe d'une dynamique de naissance et mort hétérogène.

Discutons maintenant une forme d'effet cohorte que l'on trouve dans la littérature. Rappelons qu'une cohorte, ou génération, désigne un ensemble de personnes nées à la même date, et par extension nées dans la même période de temps. On parle d'*effet cohorte* lorsque l'une de ces générations présente des caractéristiques de longévité très différentes des générations précédentes et suivantes. Ces caractéristiques de longévité particulières peuvent être mesurées de diverses manières, par exemple sur l'espérance de vie, les forces de mortalité ou les probabilités annuelles de décès. Néanmoins, pour pouvoir comparer les générations entre elles, il est préférable de mesurer le *taux d'amélioration* de la mortalité, qui n'est rien d'autre que la différence relative entre la mortalité de la génération à une date donnée et celle à la date précédente. Des taux d'amélioration particulièrement élevés ont été mesurés pour ce que l'on appelle la *Golden Cohort*, toujours au Royaume-Uni, qui est constituée des personnes nées autour de 1930. Plusieurs explications démographiques ont été avancées, concernant majoritairement des effets bénéfiques pour cette *Golden Cohort* et des effets limitants pour les générations précédentes et suivantes (voir Willets (2004)). Également, Willets (2004) a remarqué que les taux de naissance dans cette période autour de 1930 ont été particulièrement bas (voir Figure 1.10), et avance qu'une conséquence possible de brusques changements dans les taux de naissance

est que la composition de la génération soit modifiée, en présence par exemple de tendances de fertilité très différentes d'une classe à une autre. Cette sensibilité aux taux de naissance est visible sur les Équations (1.8) et (1.9), qui en se combinant expliquent l'impact de changements dans les taux de naissance par caractéristique sur la mortalité agrégée. Ceci fournit un cadre expérimental pour mettre en oeuvre des scénarios de mortalité, dont un exemple sera donné dans la Partie 1.2 et le Chapitre 2. En pratique, il est malheureusement difficile d'avoir accès à des taux de naissance fiables par caractéristiques, et également de pouvoir calibrer les noyaux de mutation : il faudrait pour cela disposer de données longitudinales extrêmement longues, au-delà par exemple de ce qui est présent dans l'*Echantillon Démographique Permanent* qui sera étudié dans le Chapitre 5 et qui constitue déjà une base de données remarquable.

FIGURE 1.10 – Taux de naissance pour 1000 individus, pour Angleterre et pays de Galles, repris de Willets (2004)



Data source: www.mortality.org

Figure 6. Crude birth rate per 1,000 population, England and Wales, 1900 to 1970

**Les trajectoires de vie** Ce que l'on observe naturellement, par exemple sur *l'Echantillon Démographique Permanent* de l'INSEE, est que les individus changent de caractéristique de naissance relative à leur période immédiate de phénomène. Il faut définir un troisième type de taux démographique, qui s'ajoute à ceux de naissance et mort. On pourra l'appeler taux d'évolution, ou taux de *suav* en référence à la littérature économique, qui est différent des changements de birth rates. On peut dire que le taux de *suav* est différent des périodes où les birth rates sont très différents. For instance, if trends in fertility vary by socio-economic class, the class mix of a population will change.

note  $e(x, a, t)$  le taux d'évolution, qui caractérise la fréquence à laquelle un individu  $(x, a, t)$  va changer de caractéristiques, et  $k_e(x, a, t, y)$  le noyau qui va donner la probabilité que celui-ci passe de la caractéristique  $x$  à la caractéristique  $y$ . Rappelons ici que si le taux  $e$  ne dépend ni de l'âge ni du temps, cela revient pour l'individu à attendre dans la classe  $x$  un temps aléatoire exponentiel de paramètre  $e(x)$ , puis de changer pour une classe  $y$  avec la règle donnée par le noyau  $k_e$ , et ainsi de suite jusqu'au décès. Dans le cas où  $e$  dépend de l'âge et/ou du temps, le principe est le même, mais le temps d'attente n'est plus exponentiel. Cette dynamique à l'échelle de l'individu représente ce que l'on appellera sa *trajectoire de vie*. Pour étudier la dynamique d'une population en présence de trajectoires de vie, nous développerons deux points de vue complémentaires. Le premier point de vue est statistique : comment estimer des forces de mortalité par caractéristique lorsque celle-ci change au cours de la vie, et quand de plus les observations sont ponctuelles ? Le deuxième point de vue est probabiliste : existe-t-il un modèle simplifié pour comprendre la dynamique hétérogène lorsque les caractéristiques changent très fréquemment ? Nous concluons ces motivations en introduisant ces deux points de vue, puis nous résumerons les différentes contributions de cette thèse.

**Les données démographiques issues du recensement** Nous pouvons prendre la mesure de l'hétérogénéité de la population nationale sur l'*Echantillon Démographique Permanent* de l'INSEE. Cette base de données contient des informations sur la situation de 992711 personnes observées à partir de 1968 et nées les 1, 2, 3 ou 4 octobre à partir de l'année 1862. Ces individus ont potentiellement été recensés aux dates 1968, 1975, 1982, 1990 et 1999, et pour chaque recensement, un ensemble de caractéristiques a été relevé. Un tel échantillon permet d'étudier le lien entre le niveau de mortalité et les caractéristiques individuelles, mais également la manière dont ces caractéristiques évoluent dans le temps. Sur la Figure 1.11, nous représentons la répartition des femmes par génération et par statut matrimonial, en 1990 d'une part et en 1999 d'autre part. Pour chaque recensement, si on fixe une année de naissance en abscisse, par exemple 1930, on observe la répartition par classe des femmes de cette génération : célibataires, mariées, veuves, divorcées. Entre les deux recensements, on observe un changement de composition : par exemple, la proportion de veuves augmente pour les générations anciennes, celle des divorcées augmente pour les générations intermédiaires, et celle des célibataires diminue pour les générations récentes. Dans l'ensemble, les changements de composition sont dus à plusieurs effets. Le premier est que les individus des différentes classes décèdent au cours du temps. Comme ceux-ci peuvent avoir des taux de décès différents, cela a pour effet de modifier la composition de chaque génération dynamiquement. Un deuxième effet, qui lui aussi modifie la composition de la population, est celui des



changements de caractéristiques au cours de la vie : par exemple, une femme mariée en 1990 peut être veuve en 1999. Dans ce cadre, le seul moyen de capter des taux de décès propres à chaque caractéristique, et également les intensités de changement de caractéristique, est de calibrer un *processus multi-états*, c'est-à-dire un modèle dans lequel un individu typique passe dans plusieurs états intermédiaires jusqu'au décès, l'état terminal. La difficulté avec ces données issues du recensement est que dans ce cas précis l'on n'observe pas ce qui s'est passé entre 1990 et 1999. Ceci est caractéristique de ce que l'on appelle les *données de panel*. En particulier, si une personne est passée de marié à veuf, on ne connaît pas le nombre de transitions qui ont eu lieu : on pourrait avoir en théorie marié  $\rightarrow$  divorcé  $\rightarrow$  marié  $\rightarrow$  veuf. Et même si l'on suppose qu'une seule transition a eu lieu sur cet intervalle de 9 ans, ce qui peut paraître raisonnable, on ne connaît pas la date de transition. On parle dans ce cas de *censure par intervalle*. La question de l'estimation statistique de modèles multi-états dans ce contexte sera abordée dans la Partie 1.5 et le Chapitre 5.

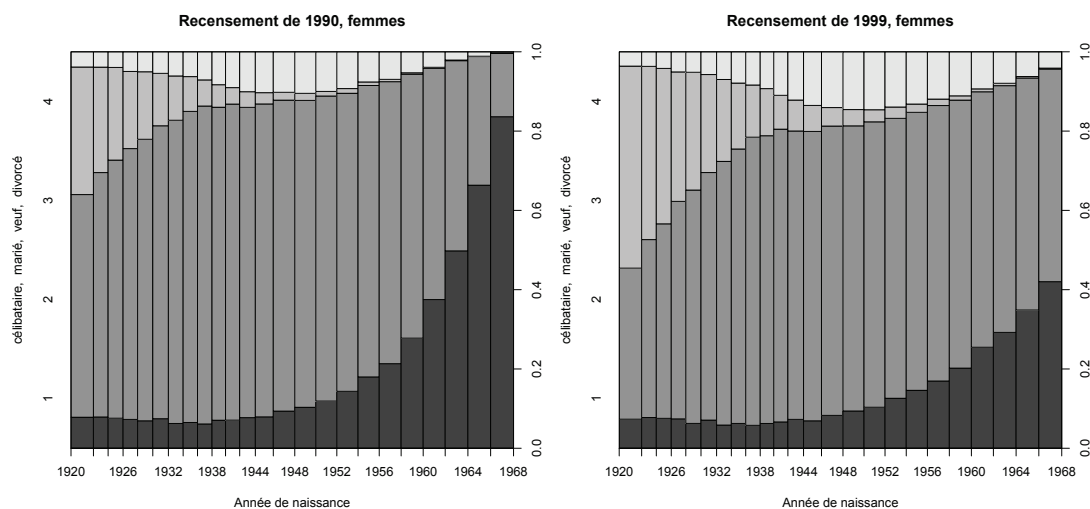


FIGURE 1.11 – Structure des recensements par statut matrimonial et par génération pour les femmes

**L'accélération des trajectoires de vie** L'estimation statistique peut fournir une analyse des populations hétérogènes, dans laquelle les individus ont des taux démographiques relatifs à leur propres caractéristiques, et quand parallèlement ces caractéristiques peuvent changer au cours de la vie. Néanmoins, le mécanisme d'observation par le recensement occasionne de la censure par intervalle, ce qui restreint l'information disponible. En particulier, si les caractéristiques changent fréquemment au cours de la vie, on ne pourra pas comprendre la dynamique en ne l'observant que ponctuellement aux dates d'inspection. Le modèle de dynamique des populations, qui sera formalisé dans la Partie 1.2 et le Chapitre 2, et les techniques de renormalisation sur un tel modèle fournissent des outils intéressants pour comprendre

les dynamiques hétérogènes dans ce cadre. Nous avons déjà mentionné que, dans la limite d'une population infinie, la dynamique aléatoire pouvait être approchée par les équations déterministes. Ces techniques de théorèmes limites vont nous permettre d'étudier un autre problème : que se passe-t-il lorsque les individus changent de caractéristiques de manière fréquente au cours de leur vie ? La renormalisation que l'on va considérer repose d'une part sur la renormalisation standard de *grande population*, qui est naturelle si l'on pense à l'ordre de grandeur des populations nationales, et d'autre part une *accélération* des changements de caractéristiques, ce qui revient à découpler cette échelle de temps de celle des temps de naissance et de mort dans la dynamique. Dans ce cadre, le comportement limite que nous obtenons est décrit par l'équation (déterministe) de McKendrick-Von Foerster (voir McKendrick (1926) et Von Foerster (1959)) dans laquelle l'âge seulement intervient, et les taux de naissance et de mort sont moyennés sur la structure stable en terme de caractéristiques. Ces résultats donnent des perspectives intéressantes en terme de compréhension de la mortalité agrégée : ceux-ci nous indiquent que, sous certaines hypothèses, il est possible de regrouper des individus, non pas parce qu'ils partagent les mêmes caractéristiques, mais parce qu'ils en *changent* de manière similaire. Ces développements seront abordés dans la Partie 1.6 et détaillés dans le Chapitre 6.

### 1.1.4 Résumé des contributions

Nous résumons ici les contributions issues de cette thèse, qui seront introduites dans les Sections 1.2 à 1.6 et détaillées dans leur intégralité dans les Chapitres 2 à 6.

- Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Bridging the gap from microsimulation practice to population models: a survey. *Work in progress*
- Boumezoued, A. 2015. Population viewpoint on Hawkes processes. HAL preprint Id: hal-01149752, To appear in *Advances in Applied Probability* 48.2 (June 2016)
- Arnold, S., A. Boumezoued, H. Labit Hardy, N. El Karoui. 2015. Cause-of-death mortality: What can be learned from population dynamics? HAL preprint Id: hal-01157900
- Boumezoued, A., N. El Karoui, S. Loisel 2015, Measuring mortality heterogeneity dynamics with interval-censored data. *Working paper*
- Boumezoued, A. 2015, Macroscopic behavior of heterogenous populations with fast random life histories. *Working paper*

**La dynamique des populations** Dans le Chapitre 2, basé sur Bensusan et al. (2010–2015), nous donnons le cadre théorique de modélisation de la dynamique d’une population, dans laquelle les individus ont un âge, des caractéristiques et vivent dans un environnement aléatoire. Nous proposons de modéliser les décès et les naissances, et également les changements de caractéristiques qui peuvent intervenir au cours de la vie. La construction du processus de population qui évolue de manière aléatoire et en temps continu se fait à l’aide de la représentation par *thinning* des processus à saut à partir de mesures ponctuelles de Poisson. L’algorithme de simulation associé est décrit, et illustré sur l’exemple d’un *effet cohorte*. Nous décrivons aussi le lien entre la dynamique aléatoire et les équations déterministes, et démontrons l’existence d’une densité pour l’équation déterministe dans notre cadre étendu aux changements de caractéristiques. Enfin, nous illustrons la décomposition entre bruit d’environnement et aléa démographique sur l’exemple d’une cohorte qui décède au cours du temps.

**La pyramide des âges des processus de Hawkes** Le Chapitre 3, tiré de Boumezoued (2015b), utilise ce modèle de population et fait le lien entre avec les processus de Hawkes avec immigrants généraux. Nous étudions la dynamique de la pyramide des âges qui dirige de tels processus. Cette approche permet de réconcilier deux définitions phares des processus de Hawkes : sous la forme d’un processus de saut avec intensité d’une part, et d’un processus de branchement d’autre part. Celle-ci nous permet d’établir un certain nombre de résultats sur la distribution d’une classe de processus de Hawkes avec immigrants généraux qui généralisent dans un cadre non-stationnaire le cas du taux de naissance exponentiel.

**Les causes de décès dans la dynamique des populations** Dans le Chapitre 4, qui reprend Arnold et al. (2015), nous analysons l’impact de changements dans la mortalité par causes de décès sur la dynamique de la pyramide des âges, et en particulier sur le ratio de dépendance qui est un indicateur crucial du vieillissement de la population. En incluant le jeu des naissances dans la dynamique, ce travail complète la littérature existante sur les causes de décès qui se focalise traditionnellement sur des indicateurs de mortalité. À ce titre, notre étude numérique, basée sur des données par causes de décès de l’Organisation Mondiale de la Santé (OMS) pour la France, montre que des populations qui présentent la même espérance de vie peuvent avoir des pyramide des âges différentes selon la cause de décès qui est réduite. Dans ce cadre, nous présentons également des analyses de sensibilité à la fertilité et aux flux de population.

**L’estimation de modèles multi-états en présence de censure par intervalle** Le Chapitre 5, basé sur Boumezoued et al. (2015), étudie l’hétérogénéité de la mortalité dans les données de l’Échantillon Démographique Permanent de l’INSEE. Il

aborde la question de l'estimation statistique de modèles multi-états en présence de censure par intervalle et de retour dans les états intermédiaires. La censure par intervalle, caractéristique des données longitudinales issues du recensement, traduit le fait que les temps de transition sont inconnus, et que la seule information dont on dispose est l'intervalle qui le contient. Les outils existants dans la littérature ne permettent pas de traiter directement notre cadre. Pour ce faire, nous proposons une méthode d'estimation paramétrique par maximum de vraisemblance qui permet de mesurer des taux de décès par caractéristiques quand celles-ci peuvent changer au cours de la vie. En appliquant une telle méthode, nous pouvons mesurer des forces de mortalité spécifiques à des catégories socio-professionnelles ou au statut matrimonial, en traitant à chaque fois des générations réelles. Ce travail nous permet d'observer plusieurs effets, incluant la loi de compensation de la mortalité aux grands âges entre les différentes sous-populations, mais aussi un certain nombre d'intuitions sur la structure en âge de la mortalité agrégée, qui dépend de la dynamique hétérogène sous-jacente. Enfin, nous étudions la distribution de flux liés à des produits de rente ou d'assurance décès stylisés, et leur modification si les changements de caractéristiques ne sont pas pris en compte.

**La limite en grande population de trajectoires de vie accélérées** Le Chapitre 6, tiré de Boumezoued (2015a), reprend le modèle général du Chapitre 2 dans lequel les individus peuvent donner naissance, changer de caractéristiques et décéder. La contribution de cette partie est l'étude de l'asymptotique de la dynamique d'une population hétérogène à la fois en grande population et avec des changements de caractéristiques accélérés, ces deux hypothèses étant motivées par la dynamique démographique de populations à l'échelle d'un pays. Dans ce cadre, le comportement limite que nous obtenons est décrit par l'équation (déterministe) de McKendrick-Von Foerster dans laquelle l'âge seulement intervient, et les taux de naissance et de mort sont moyennés sur la structure stable en terme de caractéristiques. Ceci fournit de manière explicite un modèle approché reposant sur des taux de naissance et mort agrégés, qui illustrent l'impact de la structure des changements de caractéristiques sur la dynamique macroscopique.

## 1.2 Modélisation dynamique d'une population

### 1.2.1 Le point de vue microscopique dans la littérature

Les modèles de systèmes structurés à l'échelle de leur entité élémentaire sont présents dans de nombreux domaines. Nous souhaitons en donner ici une brève description, et nous renvoyons au Chapitre 2 pour plus de détails. On trouve d'abord de nombreuses contributions dans le champ des modèles d'agents ou de micro-simulation, orientés vers des applications à l'économie. Un travail fondateur est celui d'Orcutt (1957), et depuis deux branches se sont majoritairement développées :

(i) Les modèles d'agents d'une part, voir par exemple Tesfatsion (2002) pour un état de l'art. L'idée est de définir une règle de comportement à l'échelle individuelle, puis de diffuser le système économique sans intervention de la part du modélisateur. L'objectif est d'expliquer les régularités ou les anomalies observées à l'échelle agrégée par des comportements individuels d'agents en interaction.

(ii) Sur des aspects davantage liés à la démographie, toujours tournés vers des considérations économiques comme le système des retraites. La plupart des pays développés disposent de leur propre modèle de micro-simulation (voir par exemple Morand et al. (2010)), dont l'objectif est d'évaluer l'impact de changements démographiques ou de politiques publiques sur des quantités économiques. En France par exemple, le modèle actuellement utilisé est DESTINIE 2 (voir Blanchet et al. (2009)). Il est basé sur un échantillon représentatif de la population française, que nous analyserons en détail dans la Partie 1.5 et le Chapitre 5.

Parallèlement, un cadre mathématique élaboré a été développé dans le domaine des modèles dits *individus-centrés*, qui sont des extensions des processus de naissance et mort. La plupart des contributions dans ce sens sont motivées par des questions liées à la théorie de l'évolution et aux dynamiques écologiques. De tels modèles individus-centrés mettent l'accent sur les traits, c'est-à-dire les caractéristiques individuelles, et également les interactions entre les individus. Ceux-ci ont été étudiés en particulier par Fournier and Méléard (2004) et Champagnat et al. (2006) dans le cas de populations structurées par caractéristiques, et également par Ferriere and Tran (2009) pour des populations structurées en âge. Ces modèles ont aussi été étudiés dans la thèse de Bensusan (2010) en vue d'applications aux populations humaines, en incluant en particulier des changements de caractéristiques au cours de la vie et un environnement aléatoire.

Nous rappelons ici la dynamique de population que nous avons décrite dans la Partie 1.1. Comme nous l'avons vu, la première grandeur démographique cruciale à l'échelle de l'individu est son âge, noté  $a$ , qui part de 0 à la naissance puis évolue

au cours du temps. En vue des applications démographiques, nous utilisons l'âge calendaire qui évolue à la même vitesse que le temps. Ainsi si un individu est né au temps  $\nu$ , son âge au temps  $t$  est  $t - \nu$ . Notons ici que cette évolution de l'âge pourrait aussi être modélisée de plusieurs autres manières en ayant recours à des vitesses de vieillissement (voir par exemple Tran (2008) pour la formalisation de ce phénomène). Si l'on observe un groupe de même âge, on distinguera en pratique des niveaux de taux démographiques différents. Ainsi, il est nécessaire d'inclure des caractéristiques, qui traduisent l'hétérogénéité de la population. On notera  $x$  le vecteur de caractéristiques associé à l'individu, qui est un élément d'un ensemble  $\mathcal{X}$  compact dans  $\mathbb{R}^d$ ,  $d \geq 1$ . Nous avons déjà décrit les différents événements démographiques qui peuvent concerner un individu de caractéristiques  $x$  et d'âge  $a$  au temps  $t$ , ainsi que les intensités (aussi appelées taux) démographiques associées. Celui-ci a un taux de naissance  $b(x, a, t)$ , une intensité de changement de caractéristiques au cours de sa vie  $e(x, a, t)$ , et un taux de décès  $\mu(x, a, t)$ . Lorsqu'il y a un événement de naissance, on doit préciser quelles sont les caractéristiques  $x'$  qui sont héritées par le nouvel individu : on définit pour cela un noyau de mutation  $k_b(x, a, t, x')m(dx')$ , où  $m$  est une mesure de probabilité diffuse de référence sur  $\mathcal{X}$ . Il en est de même pour les événements de changements de caractéristiques, auxquels l'individu passe des caractéristiques  $x$  à  $x'$  avec un noyau d'évolution  $k_e(x, a, t, x')m(dx')$ .

Avant de décrire la représentation aléatoire de la population et sa dynamique, revenons sur l'évolution déterministe de la population. Nous avons noté  $g(x, a, t)$  la quantité d'individus de caractéristiques  $x$  et d'âge  $a$  au temps  $t$ . Les Équations (1.7) et (1.9) nous donnent la dynamique déterministe en présence de naissances et de morts, mais en l'absence de taux de changements de caractéristiques. Etant donnés les taux démographiques, la dynamique déterministe globale qui prend en compte les naissances, les morts et les changements de caractéristiques au cours de la vie est donnée par l'équation de transport-renouvellement :

$$\begin{aligned} (\partial_a + \partial_t)g(x, a, t) &= -(\mu(x, a, t) + e(x, a, t))g(x, a, t) \\ &+ \int_{\mathcal{X}} g(x', a, t)e(x', a, t)k_e(x', a, t, x)m(dx'), \\ g(x, 0, t) &= \int_{\mathcal{X} \times \mathbb{R}_+} g(x', a, t)b(x', a, t)k_b(x', a, t, x)m(dx')da, \end{aligned} \tag{1.10}$$

avec la condition initiale  $g(x, a, 0) = g_0(x, a)$ . Celle-ci s'interprète de la façon suivante : une cohorte de caractéristique donnée évolue suivant les décès d'une part et les changements de caractéristiques d'autre part. Ainsi, des individus quittent la classe  $x$  (décès ou changement) mais d'autres individus qui avaient une caractéristique  $x'$  en changent pour rejoindre la classe  $x$ . Le terme de renouvellement traduit quant à lui le nombre de naissances de caractéristiques  $x$  à une date  $t$  fixée : celui-ci se calcule en sommant sur tous les parents  $(x', a)$  possibles, et en regardant la

fraction de ceux qui ont donné naissance à des individus de caractéristique  $x$ .

### 1.2.2 Représentation mesure de la population

En réalité, même si les taux démographiques sont fixés, il subsiste un aléa dans la dynamique dû à la taille finie de la population dans laquelle les événements démographiques ont lieu aléatoirement. Nous avons introduit cette notion de *risque idiosyncratique* ou *risque démographique*. Celui-ci est d'autant plus important que la population est petite. C'est surtout le cas par exemple pour des portefeuilles d'assurés, mais même pour une population nationale, on peut avoir un nombre réduit de personnes par classe d'âge et donc un risque démographique non-négligeable.

Pour avancer dans cette direction, il faut d'abord représenter l'état d'une population décrite à l'échelle individuelle. Celui-ci n'est plus représenté par les valeurs réelles  $g(x, a, t)$  pour tous  $a$  et  $x$ , mais plutôt par un nombre d'individus.

Nous avons recours pour cela au formalisme de la mesure : un individu  $(x_0, a_0) \in \mathcal{X} \times \mathbb{R}_+$  est représenté par une masse de Dirac  $\delta_{(x_0, a_0)}(dx, da)$ , qui met un poids en la caractéristique et l'âge de l'individu. C'est une mesure, réduite à un atome, sur l'espace  $\mathcal{X} \times \mathbb{R}_+$ . Si on se donne une collection d'individus  $(x_i, a_i)$  pour  $1 \leq i \leq n$ ,  $n$  entier fixé, la population correspondante est aussi une mesure, qui est la somme des mesures individuelles :

$$Z(dx, da) = \sum_{i=1}^n \delta_{(x_i, a_i)}(dx, da). \quad (1.11)$$

On va noter  $\mathcal{M}$  l'ensemble des mesures positives et finies sur  $\mathcal{X} \times \mathbb{R}_+$ , auquel  $Z$  appartient. Cette représentation sous forme de mesure présente de nombreux intérêts. Un de ceux-ci est de faire vivre des populations très différentes sur le même espace  $\mathcal{M}$ , quelle que soit leur taille ou leur composition. Un formalisme de vecteur n'aurait par exemple pas cet avantage. Aussi, on va pouvoir faire le lien avec les équations déterministes car en réalité  $\xi_t(dx, da) := g(x, a, t)m(dx)da$  donné par (1.10) est également un élément de  $\mathcal{M}$ . Enfin, ce formalisme est adapté pour mesurer des quantités d'intérêt sur la population. Si on se donne  $f$  une fonction de l'âge et des caractéristiques, par exemple  $f(x, a) = \mathbf{1}_{[0, \bar{a}]}(a)\mathbf{1}_{\mathcal{X}_0}(x)$  qui vaut 1 si l'individu est d'âge plus petit que  $\bar{a}$  et de caractéristiques dans  $\mathcal{X}_0 \subset \mathcal{X}$ , on peut l'évaluer sur toute la population par l'intégration

$$\langle Z, f \rangle := \int_{\mathcal{X} \times \mathbb{R}_+} f(x, a)Z(dx, da) = \sum_{i=1}^n f(x_i, a_i), \quad (1.12)$$

où la dernière égalité vient d'après (1.11). Si par exemple on choisit  $f \equiv \mathbf{1}$  la fonction constante égale à 1, on retrouve la taille de la population, i.e.  $\langle Z, \mathbf{1} \rangle = n$ . On peut

représenter de manière plus naturelle la population  $Z$  sous forme de *configuration*, c'est-à-dire sous forme de l'ensemble de points  $Z = \{(x_i, a_i), 1 \leq i \leq n\}$ , voir Chapitre 2. Nous utilisons la bijection entre l'espace des configurations et celle des mesures ponctuelles finies, et en tirons avantage pour noter indifféremment :

$$\langle Z, f \rangle = \sum_{(x,a) \in Z} f(x, a).$$

### 1.2.3 Construction de la dynamique aléatoire

À l'aide des intensités démographiques de naissance, de changements de caractéristiques, et de décès, nous sommes en mesure de donner une description de la réalisation du prochain événement démographique pour un individu donné. Au-delà, il y a un réel intérêt à décrire la dynamique aléatoire globale du processus de population, sur la base de sa propre intensité. Nous allons construire la dynamique d'une population dont l'état est représenté par une mesure, comme décrit dans la partie précédente. Nous commençons par décrire la construction par *thinning* des processus de comptage.

#### La représentation *thinning* des processus de comptage

a) Considérons dans un premier temps  $\bar{T}_1 < \bar{T}_2 < \dots < \bar{T}_n < \dots$  les instants de saut ordonnés d'un processus de Poisson  $\bar{N}$  d'intensité  $\bar{\lambda}$ . Dans ce cadre, les temps inter-arrivées  $(\bar{T}_{n+1} - \bar{T}_n)$  sont iid de loi exponentielle de paramètre  $\bar{\lambda}$ . Ensuite, marquons chaque instant de saut  $\bar{T}_n$  à l'aide d'une variable aléatoire indépendante  $Y_n$  à valeur dans un espace  $E$  muni d'une tribu  $\mathcal{E}$ , de loi  $\bar{\mu}(dy)$ , et tel que la suite des  $(Y_n)$  soit indépendante des  $(\bar{T}_n)$ . Un tel processus est appelé *processus de Poisson marqué* sur  $\mathbb{R}_+ \times E$  de mesure intensité  $q(ds, dy) = \bar{\lambda} ds \bar{\mu}(dy)$ . Ce processus peut être représenté par une *mesure ponctuelle* sur  $\mathbb{R}_+ \times E$ , notée

$$Q(ds, dy) = \sum_{n \geq 1} \delta_{(\bar{T}_n, Y_n)}(ds, dy).$$

Remarquons ici que le processus de Poisson de départ peut être retrouvé en écrivant

$$\bar{N}_t = \sum_{n \geq 1} \mathbf{1}_{\bar{T}_n \leq t} = \sum_{n \geq 1} \mathbf{1}_{\bar{T}_n \leq t} \mathbf{1}_{Y_n \in E} = \int_0^t \int_E Q(ds, dy).$$

b) Une bonne propriété de la mesure ponctuelle  $Q$  consiste en la possibilité de construire des martingales associées. Notons  $(\mathcal{G}_t)$  la filtration canonique engendrée



par  $Q$  et considérons une fonction mesurable bornée  $f(s, y)$  sur  $\mathbb{R}_+ \times E$ . Alors nous avons la propriété suivante :  $\langle Q, f \rangle_t - \langle q, f \rangle_t$  est une  $(\mathcal{G}_t)$ -martingale, où l'on note

$$\langle Q, f \rangle_t = \int_0^t \int_E f(s, y) Q(ds, dy) = \sum_{n \geq 0} f(T_n, Y_n). \quad (1.13)$$

Cette propriété reste vraie si  $f(s, y)$  est aléatoire et (i) est  $P(\mathcal{G}_t) \times \mathcal{E}$ -mesurable, où  $P(\mathcal{G}_t)$  représente la tribu prévisible associée à  $\mathcal{G}_t$  et (ii) tant que  $\mathbb{E}[\langle q, |f| \rangle] < +\infty$ . En fait, si l'on a seulement  $\langle q, |f| \rangle < +\infty$  p.s., on n'a plus une martingale mais seulement une martingale locale.

c) À l'aide des points a) et b), nous sommes en mesure de représenter une classe variée de processus de comptage. Considérons d'abord un processus de Poisson non-homogène d'intensité (déterministe)  $\lambda_t$ , c'est-à-dire un processus qui saute de 1 et vérifie le fait que  $N_t - \int_0^t \lambda_s ds$  est une martingale. Supposons que  $\lambda_t \leq \bar{\lambda}$  pour une certaine constante positive  $\bar{\lambda}$ . On se donne la mesure ponctuelle  $Q(ds, d\theta)$  en spécifiant  $E = [0, 1]$  et  $\mu$  la mesure de Lebesgue, que l'on note  $d\theta$ , et en notant dorénavant  $\Theta_n$  les marques associées. D'après ce qui précède, il s'agit de trouver une bonne fonction  $f$  pour laquelle d'une part  $N_t = \langle Q, f \rangle_t$  et d'autre part  $\lambda_t = \int_{[0,1]} f(t, \theta) \bar{\lambda} d\theta$ . Ceci nous pousse à considérer  $f(s, \theta) = \mathbf{1}_{[0, \lambda_s/\bar{\lambda}]}(\theta)$ , ce qui nous donne la construction suivante

$$N_t = \int_0^t \int_{[0,1]} \mathbf{1}_{[0, \lambda_s/\bar{\lambda}]}(\theta) Q(ds, d\theta) = \sum_{n \geq 1} \mathbf{1}_{\bar{T}_n \leq t} \mathbf{1}_{\Theta_n \leq \lambda_{\bar{T}_n}/\bar{\lambda}}.$$

Ceci donne un procédé de simulation très naturel pour le processus  $N$  : on simule d'abord les instants de saut du processus plus simple  $\bar{N}$ , et on conserve le saut en  $\bar{T}_n$  avec probabilité  $\lambda_{\bar{T}_n}/\bar{\lambda}$ . Deux exemples de simulation de processus de Poisson non-homogènes sont représentés sur la Figure 1.12.

Supposons maintenant que l'on se donne un processus  $\lambda_t$  positif,  $(\mathcal{G}_t)$ -prévisible, et dominé par  $\bar{\lambda}$ , et donc tel que  $\int_0^t \lambda_s ds < +\infty$  p.s.. On peut alors construire de la même façon le processus de comptage associé, puisque l'on conserve la propriété de martingale (locale). À ce stade, il nous reste à éclaircir deux points. Le premier consiste à décrire comment construire un tel processus lorsque l'intensité  $\lambda_t$  que l'on s'est donnée a priori n'est pas bornée. Le second concerne la construction des processus de comptage dont l'intensité n'est pas donnée a priori mais seulement décrite de manière récursive à partir du processus lui-même.

d) Comment construire le processus de comptage associé lorsque l'intensité, donnée a priori, n'est pas bornée ? En fait, cela requiert que l'espace des marques  $\theta$  soit  $\mathbb{R}_+$  (au lieu de  $[0, 1]$ ). Dans ce cadre, on va perdre l'interprétation en terme de temps de saut ordonnés et de marques tirées suivant une distribution donnée. Cependant,

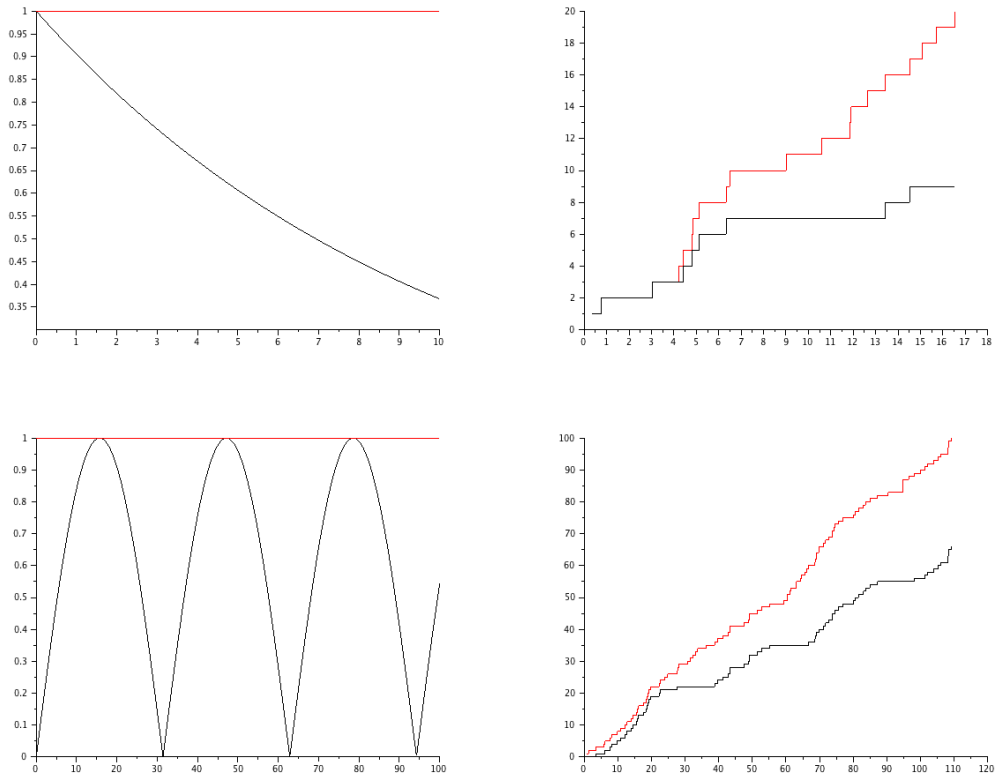


FIGURE 1.12 – Exemples de simulation d'un processus de comptage par thinning. À gauche, l'intensité est représentée en noir et la borne associée en rouge. À droite, on représente une trajectoire du processus de référence obtenue par thinning (en noir) à partir de la trajectoire du processus de Poisson (en rouge) de paramètre la borne rouge.

dès que l'on va restreindre cet espace de marques, on retombera sur la notion de processus de Poisson marqué défini au point a). Plus précisément, on définit l'objet suivant : on appelle *mesure ponctuelle de Poisson (générale)*  $Q(ds, dy)$  de mesure intensité  $q(ds, d\theta) = ds\mu(dy)$  sur  $\mathbb{R}_+ \times E$ , où (nous insistons là-dessus) la mesure  $\mu$  est seulement  $\sigma$ -finie, la mesure aléatoire à valeur dans  $\mathbb{N} \cup \{+\infty\}$  satisfaisant

- (i) pour tous ensembles mesurables disjoints  $B_1, \dots, B_k$  de  $\bar{E} = \mathbb{R}_+ \times E$ , les variables aléatoires  $Q(B_1), \dots, Q(B_k)$  sont indépendantes (les variables infinies étant, en tant que constantes, indépendantes des autres),
- (ii) pour tout ensemble mesurable  $B \subset \bar{E}$  tel que  $q(B) < +\infty$ ,  $Q(B) \sim \mathcal{Poisson}(q(B))$ .
- (iii)  $Q(\{0\} \times E) = 0$  (ceci assure qu'il n'y a pas de saut au temps 0).

La mesure ponctuelle de Poisson ainsi construite définit une mesure de comptage sur  $\mathbb{R}_+^*$  au sens où pour tout  $t > 0$ ,  $Q(\{t\} \times E) \in \{0, 1\}$  (voir par exemple Çinlar (2011), Théorème 2.17).

La propriété cruciale de martingalité locale est conservée pour ces mesures : pour toute fonction  $f(s, y)$   $P(\mathcal{G}_t) \times \mathcal{E}$ -mesurable telle que p.s.  $\forall t \geq 0, \langle q, |f| \rangle_t < +\infty$ , alors  $\langle Q, f \rangle_t - \langle q, f \rangle_t$  est une martingale locale.

On considère maintenant que  $E = \mathbb{R}_+$  et  $\mu$  est la mesure de Lebesgue sur  $\mathbb{R}_+$ , notée  $d\theta$ , de sorte que la mesure ponctuelle de Poisson  $Q(ds, d\theta)$  a pour mesure intensité  $q(ds, d\theta) = dsd\theta$ . Si on se donne un processus  $(\mathcal{G}_t)$ -prévisible positif  $\lambda$ , tel que p.s. pour tout  $t$ ,  $\int_0^t \lambda_s ds < +\infty$ , on peut alors construire le processus de comptage associé par

$$N_t = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{[0, \lambda_s]}(\theta) Q(ds, d\theta) = \sum_{n \geq 1} \mathbf{1}_{\bar{T}_n \leq t} \mathbf{1}_{\Theta_n \leq \lambda_{\bar{T}_n}}. \quad (1.14)$$

On obtient bien par propriété de  $Q$  que  $N_t - \int_0^t \lambda_s ds$  est une  $(\mathcal{G}_t)$ -martingale locale. Il nous reste à aborder la construction d'un processus de comptage dont l'intensité est définie à partir du processus lui-même.

e) Nous arrivons à notre cadre d'intérêt. Prenons l'exemple du processus de naissance pur à taux constant  $\bar{b}$ , partant d'une population de taille  $N_0$ . Chaque individu dans la population donne naissance après un temps exponentiel de paramètre  $\bar{b}$ , ainsi le premier temps de naissance dans la population est le minimum des  $N_0$  variables exponentielles. Si l'on suppose que les instants démographiques des individus sont indépendants, ce minimum est distribué suivant une variable exponentielle de paramètre  $\bar{b}N_0$ , et ainsi de suite. Ainsi, ce que l'on souhaite traduire est que l'intensité du processus  $(N_t)$  qui compte le nombre d'individus dans la population n'est rien d'autre que  $\lambda_t = \bar{b}N_{t-}$ . Ceci constitue un cas particulier très simple pour lequel l'intensité s'exprime *en fonction du processus lui-même*. Plus généralement, on souhaite définir un processus de comptage  $(N_t)$  à partir de son intensité  $\lambda_t = \Psi(N_s, 0 \leq s < t)$

qui est une fonction du passé de  $N$  jusqu'en  $t^-$  et qui est donc bien  $(\mathcal{F}_t^N)$ -prévisible, où  $\mathcal{F}^N$  est la filtration canonique associée au processus  $N$  lui-même. L'idée, basée sur la représentation (1.14) est de *définir* le processus  $N$  comme *solution de l'équation*

$$N_t = N_0 + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{[0, \Psi(N_u, 0 \leq u < s)]}(\theta) Q(ds, d\theta). \quad (1.15)$$

La représentation *thinning* de processus de comptage est classique dans la littérature ; elle a été étudiée à travers différents points de vue de façon épisodique. À notre connaissance, une telle représentation mathématique remonte à Kerstan (1964) et Grigelionis (1971). On trouve dans Lewis and Shedler (1978) et Ogata (1981) la description du processus de simulation associé à laquelle il est souvent fait référence dans la littérature. Toujours dans le contexte des processus de comptage, cette construction est détaillée dans Brémaud and Massoulié (1996) et Massoulié (1998) (voir aussi les livres de Daley and Vere-Jones (2008) et Çinlar (2011)). Dans la suite, nous étudions la représentation *thinning* pour des processus de population ; les questions d'existence et d'unicité seront abordées dans ce cadre.

**Remarque 1.** *Nous avons discuté la définition d'un processus de comptage comme la solution à un problème de thinning : partant d'une mesure ponctuelle de Poisson  $Q$ , on construit un processus de comptage  $N$ . Inversement, si on se donne un tel processus de comptage bien défini, de filtration canonique  $\mathcal{F}^N$  et d'intensité  $\mathcal{F}^N$ -prévisible  $(\lambda_t)$ , on peut se demander si l'on est capable de le représenter sous la forme thinning. Effectivement, il est alors possible de construire, sur un espace de probabilité élargi, une mesure ponctuelle de Poisson  $Q$  telle que*

*$N_t = \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{[0, \lambda_s]}(\theta) Q(ds, d\theta)$ . Ceci donne alors l'équivalence entre d'un côté le problème de martingale et de l'autre la représentation comme solution à une équation différentielle stochastique. Les résultats généraux associés peuvent être trouvés dans Jacod (1979) (Chapitre 14), et dans le cas des processus de Hawkes par exemple dans Massoulié (1998) et Daley and Vere-Jones (2008) (Proposition 14.7.I). Voir aussi sur ce thème Delattre et al. (2014) pour le cas de processus de Hawkes en interaction sur un graphe et une discussion générale.*

### Processus de naissance et mort avec âge

a) Donnons tout de suite la construction de la dynamique d'une population dans laquelle les individus ont un âge, formalisée en tant que mesure tel que décrit dans la Partie 1.2.2. Concentrons nous d'abord sur une description du processus de naissance pur avec âge, proche de celle donnée dans la Partie 1.1.2 pour introduire les processus

de Hawkes. Nous partons d'une pyramide des âges initiale (déterministe)  $Z_0(da) = \sum_{i=1}^{N_0} \delta_{a_i}(da)$  composée de  $N_0$  individus d'âges  $a_1, a_2, \dots, a_{N_0}$ . Dans ce modèle, chaque individu d'âge  $a$  donne naissance à un nouvel individu (d'âge 0) au taux  $b(a)$ , et les individus sont supposés indépendants. Ainsi, si l'on note  $Z_t(da)$  la pyramide des âges (aléatoire) au temps  $t$ , qui met un poids en l'âge de chaque individu, l'intensité de naissance globale, i.e. la fréquence d'arrivée d'un nouvel individu dans la population, s'écrit avec la notation (1.12),

$$\lambda_t := \sum_{x \in Z_{t-}} b(x) = \langle Z_{t-}, b \rangle.$$

Notons que si le taux de naissance est indépendant de l'âge, i.e.  $b(a) = \bar{b}$ , on retrouve l'intensité du processus de naissance pur  $\lambda_t = \bar{b}N_{t-}$ . Rappelons que l'on s'est donné  $Q(ds, d\theta)$  mesure ponctuelle de Poisson sur  $\mathbb{R}_+ \times \mathbb{R}_+$  de mesure intensité  $d_s d\theta$ . On définit alors le processus de naissance pur  $Z_t(da)$  comme la solution de l'équation

$$Z_t(da) = Z_0^t(da) + \int_0^t \int_{\mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\theta \leq \langle Z_{s-}, b \rangle} Q(ds, d\theta), \quad (1.16)$$

où  $Z_0^t(da)$  est la mesure (déterministe) obtenue en ajoutant  $t$  à tous les âges initiaux, c'est-à-dire  $Z_0^t(da) = \sum_{i=1}^{N_0} \delta_{a_i+t}(da)$ . L'Équation (1.16) s'interprète simplement : la pyramide des âges au temps  $t$  consiste en l'ensemble des individus initiaux ayant vieilli d'une durée  $t$ , auquel on ajoute les individus nés dans la population à fréquence  $\langle Z_{s-}, b \rangle$ ; si un individu est arrivé au temps  $s$ , son âge au temps  $t$  est  $t - s$ . À noter que sous la forme (1.16), le processus de pyramide des âges n'est pas immédiatement différentiable. Ceci est discuté en détail dans le Chapitre 2. Pour obtenir des équations différentielles, il faut tester l'Équation (1.16) contre des fonctions de classe  $\mathcal{C}^1(\mathbb{R}_+)$ , et après intégration par parties, on obtient

$$d\langle Z_t, f \rangle = f(0)d\langle Z_t, \mathbf{1} \rangle + \langle Z_t, f' \rangle dt. \quad (1.17)$$

Cette décomposition fait apparaître le terme de saut pur (arrivées d'individus d'âge zéro) et le terme de vieillissement. Cette décomposition est utile à plusieurs titres. D'abord, celle-ci permet ensuite de compenser la partie saut pour obtenir la martingale associée, ce qui est une formulation de base pour ensuite développer les théorèmes limites sur de tels processus et caractériser l'équation limite; un résultat classique important d'approximation en grande population est présenté dans cette partie et détaillé dans le Chapitre 2. Ce type de décomposition sera également exploité dans le Chapitre 3 dans l'étude des processus de Hawkes pour étudier leur dynamique, ainsi dans le Chapitre 6 pour l'établissement d'un théorème de convergence en grande population en présence de plusieurs échelles de temps. On remarquera dans le Chapitre 3 que si (i) le taux de naissance  $b$  est suffisamment régulier, on peut appliquer la décomposition (1.17) récursivement à  $\langle Z_t, b^{(n)} \rangle$ , puis

si (ii)  $b$  est solution d'une équation différentielle ordinaire (à coefficients constants) d'ordre quelconque, il est possible d'obtenir un système fini d'équations, autrement dit une dynamique Markovienne de dimension finie.

b) Incluons maintenant les décès dans la dynamique. On passe alors d'une vision comptage (un événement revient à ajouter un individu) à une vision population ou processus de naissance et mort : on ajoute un individu si une naissance a lieu, et on retire un individu si l'on a un décès dans la population. Il y a une asymétrie fondamentale entre les événements de naissance et de mort en présence de l'âge. En effet, un événement de naissance revient toujours à ajouter un individu d'âge 0. En revanche, un événement de mort revient à retirer l'individu décédé de la population, plus précisément à retirer l'âge de l'individu en question de la pyramide des âges. Cette asymétrie fondamentale a des conséquences importantes dans différents aspects de modélisation. Par exemple, la régularité systématique des événements de naissance est un point essentiel dans l'analyse des processus de Hawkes (voir Partie 1.3 et Chapitre 3). Également, dans le champ des théorèmes limites sur les processus de population, l'accélération des événements de naissance et de mort dans les modèles avec âge posent des difficultés techniques, voir ce sujet Méléard and Tran (2012) ; les méthodes pour les surmonter sont d'ailleurs utilisées dans le Chapitre 6 dans un contexte différent.

Pour modéliser des événements de mort dans un modèle en âge, le formalisme précédent donné en (1.16) n'est a priori pas adapté ; en effet, celui-ci ne permet pas d'identifier quel individu est concerné par un événement démographique. Ce que nous allons utiliser est une numérotation unique des individus, qui revient dans ce cas des processus de naissance pur à les ordonner par ordre d'arrivée dans la population. On construit pour cela une suite  $(A_i(t))_{i \in \mathbb{N}^*}$  de fonctions telle que  $A_i(t)$  est l'âge potentiel au temps  $t$  de l'individu arrivé en  $i^{\text{ième}}$  position (incluant la population initiale qui a été réordonnée). Cette construction est détaillée dans le Chapitre 2, Partie 2.3.3. Adaptons d'abord ce formalisme dans le cas du processus de naissance pur avec âge. Pour cela, remarquons la forme très spécifique de l'intensité  $\lambda_t = \sum_{x \in Z_{t-}} b(x) = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{A_i(t) \in Z_{t-}} b(A_i(t))$ , qui s'écrit comme une somme sur les individus, dû au fait que les événements démographiques de ceux-ci sont supposés indépendants. Ainsi, afin d'identifier l'individu concerné par un événement démographique, on va se donner une mesure ponctuelle de Poisson (toujours) notée  $Q(ds, di, d\theta)$  sur  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}_+$  de mesure intensité  $q(ds, di, d\theta) = ds n(di) d\theta$ , où  $n$  est la mesure de comptage sur  $\mathbb{N}^*$ , c'est-à-dire que pour tout  $A \subset \mathbb{N}^*$ ,  $n(A)$  est le nombre d'éléments de  $A$ .

Dans ce cadre, on peut de manière équivalente à (1.16) définir le processus de

pyramide des âges comme solution à l'équation

$$Z_t(da) = Z_0^t(da) + \int_0^t \int_{\mathbb{N}^* \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{A_i(s) \in Z_{s-}} \mathbf{1}_{\theta \leq b(A_i(s))} Q(ds, di, d\theta). \quad (1.18)$$

Cette écriture caractérise le fait que, si un événement de naissance a eu lieu au temps  $s$ , c'est l'individu  $i$  qui a donné naissance. On peut se convaincre tout de suite qu'il s'agit de la bonne intensité : en effet,

$$\int_{\mathbb{N}^* \times \mathbb{R}_+} \mathbf{1}_{A_i(t) \in Z_{t-}} \mathbf{1}_{\theta \leq b(A_i(t))} n(di) d\theta = \sum_{i \in \mathbb{N}^*} \mathbf{1}_{A_i(t) \in Z_{t-}} b(A_i(t)) = \langle Z_{t-}, b \rangle.$$

En terme de simulation, l'Équation (1.18) donne une interprétation légèrement différente : on sélectionne un individu  $i$  uniformément dans la population, puis une variable additionnelle  $\theta$  détermine si celui-ci donne naissance ou non. Attention, ici l'espace  $\mathbb{R}_+$  pour  $\theta$  n'étant pas borné, la description précédente n'est pas utilisable en pratique pour la simulation. Il va falloir pour cela borner l'espace de  $\theta$ , ce qui est en fait possible si les taux démographiques sont bornés. Ceci sera discuté dans la suite.

Il nous reste à construire le processus de naissance et mort avec âge. Ceci revient à ajouter à l'Équation (1.18) le fait que si un événement de mort a eu lieu au temps  $s$  et que cet événement concerne l'individu  $i$ , alors la pyramide des âges au temps  $t$  sera obtenue en retirant l'âge qu'aurait eu cet individu à cette date, c'est-à-dire  $A_i(s) + t - s$ . En d'autres termes, le processus de pyramide des âges est défini comme la solution de l'équation

$$Z_t(da) = Z_0^t(da) + \int_0^t \int_{\mathbb{N}^* \times \mathbb{R}_+} \mathbf{1}_{A_i(s) \in Z_{s-}} \left( \delta_{t-s}(da) \mathbf{1}_{\theta \leq b(A_i(s))} - \delta_{A_i(s)+t-s}(da) \mathbf{1}_{b(A_i(s)) < \theta \leq b(A_i(s))+d(A_i(s))} \right) Q(ds, di, d\theta).$$

La procédure de simulation associée repose aussi sur le tirage d'un individu  $i$  uniformément dans la population, et d'une variable additionnelle  $\theta$  qui détermine, en fonction des taux démographiques de cet individu, si un événement a eu lieu (le cas échéant, naissance ou décès).

### Inclure des caractéristiques dans la dynamique

a) Nous souhaitons enrichir la dynamique de naissance et mort avec âge en ajoutant à chaque individu une caractéristique. Pour cela, étendons d'abord la construction des processus de comptage par *thinning* que nous avons décrite précédemment. Les mesures ponctuelles de Poisson peuvent être utilisées pour construire des processus de comptage, mais plus généralement des mesures ponctuelles aléatoires sur

$\mathbb{R}_+ \times \mathcal{X}$ , disons  $\Gamma(ds, dx) = \sum_{n \geq 1} \delta_{(T_n, X_n)}(ds, dx)$ . On peut voir  $T_n$  comme un événement démographique (par exemple une naissance), et  $X_n$  la caractéristique associée (par exemple celle du nouveau-né). On se donne une mesure de probabilité de référence  $m(dx)$  sur l'espace  $\mathcal{X}$  des caractéristiques. Construisons cette mesure aléatoire ponctuelle  $\Gamma(ds, dx)$ , en supposant que sa mesure intensité admet une densité par rapport à la mesure de référence, i.e.  $\gamma(ds, dx) = \gamma(s, x) ds m(dx)$ . Remarquons que dans ce type de modèle, la fréquence des événements est donnée par l'intensité  $s \mapsto \int_{x \in E} \gamma(s, x) m(dx)$ , et si un événement a eu lieu à la date  $T_n$ , les caractéristiques  $X_n$  sont tirées suivant la loi de densité  $\frac{\gamma(T_n, x) m(dx)}{\int_{y \in \mathcal{X}} \gamma(T_n, y) m(dy)}$ . On se donne une mesure ponctuelle de Poisson  $Q(ds, dx, d\theta)$  sur  $\mathbb{R}_+ \times \mathcal{X} \times \mathbb{R}_+$  de mesure intensité  $ds m(dx) d\theta$ . On note  $(\mathcal{F}_t^Q)$  la filtration canonique générée par  $Q$ , et on introduit  $P(\mathcal{F}_t^Q)$  la tribu prévisible associée à  $\mathcal{F}_t^Q$ . On suppose ici que l'on s'est donné la quantité  $\gamma(t, x)$  qui est  $P(\mathcal{F}_t^Q) \times \mathcal{E}$ -mesurable ( $\mathcal{E}$  est ici la tribu associée à  $\mathcal{X}$ ) et aussi que  $\int_0^t \int_{\mathcal{X}} \gamma(s, x) ds m(dx) < +\infty$  p.s.. On construit alors la mesure  $\Gamma$  comme (insistons sur le fait que ce n'est pas ici une équation sur  $\Gamma$  car  $\gamma$  est donné dans la filtration de  $Q$ )

$$\Gamma(ds, dx) = \int_{\mathbb{R}_+} \mathbf{1}_{[0, \gamma(s, x)]}(\theta) Q(ds, dx, d\theta). \quad (1.19)$$

On retrouve bien une mesure ponctuelle et la propriété de martingale pour  $Q$  assure que la mesure intensité est bien  $\gamma(s, x) ds m(dx)$ . Une telle construction est détaillée par Massoulié (1998) auquel nous renvoyons le lecteur pour plus de développements.

**Remarque 2.** (*Lien avec la méthode d'acceptation-rejet*) Nous souhaitons faire le lien entre la méthode du thinning et celle dite d'acceptation-rejet de Von Neumann (1951). On se donne deux fonctions mesurables  $f$  et  $g$  sur  $E$ , strictement positives  $m$ -presque partout, telles que  $g$  est une densité de probabilité pour  $m$  et il existe une constante  $c$  telle que  $f(x) \leq cg(x)$ . On suppose que l'on sait simuler selon  $g(x)m(dx)$  et sur cette base, on souhaite générer des variables aléatoires iid de loi  $\frac{f(x)}{\langle m, f \rangle} m(dx)$ , en particulier sans avoir recours au calcul de l'intégrale  $\langle m, f \rangle = \int_E f(x) m(dx)$ . La méthode d'acceptation-rejet revient à simuler une suite de variables aléatoires iid  $(\Theta_k, X_k)$  où les  $\Theta_k$  sont uniformes sur  $[0, 1]$  et les  $X_k$  sont générées suivant  $g(x)m(dx)$ . Alors l'ensemble des  $X_k$  tels que  $\Theta_k cg(X_k) \leq f(X_k)$  forme un échantillon iid de loi  $\frac{f(x)}{\langle m, f \rangle} m(dx)$ . Pour prouver immédiatement ce résultat, on se donne  $Q(ds, dx, d\theta)$  une mesure ponctuelle de Poisson sur  $\mathbb{R}_+ \times E \times [0, 1]$  de mesure intensité  $q(ds, dx, d\theta) = ds \cdot cg(x) m(dx) d\theta$ . On définit ensuite la mesure  $\Gamma(ds, dx)$  par

$$\Gamma(ds, dx) = \int_{[0, 1]} \mathbf{1}_{\theta \leq \frac{f(x)}{cg(x)}} Q(ds, dx, d\theta),$$

qui d'après ce qui précède a pour intensité  $ds f(x) m(dx) = \langle m, f \rangle ds \frac{f(x)}{\langle m, f \rangle} m(dx)$ . On obtient bien une mesure dont les marques sont distribuées selon la loi voulue. On



peut donc voir la méthode d'acceptation-rejet comme un cas particulier de thinning pour lequel l'espace de restriction ne dépend pas du temps.

b) Introduisons maintenant l'équation pour le processus de population général. On se donne  $Q(ds, di, dx, d\theta)$  une mesure ponctuelle de Poisson sur  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathcal{X} \times \mathbb{R}_+$  de mesure intensité  $q(ds, di, dx, d\theta) = ds n(di) m(dx) d\theta$ . Afin de définir le processus de population général, introduisons l'environnement aléatoire. On se donne  $(Y_t)_{t \geq 0}$  un processus càdlàg à valeur dans  $\mathbb{R}^l$ , indépendant de la mesure ponctuelle de Poisson  $Q$ . Les taux démographiques sont supposées dépendre de la caractéristique et de l'âge de l'individu, mais également du temps et de l'environnement aléatoire. On se donne de même une suite de points  $(X_i, A_i(t))$  correspondant aux individus par ordre d'arrivée, voir de nouveau le Chapitre 2, Partie 2.3.3 pour plus de détails.

On note  $\lambda_c(Z_{t-}, t, i, x)$  l'intensité au temps  $t$  pour l'individu  $i$  dans la population  $Z_{t-}$  correspondant à l'événement démographique  $c \in \{b, d, e\}$ . Ces intensités sont données par :

$$\begin{aligned}\lambda_b(Z_{t-}, t, i, x) &= b(X_i, A_i(t), t, Y_t) k_b(X_i, A_i(t), t, Y_t, x), \\ \lambda_d(Z_{t-}, t, i, x) &= d(X_i, A_i(t), t, Y_t), \\ \lambda_e(Z_{t-}, t, i, x) &= e(X_i, A_i(t), t, Y_t) k_e(X_i, A_i(t), t, Y_t, x).\end{aligned}\tag{1.20}$$

On introduit les indicatrices suivantes :

- *Naissance* :  $J^b(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\theta \leq \lambda_b(Z_{t-}, t, i, x)}$ ,
- *Mort* :  $J^d(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\lambda_b(Z_{t-}, t, i, x) < \theta \leq (\lambda_b + \lambda_d)(Z_{t-}, t, i, x)}$ ,
- *Swap* :  $J^e(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{(\lambda_b + \lambda_d)(Z_{t-}, t, i, x) < \theta \leq (\lambda_b + \lambda_d + \lambda_e)(Z_{t-}, t, i, x)}$ .

Nous introduisons pour la suite la notation  $\mathbf{S} = \mathcal{X} \times \mathbb{R}_+$ , ainsi que l'espace  $C_b^{0,1}(\mathbf{S})$  des fonctions continues bornées sur  $\mathbf{S}$  et dérivable sur  $\mathbb{R}_+$  et dont la dérivée partielle en âge est bornée. Sur cette base, le processus de naissance-mort-swap est défini comme la solution au système d'équations différentielles : pour toute  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$\begin{aligned}d\langle Z_t, f \rangle &= \int_{i \in \mathbb{N}^*} \int_{x \in \mathcal{X}} \int_{\theta \in \mathbb{R}_+} \left\{ J^b(Z_{t-}, t, i, x, \theta) f(x, 0) - J^d(Z_{t-}, t, i, x, \theta) f(X_i, A_i(t)) \right. \\ &\quad \left. + J^e(Z_{t-}, t, i, x, \theta) [f(x, A_i(t)) - f(X_i, A_i(t))] \right\} Q(ds, di, dx, d\theta) + \langle Z_t, \partial_a f \rangle dt.\end{aligned}\tag{1.21}$$

Afin d'obtenir l'existence forte et l'unicité trajectorielle, nous introduisons l'hypothèse ci-dessous ; celle-ci indique que les taux démographiques individuels et les noyaux de mutation sont bornés, ce qui est naturel pour les applications qui nous concernent.

**Hypothèse 1.** *Les taux démographiques et les noyaux de mutation sont continus, et il existe des constantes strictement positives  $\bar{k}$ ,  $\bar{b}$ ,  $\bar{d}$  and  $\bar{e}$  telles que pour tous  $(x, a, t, y, x') \in E \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l \times E$ ,  $k^e(x, a, t, y, x') \leq \bar{k}$ ,  $k^b(x, a, t, y, x') \leq \bar{k}$ ,  $b(x, a, t, y) \leq \bar{b}$ ,  $d(x, a, t, y) \leq \bar{d}$  et  $e(x, a, t, y) \leq \bar{e}$ .*

**Proposition 1.** *Notons  $N_t = \langle Z_t, \mathbf{1} \rangle$  et supposons que  $\mathbb{E}[N_0] < +\infty$ . Sous l'Hypothèse 1, l'Équation (1.21) admet une unique solution forte càdlàg  $(Z_t)$  telle que pour tout  $T > 0$ ,  $\mathbb{E}[\sup_{t \in [0, T]} N_t] < +\infty$ .*

Il est important de noter que conditionnellement à l'environnement  $Y$ , le processus  $Z$  à valeur mesure construit en (1.21) est un processus de Markov (non-homogène). Cette propriété est cruciale car elle indique que dans ce modèle, toute l'information sur la dynamique est contenue dans la structure en âges et caractéristiques de la population. Ceci sera exploité en particulier dans la Partie 1.3 et le Chapitre 3 pour l'étude des processus de Hawkes. Notons également, que l'action du générateur infinitésimal de ce processus peut être lue sur le problème des martingales, voir Equation (2.11), Section 2.5.

### 1.2.4 Lien avec les équations déterministes

Dans le champ de la microsimulation, on s'intéresse particulièrement au bruit démographique, c'est-à-dire à l'aléa dû à la taille finie de la population. Il est souvent évoqué que l'on peut réduire ce bruit soit (i) en moyennant un grand nombre de trajectoires aléatoires (Monte Carlo) ou (ii) en augmentant la taille de l'échantillon de population initial. En réalité, l'équivalence entre (i) et (ii) est loin d'être évidente, et celle-ci n'est en général pas vérifiée pour les modèles dits *non-linéaires* qui incluent des interactions entre les individus. Dans tous les cas, les deux points (i) et (ii) proposent deux points de vue différents sur la notion de *macroscopique*, et en particulier l'analyse du point (ii) nécessite des outils mathématiques avancés.

(i) La première approche consiste à étudier la moyenne de trajectoires microscopiques; les outils mathématiques associés sont connus sous le nom de *méthode des moments*. De telles techniques sont utilisées dans de nombreux domaines (voir par exemple un état de l'art dans Kuehn (2015)), et plus récemment et de manière active, c'est dans le champ de l'écologie mathématique que l'on trouve de nombreux développements depuis l'introduction des méthodes dans ce domaine par Bolker and Pacala (1997) et Dieckmann and Law (2000). La plus grande difficulté dans cette approche est le fait de devoir fermer le système d'équations sur les moments, voir la discussion dans Champagnat et al. (2006), ce qui nécessite des hypothèses de ferme-

ture de moments. L'autre point de vue développé propose d'étudier rigoureusement les asymptotiques en grande population.

(ii) Dans une perspective *grande population*, il s'agit d'étudier la dynamique d'une seule trajectoire basée sur un échantillon initial de plus en plus grand. D'un point de vue théorique, on considère une suite de processus de population indexée par un paramètre d'échelle qui représente l'ordre de grandeur de la population initiale. Ce paramètre d'échelle est aussi utilisé pour renormaliser (ici diviser) le poids de chaque individu dans la population afin d'obtenir des limites non-dégénérées. Cette renormalisation n'est pas anodine car elle implique une décroissance des interactions vers zéro ; ceci illustre la différence de point de vue avec la moyennisation des trajectoires à population initiale finie, pour lesquelles les interactions sont toujours présentes. La renormalisation que nous venons de décrire est celle de base qui permet d'obtenir une équation limite déterministe. Au-delà, un avantage de cette approche est de pouvoir inclure d'autres types de renormalisation, comme par exemple l'accélération des naissances et des morts, ou la raréfaction des mutations en écologie mathématique, et dans ce cas des modèles différents apparaissent à la limite (voir Fournier and Méléard (2004) pour plusieurs limites dans le cas sans âge, Tran (2008) ainsi que Méléard and Tran (2012) et le Chapitre 6 pour des limites d'échelle de populations structurées en âge). Un autre avantage de ces approches est que l'on peut obtenir des théorèmes centraux limites qui donnent de l'information sur la qualité de l'approximation, voir par exemple Wang (1977), Oelschläger (1990) et Tran (2006) pour de tels résultats pour des populations structurées en âge.

Ces deux approches (i) et (ii) diffèrent par plusieurs aspects, et surtout :

- a) en général, le comportement moyen ne coïncide pas avec l'approximation en grande population, dû à la présence d'interactions dans le modèles, et
- b) même en présence d'un échantillon initial très grand, le modèle microscopique présente des comportements différents comparé à son approximation grande population. Par exemple, le modèle microscopique peut quitter l'équilibre du modèle limite (voir par exemple Tran (2008)).

Dans la suite, nous considérons d'abord le modèle linéaire pour lequel nous détaillons la dynamique moyenne et l'existence d'une densité pour le modèle macroscopique associé. Dans un deuxième temps, nous présentons le cadre mathématique des limites en grande population et son application au modèle non-linéaire prenant en compte des interactions entre les individus.

### **Modèle linéaire, existence de densité**

Si l'on se fixe l'environnement  $Y$ , les trajectoires de population partant d'une population finie  $Z_0$  présentent un aléa que l'on appelle *bruit démographique*. Un première

quantité d'intérêt concerne la trajectoire moyenne de la population, qui consiste ici à étudier la *mesure moyenne conditionnelle*  $\xi_t^Y(dx, da)$  définie par p.s. pour toute  $f \in \mathcal{C}_b^{0,1}(E \times \mathbb{R}_+)$ ,  $\langle \xi_t^Y, f \rangle = \mathbb{E}[\langle Z_t, f \rangle | Y]$ . Formellement, on peut écrire  $\xi_t(dx, da) = \mathbb{E} \left[ \sum_{(x', a') \in Z_t} \delta_{(x', a')}(dx, da) | Y \right]$ . Nous allons donner ici sa dynamique, qui s'écrit aisément en l'absence d'interactions. Celle-ci s'obtient directement en prenant l'espérance conditionnelle dans l'Équation (1.21). On rappelle pour la suite la notation  $\mathbf{S} = \mathcal{X} \times \mathbb{R}_+$ .

**Proposition 2.** *La mesure moyenne conditionnelle  $\xi_t^Y$  satisfait l'équation : pour toute fonction  $f : (x, a) \mapsto f(x, a)$ ,  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$  et pour tout  $t \in [0, T]$ ,*

$$\langle \xi_t^Y, f \rangle = \langle \xi_0^Y, f \rangle + \int_0^t \langle \xi_s^Y, \Psi_s^f \rangle ds, \quad (1.22)$$

avec  $\Psi$  définie par

$$\begin{aligned} \Psi_t^f(x, a) &= \partial_a f(x, a) + b(x, a, t, Y_t) \int_{\mathcal{X}} f(x', 0) k^b(x, a, t, Y_t, x') m(dx') - d(x, a, t, Y_t) f(x, a) \\ &+ e(x, a, t, Y_t) \int_{\mathcal{X}} (f(x', a) - f(x, a)) k^e(x, a, t, Y_t, x') m(dx'). \end{aligned}$$

L'intérêt de cette formulation est de retrouver directement la dynamique de population déterministe de l'Équation (1.10), soit sous forme faible sans plus d'hypothèses, soit sous forme forte si l'on suppose que la structure de la population initiale admet une densité par rapport à la mesure de référence, et également une certaine régularité. Cette hypothèse permet de montrer qu'il n'y a par la suite pas d'effet de concentration en âges ou en caractéristiques. Ces résultats sont données dans la proposition suivante.

**Proposition 3.** (i) *Supposons que  $\xi_0^Y$  admet une densité  $g_0^Y(x, a) = g^Y(x, a, 0)$  par rapport à la mesure de référence  $m(dx) \otimes da$  sur  $\mathbf{S}$ . Alors la solution  $\xi^Y$  de l'Équation (1.22) admet à chaque temps  $t$  une densité  $g_t^Y(x, a)$ .*  
(ii) *Si de plus p.s.  $g^Y \in C_b^{0,1,1}(\mathbf{S} \times [0, T])$ , alors  $g^Y$  est solution forte de l'équation*

$$\begin{aligned} (\partial_a g^Y + \partial_t g^Y)(x, a, t) &= - [d(x, a, t, Y_t) + e(x, a, t, Y_t)] g^Y(x, a, t) \\ &+ \int_{\mathcal{X}} g^Y(x', a, t) e(x', a, t, Y_t) k^e(x', a, t, Y_t, x) m(dx'), \\ g^Y(x, 0, t) &= \int_{\mathbf{S}} g^Y(x', a, t) b(x', a, t, Y_t) k^b(x', a, t, Y_t, x) m(dx') da, \\ g^Y(x, a, 0) &= g_0^Y(x, a). \end{aligned}$$

**Modèle non-linéaire, limite en grande population**

Les non-linéarités apparaissent naturellement dans la microsimulation et les applications connexes, et cela est en fait le principal attrait pour ces méthodes : un ensemble d'outils de simulation flexibles permettant de faire des liens entre les individus, en d'autres termes d'inclure des *interactions*. Dans l'histoire du développement des modèles déterministes, l'extension des modèles standards à la prise en compte de non-linéarités a été mise en oeuvre pour corriger l'incapacité des premiers modèles à reproduire les types de croissances démographiques observées. En particulier dans le cas sans âge, Verhulst a étendu les équations de Malthus en faisant dépendre les taux de naissance et mort de toute la population, ce qui a ensuite poussé Gurtin et MacCamy à introduire des non-linéarités similaires dans le modèle en âge de McKendrick-Von Foerster. Par la suite, plusieurs travaux ont porté sur l'extension de ces modèles à la prise en compte des deux sexes dans la population, et d'inclure d'autres types de non-linéarités liant les taux de naissance des hommes et des femmes (voir par exemple Kendall (1949)). Plusieurs contributions ont par la suite amélioré la prise en compte des liens entre les populations des deux sexes, par exemple concernant l'étude de l'impact des équilibres hommes-femmes sur le nombre de mariages (voir par exemple Schoen (1983) et Guilmoto (2012)). Si les considérations démographiques se sont beaucoup attachées à la prise en compte des non-linéarités dans les taux de naissance individuels, il est clair qu'inclure ces non-linéarités au niveau des forces de mortalité semble crucial au regard de nombreuses applications, comme par exemple l'épidémiologie (voir par exemple Cléménçon et al. (2008) et les références qui y sont données).

**Comment modéliser ces interactions ?** Nous décrivons ici le cadre mathématique permettant d'inclure des interactions dans la dynamique. Nous nous inspirons ici des travaux de Fournier and Méléard (2004), Champagnat et al. (2006) et Tran (2008). Nous introduisons pour les individus d'une population  $Z$  des taux individuels de naissance, de mort, et de changement de caractéristiques qui dépendent de la population  $Z$  elle-même à travers un *noyau d'interaction*. Notons  $\mathbf{x} = (x, a)$  un élément de l'ensemble  $\mathbf{S} = E \times \mathbb{R}_+$ . Un noyau d'interaction est une fonction  $U : \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{R}$  où  $U(\mathbf{x}, \mathbf{x}')$  mesure l'impact d'un individu  $\mathbf{x}'$  sur l'individu  $\mathbf{x}$ . Ainsi, l'impact de toute la population  $Z(dx, da)$  sur l'individu  $\mathbf{x}$  s'écrit  $\langle Z, U_{\mathbf{x}} \rangle = \sum_{\mathbf{x}' \in Z} U(\mathbf{x}, \mathbf{x}')$ . Nous introduisons ici des noyaux d'interaction pour les trois types d'événements démographiques, notés  $U^b$ ,  $U^d$  et  $U^e$ . Les taux démographiques pour l'individu  $\mathbf{x} = (x, a)$  en vie dans la population  $Z_t$  au temps  $t$  sont alors  $b(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^b \rangle)$ ,  $d(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^d \rangle)$  et  $e(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^e \rangle)$ . La construction du processus de population comme solution à un problème de *thinning* est similaire, et les résultats d'existence forte et d'uni-

ité trajectorielle sont satisfaits dès lors que les taux démographiques individuels sont uniformément bornés, voir le point (i) de l'Hypothèse 2 ci-dessous. Pour établir l'asymptotique en grande population, et en particulier pour garantir l'unicité à l'Équation (1.23), nous avons besoin de conditions de Lipschitz pour les noyaux d'interaction, voir le point (ii) de l'Hypothèse 2 ci-dessous.

**Hypothèse 2.** (i) Il existe des constantes  $\bar{k}, \bar{b}, \bar{d}$  et  $\bar{e}$  telles que pour tous  $(x, a, t, y, x', u) \in \mathcal{X} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l \times \mathcal{X} \times \mathbb{R}$ ,  $k^e(x, a, t, y, x') \leq \bar{k}$ ,  $k^b(x, a, t, y, x') \leq \bar{k}$ ,  $b(x, a, t, y, u) \leq \bar{b}$ ,  $d(x, a, t, y, u) \leq \bar{d}$  et  $e(x, a, t, y, u) \leq \bar{e}$ .

(ii) Pour tout type d'événement démographique  $r \in \{b, d, e\}$ , il existe une constante  $L_r$  telle que pour tous  $(x, a, t, y) \in \mathcal{X} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$  et  $(u, v) \in \mathbb{R}^2$ ,

$$|r(x, a, t, y, u) - r(x, a, t, y, v)| \leq L_r |u - v|.$$

L'étude des asymptotiques en grande population repose sur un paramètre d'échelle  $n \in \mathbb{N}^*$  que l'on fait tendre vers l'infini. Ce paramètre d'échelle est utilisé comme ordre de grandeur pour la taille de la population initiale, et également pour diminuer le poids de chaque individu de sorte à observer des dynamiques non-dégénérées à la limite. On note  $\Gamma(\mathbf{S})$  l'espace des mesures ponctuelles finies sur  $\mathbf{S} = \mathcal{X} \times \mathbb{R}_+$ . Pour  $n \in \mathbb{N}^*$ , on introduit l'espace des mesures finies renormalisées

$$\Gamma^n(\mathbf{S}) = \left\{ \frac{1}{n} \nu \mid \nu \in \Gamma(\mathbf{S}) \right\},$$

et on considère une variable aléatoire  $Z_0^n$  à valeur dans  $\Gamma^n(\mathbf{S})$ . On fait apparaître ici le changement de poids de chaque individu ; des hypothèses seront faites sur la suite des mesures initiales  $(Z_0^n)_n$  pour traduire le fait que l'ordre de grandeur de la taille de l'échantillon initial  $N_0^n = n \langle Z_0^n, \mathbf{1} \rangle$  est  $n$ .

Avant de donner cette hypothèse de convergence des populations initiales, nous clarifions l'espace des mesures et leur topologie. Introduisons  $\mathcal{M}(\mathbf{S})$  l'espace des mesures positives finies sur  $\mathbf{S}$  ; notons que pour tout  $n \geq 1$ ,  $\Gamma^n(\mathbf{S}) \subset \mathcal{M}(\mathbf{S})$ , de sorte que toutes les populations renormalisées vivent dans le même espace. L'espace  $\mathcal{M}(\mathbf{S})$  peut être muni de la topologie de la convergence vague ou de celle de la convergence étroite. La convergence étroite dans  $\mathcal{M}(\mathbf{S})$  est définie par  $\nu_n \xrightarrow{weak} \nu$  si  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$  pour toute fonction  $f$  continue et bornée. La convergence vague est définie par  $\nu_n \xrightarrow{vague} \nu$  si  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$  pour toute fonction  $f$  continue à support compact. Notons que si  $\mathbf{S}$  était compact, les deux topologies seraient équivalentes, mais ici comme  $\mathbf{S} = \mathcal{X} \times \mathbb{R}_+$ , celles-ci sont strictement incluses.

Dans la suite, on note  $\mathbb{D}([0, T], (\mathcal{M}(\mathbf{S}), w))$  (resp.  $\mathbb{C}([0, T], (\mathcal{M}(\mathbf{S}), w))$ ) l'espace des processus càdlag (resp. continus) sur  $[0, T]$  à valeur dans  $\mathcal{M}(\mathbf{S})$  muni de la topologie de la convergence étroite.

**Hypothèse 3.** (Convergence de la suite des populations initiales)

Il existe  $\xi_0 \in \mathcal{M}(\mathbf{S})$  tel que la suite de mesures aléatoires  $(Z_0^n)_n$  converge en loi et

pour la topologie de la convergence étroite dans  $\mathcal{M}(\mathbf{S})$  vers la mesure déterministe  $\xi_0$ .

Sous l'hypothèse précédente, on a en particulier la convergence de  $\langle Z_0^n, \mathbf{1} \rangle = nN_0^n$  vers  $\langle \xi_0, \mathbf{1} \rangle$ , ce qui traduit le fait que la taille de l'échantillon initial est d'ordre  $n$ . La dernière hypothèse requise est une hypothèse de moments sur la suite des populations initiales :

**Hypothèse 4.** (Moments d'ordre strictement plus grand que 1) Il existe  $\alpha > 0$  tel que

$$\sup_{n \in \mathbb{N}^*} \mathbb{E} [\langle Z_0^n, \mathbf{1} \rangle^{1+\alpha}] < +\infty.$$

La convergence en loi de la suite des processus de population renormalisés est donnée ci-dessous.

**Théorème 1.** Sous les Hypothèses 2, 3 et 4, la suite  $(Z^n)_{n \in \mathbb{N}^*}$  converge en loi dans  $\mathbb{D}([0, T], (\mathcal{M}(\mathbf{S}), w))$  vers l'unique solution déterministe  $\xi \in \mathbb{C}([0, T], (\mathcal{M}(\mathbf{S}), w))$  de l'équation non-linéaire : pour toute  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$\langle \xi_t, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \xi_s, \partial_a f + \Phi_s^f(\xi_s) \rangle ds, \quad (1.23)$$

avec

$$\begin{aligned} \Phi_t^f(\chi)(x, a) &= \partial_a f(x, a) + b(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^b \rangle) \int_{\mathcal{X}} f(x', 0) k^b(x, a, t, Y_t, x') m(dx') \\ &\quad - d(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^d \rangle) f(x, a) + e(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^e \rangle) \int_{\mathcal{X}} (f(x', a) - f(x, a)) k^e(x, a, t, Y_t, x') m(dx'). \end{aligned}$$

### 1.2.5 Illustrations numériques

#### Algorithme pour la simulation

L'algorithme de simulation du processus de population est décrit ci-dessous. Celui-ci est discuté en détail dans le Chapitre 2. On pose pour la suite  $\bar{\lambda} = \bar{b}\bar{k} + \bar{d} + \bar{e}\bar{k}$ .

**Algorithme 1.** La population en  $T = 0$  est  $Z_0$  de taille  $N_0$ .

**Étape 1.** Simuler une v.a.  $\tau \sim \text{Exp}(\bar{\lambda}N_T)$  ;  $T + \tau$  est le prochain temps potentiel d'événement.

**Étape 2.** Entre  $T$  et  $(T + \tau)^-$ , la population est seulement modifiée par le vieillissement : on construit la population  $Z_{(T+\tau)^-}$  en ajoutant  $\tau$  à tous les âges dans  $Z_T$ .

**Étape 3.** Sélectionner un individu  $i$  uniformément dans la population  $Z_T$  et générer  $x \sim m(dx)$ .

**Étape 4.** Générer une v.a.  $\theta$  uniformément dans  $[0, \bar{\lambda}N_T]$ .

**Étape 5.** Déterminer si un événement démographique a lieu ou non :

(i) Si  $\theta \leq \lambda_b(Z_{(T+\tau)^-}, T + \tau, i, x)$  (naissance), construire la population  $Z_{T+\tau}$  en

ajoutant  $(x', 0)$  à  $Z_{(T+\tau)^-}$ .

(ii) Si  $\lambda_b(Z_{(T+\tau)^-}, T + \tau, i, x) < \theta \leq (\lambda_b + \lambda_d)(Z_{(T+\tau)^-}, T + \tau, i, x)$  (décès), construire la population  $Z_{T+\tau}$  en supprimant  $(X_i, A_i(T + \tau))$  dans  $Z_{(T+\tau)^-}$ .

(iii) Si  $(\lambda_b + \lambda_d)(Z_{(T+\tau)^-}, T + \tau, i, x) < \theta \leq (\lambda_b + \lambda_d + \lambda_e)(Z_{(T+\tau)^-}, T + \tau, i, x)$  (swap), construire la population  $Z_{T+\tau}$  comme  $Z_{(T+\tau)^-}$  dans laquelle  $(X_i, A_i(T + \tau))$  a été remplacé par  $(x, A_i(T + \tau))$ .

(iv) Sinon, aucun événement démographique ne se produit.

Poser  $T \leftarrow T + \tau$  et aller à l'Étape 1.

### Un exemple d'effet cohorte

Les bases de données de mortalité sont pour la plupart fournies par les organismes nationaux et sont basées sur la population dans son ensemble. En général, ces données regroupent des estimations de taux de mortalité estimés par tranche d'âges et par plages d'années, en distinguant les populations hommes ou femmes. L'évolution dans le temps de ces indicateurs agrégés sont d'intérêt pour les politiques publiques et le secteur de l'assurance par exemple. Dans ce contexte, une "amélioration de la longévité" concerne une évolution à la baisse de ces indicateurs sur une période de temps. Néanmoins, on peut se demander si ces améliorations sont aussi vraies à l'échelle des sous-groupes qui composent la population nationale, et également si les dynamiques hétérogènes ne peuvent pas expliquer en partie un certain nombre de variations. Nous souhaitons illustrer ces effets sur un exemple simple, en développant en parallèle le cas de l'effet cohorte au Royaume-Uni.

La génération appelée *Golden Cohort* est celle des personnes nées entre 1925 et 1945 en Angleterre. Celle-ci a connu des taux d'amélioration relative de mortalité particulièrement élevés, voir Figure 1.13 reprise de Cairns et al. (2009). Plusieurs explications possibles ont été avancées par Willets (2004) : par exemple l'impact de la Seconde Guerre Mondiale sur les générations précédentes, la consommation de tabac pour les générations suivantes, ou l'impact positif pour la *Golden Cohort* des réformes sociales d'après-guerre. Si une partie de ces effets liés aux modifications des durées de vie a pu avoir un impact sur les variations observées, il est intéressant de noter la remarque de Willets (2004) concernant le jeu de la fertilité, basée sur l'observation de la Figure 1.14 :

*One possible consequence of rapidly changing birth rates is that the 'average' child is likely to be different in periods where birth rates are very different. For instance, if trends in fertility vary by socio-economic class, the class mix of a population will change.*



## 1.2. MODÉLISATION DYNAMIQUE D'UNE POPULATION

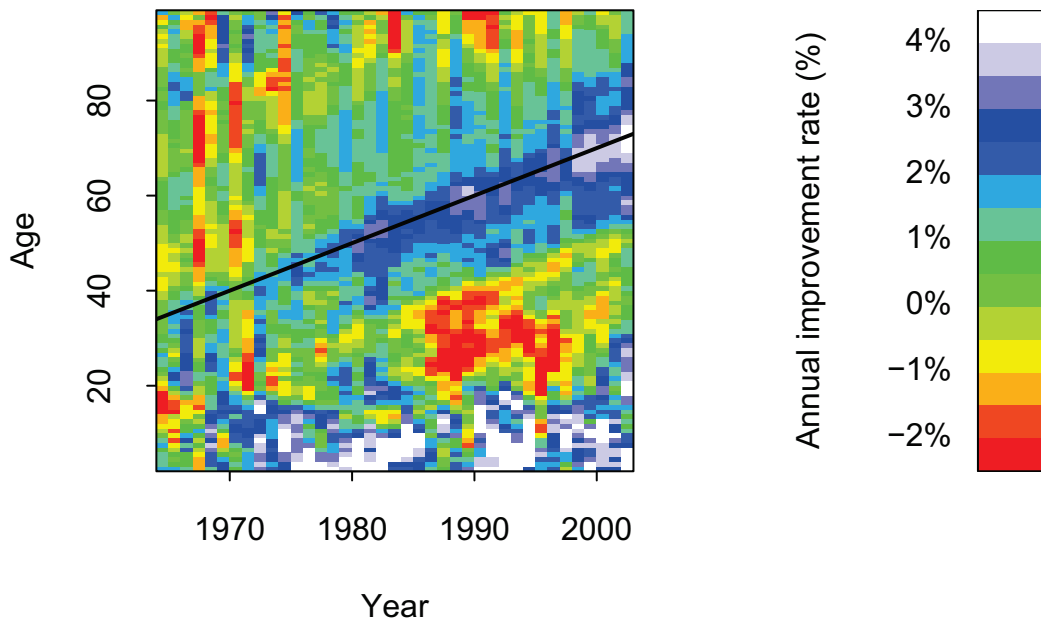


FIGURE 1.13. Graphique repris de Cairns et al. (2009). Données pour l'Angleterre et pays de Galles: l'amélioration relative des formes de mortalité pour chaque âge entre une année donnée et celle de l'année qui précède. La diagonale en noir suit la génération des personnes nées en 1930.

**2.1.2 The cohort effect** Pour illustrer de tels effets, nous proposons un modèle de population structuré en deux classes.

(i) On se donne un taux de mortalité de référence  $\bar{d}(a) = A \exp(Ba)$ , avec les paramètres  $A \sim 0.0004$  et  $B \sim 0.013$  qui reproduisent grossièrement les taux de mortalité de la période d'intérêt.

(ii) On définit un premier groupe avec des taux de décès indépendants du temps  $d_1(a) = \bar{d}(a)$  et un taux de naissance  $b_1(a) = c \mathbf{1}_{[20,40]}(a)$  ( $c=0.1$ )

(iii) On définit un second groupe avec un taux de décès indépendant du temps  $d_2(a) = \alpha \bar{d}(a)$  et un taux de naissance  $b_2(a) = c \mathbf{1}_{[20,40]}(a)$  ( $c=0.1$ )

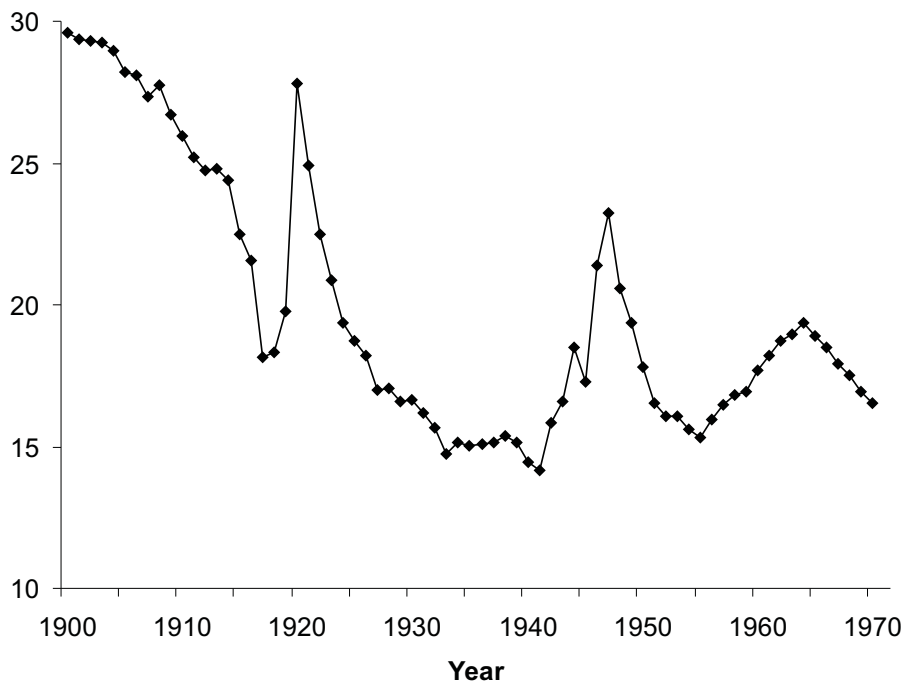
On suppose dans ce modèle une hérédité parfaite, de sorte que chaque individu

donne naissance à des personnes du même groupe que lui. There are other ways to illustrate the cohort effect and these can be found in Appendix A.

de ces variations de fertilité sur l'espérance de vie à la naissance, calculée de façon agrégée. La simulation est basée sur un échantillon de 20000 individus répartis uniformément entre les groupes 1 et 2. Les résultats sont représentés en Figure 1.15.

<sup>¶</sup>L'objectif de cette première expérience est de voir si ce cadre de modélisation est capable de reproduire ces faits stylisés. À cette fin, les paramètres sont fixés à des valeurs arbitraires qui donnent ici de premiers résultats intéressants. Cette direction de recherche constitue un travail en cours.

FIGURE 1.14 – Taux de naissance pour 1000 individus, pour Angleterre et pays de Galles, repris de Willets (2004)



Data source: www.mortality.org

Figure 6. Crude birth rate per 1,000 population, England and Wales, 1900 to 1970

Sur le graphique de gauche, on observe comme attendu que le nombre total de naissances dans la population est réduit entre les années 10 et 20 de projection. Dans le même temps, on observe sur le graphique de droite un effet cohorte notable que l'on observe sur le graphique de gauche (i.e. 1925-45) coincider avec a trough of births relative to periods immediately before and after. The fact that the birth rate increased dramatically between 1941 and 1944 (by 31%) may be particularly relevant.

One possible consequence of rapidly changing birth rates is that the 'average' child is likely to be different in periods where birth rates are very different. For instance, if trends in fertility vary by socio-economic class, the class mix of a population will change.

En retour, cela nous indique comment les taux de naissance donnent de l'information sur les variations de mortalité agrégée.

## 7. U.K. TRENDS BY CAUSE OF DEATH

Ces premiers résultats amorcent des perspectives intéressantes, notamment parce qu'ils soulignent l'importance de l'ensemble de la dynamique de population.

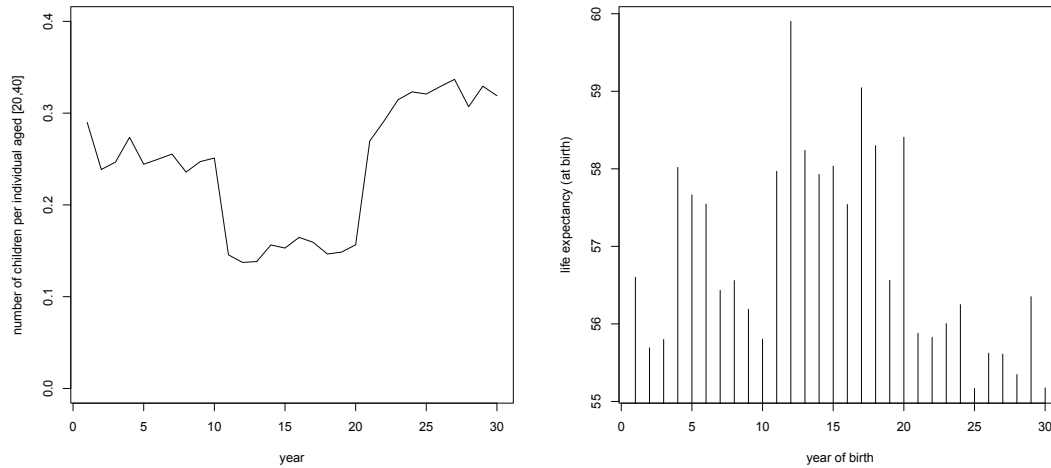
### 7.1 Cohort Trends by Cause of Death

7.1.1 Further insight into the potential causes of the U.K. cohort effect can be obtained through an investigation of mortality trends for specific causes of death.

Les études actuarielles qui évaluent les risques de longévité (mortalité pour les compagnies d'assurance ou la question des retraites se concentrent traditionnellement sur les âges à partir de 30 ou 40 ans. Ceci est de plus encouragé par le fait que les taux de mortalité ont une structure en âge plus complexe avant 30 ans environ, voir Figure 1.3, qui inclut la mortalité infantile et la bosse à accidents. Cette approche met de côté l'information cruciale sur les naissances qui déterminent la composition initiale des différentes générations en terme de caractéristiques, comme nous l'illustrons ici,

ou également sur la répartition des naissances dans l'année, voir la discussion dans la Partie 1.1.2.

FIGURE 1.15 – Une trajectoire du processus de population. À gauche : nombre d'enfants par individu d'âge dans [20-40]. À droite : estimation de l'espérance de vie à la naissance.



### Bruit d'environnement et aléa démographique

Nous venons d'illustrer sur un exemple simple l'impact de taux démographiques hétérogènes sur la mortalité agrégée. Nous souhaitons illustrer ici l'impact de deux sources majeures de bruit qui interviennent dans le champ de la microsimulation : le bruit d'environnement et l'aléa démographique. On se place dans le cas d'une population de taille initiale  $N_0 \in \mathbb{N}^*$  dans laquelle les seuls événements démographiques sont les décès. Chaque individu au temps  $t$  a un taux de décès  $\mu(Y_t) = Y_t$ , où  $Y$  est un processus continu et positif. Dans cet exemple, l'environnement est identifié à la force de mortalité stochastique de chaque individu dans la population, que l'on peut voir comme une cohorte qui décède au cours du temps. On se donne  $Q(ds, d\theta)$  une mesure ponctuelle de Poisson indépendante de  $Y$  et de mesure intensité  $q(ds, d\theta) = dsd\theta$  sur  $\mathbb{R}_+ \times \mathbb{R}_+$ . On note  $N_t$  le nombre d'individus encore en vie dans la population, défini comme solution de l'équation

$$N_t = N_0 - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq Y_s N_{s-}} Q(ds, d\theta). \quad (1.24)$$

La décomposition classique de la variance nous permet d'écrire

$$\text{Var}(N_t) = \mathbb{E}[\text{Var}(N_t | Y)] + \text{Var}(\mathbb{E}[N_t | Y]). \quad (1.25)$$

La première composante  $\mathbb{E}[\text{Var}(N_t^Y | Y)]$  est souvent appelée stochasticité démographique : celle-ci correspond à l'aléa dû à la taille finie de l'échantillon, moyenné sur tous les scénarios possibles d'environnement. La deuxième composante  $\text{Var}(\mathbb{E}[N_t | Y])$  correspond au bruit d'environnement, qui subsiste seul lorsque la taille de l'échantillon tend vers l'infini. Il faut noter que cette décomposition n'est pas unique ; nous aurions pu conditionner par rapport à l'information donnée par la mesure ponctuelle de Poisson  $Q$ . Dans ce cas cependant, les calculs apparaissent complexes, car il s'agit d'évaluer la loi de l'environnement conditionnellement à un scénario démographique. Calculons maintenant séparément chaque composante de l'Équation (1.25).

(i) Le bruit d'environnement s'écrit  $\text{Var}(\mathbb{E}[N_t | Y]) = \mathbb{E}[\mathbb{E}[N_t | Y]^2] - \mathbb{E}[\mathbb{E}[N_t | Y]]^2$ . Notons  $g^Y(t) := \mathbb{E}[N_t | Y]$ . En prenant l'espérance conditionnelle dans (1.24), nous obtenons  $g^Y(t) = N_0 - \int_0^t Y_s g^Y(s) ds$ , dont la solution est donnée par  $g^Y(t) = N_0 \exp\left(-\int_0^t Y_s ds\right)$ . Ainsi, le bruit d'environnement s'écrit

$$\text{Var}(\mathbb{E}[N_t | Y]) = N_0^2(B_t - A_t^2),$$

avec  $A_t = \mathbb{E}\left[e^{-\int_0^t Y_s ds}\right]$  et  $B_t = \mathbb{E}\left[e^{-2\int_0^t Y_s ds}\right]$ .

(ii) Calculons maintenant la partie correspondant à l'aléa démographique. D'abord,

$$\mathbb{E}[\text{Var}(N_t | Y)] = \mathbb{E}[\mathbb{E}[N_t^2 | Y]] - N_0^2 B_t.$$

Calculons le terme  $h^Y(t) := \mathbb{E}[N_t^2 | Y]$ . On peut appliquer le calcul différentiel à  $N$  pour obtenir  $N_t^2 = N_0^2 + \int_0^t \int_{\mathbb{R}_+} ((N_{s-} - 1)^2 - N_{s-}^2) \mathbf{1}_{\theta \leq Y_s N_{s-}} Q(ds, d\theta)$ . En simplifiant les termes et en prenant l'espérance conditionnelle, il vient  $h^Y(t) = N_0^2 - 2 \int_0^t Y_s h^Y(s) ds + \int_0^t Y_s g^Y(s) ds$  ce qui donne

$$h^Y(t) = N_0(N_0 - 1)e^{-2\int_0^t Y_s ds} + N_0 e^{-\int_0^t Y_s ds}.$$

Ainsi, on obtient

$$\mathbb{E}[\text{Var}(N_t^Y | Y)] = N_0(A_t - B_t).$$

**Dynamique de l'environnement** On suppose que  $(Y_t)_t$  est une diffusion de Feller (ou un processus CIR) dont la dynamique est donnée par

$$dY_t = a(b - Y_t)dt + \rho\sqrt{Y_t}dW_t.$$

Le principal avantage de ce modèle est de proposer des taux de décès qui restent positifs. De nombreux autres exemples de modèles affines de mortalité stochastique ont été introduits par Biffis (2005). Nous souhaitons illustrer ici que ce cadre de modélisation est également intéressant pour étudier le risque idiosyncratique et le comparer au risque systématique. Dans cette dynamique,  $b$  est le paramètre de

retour à la moyenne ; dans cet exemple particulier, la quantité  $b$  représente le taux de mortalité cible aux âges très élevés. Le paramètre  $a$  représente la vitesse à laquelle le taux de mortalité tend vers sa valeur moyenne cible. Enfin, le paramètre  $\rho$  correspond à la volatilité du taux de mortalité. Le processus part d'une valeur  $y_0 < b$ , ce qui nous permet de traduire que les individus à l'âge zéro vieillissent au cours du temps et ont donc en moyenne un taux de mortalité qui croît. Cette moyenne du taux de mortalité vaut en effet  $\mathbb{E}[Y_t] = b + (y_0 - b)e^{-at}$  qui part de  $y_0$  et tend vers  $b$  en croissant, et sa variance s'exprime sous la forme  $\text{Var}(Y_t) = \frac{\rho^2}{a^2}(1 - e^{-at}) \left[ y_0 e^{-at} + \frac{b}{2}(1 - e^{-at}) \right]$ , qui part de zéro et tend vers  $\frac{\rho^2 b}{2a^2}$  en croissant (car  $y_0 < b$ ).

Ce cadre élémentaire va déjà nous permettre d'illustrer des équilibres intéressants entre bruit d'environnement et aléa démographique, dus à deux effets en compétition : au cours du temps,

- (i) la taille de la population décroît avec les décès, ce qui fait augmenter la part des fluctuations dues à la taille finie de la population,
- (ii) la variance du taux de décès lui-même augmente, ce qui est naturel pour modéliser le fait que des projections de long terme sont plus incertaines, vues d'aujourd'hui, que les valeurs à court-terme.

Rappelons ici les résultats classiques dont nous avons besoin concernant le calcul des quantités  $A_t$  et  $B_t$  en fonction des paramètres du modèle :

$$A_t = \mathbb{E} \left[ \exp \left( - \int_0^t Y_s ds \right) \right] = \Phi(\gamma, T) \exp(-y_0 \Psi(\gamma, T)),$$

avec  $\gamma = \sqrt{a^2 + 2\rho^2}$  et

$$\Psi(\gamma, T) = \frac{2(e^{\gamma T} - 1)}{(\gamma + a)(e^{\gamma T} - 1) + 2\gamma}, \quad \Phi(\gamma, T) = \left[ \frac{2\gamma e^{(\gamma+a)\frac{T}{2}}}{(\gamma + a)(e^{\gamma T} - 1) + 2\gamma} \right]^{\frac{2ab}{\rho^2}}.$$

Sur cette base, on peut en outre établir que

$$B_t = \mathbb{E} \left[ \exp \left( -2 \int_0^t Y_s ds \right) \right] = \Phi(\sqrt{\gamma^2 + 2\rho^2}, T) \exp \left( -2y_0 \Psi(\sqrt{\gamma^2 + 2\rho^2}, T) \right).$$

**Illustration numérique** Pour cette illustration numérique, nous fixons un taux de décès de départ à  $y_0 = 0.001$  (l'ordre de grandeur des taux de décès pour des individus âgés de 40 ans dans les pays de l'OCDE), ainsi qu'un taux de décès cible égal à  $b = 0.8$  (l'ordre de grandeur des taux de décès aux âges très élevés). Dans la suite, nous nous intéressons à l'impact du paramètre de volatilité  $\rho$  et de la vitesse de retour à la moyenne  $a$  sur la proportion du bruit d'environnement dans la variance totale. Nous définissons formellement cette proportion  $p_t$  pour  $t > 0$  par

$$p_t = \frac{\text{Var}(\mathbb{E}[N_t | Y])}{\text{Var}(N_t)}.$$

(i) Analysons d'abord l'impact de la volatilité  $\rho$ , et fixons une vitesse de retour à la moyenne  $a = 0.01$  pour cette première illustration. Sur la Figure 1.16, on représente le comportement de la variance du taux de décès en fonction du temps  $t$  et de la volatilité  $\rho$ . La proportion  $p_t$  de bruit d'environnement dans la variance totale est représentée en Figure 1.17 pour différentes tailles de populations initiales  $N_0$ . Comme attendu, lorsque l'on fait varier la taille  $N_0$  de 100 à 100 000, Figures (a) à (d), la proportion  $p_t$  croît, les paramètres  $t$  et  $\rho$  étant fixés. De même, pour un temps  $t$  et une taille  $N_0$  fixés, cette proportion est croissante en la volatilité. L'évolution en temps, qui se fait de manière complexe, illustre de manière intéressante l'équilibre entre les deux sources de risque : dans cette configuration de paramètres, la proportion commence par croître avec le temps, ce qui traduit le fait que l'incertitude due à l'environnement l'emporte, puis celle-ci se met à décroître, rendant compte de la vitesse plus importante des événements de décès, ce qui rend le bruit d'environnement moins important sur le long terme dans la plupart des configurations.

(ii) Observons maintenant l'impact de la vitesse  $a$  de retour à la moyenne, et fixons  $\rho = 0.005$  pour cette illustration. Son effet sur la variance du taux de décès est représenté sur la Figure 1.16 ; comme attendu, le bruit d'environnement augmente lorsque la vitesse de retour à la moyenne décroît. On représente sur la Figure 1.18 l'impact du paramètre  $a$  sur la proportion  $p_t$  à différentes dates  $t$ . Dans ce cas, l'augmentation de la taille  $N_0$  initiale crée des effets similaires au cas précédent, et également, la proportion du bruit d'environnement s'accroît lorsque le paramètre  $a$  diminue. On retrouve sur cette figure le comportement intéressant en temps pour des paramètres  $a$  dans une zone déterminée. Sur le moyen terme, le bruit d'environnement est important, mais sur les court et long termes, c'est bien l'aléa démographique qui l'emporte. Notons tout de même que pour de faibles valeurs de vitesse de retour à la moyenne, l'environnement constitue la source majeure de risque.

Sur cet exemple simple, nous avons mis en avant la complexité de l'équilibre des sources de risque, et en particulier que l'analyse de celles-ci dépend fortement de l'horizon et des années d'intérêt. Cela montre également que les valeurs de paramètres et de la taille initiale sont cruciales et que celles-ci donnent des informations *a priori* sur les ordres de grandeur de ces risques.

FIGURE 1.16 – Effet conjoint du temps  $t$  et, ou bien la volatilité  $\rho$  avec  $a = 0.01$  (gauche), ou bien la vitesse  $a$  de retour à la moyenne avec  $\rho = 0.005$  (droite), sur la variance  $\text{Var}(Y_t)$  du taux de décès

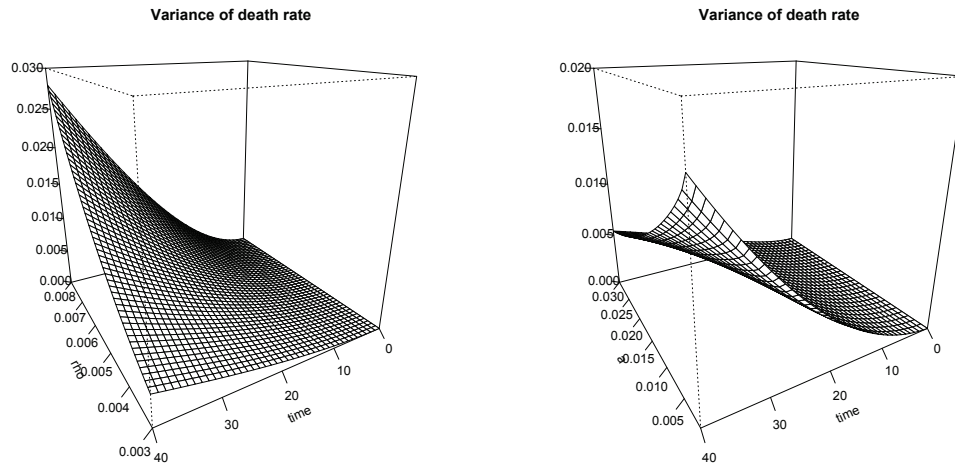


FIGURE 1.17 – Effet conjoint du temps  $t$  et de la volatilité  $\rho$  sur la proportion  $p_t$  de bruit d'environnement dans la variance totale

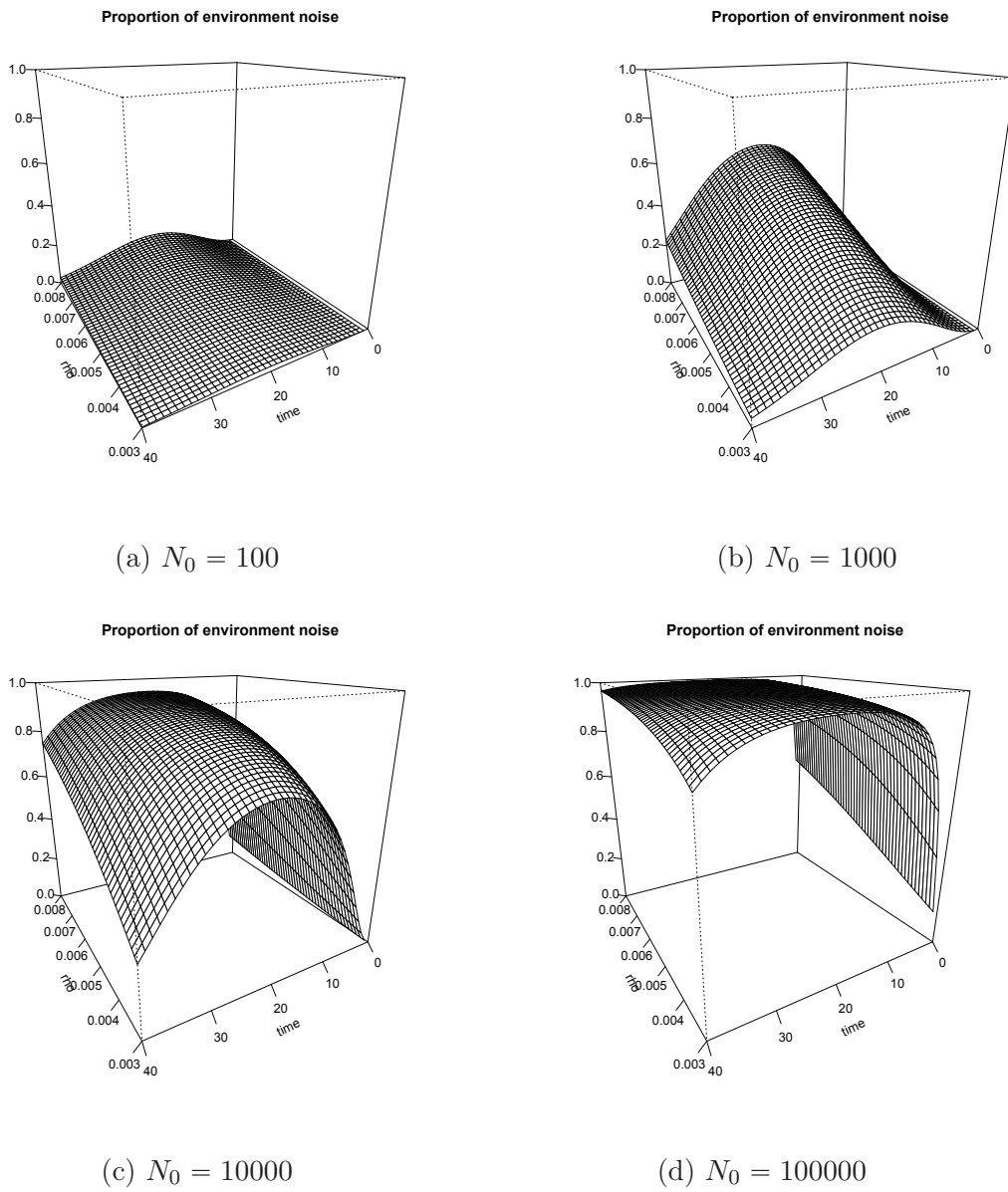
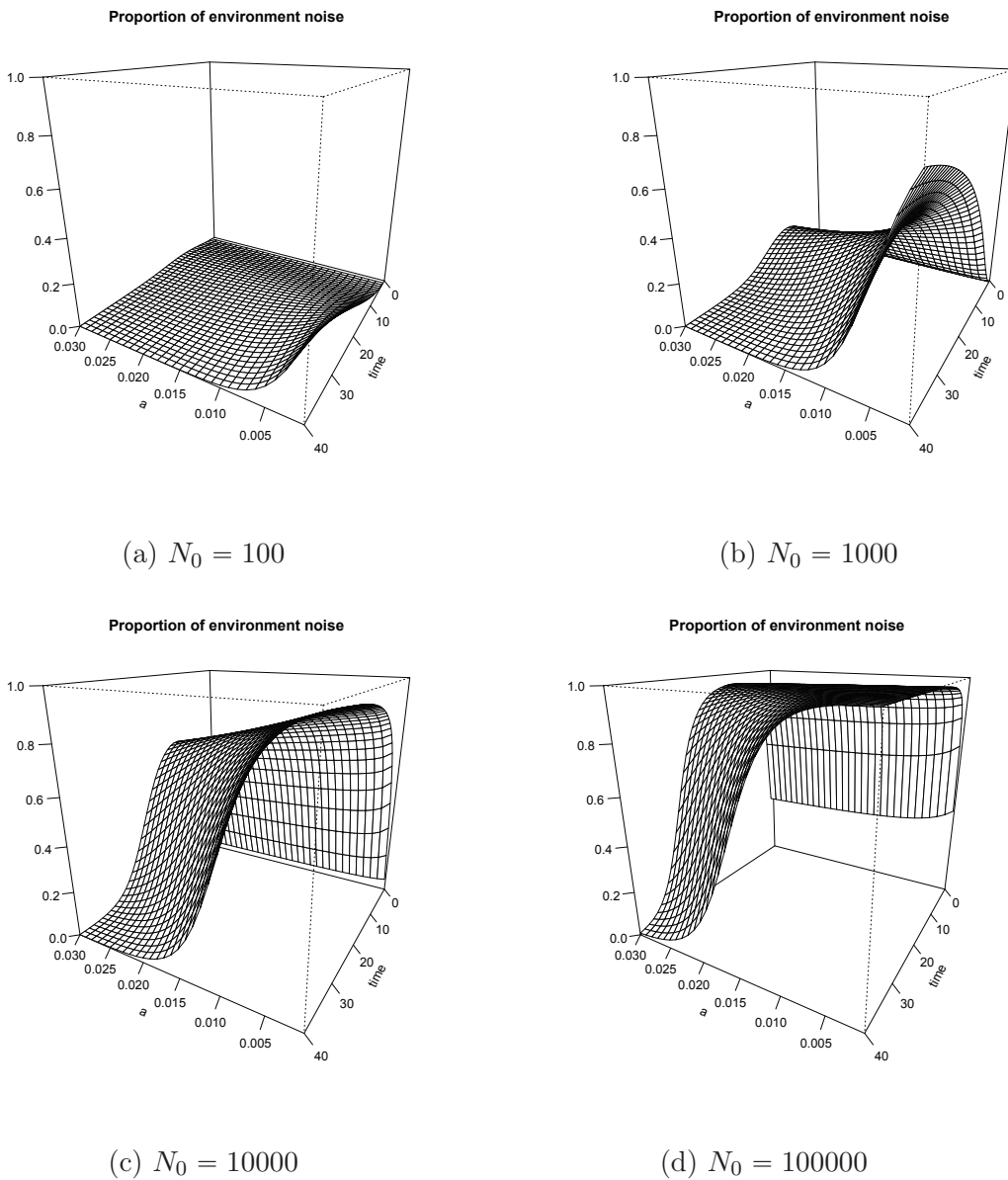




FIGURE 1.18 – Effet conjoint du temps  $t$  et de la vitesse  $a$  de retour à la moyenne sur la proportion  $p_t$  de bruit d'environnement dans la variance totale



### 1.2.6 Conclusion et perspectives

Les modèles microscopiques de dynamique des populations nous donnent un cadre mathématique de référence pour analyser les dynamiques démographiques complexes. La construction des processus de population par la représentation *thinning* fournit un outil mathématique flexible pour traiter telles dynamiques, et la méthode de simulation associée fait le lien avec le champ des méthodes de microsimulation qui simulent l'évolution au cours du temps de populations à l'échelle de l'individu. Les premières expériences numériques associées à ce travail confirment d'une part

l'importance de la dynamique hétérogène pour la compréhension des variations des indicateurs agrégés (déclinée ici sous la forme d'un *effet cohorte*), et d'autre part la nécessité de prendre en compte les différentes sources de risque que sont le bruit d'environnement et la stochasticité démographique. Ce qui ressort du cadre de modélisation et de ces premières expériences est que la structure de la population, c'est-à-dire sa pyramide des âges et son hétérogénéité, sont des composantes cruciales qui permettent de calculer les indicateurs agrégés; ceci inclut en particulier les tables de mortalité nationales.

Ce travail de revue de la modélisation fait apparaître un ensemble de directions de recherche; nous proposons dans les parties suivantes et les chapitres associés de décliner ce cadre de modélisation général afin d'étudier des problèmes spécifiques.

(i) D'abord, la formalisation du vieillissement dans la dynamique de pyramide des âges, voir l'exemple de l'Équation (1.17), est à exploiter d'un point de vue théorique. Ceci sera abordé dans la Partie 1.3 et le Chapitre 3 dans l'étude des dynamiques d'immigration-naissance qui sous-tendent les processus de Hawkes.

(ii) Une autre direction de recherche concerne la possibilité de simuler des pyramides des âges en temps continu, et d'étudier le modèle déterministe associé. La Partie 1.4 et le Chapitre 4 correspondant, basés sur des données de mortalité par causes de décès, proposent d'étudier la dynamique de pyramide des âges en cas de réduction ou de suppression de certaines causes de décès.

(iii) L'estimation statistique d'une dynamique de population complète à partir de données démographiques est un problème difficile. Ce que nous proposons dans la Partie 1.5 et le Chapitre 5 associé consiste en l'estimation statistique d'un sous-modèle de population, qui modélise les trajectoires de vie d'individus soumis à des taux de décès par caractéristique et des intensités de changement de ces caractéristiques. Au-delà, l'estimation de taux de naissance hétérogènes et de lois de mutation à la naissance est un enjeu crucial; malheureusement, peu de données sont aujourd'hui disponibles pour traiter cette question.

(iv) Enfin, les outils de théorèmes limites sur les processus, qui servent ici à établir les approximations en grande population des dynamiques démographiques, permettent d'inclure diverses renormalisations. Il est en particulier possible de séparer les échelles de temps en faisant tendre vers l'infini certains taux démographiques à l'aide du paramètre de renormalisation. Ceci sera développé dans la Partie 1.6 et le Chapitre 6 correspondant dans l'étude des dynamiques macroscopiques de populations dans lesquelles les individus changent fréquemment de caractéristiques.

D'autres orientations de recherche seraient à envisager. Ceci inclut par exemple l'étude d'impact de possibles non-linéarités dans le modèle, en particulier la quantification de l'écart entre la trajectoire moyenne et la dynamique de l'approximation en grande population, ainsi que l'estimation statistique de ces non-linéarités sur

la base de données démographiques. Aussi, le développement de méthodes numériques avancées pour de tels modèles est un enjeu important ; on peut par exemple penser à l'utilisation de méthodes type *importance sampling* pour l'estimations de probabilités d'événements rares (par exemple des scénarios de longévité extrêmes), et également à l'utilisation de l'approximation grande population en *variable de contrôle* pour accélérer la simulation du processus.

## 1.3 La pyramide des âges des processus de Hawkes

Les processus de Hawkes sont des processus de comptage qui constituent des exemples de processus auto-régressifs en temps continu. Leur évolution au cours du temps dépend de leur passé sous une forme qui modélise l'auto-excitation : si un événement a eu lieu il y a peu de temps, la probabilité de survenance d'un nouvel événement est augmentée d'une quantité, celle-ci étant décroissante au cours du temps ; on modélise ainsi un phénomène de mémoire. Les processus de Hawkes contiennent donc deux propriétés importantes (l'auto-excitation et la mémoire) qui en font des modèles naturels pour de nombreux champs d'application, incluant la sismologie, les neurosciences, l'épidémiologie, la finance et l'assurance, pour n'en citer que quelques uns. Une autre propriété intéressante de ces processus est leur interprétation en terme de processus de population avec âge, dans lequel des individus peuvent immigrer et donner naissance au cours du temps. Le modèle général présenté dans la partie précédente peut en fait être décliné sous cette forme. Sur cette base, la contribution de cette partie est l'introduction du concept de pyramide des âges pour les processus de Hawkes et son étude pour proposer de nouveaux résultats pour une classe de ces processus.

### 1.3.1 Les processus de Hawkes

Hawkes (1971) a introduit un processus de comptage  $(N_t)$  dont l'intensité est de la forme suivante, appelée *shot noise* :

$$\lambda_t = \mu + \sum_{T_n < t} \phi(t - T_n), \quad (1.26)$$

où les  $T_n$  sont les temps de saut du processus  $N$  lui-même,  $\mu > 0$  et  $\phi$  est une fonction positive. Nous appelons dans la suite ce processus le processus de Hawkes standard. Dans ce modèle, lorsqu'un événement se produit au temps  $T_n$ , l'intensité augmente d'une quantité  $\phi(t - T_n)$  : ceci modélise la propriété d'auto-excitation. En outre, dans la plupart des applications, la fonction  $\phi$  tend vers zéro, ce qui fait que l'excitation s'estompe avec le temps et que le processus présente un phénomène de mémoire. Dans la littérature, on trouve des contributions qui se sont intéressées à des processus de Hawkes présentant en plus de l'auto-excitation une composante d'excitation externe. À notre connaissance, les processus de Hawkes avec immigrants généraux ont été introduits par Brémaud and Massoulié (2002), et l'on trouve des formes particulières de ce processus général dans des études récentes motivées par des applications financières, notamment Dassios and Zhao (2011), Wheatley et al. (2014) et Rambaldi et al. (2014). Dans ces études, le processus permet de prendre en compte des chocs exogènes, l'arrivée d'informations sur les marchés financiers ou le phénomène de contagion. Dans cette partie, nous nous intéressons à une classe

de processus de Hawkes avec immigrants généraux (voir Brémaud and Massoulié (2002)), dont l'intensité est de la forme

$$\lambda_t = \mu(t) + \sum_{T_n < t} \Phi_t(t - T_n, X_n) + \sum_{S_k < t} \Psi_t(t - S_k, Y_k).$$

Dans ce modèle, les  $T_n$  sont les temps de saut de  $N$  : si un événement a eu lieu pour le système à la date  $T_n$ , l'intensité augmente d'une quantité  $\Phi_t(t - T_n, X_n)$ , où  $X_n$  est une marque aléatoire. Cette partie modélise l'auto-excitation. En parallèle, des chocs externes surviennent aux temps  $S_k$  et accroît l'intensité d'une quantité  $\Psi_t(t - S_k, Y_k)$  : ceci constitue la composante d'excitation externe.

Parmi les propriétés intéressantes des processus de Hawkes, l'une d'elles est basée sur l'écriture de l'intensité sous la forme *shot noise*. Il s'agit de la représentation en cluster, autrement dit en terme de processus de branchement. Elle utilise la remarque suivante : si un événement a eu lieu au temps  $T_n$ , alors  $t - T_n$  n'est rien d'autre que l'âge de cet événement au temps  $t$ . Quelques années après l'innovation de Hawkes (1971), Hawkes and Oakes (1974) ont proposé la représentation en cluster du processus auto-excitant. Dans ce papier, il est montré que sous des conditions de stationnarité, le processus de Hawkes peut être représenté comme un cluster de Poisson, qui est en fait un processus d'immigration-naissance avec âge. On peut trouver une définition sous la forme de processus de cluster par exemple dans Dassios and Zhao (2011) dans le cas où le processus possède également une composante d'excitation externe.

Cette représentation en terme de processus de naissance avec immigration se décrit à partir de la forme de l'intensité (1.26) de la façon suivante : des immigrants arrivent au temps d'un processus de Poisson de paramètre  $\mu$ . Chaque immigrant donne ensuite naissance à des individus avec un taux de naissance  $\phi(a)$  où  $a$  est l'âge de l'immigrant, puis chaque individu donne à son tour naissance à d'autres individus avec le même taux de naissance  $\phi$ . La représentation en cluster de Hawkes and Oakes (1974) requière que le nombre moyen d'enfants par individu qui est en fait  $\|\phi\| = \int_0^\infty \phi(a) da$  vérifie  $\|\phi\| < 1$ . Une première contribution de ces travaux de thèse sur les processus de Hawkes, détaillée dans le Chapitre 3, est de présenter le cadre non-stationnaire permettant de construire le processus d'immigration avec âge correspondant. Nous l'utiliserons ensuite pour étudier la dynamique d'une classe de processus de Hawkes et quelques propriétés statistiques (transformées de Laplace, moments) qui généralisent des résultats récents.

Dans la littérature, les propriétés statistiques des processus de Hawkes ont été étudiées différemment dans le cadre stationnaire d'une part, et non stationnaire d'autre part.

(i) Dans le cas stationnaire, Hawkes (1971) a étudié les moments d'ordre deux, tandis qu' Adamopoulos (1975) a étudié la fonction génératrice à l'aide de la représentation

en cluster stationnaire de Hawkes and Oakes (1974). Une contribution importante a été l'introduction par Brémaud and Massoulié (2002) du spectre de Bartlett pour calculer les moments du processus de Hawkes stationnaire. Nous mentionnons également deux contributions récentes qui utilisent la représentation de Hawkes and Oakes (1974) : celui de Saichev and Sornette (2011) sur la fonction génératrice des moments et Jovanović et al. (2014) sur une méthode graphique pour le calcul des cumulants (des quantités équivalentes aux moments).

(ii) La caractérisation de la distribution des processus de Hawkes dans le cas non-stationnaire a eu un regain d'intérêt dans la littérature récente. Les techniques utilisées diffèrent du cadre stationnaire, et en particulier ne reposent pas sur la représentation en cluster. En fait, la technique commune à ces contributions récentes (voir Errais et al. (2010), Aït-Sahalia et al. (2010), Dassios and Zhao (2011), et Da Fonseca and Zaatour (2014)) est l'utilisation du fait que lorsque la fonction de fertilité  $\phi$  possède la forme particulière  $\phi(a) = \alpha e^{\beta a}$ , le processus intensité  $\lambda_t$  est Markovien, ce qui avait été établi par Oakes (1975). La compréhension de la dynamique de la pyramide des âges va nous permettre d'étudier les propriétés d'une classe de processus de Hawkes plus générale, pour laquelle le processus d'intensité n'est pas Markovien.

Dans la suite, nous détaillons les deux premières des trois principales contributions concernant les processus de Hawkes avec immigrants généraux : (i) le concept de pyramide des âges et sa dynamique, (ii) l'utilisation de cette pyramide des âges pour calculer la distribution de processus de Hawkes généraux pour une classe de fonctions de fertilité qui généralise le cas exponentiel (iii) la représentation trajectorielle des processus de Hawkes avec immigrants généraux et du processus de naissance avec immigration sous-jacent, qui découle du modèle général présenté dans la Partie 1.2 précédente et sera détaillée dans le Chapitre 3. Nous présentons d'abord cette démarche pour le processus de Hawkes standard et étendons ce cadre aux processus de Hawkes avec immigrants généraux dans un deuxième temps.

### 1.3.2 La pyramide des âges des processus de Hawkes

La définition du processus de Hawkes standard est donnée ci-dessous. Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité satisfaisant les conditions habituelles. On rappelle que le processus d'intensité  $(\lambda_t)$  d'un processus de comptage  $(N_t)$  est le processus  $(\mathcal{F}_t^N)$ -prévisible tel que  $N_t - \int_0^t \lambda_s ds$  est une  $(\mathcal{F}_t^N)$ -martingale locale, où  $(\mathcal{F}_t^N)$  est la filtration canonique associée à  $(N_t)$ .

**Définition 1.** *Soit  $\phi$  une fonction continue et positive. Un processus de Hawkes  $(N_t)$  avec fonction de fertilité  $\phi$  est un processus de comptage avec filtration canonique  $(\mathcal{F}_t^N)$  qui admet une intensité de la forme*

$$\lambda_t = \mu + \sum_{T_n < t} \phi(t - T_n) = \mu + \int_{(0,t)} \phi(t - s) dN_s, \quad (1.27)$$

où  $\mu > 0$ , et les  $(T_n)$  sont les temps de saut de  $(N_t)$ .

Cette définition donne la représentation de l'intensité en fonction du processus lui-même, mais en réalité la majeure partie de l'information sur la population est perdue lorsque l'on ne se base que sur cette forme. À ce titre, il est intéressant de revenir à la dynamique de populations sous-jacente : des immigrants arrivent dans la population au taux  $\mu$ , et chaque individu d'âge  $a$  donne naissance à de nouveaux individus avec un taux  $\phi(a)$ . Cette dynamique est cruciale pour l'étude des processus de Hawkes ; l'objectif maintenant est de garder trace de tous les âges dans la population. Une façon de le faire est de reprendre le cadre du modèle de dynamique des populations. Si l'on veut compter le nombre  $Z_t([0, \bar{a}])$  d'individus qui ont un âge plus petit que  $\bar{a} > 0$  au temps  $t$ , il suffit par exemple de regarder tous les individus qui sont arrivés jusqu'en  $t$ , et d'y soustraire ceux arrivés avant  $t - \bar{a}$ , soit

$$Z_t([0, \bar{a}]) = N_t - N_{t-\bar{a}} = \int_{(0,t]} \mathbf{1}_{t-\bar{a} \leq s} dN_s = \int_{(0,t]} \mathbf{1}_{t-s \leq \bar{a}} dN_s.$$

On retrouve ici le cadre du modèle général : à un instant  $t$ , on a défini une mesure que l'on appelle pyramide des âges et qui s'écrit

$$Z_t(da) = \int_{(0,t]} \delta_{t-s}(da) dN_s = \sum_{n=1}^{N_t} \delta_{t-T_n}(da). \quad (1.28)$$

Comme dans le cadre général, on peut calculer une fonction  $f$  sur toute la pyramide des âges par

$$\langle Z_t, f \rangle = \int_{\mathbb{R}_+} f(a) Z_t(da) = \int_{(0,t]} f(t-s) dN_s. \quad (1.29)$$

Par exemple, on retrouve le processus de Hawkes en calculant la taille de la population par  $N_t = \langle Z_t, \mathbf{1} \rangle$ . On retrouve également le processus d'intensité de l'Équation (1.27) par

$$\lambda_t = \mu + \langle Z_{t-}, \phi \rangle.$$

Dans la Partie 1.2, nous avons mentionné le fait que le processus à valeur mesure  $(Z_t)$  est un processus de Markov. Ceci suggère que toute l'information nécessaire est contenue dans cette pyramide des âges qui évolue au cours du temps. Malheureusement, cette information est trop vaste pour permettre de calculer des distributions du processus de Hawkes non-stationnaire dans le cas général. Dans la suite, nous montrons comment identifier un certain nombre de composantes minimales à ajouter au processus de Hawkes et à son intensité pour conserver une dynamique Markovienne. Avant cela, nous aurons besoin d'étudier la dynamique de la pyramide des âges. Celle-ci est présentée dans le lemme suivant, qui est un cas particulier de la dynamique (1.21) décrite dans la Partie 1.2.

**Lemme 1.** Pour toute fonction  $f \in \mathcal{C}^1(\mathbb{R}_+)$ ,

$$\langle Z_t, f \rangle = f(0)\langle Z_t, \mathbf{1} \rangle + \int_0^t \langle Z_s, f' \rangle ds. \quad (1.30)$$

Le premier terme de l'Équation (1.30) est un terme de saut pur qui traduit les arrivées d'individus d'âge 0 (naissance ou immigration), tandis que le second terme de transport illustre le phénomène de vieillissement au cours du temps. En particulier, on peut retrouver directement à partir de cette représentation que le processus d'intensité  $\lambda_t = \mu + \langle Z_{t-}, \phi \rangle$  est Markovien dans le cas où le taux de naissance est exponentiel, i.e.  $\phi(a) = \alpha e^{\beta a}$  (voir Oakes (1975)). En effet, dans ce cas  $\phi' = \beta\phi$ , et l'Équation (1.30) avec  $f \equiv \phi$  permet d'obtenir la dynamique

$$d\langle Z_t, \phi \rangle = \alpha dN_t + \beta \langle Z_t, \phi \rangle dt.$$

Comme  $dN_t$  ne dépend du passé de  $(\lambda_t)$  qu'à travers la valeur courante  $\lambda_t$ , cela montre la propriété de Markov.

### 1.3.3 La généralisation du cas exponentiel

Dans cette partie, nous présentons les principales contributions à l'étude des processus de Hawkes standards, en particulier le calcul des premiers moments et de la transformée de Laplace. La généralisation aux processus de Hawkes avec immigrants généraux sera présentée dans la Partie 1.3.4. Nous introduisons tout de suite les hypothèses sur le taux de naissance qui généralisent de manière naturelle le cas exponentiel.

#### Hypothèse sur le taux de naissance

**Hypothèse 5.** La fonction  $a \in \mathbb{R}_+ \mapsto \phi(a)$  est positive, de classe  $\mathcal{C}^n(\mathbb{R}_+)$ , et il existe  $c = (c_{-1}, \dots, c_{n-1}) \in \mathbb{R}^{n+1}$  tel que  $\phi$  vérifie

$$\phi^{(n)} = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}, \quad (1.31)$$

avec les conditions initiales  $\phi^{(k)}(0) = m_k$ , pour  $0 \leq k \leq n-1$ .

L'Hypothèse 5 inclut des taux de naissance qui sont d'intérêt dans de nombreuses applications. Pour  $n = 1$ , on retrouve bien sûr le cas exponentiel, et à l'ordre supérieur  $n = 2$  nous avons déjà accès à des fonctions d'intérêt pour les applications comme par exemple celles du type  $\phi(a) = \alpha a e^{\beta a}$ . Sur la Figure 1.19, nous représentons deux comportements intéressants à l'ordre  $n = 2$  : un retard à l'excitation sur la figure de gauche, ou un changement de régime pour la figure de droite.



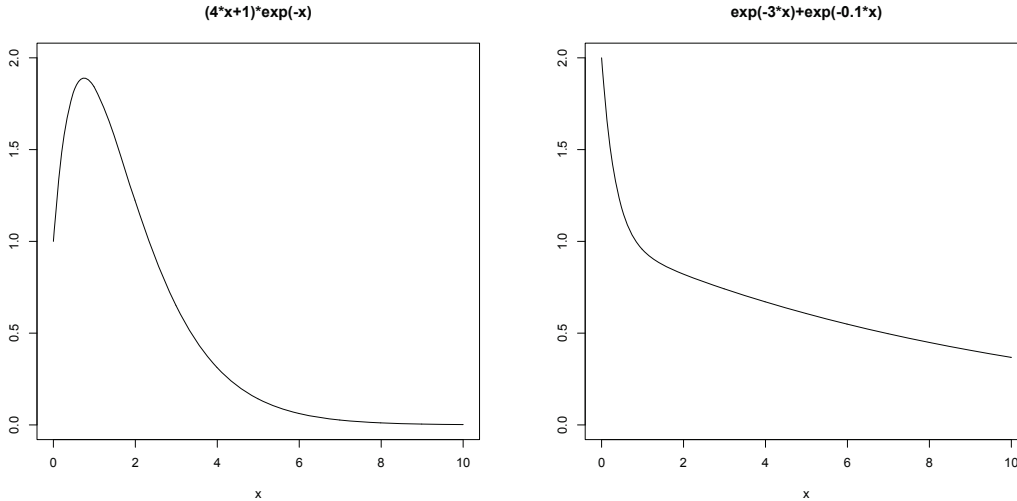


FIGURE 1.19 – Exemples de taux de naissance vérifiant l'Équation (1.31) ( $n=2$ )

Caractérisons davantage ces taux de naissance. Introduisons d'abord le vecteur

$$m = (1, m_0, \dots, m_{n-1})^T, \quad (1.32)$$

et la matrice

$$C(c) = \begin{pmatrix} 0 & 0 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ c_{-1} & c_0 & \cdots & c_{n-2} & c_{n-1} \end{pmatrix}. \quad (1.33)$$

L'Équation (1.31) peut s'écrire  $\Phi' = C\Phi$  avec  $\Phi = (1, \phi, \dots, \phi^{(n-1)})^T$ , dont la solution est donnée par  $\Phi(a) = e^{aC}m$ . La fonction  $\phi$  est donc la deuxième coordonnée du vecteur  $\Phi(a)$ . En particulier, si le polynôme  $P(y) = y^n - \sum_{k=0}^{n-1} c_k y^k$  est scindé, dont les racines distinctes sont  $y_1, \dots, y_p$  et les multiplicités associées  $n_1, \dots, n_p$ , alors  $\phi$  peut être écrite à une constante près comme  $\sum_{i=1}^p P_i(a)e^{y_i a}$  où  $P_i$  est un polynôme de degré au plus  $n_i - 1$ .

**Dynamique de la pyramide des âges** Revenons à la dynamique de la pyramide des âges au cours du temps. Celle-ci nous permet de capturer les composantes minimales à ajouter au processus de Hawkes et à son intensité pour obtenir une dynamique Markovienne. Ceci est donné dans la proposition suivante.

**Proposition 4.** *Sous l'Hypothèse 5, le processus  $X_t = (\langle Z_t, 1 \rangle, \langle Z_t, \phi \rangle, \dots, \langle Z_t, \phi^{(n-1)} \rangle)^T$  vérifie la dynamique*

$$X_t = N_t m + \int_0^t C X_s ds, \quad (1.34)$$

où le vecteur  $m$  et la matrice  $C$  sont définis en (1.32) et (1.33) respectivement. En particulier,  $X$  est un processus de Markov.

**Calcul des premiers moments** La dynamique (1.34) donne immédiatement le comportement moyen du processus de Hawkes.

**Proposition 5.** *Sous l'Hypothèse 5, la fonction (vectorielle)  $u(t) := \mathbb{E}[X_t]$  est la solution de*

$$u'(t) = \mu m + Au(t), \quad (1.35)$$

où la matrice  $A$  de taille  $(n + 1) \times (n + 1)$  est donnée par

$$A = C + mJ, \quad (1.36)$$

avec

$$J = (0, 1, 0, \dots, 0), \quad (1.37)$$

et le vecteur  $m$  et la matrice  $C$  sont définis en (1.32) et (1.33) respectivement.

Nous donnons ci-dessous la résolution de l'Équation (1.35) pour le premier moment pour le cas exponentiel  $\phi(a) = e^{-ca}$  (également donné par exemple par Dassios and Zhao (2011)), ainsi qu'à l'ordre  $n = 2$  pour le cas  $\phi(a) = \alpha^2 a e^{-\beta a}$ . Notons dans ces exemples la différence de comportement dans les deux cas, et également pour chaque cas les variations entre le régime critique tel que  $\int_0^\infty \phi(a) da = 1$  et les régimes sous- et sur-critiques.

**Corollaire 1.** *Pour le processus de Hawkes avec  $\phi(a) = e^{-ca}$ ,  $c > 0$ ,*

$$\begin{aligned} \mathbb{E}[N_t] &= \mu \left( t + \frac{t^2}{2} \right) \quad \text{if } c = 1, \\ \mathbb{E}[N_t] &= \frac{\mu}{1-c} \left( \frac{e^{(1-c)t} - 1}{1-c} - ct \right), \quad \text{if } c \neq 1. \end{aligned}$$

**Corollaire 2.** *Pour le processus de Hawkes avec  $\phi(a) = \alpha^2 a e^{-\beta a}$ ,  $\alpha, \beta > 0$ ,*

$$\begin{aligned} \mathbb{E}[N_t] &= \frac{\mu}{8\beta} (1 - e^{-2\beta t}) + \frac{3\mu}{4} t + \frac{\beta\mu}{4} t^2, \quad \text{si } \alpha = \beta, \\ \mathbb{E}[N_t] &= \frac{\mu\beta^2}{\beta^2 - \alpha^2} t + \frac{\alpha\mu}{2} \left( \frac{e^{(\alpha-\beta)t} - 1}{(\alpha - \beta)^2} - \frac{e^{-(\alpha+\beta)t} - 1}{(\alpha + \beta)^2} \right), \quad \text{si } \alpha \neq \beta. \end{aligned}$$

Nous nous intéressons maintenant au calcul des moments d'ordre deux. Pour cela, nous étudions la dynamique de la matrice  $V_t = X_t \bar{X}_t$  où  $X_t := (N_t, \langle Z_t, \phi \rangle, \dots, \langle Z_t, \phi^{(n-1)} \rangle)^T$  et  $\bar{X}_t$  représente la transposée du vecteur  $X_t$ . Nous basons notre méthode sur le calcul différentiel pour les processus à variation finie. Celle-ci pourrait être étendue aux moments d'ordre supérieur ; cet aspect est laissé en perspective de recherche.

**Proposition 6.** *Le processus (matriciel)  $(V_t)$  vérifie*

$$dV_t = dN_t (X_{t-}\bar{m} + m\bar{X}_{t-} + m\bar{m}) + dt (V_t\bar{C} + CV_t).$$

*En particulier, la matrice de variance-covariance  $v(t) = \mathbb{E}[V_t]$  vérifie l'équation différentielle ordinaire*

$$v'(t) = v(t)\bar{A} + Av(t) + \mu(m\bar{m} + u(t)\bar{m} + m\bar{u}(t)) + Ju(t)m\bar{m}, \quad (1.38)$$

*où  $u(t)$  est solution de (1.35) et la matrice  $A$  est définie en (1.36).*

Nous donnons deux exemples de formules explicites pour le cas  $\phi(a) = e^{-ca}$  et à un ordre supérieur pour le cas  $\phi(a) = \beta^2 a e^{-\beta a}$ , qui correspond au cas critique. Ces formules sont obtenues en résolvant l'Équation (1.38).

**Corollaire 3.** *Pour le processus de Hawkes avec  $\phi(a) = e^{-ca}$ ,*

$$\text{Var}(N_t) = \mu t \left( 1 + \frac{3}{2}t + \frac{2}{3}t^2 + \frac{1}{12}t^3 \right) \text{ si } c = 1,$$

$$\text{Var}(N_t) = \frac{\mu}{(1-c)^3} \left[ \frac{1-c/2}{1-c} e^{2(1-c)t} + \left( \frac{3c^2-1}{1-c} - 2ct \right) e^{(1-c)t} - c^3 t + \frac{c(1/2-3c)}{1-c} \right], \text{ si } c \neq 1.$$

**Corollaire 4.** *Pour le processus de Hawkes avec  $\phi(a) = \beta^2 a e^{-\beta a}$ , la variance de l'intensité est donnée par*

$$\text{Var}(\lambda_t) = \beta\mu \left( -\frac{7}{128} + \frac{3\beta}{32}t + \frac{\beta^2}{16}t^2 + \frac{1-\beta t}{8}e^{-2\beta t} - \frac{9}{128}e^{-4\beta t} \right).$$

**Transformée de Laplace** Le résultat principal concernant les processus de Hawkes standards, qui sera également étendu aux processus de Hawkes avec immigrants généraux, concerne la transformée de Laplace. Nous l'illustrerons numériquement dans la Partie 1.3.5.

**Proposition 7.** *Sous l'Hypothèse 5, la transformée de Laplace du processus de Hawkes est donnée par*

$$\mathbb{E}[\exp(\theta N_T)] = \exp \left\{ -\mu \left( (-1)^n G^{(n)}(0) + \sum_{k=0}^{n-1} (-1)^{k+1} c_k G^{(k)}(0) \right) \right\}, \quad (1.39)$$

*où  $G$  satisfait l'équation différentielle non-linéaire : pour tout  $0 \leq t \leq T$ ,*

$$(-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) + \exp \left( \theta - c_{-1} G(t) + \sum_{k=0}^{n-1} b_k G^{(k+1)}(t) \right) - 1 = 0,$$

*avec conditions terminales  $G^{(k)}(T) = 0$  pour  $0 \leq k \leq n$ , et pour  $0 \leq k \leq n-1$ ,  $b_k = (-1)^k (m_{n-1-k} - \sum_{l=k+1}^{n-1} m_{n-1-l} c_{n-l+k})$ .*

### 1.3.4 Les processus de Hawkes avec immigrants généraux

Dans cette partie, nous étudions le processus de Hawkes avec immigrants généraux (voir Brémaud and Massoulié (2002)), défini ci-dessous.

**Définition 2.** *Un processus de Hawkes avec immigrants généraux est un processus de comptage  $(N_t)$  dont l'intensité est donnée par*

$$\lambda_t = \mu(t) + \sum_{T_n < t} \Phi_t(t - T_n, X_n) + \sum_{S_k < t} \Psi_t(t - S_k, Y_k), \quad (1.40)$$

où les  $T_n$  sont les temps de saut de  $N$ , les  $S_k$  sont les temps de saut d'un processus de comptage d'intensité déterministe  $\rho(t)$  et les  $X_n$  (resp.  $Y_k$ ) sont des variables aléatoires positives, iid de distribution  $G$  (resp.  $H$ ). Les  $(S_k)$ ,  $(Y_k)$  et  $(X_n)$  sont supposés indépendants les uns des autres.

Rappelons la dynamique dans ce modèle. Les  $T_n$  sont les temps de saut du processus  $(N_t)$  : si un événement survient pour le système à la date  $T_n$ , l'intensité augmente d'une quantité  $\Phi_t(t - T_n, X_n)$  due à l'auto-excitation, où  $X_n$  est une marque tirée indépendamment de  $T_n$ . En parallèle, des chocs externes surviennent aux temps  $S_k$  et excitent le système en augmentant l'intensité d'une quantité  $\Psi_t(t - S_k, Y_k)$ . On peut retrouver le processus de Hawkes standard étudié précédemment en posant  $\Phi_t(a, x) = \phi(a)$  et  $\Psi_t(a, x) = 0$ . Les processus de Hawkes avec immigrants généraux ont été introduits par Brémaud and Massoulié (2002) et étudiés dans un cadre stationnaire. Leurs bonnes propriétés, flexibilité et facilité d'interprétation en particulier, leur ont fait connaître un regain d'intérêt dans une littérature récente motivée par des applications financières (voir Dassios and Zhao (2011), Wheatley et al. (2014) et Rambaldi et al. (2014)). En particulier, la distribution de tels processus a été étudiée par Dassios and Zhao (2011) dans le cas  $\Phi_t(a, x) = \Psi_t(a, x) = xe^{-\delta a}$ , cadre dans lequel le processus d'intensité est Markovien. Nous illustrons dans cette partie la contribution de ces travaux de thèse à l'étude des processus de Hawkes avec immigrants généraux non-stationnaires pour une classe plus importante de fonctions  $\Phi_t$  et  $\Psi_t$ , ce qui généralise en particulier les travaux de Dassios and Zhao (2011) dans cette direction. Pour ce faire, nous décrivons d'abord la dynamique de naissance avec immigration avec âges et caractéristiques.

**Description de la dynamique de naissance avec immigration** La dynamique que l'on décrit ici est celle d'un processus d'immigration-naissance avec âges et caractéristiques. Nous représentons ce processus à l'aide de deux populations. Dans chacune des populations, chaque individu a un âge  $a$  évoluant au cours du temps et une caractéristique  $x \in \mathbb{R}_+$ . La première population représente les chocs externes, tandis que la seconde représente les événements (internes) pour le processus de Hawkes. Chaque population  $(i)$ ,  $i = 1$  ou  $2$ , est représentée au temps  $t$  comme une

mesure qui met un poids en l'âge et en la caractéristique de chaque individu dans la population, notée  $Z_t^{(i)}(da, dx)$ . Les deux populations sont définies sur la base de la Définition 2 comme

$$Z_t^{(1)}(da, dx) = \sum_{S_k \leq t} \delta_{(t-S_k, Y_k)}(da, dx) \text{ et } Z_t^{(2)}(da, dx) = \sum_{T_n \leq t} \delta_{(t-T_n, X_n)}(da, dx). \quad (1.41)$$

On peut appeler  $Z_t^{(i)}$  *structure de population* dû au fait que les âges mais aussi les caractéristiques sont représentées, au lieu de *pyramide des âges* que nous réservons ici aux mesures en âge uniquement. Comme pour le modèle général et le processus de Hawkes standard, on peut calculer une fonction de la structure globale de la population, fonction qui peut en outre dépendre du temps. Soit une fonction  $f_t(a, x)$  du temps, de l'âge et de la caractéristique des individus. On peut la calculer sur toute la population par

$$\langle Z_t^{(i)}, f_t \rangle = \int_{\mathbb{R}_+ \times \mathbb{R}_+} f_t(a, x) Z_t^{(i)}(da, dx), \quad (1.42)$$

pour  $i = 1$  ou  $i = 2$ . Par exemple, le processus de Hawkes avec immigrants généraux est  $N_t^{(2)} = \langle Z_t^{(2)}, \mathbf{1} \rangle$ . De plus, son intensité  $\lambda_t$  définie dans l'Équation (1.40) se réécrit

$$\lambda_t = \mu(t) + \langle Z_{t-}^{(2)}, \Phi_t \rangle + \langle Z_{t-}^{(1)}, \Psi_t \rangle.$$

Cette formulation montre le fonctionnement de la dynamique de naissance avec immigration.

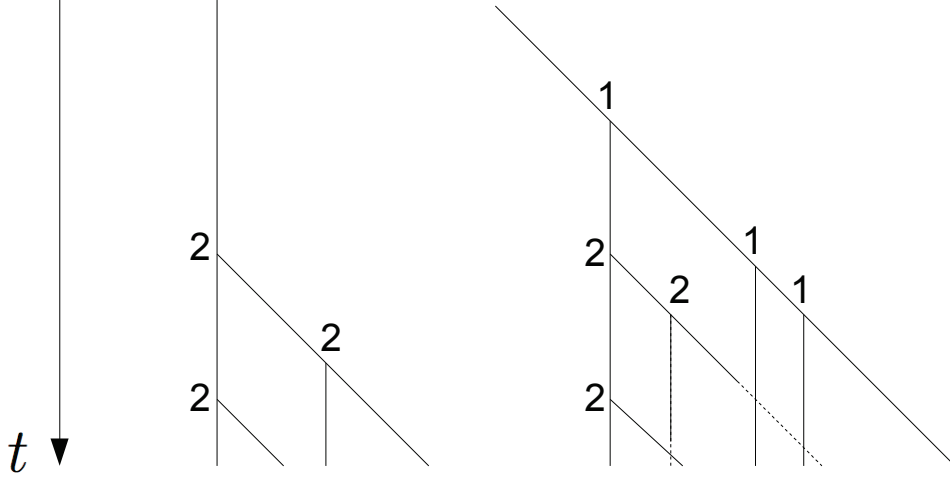
(i) Décrivons d'abord la population (1) des chocs externes. Celle-ci est constituée d'immigrants qui arrivent dans la population (1) au taux  $\rho(t)$ ; à l'arrivée, il a l'âge 0 et une caractéristique  $x$  tirée suivant la distribution  $H$ . Un individu  $(a, x)$  qui vit dans la population (1) au temps  $t$  donne naissance au taux  $\Psi_t(a, x)$ . Le descendant appartient à la population (2); il a l'âge 0 et une caractéristique tirée suivant la distribution  $G$ .

(ii) Complétons maintenant la description de la population (2). En plus des naissances dues à des individus de la population (1), la population (2) évolue suivant deux autres types d'événement : immigration et naissances internes. Les immigrants arrivent dans la population (2) au taux  $\mu(t)$  avec l'âge 0 et une caractéristique tirée suivant la distribution  $G$ . Chaque individu  $(a, x)$  vivant dans la population (2) au temps  $t$  donne naissance au taux  $\Phi_t(a, x)$ . Le descendant appartient également à la population (2); il a l'âge 0, et une caractéristique tirée suivant la distribution  $G$ .

Cette dynamique est illustrée sur la Figure 1.20.

Comme pour l'étude du processus de Hawkes standard, une étape cruciale est l'étude de la déformation de la structure de la population au cours du temps. En d'autres termes, quelle est la dynamique du processus  $\langle Z_t^{(i)}, f_t \rangle$  pour  $i = 1$  ou  $2$ ? Ceci est étudié dans le lemme suivant.

FIGURE 1.20 – Dynamique du processus de naissance avec immigration : des immigrants arrivent dans la population (1) (chocs externes). Puis chaque individu 1 donne naissance à des individus 2 (événements dus à des chocs externes). En parallèle, des immigrants arrivent dans la population (2) (événements dus à l'intensité de base). Finalement, chaque individu 2 se reproduit (auto-excitation). Le processus de Hawkes avec immigrants généraux est donné par le nombre d'individus 2.



**Lemme 2.** Pour toute fonction  $f : (t, x, a) \mapsto f_t(a, x)$  dérivable en  $t$  et  $a$ , la dynamique du processus  $\langle Z_t^{(i)}, f_t \rangle$  pour  $i = 1$  ou  $2$  est donnée par

$$d\langle Z_t^{(i)}, f_t \rangle = \int_{\mathbb{R}_+} f_t(0, x) N^{(i)}(dt, dx) + \langle Z_t^{(i)}, (\partial_a + \partial_t) f_t \rangle dt,$$

où les mesures ponctuelles  $N^{(1)}$  et  $N^{(2)}$  sont données par

$$N^{(1)}(dt, dx) = \sum_{k \geq 1} \delta_{(S_k, Y_k)}(dt, dx) \text{ et } N^{(2)}(dt, dx) = \sum_{n \geq 1} \delta_{(T_n, X_n)}(dt, dx). \quad (1.43)$$

Comme le Lemme 1, cette formulation détaille la partie de saut pur (terme de gauche) et la partie vieillissement et dépendance en temps (terme de droite). Le fait que le terme de transport dépende à la fois de  $\langle Z_t^{(i)}, \partial_a f_t \rangle$  et  $\langle Z_t^{(i)}, \partial_t f_t \rangle$  est le point de départ du résultat que nous présentons dans la suite.

**Résultat principal** Nous introduisons les hypothèses nécessaires pour obtenir une dynamique Markovienne (non-homogène) de dimension finie.

**Hypothèse 6.** (i) Les taux de naissance  $\Phi$  et  $\Psi$  sont positifs et vérifient  $\Phi_t(a, x) = v(t)\phi(a, x)$  et  $\Psi_t(a, x) = w(t)\psi(a, x)$ , où

$$\phi^{(n)}(a, x) = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}(a, x) \text{ et } v^{(p)}(t) = d_{-1}(t) + \sum_{l=0}^{p-1} d_l(t) v^{(l)}(t),$$

avec  $n, p \geq 1$  et les conditions initiales  $\phi^{(k)}(0, x) = \phi_0^{(k)}(x)$ , et

$$\psi^{(m)}(a, x) = r_{-1} + \sum_{k=0}^{m-1} r_k \psi^{(k)}(a, x) \text{ et } w^{(q)}(t) = k_{-1}(t) + \sum_{l=0}^{q-1} k_l(t) w^{(l)}(t),$$

avec  $m, q \geq 1$  et les conditions initiales  $\psi^{(k)}(0, x) = \psi_0^{(k)}(x)$ . Insistons ici sur le choix de notation  $f^{(k)}(a, x) = \partial_a^k f(a, x)$ .

(ii) Les fonctions  $(d_l)_{-1 \leq l \leq p-1}$  et  $(k_l)_{-1 \leq l \leq q-1}$  sont continues.

L'Hypothèse 6 définit une classe variée de fonctions de fertilité de la forme  $\Phi_t(a, x) = v(t)\phi(a, x)$ . Pour caractériser la partie  $\phi(a, x)$  indépendante du temps, introduisons le vecteur  $F(a, x)$  tel que  $F = (1, \phi, \dots, \phi^{(n-1)})^T$ . Dans ce cas,  $F' = CF$ , où  $C$  est définie en (1.33). En particulier, si le polynôme  $P(y) = y^n - \sum_{k=0}^{n-1} c_k y^k$  est scindé avec racines distinctes  $y_1, \dots, y_p$  et les multiplicités associées  $n_1, \dots, n_p$ , alors  $\phi$  peut s'écrire à une constante additive près comme  $\sum_{i=1}^p P_i(x, a) e^{y_i a}$  où  $P_i$  est un polynôme en  $a$  de degré au plus  $n_i - 1$  et dont les coefficients peuvent dépendre de  $x$ . Ceci inclut en particulier le cadre du processus de contagion dynamique de Dassios and Zhao (2011) où  $\Phi_t(a, x) = \Psi_t(a, x) = x e^{-\delta a}$ . L'Hypothèse 6 prend également en compte des taux de naissance dépendant du temps, ce qui est d'intérêt de manière générale pour l'étude de certains processus de Hawkes non-stationnaires, et en particulier pour inclure de la saisonnalité. Par exemple, on peut penser à un taux de naissance de la forme  $\cos^2(\alpha t)\phi(a, x)$  où  $v(t) = \cos^2(\alpha t)$  satisfait  $v'' = 4\alpha^2(1 - v)$ .

Détaillons maintenant la contribution principale à l'étude des processus de Hawkes avec immigrants généraux : la martingale exponentielle et la transformée de Laplace. Pour assurer l'existence de solutions, nous avons besoin de l'hypothèse suivante :

**Hypothèse 7.** Pour tout  $\lambda > 0$ ,

$$\int_{\mathbb{R}_+} \exp\left(\lambda \max_{0 \leq k \leq n-1} \phi_0^{(k)}(x)\right) G(x) dx < +\infty.$$

Le résultat principal est donné ci-dessous. Remarquons que la trace de la matrice  $\bar{u}M$ , donnée par  $\text{Tr}(\bar{u}M) = \sum_{k,l} u_{k,l} M_{k,l}$  calcule une combinaison linéaire des composantes de la matrice  $M$ , et rappelons que  $\bar{u}$  désigne la transposée de la matrice  $u$ .

**Théorème 2.** Définissons pour  $-1 \leq k \leq n-1$  et  $-1 \leq l \leq p-1$ ,

$$X_t^{k,l} := \langle Z_t^{(2)}, \partial_a^k \partial_t^l \Phi_t \rangle \text{ et pour } -1 \leq k \leq m-1 \text{ et } -1 \leq l \leq q-1,$$

$$Y^{k,l} := \langle Z_t^{(1)}, \partial_a^k \partial_t^l \Psi_t \rangle. \text{ Introduisons les deux processus à valeur matricielle}$$

$$M_t^{(2)} = \left( X_t^{(k,l)} \right)_{-1 \leq k \leq n-1, -1 \leq l \leq p-1} \text{ et } M_t^{(1)} = \left( Y_t^{(k,l)} \right)_{-1 \leq k \leq m-1, -1 \leq l \leq q-1}.$$

Soit  $\mathcal{F}^M$  la filtration canonique engendrée par  $(M^{(1)}, M^{(2)})$ . Sous l'Hypothèse 6, nous avons les résultats suivants.

(i) Pour toute fonction dérivable (matricielle)  $(A_t^{(1)})$  et  $(A_t^{(2)})$  de dérivées notées  $(\dot{A}_t^{(1)})$  et  $(\dot{A}_t^{(2)})$ , le processus ci-dessous est une  $\mathcal{F}^M$ -martingale :

$$\begin{aligned} & \exp \left\{ \sum_{i=1}^2 \int_0^t \text{Tr} \left( A_s^{(i)} M_s^{(i)} \right) - \int_0^t \text{Tr} \left( A_s^{(i)} C^{(i)} M_s^{(i)} + A_s^{(i)} M_s^{(i)} \bar{D}_s^{(i)} + \dot{A}_s^{(i)} M_s^{(i)} \right) ds \right. \\ & - \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(1)} W^{(1)}(s,x))} - 1 \right) \rho(s) H(x) dx ds \\ & \left. - \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(2)} W^{(2)}(s,x))} - 1 \right) (\mu(s) + M_s^{(1)}[0,0] + M_s^{(2)}[0,0]) G(x) dx ds \right\}, \end{aligned} \quad (1.44)$$

avec

- $W_{k,l}^{(1)}(t, x) = w^{(l)}(t) \psi_0^{(k)}(x)$  pour  $-1 \leq k \leq m-1$  et  $-1 \leq l \leq q-1$ ,
- $W_{k,l}^{(2)}(t, x) = v^{(l)}(t) \phi_0^{(k)}(x)$  pour  $-1 \leq k \leq n-1$  et  $-1 \leq l \leq p-1$ ,
- $C^{(1)} = C(r)$ ,  $C^{(2)} = C(c)$ ,  $D_t^{(1)} = C(k(t))$  et  $D_t^{(2)} = C(d(t))$  où la fonction  $C(\cdot)$  qui transforme un vecteur en une matrice est définie en (1.33).

(ii) Pour toutes matrices réelles  $u$  et  $v$  de dimensions  $(n+1)(p+1)$  et  $(m+1)(q+1)$  respectivement, la transformée de Laplace jointe s'écrit

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \text{Tr}(\bar{u} M_t^{(1)} + \bar{v} M_t^{(2)}) \right) \right] &= \exp \left\{ \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(1)} W^{(1)}(s,x))} - 1 \right) \rho(s) H(x) dx ds \right. \\ & \left. + \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(2)} W^{(2)}(s,x))} - 1 \right) \mu(s) G(x) dx ds \right\}, \end{aligned}$$

où

$$\text{pour } i \in \{1, 2\}, \quad \dot{A}_t^{(i)} + A_t^{(i)} C^{(i)} + \bar{D}_t^{(i)} A_t^{(i)} = \left\{ \int_{\mathbb{R}_+} \left( 1 - e^{\text{Tr}(A_t^{(2)} W^{(2)}(t,x))} \right) G(x) dx \right\} K, \quad (1.45)$$

avec conditions terminales

$$A_T^{(1)} = \bar{u} \text{ et } A_T^{(2)} = \bar{v}, \quad (1.46)$$

où la matrice  $K$  est donnée par  $K = \bar{J} J$  et  $J$  est définie en (1.37). De plus, sous l'Hypothèse 7, il existe une unique solution aux Équations (1.45)-(1.46).

### 1.3.5 Illustrations numériques

L'interprétation des processus de Hawkes avec immigrants généraux en terme de population permet d'utiliser l'Algorithme 1 décrit dans la Partie 1.2 pour la simulation. Ce procédé de *thinning* a l'avantage de s'appliquer à des taux de naissance généraux (voir une autre méthode Dassios et al. (2013) dans le cas de taux de naissance exponentiels, et plus efficace dans ce cadre). Au-delà de la mise en oeuvre la procédure de *thinning*, nous mettrons ici l'accent sur l'illustration des formules explicites et semi-explicites que nous avons établies.



**Description de quelques trajectoires** Nous proposons ici d'illustrer les différents comportements d'un processus de Hawkes lorsque l'on fixe le même nombre moyen d'enfants par individu. Prenons par exemple les taux de naissance  $\phi_1(a) = e^{-a}$  et  $\phi_2(a) = ae^{-a}$  (voir Figure 1.21), qui sont tels que leur intégrale vaut 1 : il s'agit donc du régime critique. Ce régime est intéressant en soi car il n'est de fait pas adressé dans l'étude des processus de Hawkes stationnaires (régime sous-critique) et se situe à la frontière. Nous représentons les résultats liés à la simulation d'une trajectoire avec taux de naissance  $\phi_1$  et  $\phi_2$  sur les Figures 1.22 et 1.23 ; nous y représentons le processus de Hawkes lui-même, le nombre d'événements par unité de temps, la processus intensité au cours du temps, et enfin la pyramide des âges finale.

Remarquons d'abord, comme attendu, le lien entre l'intensité du processus et le nombre d'événements par unité de temps, et au-delà la propriété dite de *clustering* du processus de Hawkes qui rend compte de l'arrivée "par paquets" des événements.

Remarquons également que dans la simulation de processus de Hawkes, la représentation de processus intensité come celui de la Figure 1.23 est plutôt peu habituelle, car en général c'est le cas exponentiel qui est représenté. Si la forme de ce processus intensité est caractéristique du retard à l'excitation, c'est aussi l'implémentation de celui-ci qui le distingue du cas exponentiel. En effet, dans le cas exponentiel, on est capable de tracer l'intensité dans un intervalle en ayant uniquement retenu les deux sauts adjacents qui en forment les bornes. En revanche, mettre en mémoire ces deux sauts n'est pas suffisant pour reconstruire l'intensité dans le cas de la Figure 1.23. Il faut alors ajouter une composante supplémentaire : ceci ne fait que traduire numériquement la contrainte de dimension minimale de la dynamique Markovienne sous-jacente.

Enfin, notons que le nombre d'événements *dans le temps* est similaire (à retournement du temps près) à la pyramide des âges *finale* ; ceci supporte l'intuition que toute l'information est contenue dans cette pyramide des âges.

**Premiers moments** Nous souhaitons maintenant illustrer les formules explicites obtenues pour les premiers moments. Remarquons que l'auto-excitation due à  $\phi_1$  étant immédiate et celle relative à  $\phi_2$  étant retardée, il est clair que ceci pousse le processus à croître plus rapidement avec  $\phi_1$ . Nous allons confirmer cette remarque en traçant sur la Figure 1.24 la moyenne explicite du processus de Hawkes, voir Corollaires 1 et 2. Sur la Figure 1.25, nous représentons également la variance de l'intensité du processus de Hawkes pour chaque fonction de naissance. Ces représentations nous donnent une première quantification des différences de comportement de ces deux modèles considérés dans le régime critique.

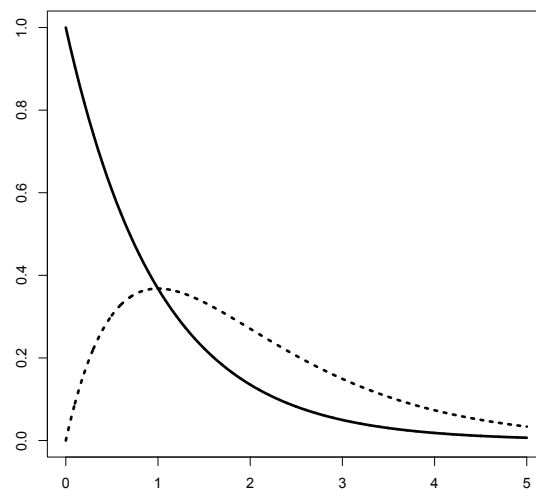


FIGURE 1.21 – Fonctions  $\phi_1(a) = e^{-a}$  (courbe pleine) et  $\phi_2(a) = ae^{-a}$  (courbe en pointillés)

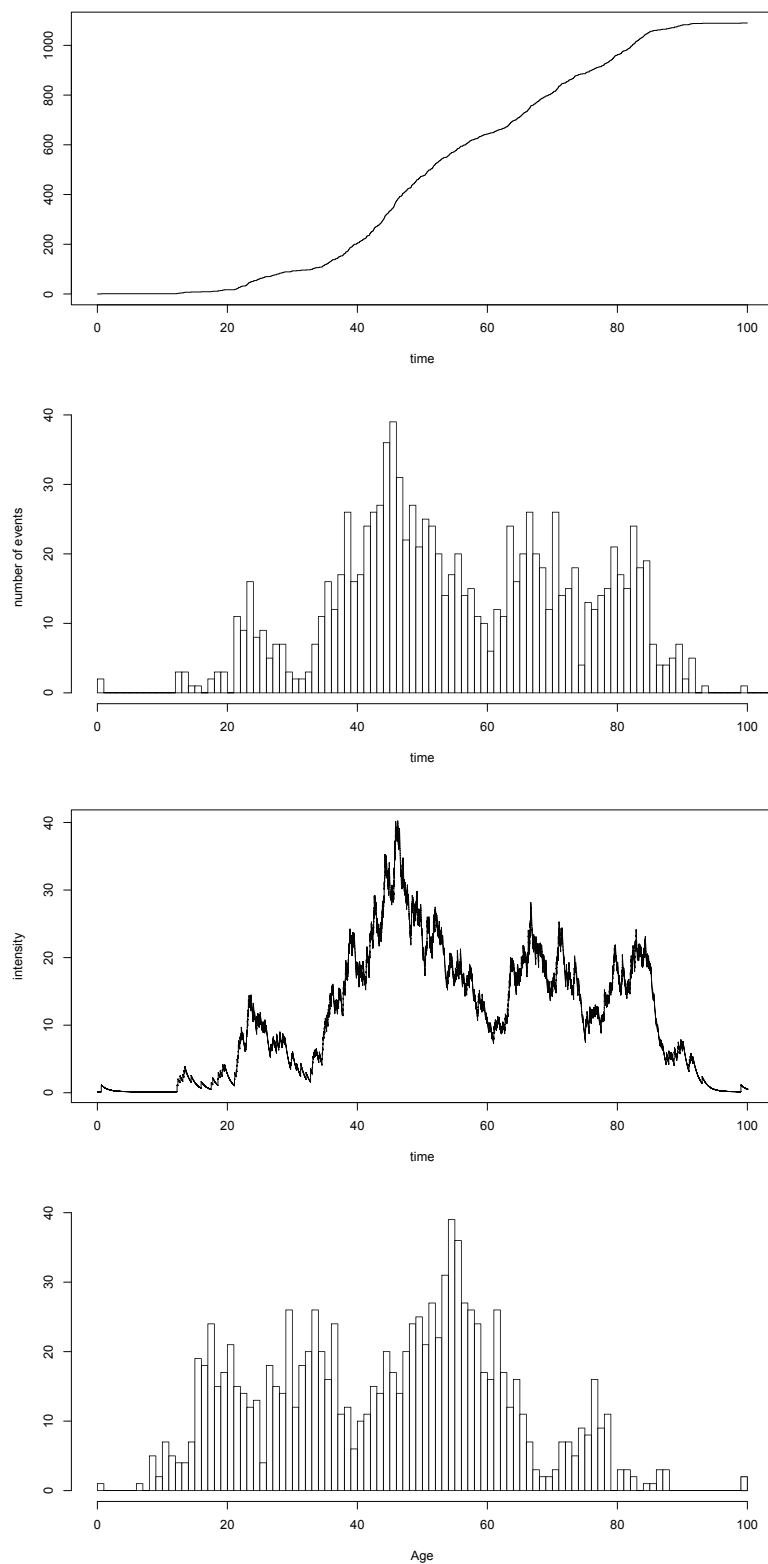


FIGURE 1.22 – De haut en bas : processus de Hawkes, nombre d'événements par unité de temps, intensité du processus de Hawkes et pyramide des âges finale avec  $\mu = 0.15$  et  $\phi_1(a) = e^{-a}$ .

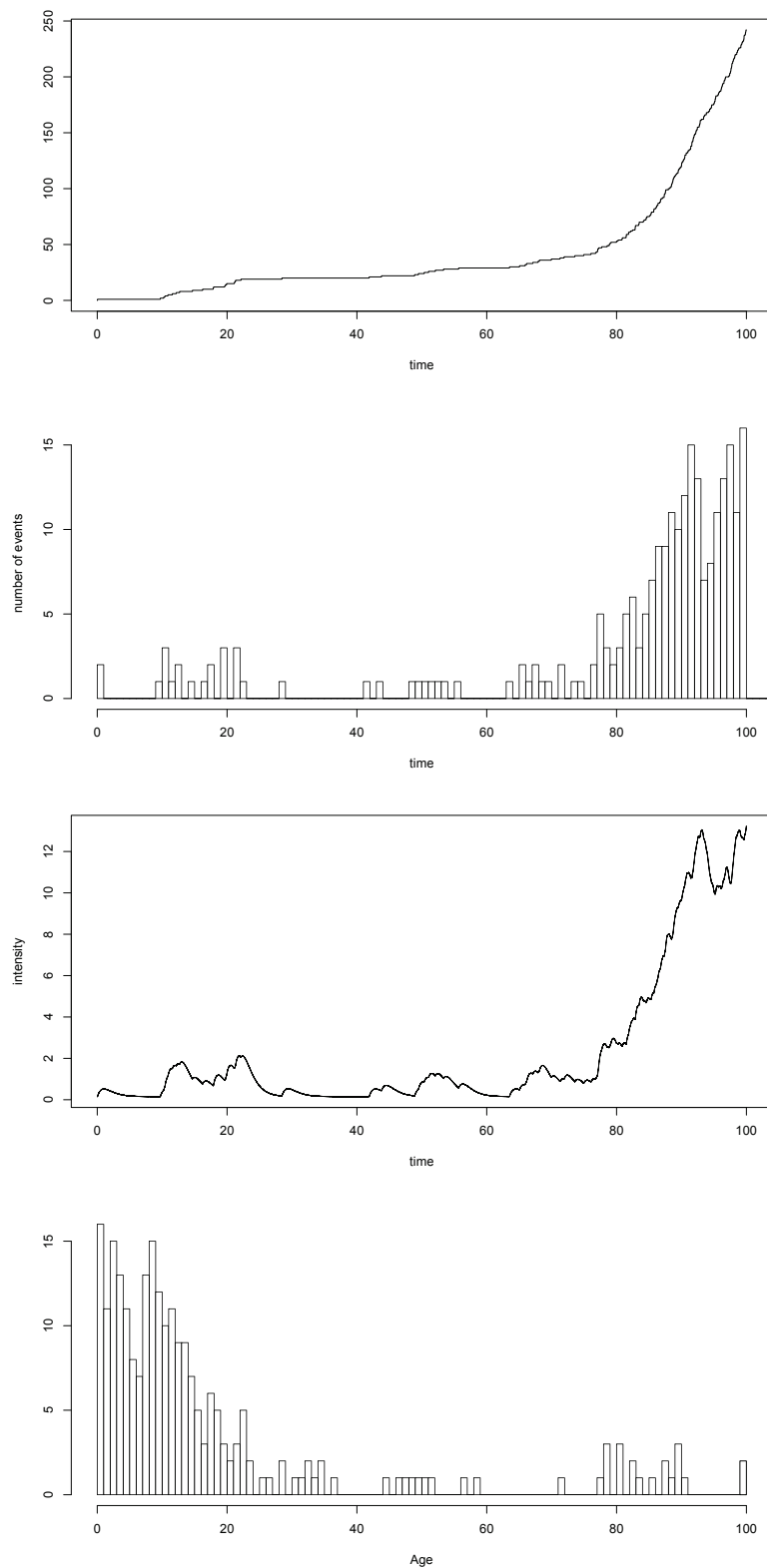


FIGURE 1.23 – De haut en bas : processus de Hawkes, nombre d'événements par unité de temps, intensité du processus de Hawkes et pyramide des âges finale avec  $\mu = 0.15$  et  $\phi_2(a) = ae^{-a}$ .

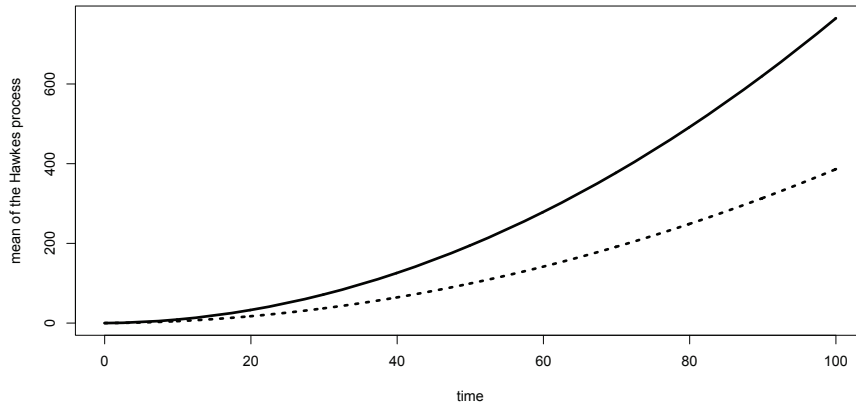


FIGURE 1.24 – Moyenne du processus de Hawkes pour les fonctions  $\phi_1(a) = e^{-a}$  (courbe pleine) et  $\phi_2(a) = ae^{-a}$  (courbe en pointillés)

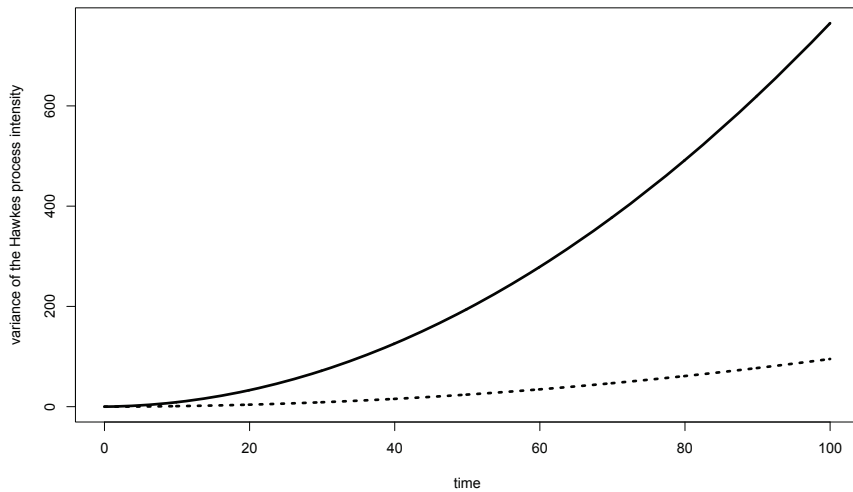


FIGURE 1.25 – Variance de l'intensité du processus de Hawkes pour les fonctions  $\phi_1(a) = e^{-a}$  (courbe pleine) et  $\phi_2(a) = ae^{-a}$  (courbe en pointillés)

**Fonction génératrice et probabilités d'atteinte de seuil** Sur la base de la Proposition 7, nous illustrons ici le calcul numérique de la fonction génératrice  $\mathbb{E}[u^{N_T}]$  (la probabilité de survie au temps  $T$  d'un système qui survit à chaque choc avec la probabilité  $u$ ), ainsi que des quantités de type  $\mathbb{P}(N_T = k) = \frac{1}{k!} \partial_u^k \mathbb{E}[u^{N_T}]|_{u=0}$  (la probabilité d'avoir subi exactement  $k$  chocs avant le temps  $T$ ). Nous avons implémenté des schémas de discrétisation pour résoudre l'équation différentielle non-linéaire satisfaite par  $G$ , et comparé les approches explicites et implicites; celles-ci nous ont fourni des résultats similaires. Nous avons également mis en oeuvre une sélection (graphique) du pas de discrétisation optimal pour le calcul des dérivées.

Les deux mêmes cas critiques sont comparés : le cas exponentiel  $\phi(a) = e^{-a}$  (cas 1) et à un ordre supérieur le cas  $\phi(a) = ae^{-a}$  (cas 2). Les résultats sont détaillés dans les Tableaux 1.1 et 1.2 avec trois chiffres significatifs. Comme suggéré dans les résultats numériques précédents, même si le nombre moyen d'enfants par individu est le même, les résultats sont assez différents car ils dépendent de la forme de chaque taux de naissance. Ceci renforce la motivation pour utiliser des taux de naissance au delà du cas exponentiel. En particulier, ces résultats permettent de mesurer l'impact du retard à l'excitation introduit dans le cas 2 par la fonction  $\phi(a) = ae^{-a}$ . Pour conclure cette expérimentation numérique, insistons sur le fait que l'approximation numérique des dérivées à des ordres plus importants nécessite des méthodes de dérivation numérique plus poussées.

$u$	0.1	0.3	0.5	0.7	0.9
Case 1, $\phi(a) = e^{-a}$	0.490	0.532	0.588	0.672	0.828
Case 2, $\phi(a) = ae^{-a}$	0.494	0.546	0.615	0.714	0.874

TABLE 1.1 – Valeurs estimées de  $\mathbb{E}[u^{N_T}]$  avec  $\mu = 0.15$  et  $T = 5$ .

$k$	0	1	2	3	4
Case 1, $\phi(a) = e^{-a}$	0.472	0.165	0.0894	0.0577	0.0407
Case 2, $\phi(a) = ae^{-a}$	0.472	0.203	0.113	0.0700	0.0451

TABLE 1.2 – Valeurs estimées de  $\mathbb{P}(N_T = k)$  avec  $\mu = 0.15$  et  $T = 5$ .

### 1.3.6 Conclusion et perspectives

La pyramide des âges semble constituer un outil intéressant pour l'étude des processus de Hawkes généraux, son principal avantage étant de garder trace de tous les événements passés grâce aux âges des individus. Ceci a permis d'étudier une classe de processus de Hawkes avec immigrants généraux dont les taux de naissance sont des généralisations du cas exponentiel en âge, pouvant également dépendre du temps. Avec la représentation trajectorielle que nous proposons dans ce travail, et qui est décrite dans le Chapitre 3, les deux définitions courantes des processus de Hawkes semblent alors réconciliées : d'une part la définition par le mécanisme d'immigration et de branchement, basée sur la représentation en cluster, et d'autre part la définition sous la forme d'un processus de comptage avec sa propre intensité.

Nous proposons ci-dessous un ensemble de perspectives sur ces travaux.

(i) Les résultats présentés pourraient être étendus. En particulier il serait immédiat de calculer les premiers moments des processus de Hawkes avec immigrants généraux en reprenant la méthode proposée pour les processus de Hawkes standards. On

pourrait également décrire les corrélations inter-temporelles avec la même démarche que pour les moments d'ordre deux.

(ii) Le lien fait ici entre les processus de Hawkes avec immigrants généraux et les modèles de processus à valeur mesure pourrait être approfondi dans la direction des théorèmes limites. En particulier, l'équation déterministe (1.10) donne le comportement macroscopique (voir Partie 1.2.4) des processus de Hawkes avec immigrants généraux pour des taux de naissance quelconques.

(iii) En terme d'application statistique, il serait intéressant d'investiguer l'estimation par la méthode des moments en se basant sur les formules explicites proposées. Il serait également intéressant d'analyser la vraisemblance du modèle pour des taux de naissance plus généraux que des sommes d'exponentielles, utilisées jusqu'ici dans la littérature.

(iv) Enfin, pour contribuer au débat actuel sur la pertinence d'inclure des chocs externes dans les modélisations en finance, il serait intéressant de mesurer le biais induit par l'estimation d'un processus de Hawkes standard sur des observations issues d'un processus de Hawkes avec immigrants généraux.

## 1.4 Les causes de décès dans la dynamique de population

La compréhension de la dynamique d'une population, en particulier sa structure en âge et son évolution au cours du temps, est réalisée dans ces travaux de thèse par une analyse des caractéristiques démographiques à un niveau plus détaillé. Nous déclinons ici cette démarche en analysant la mortalité nationale par causes de décès. Par cause nous entendons ici la chose suivante : dans les données, lorsqu'un décès est renseigné, en plus de la date de décès nous disposons de la cause (principale) ayant été identifiée. Par exemple, il peut s'agir d'une mort accidentelle, ou due à un cancer, ou à une autre maladie identifiée. Cette classification sera à la base de notre analyse, mais mentionnons ici qu'elle peut faire l'objet de nombreux débats. D'abord aux âges élevés, pour lesquels il est parfois difficile de proposer une cause principale. Également, la mention d'une unique cause pose problème à tous les âges car elle ne permet pas de capturer la dépendance qu'il peut y avoir entre les différentes causes de décès.

La classification internationale des maladies (*International Classification of Diseases, ICD*) est utilisée par l'Organisation Mondiale de la Santé (OMS) pour coder de manière commune les décès par cause depuis les années 1950 et pour plus de 200 pays. Les détails des causes sont nombreuses et peuvent être regroupées en grandes familles. Par exemple, pour la France on dispose de 4563 codes de causes de décès différents. Dans la littérature, les études sur la mortalité par causes de décès ont largement contribué à la compréhension de la structure en âge de la mortalité, et également à un certain nombre d'intuitions sur son évolution dans le temps. En particulier, une grande question abordée dans la littérature concerne l'impact d'une suppression de cause de décès sur la mortalité générale, par exemple pour modéliser une innovation médicale qui éradiquerait certains cancers. Les modèles de dynamique des populations permettent d'approfondir cette question. En effet, ceux-ci donnent l'évolution de la structure en âge de la population au cours du temps sur la base de taux de naissance et de mort qui dépendent de l'âge. Alors que les études actuelles sur les causes de décès s'intéressent à l'impact de suppression de cause sur la mortalité agrégée, la contribution de cette partie est de combiner ces deux domaines pour tenter de répondre à la question suivante : quel serait l'impact d'une réduction ou suppression de cause de décès sur la pyramide des âges ? La pyramide des âges est une quantité cruciale pour l'analyse démographique, mais également pour les systèmes de retraite par répartition basés sur des équilibres inter-générationnels. Quant aux assureurs, cette pyramide des âges leur donne une information cruciale pour évaluer la façon dont les institutions produisent les données de mortalité nationales de référence. En particulier, la dynamique de population permet dans ce cadre



de montrer comment les taux de naissance influent sur la pyramide des âges et le calcul des expositions au risque (voir la discussion de la Partie 1.1.2 sur ce thème).

### 1.4.1 La mortalité par causes de décès

Dans notre application aux populations humaines, nous nous intéressons à des individus caractérisés par un âge et un sexe, deux variables qui sont connues comme discriminantes concernant la mortalité, et en particulier la mortalité par cause de décès. Dans la littérature actuarielle et biostatistique, un cadre très répandu de modélisation considère que chaque individu, femme ou homme, est exposé à  $d$  causes de décès en compétition. Par exemple, celles-ci peuvent être le cancer, les maladies d'un type donné (circulatoires, respiratoires,...) ou les accidents. Nous décrivons ce modèle dit à *risques compétitifs* dans ce qui suit.

**Le modèle à risques compétitifs** On représente la durée de vie d'un individu de sexe  $\epsilon$  ( $\epsilon = f$  pour les femmes et  $\epsilon = m$  pour les hommes) par la variable aléatoire  $\tau^\epsilon$ , de fonction de survie  $S^\epsilon(a) = \mathbb{P}(\tau^\epsilon > a)$ . On rappelle que l'on retrouve le taux de hasard, ou force de mortalité, par

$$\mu^\epsilon(a) = \frac{\mathbb{P}(a < \tau^\epsilon \leq a + da \mid \tau^\epsilon > a)}{da} = -\frac{d \ln(S^\epsilon(a))}{da}. \quad (1.47)$$

Dans le modèle à risques compétitifs, on attache à chaque individu de sexe  $\epsilon$  un vecteur de durées de vie potentielles  $(\tau_1^\epsilon, \tau_2^\epsilon, \dots, \tau_d^\epsilon)$ , dont chaque coordonnée correspond à chacune des  $d$  causes de décès. Le décès de l'individu a lieu lorsque l'une de ces horloges en compétition sonne. En d'autres termes, la durée de vie  $\tau^\epsilon$  est définie comme le minimum entre toutes les durées de vie en compétition :  $\tau^\epsilon = \min_{1, \dots, d} \tau_i^\epsilon$ . La fonction de survie s'écrit alors  $S^\epsilon(a) = \mathbb{P}(\tau_1^\epsilon > a, \dots, \tau_d^\epsilon > a)$ . Etant donné ce cadre de modélisation, un intérêt réside dans l'évaluation de l'impact de la suppression ou de la diminution d'une ou plusieurs causes de décès. Pour cela, il faut avoir accès à la loi des variables  $\tau_i^\epsilon$ , qui sont les durées associées à chaque cause. Or en pratique, on n'observe que la durée de vie  $\tau^\epsilon$ , et également la cause qui a engendré le décès, c'est à dire que l'on connaît quel  $\tau_i^\epsilon$  a réalisé le minimum. Le problème d'identifiabilité du modèle à risques compétitifs est bien connu (voir (Tsiatis (1975))) : on ne peut pas, à partir de telles données, identifier les lois marginales des  $\tau_i^\epsilon$  et leur structure de dépendance. Ainsi, pour déterminer les lois marginales à partir des observations, il est nécessaire de faire une hypothèse sur la structure de dépendance des durées de vie associées à chaque cause. Ces hypothèses et la littérature associée sont discutées plus en détail dans le Chapitre 4. Dans ce qui suit, nous introduisons l'hypothèse d'indépendance et ses conséquences immédiates permettant de simplifier l'étude dans cette direction.

**L'hypothèse d'indépendance** Sous l'hypothèse d'indépendance entre les causes de décès, la fonction de survie s'écrit  $S^\epsilon(a) = \prod_{i=1}^d S_i^\epsilon(a)$  et le taux de décès est la somme des taux associés à chaque cause, i.e.  $\mu^\epsilon(a) = \sum_{i=1}^d \mu_i^\epsilon(a)$ , où  $\mu_i^\epsilon$  est tel que  $\mathbb{P}(\tau_i^\epsilon > a) = \exp\left(-\int_0^a \mu_i^\epsilon(s) ds\right)$ . La quantité  $\mu_i^\epsilon(a)$  est appelée la *force de mortalité nette* pour la cause  $i$ , le sexe  $\epsilon$  et l'âge  $a$ . Formellement, elle représente la probabilité de décéder instantanément sachant que l'on est en vie à l'âge  $a$  dans un monde où seule la cause  $i$  existe :

$$\mu_i^\epsilon(a) = \frac{\mathbb{P}(a < \tau_i^\epsilon \leq a + da \mid \tau_i^\epsilon > a)}{da}. \quad (1.48)$$

Comme nous l'avons décrit précédemment, en pratique les durées de vie associées à chaque cause ne sont pas totalement observées. Cependant, ce que l'on peut estimer à partir des données est la distribution de  $\tau_i$  pour les individus qui sont décédés de cette cause  $i$ , car dans ce cas précis la durée  $\tau_i$  est bien observée. En d'autres termes, il est possible d'estimer la probabilité  $\mathbb{P}(\tau_i \in [a, a + da), \tau = \tau_i)$ . Par indépendance entre les durées associées à chaque cause, cette quantité s'écrit

$$\mathbb{P}(\tau_i \in [a, a + da), \tau = \tau_i) = S_i(a)\mu_i(a)da \prod_{k \neq i} S_k(a) = S(a)\mu_i(a)da.$$

Ainsi, il est possible de retrouver la force de mortalité nette par

$$\mu_i(a) = \frac{\mathbb{P}(\tau_i \in [a, a + da), \tau = \tau_i \mid \tau > a)}{da}.$$

Dans la littérature actuarielle, la quantité de droite est appelée *force de mortalité brute*, et peut être estimée directement à partir des données de décès et des expositions au risque par cause. Une conséquence cruciale de l'hypothèse d'indépendance est donc que cette estimation capture la force de mortalité nette, en d'autres termes que les marginales peuvent être estimées directement à partir des données. Pour d'autres structures de dépendance, on peut trouver des équations liant les quantités *brutes* aux quantités *nettes* (voir par exemple Carriere (1994) et la discussion dans le Chapitre 4).

**Les données de l'OMS** Nous détaillons ici les données de mortalité par cause de décès que nous utiliserons pour les applications. Nous nous basons sur les données de l'OMS pour la France et l'année 2008. Nous regroupons les causes de décès élémentaires (au nombre de 4563) en 6 grandes classes. Ces 6 classes sont les cancers, les maladies de l'appareil circulatoire, celles de l'appareil respiratoire, les causes externes, les maladies infectieuses et parasitaires, et enfin les autres causes regroupées dans un même classe. La proportion des décès associée à chaque cause dans la population française en 2008 est donnée dans le Tableau 1.3. Les trois causes les plus importantes sont les cancers, les maladies de l'appareil circulatoire et la catégorie

des autres causes. Décrivons un peu plus en détail cette dernière : on y retrouve d'abord les maladies du système nerveux (21% des autres causes), du métabolisme (16%), de l'appareil digestif (13%) et les troubles mentaux (12%). Dans le total des décès, les maladies infectieuses et parasitaires, celles du système respiratoire et les causes externes sont moins représentées. Celles-ci peuvent néanmoins constituer une proportion importante des décès à certains âges. Par exemple, les causes externes représentent 55% des décès pour les individus d'âge entre 10 et 20 ans, et 46% des décès pour ceux entre 20 et 40 ans, ce qui est caractéristique de la *bosse à accidents*.

Pour approfondir l'analyse de la mortalité aux différents âges, revenons à l'estimation des taux de mortalité par cause. Ceux-ci serviront de paramètres de base au modèle dynamique de population. Pour chacune des causes (élémentaires ou agrégées), chaque sexe et chaque classe d'âge, on dispose du nombre de décès et de l'exposition au risque estimée. Nous estimons alors le taux de mortalité pour chaque classe d'âge par le ratio du nombre de décès et de l'exposition au risque. En fait, les données relatives au nombre de décès sont données par classes d'âge de cinq ans, excepté les classes "0", "1-4" et "95+". Les données d'exposition au risque sont quand à elles indiquées pour ces mêmes classes d'âge, sauf la dernière classe "85+". Afin d'utiliser l'information précieuse de la mortalité aux grands âges, nous utilisons les expositions au risque estimées de la base de données HMD (Human Mortality Database) pour les classes d'âge "85-89", "90-94" et "95+". Pour avoir une intuition sur la faisabilité de cette approximation entre les deux bases de données, nous avons comparé les expositions au risque OMS et HMD, et avons mesuré des différences relatives toujours plus faibles que 2%. Les taux de décès ainsi estimés par cause sont représentés sur la Figure 1.26. Une première remarque est que la forme de la mortalité en fonction de l'âge est très différente selon la cause considérée. Ceci permet d'analyser la mortalité générale de manière plus fine. En particulier, l'exemple le plus frappant est la mortalité due aux causes externes (qui représentent essentiellement les accidents) pour laquelle on récupère la "bosse à accidents" de la mortalité générale. Ces différentes formes suggèrent qu'une diminution de cause modifiera la population de manière très spécifique.

Intitulé de la classe	%
Cancers	30 %
Autres causes	28 %
Maladies de l'appareil circulatoire	27 %
Maladies de l'appareil respiratoire	6 %
Causes externes	7 %
Maladies infectieuses et parasitaires	2 %
Total	100 %

TABLE 1.3 – Répartition des décès par classe de cause pour la France en 2008

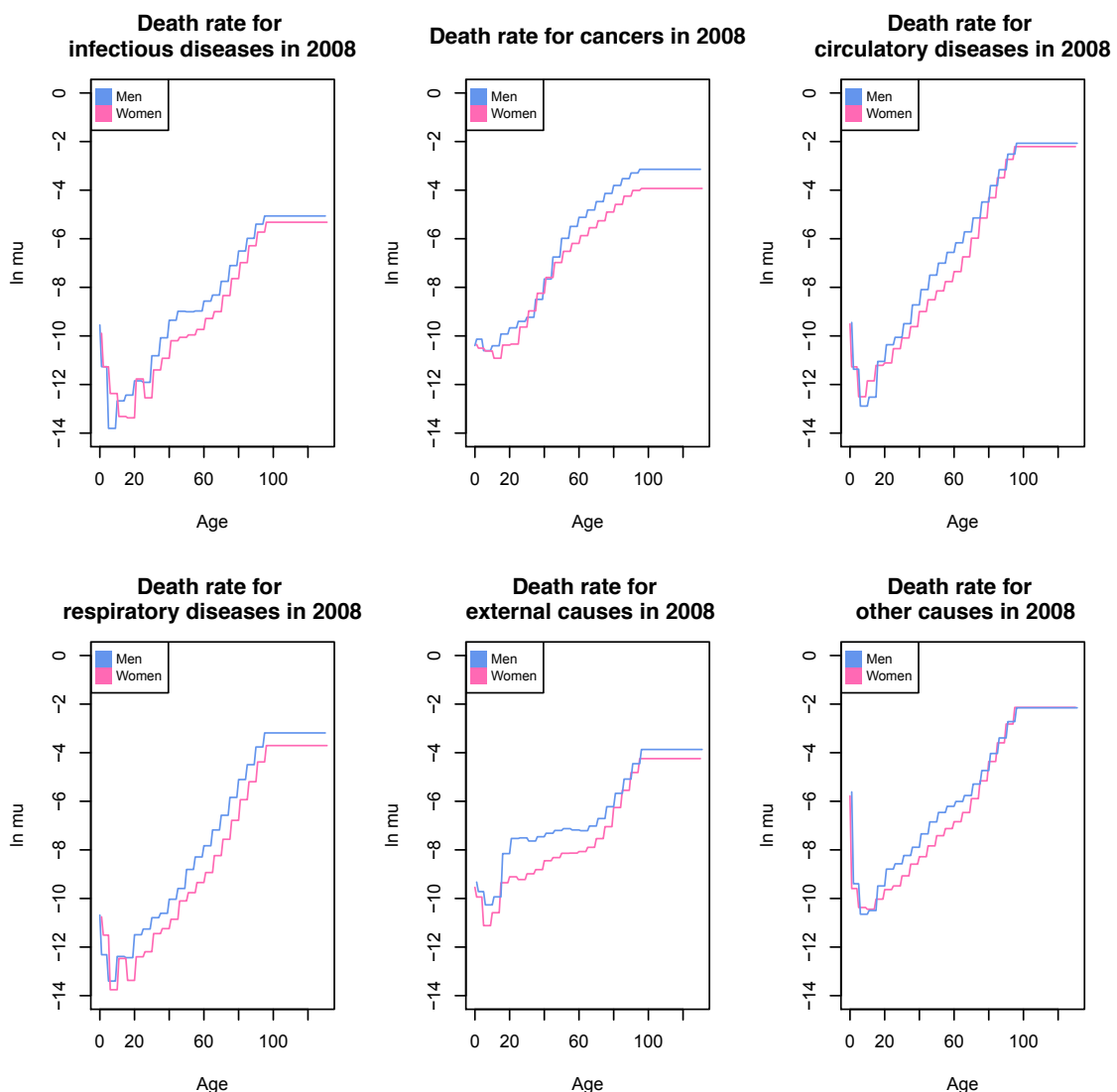


FIGURE 1.26 – Taux de mortalité par cause de décès estimés pour la population française en 2008

**Comment supprimer une cause de décès ?** Le cadre du modèle à risques compétitifs permet en théorie de tester comment se déforme la mortalité si l'on "supprime" une cause, par exemple pour modéliser qu'un médicament permet de soigner les maladies d'un type donné. Pour modéliser une telle suppression, on a besoin d'une définition rigoureuse. Dans la littérature des modèles à risques compétitifs, deux définitions ont été partiellement étudiées et testées : *ignorer* et *éliminer* (voir Elandt-Johnson (1976) et Dimitrova et al. (2013)).

(i) **IGNORER** On dit que la cause  $k$  est ignorée si la durée de vie  $\tau^\epsilon$  est remplacée par  $\tau_{(-k)}^\epsilon = \min_{1 \leq i \leq n, i \neq k} \tau_i^\epsilon$ . Avec cette définition, la variable aléatoire  $\tau_{(-k)}^\epsilon$  s'interprète comme la durée de vie des individus de sexe  $\epsilon$  dans un monde où la cause  $k$  n'existe plus.

(ii) **ELIMINER** Une autre définition est l'élimination de la cause  $k$ . Dans ce cas, la durée de vie est conditionnée au fait que la durée de vie  $\tau_k^\epsilon$  associée à la cause  $k$  tend vers l'infini. La fonction de survie devient donc  $\mathbb{P}(\tau^\epsilon > t \mid \tau_k^\epsilon \rightarrow +\infty)$ . Avec cette définition, on modélise le fait que la durée de vie associée à la cause  $k$  est repoussée à une valeur arbitrairement grande.

Malgré leurs différences en terme d'interprétation, ces deux définitions sont équivalentes lorsque les durées de vie associées à chaque cause sont indépendantes les unes des autres. On utilisera alors le même terme de *suppression* pour faire référence à l'une ou l'autre de ces définitions. Une autre conséquence de l'indépendance, autre que l'estimation directe des marginales à partir des données, est donc de faire correspondre ces deux définitions et ainsi de simplifier l'étude dans cette direction.

La suppression complète d'une cause de décès ayant un poids important pouvant paraître irréaliste, nous aurons besoin de tester la diminution d'une cause donnée. On dira que la cause  $k$  est réduite d'un facteur  $\alpha^\epsilon \in [0, 1]$  pour les individus de sexe  $\epsilon$  si le taux de décès  $\mu^\epsilon(a)$  devient

$$(1 - \alpha^\epsilon)\mu_k^\epsilon(a) + \sum_{1 \leq i \leq n, i \neq k} \mu_i^\epsilon(a).$$

### 1.4.2 Dynamique de pyramide des âges et ratio de dépendance

Les taux de mortalité par cause de décès permettent de comprendre les durées de vie individuelles des hommes et des femmes dans une population statique, dans le cas d'une réduction ou d'une élimination de cause de décès. Cependant, cette étude statique basée sur la mortalité ne donne pas d'information sur l'évolution de la population ni l'interdépendance entre la réduction de causes d'une part et le niveau de fertilité ou de migration d'autre part. Pour comprendre l'évolution de la population, en particulier la dynamique de la pyramide des âges, il est nécessaire de reprendre le modèle de population introduit dans la Partie 1.2, décliné ici sous la

forme d'un modèle à deux sexes structuré en âge. Pour comprendre cette dynamique, il faut s'intéresser à la structure en âge de la fertilité, ce que nous discutons dans la suite.

**Taux de naissance** Dans le modèle de population, chaque individu de sexe  $\epsilon$  et d'âge  $a$  donne naissance au taux  $b^\epsilon(a)$ . Rappelons que dans la dynamique aléatoire, cela se traduit par le fait que le premier temps de naissance  $T_b^\epsilon$  est donné par sa fonction de survie  $\mathbb{P}(T_b^\epsilon > a) = \exp(-\int_0^a b^\epsilon(s)ds)$ , qui représente la probabilité que l'individu donne pour la première fois naissance après l'âge  $a$ . Cet individu peut donner naissance à plusieurs enfants : les temps successifs de naissance sont distribués suivant les temps de saut d'un processus de Poisson non-homogène  $(N_a^b)_{a \geq 0}$  d'intensité  $a \mapsto b^\epsilon(a)$ , arrêté au décès de l'individu au bout d'une durée  $\tau^\epsilon$ . Suite à cette remarque, on peut par exemple calculer le nombre moyen d'enfants par individu :

$$\mathbb{E}[N_{\tau^\epsilon}^b] = \int_0^{+\infty} b^\epsilon(a) e^{-\int_0^a \mu^\epsilon(s)ds} da.$$

À la naissance, le nouvel individu est une femme avec probabilité  $p$  et un homme avec probabilité  $1 - p$ . Dans notre application, nous supposons que seules les femmes peuvent donner naissance. Ce choix est guidé par l'absence de données fiables sur l'âge du père à la naissance, et donc sur l'estimation de potentiels taux de naissance pour les hommes. Inclure un taux de naissance pour les hommes serait intéressant, par exemple pour modéliser les chocs de naissance autour des périodes de guerre. En outre, on pourrait faire dépendre le taux de naissance individuel pour les femmes par exemple de la taille totale de la population, ou des tailles respectives des populations hommes ou femmes. Nous avons réalisé plusieurs tests d'estimation dans ce sens, mais ceux-ci n'ont exhibé aucune relation exploitable. Ces aspects sont laissés pour des recherches futures.

**Dynamique déterministe** Nous commençons par rappeler la dynamique déterministe de la population, qui est un cas particulier de l'approximation grande population du modèle présenté dans la Partie 1.2. Les modèles à deux sexes font toutefois l'objet d'une littérature bien spécifique, en particulier en démographie (voir par exemple Iannelli et al. (2005)). Cette littérature est discutée dans le Chapitre 4, auquel nous renvoyons le lecteur pour plus de détails. Notons  $g(\epsilon, a, t)$  la quantité d'individus de sexe  $\epsilon$  and age  $a$  at time  $t$ , et introduisons le vecteur  $\mathbf{g}(a, t) = \begin{pmatrix} g(f, a, t) \\ g(m, a, t) \end{pmatrix}$ . Sa dynamique est donnée par l'équation aux dérivées partielles suivante :

$$(\partial_a + \partial_t)\mathbf{g}(a, t) = -D(a)\mathbf{g}(a, t), \tag{1.49}$$

$$\mathbf{g}(0, t) = \left( \int_{\mathbb{R}_+} g(f, a, t) b(a) da \right) \begin{pmatrix} p \\ 1 - p \end{pmatrix}, \quad (1.50)$$

avec  $D(a) = \begin{pmatrix} \mu^f(a) & 0 \\ 0 & \mu^m(a) \end{pmatrix}$ , et une condition initiale donnée  $\mathbf{g}(a, 0) = \mathbf{g}_0(a)$  pour tout  $a \geq 0$ . L'interprétation d'une telle équation a été faite dans la Partie 1.2. Rappelons seulement ici que nous travaillons avec une *quantité* d'individus au sens où  $g(f, a, t)$  et  $g(m, a, t)$  sont des nombres réels, qui représentent la population *moyenne*. Notons que dans notre application particulière, ces équations illustrent le comportement autonome de la population des femmes, tandis que renouvellement de la population des hommes est calculé à chaque instant à partir de la population de femmes. L'avantage de cette dynamique déterministe est d'être rapide à mettre en œuvre d'un point de vue numérique, grâce à un schéma de discrétisation classique qui est décrit dans le Chapitre 4. Ceci nous permettra d'avoir des premières intuitions sur la dynamique de la pyramide des âges au cours du temps.

**De la dynamique déterministe à la dynamique aléatoire** L'équation déterministe nous donne de l'information sur l'évolution de la pyramide des âges *moyenne*. En pratique, on s'intéresse à cette pyramide des âges mais également à des quantités non-linéaires comme le ratio de dépendance. Celui-ci est défini ici comme la population des personnes âgées d'au moins 65 ans divisé par la population des personnes entre 15 et 65 ans. Notons  $g_t(a) = g(m, a, t) + g(f, a, t)$  la quantité totale d'individus d'âge  $a$  au temps  $t$ . Reprenons également la notation  $g_t([a_1, a_2]) = \int_{[a_1, a_2]} g_t(a) da$  pour la quantité d'individus d'âge  $a \in [a_1, a_2]$  au temps  $t$ . Formellement, le ratio de dépendance  $r_t$  peut être calculé sur la pyramide des âges déterministe de la façon suivante :

$$r_t = \frac{g_t([65, \infty))}{g_t([15, 65))}.$$

L'hypothèse d'une dynamique déterministe de la population n'est de manière évidente pas réaliste. En pratique, on analyse des populations de taille finie dans lesquelles les individus ont des temps aléatoires de mort et de naissance. Une telle dynamique aléatoire a été décrite dans la Partie 1.2 et sera détaillée dans son cadre général dans le Chapitre 2. Analysons le ratio de dépendance stochastique, noté  $R_t$ , résultat de la dynamique aléatoire. Notons ici que très peu d'information est donnée par la dynamique déterministe. En particulier, le ratio de dépendance moyen  $\mathbb{E}[R_t]$  ne peut pas être directement calculé à partir de l'équation déterministe, celui-ci n'étant pas égal au ratio  $r_t$  des populations moyennes. Pour cela, il est nécessaire d'inclure un cadre stochastique à l'étude sur les causes de décès, qui pourra servir dans l'analyse des quantités non-linéaires. Rappelons les notations de la dynamique aléatoire : soit  $Z_t^\epsilon([a_1, a_2])$  le nombre d'individus d'âge  $a \in [a_1, a_2]$  et de sexe  $\epsilon$  au temps  $t$ . Rappelons que l'une des différences par rapport à la dynamique déterministe est que l'on

regarde des nombres (aléatoires) d'individus. La population  $Z_t^\varepsilon$ , qui est une mesure sur l'espace des âges, évolue suivant les taux de naissance et de mort introduits précédemment et suivant également la dynamique aléatoire décrite dans la Partie 1.2. Introduisons le nombre total d'individus  $Z_t([a_1, a_2]) = Z_t^f([a_1, a_2]) + Z_t^m([a_1, a_2])$ . Dans le modèle microscopique, le ratio de dépendance stochastique est défini formellement par

$$R_t = \frac{Z_t([65, \infty))}{Z_t([15, 65))}.$$

Comme détaillé en Section 1.2, étant donné que les taux individuels de naissance et mort ne dépendent pas de la population, la densité moyenne de population satisfait  $\mathbb{E}[Z_t([a_1, a_2])] = g_t([a_1, a_2])$ , sous la condition d'existence d'une telle densité. Cependant, le ratio de dépendance  $R_t$  est une quantité non-linéaire. En particulier, par la définition de la covariance,

$$r_t = \mathbb{E}[R_t] + \text{Cov}\left(R_t, \frac{Z_t([15, 65))}{g_t([15, 65))}\right).$$

En d'autres termes, le ratio de dépendance déterministe  $r_t$  est le ratio de dépendance moyen  $\mathbb{E}[R_t]$  plus un terme correctif relatif à la covariance entre le ratio de dépendance et la déviation du nombre d'individus d'âge dans  $[15, 65)$  par rapport à sa moyenne. Cela montre en particulier que les quantités d'intérêt de base calculées sur le ratio de dépendance stochastique ne sont pas bien décrites par l'équation déterministe (1.49)-(1.50). Cependant, nous pouvons être plus précis concernant le lien entre la dynamique microscopique et l'équation déterministe (1.49)-(1.50). Ceci a été décrit dans la Section 1.2, et sera détaillé dans le Chapitre 2 : à l'échelle microscopique, les individus se comportent aléatoirement, tandis qu'à l'échelle macroscopique, la dynamique est moyennée et suit l'équation déterministe. Le lien micro-macro sera également exploré dans le Chapitre 6 lorsque les trajectoires de vie sont accélérées.

### 1.4.3 Quelques résultats

Nous présentons ici une synthèse des résultats obtenus, qui seront repris et détaillés dans le Chapitre 4. Dans un premier temps, nous présentons les scénarios obtenus en considérant l'ensemble des causes de décès. Ceux-ci servent de référence pour comparer les simulations dans le cas de réduction ou élimination de cause de décès. Dans un deuxième temps, nous présentons les scénarios obtenus en supprimant chaque cause de décès séparément. Nous illustrons en particulier la dynamique modifiée du ratio de dépendance dans ce cadre. Sur ces scénarios, on montrera la difficulté que l'on a à comparer les causes de décès entre elles, celles-ci pouvant avoir des ordres de grandeur très différents. À cette fin, nous considérerons deux causes particulières



(causes externes d'une part, et cancers d'autre part) et comparerons leur impact respectif sur le ratio de dépendance en conservant la même amélioration d'espérance de vie.

**Scénarios de référence** La population est simulée de 2008 à 2108 en prenant en compte la mortalité toutes causes. Les Figures 1.27 et 1.28 représentent la réalisation d'une trajectoire stochastique partant de 100 000 individus. La population initiale, qui fait apparaître le *Baby Boom* (les personnes entre 40 et 70 ans), est basée sur les proportions par classe d'âge de la population française en 2008. Deux premières intuitions apparaissent grâce à cette expérience numérique. La première concerne la forme de la pyramide des âges au bout d'un temps long : celle-ci se caractérise par une répartition presque uniforme jusqu'à 60 ans, puis décroît exponentiellement. En fait, celle-ci dépend à la fois des taux de mort mais aussi des taux de naissance de la population, et l'on dispose de sa forme de manière semi-explicite (voir Chapitre 4). La deuxième intuition concerne l'évolution du ratio de dépendance dans le temps. En particulier, on mesure avec cette expérience l'impact dans le temps de la configuration initiale de la pyramide des âges : le ratio de dépendance augmente jusqu'à l'extinction des générations du *Baby Boom*, puis se stabilise. Cette expérience numérique nous permet donc d'illustrer le caractère dynamique de la population : même sur la base de taux de naissance et de mort indépendants du temps, on peut exhiber une dynamique temporelle cruciale, comme celle du ratio de dépendance sur les 20 premières années.

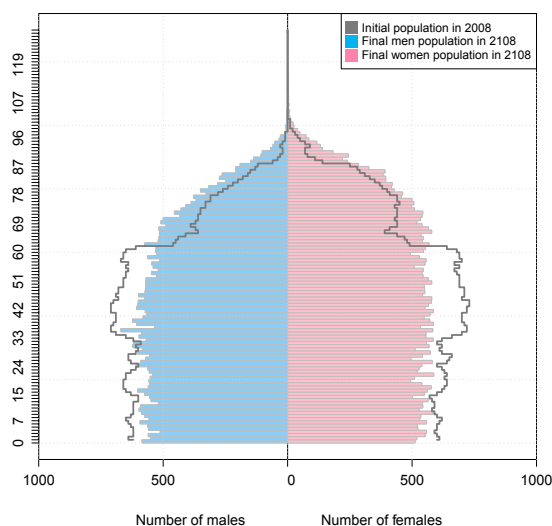


FIGURE 1.27 – Pyramide des âges initiale en ligne grise (100 000 individus) et pyramide des âges au bout de 100 ans, en couleur.

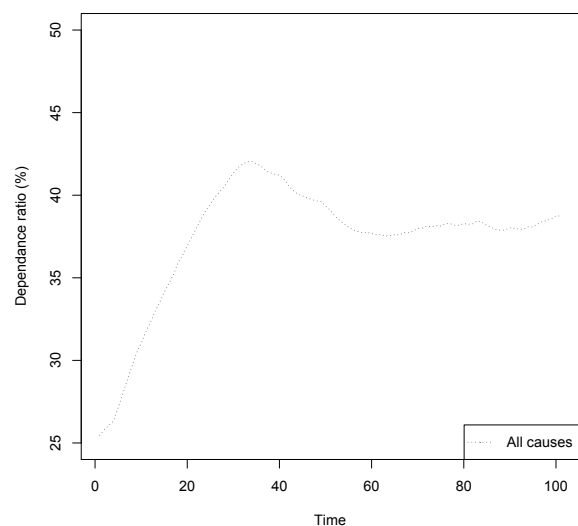


FIGURE 1.28 – Trajectoire du ratio de dépendance simulé de 2008 à 2108

Les Équations (1.49)-(1.50) qui sont valides en grande population permettent d'analyser la dynamique du ratio de dépendance déterministe. Ici, il n'est plus possible d'analyser le comportement aléatoire mais l'avantage de la dynamique déterministe est de pouvoir se calculer plus rapidement, comme ceci revient à discrétiser une équation aux dérivées partielles. On représente sur la Figure 1.29 l'évolution du ratio de dépendance sur 100 ans. Ceci correspond, à l'erreur de discrétisation près, à l'évolution du ratio de dépendance pour une population infinie. Notons la grande similitude de trajectoire avec la dynamique aléatoire pour 100 000 individus initiaux. Remarquons également la différence de trajectoire sur la fin de la simulation. Ceci s'explique par la taille de la population, qui ici décroît au cours du temps dû au fait que les flux de population ne sont pas encore pris en compte (ceci sera discuté dans le Chapitre 4). Ainsi, si l'approximation déterministe semble plutôt valide au début de la simulation, cette propriété s'estompe à mesure que la taille de la population diminue. Néanmoins, la forte similitude entre les deux laisse penser que pour des ordres de grandeur réels de populations nationales comme la France, l'approximation déterministe est justifiée. Tester ce constat nécessiterait des ressources informatiques très importantes, ce que nous n'avons pas pu mettre en oeuvre à ce stade pour ce projet. Ces aspects sont laissés pour des travaux de recherche futurs.

À partir de cette équation, il est également possible de simuler la trajectoire sur le très long terme, comme 300 ans. Bien entendu, cet horizon de temps n'a pas de sens en soi, en particulier si l'on fait l'hypothèse de taux de naissance et mort constants d'une année à l'autre. Néanmoins, cette expérience numérique illustre la stabilisation de la pyramide des âges dans les modèles de population, sur la base ici d'ordres de grandeur réels. En particulier, on peut mesurer le temps qu'il faut pour que la convergence s'effectue. Cette stabilisation est un fait bien connu en démographie. Dans le Chapitre 4, nous discutons les principaux résultats connus sur cette pyramide stable.

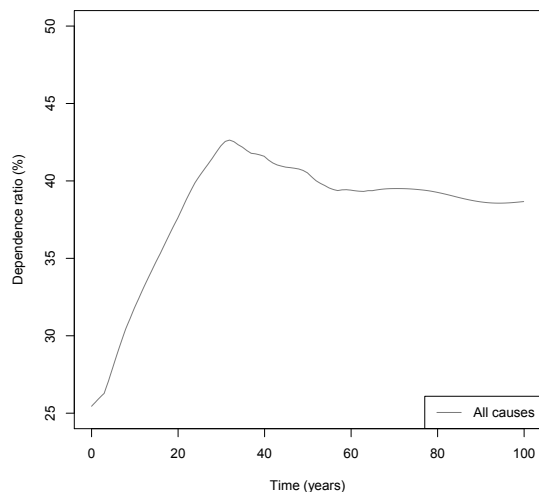


FIGURE 1.29 – Ratio de dépendance déterministe simulé de 2008 à 2108

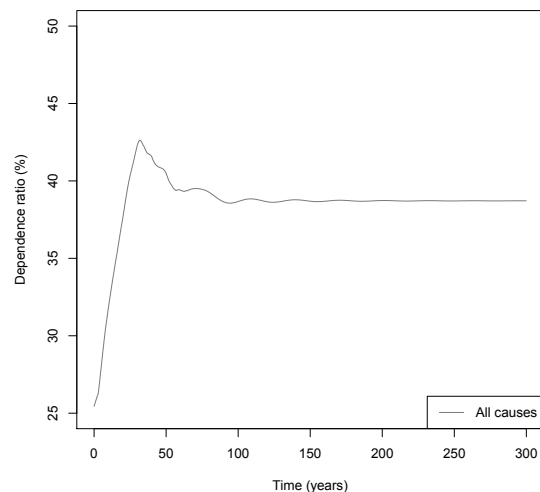
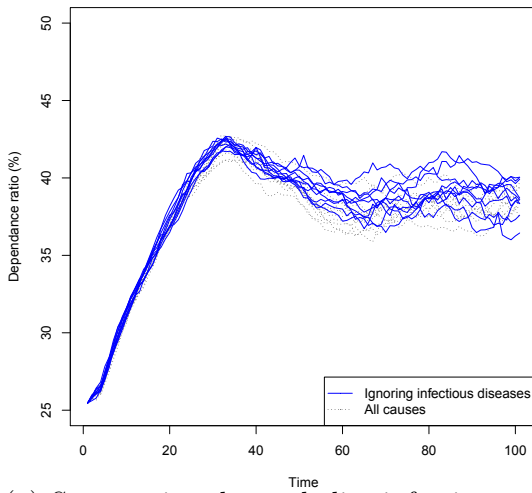


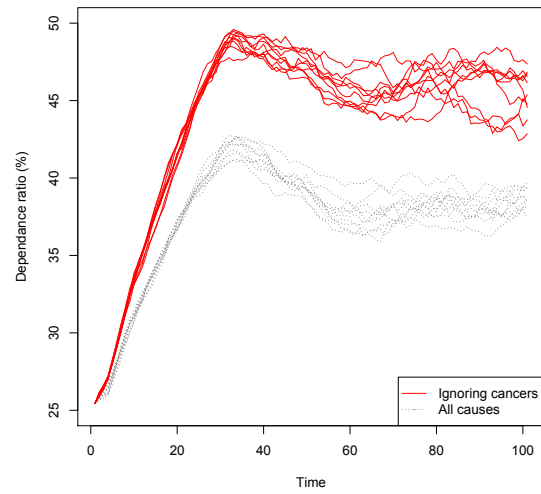
FIGURE 1.30 – Ratio de dépendance déterministe simulé de 2008 à 2308

**Scénarios de suppression de cause** Comment se comporte le ratio de dépendance si l'on supprime une cause de décès? Avec la méthode décrite dans la section précédente, nous testons l'impact de suppression de chacune des six causes de décès sur la dynamique du ratio de dépendance. La population initiale est composée de 10 000 individus et sa répartition en classes d'âges et par sexe est la même que la population française en 2008. Nous représentons des trajectoires stochastiques du ratio de dépendance sur les Figures 1.31a à 1.31f. La suppression d'une cause implique, pour un bruit aléatoire donné, un vieillissement de la population et donc une augmentation du ratio de dépendance. Comme attendu, les causes ayant un taux de mortalité faible (infectieuses et parasitaires, maladies de l'appareil respiratoire, causes externes) donnent des ratios de dépendance moins élevés que celles avec un taux de décès important. Notons que par la nature stochastique de la dynamique, il est néanmoins possible à ce stade d'observer des trajectoires du ratio de dépendance en dessous d'une trajectoire de référence (voir par exemple les Figures 1.31a et 1.31e). Ainsi, ces différences peuvent se trouver atténuées ou accentuées par le bruit démographique. Sur ces graphiques, il apparaît que la comparaison entre les causes de décès n'est pas évidente car les ordres de grandeur de chaque cause sont très différents. De plus, l'élimination complète de causes de décès peut paraître irréaliste, en particulier pour celles qui présentent un taux de décès important. Dans la suite, nous nous focalisons sur le cancer d'une part, et les causes externes d'autre part et nous nous intéressons à l'impact sur la pyramide des âges lorsque chaque cause est modifiée de sorte à répliquer la même amélioration de l'espérance de vie.

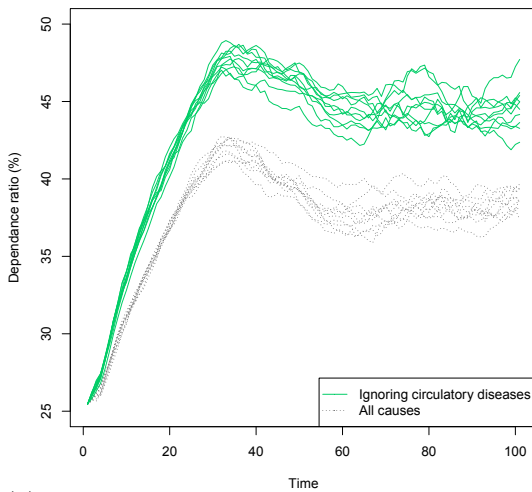
FIGURE 1.31 – Ratios de dépendance simulés de 2008 à 2108 dans le modèle stochastique



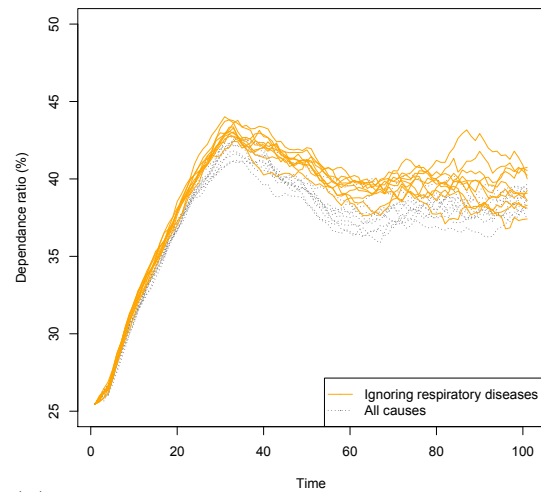
(a) Suppression des maladies infectieuses et parasitaires



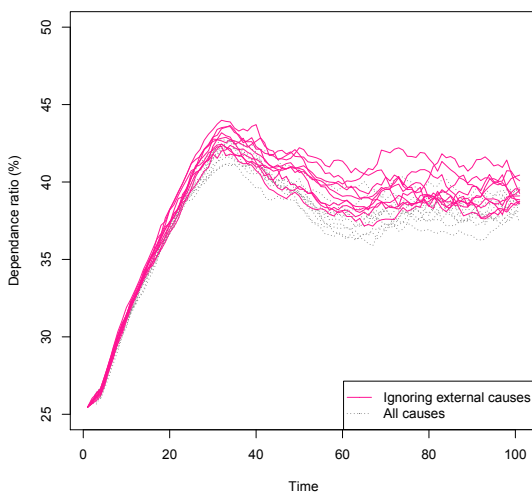
(b) Suppression des cancers



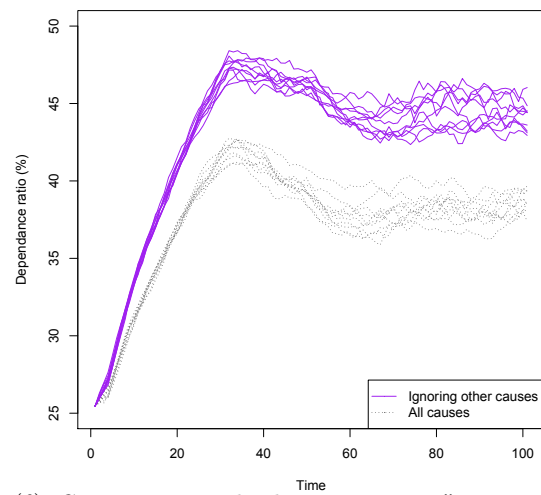
(c) Suppression des maladies de l'appareil circulatoire



(d) Suppression des maladies de l'appareil respiratoire



(e) Suppression des causes externes



(f) Suppression de la catégorie "autres causes"

**Projections avec conservation de l'espérance de vie** Avec les simulations précédentes, il apparaît difficile de comparer l'impact de deux causes de décès dont les ordres de grandeur sont très différents. Ce que l'on souhaite étudier ici est la spécificité de la structure en âge de différentes causes de décès à travers la modification de pyramide des âges qu'elle engendre. Pour ce faire, nous adoptons un critère de comparaison commun qui est l'espérance de vie résultant de la réduction de cause. Ainsi, nous allons tenter de détecter si la structure en âge de causes de décès est suffisamment différente pour qu'à espérance de vie fixée, la pyramide des âges soit modifiée de manière distincte.

Pour cette expérience numérique, nous nous concentrons sur deux grandes causes connues pour avoir des structures en âge très différentes : le cancer et les causes externes. Le cancer concerne majoritairement les âges au dessus de 40 ans, tandis que les causes externes, qui capturent la bosse à accidents, ont de l'impact sur les âges autour de 25 ans (voir Figure 1.26). Nous souhaitons comparer les trois configurations suivantes :

(i) La première est la configuration de référence, simulée à partir des taux de naissance et mort actuels. L'espérance de vie à la naissance dans ce scénario est de 84.4 ans pour les femmes et de 77.7 ans pour les hommes.

(ii) La deuxième configuration est obtenue en supprimant les causes externes. Sous l'hypothèse d'indépendance entre les causes de décès, l'espérance de vie à la naissance devient 85.1 pour les femmes et 79.1 ans pour les hommes.

(iii) La dernière configuration est obtenue en réduisant le cancer de la façon suivante : chaque facteur de réduction pour les femmes et les hommes  $\alpha^f$  et  $\alpha^m$  (voir Partie 1.4.1) est calculé de sorte à répliquer les espérances de vie obtenues en supprimant les causes externes. On calibre ainsi les facteurs de réduction du cancer  $\alpha^f = 24.6\%$  et  $\alpha^m = 35\%$ .

En comparant (ii) et (iii), on souhaite étudier la différence de dynamique de pyramide des âges et en particulier quantifier l'impact sur le ratio de dépendance.

Les scénarios déterministes du ratio de dépendance correspondant aux configurations (i), (ii) et (iii) sont représentés sur la Figure 1.32. La réduction ou la suppression de cause de décès donne ici des ratios de dépendance plus élevés par rapport à la configuration de référence. Ceci peut paraître naturel, mais on pourrait en théorie construire des exemples pour lesquels une diminution de la mortalité peut donner une population plus jeune (au sens d'un ratio de dépendance plus faible). On peut trouver quelques études démographiques qui traitent de cet aspect (voir par exemple Keyfitz (1968)), que nous n'avons toutefois pas observé avec ces données et cet exemple précis. À notre niveau, nous allons voir en comparant les configurations (ii) et (iii) qu'avec la même espérance de vie, une réduction de cause peut donner une population "moins vieille que prévue". Ceci est représenté sur la Figure 1.32,

sur laquelle on observe que la suppression des causes externes limite l'augmentation du ratio de dépendance. Comme cette cause concerne des âges auxquels les individus peuvent donner naissance, les individus qui survivent grâce à la suppression de cause assurent une partie du renouvellement de la population. Quant au cancer, sa réduction donne un ratio de dépendance plus élevé, car les personnes concernées survivent à des âges ne permettant plus de donner naissance. Ainsi, l'impact sur la pyramide des âges dépend fortement des âges auxquels la mortalité a été réduite, ce qui est caractéristique de chaque cause de décès. Rappelons ici que les causes externes concernent majoritairement les hommes (voir Figure 1.26), alors que nos hypothèses (qui reflètent les données) ne font dépendre la fertilité que de la population de femmes. Ainsi, en cas de réduction des causes externes, le nombre de naissances augmentera dû à la population de femmes qui survit aux causes externes, mais cet effet pourrait être renforcé si nous avions défini un taux de naissance pour les hommes.

Nous représentons sur la Figure 1.33 l'évolution dans le temps de la taille de la population dans les différentes configurations. Il est intéressant de remarquer qu'alors que la répartition en âges se stabilise, la taille de la population décroît. Ceci n'est pas vraiment cohérent avec les ordres de grandeur démographiques actuels, mais nous l'expliquons par le fait que les flux de population n'ont pas été pris en compte dans ces simulations (voir le Chapitre 4 pour la prise en compte de ces flux). Sur le court terme, c'est le scénario de réduction du cancer qui donne une taille de population plus importante car l'effet de cette réduction est immédiat et important aux âges élevés. En revanche, comme on peut s'y attendre, la taille moyenne de la population est plus importante sur le long terme dans le scénario de suppression des causes externes dû à l'effet supplémentaire d'augmentation du nombre de naissances.

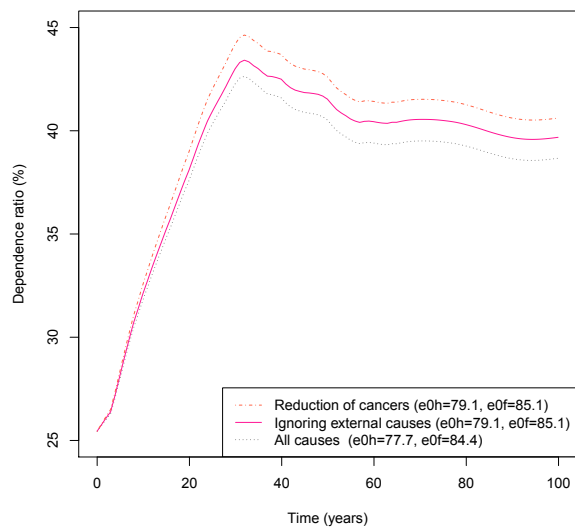


FIGURE 1.32 – Ratio de dépendance simulé à partir de l'équation déterministe (1.49)-(1.50) sur 100 ans

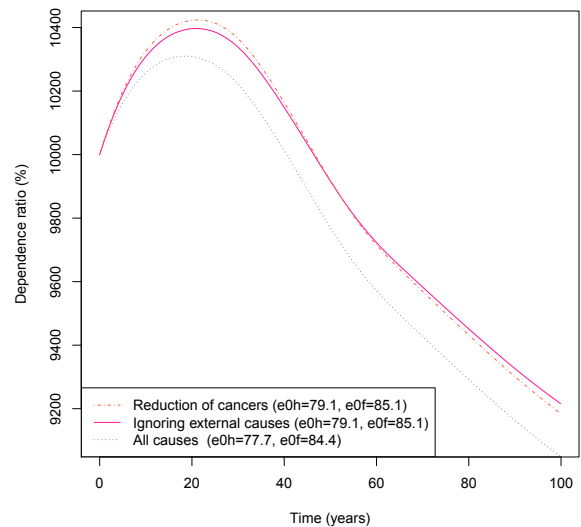


FIGURE 1.33 – Taille de la population simulée à partir de l'équation déterministe (1.49)-(1.50) sur 100 ans

#### 1.4.4 Conclusion et perspectives

Ces premières analyses de l'impact de changements de cause de décès au niveau de la population mettent en évidence l'information contenue par la pyramide des âges. En particulier, pour deux scénarios qui ont une espérance de vie similaire, il est possible de distinguer le type de cause qui a été réduite en observant la pyramide des âges et ses équilibres internes. Nous proposons ci-dessous un certain nombre de perspectives de recherche associées à ces travaux.

(i) Une première perspective concerne la prise en compte de la dimension temporelle de la mortalité par causes de décès. Ceci aurait un intérêt, dans un premier temps en reprenant les taux de mortalité passés et en comparant des scénarios de réduction de cause à la pyramide des âges effectivement réalisée. Une vraie difficulté dans cette direction est la prise en compte des changements de classification de causes de décès qui ont eu lieu à certaines dates et occasionnent des ruptures dans les historiques de données (voir par exemple Rooney et al. (2002)).

(ii) Une deuxième perspective de recherche concerne la dépendance entre les causes de décès. Dans le modèle à risques compétitifs, la structure de dépendance entre les différentes causes n'est pas identifiable, ce qui est discuté plus en détail dans le Chapitre 4. Néanmoins, il pourrait être intéressant de tester la sensibilité de nos conclusions à la dépendance entre les causes en testant différents types de structures et de niveaux. Dans un cadre statique, il pourrait être intéressant de partir des travaux de Carriere (1994) et Dimitrova et al. (2013) pour tester différentes copules. La difficulté supplémentaire dans un cadre dynamique réside dans la définition d'une

structure de dépendance qui évolue dans le temps.

(iii) Enfin, toujours sur la question de la dépendance entre les causes de décès, et à la vue des conclusions de ce travail, il serait intéressant de tenter d'estimer la structure de dépendance entre les causes de décès dans un modèle à risques compétitifs, par l'observation jointe de la pyramide des âges d'une part, et des grandes dates de réduction de causes (innovation médicale) d'autre part. Ce travail d'estimation statistique serait peut-être à envisager dans le cadre de la démarche de Hoffmann and Olivier (2014) qui s'intéresse à l'estimation non-paramétrique du taux de division (par exemple cellulaire) dans des modèles de populations structurées en âge (voir aussi par exemple Doumic et al. (2012) et Doumic et al. (2015)).



## 1.5 Estimation de l'hétérogénéité de la mortalité en présence de censure par intervalle

L'étude de la dynamique des populations structurées en âge nous donne de nombreuses intuitions sur le lien entre la structure en âge des taux démographiques et la dynamique de la pyramide des âges. Ce point de vue a été décliné dans un cadre théorique à l'étude de processus d'immigration-naissance avec âge (les processus de Hawkes, Partie 1.3), et dans une application à des problématiques démographiques par l'étude de l'impact de réduction de causes de décès au niveau individuel sur la dynamique globale de la répartition des générations (Partie 1.4).

Nous souhaitons dans cette partie aborder la question de l'estimation statistique de l'hétérogénéité de la mortalité. L'objectif est d'estimer des forces de mortalité spécifiques à des caractéristiques propres dans le cas général où celles-ci peuvent changer au cours de la vie. Cette étude se base sur des données longitudinales issues du recensement, et l'on fait donc face à un schéma d'observation dit *censuré par intervalle* car les individus ne sont observés qu'à des temps isolés : ainsi les dates de changements de caractéristiques sont inconnus. Nous nous intéressons dans ce cadre à la dynamique de générations avec une composition spécifique en terme de statut matrimonial ou de catégorie socio-professionnelle, ainsi que son impact sur des quantités agrégées comme la structure en âge de la mortalité globale observée, mais également dans une optique actuarielle sur des produits types d'assurance-vie.

### 1.5.1 Description des données longitudinales de l'INSEE

Comme nous l'avons vu, la mortalité dépend au niveau individuel d'un certain nombre de facteurs observables comme l'âge et le sexe. À un niveau plus fin, on peut avoir accès dans certaines bases de données à des caractéristiques individuelles supplémentaires comme par exemple des caractéristiques socio-économiques. Nous avons en particulier accès à une partie de la base de l'INSEE, appelée *Échantillon Démographique Permanent\**, que nous avons en partie décrite dans la Partie 1.1.3. Nous proposons de compléter ici la description de ces données auxquelles nous avons accès, en rappelant d'abord quelques éléments.

Cette base de données contient des informations sur la situation de 992711 personnes observées à partir de 1968 et nées les 1, 2, 3 ou 4 octobre à partir de l'année 1862. Ces individus ont potentiellement été recensés aux dates 1968, 1975, 1982, 1990 et 1999 et, pour chaque recensement, un ensemble de caractéristiques a été relevé. Un tel échantillon permet d'étudier le lien entre le niveau de mortalité et les

---

\*Nous remercions H. Bensusan ainsi que A. Trognon et A. Frachot à l'INSEE pour l'accès à cette base de données.

### 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

caractéristiques individuelles, mais également la manière dont ces caractéristiques évoluent dans le temps.

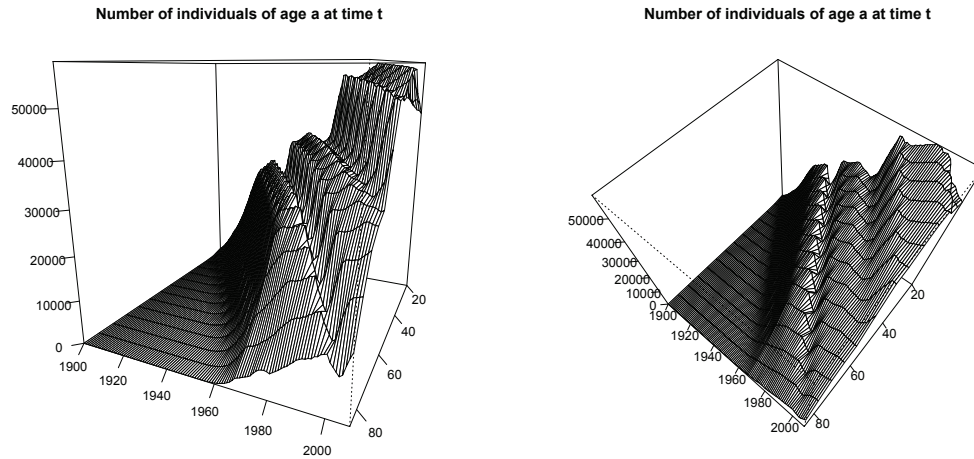


FIGURE 1.34 – Nombre d'individus par classes d'âge au cours du temps

On peut dans un premier temps s'intéresser à la taille de cet échantillon. Sur la Figure 1.34, on représente le nombre d'individus de l'échantillon appartenant à une classe d'âge de 5 ans en début d'année  $t$ , sans distinction de caractéristique individuelle. Sur ce graphique, on distingue en particulier deux cohortes sous-représentées dans l'échantillon, correspondant respectivement à celle des combattants de la Seconde Guerre Mondiale et celle des personnes nées pendant cette même guerre. On note également le grand nombre d'individus de la génération du Baby Boom.

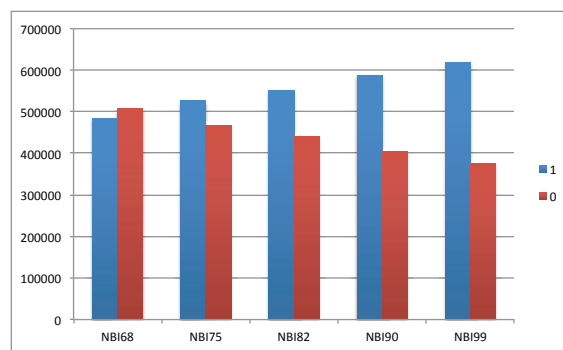


FIGURE 1.35 – Nombre d'individus recensés (1) et non recensés (0) aux dates 1968, 1975, 1982, 1990 et 1999

Le recensement régulier permet de suivre l'évolution des caractéristiques individuelles dans le temps. Sur la Figure 1.35, on représente le nombre d'individus vivants à la date de recensement qui ont été recensés (1) ou non (0) ; il apparaît que la proportion d'individus recensés augmente avec la date de recensement. À l'aide

de ces recensements, nous allons étudier en partie l'hétérogénéité de la population et son évolution dans le temps. L'analyse de cette évolution de la structure de la population entre deux recensements est complexe. En effet, pour une génération donnée, les changements de proportions entre les différentes classes au cours du temps sont dus à différents effets : d'abord le fait que la mortalité diffère selon les catégories, ensuite le fait que les individus peuvent changer de caractéristiques au cours de la vie ; la quantification de ces effets sera d'intérêt dans la suite. Enfin, ajoutons l'effet des départs et arrivées. Les arrivées de personnes au cours du temps se traduisent dans la base par des recensements non effectués précédents des recensements observés. Les départs, quant à eux, induisent des dates de décès inconnues avec le plus souvent des recensements non observés.

**Observations incomplètes** Prenons le temps de rappeler à travers des exemples les différentes formes d'observations incomplètes (voir par exemple Andersen et al. (1993)).

a) On parle de *censure à droite* lorsqu'il y a perte d'information due à l'arrêt de l'observation ; dans ce cas, on ne connaît pas la date de décès des individus encore en vie à cette date. Ainsi, on n'observe pas les dates de décès qui ont eu lieu *après la fin de l'observation*. C'est un type de censure qui est structurel en analyse de survie. La *censure à gauche* est moins commune mais peut se produire pour des procédures d'observation spécifiques. On peut reprendre l'exemple des babouins de Andersen et al. (1993) : ceux-ci dorment dans les arbres et descendent à un moment de la journée pour ne remonter que plus tard. Si on commence l'observation de l'arbre à un instant donné, la seule information dont on dispose pour les babouins manquants est que leur descente a eu lieu *avant le début de l'observation*. Qu'elle soit à droite ou à gauche, la censure correspond à une observation partielle, au sens où l'on sait que l'événement a lieu (décès ou descente) mais on ne dispose que d'une borne sur la date d'occurrence (inférieure ou supérieure).

b) Donnons maintenant deux exemples de troncature à gauche ou à droite. On parle de *troncature à gauche* lorsqu'il y a perte d'information due au début de l'observation : en analyse de survie par exemple, on n'observe pas les individus décédés *avant le début de l'observation*. La *troncature à droite* est moins commune en analyse de survie, mais se retrouve aussi pour certains types d'observation. Prenons l'exemple de la durée d'incubation d'un virus : si nous observons un sujet à partir de l'infection et que nous arrêtons l'observation, nous n'aurons alors pas accès à la durée d'incubation si les symptômes se manifestent *après la fin de l'observation*. Dans tous les cas, la troncature est relative à une absence d'information, au sens où la quantité d'intérêt (individu ou symptôme) n'est observée que si celle-ci est présente dans l'intervalle d'observation.

c) Décrivons enfin des mécanismes d'observations incomplètes plus complexes, qui seront structurels dans nos données. Nous allons dans la première partie de ce qui suit utiliser le vocabulaire de Andersen et al. (1993) et statuer sur ce point en fin de paragraphe. Les cas de censure/troncature à droite/gauche que nous avons évoqués portent tous sur une observation sous forme d'intervalle. Lorsqu'un phénomène est observé sur un ou plusieurs intervalles, on parle en fait de censure par intervalle. Il existe un cas extrême où ces intervalles sont réduits à des points : c'est le cas si on observe le phénomène à des temps isolés. Ceci est structurel dans les données issues du recensement. Dans ce cas, on trouve dans Andersen et al. (1993) la notion de *censure par intervalle extrême*. Cette notion est souvent reprise dans la littérature statistique sous le nom de *censure par intervalle*. Pour les processus multi-états en particulier, on peut aussi trouver la notion de *données de panel* lorsque l'observation est faite à des temps isolés et que de plus on ne connaît pas le nombre de transitions dans chaque intervalle. Enfin, certaines contributions précisent que la notion de censure par intervalle ne s'applique que dans le cas particulier où une seule transition peut avoir lieu dans l'intervalle. Nous donnerons dans la Partie 1.5.2 les références des contributions statistiques à la base de notre approche, qui utilisent ces dénominations. **Pour conclure** : dans la suite, nous emploierons le terme *censure par intervalle* pour désigner l'observation du phénomène à des temps isolés, et nous préciserons *données de panel* lorsque nous citerons une méthode qui met l'accent sur la prise en compte de plusieurs transitions possibles entre deux temps isolés d'observation.

Revenons à la description des données à disposition, en illustrant dans le temps l'observation des caractéristiques à disposition.

**Catégorie socio-professionnelle dans le temps** Chaque recensement fournit une photographie de la population à une date précise. On représente sur les Figures 1.36 et 1.37 la structure de ces recensements par rapport à la catégorie socio-professionnelle et selon les différentes générations, pour les femmes et les hommes respectivement. Pour cette illustration, nous choisissons de regrouper les catégories socio-professionnelles en quatre grandes classes. La première, notée 1, contient la majorité des sous-catégories de l'INSEE : agriculteurs exploitants, artisans, commerçants et assimilés, professions intermédiaires, employés, ouvriers, personnes sans activité professionnelle. La deuxième catégorie, notée 2, regroupe les cadres et professions intellectuelles supérieures, ainsi que les chefs d'entreprise de 10 salariés ou plus. La troisième catégorie, notée 3, regroupe des inactifs divers de tous âges. La quatrième catégorie, notée 4, regroupe les élèves et étudiants.

Analysons d'abord les tendances temporelles communes aux hommes et aux

femmes. Observons d'abord les générations les plus récentes, c'est-à-dire les personnes nées à partir des années 1950 (en abscisse). Au cours du temps, ces générations passent de l'état inactif (1968, 1975) à l'état élève/étudiant (1982). Puis on distingue trois classes au recensement de 1990 : étudiants, CSP-, CSP+. Notons qu'à ce recensement, les CSP+ sont sous représentées et l'on peut vérifier dans les données que ce sont surtout les personnes encore étudiantes à ce moment là qui gonfleront leurs rangs. Observons maintenant les générations les plus anciennes nées dans les années 1920. Celles-ci ne présentent naturellement plus d'étudiants ; leur état est soit inactif, soit CSP-, soit CSP+. Au fil des recensements, la composition évolue due aux changements de caractéristiques d'une part et à la mortalité d'autre part. Pour une génération donnée (hommes ou femmes), il semble que la proportion de CSP+ augmente avec le temps, tandis que celle des personnes inactives ou CSP- diminue. Même si l'effet de changement de caractéristiques a un impact, ces variations grossières suggèrent que la mortalité des CSP+ est plus faible que celle des CSP-. C'est ce type de différentiels de mortalité que nous tâcherons de quantifier dans la suite. Les différences entre hommes et femmes apparaissent de manière flagrante en comparant les Figures 1.36 et 1.37. En particulier, on retrouve le fait que les inactifs (dénomination de l'INSEE) sont particulièrement représentés chez les femmes, tandis que la CSP+ y est sous-représentée.

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

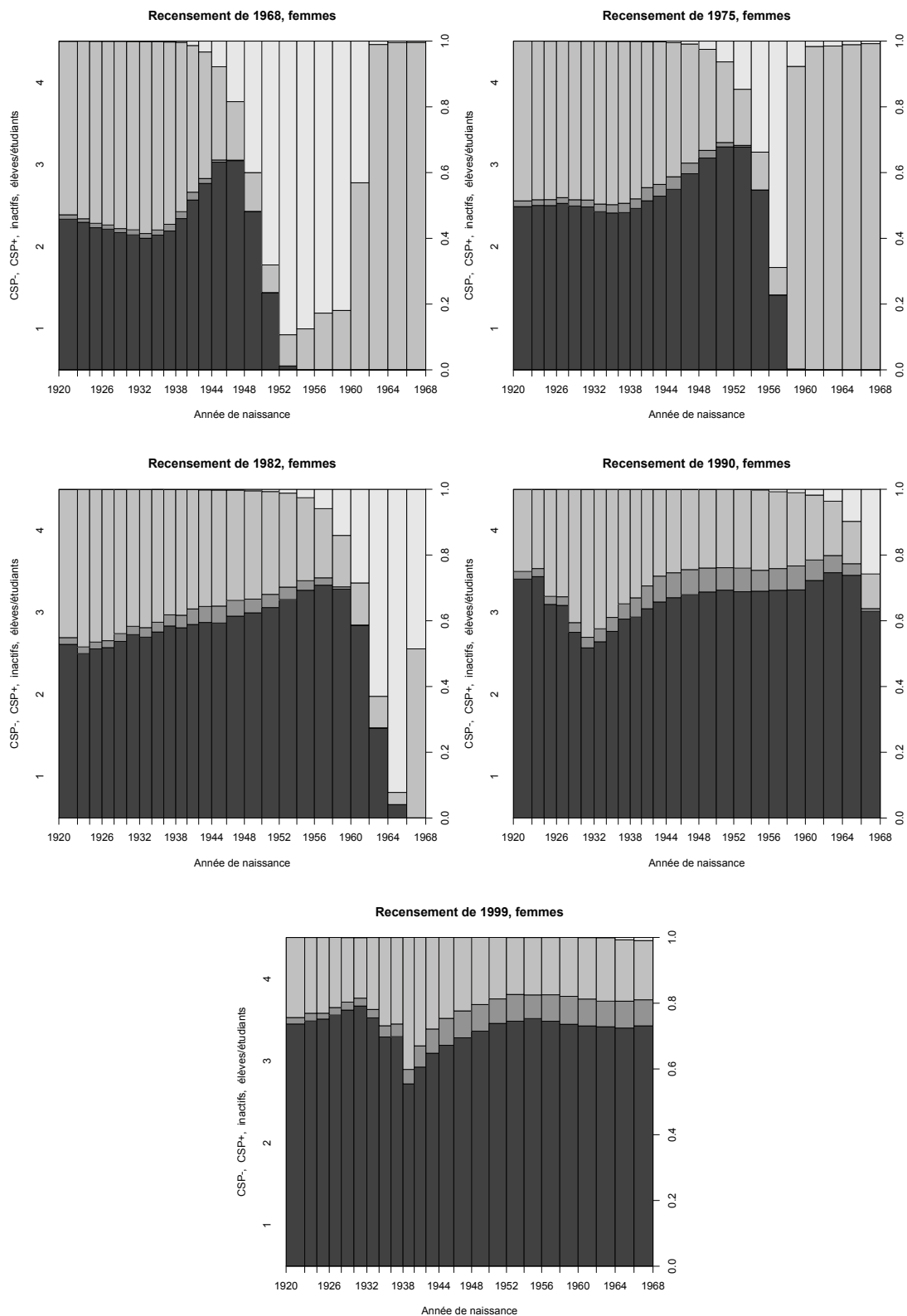


FIGURE 1.36 – Structure des recensements par catégorie socio-professionnelle et par génération pour les femmes

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

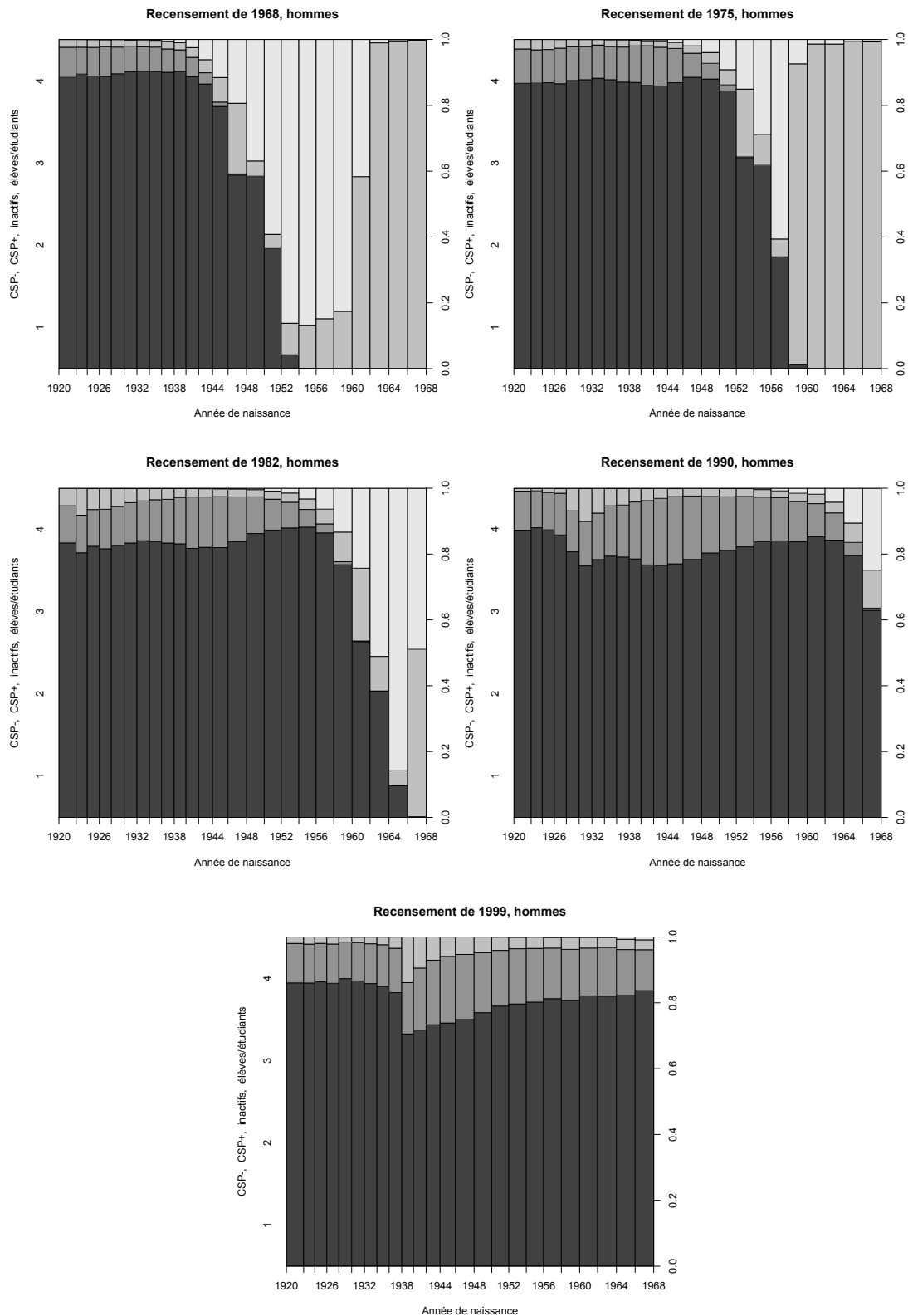


FIGURE 1.37 – Structure des recensements par catégorie socio-professionnelle et par génération pour les hommes

**Statut matrimonial dans le temps** Nous proposons la même représentation concernant cette fois le statut matrimonial, présentée en Figures 1.38 et 1.39. Nous conservons la classification de l'INSEE, codée de la façon suivante : 1 célibataire, 2 marié, 3 veuf, 4 divorcé. Pour hommes et femmes, les dynamiques d'évolution des différentes classes présentent un profil similaire. Notons bien la déformation intéressante du profil de composition au cours du temps. En particulier, en ce qui concerne les célibataires et les personnes divorcées, la dynamique est très proche. En revanche, on notera tout de même une grande différence dans la proportion restante entre les personnes veuves d'une part et mariées d'autre part. On retrouve en particulier le fait que, les femmes vivant plus longtemps, la proportion de veuves est plus importante. Notons par exemple que pour la génération des personnes nées en 1920 et recensées en 1999, les femmes sont pour environ la moitié veuves, tandis que les veufs ne représentent qu'une fraction faible des hommes de la même génération et recensés la même année.



## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

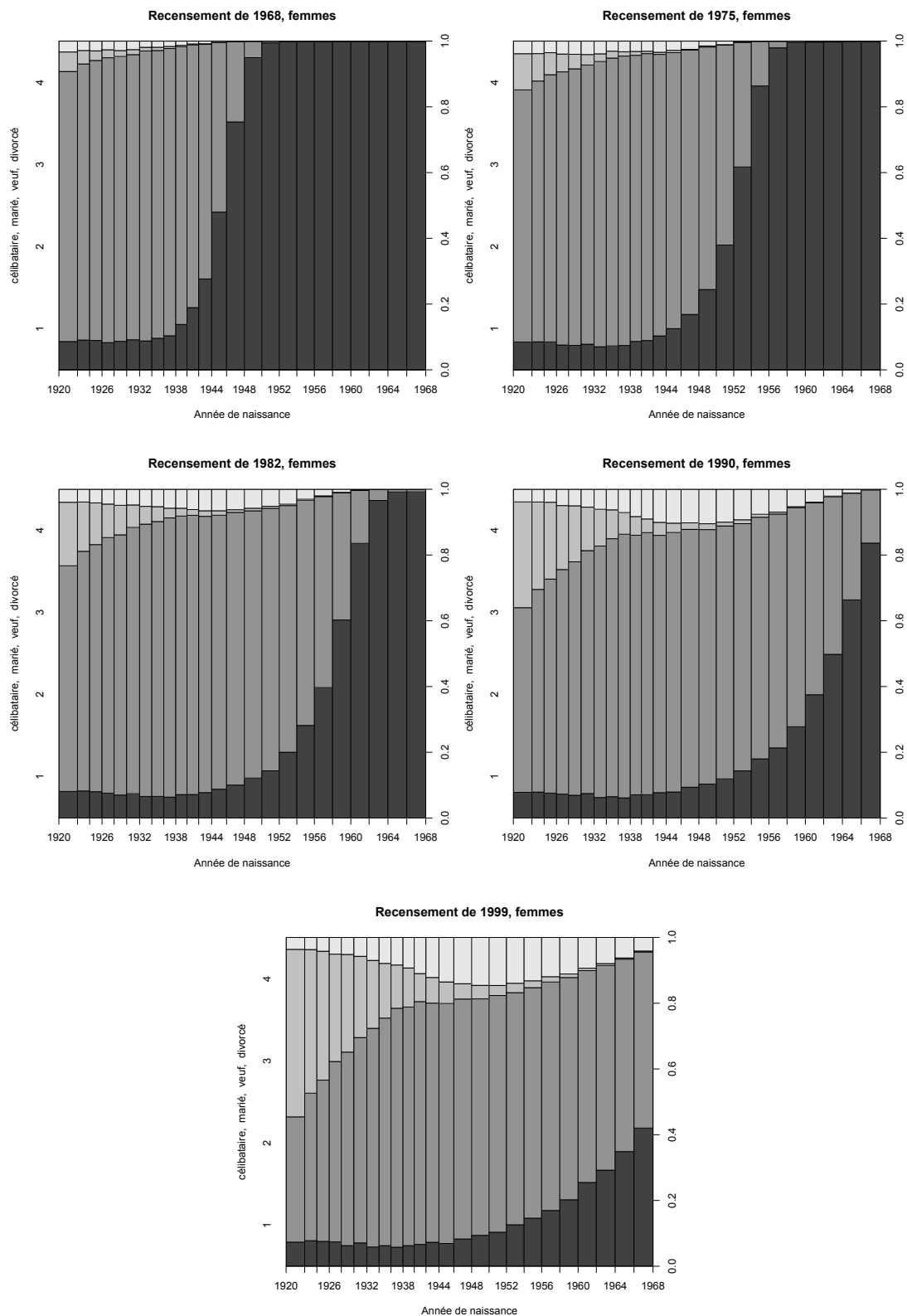


FIGURE 1.38 – Structure des recensements par statut matrimonial et par génération pour les femmes

1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

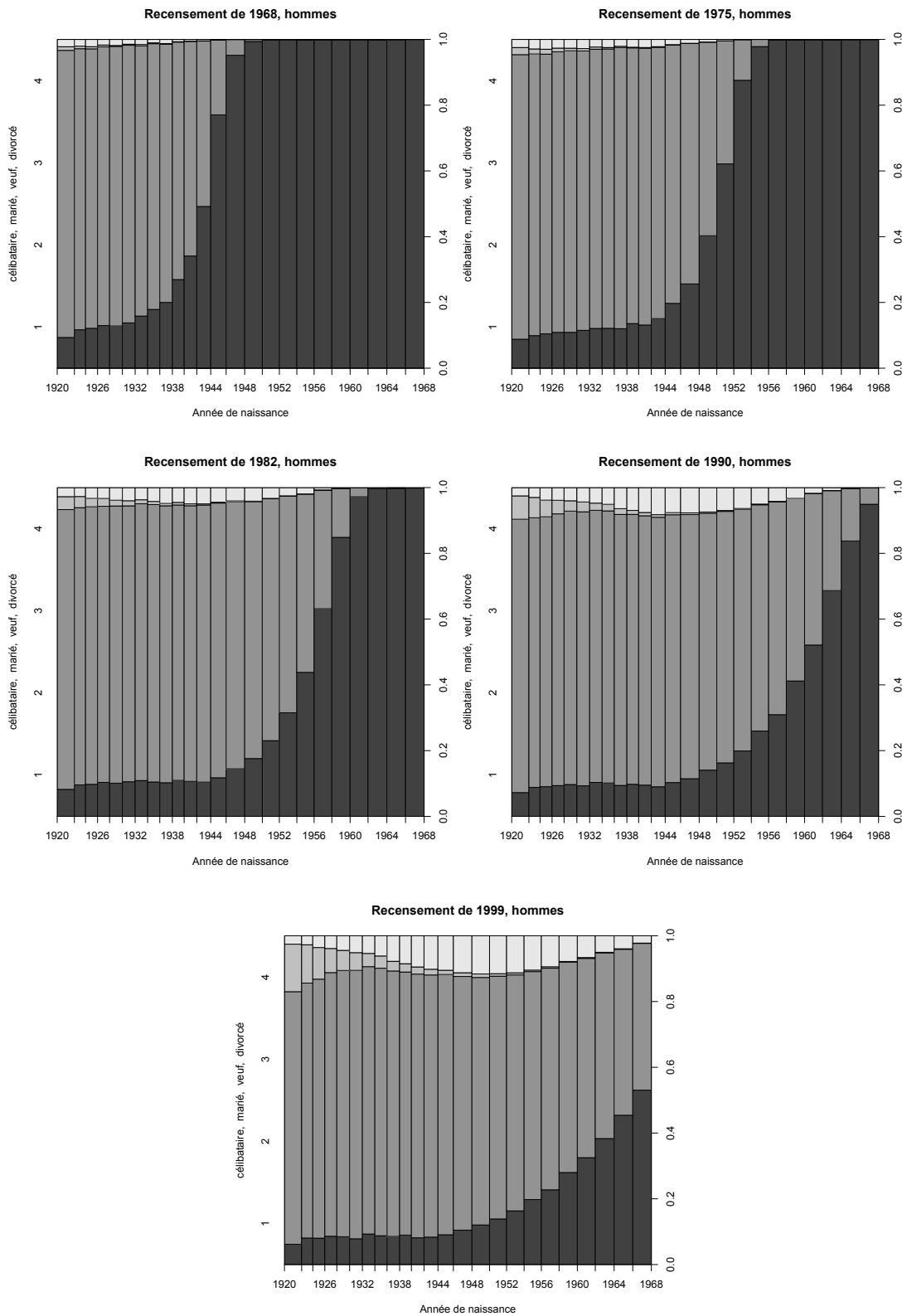


FIGURE 1.39 – Structure des recensements par statut matrimonial et par génération pour les hommes

**Niveau de diplôme dans le temps** Enfin, en Figures 1.40 et 1.41, nous représentons les recensements par niveau de diplôme, pour les hommes et les femmes. Nous présentons uniquement les recensements de 1990 et 1999 car les autres possèdent une nomenclature incomplète dans nos données. Nous regroupons les personnes de la façon suivante : 1 pour celles qui n'ont pas de diplôme ou un diplôme inférieur ou égal au Baccalauréat, et 2 pour les personnes ayant un niveau de diplôme strictement supérieur. Nous distinguons ici aussi hommes et femmes dans cette classification. Il apparaît que pour les générations récentes, la proportion de personnes diplômées du supérieur est comparable chez les hommes et les femmes, tandis qu'elle est pour les générations les plus anciennes environ deux fois moins importante chez les femmes. Au global, cette proportion tend à augmenter au fil des générations. Notons ici la rupture de tendance commune aux deux populations (homme ou femme), et qui s'opère pour les générations nées dans les années 1930. Par exemple, on peut comparer la proportion de personnes de la classe 2 pour les générations nées en 1930 ou nées en 1940. En dix ans, cette proportion a presque doublé pour les hommes mais plus que doublé pour les femmes.

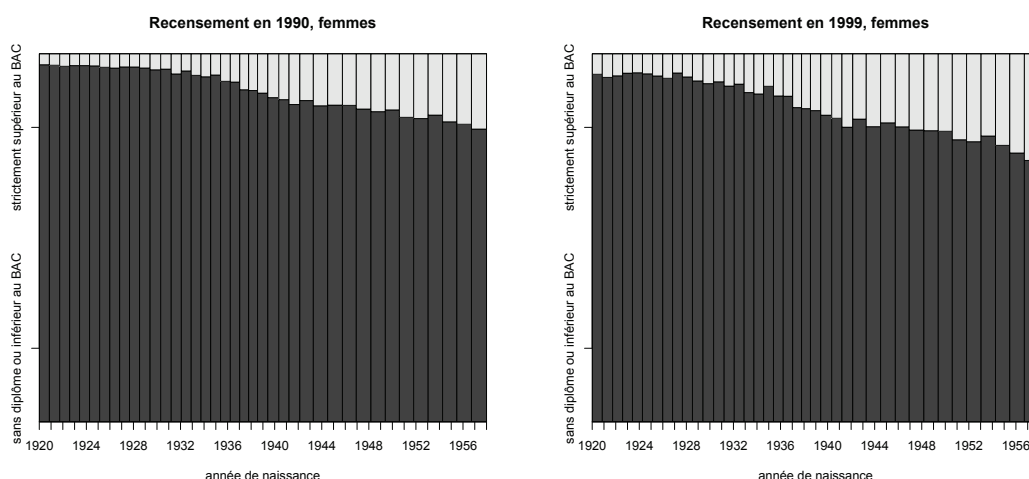


FIGURE 1.40 – Structure des recensements par niveau de diplôme et par génération pour les femmes

**Périmètre de l'étude** Cette partie de l'Échantillon Démographique Permanent à laquelle nous avons accès constitue une base de données remarquable, en particulier compte tenu de son exhaustivité, du suivi de générations réelles, et du jeu de caractéristiques recensées. Aussi, l'analyse fine des niveaux de mortalité spécifiques à des sous-populations et des changements de caractéristiques au cours de la vie est un travail statistique qui pose deux contraintes majeures :

a) La taille des sous-populations considérées doit être suffisante, ce qui va nous orienter d'abord vers le regroupement de plusieurs modalités de la classification de

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

---

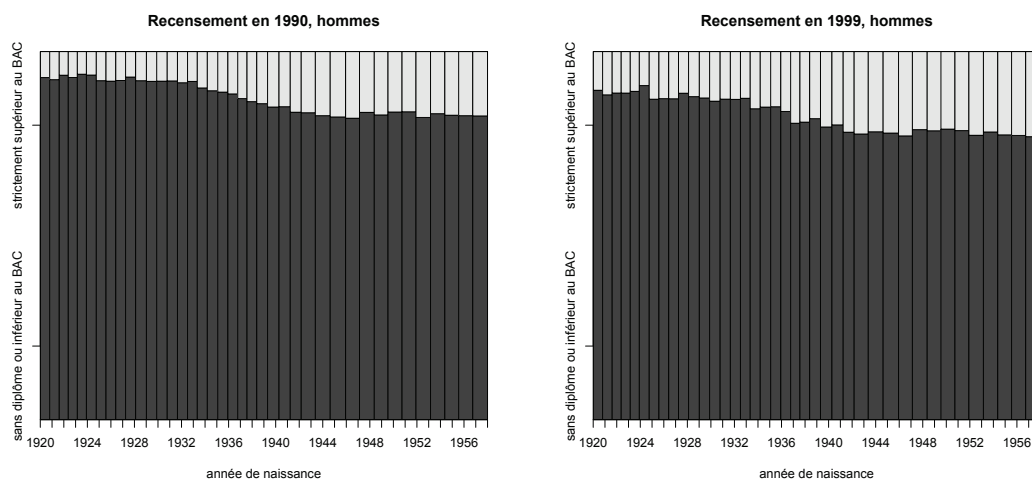


FIGURE 1.41 – Structure des recensements par niveau de diplôme et par génération pour les hommes

l'INSEE, ensuite vers la prise en compte d'un type de caractéristique à la fois.

b) L'observation des trajectoires de vie, bien que ponctuelle, doit être suffisamment répétée, ce qui va nous pousser à considérer des générations pour lesquelles les cinq recensements disponibles (1968, 1975, 1982, 1990, 1999) sont exploitables compte tenu de la caractéristique considérée.

Enfin, notons que le mécanisme d'observation ponctuel est particulier car commun à tous les individus ; ainsi l'âge auquel un individu est recensé dépend de sa date de naissance. On est ici dans un cas de *censure par intervalle dépendante* qui est un sujet en soi, et qui constitue une étape statistique supplémentaire. Nous aurons l'occasion de revenir sur ces perspectives de recherche et de souligner les défis statistiques associés.

**En conclusion.** Nous allons explorer séparément l'impact des deux caractéristiques fiables auxquelles nous avons accès : la catégorie socio-professionnelle et le statut matrimonial. Pour chacune, nous nous concentrons sur une génération rendant les cinq recensements exploitables. Ce travail, qui constitue une première étape, nous permettra d'illustrer notre méthode statistique et de déjà mettre en évidence un certain nombre d'effets démographiques cruciaux. Nous détaillons ci-dessous les données concernées dans l'étude de chacune des caractéristiques.

### Catégories socio-professionnelles

Pour chaque année de recensement, nous avons accès à la catégorie socio-professionnelle des individus observés. Nous adoptons ici une classification spécifique qui permet de ne pas avoir de classe d'effectif initial trop faible. Afin de clarifier cette étape, notons que cette classification est différente de la description des données présentée

ci-dessus. Nous regroupons ainsi les professions de la façon suivante<sup>‡</sup> :

- Le Groupe 1 rassemble les agriculteurs exploitants, les artisans, les commerçants, les ouvriers et les personnes sans activité professionnelle. Par abus de dénomination, nous désignerons ce groupe "CSP-".
- Le Groupe 2 rassemble les cadres et professions intellectuelles supérieures, les chefs d'entreprise de 10 salariés ou plus, les professions intermédiaires et les employés. Par abus de dénomination, nous désignerons ce groupe par "CSP+".

Cette classification, bien qu'arbitraire, va nous permettre de capturer des différentiels de mortalité intéressants, en particulier en vue d'applications actuarielles. Nous nous concentrons sur la génération des hommes nés en 1930 : c'est la première génération à présenter un nombre négligeable d'étudiants au recensement de 1968, ce qui permet d'exploiter l'ensemble des recensements disponibles, et considérer la dynamique multi-états représentée sur la Figure 1.42. La population de cette génération est composée de 4266 individus au premier recensement de 1968. Le Groupe 1 représente environ 70% de la population observée et le Groupe 2 son complémentaire environ 30%. Sur la Figure 1.43, nous représentons pour chaque recensement la proportion du Groupe 1 parmi les individus observés. Rappelons qu'au cours du temps, la composition de la génération évolue dû aux décès d'une part, et des changements de caractéristiques d'autre part. L'estimation de cette dynamique sera d'intérêt dans la suite.

#### Statut matrimonial

Décrivons maintenant le périmètre de notre étude sur le statut matrimonial. Le statut matrimonial est une caractéristique qui est cruciale à analyser aux grands âges, en particulier dû aux transitions du mariage vers le veuvage, particulièrement importantes dans la population de femmes. Nous nous focalisons sur la génération de femmes de l'échantillon qui ont 60 ans au premier recensement de 1968, soit nées en 1907. Ceci donne un échantillon de 3038 individus en vie au premier recensement de 1968, observés ou non. Nous regroupons la classification de l'INSEE de la façon suivante<sup>§</sup> :

- Le Groupe 1 rassemble les personnes "célibataires", "divorcées" et "veuves". On désignera leur caractéristique commune par "non-mariées".

---

<sup>‡</sup>Nous avons effectué une étude préliminaire en considérant la classification de la présentation des données qui précède, mais le nombre trop faible de personnes dans la classe des catégories professionnelles supérieures a donné des intervalles de confiance inexploitable sur les forces de mortalité.

<sup>§</sup>Ici aussi, nous avons effectué une étude préliminaire en considérant la classification détaillée avec les quatre groupes, ce qui n'a pas donné de résultat, dû au nombre trop faible de données dans certaines classes.

- Le Groupe 2 rassemble les individus "mariés".

La dynamique multi-états considérée est également représentée sur la Figure 1.42. Sur la Figure 1.44, on représente la proportion des individus du Groupe 1 au cours du temps. Si les âges considérés entre les deux applications diffèrent, c'est aussi le cas pour le statut matrimonial en particulier, au dernier recensement. Les modèles dédiés à l'estimation de

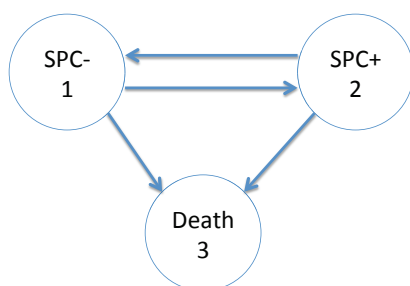


FIGURE 1.42 – Modèles multi-états considérés pour la catégorie socio-professionnelle (gauche) et le statut matrimonial (droite)

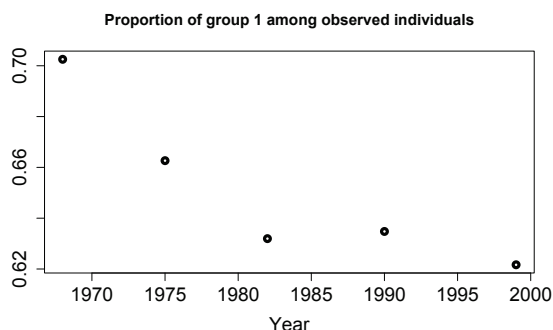


FIGURE 1.43 – Proportion d'individus du Groupe 1 (CSP-) à chaque recensement parmi les individus observés de la cohorte née en 1930.

## 1.5.2 Modèles multi-états et estimation statistique

Des outils statistiques de survie univariée peuvent être utilisés lorsque les caractéristiques des individus restent stables au cours de la vie. En revanche, quand celles-ci varient dans le temps, comme cela est le cas dans les données décrites précédemment, les modèles à adopter sont dits *multi-états* : un individu alterne entre différentes caractéristiques jusqu'au décès. Dans ce cadre, l'utilisation des estimateurs classiques

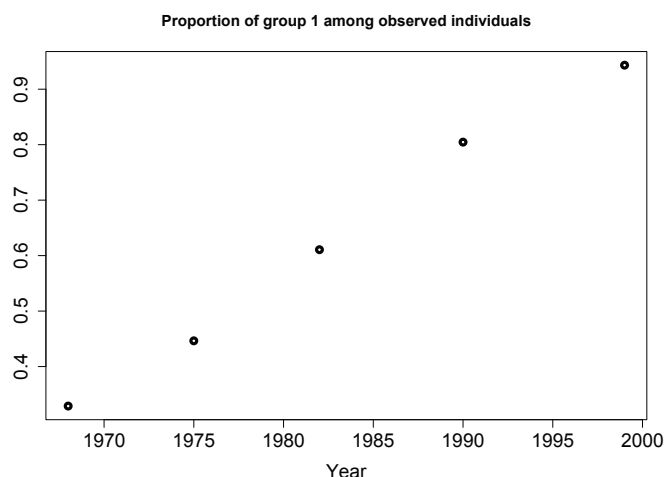


FIGURE 1.44 – Proportion d’individus du Groupe 1 (non-mariés) à chaque recensement parmi les individus observés de la cohorte née en 1907.

en survie univariée (par exemple Kaplan-Meier pour la fonction de survie ou Nelson-Aalen pour l’intensité cumulée) souffre d’un certain nombre de biais ; ceux-ci peuvent être dus au fait que l’utilisateur va procéder à une répartition arbitraire des individus selon leur caractéristique à un instant donné, donc fixée une fois pour toutes, et/ou parce que l’événement terminal (i.e. le décès la plupart du temps) sera traité comme censure indépendante lors de l’estimation des intensités de transition entre les états intermédiaires (voir par exemple Joly et al. (2002) pour la mise en évidence d’un tel biais).

Un premier apport de l’estimation de modèles multi-états est de capturer des taux de mortalité spécifiques mais également les intensités de transition entre les différentes caractéristiques : ceci donne une estimation du *niveau* des intensités de transition entre les différents états. Un autre avantage concerne l’*erreur d’estimation* : par une telle approche, on peut aussi quantifier la variance des intensités de transitions ainsi que leurs corrélations. Ainsi, si l’on s’intéresse à la dynamique d’une population à la composition spécifique, il est alors possible de prendre en compte l’impact de l’erreur d’estimation sur la trajectoire globale.

L’estimation statistique de tels modèles observés continûment, peut être réalisée à l’aide de l’estimateur standard non-paramétrique d’Aalen-Johansen (voir Aalen and Johansen (1978) et le livre de Andersen et al. (1993)). Bien que l’estimateur d’Aalen-Johansen puisse, en tant que généralisation dans un cadre multi-états de l’estimateur de Nelson-Aalen, prendre en compte plusieurs cas d’observations incomplète comme la censure et troncature classiques, son utilisation requiert de connaître exactement les instants des transitions observées. Or en pratique, dans les données longitudinales en général et particulièrement dans l’échantillon que nous avons décrit dans la partie précédente, les caractéristiques des individus ne sont observées

qu'aux instants isolés auxquels ils ont été recensés. Même si cela donne de l'information, les instants de transition restent inconnus : on parle dans ce cas de *censure par intervalle*<sup>†</sup>.

Remarquons également qu'en général les instants de décès sont quant à eux connus, comme c'est le cas dans la base de données qui nous intéresse, mais qu'en même temps la caractéristique au décès est inconnue. On peut tout de même parler ici de schéma d'observation *mixte* concernant les temps de transition : inconnus pour les états intermédiaires et connu pour l'état terminal. À notre connaissance, les contributions actuarielles et démographiques sur ces questions contournent le problème de la censure par intervalle par exemple en se donnant a priori les instants de transition, ou en restreignant l'étude à un modèle à pas discret.

On peut trouver dans la littérature actuarielle et démographique de nombreuses contributions à l'estimation de modèles multi-états à temps discret, ou en temps continu dans le cas où l'on connaît exactement les instants des transitions, et en présence de troncature et censure classiques (i.e. à droite ou à gauche). Nombre de ces contributions s'intéressent aux applications à la dépendance (état d'incapacité de l'individu) et à l'assurance santé. Gaüzère et al. (1999) estime ce que l'on appelle un processus *illness-death* (maladie-décès) irréversible, c'est-à-dire un processus dont les états sont 1=sain, 2=malade et 3=décédé, et pour lequel seules les transitions 1→2, 1→3 et 2→3 sont possibles. Gaüzère et al. (1999) suppose que les instants de transition ont lieu au milieu de la période d'observation censurée par intervalle, et utilise alors le cadre non-paramétrique classique d'Aalen-Johansen. Czado and Rudolph (2002) estime les intensités de transition d'un modèle multi-états avec des intensités à hasard proportionnel avec les temps de transition connus, tandis que Helms et al. (2005) propose dans ce cadre d'estimer directement les intensités cumulées de transition à l'aide de l'estimateur d'Aalen-Johansen. On peut également mentionner le travail de Levantesi and Menzietti (2012) qui se focalise sur un modèle multi-états irréversible à temps discret, et étudie comment les probabilités de transition évoluent d'une année sur l'autre, ce qui donne une analyse du *risque systématique*<sup>‡</sup>. En dehors des problématiques liées à la dépendance, l'étude de l'hétérogénéité de la mortalité connaît un regain d'intérêt, motivé par une meilleure compréhension de la mortalité nationale et également de la gestion des risques de mortalité/longévité. On peut citer par exemple le travail de Kwon and Jones (2006) et Kwon and Jones (2008) qui calibrent un modèle multi-états à temps discret sur une base de donnée longitudinale canadienne (*Canadian National Population Health Survey*), et étudient

---

<sup>†</sup>Rappelons que l'on parle plus généralement de données de panel lorsque plusieurs instants de transition sont manquants dans l'intervalle. Ici, on se placera par hypothèse dans le cadre de la censure par intervalle au sens strict, qui est un cas particulier de données de panel, en supposant qu'au plus une transition peut avoir lieu dans l'intervalle de recensement.

<sup>‡</sup>Voir Partie 1.1.2



l'impact des différentiels de mortalité sur des mécanismes typiques en assurance.

À notre connaissance, il n'existe pas d'études actuarielles sur l'estimation de l'hétérogénéité de la mortalité en présence de censure par intervalle, bien que ce schéma d'observation soit typique des données longitudinales. Dans ce contexte, nous nous inspirons de contributions statistiques récentes pour développer une méthode d'estimation paramétrique par maximum de vraisemblance qui prend en compte à la fois la censure par intervalle et les retours dans les états intermédiaires.

### **L'estimation de modèles multi-états en présence de censure par intervalle**

On trouve dans la littérature statistique de nombreuses méthodes pour l'estimation de modèles multi-états en présence de censure par intervalle. Dans l'ensemble, il est possible de les classer selon le type de méthode (paramétrique, semi-paramétrique, non-paramétrique), la classe des modèles multi-états considérée (*illness-death* réversible ou non, risques compétitifs, multi-états général) et l'hypothèse Markovienne (Markov, semi-Markov, non Markovien). Depuis le travail fondateur de Kalbfleisch and Lawless (1985) dans le cas de processus Markoviens à intensités constantes, plusieurs méthodes ont été développées pour les processus Markoviens *illness-death* irréversibles. Dans ce cadre, on peut par exemple trouver les approches non-paramétriques de Frydman (1995) et Frydman and Szarek (2009), et les approches paramétriques et semi-paramétriques de Joly et al. (2002), Commenges and Gégout-Petit (2007), Commenges et al. (2007) et Touraine (2013) (voir aussi Foucher et al. (2007) pour le cas semi-Markovien). Dans le cas de processus réversibles, il existe plusieurs approches dédiées à des cas particuliers de processus multi-états, voir par exemple Kang and Lagakos (2007), Titman and Sharples (2010) et Wei (2015). Dans le cadre de données de panel, Kang and Lagakos (2007) propose une méthode par maximum de vraisemblance pour l'estimation de modèles semi-Markoviens homogènes qui suppose (et exploite) le fait qu'au moins une intensité de transition est constante (et ne dépend donc pas de la durée passée dans l'état). Titman and Sharples (2010) développe une méthode alternative combinant des temps de séjour de type phase-type et les chaînes de Markov cachées. Enfin, Wei (2015) propose une approche basée sur la méthode quasi-Monte Carlo en supposant qu'il n'y a pas de transition non-observée, autrement dit que l'on est dans le cas de censure par intervalle et non de données de panel. Celle-ci s'applique à des processus semi-Markoviens indépendants du temps.

L'avantage des méthodes non-paramétriques est de s'affranchir d'hypothèses sur la structure en âge des intensités de transition, ce qui est aussi le cas dans une certaine mesure pour les méthodes semi-paramétriques car les fonctions régulières de base ne sont pas spécifiées. Ceci est particulièrement intéressant concernant les applications pour lesquelles nous n'avons aucune idée de la structure en âge de ces intensités. Malheureusement, dans le cas de petits échantillons et/ou un grand

nombre d'observations censurées par intervalle, les méthodes non-paramétriques ou semi-paramétriques peuvent se révéler instables : l'utilisation de méthodes paramétriques permet alors de capturer, en présence de telles données, un maximum d'information sur la dynamique (voir par exemple Foucher et al. (2007)). Insistons sur le fait que pour de nombreuses études statistiques, il est plus raisonnable d'implémenter une méthode paramétrique prenant en compte la censure par intervalle plutôt que d'utiliser une approche non-paramétrique en supposant les instants de transition connus (voir par exemple Touraine et al. (2013)). En effet, modifier les données pour fixer les temps de transition revient à figer la structure en âge des intensités de transition entre les états : avoir des sauts dans l'intensité pour des temps au milieu de l'intervalle, ou supposer cette intensité constante par morceaux si l'on simule ces temps uniformément dans l'intervalle de recensement. Concernant notre application en particulier, une approche paramétrique permet d'inclure l'information que l'on a sur la structure des intensités, par exemple une forme de type Gompertz pour les forces de mortalité à des âges raisonnables. Nous allons nous inspirer des approches paramétriques de Joly et al. (2002), Commenges et al. (2007), Foucher et al. (2007) et Touraine et al. (2013) qui portent sur des processus irréversibles, pour les étendre à notre cadre et en particulier prendre en compte le retour dans les états intermédiaires. Cette méthode est développée dans la partie suivante.

**Remarque 3.** *(Sur les outils existants) Plusieurs packages R sont disponibles pour l'estimation de modèles multi-états Markoviens en présence de censure par intervalle, mais aucun ne permet de traiter notre cadre. Nous prenons tout de même le temps de les décrire ci-après à l'attention du lecteur intéressé. Le package **msm** (voir Jackson et al. (2011)) traite le cas de données de panel. Il permet d'estimer des modèles multi-états Markoviens non-homogènes dont les intensités sont constantes par morceaux dans le temps<sup>§</sup>. Le package **SmoothHazard**, basé sur Touraine (2013), propose deux approches pour des modèles illness-death irréversibles : une paramétrique avec des intensités de transition de type Weibull, et une semi-paramétrique avec vraisemblance pénalisée donnant un estimateur lisse qui s'exprime sur une base de splines. Enfin, le package **coxinterval** permet, toujours dans le cas de processus illness-death irréversibles, d'estimer un modèle semi-paramétrique de Cox par maximum de vraisemblance.*

---

<sup>§</sup>Nous avons testé le package **msm** au cas de notre processus avec retour possible et des intensités de transition supposées constantes par morceaux. Comme le nombre de chemins possibles est très grand, et le nombre de paramètres élevé, la méthode s'est révélée instable et nous avons obtenu des intervalles de confiance très importants.

### 1.5.3 Méthode paramétrique par maximum de vraisemblance

**Le processus Markovien** On considère les processus Markoviens multi-états dont les dynamiques sont représentées sur la Figure 1.42, dont les états intermédiaires sont 1 et 2, et l'état terminal est 3. On note  $(X_t)$  un tel processus à espace d'états  $\{1, 2, 3\}$ , et  $\alpha_{kl}(t)$  l'intensité de transition à partir de l'état  $k$  vers l'état  $l$ . Insistons sur le fait que la composante temporelle  $t$  représente l'âge de l'individu. Le processus démarre à l'âge initial  $t = \underline{a}$  dans l'état 1 avec une probabilité notée  $p$  et dans l'état 2 avec la probabilité  $1 - p$ . Définissons également  $\alpha_{11}(t) := \alpha_{12}(t) + \alpha_{13}(t)$  et  $\alpha_{22}(t) := \alpha_{21}(t) + \alpha_{23}(t)$  les intensités d'attente dans les états 1 et 2 respectivement. À partir de l'intensité  $\alpha_{kl}(t)$ , on définit l'intensité cumulée  $A_{kl}(s, t) = \int_s^t \alpha_{kl}(u) du$ . On s'intéresse à la dynamique du processus jusqu'à ce qu'il atteigne l'état absorbant 3, à un âge aléatoire noté  $T$  (qui correspond donc à la durée de vie).

**Le schéma d'observation** Un individu de trajectoire de vie  $(X_t)$  est potentiellement observé aux âges  $R_1, \dots, R_5$ , correspondant à chacun des cinq recensements. Soit  $\delta_i$  l'indicateur de recensement :  $\delta_i = 1$  si l'individu a été observé à l'âge  $R_i$ , et  $\delta_i = 0$  sinon. On suppose que l'observation  $\delta_i$  est indépendante du processus  $(X_t)$ . Notons ici que le dernier recensement potentiel d'un individu donné est celui qui précède son âge au décès (i.e. la durée de vie)  $T$ . Rappelons aussi que les caractéristiques individuelles ne sont pas observées au décès de l'individu. Définissons  $\tau$  l'âge à partir duquel l'individu n'est plus observé ; c'est l'âge au décès si  $T = \tau$ , dans ce cas on notera  $d = 1$ , ou c'est un temps de censure à droite si  $\tau < T$ , auquel cas nous avons  $d = 0$ . Décrivons maintenant l'ensemble des recensements observés. On note  $I$  l'ensemble des indices  $i$  tels que  $\delta_i = 1$  ; pour  $i \in I$ , on a  $R_i \leq \tau$ . On note  $\bar{I}$  le plus grand élément de  $I$ , c'est-à-dire l'indice du dernier recensement pour lequel l'individu a été observé, et  $\underline{I}$  le plus petit, c'est-à-dire le premier recensement que l'on peut exploiter pour l'individu en question. Enfin, pour un indice  $i \in I$ , on note  $i_+$  l'élément suivant. Par exemple, si l'individu a été observé au recensement  $i + 1$ , on a  $i_+ = i + 1$ .

**La vraisemblance** Pour une trajectoire de vie observée suivant le schéma décrit précédemment, dont nous notons  $x_i$  l'état au recensement  $R_i$  pour tout  $i \in I$ , la contribution s'écrit de la façon suivante en utilisant la propriété de Markov :

$$\mathbb{P}(X_{R_{\underline{I}}} = x_{\underline{I}}) \prod_{i \in I, i < \bar{I}} \mathbb{P}(X_{R_{i_+}} = x_{i_+} \mid X_{R_i} = x_i) \mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}) \mathbb{P}(T = \tau \mid T \geq \tau, X_{R_{\bar{I}}} = x_{\bar{I}})^d. \quad (1.51)$$

Comme nous faisons l'hypothèse que les trajectoires de vie des individus sont indépendantes, la vraisemblance globale s'écrit comme le produit de chaque contribution individuelle (1.51). Le calcul de chaque terme élémentaire de l'Équation (1.51) en

fonction des intensités de transition  $\alpha_{kl}(t)$  et des intensités cumulées  $A_{kl}(s, t)$  est détaillé dans le Chapitre 5. Nous donnons ici simplement un exemple typique dans le cadre de la censure par intervalle, celui de la contribution de l'observation de l'individu à deux recensements consécutifs :  $\mathbb{P}(X_{R_{i_+}} = x_{i_+} \mid X_{R_i} = x_i)$ . Nous illustrons ce calcul dans le cas où l'individu a été observé au recensement suivant  $R_i$ , c'est-à-dire  $i_+ = i + 1$ . Traitons d'abord le cas où les deux caractéristiques successivement observées sont différentes, c'est-à-dire  $x_i \neq x_{i_+}$ . **Hypothèse :** L'hypothèse que nous faisons sur la base de données est qu'entre deux recensements consécutifs, au plus une transition peut avoir lieu. Ceci peut paraître comme une hypothèse forte sur la base de données, mais en réalité il semble raisonnable de penser que sur un intervalle d'environ 7 ans, il n'est pas possible que deux transitions ou plus aient lieu avec la classification choisie en terme de catégorie socio-professionnelle ou de statut matrimonial. Cette hypothèse ne peut pas être testée dans nos données, mais nous insistons sur le fait que c'est une étape supplémentaire qui nous affranchit ici d'hypothèses sur les temps de transition. Ainsi, nous supposons que dans nos données, les trajectoires de vie sont irréversibles par morceaux.

Dans ce cadre, la contribution à la vraisemblance qui prend en compte la censure par intervalle revient à intégrer sur tous les temps de transition possibles  $u \in [R_i, R_{i+1}]$ . Elle exprime le fait que l'individu reste dans l'état  $x_i$  dans l'intervalle de temps  $[R_i, u]$ , qu'il transite instantanément au temps  $u$ , puis qu'il reste dans l'état  $x_{i+1}$  dans l'intervalle de temps  $[u, R_{i+1}]$ , c'est-à-dire

$$\mathbb{P}(X_{R_{i+1}} = x_{i+1} \mid X_{R_i} = x_i) = \int_{R_i}^{R_{i+1}} e^{-A_{x_i x_i}(R_i, u)} \alpha_{x_i x_{i+1}}(u) e^{-A_{x_{i+1} x_{i+1}}(u, R_{i+1})} du.$$

Traitons également le cas où les deux observations sont les mêmes, c'est-à-dire telles que  $x_{i+1} = x_i$ . Sous l'hypothèse qu'au plus une transition peut avoir lieu, la seule possibilité est que l'individu reste dans cet état, c'est-à-dire

$$\mathbb{P}(X_{R_{i+1}} = x_i \mid X_{R_i} = x_i) = e^{-A_{x_i x_i}(R_i, R_{i+1})}.$$

Les calculs des autres cas et briques élémentaires de l'Équation (1.51) sont détaillés dans la Partie 5.3.1 du Chapitre 5.

**Paramétrisation des intensités de transition** Nous détaillons ici les formes paramétriques choisies pour les intensités de transition, c'est-à-dire sur la structure en âge des forces de mortalité par caractéristique et également des transitions entre les différents états.

Concernant les taux de mortalité spécifiques à chaque caractéristique, guidés par l'observation des taux de mortalité nationaux, nous proposons une forme de type Gompertz (voir Gompertz (1825)) étant donné que des âges autour de 60 et 100 ans

pour le statut matrimonial, et 40 et 70 ans pour la catégorie socio-professionnelle sont concernés pour l'estimation. Ceci s'écrit, pour  $(k, l) \in \{(1, 3), (2, 3)\}$ ,

$$\alpha_{kl}(t) = c_{kl} \exp(d_{kl}t).$$

Nous détaillons ci-dessous les formes paramétriques choisies pour les transitions entre les états intermédiaires. Pour chacune des deux applications (catégorie socio-professionnelle ou statut matrimonial), nous nous attacherons à comparer deux types de paramétrisations.

Concernant les transitions entre les états intermédiaires représentant les catégories socio-professionnelles, contrairement à la mortalité, nous n'avons pas de données de référence pour nous guider sur la structure en âge des intensités. Nous avons testé une approche non-paramétrique en supposant les temps de transition connus, mais les incréments de l'estimateur de Nelson-Aalen pour les intensités cumulées étaient trop erratiques pour se forger une intuition sur la structure en âge de ces intensités de transition. Néanmoins, un cadre raisonnable et largement utilisé est celui des intensités monotones de type Weibull, qui nous donnent la possibilité d'avoir une augmentation ou diminution de l'intensité de passage d'une classe à une autre au cours de la vie, voir par exemple Joly et al. (2002), Commenges et al. (2007), Foucher et al. (2007) et Touraine et al. (2013). Un autre avantage est qu'un cas particulier est celui de l'intensité constante, ce qui va nous permettre de comparer des modèles imbriqués à l'aide de critères statistiques. Pour  $(k, l) \in \{(1, 2), (2, 1)\}$ , on écrit la paramétrisation de type Weibull de l'intensité  $\alpha_{kl}(t)$  comme

$$\alpha_{kl}(t) = a_{kl} b_{kl}^{-a_{kl}} t^{a_{kl}-1}.$$

Dans notre application aux catégories socio-professionnelles, nous devons choisir un âge à partir duquel la caractéristique est stable. On choisit ici l'âge 60 ans qui, bien qu'arbitraire, nous permet de prendre en compte les changements possibles en fin de carrière et en même temps d'éviter des effets de bord structurels. Ainsi, l'intensité précédente est mise à zéro pour  $t > 60$ , et l'intensité cumulée est modifiée en ce sens.

Les hypothèses paramétriques sur la structure en âge des transitions entre les statuts matrimoniaux sont cruciales. En effet, à la différence de la catégorie socio-professionnelle, le statut matrimonial peut changer aux grands âges, surtout dû à la transition mariage  $\rightarrow$  veuvage. Aussi, en se basant sur les recensements observés et la contrainte de structure en âge, il est possible de capturer un niveau de transition dans la tranche d'âge à laquelle aucune observation n'est possible, ici entre le dernier recensement (année 1999) et la fin du suivi de la base de données (année 2008). Nous avons réalisé une étude préliminaire et obtenu de premières intuitions sur la forme possible des intensités de transition de marié vers non-marié aux grands âges, en considérant que celle-ci correspond à la probabilité de décès instantanée de l'époux,

moyennée sur tous ses âges possibles. Sur la base de données nationales pour la France sur la mortalité et sur la répartition des écarts d'âge entre les couples, nous avons calculé une telle quantité et exhibé sa structure en âge. Cette observation se situe à la frontière d'une croissance exponentielle (modèle Gompertz) et puissance (modèle Weibull), et nous allons donc tester les deux familles paramétriques pour de telles transitions. La première sera notée

$$\alpha_{21}(t) = a_{21} \exp(b_{21}t),$$

et la seconde

$$\alpha_{21}(t) = a_{21} b_{21}^{a_{21}} t^{a_{21}-1}.$$

Concernant la transition de non-marié à marié, étant donné le faible nombre de transitions, nous proposons d'inclure une intensité constante  $\alpha_{12}(t) = b_{12}$ , supposée nulle pour  $t > 90$  ans.

Pour conclure, le jeu de paramètres sur lequel la vraisemblance devra être maximisée dans chaque application est au plus

$$\theta = (p, a_{12}, a_{21}, b_{12}, b_{21}, c_{13}, c_{23}, d_{13}, d_{23}). \quad (1.52)$$

**Incertitude sur les paramètres** Une fois la vraisemblance maximisée, nous obtenons un estimateur des différents paramètres, noté  $\hat{\theta}$ . Nous nous intéressons également dans nos applications à l'erreur d'estimation, c'est-à-dire aux fluctuations des paramètres estimés autour de leur vraie valeur. Cette erreur est ici due à la taille finie de l'échantillon et au schéma d'observation présentant à la fois de la censure par intervalle et des données manquantes. Un résultat standard sur les estimateurs du maximum de vraisemblance est la normalité asymptotique : lorsque la taille de l'échantillon tend vers l'infini, le jeu aléatoire de paramètres suit une loi Gaussienne multi-variée, centrée autour de leur vraie valeur  $\theta$  et de matrice de variance-covariance donnée par l'opposé de la matrice hessienne inverse évaluée en  $\theta$ . On estime classiquement la moyenne par l'estimateur  $\hat{\theta}$  et la matrice de variance-covariance par l'estimation numérique de l'opposé de la matrice hessienne inverse prise en  $\hat{\theta}$ , c'est-à-dire l'estimation numérique de  $\hat{\Sigma} = -H(\hat{\theta})$  où  $H_{i,j}(\theta) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\theta)$ . Nous supposons ici que nous sommes dans le régime de la normalité asymptotique. Nous pouvons alors réaliser des simulations jointes des paramètres, et ensuite calculer la distribution de différentes quantités d'intérêt en les évaluant dans chaque scénario.

#### 1.5.4 Résultats

La maximisation de fonction en dimension relativement grande est toujours un problème délicat, surtout lorsqu'une seule évaluation de celle-ci demande des ressources

informatiques non-négligeable (car chaque contribution individuelle doit être calculée). Nous avons isolé les termes communs aux individus pour en réaliser le calcul une seule fois, et mis en oeuvre des procédures parallèles sur les coeurs de calcul du LPMA<sup>¶</sup>. Nous avons testé plusieurs jeux de paramètres initiaux ; en particulier, les forces de mortalité ont été initialisées avec des paramètres que nous avons calibré sur les données nationales.

### Résultats pour la catégorie socio-professionnelle

On souhaite comparer deux modèles : le Modèle 1 pour lequel les intensités de transition entre les catégories socio-professionnelles sont constantes, et le Modèle 2 pour lequel ces intensités sont de type Weibull. Nous obtenons dans le Modèle 1 une log-vraisemblance égale à -14743.74, et dans le Modèle 2 égale à -14727.55. Les deux modèles étant imbriqués, il est naturel d'obtenir une vraisemblance plus élevée dans le Modèle 2 pour lequel le nombre de paramètres est plus important. Pour voir si l'ajout des paramètres  $a_{12}$  et  $a_{21}$  est utile, nous comparons les deux modèles dans le Chapitre 5 à l'aide des critères AIC et BIC, indiquant tous deux le choix du Modèle 2. Dans la suite, nous donnons néanmoins les résultats pour les deux modèles. Les paramètres estimés sont donnés dans le Tableau 1.4, et les intensités de transitions entre les groupes et vers le décès sont représentées sur les Figures 1.45 et 1.46 respectivement, avec les intervalles de confiance (ponctuels) à 95% associés (calculés sur la base de la normalité asymptotique de l'estimateur, comme décrit précédemment).

Paramètre	Modèle 1	Modèle 2
$p$	0.703	0.703
$a_{12}$	1	0.621
$a_{21}$	1	3.17
$b_{12}$	80.1	50.8
$b_{21}$	56.3	74.0
$c_{13}$	0.000191	0.000189
$c_{23}$	0.000152	0.000173
$d_{13}$	0.0721	0.0723
$d_{23}$	0.0703	0.0683

TABLE 1.4 – Paramètres estimés pour les Modèles 1 et 2. Rappelons que pour le Modèle 1, le nombre de paramètres est réduit car  $a_{12}$  et  $a_{21}$  sont fixés à la valeur 1.

---

<sup>¶</sup>Nous remercions chaleureusement Jacques Portes pour son aide et ses cours particuliers de *parallélisme du pauvre* !

### 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.45 – Intensités de transition estimées entre les états intermédiaires pour le Modèle 1 (gauche) et le Modèle 2 (droite) et les intervalles de confiance ponctuels à 95 % associés

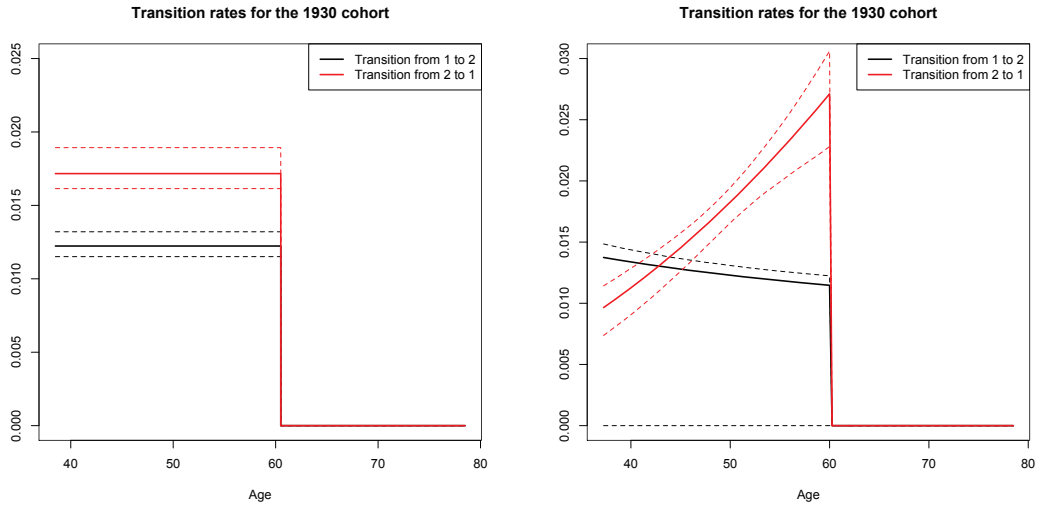
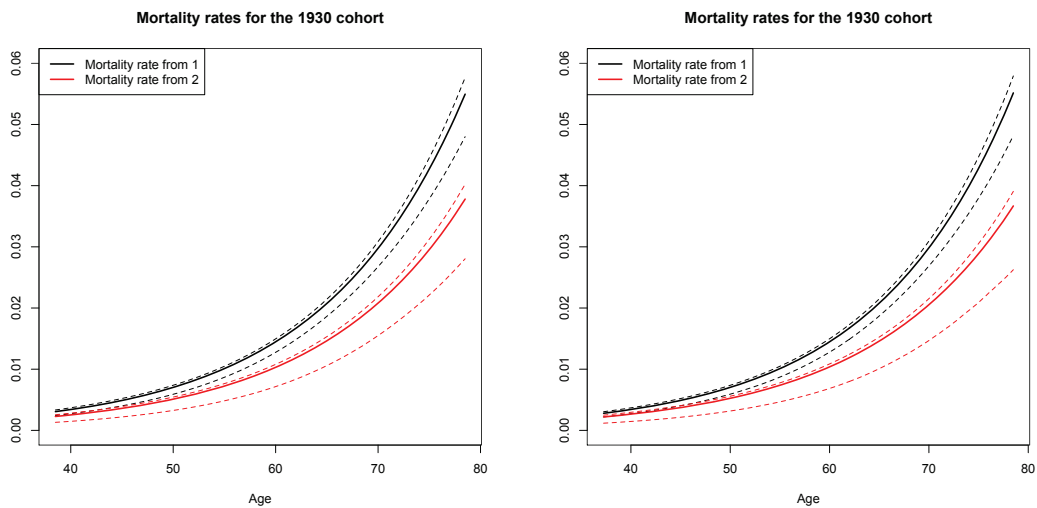


FIGURE 1.46 – Estimation des forces de mortalité par caractéristique pour le Modèle 1 (gauche) et le Modèle 2 (droite) et les intervalles de confiance ponctuels à 95 % associés



Ces résultats nous donnent de premières intuitions sur les phénomènes en jeu. D'abord, comme attendu, les forces de mortalité entre les deux groupes sont significativement différentes, celles du premier groupe étant plus élevées. Les résultats concernant ces taux de décès sont très proches entre les deux modèles, tant en terme de niveau qu'en terme d'incertitude. En revanche, si les intensités entre les états intermédiaires sont d'un ordre de grandeur similaire dans le Modèle 1, l'estimation du Modèle 2 montre que la structure en âge de ces intensités sont très différentes.



Comme cette approche nous donne à la fois une estimation du niveau et de l'erreur associée, il apparaît que ces deux quantités sont spécifiques aux caractéristiques considérées. Ainsi, on s'attend à des comportements différents pour la dynamique d'une population avec sa composition propre. Ceci sera étudié dans la Partie 1.5.5.

### Résultats pour le statut matrimonial

On désigne par Modèle 3 le modèle pour lequel les transitions entre les statuts matrimoniaux est de type Weibull, et Modèle 4 le modèle dans lequel ces transitions sont de type Gompertz. La vraisemblance obtenue dans le Modèle 3 est -14061.86 et dans le Modèle 4, celle-ci est égale à -14062.51. Les paramètres estimés sont donnés dans le Tableau 1.5. On estime la matrice de variance-covariance des paramètres comme précédemment, à la fois pour les Modèles 3 et 4. Sur cette base, on représente les intensités estimées et leurs intervalles de confiance (ponctuels) à 95%, sur les Figures 1.47 et 1.48 pour les transitions entre les statuts matrimoniaux et les forces de mortalité respectivement.

Paramètre	Modèle 3	Modèle 4
$p$	0.329	0.329
$a_{21}$	5.99	0.000397
$b_{12}$	0.00194	0.00193
$b_{21}$	0.0131	0.0693
$c_{13}$	1.22e-05	1.23e-05
$c_{23}$	2.96e-06	3.11e-06
$d_{13}$	0.103	0.103
$d_{23}$	0.119	0.118

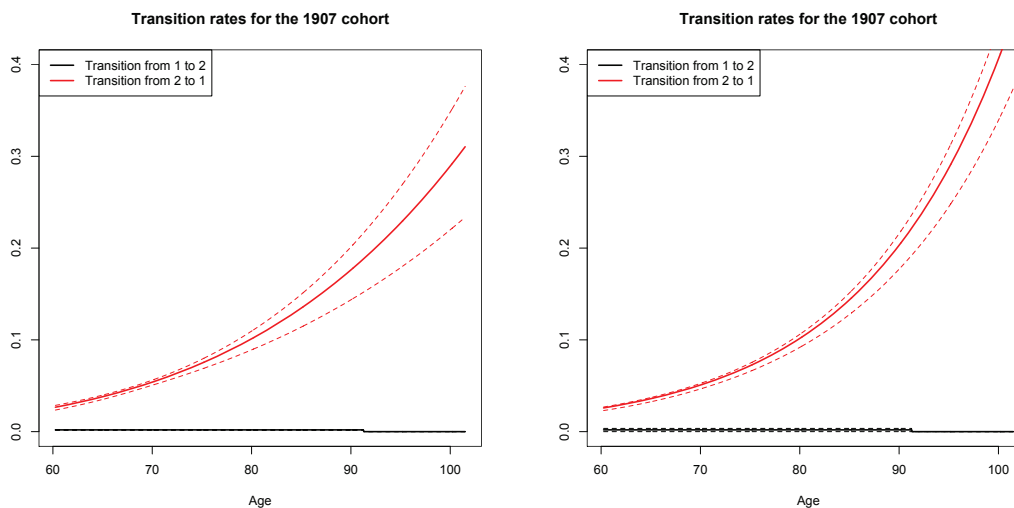
TABLE 1.5 – Paramètres estimés pour les Modèles 3 et 4

Concentrons nous d'abord sur les transitions entre les statuts matrimoniaux (marié ou non-marié), voir Figure 1.47. Comme l'on s'y attend en observant le nombre de transitions entre les états, l'intensité de transition  $\alpha_{12}(t) = b_{12}$  de non-marié à marié est très faible en comparaison de la transition inverse  $2 \rightarrow 1$ , bien que celle-ci soit non-nulle avec une probabilité non-négligeable (voir Tableau 1.5 et Figure 1.47). On est ici très proche d'un modèle irréversible, mais nous verrons dans la Partie 1.5.5 que cette analogie n'est pas parfaitement valide car un impact, même faible, réside dans le calcul de produits d'assurance-vie typiques.

**Loi de compensation de la mortalité** Observons maintenant les résultats obtenus sur la mortalité spécifique à chaque groupe (marié ou non-marié), plus précisément le logarithme des forces de mortalité, voir Figure 1.48. Il apparaît que pour

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.47 – Intensités de transition estimées entre les états intermédiaires pour le Modèle 3 (gauche) et le Modèle 4 (droite) et les intervalles de confiance ponctuels à 95 % associés

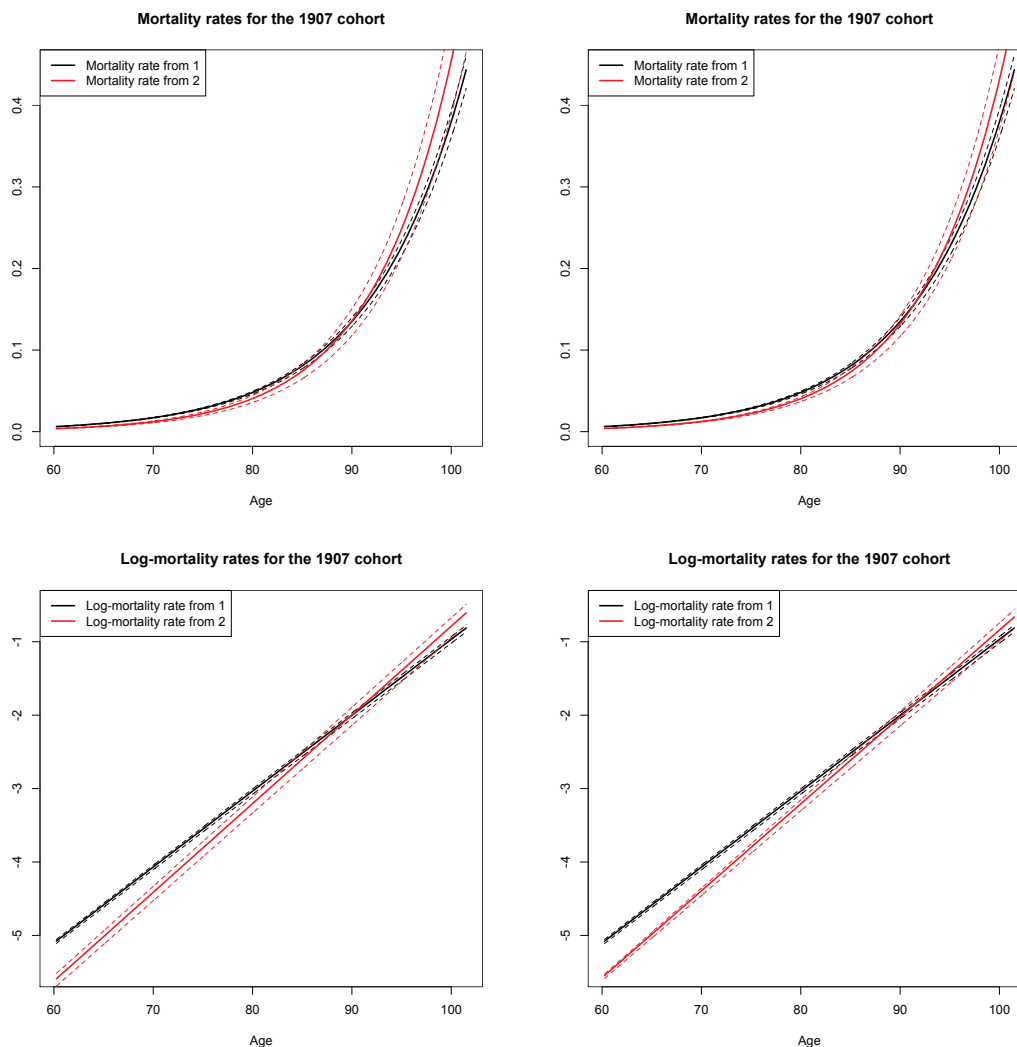


les âges entre environ 60 et 85 ans, la force mortalité est plus faible pour les individus mariés, ce qui est un résultat attendu dans ce sens. Mais après un âge pivot autour de 90 ans, la situation s'inverse : aux grands âges, la mortalité des personnes mariées devient la plus élevée. Ceci peut paraître surprenant à première vue, mais en réalité il s'agit précisément d'un fait bien connu en démographie que l'on appelle *loi de compensation de la mortalité*. Celle-ci apparaît lorsque l'on compare plusieurs sous-populations au même moment, par exemple des sous-groupes d'un même pays (ce qui est le cas ici) ou plusieurs pays entre eux, ou bien lorsque l'on compare une même population à des dates successives<sup>||</sup>. Ce phénomène revient à dire que dans le modèle de Gompertz, si le facteur multiplicatif est plus élevé, ici  $c_{23} < c_{13}$ , alors nécessairement le coefficient en âge est plus élevé, c'est-à-dire ici  $d_{23} > d_{13}$  (voir Tableau 1.5). Plus généralement, ce phénomène reflète le fait qu'une diminution de la force de mortalité aux âges intermédiaires conduit à une augmentation de celle aux grands âges. Ce phénomène s'explique par le fait que la *durée de vie humaine*, au sens de la durée de vie biologique maximale que l'on peut atteindre en ne considérant que le vieillissement pur, n'est pas modifiée sur des périodes de temps raisonnables (voir par exemple Strulik and Vollmer (2013) et la discussion dans la Partie 1.1.1). Cet effet de compensation de la mortalité peut être mesuré de deux autres façons, avec chacune leur dénomination respective : la compression de la distribution des

<sup>||</sup>Nous avons vu dans la Partie 1.1.1 que dans l'estimation de modèles de mortalité stochastique qui possèdent une formulation très proche de Gompertz, tels que Cairns et al. (2006), les séries temporelles de pente et d'ordonnée à l'origine sont corrélées négativement, comme l'ont déjà formulé par exemple Strehler and Mildvan (1960).

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.48 – Estimation des forces de mortalité par caractéristique (haut) et de leur logarithme (bas) pour le Modèle 3 (gauche) et le Modèle 4 (droite) et les intervalles de confiance ponctuels à 95 % associés



âges au décès et la rectangularisation de la courbe de survie. Pour plus de détails sur ces effets, nous renvoyons le lecteur à Strehler and Mildvan (1960), Fries (1980), Gavrilov and Gavrilova (1991), Wilmoth and Horiuchi (1999) et Strulik and Vollmer (2013).

Remarquons ici que les modèles à hasard proportionnel, qui supposent que la force de mortalité d'un sous-groupe est donnée à un facteur multiplicatif près par la force de mortalité de référence (par exemple nationale) ne permettent pas de tenir compte de tels effets. Cet effet de compensation va nous donner des intuitions sur le comportement de la mortalité agrégée d'un échantillon que l'on fait évoluer dynamiquement avec les taux de transition estimés. Nous étudions ce point dans la partie suivante.

### 1.5.5 Impact de l'hétérogénéité sur la mortalité agrégée et des produits d'assurance-vie

Dans cette partie, nous discutons l'impact de l'hétérogénéité sur (i) la mortalité agrégée et (ii) des produits d'assurance typiques de rentes ou de décès. Nous décrivons d'abord la dynamique d'une population hétérogène.

**Dynamique de population hétérogène** Rappelons ici la dynamique déterministe d'une cohorte hétérogène, prenant en compte les décès mais également les changements de caractéristiques. On considère les intensités de transition  $\alpha_{kl}(t)$  qui dépendent de l'âge  $t$  comme étant aléatoires car elles dépendent de la réalisation du vecteur de paramètres  $\theta$ , voir (1.52), dont la distribution peut être approchée par la méthode décrite dans la Partie 1.5.3. Étant donnés les paramètres, nous travaillons donc ici sur une trajectoire *moyenne* de population, ainsi le bruit démographique\*\*, qui est présent pour une trajectoire de population de taille finie, n'est pas pris en compte en tant que tel. En revanche, ce que nous mesurons avec la distribution estimée des paramètres concerne les fluctuations dues à *la taille finie de la population dans les données* et également au schéma d'observation censuré par intervalle. Notons  $g_1(\underline{a})$  et  $g_2(\underline{a})$  les grandeurs réelles qui fixent la composition à l'âge  $\underline{a}$  de la génération en chaque groupe 1 et 2; nous fixerons l'âge de départ selon la plage d'âge d'estimation de l'application considérée ( $\underline{a} = 37.25$  pour la catégorie socio-professionnelle et  $\underline{a} = 60.25$  pour le statut matrimonial). On note pour  $t \geq \underline{a}$ ,  $G(t) = \begin{pmatrix} g_1(t) \\ g_2(t) \end{pmatrix}$  la répartition entre les groupes à l'âge  $t$ . Étant donnés les temps de transition, cette composition évolue dans le temps suivant les décès et les changements de caractéristiques. Ceci correspond aux équations classiques de Chapman-Kolmogorov du processus de Markov  $(X_t)$ , que l'on peut voir aussi comme un cas particulier des équations déterministes (moyennes ou en grande population) associées au modèle de population général présenté dans la Partie 1.2 et détaillé dans le Chapitre 2 :

$$G'(t) = K(t)G(t), \quad (1.53)$$

où la matrice de transition  $K(t)$  est donnée par

$$K(t) = \begin{pmatrix} -\alpha_{13}(t) - \alpha_{12}(t) & \alpha_{21}(t) \\ \alpha_{12}(t) & -\alpha_{23}(t) - \alpha_{21}(t) \end{pmatrix}.$$

Remarquons qu'ici apparaît la complexité due à la prise en compte de transitions réversibles; dans ce cas en général, aucune solution analytique n'est disponible pour  $G$  (voir par exemple la discussion dans Andersen et al. (1993)). C'est bien le cas dans

---

\*\*Voir Partie 1.1.2

notre cadre paramétrique des intensités, car les matrices  $K(s)$  et  $K(t)$  ne commutent pas<sup>††</sup>. En pratique, l'hypothèse que la transition inverse est égale à zéro est souvent faite lorsque un nombre faible de telles transitions est observé (voir par exemple Czado and Rudolph (2002)). Comme nous ne disposons pas de formule fermée, nous utilisons un schéma numérique classique pour discrétiser l'Équation (1.53).

**Mortalité agrégée** Nous souhaitons illustrer ici la dynamique du taux de mortalité agrégé basé sur la représentativité de l'échantillon. On définit la quantité moyenne (nombre réel) d'individus par

$$g(t) = g_1(t) + g_2(t). \quad (1.54)$$

Le taux de mortalité agrégé à l'âge  $t$  est défini par

$$\mu(t) = -\frac{g'(t)}{g(t)} = \frac{g_1(t)\alpha_{13}(t) + g_2(t)\alpha_{23}(t)}{g(t)}, \quad (1.55)$$

ce qui n'est rien d'autre qu'une moyenne pondérée des taux de mortalité spécifiques, dont les poids dépendent de la dynamique hétérogène sous-jacente.

Lorsque l'on s'intéresse à l'hétérogénéité, un phénomène connu est celui du *rat-trapage* : la mortalité agrégée rejoint la plus faible, étant donné que le groupe qui possède cette force de mortalité plus faible soit de plus en plus représenté dans la génération. Ce fait est observé lorsque les caractéristiques restent stables au cours de la vie. Lorsque les changements de caractéristique sont pris en compte, la façon dont la mortalité agrégée évolue avec l'âge est potentiellement différente, et il est possible d'observer l'effet inverse. Les deux applications que nous détaillons ci-après permettent d'illustrer de tels effets.

Les résultats pour les catégories socio-professionnelles sont représentés sur la Figure 1.49. On observe que la mortalité agrégée rejoint (lentement) la mortalité du groupe 2, et mesure la vitesse à laquelle ceci se produit (ce qui dépend des différentes intensités de transition). À l'aide des distributions estimées pour les paramètres, il est possible de calculer les intervalles de confiance à 95% pour la mortalité agrégée, ce qui est représenté sur la Figure 1.50.

Pour valider le cadre de modélisation et les hypothèses paramétriques, il est intéressant d'observer la fonction de survie agrégée obtenue à partir de la mortalité agrégée ci-dessus, et de la comparer à la fonction de survie calculée directement sur l'échantillon, sans tenir compte des caractéristiques individuelles, par exemple par l'estimateur non-paramétrique de Kaplan-Meier. Cette comparaison est donnée dans la Figure 1.51 pour les Modèles 1 et 2. Remarquons que les deux courbes sont relativement similaires, bien qu'elles puissent différer aux grands âges dans plusieurs scénarios de paramètres. Ceci pourrait s'expliquer par les hypothèses de mortalité

---

<sup>††</sup>Cette vérification est laissée au lecteur.

de type Gompertz qui est une approximation moyenne pour les âges autour de 40 ans, ce qui peut entraîner une légère sous-estimation de la mortalité. Nous pensons néanmoins que ceci constitue la meilleure option de forme paramétrique à deux paramètres, étant donné le faible nombre de données. Notons également que malgré ces différences dans certains scénarios de paramètres, nous capturons d'un autre côté plus d'information sur la dynamique hétérogène, ce qui est au coeur de notre étude.

Intéressons-nous maintenant aux résultats sur le statut matrimonial. Sur la Figure 1.52, on représente les logarithmes des forces de mortalité pour chaque groupe 1 ou 2 et agrégée, pour les Modèles 3 et 4, ainsi que sur la Figure 1.53 les intervalles de confiance à 95% associés. Décrivons la structure en âge de la mortalité agrégée : aux âges autour de 60 ans, celle-ci est proche de la force de mortalité des personnes mariées, elle rejoint ensuite la mortalité des personnes non-mariées à l'âge de 85 ans environ, puis reste à ce niveau de mortalité jusqu'à la fin de la plage d'âge considérée pour l'estimation. Ceci donne un exemple où le phénomène de rattrapage n'a pas lieu, dû aux changements de caractéristiques. Ici en effet, la mortalité agrégée rejoint d'abord celle du groupe ayant la plus élevée (aux âges intermédiaires), puis reste au niveau de la mortalité des personnes "non-mariées" car celles-ci sont sur-représentées aux grands âges dû aux changements de plus en plus fréquents marié  $\rightarrow$  non-marié. Pour cette application, nous comparons également d'une part la fonction de survie obtenue sur la base des forces de mortalité décrites précédemment, et d'autre part l'estimation par Kaplan-Meier de la fonction de survie sur l'échantillon sans tenir compte des caractéristiques individuelles. Ces résultats sont représentés sur la Figure 1.54. Les deux courbes de survie sont très similaires pour chaque Modèle (3 ou 4), ce qui va dans le sens des choix paramétriques qui ont été faits et de la vraisemblance obtenue. Ainsi pour cette application, à la fois la dynamique hétérogène est captée, mais la mortalité agrégée de l'échantillon est également répliquée.

### 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.49 – Forces de mortalité spécifiques à chaque caractéristique et agrégée (gauche) et leur logarithme (droite) pour le Modèle 1.

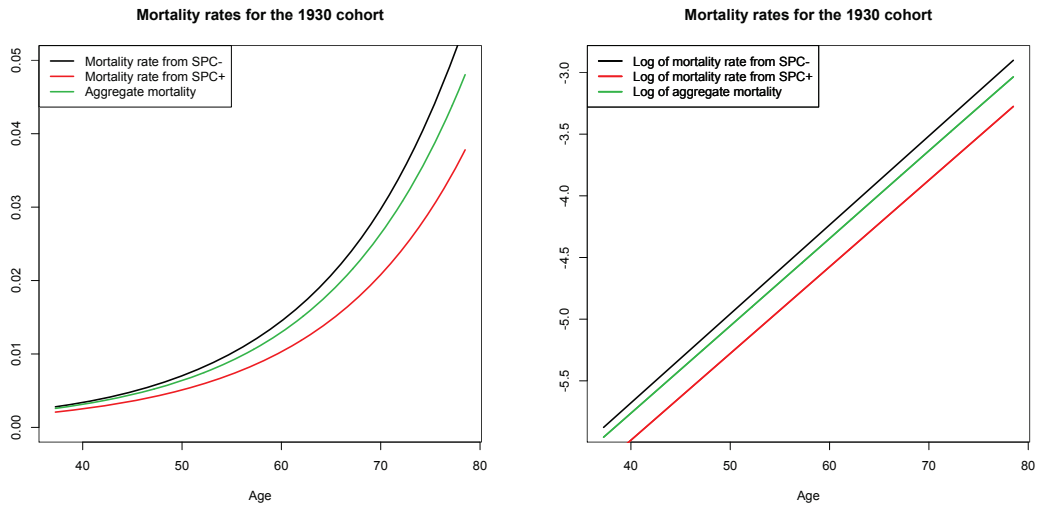
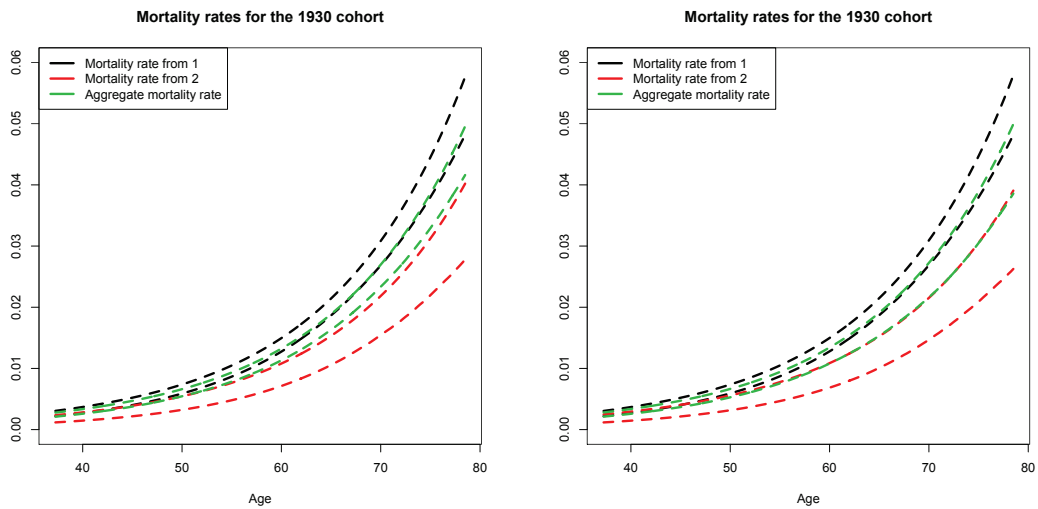


FIGURE 1.50 – Intervalles de confiance ponctuels à 95% pour les forces de mortalité spécifiques à chaque caractéristique et agrégée pour le Modèle 1 (gauche) et le Modèle 2 (droite).



1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

---

FIGURE 1.51 – Courbes de survie agrégée et intervalles de confiance ponctuels à 95% obtenues dans le Modèle 1 (gauche) et le Modèle 2 (droite) comparé à l'estimateur Kaplan-Meier

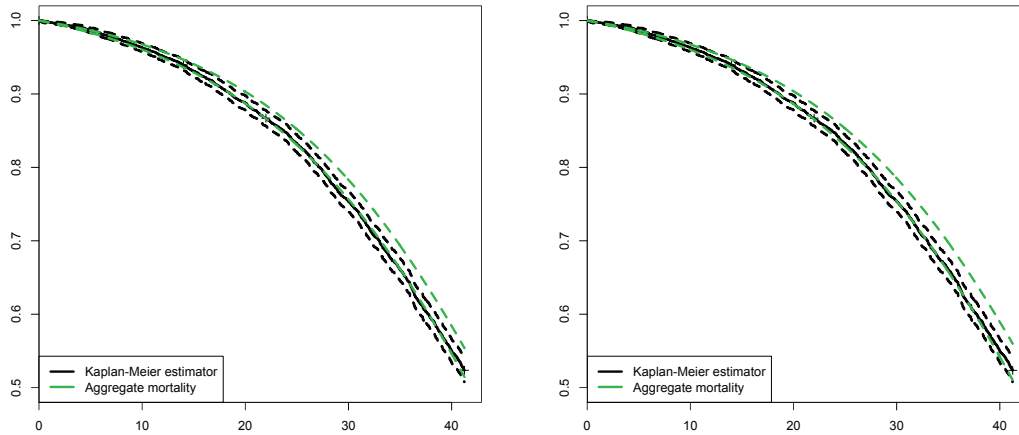
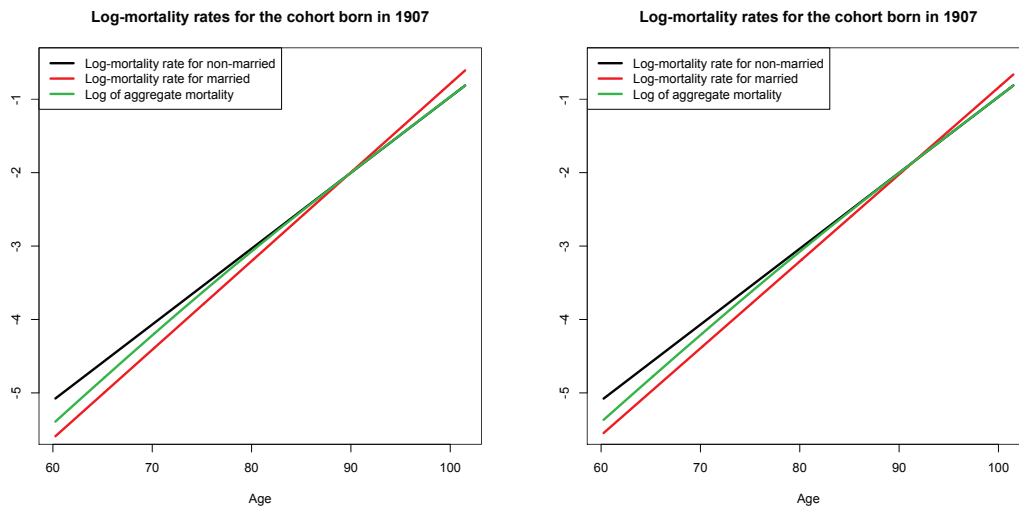


FIGURE 1.52 – Logarithme des forces de mortalité spécifiques à chaque caractéristique et agrégée pour le Modèle 3 (gauche) et le Modèle 4 (droite).





## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.53 – Intervalles de confiance ponctuels à 95% pour le logarithme des forces de mortalité spécifiques à chaque caractéristique et agrégée pour le Modèle 3 (gauche) et le Modèle 4 (droite).

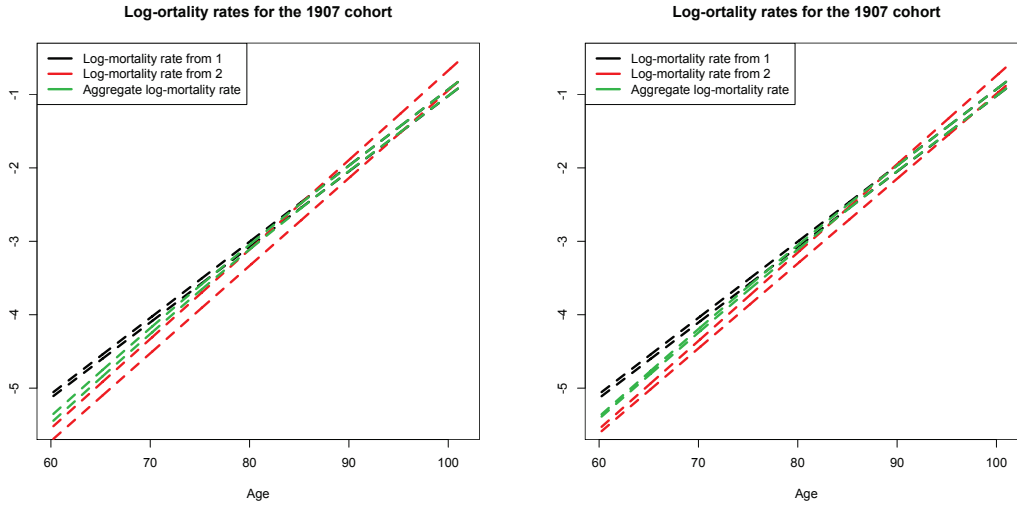
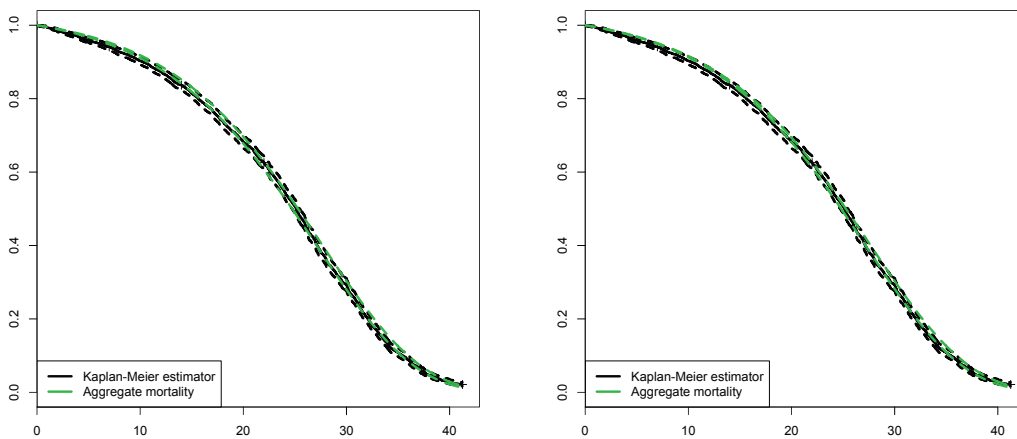


FIGURE 1.54 – Courbes de survie agrégée et intervalles de confiance ponctuels à 95% obtenues dans le Modèle 3 (gauche) et le Modèle 4 (droite) comparé à l'estimateur Kaplan-Meier



**Impact sur des produits d'assurance** Pour conclure cette étude, nous souhaitons étudier l'impact de l'hétérogénéité sur des produits d'assurance stylisés, en mettant l'accent sur les changements de caractéristiques et l'erreur d'estimation. On considère une population assurée de composition initiale  $G(\underline{a}) = \begin{pmatrix} g_1(\underline{a}) \\ g_2(\underline{a}) \end{pmatrix}$ , et on définit deux quantités actuarielles stylisées relatives à des rentes ou une assurance décès, qui démarre à l'âge  $\underline{a}$  et se termine à l'âge  $\bar{a}$ . On considère un taux d'actualisation constant  $r$ . Le produit de rentes (*Life Benefit*) consiste à payer 1 par unité

de temps aux individus encore en vie dans la population assurée, c'est-à-dire

$$LB := \int_{\underline{a}}^{\bar{a}} e^{-r(t-\underline{a})} g(t) dt,$$

où  $g(t)$  est donné par (1.54). Une autre quantité d'intérêt est l'assurance décès (*Death Benefit*) qui paye 1 au moment du décès d'un individu. Formellement, le nombre de décès dans l'intervalle de temps  $[t, t + dt)$  est  $g(t + dt) - g(t) \approx -g'(t)dt$ ; on définit ainsi le flux moyen payé par l'assureur par

$$DB := - \int_{\underline{a}}^{\bar{a}} e^{-r(t-\underline{a})} g'(t) dt.$$

Dans la suite, nous nous intéressons à l'impact sur ces deux types de quantités de (i) la composition de la population initiale, (ii) la prise en compte ou non des changements de caractéristiques et (iii) deux valeurs possibles du taux d'actualisation  $r$ .

#### Catégories socio-professionnelles

On se donne  $\underline{a} = 37.25$  et  $\bar{a} = 78.5$  qui correspondent à la plage d'âges étudiée pour l'estimation, et qui est plutôt typique d'un produit de bénéfice au décès. Les paramètres utilisés pour les simulations sont les suivants :

- (i) La population initiale à l'âge  $\underline{a} = 37.25$  est composée de CSP+ (groupe 2) seulement ( $p = 0$ ), ou bien représentative de l'échantillon initial ( $p = 0.703$ ), ou bien composée uniquement de CSP- (groupe 1) ( $p = 1$ ).
- (ii) La distribution aléatoire des flux liés au bénéfice au décès est simulée suivant le modèle complet de l'Équation (1.53), et comparée à ce même modèle lorsque l'on ne tient pas compte des transitions entre les groupes (dans ce cas  $\alpha_{12}(t) = \alpha_{21}(t) = 0$ ).
- (iii) Deux valeurs du taux d'actualisation  $r$  sont testées :  $r = 1\%$  en Figure 1.55 et  $r = 3\%$  en Figure 1.56.

Ces résultats sont représentés pour les deux Modèles 1 et 2.

Observons d'abord les résultats pour une composition initiale représentative de l'échantillon de départ ( $p = 0.703$ ), voir les colonnes médianes en Figures 1.55 et 1.56. Il apparaît dans ces simulations que considérer les caractéristiques stables au cours de la vie revient à légèrement sur-estimer les flux de bénéfice au décès, autrement dit à sous-estimer le nombre de personnes en vie, ou encore à sur-estimer la mortalité. Ainsi, il semble que la proportion d'individus du groupe 1 soit sur-estimée lorsque les changements de caractéristiques ne sont pas pris en compte. Rappelons que l'intensité de transition de 1 vers 2 est au global plus importante que celle de 2 vers 1, voir Figure 1.45. Mais également, que l'intensité totale des changements de 1 vers 2 est le produit entre l'intensité individuelle de transition  $1 \rightarrow 2$  et le nombre d'individus du groupe 1. Ainsi, lorsque la proportion d'individus du groupe 1 est suffisamment élevée, le nombre de transitions  $1 \rightarrow 2$  devient plus grand que celui des

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

---

transitions  $2 \rightarrow 1$ , et dans ce cas les flux de l'assurance décès sont bien sur-estimés lorsque les changements de caractéristiques ne sont pas pris en compte. On peut illustrer ce phénomène en partant d'une population initiale composée uniquement d'individus du groupe 1 ( $p = 1$ ), voir les colonnes de droite en Figures 1.55 et 1.56. En revanche, si la proportion d'individus du groupe 1 est suffisamment faible, les flux de bénéfice au décès sont sous-estimés si l'on considère que les caractéristiques sont stables au cours de la vie. On peut l'illustrer en partant cette fois-ci d'une population exclusivement composée d'individus du groupe 2 ( $p = 0$ ), voir les colonnes de gauche en Figures 1.55 et 1.56. Remarquons que comme attendu, ces flux sont plus faibles en partant d'individus du groupe 2, et cette approche permet au-delà d'illustrer que *toute la distribution* de ces flux dépend de la population considérée (tant en niveau qu'en terme d'incertitude due à l'erreur sur les paramètres). Ces conclusions sont similaires pour chacun des Modèles 1 et 2. Finalement, on peut tester la sensibilité au taux d'actualisation en comparant les Figures 1.55 et 1.56. Comme attendu, la valeur diminue avec l'augmentation du taux mais les conclusions d'impact sont similaires.

FIGURE 1.55 – Distribution des flux d'assurance décès avec  $r = 1\%$  et  $p = 0$  (gauche),  $p = 0.703$  (milieu) et  $p = 1$  (droite) pour le Modèle 1 (haut) et le Modèle 2 (bas)

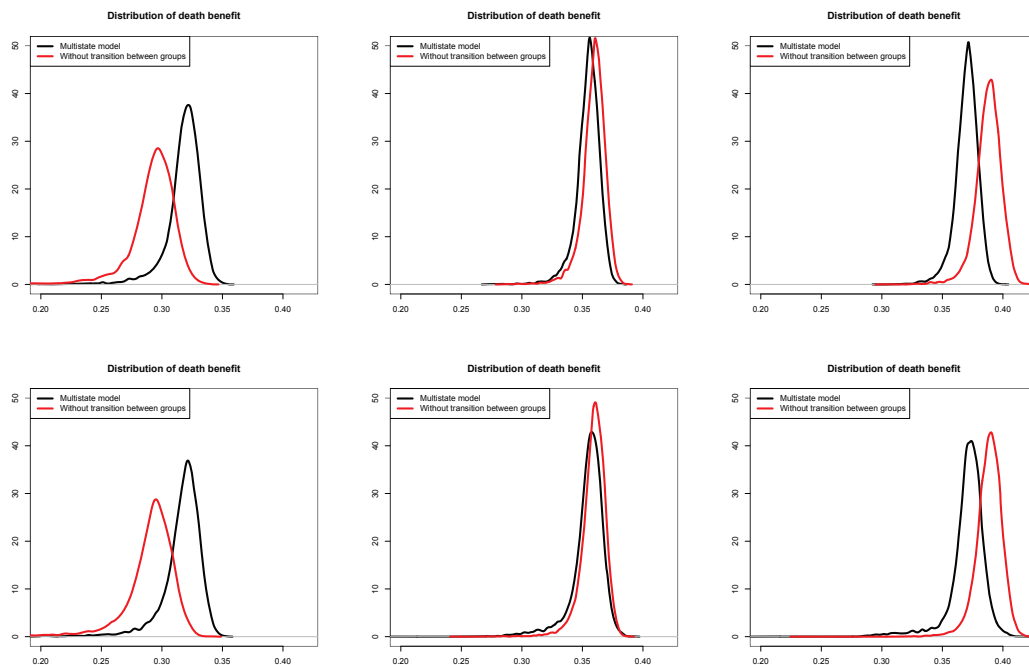
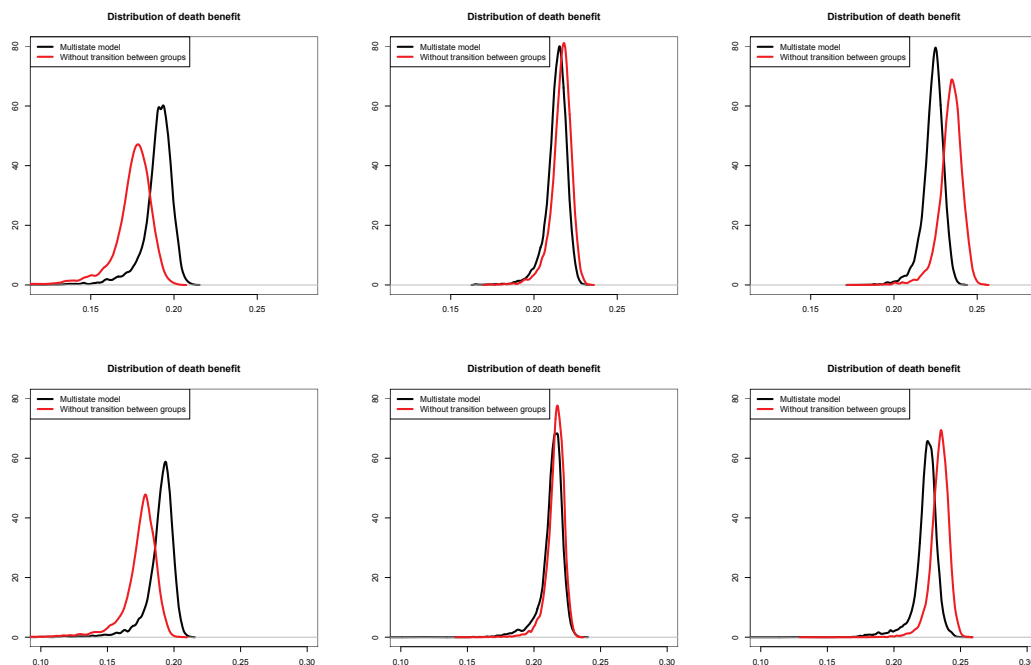


FIGURE 1.56 – Distribution des flux d'assurance décès avec  $r = 3\%$  et  $p = 0$  (gauche),  $p = 0.703$  (milieu) et  $p = 1$  (droite) pour le Modèle 1 (haut) et le Modèle 2 (bas)



### Statut matrimonial

Étudions maintenant l'hétérogénéité due au statut matrimonial, qui donne un certain nombre d'intuitions complémentaires. On se donne  $\underline{a} = 60.25$  et  $\bar{a} = 101.5$ , qui correspondent à la plage d'âge considérée pour l'estimation, qui est plutôt typique d'un produit de rentes. Pour ces simulations, on se donne les paramètres suivants :

- (i) La population initiale à l'âge  $\underline{a} = 60.25$  est constituée d'individus mariés (groupe 2) uniquement ( $p = 0$ ), ou bien représentative de l'échantillon de données ( $p = 0.329$ ), ou bien constituée d'individus non mariés (groupe 1) uniquement ( $p = 1$ ).
- (ii) La distribution aléatoire des flux de bénéficiaires au décès est simulée sur la base du modèle complet de population (1.53) et comparée au même modèle mais pour lequel les changements de caractéristiques ne sont pas pris en compte (c'est-à-dire  $\alpha_{12}(t) = \alpha_{21}(t) = 0$ ).
- (iii) Deux valeurs du taux d'actualisation sont comparées :  $r = 1\%$  en Figure 1.57 et  $r = 3\%$  en Figure 1.58.

Ces résultats sur le produit de rentes sont représentés pour les Modèles 3 et 4.

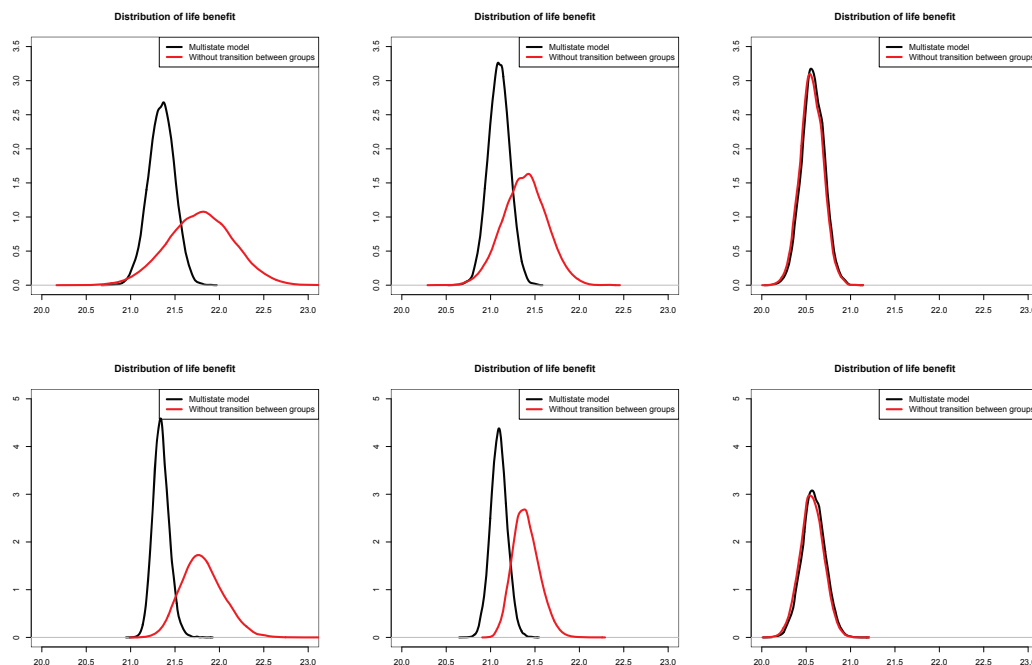
Avec ces résultats, l'impact de la transition de marié vers non-marié est clairement visible (étant majoritairement constitué des transitions vers le veuvage aux grands âges). Cet effet est visible sur les Figures 1.57 et 1.58 en partant d'individus mariés uniquement (colonne de gauche) ou d'une population représentative (colonne

## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

du milieu). En revanche, comme la transition de non-marié à marié est faible, c'est ici avec une population composée uniquement d'individus du groupe 1 que l'effet est le moins visible (c'était plutôt le cas de la population représentative concernant les résultats sur les catégories socio-professionnelles). Remarquons tout de même qu'une différence persiste, qui peut avoir un impact sur des montants de rente importants. Si l'on compare les Modèles 3 et 4, la plus grande différence réside dans la variance de la distribution des flux de rente, ceux-ci étant plus faibles dans le Modèle 4 ; ceci s'explique par le fait que dans ce modèle, l'incertitude sur les intensités de transition est plus faible tant pour les forces de mortalité (voir Figure 1.48) que pour les transitions entre les états intermédiaires (voir Figure 1.47).

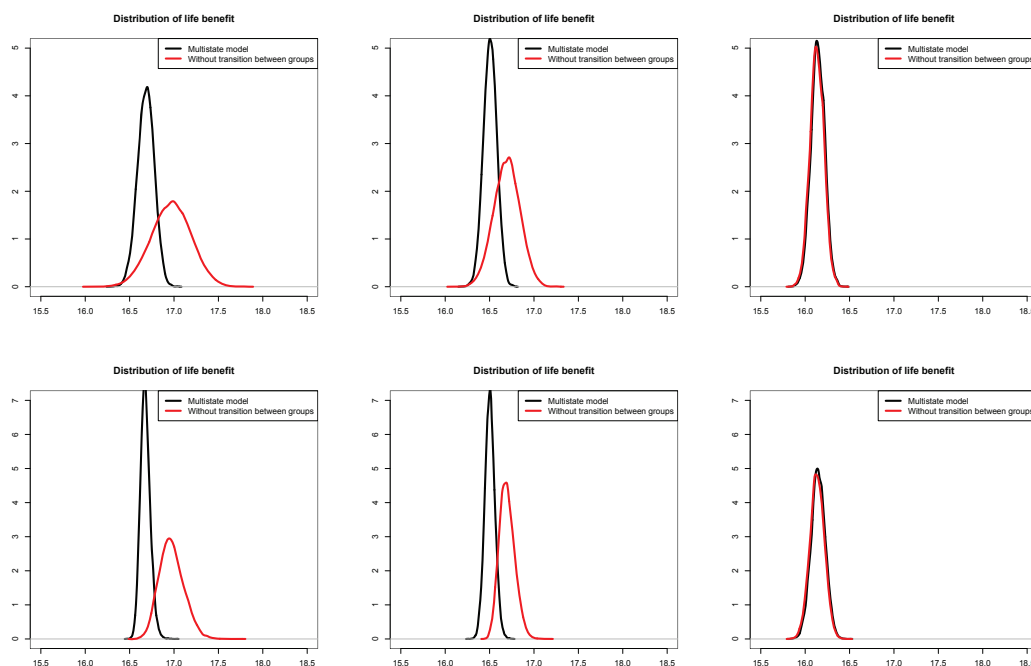
En conclusion de ces résultats numériques concernant l'hétérogénéité en terme de catégorie socio-professionnelle ou de statut matrimonial, insistons sur le fait que l'impact de la prise en compte des changements de caractéristiques dépend à la fois de la composition initiale de la population et des niveaux et fluctuations des intensités de transition entre les états intermédiaires.

FIGURE 1.57 – Distribution des flux de rentes avec  $r = 1\%$  et  $p = 0$  (gauche),  $p = 0.329$  (milieu) et  $p = 1$  (droite) pour le Modèle 3 (haut) et le Modèle 4 (bas)



## 1.5. ESTIMATION DE L'HÉTÉROGÉNÉITÉ DE LA MORTALITÉ EN PRÉSENCE DE CENSURE PAR INTERVALLE

FIGURE 1.58 – Distribution des flux de rentes avec  $r = 3\%$  et  $p = 0$  (gauche),  $p = 0.329$  (milieu) et  $p = 1$  (droite) pour le Modèle 3 (haut) et le Modèle 4 (bas)



### 1.5.6 Conclusion et perspectives

Nous avons proposé une méthode paramétrique par maximum de vraisemblance qui permet de capturer l'hétérogénéité de la mortalité en présence de changements de caractéristiques et de schéma d'observation censuré par intervalle, spécifique aux données démographiques longitudinales issues du recensement. En effet, pour de telles données, les individus ne sont observés qu'aux dates isolées du recensement, et les instants de changements de caractéristiques restent inconnus.

En appliquant une telle méthode à l'Échantillon Démographique Permanent de l'INSEE, qui présente à la fois un faible nombre d'individus mais également une censure par intervalle systématique, nous avons pu mesurer des forces de mortalité spécifiques à des catégories socio-professionnelles ou au statut matrimonial, en traitant à chaque fois des générations réelles. Nous avons également mis l'accent sur l'erreur d'estimation qui dépend de la caractéristique et de la transition considérées.

Nous avons observé plusieurs effets, incluant la loi de compensation de la mortalité aux grands âges entre les différentes sous-populations. Nous avons également étudié la dynamique de générations avec une composition spécifique, ce qui nous a permis d'avoir un certain nombre d'intuitions sur la structure en âge de la mortalité agrégée. Enfin, nous avons étudié la distribution de flux liés à des produits de rente ou d'assurance décès stylisés, et leur modification si les changements de caractéristiques ne sont pas pris en compte. En conclusion, il s'avère que le niveau

et les fluctuations de ces flux dépendent à la fois de la composition de la population initiale et des intensités de transition ; au-delà de la quantification de ces effets, ces intuitions peuvent être utiles pour la gestion des risques de mortalité/longévité.

Dans le champ des études démographiques et actuarielles, l'estimation de l'hétérogénéité de la mortalité à partir de données réelles complexes est un sujet difficile. Nous proposons ci-dessous plusieurs perspectives de travail.

(i) Ces estimations pourraient être mises en oeuvre en donnant la possibilité de plusieurs transitions dans chaque intervalle de temps et/ou en prenant en compte davantage de caractéristiques dans le modèle multi-états (par exemple en distinguant dans "non-marié" : "célibataire", "veuf", "divorcé"), et en essayant de tirer parti du peu de données alors disponibles.

(ii) Il serait intéressant de proposer un modèle général permettant d'estimer conjointement plusieurs générations. Un tel problème est assez délicat à plusieurs niveaux : d'abord le traitement de la censure par intervalle doit être cohérent entre les générations, ensuite la maximisation de vraisemblance doit être réalisée en prenant en compte la grande dimension du problème, augmentée par le fait qu'il est préférable d'éviter des hypothèses de hasard proportionnel (afin de répliquer la loi de compensation), et enfin prenant en compte la dépendance structurelle entre la date de recensement et la date de naissance des individus (les dates de recensement étant communes à tous, les âges à l'observation sont spécifiques à la cohorte considérée). Sur ce thème, le développement de méthodes d'estimation de modèles multi-états en présence de censure par intervalle *dépendante* paraît être un sujet statistique en soi.

(iii) Dans l'optique *population*, l'estimation statistique de taux de naissance hétérogènes est un enjeu crucial. Malheureusement, les données à disposition ne permettent pas de traiter cet aspect. À l'avenir, d'autres tests sur les taux de naissance hétérogènes pourraient compléter ceux du Chapitre 2 en reprenant par exemple les estimations de mortalité réalisées dans ce travail<sup>‡</sup>.

---

<sup>‡</sup>Ces aspects constituent des réflexions en cours avec Nicole El Karoui et Stéphane Loisel dans l'équipe de l'ANR Lolita, et également déclinées dans un contexte de mortalité par causes de décès et caractéristiques individuelles dans le projet PHC Germaine de Staël, joint avec Séverine Arnold et Héloïse Labit-Hardy du département de sciences actuarielles de l'Université de Lausanne.

## 1.6 Comportement macroscopique de trajectoires de vie accélérées

### 1.6.1 Motivations

L'étude de la mortalité par caractéristiques individuelles est un sujet délicat, surtout lorsque celles-ci évoluent au cours de la vie. Nous avons traité dans la section précédente un aspect de l'estimation statistique de modèles dits *multi-états* qui représentent des trajectoires de vie : l'individu passe par un certain nombre d'états jusqu'au décès. En parallèle de cette première approche, et pour avancer dans la compréhension des phénomènes démographiques, on peut se demander si l'évolution d'une population dans laquelle les individus changent de caractéristiques au cours de la vie ne peut pas être *approchée* par un modèle plus simple, dans lequel les paramètres seraient *moyennés* sur la population. En particulier, parmi les caractéristiques connues comme ayant un impact sur la mortalité, nombre d'entre elles peuvent varier très fréquemment en comparaison de l'échelle de temps d'une vie : on peut penser à l'exposition ou non à un contaminant, l'alternance entre une activité dangereuse et une activité sécurisée ou l'état de santé qui oscille entre les périodes de maladie et de rémission. Ces caractéristiques ont aussi un impact sur la fertilité, ce qui fait qu'au global la dynamique de population est difficile à analyser. On pense toujours dans ces travaux à l'évolution de la pyramide des âges au cours du temps, qui donne de l'information tant sur le calcul des indicateurs de mortalité agrégés que sur la répartition des générations pour la question des retraites.

Nous reprenons le modèle microscopique de dynamique de population introduit dans la Partie 1.2, dans lequel les individus peuvent donner naissance, changer de caractéristiques et décéder. Nous avons déjà mentionné dans la Partie 1.2 l'asymptotique en grande population de ce modèle : lorsque le nombre d'individus tend vers l'infini, le comportement macroscopique est donné par une équation aux dérivées partielles qui est une forme générale de l'équation de McKendrick Von Foerster (voir McKendrick (1926) et Von Foerster (1959)). Celle-ci donne l'évolution au cours du temps du nombre *moyen* d'individus d'un âge et de caractéristiques donnés. La contribution de cette partie est l'étude de l'asymptotique de ce modèle général à la fois en grande population et avec des changements de caractéristiques accélérés. En d'autres termes, il s'agit d'étudier la dynamique macroscopique du modèle lorsque deux échelles de temps sont présentes : les caractéristiques changent à l'échelle rapide tandis que le vieillissement, les naissances et les morts évoluent à l'échelle lente. Dans ce cadre, sous une hypothèse d'invariance sur les changements de caractéristiques, le comportement limite que nous obtenons est décrit par l'équation (déterministe) de McKendrick-Von Foerster dans laquelle l'âge seulement intervient, et les taux



de naissance et de mort sont moyennés sur la structure stable en terme de caractéristiques. Ainsi, une population hétérogène dans laquelle les individus changent de caractéristiques fréquemment a un comportement similaire à une population homogène dont les taux démographiques sont obtenus à partir de la mesure invariante des changements de caractéristiques.

On trouve dans la littérature de nombreux travaux sur les théorèmes limites pour les processus en présence de différentes échelles de temps. Quant au cadre qui nous intéresse, qui concerne plus particulièrement les processus de population à valeur mesure, on peut noter des contributions variées prenant en compte deux échelles de temps, par exemple concernant les mécanismes d'évolution écologique avec des mutations rares ou accélérées pour des populations structurées par caractéristiques (voir par exemple Fournier and Méléard (2004), Champagnat et al. (2006), Bovier and Wang (2013) et Billiard et al. (2014)), les modèles proies-prédateurs (voir par exemple Costa et al. (2015) et Costa (2015)), ainsi que les populations structurées avec vieillissement accéléré (voir Méléard and Tran (2012)). À notre connaissance, il n'y a pas de contribution traitant dans ce cadre de processus de population à valeur mesure avec changements de caractéristiques accélérés et taux de naissance et de mort dépendants de l'âge. L'objet de notre étude est de développer un tel cadre et les résultats correspondants, ainsi que d'illustrer l'apport de cette modélisation d'un point de vue démographique. La preuve est basée sur les outils et la démarche de Méléard and Tran (2012) et Gupta et al. (2014), travaux dans lesquels le vieillissement est accéléré. Nous reviendrons sur ces travaux à l'occasion d'une discussion dans la partie suivante. Dans la suite, nous détaillons le modèle renormalisé, les hypothèses ainsi que le résultat principal associé. Nous terminons par quelques exemples illustratifs de distributions stables de caractéristiques et de taux démographiques équivalents pour le modèle approché. On trouvera dans le Chapitre 6 le détail de ce travail, en particulier l'ensemble des résultats et des preuves, ainsi que les exemples et illustrations numériques.

### 1.6.2 Le modèle

Nous reprenons le modèle microscopique de dynamique de population introduit dans la Partie 1.2, avec quelques spécifications que nous détaillons dans la suite. Chaque individu a un âge (positif)  $a$  et une caractéristique  $x \in \mathcal{X}$ , où  $\mathcal{X}$  est un compact de  $\mathbb{R}^d$ . On rappelle que dans la population, trois types d'événements peuvent se produire :

- (i) Une naissance, c'est-à-dire l'arrivée d'un individu d'âge zéro,
- (ii) Un décès, c'est-à-dire la disparition d'un individu,
- (iii) Un *swap*, c'est-à-dire le changement de caractéristiques d'un individu.

L'objectif est de modéliser le fait que les individus changent de caractéristiques

de manière très fréquente relativement aux temps de naissance et de mort. Pour ce faire, nous spécifions les intensités d'événements de la façon suivante. Chaque individu de caractéristiques  $x$  et d'âge  $a$  donne naissance au taux  $b(x, a)$ , décède au taux  $d(x, a)$ , et change de caractéristiques aux temps donnés par le taux de *swap*  $n.e(x, a)$ , où  $n$  est le paramètre d'échelle que nous ferons tendre vers l'infini. Comme nous nous intéressons aussi au comportement macroscopique, le paramètre d'échelle  $n$  sera utilisé comme l'ordre de grandeur de la taille de la population. Il nous reste à spécifier les changements de caractéristiques qui ont lieu au moment du *swap*, et également les mutations possibles à la naissance. Pour cela, nous considérons une mesure de probabilité de référence  $m(dx)$  sur l'espace  $\mathcal{X}$  des caractéristiques. Au moment d'un *swap*, les caractéristiques  $x$  d'un individu d'âge  $a$  sont remplacées par de nouvelles caractéristiques  $x'$  tirées suivant le noyau  $k_e(x, a, x')m(dx')$ . À la naissance d'un individu de parent  $(x, a)$ , des mutations se produisent suivant le noyau  $k_b(x, a, x')m(dx')$ .

Pour un entier  $n$  donné, nous construisons un processus de population  $(Z_t^n(dx, da))_{t \geq 0}$  qui part à l'instant initial d'une population aléatoire  $Z_0^n(dx, da)$ . À un instant  $t$ , la mesure  $Z_t^n(dx, da)$  met un poids  $(x, a)$  en les caractéristiques et l'âge de chaque individu, qui évolue suivant les taux  $b(x, a)$ ,  $d(x, a)$  et  $n.e(x, a)$ . Ce type de construction est détaillé dans la Partie 1.2 et le Chapitre 2, et explicité dans ce cadre précis dans le Chapitre 6. Pour étudier le comportement macroscopique, on définit pour chaque  $n \in \mathbb{N}^*$  le processus de population renormalisé  $(\tilde{Z}_t^n)_{t \geq 0}$  par

$$\tilde{Z}_t^n(dx, da) := \frac{1}{n} Z_t^n(dx, da),$$

qui correspond à la population dans laquelle chaque individu a le poids  $\frac{1}{n}$ . Nous travaillerons sous des hypothèses permettant la stabilité de la composition en terme de caractéristiques, et nous serons alors intéressés par le comportement de la suite des processus renormalisés de pyramide des âges lorsque le paramètre d'échelle  $n$  tend vers l'infini. Pour cela, nous introduisons la suite de processus à valeur mesure  $(\bar{Z}_t^n(da))_{t \geq 0}$  définis par la marginale en âge par : pour toute  $f \in C_b(\mathbb{R}_+)$  (continue et bornée sur  $\mathbb{R}_+$ ),  $\int_{\mathbb{R}_+} f(a) \bar{Z}_t^n(da) = \int_{\mathcal{X} \times \mathbb{R}_+} f(a) \tilde{Z}_t^n(dx, da)$  (soit avec nos notations  $\langle \bar{Z}_t^n, f \rangle = \langle \tilde{Z}_t^n, f \rangle$ ).

On s'intéresse à la limite quand  $n$  croît vers l'infini, donc la fréquence  $n.e(x, a)$  des changements de caractéristiques augmente, tandis que les taux de naissance et de mort restent les mêmes. L'idée, comme nous l'avons déjà évoquée, est de séparer les échelles de temps des changements de caractéristiques d'une part, et des événements démographiques (vieillessement, naissances, morts) d'autre part. Dans le même temps, la taille de la population tendra vers l'infini. Au total, deux effets se combinent :

(i) Comme la population tend vers l'infini, le phénomène aléatoire se moyennise en

une dynamique déterministe,

(ii) En parallèle, les événements de *swap* se produisent de plus en plus fréquemment, ce qui amène la répartition des caractéristiques à un équilibre dans la population globale.

Ainsi, dans la limite où  $n \rightarrow +\infty$ , le processus de pyramide des âges devrait satisfaire une équation déterministe en âge, dans laquelle les paramètres démographiques (fertilité et mortalité) sont moyennés sur la répartition stable des caractéristiques.

### 1.6.3 Hypothèses

Nous commençons par détailler les hypothèses sur le contrôle des paramètres démographiques ainsi que le contrôle des populations initiales.

**Hypothèse 8.** (*Taux et noyaux bornés*) *Les taux démographiques et les noyaux de changement sont continus en chaque variable, et il existe des constantes positives  $\bar{k}$ ,  $\bar{b}$ ,  $\bar{d}$  et  $\bar{e}$  telles que pour tout  $(x, a, x') \in \mathcal{X} \times \mathbb{R}_+ \times \mathcal{X}$ ,  $k^e(x, a, x') \leq \bar{k}$ ,  $k^b(x, a, x') \leq \bar{k}$ ,  $b(x, a) \leq \bar{b}$ ,  $d(x, a) \leq \bar{d}$  et  $e(x, a) \leq \bar{e}$ .*

**Hypothèse 9.** (*Contrôle de moments pour les populations initiales aléatoires*)

$$\exists \alpha > 0, \sup_n \mathbb{E} \left[ \left( \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right)^{1+\alpha} \right] < +\infty.$$

**Remarque 4.** *La preuve peut être adaptée (et même légèrement simplifiée) dans le cas où les âges évoluent dans un compact, par exemple de la forme  $[0, \bar{a}]$ , où  $\bar{a}$  est un âge limite. Dans ce cadre, on peut vérifier que l'hypothèse  $\sup_n \mathbb{E} \left[ \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right] < +\infty$  est suffisante. Néanmoins, il est dommage de contraindre le modèle dans cette direction (dont la dynamique présenterait d'ailleurs des modifications) et également de perdre la possibilité d'avoir des âges dans  $\mathbb{R}_+$ , ce qui est d'intérêt pour de nombreuses applications. En particulier, si l'on pense à la possibilité d'un plateau pour le taux de mortalité, qui reste donc fini et invalide la présence d'un âge limite, ou si l'on souhaite inclure un âge limite qui évolue au cours du temps. C'est également le cas si l'on souhaite appliquer ces limites d'échelles aux processus de Hawkes généraux (voir Partie 1.3 et Chapitre 3), pour lesquels la dynamique en âge n'est pas contrainte.*

L'hypothèse cruciale est donnée ci-dessous. Celle-ci indique qu'à chaque âge, la dynamique individuelle de changement de caractéristiques suit un processus qui possède une mesure invariante. Ainsi, si l'on accélère ce processus de sorte à ce qu'il atteigne sa mesure invariante, alors on connaît pour cet âge la distribution possible des caractéristiques; cette dynamique individuelle va se traduire au niveau macroscopique par une stabilisation de la répartition des caractéristiques à chaque âge.

**Hypothèse 10.** (*Mesure invariante*) Pour tout  $a \in \mathbb{R}_+$ , il existe une solution non nulle  $x \mapsto g(x, a)$  telle que  $\int_{\mathcal{X}} g(x, a) l(dx) = 1$  à l'équation de Fredholm : m-p.p.,

$$e(x, a)g(x, a) = \int_{\mathcal{X}} g(y, a)e(y, a)k_e(y, a, x)m(dy). \quad (1.56)$$

Cette hypothèse s'interprète classiquement de manière probabiliste. Nous omettons l'âge pour la discussion suivante, et nous notons  $g(x) = g(x, a)$ . Considérons un processus de Markov à valeurs dans  $\mathcal{X}$ , qui saute d'un état  $x$  à un état  $y$  au taux  $e(x)k_e(x, y)$ . Ce processus représente les changements de caractéristiques au cours de la vie d'un individu. Le générateur infinitésimal de ce processus est défini sur les fonctions continues  $f$  par

$$\mathcal{A}f(x) = e(x) \left( \int_{y \in \mathcal{X}} k_e(x, y)f(y)m(dy) - f(x) \right).$$

On voit que sous l'Hypothèse 10, pour toute fonction  $f$  continue,  $\int_{\mathcal{X}} \mathcal{A}f(x)g(x)m(dx) = 0$ . Dans le cas où le processus de Markov est continu à droite, ceci caractérise le fait que son unique mesure invariante (de probabilité) n'est rien d'autre que  $\nu(dx) = g(x)m(dx)$ .

Il reste à donner l'hypothèse sur la convergence de la suite de populations initiales. Pour cela, nous rappelons ici l'espace des mesures considéré et sa topologie, ce qui a été déjà fait dans la Partie 1.2.4. Nous considérons ici des mesures aléatoires à valeur dans  $\mathcal{M}(E)$ , l'espace des mesures positives finies sur un ensemble  $E$ , et dans notre cadre  $E$  sera  $\mathbb{R}_+$  ou  $\mathbf{S} = \mathcal{X} \times \mathbb{R}_+$ . L'espace  $\mathcal{M}(E)$  peut être muni de la topologie de la convergence vague ou étroite. Rappelons que la convergence étroite dans  $\mathcal{M}(E)$  est définie par  $\nu_n \xrightarrow{weak} \nu$  si pour toute fonction  $f$  continue et bornée,  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$ . La convergence vague est définie par  $\nu_n \xrightarrow{vague} \nu$  si pour toute fonction  $f$  continue à support compact,  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$ . Si  $E$  était un ensemble compact, ces deux topologies seraient les mêmes, mais dans notre cas où  $E$  n'est pas compact, celles-ci présentent une inclusion stricte. L'hypothèse de convergence de la population initiale est donnée ci-dessous.

**Hypothèse 11.** (*Convergence de la population initiale*) Il existe une mesure déterministe  $\xi_0 \in \mathcal{M}(\mathbf{S})$  telle que la suite de mesures aléatoires  $(\tilde{Z}_0^n(dx, da))_{n \in \mathbb{N}^*}$  converge en loi et pour la topologie de la convergence étroite dans  $\mathcal{M}(\mathbf{S})$  vers  $\xi_0(dx, da)$ .

### 1.6.4 Résultat principal

Le résultat principal est donné ci-dessous, la preuve quant à elle est détaillée dans le Chapitre 6. Rappelons les notations :  $\mathbb{D}([0, T], (\mathcal{M}(\mathbb{R}_+), w))$  (resp.  $\mathbb{C}([0, T], (\mathcal{M}(\mathbb{R}_+), w))$ ) est l'espace des processus càdlàg (resp. continus) sur  $[0, T]$  à valeur dans  $\mathcal{M}(\mathbb{R}_+)$  muni de la topologie de la convergence étroite.

**Théorème 3.** (i) *Sous les Hypothèses 8, 9, 10, et 11, la suite des processus de pyramide des âges  $(\bar{Z}^n(da))_n$  converge en loi dans  $\mathbb{D}([0, T], (\mathcal{M}(\mathbb{R}_+), w))$  vers le processus à valeur mesure déterministe  $\xi \in \mathbb{C}([0, T], (\mathcal{M}(\mathbb{R}_+), w))$ , unique solution à l'équation*

$$\langle \xi_t, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \xi_s, \partial_a f + \hat{b} - \hat{d} \rangle ds, \quad (1.57)$$

où

$$\hat{b}(a) = \int_{\mathcal{X}} b(x, a)g(x, a)m(dx), \quad \hat{d}(a) = \int_{\mathcal{X}} d(x, a)g(x, a)m(dx).$$

(ii) *Pour chaque  $t > 0$  fixé, la suite de pyramide des âges  $(\tilde{Z}_t^n(dx, da))_n$  converge en loi dans  $(\mathcal{M}(\mathbf{S}), w)$  vers  $\tilde{\xi}_t(dx, da) := g(x, a)\xi_t(da)m(dx)$ .*

L'Équation (1.57) indique que le comportement macroscopique de la pyramide des âges est dirigé par les taux de naissance et mort  $\hat{b}(a)$  and  $\hat{d}(a)$ , dont la valeur à chaque âge est la moyenne sur la répartition stable en terme de caractéristiques. Cela revient à réduire le nombre de paramètres : on remplace les taux de naissance et mort par caractéristiques par un équivalent et l'on obtient ainsi un modèle simplifié. De plus, le point (ii) montre que la dépendance en temps ne concerne que la pyramide des âges, et qu'à chaque instant, la composition des caractéristiques pour la classe d'âge (infinitésimale)  $a$  est donnée par  $g(x, a)m(dx)$ . Notons qu'en présence des différences d'échelle entre les événements de naissance et de changements de caractéristiques, le noyau de mutation à la naissance n'intervient pas dans la dynamique macroscopique. Mentionnons enfin qu'il n'est pas possible d'améliorer le résultat du point (ii) car la suite de processus  $(\tilde{Z}^n)$  n'est pas tendue dans  $\mathbb{D}([0, T], (\mathcal{M}(\mathcal{X} \times \mathbb{R}_+), v/w))$  : ceci sera discuté dans le Chapitre 6, voir Proposition 6.2.

Avant de présenter des illustrations numériques associées à ce résultat, nous faisons ci-dessous le lien avec la démarche de Méléard and Tran (2012) qui est à la base de la preuve, celle-ci étant détaillée dans le Chapitre 6.

**Le cadre de Méléard and Tran (2012)** Dans Méléard and Tran (2012), la limite de suite de processus de naissance et mort est étudiée en présence de composante allométrique, c'est-à-dire, en plus de l'évolution lente d'une population soumise à des naissances et des morts, des naissances et des morts additionnelles rapides, de même intensité. Cette modélisation représente la cohabitation de deux populations au niveau écologique, dont les dynamiques démographiques sont très différentes (par exemple des hôtes éléphants et des parasites insectes). L'inclusion de l'âge dans la modélisation rend l'étude des dynamiques allométriques difficiles. Ceci est dû à l'asymétrie entre les deux événements rapides de même intensité : la naissance qui ajoute un individu d'âge zéro d'une part, et la mort qui retire un individu d'âge strictement positif; ce qui a pour conséquence de dégénérer la mesure qui garde

trace des âges dans la population (pyramide des âges). Ce problème technique peut être surmonté en supposant que le vieillissement se déroule à l'échelle rapide, puis en se basant sur des techniques de moyennisation de Kurtz (1992). Dans la limite de naissances et morts accélérées, la pyramide des âges atteint un équilibre à chaque instant. Ainsi, la limite obtenue par Méléard and Tran (2012) indique que l'évolution en âge est perdue (composante rapide), et qu'alors l'évolution de la structure de la population en terme de caractéristiques, qui évolue à l'échelle lente, est décrite par une dynamique aléatoire donnée par une diffusion de Feller. La dynamique qui nous intéresse présente de nombreuses analogies et nous reprenons dans la preuve la démarche de Méléard and Tran (2012) (voir aussi Gupta et al. (2014)). Dans notre cadre cependant, la situation est inversée : le vieillissement est à l'échelle lente, tandis que les caractéristiques évoluent à l'échelle rapide. De manière analogue, à la limite la structure en caractéristiques est stabilisée et une équation décrit la dynamique de la pyramide des âges. Cependant, la particularité du phénomène de changement de caractéristiques donne à la limite une équation déterministe. Ceci s'explique de la façon suivante : alors que les temps de naissances et les morts (indépendants) ne peuvent se produire au même moment, il reste un bruit à la limite, tandis que dans notre cas, les changements de caractéristiques sont simultanés. En effet, de manière artificielle, on enlève l'ancienne caractéristique et on ajoute la nouvelle au même instant.

### 1.6.5 Illustration numérique

Afin d'illustrer les hypothèses et le résultat principal, nous proposons de calculer un exemple de solution à l'Équation (1.56) pour la composition stable (un exemple supplémentaire est donné dans le Chapitre 6), et d'illustrer numériquement l'impact de différents choix de fréquences de changements de caractéristiques sur les taux de naissance et mort agrégés. Nous traitons le cas d'un mécanisme de saut au plus proche voisin, que nous décrivons ci-après. Nous ne mentionnons pas la dépendance en âge pour simplifier les notations et l'application, mais il faut garder à l'esprit que l'intensité et le noyau de changement de caractéristiques peuvent en dépendre.

Considérons  $n$  classes de caractéristiques, soit  $\mathcal{X} = \{x_1, \dots, x_n\}$ . Dans ce cadre, l'équation de la mesure invariante est donnée de manière discrète par

$$\forall i \in \{1, \dots, n\}, e(x_i)g(x_i) = \sum_{j \neq i} g(x_j)e(x_j)k_e(x_j, x_i).$$

Si un changement a lieu pour un individu de caractéristiques  $x_i, i \in \{2, \dots, n-1\}$ , alors ses nouvelles caractéristiques sont choisies uniformément entre  $x_{i-1}$  et  $x_{i+1}$ , en d'autres termes  $k_e(x_i, x')m(dx') = \frac{1}{2}\delta_{x_{i-1}}(dx') + \frac{1}{2}\delta_{x_{i+1}}(dx')$ . Si la caractéristique est  $x_i = x_1$  ou  $x_i = x_n$ , le voisin est choisi avec probabilité 1, c'est-à-dire

$k_e(x_1, x')m(dx') = \delta_{x_2}(dx')$  et  $k_e(x_n, x')m(dx') = \delta_{x_{n-1}}(dx')$ . Définissons la matrice  $A$  par

$$A_{i,j} = e(x_j)k_e(x_j, x_i)\mathbf{1}_{i \neq j} - e(x_i)\mathbf{1}_{i=j}.$$

Pour résoudre l'équation de la mesure invariante, on cherche les vecteurs propres de la matrice  $A$  associés à la valeur propre 0. Comme les changements de caractéristiques ne peuvent se produire qu'aux plus proches voisins, la matrice  $A$  est tri-diagonale. Aussi, les sur- et sous-diagonales ont le même signe, ce que qui montre que  $A$  est diagonalisable. On peut alors montrer que 0 est une valeur propre pour  $A$  (de multiplicité 1) et qu'un vecteur propre  $Y$  est donné par  $Y_1 = \prod_{i \neq 1} e(x_i)$ ,  $Y_n = \prod_{i \neq n} e(x_i)$  et pour  $k \in \{2, \dots, n-1\}$ ,  $Y_k = 2 \prod_{i \neq k} e(x_i)$ . Ainsi, la solution  $g$  est donnée par

$$g(x_i) = \frac{Y_i}{\sum_{1 \leq k \leq n} Y_k}.$$

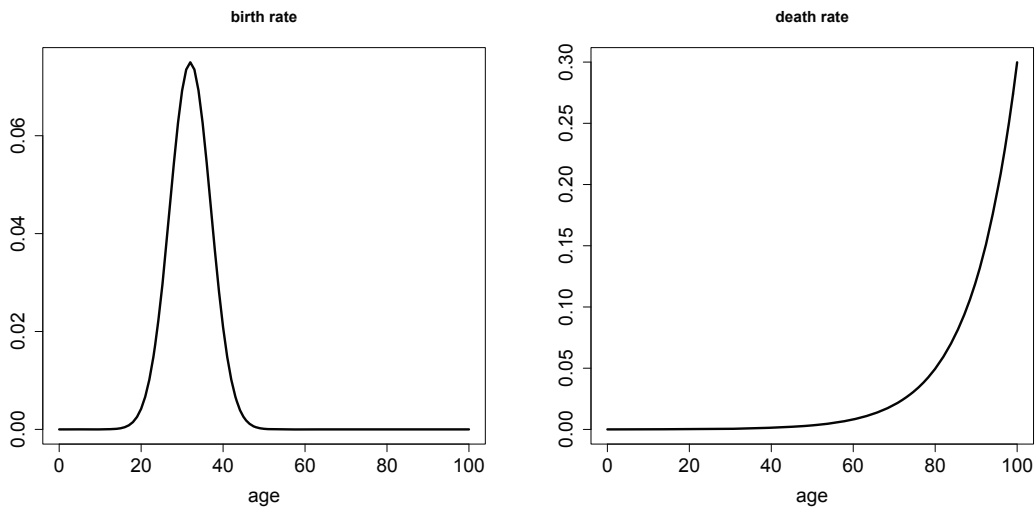
Pour mettre en œuvre les calculs numériques, nous proposons de considérer un modèle à hasard proportionnel du type  $d(x, a) := \alpha(x)\bar{d}(a)$  et  $b(x, a) := \beta(x)\bar{b}(a)$ , avec  $\bar{d}(a)$  et  $\bar{b}(a)$  des taux de naissance et mort de référence, dont nous capturons la structure en âge sur des données démographiques réelles. Dans le modèle macroscopique, le taux de mort agrégé est  $\hat{d}(a) = \int_{\mathcal{X}} d(x, a)g(x, a)m(dx) = \hat{\alpha}(a)\bar{d}(a)$ , où  $\hat{\alpha}(a) := \int_{\mathcal{X}} \alpha(x)g(x, a)m(dx)$ . De même, on a  $\hat{b}(a) = \hat{\beta}(a)\bar{b}(a)$ . Il est intéressant de noter que dans le cadre général, on n'obtient donc pas des taux de mort et naissance agrégés qui sont proportionnels aux taux de référence, dû au fait que la structure stable en terme de caractéristiques peut dépendre de l'âge considéré. Dans l'exemple numérique cependant, comme les taux de swap et donc la structure stable ne dépendent pas de l'âge, on restera pour les taux agrégés dans un modèle à hasard proportionnel. Pour le taux de mort de référence, nous proposons une forme stylisée de type Gompertz :  $\bar{d}(a) = A_1 e^{B_1 a}$ , et un taux de naissance qui reproduit la forme de ceux des populations humaines, soit ici  $\bar{b}(a) = A_2 e^{-B_2(a-\bar{a})^2}$ . Nous choisissons des valeurs qui reproduisent globalement le niveau des taux démographiques pour la France en 2008\*, soit ici :

$A_1 = 3, 7.10^{-5}$ ,  $B_1 = 0.09$ ,  $\bar{a} = 32$ ,  $A_2 = 0.075$  et  $B_2 = 0.02$ . Les taux de référence basés sur ces paramètres sont représentés sur la Figure 1.59.

---

\*Pour les taux de naissance, nous utilisons les ordres de grandeur des taux par femme et divisons par deux pour récupérer grossièrement un taux de naissance individuel (sans sexe spécifié).

FIGURE 1.59 – Taux de naissance de référence  $\bar{b}(a)$  et de décès de référence  $\bar{d}(a)$  en fonction de l'âge  $a$ .

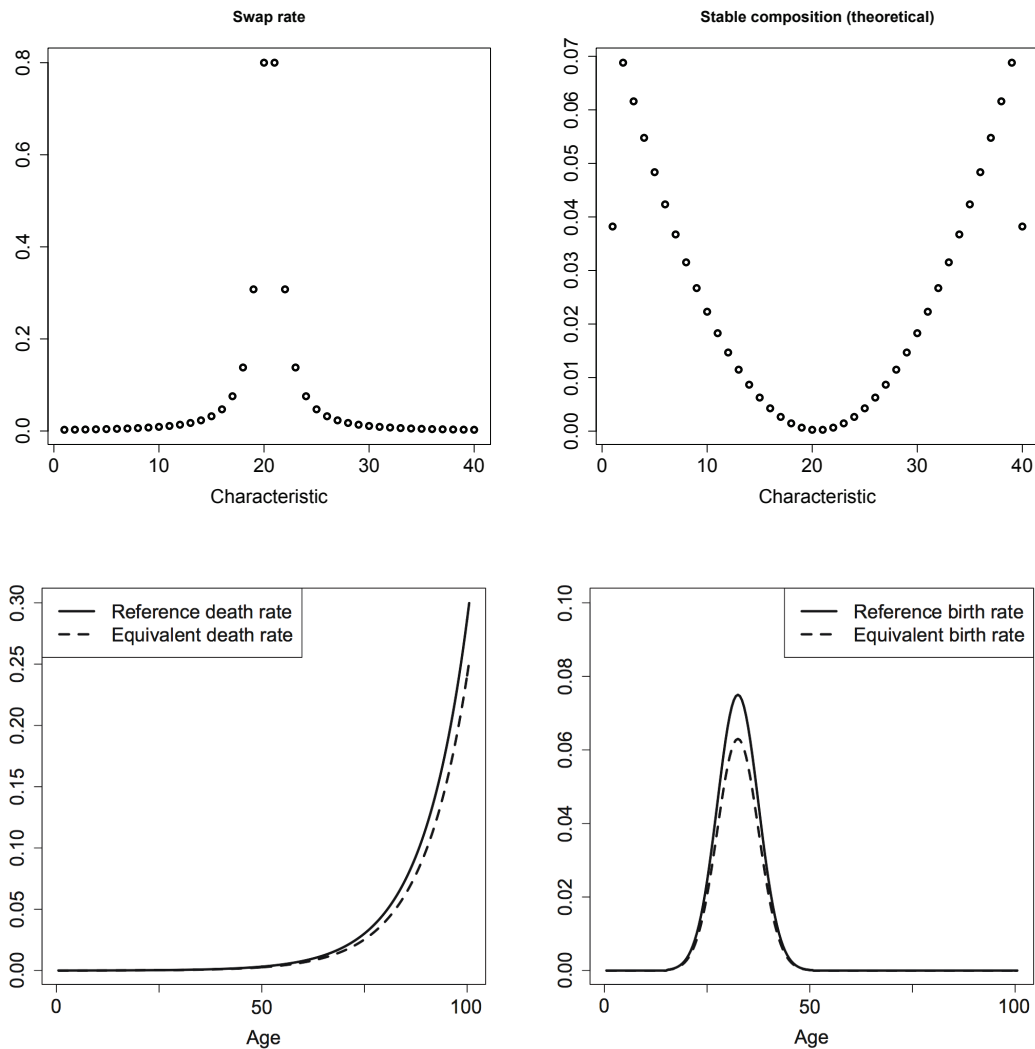


Pour l'espace des caractéristiques, on choisit  $n = 40$  classes discrètes et on pose  $\mathcal{X} = \{1, \dots, 40\}$ . Nous proposons de considérer qu'une caractéristique  $x$  élevée implique de vivre plus longtemps mais de donner moins d'enfants : posons  $\alpha(x) = \beta(x) = 1 - \frac{x-4,5}{100}$ . Nous choisissons quatre formes possibles pour le taux de swap :  $e(x) = \frac{1}{1-(x-20,5)^2}$ ,  $e(x) = \frac{1}{x}$ ,  $e(x) = x/40$  and  $e(x) = \sqrt{x/40}$ . Sur les Figures 1.60 à 1.63, nous représentons pour chacune des quatre formes : le taux de swap, la structure stable en terme de caractéristiques, et les taux de naissance et morts agrégés. Insistons sur le fait que dans chacune des quatre configurations, les taux individuels de naissance et mort  $b(x, a)$  et  $d(x, a)$  sont les mêmes, ainsi que la façon dont les nouvelles caractéristiques sont choisies au moment d'un événement de swap (saut au plus proche voisin). Il s'agit donc ici de faire varier la fréquence des événements de changements de caractéristiques. Concentrons nous d'abord sur les deux graphiques supérieurs pour chacune des quatre expériences. Ceux-ci représentent l'impact de la forme du taux de swap (en haut à gauche) sur la structure stable en terme de caractéristiques (en haut à droite). On remarque en particulier que comme attendu, les caractéristiques pour lesquelles le taux de swap est élevé sont moins représentées dans la structure stable de population, car en effet les individus avec de telles caractéristiques sont "forcés" à quitter leur classe plus souvent. Remarquons également l'effet de bord dans la composition stable, due au choix du mécanisme de saut au plus proche voisin : les caractéristiques sur le bord ont une probabilité plus faible d'être choisies, n'ayant qu'un seul voisin (au lieu de deux pour les autres). Observons maintenant les taux de naissance et mort équivalents. Remarquons d'abord qu'ils sont toujours plus faibles que les taux de référence (dont l'objectif est uniquement de répliquer une structure en âge réaliste à partir des don-



nées) : ceci est dû au fait que les coefficients de proportionnalité  $\alpha(x)$  et  $\beta(x)$  sont inférieurs à 1 pour la plupart des caractéristiques. On observe sur ces graphiques les effets de la fréquence des changements de caractéristiques en comparant les taux agrégés des Figures 1.60 à 1.63. Par exemple, le choix d'un taux de swap  $e(x) = x/40$  donne des taux de naissance et mort agrégés plus élevés que dans les trois autres configurations ; ceci s'explique, comme détaillé précédemment, par le fait que les caractéristiques dont les taux sont les plus proches du taux de référence sont sur-représentées. Ceci illustre comment la fréquence des changements de caractéristiques influence les taux de naissance et mort observés à l'échelle macroscopique.

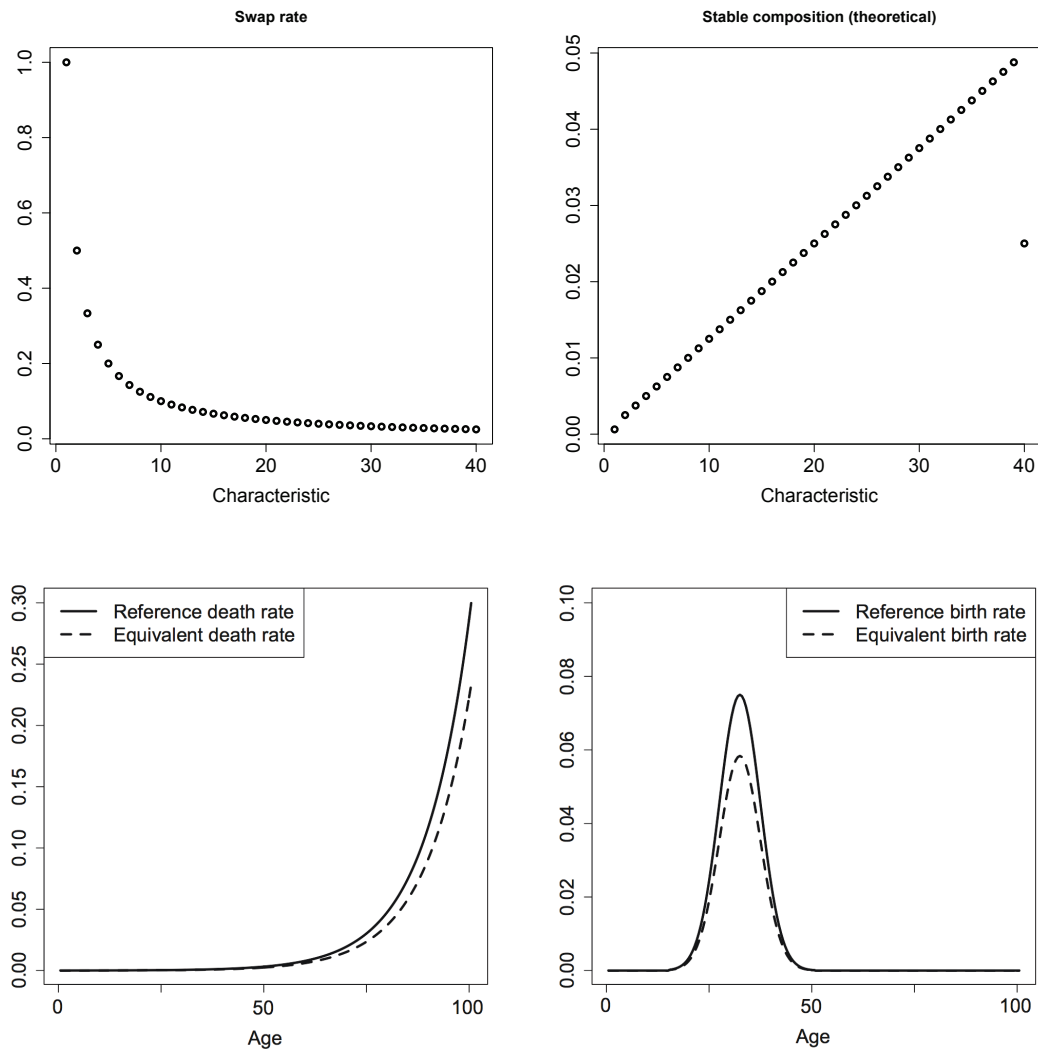
FIGURE 1.60 – Taux de swap, composition stable et taux de mort et naissance agrégés avec  $e(x) = \frac{1}{1-(x-20,5)^2}$



## 1.6. COMPORTEMENT MACROSCOPIQUE DE TRAJECTOIRES DE VIE ACCÉLÉRÉES

---

FIGURE 1.61 – Taux de swap, composition stable et taux de mort et naissance agrégés avec  $e(x) = \frac{1}{x}$



## 1.6. COMPORTEMENT MACROSCOPIQUE DE TRAJECTOIRES DE VIE ACCÉLÉRÉES

---

FIGURE 1.62 – Taux de swap, composition stable et taux de mort et naissance agrégés avec  $e(x) = x/40$

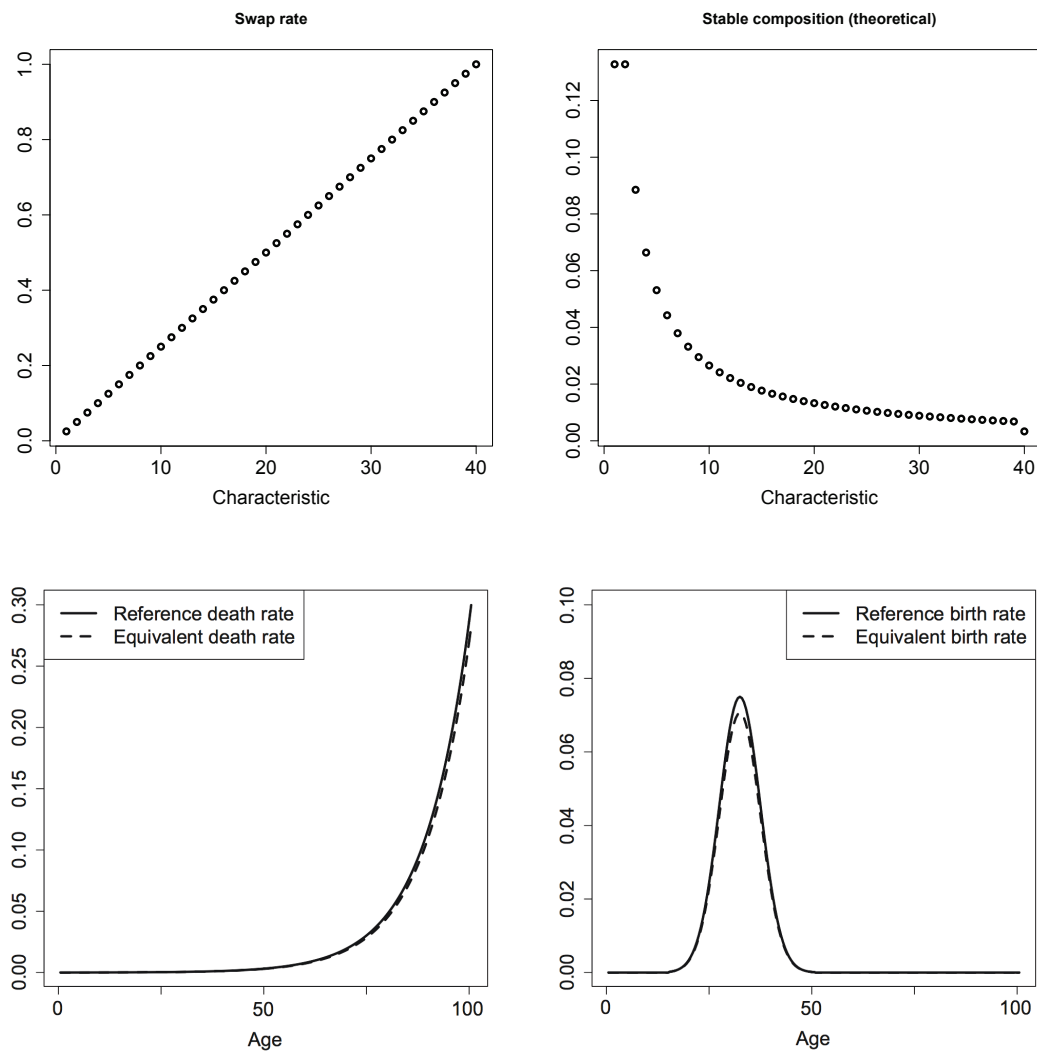
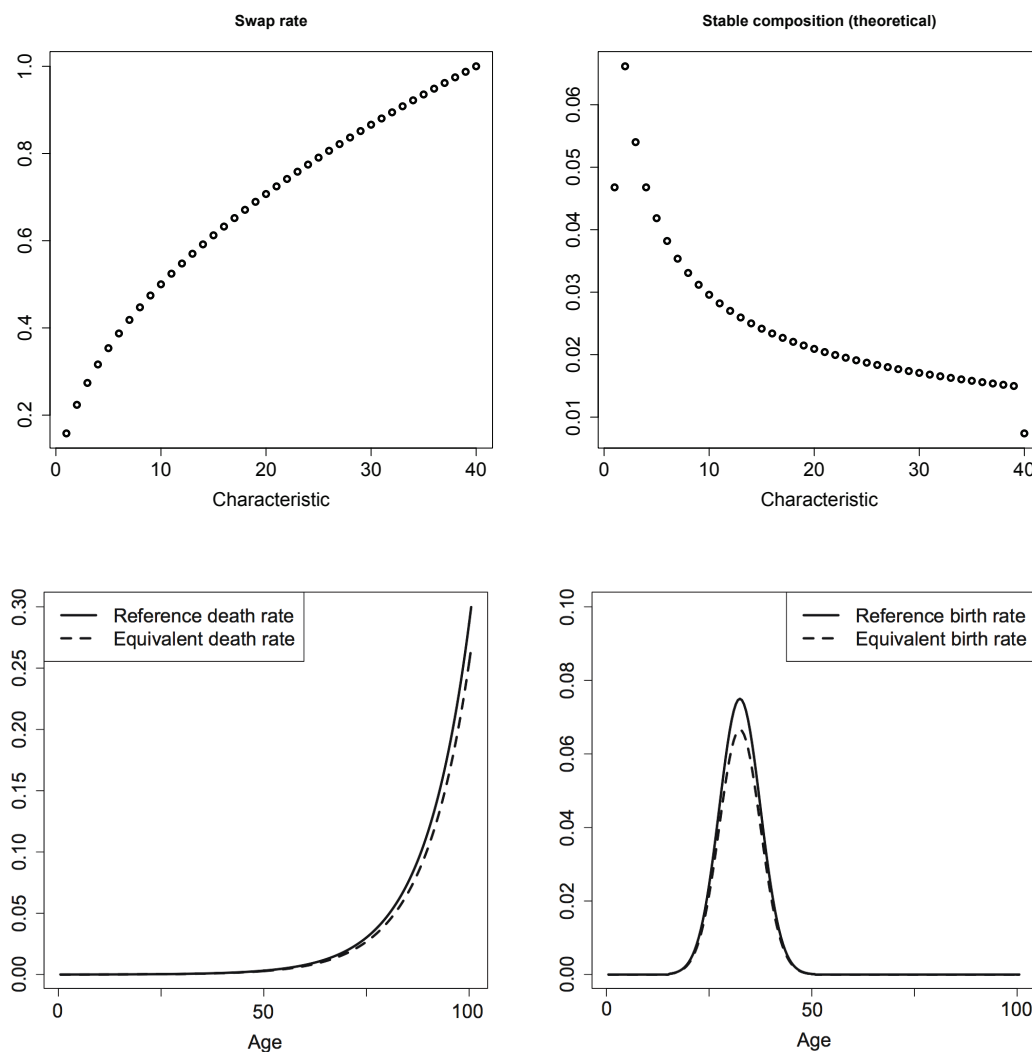


FIGURE 1.63 – Taux de swap, composition stable et taux de mort et naissance agrégés avec  $e(x) = \sqrt{x/40}$



### 1.6.6 Conclusion et perspectives

Nous avons proposé d'étudier la limite d'échelle d'un modèle dynamique de population microscopique, dans lequel les individus ont un âge et des caractéristiques, prenant en compte des naissances, des morts et des changements de caractéristiques au cours de la vie. La limite d'échelle repose sur deux hypothèses : celle de grande population d'une part, et celle de changements rapides de caractéristiques d'autre part. Ces hypothèses sont motivées par l'étude de dynamiques de populations humaines observées à une échelle macroscopique comme celle d'un pays. Le résultat que nous obtenons, sous l'hypothèse d'invariance du processus de changements de caractéristiques, concerne la convergence des suites de processus de pyramide des âges vers la solution déterministe de l'équation de transport-renouvellement classique, qui évolue suivant des taux de naissance et mort agrégés obtenus par la moyenne des taux

microscopiques sur la structure de caractéristiques stable. Nous avons illustré numériquement l'impact de la fréquence des changements de caractéristique sur les taux de naissance et mort agrégés, montrant en particulier qu'à taux démographiques microscopiques fixés, et règle de swap inchangée, la fréquence de ces changements caractérise de manière fondamentale l'hétérogénéité de la population et ainsi les niveaux des taux démographiques agrégés observés à l'échelle macroscopique. Nous proposons ci-dessous des perspectives de recherche sur ces travaux.

(i) Ce que le modèle équivalent nous indique en particulier, c'est qu'il est possible sous certaines hypothèses de regrouper les individus, non pas parce qu'ils partagent une même caractéristique, mais parce qu'ils en changent de la même manière. Ceci laisse penser que des applications sont possibles tant en démographie qu'en assurance, lorsqu'il s'agit de réduire l'information d'un groupe d'individus à un représentant, aussi appelé *model point*. Dans le cadre de grandes données individuelles, lorsque les caractéristiques individuelles changent trop fréquemment, l'étude de ces changements et du modèle agrégé correspondant pourrait être une direction d'application prometteuse.

(ii) Dans le champ des théorèmes limites, il y a un véritable intérêt à explorer le théorème central limite associé, et en particulier à faire le lien avec la vitesse à laquelle le phénomène d'invariance atteint son équilibre. Ceci pourrait s'avérer également utile d'un point de vue numérique, pour évaluer la pertinence d'utiliser plutôt le modèle équivalent que de simuler la population au niveau microscopique avec les changements fréquents.

## Chapter 2

# Bridging the gap from microsimulation practice to population models: a survey<sup>†</sup>

### Abstract

This paper aims at bridging the gap between microsimulation practice widely spread in demographics and social sciences, and population theory. This takes the form of an interdisciplinary survey which starts by highlighting the complexity of the *demographic transition* at several levels, in particular due to population heterogeneity and interplay between mortality and fertility. We propose an additional review of microsimulation methods widely spread in social sciences, and in particular dedicated to demographic applications. This shows what are the main components of a *microsimulation exercise* to study heterogeneous populations with possible individual interactions, in the form of a complex algorithmic approach based on several kinds of data. This leads us to argue that a proper mathematical framework is needed in order to unify the practical uses as well as to provide a reference framework to think jointly the several demographic effects. We introduce such framework as a microscopic stochastic birth-death-swap population process modeling the evolution of an heterogeneous population at the scale of the individual. On this basis, we show that it is crucial to take into account both population heterogeneity as well as the uncertainty on external environment in order to properly understand demographic patterns. We also discuss the macroscopic behavior of linear and non-linear dynamics, including the method of moments and large population asymptotics.

**Keywords:** Microsimulation, Demography, Age pyramid, Heterogeneity, Poisson point measure, Point processes, Birth-death processes, Thinning.

---

<sup>†</sup>Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Bridging the gap from microsimulation practice to population models: a survey. *Work in progress*

## 2.1 Introduction

In the past decades, human societies experienced major demographic changes, often referred to as *demographic transition*: an increase in life expectancy combined with a decline in fertility, as well as a general economic growth and health improvements at the individual level. The complexity of demographic phenomena appears at several levels since mainly (i) there is a structural interplay between mortality and fertility, possibly with delay, (ii) the age structure of mortality presents strong age regularities but its time pattern is complex and the existence of a limiting age is still a debate, and (iii) at the microscopic level, hidden underlying population heterogeneity creates non-trivial dynamics and unexpected fluctuations at the macroscopic scale.

In social sciences, microsimulation methods have been developed, which aims at simulating a collection of (possibly interacting) individuals over time, and observing the impact of microscopic rules at the macroscopic level. Computational micro-macro approaches are widely used in demographics (see e.g. Tesfatsion (2002), Morand et al. (2010) and Silverman et al. (2011)) and such microsimulation techniques go back to the work of Orcutt (1957). Since then, as mentioned in Morand et al. (2010), most statistical or demographical government bodies in developed countries use microsimulation methods. For instance, the MicMac project funded by the European Commission and implemented by a consortium of research centers in Europe develops the general micro-macro model of the European Union that is documented in the survey of Willekens (2005). In France in particular, the national institute INSEE is currently using the microsimulation model *Destinie 2* (see Blanchet et al. (2009)), which is used to measure the efficiency of reforms on the national pay-as-you-go pension system.

Such computational approaches allow to capture many effects and get insights on demographic patterns. However, due to their implementation and description diversity, as well as the several data sources, such approach are constrained by the specific implementation methods and underlying data. Therefore, a proper mathematical framework is needed in order to unify the practical uses as well as to provide a reference framework to think jointly the several demographic effects. The micro-macro mathematical setting is inspired by Fournier and Méléard (2004), Champagnat et al. (2006), and Ferriere and Tran (2009) in mathematical ecology, and here developed for demographic purposes. In this context, we propose further insights on micro-macro population modeling in a demographic context. In particular, we show that it is crucial to take into account both the environment and the heterogeneity in the whole population dynamics to properly understand demographic patterns. This modeling framework is also seen as a complementary approach to the traditional models in demography and actuarial science based on aggregate mortality indicators, such as proposed by Lee and Carter (1992), Cairns et al. (2006) and Cairns

et al. (2009).

The model we introduce focuses on heterogenous population dynamics, providing the evolution over time of a given population and its specific composition. This evolution in time is made at the scale of the individual. In the population, each individual is characterized by its age and also personal features (e.g. geographic location, family structure, gender, level of income, socio-professional category, diploma level or health status). The population evolves in a stochastic environment, seen as the dynamic uncertainty on the demographic rates. In the population, each individual can die, change its characteristics (event called *swap*) or give birth. Such a microscopic model provides a useful mathematical representation as well as a flexible simulation method for the evolution of the population in a continuous-time setting. Considering Poisson measures as starting models, we enrich it and detail the construction of the so-called birth-death-swap process as the solution to a stochastic differential equation driven by Poisson measures using the *thinning* representation. The crucial property of such microscopic model concerns its micro-macro consistency, which means that at the macroscopic scale, the population behavior is described by classical deterministic equations well known in demography. For standard linear models, the mean behavior is well described by such deterministic equations, but microscopic models involving interactions need to be studied under large population asymptotics to recover the corresponding non-linear deterministic dynamics. In this paper, particular attention is given to these two points of view on the meaning of *macroscopic* as well as the mathematical tools that are needed.

The remainder of this paper is organized as follows. In Section 2.2, we confront several points of view on the demographic transition, and we describe the microsimulation approach used in demography and other applications in social sciences. In Section 2.3, we develop the mathematical formalism modelling heterogenous populations with age, and we introduce the birth-death-swap process. Section 2.4 details the numerical microsimulation algorithm as well as an important example regarding the "cohort effect", and a discussion on the impact of the random environment. Finally, in Section 2.5 we discuss the several points of view about macroscopic approximations, both for linear and non-linear models.

## 2.2 Complexity and heterogeneity in demographic transition analysis

Since the nineteenth century, human society has experienced remarkable changes, often referenced to as demographic transition, that can be subdivided into three most directions (see Canning (2011)):

- (i) *Mortality transition*: a substantial reduction in mortality rates (that caused life



expectancy at birth to grow by nearly 30 years in the last century) followed by a reduction in fertility leading to "approximately stable" population sizes.

(ii) *Economic and social transition*: an economic growth, with the emergence of sustained increases in income per head, together with an increase in social and political equality, particularly between men and women and the adoption of democracy in many countries.

(iii) *Health transition*: a remarkable feature of the modern era is the increase in the "quality" of people, taken in a broader sense than the human capital in economics, often synonymous of education. This refers in particular to a substantial increase in the physical and cognitive development of children, not associated with education features, as well as health improvements at several ages, in particular for old age classes.

Nevertheless, there exists an observed heterogeneity in this dynamics: reduction of mortality does not concern all ages in the same manner, the major improvements concerning younger ages. One also observes different mortality levels in several countries and by gender, as well as across groups in a given country.

### 2.2.1 Determinants of mortality improvement

Therefore, it is useful to improve the analysis of these evolutions, with different perspectives including demographics, economics, biology of aging, and this from different review articles as for example the survey of Cutler et al. (2006).

#### Surveys

**On the determinants of the mortality transition** The remarkable study on the determinants of mortality by Cutler et al. (2006) starts by laying out the historical decline in mortality, then moves to mortality differences between rich and poor countries, and finally discusses differences in mortality within countries. Let us develop only the example of *infectious diseases* and the role of the *economic growth* to illustrate the complexity of the analysis, leaving the reader to refer the full version of the paper by Cutler et al. (2006).

(i) The decline in infectious disease explained 60 percent of deaths in England and Wales in 1848, when in 1971 the infectious disease mortality declined by 95 percent, more specially for the most vulnerable population, infants and children. The sources of these reductions have not found some consensus in the demographic community. The debate started with the work of McKeown (1976) in favor of the impact of *improved nutrition*, followed by the work of Fogel (1997) on the body mass, both thinking that nutritional improvement explained a large part (more than 50 percent) of the improvement in mortality in the period 1850-1950. The magnitude of

the effects did not convince everyone. As a result, some argued that it was disease, not nutrition, that was the primary exogenous influence, and that disease burdens changed most strongly as a result of *public health*.

(ii) *Macro public health* involves big public works projects: filtering and chlorinating water supplies, building sanitation systems, draining swamps, pasteurizing milk and undertaking mass vaccination campaigns. For instance, the water and food-borne diseases after that time – typhoid, cholera, dysentery, and non-respiratory tuberculosis, are now eradicated in rich countries. Cutler and Miller (2005) estimate that the water purification explains half of the mortality reduction in the US in the first third of the twentieth century. Vaccination campaigns were starting at the early 20<sup>th</sup> century, but except for tuberculosis, they contribute only for three percents of the total mortality reduction.

(iii) Very detailed other questions are considered in the survey. Progress in medicine, progress in public and private health, economic growth, contribute to the decline to mortality over time, but their impacts are complex to analyze, with difference by age, by type of sub-populations, country, and sometimes delayed in time (timing issues), or with possible reverse effects. The quantitative aspects of these relationships remain imprecise and mostly unknown, even if education (in particular women's education) seems to be the most likely important common factor in this heterogeneous evolution. A large part of the study is dedicated to the links between health and wealth, with a discussion on the extent of the impact of individual income. This last point supports some complementary explanations.

**Economic growth and mortality** There is also a strong relationship between income and life expectancy across countries over time and at a point in time.

(i) In 1975, Preston examined the relationship between life expectancy and income in three different decades: the 1900s, 1930s and 1960s. In each decade the association between the two measures held true; The Preston Curves (see Preston (1975)) show that as per capita income increases life expectancy increases as well, although the increase is much larger for initially poor countries (see Figure 2.1). The estimates imply that increases in income are responsible for about 20% of the increases in life expectancy (Bloom and Canning 2007). The relationship is non-linear, and the curvature (concave) of the relationship between income and health suggests that a policy of redistributing income from the rich to the poor will improve average health outcomes; but the policy depends on the causality of the relation. Other recent studies show that link between income and life-expectancy is more and more verified except in Sub-Saharan Africa with the AIDS epidemic in recent years), which is an argument in favor of a relation of causality. The link between income and health holds for individuals as well as at the level of countries.

Nevertheless, it seems that at the individual level this is the *relative* and not the

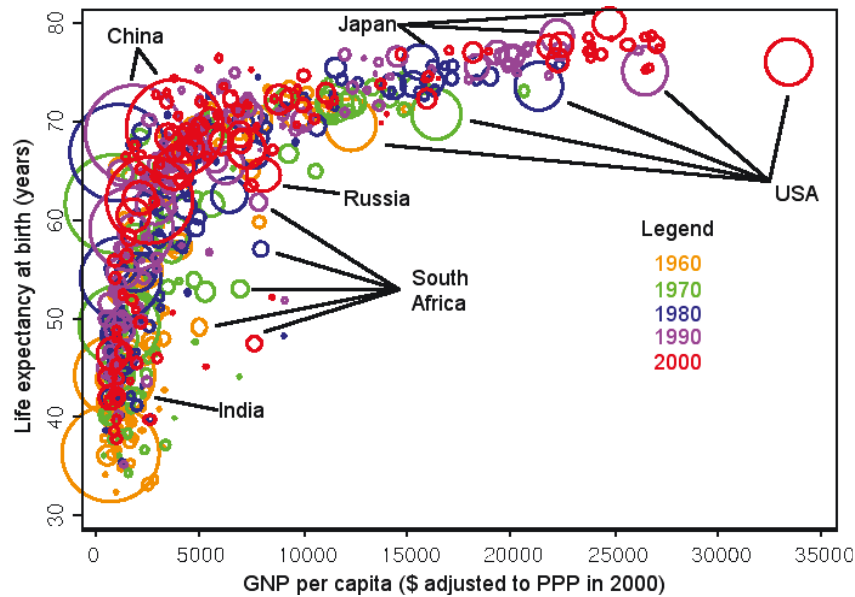


Figure 2.1: Preston curves superimposed from 1960 to 2000. Source: <http://www.ganfyd.org>

absolute income that matters. In addition, there is new evidence of intergenerational transmission of health. So improvements for a cohort can benefit future cohorts with potentially long delays.

(ii) Let us remark that the role of context and environment to explain part of health modification can be explored by so-called multilevel statistical methods (see e.g. Bryk and Raudenbush (1992) and Goldstein (2011)). The core of multilevel statistical methods is to account for the natural hierarchies that are present in observed data, by grouping individuals by *clusters*. It therefore leads to correct the analysis of the several group effects. Such multilevel analysis has indeed been successfully applied to public health research (see e.g. Subramanian et al. (2003)) as well as to a great variety of other topics.

**Debate on the limiting age** There are numerous biological theories of aging in the field of the so-called bio-gerontology, and no consensus arises about the existence of a fixed human life span, that is the *maximum biological life length driven by the intrinsic rate of bodily deterioration*. However, several studies based on real data reported the fact that when comparing a given population at successive time periods, or several subgroups at the same time, a phenomenon called *mortality compensation law\** appears (see e.g. Strehler and Mildvan (1960), Fries (1980), Gavrilov and Gavrilova (1991), Wilmoth and Horiuchi (1999)). It states that a decrease of

\*Note that the compensation law can be also measured in two other ways, with their corresponding denominations: the compression of the distribution of ages at death, and the rectangularization of the survival curve.

mortality rates at the lowest ages in the age range leads to an increase of mortality at highest ages; in other words, when comparing two populations, there is an age after which the ordering between mortality forces is reversed. This can be easily addressed when considering the famous Gompertz model (see Gompertz (1825)) assuming the exponential increase of the mortality force at reasonable ages: indeed, for stochastic mortality models with a formulation closely related to Gompertz, such as Cairns et al. (2006), the times series of slope and intercept of log-mortality rates appear to be negatively correlated, as already highlighted e.g. in Strehler and Mildvan (1960). This effect can be explained by the existence of a limited human life span varying *slowly*, or being stable in reasonable time periods (see e.g. Strulik and Vollmer (2013)). These insights suggest that an heterogenous collection of individuals may be affected by external factors that are common to all of them; in particular, it seems crucial to include a dynamic and stochastic *human life span* as part of a time-varying and stochastic environment.

**Demographic data and forecasting** There is increasing concern about poor accuracy of demographic population projections at national level as backtesting studies show large differences between projections and realized population (see e.g. Mayhew and Smith (2013)). At the core of forecasting issues are the problems of unreliable data as inputs of population models. In this context, the way mortality rates are computed at the national level is of particular importance, as they are the basis of the wide variety of *stochastic mortality models* (see e.g. Plat (2009)) which aim at producing forecasts of mortality levels. The most famous mortality data provider is the *Human Mortality Database* (see HMD (2014)) which regroups national data by gender and age classes for developed countries. To get some insights on such data, we propose to depict in Figure 2.2 what are called (estimated) mortality rates, given by years and age classes; both time windows and age classes may vary from one to five years, and here depicted for one year age classes and fixed observation year 2008. Several issues are related to the computation of such mortality data; in particular, it is suggested in several studies that these data can suffer from biases due to particular birth patterns, see e.g. Richards (2008). This idea has been worked by Cairns et al. (2014) who developed some statistical approach which aims at detecting and correcting such data by using fertility rates at a refined time scale, e.g. recorder monthly or quarterly. In this direction, we argue that understanding the underlying age pyramid dynamics is crucial to analyze past data; in particular that the use of a proper population dynamics model is promising in this context. From a more general perspective, there is a need to escape from the "tyranny of data", as claimed by Silverman et al. (2011), who also argue that there is a need for theoretical investigation of demographical dynamics.

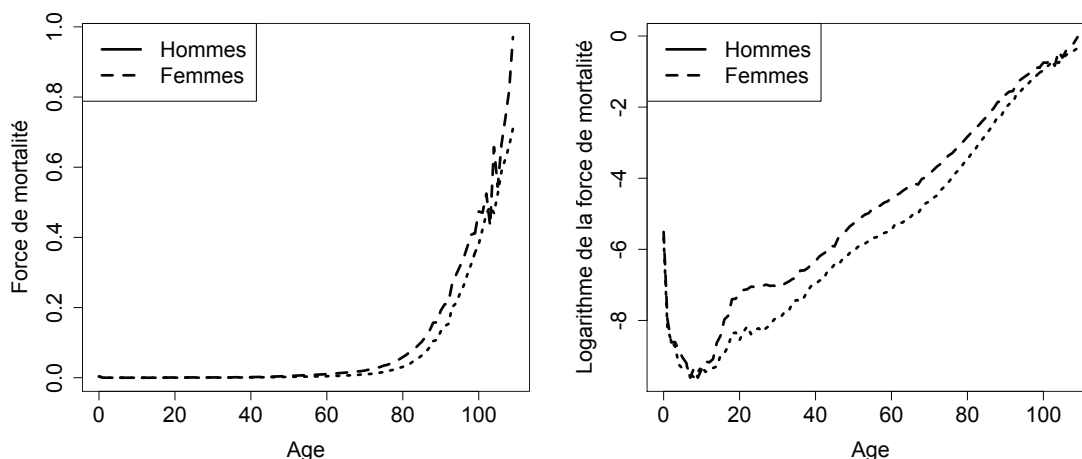


Figure 2.2: Mortality rates by gender and age (left) and their logarithm (right) for France in year 2008

**Interplay between mortality and fertility: summary** Several previous considerations highlighted the fact that demographic dynamics is the result of a complex interplay between mortality and fertility at several levels. We aim at summarizing here these three important levels.

(i) The demographic transition has seen important improvements in longevity soon followed by a reduction in fertility, leading to controlled population growth (see Canning (2011)). These indicators seem to be on the whole negatively correlated.

(ii) Aggregate national mortality indicators are computed based on the age pyramid, using several approximations. As we detailed previously, the quality of these approximations directly relies on mortality and fertility patterns, and it is possible to detect and correct to some extent such kind of data by using birth records at a refined time scale (see Richards (2008), Cairns et al. (2014)).

(iii) Aggregate mortality is the result of the dynamics of population *heterogeneity*, of which death and birth rates for specific characteristics drive the evolution over time. Therefore, a detailed knowledge of such dynamics helps to understand how standard indicators evolve over time. As an example, abrupt changes in birth patterns which may be incurred to specific subgroups seem to be the cause of the famous "cohort effect" for the Golden Cohort in UK (see Willets (2004)); this example is discussed in more details in Subsection 2.4.2.

Let us conclude this section with Cohen (2003) for a more global perspective on population dynamics.

*So, it is a convenient but potentially dangerous fiction to treat mondial population projections as exogenous inputs to economic, environmental, cultural, and politi-*

*cal scenarios, as if population processes were autonomous. Belief in this fiction is encouraged by conventional population projections... In fact, the absence from population projection algorithms of influential external variables indicates scientific ignorance of how external variables influence demographic rates rather than any lack of influence.*

### 2.2.2 Micro-macro simulation in population dynamics

Cohen's perspective on modeling should be nuanced by the dramatic development of microsimulation (dynamic) models in demography and in social sciences. From the previous considerations, it clearly appears that the complexity in the demographic transition cannot be apprehended directly at the aggregated level.

#### Micro-simulation issues

**Microsimulation in social science** So, there is an intrinsic interest in describing the variability and heterogeneity in the population from the individual level, in view to obtain *macro*-outcomes by aggregation.

(i) As any model (in social science), the process is both a modeling exercise, and an exercise to run the model and experiment with it. As suggested by Spielauer (2011), in addition to helping testing theory or picturing the future, the exercise may be used for training (as a simulator) by policy makers, or citizens on the impacts of policy effects. The results of such microsimulation are population *projections* rather than *forecasts*, since they do not tell us how the future will be, but rather what would be the scenario like if the assumptions are correct. In this sense, making projections is more a prospective exercise to support public policies and nourish public debate than to predict the future.

(ii) Dynamic microsimulation in social sciences goes back to the work of Orcutt (1957), who developed the so-called data-driven dynamic microsimulation model, mainly motivated by projection and policy recommendation. The model is parametrized with micro-econometrics and statistical methods. Orcutt stressed the interest to In contrast are the Agent Based Models (ABM) whose main purpose is to explain global economic regularities by dynamical systems of interacting agents (see e.g. Tesfatsion (2002) for an overview of ABM), in a so-called bottom-up approach; they also differ in the modeling of the micro behavior with interacting rules based on theory. In demography, these ideas are used by Diaz et al. (2007) to solve the difficult problem of age-specific projection of fertility rates.

**Implemented systems** Orcutt soon afterwards was in charge of Dynasim (1976), the first large-scale US microsimulation model of the socioeconomic status and behavior of individuals and families. Since then, as mentioned in Morand et al. (2010),

most statistical or demographical government bodies in developed countries use their own microsimulation model, for instance DYNASIM in the US, DYNACAN in Canada, SESIM in Sweden, or DYNAMOD in Australia, as well as the MicMac project funded by the European Commission, and implemented by a consortium of research centers in Europe, documented in the survey of Willekens (2005). In France in particular, the national institute INSEE, as part of the MicMac project funded by the European Commission, developed different versions of a microsimulation model whose current version is DESTINIE2 (see Blanchet et al. (2009)). It is used e.g. to measure the efficiency of reforms on state pension systems, and is based on a representative sample of the national population.

### A dynamic microsimulation exercise

Any modeling exercise starts with a clear specification of the objectives of the study: for instance the specific objective of MicMac is to provide demographic projections by detailed population categories that are required in the development of sustainable (elderly) health care and pension systems in European Union. Methods, assumptions and scenarios are chosen with a view to achieving the objective.

(i) *States Variables* In the center of a microsimulation method is a population data-base, storing the informations on all members (individuals) of the heterogeneous population; this information at the individual level is at the core of microsimulation, and is a main difference with so-called *macrosimulation*. The first step is the definition of the state space, that is the specification of the variables concerning individuals of interest for the study. At the individual level, the state space consists of a number of characteristics or attributes, each of which can take a certain value; at the population level, the state space consists of all possible combinations of individual attribute values.

(ii) *Dynamic simulation* includes times and so demands to *predict the future state* of the population, by making the distinction between the events influencing the population itself and those affected by it (population aging, concentration of wealth, sustainability of social reforms...).

(a) In the microsimulation approach, we have to model individual changes over the time during the life course, producing a simulated database with longitudinal information, e.g. in the form of individual virtual biographies, viewed as a sequence of state variables. The aggregation of the individual bibliographies yields a bottom-up estimate of the cohort biography. Nevertheless, in presence of interactions, all the biographies have to be simulated simultaneously, which is challenging and computationally demanding for large populations.

(b) Another representation is given by cross-sectional counts of groups by characteristics in very flexible ways, as in the two-sex age pyramid representation of the

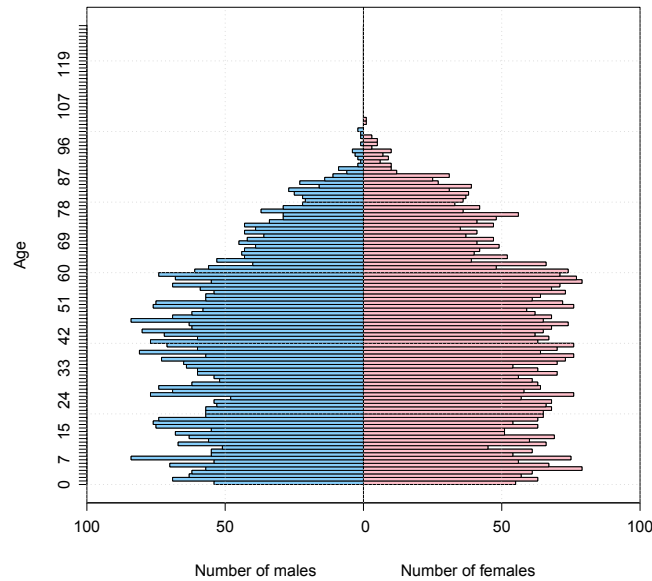


Figure 2.3: Age pyramid generated on the basis on French data for year 2008

census data (see Figure 2.3 for an illustration). Even if this representation is very close to the one of aggregated population level, it is not the case for their dynamics. (c) The common approach to address the microscopic evolution is based on individual hazard functions, or individual transition rates, estimated conditionally on support variables, and other risk factors that affect the rate of occurrence of some event.

If the state space is large, there should be a fair degree of independence between the variables involved; if not, either the estimation problem would be very complicated, which makes it tempting to disregard interaction effects in the estimation phase, a practice which may lead to quite misleading results or microsimulation ignoring the dependence would give a false sense of security. Following Van Imhoff and Post (1998), microsimulation is particularly appropriate if the results of the process under study are complex but the driving forces of the process are simple. Indeed, if more explanatory variables are included, then the outcomes are subject to a higher degree of randomness: first since the size of the estimation process is higher, leading to a higher degree of uncertainty on the parameters, and second by the need for more stochastic simulations in the biographies which contain more possible variables.

(iii) *Internal consistency, Interaction*

In many cases, behaviors are more stable at the micro-level at which there is no interference from composition effects. Even complete stability at the micro level does not automatically correspond to stability at more macro level. For instance, according to Spielauer (2011), an educational expansion at the macro level can lie entirely in the changing composition of the parents generations, and not necessarily in a change of micro behavior.



Microsimulation methods are well suited to make links between individuals, in other words to include *interactions*, and this aspect is one of their appealing features.

**Sources of randomness** Let us describe the several sources of uncertainty arising in microsimulation models, see Van Imhoff and Post (1998).

(i) The first one is the *inherent randomness* to Monte Carlo methodology, also called *demographic stochasticity* or *idiosyncratic risk* in other areas as actuarial science, than can be reduced by taking average over the a large number of model runs, or by increasing the size of the database (the number of individual in the sample); the equivalence between these two points of view is not as obvious as it looks, and will be discussed in the theoretical Section 2.5. Moreover, various techniques to accelerate Monte Carlo simulation coupled with variance reduction have been developed for many areas, in particular dedicated to so-called spatial point processes. These numerical techniques are beyond the scope of the present paper.

(ii) The second source of randomness is the so-called *starting-population randomness*. The starting population, which is the initial database for the microsimulation model, is a subsample of the total population and is thus subject to random variation. Any deviation of the sample distribution from the reference population distribution may lead to several biases in the outputs. This source of randomness can be kept under control by using sizeable database and are probably of minor importance when compared to the following source of randomness.

(iii) The third source of uncertainty is the *specification randomness*. We have seen that the more explanatory variables are included in a micro simulation model, the larger is the degree of randomness to which the model outcomes are subject. This source of uncertainty is also referred to as *systematic risk* or *estimation error* in other areas. To reduce the *specification randomness*, a recent approach is the sorting method, selecting the simulated trajectories in such a way that the micro model respects some macro properties, including the property of producing the expected values. For example, for DESTINIE 2 (see Blanchet et al. (2009)) this alignment is ensured by making the individual probabilities of event consistent with obtained annual number of births or deaths by gender, as well as migration flows.

**Multilevel analysis and Environment randomness** From the analysis of the determinants of mortality, we have seen that there is an intrinsic interest in describing the variability and heterogeneity in the population from the individual level, but also from the level contexts, as neighborhoods, regions or states, or countries.

(i) This distinction is relevant when the aggregate effect differs from the sum of the individual effects from the group (the whole may be more than the sum of its parts). The Glasgow effect is a typical example of this type of analysis: the mortality rate in Glasgow significantly differs from those of other cities with the same "deprivation"

level as Liverpool or Edinburgh (see e.g. Gray and Leyland (2009) and Donnelly (2010)).

(ii) *Neighborhoods, clusters or groups level.* The multilevel framework with its simultaneous examination of the characteristics of the individuals at one level (standard) and the context in which they are "grouped" at another level, offers a comprehensive framework for understanding the ways in which groups can affect people (contextual, or *design effect*) or, alternatively, people can affect the cluster *composition effect*. Much documentation exists on measuring design effect and correcting for it. The group designs (from individual to the group) is in general considered as a nuisance, in traditional analysis, but not in multilevel model. Ignoring this multilevel structure of variations does not simply risk overlooking the importance of neighborhood effects; it has implications for statistical validity.

Clusters are treated as levels because they are a family from which we have observed a random sample, and their evolution may be integrated as a part of the random environment of the population.

## 2.3 Point and population processes

### 2.3.1 Static population as a random point measure

In this part, we introduce the abstract representation of what we call "population". We begin by a reference example that is using Poisson point measures.

**Classical example: Poisson point measure** A *marked population*  $\xi$  is a finite random set whose members (elements, points, individuals, particles) are characterized by quantitative attributes (location, characteristics, traits, indicators) taking values in a multi-dimensional space  $E$ .

The uncertainty concerns together the number of individuals in  $\xi$  ( $\text{card}(\xi)$ ) and the "vector" of attributes ("location") in the space  $E$ . Formally, the marked population is described by a random set  $\xi = \{(X_n)_{n=1}^\nu\}$ , where  $\nu = \text{card}(\xi)$ ,  $\text{card}(\emptyset) = 0$ , and  $X_n$  is the random "location" (characteristics) in  $E$  of the individual  $n$ . By assumption, the population may be viewed as a "cloud", where changing the numeration of individuals has no effect.

(i) *Poisson population:* A common way to generate such population is to sample a probability distribution  $\mu(dx)$  on  $E$  with an iid sequence  $(X_n)_{n \geq 1}$ , and to stop randomly after  $\nu$  drawings, where  $\nu$  is an independent Poisson variable with mean  $\lambda$  ( $\mathcal{P}(\lambda)$ ). The marked population  $\xi = \{(X_n)_{n=1}^\nu\}$  thus constructed is called *Poisson population* with mean measure  $\eta(dx) = \lambda\mu(dx)$ .

a) The "spatial" distribution is given by counting the individuals whose characteristics lie in a subset  $B \subset E$ . The new population, restricted to leave in  $B$ ,

$\xi^B = \cup_{n=1}^{\nu} \{X_n\} \cap B$ , is still a Poisson population with distribution measure  $\mu^B(dx) = \mu(dx|B) = \mathbf{1}_B(x)\mu(dx)/\mu(B)$ , and Poisson parameter  $\lambda^B = \lambda\mu(B)$ . The size of the new population  $\nu^B$  is exactly  $N(B) = \sum_{n=1}^{\nu} \mathbf{1}_B(X_n)$ . Note that  $\lambda^B\mu^B(dx) = \lambda\mu(dx) = \eta(dx)$  and  $\mathbb{E}(N(B)) = \eta(B)$ .

b) The Poisson population satisfies all the standard properties of what is called a *finite Poisson point measure* with finite mean measure  $\eta$ :

– for any set  $B$ ,  $N(B)$  is distributed as a Poisson random variable with intensity  $\eta(B)$ , so that  $\mathbb{P}(N(B) = k) = e^{-\eta(B)} \frac{(\eta(B))^k}{k!}$ ,

– for any finite family of disjoint sets  $(B_j)_{j=1}^k$  the (Poisson) random variables  $(N(B_j))_{j=1}^k$  are independent.

It is well-known that any finite Poisson point measure with these two properties is a Poisson population in the previous sense.

(iii)  *$\sigma$ -finite mean measure* We still reserve the name of Poisson point measure to *countable* point measures with  $\sigma$ -finite mean measure, whose restriction to the *admissible sets*  $B$  (a Borelian subset with finite mass  $\eta(B)$ ) is a Poisson process with finite mean measure  $\eta^B$ . The example of temporal Poisson process whose mean measure is the Lebesgue measure on  $\mathbb{R}^+$  is detailed in the next subsection.

If  $\eta$  is only  $\sigma$ -finite, we have to generate more iid random variables, based on a two step process:

first select an *increasing* family  $(W_k)$  of Borel windows ( for example increasing balls on  $\mathbb{R}^d$ ) with limit  $E$ , with  $\eta(W_k) < \infty$ , where the essential information is given by  $W_1 = E_1$ ; then we define a partition  $(E_k = W_k/W_{k-1}), k \leq 2$  with  $\eta(E_k) \rightarrow 0, k \rightarrow \infty$ . The Point Poisson processes with data  $(E_k, \eta^k)$  are generated by *independent* families  $\chi^k = \{(X_n^k)_{n=1}^{\nu^k}\}$  with independent random measures  $N^k(B \cap E_k)$ . The "countable" reunion of these populations  $\chi = \bigcup_k \{(X_n^k)_{n=1}^{\nu^k}\}$  generates a countable measure which is on the admissible set  $B$  a finite counting measure  $N(\cdot|B) = \sum_{k=1}^{\infty} N^k(\cdot|B)$ , and then a *finite Poisson point process*. Note that in practice, it is not possible to simulate the whole Poisson point measure. This can only be achieved on admissible sets, and even in this case this can be very expensive in terms of computational time.

**Abstract representation of marked population** We first present the basic concepts of marked population (spatial point measure) and its distribution on the space of finite populations as a theoretical tool in the analysis of starting population randomness, with only technical details useful in the sequel. Their mathematical description is given both as random set or in terms random counting measure on a complete probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , first without temporal evolution. (More complementary details on the mathematical formalism may be found in the book of Çınlar (2011), Chap VI, or in the lectures of Baddeley (2007).)

(i) A *marked population* is the probabilistic representation of *countable* random population (set, pattern), whose members (individuals, elements, points) are char-

acterized by a finite number of quantitative attributes (characteristics, features, location) taking values in a multi-dimensional space  $E$ , (often a compact subset of  $\mathbb{R}^d$ , or more generally a complete, separable metric space). A typical example is the *starting population* in a microsimulation process.

The space of outcomes, (the data-base storing the information on the population) is the so-called population (configuration) space  $\Gamma(E)$  whose elements are the populations  $\xi$  whose only a finite number of individuals have their characteristics in a compact  $K$  in  $E$ , ( $\text{card}(\xi \cap K) < \infty$ ).

(ii) To study the "spatial" distribution of the individuals, we define two useful family of random variables, indexed by the subsets  $B$  as

–  $N(B)(\xi)$  the number of individuals of the population  $\xi$  in  $B$ , that is  $N(B)(\xi) = \text{card}(\xi \cap B)$ ,

–  $V(B)(\xi)$  the *vacancy indicator*, equal to 1 if there is no individual of  $\xi$  in  $B$  and 0 if not. More formally,  $V(B)$  is the indicator function of the vacancy set  $V(B) = \{\xi : N(B)(\xi) = 0\} = \{\xi : \text{no points in } B\}$ .

a) The first one is well adapted for exploring additive properties of marked population. For example, the reunion of two marked populations without common individual, and distinct colored symbols, (say red ( $R$ ) or yellow ( $Y$ )) is still a marked population called *surimposed* population and  $N_{R \cup Y}(B) = N_R(B) + N_Y(B)$ .

It gives the *cross-sectional representation* of the population by counts of the individuals by "groups" of attributes. The first moment, called *mean measure*,  $\eta(B) = \mathbb{E}(N(B))$  (which is an additive operator assumed to be a positive measure on the Borel  $\sigma$ -field  $\mathcal{B}(E)$ ) is often considered by practitioners as the main important information on the population.

b) The second one is well-adapted to multiplicative properties, since the vacation indicator of a surimposed population is  $V_{R \cup Y}(B) = V_R(B) \cdot V_Y(B)$ , since  $V_{R \cup Y}(B) = \{N_R(B) + N_Y(B) = 0\} = \{N_R(B) = 0\} \cap \{N_Y(B) = 0\}$ . The expectation of the vacation indicator induces a "capacity" operator  $T(B)$  on  $\mathcal{B}(E)$  by  $1 - T(B) = \bar{T}(B) = \mathbb{E}(V(B)) = \mathbb{P}(N(B) = 0)$ . If the set  $B$  is identified with a "trap",  $\bar{T}(B)$  is the survival operator.

**Comment:** *The example of Poisson population shows that the mean measure  $\eta(B) = \mathbb{E}(N(B))$  is a very poor information in comparison to the vacancy indicator  $\bar{T}(B) = \mathbb{P}(N(B) = 0) = e^{-\eta(B)}$ , since it is possible to deduce from the conditions  $\bar{T}(B) = e^{-\eta(B)}$  that a marked population is a Poisson population under the assumption that  $\eta$  is a  $\sigma$ -finite and diffuse measure.*

**Distributional Point of view** A Poisson point measure with finite and diffuse mean measure  $\eta$  can be described as a finite counting measure, based on Dirac

measures  $\delta_z(dx)$  as

$$\xi(dx) = \sum_{i=1}^{\nu} \delta_{X_n}(dx).$$

When the measure  $\eta$  is diffuse, it is easy to describe the distribution of this random measure on the space of the configurations  $\Gamma(E)$  identified to that of Radon counting measures (finite on the compact sets of  $E$ ), see e.g. Jensen and Møller (1991). This amounts to identify  $N$  with the *unordered (symmetric)* set  $\{X_1, \dots, X_\nu\}$ , and to express the probability by conditioning on the size of the measure, that is for any  $A \subset \Gamma(E)$ ,

$$\mathbf{P}^\eta(A) = e^{-\eta(E)} \left[ \mathbf{1}_A(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{E^k} \mathbf{1}_A(\{x_1, \dots, x_k\}) \eta^{\otimes k}(dx_1, \dots, dx_k) \right].$$

In the following, we denote indifferently  $\xi$  the configuration or the measure, so that it is possible to write

$$\xi(dx) = \sum_{x' \in \xi} \delta_{x'}(dx).$$

**Integral notation** As several integrals are involved in the following, we introduce the "physicist" notation

$$\langle m, f \rangle = \int_{x \in E} f(x) m(dx)$$

for any measure  $m$  and function  $f$  integrable w.r.t.  $m$ . When several parameters appear, we put as an index the parameters which are not involved in the integration. More precisely, consider a measure  $m(dx, dy)$  on a space  $E \times F$  and a function  $f(x, y, z)$  integrable w.r.t  $m$  ; then

(i) we denote

$$z \mapsto \langle m, f_z \rangle = \int_{x \in E} \int_{y \in F} f(x, y, z) m(dx, dy)$$

the function of  $z$  with values in the same space as  $f$ ,

(ii) we denote

$$z \mapsto \langle m, f_{y,z} \rangle(dy) = \int_{x \in E} f(x, y, z) m(dx, dy)$$

the function of  $z$  with values in the space of measures on  $F$  (in general).

### 2.3.2 Dynamic marked population

Given its importance in population dynamics, we superpose a temporal component in the model of population. The most popular example is the dynamic Poisson point process.

### Dynamic Poisson point process

**Temporal Poisson process** The temporal Poisson process (or Poisson process on the line) is a Poisson point process  $N$  on  $\mathbb{R}^+$ , with  $\sigma$ -finite mean measure proportional to the Lebesgue measure  $l(dt) = dt$ ,  $\eta(dt) = \lambda dt$ .

(i) We can restrict the process to leave only on admissible set  $[0, T]$ , with mean measure  $\lambda T$ , so that conditionally to the event  $\{N_T = n\}$ , the  $n$ -event dates  $\theta_i^T$  are independent and uniformly distributed on  $[0, T]$ , and independent of the Poisson variable  $N_T$ . The main drawback of this representation is in its dependence on  $T$ .

(ii) In fact, the classical presentation of the temporal Poisson process is based on the same definition, but with using the order structure of  $\mathbb{R}^+$ . The subject of the study is the increasing process  $t \mapsto N_t = N([0, t])$ , whose paths are non decreasing functions, only increasing by jumps of size 1. The description is made in terms of *waiting times* between two jumps ( $T_1, \tau_1 = T_2 - T_1 \dots \tau_n = T_n - T_{n-1}$ ).  $T_1$  is the first time where the vacancy set  $V_t = \{N_t = 0\}$  is not empty, so  $\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$  and  $T_1$  is an exponential variable with mean  $1/\lambda$ . Its is well-known that this property is extended to the waiting times between two jumps ( $T_1, \tau_1 = T_2 - T_1 \dots \tau_n = T_n - T_{n-1}$ ) that are iid random variables with exponential distribution with parameter  $\lambda$ . This new decomposition is intrinsic, although the mean measure is  $\sigma$ -finite and it is use to write  $N_t = \sum_{n=1}^{\infty} \mathbf{1}_{T_n \leq t}$ , or in a more abstract form  $N(dt) = \sum_{n \geq 1} \delta_{T_n}(dt)$ .

(iii) We can say more on the temporal component  $(T_n)$  considered as a "mark" of the Poisson process  $N_t$ , by studying the product measure on  $\mathbb{R}_+ \times \mathbb{R}_+ = \mathbb{R}_+^2$ , where the second component is associated with  $(T_n)$ . Since the sequence  $(T_n)$  is increasing, we cannot expect to define a Poisson point process in the classical sense, nevertheless useful properties still hold. Let us define for any Borel set  $K$  with finite Lebesgue measure  $\bar{N}([0, t] \times K) = N_t(K)$  which is a temporal Poisson process constrained to live in the interval  $B$  with intensity measure  $\lambda \mathbf{1}_B(t) dt$ .

As a bi-dimensional measure on  $\mathbb{R}_+^2$ ,  $\bar{N}(dt, du) = \sum_{n \geq 1} \delta_{T_n}(dt) \delta_{T_n}(du)$  is a measure supported by the diagonal of  $\mathbb{R}_+^2$ , so that the mean measure of  $\bar{N}$  is proportional to the Lebesgue measure on the diagonal of  $\mathbb{R}_+^2$  often denoted  $\Delta(dt, du) = dt \delta_t(du)$

In what follows, we add the name of "dynamic" or "process" to marked populations with a temporal evolution.

**Poisson point process** A *Poisson point process* models in addition to the temporal events (described by  $N_t$ ) the random "location" of individuals (particles) in the space  $E$ , equipped with a measure  $m$ .

(i) When the total mass  $m(E) = \lambda$  is finite, we extend the static construction based on a countable sampling  $(X_n)$  of  $\mu(dx) = m(dx)/m(E)$  in defining a time evolving population  $(\xi_t)$  only based on the dynamics of the population size  $N_t$ ,

that is  $\xi_t = ((X_n)_{n=1}^{N_t})$ . By extending the static case, the associated counting measure  $N_t(B) = \sum_{T_n \leq t} \mathbf{1}_B(X_n)$  is a temporal Poisson process with temporal intensity  $\lambda\mu(B) = m(B)$ ; moreover, the temporal Poisson processes  $N_t(B)$  and  $N_t(C)$ , where B and C are disjoint subsets have no common jumps and they are therefore independent. Later, we will interpret this construction as a thinning procedure.

(ii) The process  $\xi_t = ((X_n)_{n=1}^{N_t})$  is a minimal representation of the Poisson population, but as for the temporal Poisson process, we need to keep in memory for any individual its date of entry (or "birthday") into the population. So, the useful representation of the population is  $\bar{\xi}_t = \{(T_n, X_n)_{n=1}^{N_t}\}$ . It provides a natural classification of individuals based on their "birthdays", and their "characteristics at birthday". The counting measure associated with this point process is the point measure on  $\mathbb{R}_+^2 \times E$ ,  $\bar{N}(dt, du, dx) = \sum_{n \geq 1} \delta_{T_n}(dt)\delta_{T_n}(du)\delta_{X_n}(dx)$  whose mean measure is not a product measure, but the measure  $\Delta(dt, du) \times m(dx)$  where the measure  $\Delta(dt, du)$  is the Lebesgue measure on the diagonal of  $\mathbb{R}_+^2$ . Nevertheless,  $(\bar{N}_t(K \times B))$  is still a Poisson point process forced to live in K, with mean measure  $\int_0^t \mathbf{1}_K(s) ds m(B)$ .

(iii) When the intensity measure  $m$  is only  $\sigma$ -finite, ( $\sigma$ -finite Poisson process) we can proceed as in the static case. The construction is then based on a double sequence of globally independent variables  $((\tau_n^{(k)}, X_n^{(k)})_{n,k})$  whose distribution depends on  $k$ . Moreover, this decomposition is not unique. Note that in this case we do not have a natural enumeration of the time of events and marks.

**Age pyramid** It is more usual to structure the populations by age rather than birthday, (the famous age pyramid in the presentation of national census data, see Figure 2.3), where the age at time  $t$  of an individual with birthday at time  $u$  is  $a = t - u$ .

(i) With this representation, the attribute "age" implies a deterministic evolution in time of this new characteristic, making it completely different from the other (location).

From the measure theory, the introduction of age corresponds to a change a variable depending on  $t$  on the mark  $T_n = u$ , as  $a \mapsto u = r_t(a) = (t - a)^+$ , so  $a \in A$  is equivalent to  $u \in r_t(A)$ . The new population at the date  $t$  is  $Z_t = \bigcup_{T_n \leq t} (t - T_n, X_n)$ , with a new counting process,  $\bar{N}_t^a(A \times B) = N_t(r_t(A) \times B)$ .

The dynamics in time of  $Z_t$  is complex, since the time  $t$  is present at two places in the definition, as a temporal parameter, and as the parameter of time reversal; then, the process  $N_t(r_t(A) \times B) = \sum_{T_n \leq t} \mathbf{1}_{t-T_n \in A, X_n \in B} = \int_0^t \mathbf{1}_A(t-s)\mathbf{1}_B(x)N(ds, dx)$  is no more a measure with respect to  $t$ . The difficulty comes from the non differentiability in time of the function  $\mathbf{1}_{t-T_n \in A}$ .

(ii) *Deterministic formula* A similar problem arises in many deterministic situations, when we are concerned with the "primitive" function with respect to general Radon

measure  $n(dt)$  on  $\mathbb{R}^+$  of some reversal function  $f$  defined by  $z_t(f) = \int_0^t f(t-s)n(ds)$ . For instance, when  $n(ds)$  is a Dirac mass at  $a_0$ ,  $z_t(f) = f(t-a_0)\mathbf{1}_{0 < a_0 \leq t}$ , and  $z_t(f)$  is not of finite variation if  $f$  is not of finite variation.

(a) To overcome this difficulty, we only consider differentiable functions  $f$ , with "nice integrability" property. The differentiability of  $f$  is used in the following identity  $f(t-s) = f(0) + \int_0^{t-s} f'(u)du = f(0) + \int_s^t f'(v-s)dv$ , where in the last integral we have used the invariance by translation of the Lebesgue measure. The integration by parts formula with respect to  $dn$  and  $dv$  yields

$$\int_{]0,t]} \int_s^t f'(v-s)dv n(ds) = \int_0^t dv \int_{]0,v]} f'(v-s)n(ds) = \int_0^t dv z_v(f').$$

The main formula is then

$$z_t(f) = f(0)z_t(1) + \int_0^t z_v(f')dv. \tag{2.1}$$

This differential rule is more easily understood from the "distribution theory".

From a distribution theory perspective, one writes

$$z_t(f) = \int_0^t f(t-s)n(ds) = \int_0^t \int_0^t n(ds)f(a)\delta_{t-s}(da).$$

We see that the differentiability in  $t$  is based on the derivative in  $t$  of the Dirac mass  $\delta_{t-s}(da)$ , to be understood in the distribution sense, that is here for each differentiable  $f$ ,  $\delta'_t(f) = f'(t)$ .

(b) Let us come back to the population problem with age, where the useful differentiation formula (2.1) becomes, for  $f$  differentiable and  $B \subset E$ ,

$$Z_t(f \times \mathbf{1}_B) = f(0)N_t(B) + \int_0^t Z_v(f' \times \mathbf{1}_B)dv. \tag{2.2}$$

The "purely stochastic part" of this equation is the Poisson process without age  $N_t(B)$ . The right hand side with finite variation is in this form typical of some "transport" phenomenon : it means here that age is transported along the temporal component (with speed 1). Such decomposition is in fact very useful as it can be applied recursively to the processes  $Z_t(f^{(n)} \times \mathbf{1}_B)$  ; when  $f$  is sufficiently regular one recovers a finite system of equations which can be used to compute several quantities on population processes with age. Such approach has been used for the computation of distribution properties of a class of Hawkes processes based on their immigration-birth representation, see Boumezoued (2015b).

(iii) *Time reversal and age process* Let us define a temporal window  $[0, T]$  for the observation of the dynamic Poisson point process  $N$ . The process may be considered as a static Poisson population with state space  $\bar{E} = [0, T] \times E$  and mean measure on  $\bar{E}$ ,  $\bar{m}(dt, dx) = dt \times m(dx)$ . The static population on the enlarged space is  $\bar{Z}^T = \bigcup_{i=1}^{N_T} \{(\theta_i^T, X_i^T)\}$ .



Given that  $N_T = n$ , the arrival times are given by the order statistic of the independent and uniform variables  $(\theta_i^T)$  on  $[0, T]$ , and the age structure is given by the order statistic of the variables  $\bar{\theta}_i^T = T - \theta_i^T$  which are also independent and uniformly distributed on  $[0, T]$ .

So, in a dynamic Poisson population, conditionally on its size  $N_T$ , the sequence of the ages or of the birthdays have the same conditional distribution, given by the order statistic of  $n$  variables uniformly distributed on  $[0, T]$ .

### General point process with stochastic intensity

In the microsimulation exercise, the timing of the events is very complex and cannot be properly described by a Poisson process. In fact, the temporal component in the Poisson point process is conventionally used to structure the available information over the time.

**Martingale point of view** (i) As usual in probability theory, the historical information is described by a increasing family of  $\sigma$ -fields  $(\mathcal{F}_t)$  (satisfying usual assumptions of completion and right continuity), with the property that for any  $t$ ,  $N_t$  is measurable with respect to  $\mathcal{F}_t$ . Then,  $N$  is said to be  $\mathcal{F}_t$  adapted, or compatible.

(a) When  $(\mathcal{F}_t)$  is richer than the minimal one generated by the counting process  $N$ , (a classical situation in microsimulation, where the environmental covariates are at the core of the modeling process) then the assumption of independence of the increments of the Poisson point process must be modified into: "for any  $t, h > 0$ ,  $N_{t+h}(B) - N_t(B)$  is independent of  $\mathcal{F}_t$ ." In particular,  $(N_t(B) - \mu(B)t)$  is a  $\mathcal{F}$ -martingale if  $\mu(B)$  is finite. More interesting is that the converse property is true.

(b) Similar martingale property is associated with the vacancy operator  $V_t(B) = \mathbf{1}_{\{N_t(B)=0\}}$  under the multiplicative form,  $e^{t\mu(B)} V_t(B)$  is a  $\mathcal{F}$ -martingale. A more conventional description is obtained from the entry time  $T^B$  into  $B$  by the population process, (which is the first jump time of the on-line point process  $N_t(B)$ ). The multiplicative martingale property is equivalent to the martingale property of  $e^{t\mu(B)} \mathbf{1}_{\{T^B > t\}}$ . As in the static case, this property (true for any admissible set  $B$ ) characterizes a dynamic Poisson process.

(ii) A convenient way to express the martingale property is to consider a so-called predictable simple process  $f(\omega, t, x)$  defined by

$$f(\omega, t, x) = \sum_{j=0}^K \mathbf{1}_{A_{t_j}}(\omega) \mathbf{1}_{C_j}(x) \mathbf{1}_{]t_j, t_{j+1}]}(t), \quad A_{t_j} \in \mathcal{F}_{t_j},$$

and its integral with respect to  $N(dt, dx)$  and  $\Lambda(dt, dx) = dt \mu(dx)$ ,

$$\langle N, f \rangle = \sum_{j=0}^K \mathbf{1}_{A_{t_j}} (N_{t_{j+1}}(C_j) - N_{t_j}(C_j)), \quad \langle \Lambda, f \rangle = \sum_{j=0}^K \mathbf{1}_{A_{t_j}} \mu(C_j) (t_{j+1} - t_j).$$

The martingale property is equivalent to the static condition

$$\mathbb{E}(\langle N, f \rangle) = \mathbb{E}(\langle \Lambda, f \rangle) \text{ if } \mathbb{E}(\langle N, f \rangle) < \infty.$$

It is useful to observe that the same equality holds true for a general predictable process  $f(\omega, t, x)$  such that  $\mathbb{E}(\langle N, |f| \rangle) < \infty$ , ( $f$  belonging to the  $\sigma$ -field generated by the step processes on  $\Omega \times \mathbb{R}^+ \times E$ , also called product of the predictable  $\sigma$ -field on  $\Omega \times \mathbb{R}^+$  and the Borelian  $\sigma$ -field on  $E$ .)

(iii) *Stochastic intensity and density assumption* In the microsimulation example, the timing of the events is quite complex ; this complexity is expressed at the microlevel by making assumptions on the intensity process.

(a) In general, the state space  $\mathbf{S} = \mathbb{R}^+ \times E$  is equipped with a product measure  $m(dt, dx) = l(dt) \otimes m(dx)$  where  $l(dt)$  is the Lebesgue measure on  $\mathbb{R}^+$  and  $m$  a diffuse  $\sigma$ -finite measure on  $E$  (for instance the Lebesgue measure on  $\mathbb{R}^d$ ).

We introduce the intensity process  $\lambda(t, x)$ , as a non negative adapted process  $\lambda(t, x)$  with enough integrability (see below). The process  $\lambda(t, x)$  may be interpreted as the intensity process of some marked dynamic population  $N$ ; when the characteristics are in  $\mathbb{R}^d$ , equipped with the Lebesgue measure  $m(dx) = dx$ , formally  $\lambda(t, x)dt dx$  is an equivalent of the probability that an event occurs between  $(t, t + dt)$ , with a characteristic in  $(x, x + dx)$ ,

$$\lambda(t, x) = \lim_{(dt, dx) \rightarrow 0} \frac{\mathbb{E}(N([t, t + dt] \times [x, x + dx]) | \mathcal{F}_{t-})}{dt dx}.$$

The notation  $\mathcal{F}_{t-}$  refers to the notion of predictable process, for which the events occurring at time  $t$  are non included. Obviously, the intensity process depends on the reference filtration  $\mathcal{F}$  (and on the probability measure  $\mathbb{P}$ ).

*Integrability condition:* We denote by  $\mathcal{L}^1$  the space  $\mathbb{L}^1(\mathbb{P} \otimes l \otimes m)$  of the predictable processes, integrable with respect to the measure  $\mathbb{P}(d\omega) \times dt \times m(dx)$ .

The integrability condition on  $\lambda$  is the existence of a non negative predictable process  $H(t, x)$  such that  $H(t, x)\lambda(t, x) \in \mathcal{L}^1$ .

(b) The precise formulation is given by a martingale property. To formulate this property in concise form, let us introduce the two family of random measures,  $N_t(ds, dx) = \mathbf{1}_{]0, t]}(s)N(ds, dx)$  and  $\Lambda_t(ds, dx) = \mathbf{1}_{]0, t]}(s)\Lambda(ds, dx)$  where  $\Lambda(ds, dx) = \lambda(s, x)m(ds, dx)$ .

Then the fundamental condition is that  $M_t^f = \langle N_t, f \rangle - \langle \Lambda_t, f \rangle$  is a martingale for all predictable process  $f$  (with  $\lambda f \in \mathcal{L}^1$ ).

Sometimes, the random measure  $M(dt, dx) = N(dt, dx) - \Lambda(dt, dx)$  is called a *martingale measure*.

A thinning procedure of a Poisson process may be used to solve the existence problem of dynamic marked population with given stochastic intensity under the density assumption.

**Thinning procedure** It is easy to generate a large family of marked dynamic populations from a given existing point process  $N$ . The thinning procedure, which

consists in clearing (eliminating) some points (individuals) of the given point process (marked dynamic population) when an event occurs, is an example of such transformation.

(i) *Thinning procedure with digital intensity* Assume the given point process to be a dynamic Poisson process  $N$  with intensity  $m(dt, dx) = dt \times m(dx)$ . Generally, the procedure consists in rejecting all the points outside of a given (predictable) subset  $D$  of  $\mathbb{R}^+ \times E$ , (in  $\mathcal{P} \times \mathcal{B}(E)$ ). From the measure theory, this procedure consists in restricting the given Poisson process to the set  $D$  (still in a counting measure). An equivalent way is to use the binary intensity process  $\mathbf{1}_D(t, x)$  of the set  $D$  as a density for the measures  $N$  and  $\Lambda$ . The thinned measures  $N^D$  and  $\Lambda^D$  are defined by:  $N^D(dt, dx) = \mathbf{1}_D(t, x)N(dt, dx)$ ,  $\Lambda^D(dt, dx) = \mathbf{1}_D(t, x)m(dt, dx)$ ; thanks to the martingale property, since  $D$  is predictable,  $\Lambda^D$  is the global intensity of  $N^D$ . The process  $N^D$  is then a point process with binary intensity process  $\mathbf{1}_D(t, x)$ .

(ii) Assume the intensity process to be a non digital process  $\lambda(t, x)$ . The idea is to introduce a new variable  $\theta \in \mathbb{R}^+$  used as a *thinning parameter* erasing all the points  $(t, x)$  such that  $\lambda(t, x) > \theta$ .

(a) On the augmented state space  $\mathbf{S} \times \mathbb{R}_+ = E \times \mathbb{R}^+ \times \mathbb{R}_+$ , with current point  $(t, x, \theta)$  and product measure  $q(dt, dx, d\theta) = dt m(dx) d\theta$ , we define a  $\sigma$ -finite Poisson point process on  $\mathbf{S} \times \mathbb{R}_+$ ,  $Q(dt, dx, d\theta)$  with intensity measure  $q(dt, dx, d\theta)$  and the thinned process  $Q^D(dt, dx, d\theta)$  on the predictable set  $D = \{\theta \leq \lambda(t, x)\}$ , with stochastic intensity measure  $q^D(dt, dx, d\theta) = \mathbf{1}_D(t, x, \theta)q(dt, dx, d\theta)$ .

Then, we construct the projection of  $Q^D(dt, dx, d\theta)$  in  $\theta$  as

$$N(dt, dx) := \int_{\mathbb{R}_+} Q^D(dt, dx, d\theta) = Q^D(dt, dx, \mathbb{R}_+).$$

Assume that for each  $t \geq 0$ ,  $\mathbb{E} \left[ \int_0^t \int_E \lambda(s, x) ds m(dx) \right] < \infty$ ; then according to the martingale property, the new random point measure  $N(dt, dx)$  has intensity measure  $q^D(dt, dx, \mathbb{R}_+) = \lambda(t, x) dt m(dx)$ .

The thinning construction goes back to the work of Kerstan (1964) and Grigelionis (1971). Lewis and Shedler (1978) used this procedure to simulate an inhomogeneous Poisson process with bounded deterministic intensity  $\lambda_t$  (see also Ogata (1981)). The proof uses the effective simulation procedure, and yields to complex calculations that disappear with the martingale method. The numerical procedure is very popular, and is very simple to derive in the case where the intensity process is bounded by a constant, in such a way that the Poisson point process  $Q(dt, dx, d\theta)$  may be restricted to a compact set for  $\theta$  and then directly and easily simulated. However, in practice these assumptions are too restrictive, and we prefer to address the simulation issue when the intensity is locally or piecewise bounded, as it is the case in this paper (see Section 2.4). The thinning construction is a good track when the intensity process depends on the process itself: in this case we have to

solve a stochastic differential equation. This is the main tool for the construction of our population process. Several contributions dealt with such thinning construction (stochastic equations driven by Poisson measures). Kurtz (1989) and Garcia (1995) used it for a long time, known as the projection method (see also Garcia and Kurtz (2006), Garcia and Kurtz (2008)), as well as for example Brémaud and Massoulié (1996), Massoulié (1998), Fournier and Méléard (2004) and Delattre et al. (2014). Let us illustrate such construction on two processes at the core of our modeling framework, namely the spatial birth process and the spatial birth process with age.

**Thinning equations for spatial birth processes** The thinning construction can be used to define a wide variety of processes as solution to stochastic equations.

(i) *Equation for spatial birth process.* Our aim here is to give a first example of such construction for the spatial birth process. Let us start with some random configuration  $Z_0(dx)$ , a random variable with values in  $\Gamma(E)$ . We want to construct a population process  $(Z_t(dx))_{t \geq 0}$  with the following dynamics: each individual  $x'$  in the configuration gives birth with rate  $b(x')$ , and in this case the new individual is located at  $x$  chosen with distribution  $k_b(x', x)m(dx)$ . It is convenient to write the total birth intensity in time and space; as individual demographic events are independent, this is the sum of individual intensities, namely

$$\beta(Z_{t-}, x) = \sum_{x' \in Z_{t-}} b(x')k_b(x', x).$$

Note here the dependence in  $Z_{t-}$ , the intensity process being predictable with respect to the population filtration. Note also that we specified a linear structure but this can be more general; this will be discussed in Section 2.5. Let  $Q_b(dt, dx, d\theta)$  be a Poisson point measure on  $\mathbb{R}_+ \times E \times \mathbb{R}_+$  with intensity measure  $q_b(dt, dx, d\theta) = dtm(dx)d\theta$ . Let us now use the previous thinning construction: the population process  $(Z_t(dx))_{t \geq 0}$  is defined as the *solution to the stochastic differential equation*

$$Z(dt, dx) = \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \beta(Z_{t-}, x)} Q_b(dt, dx, d\theta),$$

where the population at time  $t$  is  $Z_t(dx) = Z_0(dx) + Z([0, t], dx)$ . The initial condition associated with this equation is  $Z_0(dx)$ . Existence and strong uniqueness issues will be discussed in the general case. Let us write this equation in a concise way with our notations: let us define  $I^b(\chi, x, \theta) := \mathbf{1}_{\theta \leq \beta(\chi, x)}$ , then the previous equation writes

$$Z(dt, dx) = \langle I_{Z_{t-}, x}^b, Q \rangle(dt, dx).$$

(ii) *Equation for spatial birth process with age.* Let us now illustrate this construction for spatial birth processes with age. It starts with a  $\Gamma(\mathbf{S})$ -valued random variable  $Z_0(dx, da)$ , where we recall that  $\mathbf{S} = E \times \mathbb{R}_+$ ; this is the initial population.

Each individual  $(x', a')$  in the population gives birth with rate  $b(x', a')$ ; at birth, the newborn has age 0 and characteristics  $x$  drawn with distribution  $k_b(x', a', x)m(dx)$ . Note that one could define a general kernel  $k_b(x', a', x, a)m(dx)da$  to make new individuals get a particular age at arrival. Let us write the total birth intensity (density) in the population as

$$\beta(Z_{t-}, x) = \sum_{(x', a') \in Z_{t-}} b(x', a')k_b(x', a', x).$$

Let us again denote  $I^b(\chi, x, \theta) := \mathbf{1}_{\theta \leq \beta(\chi, x)}$ , and introduce  $C_b^{0,1}(\mathbf{S})$  the space of continuous and bounded functions on  $\mathbf{S} = E \times \mathbb{R}_+$  with continuous and bounded partial derivative on  $\mathbb{R}_+$ . As age is involved, we use the differentiation formula (2.2) to define the spatial birth process by means of proper differential equations. More precisely, the spatial birth process is defined as the solution to: for each  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$d\langle Z_t, f \rangle = \langle I_{Z_{t-}}^b f(\cdot, 0), Q_b \rangle(dt) + \langle Z_t, \partial_a f \rangle dt,$$

where we recall that

$$\langle I_{Z_{t-}}^b f(\cdot, 0), Q_b \rangle(dt) = \int_{x \in E} \int_{\theta \in \mathbb{R}_+} I^b(Z_{t-}, x, \theta) f(x, 0) Q_b(dt, dx, d\theta).$$

### 2.3.3 Birth-death-swap process

#### Numbering

In the previous part, we detailed the construction of spatial birth processes with age. Let us now detail the construction of death processes, which will then lead us to introduce the full birth-death-swap population process. There is a fundamental asymmetry between births and deaths: indeed, a birth amounts to add a new individual to the population, that is a zero age particle with characteristics generated by the Poisson measure, whereas a death amounts to remove an individual (its age and characteristics) which pre-exists in the population.

Let us consider a simple death process starting at  $Z_0(dx)$  and for which each individual  $x$  has death rate  $d(x)$ . Our aim is to reproduce the total death intensity at time  $t$ ,  $\sum_{x \in Z_{t-}} d(x)$ . Assume that it is possible to define a Poisson point measure  $\tilde{Q}(dt, dx, d\theta)$  with intensity measure  $\tilde{q}(dt, dx, d\theta) = dt c(dx) d\theta$ , where  $c$  is the counting measure on  $E$ , that is for any measurable subset  $B$ ,  $c(B)$  is the number of elements in  $B$ . At first sight, one may propose the following differential equation:

$$Z(dt, dx) = - \int_{\mathbb{R}_+} \mathbf{1}_{x \in Z_{t-}} \mathbf{1}_{\theta \leq d(x)} \tilde{Q}(dt, dx, d\theta),$$

because in such case the compensator of  $Z([0, t], E)$  would be

$$- \int_0^t \int_{x \in E} \mathbf{1}_{x \in Z_{s-}} d(x) c(dx) ds = - \int_0^t \left( \sum_{x \in Z_{s-}} d(x) \right) ds. \quad (2.3)$$

However, since  $c$  is not a  $\sigma$ -finite measure, such construction faces measurability issues (see the discussion in Bezborodov (2015), p16). Alternatively, one may propose  $c$  to be the Lebesgue measure that is  $\sigma$ -finite but in this case the compensator in the left hand side of (2.3) would be zero since  $Z_{t-}$  is a point measure and  $c$  charges no point.

Therefore it is needed to identify the individuals in the population, in a way that preserves the point property of the population, and by means of a  $\sigma$ -finite intensity measure. To this aim, we use a Poisson point measure  $Q_d(ds, di, d\theta)$  on  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}_+$  with intensity measure  $q_d(ds, di, d\theta) = ds n(di) d\theta$ , where  $n$  is the counting measure on  $\mathbb{N}^*$ , that is for any  $A \subset \mathbb{N}^*$ ,  $n(A)$  is the number of elements in  $A$ . Then, we build a sequence of random marks  $(X_k)$  (random variables with values in  $E$ ), as well as  $\mathbb{R}$ -valued random maps  $(A_k(\cdot))$  which are intrinsic to the population process; they are a countable "reserve" of marks and ages which can then be (randomly) picked in the population by the (random) choice of a number.

(i) Let us first consider the random population  $\xi_0$  with values in  $\Gamma(\mathbf{S})$ ,  $\mathbf{S} = E \times \mathbb{R}_+$ ; since this is a finite point measure, this can be ordered: we denote  $\xi_0 = \bigcup_{l=1}^{N_0} \{(X_l, A_l^*)\}$  where  $A_{N_0}^* < A_{N_0-1}^* < \dots < A_1^*$  and we introduce for  $1 \leq l \leq N_0$  the random maps  $A_l : t \mapsto A_l^* + t$ .

(ii) Now, let us consider a càdlàg population process  $(\xi_t(dx, da))_{t \geq 0}$  starting at  $\xi_0$ , with values in  $\Gamma(\mathbf{S})$ . Let us assume that a.s. for each  $t > 0$ ,  $\langle \bar{\xi}_t, \mathbf{1} \rangle < +\infty$ , where the *envelop process*  $(\bar{\xi}_t(dx))$  with values in  $\Gamma(E)$  is defined as  $\bar{\xi}_t = \bigcup_{s \in [0, t]} \tilde{\xi}_s$  with  $\tilde{\xi}_s(dx) = \xi_s(dx, \mathbb{R}_+)$ ; note that the envelop process jumps when a new characteristic which never existed in the past is added. In this case, the envelop process has a finite number of jumps on any finite time interval: let us denote  $(S_k)_{k \geq 0}$  the ordered times of jump of the envelop process  $(\bar{\xi}_t)$ , with convention  $S_0 = 0$ ; this is a subset of the times of jump of the population process  $(\xi_t)$ .

Let us further assume that for each  $k \in \mathbb{N}$ ,  $\langle \bar{\xi}_{S_{k+1}}, \mathbf{1} \rangle = \langle \bar{\xi}_{S_k}, \mathbf{1} \rangle + 1$  so that one individual is added at each jump. Then construct the sequence  $(X_k, A_k(\cdot))_{k \geq N_0+1}$  such that for each  $k \geq N_0 + 1$ ,  $(X_k, A_k^*) = \xi_{S_k - N_0} \setminus \xi_{S_k - N_0}^-$  and  $A_k(t) = A_k^* + t - S_k$ ;  $(X_k, A_k^*)$  is the individual added at time  $S_k - N_0$  with its own characteristics and age. This approach is closely related to that of Bezborodov (2015) (see also Fournier and Méléard (2004)) who focuses on the construction of spatial birth-death processes. In the following, we work the construction of the spatial death process as the solution to a thinning problem, and then include aging in the dynamics.

### Thinning equations for spatial death processes

Recall that the Poisson point measure  $Q_d(ds, di, d\theta)$  on  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}_+$  has intensity measure  $q_d(ds, di, d\theta) = ds n(di) d\theta$ , where  $n$  is the counting measure on  $\mathbb{N}^*$ , that is

for any  $A \subset \mathbb{N}^*$ ,  $n(A)$  is the number of elements in  $A$ .

(i) *Equation for spatial death process.* Let us begin by the construction of the spatial death process  $Z_t(dx)$  starting at  $Z_0(dx)$ , a random variable with values in  $\Gamma(E)$ . Any individual  $x \in Z_0$  has a death rate  $d(x)$ . Let us introduce

$$I^d(Z_{t-}, i, \theta) = \mathbf{1}_{X_i \in Z_{t-}} \mathbf{1}_{\theta \leq d(X_i)}.$$

Then the spatial death process is defined as the solution to the stochastic differential equation

$$Z(dt, dx) = - \int_{i \in \mathbb{N}^*} \int_{\theta \in \mathbb{R}_+} I^d(Z_{t-}, i, \theta) \delta_{X_i}(dx) Q_d(dt, di, d\theta),$$

where the population at time  $t$  is  $Z_t(dx) = Z_0(dx) + Z([0, t], dx)$ . If well defined, one checks that the compensator of  $Z_t(E)$  is

$$- \int_0^t \int_{\mathbb{N}^*} \mathbf{1}_{X_i \in Z_{s-}} d(X_i) n(di) ds = - \int_0^t \left( \sum_{x \in Z_{s-}} d(x) \right) ds.$$

(ii) *Equation for spatial death process with age.* Let us start with a random initial population  $Z_0(dx, da)$ , a random variable with values in  $\Gamma(\mathbf{S})$ ,  $\mathbf{S} = E \times \mathbb{R}_+$ . Let us denote  $I^d(Z_{t-}, t, i, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\theta \leq d(X_i, A_i(t))}$ ; note that the indicator now depends on time to keep track of the set of ages. As for the spatial birth process with age, we use the differentiation formula (2.2) to define the spatial death process with age by means of (an infinite system of) proper differential equations. Then the spatial birth process with age can be defined as the solution to: for each  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$d\langle Z_t, f \rangle = - \langle I_{Z_{t-}, t}^d f(X., A.(t)), Q_d \rangle(dt) + \langle Z_t, \partial_a f \rangle dt,$$

where we denote

$$\langle I_{Z_{t-}, t}^d f(X., A.(t)), Q_d \rangle(dt) = \int_{i \in \mathbb{N}^*} \int_{\theta \in \mathbb{R}_+} I^d(Z_{t-}, t, i, \theta) f(X_i, A_i(t)) Q_d(dt, di, d\theta).$$

### The general birth-death-swap process

Our aim here is to build the general birth-death-swap process using the thinning techniques described above. At the scale of the whole population, we want to include an external environment, modeled by a càdlàg process  $(Y_t)_{t \geq 0}$  with values in  $\mathbb{R}^l$ . Let us now reconsider the full set of demographic rates at the individual level.

**Individual hazard rates** (i) Modeling mortality patterns is crucial for demographic purposes. In particular, the aim is to take into account some trends in mortality (e.g. its overall reduction over time) but also to account for its heterogeneity (different individuals can have contrasting mortality forces) and random variability

(modeled by the impact of the stochastic environment). Thus, the impact of various parameters is modeled through the intensities defining the way demographic events occur. To model such feature for an individual of characteristics  $x \in E$  and age  $a \in \mathbb{R}_+$  at time  $t$ , we define a stochastic mortality force (or death intensity) depending on  $x, a$  but also on time and the environment, i.e.  $d(x, a, t, Y_t)$ .

(ii) To understand the dynamics of the population composition, it is necessary to account for specific birth patterns in the population. In this model, each individual can give birth to a new individual that inherits of some characteristics linked to those of its ancestor. More precisely, an individual of characteristics  $x$  and age  $a$  at time  $t$  gives birth to an individual with intensity  $b(x, a, t, Y_t)$ . In this model, children inherit of some characteristics of their parents with possible little modifications (e.g. way of life, geographic location, social class or health status), so that the mutation rule depends on the characteristics of the parents, but also their age. This is modeled by the following mechanism: when a birth occurs at time  $t$ , where the ancestor is  $(x, a)$ , a new individual with age 0 is added to the population and its characteristics  $x' \in E$  are drawn with a density  $k_b(x, a, t, Y_t, x')$  w.r.t.  $m(dx')$ . This way, the characteristics of the new individual depend on the age and the characteristics of its ancestor, but also on the time of birth and on the environment. The total birth intensity density for a population process  $(Z_t)$  is denoted

$$\beta(Z_{t-}, t, x) = \sum_{(x', a') \in Z_{t-}} b(x', a', t, Y_t) k_b(x', a', t, Y_t, x).$$

(iii) If we take only into account birth and death events, the individuals in the population have the same characteristics during life. In order to account for possible changes in the individual life histories, we allow to model *swap events* (i.e. changes of characteristics) of an individual  $(x, a)$  at time  $t$  through an intensity  $e(x, a, t, Y_t)$ . If a swap event occurs, the characteristics of the individual change from  $x$  to  $x'$  following the density  $k_e(x, a, t, Y_t, x')$  w.r.t.  $m(dx')$ . The fact that individual characteristics can randomly evolve through time is natural when dealing with human populations, and implies an additional interesting complexity in the demographic dynamics.

**Thinning equation** Let us introduce three independent Poisson point measures:

- *Birth*:  $Q_b(dt, dx, d\theta)$  on  $\mathbb{R}_+ \times E \times \mathbb{R}_+$  with intensity measure

$$q_b(dt, dx, d\theta) = dtm(dx)d\theta,$$

- *Death*:  $Q_d(ds, di, d\theta)$  on  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}_+$  with intensity measure

$$q_d(ds, di, d\theta) = dsn(di)d\theta,$$

- *Swap*:  $Q_e(ds, di, dx, d\theta)$  on  $\mathbb{R}_+ \times \mathbb{N}^* \times E \times \mathbb{R}_+$  with intensity measure

$$q_e(ds, di, dx, d\theta) = dsn(di)m(dx)d\theta.$$

We assume that  $Q_b, Q_d$  and  $Q_e$  are independent of the environment process  $Y$ . Let us introduce the following indicator functions:



- Birth:  $I^b(Z_{t-}, t, x, \theta) = \mathbf{1}_{\theta \leq \beta(Z_{t-}, t, x)}$ ,
- Death:  $I^d(Z_{t-}, t, i, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\theta \leq d(X_i, A_i(t), t, Y_t)}$ ,
- Swap:  $I^e(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\theta \leq e(X_i, A_i(t), t, Y_t) k_e(X_i, A_i(t), t, Y_t, x)}$ .

The birth-death-swap process  $(Z_t(dx, da))_{t \geq 0}$  is defined as the solution to the following system of stochastic differential equations:

for each  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$\begin{aligned} d\langle Z_t, f \rangle &= \langle I_{Z_{t-}, t}^b f(\cdot, 0), Q_b \rangle(dt) - \langle I_{Z_{t-}, t}^d f(X_\cdot, A_\cdot(t)), Q_d \rangle(dt) \\ &+ \langle I_{Z_{t-}, t}^e [f(\cdot, A_\cdot(t)) - f(X_\cdot, A_\cdot(t))], Q_e \rangle(dt) + \langle Z_t, \partial_a f \rangle dt. \end{aligned} \quad (2.4)$$

In order to get strong existence and pathwise uniqueness, we make the assumption below. It states that individual demographic rates are bounded, which is natural for the applications we have in mind.

**Assumption 2.1.** (*Bounded intensities and mutation densities*) Demographic rates and kernels are continuous and there exists positive constants  $\bar{k}$ ,  $\bar{b}$ ,  $\bar{d}$  and  $\bar{e}$  such that for all  $(x, a, t, y, x') \in E \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l \times E$ ,  $k_e(x, a, t, y, x') \leq \bar{k}$ ,  $k_b(x, a, t, y, x') \leq \bar{k}$ ,  $b(x, a, t, y) \leq \bar{b}$ ,  $d(x, a, t, y) \leq \bar{d}$  and  $e(x, a, t, y) \leq \bar{e}$ .

The following result states strong existence and pathwise uniqueness of the population process.

**Proposition 2.1.** Denote  $N_t = \langle Z_t, \mathbf{1} \rangle$  and assume that  $\mathbb{E}[N_0] < +\infty$ . Under Assumption 2.1, Equation (2.4) admits a strong càdlàg solution  $(Z_t)$  such that for each  $T > 0$ ,  $\mathbb{E}[\sup_{t \in [0, T]} N_t] < +\infty$ . Moreover, there is pathwise uniqueness for Equation (2.4).

**Remark 2.1.** Note that conditionally on the environment process  $Y$ , the process  $(Z_t)$  is by construction (2.4) a (non-homogenous) Markov process. The action of its generator on simple functions  $f$  can be read in the martingale problem (2.11), see Section 2.5.

### Proof of Proposition 2.1

Let us first prove existence of the population process in two steps.

**Step 1.** We first show that  $\mathbb{E}[\sup_{t \in [0, T]} N_t] < +\infty$ .

Note that  $\beta(Z_{t-}, t, x) \leq \bar{b}\bar{k}N_{t-}$  then a.s. for each  $t$ ,  $N_t \leq \tilde{N}_t$ , where  $\tilde{N}_t$  is the size of the population starting at  $N_0$  for which only births occur with rate  $\bar{b}\bar{k}$ , verifying the equation  $\tilde{N}_t = N_0 + \langle \mathbf{1}_{\theta \leq \bar{b}\bar{k}\tilde{N}_{s-}}, Q_b \rangle_t$ . Note that this is a simple example of coupling argument. It is thus sufficient to prove that  $\mathbb{E}[\sup_{t \in [0, T]} \tilde{N}_t] < +\infty$ . Define  $\tilde{\tau}_n = \inf\{t : \tilde{N}_t \geq n\}$  the hitting time of size  $n$ . It follows that

$$\sup_{u \in [0, t \wedge \tilde{\tau}_n]} \tilde{N}_u = \tilde{N}_{t \wedge \tilde{\tau}_n} = N_0 + \langle \mathbf{1}_{\theta \leq \bar{b}\bar{k}\tilde{N}_{s- \wedge \tilde{\tau}_n}}, Q_b \rangle_t \leq N_0 + \langle \mathbf{1}_{\theta \leq \bar{b}\bar{k} \sup_{u \in [0, s \wedge \tilde{\tau}_n]} \tilde{N}_u}, Q_b \rangle_t. \quad (2.5)$$

Since  $\mathbb{E} \left[ \langle \mathbf{1}_{\theta \leq \bar{b}k \sup_{u \in [0, s \wedge \tilde{\tau}_n]} \tilde{N}_{u-}}, q_b \rangle_t \right] = \bar{b}k \mathbb{E} \left[ \int_0^t \sup_{u \in [0, s \wedge \tilde{\tau}_n]} \tilde{N}_{u-} ds \right] \leq \bar{b}knt < +\infty$ , one has by taking the expectation in (2.5) and using Fubini's theorem

$$\mathbb{E} \left[ \sup_{u \in [0, t \wedge \tilde{\tau}_n]} \tilde{N}_u \right] \leq \mathbb{E}[N_0] + \bar{b}k \int_0^t \mathbb{E} \left[ \sup_{u \in [0, s \wedge \tilde{\tau}_n]} \tilde{N}_{u-} \right] ds.$$

Grönwall's lemma thus leads to  $\mathbb{E} \left[ \sup_{u \in [0, T \wedge \tilde{\tau}_n]} \tilde{N}_u \right] \leq \mathbb{E}[N_0] \exp(\bar{b}kT)$ . The process  $\tilde{N}$  is a birth process with constant birth rates, it is thus non-explosive. Then

$\tilde{\tau}_n \rightarrow +\infty$  as  $n \rightarrow +\infty$  and Fatou's lemma applied to positive r.v. leads to

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \tilde{N}_t \right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E} \left[ \sup_{t \leq T \wedge \tilde{\tau}_n} \tilde{N}_t \right] \leq \mathbb{E}[N_0] e^{\bar{b}kT} < +\infty.$$

**Step 2.** We prove the existence of the population process. Denote  $(T_k)_{k \geq 0}$  the times of jump of the process  $(Z_t)_{t \in [0, T]}$ . Thanks to Algorithm 2.1, see next section, one can construct the process from 0 to  $T_\infty \wedge T$  where  $T_\infty := \lim_{k \rightarrow +\infty} T_k$ . It remains to prove that a.s.  $T_\infty \wedge T = T$ . Denote  $\bar{N}_t$  the number of events which occurred until time  $t$  with rate  $\bar{\lambda} = \bar{b} + \bar{d} + \bar{e}$  for each individual, verifying  $\bar{N}_t = N_0 + \langle \mathbf{1}_{\theta \leq \bar{\lambda} \bar{N}_{s-}}, Q \rangle_t$ : this can be seen as the number of individuals in a population in which only births occur with maximum rate  $\bar{\lambda}$ . Denote  $(\bar{T}_k)_{k \geq 0}$  the times of jump of  $\bar{N}$ . Using the same argument as in the previous step for the birth process with constant rates, one has  $\bar{T}_\infty = \infty$ . Since a.s. for each  $t \in [0, T]$ ,  $N_t \leq \bar{N}_t$  and using the definition of  $\bar{N}_t$ , we deduce that the times of jump of  $N$  are a subset of those of  $\bar{N}$ , which shows that a.s.  $T_k \geq \bar{T}_k$  and  $T_\infty = \infty$ .

Let us now address the proof of pathwise uniqueness. To this aim, it is possible to directly rely on the corresponding algorithm, since demographic rates are bounded (see again Algorithm 2.1): starting from an initial random population  $Z_0$ , the process is recursively uniquely determined by  $Z_0$  and  $Q$ .  $\diamond$

## 2.4 Microsimulation

### 2.4.1 Simulation algorithm based on an equivalent representation

Since the Poisson point processes  $Q_b$ ,  $Q_d$  and  $Q_e$  are independent, it is possible to use a single Poisson point process  $Q$  whose restriction to specific domains of the thinning parameter correspond to a specific kind of demographic event. Let  $Q(ds, di, dx, d\theta)$  be a Poisson point process on  $\mathbb{R}_+ \times \mathbb{N}^* \times E \times \mathbb{R}_+$  with intensity measure  $q(ds, di, dx, d\theta) = ds n(di) m(dx) d\theta$ . We denote  $\lambda_c(Z_{t-}, t, i, x)$  the intensity at time  $t$  for individual  $i$  in population  $Z_{t-}$  corresponding to demographic event

$c \in \{b, d, e\}$ . These intensities are given by:

$$\begin{aligned}\lambda_b(Z_{t-}, t, i, x) &= b(X_i, A_i(t), t, Y_t)k_b(X_i, A_i(t), t, Y_t, x), \\ \lambda_d(Z_{t-}, t, i, x) &= d(X_i, A_i(t), t, Y_t), \\ \lambda_e(Z_{t-}, t, i, x) &= e(X_i, A_i(t), t, Y_t)k_e(X_i, A_i(t), t, Y_t, x).\end{aligned}\tag{2.6}$$

Let us us define the following indicator functions with disjoint support:

- *Birth*:  $J^b(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\theta \leq \lambda_b(Z_{t-}, t, i, x)}$ ,
- *Death*:  $J^d(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{\lambda_b(Z_{t-}, t, i, x) < \theta \leq (\lambda_b + \lambda_d)(Z_{t-}, t, i, x)}$ ,
- *Swap*:  $J^e(Z_{t-}, t, i, x, \theta) = \mathbf{1}_{(X_i, A_i(t)) \in Z_{t-}} \mathbf{1}_{(\lambda_b + \lambda_d)(Z_{t-}, t, i, x) < \theta \leq (\lambda_b + \lambda_d + \lambda_e)(Z_{t-}, t, i, x)}$ .

On this basis, the birth-death-swap process can be defined as the solution to: for each  $f \in C_b^{0,1}(\mathbf{S})$ ,

$$\begin{aligned}d\langle Z_t, f \rangle &= \langle J_{Z_{t-}, t}^b f(\cdot, 0) - J_{Z_{t-}, t}^d f(X_\cdot, A(\cdot)) + J_{Z_{t-}, t}^e [f(\cdot, A(\cdot)) - f(X_\cdot, A(\cdot))] \rangle (dQ) \\ &+ \langle Z_t, \partial_a f \rangle dt.\end{aligned}\tag{2.7}$$

For simulation purposes, the random points of  $Q$  are seen as observation times for the system, then a procedure determines if an event occurs or not. Since the time-independent part of the intensity measure of  $Q$ , that is  $n(di)m(dx)d\theta$ , is  $\sigma$ -finite, it is not possible in practice to generate all points of  $Q$ . But first, as demographic rates and mutation kernels are bounded, one can restrict the parameter  $\theta$  to live in the compact set  $[0, \bar{\lambda}]$ , where  $\bar{\lambda} = \bar{b}\bar{k} + \bar{d} + \bar{e}\bar{k}$ . As well, one would attempt to restrict the set of numbers  $i$  to be below some fixed value  $i_0$ , but since the number of new individuals in the population is not bounded, this would lead to a bias in the simulation procedure, since the random numbers that are greater than  $i_0$  are missed. The idea of the algorithm is to center the simulation around the true jumps of the population. The obverse is that the Poisson property of candidate random points is lost. In this approach, the frequency is adjusted to be as close as possible to the true rate and to fulfill two constraints: try to limit the additional information to be computationally efficient and at the same time propose a number of candidate times large enough to simulate the process exactly. Denote  $(T_n)$  the times of jump for the population; between  $T_n$  and  $T_{n+1}^-$ , the set of individual numbers that are involved in the population is fixed at the set of  $i$  such that  $X_i \in Z_{T_n}(dx, \mathbb{R}_+)$ , which is finite since  $Z$  has a finite number of jumps in any compact time interval. Then between  $T_n$  and  $T_{n+1}^-$ , the Poisson point measure  $Q$  is restricted to this fixed set. The algorithm thus derived is described below.

**Algorithm 2.1.** *We start at time  $T = 0$  with a given population  $Z_0$  of size  $N_0$ .*

**Step 1.** *Simulate a r.v.  $\tau \sim \text{Exp}(\bar{\lambda}N_T)$  and set  $T + \tau$  as the next candidate time.*

**Step 2.** *Since demographic events (birth, swap or death) can only occur at candidate times, between  $T$  and  $T + \tau$ , the population is only modified by aging: form the population  $Z_{(T+\tau)-}$  by adding  $\tau$  to all ages in  $Z_T$ .*

**Step 3.** *Select an individual  $i$  uniformly in the population  $Z_T$  and generate  $x \sim m(dx)$ .*

**Step 4.** *Generate a random variable  $\theta$  uniformly in  $[0, \bar{\lambda}N_T]$*

**Step 5.** *Determine if a demographic event occurs or not as:*

(i) *If  $\theta \leq \lambda_b(Z_{(T+\tau)-}, T + \tau, i, x)$  (birth), then form the population  $Z_{T+\tau}$  by adding  $(x, 0)$  to  $Z_{(T+\tau)-}$ .*

(ii) *If  $\lambda_b(Z_{(T+\tau)-}, T + \tau, i, x) < \theta \leq (\lambda_b + \lambda_d)(Z_{(T+\tau)-}, T + \tau, i, x)$  (death), then form the population  $Z_{T+\tau}$  by removing  $(X_i, A_i(T + \tau))$  from  $Z_{(T+\tau)-}$ .*

(iii) *If  $(\lambda_b + \lambda_d)(Z_{(T+\tau)-}, T + \tau, i, x) < \theta \leq (\lambda_b + \lambda_d + \lambda_e)(Z_{(T+\tau)-}, T + \tau, i, x)$  (swap), then define the population  $Z_{T+\tau}$  as  $Z_{(T+\tau)-}$  where  $(X_i, A_i(T + \tau))$  has been replaced by  $(x, A_i(T + \tau))$ .*

(iv) *Else, no event occurs.*

*Let  $T \leftarrow T + \tau$  and go to **Step 1**.*

**Remark 2.2.** *The procedure can be improved for example if characteristics can be easily generated by the kernels  $k_b$  and  $k_e$ . In this case, the maximum frequency can be replaced by  $\bar{\lambda} = \bar{b} + \bar{d} + \bar{e}$ . Then at step 5, the occurrence of a demographic event can be assessed by testing if a uniform random variable  $\theta$  generated in  $[0, \bar{\lambda}]$  lies in the interval  $[0, b())$  or  $]b(), b() + d())$  or  $]b() + d(), b() + d() + e())$ , and then if a birth or a swap occurs, the new characteristics can be directly generated from  $k_b$  and  $k_e$ .*

### 2.4.2 The cohort effect

The largest datasets in demographics are most often provided by government bodies and are on a national basis. As for mortality data (see e.g. HMD (2014)), they provide estimated life expectancy or estimated mortality rates by age groups, both for males and females. Such aggregate indicators are of interest, and are often the basis for decision making for state pension reforms, insurance and regulatory reserving policies, as well as pricing for some financial products indexed on national mortality indices. For example in this context, an "improvement in longevity" always refers to a decrease in these indicators over a time period. But these quantities are computed at each time by looking at all individuals by gender, without taking into account the heterogeneity of the population at a given time. In this way, comparisons between different periods are made without correcting the underlying dynamic composition. The model presented in this paper can be of great interest for such purpose, and for example to explain part of observed particular variations of aggregate mortality indicators. In the following, we illustrate this potential impact

on a case study: the cohort effect in United Kingdom. It focuses on the impact of heterogeneous birth rates on aggregate mortality.

**Cohort effect and impact of birth patterns** A birth cohort for a time period  $[t_1, t_2]$  is defined as the group of individuals who were born during this period. As they live in the same environment at the same time, but also as they are born from the same population (see below), they share some common demographic characteristics. For example, one observes specific longevity features for the individuals born between 1925 and 1945 in UK, known as the *golden cohort*. The *golden cohort* has experienced more rapid improvements than previous and next generations, as depicted in Figure 2.4. Some possible explanations are proposed and some are studied by Willets (2004): for example the impact of World War II on previous generations, tobacco consumption in next generations, positive impact of diet in early life and the post World War II welfare state. All these remarks are dedicated to factors driving mortality.

But Willets (2004) also remarks that during the period  $[1925, 1945]$ , the overall fertility is particularly low (see Figure 2.5) and that in fact it can have an impact on the composition of the population itself: *"One possible consequence of rapidly changing birth rates is that the 'average' child is likely to be different in periods where birth rates are very different. For instance, if trends in fertility vary by socio-economic class, the class mix of a population will change."*

## 2 DATA

7

Figure 2.4: Improvement rates in mortality for England & Wales by calendar year and age relative to mortality rates at the same age in the previous year. The black diagonal follows the 1930 cohort. Taken from Cairns et al. (2009).

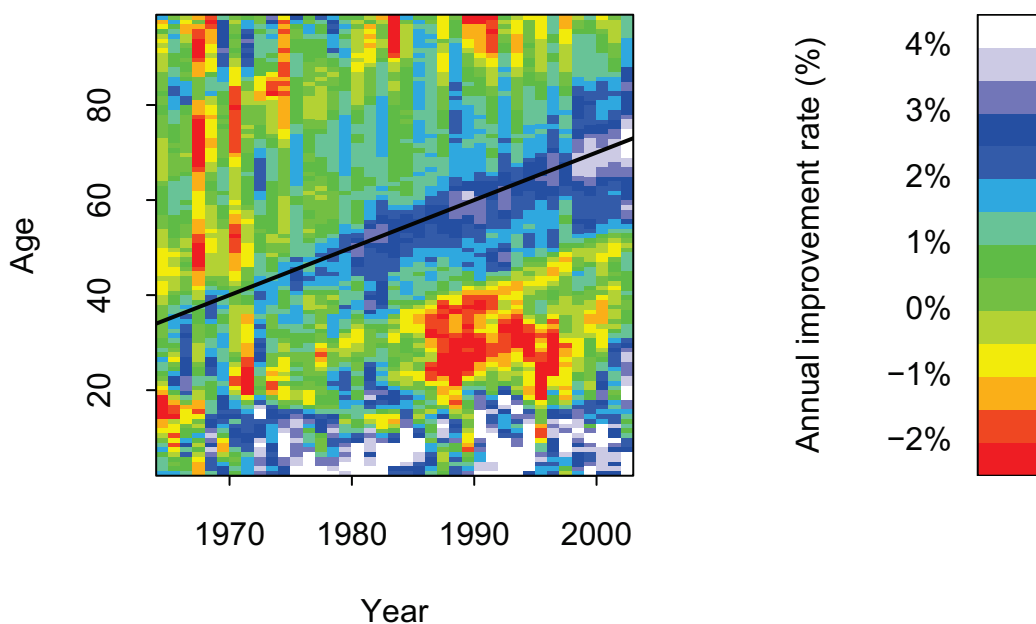
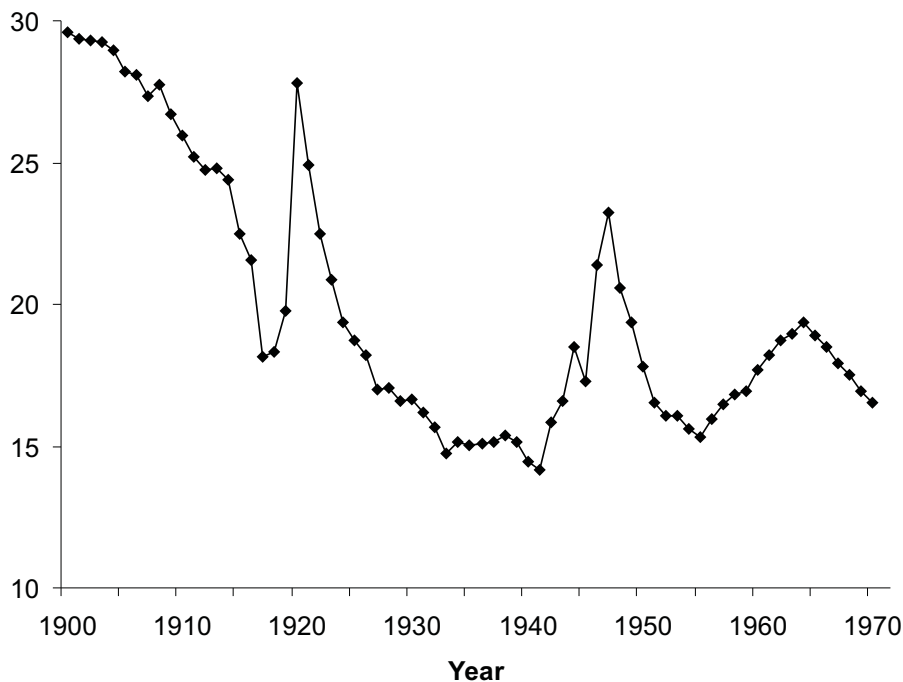


Figure 3: Improvement rates in mortality for England & Wales by calendar year and age relative to mortality rates at the same age in the previous year. Red cells imply that mortality is deteriorating; green small rates of improvement, and blue and white strong rates of improvement. The black diagonal line follows the progress

Figure 2.5: Crude birth rates per 1000 individuals in England & Wales. Taken from Willets (2004) *The Cohort Effect: Insights and Explanations*



Data source: www.mortality.org

Figure 6. Crude birth rate per 1,000 population, England and Wales, 1900 to 1970

**Toy model**<sup>¶</sup> To illustrate such effects, we propose a toy model parametrized with then again just after World War II and finally in the mid-1960s. It may be notable that the 'high improvement birth cohort' (i.e. 1925-45) coincided

(i) We compute of births relative to periods (i) immediately before and after the fact that the birth rate increased dramatically between 1941 and 1944 (by 31%) may be particularly relevant.

(ii) We average child stylized economic periods with where birth rates are very different. For instance, if trends in fertility vary by socio-economic class, the class mix of a population will change.

(iii) We also consider some stylized economic "Lower class" with time independent death rate  $d_2(a) = 2\bar{d}(a)$  but birth rate

7. U.K. TRENDS BY CAUSE OF DEATH

7.1 Cohort Trends by Cause of Death  $b(a,t) = 4c1_{[20,40]}(a)1_{(t_2,\infty)}(t) + 2c1_{[20,40]}(a)1_{[t_1,t_2]}(t)$ .

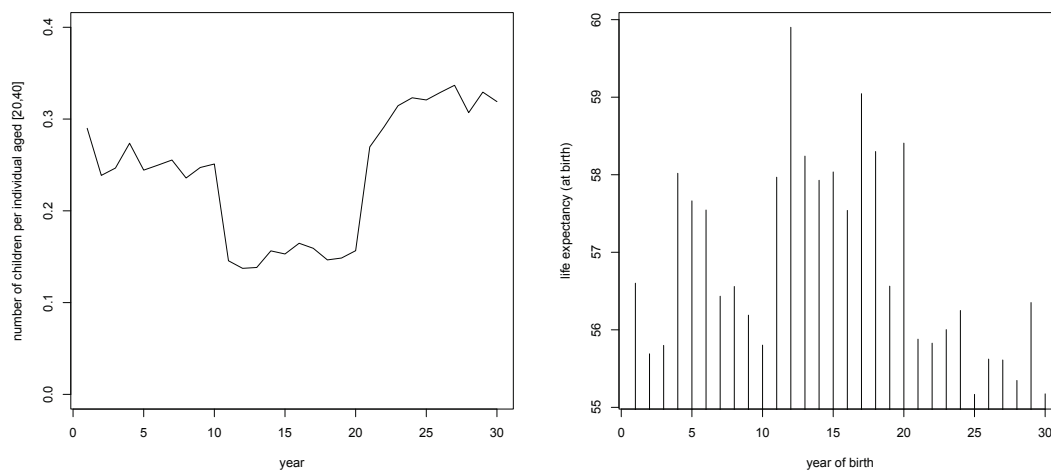
7.1.1 Further insight into the potential causes of the U.K. cohort effect Weigh high obtained through an investigation of mortality trends and specific causes of death to compute life expectancy at birth, and illustrate that the decrease in fertility can have a real impact on it.

<sup>¶</sup>Our aim with this experiment was to see if such modeling framework were able to reproduce such important stylized facts. To this aim, parameters are set to arbitrary values, which at this stage allow to get first interesting outputs. This research direction is still a work in progress, and further improvements will be considered.

**Results** The simulation is based on an initial population with 20000 individuals (randomly and uniformly) split between the upper and the lower class. The results are plotted in Figure 2.6. It is clear that aggregate fertility is reduced between times  $t_1 = 10$  and  $t_2 = 20$ . At the same time, a cohort effect appears in terms of life expectancy. Individuals born between times 10 and 20 seem to experience higher life expectancies at birth. Note that in this model death rates do not depend on time, thus one can not conclude that an improvement in longevity occurred at the individual level. In return, this shows us how birth rates provide information on changes in aggregate mortality.

These first results are interesting prospects because they stress the importance of the overall population dynamics. The actuarial studies that assess longevity/-mortality risk for insurance companies or state pension issues traditionally focus on the ages from 30/40 years. This is further encouraged by the fact that the mortality rates have a more complex structure before age 30, see Figure 2.2, including infant mortality and the accident hump at ages around 25. This approach puts aside crucial information on births that determine the initial composition of the different cohorts in terms of characteristics, as we illustrate here, but also on the distribution of births in the year, see the discussion in Subsection 2.2.1 on the reliability of mortality data and interplay with fertility.

Figure 2.6: One trajectory of the population process. Left: Number of children per individual aged [20,40]. Right: Life expectancy at birth by year of birth.



### 2.4.3 On the random environment

Our purpose here is to give insights on the relative impact of two main sources of uncertainty which arise in practice in the field of microsimulation. Consider a population starting at  $N_0 \in \mathbb{N}^*$  in which only deaths occur. Each individual alive at time  $t$  has a death rate  $\mu(Y_t) = Y_t$ , where  $Y$  is a non-negative and continuous process. In this example, the environment is therefore identified with the stochastic mortality rate experienced by individuals dying over time. Denote  $N_t$  the number of individuals at time  $t$ , defined as the solution to the stochastic equation:

$$N_t = N_0 - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq Y_s N_{s-}} Q(ds, d\theta), \quad (2.8)$$

where  $Q$  is a Poisson point measure on  $\mathbb{R}_+^2$  with intensity measure  $q(ds, d\theta) = dsd\theta$ , assumed to be independent of  $Y$ .

**Environment noise and demographic stochasticity** The classical variance decomposition allows us to split the overall stochasticity in the following way

$$\text{Var}(N_t) = \mathbb{E} [\text{Var}(N_t | Y)] + \text{Var} (\mathbb{E}[N_t | Y]).$$

The first component  $\mathbb{E} [\text{Var}(N_t^Y | Y)]$  is called demographic stochasticity: this is the randomness due to the finite size of the sample, averaged over all environment scenarios. The second component  $\text{Var} (\mathbb{E}[N_t | Y])$  is called environment noise, which remains alone in the large population framework. Note that this decomposition and the corresponding interpretation is not unique ; indeed one could have conditioned with respect to the Poisson point measure  $Q$ . However, such computations are not straightforward. Let us now express each component separately.

(i) Environment noise is given as  $\text{Var} (\mathbb{E}[N_t | Y]) = \mathbb{E} [\mathbb{E}[N_t | Y]^2] - \mathbb{E} [\mathbb{E}[N_t | Y]]^2$ . Denote  $g^Y(t) := \mathbb{E}[N_t | Y]$ . Then taking conditional expectation in Equation (2.8) leads to  $g^Y(t) = N_0 - \int_0^t Y_s g^Y(s) ds$ , which solution is given by  $g^Y(t) = N_0 \exp \left( - \int_0^t Y_s ds \right)$ . Therefore environment noise writes

$$\text{Var} (\mathbb{E}[N_t | Y]) = N_0^2 (B_t - A_t^2),$$

where  $A_t = \mathbb{E} \left[ e^{-\int_0^t Y_s ds} \right]$  and  $B_t = \mathbb{E} \left[ e^{-2 \int_0^t Y_s ds} \right]$ .

(ii) Let us now compute demographic stochasticity. First,

$$\mathbb{E} [\text{Var}(N_t | Y)] = \mathbb{E} [\mathbb{E}[N_t^2 | Y]] - N_0^2 B_t.$$

Let us compute the term  $h^Y(t) := \mathbb{E}[N_t^2 | Y]$ . Differential calculus applied to the point process  $N$  leads to  $N_t^2 = N_0^2 + \int_0^t \int_{\mathbb{R}_+} ((N_{s-} - 1)^2 - N_{s-}^2) \mathbf{1}_{\theta \leq Y_s N_{s-}} Q(ds, d\theta)$ .



Let us simplify the terms and take conditional expectation to get  $h^Y(t) = N_0^2 - 2 \int_0^t Y_s h^Y(s) ds + \int_0^t Y_s g^Y(s) ds$  which gives  $h^Y(t) = N_0(N_0 - 1)e^{-2 \int_0^t Y_s ds} + N_0 e^{-\int_0^t Y_s ds}$ . This shows that demographic stochasticity writes

$$\mathbb{E} [\text{Var}(N_t^Y | Y)] = N_0(A_t - B_t).$$

**Environment dynamics** We assume that  $(Y_t)_t$  is a Feller diffusion (or CIR process) with dynamics

$$dY_t = a(b - Y_t)dt + \rho\sqrt{Y_t}dW_t. \quad (2.9)$$

The main advantage of such modeling framework, among others, is to avoid negative death rates. One can find several other examples of affine stochastic mortality models introduced in Biffis (2005). We want to underline here that this is also a suitable framework to address the relative impact of systematic and idiosyncratic risk. In the dynamics (2.9),  $b$  is the mean reverting value; since we consider here a group of individuals dying over time, the quantity  $b$  can be seen as the target mortality rate at very high ages. The parameter  $a$  represents the speed at which the mortality rate reaches its mean reverting value. Finally, parameter  $\rho$  refers to the volatility of the mortality rate. The process starts at a (given deterministic) level  $y_0 < b$ , modelling the fact that individuals at time 0 are aging over time, so in average their mortality rate increases with time. Expectation of the death rate at a given time can be computed as  $\mathbb{E}[Y_t] = b + (y_0 - b)e^{-at}$  which start from  $y_0$  and increases to  $b$  as time goes, and its variance can be expressed as

$\text{Var}(Y_t) = \frac{\rho^2}{a^2}(1 - e^{-at}) \left[ y_0 e^{-at} + \frac{b}{2}(1 - e^{-at}) \right]$ , which starts at 0 and increases (since  $y_0 < b$ ) to  $\frac{\rho^2 b}{2a^2}$ .

This setting, although elementary, allows us to illustrate an interesting tradeoff between environment noise and demographic stochasticity, due to two competing effects: as time goes,

- (i) the population size decreases due to deaths, therefore demographic fluctuations increase, and
- (ii) the uncertainty on the death rate itself increases as well, which is natural to model the fact that long term forecasts are more uncertain than short term ones.

Let us recall the results we need for the computation of quantities  $A_t$  and  $B_t$ :

$$A_t = \mathbb{E} \left[ \exp \left( - \int_0^t Y_s ds \right) \right] = \Phi(\gamma, T) \exp(-y_0 \Psi(\gamma, T)),$$

where  $\gamma = \sqrt{a^2 + 2\rho^2}$  and

$$\Psi(\gamma, T) = \frac{2(e^{\gamma T} - 1)}{(\gamma + a)(e^{\gamma T} - 1) + 2\gamma}, \text{ and } \Phi(\gamma, T) = \left[ \frac{2\gamma e^{(\gamma+a)\frac{T}{2}}}{(\gamma + a)(e^{\gamma T} - 1) + 2\gamma} \right]^{\frac{2ab}{\rho^2}}.$$

In addition, one can see that

$$B_t = \mathbb{E} \left[ \exp \left( -2 \int_0^t Y_s ds \right) \right] = \Phi(\sqrt{\gamma^2 + 2\rho^2}, T) \exp \left( -2y_0 \Psi(\sqrt{\gamma^2 + 2\rho^2}, T) \right).$$

**Numerical illustration** Recall that the purpose of this example is to illustrate the interesting tradeoffs which appear when assessing the relative impact of environment noise and demographic stochasticity. As on one hand the intrinsic variance of the environment increases, and on the other hand the population size decreases, two effects are competing. Let us fix the initial death rate to be  $y_0 = 0.001$  (the order of magnitude of death rates for individuals aged 40 in developed countries), as well a target death rate  $b = 0.8$  (the order of magnitude of death rates at very high ages). In the following, we focus on the impact of the volatility parameter  $\rho$  and the mean reverting speed  $a$ . We formally define  $p_t$  for  $t > 0$  to be the proportion of environment noise among the total variance, that is

$$p_t = \frac{\text{Var}(\mathbb{E}[N_t | Y])}{\text{Var}(N_t)}.$$

(i) Let us first analyze the impact of volatility parameter  $\rho$ , and fix a mean reverting speed  $a = 0.01$  for the purpose of this first illustration. In Figure 2.7, we depict the behavior of the variance of the death rate with respect to time  $t$  and volatility  $\rho$ . The proportion  $p_t$  of environment noise among the total variance is depicted in Figure 2.8 for several initial sizes  $N_0$ . As expected, when varying the initial sample size  $N_0$  from 100 to 100000, from (a) to (d), the proportion  $p_t$  increases, parameters  $t$  and  $\rho$  being fixed. As well, with fixed time  $t$  and initial size  $N_0$ , this proportion increases when volatility  $\rho$  increases. More interestingly, the complex time pattern illustrates the balance between the two sources of risk: in this parameter configuration, the proportion first increases as time goes, which turns out that the uncertainty of the environment beats the decrease of the population size, therefore making environment noise the main source of risk in many cases. But then, this proportion starts to decrease due to the increasing (relative) number of deaths in the population; this makes environment noise less important in the long term in most configurations.

(ii) Let us now address the impact of the mean reverting speed  $a$ , and let us fix  $\rho = 0.005$  for this experiment. The effect of the mean-reverting parameter  $a$  on the variance of the death rate is depicted in Figure 2.7 as well; this shows as expected that environment uncertainty increases when the mean-reverting parameter decreases. In Figure 2.9, the impact of the mean-reverting parameter on the proportion  $p_t$  of environment noise is illustrated. As well, varying the initial size creates similar patterns, and as expected decreasing the mean-reverting parameter makes environment noise more important. In this experiment also, the interesting

time pattern is exhibited: for mean-reverting parameter in a reasonable range of values, the same effect appears. In the mid-term indeed, environment noise becomes important but this vanishes in the long term. Remark that for very small values of the mean-reverting speed  $a$ , environment noise remains important.

These considerations highlight the complex time patterns of several sources of risk, even in a simple framework. This shows that parameter values are crucial to get insights on the relative importance of environment noise over time, as well as it highlights for each value the population size for which external environment stochasticity is crucial to take into account.

Figure 2.7: Joint effect of time  $t$  and either volatility  $\rho$  with  $a = 0.01$  (left) or mean-reverting-speed  $a$  with  $\rho = 0.005$  (right) on the variance of the death rate  $\text{Var}(Y_t)$

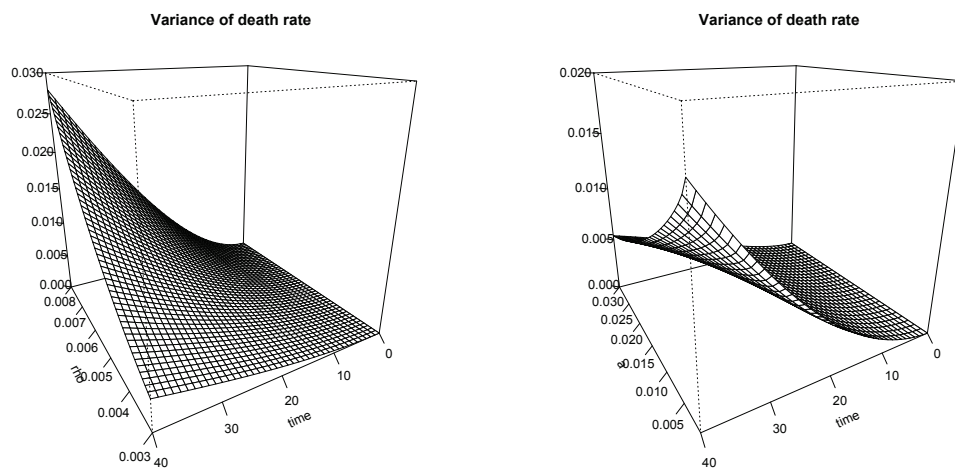


Figure 2.8: Joint effect of time  $t$  and volatility  $\rho$  on the proportion  $p_t$  of environment noise among the total variance

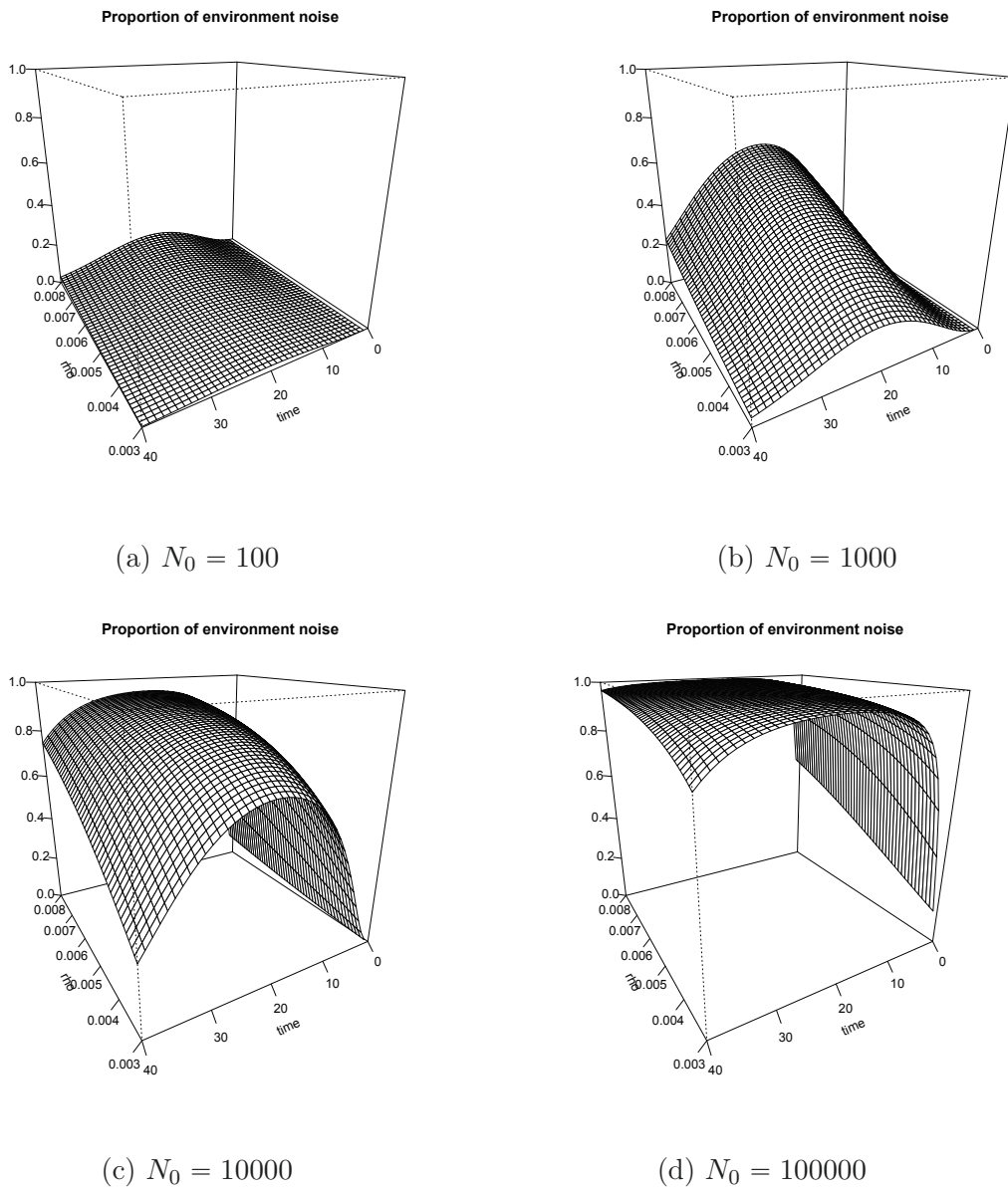
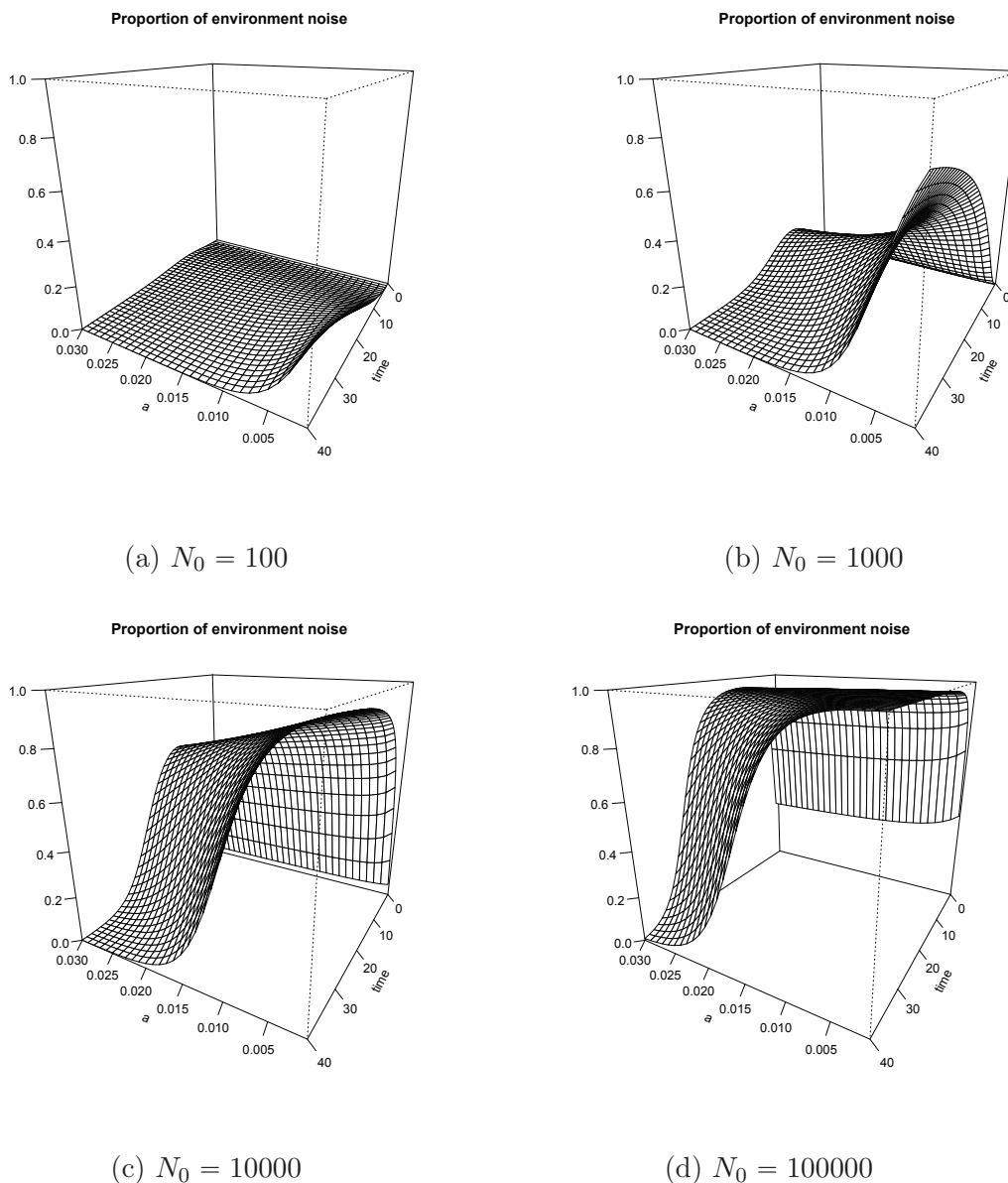


Figure 2.9: Joint effect of time  $t$  and mean reverting speed  $a$  on the proportion  $p_t$  of environment noise among the total variance



## 2.5 Macroscopic behavior of linear and non-linear dynamics

In the field of microsimulation, particular attention is given to the inherent randomness (or demographic stochasticity); it is often highlighted that this can be reduced either by (i) taking the average over a large number of Monte Carlo runs, or (ii) by increasing the size of the number of individuals in the sample, see Van Imhoff and Post (1998). In fact, the equivalence between (i) and (ii) is far from obvious, and it

is in general not true for *non-linear models*, that are models including interactions. In all cases, the two points (i) and (ii) traduce different points of view on the meaning of *macroscopic*, and in particular the theoretical analysis of point (ii) involves advanced mathematical techniques.

(i) The first philosophy is related to the study of the average of *multiple microscopic scenarios*; the associated theoretical techniques are referred to as *moment closure techniques* based on moment equations. Such techniques are used in many fields (see e.g. the review in Kuehn (2015)), among them (more recently and actively) mathematical biology where they have been introduced by Bolker and Pacala (1997) and Dieckmann and Law (2000). Several issues are related to such approaches, the most important being the difficulty to "close" the equation systems and to find explicit expressions for the mean, see the discussion in Champagnat et al. (2006). To this aim, moment closure assumptions are made. For example, *mean field* approaches focusing on the average behavior use such assumptions for the second moment, whereas for other models as *spatial-moment models* focus on spatial correlations and therefore assume moment closure for third order moments or higher (see e.g. Plank and Law (2014)). To tackle these issues, the other point of view proposes to explore rigorous derivations about large population asymptotics.

(ii) From a large population perspective, the idea is the study the dynamics of a *single scenario* based on a sample with more and more individuals. From a theoretical perspective, one considers a sequence of population processes, indexed by a scaling parameter which represents the order of magnitude of the initial sample size, related to the concept of "system size" in the field of adaptative dynamics, see Metz et al. (1996). This scaling parameter is also used to divide the weight of each individuals, allowing to recover non-degenerate behavior on the limit. This scaling is not insignificant, because its consequence is to make interactions between individuals vanish proportionally to the scaling parameter ; this suggests the different point of view compared to averaging Monte Carlo runs. Such rescaling approach leading to large number asymptotics is the first basic possible one and leads to a deterministic macroscopic approximation, but the theoretical toolkit is flexible enough to introduce more scalings, related to e.g. birth/death accelerations or rare mutations in theoretical biology, and in this case macroscopic models of several kinds appear, even stochastic: see e.g. Fournier and Méléard (2004) for several cases without age, and Tran (2008) as well as Méléard and Tran (2012) for two key large population limits with age-structured populations. Even if this large population framework involves advanced mathematical techniques, one advantage is the possibility to get complementary central limit theorems assessing the quality of the approximation, see e.g. Wang (1977), Oelschläger (1990) and Tran (2006) for such results for age-

structured populations.

These two approaches (i) and (ii) differ by several aspects, mainly

- a) in general, the average behavior does not coincide with the large population approximation, and
- b) even with (very) large sample size, the microscopic model also presents some different behavior compared to its large population asymptotics. For example, the microscopic model may exit the equilibrium of the macroscopic model (see e.g. Tran (2008)). Both aspects are linked to the different treatment of the interactions in the model.

In the following, we first focus on the linear birth-death-swap process for which we describe the moment equation and discuss the existence of a density. In a second step, we present the large population framework as well as its application to general non-linear models.

### 2.5.1 Linear model and moment equation

#### Population mean measure

Let us consider the linear model, that is the model in which individual birth-death-swap rates are given by  $b(x, a, t, Y_t)$ ,  $d(x, a, t, Y_t)$  and  $e(x, a, t, Y_t)$  and do not depend on the population itself.

**Proposition 2.2.** *For any function  $f \in C_b^{0,1}(\mathbf{S})$ , denote*

$$\begin{aligned} \Psi_t^f(x, a) &= \partial_a f(x, a) + b(x, a, t, Y_t) \int_E f(x', 0) k_b(x, a, t, Y_t, x') m(dx') - d(x, a, t, Y_t) f(x, a) \\ &+ e(x, a, t, Y_t) \int_E (f(x', a) - f(x, a)) k_e(x, a, t, Y_t, x') m(dx'), \end{aligned} \quad (2.10)$$

then for each  $t \geq 0$ ,

$$\langle Z_t, f \rangle = \langle Z_0, f \rangle + \int_0^t \langle Z_s, \Psi_s^f \rangle ds + M_t^f, \quad (2.11)$$

where  $M^f$  is a square integrable martingale with quadratic variation  $\langle M^f \rangle_t = \int_0^t \langle Z_s, H_s^f \rangle ds$  and

$$\begin{aligned} H_t^f(x, a) &= b(x, a, t, Y_t) \int_E f^2(x', 0) k_b(x, a, t, Y_t, x') m(dx') + d(x, a, t, Y_t) f^2(x, a) \\ &+ e(x, a, t, Y_t) \int_E (f(x', a) - f(x, a))^2 k_e(x, a, t, Y_t, x') m(dx'). \end{aligned} \quad (2.12)$$

**Proof of Proposition 2.2.** Equation (2.11) can be immediately obtained by compensating the Poisson point measures  $Q_b$ ,  $Q_d$  and  $Q_e$  in (2.4) by their intensity measures to obtain a local martingale, whose bracket can be computed. It remains to prove that  $M^f$  is a square integrable martingale: for each  $t \in [0, T]$ , we get

$$\mathbb{E} [\langle M^f \rangle_t] \leq \|f\|_\infty^2 ((\bar{b} + 2\bar{e})\bar{k} + \bar{d}) \mathbb{E} \left[ \sup_{0 \leq s \leq t} N_s \right],$$

which is finite according to Proposition 2.1. This concludes the proof.  $\diamond$

Define the conditional mean empirical measure  $\xi_t^Y(dx, da)$  by: a.s. for each  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$ ,  $\langle \xi_t^Y, f \rangle = \mathbb{E} [\langle Z_t, f \rangle | Y]$ . Formally, this can be denoted  $\xi_t^Y(dx, da) = \mathbb{E} \left[ \sum_{(x', a') \in Z_t} \delta_{(x', a')}(dx, da) | Y \right]$ . The following result states the behavior of the average population: given the environment, the mean measure is given as a solution to some deterministic equation.

**Theorem 2.1.** *The conditional mean measure  $\xi_t^Y$  satisfies the following equation: a.s. for each  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$  and each  $t \in [0, T]$ ,*

$$\langle \xi_t^Y, f \rangle = \langle \xi_0^Y, f \rangle + \int_0^t \langle \xi_s^Y, \Psi_s^f \rangle ds, \quad (2.13)$$

where  $\Psi$  is defined in Equation (2.10).

**Proof of Theorem 2.1.** Let us take conditional expectation in Equation (2.11) to get for each  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$  and each  $t \in [0, T]$ ,

$$\langle \xi_t^Y, f \rangle = \langle \xi_0^Y, f \rangle + \mathbb{E} \left[ \int_0^t \langle Z_s, \Psi_s^f \rangle | Y \right] ds. \quad (2.14)$$

One can invert expectation and integral thanks to Assumption 2.1, Proposition 2.1 and properties on  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$ . This concludes the proof.  $\diamond$

### Existence of a density and strong formulation

At this stage, we highlighted the fact that for the linear model, the average of Monte-Carlo simulations satisfies the deterministic Equation (2.13). It will be clear in the following that this equation corresponds to classical equations in demography. This will be the purpose of the following developments to detail such equations. In parallel, one question arises concerning the existence of a density of the population mean measure, that is a density  $g_t(x, a)$  such that the mean measure satisfies  $\xi_t(dx, da) = g_t(x, a)m(dx)da$ . Such property would ensure no concentration of population mass on particular ages or characteristics.

The result providing the strong PDE representation is stated below. Note that without further assumption, the mean measure is a weak solution of the system



(2.15) detailed below. Here we assume the existence of a density at time 0, which allows to get the existence for any time. This ensures that no concentration effect around age or characteristic arise if the initial condition admits a density. In addition, a regularity assumption is used to prove the strong solution. In the following result,  $g_t^Y(x, a)$  can be interpreted as the averaged (conditionally on  $Y$ ) quantity of individuals of characteristics in  $[x, x + dx)$  and age in  $[a, a + da)$  in the population. The proof is given after some interpretations and remarks. Note that the technique is similar to e.g. Tran (2006), but we rely on a different disintegration argument in order to overcome the swap pattern complexity.

**Theorem 2.2.** (i) *Suppose that a.s.  $\xi_0^Y$  admits a density  $g_0^Y(x, a) = g^Y(x, a, 0)$  w.r.t. the measure  $m(dx) \otimes da$  on  $\mathbf{S}$ . Then the solution  $\xi^Y$  of Equation (2.13) admits at each time  $t \in [0, T]$  a density  $g_t^Y(x, a) = g^Y(x, a, t)$ .*

(ii) *Moreover, under the assumption that a.s.  $g^Y \in C_b^{0,1,1}(\tilde{\mathcal{X}} \times [0, T])$ , then  $g^Y$  is (strong) solution to the following equation system: a.s.,  $m(dx) \otimes da$ -a.e.,  $\forall t \in [0, T]$ ,*

$$\begin{aligned} & (\partial_a g^Y + \partial_t g^Y)(x, a, t) = -[d(x, a, t, Y_t) + e(x, a, t, Y_t)] g^Y(x, a, t) \\ & + \int_E g^Y(x', a, t) e(x', a, t, Y_t) k_e(x', a, t, Y_t, x) m(dx'), \\ & g^Y(x, 0, t) = \int_{\mathbf{S}} g^Y(x', a, t) b(x', a, t, Y_t) k_b(x', a, t, Y_t, x) m(dx') da, \\ & g^Y(x, a, 0) = g_0^Y(x, a). \end{aligned} \tag{2.15}$$

**Interpretation of the mean population equation** We propose here the description the mechanism involved in Equation (2.15), from the point of view of classical deterministic models. Malthus' equation can be written as a linear PDE in the following way:

$$\partial_t n(x, t) = \int_E n(x', t) b(x') k_b(x', x) dx' - d(x) n(x, t), \quad n(x, 0) = n_0(x), \tag{2.16}$$

In this model, the initial distribution of characteristics in the "average" population is described by  $n_0(x)$ . The density of individuals of characteristic  $x$  then evolves through time from  $n(x, t)$  to  $n(x, t + dt)$  by including all individuals who were born from any parent of characteristic  $x'$  and inherited of a characteristic  $x$  according to the mutation law  $k_b(x', \cdot)$ , and subtracting the individuals who died in the small time interval  $[t, t + dt]$ .

McKendrick-Von Foerster's equation models age-structured populations without interactions (so that the PDE is linear, as it is the case for Equation (2.15)):

$$\partial_t n(a, t) + \partial_a n(a, t) = -d(a) n(a, t), \quad n(0, t) = \int_0^{+\infty} b(a) n(a, t) da, \quad n(a, 0) = n_0(a). \tag{2.17}$$

The interpretation of such transport PDE in which each particle or individual with age  $a$  is transported to age  $a + dt$  through time is the following: start with an initial population whose distribution of ages is  $n(a, 0) = n_0(a)$ . The density of individuals of age  $a > 0$  only evolves due to the deaths occurring through time (because no individual arrives in the population with a strictly positive age). Individuals are also aging, which is modeled by the transport term  $\partial_a n(a, t)$ . Compared to Equation (2.16), the PDE (2.17) needs an initial condition for ages: the density of individuals of age 0 at each time  $t$ ,  $n(0, t)$ , is given by counting the number of birth at time  $t$ . This is known as the renewal or mass conservation condition.

The limit Equation (2.15) combines these two types of PDEs and can be interpreted in the same way. One can notice the impact of the additional swap pattern, as well as the random environment  $Y$ .

**Proof of Theorem 2.2.** For the whole proof, we use the concise notations  $k_{b,e}(x, a, t, Y_t, x') \equiv k_{b,e}(x, a, x')$  as well as  $g^Y \equiv g$  and  $\xi^Y \equiv \xi$ . Suppose that  $\xi_0$  has a density  $g_0(x, a) = g(x, a, 0)$  w.r.t.  $m(dx)da$  on  $\mathbf{S}$ .

**Step 1:** Let  $t \in [0, T]$  and let us prove that  $\xi_t$  has a density  $g_t(x, a)$  w.r.t.  $m(dx)da$  on  $\mathbf{S}$ .

**Step 1.a.** We first prove the absolute continuity of the marginal in  $a$  of  $\xi_t$ , denoted  $\hat{\xi}_t(da) := \xi_t(E, da)$ . Our aim is to prove that there is some measurable and integrable function  $F$  on  $[0, T] \times \mathbb{R}_+$  such that for all  $\Psi$  measurable, non-negative and bounded  $\hat{\xi}_t - a.e.$  on  $\mathbb{R}_+$ ,

$$\int_{\mathbb{R}_+} \Psi(a) \hat{\xi}_t(da) \leq \int_{\mathbb{R}_+} \Psi(a) F(t, a) da. \quad (2.18)$$

Using a density argument, it is sufficient to prove the result for  $\Psi \in C_b^1(\mathbb{R}_+, \mathbb{R}_+)$ . First, from Equation (2.13), one deduces by integration by parts that for each  $f_t(x, a)$  differentiable in  $t$  and  $a$ , with continuous partial derivatives,

$$\langle \xi_t, f_t \rangle = \langle \xi_0, f_0 \rangle + \int_0^t \langle \xi_s, \partial_t f_s + \Psi_s^{f_s} \rangle ds. \quad (2.19)$$

Let us use Equation (2.19) with for each  $s \in [0, t]$ ,  $f_s(x, a) = \Psi(a + t - s)$ , which simplifies the derivatives and the evolutionary term, leading to

$$\begin{aligned} \int_{\mathbb{R}_+} \Psi(a) \hat{\xi}_t(da) &= \int_{\mathbf{S}} \Psi(a) \xi_t(dx, da) \\ &= \int_{\mathbf{S}} \Psi(a + t) \xi_0(dx, da) \\ &+ \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) \left[ b(x, a, s, Y_s) \Psi(t - s) - d(x, a, s, Y_s) \Psi(a + t - s) \right] \\ &\leq \int_{\mathbf{S}} \Psi(a + t) \xi_0(dx, da) + \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) b(x, a, s, Y_s) \Psi(t - s), \end{aligned}$$

where the last inequality comes from the fact that  $\Psi$  is non-negative.

Let us use the fact that  $\xi_0$  has a density  $g_0(x, a)$  w.r.t.  $m(dx)da$  on  $\mathbf{S}$  and the change of variable  $\alpha := a + t$  to get for the first term

$$\int_{\mathbf{S}} \Psi(a + t) \xi_0(dx, da) = \int_{\mathbb{R}_+} \Psi(\alpha) \left\{ \int_E g_0(x, \alpha - t) \mathbf{1}_{\alpha \geq t} m(dx) \right\} d\alpha.$$

With the change of variable  $\alpha := t - s$ , the second term can be written

$$\begin{aligned} & \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) b(x, a, s, Y_s) \Psi(t - s) \\ &= \int_{\mathbb{R}_+} \Psi(\alpha) \left\{ -\mathbf{1}_{\alpha \in [0, t]} \int_{\mathbf{S}} \xi_{t-\alpha}(dx, da) b(x, a, t - \alpha, Y_{t-\alpha}) \right\} d\alpha. \end{aligned}$$

This proves Equation (2.18) with

$$F(t, \alpha) = \int_E g_0(x, \alpha - t) \mathbf{1}_{\alpha \geq t} m(dx) - \mathbf{1}_{\alpha \in [0, t]} \int_{\mathbf{S}} \xi_{t-\alpha}(dx, da) b(x, a, t - \alpha, Y_{t-\alpha}),$$

which is integrable since

$$\int_{\mathbb{R}_+} \Psi(a) F(t, a) da = \int_{\mathbf{S}} \Psi(a+t) \xi_0(dx, da) + \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) b(x, a, s, Y_s) \Psi(t-s) < +\infty.$$

**Step 1.b.** We then show the absolute continuity w.r.t.  $m(dx)da$ . We prove that there is some measurable and integrable function  $H$  on  $[0, T] \times \mathbf{S}$  such that for each  $t \in [0, T]$  and each  $\Phi$  measurable, non-negative and bounded  $\xi_t - a.e.$  on  $\mathbf{S}$ ,

$$\int_{\mathbf{S}} \Phi(x, a) \xi_t(dx, da) \leq \int_{\mathbf{S}} \Phi(x, a) H(t, x, a) m(dx) da. \quad (2.20)$$

Since  $C_b^{0,1}(\mathbf{S}, \mathbb{R}_+)$  is dense in the set of measurable, non-negative and bounded  $\xi_t - a.e.$  maps, it is sufficient to prove this inequality for  $\Phi \in C_b^{0,1}(\mathbf{S}, \mathbb{R}_+)$ . As previously, we use Equation (2.19) with for each  $s \in [0, t]$ ,  $f_s(x, a) = \Phi(x, a + t - s)$ . This leads to

$$\begin{aligned} & \int_{\mathbf{S}} \Phi(x, a) \xi_t(dx, da) \\ & \leq \int_{\mathbf{S}} \Phi(x, a + t) \xi_0(dx, da) + \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) \left[ b(x, a, s, Y_s) \int_E \Phi(x', t - s) k_b(x, a, x') m(dx') \right. \\ & \quad \left. + e(x, a, s, Y_s) \int_E \Phi(x', a + t - s) k_e(x, a, x') m(dx') \right]. \end{aligned}$$

As for the case with age only, the first dominating term can be written, with the change of variable  $\alpha = a + t$ ,

$$\begin{aligned} \int_{\mathbf{S}} \Phi(x, a + t) \xi_0(dx, da) &= \int_{\mathbf{S}} \Phi(x, a + t) g_0(x, a) m(dx) da \\ &= \int_{\mathbf{S}} \Phi(x, \alpha) \{g_0(x, \alpha - t) \mathbf{1}_{\alpha \geq t}\} m(dx) d\alpha. \end{aligned}$$

As for the second term, with the change of variable  $\alpha = t - s$ , we get

$$\begin{aligned} & \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) b(x, a, s, Y_s) \int_E \Phi(x', t - s) k_b(x, a, x') m(dx') \\ &= - \int_0^t d\alpha \int_{\mathbf{S}} \xi_{t-\alpha}(dx, da) b(x, a, t - \alpha, Y_{t-\alpha}) \int_E \Phi(x', \alpha) k_b(x, a, x') m(dx') \\ &= \int_0^t \int_{x' \in E} \Phi(x', \alpha) \left\{ - \int_{\mathbf{S}} \xi_{t-\alpha}(dx, da) b(x, a, t - \alpha, Y_{t-\alpha}) k_b(x, a, x') \right\} m(dx') d\alpha \end{aligned}$$

where the third equality comes from Fubini theorem for non-negative functions.

It remains to express the evolutionary term in a convenient way. We have:

$$\tilde{\xi}_t(dx, da) = \hat{\xi}_t(da) p_t(a, dx),$$

where  $p_t(a, dx)$  is a version of the measure  $\tilde{\xi}_t(dx, da)$  conditioned to the fact that age is equal to  $a$ . Using the absolute continuity of  $\hat{\xi}_t(da)$  proved in **Step 1.a.**, for all  $t \in \mathbb{R}_+$ , there exists  $q_t$  such that

$$\tilde{\xi}_t(dx, da) = q_t(a) p_t(a, dx) da.$$

The evolutionary term can be thus written as

$$\begin{aligned} & \int_0^t ds \int_{\mathbf{S}} \xi_s(dx, da) \left\{ e(x, a, s, Y_s) \int_E \Phi(x', a + t - s) k_e(x, a, x') m(dx') \right\} \\ &= \int_{\mathbb{R}_+} ds \mathbf{1}_{s \in [0, t]} \int_{\mathbf{S}} q_s(a) p_s(a, dx) da \left\{ e(x, a, s, Y_s) \int_E \Phi(x', a + t - s) k_e(x, a, x') m(dx') \right\}. \end{aligned}$$

We make the change of variable  $c_t : (a, s) \mapsto (\alpha, \beta) = (a + t - s, s)$  continuous and bijective from  $\mathbb{R}_+^2$  to  $\mathbb{R}_+^2$ , with inverse  $c_t^{-1} : (\alpha, \beta) \mapsto (a, s) = (\alpha - t + \beta, \beta)$ . The inverse  $c_t^{-1}$  has a Jacobian matrix with determinant 1, therefore the previous term is equal to

$$\begin{aligned} & \int_{\mathbb{R}_+} ds \mathbf{1}_{s \in [0, t]} \int_{\mathbf{S}} q_s(a) p_s(a, dx) da \left\{ e(x, a, s, Y_s) \int_E \Phi(x', a + t - s) k_e(x, a, x') m(dx') \right\} \\ &= \int_{\mathbb{R}_+} d\beta \mathbf{1}_{\beta \in [0, t]} \int_{\mathbf{S}} q_\beta(\alpha - t + \beta) p_\beta(\alpha - t + \beta, dx) d\alpha \times \\ & \left\{ e(x, \alpha - t + \beta, \beta, Y_\beta) \int_E \Phi(x', \alpha) k_e(x, \alpha - t + \beta, x') m(dx') \right\}. \end{aligned}$$

According to Fubini theorem, this formula is equal to

$$\int_{\mathbb{R}_+} d\alpha \int_{x' \in E} m(dx') \Phi(x', \alpha) H(t, x', \alpha),$$

where

$$\begin{aligned} H(t, x', \alpha) := & \int_{\mathbb{R}_+} d\beta \int_{x \in E} p_\beta(\alpha - t + \beta, dx) \times \\ & \left\{ \mathbf{1}_{\beta \in [0, t]} q_\beta(\alpha - t + \beta) e(x, \alpha - t + \beta, \beta, Y_\beta) k_e(x, \alpha - t + \beta, x') \right\}. \end{aligned}$$

**Step 2.** We now prove that Equation (2.15) is verified by  $g$ . Let us differentiate Equation (2.13) to get

$$\begin{aligned} \partial_t \left\{ \int_{\mathbf{S}} f(x, a) g(x, a, t) m(dx) da \right\} &= \int_{\mathbf{S}} g(x, a, t) m(dx) da \left[ \partial_a f(x, a) - d(x, a, t, Y_t) f(x, a) \right. \\ &\quad + b(x, a, t, Y_t) \int_E f(x', 0) k_b(x, a, x') m(dx') \\ &\quad \left. + e(x, a, t, Y_t) \int_E (f(x', a) - f(x, a)) k_e(x, a, x') m(dx') \right]. \end{aligned} \quad (2.21)$$

**Step 2.a.** We prove that the partial differential equation of (2.15) is verified by  $g$ . We choose  $f \in C^{0,1}(\mathbf{S})$  and such that  $f$  has compact support in  $E \times ]0, \infty[$ . This implies that for each  $x \in E$ ,  $f(x, 0) = 0$  and the birth term vanishes in Equation (2.21). Moreover, by Fubini theorem and integration by parts,  $\langle g_t, \partial_a f \rangle = -\langle \partial_a g_t, f \rangle$ , where we use the fact that  $f$  has compact support in  $E \times ]0, \infty[$ . Thus, Equation (2.21) becomes

$$\begin{aligned} &\partial_t \int_{\mathbf{S}} g(x, a, t) f(x, a) m(dx) da + \int_{\mathbf{S}} \partial_a g(x, a, t) f(x, a) m(dx) da \\ &= - \int_{\mathbf{S}} f(x, a) \{d(x, a, t, Y_t) + e(x, a, t, Y_t)\} g(x, a, t) m(dx) da \\ &\quad + \int_{\mathbf{S}} g(x, a, t) e(x, a, t, Y_t) \left\{ \int_E f(x', a) k_e(x, a, x') m(dx') \right\} m(dx) da. \end{aligned}$$

Assumption on  $g$  in the l.h.s. and Fubini's theorem in the r.h.s. thus leads to

$$\begin{aligned} &\int_{\mathbf{S}} \partial_t g(x, a, t) f(x, a) m(dx) da + \int_{\mathbf{S}} \partial_a g(x, a, t) f(x, a) m(dx) da \\ &= - \int_{\mathbf{S}} f(x, a) \{d(x, a, t, Y_t) + e(x, a, t, Y_t)\} g(x, a, t) m(dx) da \quad (2.22) \\ &\quad + \int_{\mathbf{S}} f(x', a) \left\{ \int_E g(x, a, t) e(x, a, t, Y_t) k_e(x, a, x') m(dx) \right\} m(dx') da. \end{aligned}$$

This proves the partial differential equation in (2.15), using the density of the set of functions  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$  having a compact support in  $E \times ]0, \infty[$  in the set of bounded  $m(dx) \otimes da$ -a.e. functions on  $\mathbf{S}$ .

**Step 2.b.** We prove that the renewal condition in Equation (2.15) is satisfied by  $g$ . Let  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$ . Integration by part formula leads to

$$\int_{\mathbf{S}} g(x, a, t) \partial_a f(x, a) m(dx) da = -f(x, 0) g(x, 0, t) - \int_{\mathbf{S}} \partial_a g(x, a, t) f(x, a) m(dx) da,$$

and by Equations (2.21) and (2.22), one has

$$\int_E f(x, 0) g(x, 0, t) = \int_{\mathbf{S}} g(x, a, t) b(x, a, t, Y_t) \left\{ \int_E f(x', 0) k_b(x, a, x') m(dx') \right\} m(dx) da.$$

Fubini's theorem and a change of notation  $x \leftrightarrow x'$  lead to

$$\int_E f(x, 0)g(x, 0, t) = \int_E f(x, 0) \left\{ \int_{\mathbf{S}} g(x', a, t)b(x', a, t, Y_t)k_b(x', a, x)m(dx')da \right\} m(dx).$$

Since  $h \in \mathcal{C}_b(E)$  where  $h : x \mapsto f(x, 0)$ , and  $\mathcal{C}_b(E)$  is dense in the set of  $m(dx)$ -a.e. bounded functions on  $E$ , the initial condition in age is proved:

$$g(x, 0, t) = \int_{\mathbf{S}} g(x', a, t)b(x', a, t, Y_t)k_b(x', a, x)m(dx')da. \diamond$$

## 2.5.2 Non-linear models: a large population perspective

### Non-linearities in population models

Non-linearities appear naturally in microsimulation and related applications, and in fact this is the main advantage of such methods: a flexible simulation toolkit allowing to make links between individuals, in other words to include *interactions*. In the early history of deterministic population models, some drawbacks of linear models such as those of Malthus and McKendrick Von-Foerster have been highlighted. The main drawback for demographic issues concerns their inability to reproduce the time behavior of population growth, due to their structural linear formulation. Based on these considerations, Verhulst extended Malthus' deterministic equation to make individual birth and death rates depend on the whole population. The work of Verhulst then led Gurtin and MacCamy to introduce such non-linearities in the more complex age-structured McKendrick-VonFoerster equation. In the following, early attempts to incorporate both sexes into existing one-sex models led to include other kinds of non-linearities particularly into linking the male and female birth rates (see e.g. Kendall (1949)). Further investigation on related topics contributed to improve the link between male and female populations, such as "marriage squeeze", i.e. the impact on marriage of an imbalance between males and females (see e.g. Schoen (1983), Guilmoto (2012)). As well, including interactions at the level of individual death rates is crucial to model a wide variety of practical issues, related for example to the modeling of the spread of pandemics (SIR models, see for example Cléménçon et al. (2008) and references therein), but more generally it seems important to make individual death rates depend on the current population.

**How to model interactions ?** Let us describe the associated mathematical framework, inspired by Fournier and Méléard (2004), Champagnat et al. (2006) and Tran (2008). We introduce individual birth, death and swap rates that are dependent on the population  $Z$  itself through an interaction kernel. Let us denote  $\mathbf{x} = (x, a)$  an element of  $\mathbf{S} = E \times \mathbb{R}_+$ . An interaction kernel is a map  $U : \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{R}$  where  $U(\mathbf{x}, \mathbf{x}')$  measures the impact of the individual  $\mathbf{x}'$  on the individual  $\mathbf{x}$ . Therefore, the impact of the whole population  $Z(dx, da)$  on individual  $\mathbf{x}$  is computed as

$\langle Z, U_{\mathbf{x}} \rangle = \sum_{\mathbf{x}' \in Z} U(\mathbf{x}, \mathbf{x}')$ . We introduce interaction kernels for each kind of event (birth, death, swap), namely  $U^b$ ,  $U^d$  and  $U^e$ .

The birth, death and swap rates for an individual  $\mathbf{x} = (x, a)$  alive in the population  $Z_t$  at time  $t$  are then  $b(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^b \rangle)$ ,  $d(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^d \rangle)$  and  $e(\mathbf{x}, t, Y_t, \langle Z_t, U_{\mathbf{x}}^e \rangle)$  respectively. The construction of the population dynamics as a thinning problem is similar, as well as the existence of the process on the whole time horizon  $\mathbb{R}_+$  which holds (i.e. non-explosion is guaranteed) under the assumption that all individual rates are uniformly bounded by a constant (see point (i) of Assumption 2.2 below). Moreover, to derive the large number asymptotics (in particular to get uniqueness for the limit equation (2.13)), we need some Lipschitz conditions with respect to the interaction kernel (see point (ii) of Assumption 2.2 below).

**Assumption 2.2.** (i) *There exists positive constants  $\bar{k}$ ,  $\bar{b}$ ,  $\bar{d}$  and  $\bar{e}$  such that for all  $(x, a, t, y, x', u) \in E \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l \times E \times \mathbb{R}$ ,  $k_e(x, a, t, y, x') \leq \bar{k}$ ,  $k_b(x, a, t, y, x') \leq \bar{k}$ ,  $b(x, a, t, y, u) \leq \bar{b}$ ,  $d(x, a, t, y, u) \leq \bar{d}$  and  $e(x, a, t, y, u) \leq \bar{e}$ .*

(ii) *For any demographic rate  $r \in \{b, d, e\}$  there exists some constant  $L_r$  such that for each  $(x, a, t, y) \in E \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^l$ , for each  $(u, v) \in \mathbb{R}^2$ ,*

$$|r(x, a, t, y, u) - r(x, a, t, y, v)| \leq L_r |u - v|.$$

### Large population asymptotics

The study of large population asymptotics relies on a scaling parameter  $n \in \mathbb{N}^*$  which is grown to infinity. It is used as the order of magnitude of the initial population size, as well as to decrease the weight of each individual in order to capture a non-degenerate dynamics in the limit. For  $n \in \mathbb{N}^*$ , we introduce the space of rescaled finite measures  $\Gamma^n(\mathbf{S}) = \{\frac{1}{n}\chi \mid \chi \in \Gamma(\mathbf{S})\}$ , and consider a random variable  $Z_0^n$  with values in  $\Gamma^n(\mathbf{S})$ . Here, the fact that each individual is weighted  $\frac{1}{n}$  appears; assumptions will then be made on random initial populations to model the fact that the order of magnitude of initial population size  $N_0^n = n\langle Z_0^n, \mathbf{1} \rangle$  is  $n$ .

Before stating this assumption of convergence of the rescaled population at time 0, we clarify the space of measures and its topology. Let us denote again  $\mathbf{S} = E \times \mathbb{R}_+$  and introduce  $\mathcal{M}(\mathbf{S})$  the space of finite positive measures on  $\mathbf{S}$ ; note that for each  $n \geq 1$ ,  $\Gamma^n(\mathbf{S}) \subset \mathcal{M}(\mathbf{S})$ , so that all rescaled population live in the same space  $\mathcal{M}(\mathbf{S})$ . The space  $\mathcal{M}(\mathbf{S})$  can be embedded with the topology of the vague or weak convergence. Recall that the weak convergence in  $\mathcal{M}(\mathbf{S})$  is defined as:  $\chi_n \xrightarrow{weak} \chi$  if  $\langle \chi_n, f \rangle \rightarrow \langle \chi, f \rangle$  for each continuous and bounded  $f$ . The vague convergence is defined as  $\chi_n \xrightarrow{vague} \chi$  if  $\langle \chi_n, f \rangle \rightarrow \langle \chi, f \rangle$  for each continuous  $f$  with compact support. If  $\mathbf{S}$  were a compact set, for example if a fixed limiting age were imposed, these topologies would be the same, which would contribute to simplify some assumptions. In our case however, since  $\mathbf{S} = E \times \mathbb{R}_+$  is not compact, these two topologies are strictly included.

In the following, we denote  $\mathbb{D}([0, T], (\mathcal{M}(\mathbf{S}), w))$  (resp.  $\mathbb{C}([0, T], (\mathcal{M}(\mathbf{S}), w))$ ) the space of càdlàg (resp. continuous) processes on  $[0, T]$  taking values in  $\mathcal{M}(\mathbf{S})$  embedded with the topology of weak convergence.

**Assumption 2.3.** (*Convergence of the initial population*) *There exists  $\xi_0 \in \mathcal{M}(\mathbf{S})$  such that  $Z_0^n$  converges in distribution, and for the weak topology on  $\mathcal{M}(\mathbf{S})$ , towards the deterministic measure  $\xi_0$ .*

Under the previous assumption, we get in particular the convergence of  $\langle Z_0^n, \mathbf{1} \rangle = nN_0^n$  towards  $\langle \xi_0, \mathbf{1} \rangle$ , which shows that the initial sample size is of order  $n$ . Let us also assume some moment condition on the initial rescaled (random) populations:

**Assumption 2.4.** (*Moments of order strictly greater than 1*) *There exists  $\alpha > 0$  such that*

$$\sup_{n \in \mathbb{N}^*} \mathbb{E} [\langle Z_0^n, \mathbf{1} \rangle^{1+\alpha}] < +\infty.$$

The convergence in distribution of the sequence of rescaled population processes is given below:

**Theorem 2.3.** *Under Assumptions 2.2, 2.3 and 2.4, the sequence  $(Z^n)_{n \in \mathbb{N}^*}$  converges in distribution in  $\mathbb{D}([0, T], (\mathcal{M}(\mathbf{S}), w))$  towards the unique (deterministic) solution  $\xi \in \mathbb{C}([0, T], (\mathcal{M}(\mathbf{S}), w))$  to the following non-linear equation: for each  $f \in \mathcal{C}_b^{0,1}(\mathbf{S})$ ,*

$$\langle \xi_t, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \xi_s, \partial_a f + \Phi_s^f(\xi_s) \rangle ds, \quad (2.23)$$

where

$$\begin{aligned} \Phi_t^f(\chi)(x, a) &= \partial_a f(x, a) + b(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^b \rangle) \int_E f(x', 0) k_b(x, a, t, Y_t, x') m(dx') \\ &- d(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^d \rangle) f(x, a) + e(x, a, t, Y_t, \langle \chi, U_{\mathbf{x}}^e \rangle) \int_E (f(x', a) - f(x, a)) k_e(x, a, t, Y_t, x') m(dx'). \end{aligned} \quad (2.24)$$

The proof can be adapted from Fournier and Méléard (2004) and Tran (2006). We detail below the main steps.

### Main steps for the proof of Theorem 2.3

- (a) The first step of the proof is dedicated to the tightness of the sequence of processes  $(Z^n)_{n \in \mathbb{N}^*}$  in  $\mathbb{D}([0, T], (\mathcal{M}(\mathbf{S}), w))$ . This mainly relies on the fact that moments are controlled due to bounded demographic rates (see point (i) of Assumption 2.2).
- (b) The tightness of the sequence then allows to extract from each subsequence a further subsequence converging in distribution; the second step aims at proving that its limit is a solution to Equation (2.13). Assumption 2.3 is used, as well as the crucial Assumption 2.4 allowing to get the convergence result with uniform integrability of a rescaled martingale sequence, which is analogous to the martingale



part in Equation (2.11).

(c) The last step is dedicated to show that there is a unique solution  $(\xi_t)$  to Equation (2.13): this is achieved using the boundedness of the demographic rates as well as their Lipschitz property with respect to the interaction kernel (see Assumption 2.3). As a result, any subsequence of  $(Z^n)_{n \in \mathbb{N}^*}$  contains a further subsequence which converges in distribution to  $\xi$ : this proves the convergence in distribution (see Billingsley (2009), Theorem 2.6).

# Chapter 3

## Population viewpoint on Hawkes processes<sup>†</sup>

### Abstract

This paper focuses on a class of linear Hawkes processes with general immigrants. These are counting processes with shot noise intensity, including self-excited and externally excited patterns. For such processes, we introduce the concept of age pyramid which evolves according to immigration and births. The virtue of this approach that combines an intensity process definition and a branching representation is that the population age pyramid keeps track of all past events. This is used to compute new distribution properties for a class of linear Hawkes processes with general immigrants which generalize the popular exponential fertility function. The pathwise construction of the Hawkes process and its underlying population is also given.

**Keywords:** Hawkes processes, branching, immigration, age pyramid, non-stationarity, laplace transform, thinning, Poisson point measure.

### 3.1 Introduction

This paper investigates the link between some population dynamics models and a class of Hawkes processes. We are interested in processes whose behavior is modified by past events, which are self-excited and externally excited. The introduction of a self-excited process with shot noise intensity is due to Hawkes (1971) and the famous Hawkes process has been used until now for a variety of applications, including seismology, neuroscience, epidemiology, insurance and finance, to name but a few.

---

<sup>†</sup>Boumezoued, A. 2015. Population viewpoint on Hawkes processes. Shorter version v2 available in HAL preprint Id: hal-01149752, To appear in Advances in Applied Probability 48.2 (June 2016)

The shot noise intensity of the Hawkes process  $(N_t)$  is expressed as

$$\lambda_t = \mu + \sum_{T_n < t} \phi(t - T_n),$$

where the  $T_n$  are the times of jump of the Hawkes process  $N$  itself,  $\mu > 0$  and  $\phi$  is a non-negative function. In the Hawkes model, when an event occurs at time  $T_n$ , the intensity grows by an amount  $\phi(t - T_n)$ : this models the self-exciting property. Also, for many modeling purposes,  $\phi$  returns to zero as  $t$  increases, so that the self-excitation vanishes after a long time. On the whole, each event excites the system as it increases its intensity, but this increase vanishes with time as it is natural to model the fact that very old events have a negligible impact on the current behavior of the process. In the literature, more recent contributions focused on processes with self-exciting behavior and also some externally exciting component. To our knowledge, the Hawkes process with general immigrants has been introduced in Brémaud and Massoulié (2002), and specific forms can also be found in recent studies motivated by financial applications, such as Dassios and Zhao (2011), Wheatley et al. (2014) and Rambaldi et al. (2014), where external shocks, news arrivals and contagion are crucial to model. In this paper, we are interested in a class of Hawkes processes with general immigrants (see Brémaud and Massoulié (2002)), whose intensity is of the form

$$\lambda_t = \mu(t) + \sum_{T_n < t} \Phi_t(t - T_n, X_n) + \sum_{S_k < t} \Psi_t(t - S_k, Y_k).$$

In this model, the  $T_n$  are the times of jump of  $N$ : if an event occurs for the system at time  $T_n$ , the intensity grows by an amount  $\Phi_t(t - T_n, X_n)$ , where  $X_n$  is some mark. This part models the self-exciting property. In parallel, external events occur at times  $S_k$  and excites the system of interest with some amount  $\Psi_t(t - S_k, Y_k)$ : this is the externally excited component.

Among the appealing properties of such models, one of them comes from the shot-noise form of the intensity. This is called the cluster (or branching) representation of the Hawkes process, and it is based on the following remark: if an event occurred at time  $T_n$ , then  $t - T_n$  is nothing but the "age" of this event at time  $t$ . Few years later after the seminal work of Hawkes (1971), Hawkes and Oakes (1974) proposed the cluster representation of the self-exciting process. They interpreted it as an immigration-birth process with age: they proved that under some stationarity conditions, it can be described as a branching Poisson process (also called Poisson cluster). Also, in Dassios and Zhao (2011), a definition of a dynamic contagion process is given through its cluster representation. Until now, most studies on the Hawkes process recalled the immigration-birth representation as follows: immigrants arrive at times given by a Poisson process with intensity  $\mu$ . Then each immigrant starts a new generation: it gives birth to new individuals with fertility function

$\phi$ , each one giving birth with same fertility function  $\phi$ . This is often used as a definition for the Hawkes process, providing a good intuition on its behavior. The cluster representation of Hawkes and Oakes (1974) requires that the mean number of children per individual which is nothing but  $\|\phi\| = \int_0^\infty \phi(a)da$  satisfies  $\|\phi\| < 1$ . In our paper, we exhibit the underlying immigration-birth dynamics which does not require the stationary assumption. The virtue of this approach that combines an intensity process definition and a branching representation is that the population age pyramid keeps track of all past events. This is used to compute new distribution properties for a class of linear Hawkes process with general immigrants.

In the literature, the distribution properties of the Hawkes process have first been studied under stationary conditions. Hawkes (1971) addressed second order stationary properties, whereas Adamopoulos (1975) derived the probability generating functional under stationarity, by using the cluster representation of Hawkes and Oakes (1974). In this work, Adamopoulos (1975) expressed the probability generating function as a solution to some functional equation. Furthermore, Brémaud and Massoulié (2002) introduced the framework for studying moments of the stationary Hawkes process by means of the Bartlett spectrum. Let us also mention two recent studies of the distribution properties under stationarity. The moment generating function has been expressed in Saichev and Sornette (2011) as a solution to some transcendental equation. In addition, Jovanović et al. (2014) proposed a graphical way to derive closed form expressions for cumulant densities, leading to the moments of the stationary Hawkes process. It is interesting to note that such recent contributions rely the stationary branching representation of Hawkes and Oakes (1974). Recently, the computation of statistical properties has gained attention under non-stationarity, both for mathematical analysis and statistical estimation techniques. However, the recent studies in this framework only focus on exponential fertility rates  $\phi(t) = \alpha e^{\beta t}$ . The tool they rely on is the infinitesimal generator of the intensity process  $(\lambda_t)$  which is Markovian for such exponential fertility rate (see Oakes (1975)). This includes the work of Errais et al. (2010), Aït-Sahalia et al. (2010), Dassios and Zhao (2011), and Da Fonseca and Zaatour (2014). Our paper generalizes these studies in a natural direction for a wider class of Hawkes processes.

**Scope of this paper** The aim of this paper is (i) to introduce the concept of age pyramid for general Hawkes processes and study its dynamics over time, (ii) to use this concept to compute new distribution properties for a class of fertility functions which generalize the popular exponential case, and (iii) to give a pathwise representation of the general Hawkes processes and its underlying immigration-birth dynamics. We represent the population as a multi-type dynamics with ages, includ-

ing immigration and births with mutations. Our population point of view that introduces the concept of age pyramid is inspired by Bensusan et al. (2010–2015) (see also Tran (2008)). As highlighted in Bensusan et al. (2010–2015), the key idea is that the population structure in terms of ages and characteristics, which keeps track of past events, provides much more information than the intensity itself and allows to study the whole system. In this way, we address the computation of distribution properties of the Hawkes process with general immigrants for a wide class of time-dependent fertility functions. We also give the pathwise construction of the age pyramid represented as a measure-valued process solution to a stochastic equation driven by a Poisson point measure, which is the theoretical counterpart of the thinning numerical procedure. Our approach seems to reconcile the two definitions of Hawkes processes, through an intensity process or a branching dynamics.

The paper is organized as follows. Section 3.2 focuses on the standard Hawkes process with time-independent fertility function. On this particular case, we give the population point of view and study the dynamics of the age pyramid over time. In Section 3.3, we use this concept to compute new distribution properties as moments and Laplace transform for a class of Hawkes processes which generalizes the popular exponential case. Section 3.4 details the pathwise construction of the standard Hawkes process and its underlying population. Our general population representation and results are given in Section 3.5, where we focus on Hawkes processes with general immigrants. In particular, we derive its dynamics and Laplace transform for a wide class of time-dependent fertility functions.

## 3.2 Population point of view

The definition of the (standard linear) Hawkes process through its intensity is given below. Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space satisfying the usual conditions. Recall that the intensity process  $(\lambda_t)$  of a counting process  $(N_t)$  is the  $(\mathcal{F}_t^N)$ -predictable process such that  $N_t - \int_0^t \lambda_s ds$  is an  $(\mathcal{F}_t^N)$ -local martingale, where  $(\mathcal{F}_t^N)$  denotes the canonical filtration of  $(N_t)$ .

**Definition 3.1.** *Let  $\phi$  be a continuous and non-negative map. A Hawkes process  $(N_t)$  with kernel  $\phi$  is a counting process with canonical filtration  $(\mathcal{F}_t^N)$  which admits an  $(\mathcal{F}_t^N)$ -predictable intensity*

$$\lambda_t = \mu + \sum_{T_n < t} \phi(t - T_n) = \mu + \int_{(0,t)} \phi(t - s) dN_s, \quad (3.1)$$

where  $\mu > 0$ , and the  $(T_n)$  are the times of jump of  $(N_t)$ .

The previous definition provides the representation of the intensity process, which is interesting in order to study the behavior of the Hawkes process. But

in fact, the whole information on the dynamics is lost. Indeed, it is interesting to go back to the branching representation of Hawkes and Oakes (1974) to have in mind the underlying population dynamics. First, immigrants arrive according to some Poisson process with parameter  $\mu$ . Then each immigrant generates a cluster of descendants with the following rule: if an individual arrived or was born at some time  $s$ , it gives birth to new individuals with rate  $\phi(t - s)$  at time  $t$ . Note that in fact,  $t - s$  is nothing but the age at time  $t$  of the individual born at time  $s$ . The birth mechanism can thus be reformulated as: any individual with age  $a$  in the population gives birth with rate  $\phi(a)$ . The whole dynamics describes an immigration-birth process with age, in which the immigration rate is  $\mu$  and the birth rate is  $\phi(a)$ .

Since the immigration-birth mechanism is crucial to understand the Hawkes dynamics, the aim now is to keep track of all ages in the population. One way to address this issue is to count the number of individuals with age below  $\bar{a} > 0$  at time  $t$ , denoted  $Z_t([0, \bar{a}])$ . This can be computed as the number of individuals arrived until time  $t$  without those arrived before  $t - \bar{a}$ , that is

$$Z_t([0, \bar{a}]) = N_t - N_{t-\bar{a}} = \int_{(0,t]} \mathbf{1}_{t-s \leq \bar{a}} dN_s.$$

The previous equation shows that with fixed  $t$ , this defines a measure on the space  $\mathbb{R}_+$  of ages that is an image of the jump measure  $dN_t$ . It can be written as

$$Z_t(da) = \int_{(0,t]} \delta_{t-s}(da) dN_s = \sum_{n=1}^{N_t} \delta_{t-T_n}(da). \quad (3.2)$$

Note that  $Z_t(da)$  charges only  $[0, t]$  since no individual born after time 0 can reach an age greater than  $t$ . Formally, the measure  $Z_t(da)$  puts a weight on the age of each individual alive at time  $t$ , therefore we call it age pyramid in reference to demographic analysis. In general, demographic studies focus on the number of individuals per age class of e.g. one year, so the quantity of interest is e.g.  $Z_t([a, a + 1))$ . The virtue of the measure representation is that one can compute a function  $f$  of the population age structure by integrating it with respect to the age pyramid. To do this, we use the notation

$$\langle Z_t, f \rangle = \int_{\mathbb{R}_+} f(a) Z_t(da) = \int_{(0,t]} f(t-s) dN_s. \quad (3.3)$$

For example, the Hawkes process can be computed as  $N_t = \langle Z_t, \mathbf{1} \rangle$ . Also, the intensity process defined in Equation (3.1) can be rewritten using (3.3) as

$$\lambda_t = \mu + \langle Z_{t-}, \phi \rangle.$$

The intensity is the sum of the migration intensity  $\mu$  and the individual birth intensities: this is indeed the intensity of an immigration-birth process with migration rate  $\mu$  and birth rate  $\phi(a)$ , in which all individuals behave independently. Viewed as

a stochastic process,  $(Z_t(da))_{t \geq 0}$  is a measure-valued process. In fact, this age pyramid process, that is the measure-valued process  $(Z_t(da))_{t \geq 0}$ , is a Markov process (see Tran (2006)). Note however that its differentiation in time is not straightforward (see Bensusan et al. (2010–2015) and Lemma 3.1 below). The Markov property of the age pyramid process shows that all the information needed is contained in the population age structure. Let us mention the seminal point of view of Harris (1963), for who "it does seem intuitively plausible that we obtain a Markov process, in an extended sense, if we describe the state of the population at time  $t$  not simply by the number of objects present but by a list of the ages of all objects." However, in practice this information is "too large" to perform tractable computations. In the next section, we illustrate how to identify some minimal components to add to the Hawkes process in order to make the dynamics Markovian. To do this, we first need to address the time evolution of the age pyramid. The following lemma details the dynamics of  $\langle Z_t, f \rangle$  in the case where  $f$  is differentiable. This is the key tool for our results in Section 3.3.

**Lemma 3.1.** *For each differentiable  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,*

$$\langle Z_t, f \rangle = f(0)\langle Z_t, \mathbf{1} \rangle + \int_0^t \langle Z_s, f' \rangle ds. \quad (3.4)$$

**Proof of Lemma 3.1.** Let us write between  $s$  and  $t$ ,  $f(t-s) = f(0) + \int_s^t f'(u-s) du$  and use it into Equation (3.3) to get  $\langle Z_t, f \rangle = f(0)\langle Z_t, \mathbf{1} \rangle + \int_0^t \left( \int_s^t f'(u-s) du \right) dN_s$ . By Fubini's theorem, the last term of the sum is equal to  $\int_0^t \left( \int_0^u f'(u-s) dN_s \right) du$ , and by Equation (3.3), this is equal to  $\int_0^t \langle Z_u, f' \rangle du$ . This concludes the proof.  $\diamond$

The decomposition (3.4) is classical in the field of measure-valued population dynamics (see Tran (2008) and Bensusan et al. (2010–2015)). The first term refers to the pure jump part of arrivals of individuals with age 0, whereas the second term of transport type illustrates the aging phenomenon: all ages are translated along the time axis. In particular, this shows why the intensity process  $\lambda_t = \mu + \langle Z_{t-}, \phi \rangle$  is Markovian in the case where the fertility function is exponential (see Oakes (1975)), that is  $\phi(a) = \alpha e^{\beta a}$ . In this case,  $\phi' = \beta \phi$ , and Equation (3.4) with  $f \equiv \phi$  leads to the differential form

$$d\langle Z_t, \phi \rangle = \alpha dN_t + \beta \langle Z_t, \phi \rangle dt.$$

Note that  $dN_t$  only depends on the past of  $(\lambda_t)$  by means of the current value  $\lambda_t$ , which proves the Markov property. This remark is the starting point of our study, which extends the exponential case in a natural setting.

### 3.3 The exponential case generalized

In this section, the aim is to use the concept of age pyramid process introduced in Section 3.2 in order to compute several distribution properties for the non-stationary Hawkes process. In particular, we provide ordinary differential equations for first and second order moments and the Laplace transform. All computations are performed under some assumption on the birth rate  $\phi$  which naturally extends the popular exponential case.

#### 3.3.1 Assumption on the birth rate

**Assumption 3.1.** *The map  $a \in \mathbb{R}_+ \mapsto \phi(a)$  is non-negative, of class  $\mathcal{C}^n(\mathbb{R}_+)$ , and there exists  $c = (c_{-1}, \dots, c_{n-1}) \in \mathbb{R}^{n+1}$  such that  $\phi$  satisfies*

$$\phi^{(n)} = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}, \quad (3.5)$$

with initial conditions  $\phi^{(k)}(0) = m_k$ , for  $0 \leq k \leq n - 1$ .

The birth rates that satisfy Assumption 3.1 include the exponential case but also some fertility functions that are interesting for a variety of applications. Let us introduce the vector

$$m = (1, m_0, \dots, m_{n-1})^T, \quad (3.6)$$

and the matrix  $C = (C_{i,j})_{-1 \leq i, j \leq n-1}$  given by  $C_{i,i+1} = 1$  for  $0 \leq i \leq n - 2$  and  $C_{n-1,j} = c_j$  for  $-1 \leq j \leq n - 1$ , all other components being zero. Since it is fully determined by the vector  $c$ , we denote

$$C(c) = \begin{pmatrix} 0 & 0 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & 0 & 1 & \\ c_{-1} & c_0 & \cdots & c_{n-2} & c_{n-1} & \end{pmatrix}. \quad (3.7)$$

Equation (3.5) can be rewritten  $\Phi' = C\Phi$  where  $\Phi = (1, \phi, \dots, \phi^{(n-1)})^T$ , whose solution is given by  $\Phi(a) = e^{aC}m$ . Then  $\phi$  can be recovered as the second component of the matrix  $\Phi(a)$ . In particular, if the polynomial  $P(y) = y^n - \sum_{k=0}^{n-1} c_k y^k$  is split with distinct roots  $y_1, \dots, y_p$  and corresponding multiplicity  $n_1, \dots, n_p$ , then  $\phi$  can be written up to some constant as  $\sum_{i=1}^p P_i(a) e^{y_i a}$  where  $P_i$  is a polynomial with degree at most  $n_i - 1$ . This is a sufficiently large set of functions to approximate any fertility function outside of the range of Assumption 3.1. As an example, the power law kernel is of importance for many applications. In the context of earthquakes, the Omori law describes the epidemic-type aftershock (ETAS) model: it corresponds to a specific



form  $\phi(a) \sim \frac{K}{a^{1+\epsilon}}$ . Also in the field of financial microstructure, recent studies (see e.g. Hardiman et al. (2013)) found that high-frequency financial activity is better described by a Hawkes process with power law kernel rather than exponential. The power law kernel with cut off can be approximated as in Hardiman et al. (2013) up to a constant by the smooth function

$$\phi(a) = \sum_{i=0}^{M-1} \frac{e^{-a/(\tau_0 m^i)}}{(\tau_0 m^i)^{1+\epsilon}} - S e^{-a/(\tau_0 m^{M-1})},$$

where  $S$  is such that  $\phi(0) = 0$ . In general, one can use approximation theory to construct a sequence of fertility functions which tends to the original one. As a result, this constructs a sequence of Hawkes processes that approximate the original Hawkes process.

### 3.3.2 Dynamics

Let us go back to the dynamics of the age pyramid over time. The key property that will allow us to compute distribution properties is that the population enables to identify the components to add to the Hawkes process and its intensity to make the dynamics Markovian. This is stated in the following proposition.

**Proposition 3.1.** *Under Assumption 3.1, the process  $X_t = (\langle Z_t, 1 \rangle, \langle Z_t, \phi \rangle, \dots, \langle Z_t, \phi^{(n-1)} \rangle)^T$  satisfies the dynamics*

$$X_t = N_t m + \int_0^t C X_s ds, \quad (3.8)$$

where we the vector  $m$  and the matrix  $C$  are given in (3.6) and (3.7) respectively. In particular,  $X$  is a Markov process.

**Proof of Proposition 3.1.** Let us use Lemma 3.1 to get for  $0 \leq k \leq n-1$ , with  $f \equiv \phi^{(k)}$ ,

$$\langle Z_t, \phi^{(k)} \rangle = m_k N_t + \int_0^t \langle Z_s, \phi^{(k+1)} \rangle ds. \quad (3.9)$$

By Assumption 3.1, we get in particular

$$\langle Z_t, \phi^{(n-1)} \rangle = m_{n-1} N_t + \sum_{k=-1}^{n-1} c_k \int_0^t \langle Z_s, \phi^{(k)} \rangle ds, \quad (3.10)$$

with convention  $\phi^{(-1)} \equiv \mathbf{1}$ . This implies the dynamics (3.8) which also shows that  $X$  is a Markov process.  $\diamond$

The dynamics (3.8) for the  $(n+1)$ -dimensional vector  $X_t$  gives a set of  $n$  equations, the first component of  $X$ , which is the Hawkes process  $N$ , being free. In

Section 3.4 we will give an equation on the Hawkes process  $N$  by means of stochastic representation based on Poisson point measures. This will provide a full system of equations for the components of  $X$  as well as a pathwise representation. For now, we are interested into deriving several distribution properties of the Hawkes process and its additional components in  $X$ .

### 3.3.3 Moments

**First order moments** The differential system of Equation (3.8) is linear and allows us to propose a straightforward differential equation for the first order moment. We also perform explicit computations for small dimensions  $n = 1$  and  $n = 2$ .

**Proposition 3.2.** *Under Assumption 3.1, the vector map  $u(t) := \mathbb{E}[X_t]$  is solution to*

$$u'(t) = \mu m + Au(t), \quad (3.11)$$

where the  $(n + 1) \times (n + 1)$  matrix  $A$  is given by

$$A = C + mJ, \quad (3.12)$$

where

$$J = (0, 1, 0, \dots, 0), \quad (3.13)$$

and the vector  $m$  and the matrix  $C$  are given in (3.6) and (3.7) respectively.

**Proof of Proposition 3.2** *Let us use the martingale property of the compensated counting process, then use Fubini's theorem and the fact that Lebesgue measure charges no point to get  $\mathbb{E}[N_t] = \int_0^t (\mu + \mathbb{E}[\langle Z_s, \phi \rangle]) ds$ . Now, let us take expectation in (3.8) and use the previous formula to get Equation (3.11).  $\diamond$*

The differential equation (3.11) allows to get explicit formulas for the expected number of events. We recall the first order moment for the popular exponential case  $\phi(a) = e^{-ca}$  (see e.g. Dassios and Zhao (2011)) and also give the explicit formulas for the birth rate  $\phi(a) = \alpha^2 a e^{-\beta a}$ . Note that this case can be useful for a variety of applications to model a smooth delay at excitation. Remark also the different behavior of the first order moments, in particular in the critical case  $\int_0^\infty \phi(a) da = 1$ , which corresponds to  $c = 1$  and  $\alpha = \beta$ . For the two examples given below, the computations are left to the reader.

**Corollary 3.1.** *For the Hawkes process with  $\phi(a) = e^{-ca}$ ,  $c > 0$ , ( $n = 1$  in Assumption 3.1),*

$$\mathbb{E}[N_t] = \mu \left( t + \frac{t^2}{2} \right) \quad \text{if } c = 1,$$

$$\mathbb{E}[N_t] = \frac{\mu}{1-c} \left( \frac{e^{(1-c)t} - 1}{1-c} - ct \right), \quad \text{if } c \neq 1.$$

**Corollary 3.2.** *For the Hawkes process with  $\phi(a) = \alpha^2 a e^{-\beta a}$ ,  $\alpha, \beta > 0$ , ( $n = 2$  in Assumption 3.1),*

$$\begin{aligned}\mathbb{E}[N_t] &= \frac{\mu}{8\beta} (1 - e^{-2\beta t}) + \frac{3\mu}{4}t + \frac{\beta\mu}{4}t^2, \quad \text{if } \alpha = \beta, \\ \mathbb{E}[N_t] &= \frac{\mu\beta^2}{\beta^2 - \alpha^2}t + \frac{\alpha\mu}{2} \left( \frac{e^{(\alpha-\beta)t} - 1}{(\alpha - \beta)^2} - \frac{e^{-(\alpha+\beta)t} - 1}{(\alpha + \beta)^2} \right), \quad \text{if } \alpha \neq \beta.\end{aligned}$$

**Second order moments** In the following, we derive the dynamics of the process  $(V_t) = (X_t \bar{X}_t)$  where  $X_t := (N_t, \langle Z_t, \phi \rangle, \dots, \langle Z_t, \phi^{(n-1)} \rangle)^T$ . As a consequence, we represent the variance-covariance matrix of the process  $(X_t)$  as the solution to a linear ordinary differential equation. Our method is based on differential calculus with the finite variation process  $(X_t)$  with dynamics (3.8) and could be extended to higher moments.

**Proposition 3.3.** *Let us introduce the matrix  $V_t = X_t \bar{X}_t$ , where  $\bar{X}_t$  denotes the transpose of  $X_t$ . Then the matrix  $V_t$  satisfies the dynamics*

$$dV_t = dN_t (X_{t-} \bar{m} + m \bar{X}_{t-} + m \bar{m}) + dt (V_t \bar{C} + CV_t).$$

*In particular, the variance-covariance matrix  $v(t) = \mathbb{E}[V_t]$  satisfies the following ordinary differential equation:*

$$v'(t) = v(t) \bar{A} + Av(t) + \mu(m \bar{m} + u(t) \bar{m} + m \bar{u}(t)) + Ju(t) m \bar{m}. \quad (3.14)$$

where  $u(t)$  is solution to (3.11) and the matrix  $A$  is defined in (3.12).

**Proof of Proposition 3.3.** Let us use the notation  $X_t = (X_t^{[-1]}, X_t^{[0]}, \dots, X_t^{[n-1]})$ . Integration by parts leads to, for  $-1 \leq l, k \leq n-1$ ,

$$d \left( X_t^{[k]} X_t^{[l]} \right) = X_{t-}^{[k]} dX_t^{[l]} + X_{t-}^{[l]} dX_t^{[k]} + m_k m_l dN_t.$$

The previous equation shows that  $dV_t = X_{t-} d\bar{X}_t + (dX_t) \bar{X}_{t-} + dN_t m \bar{m}$ . By Proposition 3.1 and since Lebesgue measure charges no point we get

$$dV_t = dN_t (X_{t-} \bar{m} + m \bar{X}_{t-} + m \bar{m}) + dt (V_t \bar{C} + CV_t).$$

Recall that  $u(t) = \mathbb{E}[X_t]$ . Now, take expectation in the previous equation to get

$$v'(t) = \mathbb{E} \left[ (\mu + X_t^{[0]}) X_t \right] \bar{m} + m \mathbb{E} \left[ (\mu + X_t^{[0]}) \bar{X}_t \right] + \left( \mu + \mathbb{E}[X_t^{[0]}] \right) m \bar{m} + v(t) \bar{C} + Cv(t).$$

Finally, note that  $X_t^{[0]} X_t = V_t \bar{J}$  where we recall that  $J$  is defined by  $J = (0, 1, 0, \dots, 0)$ , which makes the previous equation reduce to (3.14).  $\diamond$

We give explicit formulas for the popular exponential fertility function  $\phi(a) = e^{-ca}$  and at a higher order for the case  $\phi(a) = \beta^2 a e^{-\beta a}$ , which corresponds to the critical case since the mean number of children per individual satisfies  $\int_0^\infty \phi(a) da = 1$ . Computations are based on the differential equation (3.14) and are left to the reader.

**Corollary 3.3.** *For the Hawkes process with  $\phi(a) = e^{-ca}$  ( $n = 1$  in Assumption 3.1),*

$$\begin{aligned} \text{Var}(N_t) &= \mu t \left( 1 + \frac{3}{2}t + \frac{2}{3}t^2 + \frac{1}{12}t^3 \right) \text{ if } c = 1, \\ \text{Var}(N_t) &= \frac{\mu}{(1-c)^3} \left[ \frac{1-c/2}{1-c} e^{2(1-c)t} + \left( \frac{3c^2-1}{1-c} - 2ct \right) e^{(1-c)t} - c^3 t + \frac{c(1/2-3c)}{1-c} \right], \text{ if } c \neq 1. \end{aligned}$$

**Corollary 3.4.** *For the Hawkes process with  $\phi(a) = \beta^2 a e^{-\beta a}$ , ( $n = 2$  in Assumption 3.1), the variance of the intensity is given by*

$$\text{Var}(\lambda_t) = \beta \mu \left( -\frac{7}{128} + \frac{3\beta}{32}t + \frac{\beta^2}{16}t^2 + \frac{1-\beta t}{8}e^{-2\beta t} - \frac{9}{128}e^{-4\beta t} \right).$$

### 3.3.4 Laplace transform

The aim is to exhibit the exponential martingale associated with the process  $X$  which consequently expresses its Laplace transform in a semi-explicit form. This is given in the following proposition. It is interesting to note that point (i) refers to some forward martingale, whereas point (ii) focuses on backward martingality. Note that based on the Laplace transform, one can recover several quantities of interest. To this aim, numerical illustration is given at the end of this part.

**Proposition 3.4.** *Let us denote  $(\mathcal{F}_t^X)$  the canonical filtration of the process  $X$  and let us work under Assumption 3.1.*

(i) *For any deterministic and differentiable  $A_t$ , the following process is an  $(\mathcal{F}_t^X)$ -martingale:*

$$\exp \left\{ A_t \cdot X_t - \int_0^t A_s \cdot (C X_s) ds - \int_0^t A'_s \cdot X_s ds - \int_0^t (e^{A_s \cdot m} - 1) \lambda_s ds \right\}. \quad (3.15)$$

(ii) *For any  $(n+1)$  real vector  $v$ ,*

$$\mathbb{E}[\exp(v \cdot X_T)] = \exp \left( -\mu \int_0^T (1 - e^{A_s \cdot m}) ds \right), \quad (3.16)$$

*where the vector map  $(A_t)$  satisfies the following non linear differential equation*

$$\bar{C} A_t + A'_t + (e^{A_t \cdot m} - 1) J = 0, \quad (3.17)$$

*with terminal condition  $A_T = v$ . Here,  $v \cdot X_T$  denotes the scalar product between  $v$  and  $X_T$ ,  $J$  is defined in (3.13), and  $\bar{C}$  is the transpose of the matrix  $C$ .*

(iii) *Moreover, there exists a unique solution to Equation (3.17).*

**Proof of Proposition 3.4** The exponential formula states that the following process is a martingale, for any deterministic  $\alpha_s$ ,

$$\exp \left\{ \int_0^t \alpha_s dN_s - \int_0^t (e^{\alpha_s} - 1) \lambda_s ds \right\}. \quad (3.18)$$

Now, by integration by parts and the use of Equation (3.8),

$$\begin{aligned} A_t \cdot X_t &= \int_0^t A_s \cdot dX_s + \int_0^t A'_s \cdot X_s ds \\ &= \int_0^t A_s \cdot m dN_s + \int_0^t A_s \cdot (CX_s) ds + \int_0^t A'_s \cdot X_s ds. \end{aligned}$$

Then by Equation (3.18) with  $\alpha_s = A_s \cdot m$ , the process in (3.15) is a martingale. To prove the second point, the aim is to find a martingale of the form  $\exp\{A_t \cdot X_t + D(t)\}$  with some deterministic  $D(t)$  and a terminal condition  $A_T = v$ . To do this, let us choose  $A$  such that the random part in the integrant in (3.15) vanish. Since  $\lambda_s = \mu + \langle Z_{s-}, \phi \rangle$ , this amounts to get for each vector  $X = (X^{[-1]}, \dots, X^{[n-1]})$ ,

$$\bar{C}A_t \cdot X + A'_t \cdot X + (e^{A_t \cdot m} - 1)(\mu + X^{[0]}) = 0. \quad (3.19)$$

Let us now identify the term in  $X$ , leading to the equation for  $A$ :  $\bar{C}A_t \cdot X + A'_t \cdot X + (e^{A_t \cdot m} - 1)X^{[0]} = 0$ , that is  $\bar{C}A_t + A'_t + (e^{A_t \cdot m} - 1)J = 0$ , where  $J$  is defined in (3.13). If we set terminal condition  $A_T = v$ , we get Equation (3.16). Finally, existence and uniqueness for Equation (3.17) arises from Cauchy-Lipschitz theorem, since the map  $Y \mapsto \bar{C}Y + (e^{Y \cdot m} - 1)J$  is of class  $\mathcal{C}^1$  on  $\mathbb{R}^{n+1}$ , and thus continuous and locally Lipschitz.  $\diamond$

The previous result can be expressed in terms of a single function, and is derived below for the Hawkes process and its intensity. The proof is given in Appendix A.

**Corollary 3.5.** *Under Assumption 3.1, the joint Laplace transform of the Hawkes process and its intensity is given for each real  $\theta_1$  and  $\theta_2$  by*

$$\mathbb{E} [\exp (\theta_1 N_T + \theta_2 \lambda_T)] = \exp \left\{ -\mu \left( (-1)^n G^{(n)}(0) + \sum_{k=0}^{n-1} (-1)^{k+1} c_k G^{(k)}(0) \right) \right\},$$

where the function  $G$  satisfies the non-linear ordinary differential equation: for each  $0 \leq t \leq T$ ,

$$(-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) + \exp \left( \theta_1 - c_{-1} G(t) + \sum_{k=0}^{n-1} b_k G^{(k+1)}(t) \right) - 1 = 0,$$

with terminal conditions  $G^{(k)}(T) = 0$  for  $0 \leq k \leq n-1$  and  $G^{(n)}(T) = (-1)^{n-1} \theta_2$ ,

$$(3.20)$$

and for  $0 \leq k \leq n-1$ ,  $b_k = (-1)^k (m_{n-1-k} - \sum_{l=k+1}^{n-1} m_{n-1-l} c_{n-l+k})$ .

**Numerical example** Let us illustrate numerically Corollary 3.5 for the computation of the generating functional  $\mathbb{E}[u^{N_T}]$ , interpreted as the survival probability at time  $T$  of a system which survives with probability  $u$  at each shock, as well as quantities as the probability to get exactly  $k$  shocks until time  $T$  as

$$\mathbb{P}(N_T = k) = \frac{1}{k!} \partial_u^k \mathbb{E}[u^{N_T}]|_{u=0}.$$

Setting  $u = e^{\theta_1}$  and  $\theta_2 = 0$ , an explicit discretization scheme has been used to solve the non-linear differential equation satisfied by  $G$  and the differentiation step for the derivatives of the generating functional has been chosen carefully. The results for the two critical cases  $\phi(a) = e^{-a}$  (case 1) and  $\phi(a) = ae^{-a}$  (case 2) are described in Tables 3.1 and 3.2 with three significative numbers. Note that even if the mean number of children per individual is one in each case, the results are different due to the shape of each birth rate  $\phi$ . This promotes the use of many kernels, beyond the exponential case. This approach gives a way of computing such quantities (numerically approximated) which is a faster alternative compared to Monte Carlo simulations. To conclude this numerical experiment, we emphasize that the computation of  $\mathbb{P}(N_t = k)$  for higher values of  $k$  will require more stable numerical differentiation methods, and are therefore beyond the scope of the present paper.

$u$	0.1	0.3	0.5	0.7	0.9
Case 1, $\phi(a) = e^{-a}$	0.490	0.532	0.588	0.672	0.828
Case 2, $\phi(a) = ae^{-a}$	0.494	0.546	0.615	0.714	0.874

Table 3.1: Computed values of  $\mathbb{E}[u^{N_T}]$  with  $\mu = 0.15$  and  $T = 5$ .

$k$	0	1	2	3	4
Case 1, $\phi(a) = e^{-a}$	0.472	0.165	0.0894	0.0577	0.0407
Case 2, $\phi(a) = ae^{-a}$	0.472	0.203	0.113	0.0700	0.0451

Table 3.2: Computed values of  $\mathbb{P}(N_T = k)$  with  $\mu = 0.15$  and  $T = 5$ .

### 3.4 Pathwise representation of Hawkes population

The aim of this section is to detail the pathwise construction of the Hawkes process and its underlying population. This is done by means of stochastic differential equations driven by Poisson point measures. The virtue of this approach is that it seems to reconcile both the intensity process definition of the Hawkes process and its branching representation. We first describe the construction of the Hawkes process with reference to a Poisson measure, then exhibit the system of equations driving the

generalized exponential case, and finally address the pathwise population dynamics for general birth rates.

**Construction of the Hawkes process** One question which arises with Definition 3.1 refers to the construction of such process and the notion of pathwise uniqueness. An answer can be given by the thinning representation, which works as follows. Consider a Poisson point measure  $Q(ds, d\theta)$  with intensity measure  $q(ds, d\theta) = dsd\theta$  on  $\mathbb{R}_+ \times \mathbb{R}_+$  (see e.g. Çınlar (2011) for a definition), and denote  $(\mathcal{F}_t^Q)$  the canonical filtration generated by  $Q$ . Note that the intensity measure  $q$  is not finite and only  $\sigma$ -finite, which makes it impossible to order the points in time of the Poisson point measure  $Q$ . But this flexible representation allows to represent a wide class of counting processes. Let  $(\lambda_t)$  be a  $(\mathcal{F}_t^Q)$ -predictable process such that a.s. for each  $t > 0$ ,  $\int_0^t \lambda_s ds < +\infty$ . Then the following process  $(N_t)$  is a counting process with  $(\mathcal{F}_t^Q)$ -predictable intensity  $\lambda_t$ :

$$N_t = \int_{(0,t]} \int_{\mathbb{R}_+} \mathbf{1}_{[0,\lambda_s]}(\theta) Q(ds, d\theta).$$

Indeed,  $N$  is clearly a counting process because each atom of  $Q$  is weighted 1 or 0. Also, since a.s.  $\int_0^t \lambda_s ds < +\infty$ , the martingale property for Poisson point measures ensures that  $N_t - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{[0,\lambda_s]}(\theta) d\theta ds = N_t - \int_0^t \lambda_s ds$  is a  $(\mathcal{F}_t^Q)$ -local martingale, which shows that the counting process  $(N_t)$  has intensity  $(\lambda_t)$ . Now, let us describe the construction of the Hawkes process. Since the intensity in (3.1) is given as a particular form of the process itself, the idea is to define the Hawkes process as the solution to the stochastic equation

$$N_t = \int_{(0,t]} \int_{\mathbb{R}_+} \mathbf{1}_{[0,\mu + \int_{(0,s)} \phi(s-u) dN_u]}(\theta) Q(ds, d\theta). \quad (3.21)$$

General results about existence and uniqueness for the Hawkes process (even non-linear) can be found in Brémaud and Massoulié (1996) and Massoulié (1998) (see also Delattre et al. (2014) and the books of Daley and Vere-Jones (2008) and Çınlar (2011)). The thinning method to represent a counting process as the solution of a stochastic equation is in fact classical. This general mathematical representation goes back to Kerstan (1964) and Grigelionis (1971). One often refers to the thinning algorithms that have been proposed by Lewis and Shedler (1978) and Ogata (1981), which are very useful to perform numerical simulations for quite complex intensity processes. A first advantage of the thinning formulation arises when one wants to show the existence of the Hawkes process. This is done by Picard iteration method (see Massoulié (1998)): one constructs a sequence  $(N^k)_{k \geq 0}$  of counting processes starting at  $N^0 \equiv 0$ , and for  $k \geq 0$ ,

$$N_t^{k+1} = \int_{(0,t]} \int_{\mathbb{R}_+} \mathbf{1}_{[0,\mu + \int_{(0,s)} \phi(s-u) dN_u^k]}(\theta) Q(ds, d\theta). \quad (3.22)$$

One can show that the sequence  $(N^k)$  is Cauchy and thus converges to the desired process. Moreover, another advantage is to give strong uniqueness. With this issue, it appears that the thinning representation has the virtue to use "one noise once for all" and thus give pathwise construction and results. This is interesting to note that this approach is used by Delattre et al. (2014) to show existence and uniqueness of an infinite graph of interacting Hawkes processes. Due to the pathwise representation and the iterative construction, one can also identify each generation in the dynamics. Indeed, one sees in the construction of the Cauchy sequence in (3.22) that  $N^1$  counts the number of immigrants, whereas  $N^2 - N^1$  counts the children of immigrants,  $N^3 - N^2$  the grandchildren of immigrants, and so on. Generally,  $N_t^{k+1} - N_t^k$  is the number of individuals in generation  $k$  born before time  $t$ . This shows another advantage of the pathwise construction: what is called the "thinning parameter"  $\theta$  gives additional information on the dynamics, making it possible in particular to study each generation separately. Before giving the representation of the age pyramid, we first go back to the extension of the exponential case.

**Exponential case generalized** We first address the particular case where the birth rate  $\phi$  satisfies Assumption 3.1. The dynamics of the  $(n + 1)$ -dimensional vector  $X_t := (N_t, \langle Z_t, \phi \rangle, \dots, \langle Z_t, \phi^{(n-1)} \rangle)^T$  is given in (3.8) by  $dX_t = dN_t m + C X_t dt$ . This gives in fact  $n$  equations, the first coordinate  $N_t$  being free. The pathwise representation (3.21) allows to derive the full system of equations by

$$dX_t = \int_{\mathbb{R}_+} m \mathbf{1}_{[0, \mu + X_t^{[0]}]}(\theta) Q(dt, d\theta) + C X_t dt,$$

where we recall the notation  $X_t = (X_t^{[-1]}, X_t^{[0]}, \dots, X_t^{[n-1]})$ .

**Immigration-birth process with general fertility functions** In the case where the birth rate is general, one has to represent the whole age pyramid, that is to give the thinning representation of the underlying immigration-birth process. In the field of population dynamics, this approach is used to construct extended birth-death processes with age in particular in Bensusan et al. (2010–2015) (see also Fournier and Méléard (2004) and Tran (2008)). From Equations (3.3) and (3.21), we get the pathwise representation

$$Z_t(da) = \int_{(0,t]} \int_{\mathbb{R}_+} \mathbf{1}_{[0, \mu + \langle Z_{s-}, \phi \rangle]}(\theta) \delta_{(t-s)}(da) Q(ds, d\theta). \quad (3.23)$$

This illustrates the fact that the population at time  $t$  is nothing but all individuals that arrived before time  $t$  (immigration or birth); if an individual arrived at time  $s$ , its age at time  $t$  is  $t - s$ . Note that in this form, the differentiation is not straightforward (see Bensusan et al. (2010–2015)). But from Lemma 3.1, one can



write the following (infinite) system of equations: for each differentiable  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$d\langle Z_t, f \rangle = f(0) \int_{\mathbb{R}_+} \mathbf{1}_{[0, \mu + \langle Z_{t-}, \phi \rangle]}(\theta) Q(dt, d\theta) + \langle Z_t, f' \rangle dt. \quad (3.24)$$

This approach allows both to keep track of the age pyramid (branching population) and to represent it as a stochastic process with its own intensity (in a pathwise way). This way, it seems to reconcile the intensity process definition of the Hawkes process with its branching representation.

Before going through the last section on the Hawkes process with general immigrants, we briefly discuss the existing cluster representation in the following remark.

**Remark 3.1.** *We recall the definition of the Hawkes process in terms of a Poisson cluster introduced in Hawkes and Oakes (1974) and surveyed in the book of Daley and Vere-Jones (2003). Let  $N_c(ds)$  be a Poisson point measure on  $\mathbb{R}_+$  with intensity measure  $\mu ds$ : this defines the cluster centers, also called ancestors. Let us introduce a family of point processes  $\{\bar{N}(dt | s), s \in \mathbb{R}_+\}$ . For each  $s$ ,  $\bar{N}(dt | s)$  defines the location of the offsprings within the cluster of an ancestor located at  $s$ . The cluster process  $\hat{N}$  counts the number of all offsprings of all immigrants by  $\hat{N}(dt) = \int_{\mathbb{R}_+} \bar{N}(dt | s) N_c(ds)$ . That is, the number of all offsprings up to time  $t$  is given by*

$$\hat{N}([0, t]) = \int_{\mathbb{R}_+} \bar{N}([0, t] | s) N_c(ds).$$

*Thus in the cluster representation, the Hawkes process can be written as the sum of the immigrants and their offsprings by*

$$N_c([0, t]) + \hat{N}([0, t]).$$

*Note that the cluster representation has shown to facilitate the study of the Hawkes process under stationarity by using results in the field of branching processes. Our population representation seems to be the non-stationary counterpart, as it allows us to derive new distribution properties in this framework. Our population representation provides not only the size of the total progeny up to time  $t$ , but also a variety of quantities of interest depending on the population age structure. This has been used in Section 3.3 in order to identify the components needed to make the dynamics Markovian. This will be also used in the following section to study a class of Hawkes process with general immigrants.*

**Remark 3.2.** *In the literature, some link has also been made between marked Hawkes processes and birth-death dynamics. In Appendix B, we rework an example Brémaud and Massoulié (2002) and in addition use some ideas by Garcia and Kurtz (2006) to further detail the way a (marked) Hawkes process can be used to construct a birth-death process.*

## 3.5 Towards more general Hawkes processes

In this section, we focus on a class of counting processes  $N_t$  named as Hawkes processes with general immigrants (see Brémaud and Massoulié (2002)), which is defined below.

**Definition 3.2.** *A Hawkes process with general immigrants is a counting process  $N_t$  whose intensity is given by*

$$\lambda_t = \mu(t) + \sum_{T_n < t} \Phi_t(t - T_n, X_n) + \sum_{S_k < t} \Psi_t(t - S_k, Y_k), \quad (3.25)$$

where the  $T_n$  are the times of jump of  $N$ , the  $S_k$  are the jumps of a counting process with deterministic intensity  $\rho(t)$  and the  $X_n$  (resp.  $Y_k$ ) are real positive iid with distribution  $G$  (resp.  $H$ ). The  $(S_k)$ ,  $(Y_k)$  and  $(X_n)$  are assumed to be independent of each other.

In this model, the  $T_n$  are the times of jump of  $N_t$ : if an event occurs for the system at time  $T_n$ , the intensity grows by an amount  $\Phi_t(t - T_n, X_n)$ , where  $X_n$  is some mark. This part models the self-exciting property. In parallel, external events occur at times  $S_k$  and excites the system of interest with some amount  $\Psi_t(t - S_k, Y_k)$ : this is the externally excited component. The standard Hawkes process that have been studied in the previous sections can be recovered by setting  $\Phi_t(a, x) = \phi(a)$  and  $\Psi_t(a, x) = 0$ . The Hawkes process with general immigrants has been introduced and studied under stationary conditions by Brémaud and Massoulié (2002). Due to their flexibility and natural interpretation, such models have gained recent attention for financial applications e.g. by Dassios and Zhao (2011), Wheatley et al. (2014) and Rambaldi et al. (2014). In particular, distribution properties of such process have been investigated by Dassios and Zhao (2011) in the case  $\Phi_t(a, x) = \Psi_t(a, x) = xe^{-\delta a}$ , in which framework the intensity process is Markovian. The aim of this section is to study the dynamics and characterize the distribution of the non-stationary Hawkes process with general immigrants for a larger class of fertility functions, possibly time-dependent, which extends the previous work of Dassios and Zhao (2011) in this direction. To do this, we use a two-population immigration-birth dynamics with ages and characteristics.

### 3.5.1 Description of the two-population dynamics

The aim is to construct populations of several individuals (or particles), each one having an age  $a$  evolving over time, and a characteristic  $x \in \mathbb{R}_+$ . We construct two populations: the first one represents external shocks, whereas the second one represents events for the Hawkes process.

Each population  $(i)$ ,  $i = 1$  or  $2$ , is represented at time  $t$  as a measure which puts a

weight on the age and characteristic of each individual, denoted  $Z_t^{(i)}(da, dx)$ . The two populations are introduced based on Definition 3.2 as

$$Z_t^{(1)}(da, dx) = \sum_{S_k \leq t} \delta_{(t-S_k, Y_k)}(da, dx) \text{ and } Z_t^{(2)}(da, dx) = \sum_{T_n \leq t} \delta_{(t-T_n, X_n)}(da, dx). \quad (3.26)$$

Since ages but also characteristics of individuals are involved, we prefer to call  $Z_t^{(i)}$  population *structure* rather than *age pyramid*, which is more specific. As for the standard Hawkes population representation, one can compute functions of the whole population structure, which can even depend on time. Consider a function  $f_t(a, x)$  depending on time, and also on age and characteristics of individuals. This can be computed on the overall population by

$$\langle Z_t^{(i)}, f_t \rangle = \int_{\mathbb{R}_+ \times \mathbb{R}_+} f_t(a, x) Z_t^{(i)}(da, dx), \quad (3.27)$$

for  $i = 1$  or  $i = 2$ . For example, the Hawkes process is  $N_t^{(2)} = \langle Z_t^{(2)}, \mathbf{1} \rangle$ . Also, the intensity  $\lambda_t$  of the Hawkes process  $N_t^{(2)}$  given in Equation (3.25) can be rewritten as

$$\lambda_t = \mu(t) + \langle Z_{t-}^{(2)}, \Phi_t \rangle + \langle Z_{t-}^{(1)}, \Psi_t \rangle.$$

This shows that the underlying population dynamics works as follows.

- (i) Let us first describe the population (1) of external shocks. It is made with immigrants that arrive in population (1) with rate  $\rho(t)$ ; at arrival, they have age 0 and some characteristic  $x$  drawn with distribution  $H$ . Any individual  $(a, x)$  at time  $t$  that belongs to population (1) gives birth with rate  $\Psi_t(a, x)$ . The newborn belongs to population (2); it has age 0, and some characteristic drawn with distribution  $G$ .
- (ii) Let us now complete the description of population (2). In addition to births from population (1), the population (2) evolves according to two other kinds of events: immigration and internal birth. Immigrants arrive in population (2) with rate  $\mu(t)$  with age 0 and a characteristic drawn with distribution  $G$ . Any individual  $(a, x)$  at time  $t$  that belongs to population (2) gives birth with rate  $\Phi_t(a, x)$ . The newborn also belongs to population (2); it has age 0, and some characteristic drawn with distribution  $G$ . This dynamics is illustrated in Figure 3.1.

As for our analysis of the standard Hawkes process, a crucial step is to study the dynamics of the population structure over time. That is, what is the dynamics of the process  $\langle Z_t^{(i)}, f_t \rangle$  for  $i = 1$  or  $2$ ? This is stated in the following lemma.

**Lemma 3.2.** *For each function  $f : (t, x, a) \mapsto f_t(a, x)$  differentiable in  $t$  and  $a$ , the dynamics of the process  $\langle Z_t^{(i)}, f_t \rangle$  for  $i = 1$  or  $2$  is given by*

$$d\langle Z_t^{(i)}, f_t \rangle = \int_{\mathbb{R}_+} f_t(0, x) N^{(i)}(dt, dx) + \langle Z_t^{(i)}, (\partial_a + \partial_t) f_t \rangle dt,$$

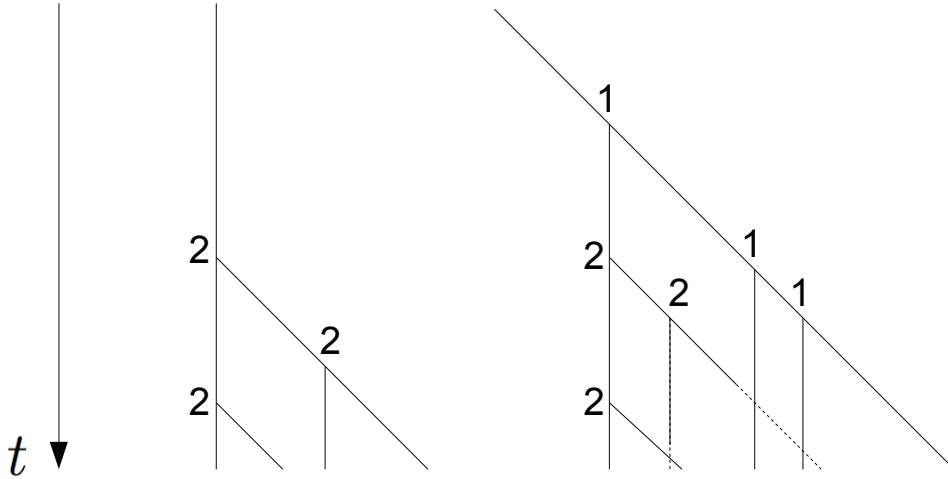


Figure 3.1: Dynamics of the immigration-birth process: immigrants arrive in population (1) (external shocks). Then each individual 1 gives birth to individuals 2 (events due to external shocks). In parallel, immigrants arrive in population 2 (events due to the baseline intensity). Finally, each individual 2 reproduce (self-excitation). The Hawkes process with general immigrants can be recovered as the number of individuals 2.

where the point measures  $N^{(1)}$  and  $N^{(2)}$  are given by

$$N^{(1)}(dt, dx) = \sum_{k \geq 1} \delta_{(S_k, Y_k)}(dt, dx) \text{ and } N^{(2)}(dt, dx) = \sum_{n \geq 1} \delta_{(T_n, X_n)}(dt, dx). \quad (3.28)$$

**Proof of Lemma 3.2** The proof is a straightforward adaptation of that of Lemma 3.1, using (3.26) and (3.27) together with the fact that

$$f_t(t-s) = f_s(0) + \int_s^t (\partial_a + \partial_t) f_u(u-s) du. \diamond$$

Analogously to Lemma 3.1, this result exhibits the pure jump part in the left-hand side, whereas the drift part illustrates the aging term and the time-dependency. The fact that the drift depends on both  $\langle Z_t^{(i)}, \partial_a f_t \rangle$  and  $\langle Z_t^{(i)}, \partial_t f_t \rangle$  is the starting point of our results derived in what follows.

### 3.5.2 Main result

In the following, we introduce the assumptions allowing to recover a finite dimensional Markovian dynamics.

**Assumption 3.2.** (i) *The birth rates  $\Phi$  and  $\Psi$  are non-negative and satisfy  $\Phi_t(a, x) = v(t)\phi(a, x)$  and  $\Psi_t(a, x) = w(t)\psi(a, x)$ , where*

$$\phi^{(n)}(a, x) = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}(a, x) \text{ and } v^{(p)}(t) = d_{-1}(t) + \sum_{l=0}^{p-1} d_l(t)v^{(l)}(t),$$

with  $n, p \geq 1$  and initial conditions  $\phi^{(k)}(0, x) = \phi_0^{(k)}(x)$ , and

$$\psi^{(m)}(a, x) = r_{-1} + \sum_{k=0}^{m-1} r_k \psi^{(k)}(a, x) \text{ and } w^{(q)}(t) = k_{-1}(t) + \sum_{l=0}^{q-1} k_l(t)w^{(l)}(t),$$

with  $m, q \geq 1$  and initial conditions  $\psi^{(k)}(0, x) = \psi_0^{(k)}(x)$ . Note that we used the notation  $f^{(k)}(a, x) = \partial_a^k f(a, x)$ .

(ii) *The maps  $(d_l)_{-1 \leq l \leq p-1}$  and  $(k_l)_{-1 \leq l \leq q-1}$  are continuous.*

**Remark 3.3.** *Assumption 3.2 defines a wide class of self and externally exciting fertility functions of the form  $\Phi_t(a, x) = v(t)\phi(a, x)$ . Let us first focus on the time-independent part and introduce  $F(a, x)$  such that  $F = (1, \phi, \dots, \phi^{(n-1)})^T$ . Then  $F' = CF$ , where  $C$  is defined in (3.7). In particular, if the polynomial  $P(y) = y^n - \sum_{k=0}^{n-1} c_k y^k$  is split with distinct roots  $y_1, \dots, y_p$  and corresponding multiplicity  $n_1, \dots, n_p$ , then  $\phi$  can be written up to some constant as  $\sum_{i=1}^p P_i(x, a)e^{y_i a}$  where  $P_i$  is a polynomial in  $a$  with degree at most  $n_i - 1$  whose coefficients may depend on  $x$ . This includes the framework of Dassios and Zhao (2011) where  $\Phi_t(a, x) = \Psi_t(a, x) = xe^{-\delta a}$ . As we also allow for time-dependency, such birth rates  $\Phi$  and  $\Psi$  that satisfy Assumption 3.2 seem also useful to define non-stationary Hawkes processes, and in particular to include seasonality. As an example, one can simply think of a kernel of the form  $\cos^2(\alpha t)\phi(a, x)$  where  $v(t) = \cos^2(\alpha t)$  satisfies  $v'' = 4\alpha^2(1 - v)$ .*

The aim of this part is to exhibit some exponential martingale which leads us to compute the Laplace transform of the whole dynamics. This provides in particular the joint Laplace transform of the Hawkes process with general immigrants and its intensity. This is the main result of our paper. We first state the following Lemma.

**Lemma 3.3.** *Let us define for  $-1 \leq k \leq n - 1$  and  $-1 \leq l \leq p - 1$ ,  $X_t^{k,l} := \langle Z_t^{(2)}, \partial_a^k \partial_t^l \Phi_t \rangle$  and for  $-1 \leq k \leq m - 1$  and  $-1 \leq l \leq q - 1$ ,  $Y_t^{k,l} := \langle Z_t^{(1)}, \partial_a^k \partial_t^l \Psi_t \rangle$ . Let us also define the two matrices*

$$M_t^{(2)} = \left( X_t^{(k,l)} \right)_{-1 \leq k \leq n-1, -1 \leq l \leq p-1} \text{ and } M_t^{(1)} = \left( Y_t^{(k,l)} \right)_{-1 \leq k \leq m-1, -1 \leq l \leq q-1}.$$

(i) *Let us recall that  $\bar{D}$  denotes the transpose of a given matrix  $D$ . The processes  $M^{(1)}$  and  $M^{(2)}$  follow the dynamics*

$$dM_t^{(i)} = \int_{\mathbb{R}_+} W^{(i)}(t, x) N^{(i)}(dt, dx) + \left( C^{(i)} M_t^{(i)} + M_t^{(i)} \bar{D}^{(i)} \right), \quad (3.29)$$

where

- $W_{k,l}^{(1)}(t, x) = w^{(l)}(t)\psi_0^{(k)}(x)$  for  $-1 \leq k \leq m-1$  and  $-1 \leq l \leq q-1$ ,
- $W_{k,l}^{(2)}(t, x) = v^{(l)}(t)\phi_0^{(k)}(x)$  for  $-1 \leq k \leq n-1$  and  $-1 \leq l \leq p-1$ ,
- $C^{(1)} = C(r)$ ,  $C^{(2)} = C(c)$ ,  $D_t^{(1)} = C(k(t))$  and  $D_t^{(2)} = C(d(t))$  where  $C(\cdot)$  is defined by Equation (3.7).

(ii) As a consequence of the dynamics (3.29),  $(M_t^{(1)}, M_t^{(2)})_{t \geq 0}$  is a Markov process.

**Proof of Lemma 3.3** We focus on the dynamics of the  $X^{k,l}$ , the problem being the same for the  $Y^{k,l}$ . From Lemma 3.2, for  $0 \leq k \leq n-2$  and  $0 \leq l \leq p-2$ ,

$$dX_t^{k,l} = v^{(l)}(t) \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(dt, dx) + (X_t^{k+1,l} + X_t^{k,l+1}) dt. \quad (3.30)$$

From Assumption 3.2,  $X_t^{n,l} = \sum_{k=-1}^{n-1} c_k X_t^{k,l}$  and  $X_t^{k,n} = \sum_{l=-1}^{p-1} d_l(t) X_t^{k,l}$ . This shows that for  $0 \leq l \leq p-2$ ,

$$dX_t^{n-1,l} = v^{(l)}(t) \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(dt, dx) + \left( \sum_{k=-1}^{n-1} c_k X_t^{k,l} + X_t^{n-1,l+1} \right) dt, \quad (3.31)$$

and for  $0 \leq k \leq n-2$ ,

$$dX_t^{k,p-1} = v^{(p-1)}(t) \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(dt, dx) + \left( X_t^{k+1,p-1} + \sum_{l=-1}^{p-1} d_l(t) X_t^{k,l} \right) dt. \quad (3.32)$$

and also that

$$dX_t^{n-1,p-1} = v^{(p-1)}(t) \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(dt, dx) + \left( \sum_{k=-1}^{n-1} c_k X_t^{k,p-1} + \sum_{l=-1}^{p-1} d_l(t) X_t^{n-1,l} \right) dt, \quad (3.33)$$

In addition, Lemma 3.2 gives for  $0 \leq l \leq p-1$ ,

$$dX_t^{-1,l} = v^{(l)}(t) dN_t^{(2)} + X_t^{-1,l+1} dt, \quad (3.34)$$

and for  $0 \leq k \leq n-2$ ,

$$dX_t^{k,-1} = \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(dt, dx) + X_t^{k+1,-1} dt. \quad (3.35)$$

Finally, by Assumption 3.2 again, we get the following two equations:

$$dX_t^{-1,p-1} = v^{(p-1)}(t) dN_t^{(2)} + \left( \sum_{l=-1}^{p-1} d_l(t) X_t^{-1,l} \right) dt, \quad (3.36)$$

and

$$dX_t^{n-1,-1} = \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(dt, dx) + \left( \sum_{k=-1}^{n-1} c_k X_t^{k,-1} \right) dt. \quad (3.37)$$

From Equations (3.30) to (3.37), one then deduces the dynamics (3.29).  $\diamond$

To ensure tractability of the Laplace transform derived in the following Theorem, we also state the following assumption.

**Assumption 3.3.** For each  $\lambda > 0$ ,

$$\int_{\mathbb{R}_+} \exp \left( \lambda \max_{0 \leq k \leq n-1} \phi_0^{(k)}(x) \right) G(x) dx < +\infty.$$

Our main result is stated below. Note that the trace of the matrix  $\bar{u}M$  given by  $\text{Tr}(\bar{u}M) = \sum_{k,l} u_{k,l} M_{k,l}$  computes a linear combination of the components of a given matrix  $M$ , and recall that  $\bar{u}$  denotes the transposition of the matrix  $u$ .

**Theorem 3.1.** Let us denote  $\mathcal{F}^M$  the canonical filtration generated by  $(M^{(1)}, M^{(2)})$ .

Under Assumption 3.2,

(i) For any deterministic and differentiable matrix-valued  $(A_t^{(1)})$  and  $(A_t^{(2)})$  with derivatives  $(\dot{A}_t^{(1)})$  and  $(\dot{A}_t^{(2)})$ , the following process is an  $\mathcal{F}^M$ -martingale:

$$\begin{aligned} & \exp \left\{ \sum_{i=1}^2 \text{Tr} \left( A_t^{(i)} M_t^{(i)} \right) - \int_0^t \text{Tr} \left( A_s^{(i)} C^{(i)} M_s^{(i)} + A_s^{(i)} M_s^{(i)} \bar{D}_s^{(i)} + \dot{A}_s^{(i)} M_s^{(i)} \right) ds \right. \\ & - \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(1)} W^{(1)}(s,x))} - 1 \right) \rho(s) H(x) dx ds \\ & \left. - \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(2)} W^{(2)}(s,x))} - 1 \right) (\mu(s) + M_s^{(1)}[0,0] + M_s^{(2)}[0,0]) G(x) dx ds \right\}, t \geq 0. \end{aligned} \quad (3.38)$$

(ii) For each matrices  $u$  and  $v$  with dimensions  $(n+1)(p+1)$  and  $(m+1)(q+1)$  respectively, the joint Laplace transform can be expressed as

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \text{Tr}(\bar{u}M_t^{(1)} + \bar{v}M_t^{(2)}) \right) \right] &= \exp \left\{ \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(1)} W^{(1)}(s,x))} - 1 \right) \rho(s) H(x) dx ds \right. \\ & \left. + \int_0^t \int_{\mathbb{R}_+} \left( e^{\text{Tr}(A_s^{(2)} W^{(2)}(s,x))} - 1 \right) \mu(s) G(x) dx ds \right\}, \end{aligned} \quad (3.39)$$

where

$$\text{for } i \in \{1, 2\}, \quad \dot{A}_t^{(i)} + A_t^{(i)} C^{(i)} + \bar{D}_t^{(i)} A_t^{(i)} = \left\{ \int_{\mathbb{R}_+} \left( 1 - e^{\text{Tr}(A_t^{(i)} W^{(i)}(t,x))} \right) G(x) dx \right\} K, \quad (3.40)$$

with terminal conditions

$$A_T^{(1)} = \bar{u} \text{ and } A_T^{(2)} = \bar{v}, \quad (3.41)$$

where the matrix  $K$  is given by  $K = \bar{J}J$  and  $J$  is given in (3.13). Moreover, solution to (3.40)-(3.41) exist and is unique provided that Assumption 3.3 is satisfied.

**Proof of Theorem 3.1** We begin by exhibiting the exponential martingale (3.38). Let us denote  $\langle N^{(i)}, H \rangle_t = \int_0^t \int_{\mathbb{R}_+} H(s, x) N^{(i)}(ds, dx)$ . For deterministic  $\alpha(t, x)$  and  $\beta(t, x)$ , then by the classical exponential formula the following process is a martingale

$$\begin{aligned} & \exp \left\{ \langle N^{(1)}, \alpha \rangle_t + \langle N^{(2)}, \beta \rangle_t - \int_0^t \int_{\mathbb{R}_+} (e^{\alpha(s, x)} - 1) \rho(s) H(x) dx ds \right. \\ & \left. - \int_0^t \int_{\mathbb{R}_+} (e^{\beta(s, x)} - 1) \left( \mu(s) + \langle Z_{s-}^{(1)}, \Psi_s \rangle + \langle Z_{s-}^{(2)}, \Phi_s \rangle \right) G(x) dx ds \right\}. \end{aligned} \quad (3.42)$$

The aim now is to compute the joint Laplace transform of the processes  $M^{(1)}$  and  $M^{(2)}$ . This remains to compute  $\mathbb{E} \left[ e^{\text{Tr}(\bar{u} \cdot M_t^{(1)} + \bar{v} \cdot M_t^{(2)})} \right]$ , since  $\text{Tr}(\bar{u} \cdot M) = \sum_{k, l} u_{k, l} M_{k, l}$ . Let us consider the two (deterministic) processes  $A_t^{(1)}$  and  $A_t^{(2)}$  with sizes  $(m+1)(q+1)$  and  $(n+1)(p+1)$  respectively. By integration by parts,

$$d \left( A_t^{(i)} M_t^{(i)} \right) = A_t^{(i)} dM_t^{(i)} + \mathring{A}_t^{(i)} M_t^{(i)} dt.$$

From (3.29), we get the dynamics

$$\begin{aligned} d \text{Tr} \left( A_t^{(i)} M_t^{(i)} \right) &= \int_{\mathbb{R}_+} \text{Tr} \left( A_t^{(i)} W^{(i)}(t, x) \right) N^{(i)}(dt, dx) \\ &+ \text{Tr} \left( C^{(i)} M_t^{(i)} + M_t^{(i)} \bar{D}_t^{(i)} + \mathring{A}_t^{(i)} M_t^{(i)} \right) dt. \end{aligned}$$

Let us now use Equation (3.42) with  $\alpha(t, x) = \text{Tr} \left( A_t^{(1)} W^{(1)}(t, x) \right)$  and  $\beta(t, x) = \text{Tr} \left( A_t^{(2)} W^{(2)}(t, x) \right)$  to get the martingale (3.38).

To get the Laplace transform, it remains to make the random part of the integrand in (3.38) vanish. To do this, let us first identify the term in  $M^{(1)}$  to get the linear equation (3.40) for  $i = 1$ . In addition, the term in  $M^{(2)}$  leads to (3.40) for  $i = 2$ . If we set terminal conditions (3.41), we get the Laplace transform (3.39) by the martingale property of (3.38).

To conclude on the existence and uniqueness, we use Cauchy-Lipschitz theorem. To show that solution of class  $\mathcal{C}^1$  to (3.40)-(3.41) exist and is unique, it is sufficient to prove that the map  $(Y, t) \mapsto \int_{\mathbb{R}_+} e^{\text{Tr}(Y W^{(2)}(t, x))} G(x) dx$  is of class  $\mathcal{C}^1$ . Since the integrand is  $\mathcal{C}^1$  by Assumption 3.2 (i) and (ii), it is sufficient to prove that its gradient given by

$$\left( e^{\text{Tr}(Y W^{(2)}(t, x))} W^{(2)}(t, x)^T, \text{Tr} \left( Y \partial_t W^{(2)}(t, x) \right) e^{\text{Tr}(Y W^{(2)}(t, x))} \right) \quad (3.43)$$

is locally bounded by some quantity that is independent of  $Y$  and  $t$ , and is integrable with respect to  $G$ . Let us use some localization argument, and define the set  $B(0, r) = \{A \text{ real } (n+1) \times (p+1) \text{ matrix such that } \|A\|_\infty \leq r\}$ , where  $r > 0$  and



$\|A\|_\infty = \max_{-1 \leq i \leq n-1} \sum_{j=-1}^{p-1} |A_{i,j}|$ . Now, for  $(Y, t) \in B(0, r) \times [0, T]$  we get

$$\begin{aligned} & \exp(\operatorname{Tr}(YW^{(2)}(t, x))) \\ & \leq \exp\left(\sum_{i=-1}^{n-1} \sum_{k=-1}^{p-1} |Y_{i,k}| \left|W_{k,i}^{(2)}(t, x)\right|\right) \\ & \leq \exp\left((n+1) \max_{-1 \leq i \leq n-1} \sum_{k=-1}^{p-1} |Y_{i,k}| \left|W_{k,i}^{(2)}(t, x)\right|\right) \\ & \leq \exp\left(r(n+1) \max_{-1 \leq l \leq p-1} \sup_{t \in [0, T]} |v^{(l)}(t)| \max_{-1 \leq k \leq n-1} \left|\phi_0^{(k)}(x)\right|\right), \end{aligned}$$

where the last inequality uses that  $Y \in B(0, r)$ . As for the first component of (3.43),

$$\left|W_{k,l}^{(2)}(t, x)\right| \leq \left|\phi_0^{(k)}(x)\right| \sup_{t \in [0, T]} |v^{(l)}(t)|,$$

and for the second component we have

$$\left|\operatorname{Tr}(Y \partial_t W^{(2)}(t, x))\right| \leq r(n+1) \max_{0 \leq l \leq p} \sup_{t \in [0, T]} |v^{(l)}(t)| \max_{-1 \leq k \leq n-1} \left|\phi_0^{(k)}(x)\right|,$$

this concludes the proof by the use of Assumptions 3.2 and 3.3.  $\diamond$

### 3.5.3 On the pathwise representation

As for the standard Hawkes process, it is possible to give a pathwise representation of the Hawkes process with general immigrants and its underlying population. To do this, let us first extend the thinning construction in Section 3.4 to point processes with marks. Poisson point measures can be used, not only to represent counting processes, but also general random point measures on  $\mathbb{R}_+ \times E$ , say  $\Gamma(ds, dy) = \sum_{n \geq 1} \delta_{(T_n, Y_n)}(ds, dy)$ , where  $(E, \mathcal{E})$  is some measurable space. As for the Hawkes process,  $T_n$  is seen as the time at which an individual arrives in the population (immigration or birth). In addition,  $E$  represents the space of characteristics and the mark  $Y_n$  refers to the characteristic inherited by the individual that arrived at time  $T_n$ . Let us construct a random point measure  $\Gamma(ds, dy)$  with general intensity measure  $\gamma(ds, dy)$  assuming that it admits a density:  $\gamma(ds, dy) = \gamma(s, y) ds \mu(dy)$ .

In this model, events occur with intensity  $s \mapsto \int_{x \in E} \gamma(s, x) \mu(dx)$ , and if a birth occurs at time  $T_n$ , then the characteristics  $Y_n$  of the newborn are drawn with distribution

$$\frac{\gamma(T_n, y) \mu(dy)}{\int_{x \in E} \gamma(T_n, x) \mu(dx)}.$$

Let  $Q(ds, dy, d\theta)$  be a Poisson point measure on  $\mathbb{R}_+ \times E \times \mathbb{R}_+$  with intensity measure  $ds \mu(dy) d\theta$ . Let us still denote  $(\mathcal{F}_t^Q)$  the canonical filtration generated by  $Q$ , and introduce  $P(\mathcal{F}_t^Q)$  the predictable  $\sigma$ -field associated with  $\mathcal{F}_t^Q$ . We further assume that  $\gamma(t, y)$  is  $P(\mathcal{F}_t^Q) \times \mathcal{E}$ -measurable and also that  $\int_0^t \int_E \gamma(s, y) ds \mu(dy) < +\infty$  a.s.. Now, define

$$\Gamma(ds, dy) = \int_{\mathbb{R}_+} \mathbf{1}_{[0, \gamma(s, y)]}(\theta) Q(ds, dy, d\theta). \quad (3.44)$$

This clearly defines a point measure and the martingale property for  $Q$  ensures that the random point measure  $\Gamma(ds, dy)$  has intensity measure  $\gamma(s, y)ds\mu(dy)$ . Such construction can be found in Massoulié (1998); we refer to this paper for more details.

We are now ready to construct the two point measures  $N^{(1)}$  and  $N^{(2)}$  given in Equation (3.28). Let us introduce two independent Poisson point measures  $Q^{(1)}(dt, dx, d\theta)$  and  $Q^{(2)}(dt, dx, d\theta)$  on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  (enlarged if necessary) with same intensity measure  $dsdx d\theta$  on  $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$ . The first point measure is immediate to construct since its intensity does not depend on it. Indeed, one can define

$$N^{(1)}(dt, dx) = \int_{\mathbb{R}_+} \mathbf{1}_{[0, \rho(t)H(x)]}(\theta) Q^{(1)}(dt, dx, d\theta).$$

We emphasize that this is not an equation on  $N^{(1)}$  since its intensity does not depend on  $N^{(1)}$  itself. As for the second point process related to the Hawkes process, the intensity is given as a particular for of the process itself. Indeed, the intensity measure of the point measure  $N^{(2)}(dt, dx)$  is given by  $\lambda_t G(x)$  where  $\lambda_t$  can be written using (3.25) as

$$\lambda_t = \mu(t) + \int_{(0,t)} \Phi_t(t-s, x) N^{(2)}(ds, dx) + \int_{(0,t)} \Psi_t(t-s, x) N^{(1)}(ds, dx).$$

Then the point measure  $N^{(2)}$  can be defined as the solution to the following equation:

$$N^{(2)}(dt, dx) = \int_{\mathbb{R}_+} \mathbf{1}_{[0, (\mu(t) + \int_{(0,t)} \Phi_t(t-s, x) N^{(2)}(ds, dx) + \int_{(0,t)} \Psi_t(t-s, x) N^{(1)}(ds, dx)) G(x)]}(\theta) Q^{(2)}(dt, dx, d\theta).$$

Let us now give the pathwise representation of the corresponding populations. From Equation (3.26), it follows that

$$Z_t^{(1)}(da, dx) = \int_{(0,t]} \int_{\mathbb{R}_+ \times \mathbb{R}_+} \mathbf{1}_{[0, \rho(s)H(x)]}(\theta) \delta_{(t-s, x)}(da, dx) Q^{(1)}(ds, dx, d\theta). \quad (3.45)$$

and

$$Z_t^{(2)}(da, dx) = \int_{(0,t]} \int_{\mathbb{R}_+ \times \mathbb{R}_+} \mathbf{1}_{[0, (\mu(s) + \langle Z_{s-}^{(2)}, \Phi_s \rangle + \langle Z_{s-}^{(1)}, \Psi_s \rangle) G(x)]}(\theta) \delta_{(t-s, x)}(da, dx) Q^{(2)}(ds, dx, d\theta). \quad (3.46)$$

Such representations are used in the field of stochastic population dynamics for populations with ages and/or characteristics (see in particular Fournier and Méléard (2004), Tran (2008) and Bensusan et al. (2010–2015)). As for the standard Hawkes process, the pathwise representation has many advantages. In particular, it allows to derive the full system of equations and to identify each generation (see Section 3.4). More importantly, this formulation makes the link between the Hawkes process literature and the field of stochastic population dynamics. To further investigate this link seems to be a promising direction for future research.

## Conclusion

We introduced the concept of age pyramid for a class of Hawkes processes with general immigrants. The virtue of this approach is to keep track of all past events. This allows tractable computations for the Hawkes process with general immigrants whose fertility functions are time dependent generalizations of the popular exponential case, providing natural extensions of the existing results in this direction. In addition, we illustrated the pathwise construction of the Hawkes dynamics and its underlying population process. On the whole, our approach seems to reconcile two definitions of Hawkes processes, through an intensity process or a branching dynamics. This framework appears to be a promising direction for further research. As an example, the large population asymptotics in the field of measure-valued population dynamics could give further insights on the macroscopic behavior of Hawkes processes.

## Acknowledgements

The author is grateful to his supervisor Nicole El Karoui for her help to improve the results and the whole paper. The author also thanks Mathieu Rosenbaum, Thibault Jaisson and Monique Jeanblanc for fruitful discussions and enlightening comments, as well as the anonymous referee for valuable advice.

## Appendix A

**Proof of Corollary 3.5.** Let us identify the terms in Equation (3.17). Let us denote  $A_t = (A_{-1}(t), \dots, A_{n-1}(t))$ . The identification of the first component gives

$$c_{-1}A_{n-1}(t) + A'_{-1}(t) = 0. \quad (3.47)$$

The second component leads to

$$A'_0(t) + c_0A_{n-1}(t) + e^{A_t \cdot m} - 1 = 0. \quad (3.48)$$

As for  $1 \leq k \leq n - 1$ , we get

$$A_{k-1}(t) + c_kA_{n-1}(t) + A'_k(t) = 0. \quad (3.49)$$

Recursive computation of (3.49) provides for  $0 \leq k \leq n - 1$ ,

$$A_k(t) = (-1)^{n-1-k}A_{n-1}^{(n-1-k)}(t) + \sum_{l=1}^{n-1-k} (-1)^l c_{k+l}A_{n-1}^{(l-1)}(t). \quad (3.50)$$

One deduces from (3.50) that

$$A'_0(t) = (-1)^{n-1} A_{n-1}^{(n)}(t) + \sum_{k=1}^{n-1} (-1)^k c_k A_{n-1}^{(k)}(t). \quad (3.51)$$

Let us introduce the function  $G(t) = \int_T^t A_{n-1}(s) ds$  and choose  $A_{-1}(t) = \theta_1 - c_{-1}G(t)$  that satisfies (3.47). Now, put (3.50) and (3.51) into (3.48) to get the following non-linear ordinary differential equation for  $G$ ,

$$\begin{aligned} & (-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) + \exp\left(\theta_1 - c_{-1}G(t)\right) \\ & + m_{n-1} G'(t) + \sum_{k=0}^{n-2} m_k \left[ (-1)^{n-1-k} G^{(n-k)}(t) + \sum_{l=1}^{n-1-k} (-1)^l c_{k+l} G^{(l)}(t) \right] - 1 = 0. \end{aligned} \quad (3.52)$$

Let us simplify the sum in the exponential. By changing variable  $k$  into  $n-1-k$ , it is equal to  $\sum_{k=1}^{n-1} m_{n-1-k} (-1)^k G^{(k+1)}(t) + \sum_{k=1}^{n-1} \sum_{l=1}^k (-1)^l m_{n-1-k} c_{n-1-k+l} G^{(l)}(t)$ .

Then exchanging the sums leads to

$$\sum_{k=1}^{n-1} m_{n-1-k} (-1)^k G^{(k+1)}(t) + \sum_{l=1}^{n-1} (-1)^l \left( \sum_{k=l}^{n-1} m_{n-1-k} c_{n-1-k+l} \right) G^{(l)}(t).$$

Finally, by setting  $l \leftarrow l+1$  and exchanging notations  $k$  and  $l$ , (3.52) becomes

$$(-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) + \exp\left(\theta_1 - c_{-1}G(t) + \sum_{k=0}^{n-1} b_k G^{(k+1)}(t)\right) - 1 = 0, \quad (3.53)$$

where for  $0 \leq k \leq n-1$ ,  $b_k = (-1)^k (m_{n-1-k} - \sum_{l=k+1}^{n-1} m_{n-1-l} c_{n-l+k})$ .

Now, let us use (3.16) with (3.48) to get

$$\begin{aligned} \mathbb{E}[\exp(v.X_T)] &= \exp\left(-\mu \int_0^T (A'_0(t) + c_0 A_{n-1}(t)) dt\right), \\ &= \exp\left(-\mu \int_0^T \left( (-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) \right) dt\right), \\ &= \exp\left(-\mu \left( (-1)^{n-1} (G^{(n)}(T) - G^{(n)}(0)) + \sum_{k=0}^{n-1} (-1)^k c_k (G^{(k)}(T) - G^{(k)}(0)) \right)\right), \end{aligned}$$

where the second equality comes from (3.51). Let us set  $A_0(T) = \theta_2$  and for  $1 \leq k \leq n-1$ ,  $A_k(T) = 0$ . One can show by (3.50) that the previous conditions are equivalent to the terminal values  $G^{(n)}(T) = (-1)^{n-1} \theta_2$  and for  $1 \leq k \leq n-1$ ,  $G^{(k)}(T) = 0$ . Note that by definition of  $G$  we also get  $G(T) = 0$ . This leads to

$$\mathbb{E}[\exp(\theta_1 N_T + \theta_2 \langle Z_T, \phi \rangle)] = \exp\left\{-\mu \left( \theta_2 + (-1)^n G^{(n)}(0) + \sum_{k=0}^{n-1} (-1)^{k+1} c_k G^{(k)}(0) \right)\right\}.$$

This concludes the proof.  $\diamond$

## Appendix B

**Link with birth-death processes** In this work, we derived and exploited the interpretation of the Hawkes process in terms of birth dynamics. In this appendix, our aim is to highlight some link between Hawkes processes and birth-death dynamics. Let us first rework an example by Brémaud and Massoulié (2002) (see Example 4). Consider a Hawkes process with intensity

$$\lambda_t = \mu + \sum_{T_n < t} h(t - T_n, X_n),$$

where the marks  $(X_n)$  are iid with distribution  $G(dx)$ . Let us interpret  $X_n$  as the lifetime of the individual born at time  $T_n$ . Take  $h(a, x) = \bar{b}\mathbf{1}_{[0,x]}(a)$  for some constant birth rate  $\bar{b}$  and let  $G(dx)$  be the exponential distribution with parameter  $\bar{d} > 0$ . Then the intensity is that of a birth-death process (with immigration), where the (constant) birth and death rates are  $\bar{b}$  and  $\bar{d}$  respectively (also, the immigration rate is  $\mu$ ). In this setting, the Hawkes process  $N_t$  still counts the number of births up to time  $t$ . To get the number of alive individuals at time  $t$ , which we denote  $N_t^a$ , one has to keep track of the number of deaths, that is the number of individuals whose age at time  $t$  exceeds their lifetime, as

$$N_t^d = \sum_{T_n \leq t} \mathbf{1}_{[X_n, \infty)}(t - T_n).$$

Then the number of alive individuals is  $N_t^a = N_t - N_t^d$ , and the intensity can be rewritten  $\lambda_t = \mu + \bar{b}N_{t-}^a$ . The good feature of the constant rate framework is that the population size  $N_t^a$  can be directly recovered from the intensity  $\lambda_t$ .

How to construct a birth-death process with age-dependent birth and death rates? For this task, we are inspired by the construction of birth-death processes by Garcia and Kurtz (2006). Let us choose  $G(dx)$  to be the exponential distribution with parameter 1, and introduce age-dependent birth and death rates  $b(a)$  and  $d(a)$  respectively. If we set

$$h(a, x) = b(a)\mathbf{1}_{[0,x]} \left( \int_0^a d(u)du \right),$$

then

$$\lambda_t = \mu + \sum_{T_n < t} b(t - T_n)\mathbf{1}_{[0, X_n]} \left( \int_0^{t-T_n} d(u)du \right)$$

is nothing but the intensity of a birth-death process (with immigration) where the age-dependent birth and death rates are  $b(a)$  and  $d(a)$  respectively. To conclude this appendix, let us mention that the number of alive individuals can be recovered using the associated age pyramid (measure-valued) process  $Z_t(da, dx)$  by  $N_t^a = \langle Z_t, f \rangle$  with  $f(a, x) = \mathbf{1}_{[0,x]} \left( \int_0^a d(u)du \right)$ .

# Chapter 4

## Cause-of-death mortality: what can be learned from population dynamics?<sup>†</sup>

### Abstract

This paper analyses cause-of-death mortality changes and its impacts on the whole population evolution. The study combines cause-of-death analysis and population dynamics techniques. Our aim is to measure the impact of cause-of-death reduction on the whole population age structure, and more specifically on the dependency ratio which is a crucial quantity for pay-as-you-go pension systems. Whereas previous studies on causes of death focused on mortality indicators such as survival curves or life expectancy, our approach provides additional information by including birth patterns. As an important conclusion, our numerical results based on French data show that populations with identical life expectancies can present important differences in their age pyramid resulting from different cause-specific mortality reductions. Sensitivities to fertility level and population flows are also given.

**Keywords:** Cause-of-Death Mortality, Population Dynamics, Birth and Death Process, Individual-Based Model, Age Pyramid, Age Dependency Ratio, Macroscopic Behavior

### 4.1 Introduction

The studies on cause-of-death mortality have provided a better understanding of the level and the evolution of aggregate mortality rates over the years. At the same

---

<sup>†</sup>Arnold, S., A. Boumezoued, H. Labit Hardy, N. El Karoui. 2015. Cause-of-death mortality: What can be learned from population dynamics? HAL preprint Id: hal-01157900

time, population dynamics models have been developed in mathematical ecology and demography to model the evolution of the population age structure over time by including both death and birth patterns. Whereas studies on cause-of-death mortality focused on the impact on mortality indicators, this paper aims to combine these two fields in order to answer the following question: what is the impact of changes in cause-of-death mortality on the whole population age structure ?

To address such question, the first step is the modeling of cause-of-death mortality. Each observed death is recorded with a primary cause, such as for example cancer, accident, or a disease related to circulatory or respiratory system. For modeling cause-of-death mortality, one famous framework in actuarial science and biostatistics is the competing risks model. In this framework, each cause of death is modeled at the individual level by a clock, and the death occurs when the first clock rings. In other words, the lifetime of an individual is modeled as the minimum between competing cause-specific lifetimes. Unfortunately, those cause-specific durations are not observable in the data: in practice we only observe the minimum between the durations related to each cause. Since information on causes of death durations are not observable, assumptions about the dependence between causes of death have to be set in order to capture the marginal distributions and to model changes in cause-of-death mortality. The seminal work of Chiang (1968) on the competing risks theory considered independent causes of death and led to further studies during the last decades. The competing risks framework has been investigated with several dependence structures between the competing lengths of time (see e.g. Carriere (1994), Kaishev et al. (2007) and Dimitrova et al. (2013)). For our study, we focus on the competing risks framework under the independence assumption, although we also briefly survey the debate on the dependence structure. This independence assumption is much used (see e.g. Prentice et al. (1978), Tsai et al. (1978), Wilmoth (1995) and Putter et al. (2007)), and allows us to already highlight interesting effects of cause removal on the population dynamics.

The previous studies in the cause-of-death literature focused on the impact of cause-of-death removal or reduction on mortality indicators: e.g. life expectancy for Prentice et al. (1978), survival curve for Dimitrova et al. (2013) and death probabilities for Alai et al. (2015). In this paper, the aim is to go further and study the impact on the population age pyramid (the number of individuals by age-class) by means of a population dynamics model. The age pyramid is a crucial quantity of interest since it gives information on how age-classes evolve and interact with each others. To our knowledge, this has not been addressed in the current literature.

To do so, we rely on both deterministic and stochastic population dynamics models. Historically, the first models that addressed the deformation of the age structure of a population over time were deterministic models. The most famous one is

called the McKendrick-Von Foerster model (see McKendrick (1926) and Von Foerster (1959)). Based on death rates by age, and also birth rates, they are able to capture the time evolution of the number of individuals by age-class. For demographic purposes, dealing with gender classification has become natural as census and national data make this distinction between males and females for a long time. It is also crucial since age but also gender is known to have a real impact on mortality. To this aim, there is now a sizable literature on two-sex population deterministic models both with and without age structure (see e.g. Keyfitz (1972) and Iannelli et al. (2005)). The deterministic point of view is useful to understand the evolution of the "average" age pyramid, and can provide first interesting insights about the impact of cause-of-death reductions on the whole population, based on a fast numerical discretization scheme. But in practice, we are not only interested in the age pyramid, but also in non-linear quantities such as the age dependency ratio. This is defined as the number of people aged above 65 divided by the number of individuals aged between 15 and 65. This indicator is a crucial quantity of interest for pure demographic analysis but also for pay-as-you-go pension systems (see e.g. Fougère and Mérette (1999), Mirkin and Weinberger (2001) and Hock and Weil (2012)). The deterministic evolution of the population and the age dependency ratio, interesting in itself, is not realistic for populations with finite size in which individuals have random lifetimes and times of birth. What is then the population dynamics ? First insights tell us that the average stochastic dependency ratio, as a non-linear quantity, cannot be equal to the ratio of the average population. Therefore a proper stochastic modeling framework is needed. In fact, in a large population asymptotics, the stochastic behavior can be approximated by its deterministic counterpart, making this approach micro-macro consistent. In other words, at the microscopic scale, the population behaves stochastically whereas if the number of individuals is "large", the behavior is described by deterministic equations that are classical in demography. The micro-macro consistency is at the core of what are called individual-based models, that have been developed in mathematical ecology (see Fournier and Méléard (2004) and Champagnat et al. (2006)), in particular to understand the evolutionary mechanism. In addition, these models have been developed to take ages into account (see Ferriere and Tran (2009)), which is crucial for our purpose.

These processes have been little used for demographic applications compared to issues linked to biology or evolutionary theory. In particular, to our knowledge none of them address the impact of cause-of-death removals on the population age structure. Importantly for our purpose, we are inspired by the work of Bensusan et al. (2010–2015) (see also Bensusan (2010)) who have extended and used population dynamics models in view of human population applications, including pension systems and longevity risk in insurance. Among the few applications to demographic pur-



poses, note also that an alternative named Markovian binary tree is considered in Hautphenne and Latouche (2012) for demographic applications. Markovian binary trees differ from individual-based models as they involve discrete age-classes and a random aging pattern to preserve the Markov property. The key feature of our approach is to take into account aging in continuous time and to use a microsimulation procedure for the stochastic scenarios. It is interesting to note that microsimulation procedures are used in a discrete time setting by several public institutions to test economic and retirement policies (see e.g. the review by Li and O'Donoghue (2013)). This includes for example the French National Institute for Statistics and Economic Studies (INSEE) with the model *Destinie* (see Bonnet et al. (1999) and Duée (2005)). Apart from government bodies, microsimulation approaches have also been used in the demographic and statistical literature. As an example, Hynman and Booth (2008) used discrete-time stochastic simulations to compute the population evolution and the age pyramid, while aggregate mortality, fertility, and migration rates are forecasted.

The paper is organized as follows. In Section 4.2, we first describe the competing risks framework for the modeling of cause-of-death mortality, and secondly introduce the deterministic and stochastic population dynamics approaches. Section 4.3 details the data we use to carry out our study: the World Health Organization data for the cause-of-death mortality rates and the data from INSEE for the birth rates and the initial age pyramid. Section 4.4 contains our numerical results regarding deterministic and stochastic patterns of the age pyramid and the dependency ratio, as well as sensitivities to fertility scenarios and also to population flows computed from the data. As an important conclusion, our numerical results show that under same life expectancy improvement, the age pyramid can evolve in many ways if different causes of death are reduced.

## 4.2 The Model

As usual when working with human population, we are concerned with population structured by gender and age, so we make the distinction between the age pyramid of females and the age pyramid of males. It is well known that gender and age induce significant differences on mortality. In the model, each individual, female or male, is exposed to  $d$  competing causes of death, such as cancers, diseases of circulatory system, diseases of respiratory system or accidents, well described in Section 4.3, but may die from any single one of the  $d$  causes. After a brief description of the classical competing risks framework in survival analysis, we introduce the debate on the dependence structure between causes of death, and describe the different points of view proposed in the literature. Afterwards, we present the meaning of removal

of a cause of death. The population dynamics modeling is described secondly.

### 4.2.1 Cause-of-Death Mortality

**Competing risks framework** Standard survival data measure the length of time from the time origin until the occurrence of some events; for human population, one is interested in the duration between the time of birth and the time of death (*frailty time* in other framework). This is called the *lifetime* and is represented by a random variable  $\tau^\epsilon$ , with survival function  $S^\epsilon(a) = \mathbb{P}(\tau^\epsilon > a)$ , for an individual with gender  $\epsilon$  ( $\epsilon = f$  for female or  $\epsilon = m$  for male). This is the probability for an individual with gender  $\epsilon$  to survive until age  $a$ . As usual, the information is often expressed in terms of *hazard rate* (also called *force of mortality* in actuarial sciences), defined as the probability to die before age  $a + da$  for an individual alive at age  $a$ :

$$\mu^\epsilon(a) = \frac{\mathbb{P}(a < \tau^\epsilon \leq a + da \mid \tau^\epsilon > a)}{da} = -\frac{d \ln(S^\epsilon(a))}{da}. \quad (4.1)$$

At this elementary stage, as in Elandt-Johnson (1976), we assume that in each group, females or males, each individual with gender  $\epsilon$  is assigned a vector of potential lifetimes  $(\tau_1^\epsilon, \tau_2^\epsilon, \dots, \tau_d^\epsilon)$  corresponding to the  $d$  causes of death. The  $d$  causes of death are modeled by  $d$  lengths of time, that is to say  $d$  competing clocks, denoted  $\tau_1^\epsilon, \tau_2^\epsilon, \dots, \tau_d^\epsilon$ . The death occurs when the first event corresponding to one of the  $d$  causes of death occurs: that is when the first clock rings. In other words, the lifetime of an individual  $\tau^\epsilon$  is defined as the minimum between the different lengths of time:  $\tau^\epsilon = \min_{1, \dots, d} \tau_i^\epsilon$ . In terms of survival function, we have:

$$S^\epsilon(a) = \mathbb{P}(\tau_1^\epsilon > a, \dots, \tau_d^\epsilon > a). \quad (4.2)$$

To study the lifetimes associated to causes of death, we need the corresponding marginal survival functions. In particular, to model changes in cause-of-death mortality, as a reduction of deaths due to cause  $i$ , the survival function has to be expressed with the marginal distribution  $S_i^\epsilon(a) = \mathbb{P}(\tau_i^\epsilon > a)$  and these have to be estimated from the data. In practice, when looking at a sample of lifetimes, we observe for a given individual, the age at death and the cause from which this individual dies. In the competing risks framework, this means that we observe the minimum of the  $d$  competing clocks, and that we know what is the clock which rang first. However, we do not observe the clocks for the other causes of death. Since marginals are not observable, dependence assumptions between causes of death have to be set.

**On the dependence structure** It is well known that, without additional assumption, it is impossible to identify the dependence structure of the joint distribution

and their marginal from the data (Tsiatis (1975)). A considerable literature has been devoted to the competing risk model for cause-of-death mortality, starting from the independent case (Chiang (1968)) to more complex dependence structures. Carriere (1994), Kaishev et al. (2007) and Dimitrova et al. (2013) have worked on the theory of copula functions to model dependence: they have been interested in modeling density function and survival curve. Alai et al. (2015) have studied causal mortality shocks with a multinomial logistic model. They have studied and modeled the death probabilities. Also models incorporating observed individual risk factors (see e.g. Manton (1986a), Rosen (2006) and Girosi and King (2006)) and unobserved ones (frailty models, see e.g. Vaupel and Yashin (1983) and Manton et al. (1986b)) have been developed. More recently, cointegration analyses have been applied to causes of death studies (see Arnold and Sherris (2013), Arnold and Sherris (2015a), Arnold and Sherris (2015b)). In that framework, cointegration is used to model dependence between the cause-of-death mortality rates.

Nonetheless, the independence assumption between cause-of-death durations is widely used (see e.g. Prentice et al. (1978), Tsai et al. (1978) and Wilmoth (1995), Putter et al. (2007)). Even if we agree with many authors on the partial irrelevance of this assumption, that can unfortunately not be tested in practice, it contributes to considerably simplify the further developments of this study. In particular, the marginals are directly estimated from the data (see the discussion below). Moreover, under the independence assumption, we can already highlight interesting effects of cause removal on the population dynamics.

**Independence assumption** Under the independence assumption between causes of death, the survival function defined in Equation (4.2) can be expressed with the marginals:  $S^\epsilon(a) = \prod_{i=1}^d S_i^\epsilon(a)$ . In this way, the hazard rate of the individual lifetime is the sum of the hazard rates by cause:  $\mu^\epsilon(a) = \sum_{i=1}^d \mu_i^\epsilon(a)$ , where  $\mu_i^\epsilon$  is such that  $\mathbb{P}(\tau_i^\epsilon > a) = \exp(-\int_0^a \mu_i^\epsilon(s) ds)$ . The rate  $\mu_i^\epsilon(a)$  is called the *net* force of mortality for cause  $i$ , gender  $\epsilon$  and age  $a$ . Formally, this corresponds to the probability to die instantaneously if only cause  $i$  exists, given the survival at age  $a$ :

$$\mu_i^\epsilon(a) = \frac{\mathbb{P}(a < \tau_i^\epsilon \leq a + da \mid \tau_i^\epsilon > a)}{da}. \quad (4.3)$$

As already described, the cause-specific durations are not totally observable in practice. However, from the data one can estimate the distribution of  $\tau_i^\epsilon$  for all individuals that died from cause  $i$ , because in this specific case the duration  $\tau_i^\epsilon$  is completely observed. In other words, it is possible to estimate the quantity  $\mathbb{P}(\tau_i^\epsilon \in [a, a + da), \tau^\epsilon = \tau_i^\epsilon)$ . By independence between the cause-specific durations,

this quantity is equal to

$$\mathbb{P}(\tau_i^\epsilon \in [a, a + da), \tau^\epsilon = \tau_i^\epsilon) = S_i^\epsilon(a) \mu_i^\epsilon(a) da \prod_{k \neq i} S_k^\epsilon(a) = S^\epsilon(a) \mu_i^\epsilon(a) da.$$

This way, it is possible to recover the *net* force of mortality by

$$\mu_i^\epsilon(a) = \frac{\mathbb{P}(\tau_i^\epsilon \in [a, a + da), \tau^\epsilon = \tau_i^\epsilon \mid \tau^\epsilon > a)}{da}.$$

In the actuarial literature, the right hand side of the previous equation is called the *crude* force of mortality. Under the independence assumption, it can be directly estimated from the data. The fact that *net* and *crude* mortality forces are equal under the independence assumption is a fundamental result which considerably simplifies the estimations.

**Cause-of-death removal** To address the impact of cause-of-death reduction or removal, a rigorous definition is needed. In the competing risks literature, two possible definitions have been introduced and tested: *ignore* or *eliminate*, see e.g. Elandt-Johnson (1976) and Dimitrova et al. (2013).

(i) **TO IGNORE** The first one says that if cause  $k$  is ignored, then the lifetime becomes  $\tau_{(-k)}^\epsilon = \min_{1 \leq i \leq d, i \neq k} \tau_i^\epsilon$ . In this case, the survival function is replaced by  $\mathbb{P}(\tau_{(-k)}^\epsilon > t)$ . Under the *ignoring* definition, the random variable  $\tau_{(-k)}^\epsilon$  can be interpreted as the lifetime in the world where cause  $k$  does not exist.

(ii) **TO ELIMINATE** The other definition is the *elimination*: in this case, the lifetime is now conditioned to the fact that  $\tau_k^\epsilon$  grows to infinity, that is the survival function is replaced by  $\mathbb{P}(\tau^\epsilon > t \mid \tau_k^\epsilon \rightarrow +\infty)$ . This models the fact that the cause still remains, but its associated lifetime is postponed to an arbitrary value.

Despite their difference in terms of interpretation, in the case where the  $\tau_i^\epsilon$  are independent, the two definitions are equivalent. This is another fundamental consequence of the independence assumption. In our case, we will use the term *remove* to refer to both definitions.

We will also refer to cause-of-death reduction: we say that cause  $k$  is reduced by a factor  $\alpha^\epsilon \in [0, 1]$  for individuals with gender  $\epsilon$  if the death rate  $\mu^\epsilon(a)$  becomes:

$$(1 - \alpha^\epsilon) \mu_k^\epsilon(a) + \sum_{1 \leq i \leq d, i \neq k} \mu_i^\epsilon(a).$$

**Debate on temporality of death rates** It is well known that mortality is improving over the years. This is captured for example by the increase in life expectancy which reflects the fact that individuals belonging to recent generations are living longer. In particular, the time evolution of mortality differs across age classes. In this paper, as a first step, we consider time-invariant death rates (and also fertility

rates), although the modeling framework can be described with time varying demographic rates. Whereas the purpose of our paper is not to forecast cause-specific mortality rates, we are interested to forecast the population itself with current fertility and cause-of-death mortality rates, and possibly population flows, in order to analyse the deformation of the age structure in this special setting. This is known as the stable population framework which is standard in demography. Although, it does not provide realistic population forecasts, this framework already allows to understand and interpret some short and long term impacts of a change in current cause-of-death mortality on the population age pyramid.

### 4.2.2 Population Dynamics

Cause-of-death mortality rates are useful to understand the distribution of individual lifetimes in a static population made of both men and women. However, this does not tell something about the time evolution of the whole population, in particular its interdependence with birth patterns and mortality reduction.

**Two-sex birth-death process with age structure** The simple one-sex population model is the well known birth-death process, in which birth and death rates are constant. While applied demographers primarily worked with one-sex population models, many sociological applications required two-sex models that explicitly address matching of females and males into couples. Some important applications such as "marriage squeeze", i.e. the impact on marriage of an imbalance between males and females (see e.g. Schoen (1983), Guilmoto (2012)), further investigated age-structured models that keep track of the age distributions of single females and single males as well as the joint distribution of ages across couples. In fact, dealing with gender classification has become natural as census and national data make this distinction for a long time. There is now a sizable literature on two-sex population models both with and without age structures (see e.g. Keyfitz (1972), Iannelli et al. (2005)). However, these two-sex population models do not embed cause-of-death analysis, which is the purpose of the present paper which focuses on both deterministic and stochastic two-sex birth-death processes with age structure and cause-of-death mortality. In order to understand the behavior of the population over time, in particular that of the age pyramid, we first need to specify how individuals give birth. The age pyramid (see Figure 4.3 for an illustration) is a crucial quantity of interest for pure demographic analysis, but also public pension pay-as-you-go systems. As for private insurers, this is crucial to understand how national mortality data is computed, especially how a biased knowledge on birth patterns and corresponding age pyramid can lead to misvaluation of reference mortality rates (see Cairns et al. (2014)). Further investigation on the impact of (heterogenous) birth

patterns is carried out in Bensusan et al. (2010–2015).

**Female fertility** In our model, each individual with gender  $\epsilon$  and age  $a$  gives birth to new individuals (of age 0) with rate  $b^\epsilon(a)$ . In a stochastic setting, the first time of birth  $T_b^\epsilon$ , for an individual with gender  $\epsilon$ , is given by its marginal distribution  $\mathbb{P}(T_b^\epsilon > a) = \exp\left(-\int_0^a b^\epsilon(s)ds\right)$ . As this individual can give birth to more than one individual, the times of birth occur as the jumps of a non-homogenous Poisson process with rate  $a \mapsto b^\epsilon(a)$ , until the parent dies. At birth, the new individual of age 0 is a female ( $f$ ) with probability  $p$  and a male ( $m$ ) with probability  $1 - p$ . In our application, we assume that only a female can be the cause of the arrival of a new individual in the population. This choice is driven by the lack of data on male birth rates, in particular the age of the father at birth of his children. So the birth rate for any individual of gender  $\epsilon$  and age  $a$  can be written  $b^\epsilon(a) = b(a)\mathbf{1}_f(\epsilon)$ , where  $\mathbf{1}_f(\epsilon)$  is the indicator function defined as  $\mathbf{1}_f(f) = 1$  and  $\mathbf{1}_f(m) = 0$ . In this model, only each female has an intrinsic birth rate  $b(a)$ , so we do not take into account the impact of the relative number of males and females in the population on the birth patterns. This has been the topic of several models and debates in demographic studies. In particular, this could be of interest to model events such as shocks in birth patterns prior and posterior to world wars. Also, the general setting can be useful to test several assumptions on the link between the birth rate and the whole population. This issue is left for further research.

**Deterministic population model** How does the population evolve over the years? In particular, what is the dynamics of the age pyramid over time? The first models which tried to address these questions were deterministic population models. They go back to McKendrick (1926) and Von Foerster (1959) and focus on the quantity  $g(\epsilon, a, t)$  of individuals with gender  $\epsilon$  and age  $a$  at time  $t$  (see also Iannelli et al. (2005) and the brief review in Tran (2008)). The population structured by gender is described by the vector

$$\mathbf{g}(a, t) = \begin{pmatrix} g(f, a, t) \\ g(m, a, t) \end{pmatrix}.$$

First, individuals with non-negative ages evolve according to deaths over time, that is

$$(\partial_a + \partial_t)\mathbf{g}(a, t) = -Q(a)\mathbf{g}(a, t), \quad (4.4)$$

where  $Q(a) = \begin{pmatrix} \mu^f(a) & 0 \\ 0 & \mu^m(a) \end{pmatrix}$ . This equation corresponds to the transport component: it states that individuals are aging over time and that the number of individuals is decreased by the number of deaths. Now, at a given time  $t$ , the number

of newborn is given by the sum over all newborns of all females with any age, and added to the gender class depending on the probability at birth, that is

$$\mathbf{g}(0, t) = \left( \int_{\mathbb{R}_+} g(f, a, t) b(a) da \right) \begin{pmatrix} p \\ 1 - p \end{pmatrix}. \quad (4.5)$$

This is called the renewal component, which sets some boundary condition on the space of ages. Initially, at time 0, the age pyramid is given by some initial condition  $\mathbf{g}(a, 0) = \mathbf{g}_0(a)$  for each  $a \geq 0$ . Equations (4.4) and (4.5) form what is called the McKendrick-Von Foerster equation. Note that this formulation exhibits an autonomous behavior of the female population, and that the renewal of the male population can be computed based on the female population at time  $t$ . Note also that we deal with "quantities" of individuals as  $g(f, a, t)$  and  $g(m, a, t)$  are real numbers: this represents the average population quantity. A first advantage of the deterministic equation is to be numerically very fast to compute through an explicit discretization scheme, and thus provides first insights on the average dynamics.

**The numerical scheme** The deterministic equation can be computed in the following way. Let  $\Delta t$  be the discretized time step. Let also  $\mathcal{G}$  denote the grid  $\{0, \Delta t, 2\Delta t, \dots\}$  on  $\mathbb{R}_+$ , and  $\mathcal{G}_+$  denote the grid  $\mathcal{G}$  without 0. We start with an initial population  $\mathbf{g}_0(a)$  for  $a \in \mathcal{G}$ , with  $\mathbf{g}_0(a)$  the quantity of individuals of age in  $[a, a + \Delta t[$  for each gender  $\epsilon$  in the initial population. The explicit numerical scheme can be written

$$\begin{aligned} \mathbf{g}(a + \Delta t, t + \Delta t) - \mathbf{g}(a, t) &= -Q(a)\mathbf{g}(a, t)\Delta t, \quad \forall a \in \mathcal{G}_+ \\ \mathbf{g}(0, t) &= \sum_{a \in \mathcal{G}_+} g(f, a, t) b(a) \Delta a \begin{pmatrix} p \\ 1 - p \end{pmatrix}, \quad \text{where } \Delta a = \Delta t, \\ \mathbf{g}(a, 0) &= \mathbf{g}_0(a), \quad \forall a \in \mathcal{G}. \end{aligned} \quad (4.6)$$

To clarify the simulation steps, we describe how to proceed recursively. Suppose that the population at time  $t$  has been computed. From the first line of (4.6), for  $a \in \mathcal{G}$ , one can compute

$$\mathbf{g}(a + \Delta t, t + \Delta t) = (I - \Delta t Q(a)) \mathbf{g}(a, t),$$

where  $I$  denotes the  $2 \times 2$  identity matrix. Then, one gets from the second line of (4.6), where  $t$  is replaced by  $t + \Delta t$ ,

$$\mathbf{g}(0, t + \Delta t) = \sum_{a \in \mathcal{G}_+} g(f, a, t + \Delta t) b(a) \Delta t \begin{pmatrix} p \\ 1 - p \end{pmatrix}.$$

This leads to the whole population at time  $t + \Delta t$ .

**From deterministic to stochastic population** The deterministic equation is useful to understand the evolution of the "average" age pyramid, in the sense that  $g(\epsilon, a, t)$  represents the average quantity of individuals with gender  $\epsilon$  and age  $a$  at time  $t$ . But in practice, we are not only interested in the age pyramid, but also in non-linear quantities such as the dependency ratio. This is defined as the number of people aged above 65 divided by the number of individuals aged between 15 and 65. Let us denote  $g_t(a) = g(m, a, t) + g(f, a, t)$  the quantity of people aged  $a$  at time  $t$ . Let us also introduce the notation  $g_t([a_1, a_2]) = \int_{[a_1, a_2]} g_t(a) da$  for the quantity of individuals with age  $a \in [a_1, a_2)$  at time  $t$ . Formally, the dependency ratio computed from the deterministic equation can be defined as

$$r_t = \frac{g_t([65, \infty))}{g_t([15, 65))}.$$

The assumption of a deterministic population pattern is obviously not realistic. In practice, one deals with finite size populations in which individuals have random lifetimes and times of birth. What is then the stochastic population dynamics? Before introducing the stochastic point of view, let us focus on the corresponding stochastic dependency ratio, say  $R_t$ . We want to underline that the average dependency ratio  $\mathbb{E}[R_t]$  cannot be directly computed from the deterministic equation, as it is not equal to the ratio of the average population,  $r_t$ . To further analyse the stochastic behavior and make the link with its deterministic counterpart, let us first describe the stochastic representation.

Analogously to the deterministic counterpart, let us denote  $Z_t^\epsilon([a_1, a_2])$  the number of individuals with age  $a \in [a_1, a_2)$  and gender  $\epsilon$  at time  $t$ . The difference is that we deal here with (random) numbers of individuals as we are in the microscopic stochastic setting. The population  $Z_t^\epsilon$ , which is in fact a measure on the space of ages, evolves over time according to random births and deaths with rates described at the beginning of this section. Let us introduce the total number  $Z_t([a_1, a_2]) = Z_t^f([a_1, a_2]) + Z_t^m([a_1, a_2])$ . In this microscopic setting, the stochastic dependency ratio is defined as

$$R_t = \frac{Z_t([65, \infty))}{Z_t([15, 65))}.$$

As mentioned in Bensusan et al. (2010–2015) (see also the discussion in Fournier and Méléard (2004)), as individual birth and death rates do not depend on the whole population, the average population density satisfies  $\mathbb{E}[Z_t([a_1, a_2])] = g_t([a_1, a_2])$ , provided that such density exists. However, the dependency ratio  $R_t$  is a non-linear quantity. In particular, one can write the definition of the covariance to get

$$r_t = \mathbb{E}[R_t] + \text{Cov} \left( R_t, \frac{Z_t([15, 65))}{g_t([15, 65))} \right).$$



That is, the deterministic dependency ratio  $r_t$  is the average dependency ratio  $\mathbb{E}[R_t]$  plus a corrective term related to the covariance between the dependency ratio and the deviation of the number of people aged  $[15,65)$  from its mean. This shows that the stochastic dependency ratio with finite population is obviously not well described by the deterministic equation. Therefore, a proper stochastic framework is needed. In addition, one can be more precise about the link between the stochastic representation and its deterministic counterpart. At the core of what are called individual-based models is the micro-macro consistency: at the microscopic scale, the population behaves stochastically whereas if the number of individuals is "large", the behavior is described by deterministic equations that are classical in demography. In the following, we briefly highlight the large population point of view and then describe the intuitive stochastic simulation procedure.

**Link between stochastic and deterministic models** One main advantage of such stochastic model is its micro-macro consistency. If the number of individuals increases to infinity, the stochastic dynamics tends to the deterministic pattern described by Equations (4.4)-(4.5). Let us introduce the vector  $\mathbf{Z}_t = \begin{pmatrix} Z_t^f \\ Z_t^m \end{pmatrix}$ , and recall that each component defines a measure on the space of ages. In particular,  $\mathbf{Z}_t([a_1, a_2))$  is the number of individuals for each gender and ages in  $[a_1, a_2)$  at time  $t$ . To get a large population limit theorem, it is assumed that the size of the initial population tends to infinity proportionally to a scaling parameter  $K \geq 1$ . We consider a sequence of population processes  $(\mathbf{Z}_t^K)_{t \in [0, T]}$  with initial population  $\mathbf{Z}_0^K$ . A non-degenerate process appears in the limit if each individual is weighted  $\frac{1}{K}$ : the renormalized population process is defined by  $\tilde{\mathbf{Z}}_t^K := \frac{1}{K} \mathbf{Z}_t^K$ . The convergence to the deterministic limit solution to (4.4)-(4.5) is stated in the following theorem. We state it in a summarized way and refer to Tran (2006) and Bensusan et al. (2010–2015) for the technical details.

**Theorem 4.1.** *Assume that there exists a finite and deterministic measure  $\mathbf{g}_0(a)da$  such that  $\tilde{\mathbf{Z}}_0^K$  converges in distribution to  $\mathbf{g}_0$  as  $K \rightarrow +\infty$  and that  $\sup_K \mathbb{E} \left[ \tilde{\mathbf{Z}}_0^K([0, \infty))^2 \right] < +\infty$ . Then the sequence of measure-valued processes  $(\tilde{\mathbf{Z}}_t^K)_{K \in \mathbb{N}^*}$  converges in distribution towards the continuous and deterministic process  $\mathbf{g}_\cdot$ , which satisfies the McKendrick-Von Foerster equation (4.4)-(4.5) (in a weak sense).*

**Remark 4.1.** *Note that one can also find in the literature central limit theorems for age-structured populations, which state that the convergence is achieved with speed  $\sqrt{K}$ , see for example Wang (1977), Oelschläger (1990) and Tran (2006).*

**Simulation of the stochastic scenarios** We detail here the simulation procedure described in Bensusan et al. (2010–2015) (see also Tran (2006)). The idea of

the simulation procedure is to generate candidate times of event for the population, and for each of them, decide if a real change happens or not. At each candidate time, we pick an individual in the population. According to this choice, we decide if an event (birth or death from a cause) occurs or not. This approach is based on the simulation of a simpler process with more jump times and some of them are deleted with a rule leading to the desired distribution for the population process. This procedure is analogous to the thinning procedure for point processes. This simulation technique is required since demographic rates depend on age which introduce time dependency.

**Assumption 4.1.** (*Bounded intensities*) Denote  $\mathcal{X} = \{f, m\}$  the set of possible gender. There exists positive constants  $\bar{b}$  and  $(\bar{\mu}_i)_{i=1}^d$  such that for all  $(\epsilon, a) \in \mathcal{X} \times \mathbb{R}_+$ ,  $b^\epsilon(a) \leq \bar{b}$ ,  $\mu_i^\epsilon(a) \leq \bar{\mu}_i$ . Let  $\bar{\mu} = \sum_{i=1}^d \bar{\mu}_i$ .

We describe below the simulation procedure.

**Algorithm 4.1.** Start at time  $T = 0$  with a given population  $\mathbf{Z}_T$  of size  $N_T$ .

**Step 1.**

Simulate a r.v.  $\tau \sim \text{Exp}((\bar{b} + \bar{\mu})N_T)$  and set  $T + \tau$  as the next candidate time.

**Step 2.**

Form the population  $\mathbf{Z}_{T+\tau}$  by adding  $\tau$  to all ages in  $\mathbf{Z}_T$ .

**Step 3.**

Pick an individual  $I$  uniformly in the population  $\mathbf{Z}_{T+\tau}$ , with gender and age  $(\epsilon^I, a^I)$ , and compute his probabilities to give birth ( $B$ ) and to die for cause  $i$  ( $D_i$ ), denoted respectively  $p_b$  and  $p_i$  for  $i$  in  $\llbracket 1, d \rrbracket$ :  $p_b = \frac{b^{\epsilon^I}(a^I)}{b + \bar{\mu}}$  and  $p_i = \frac{\mu_i^{\epsilon^I}(a^I)}{b + \bar{\mu}}$ . Let also define  $p_o = 1 - p_b - \sum_{i=1}^d p_i$  the "no event" probability.

**Step 4.**

Generate a potential event  $\mathcal{E} \in \{B, D_1, \dots, D_d, O\}$  with probability  $p_b, p_1, \dots, p_d, p_o$ :

(i) If  $\mathcal{E} = B$  (birth) then generate  $\epsilon'$  in  $\{f, m\}$  with probability  $(p, 1 - p)$  and add the individual  $(\epsilon', 0)$  to  $\mathbf{Z}_{T+\tau}$ .

(ii) If  $\mathcal{E} = D_i$  (death for cause  $i$ ) then remove the individual  $(\epsilon^I, a^I)$  from  $\mathbf{Z}_{T+\tau}$ .

(iii) If  $\mathcal{E} = O$  (no event) then nothing happens.

Let  $T \leftarrow T + \tau$  and go to **Step 1**.

This algorithm is simple and only involves the computation of uniform random variables which can be used to generate  $\tau, I, \mathcal{E}$  and  $\epsilon'$ . Note also that at each candidate time of event, only one individual is checked so that only one intensity function is computed. Remark that for numerical efficiency, the four random variables  $\tau, I, \mathcal{E}$  and  $\epsilon'$  needed can be generated using only one uniform random variable.

## 4.3 Data

We choose to apply our model to the French population with birth and death rates for the year 2008. We use data from the World Health Organization (WHO) for the cause-of-death mortality rates and the French National Institute for Statistics and Economic Studies (INSEE) for the birth rates and the initial age pyramid. The WHO has developed a tool to classify causes of death: the International Classification of Diseases (ICD). The database assembles deaths by cause of death, year, age and gender since approximately 1950 in about 225 countries. In this section, we present the data and the estimation methods for the mortality and fertility rates.

**Cause-of-death mortality rates** WHO provides the number of deaths by cause, gender and five-year age-class, and the exposure to risk by gender and age-class. We estimate death rates by the number of deaths over the exposure to risk (see e.g. Delwarde and Denuit (2006)). The exposure to risk corresponds to an estimation of the average time during which the population of interest is alive, here, the mid-year population. We denote  $D_i^\epsilon(a)$  the number of deaths due to cause  $i$  with gender  $\epsilon$  at age  $a$ , and  $ER^\epsilon(a)$  the exposure to risk for gender  $\epsilon$  and age  $a$ . Under the assumption of independence between cause-specific durations (see Section 4.2.1), the *net* force of mortality can be estimated from the data by

$$\hat{\mu}_i^\epsilon(a) = \frac{D_i^\epsilon(a)}{ER^\epsilon(a)}.$$

Data on deaths are provided for age-classes of five years, except classes "0", "1-4" and "95+", therefore we assume the mortality to be constant within each age-class. Moreover, the number of deaths by cause of death are classified until the age-class "95 years and more" while the data on exposures to risk are classified until the age-class "85 years and more". To use information on the number of deaths for age-classes "85-89", "90-94" and "95 years and more", we approximate the cause-of-death mortality rates using exposures to risk from the Human Mortality Database (HMD) from the corresponding age-classes\*.

In the International Classification of Diseases (ICD) for France, deaths are classified according to 4563 different cause labels. Based on previous papers (Alai et al. (2015), Arnold and Sherris (2015a)), we have chosen to regroup causes of death in six categories: cancers, diseases of the circulatory system, diseases of the respiratory system, external causes, infectious and parasitic diseases (in short: I&P) and all the others causes brought together, see details on Table 4.1. The three main causes of death are cancers, diseases of circulatory system and the category other causes.

---

\*We compared exposures to risk from HMD and WHO for the entire age range and year 2008. The relative difference between the exposures to risk is lower than 2%.

More precisely, other causes regroup mainly the diseases of the nervous system (21% of other causes), metabolic diseases (16%), diseases of the digestive system (13%) and mental disorders (12%). Infectious and parasitic diseases, diseases of respiratory system and external causes are less represented in the total deaths but represent an important part of deaths for some ages. For example, external causes represent 55% of deaths for individuals between ages 10 and 20, and 46% of deaths for individuals between ages 20 and 40. By comparison, death rates for external causes impact more specifically ages from ages 10 to 40, while cancers impact more generally ages above 40.

Cause label	%
Cancers	30 %
Other causes	28 %
Diseases of the circulatory system	27 %
Diseases of the respiratory system	6 %
External causes	7 %
Infectious and parasitic diseases	2 %
Total	100 %

Table 4.1: Deaths repartition by cause for the French population in 2008

On the whole, the different categories of causes of death impact differently the age-classes and genders. It is illustrated in Figure 4.1, which represents the death rates by cause, gender and age. As external causes mainly represent the accidents, it is interesting to note that the accident hump of aggregate mortality, which impacts more males, is essentially reflecting these external causes. More generally, it is clear that causes of death impact very different age ranges. This suggests that a drop in some cause-of-death mortality would modify the population age structure in a very specific manner. This is studied in the next section.

**Fertility rates** Fertility rates by age for females are provided by the INSEE. As for the mortality rates, they are estimated by the ratio between the number of births by the age of the mother at birth and the exposure to risk (see INSEE (2011)). Let us denote  $B(f, a)$  the number of births by age  $a$  of the mother at birth and  $ER(f, a)$  the female exposure to risk at age  $a$ . The fertility rate is estimated by the INSEE as follows:

$$\hat{b}(a) = \frac{B(f, a)}{ER(f, a)}.$$

Birth rates by age of the mother at birth in 2008 for France are represented in Figure 4.2. We observe the particular form of the birth rates: the intensity is maximal around age 30 and the distribution is almost symmetric. On the whole,

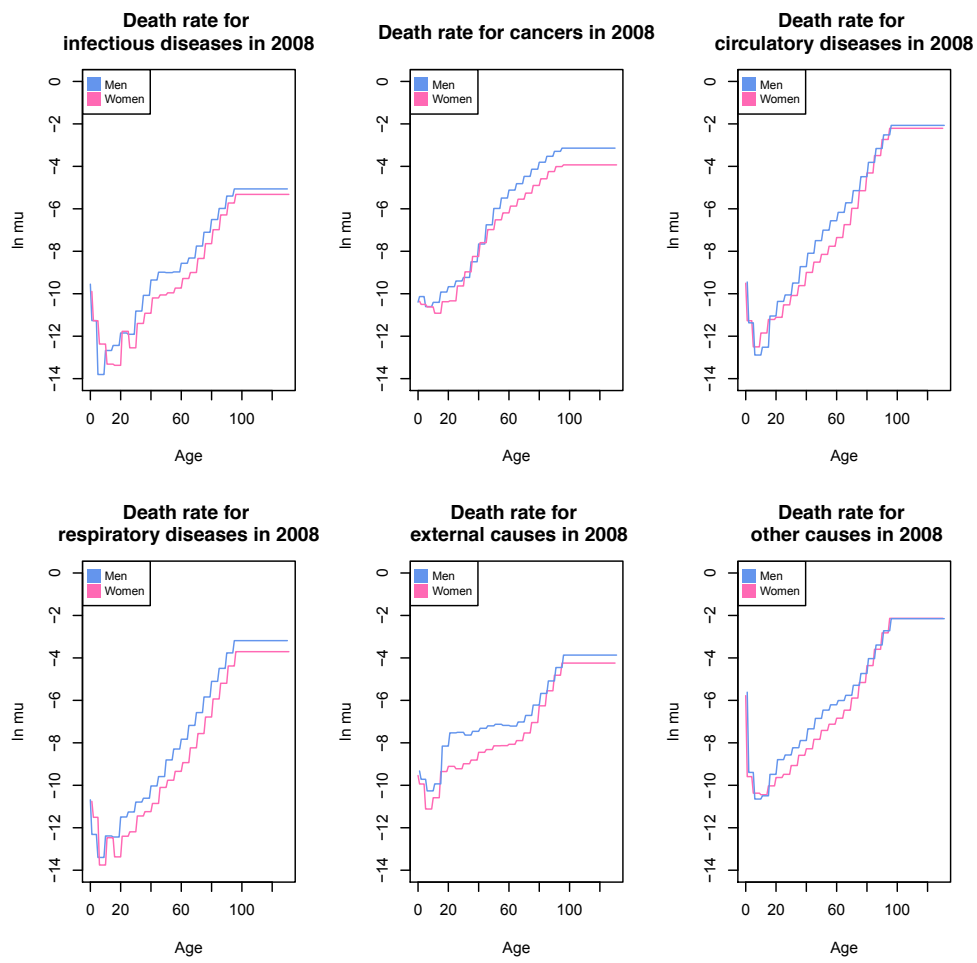


Figure 4.1: French death rates by cause, gender and age for year 2008

the age range to give birth is between ages 15 and 50.

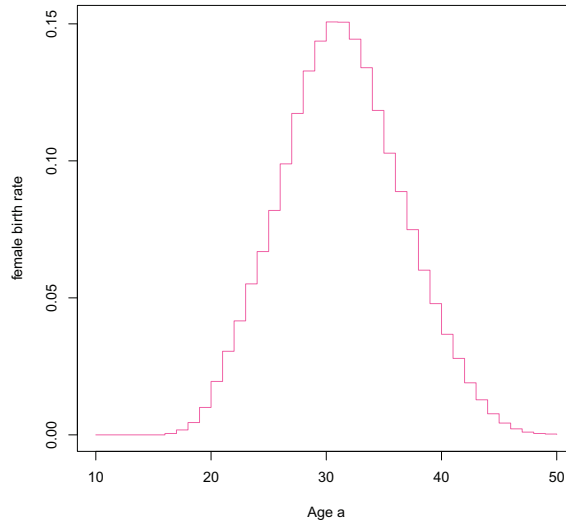


Figure 4.2: Birth rates by age of the mother in 2008

As detailed in Section 4.2, a newborn can be a female with probability  $p$  and a male with probability  $1-p$ . The probability  $p$  is derived from the data on the French population age pyramid, provided by the INSEE. We estimate the probability  $p$  as the ratio between the number of female births in year 2008 and the total number of births in year 2008. We estimate that in France, in 2008, the probability to be born as a female is equal to  $\hat{p} = 0.4886$ .

## 4.4 Results

In this section, we present the results of the simulations of the stochastic scenarios and the deterministic large population model. In the first part, we simulate the population dynamics while considering all causes of death in order to get some reference scenario. The time evolution of the dependency ratio (i.e. the number of individuals aged above 65 divided by the individuals between 15 and 65) already provides many insights regarding the long term impact of the initial population and the level of current birth and death rates. In the second part, we remove each cause of death to test the impact on the population age structure. As expected, the higher the cause-of-death mortality, the bigger the impact on the age pyramid. Finally, in the third part, we study the impact of a cause-specific reduction on the population age structure. We focus in particular on cancer and external causes and investigate how the age pyramid is affected if each cause is modified to get the same improvement in life expectancy.

### 4.4.1 Reference Scenario

**Stochastic scenario** First, we model the population dynamics from 2008 to 2108, considering the general mortality. We perform the simulation of a single stochastic scenario from the microscopic model described in Section 4.2.2, based on Algorithm 4.1 and an initial population with size 100 000. On Figure 4.3, we represent the 2008 initial population with grey lines and the final 2108 population in color (males in blue, females in pink). As explained in Section 4.3, the initial population is based on the structure of the French population. In the initial population, we observe more individuals between ages 40 and 70 representing the "Baby Boom" generation. In 2108, we see that the population is almost globally uniformly distributed up to age 70, and then decreases in an exponential form. This shape is stable and will be analysed in the deterministic scenarios section. The aging of the population is visible on Figure 4.4 which represents a path of the age dependency ratio with an initial population of 100 000 individuals. The age dependency ratio is increasing until the individuals from the "Baby Boom" generation die and then becomes stable due to time-independent birth and death rates. With this first analysis, we capture the order of magnitude of the dependency ratio and we address the impact in time of the "Baby Boom".

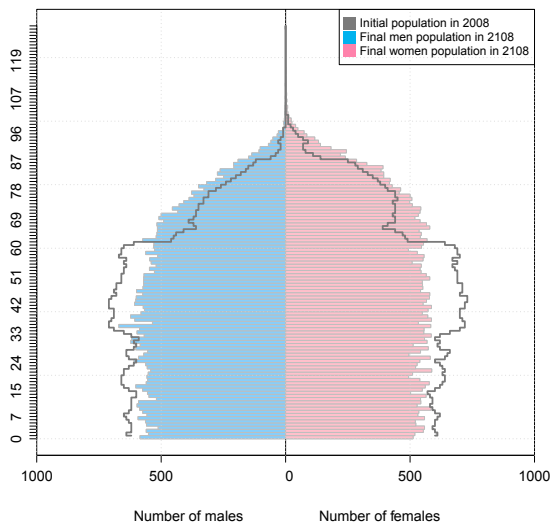


Figure 4.3: Initial (100 000 individuals) and final population pyramids

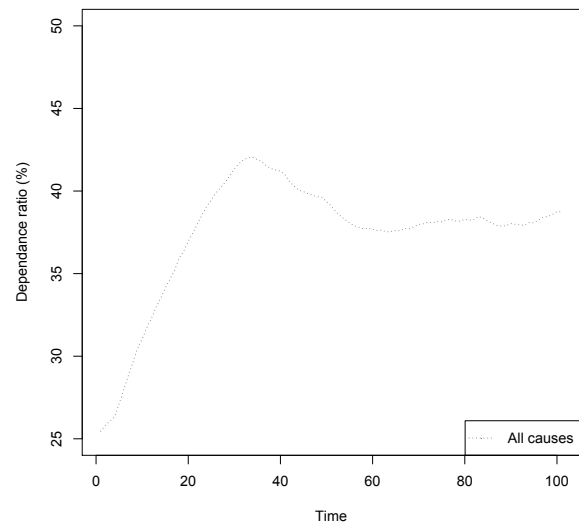


Figure 4.4: Age dependency ratio from 2008 to 2108

**Deterministic scenario** The large population equation (4.4)-(4.5) allows to compute the deterministic evolution of the population, in particular the whole age pyramid and the old-age dependency ratio. Here, numerical insights about the stochastic behavior cannot be addressed, but the virtue is that the numerical computation is far faster, as it amounts to discretize a partial differential equation (see Section 4.2.2).

We first plot in Figure 4.5 the evolution of the dependency ratio for 100 years. This represents, up to the discretization error, the evolution of the dependency ratio under the large population framework. Note that it is the dependency ratio for an "infinite" population, and not the average dependency ratio of any finite-size population. From the large population equation, it is possible to draw the deterministic scenario in the long time. On Figure 4.6, we represent the deterministic evolution of the dependency ratio over 300 years. Obviously, this time horizon does not make sense in itself, in particular regarding the assumption of time-independent death rates. But it provides the value of the asymptotically stable age dependency ratio. The long-term stabilization of the age pyramid is a well known fact in demography in a time-invariant rates setting. In the Appendix, we sketch the main results on the stable age pyramid in the deterministic model and the closed-formulas for its shape.

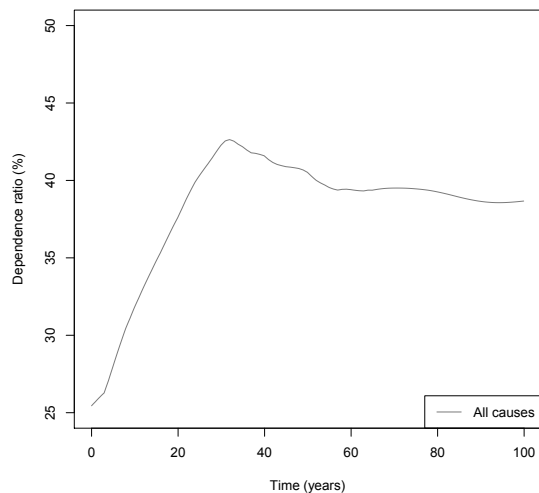


Figure 4.5: Age dependency ratio from the limit equation from 2008 to 2108

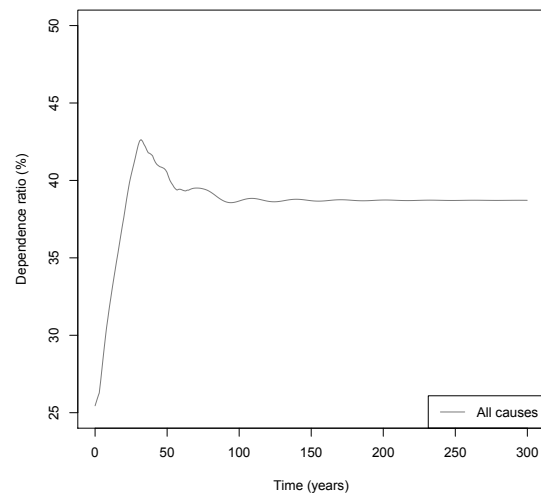


Figure 4.6: Age dependency ratio from the limit equation from 2008 to 2308

#### 4.4.2 Cause-of-Death Removal

Based on numerical simulations of the microscopic model, we test the impact of removing each of the six causes of death. We now start with an initial population of size 10 000 to allow the computation of several stochastic scenarios and thus gain insights on the demographic fluctuations. The impacts of cause-removal on the old-age dependency ratio are plotted in Figures 4.7a to 4.7f, with 10 scenarios. As expected, the causes with low death rates (infectious and parasitic diseases, respiratory diseases, external causes) lead to lower differences with respect to the reference scenarios. Besides, these differences are partly hidden with the demographic stochasticity. On the whole, comparisons between cause removal scenarios are not easy since the importance of each cause is quite different. Moreover, complete re-



removal of one important cause can be considered as unrealistic. In the following, we focus on cancer and external causes and investigate how the age pyramid is affected if each cause is modified to get the same improvement in life expectancy.

### 4.4.3 Cause-of-Death Reduction

For this numerical experiment, we focus on the two following causes: cancer and external causes. They are known to have a different mortality structure: cancer has an important impact on ages above 40, whereas external causes embed the accident hump and impacts mainly ages around 25 (see Figure 4.1). Therefore, we want to compare several scenarios:

(i) The first one is the reference scenario, computed with current birth and cause-of-death rates. Life expectancies at birth are 84.4 and 77.7 years for females and males respectively.

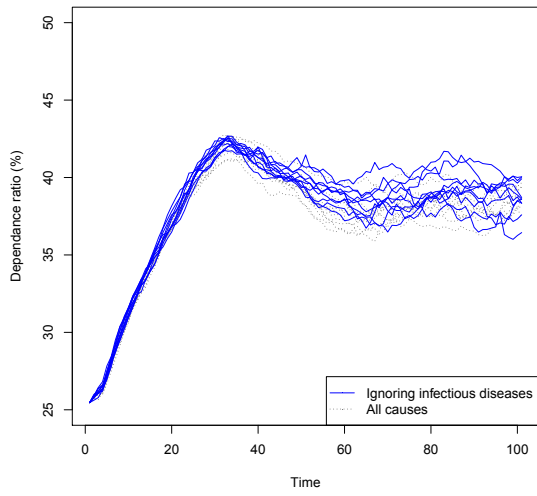
(ii) The second one is obtained by removing external causes. Under the assumption of independent causes of death, life expectancies at birth are 85.1 and 79.1 years for females and males respectively.

(iii) The last scenario is obtained by cancer reduction in the following way: each female and male reduction factor,  $\alpha^f$  and  $\alpha^m$  respectively (see Section 4.2.1), is computed so as to replicate the same life expectancy as in the scenario of external causes removal. This implies a reduction of cancer death rates by  $\alpha^f = 24.6\%$  and  $\alpha^m = 35\%$  for females and males respectively.

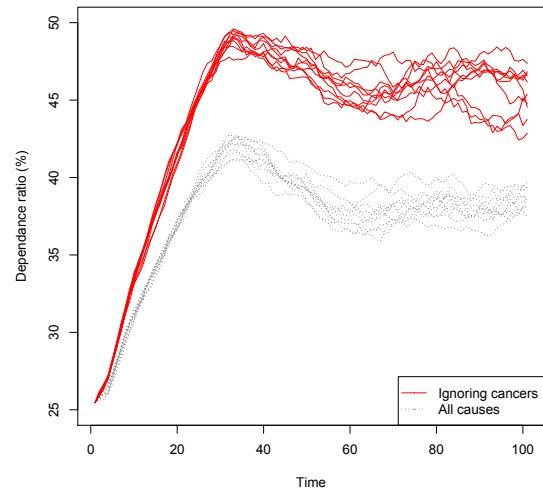
Comparing (ii) and (iii), we are interested in the evolution of the age pyramid knowing that life expectancies are equivalent in the two scenarios.

**Deterministic scenarios** The deterministic scenario can be computed using the large population equation. In Figure 4.8, the evolution of the age dependency ratio is plotted under the three parameter configurations (i), (ii) and (iii). The scenario (ii) removing the external causes is presented, and compared to the scenario (iii) of cancer reduction. Recall that males and females life expectancies are equivalent in both scenarios (ii) and (iii). We are then interested in analysing the impact of each scenario on the population age pyramid. In all cases, the age dependency ratio is higher than the reference scenario. However, the cancer reduction scenario leads to a higher dependency ratio compared to that of external causes reduction. This is explained by the fact that mortality due to cancer impacts mainly higher ages, especially ages at which people do not give birth anymore, so that the effect is only to extend life. As for the mortality due to external causes, it mainly impacts ages at which people can give birth. The removal of external causes has two main effects: as for cancer, the first one is to make people live longer, but at the same time many people who did not die from external causes (e.g. accidents) at young ages will give

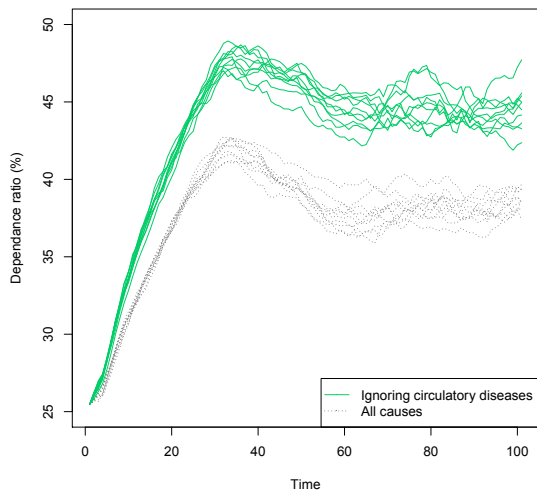
Figure 4.7: Age dependency ratio from 2008 to 2108



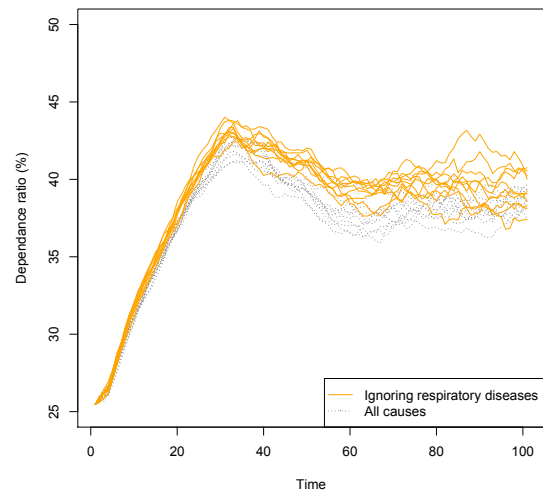
(a) I&P diseases removal



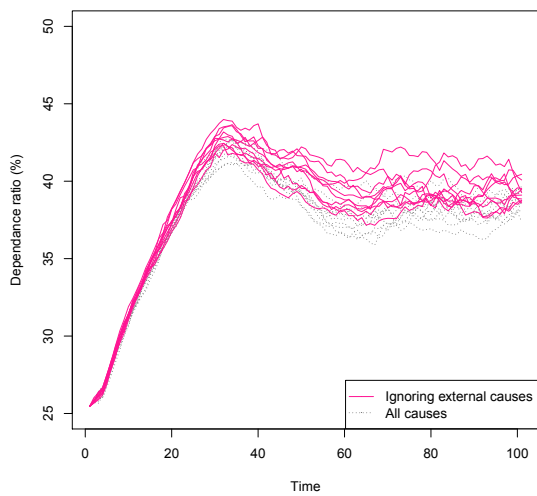
(b) Cancers removal



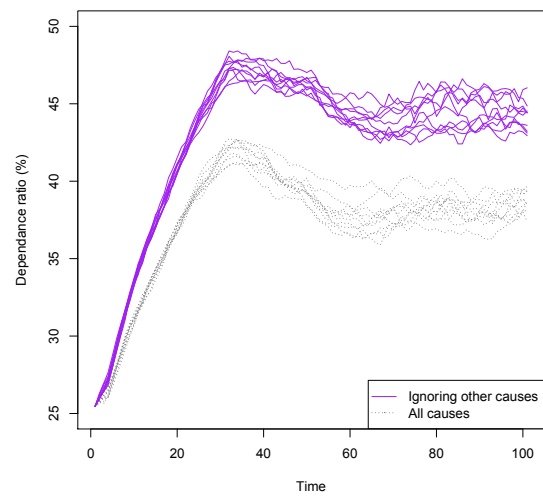
(c) Circulatory diseases removal



(d) Respiratory diseases removal



(e) External causes removal



(f) Other causes removal

birth. This population renewal increases the proportion of young people and makes the age dependency ratio lower than that with a reduction in cancer. One can find other demographic studies dealing with the counter-intuitive fact that a reduction in mortality do not necessarily lead to older populations; it can even lead to younger populations (see e.g. Keyfitz (1968)). As we see, the impact on the whole population depends on the ages at which mortality decreases. Different causes impact different ages and thus different populations will result when some causes are reduced or removed.

**Remark 4.2.** *Note that external causes impact mainly men, whereas our model assumptions result in a number of births depending only on the female population. In case of removal of external causes, the number of births will increase due to the part of the female population that do not die from external causes, but this could be reinforced if a birth rate would have been defined for males.*

In Figure 4.9, the time evolution of the population size under these three scenarios is plotted. It is interesting to note that, even if the age composition becomes stable over time, the population size is still decreasing. This is not consistent with real demographic insights but this can be explained by the fact the migration flows are not taken into account in these simulations. This is tested and discussed in the following subsection. As expected, the population size is higher under cause removal scenarios. In the short term, cancer reduction leads to a bigger population since more individuals survive. However, in the medium-long term, the removal of external causes leads to a higher number of individuals due to the additional effect of the increasing number of births. On the whole, the size evolution depends on the cause which is reduced.

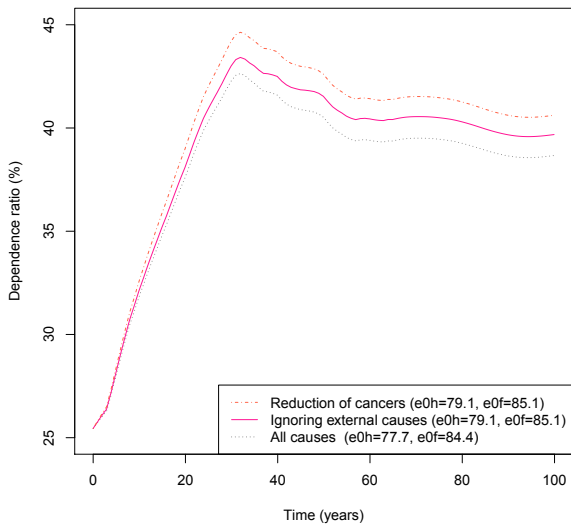


Figure 4.8: Age dependency ratio over 100 years from the limit equation

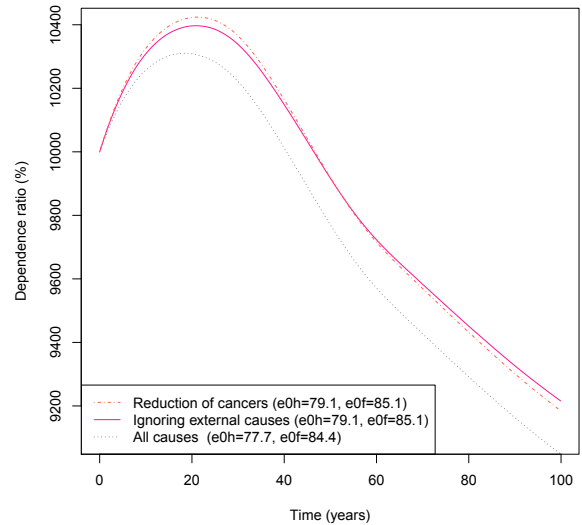


Figure 4.9: Population size for 100 years from the limit equation

**Stochastic scenarios** Stochastic scenarios are interesting to get additional insights on the random pattern of the dependency ratio and the magnitude of the demographic noise linked to the finite size of the population. Scenario (ii) of external causes of death removal is depicted in Figure 4.10, whereas scenario (iii) (cancer reduction) is added in Figure 4.11. All stochastic scenarios are computed starting with an initial population with size 10 000. For such size, to distinguish both scenarios is a hard task, making previous conclusions less striking with small samples. However, we argue that real national population sizes are more in accordance with "large population" patterns (see Figure 4.4 for an initial population size of 100 000). Unfortunately, testing it numerically would require computer resources that we do not have, and is beyond the scope of the present paper.

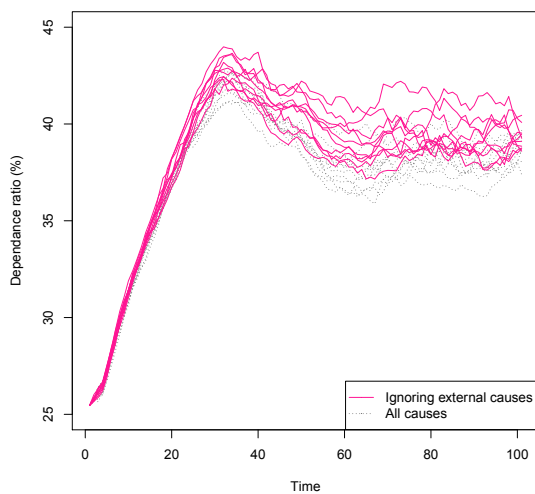


Figure 4.10: Age dependency ratio from 2008 to 2108: External causes removal (10 scenarios)

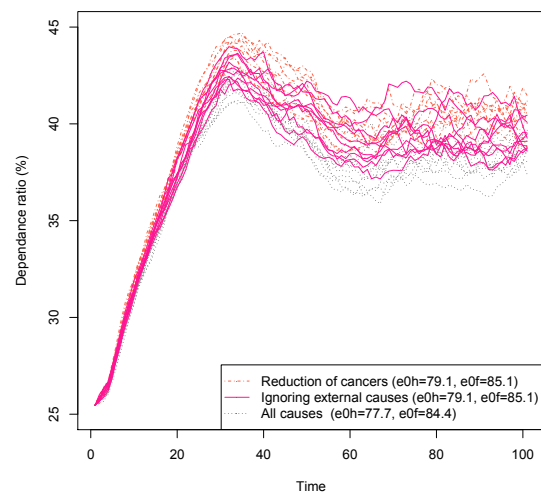


Figure 4.11: Age dependency ratio from 2008 to 2108: reducing cancers and external causes removal (10 scenarios)

#### 4.4.4 Sensitivity to Population Flows and Fertility

We address here the sensitivity of our results to several assumptions on migration flows and fertility rates. We emphasize that population flows are computed here in a first basic approach, therefore we keep in mind that what we call "migration" is not measured properly in terms of pure demographic analysis. However, our approach is made so as to replicate the data and allows us to observe several interesting patterns. We first detail how the migration rates are computed and then present nine scenarios with several fertility and migration assumptions. On the whole, our results persist with the inclusion of migration at several levels: the dependency ratios differ between the two scenarios (cancer reduction and external causes removal) even in the short term, making our findings stronger when population flow effects are taken

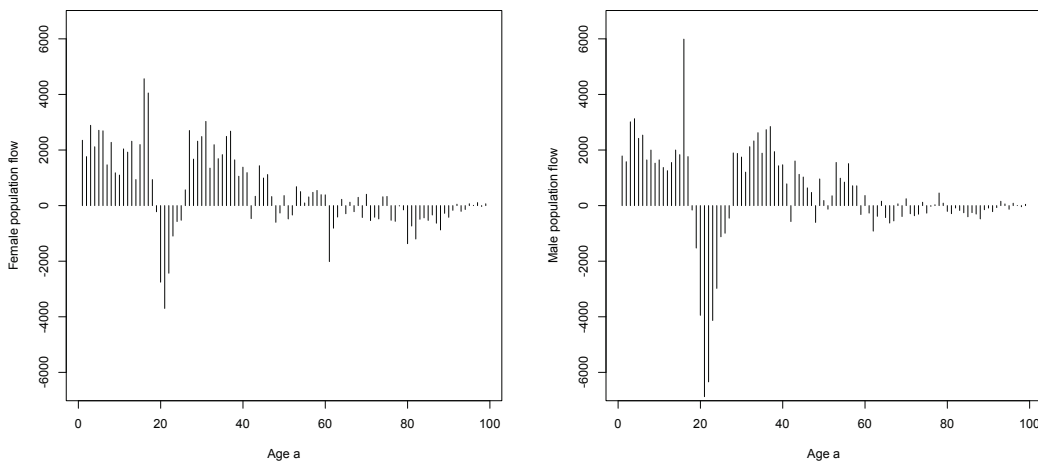
into account.

**Population flow estimation** As for the whole study, we focus on year 2008 for the computation of migration indicators. Migration flows are computed as the difference between the population at the end and the beginning of 2008, augmented by the number of deaths during this year. We approximate the population at the end of the year 2008 by the population at the beginning of year 2009. Let us denote  $M^\epsilon(a)$  the population flow per gender  $\epsilon$  and age  $a$  in year 2008. It is computed as follows:

$$M^\epsilon(a) = N_{2009}^\epsilon(a+1) - N_{2008}^\epsilon(a) + D^\epsilon(a),$$

where  $N_t^\epsilon(a)$  is the number of individuals with gender  $\epsilon$  and age  $a$  at the beginning of year  $t$ , and  $D^\epsilon(a)$  is the number of deaths with gender  $\epsilon$  and age  $a$  during the year 2008. Since data on deaths are available until the one-year age-class 98\*, we perform this computation for ages below 98 and assume that ages greater than 99 do not participate into population flows. This is a reasonable assumption since at old ages, above 95, population flows are negligible. Estimated population flows  $M^\epsilon(a)$  by age class can be positive or negative. These are depicted in Figure 4.12. Let us denote  $M_+^\epsilon(a)$  (resp.  $M_-^\epsilon(a)$ ) the population flow if  $M^\epsilon(a)$  is positive (resp. negative), so that  $M^\epsilon(a) = M_+^\epsilon(a)\mathbf{1}_{M^\epsilon(a)\geq 0} - M_-^\epsilon(a)\mathbf{1}_{M^\epsilon(a)\leq 0}$ . This way, for each age-class, we do not capture the real number of emigrants or immigrants but only the surplus or deficit. By abuse of definition, we will call  $M_+^\epsilon(a)$  (resp.  $M_-^\epsilon(a)$ ) the number of immigrants (resp. emigrants). For the entire population in year 2008, population flows implied a population surplus of 97369 individuals (males and females).

Figure 4.12: Estimated population flows in 2008 for females (left) and males (right)



For the use of the population dynamics model, it remains to transform the population numbers  $M_+^\epsilon(a)$  and  $M_-^\epsilon(a)$  into intensities. Let us first consider emigration:

\*INSEE: [http://www.insee.fr/fr/themes/detail.asp?ref\\_id=ir-irsocsd20133](http://www.insee.fr/fr/themes/detail.asp?ref_id=ir-irsocsd20133)

we assume that the emigrant number  $M_-^\epsilon(a)$  is the result of the sum of emigration intensities of each individual in the population. That is, the individual emigration rate  $m^\epsilon(a)$  is estimated as  $\hat{m}^\epsilon(a) = \frac{M_-^\epsilon(a)}{ER^\epsilon(a)}$ , where  $ER^\epsilon(a)$  is the exposure to risk at age  $a$  for individuals with gender  $\epsilon$  in the 2008 population. As for immigration, we want to capture a total intensity, say  $I_{nat}^\epsilon(a)$ , of arrival of an individual with age  $a$  and gender  $\epsilon$  in the reference national 2008 population. Since the intensity  $I_{nat}^\epsilon(a)$  represents the average number of arrivals of individuals in age-class  $a$  over one year, an estimator for this quantity is nothing but the number of immigrants in the year 2008, that is  $\hat{I}_{nat}^\epsilon(a) = M_+^\epsilon(a)$ . However, it is natural to think that the immigration flows depend on the size of the population at destination. Therefore we assume a linear link with the initial population, so that if we start with an initial population with size  $N^\epsilon$  for gender  $\epsilon$ , we use the immigration rate  $\hat{I}^\epsilon(a) = \frac{N^\epsilon}{N_{nat}^\epsilon} \hat{I}_{nat}^\epsilon(a)$ , where  $N_{nat}^\epsilon$  is the size of the 2008 national population. Note here that the total immigration rate is adjusted once at the beginning of the simulation and does not vary with time: this is consistent to capture the proper order of magnitude but also not to make it depend on the year to year variations of the population size, since in practice there is no obvious demographic reason to do so.

**Large population equation** The sensitivities will be studied with the large population equation, including population flows. This is written as

$$\begin{aligned} (\partial_a + \partial_t)\mathbf{g}(a, t) &= -\tilde{Q}(a)\mathbf{g}(a, t) + \begin{pmatrix} I^f(a) \\ I^m(a) \end{pmatrix}, \\ \mathbf{g}(0, t) &= \left( \int_{\mathbb{R}_+} g(f, a, t)b(a) da \right) \begin{pmatrix} p \\ 1 - p \end{pmatrix}, \end{aligned}$$

where  $\tilde{Q}(a) = \begin{pmatrix} \mu^f(a) + m^f(a) & 0 \\ 0 & \mu^m(a) + m^m(a) \end{pmatrix}$ .

In this setting, the aging population is decreased by the number of deaths and the number of emigrants, and is increased by the number of immigrants. The dynamics of the births is the same as the case without population flows: the number of newborn is computed as the sum of all newborn of all females, weighted by gender probability at birth. The numerical scheme is a straightforward adaptation of that described in Section 4.2.2.

**Results** We present nine deterministic scenarios with several fertility and migration assumptions. The scenario with current fertility and migration is displayed in Figure 4.13e. In this reference scenario, the age dependency ratio stabilizes around 26% in the mid-term, and then increases to 30% in the long term. Note that previously, without population flows, the dependency ratio was around 40% in the mid

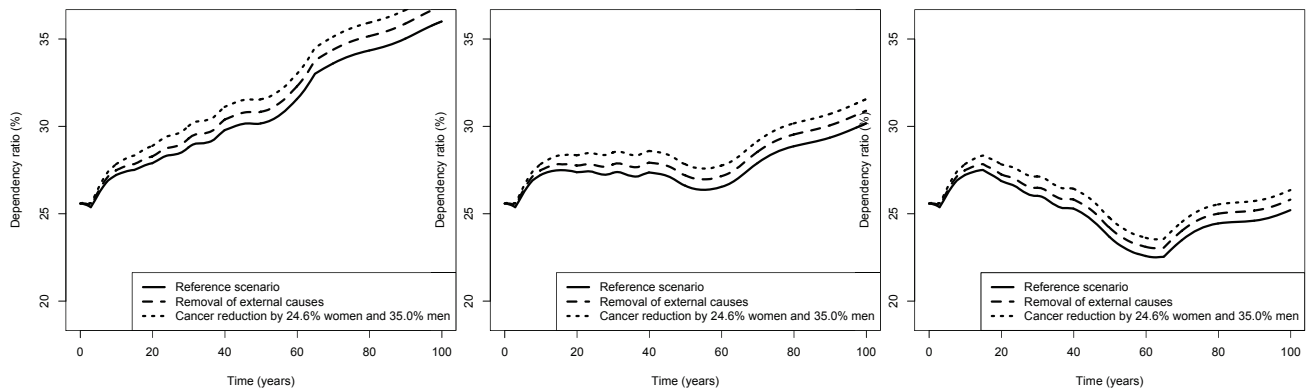
and long-term (see Figure 4.8). In Figure 4.13e, the two other scenarios (ii) and (iii) are also depicted. Interestingly, cause reduction scenarios exhibit different dependency ratios levels in the short term, therefore making our findings more relevant when population flows replicating the data are taken into account. The same conclusion holds for the other graphs, which present variations of fertility and migration rates (variation of 25% of the current rates). Fertility scenarios can be read from left to right: 75% , 100% and 125% of the current fertility rate, whereas the migration scenarios are displayed from top to bottom: 75%, 100% and 125% of the current migration rates. As expected, the higher are the fertility and migration, the lower is the dependency ratio (see Figure 4.13i). Moreover, we note a sensitivity more important on fertility (left to right) compared to migration (top to bottom). On the whole, these results including population flows seems to be more in accordance with real demographic insights.

## 4.5 Concluding Remarks

In this paper, we studied the impact of a decrease in cause-of-death mortality on the whole population dynamics, including birth patterns and population flows. The study of the whole population structure gives additional information compared to previous studies on causes of death focusing only on mortality indicators. We based our study on a stochastic individual based model which is micro-macro consistent. This allowed us to compute stochastic and deterministic scenarios, and to study both the demographic noise in the scenarios and the stable average age pyramid. Interestingly, as causes of death impact different age ranges, the impact on the age pyramid can be different. Numerical simulations based on French data illustrate that even if life expectancy improves in the same way, the reduction of some causes of death can lead to populations that are older (cancer) or younger (external causes) than expected. One can think of many applications of our model, including the study of pension systems. Indeed, the evolution of the age dependency ratio impacts directly pay-as-you-go pension systems, whose stability is often monitored through the age dependency ratio. This model can also help to study impacts on the population of public awareness campaigns on cancers or road safety for example. Our modeling framework is rather general and flexible, and it is possible at this step to further investigate the impact of several assumptions regarding cause-of-death mortality, fertility and migration. In particular, further improvements of our study could be to take into account different dependence assumptions between causes of death at the individual level.

#### 4.5. CONCLUDING REMARKS

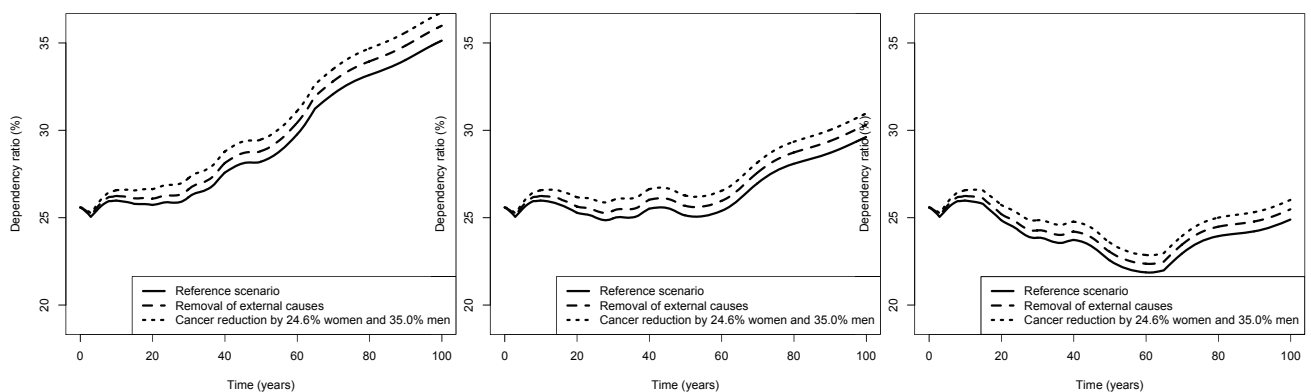
Figure 4.13: Age dependency ratio from 2008 to 2108 with variation of fertility and migration



(a) Fertility 75%, Migration 75%

(b) Fertility 100%, Migration 75%

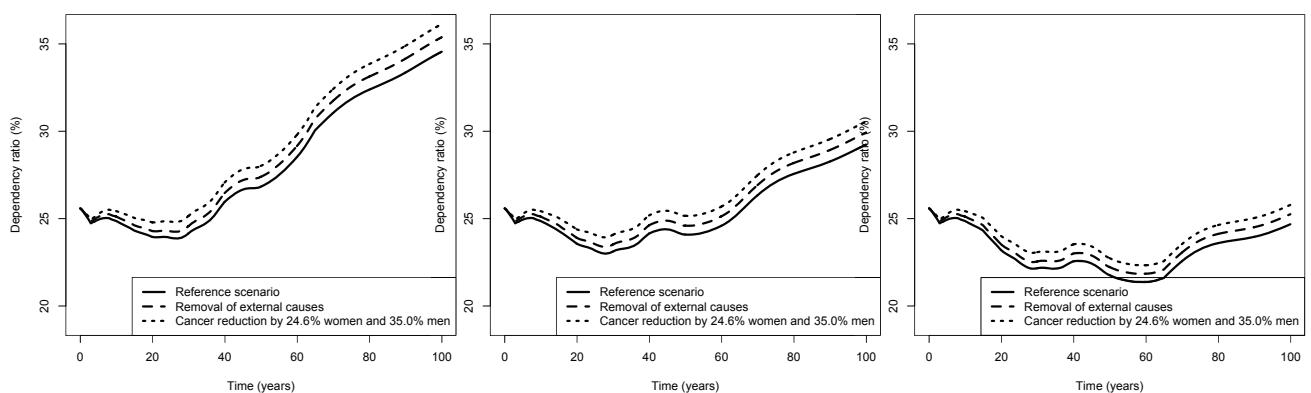
(c) Fertility 125%, Migration 75%



(d) Fertility 75%, Migration 100%

(e) Fertility 100%, Migration 100%

(f) Fertility 125%, Migration 100%



(g) Fertility 75%, Migration 125%

(h) Fertility 100%, Migration 125%

(i) Fertility 125%, Migration 125%



## Acknowledgements

The authors thank Frédéric Vautrain, founder of Isthma, for the free use of the tool Deltametric that helped to analyse the data. The authors benefited from the financial support of the "Germaine de Staël" program *Modélisation des causes de décès dans la dynamique des populations* (Gds 2014-14). Alexandre Boumezoued and Nicole El Karoui also benefited from the financial support by ANR project *Dynamic models for longevity with lifestyle adjustments*, Lolita (ANR-13-BS01-0011).

## Appendix

We present the asymptotic behavior of the deterministic age pyramid (see Iannelli et al. (2005)). Let us first focus on the dynamics of the number of females derived from Equations (4.4)-(4.5):

$$\begin{aligned}(\partial_a g + \partial_t g)(f, a, t) &= -\mu^f(a) g(f, a, t) \\ g(f, 0, t) &= p \int_{\mathbb{R}_+} g(f, a, t) b(a) da \\ g(f, a, 0) &= g_0(f, a).\end{aligned}$$

In the deterministic model, the number of children per female is

$$R_0^f = \int_{\mathbb{R}_+} b(a) e^{-\int_0^a \mu^f(s) ds} da.$$

Then, the mean number of children per individual is given by  $R_0 = pR_0^f$ . With our data, in the reference scenario, we find  $R_0^f = 1.97$  and  $R_0 = 0.964$ . Note that the first value  $R_0^f$  is of the same order of magnitude as the fertility statistics for France. Note also that we are in the case  $R_0 < 1$  so that the average population size decreases after some time and goes to zero as time goes to infinity. In Figure 4.9, one can see that the speed of decrease for the population size depends on the cause of death which is under reduction. But up to some scaling, one can be more precise about the age pyramid long term behavior. Let us introduce the characteristic equation in  $\lambda$ ,

$$p \int_{\mathbb{R}_+} b(a) e^{-\lambda a - \int_0^a \mu^f(s) ds} da = 1.$$

The unique real solution  $\lambda^*$  (say) is called the population intrinsic growth rate. Since we have  $R_0 < 1$ , then  $\lambda^* < 0$ . Numerically, we find  $\lambda^* = -0.00123$  (reference scenario) and we are in the sub-critical regime. It can be interpreted as follows: this is the constant to be added to the death rate to get exactly one children per individual, in other words to make the dynamics critical. One can now focus on the long term behavior of  $e^{-\lambda^* t} g(f, a, t)$ : this is known as the Sharpe-Lotka theorem.

It states that the long-term age pyramid density for the female population can be written, up to a renormalizing constant, as

$$\phi(f, a) = e^{-\lambda^* a - \int_0^a \mu^f(s) ds}.$$

Now, from Equations (4.4)-(4.5), it can be shown that  $e^{-\lambda^* t} g(m, a, t)$  converges up to a normalizing constant to  $(1 - p) \int_{\mathbb{R}_+} \phi(f, a) b(a) da$ , and since  $g(m, a, t + a) = g(m, 0, t) e^{-\int_0^a \mu^m(s) ds}$ , we formally get

$$e^{-\lambda^*(t+a)} g(m, a, t + a) = e^{-\lambda^* t} g(m, 0, t) e^{-\lambda^* a - \int_0^a \mu^m(s) ds},$$

which converges as  $t \rightarrow +\infty$  up to a constant to  $e^{-\lambda^* a - \int_0^a \mu^m(s) ds}$ . In other words, the age pyramid density for males is given up to a constant by

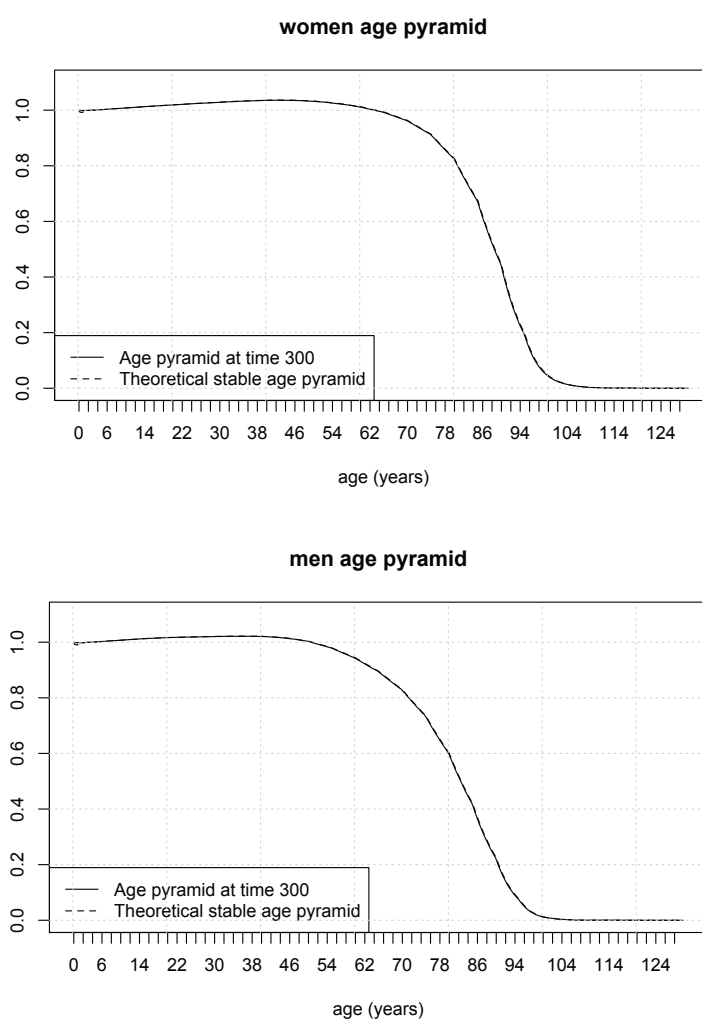
$$\phi(m, a) = e^{-\lambda^* a - \int_0^a \mu^m(s) ds}.$$

So the age pyramid for males and female is governed both by mortality and natality, since the solution  $\lambda^*$  depends on these two. To conclude, the overall average age pyramid density is given up to a renormalizing constant by

$$p\phi(m, a) + (1 - p)\phi(f, a).$$

The comparison between the average age pyramid at time 300 years and the theoretical age pyramid is given in Figure 4.14.

Figure 4.14: Comparison between the theoretical age pyramid and the age repartition at time 300 for females and males.



## Chapter 5

# Measuring mortality heterogeneity dynamics with interval-censored data<sup>†</sup>

### Abstract

In this paper, our aim is to measure mortality rates which are specific to individual observable factors when these can change during life. The study is based on longitudinal data recording marital status and socio-professional features at census times, therefore the observation scheme is interval-censored since individual characteristics are only observed at isolated dates and transition times remain unknown. To this aim, we develop a parametric maximum likelihood estimation procedure for multi-state models that takes into account both interval-censoring and reversible transitions. This method, inspired by recent advances in the statistical literature, allows us to capture characteristic-specific mortality rates, in particular to recover the mortality compensation law at high ages, but also to capture the age pattern of characteristics changes. The dynamics of several population compositions is addressed, and allows us to give explanations on the pattern of aggregate mortality, as well as on the impact on typical life insurance products. Particular attention is devoted to characteristics changes and parameter uncertainty that are both crucial to take into account.

**Keywords:** Mortality heterogeneity, longevity risk, multi-state models, interval-censoring, parametric maximum likelihood.

---

<sup>†</sup>Boumezoued, A., N. El Karoui, S. Loisel 2015, Measuring mortality heterogeneity dynamics with interval-censored data. *Working paper*

## 5.1 Introduction

Mortality at the individual level is known to depend on several observable factors, among the first ones age and gender. At a deeper level, one can have access to datasets recording other observable characteristics, such as socio-economic features or marital status. In this context, to capture mortality differentials depending on several characteristics is a key issue to better assess longevity/mortality risk at several levels. This paper is indeed motivated by reserving purposes regarding specific insurance portfolios, but also by a *population dynamics* point of view. This point of view, detailed in Bensusan et al. (2010–2015), allows to better understand how aggregate mortality and other demographic quantities as the age pyramid evolve with an underlying dynamic heterogeneity, based on populations evolving due to characteristic-specific mortality rates but also birth rates, as well as characteristics changes during life (see also Boumezoued (2015a)).

From a statistical point of view, in the case where these characteristics are stable during life, classical survival analysis can be used to measure characteristic-specific mortality rates. However, when these characteristics can change over time, the model to be used is a multi-state model: a given individual switches from a state to another until death, which is an absorbing state. The first advantage of such model is to capture the characteristic-specific mortality rates but also the transition intensities from a characteristic to another. This provides a measure of the *level* of each possible transition in the model. Another advantage of such approach concerns *parameter uncertainty*: as the whole dynamics is captured, this approach provides the variance of the transition rates, as well as their correlations. This way, it is possible to take into account estimation error in the life trajectories of a given group of individuals.

To study life trajectories embedding characteristics changes, it is not possible to use the classical survival analysis estimators as Kaplan-Meier for the survival curve or Nelson-Aalen for the cumulative intensity. Indeed, such estimators will suffer from several statistical biases when applied to life trajectories with several intermediate states, since for example the user will proceed to an arbitrary distribution of individuals by mode, set once for all, or because the terminal event (death in most cases) is treated as independent censoring (see e.g. Joly et al. (2002)). The statistical estimation of continuously observed multi-state models, involving classical censoring and truncation, can be performed by the standard non-parametric estimator of Aalen-Johansen (see Aalen and Johansen (1978) and the book of Andersen et al. (1993)). Although the Aalen-Johansen estimator, as a generalization of the Nelson-Aalen estimator in a multi-state framework, can be used with several kinds of incomplete observations, its use requires to know the exact transition dates. However in practice, the multi-state process is most often observed at some isolated

times: indeed, in so-called longitudinal data related to medical visits or times of census, the state of the individual (marital status, social class, ...) is known at inspection dates only. Even if this gives some information, the transition times remain unknown: such observation scheme is called *interval censoring*. Note also that times of death are often exactly recorded, so that in fact the observation scheme is mixed: interval-censored intermediate states and exactly know absorbing state. Several demographic and actuarial studies in this field tackle this issue by limiting the scope to a discrete-time model, or making assumptions about the transition times in each interval.

In the demographic and actuarial literature, one can find several studies focusing on the statistical estimation of multi-state models in discrete time, or continuous time with exactly known transition times. Many of them focus on health insurance and Long Term Care (LTC). Gaüzère et al. (1999) estimate what is called an irreversible illness-death process, a process with states 1=healthy, 2=illness and 3=death, where only transitions  $1 \rightarrow 2$ ,  $1 \rightarrow 3$  and  $2 \rightarrow 3$  are allowed. Gaüzère et al. (1999) assume that transitions occur at the middle of the censoring interval, therefore uses the classical non-parametric framework of Aalen-Johansen since transition times are given by data modification. Czado and Rudolph (2002) estimates the transition intensities of multi-state model with Cox-proportional hazard model with known transition times, whereas Helms et al. (2005) proposes to directly compute the estimated transition probabilities by means of Aalen-Johansen estimator in this context. Also, Levantesi and Menziatti (2012) focus on a discrete-time irreversible illness-death model to study how transition probabilities change over time and measure the so-called systematic risk. Apart from specific LTC and health insurance issues, the study of mortality differentials in general life insurance has also gained recent attention for a better understanding of national mortality and a better assessment of longevity/mortality risk. For example, Kwon and Jones (2006) and Kwon and Jones (2008) calibrate a discrete-time multi-state model from Canadian health longitudinal data (Canadian National Population Health Survey), and study the impact of mortality differentials on typical life and health insurance mechanisms.

To our knowledge, no actuarial studies focused on the assessment of mortality differentials in the presence of interval censoring, despite this is the main characteristic of longitudinal data. In this context, our aim is to capture mortality differentials with continuous age by means of reversible multi-state models when the data is interval-censored. We first develop a parametric maximum likelihood estimation procedure, which is inspired by recent advances in the statistical literature, and second apply it to a representative sample of the French national population made by the French institute INSEE, called *Échantillon Démographique Permanent\**.

---

\*The authors thank H. Bensusan, as well as A. Frachot and A. Trognon for the access to a part

In the statistical literature, several contributions focused on multi-state model estimation with interval-censored data. On the whole, they can be classified depending on the type of method (parametric, semi-parametric, non-parametric), the class of multi-state models considered (illness-death reversible or not, competing risk, general multi-state) and the Markov assumption (Markov, semi-Markov, non-Markov). Since the seminal work of Kalbfleisch and Lawless (1985) in the case of constant intensities, methods have been developed for irreversible illness-death Markov models. In this context, one can find non-parametric approaches in e.g. Frydman (1995) and Frydman and Szarek (2009), and both semi-parametric and parametric methods in e.g. Joly et al. (2002), Commenges and Gégout-Petit (2007) and Commenges et al. (2007) (see also Foucher et al. (2007) and Touraine (2013)). As for reversible processes, one can find in the literature methods dedicated to special cases of multi-state models and intensities, as Kang and Lagakos (2007) and Titman and Sharples (2010) (see also Wei (2015)). Kang and Lagakos (2007) focused on maximum likelihood estimation for homogenous semi-Markov multi-state models, assuming that at least one transition intensity is constant. Titman and Sharples (2010) developed an alternative method based on phase-type waiting times and hidden Markov chain models. Recently, Wei (2015) introduced a method based on quasi-Monte Carlo methods applied to time-independent semi-Markov models.

The advantage of non-parametric methods is to overcome assumptions about the shape of the transition intensities, which is also the case to some extent for semi-parametric methods as smooth basis functions remain unspecified. This is particularly useful for applications for which we have no idea of the age structure of these intensities. Unfortunately, in the case of small samples and/or high number of interval-censored transition times, non-parametric methods may be unstable: the use of parametric methods then captures, for such kind of data, a maximum amount of information on the dynamics (see e.g. Foucher et al. (2007)). Let us emphasize that for several statistical studies, it is more reasonable to implement a parametric method taking into account interval censoring, rather than using a nonparametric method assuming that transition times are known (see e.g. Touraine et al. (2013)). Indeed, modifying the data by setting middle-interval transition times (or even uniformly distributed) leads to fix the age pattern of the transition intensities between groups: to have intensities with jumps at fixed points for middle-interval assumption, or to be constant on the interval for uniformly simulated transition times. In this context, parametric methods allows to capture the age pattern of transition rates while avoiding several biases due to data modification. Parametric methods are also relevant to include information that we have on the shape of the mortality rates, for example a Gompertz-type mortality rate at reasonable ages. In this con-

---

of this dataset.

text, and given the sample we want to study, we propose here a parametric approach that will allow us to include a maximum of a priori information on the shape of the intensity and reduce at best the dimension of the problem.

In the current statistical literature, parametric maximum likelihood procedures have not been developed for interval-censored data, when intensities depend on age or time, concerning Markov multi-state models which can have reversible transitions. So first, closely to e.g. Joly et al. (2002), Commenges et al. (2007), Foucher et al. (2007) and Touraine et al. (2013) who focus on irreversible processes, we develop such method. Second, we apply it to the French representative sample *Échantillon Démographique Permanent* to capture mortality forces which depend on socio-economic features or marital status, as well as transition intensities between the several groups. Special attention is given to parameter uncertainty, which can be different depending on the socio-economic group and the transition rate considered. This way, we are able to analyze the age-pattern of aggregate mortality which depends on the underlying sub-populations dynamics, as well as the impact on typical insurance products of heterogeneity and the associated level and uncertainty.

The remainder of this paper is organized as follows. In Section 5.2, we describe the longitudinal data we use as a basis for our estimation procedure. The parametric maximum likelihood method that takes into account interval-censoring and reversible transitions in the multi-state model is detailed in Section 5.3, and the results we obtain are described in Section 5.4. Finally, in Section 5.5 we study the age pattern of aggregate mortality and we analyze the impact of heterogeneity on typical life insurance products.

## 5.2 Longitudinal data

In this section, we describe the data we want to analyze, which makes the statistical method developed in the next section conditional upon it. The Permanent Demographic Sample (*Échantillon Démographique Permanent*) of the French institute INSEE is a longitudinal dataset which aims at observing a representative sample of the national population while recording several individual characteristics over time. It contains information of about 992,711 individuals observed since 1968 and born from 1862, the 1, 2, 3 or 4 of October. These individuals have been (potentially) observed at the census dates 1968, 1975, 1982, 1990 and 1999, and for each census, a set of characteristics was recorded: we here have access to socio-professional categories as well as marital status. Such a sample is useful to study the link between the level of mortality and individual characteristics, but also the way these characteristics change over time.



**Scope of the study** This part of the Permanent Demographic Sample that we have access to is a remarkable database, especially given its completeness, its focus on real cohorts and regarding the set of characteristics involved. Also, the detailed analysis of specific levels of mortality in sub-populations and changes in their characteristics during life is a statistical task that poses two major constraints:

a) The size of the subpopulations considered has to be sufficient, which will guide us first to regroup several modalities of the classification of INSEE, and second to include one type of characteristics (socio-professional category or marital status) at a time.

b) The observation of the life trajectories must be sufficiently repeated over time, which will push us to consider specific cohorts making the five censuses (1968, 1975, 1982, 1990, 1999) available and exploitable given the characteristic considered.

Finally, note that the interval-censoring mechanism involved here has the particularity to be common to all individuals. Therefore the age at which a given individual is observed depends on his/her date of birth. We are here in the case of *dependent interval-censoring* which is a statistical framework that is beyond the scope of the present paper. Therefore, we focus on specific cohorts and taking into account several cohorts at a time is left for further research. In the following, we detail the dataset we are interested in regarding the socio-professional groups as well as marital status.

**Socio-professional categories** For each year of census, (1968, 1975, 1982, 1990 and 1999), we have access to the socio-professional category of each observed individual. It is classified by the INSEE in a detailed way, therefore we choose to split them in two groups. The group 1 includes farmers, craftsmen, salesmen, low-skilled workers and people without work. The group 2 includes directors of 10 employees or more, managers, higher intellectual professions, middle management and employees. This classification, although arbitrary, will be useful to illustrate our methodology as well as to give interesting insights on the mortality differentials that are of interest for insurance purposes. In this study, we first focus on the male population of the 1930 birth cohort, which is the first to present a negligible number of students in 1968, which makes all times of census exploitable with the multi-state model described in Figure 5.1. This leads to a sample of 4266 individuals alive at the first census, observed or not. Group 1 of the 1930 birth cohort represents about 70 % of the population in 1968, and group 2 its complementary 30 %. In Figure 5.2, we represent the proportion of group 1 at each census. Over time, the composition of the cohort evolves according to two effects: first due to the fact that mortality for each group may be different, second due to characteristics changes during the

lifetime of individuals<sup>¶</sup>.

**Marital status** As well, we focus on marital status. Since we want to explore the impact of this characteristic at highest ages, we focus on the 1907 female birth cohort leading to observe individuals with age 60 at the first census. This leads to a sample of 3038 individuals alive at the first census, observed or not. The focus on marital status leads us to consider two groups. The first characteristic label regroups "single", "divorced" or "widowed", and the second characteristic label is "married". This choice is driven by the data which would not have been sufficient if we wanted to analyze transitions between "non-married" sub-groups. Although we lose some information on the original dataset, we are able to analyze the impact of being married or not, and the transitions between both, all of them being crucial for insurance purposes. The multi-state dynamics therefore considered is depicted in Figure 5.1. In Figure 5.3, we represent the proportion of non-married individuals at each census. As well, recall that the cohort composition evolves first due to the fact that mortality for each group may be different, and second due to characteristics changes during the lifetime of individuals.

With these two possible classifications, we want to illustrate two facts: heterogeneity implies different mortality levels, and also different orders of magnitude of uncertainty around these levels. By the study of the dynamics of life trajectories within a cohort, it is possible to get further insights on the age pattern of aggregate mortality. We also want to illustrate that this is crucial to take into account characteristics changes during life, in particular when computing the price of typical life insurance products. Note that we estimate mortality forces by age for a *real* cohort, which contrasts with many actuarial studies that compute transition probabilities based on a mix of all age classes at a given year, therefore does not quantify realistic age patterns.

**Remark 5.1.** *(On missing data) This dataset mentions all observed deaths before the year 2008, the end of the follow up. Therefore, if an individual is observed at the last census, and no date of death is given, then it is reasonable to think that he/she is alive at the end of the follow up. However, if both date of death and last census are missing, there is a small probability that the individual trajectory has been lost. In the case where no date of death is mentioned and several censuses at the end are missing, then to avoid any bias we assume that the trajectory is right-censored at the last observed census. This data modification is needed to make sure that no wrong information is added in the model, although it slightly increases the number*

---

<sup>¶</sup>As well, migration flows may have a small impact in such a way that they make some census unavailable, and these are treated as independent classical censoring or truncation.

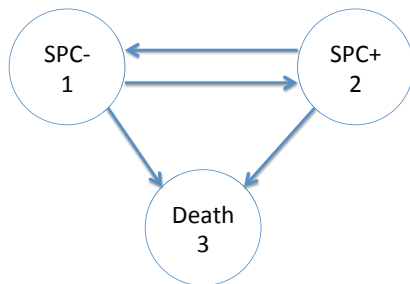


Figure 5.1: Multi-state models considered, for Socio-Professional Categories (left) or marital status (right)

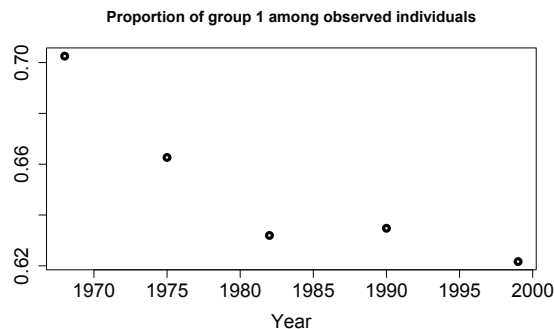


Figure 5.2: Proportion of individuals labelled 1 (SPC-) among observed individuals at each census in the 1930 birth cohort

### 5.3 Parametric maximum likelihood method for interval-censored data

In this section, the aim is to express the likelihood associated with the interval-censored observation of the Markov multi-state models depicted in Figure 5.1.

#### 5.3.1 Likelihood derivation

**The Markov process** Let us consider the Markov multi-state models depicted in Figure 5.1 with intermediate states 1 and 2, and absorbing state 3. Let us denote  $(X_t)$  such process with state space  $\{1, 2, 3\}$ , and let  $\alpha_{kl}(t)$  be the transition intensity from state  $k$  to state  $l$ . Note that the time component  $t$  represents the age of the individual. The process starts at initial age  $t = \underline{a}$  in state 1 with probability  $p$ , in state 2 with probability  $1 - p$ . Let us also define  $\alpha_{11}(t) := \alpha_{12}(t) + \alpha_{13}(t)$  and

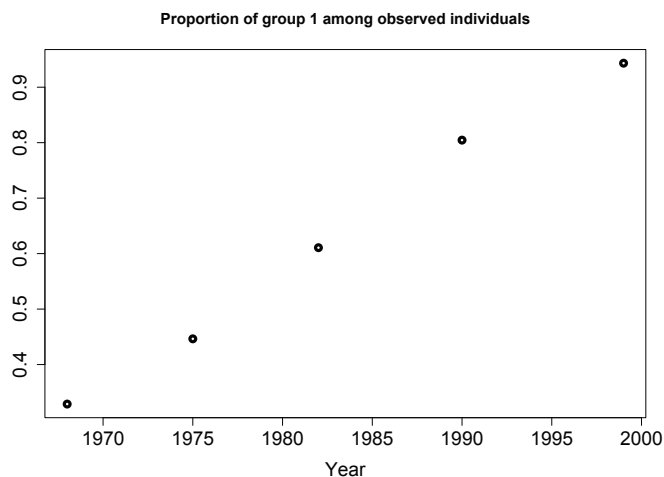


Figure 5.3: Proportion of individuals labelled 1 (non-married) among observed individuals at each census in the 1907 birth cohort

$\alpha_{22}(t) := \alpha_{21}(t) + \alpha_{23}(t)$  the waiting intensities in states 1 and 2 respectively. From any intensity  $\alpha_{kl}(t)$ , one defines the cumulative intensity as  $A_{kl}(s, t) = \int_s^t \alpha_{kl}(u) du$ . We are interested in the dynamics of the process until it reaches the absorbing state 3 at a random age denoted  $T$  (which is the lifetime).

**Observation scheme** The individual with life trajectory  $(X_t)$  is (potentially) observed at ages  $R_1, \dots, R_5$ , which correspond to times of census that are common to all individuals. Let  $\delta_i$  be the census indicator:  $\delta_i = 1$  if the individual has been observed at age  $R_i$ , whereas  $\delta_i = 0$  if not. We assume that the observation  $\delta_i$  is independent from the process  $(X_t)$ . Note that the last potential census time for a given individual is the one that precedes his/her age of death (i.e. his/her lifetime), denoted  $T$ . Note also that individual characteristics are not observed at time of death in the dataset. Let us denote  $\tau$  the age corresponding to the end of the follow-up: this is the age of death if  $T = \tau$ , in this case we denote  $d = 1$ , or this is a right-censoring time if  $\tau < T$ , which case we denote  $d = 0$ . Finally, let us characterize the set of exploitable censuses. Denote  $I$  the set of  $i$  such that  $\delta_i = 1$ ; note that for  $i \in I$ , we have  $R_i \leq \tau$ . Let  $\bar{I}$  the biggest element of  $I$ , that is the index of the last census for which the individual is observed, and  $\underline{I}$  the smallest, that is the index of the first exploitable census. Lastly, for an index  $i \in I$ , we denote  $i_+$  the following element. For example, if the individual is observed at the census  $i + 1$ , one has  $i_+ = i + 1$ .

**The likelihood** For a life-trajectory observed according to the observation scheme described above, whose state at time  $R_i$  is denoted  $x_i$  for each  $i \in I$ , the contribution

to the likelihood is, using the Markov property,

$$\mathbb{P}(X_{R_{\underline{I}}} = x_{\underline{I}}) \prod_{i \in I, i < \bar{I}} \mathbb{P}(X_{R_{i_+}} = x_{i_+} \mid X_{R_i} = x_i) \mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}) \mathbb{P}(T = \tau \mid T \geq \tau, X_{R_{\bar{I}}} = x_{\bar{I}})^d. \quad (5.1)$$

In the end, as we assume that life trajectories of individuals are independent, the total likelihood will be given as the product of all individual contributions. Let us now compute each elementary term in (5.1) separately.

(i) Let us start with the terms  $\mathbb{P}(X_{R_{i_+}} = x_{i_+} \mid X_{R_i} = x_i)$  which reflect the contribution to the likelihood of a trajectory between two observed censuses. Let us first treat the case where the individual was seen at the census following that in  $R_i$ , that is  $i_+ = i + 1$ , and where also the two censuses are different, that is  $x_i \neq x_{i_+}$ . For two consecutive censuses, we assume that only one transition may have occurred. This may seem as a strong assumption, but in fact, this seems reasonable to think that in time intervals of about 7 years, it is not likely that two or more transitions occur regarding the socio-professional category or marital status. Of course, this can not be tested in our data, but we argue that this approach is a step forward for such purpose, since it does not make any assumption about the transition times themselves. As such, we assume that in the dataset the life trajectory is piecewise irreversible.

In this context, the contribution to the likelihood which takes into account interval-censoring amounts to integrate over all possible transition times  $u \in [R_i, R_{i+1}]$ . It reflects the fact that it remains in the state  $x_i$  during the time interval  $[R_i, u]$ , that it instantaneously jumps at time  $u$ , and that it stays in state  $x_{i+1}$  during  $[u, R_{i+1}]$ , that is

$$\mathbb{P}(X_{R_{i+1}} = x_{i+1} \mid X_{R_i} = x_i) = \int_{R_i}^{R_{i+1}} e^{-A_{x_i x_i}(R_i, u)} \alpha_{x_i x_{i+1}}(u) e^{-A_{x_{i+1} x_{i+1}}(u, R_{i+1})} du. \quad (5.2)$$

Let us also treat the other case where the two consecutive observations are the same, i.e.  $x_{i+1} = x_i$ . Since we do not allow two transitions during a census interval, the only possibility is that the individual stayed in this state, that is

$$\mathbb{P}(X_{R_{i+1}} = x_i \mid X_{R_i} = x_i) = e^{-A_{x_i x_i}(R_i, R_{i+1})}. \quad (5.3)$$

Now, let us focus on the case where a census is missing between the two observed censuses  $i$  and  $i_+$ . In this configuration, we reduce to the previous cases by summing over all the possible values taken by the process at intermediate census. For example, if  $i_+ = i + 2$ , we get by conditioning and using the Markov property,

$$\mathbb{P}(X_{R_{i+2}} = x_{i+2} \mid X_{R_i} = x_i) = \sum_{k \in \{1, 2\}} \mathbb{P}(X_{R_{i+2}} = x_{i+2} \mid X_{R_{i+1}} = k) \mathbb{P}(X_{R_{i+1}} = k \mid X_{R_i} = x_i),$$

whereas if  $i_+ = i + 3$ , we get

$$\sum_{k \in \{1,2\}} \sum_{l \in \{1,2\}} \mathbb{P}(X_{R_{i+3}} = x_{i+3} \mid X_{R_{i+2}} = l) \mathbb{P}(X_{R_{i+2}} = l \mid X_{R_{i+1}} = k) \mathbb{P}(X_{R_{i+1}} = k \mid X_{R_i} = x_i),$$

which extends in the same way to the case  $i_+ = i + 4$ , the maximum case in our study.

(ii) Let us now focus on the term  $\mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}})$ , and first assume that  $\bar{I} = 5$  (case 1), i.e. that we are at the last possible census. We have to distinguish here between our two applications.

Let us start with socio-professional category. In our study, the retirement age is attained before the last census, therefore characteristics of individuals remain stable until death. The contribution to the likelihood amounts to compute the probability that the terminal event is not reached during the time interval  $[R_5, \tau]$ , that is

$$\mathbb{P}(T \geq \tau \mid X_{R_5} = x_5) = e^{-A_{x_5 3}(R_5, \tau)}. \quad (5.4)$$

Let us now focus on marital status. In this context, it is crucial to take into account the fact that at high ages after the last census, individuals can switch from married (group 2) to non-married (group 1) due to widow. Since transitions  $2 \rightarrow 1$  are allowed until death, and that in addition marital status at death is unknown, we have to explore two possibilities: either the individual stayed married and died, or switches to group 1 before death. Also, we assume that after the last census in 1999, no transition  $1 \rightarrow 2$  occurs (which is natural since individuals have age 91). The two possible contributions thus write

$$\mathbb{P}(T \geq \tau \mid X_{R_5} = 2) = e^{-A_{22}(R_5, \tau)} + \int_{R_5}^{\tau} e^{-A_{22}(R_5, u)} \alpha_{21}(u) e^{-A_{13}(u, \tau)} du, \quad (5.5)$$

and

$$\mathbb{P}(T \geq \tau \mid X_{R_5} = 1) = e^{-A_{13}(R_5, \tau)}. \quad (5.6)$$

Let us now assume that  $\bar{I} \leq 4$  (case 2). We do not distinguish between the two applications anymore. Remark that the time of the follow-up end  $\tau$  is always greater than  $R_{\bar{I}}$ . Let us first assume that  $\tau$  belongs to the time interval  $[R_{\bar{I}}, R_{\bar{I}+1}]$ . In this case, there are two possible trajectories: either the individual stayed in his/her state  $x_{\bar{I}}$  until time  $\tau$ , or he/she switched to another intermediate state before  $\tau$ . Let us denote  $y_{\bar{I}}$  the intermediate state which is different from  $x_{\bar{I}}$ , that is  $y_{\bar{I}} = 2$  if  $x_{\bar{I}} = 1$  and  $y_{\bar{I}} = 1$  if  $x_{\bar{I}} = 2$ . The contribution to the likelihood can then be written as

$$\mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}) = e^{-A_{x_{\bar{I}} x_{\bar{I}}}(R_{\bar{I}}, \tau)} + \int_{R_{\bar{I}}}^{\tau} e^{-A_{x_{\bar{I}} x_{\bar{I}}}(R_{\bar{I}}, u)} \alpha_{x_{\bar{I}} y_{\bar{I}}}(u) e^{-A_{y_{\bar{I}} y_{\bar{I}}}(u, \tau)} du. \quad (5.7)$$

Let us now assume that  $\tau$  is greater than  $R_{\bar{I}+2}$ . Then the contribution writes

$$\mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}) = \sum_{k \in \{1,2\}} \mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}+1}} = k) \mathbb{P}(X_{R_{\bar{I}+1}} = k \mid X_{R_{\bar{I}}} = x_{\bar{I}}),$$

and this reduces to the computation of elementary terms of the form (5.2), (5.3), (5.4), (5.5), (5.6) and (5.7).

In practice with our data, the right-censoring times (that is the end of follow-up of the individual) are either the end of the follow-up for the whole sample, which is after the last possible census, or a census time at which the individual has been observed (see Remark 5.1 in Section 5.2). Therefore, formula (5.7) is only used here to compute the contribution of an observed death, with the last term detailed below.

(iii) In the case where  $d = 1$ , we have to compute

$$\mathbb{P}(T \geq \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}})\mathbb{P}(T = \tau \mid T \geq \tau, X_{R_{\bar{I}}} = x_{\bar{I}}) = \mathbb{P}(T = \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}).$$

The term  $\mathbb{P}(T = \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}})$  can be computed using the same reasoning: if  $\bar{I} = 5$ , we add a multiplicative factor  $\alpha_{x_5 3}(\tau)$  of instantaneous death to Equations (5.4) and (5.6) as

$$\mathbb{P}(T = \tau \mid X_{R_5} = x_5) = e^{-A_{x_5 3}(R_5, \tau)} \alpha_{x_5 3}(\tau),$$

or to Equation (5.5) as

$$\mathbb{P}(T = \tau \mid X_{R_5} = 2) = e^{-A_{22}(R_5, \tau)} \alpha_{23}(\tau) + \int_{R_5}^{\tau} e^{-A_{22}(R_5, u)} \alpha_{21}(u) e^{-A_{13}(u, \tau)} \alpha_{13}(\tau) du.$$

If  $\bar{I} \leq 4$  and  $\tau$  lies in the interval  $[R_{\bar{I}}, R_{\bar{I}+1}]$ , we add the instantaneous death probability to Equation (5.7) in order to get

$$\mathbb{P}(T = \tau \mid X_{R_{\bar{I}}} = x_{\bar{I}}) = e^{-A_{x_{\bar{I}} x_{\bar{I}}}(R_{\bar{I}}, \tau)} \alpha_{x_{\bar{I}} 3}(\tau) + \int_{R_{\bar{I}}}^{\tau} e^{-A_{x_{\bar{I}} x_{\bar{I}}}(R_{\bar{I}}, u)} \alpha_{x_{\bar{I}} y_{\bar{I}}}(u) e^{-A_{y_{\bar{I}} y_{\bar{I}}}(u, \tau)} \alpha_{y_{\bar{I}} 3}(\tau) du.$$

The other cases are easily obtained by the three previous elementary terms.

(iv) Finally, let us compute the term  $\mathbb{P}(X_{R_{\underline{I}}} = x_{\underline{I}})$ . If  $\underline{I} = 1$ , this corresponds to the initial distribution of the process, that is

$$\mathbb{P}(X_{R_1} = x_1) = p \mathbf{1}_{x_1=1} + (1-p) \mathbf{1}_{x_1=2}. \quad (5.8)$$

In the case where  $\underline{I} \geq 2$ , it reduces to the computation of elementary terms of the form (5.2)-(5.3) and (5.8) by

$$\begin{aligned} & \mathbb{P}(X_{R_{\underline{I}}} = x_{\underline{I}}) \\ &= \sum_{j_1 \in \{1,2\}} \dots \sum_{j_{\underline{I}-1} \in \{1,2\}} \mathbb{P}(X_{R_{\underline{I}}} = x_{\underline{I}} \mid X_{R_{\underline{I}-1}} = j_{\underline{I}-1}) \prod_{i=1}^{\underline{I}-2} \mathbb{P}(X_{R_{i+1}} = x_{i+1} \mid X_{R_i} = x_i) \mathbb{P}(X_{R_1} = x_1). \end{aligned}$$

This concludes the computation of the contributions in Equation (5.1).

### 5.3.2 Parametric framework

We detail here the parametric assumptions we make on the shape of the transition intensities, that is on the age pattern of characteristic-specific mortality rates as well as transitions between states. As previously mentioned, parametric methods are relevant to include information that we have on the shape of the intensities, for example a Gompertz-type mortality rate at reasonable ages. Given the sample we want to study, involving small samples coupled with interval-censored observations, such approach allows us to include a maximum of information and to reduce at best the dimension of the problem.

**Mortality rates** For the characteristic-specific mortality rates, which are nothing but the transition rates from the intermediate states to the absorbing state, we assume a Gompertz-type setting (see Gompertz (1825)), since ages between around 40 and 70 years are involved for socio-economic groups and between around 60 to 100 for marital status. This can be written, for  $(k, l) \in \{(1, 3), (2, 3)\}$ ,

$$\alpha_{kl}(t) = c_{kl} \exp(d_{kl}t).$$

The cumulative intensity can thus be written

$$A_{kl}(s, t) = \int_s^t \alpha_{kl}(u) du = \frac{c_{kl}}{d_{kl}} (\exp(d_{kl}t) - \exp(d_{kl}s)).$$

**Transitions between socio-economic groups** As for the transitions between intermediate states which represent different socio-economic groups, unlike mortality we do not have baseline data to guide us on the parametric form of the age-dependent transition intensities. Also, we tested a non-parametric method assuming known middle-interval transition times, but unfortunately the Nelson-Aalen increments were so erratic that it was not possible to get insights on possible monotonicity of even the shape of the transition intensities. Nevertheless, a reasonable framework and widely used is the monotonous transition intensities of Weibull type that will help in our study to give the possibility of an increase or decrease of the transition intensities during life, see e.g. Joly et al. (2002), Commenges et al. (2007), Foucher et al. (2007) and Touraine et al. (2013). Another advantage is that a particular case is the constant intensity framework, which will allow us to compare nested models with some statistical criteria. For  $(k, l) \in \{(1, 2), (2, 1)\}$ , the Weibull-type parametrization of the intensity  $\alpha_{kl}(t)$  writes

$$\alpha_{kl}(t) = a_{kl} b_{kl}^{-a_{kl}} t^{a_{kl}-1}, \quad (5.9)$$

whose cumulative intensity is

$$A_{kl}(s, t) = \int_s^t \alpha_{kl}(u) du = b_{kl}^{-a_{kl}} (t^{a_{kl}} - s^{a_{kl}}). \quad (5.10)$$



In our application regarding socio-professional category, we have to choose an age after which the characteristic is stable. Here we choose this age to be 60, just before common retirement age. Although arbitrary, this allows us to take into account the job status at the very end of the career while avoiding edge effects due to a common retirement age. Then the intensity (5.9) and its cumulative (5.10) are modified in this way. In the following, we denote Model 1 the model such that  $a_{kl} = 1$  and Model 2 the model with the Weibull parametrization.

**Transitions between marital status** Parametric assumptions about the age pattern of marital status transition rates are crucial. Indeed, contrary to socio-professional category, marital status can change at very high ages mainly due to married  $\rightarrow$  widowed transitions. Also, based on observed census, the parametric assumption combined with the maximum likelihood procedure allows us to capture an estimated value of these transition rates in the age range where no observation is available, that is between the last possible census (year 1999) and the end of the follow-up (year 2008). We performed preliminary studies and got insights on the shape of transition rates from married to widowed at high ages: at first sight, for a women at a given (high) age  $t$ , the transition rate  $\alpha_{21}(t)$  corresponds to the mortality rate of the spouse averaged over all its possible ages. With the use of national mortality data and also the observed distribution of ages between the members of a given couple, we computed such quantities and exhibited the possible corresponding age pattern. This one was in-between an exponential growth and a power law behavior, therefore we aim at testing two parametric families for such transitions. The first one is a Gompertz type shape denoted  $\alpha_{21}(t) = a_{21} \exp(b_{21}t)$ , and the second one is a Weibull type distribution  $\alpha_{kl}(t) = a_{21} b_{21}^{a_{21}} t^{a_{21}-1}$ . As for the transition rate  $\alpha_{12}(t)$ , due to the very small number of transitions from non-married to married in the age range considered, we propose to include a constant transition intensity  $\alpha_{12}(t) = b_{12}$ . In the following, we denote Model 3 the model in which the transition intensity from married to non-married is of Weibull type and Model 4 in the Gompertz-type parametrization.

To conclude, the set of parameters over which the likelihood  $L(\theta)$  has to be maximized is at most

$$\theta = (p, a_{12}, a_{21}, b_{12}, b_{21}, c_{13}, c_{23}, d_{13}, d_{23}). \quad (5.11)$$

**Parameter uncertainty** Once the likelihood is maximized, we obtain the value of estimated parameters. In the applications, one is also interested in parameter uncertainty, that is the fluctuations of the estimated parameters around the true (unknown) value. This way, it is possible to measure for a given model the error we commit on the parameters due to the finite size of the sample and the particulari-

ties of the interval-censored observation scheme. A standard result about maximum likelihood estimators is their asymptotic normality: as the size of the sample goes to infinity, the random set of parameters is multivariate normal, centered around the true value  $\theta$ , whose estimated value is denoted  $\hat{\theta}$ , with variance-covariance matrix given by the opposite of the hessian inverse at (unknown)  $\theta$ , which is estimated by the same matrix taken at  $\hat{\theta}$ , that is  $\hat{\Sigma} = -H(\hat{\theta})$  where  $H_{i,j}(\theta) = \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log L(\theta)$ . Note that as usual in such statistical analysis, it is assumed that we are in the asymptotic normality regime, for which we estimate numerically the mean and covariance matrix as previously detailed. This makes us able to perform joint simulations of the parameters, and then compute the distribution of several quantities of interest as transition intensities by evaluating it in each simulation.

## 5.4 Results

### 5.4.1 Results for socio-economic groups

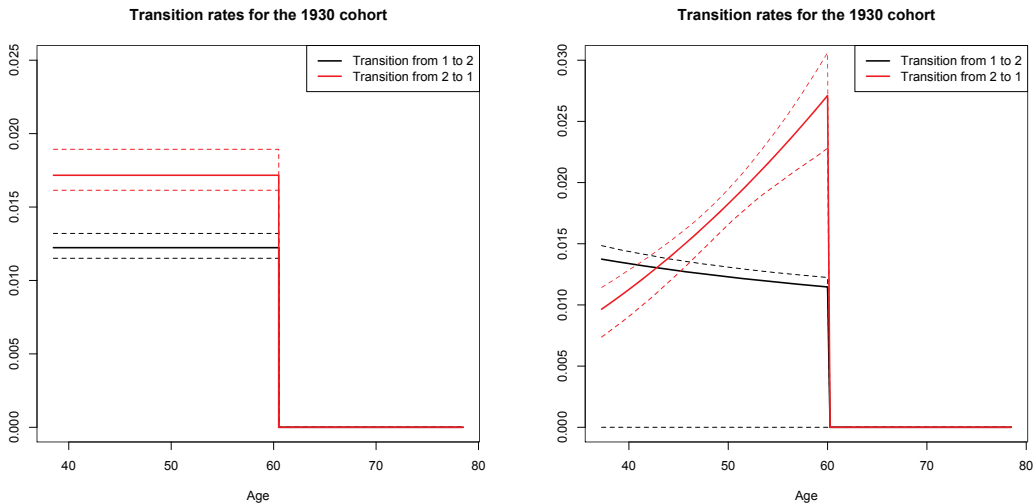
**Comparison of the two nested models** We first perform the maximization of the likelihood in the model where the transition intensities between intermediate states are constant (Model 1). We tested several initial parameters, and in particular we chose parameters that were close to a fitted Gompertz curve on national data. We emphasize that the maximization step requires a lot of computer resources, since for a given evaluation of the (log-)likelihood we have to compute each individual contribution. We obtain a log-likelihood of -14743.74 by means of the Nelder-Mead algorithm. The estimated parameters are shown in Table 5.1 with three significant numbers. In addition, we also perform the maximization under the Weibull parametrization for the transitions between intermediate states (Model 2), and we obtain a log-likelihood of -14727.55. The parameters are given in Table 5.1 as well. As expected, since the models are nested, the Weibull parametrization leads to a higher log-likelihood. With the methodology described in the previous section, we estimate the variance-covariance matrix of the intensity parameters both for Models 1 and 2. Finally, the characteristic-specific mortality forces as well as the transition rates between groups are depicted in Figures 5.4 and 5.5 with their 95% confidence intervals. With these graphs, we get several interesting insights. First, as expected, the specific mortality rates of the two groups are significantly different (in a statistical sense), and in particular that of group 1 individuals is higher. Also, transitions between groups are quite similar in the constant intensity Model 1, but the use of Model 2 shows that the age pattern of these transition rates are really different. Since the method quantifies the magnitude and fluctuations around transition intensities, we observe that both level and uncertainty are characteristic-specific, therefore we expect different behavior if we forecast a sample of this cohort

with a specific composition. This issue is addressed in Section 5.5. As well, let us remark that the level and uncertainty of the transition intensities between groups is different. In this context, we are interested in quantifying the impact of these characteristics changes on the dynamics.

Parameter	Model 1	Model 2
$p$	0.703	0.703
$a_{12}$	1	0.621
$a_{21}$	1	3.17
$b_{12}$	80.1	50.8
$b_{21}$	56.3	74.0
$c_{13}$	0.000191	0.000189
$c_{23}$	0.000152	0.000173
$d_{13}$	0.0721	0.0723
$d_{23}$	0.0703	0.0683

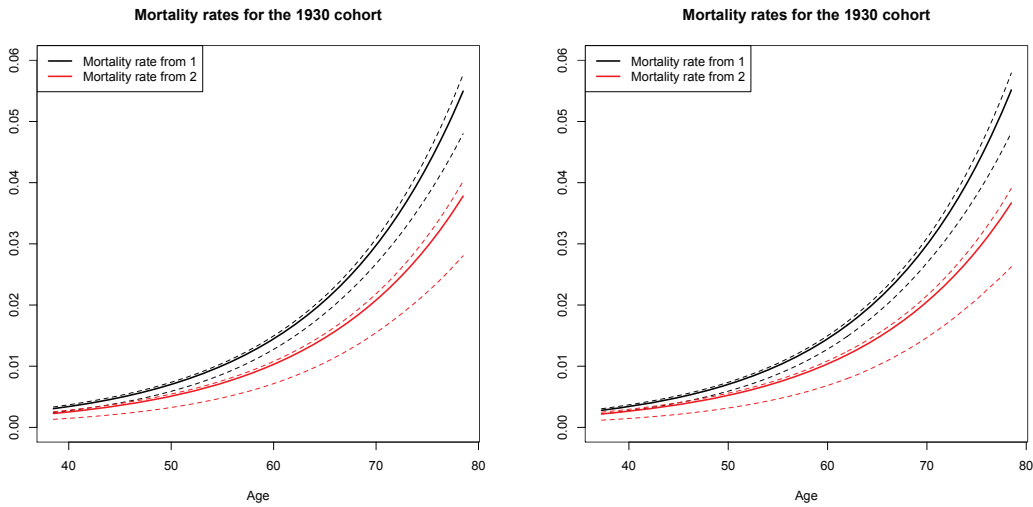
Table 5.1: Estimated parameters for the two models. Recall that for model 1, the set of parameters is reduced since  $a_{12}$  and  $a_{12}$  are set at value 1.

Figure 5.4: Estimated transition rates between groups for Model 1 (left) and Model 2 (right) with their 95 % confidence intervals



**Model comparison** As usual when comparing nested models, the issue is to determine if the two additional parameters  $a_{12}$  and  $a_{21}$  increase the likelihood significantly enough compared to the fact that the dimension of the problem is augmented. Several possible criteria can help to identify the "best" model. We focus here on both the classical Akaike Information Criterion (AIC) and also the Bayesian Information

Figure 5.5: Estimated characteristic-specific death rates for Model 1 (left) and Model 2 (right)



Criterion (BIC), which is more parsimonious than the latter regarding the problem dimension. The AIC can be computed as

$$AIC = -2 \log L(\hat{\theta}) + 2 \times k,$$

where  $k$  is the number of parameters involved. Also the BIC can be computed as

$$BIC = -2 \log L(\hat{\theta}) + \log(n) \times k,$$

where  $n$  is the number of observations. Note that the quantity  $n$  is in fact not obvious to determine, since it can be the number of individual observed, or the number of observed (uncensored) transitions. One can find discussions on this issue in the literature, see e.g. Volinsky and Raftery (2000). For our application, we choose  $n$  to be the number of observed points, that is the number of observed death plus the number of recorded censuses, which is 18114. This makes the BIC the most parsimonious as possible, since at most one transition can occur between two censuses.

In the end, for each criterion, the model with the lowest value is indicated. The results are given in Table 5.2 with five significative numbers. In each case, even with a parcimonious BIC, Model 2 is chosen. Note however that it presents wider confidence intervals for transition intensity  $1 \rightarrow 2$ . For the sake of comparison, we think that is it interesting to develop the numerical results for both models, which is done in the next section.

Criterion	Model 1	Model 2
AIC	29501	29473
BIC	29556	29543

Table 5.2: AIC and BIC computed for models 1 and 2.

### 5.4.2 Results for marital status

In the Weibull-type marital Model 3 we obtain a log-likelihood of -14061.86 by means of the Nelder-Mead algorithm, as for the Gompertz-type marital Model 4, a log-likelihood of -14062.51 is obtained. The parameters are displayed in Table 5.3. We estimate the variance-covariance matrix of the parameters as before, both for Models 3 and 4. On this basis, we represent the estimated intensities and their associated pointwise 95%-confidence intervals in Figures 5.6 and 5.7 for transition between groups and characteristic-specific death rates respectively. Let us first focus on the transition intensities between groups, see Figure 5.6. As expected when counting the number of observed transitions, the transition intensity  $\alpha_{12}(t) = b_{12}$  from non-married to married is very low compared to the reverse transition  $2 \rightarrow 1$ , although it is not zero (see Table 5.3). We are here close to an irreversible multi-state model, but it is shown in Section 5.5 that the analogy is not completely valid since a small impact remains on stylized life insurance products.

**On the mortality compensation law** Let us now focus on characteristic-specific log-mortality rates, see Figure 5.7. These show that for ages between around 60 and 85, mortality rates are lower for married individuals, which is intuitive. But after some age around 90, the situation is reversed, as for higher ages the mortality of married individuals becomes higher. This may be a surprising result at first sight, but in fact, this illustrates a well known demographic phenomenon referred to as the *mortality compensation law*. This appears when one compares several sub-populations at the same time, e.g. specific groups in a country or several national population, or the same population in successive time periods. It states that if the mortality factor is lower, namely here  $c_{23} < c_{13}$ , then necessarily the age coefficient is higher, that is  $d_{23} > d_{13}$  (see Table 5.3). In other words, a decrease of mortality at the lowest ages in the age range leads to an increase of mortality at highest ages. Dynamically, for stochastic mortality models with a formulation closely related to Gompertz, such as Cairns et al. (2006), the times series of slope and intercept of log-mortality rates appear to be negatively correlated, as already highlighted in Strehler and Mildvan (1960). This phenomenon can be explained by the fact that the human life span, that is the *maximum biological life length driven by the intrinsic rate of bodily deterioration*, does not change much for reasonable time periods (see e.g. Strulik and

Vollmer (2013)). Note that the compensation law can be also measured in two other ways, with their corresponding denominations: the compression of the distribution of ages at death, and the rectangularization of the survival curve. For further investigation of the compensation law and related effects, the reader is referred to e.g. Strehler and Mildvan (1960), Fries (1980), Gavrilov and Gavrilova (1991), Wilmoth and Horiuchi (1999) and Strulik and Vollmer (2013). Note that as we avoid proportional hazard assumptions, we are able to reproduce such effects. This leads to several insights on the shape of aggregate mortality, as well as on the impact of life insurance products. This is developed in the next section.

Parameter	Model 3	Model 4
$p$	0.329	0.329
$a_{21}$	5.99	0.000397
$b_{12}$	0.00194	0.00193
$b_{21}$	0.0131	0.0693
$c_{13}$	1.22e-05	1.23e-05
$c_{23}$	2.96e-06	3.11e-06
$d_{13}$	0.103	0.103
$d_{23}$	0.119	0.118

Table 5.3: Estimated parameters for Models 3 and 4.

Figure 5.6: Estimated transition rates between groups for Model 3 (left) and Model 4 (right) with their 95 % confidence intervals

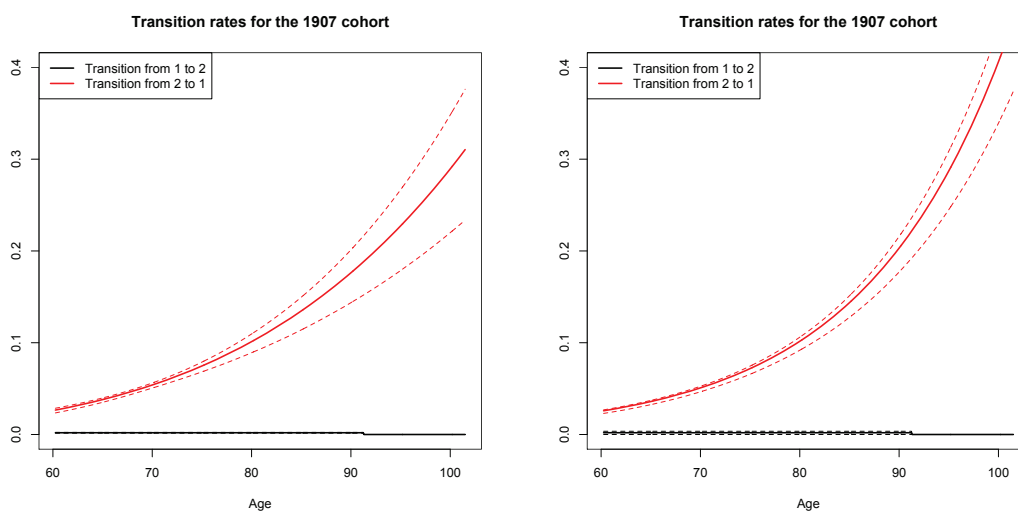
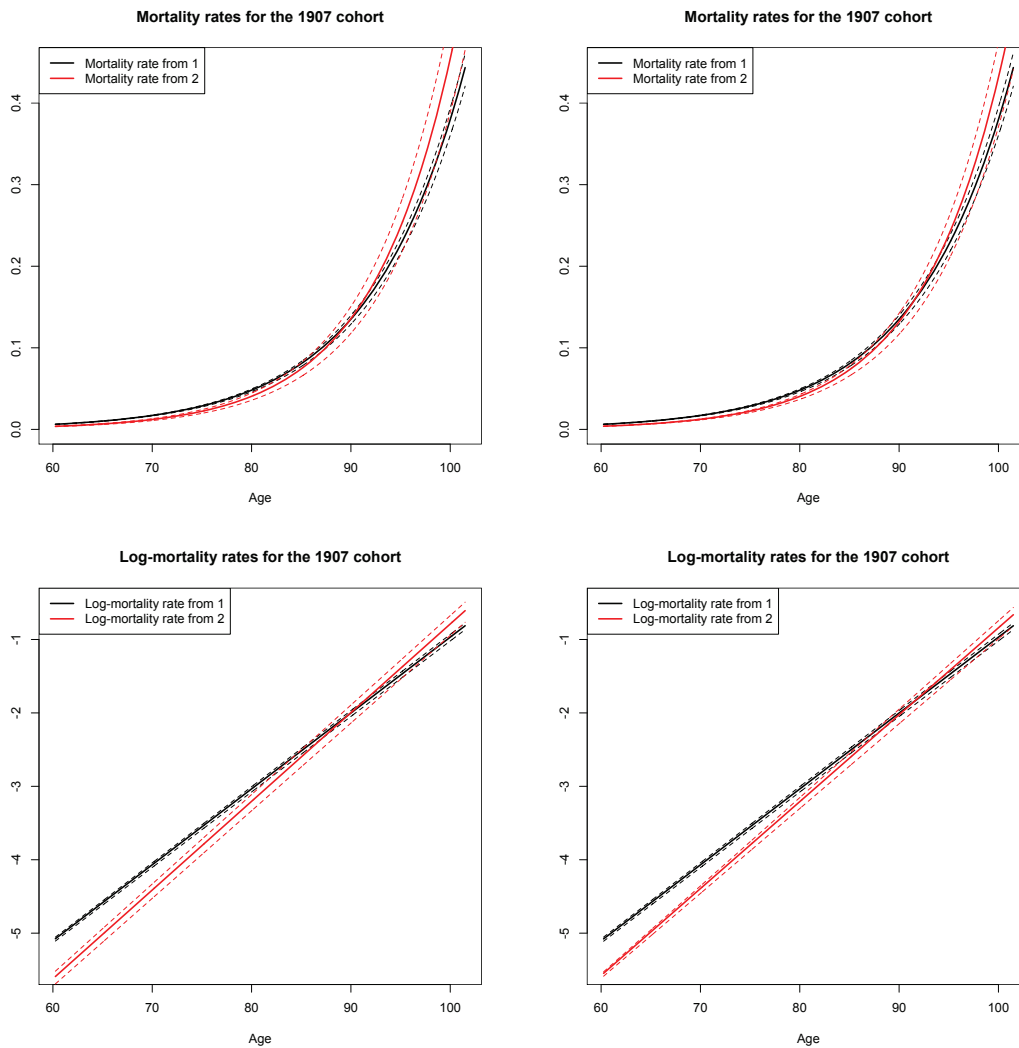


Figure 5.7: Estimated characteristic-specific death rates (top) and their logarithm (bottom) both for Model 3 (left) and Model 4 (right)



## 5.5 Aggregate mortality and the impact of heterogeneity on life insurance products

The aim of this section is to address the impact of the population dynamic heterogeneity on (i) the aggregate mortality rate and (ii) typical life insurance products embedding life or death benefits. We begin by describing the dynamics of an heterogeneous population with specific composition.

### 5.5.1 Population heterogeneity dynamics

To understand the impact of the dynamic composition on several quantities of interest, one has to detail how a particular sample evolves through time, taking into account deaths but also characteristics changes during life. We see transition in-

tensities  $\alpha_{kl}(t)$  depending on age  $t$  to be stochastic quantities as they depend on the realized parameter vector  $\theta$ , see (5.11), whose distribution can be approximated as detailed at the end of Section 5.3. We deal here with the mean trajectory of a given sample, therefore demographic stochasticity (which is a consequence of finite sample sizes) is not taken into account in itself. Let us recall that we are interested here in the uncertainty in the parameters due to the finite sample data size and the interval-censoring mechanism. Let us denote  $g_1(\underline{a})$  and  $g_2(\underline{a})$  the initial deterministic quantities (real numbers) of individuals in each group 1 or 2 with age  $\underline{a}$ , which is nothing but the age of our sample of interest at the first census (here  $\underline{a} = 37.25$ ). Now, let us denote for  $t \geq \underline{a}$ ,  $G(t) = \begin{pmatrix} g_1(t) \\ g_2(t) \end{pmatrix}$  the repartition between the two groups. Given the transition rates, this evolves (deterministically) according to deaths and characteristic changes. The classical Chapman-Kolmogorov equation gives us the average dynamics of our Markov process as

$$G'(t) = K(t)G(t), \quad (5.12)$$

where the transition matrix  $K(t)$  is given by

$$K(t) = \begin{pmatrix} -\alpha_{13}(t) - \alpha_{12}(t) & \alpha_{21}(t) \\ \alpha_{12}(t) & -\alpha_{23}(t) - \alpha_{21}(t) \end{pmatrix}.$$

Note that the complexity of taking into account reversible transitions arises due to the fact that in general no analytical solutions are available for  $G$  (see e.g. Andersen et al. (1993)). This is also the case for the particular parametric forms we have chosen for the intensities. Indeed, since the matrices  $K(s)$  and  $K(t)$  do not commute\*, we are not able to write an explicit formula for  $G(t)$ . In practice, the assumption that the reverse intensity is zero is often used when a few number of such transitions occurred (see e.g. Czado and Rudolph (2002)). Although analytical solutions are not available in our case, we are able to compute the ordinary differential equation (5.12) numerically using a standard discretization scheme.

### 5.5.2 Aggregate mortality

In this part, our aim is to illustrate several patterns of aggregate mortality and explain them based on the analysis of the underlying population structure. Let us define the total number of individuals in the sample as

$$g(t) = g_1(t) + g_2(t). \quad (5.13)$$

---

\*The proof that  $K(s)$  and  $K(t)$  do not commute is left to the reader.



Note again that this is in fact a quantity in  $\mathbb{R}_+$ . Formally, the equivalent death rate (in continuous time) is defined as

$$\mu(t) = -\frac{g'(t)}{g(t)} = \frac{g_1(t)\alpha_{13}(t) + g_2(t)\alpha_{23}(t)}{g(t)}, \quad (5.14)$$

which is nothing but an average of characteristic-specific death rates, whose weights depend on the dynamics of the underlying population composition.

When looking at heterogeneity, a well known phenomenon is the fact that mortality rejoins the lowest one, more precisely: the aggregate death rate, that is the "equivalent" mortality force, gets closer to that of the group whose mortality the lowest. This fact is naturally observed when individuals do not change between the groups. When characteristics changes are taken into account, the way aggregate mortality evolves through age may be different, and in particular even the reverse effect can be observed. The two applications we further develop below are able to capture such facts.

**Socio-economic groups** Aggregate mortality of Equation (5.14) is computed using the dynamics (5.12), and the results are depicted in Figure 5.8 for the age range concerned with the estimation procedure. In this case, based on the initial sample proportions, and given the level of estimated transition intensities, aggregate mortality slowly rejoins mortality of group 2, and the speed at which this effect occurs is quantified. This is crucial because it shows how a quantification of mortality at a deeper level leads to several insights on the shape of aggregate mortality. In our study, we were in addition able to capture the variance-covariance matrix of the parameters. Also, we can compute the associated 95%-confidence intervals, shown in Figure 5.9.

To further validate the modeling framework and the parametric assumptions, we can compare one hand the aggregate survival function obtained by taking into account the interval censoring scheme and the underlying heterogenous dynamics, and on the other hand the survival function computed on the sample regardless of individual characteristics, e.g. by Kaplan-Meier estimator. This comparison is shown in Figure 5.10 for Models 1 and 2. Note that the two curves shapes and levels are very similar, although they may differ at high ages for extreme scenarios. This could be explained by the Gompertz assumptions for mortality which is not a good approximation for ages around 40 years, which may result in a slight underestimation of mortality. Nevertheless, we believe that this is the best parametric option for the whole age range and given the small amount of data. Note also that despite these differences in some parameter scenarios, we capture on the other side more information on the heterogenous dynamics, which is at the core of our study.

**Marital status** The study of the underlying population composition gives us many insights on the age pattern of aggregate mortality. This is particularly the case for the focus on marital status. In Figure 5.11, we depict the characteristic-specific and aggregate log-mortality rates for Models 3 and 4, as well as their corresponding 95%-confidence intervals in Figure 5.12. Let us describe the age pattern of aggregate log-mortality: at age around 60, this quantity is closer to the "married" mortality, then rejoins the "non-married" mortality at age around 85 and remains at this level until the end of the age range considered. This fact can not be described as "aggregate mortality rejoins the lowest one", due to the impact of the evolving population composition taking characteristics changes into account. Here, aggregate mortality first rejoins the highest one, and then stays at the level of the "non-married" group since it corresponds to the characteristic which is more and more represented in the population due to characteristic changes at high ages.

For this study on marital status also, to further validate the modeling framework and the parametric assumptions, we plot in Figure 5.13 the comparison between the aggregate survival function obtained by taking into account the interval censoring scheme and the underlying heterogeneity, and also the survival function computed on the sample regardless of individual characteristics, e.g. by Kaplan-Meier estimator. In this case, survival curves are really similar, showing that our method captures the underlying heterogeneity that furthermore replicates the overall mortality observed in the sample.

Figure 5.8: Characteristic-specific and aggregate mortality rate (left) and their logarithm (right) for Model 1.

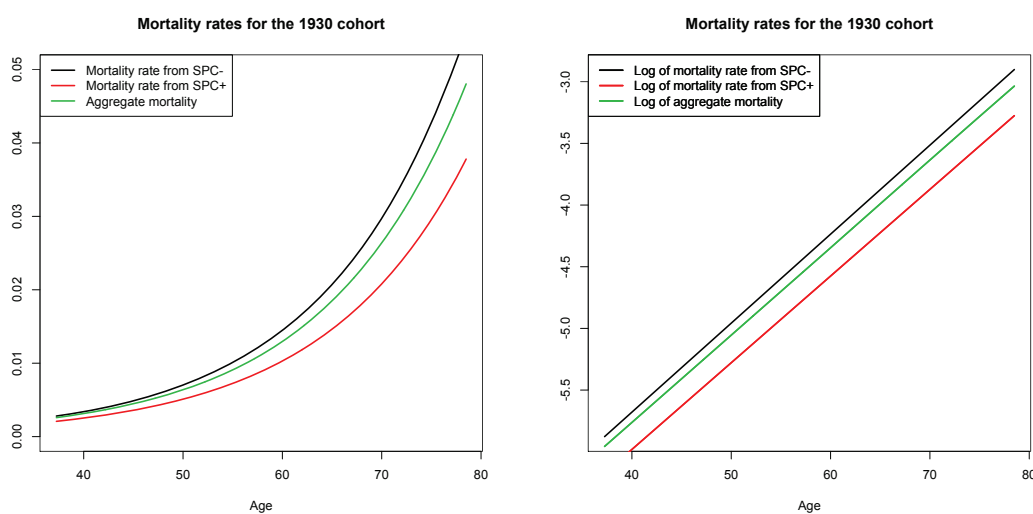
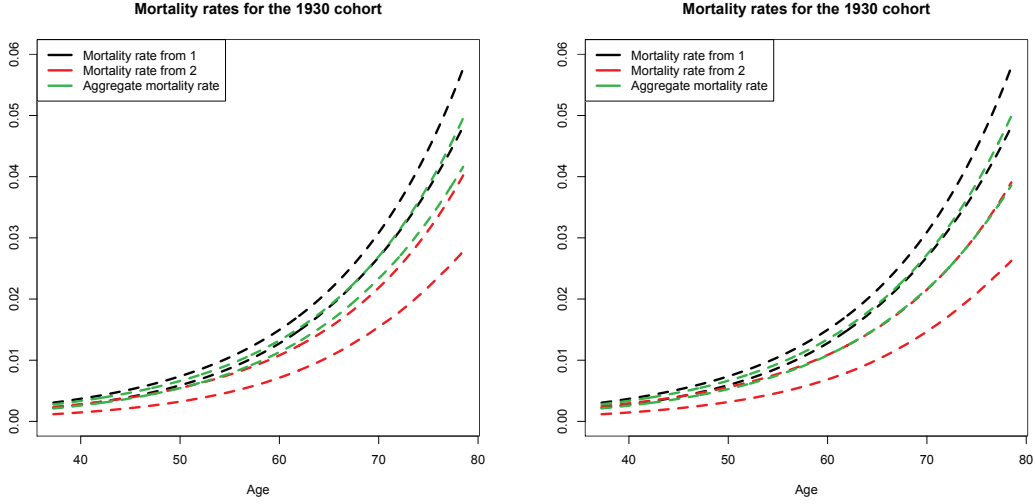


Figure 5.9: Characteristic-specific and aggregate mortality forces 95%-confidence intervals for Model 1 (left) and Model 2 (right).



### 5.5.3 Impact on life insurance products

The aim of this section is to assess the impact of heterogeneity on life insurance products as death or life benefits. Particular attention is devoted to characteristics changes and parameter uncertainty. Let us start with a portfolio with some composition  $G(\underline{a}) = \begin{pmatrix} g_1(\underline{a}) \\ g_2(\underline{a}) \end{pmatrix}$ , and focus on two simplified actuarial quantities as the life benefit and the death benefit, that start at age  $\underline{a}$  and end at age  $\bar{a}$ . Let us consider a constant discount rate  $r$ . The life benefit, also known as annuity contract, amounts to pay 1 per time unit to still alive individuals until age  $\bar{a}$ , that is

$$LB := \int_{\underline{a}}^{\bar{a}} e^{-r(t-\underline{a})} g(t) dt,$$

where  $g(t)$  is defined in (5.13). Another quantity of interest is the death benefit, which pays 1 at each death before age  $\bar{a}$ . Note that formally, the number of deaths in the time interval  $[t, t + dt)$  is  $g(t) - g(t + dt) \approx -g'(t)dt$ . Then the death benefit, also known as term insurance contract, is rigorously defined as

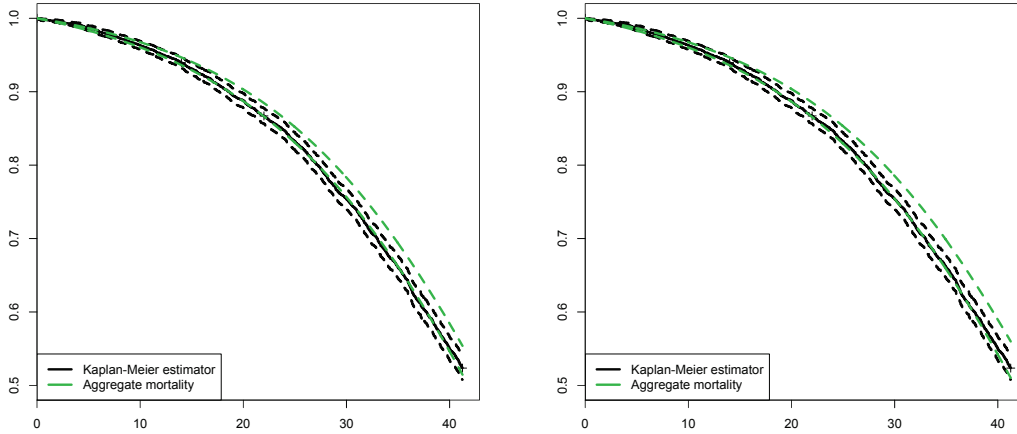
$$DB := - \int_{\underline{a}}^{\bar{a}} e^{-r(t-\underline{a})} g'(t) dt.$$

After integration by parts, it can also be rewritten in terms of the life benefit as

$$DB = g(\underline{a}) - e^{-r(\bar{a}-\underline{a})} g(\bar{a}) - rLB.$$

This gives another interpretation of the death benefit: it amounts for the insurer to pay 1 at each individual in the portfolio (first term), then take back 1 from each

Figure 5.10: Survival curves 95%-confidence intervals obtained in Model 1 (left) and Model 2 (right) compared to the Kaplan-Meier estimate



individual which is alive at the end at age  $\bar{a}$  (second term), and finally that each alive individual pays back  $r$  per time unit to the insurer (third term).

The aim now is to compute the distribution of the life and death benefits, which depends on the (random) set of parameters, under several configurations. In particular, we measure sensitivities due to (i) the composition of the initial population, (ii) the fact that characteristics changes are taken into account or not, and (iii) several values for the interest rate  $r$ .

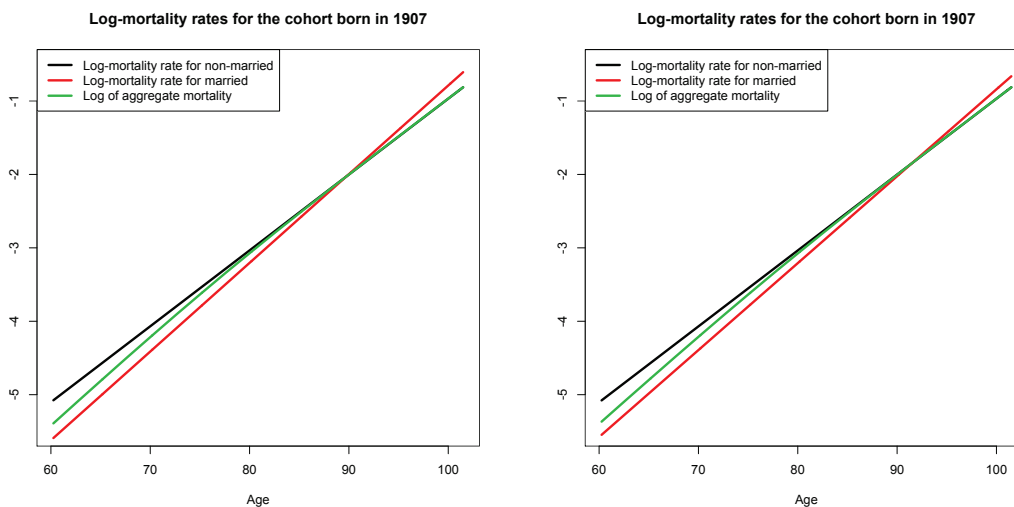
**Impact of socio-economic heterogeneity** Let us fix  $\underline{a} = 37.25$  and  $\bar{a} = 78.5$ , corresponding to the age range considered in the estimation procedure. This is typical of death benefits, therefore numerical results are given in this context. The set of parameters is given as follows:

- (i) the initial population at age  $\underline{a} = 37.25$  is made of SPC+ (group 2) only ( $p = 0$ ), or representative of the initial sample ( $p = 0.703$ ), or made of SPC- (group 1) only ( $p = 1$ ).
- (ii) the random distribution of the value of the death benefit is represented using the full model (5.12) and compared to the same where characteristics changes are not taken into account (that is  $\alpha_{12}(t) = \alpha_{21}(t) = 0$ ).
- (iii) two values for the discount rate are tested, namely  $r = 1\%$  in Figure 5.14 and  $r = 3\%$  in Figure 5.15.

These results are depicted for both Models 1 and 2.

Let us first focus on the initial population that is representative of the original sample composition ( $p = 0.703$ ), see the middle column in Figures 5.14 and 5.15. From this test, it appears that considering characteristics to be stable slightly over-

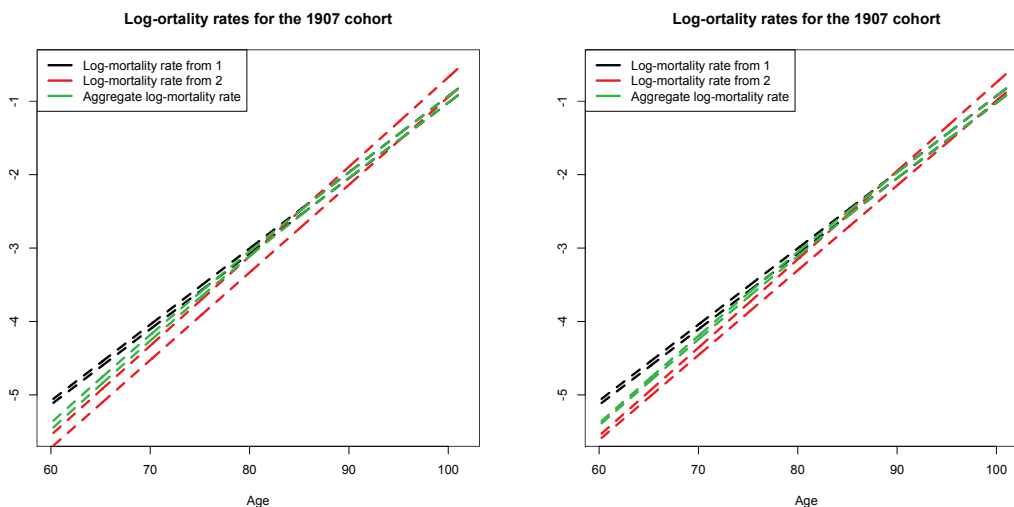
Figure 5.11: Characteristic-specific and aggregate log-mortality rate for Model 3 (left) and Model 4 (right).



estimates the value of the death benefit, in other words under-estimates the sample size, or equivalently over-estimates mortality. Therefore, it seems that the proportion of individuals in group 1 is over-estimated when characteristics changes are omitted. Yet, let us recall that the transition intensity from 1 to 2 is on the whole lower than 2 to 1, see Figure 5.4. But in fact, the total intensity of characteristics changes from group 1 to group 2 is the product of the  $1 \rightarrow 2$  transition intensity times the number of individuals in group 1. Therefore, when the initial proportion of group 1 is sufficiently large, the number of changes from 1 to 2 becomes higher than those from 2 to 1, and the death benefit is over-estimated if changes are not taken into account. This can be illustrated if we start from a population initially made of group 1 individuals only ( $p = 1$ ), see the right column in Figures 5.14 and 5.15. On the contrary, if the proportion of group 1 individuals is lower enough, the death benefit is under-estimated when considering characteristics stable during life. This is illustrated when starting with group 2 individuals only ( $p = 0$ ), see the left column in Figures 5.14 and 5.15. Note that of course, the death benefit is naturally lower with individuals in group 2 compared to group 1, and our modeling framework quantifies the difference between the two.

As for the comparison between the two Models 1 and 2, results are very similar, both in terms of overall level and uncertainty. Finally, comparing Figures 5.14 and 5.15 allows us to test the sensitivity to the interest rate in each scenario. As expected, the value of the death benefit (x-axis) decreases as  $r$  increases, but in each scenario taking into account transitions between groups still has a huge impact.

Figure 5.12: Characteristic-specific and aggregate log-mortality forces 95%-confidence intervals for Model 3 (left) and Model 4 (right).

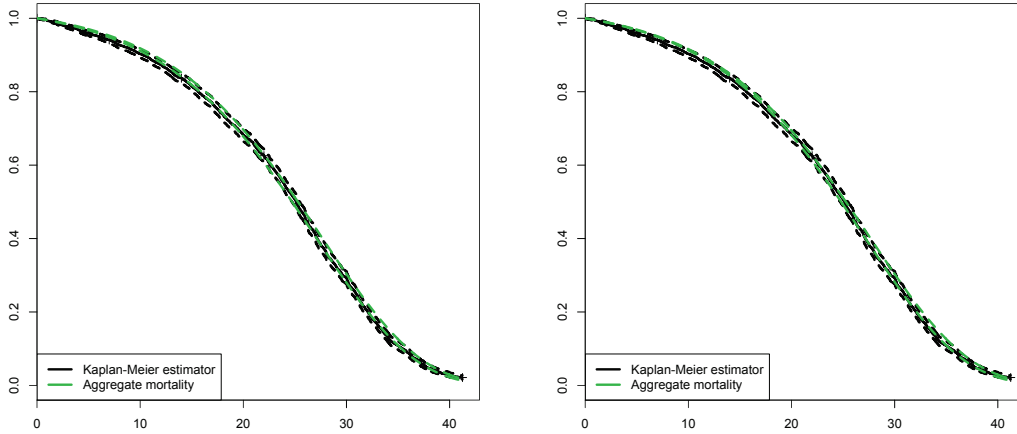


**Impact of marital status heterogeneity** Concerning marital status heterogeneity, numerical results also show that it is crucial to take into account both characteristic-specific mortality and transitions between states. Let us fix  $\underline{a} = 60.25$  and  $\bar{a} = 101.5$ , corresponding to the age range considered in the estimation procedure, typical of a life benefit. For these results, the set of parameters is given as follows:

- (i) the initial population at age  $\underline{a} = 60.25$  is made of married individuals (group 2) only ( $p = 0$ ), or representative of the initial sample ( $p = 0.329$ ), or made of non-married individuals (group 1) only ( $p = 1$ ).
- (ii) the random distribution of the value of the death benefit is represented using the full model (5.12) and compared to the same where characteristics changes are not taken into account (that is  $\alpha_{12}(t) = \alpha_{21}(t) = 0$ ).
- (iii) two values for the discount rate are tested, namely  $r = 1\%$  in Figure 5.16 and  $r = 3\%$  in Figure 5.17.

These results on life benefits are depicted for both Models 3 and 4. The impact of transitions out of marriage is clear, being mostly at these ages transitions from married to widowed. This is depicted in Figures 5.16 and 5.17 when starting with married individuals (left column) or a population which is representative of the original sample (middle column). On the contrary, as the transition intensity from non-married to married is small, the impact is not so important when starting from non-married individuals only (right column). But let us emphasize that a difference still remains, which can have a real impact when multiplied by a huge amount of pensions. Also here, when comparing Models 3 and 4, the main difference appears to be the uncertainty about the life benefit value, this one being lower in Model 4; in this model, also less performing than Model 3 in terms of likelihood (see Subsection

Figure 5.13: Survival curves 95%-confidence intervals obtained in Model 3 (left) and Model 4 (right) compared to the Kaplan-Meier estimate



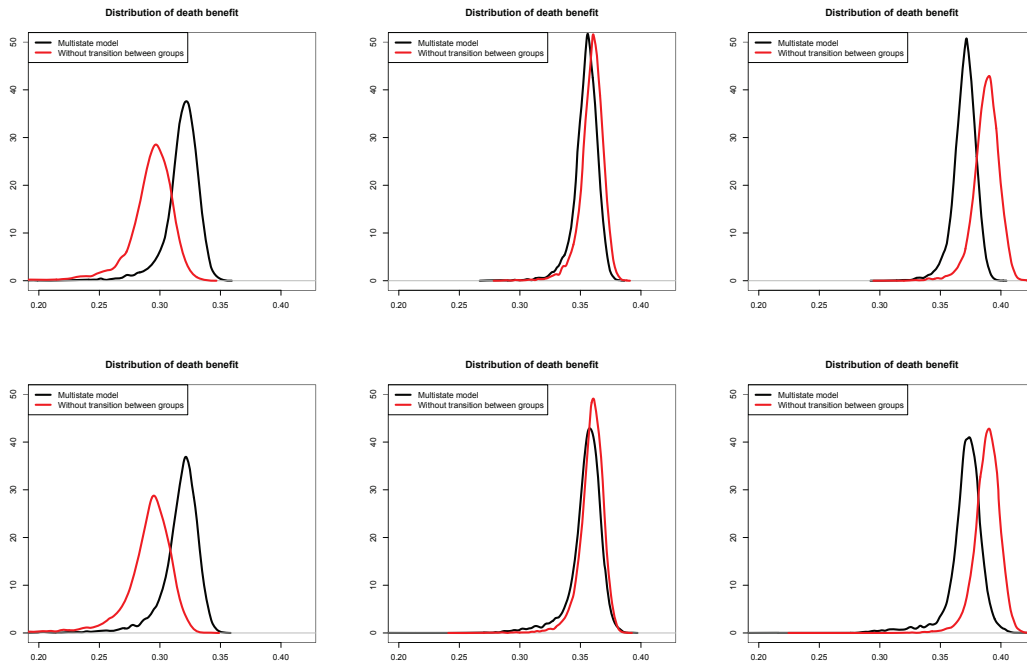
5.4.2), confidence intervals are in fact lower both for mortality rates (see Figure 5.7) and transitions rates between states (see Figure 5.6).

To conclude these numerical results assessing the impact of socio-economic or marital status heterogeneity on typical life insurance products, let us emphasize that the impact of taking into account characteristics changes depends on both the initial population composition and the level of the several possible transition intensities. The maximum likelihood procedure developed in this paper that takes into account interval censoring allows us to quantify such effect based on longitudinal data.

## Conclusion

In this paper, we have developed a parametric maximum likelihood method for measuring mortality heterogeneity when characteristics changes are interval-censored, which is a particularity of longitudinal data based on censuses. Indeed, for such data individuals are often observed at isolated points in time so the times at which characteristics change during life remain unknown. While applying such a method to a representative sample of the French national population, which presents both a small number of individuals and a systematic interval-censored observation mechanism, we were still able to capture characteristic-specific mortality rates linked to socio-economic groups or marital status, as well as to measure the transition rates between the groups. Special attention was devoted to parameter uncertainty whose magnitude differed depending on the characteristics and the transition rates considered.

Figure 5.14: Death benefit distribution with  $r = 1\%$  and  $p = 0$  (left),  $p = 0.703$  (middle) and  $p = 1$  (right) for Model 1 (top) and Model 2 (bottom)

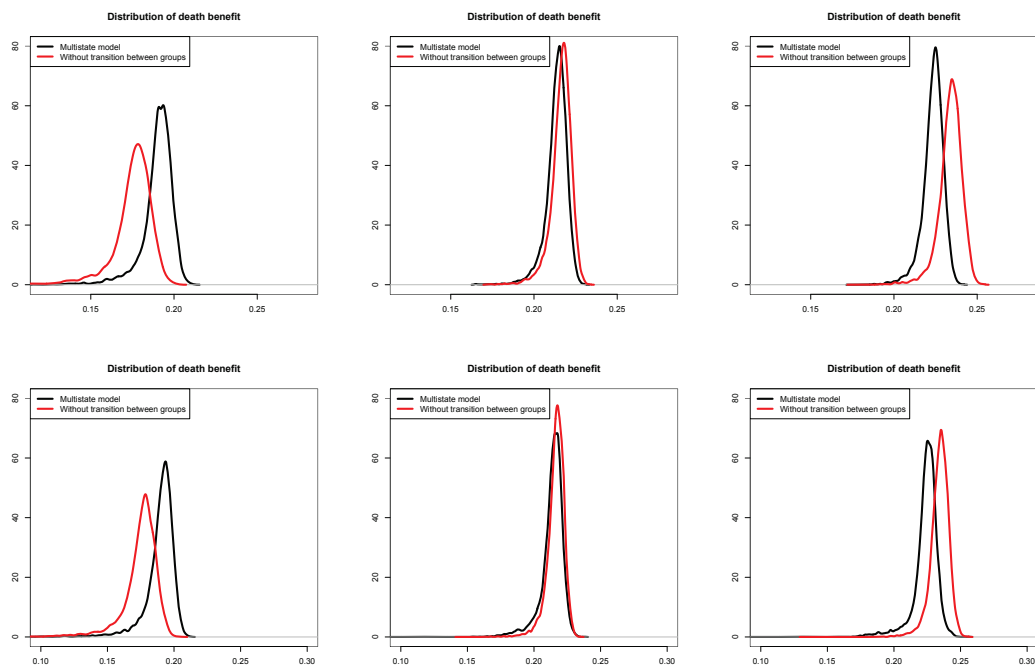


Based on such estimates, we highlighted effects such that the mortality compensation law between groups. We also addressed the dynamic evolution of several populations with specific compositions, leading us to get insights on the age pattern of aggregate mortality related to the underlying dynamic heterogeneity. We were also able to quantify the impact on typical life insurance products of considering characteristics stable during life instead of taking into account their variability. Interestingly, this impact depends on both the initial population composition and the level of the transition intensities between groups; these insights are crucial for longevity/mortality risk management.

The statistical estimation of mortality heterogeneity based on complex data remains a challenging field of actuarial research. Further improvements would concern higher dimensional multi-state models with more groups, as well as the estimation of several cohorts simultaneously. To this aim, "larger" datasets are required, first including more individuals to allow to increase the number of possible states, and second with more frequent and deep records leading to reasonable observation schemes and time windows, since particularly the study of the age pattern of heterogeneous mortality within *real* cohorts requires very long historical data.



Figure 5.15: Death benefit distribution with  $r = 3\%$  and  $p = 0$  (left),  $p = 0.703$  (middle) and  $p = 1$  (right) for Model 1 (top) and Model 2 (bottom)



## Aknowledgements

The authors are grateful to Philippe Saint-Pierre for several enlightening discussions on this work. The authors also thank Jacques Portes for his help on using computer resources at LPMA, as well as Catherine Mathias, Olivier Lopez, Xavier Milhaud, Quentin Guibert and Julien Tomas for fruitful discussions on several topics related to this work.

## 5.5. AGGREGATE MORTALITY AND THE IMPACT OF HETEROGENEITY ON LIFE INSURANCE PRODUCTS

Figure 5.16: Life benefit distribution with  $r = 1\%$  and  $p = 0$  (left),  $p = 0.329$  (middle) and  $p = 1$  (right) for Model 3 (top) and Model 4 (bottom)

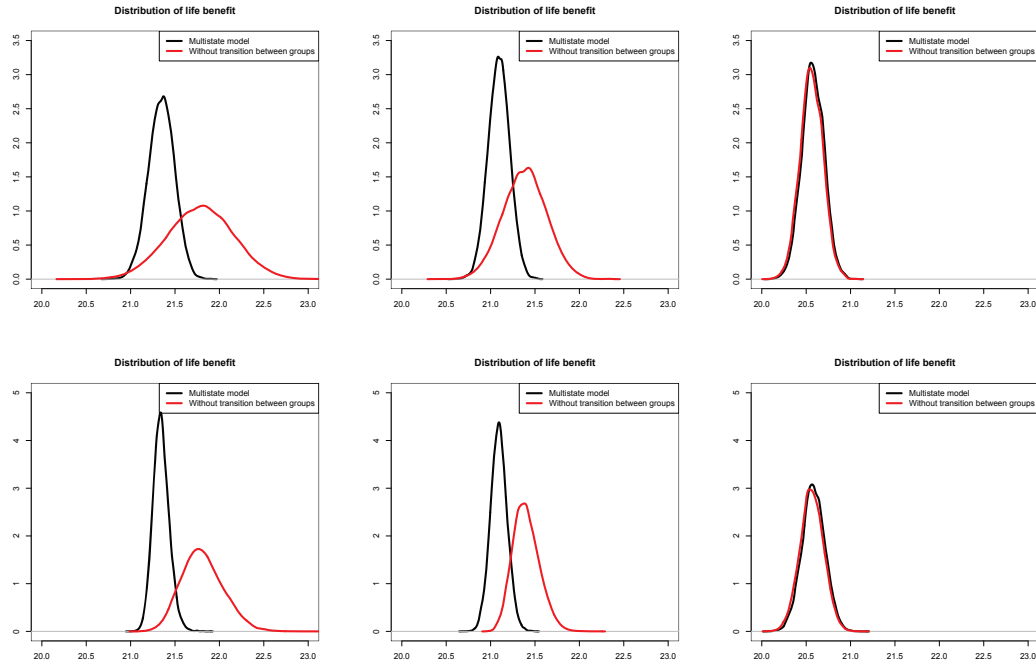
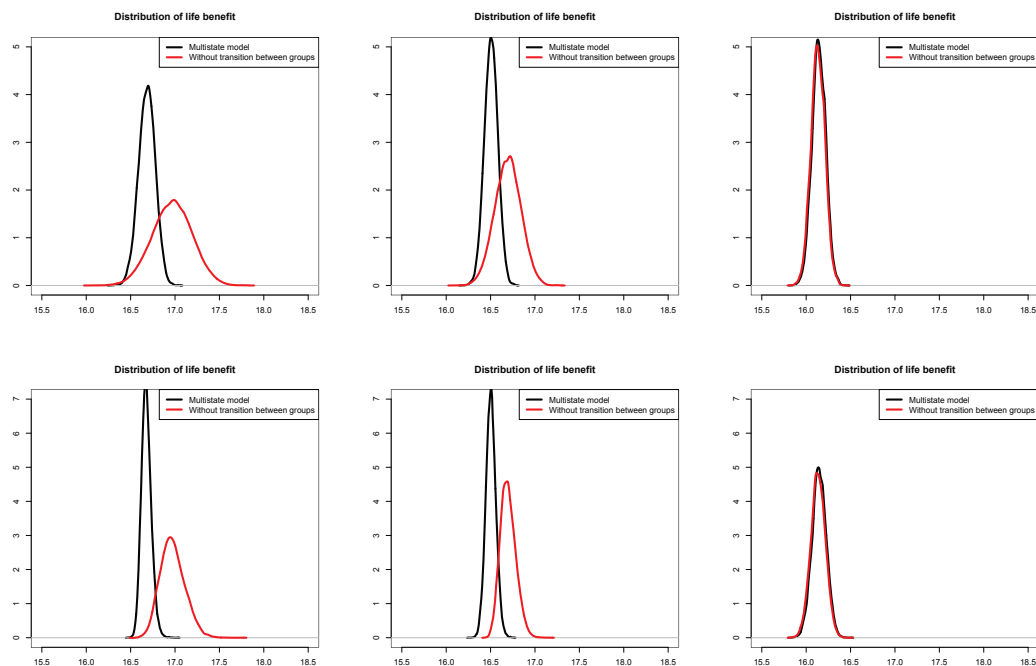


Figure 5.17: Life benefit distribution with  $r = 3\%$  and  $p = 0$  (left),  $p = 0.329$  (middle) and  $p = 1$  (right) for Model 3 (top) and Model 4 (bottom)



# Chapter 6

## Macroscopic behavior of heterogenous populations with fast random life histories<sup>†</sup>

### Abstract

In this paper, we consider the large population limit of an age and characteristic-structured stochastic population model evolving according to individual birth, death and fast characteristics changes during life. Both the large population framework and the fast characteristics changes assumption are motivated by demographic patterns of human populations at the scale of a given country. When rescaling the population process, and under some invariance assumption about the characteristics changes dynamics, the classical deterministic transport-renewal McKendrick-Von Foerster equation appears, that describes the time evolution of the age pyramid driven by equivalent birth and death rates. The proof follows the work of Méléard and Tran (2012) and Gupta et al. (2014) in which analogous mathematical issues are encountered. We further prove that the sequence of processes taking track of the characteristics distribution is not tight even in the presence of age-independent demographic rates. To illustrate the use of the limiting model, a set of computable invariant distributions is given, as well as numerical implementation of equivalent birth and death rates which mimics real demographic data. These results highlight the fact that characteristics changes frequencies are crucial to understand aggregate demographic rates at the macroscopic scale.

---

<sup>†</sup>Boumezoued, A. 2015, Macroscopic behavior of heterogenous populations with fast random life histories. *Working paper*

## 6.1 Introduction

The study of the dynamics of mortality rates at the national level is still a major issue both in demographics and actuarial science. Whereas national data provides estimation of death rates for men and women by age over time, a growing interest concerns the dynamics of mortality rates at a deeper level, looking at other individual characteristics. This is crucial to understand the dynamics of mortality, both for state pension issues and insurance risk assessment. In particular, the statistical estimation of death rates by individual characteristics can be carried out when looking at a sample of individuals who share some characteristics which are quite stable during their life. However in practice, many characteristics of individuals can vary over time. A possible approach is to estimate a "larger" model including death rates but also transitions between characteristics or states. The statistical techniques dedicated to so-called multi-state models address this issue for a class of processes and observation schemes (see e.g. Boumezoued et al. (2015) and references therein). In this field however, the statistical estimation remains very challenging, partly due to the lack of data and the censoring scheme, but also due to the high number of involved parameters. Another approach could be to deal with an approximate model. Indeed, among the characteristics which are known to have a real impact on longevity, many vary very frequently over time: one can think to exposition or not to some contaminant, alternating a dangerous activity with a safe one, the precise income, the health status with alternate periods of illness. Those can also have a huge impact on birth patterns, making the dynamics of the whole population difficult to analyze. As an example, the time evolution of the age pyramid is a crucial quantity of interest for decision making in public pension systems. In this context of fast changing characteristics, it seems difficult to keep track of the population evolution at the microscopic level and one could be interested in the right "approximation" of the dynamics, that is, to replace all birth and death rates by characteristics by an aggregate death rate which depends on the microscopic rates but also on some stable population composition.

In this paper, we consider a stochastic population model in which individuals have an age and characteristics, and can give birth, change their characteristics (event called swap) and/or die. We study the asymptotics of the stochastic individual-based model under both large population and fast swap patterns; in particular characteristics change at the fast time scale whereas aging remain at the slow time scale. Both the large population framework and the frequent characteristics changes appear naturally when focusing on the demographic evolution of a human population at the scale of a given country. In this context, and under some invariance assumption on the swap patterns, the macroscopic behavior is described by a McKendrick-Von Foerster deterministic equation (see McKendrick (1926) and

Von Foerster (1959)) in which only age is involved and parameters are averaged over the stable characteristics distribution. Our probabilistic setting is inspired by Fournier and Méléard (2004), Champagnat et al. (2006), Tran (2008), Ferriere and Tran (2009) and Bensusan et al. (2010–2015). In particular, the birth-death-swap process representation is based on Bensusan et al. (2010–2015) (see also Bensusan (2010)).

In the literature, limit theorems for stochastic processes involving several time scales have been widely studied. As for our framework of interest which concerns measure-valued population processes, one can find several studies involving two time scales, for example related to evolutionary mechanisms with rare or accelerated mutations in characteristics-structured (called trait-structured) population models (see e.g. Fournier and Méléard (2004), Champagnat et al. (2006), Bovier and Wang (2013) and Billiard et al. (2014)), prey-predator models (see e.g. Costa et al. (2015) and Costa (2015)), as well as fast aging structured populations (see e.g. Méléard and Tran (2012) and Gupta et al. (2014)). To our knowledge, no contribution focused on measure-valued age and trait-structured population models with fast changing characteristics and age-dependent birth and death rates. This is the purpose of the present paper to develop such modeling framework, as well as to highlighting its contribution to demographic purposes. This work can be seen as the probabilistic counterpart of aggregation methods for deterministic equations based on time scale separation techniques (see e.g. Auger et al. (2012) for a review). Concerning the rescaling techniques, we are specifically interested in Méléard and Tran (2012) and Gupta et al. (2014) in which analogous mathematical issues are encountered, and whose techniques are used in the present paper. The link with our result will be further detailed in the corresponding section.

The remainder of this paper is organized as follows. In Section 6.2, the birth-death-swap population process is introduced, as well as the assumptions and our results. Proofs are given in Section 6.3. Finally, Section 6.4 details some examples and numerical illustration, and we give some concluding remarks in Section 6.5.

## 6.2 Setting and main results

We are interested in the evolution of a population in which each individual has characteristics  $x \in \mathcal{X}$ , where  $\mathcal{X}$  is a compact set of  $\mathbb{R}^d$ , and an age  $a \in \mathbb{R}_+$ . In the population, three kinds of events can occur:

- (i) A birth, that is the arrival of an individual with age zero,
- (ii) A death, that is the removal of an individual,
- (iii) A swap, that is a change of individual's characteristics.

We want to model the fact that each individual changes its characteristics very

often compared to the times at which it gives birth and dies. This is motivated by human populations for which one can consider individuals who change their income, health status, or food condition very often during their life. Each individual with characteristics  $x$  and age  $a$  gives birth at rate  $b(x, a)$ , dies with rate  $d(x, a)$ , and changes its characteristics at times given by the swap rate  $n.e(x, a)$ , where  $n$  is the scale parameter which will be grown to infinity. We are interested in the macroscopic behavior of such population evolution, so the scale parameter  $n$  is also used as the order of magnitude of the population size.

We consider a reference probability measure  $m(dx)$  on the space  $\mathcal{X}$  of characteristics. At birth, mutations occur thanks to a kernel  $k_b(x, a, x')m(dx')$ . At a time of swap, the characteristics  $x$  of the individual with age  $a$  are replaced by new characteristics  $x'$  drawn according to the kernel  $k_e(x, a, x')m(dx')$ .

### 6.2.1 Construction

We construct the population processes as solutions to a thinning problem. This construction for measure-valued birth-death processes can be found in Fournier and Méléard (2004), Champagnat et al. (2006), Tran (2008), Ferriere and Tran (2009), as well as in particular Bensusan et al. (2010–2015) for the detailed construction of a birth-death-swap process. Let  $(\Omega, \mathcal{A}, \mathbb{P})$  a probability space satisfying the usual conditions. On this probability space, let  $Q(ds, di, dx', d\theta)$  be a Poisson point measure on  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathcal{X} \times \mathbb{R}_+$  with intensity measure  $dsn(di)m(dx')d\theta$ , where  $n(\cdot)$  is the counting measure on  $\mathbb{N}^*$ , that is  $n(A)$  is the number of elements in  $A$ . On this probability space, let us also introduce for each  $n \in \mathbb{N}^*$  the initial population, represented by a random point measure  $Z_0^n(dx, da) = \sum_{i=1}^{N_t^0} \delta_{X^i(Z_0), A^i(Z_0)}(dx, da)$  on the space  $\mathcal{X} \times \mathbb{R}_+$  which puts a weight on characteristics and ages of individuals present at time 0. Let  $(\mathcal{F}_t)$  be the canonical filtration generated by  $Z_0$  and  $Q$ , which will be the reference filtration in this paper. The population at time  $t$  indexed by the scale parameter  $n \in \mathbb{N}^*$  is denoted  $Z_t^n(dx, da) = \sum_{i=1}^{N_t^n} \delta_{X^i(Z_t), A^i(Z_t)}(dx, da)$ , where at any time individuals are ordered by age. The virtue of the measure representation is that one can compute a function of the whole population structure using the notation

$$\langle Z_t^n, f \rangle = \int_{\mathcal{X} \times \mathbb{R}_+} f(x, a) Z_t^n(dx, da) = \sum_{i=1}^{N_t^n} f(X^i(Z_t), A^i(Z_t)).$$

For example, the population size is  $N_t^n = \langle Z_t^n, \mathbf{1} \rangle$ , whereas the total population birth intensity is  $\langle Z_{t-}^n, b \rangle$ . The population process is constructed as the solution to a *thinning* problem which can be interpreted as follows: the population at time  $t$  is computed as the population  $Z_0$  corrected by all random demographic events which happened between 0 and  $t$ , namely birth, death and swap events. More precisely, for each  $n \in \mathbb{N}^*$ , the measure-valued process  $Z_t^n(dx, da)$  is defined as the solution to

the following equation:

$$\begin{aligned}
 Z_t^n(dx, da) &= \sum_{j=1}^{\langle Z_0^n, \mathbf{1} \rangle} \delta_{(X^i(Z_0^n), A^i(Z_0^n)+t)}(dx, da) \\
 &+ \int_0^t \int_{N^*} \int_{\mathcal{X}} \int_{\mathbb{R}_+} \mathbf{1}_{i \leq \langle Z_{s-}^n, \mathbf{1} \rangle} \left( \mathbf{1}_{0 \leq \theta < m_1(Z_{s-}^n, i, x')} \delta_{(x', t-s)}(dx, da) \right. \\
 &+ \mathbf{1}_{m_1(Z_{s-}^n, i, x') \leq \theta < m_2(Z_{s-}^n, i, x')} \left( \delta_{(x', A^i(Z_{s-}^n)+t-s)}(dx, da) - \delta_{(X^i(Z_{s-}^n), A^i(Z_{s-}^n)+t-s)}(dx, da) \right) \\
 &\left. - \mathbf{1}_{m_2(Z_{s-}^n, i, x') \leq \theta < m_3(Z_{s-}^n, i, x')} \delta_{(X^i(Z_{s-}^n), A^i(Z_{s-}^n)+t-s)}(dx, da) \right) Q(ds, di, dx', d\theta),
 \end{aligned} \tag{6.1}$$

where  $m_1(Z_{s-}^n, i, x') = b(X^i(Z_{s-}^n), A^i(Z_{s-}^n))k_b(X^i(Z_{s-}^n), A^i(Z_{s-}^n), x')$ ,  
 $m_2(Z_{s-}^n, i, x') = m_1(Z_{s-}^n, i, x') + n.e(X^i(Z_{s-}^n), A^i(Z_{s-}^n))k_e(X^i(Z_{s-}^n), A^i(Z_{s-}^n), x')$ , and  
 $m_3(Z_{s-}^n, i, x') = m_2(Z_{s-}^n, i, x') + d(X^i(Z_{s-}^n), A^i(Z_{s-}^n))$ .

Existence and uniqueness results are given in Subsection 6.2.3. Note that by construction the measure-valued process has the Markov property, since the total intensity at time  $t$  is fully determined by the population  $Z_{t-}$ . In order to study the macroscopic behavior, let us first define the renormalized measure

$$\tilde{Z}_t^n(dx, da) := \frac{1}{n} Z_t^n(dx, da),$$

that is, the population in which each individual has weight  $1/n$ . The corresponding age pyramid processes of interest are defined below.

**Definition 6.1.** (*Age pyramid*) *The sequence of measure-valued process  $(\bar{Z}_t^n(da))_{t \geq 0}$  defined as the age marginal by: for each  $f \in C_b(\mathbb{R}_+)$  (continuous and bounded on  $\mathbb{R}_+$ ),  $\int_{\mathbb{R}_+} f(a) \bar{Z}_t^n(da) = \int_{\mathcal{X} \times \mathbb{R}_+} f(a) \tilde{Z}_t^n(dx, da)$  (denoted  $\langle \bar{Z}_t^n, f \rangle = \langle \tilde{Z}_t^n, f \rangle$ ) is called sequence of age pyramid processes.*

We are interested in the limit as  $n$  grows to infinity, so the frequency of swaps  $n.e(x, a)$  increases to infinity whereas the birth and death rates remain the same. The idea is to separate the time scale of demographic events (birth and death), and that of characteristics changes. At the same time, the population will be grown to infinity with  $n$  (see Assumption 6.4 below), so that two effects remain:

- (i) As the population grows, the stochastic dynamics averages to a deterministic pattern,
- (ii) As swap events occur more and more frequently, under some invariance assumption, some stable distribution of the characteristics is reached by the whole population.

Then, in the limit  $n \rightarrow +\infty$ , the age pyramid process should solve a deterministic equation in which demographic parameters are averaged over the stable distribution of the characteristics.

## 6.2.2 Assumptions

We detail here our assumptions and we discuss their use and interpretation.

**Assumption 6.1.** (*Bounded intensities and mutation densities*) Demographic rates and mutation kernels are continuous and there exists positive constants  $\bar{k}$ ,  $\bar{b}$ ,  $\bar{d}$  and  $\bar{e}$  such that for all  $(x, a, x') \in \mathcal{X} \times \mathbb{R}_+ \times \mathcal{X}$ ,  $k_e(x, a, x') \leq \bar{k}$ ,  $k_b(x, a, x') \leq \bar{k}$ ,  $b(x, a) \leq \bar{b}$ ,  $d(x, a) \leq \bar{d}$  and  $e(x, a) \leq \bar{e}$ .

**Assumption 6.2.** (*Control of moments*) There exists  $\alpha > 0$  such that

$$\sup_{n \geq 1} \mathbb{E} \left[ \left( \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right)^{1+\alpha} \right] < +\infty.$$

**Remark 6.1.** The proof can be adapted (and simplified) in the case where ages lie in  $[0, \bar{a}]$ , where  $\bar{a}$  is some limiting age. In this framework, one can check that only the assumption that  $\sup_n \mathbb{E} \left[ \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right] < +\infty$  is required. However, such limit age would constraint the model and is in fact restrictive for many applications. Therefore, our aim is to detail a more general convergence result for an unconstrained age space. This aspect will be discussed in the proof Section 6.3.

The crucial assumption about the invariance of the swap pattern is given below. It states that at the individual level, the way characteristics change during life admits an invariant distribution. Moreover, we propose here a general framework in which this invariant pattern may depend on age.

**Assumption 6.3.** (*Invariant measure*) For each  $a \in \mathbb{R}_+$ , there exists a positive solution  $x \mapsto g(x, a)$  such that  $\int_{\mathcal{X}} g(x, a) l(dx) = 1$  to the Fredholm equation:  $l(dx)$ -a.e.,

$$e(x, a)g(x, a) = \int_{\mathcal{X}} g(y, a)e(y, a)k_e(y, a, x)m(dy). \quad (6.2)$$

**Remark 6.2.** (*Probabilistic interpretation of Assumption 6.3*) We omit age for the discussion here. Consider a Markov process with values in  $\mathcal{X}$ , which jumps from a state  $x$  to a state  $y$  with rate  $e(x)k_e(x, y)$ . Its infinitesimal generator is given by  $\mathcal{A}f(x) = e(x) \left( \int_{y \in \mathcal{X}} k_e(x, y)f(y)l(dy) - f(x) \right)$  for each continuous  $f$ . It is easy to see that under Assumption 6.3, for each continuous  $f$ ,  $\int_{\mathcal{X}} \mathcal{A}f(x)g(x)l(dx) = 0$ . If the Markov process is right-continuous, then the measure  $\nu(dx) = g(x)l(dx)$  is its unique invariant measure (up to scaling).

**Remark 6.3.** Equation (6.2) is of the form  $\Psi(x) = \int_{y \in \mathcal{X}} \Psi(y)k_e(y, x)m(dy)$  (for each age) and is called homogenous Fredholm equation of the second kind. For the study of the solutions to such equations, we refer to Zemyan (2012). In Section 6.4, we provide examples of solutions both for a mixture kernel and a model with swaps to the nearest neighbor.



Before stating the assumption of convergence of the rescaled population at time 0, we clarify the space of measures and its topology. We will deal with random measures taking values in  $\mathcal{M}_F(E)$ , the space of finite positive measures on  $E$ , where mainly  $E = \mathbb{R}_+$  or  $E = \mathcal{X} \times \mathbb{R}_+$ . The space  $\mathcal{M}_F(E)$  can be embedded with the topology of the vague or weak convergence. Recall that the weak convergence in  $\mathcal{M}_F(E)$  is defined as:  $\nu_n \xrightarrow{weak} \nu$  if  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$  for each continuous and bounded  $f$ . The vague convergence is defined as  $\nu_n \xrightarrow{vague} \nu$  if  $\langle \nu_n, f \rangle \rightarrow \langle \nu, f \rangle$  for each continuous  $f$  with compact support. In the case where  $E$  is a compact set, for example if a fixed limiting age is imposed, these topologies are the same, which contributes to simplify some assumptions (see Remark 6.1). In our case however, since  $E = \mathcal{X} \times \mathbb{R}_+$  is not compact, these two topologies are strictly included.

In this paper, we denote  $\mathbb{D}([0, T], (\mathcal{M}_F(E), w))$  (resp.  $\mathbb{C}([0, T], (\mathcal{M}_F(E), w))$ ) the space of càdlàg (resp. continuous) processes on  $[0, T]$  taking values in  $\mathcal{M}_F(E)$  embedded with the topology of weak convergence.

The last assumption relates to the convergence of the sequence of initial populations towards a deterministic measure, assessing the repartition of ages and characteristics at initial time 0. This is stated below.

**Assumption 6.4.** (*Convergence of the initial population*) *There exists a deterministic measure  $\tilde{\xi}_0 \in \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$  such that the sequence of random measures  $\tilde{Z}_0^n(dx, da)$  converges in distribution and for the weak topology on  $\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$  to  $\tilde{\xi}_0(dx, da)$ .*

Note that in the case of Assumption 6.4, we have in particular  $\langle \tilde{Z}_0^n, \mathbf{1} \rangle = \frac{1}{n} \langle Z_0^n, \mathbf{1} \rangle$  converges to  $\langle \tilde{\xi}_0, \mathbf{1} \rangle$  (in distribution thus in probability since the limit is deterministic), therefore the initial sample size  $\langle Z_0^n, \mathbf{1} \rangle$  is assumed to be of order  $n$ .

### 6.2.3 Results

Here are detailed the main results of this paper. Proofs are given in Section 6.3. We before state the result on existence and strong uniqueness for the stochastic Equation (6.1). The reader is referred to Propositions 2.2.5 and 2.2.6 in Tran (2006) for the proof.

**Proposition 6.1.** *Under Assumptions 6.1 and 6.2, for each  $n \in \mathbb{N}^*$  and  $T > 0$ , there exists a unique strong solution  $\tilde{Z}^n(dx, da) \in \mathbb{D}([0, T], (\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+), w))$  to Equation (6.1).*

Our main result is stated below.

**Theorem 6.1.** (i) *Under Assumptions 6.1, 6.2, 6.3 and 6.4, the sequence of age pyramid processes  $(\bar{Z}^n(da))_n$  (see Definition 6.1) converges in distribution in  $\mathbb{D}([0, T], (\mathcal{M}_F(\mathbb{R}_+), w))$  towards the unique (deterministic) measure-valued process*

$\xi \in \mathbb{C}([0, T], (\mathcal{M}_F(\mathbb{R}_+), w))$  solution to: for each differentiable  $f$  on  $\mathbb{R}_+$ , with continuous derivatives,

$$\langle \xi_t, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \xi_s, \partial_a f + \hat{b} - \hat{d} \rangle ds, \quad (6.3)$$

with initial condition  $\xi_0(da) = \tilde{\xi}_0(\mathcal{X}, da)$ , and where

$$\hat{b}(a) = \int_{\mathcal{X}} b(x, a)g(x, a)m(dx), \quad \hat{d}(a) = \int_{\mathcal{X}} d(x, a)g(x, a)m(dx).$$

(ii) For each  $t > 0$ , the sequence of random measures  $(\tilde{Z}_t^n(dx, da))_n$  converges in distribution in  $(\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+), w)$  to  $\tilde{\xi}_t(dx, da) := g(x, a)\xi_t(da)m(dx)$ .

Equation (6.3) states that the macroscopic behavior of the age pyramid evolves as in a population with ages only, with birth and death rates given by  $\hat{b}(a)$  and  $\hat{d}(a)$ . The weak formulation of Equation (6.3) is indeed the classical transport-renewal equation, known as McKendrick-Von Foerster,

$$\begin{aligned} (\partial_a + \partial_t)\xi_t(a) &= -\hat{d}(a)\xi_t(a), \\ \xi_t(0) &= \int_0^\infty \xi_t(a)\hat{b}(a)da. \end{aligned}$$

In this model derived in point (i), the transport component states that each generation is aging and its size is decreased by the number of deaths over time, whereas the renewal component (initial condition in age) computes the number of newborns at time  $t$  based on the whole population alive at time  $t$ . Moreover, point (ii) claims first that the remaining time dependency only concerns the evolving age pyramid, and second that at each time  $t$  and age  $a$ , the characteristics distribution is given by  $g(x, a)m(dx)$ . Note also that due to different time scales between birth and swap events, the mutation kernel at birth does not appear in the macroscopic dynamics. Let us emphasize that in the macroscopic model, the set of parameters is reduced: the birth, death and swap rates are "replaced" by birth and death rates which do not depend on characteristics anymore, driving the evolution of the age pyramid over time.

Let us describe the analogy with the work of Méléard and Tran (2012) in the following remark, which is used to establish our result.

**Remark 6.4.** *In Méléard and Tran (2012), the scaling limit of birth-death measure-valued processes is considered under some allometric component, that is additional fast births and deaths with same rate. The asymetry between birth (add an individual with age 0) and death events (remove an individual with positive age) leads to a major technical issue, and the problem can be tackled by assuming some fast aging*

phenomenon and extending averaging techniques by Kurtz (1992). In the limit, with fast birth and death events, the age pyramid reaches an equilibrium at each time. So, in the limit, the fast age component is stable and a Feller diffusion describes the evolution of the population structure in terms of characteristics which evolve at the slow time scale. The tools of Méléard and Tran (2012) are used in the proof of Theorem 6.1. In our framework including additional swap patterns, the situation is reversed: the aging component is at the slow time scale of the system, whereas characteristics evolve at the fast time scale. Analogously, in the limit, the characteristics structure is stable and a specific equation describes the evolution of the age pyramid. In our case also, the particularity of the swap phenomenon, which can be seen as special simultaneous birth and death events, leads to a limiting equation which is deterministic.

A question which arises with point (ii) that states a convergence for each fixed time, refers to the possibility to get the convergence in distribution of the whole sequence of processes. In fact, it is suggested in Méléard and Tran (2012) that the sequence of processes can not be tight. The following result is a statement about the non-tightness of the sequence of the measure-valued processes. This shows that for reasonable parameters, it is not possible to improve the result of Theorem 6.1. We state this results with characteristics only, and we make the following assumption:

**Assumption 6.5.** *Let us work with age-independent demographic rates and kernels, namely  $b(x, a) \equiv b(x)$ ,  $d(x, a) \equiv d(x)$ ,  $e(x, a) \equiv e(x)$ , and  $k_e(x, a, x') \equiv k_e(x, x')$ . We consider  $\Delta$  to be the euclidian distance on  $\mathbb{R}^d$  and for a given set  $A \subset \mathcal{X}$ , we denote  $\overset{\circ}{A}$  its interior for the induced topology on  $\mathcal{X}$ . Let us assume that there exists two measurable non-empty and disjoint subsets  $A$  and  $B$  in  $\mathcal{X}$  such that*

- (i)  $\Delta(A, B) = \inf_{x \in A, y \in B} \Delta(x, y) > 0$ ,
- (ii)  $\langle \xi_0, e \mathbf{1}_{\overset{\circ}{A}} \rangle > 0$ , where  $\mathbf{1}_{\overset{\circ}{A}}$  is the indicator of the interior of  $A$ ,
- (iii)  $k_e(A, B) = \int_{x \in A} \int_{y \in B} k_e(x, y) m(dx) m(dy) > 0$ .

These assumptions are quite natural: to sum up, they state that there exists some sub-population in the initial macroscopic population whose swap parameters allow to "escape" from the current characteristics. The non-tightness result is derived below.

**Proposition 6.2.** *Under Assumptions 6.4 and 6.5, the measure-valued process  $\tilde{Z}^n(dx, da)$  is not tight in  $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$ .*

## 6.3 Proofs

### 6.3.1 Proof of Theorem 6.1

**Semi-martingale decomposition.** The semi-martingale decomposition and the control of quadratic variations are key tools for limit theorems. In the following, decompositions are performed by compensation of the Poisson Point Measure, which give further insights on the behavior of the sequence of processes. In the following, we denote  $C_b^{0,1}(\mathcal{X} \times \mathbb{R}_+)$  the set of bounded functions, continuous on  $\mathcal{X}$  and differentiable on  $\mathbb{R}_+$  with continuous and bounded partial derivative.

**Lemma 6.1.** *a) For each  $f \in C_b^{0,1}(\mathcal{X} \times \mathbb{R}_+)$ , the semi-martingale decomposition of the process  $\langle \tilde{Z}_t^n, f \rangle$  is given by*

$$\langle \tilde{Z}_t^n, f \rangle = \langle \tilde{Z}_0^n, f \rangle + \int_0^t \langle \tilde{Z}_s^n, \partial_a f + H^{f,n} \rangle ds + M_t^{f,n}, \quad (6.4)$$

where

$$\begin{aligned} H^{f,n}(x, a) &= b(x, a) \int_{\mathcal{X}} f(x', 0) k_b(x, a, x') m(dx') - d(x, a) f(x, a) \\ &+ n e(x, a) \int_{\mathcal{X}} (f(x', a) - f(x, a)) k_e(x, a, x') m(dx') \end{aligned} \quad (6.5)$$

and  $M^{f,n}$  is the local martingale (starting at zero) corresponding to the compensated Poisson point measure.

b) Let  $\tau_k = \inf\{t : \langle \tilde{Z}_t^n, 1 \rangle \geq k\}$ . Then  $M_{\cdot \wedge \tau_k}^{f,n}$  is a square-integrable martingale with quadratic variation

$$\begin{aligned} \langle M^{f,n} \rangle_{t \wedge \tau_k} &= \frac{1}{n} \int_0^{t \wedge \tau_k} ds \int_{\mathcal{X}} \tilde{Z}_s^n(dx, da) \left\{ b(x, a) \int_{\mathcal{X}} f(x', 0)^2 k_b(x, a, x') m(dx') \right. \\ &\left. + d(x, a) f(x, a)^2 + n e(x, a) \int_{\mathcal{X}} (f(x', a) - f(x, a))^2 k_e(x, a, x') m(dx') \right\}. \end{aligned} \quad (6.6)$$

**Proof of Lemma 6.1** The result is obtained by applying Equation (6.1) to a function  $f \in C_b^{0,1}(\mathcal{X} \times \mathbb{R}_+)$ . After integration by parts, the Poisson point measure  $Q$  can be compensated to get a local martingale, and the bracket can be computed. See Bensusan et al. (2010–2015) and Tran (2008) for more details.  $\diamond$

As stated in Proposition 6.2, which is proved at the end of this section, the sequence of measure-valued processes  $\tilde{Z}_t^n(dx, da)$  is not tight. This is in fact suggested by the term of order  $n$  in Equation (6.5). The remaining results we hope are

- (i) the convergence for fixed  $t$  of the sequence  $\tilde{Z}_t^n(dx, da)$ ,
- (ii) the convergence of the sequence of age pyramid processes  $\tilde{Z}_t^n(da)$ .

We now forget characteristics and focus on the age pyramid. By Lemma 6.1, the sequence of age pyramid processes verify for each bounded and differentiable  $h(a)$ ,

$$\langle \bar{Z}_t^n, h \rangle = \langle \bar{Z}_0^n, h \rangle + \int_0^t \langle \bar{Z}_s^n, \partial_a h + h(0)b - h d \rangle ds + M_t^{h,n}, \quad (6.7)$$

where  $M_{\cdot \wedge \tau_k}^{h,n}$  is a square-integrable martingale with quadratic variation

$$\langle M^{h,n} \rangle_{t \wedge \tau_k} = \frac{1}{n} \int_0^{t \wedge \tau_k} \langle \bar{Z}_s^n, h(0)^2 b + h^2 d \rangle ds. \quad (6.8)$$

Heuristically, the behavior of the quadratic variation in (6.8) is of order  $\frac{1}{n}$ , suggesting that the noise will vanish in the limit. Let us also remark the link, although not so obvious, between Equation (6.7) setting the noise to be zero, and the limit Equation (6.3) we want to derive.

**Proof of Theorem 6.1.** The proof is divided in seven steps. We follow the reasoning of Méléard and Tran (2012).

(i) Let us show that under Assumptions 6.1 and 6.2,

$$\sup_n \mathbb{E} \left[ \sup_{t \leq T} \langle \tilde{Z}_t^n, \mathbf{1} \rangle \right] < +\infty.$$

Recall that  $\tau_k = \inf\{t : \langle \tilde{Z}_t^n, \mathbf{1} \rangle \geq k\}$ . Equation (6.1) implies that

$$\langle \tilde{Z}_{t \wedge \tau_k}^n, \mathbf{1} \rangle \leq \langle \tilde{Z}_0^n, \mathbf{1} \rangle + \frac{1}{n} \int_0^t \int_{N^*} \int_{\mathcal{X}} \int_{\mathbb{R}_+} \mathbf{1}_{i \leq \sup_{0 \leq u \leq s \wedge \tau_k} \langle \tilde{Z}_{s-}^n, \mathbf{1} \rangle} \mathbf{1}_{0 \leq \theta < m_1(Z_{s-}^n, i, x')} Q(ds, di, dx', d\theta).$$

Take the supremum, then expectation and isometry formula, and finally use Assumption 6.1 to get

$$\mathbb{E} \left[ \sup_{0 \leq u \leq t} \langle \tilde{Z}_{t \wedge \tau_k}^n, \mathbf{1} \rangle \right] \leq \mathbb{E} \left[ \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right] + \bar{b} \mathbb{E} \left[ \int_0^t \sup_{0 \leq u \leq s \wedge \tau_k} \langle \tilde{Z}_{s-}^n, \mathbf{1} \rangle ds \right].$$

Fubini's theorem and Grönwall's lemma thus leads to

$$\mathbb{E} \left[ \sup_{0 \leq u \leq t} \langle \tilde{Z}_{t \wedge \tau_k}^n, \mathbf{1} \rangle \right] \leq \mathbb{E} \left[ \langle \tilde{Z}_0^n, \mathbf{1} \rangle \right] \exp(\bar{b}t).$$

Under Assumption 6.2, the right hand side is dominated by a constant which does not depend on  $n$ , that is for each  $n$ ,  $\mathbb{E} \left[ \sup_{0 \leq u \leq t} \langle \tilde{Z}_{t \wedge \tau_k}^n, \mathbf{1} \rangle \right] \leq C \exp(\bar{b}t)$ . This shows that  $\tau_k \rightarrow +\infty$  a.s.. Then Fatou's lemma leads to  $\mathbb{E} \left[ \sup_{0 \leq u \leq t} \langle \tilde{Z}_t^n, \mathbf{1} \rangle \right] \leq C \exp(\bar{b}t)$ . As the r.h.s. does not depend on  $n$ , we conclude that  $\sup_n \mathbb{E} \left[ \sup_{t \leq T} \langle \tilde{Z}_t^n, \mathbf{1} \rangle \right] < +\infty$ .

A direct corollary is that the local martingale  $(M_t^{f,n})$  with quadratic variation given by (6.6) is a square integrable martingale. Indeed,

$$\mathbb{E} \left[ \langle M^{f,n} \rangle_{t \wedge \tau_k} \right] \leq \|f\|_\infty^2 (\bar{b} + \bar{d} + 4\bar{e}) t \sup_n \mathbb{E} \left[ \sup_{0 \leq s \leq t} \langle \tilde{Z}_s^n, \mathbf{1} \rangle \right] < +\infty.$$

The last term being independent of  $k$  we get  $\mathbb{E} [\langle M^{f,n} \rangle_t] < +\infty$  by Fatou's lemma, using that  $\lim_{k \rightarrow +\infty} \tau_k = +\infty$ .

A second corollary is that for fixed  $t$  the family of random variables  $(M_t^{f,n})_n$  is uniformly integrable. Indeed, the previous inequality leads to

$$\sup_n \mathbb{E} \left[ (M_t^{f,n})^2 \right] = \sup_n \mathbb{E} [\langle M^{f,n} \rangle_t] \leq \|f\|_\infty^2 (\bar{b} + \bar{d} + 4\bar{e}) t \sup_n \mathbb{E} \left[ \sup_{0 \leq s \leq t} \langle \tilde{Z}_s^n, 1 \rangle \right] < +\infty.$$

(ii) We want to show that the sequence of age pyramid processes  $\bar{Z}^n(da) = \int_{\mathcal{X}} \tilde{Z}^n(dx, da)$  is tight in  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{R}_+))$  where  $\mathcal{M}_F(\mathbb{R}_+)$  is embedded with the topology of vague convergence. The extension to weak convergence is carried out in step (iii) .

According to Roelly-Coppoletta (1986) (Theorem 2.1), it is sufficient to prove that for each  $f \in \Theta \cup \{1\}$ ,  $\langle \bar{Z}^n, f \rangle$  is tight in  $\mathbb{D}([0, T], \mathbb{R})$  where  $\Theta$  is a dense subset of the space  $C_0(\mathbb{R}_+, \mathbb{R})$ , the space of continuous maps vanishing at infinity, for the topology of uniform convergence. By Tran (2006), Appendix A.2, the set  $\Theta = C_0(\mathbb{R}_+, \mathbb{R}) \cap C_b^1(\mathbb{R}_+, \mathbb{R})$  is dense in  $C_0(\mathbb{R}_+, \mathbb{R})$ .

Let us show that for  $f \in C_0(\mathbb{R}_+, \mathbb{R}) \cap C_b^1(\mathbb{R}_+, \mathbb{R})$ , the sequence of  $\mathbb{R}$ -valued processes  $(\langle \bar{Z}^n, f \rangle)_{n \in \mathbb{N}^*}$  is tight in  $\mathbb{D}([0, T], \mathbb{R})$ . To do this, we use Aldous-Rebolledo criterion (see Aldous (1978) and Joffe and Métivier (1986)):

since  $\langle \bar{Z}^n, f \rangle = V^{n,f} + M^{n,f}$  is a semi-martingale (with decomposition given in (6.7)), it sufficient to prove that

1.  $\forall t \in [0, T]$ ,  $(\langle M^{n,f} \rangle_t)_{n \in \mathbb{N}^*}$  and  $(V_t^{n,f})_{n \in \mathbb{N}^*}$  are uniformly tight in  $\mathbb{R}$ ,
  2.  $\forall \epsilon > 0, \forall \eta > 0, \exists \delta > 0, \exists n_0 \in \mathbb{N}^*$  such that for any sequence of stopping times  $(S_k)_{k \in \mathbb{N}^*}$  and  $(T_k)_{k \in \mathbb{N}^*}$  verifying a.s.  $\forall k \in \mathbb{N}^*, S_k \leq T_k \leq T$ ,
- $$\begin{aligned} \sup_{n \geq n_0} \mathbb{P} \left( \left| \langle M^{n,f} \rangle_{T_k} - \langle M^{n,f} \rangle_{S_k} \right| \geq \eta, T_k < S_k + \delta \right) &\leq \epsilon, \\ \sup_{n \geq n_0} \mathbb{P} \left( \left| V_{T_k}^{n,f} - V_{S_k}^{n,f} \right| \geq \eta, T_k < S_k + \delta \right) &\leq \epsilon. \end{aligned} \tag{6.9}$$

To prove the first point, it is sufficient to show that

$$\sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \langle M^{n,f} \rangle_t \right] < +\infty \text{ and } \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| V_t^{n,f} \right| \right] < +\infty.$$

According to (6.8) and step (i),

$$\sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \langle M^{n,f} \rangle_t \right] \leq \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \langle \tilde{Z}_t^n, 1 \rangle \right] T(\bar{b} + \bar{d}) \|f\|_\infty^2 < +\infty,$$

and from (6.7),

$$\begin{aligned} \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| V_t^{n,f} \right| \right] &\leq \|f\|_\infty \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \langle \tilde{Z}_0^n, 1 \rangle \right] \\ &\quad + \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \langle \tilde{Z}_t^n, 1 \rangle \right] T \left( (\bar{b} + \bar{d}) \|f\|_\infty + \|f'\|_\infty \right) < +\infty. \end{aligned}$$

To prove the second point (Equation (6.9)), let  $\epsilon > 0, \eta > 0$  and  $\forall k \in \mathbb{N}^*, S_k(\omega) \leq T_k(\omega) \leq T$  stopping times such that a.s.  $S_k \leq T_k \leq S_k + \delta$ . For all  $n_0 \in \mathbb{N}^*$ ,

$$\begin{aligned} & \sup_{n \geq n_0} \mathbb{P} \left( \left| \langle M^{n,f} \rangle_{T_k} - \langle M^{n,f} \rangle_{S_k} \right| \geq \eta \right) \\ & \leq \frac{1}{\eta} \sup_{n \geq n_0} \mathbb{E} \left[ \left| \langle M^{n,f} \rangle_{T_k} - \langle M^{n,f} \rangle_{S_k} \right| \right] \\ & \leq \frac{\delta}{\eta} \|f\|_\infty^2 (\bar{b} + \bar{d}) \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \langle \tilde{Z}_t^n, \mathbf{1} \rangle \right] < +\infty. \end{aligned}$$

Thus one can choose  $\delta > 0$  such that

$$\sup_{n \geq n_0} \mathbb{P} \left( \left| \langle M^{n,f} \rangle_{T_k} - \langle M^{n,f} \rangle_{S_k} \right| \geq \eta \right) \leq \epsilon.$$

This shows that the sequence of processes  $\langle \tilde{Z}^n, f \rangle$  is tight in  $\mathbb{D}([0, T], \mathbb{R})$ . Note that by the same reasoning, one can also show the tightness of  $\langle \tilde{Z}^n, \mathbf{1} \rangle$ . This concludes the proof of step (ii).

(iii) Let us now prove the tightness of  $(\bar{Z}^n(da))_n$  in  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{R}_+))$  where  $\mathcal{M}_F \equiv \mathcal{M}_F(\mathbb{R}_+)$  is embedded with the topology of weak convergence. To do this, we use the following criterion (see Méléard and Roelly (1993)). In the following, we denote  $\Rightarrow$  the convergence in distribution of a sequence of processes.

**Theorem 6.2.** *Let  $(Z^n)$  be a sequence of processes in  $\mathbb{D}([0, T], (\mathcal{M}_F, w))$  and  $\xi$  a process in the space  $\mathbb{C}([0, T], (\mathcal{M}_F, w))$ . Then the following statements are equivalent:*

- $Z^n \Rightarrow \xi$  in  $\mathbb{D}([0, T], (\mathcal{M}_F, w))$ ,
- $Z^n \Rightarrow \xi$  in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$  and  $\langle Z^n, \mathbf{1} \rangle \Rightarrow \langle \xi, \mathbf{1} \rangle$  in  $\mathbb{D}([0, T], \mathbb{R})$ .

Recall that the tightness of  $(\langle \bar{Z}^n, \mathbf{1} \rangle)$  in  $\mathbb{D}([0, T], \mathbb{R})$  has been derived in the previous step. Now, since  $(\bar{Z}^n(da))$  is tight in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$  and  $(\langle \bar{Z}^n, \mathbf{1} \rangle)$  is tight in  $\mathbb{D}([0, T], \mathbb{R})$ , then one can choose a subsequence  $\phi(n)$  such that

$$\bar{Z}^{\phi(n)} \Rightarrow \xi^\phi \text{ in } \mathbb{D}([0, T], (\mathcal{M}_F, v)) \text{ and } \langle \bar{Z}^{\phi(n)}, \mathbf{1} \rangle \Rightarrow Y \text{ in } \mathbb{D}([0, T], \mathbb{R}),$$

where  $(\xi_t^\phi)_{t \in [0, T]}$  is some process in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$  and  $(Y_t)_{t \in [0, T]}$  is some càdlàg real valued process. The aim now is to show that

- a)  $Y = \langle \xi^\phi, \mathbf{1} \rangle$ , and
- b)  $\xi^\phi \in \mathbb{C}([0, T], (\mathcal{M}_F, w))$ ,

which will prove the convergence of  $(\bar{Z}^{\phi(n)}(da))$  in  $\mathbb{D}([0, T], (\mathcal{M}_F, w))$ , with the help of the criterion in Theorem 6.2. This will show, by definition, that the sequence  $(\bar{Z}^n(da))$  is tight in  $\mathbb{D}([0, T], (\mathcal{M}_F, w))$ .

a) Let us first prove that  $\langle \bar{Z}^{\phi(n)}, \mathbf{1} \rangle \Rightarrow \langle \xi^\phi, \mathbf{1} \rangle$  in  $\mathbb{D}([0, T], \mathbb{R})$ .

Let  $F$  Lipschitz continuous and bounded function from  $\mathbb{D}([0, T], \mathbb{R})$  to  $\mathbb{R}$ . Let us show that

$$\limsup_{n \rightarrow +\infty} \left| \mathbb{E} \left[ F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} \rangle) - F(\langle \xi^\phi, \mathbf{1} \rangle) \right] \right| = 0.$$

To do this, we use the ideas of Jourdain et al. (2012) also used in Méléard and Tran (2012), in particular the following Lemma (see Lemma 4.3 in Jourdain et al. (2012)). The proof is postponed at the end of this section.

**Lemma 6.2.** *Introduce the functions  $f_k(a) := \Psi(0 \vee (|a| - (k - 1)) \wedge 1)$ , where  $\Psi(y) = 6y^5 - 15y^4 + 10y^3$ , which are continuous approximations of the indicator function  $\mathbf{1}_{a \geq k}$ . Then under Assumptions of Theorem 6.1,*

$$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \bar{Z}^{\phi(n)}, f_k \rangle \right] = 0,$$

and

$$\lim_{k \rightarrow +\infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \xi_t^\phi, f_k \rangle \right] = 0.$$

Now, let us introduce the terms  $F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} - f_k \rangle)$  and  $F(\langle \xi^\phi, \mathbf{1} - f_k \rangle)$  to dominate by

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} \left| \mathbb{E} \left[ F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} \rangle) - F(\langle \xi^\phi, \mathbf{1} \rangle) \right] \right| \\ & \leq \limsup_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \left| \mathbb{E} \left[ F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} \rangle) - F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} - f_k \rangle) \right] \right| \\ & + \limsup_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \left| \mathbb{E} \left[ F(\langle \bar{Z}^{\phi(n)}, \mathbf{1} - f_k \rangle) - F(\langle \xi^\phi, \mathbf{1} - f_k \rangle) \right] \right| \\ & + \limsup_{k \rightarrow +\infty} \left| \mathbb{E} \left[ F(\langle \xi^\phi, \mathbf{1} - f_k \rangle) - F(\langle \xi^\phi, \mathbf{1} \rangle) \right] \right|. \end{aligned}$$

For the second term, since  $\bar{Z}^{\phi(n)} \Rightarrow \xi^\phi$  in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$  and  $(\mathbf{1} - f_k) \in C_K(\mathbb{R}_+)$ , the continuous mapping theorem (see e.g. Billingsley (2009)) implies that the second term is zero. Also, by the Lipschitz property,

$$|F(\langle \nu, \mathbf{1} - f_k \rangle) - F(\langle \nu, \mathbf{1} \rangle)| \leq [F]_{lip} \sup_{t \in [0, T]} \langle \nu_t, f_k \rangle,$$

so according to Lemma 6.2, the first term and third terms are equal to zero.

b) Let us now prove that  $\xi^\phi \in \mathbb{C}([0, T], (\mathcal{M}_F, v))$  (vague topology). Since  $\bar{Z}^{\phi(n)} \Rightarrow \xi$  in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$  and for each  $f \in C_K(\mathbb{R}_+)$ , the map  $\nu \in D([0, T], (\mathcal{M}_F(\mathbb{R}_+), v)) \mapsto \sup_{t \in [0, T]} |\langle \nu_t, f \rangle - \langle \nu_{t-}, f \rangle|$  is continuous, the continuous mapping theorem implies that

$$\sup_{t \in [0, T]} \left| \langle \bar{Z}_t^{\phi(n)}, f \rangle - \langle \bar{Z}_{t-}^{\phi(n)}, f \rangle \right| \Rightarrow \sup_{t \in [0, T]} \left| \langle \xi_t^\phi, f \rangle - \langle \xi_{t-}^\phi, f \rangle \right|.$$



Moreover, by construction if there is a jump the quantity  $\langle \bar{Z}_t^{\phi(n)}, f \rangle$  increases at most by the amount  $\frac{\|f\|_\infty}{\Phi(n)}$ , which turns out that  $\sup_{t \in [0, T]} \left| \langle \bar{Z}_t^{\phi(n)}, f \rangle - \langle \bar{Z}_{t-}^{\phi(n)}, f \rangle \right| \leq \frac{\|f\|_\infty}{\Phi(n)}$ , and in particular  $\sup_{t \in [0, T]} \left| \langle \bar{Z}_t^{\phi(n)}, f \rangle - \langle \bar{Z}_{t-}^{\phi(n)}, f \rangle \right| \Rightarrow 0$ . This shows that for each  $f \in C_K(\mathbb{R}_+)$ , a.s.  $\sup_{t \in [0, T]} \left| \langle \xi_t^\phi, f \rangle - \langle \xi_{t-}^\phi, f \rangle \right| = 0$  thus  $\xi^\phi \in \mathbb{C}([0, T], (\mathcal{M}_F, v))$ .

Now, we prove that  $\xi^\phi \in \mathbb{C}([0, T], (\mathcal{M}_F, w))$  (weak topology). Let  $h \in C_b(\mathbb{R}_+)$  and the  $(f_k)$  as in Lemma 6.2. Then

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \langle \xi_t^\phi, h \rangle - \langle \xi_{t-}^\phi, h \rangle \right| \right] &\leq \|h\|_\infty \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \langle \xi_t^\phi, f_k \rangle \right| \right] \\ &+ \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \langle \xi_t^\phi, h(\mathbf{1} - f_k) \rangle - \langle \xi_{t-}^\phi, h(\mathbf{1} - f_k) \rangle \right| \right] \\ &+ \|h\|_\infty \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \langle \xi_{t-}^\phi, f_k \rangle \right| \right]. \end{aligned}$$

Since  $\xi^\phi \in \mathbb{C}([0, T], (\mathcal{M}_F, v))$  (vague topology) and  $h(\mathbf{1} - f_k) \in C_K(\mathbb{R}_+)$ , the second term is zero. In addition, by the use of Lemma 6.2, the first and third terms are zero by letting  $k \rightarrow +\infty$ .

(iv) Let us prove that for each  $t > 0$ , the sequence of measures  $(\tilde{Z}_t^n(dx, da))_n$  is uniformly tight in  $\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$ .

To show the tightness of a sequence of measures, we use the following result (see Kurtz (1992), Lemma 1.1):

**Lemma 6.3.** *Let  $(\mu_n)$  be a sequence of random variables with values in  $(\mathcal{M}_F(S), w)$ , with  $(S, d)$  a complete separable metric space. Then  $(\mu_n)$  is relatively compact in  $(\mathcal{M}_F(S), w)$  if and only if the sequence  $(\mu_n(S))$  is relatively compact in  $\mathbb{R}$  and for each  $\epsilon > 0$  there exists a compact set  $K \subset S$  such that  $\sup_n \mathbb{P}(\mu_n(K^c) > \epsilon) < \epsilon$ .*

First note that  $\tilde{Z}_t^n(\mathcal{X} \times \mathbb{R}_+) = \langle \tilde{Z}_t^n, \mathbf{1} \rangle$ . Let  $\epsilon > 0$ , then for  $k$  large enough,

$$\sup_{n \geq 1} \mathbb{P}(\langle \tilde{Z}_t^n, \mathbf{1} \rangle \notin [0, k]) \leq \frac{1}{k} \sup_{n \geq 1} \mathbb{E}[\langle \tilde{Z}_t^n, \mathbf{1} \rangle] \leq \epsilon,$$

so that the first condition is matched. For the second condition, let  $A > 0$  and consider the compact set  $K := \mathcal{X} \times [0, A]$ . Obviously,  $\tilde{Z}_t^n(K^c) = 0$  since all characteristics lie in  $\mathcal{X}$ : this proves that the second condition is satisfied.

The tightness of the sequence of the age pyramid processes and the time marginal of the population process has been established. To prove the convergence in distribution, it is needed to identify the limiting values. In our framework, this issue is technical due to the fast swap pattern: it is not possible to keep track of the characteristics in the population over time. The issue has an analogy with that in Méléard and Tran (2012) and Gupta et al. (2014) in which fast aging is considered

(see Remark 6.4). The main technical tool they use for the proof is an extension of averaging techniques by Kurtz (1992). This is used in the following step: the idea is to identify the limiting values of the sequence of occupation measures  $(\Gamma^n(dx, da, dt))_n := (\tilde{Z}_t^n(dx, da)dt)_n$ .

(v) To show that the sequence of measures  $(\Gamma^n(dx, da, dt))_n$  is tight in  $(\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+ \times [0, T]), w)$ , we use the same result as for the previous step. As previously, the second condition is straightforward with  $\tilde{K} := \mathcal{X} \times [0, A] \times [0, T]$ . As for the first condition, we have

$$\sup_n \mathbb{P}(\Gamma^n(\mathcal{X} \times \mathbb{R}_+ \times [0, T]) \notin [0, k]) = \sup_n \mathbb{P}\left(\int_0^T \langle \tilde{Z}_t^n, \mathbf{1} \rangle dt > k\right) \leq \frac{T}{k} \sup_n \mathbb{E}\left[\sup_{0 \leq t \leq T} \langle \tilde{Z}_t^n, \mathbf{1} \rangle\right],$$

which can be made smaller than  $\epsilon$  for  $k$  large enough.

Now, from (iv) and (v), the sequence  $(\Gamma^n, \bar{Z}^n(da))_n$  is tight. Due to Prohorov theorem, from any subsequence, one can extract a further sub-subsequence converging in distribution. Denote  $(\Gamma^{\Phi(n)}, \bar{Z}^{\Phi(n)}(da))$  such sub-subsequence and  $(\bar{\Gamma}^\Phi, \bar{X}^\Phi(da))$  the corresponding limit (the first component being a measure and the second component a measure-valued process). The last two steps of the proof are dedicated to the characterization of this limit. If it is unique (i.e. does not depend on  $\Phi$ ), we get the convergence in distribution (see Billingsley (2009), Theorem 2.6). Step (vi) focuses on the marginal in characteristics of  $\bar{\Gamma}^\Phi$ , whereas Step (vii) concentrates on the age and time marginal.

(vi) By definition of  $\Gamma^{\Phi(n)}$ , for each continuous and bounded  $f(a, s)$ ,

$$\int_0^t \int_{\mathbb{R}_+} \int_{\mathcal{X}} f(a, s) \Gamma^{\Phi(n)}(dx, da, ds) = \int_0^t \int_{\mathbb{R}_+} f(a, s) \bar{Z}_s^{\Phi(n)}(da) ds.$$

Since  $\Gamma \mapsto \int_0^t \int_{\mathbb{R}_+} \int_{\mathcal{X}} f(a, s) \Gamma(dx, da, ds)$  and  $Z \mapsto \int_0^t \int_{\mathbb{R}_+} f(a, s) Z_s(da) ds$  are both continuous, the continuous mapping theorem implies that  $\bar{X}_s^\Phi(da) ds$  is necessarily the marginal measure of  $\bar{\Gamma}^\Phi$  on  $\mathbb{R}_+ \times [0, T]$  up to a null-measure set.

By Kurtz (1992), Lemma 1.4, there exists a predictable probability-valued process  $\gamma_{a,s}^\Phi(dx)$  such that a.s.,  $dt - ae$ ,  $\bar{\Gamma}^\Phi(dx, da, ds) = \gamma_{a,s}^\Phi(dx) \bar{X}_s^\Phi(da) ds$ .

Now, the aim is to characterize  $\gamma_{a,s}^\Phi(dx)$ . From (6.4),  $(\frac{1}{\Phi(n)} M_t^{f, \Phi(n)})_t$  is a martingale, converging in distribution to (continuous mapping theorem for  $(\Gamma^{\Phi(n)})_n$ )

$$\bar{M}_t^\Phi := \int_0^t \int_{\mathbb{R}_+} \int_{x \in \mathcal{X}} e(x, a) \left( \int_{\mathcal{X}} (f(x', a) - f(x, a)) k_e(x, a, x') m(dx') \right) \gamma_{a,s}^\Phi(dx) \bar{X}_s^\Phi(da) ds.$$

For a given  $t > 0$  the family  $(\frac{1}{n} M_t^{f, n})_n$  is uniformly integrable since from (i),

$$\sup_n \mathbb{E}\left[(M_t^{f, n})^2\right] = \sup_n \mathbb{E}\left[\langle M^{f, n} \rangle_t\right] \leq \|f\|_\infty^2 (\bar{b} + \bar{d} + 4\bar{e}) t \sup_n \mathbb{E}\left[\sup_{0 \leq s \leq t} \langle \tilde{Z}_s^n, \mathbf{1} \rangle\right] < +\infty.$$

This implies that  $\bar{M}^\Phi$  is a martingale. But it is by construction a finite-variation process which is also continuous so it is the null process (up to indistinguishability). We get *a.s.*,  $X_s^\Phi(da) - a.e.$  and  $dt - a.e.$ ,

$$\int_{x \in \mathcal{X}} f(x, a) e(x, a) \gamma_{a,s}^\Phi(dx) = \int_{x' \in \mathcal{X}} f(x', a) \left( \int_{x \in \mathcal{X}} e(x, a) k_e(x, a, x') \gamma_{a,s}^\Phi(dx) \right) m(dx').$$

This shows that (*a.s.*,  $X_s^\Phi(da) - a.e.$  and  $dt - a.e.$ )  $\gamma_{a,s}^\Phi$  is absolutely continuous w.r.t.  $m$ ,  $\gamma_{a,s}^\Phi(dx) = \gamma_{a,s}^\Phi(x) m(dx)$ , with

$$\gamma_{a,s}^\Phi(x) e(x, a) = \int_{y \in \mathcal{X}} e(y, a) k_e(y, a, x) \gamma_{a,s}^\Phi(y) m(dy),$$

and in addition  $\gamma_s^\Phi$  is a probability measure. Then under Assumption 6.3, we get  $\gamma_{a,s}^\Phi(x) = g(x, a)$ . We just characterized that all limiting values of  $\Gamma^{\Phi(n)}$  are of the form  $\bar{\Gamma}^\Phi(dx, da, ds) = g(x, a) m(dx) \bar{X}_s^\Phi(da) ds$ . In the last step, we identify  $\bar{X}^\Phi$  as the solution to a deterministic equation.

(vii) We want to show that *a.s.* for each  $t$  and  $f \in C_b^1(\mathbb{R}_+)$ ,

$$\langle \bar{X}_t^\Phi, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \bar{X}_s^\Phi, \partial_a f + \hat{b} - \hat{d} \rangle ds,$$

where  $\hat{b}(a) = \int_{\mathcal{X}} b(x, a) g(x, a) m(dx)$  and  $\hat{d}(a) = \int_{\mathcal{X}} d(x, a) g(x, a) m(dx)$ . Let  $\bar{H}_t^\Phi = \langle \bar{X}_t^\Phi, f \rangle_t - \langle \xi_0, f \rangle - \int_0^t \langle \bar{X}_s^\Phi, \partial_a f + \hat{b} - \hat{d} \rangle ds$ . To prove that  $\mathbb{E}[\bar{H}_t^\Phi] = 0$ , we use the following three facts:

a) From Equation (6.7), one has  $M_t^{f, \Phi(n)} \Rightarrow \bar{H}_t^\Phi$ , using that  $\Gamma^{\Phi(n)} \Rightarrow \bar{\Gamma}^\Phi$ , Assumption 6.4 and the continuous mapping theorem.

b) One can also write

$$\begin{aligned} \mathbb{E} \left[ \left| M_t^{f, \Phi(n)} \right|^2 \right] &\leq \mathbb{E} \left[ \left( M_t^{f, \Phi(n)} \right)^2 \right] = \mathbb{E} \left[ \langle M^{f, \Phi(n)} \rangle_t \right] \\ &\leq \frac{1}{\Phi(n)} t \|f\|_\infty^2 (\bar{b} + \bar{d}) \sup_{n \geq 1} \mathbb{E} \left[ \sup_{0 \leq s \leq t} \langle \tilde{Z}_s^n, \mathbf{1} \rangle \right], \end{aligned}$$

which shows that  $\lim_{n \rightarrow +\infty} \mathbb{E} \left[ \left| M_t^{f, \Phi(n)} \right|^2 \right] = 0$  since  $\Phi(n) \rightarrow +\infty$ .

c) From the second corollary of (i), for fixed  $t$ , the sequence  $(M_t^{f, \Phi(n)})_n$  is uniformly integrable.

From these facts we get  $\mathbb{E}[\bar{H}_t^\Phi] = \lim_{n \rightarrow +\infty} \mathbb{E} \left[ M_t^{f, \Phi(n)} \right] = 0$ .

Using Grönwall lemma, one can prove the uniqueness for Equation (6.3). This concludes the proof.

**Proof of Lemma 6.2.** The proof is similar to Jourdain et al. (2012), exposed in order to illustrate that here only the moments of order  $1 + \alpha$  of Assumption

6.2 are needed (order 1 suffices elsewhere in the proof of Theorem 6.1). Note that since  $f_k \notin C_K(\mathbb{R}_+)$ , it is not possible to use the fact that  $(\bar{Z}^{\Phi(n)}(da))$  converges in distribution in  $\mathbb{D}([0, T], (\mathcal{M}_F, v))$ . We rather use Equation (6.7) with  $h \equiv f_k$  for  $k \geq 1$  leads to (since  $f_k(0) = 0$ ),

$$\langle \bar{Z}_t^{\Phi(n)}, f_k \rangle = \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle + \int_0^t \langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k - f_k d \rangle ds + M_t^{f_k, \Phi(n)}. \quad (6.10)$$

Dominate by omitting the death term, take the supremum, then expectation and finally use Doob inequality to get

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \langle \bar{Z}_t^{\Phi(n)}, f_k \rangle \right] \leq \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \right] + 2 \left( \mathbb{E} \left[ \langle M^{f_k, \Phi(n)} \rangle_T \right] \right)^{1/2} + \int_0^T \mathbb{E} \left[ \langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k \rangle \right] ds.$$

The limits of the three terms are studied separately.

(i) Since  $\tilde{Z}_0^{\phi(n)} \Rightarrow \xi_0$  in  $(\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+), w)$  and  $f_k \in C_b(\mathcal{X} \times \mathbb{R}_+)$ , then  $\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \Rightarrow \langle \xi_0, f_k \rangle$ . We also have  $f_k \leq 1$  and under Assumption 6.2,  $\sup_n \mathbb{E} \left[ (\langle \bar{Z}_0^n, \mathbf{1} \rangle)^{1+\alpha} \right] < +\infty$  so the sequence is uniformly integrable. We thus get  $\lim_{n \rightarrow +\infty} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \right] = \langle \xi_0, f_k \rangle$ . Finally, since  $\xi_0 \in \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$ , that is  $\langle \xi_0, \mathbf{1} \rangle < +\infty$ , we get  $\lim_{k \rightarrow +\infty} \langle \xi_0, f_k \rangle = 0$ , then

$$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \right] = 0.$$

(ii) From (6.8), since  $f_k(0) = 0$ ,

$$\mathbb{E} \left[ \langle M^{f_k, \Phi(n)} \rangle_T \right] = \frac{1}{\Phi(n)} \int_0^t \langle \bar{Z}_s^{\Phi(n)}, f_k^2 d \rangle ds \leq \frac{\bar{d}T}{\Phi(n)} \sup_{n \geq 1} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \tilde{Z}_s^n, \mathbf{1} \rangle \right],$$

which shows that  $\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E} \left[ \langle M^{f_k, \Phi(n)} \rangle_T \right] = 0$ .

(iii) Note that

$$\partial_a f_k = \mathbf{1}_{k-1 \leq a \leq k} \Psi'(a - (k-1)) \leq \mathbf{1}_{k-1 \leq a \leq k} \sup_{y \in [0, 1]} \Psi'(y) \leq \mathbf{1}_{k-1 \leq a} \sup_{y \in [0, 1]} \Psi'(y).$$

Since  $\mathbf{1}_{k-1 \leq a} \leq f_{k-1}(a)$ , this shows that  $\partial_a f_k(a) \leq f_{k-1}(a) \sup_{y \in [0, 1]} \Psi'(y)$ . To show that  $\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \int_0^T \mathbb{E} \left[ \langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k \rangle \right] ds = 0$ , it is thus sufficient to prove that for  $t \in [0, T]$ ,  $\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E} \left[ \langle \bar{Z}_t^{\Phi(n)}, f_k \rangle \right] = 0$ . From (6.10), for each  $k \geq 1$ , with  $C := \sup_{y \in [0, 1]} \Psi'(y)$ ,

$$\mathbb{E} \left[ \langle \bar{Z}_t^{\Phi(n)}, f_k \rangle \right] \leq \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \right] + C \int_0^t \mathbb{E} \left[ \langle \bar{Z}_s^{\Phi(n)}, f_{k-1} \rangle \right] ds, \quad (6.11)$$

and for  $k = 0$ ,  $f_0 \equiv \mathbf{1}$  then

$$\mathbb{E} \left[ \langle \bar{Z}_t^{\Phi(n)}, f_0 \rangle \right] \leq \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_0 \rangle \right] + \bar{b} \int_0^t \mathbb{E} \left[ \langle \bar{Z}_s^{\Phi(n)}, f_0 \rangle \right] ds,$$

so Grönwall's lemma leads to

$$\mathbb{E} \left[ \langle \bar{Z}_t^{\Phi(n)}, f_0 \rangle \right] \leq \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_0 \rangle \right] \exp(\bar{b}t). \quad (6.12)$$

From (6.11) and (6.12) we get

$$\mathbb{E} \left[ \langle \bar{Z}_t^{\Phi(n)}, f_k \rangle \right] \leq \sum_{j=0}^{k-1} \frac{C^j t^j}{j!} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{k-j} \rangle \right] + \frac{C^k t^k}{k!} \exp(\bar{b}t) \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, \mathbf{1} \rangle \right].$$

Under Assumption 6.2, we have  $\frac{C^k t^k}{k!} \exp(\bar{b}t) \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, \mathbf{1} \rangle \right] \leq \frac{C^k t^k}{k!} \exp(\bar{b}t) \sup_n \mathbb{E} \left[ \langle \bar{Z}_0^n, \mathbf{1} \rangle \right]$ , which tends to zero as  $k \rightarrow +\infty$ . As for the first term, this can be split into

$$\begin{aligned} \sum_{j=0}^{k-1} \frac{C^j t^j}{j!} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{k-j} \rangle \right] &= \sum_{j=0}^{[k/2]} \frac{C^j t^j}{j!} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{k-j} \rangle \right] + \sum_{j=[k/2]+1}^{k-1} \frac{C^j t^j}{j!} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{k-j} \rangle \right] \\ &\leq \exp(Ct) \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{k-[k/2]} \rangle \right] + \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, \mathbf{1} \rangle \right] \sum_{j=[k/2]+1}^{k-1} \frac{C^j t^j}{j!}. \end{aligned}$$

Since  $[k/2] \leq k - [k/2]$  then  $f_{k-[k/2]} \leq f_{[k/2]}$ . As in (i), by the convergence in distribution and uniform integrability (Assumption 6.2), one gets

$\lim_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{E} \left[ \langle \bar{Z}_0^{\Phi(n)}, f_{[k/2]} \rangle \right] = 0$ . As for the second term, this converges to zero since the first component is bounded (Assumption 6.2) and the sum converges to zero.  $\diamond$

### 6.3.2 Proof of Proposition 6.2

Let us recall below the assumptions and the corresponding result about the non-tightness of the sequence of population processes.

**Assumption 6.5** *Let us work with age-independent demographic rates and kernels, namely  $b(x, a) \equiv b(x)$ ,  $d(x, a) \equiv d(x)$ ,  $e(x, a) \equiv e(x)$ , and  $k_e(x, a, x') \equiv k_e(x, x')$ . We consider  $\Delta$  to be the euclidian distance on  $\mathbb{R}^d$  and for a given set  $A \subset \mathcal{X}$ , we denote  $\overset{\circ}{A}$  its interior for the induced topology on  $\mathcal{X}$ . Let us assume that there exists two measurable non-empty and disjoint subsets  $A$  and  $B$  in  $\mathcal{X}$  such that*

- (i)  $\Delta(A, B) = \inf_{x \in A, y \in B} \Delta(x, y) > 0$ ,
- (ii)  $\langle \xi_0, e \mathbf{1}_{\overset{\circ}{A}} \rangle > 0$ , where  $\mathbf{1}_{\overset{\circ}{A}}$  is the indicator of the interior of  $A$ ,
- (iii)  $k_e(A, B) = \int_{x \in A} \int_{y \in B} k_e(x, y) m(dx) m(dy) > 0$ .

**Proposition 6.2** *Under Assumptions 6.4 and 6.5, the measure-valued process  $\tilde{Z}^n(dx, da)$  is not tight in  $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$ .*

#### Proof of Proposition 6.2

Aim is to give here a proof which does not need the main result of Theorem 6.1.

Let us first construct a function  $f \in \mathcal{C}_b(\mathcal{X})$  such that

$$\inf_{x \in A, y \in B} |f(y) - f(x)| > 0.$$

To this aim, consider for example the map  $f : x \mapsto \Delta(x, A)$  with support in  $\mathcal{X} \setminus A$ , that satisfies the previous equation according to point (i). It is moreover continuous on the compact  $\mathcal{X}$  so it is bounded. To prove that the sequence of measure-valued processes  $(\tilde{Z}^n(dx, da))_n$  is not tight in  $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$ , it is sufficient to prove that the process  $X^n := \langle \tilde{Z}^n, f \rangle$  is not tight in  $\mathbb{D}([0, T], \mathbb{R})$ . Recall the following equivalence derived from compactness characterization in the space of càdlàg functions (see Billingsley (2009), Theorem 13.2):

The sequence  $(X^n)_n$  is tight  $\mathbb{D}([0, T], \mathbb{R})$  if and only if

- a) the sequence  $\sup_{0 \leq s \leq T} |X_t^n|$  is tight in  $\mathbb{R}$  and
- b) (equicontinuity)  $\forall \eta > 0, \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow +\infty} \mathbb{P}(w'(X^n, \delta) > \eta) = 0$ , where  $w'(X^n, \delta) = \inf_{(t_i)_{>\delta}} \sup_i \sup_{s, t \in [t_{i-1}, t_i[} |X_t^n - X_s^n|$  and  $\inf_{(t_i)_{>\delta}}$  denotes the infimum over all subdivisions of  $[0, T]$  with step strictly greater than  $\delta$ .

To prove the result, we show that the equicontinuity criterion is not satisfied, that is we show that:

$$\exists \eta > 0, \exists \epsilon > 0, \forall \delta > 0, \limsup_{n \rightarrow +\infty} \mathbb{P} \left( \inf_{(t_i)_{>\delta}} \sup_i \sup_{s, t \in [t_{i-1}, t_i[} |X_t^n - X_s^n| > \eta \right) > \epsilon.$$

First remark that for each  $\delta > 0$  and  $\eta > 0$ ,

$$\mathbb{P} \left( \inf_{(t_i)_{>\delta}} \sup_i \sup_{s, t \in [t_{i-1}, t_i[} |X_t^n - X_s^n| > \eta \right) \geq \mathbb{P} \left( \sup_{t \in [0, \delta]} |X_t^n - X_{t-}^n| > \eta \right).$$

Second, denote  $\tau^n$  the first time of jump of  $Z^n$ . Conditionally on  $Z_0^n$ ,  $\tau^n$  is exponentially distributed with parameter  $n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle$ . Then

$$\mathbb{P}(\tau^n \leq \delta \mid Z_0^n) = 1 - e^{-\delta(n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle)}.$$

Let  $\tau_e^n$  be the first time of swap for  $Z^n$ . By independence between swap, birth and death events, the probability that the first time of event is a swap is given by (with convention  $\frac{0}{0} = 0$ )

$$\mathbb{P}(\tau^n = \tau_e^n \mid Z_0^n) = \frac{n\langle Z_0^n, e \rangle}{n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle},$$

and the clock lemma for independent exponentially distributed random variables assesses independence, conditionally on  $Z_0^n$ , between the events  $\{\tau^n = \tau_e^n\}$  and  $\{\tau^n \leq \delta\}$ .

Let  $Y$  be the random variable "characteristics of the individual for which the event

occurs at time  $\tau^n$ , and  $A$  and  $B$  as in Assumption 6.5. The probability that the individual has characteristics in  $A$  conditionally on a swap event is

$$\mathbb{P}(Y \in A \mid Z_0^n, \tau^n = \tau_e^n) = \frac{\langle Z_0^n, e \mathbf{1}_A \rangle}{\langle Z_0^n, e \rangle}.$$

Denote  $Y'$  the new characteristics (at birth or swap). In case of a swap, the characteristics  $Y$  are replaced by the new characteristics  $Y'$ , drawn with distribution  $k_e(Y, \cdot)$ . Then

$$\mathbb{P}(Y' \in B \mid Y \in A, \tau^n = \tau_e^n, Z_0^n) = \frac{k_e(A, B)}{m(A)}.$$

Finally, remark that  $Y$  and  $Y'$  are independent of  $\tau^n$  and let us take

$\eta = \frac{1}{2} \inf_{x \in A, y \in B} |f(y) - f(x)|$  which is positive by construction of  $f$ . Then

$$\begin{aligned} & \mathbb{P} \left( \sup_{t \in [0, \delta]} |X_t^n - X_{t-}^n| > \eta \right) \\ & \geq \mathbb{P}(|X_{\tau^n \wedge \delta}^n - X_0^n| > \eta) \\ & \geq \mathbb{P}(\tau^n \leq \delta, \tau^n = \tau_e^n, Y \in A, Y' \in B) \\ & = \mathbb{E} [\mathbb{P}(\tau^n \leq \delta \mid Z_0^n) \mathbb{P}(\tau^n = \tau_e^n, Y' \in B, Y \in A \mid Z_0^n)] \\ & = \mathbb{E} [\mathbb{P}(\tau^n \leq \delta \mid Z_0^n) \mathbb{P}(Y' \in B \mid \tau^n = \tau_e^n, Y \in A, Z_0^n) \mathbb{P}(Y \in A \mid \tau^n = \tau_e^n, Z_0^n) \mathbb{P}(\tau^n = \tau_e^n \mid Z_0^n)] \\ & = \frac{k_e(A, B)}{m(A)} \mathbb{E} \left[ \left( 1 - e^{-\delta(n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle)} \right) \frac{\langle Z_0^n, e \mathbf{1}_A \rangle}{\langle Z_0^n, e \rangle} \frac{n\langle Z_0^n, e \rangle}{n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle} \right] \end{aligned}$$

where the last two equalities are due to recursive conditioning and the use of the previous computations respectively. Now, let us construct  $h_A \in \mathcal{C}_b(\mathcal{X})$  such that  $h_A \leq \mathbf{1}_A$  and  $\langle \xi_0, e h_A \rangle > 0$ . Let us define

$$h_A : x \mapsto \frac{\Delta(x, \mathcal{X} \setminus A)}{\sup_{y \in A} \Delta(y, \mathcal{X} \setminus A)},$$

which is continuous on  $\mathcal{X}$  and positive on  $\dot{A}$ ; according to point (ii) in Assumption 6.5, we get  $\langle \xi_0, e h_A \rangle > 0$ . Since  $\langle \xi_0, e h_A \rangle \leq \langle \xi_0, e \mathbf{1}_A \rangle$ , we have

$$\begin{aligned} & \mathbb{P} \left( \sup_{t \in [0, \delta]} |X_t^n - X_{t-}^n| > \eta \right) \\ & \geq \frac{k_e(A, B)}{m(A)} \mathbb{E} \left[ \left( 1 - e^{-\delta(n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle)} \right) \frac{n\langle Z_0^n, e h_A \rangle}{n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle} \right] \\ & \geq \frac{k_e(A, B)}{m(A)} \mathbb{E} \left[ \frac{n\delta\langle Z_0^n, e h_A \rangle}{1 + \delta(n\langle Z_0^n, e \rangle + \langle Z_0^n, b \rangle + \langle Z_0^n, d \rangle)} \right] \end{aligned}$$

where the last inequality comes from the fact that  $1 - \exp(-x) \geq \frac{x}{1+x}$ . Renormalization by  $\delta n^2$  leads to the final inequality

$$\mathbb{P} \left( \sup_{t \in [0, \delta]} |X_t^n - X_{t-}^n| > \eta \right) \geq \frac{k_e(A, B)}{m(A)} \mathbb{E} \left[ \frac{\langle \tilde{Z}_0^n, e h_A \rangle}{1/(\delta n^2) + \langle \tilde{Z}_0^n, e \rangle + \langle \tilde{Z}_0^n, b \rangle/n + \langle \tilde{Z}_0^n, d \rangle/n} \right].$$

Under Assumption 6.4, one has that  $(\langle \tilde{Z}_0^n, eh_A \rangle, \langle \tilde{Z}_0^n, e \rangle, \langle \tilde{Z}_0^n, d \rangle, \langle \tilde{Z}_0^n, b \rangle)$  converges to the deterministic limit  $(\langle \xi_0, eh_A \rangle, \langle \xi_0, e \rangle, \langle \xi_0, d \rangle, \langle \xi_0, b \rangle)$  in probability thus one can get the a.s. convergence for a subsequence  $\phi(n)$ .

Moreover, since  $\frac{\langle \tilde{Z}_0^n, eh_A \rangle}{1/(\delta n^2) + \langle \tilde{Z}_0^n, e \rangle + \langle \tilde{Z}_0^n, b \rangle/n + \langle \tilde{Z}_0^n, d \rangle/n} \leq 1$ , the dominated convergence theorem leads to

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} \mathbb{P} \left( \sup_{t \in [0, \delta]} |X_t^n - X_{t-}^n| > \eta \right) \\ & \geq \frac{k_e(A, B)}{m(A)} \lim_{n \rightarrow +\infty} \mathbb{E} \left[ \frac{\langle \tilde{Z}_0^{\phi(n)}, eh_A \rangle}{1/(\delta \phi(n)^2) + \langle \tilde{Z}_0^{\phi(n)}, e \rangle + \langle \tilde{Z}_0^{\phi(n)}, b \rangle/\phi(n) + \langle \tilde{Z}_0^{\phi(n)}, d \rangle/\phi(n)} \right] \\ & = \frac{k_e(A, B)}{m(A)} \frac{\langle \xi_0, eh_A \rangle}{\langle \xi_0, e \rangle}. \end{aligned}$$

The choice  $\epsilon = \frac{1}{2} \frac{k_e(A, B)}{m(A)} \frac{\langle \xi_0, eh_A \rangle}{\langle \xi_0, e \rangle}$ , which is positive by Assumption 6.5 (iii), concludes the proof.  $\diamond$

## 6.4 Examples and numerical illustration

### 6.4.1 Examples

We propose here to compute explicit solutions to Equation (6.2) for the stable composition. We analyze two examples: the two-dimensional mixture and the swap to the nearest neighbor. We omit here the dependence in age for clarity.

**Mixture kernel.** We want to derive explicit solutions to (6.2) in the case where  $k_e(y, x)$  is a two dimensional mixture, that is  $k_e(y, x) = b_1(y)f_1(x) + b_2(y)f_2(x)$  where  $b_1$  and  $b_2$  are continuous functions from  $\mathcal{X}$  to  $(0, 1)$  such that  $b_1 + b_2 \equiv \mathbf{1}$ , and  $f_1$  and  $f_2$  are probability densities on  $\mathcal{X}$ . In this model, new characteristics are chosen based on two probability densities which are the same for all individuals but weights depend on the old characteristics of the individual. This is a particular case of a separable kernel with  $n = 2$ , and we illustrate the methodology described in Zemyan (2012) for the computations. We first derive the characteristic polynomial in  $\lambda \in \mathbb{R}$  of the matrix

$$A = \begin{pmatrix} \int b_1 f_1 & \int b_1 f_2 \\ \int b_2 f_1 & \int b_2 f_2 \end{pmatrix} = \begin{pmatrix} d_1 & 1 - d_2 \\ 1 - d_1 & d_2 \end{pmatrix},$$

where  $d_1 = \int b_1 f_1$  and  $d_2 = \int b_2 f_2$ . Then

$$\begin{aligned} P(\lambda) & := |I - \lambda A| = 1 - \lambda d_2 - \lambda d_1 - \lambda^2 + \lambda^2 d_2 + \lambda^2 d_1, \\ & = (1 - \lambda)(1 + \lambda - \lambda d_2 - \lambda d_1). \end{aligned}$$

Since  $P(1) = 0$ , there exists non-trivial solutions to  $(I - A)c = 0$ . Let  $Q(\lambda) = P(\lambda)/(1 - \lambda)$ . Then  $Q(1) = 2 - d_1 - d_2 > 0$  since  $d_1 = \int f_1 b_1 < \int f_1 = 1$  (we assumed



that  $b_1, b_2 : \mathcal{X} \rightarrow (0, 1)$ ). We deduct that the space of all solutions is of dimension  $p = 1$ . Solutions of  $(I - A)c = 0$  are of the form  $c = \beta(1 - d_2, 1 - d_1)^T$ ,  $\beta \in \mathbb{R}$ . So that solution  $g$  to (6.2) is

$$g(x) = \frac{1}{\beta e(x)} ((1 - d_2)f_1(x) + (1 - d_1)f_2(x)), \quad (6.13)$$

where  $\beta = (1 - d_2) \int_{\mathcal{X}} \frac{f_1(x)}{e(x)} m(dx) + (1 - d_1) \int_{\mathcal{X}} \frac{f_2(x)}{e(x)} m(dx)$ .

**Swap to the nearest neighbor.** We consider  $n$  classes,  $\mathcal{X} = \{x_1, \dots, x_n\}$ . If a swap occurs for individual  $i$  with characteristic  $x_i \in \{2, \dots, n - 1\}$ , we suppose that its new one is uniformly chosen between  $x_{i-1}$  and  $x_{i+1}$ , that is  $k_e(x_i, x')m(dx') = \frac{1}{2}\delta_{x_{i-1}}(dx') + \frac{1}{2}\delta_{x_{i+1}}(dx')$ . If the characteristics are  $x_i = x_1$  or  $x_i = x_n$ , the neighbor is chosen with probability one, that is  $k_e(x_1, x')m(dx') = \delta_{x_2}(dx')$  and  $k_e(x_n, x')m(dx') = \delta_{x_{n-1}}(dx')$ . The discrete formulation of Equation (6.2) is given by

$$\forall i \in \{1, \dots, n\}, e(x_i)g(x_i) = \sum_{j \neq i} g(x_j)e(x_j)k_e(x_j, x_i).$$

Define the matrix  $A$  by  $A_{i,j} = e(x_j)k_e(x_j, x_i)\mathbf{1}_{i \neq j} - e(x_i)\mathbf{1}_{i=j}$ . We then want to find the eigenvectors for  $A$  associated with eigenvalue 0. Since characteristics changes can only lead to neighbor characteristics, the matrix  $A$  is tridiagonal. Remark also that upper and lower diagonals of  $A$  have same sign, which shows that it is diagonalizable. Let us write  $A = PDP^{-1}$  the canonic decomposition with  $D$  diagonal. It is then straightforward to show that 0 is an eigenvalue for  $A$  (with multiplicity 1) and that an eigenvector  $Y$  is given by  $Y_1 = \prod_{i \neq 1} e(x_i)$ ,  $Y_n = \prod_{i \neq n} e(x_i)$  and if  $k \in \{2, \dots, n - 1\}$ ,  $Y_k = 2 \prod_{i \neq k} e(x_i)$ . So the solution  $g$  is given by

$$g(x_i) = \frac{Y_i}{\sum_{1 \leq k \leq n} Y_k}. \quad (6.14)$$

## 6.4.2 Numerical illustration

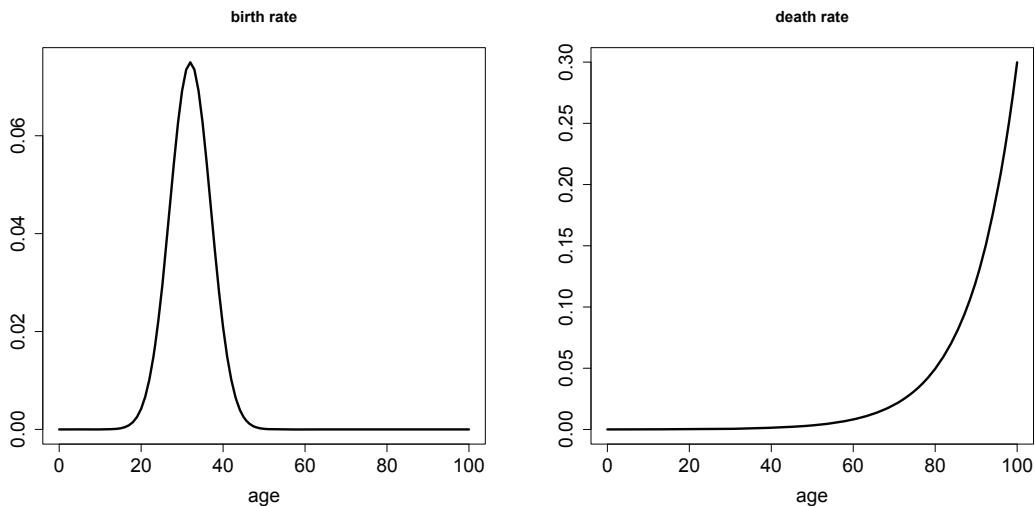
Numerical examples will be performed under the proportional hazard framework:  $d(x, a) := \alpha(x)\bar{d}(a)$  and  $b(x, a) := \beta(x)\bar{b}(a)$ , with  $\bar{d}(a)$  and  $\bar{b}(a)$  some reference death and birth rate. In the macroscopic model, the death rate is

$$\hat{d}(a) = \int_{\mathcal{X}} d(x, a)g(x, a)m(dx) = \hat{\alpha}(a)\bar{d}(a),$$

where  $\hat{\alpha}(a) := \int_{\mathcal{X}} \alpha(x)g(x, a)m(dx)$ . In the same way,  $\hat{b}(a) = \hat{\beta}(a)\bar{b}(a)$ . It is interesting to note that in the general framework where the invariant swap pattern is age-dependent, the equivalent model is not proportional hazard with an age-independent factor anymore.

For the reference death rate, we choose a stylized Gompertz form  $\bar{d}(a) = A_1 e^{B_1 a}$ , and a stylized birth rate  $\bar{b}(a) = A_2 e^{-B_2(a-\bar{a})^2}$ . We choose arbitrary values that mimic the shape and the order of magnitude of human populations, inspired from death and birth data for France in 2008. For the birth data, the aim is to reproduce the shape of female birth rates and divide by two to approximately recover the intensity for an arbitrary individual (male or female). We take  $A_1 = 3, 7.10^{-5}$ ,  $B_1 = 0.09$ ,  $\bar{a} = 32$ ,  $A_2 = 0.075$  and  $B_2 = 0.02$ . The reference rates  $\bar{b}(a)$  and  $\bar{d}(a)$  are plotted in Figure 6.1.

Figure 6.1: Stylized reference birth rate  $\bar{b}(a)$  and death rate  $\bar{d}(a)$  as a function of age.

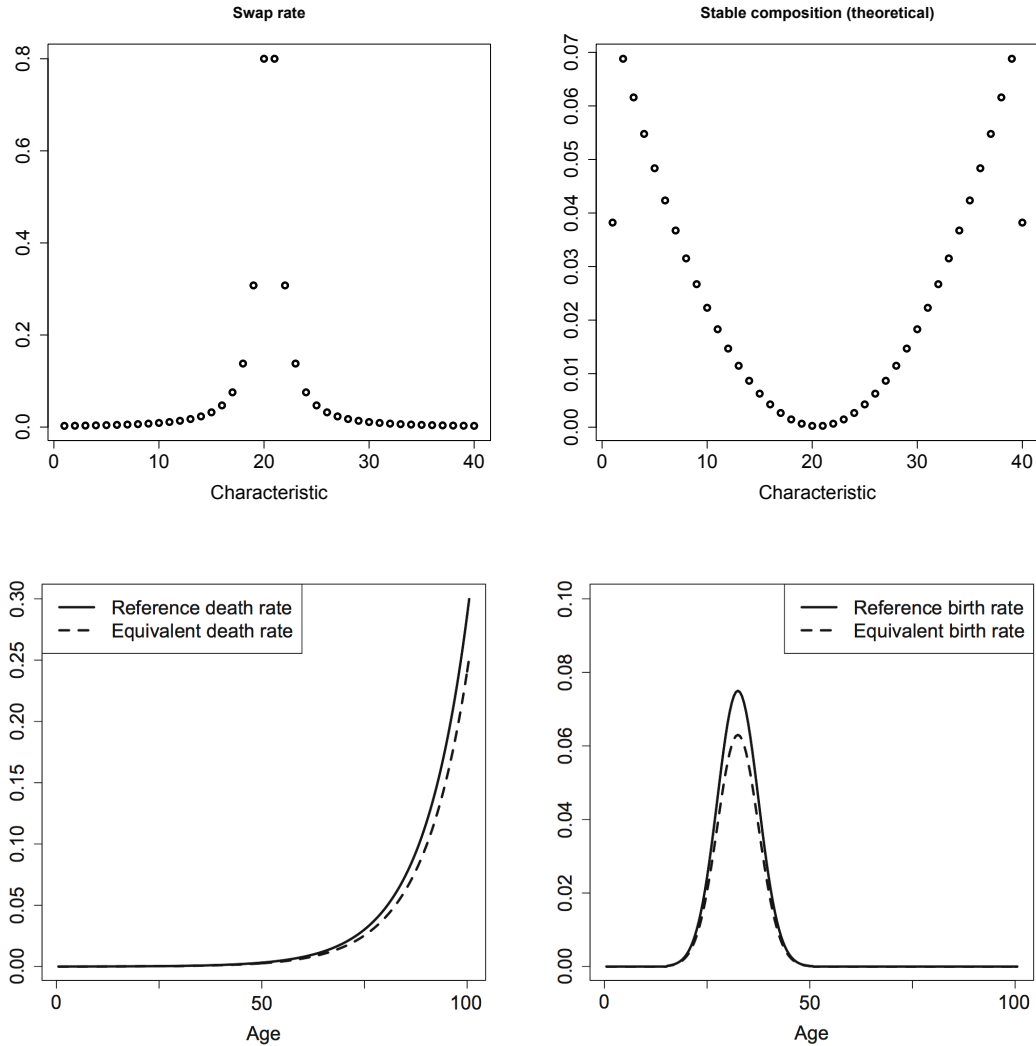


We illustrate numerically the swap to the nearest neighbor. We consider  $n = 40$  discrete classes and set  $\mathcal{X} = \{1, \dots, 40\}$ . We choose the following dynamics : the characteristic  $x$  is such that high value implies living longer but having less children, whereas a low value means having more children: let  $\alpha(x) = \beta(x) = 1 - \frac{x-4.5}{100}$ . The swap rate takes four possible forms, namely  $e(x) = \frac{1}{1-(x-20,5)^2}$ ,  $e(x) = \frac{1}{x}$ ,  $e(x) = x/40$  and  $e(x) = \sqrt{x/40}$ . The swap rate, the theoretical stable composition and the equivalent birth and death rates are given for the four configurations from Figure 6.2 to 6.5. Let us emphasize that in the four experiments, the age and characteristic-dependent birth and death rates are the same, and also that the way the new characteristics are chosen at the time of swap is fixed (namely the swap to the nearest neighbor). We only vary the characteristics-dependent frequency of swap events, driving the way characteristics changes occur at the individual level. Let us first focus on the two upper graphs of each one of the four experiments. These show how the stable composition in terms of characteristics (right-upper graph) is linked to the swap pattern (left-upper graph). In each case, one can notice the interesting link between the shape of the swap rate and that of the stable composition

given by the invariant measure. In particular, as expected, the characteristics for which the swap rate is high are less represented in the population composition, since individuals are forced to escape the corresponding class. Also, let us remark the side effect of the stable composition: this is due to the choice of the swap to the nearest neighbor mechanisms, since characteristics on the side are less likely to be chosen (they have one neighbor instead of two). Now, let us focus on the equivalent death and birth rates. First notice that these are always lower than the reference demographic rates, which are only used here to set a realistic age pattern; this is due to the fact that the proportional parameters  $\alpha(x)$  and  $\beta(x)$  are lower than one for almost all characteristics. One then gets insights on the impact of characteristics changes frequencies when comparing one by one the Figures 6.2 to 6.5. For example, the choice of the swap rate  $e(x) = x/40$  leads to aggregate death and birth rates that are higher than in the three other configurations; this can be explained, as previously detailed, by the fact that small characteristics values, whose birth and death rates are higher, are over-represented. This shows how, with fixed swap rules and characteristic-specific demographic rates, the characteristic-dependent swap frequencies impacts what one observes at the macroscopic level.

Figure 6.2: Swap rate, stable composition and equivalent death and birth rates for

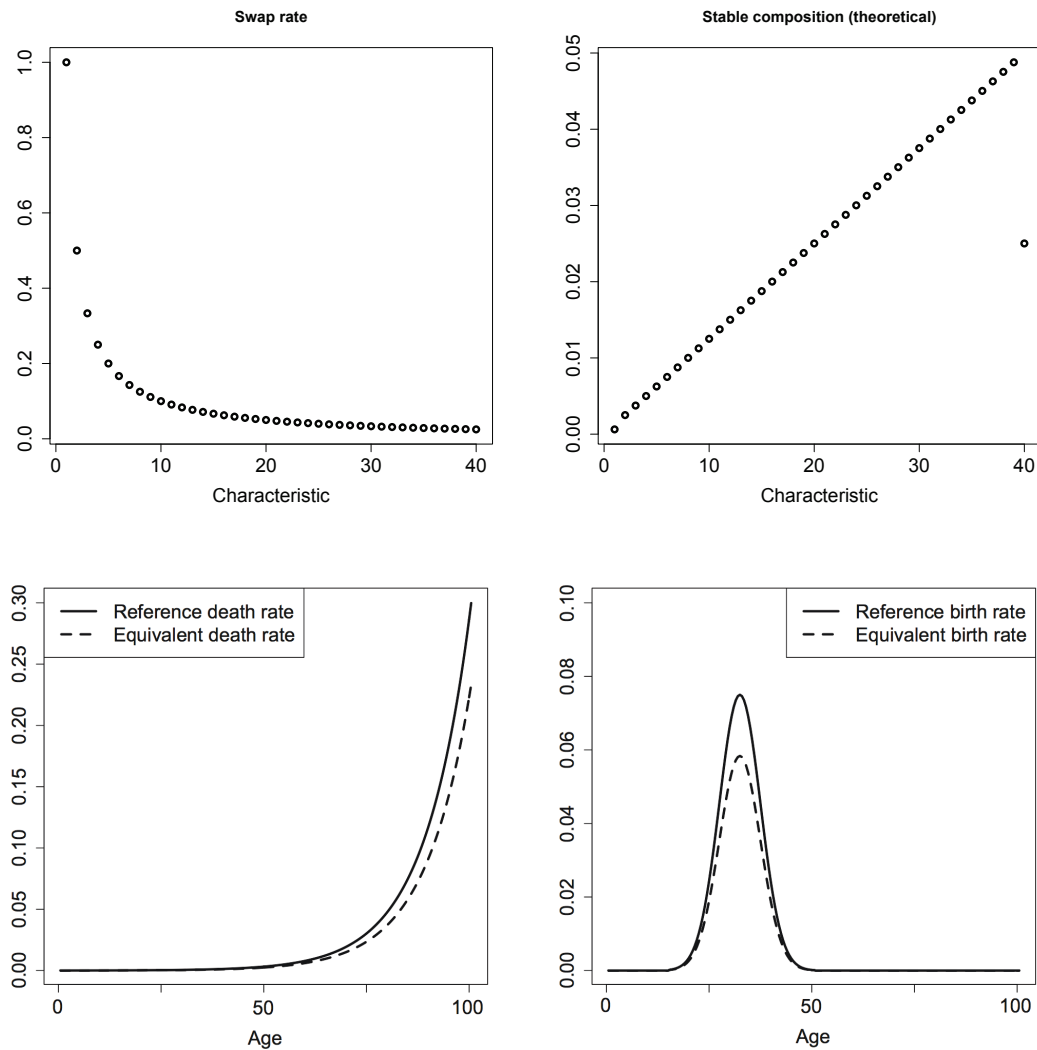
$$e(x) = \frac{1}{1-(x-20,5)^2}$$



## 6.5 Conclusion

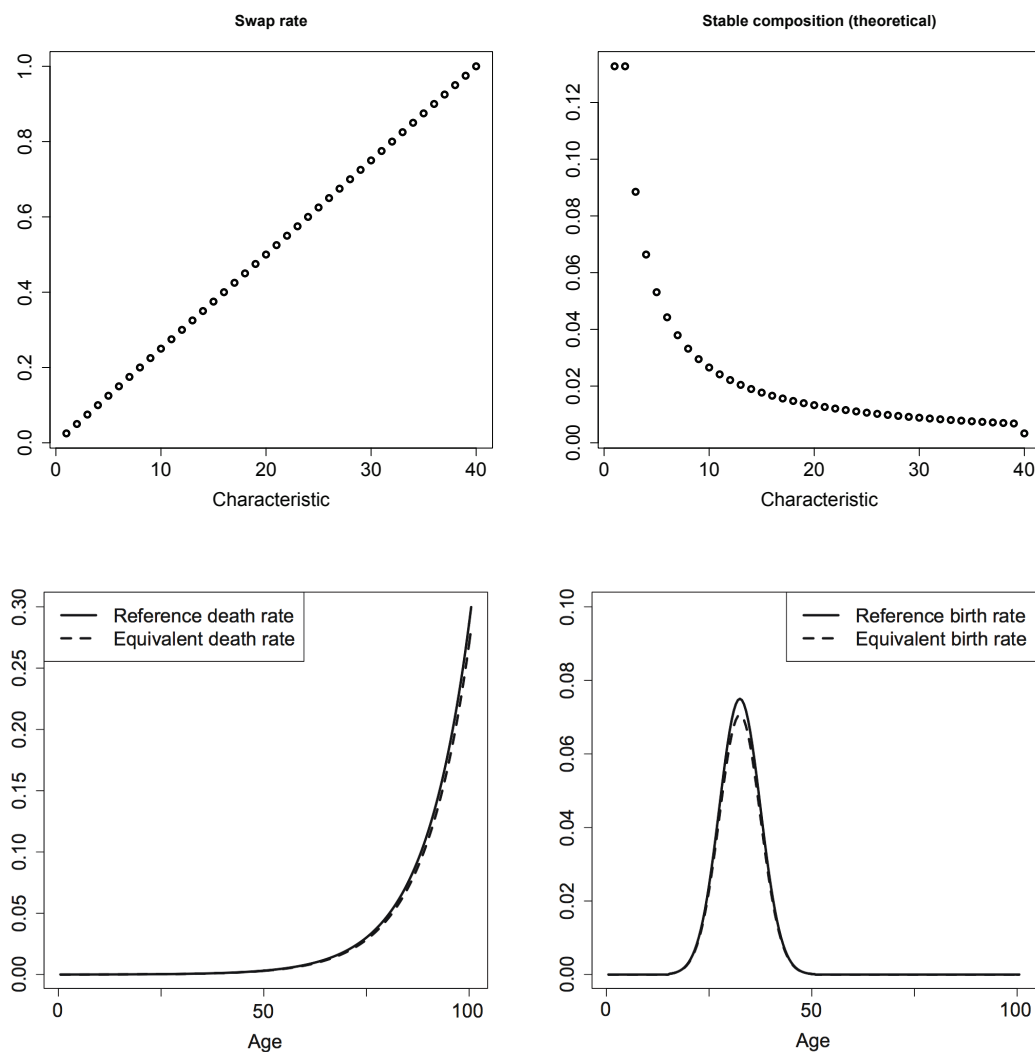
In this paper, we considered the large population limit of an age and characteristics-structured population model evolving according to individual birth, death and fast characteristic changes during life. Both the large population framework and the frequent characteristics changes appear naturally when focusing on the demographic evolution of a human population at the scale of a given country. When rescaling the population process, the classical deterministic transport-renewal McKendrick-Von Foerster equation appears, that describes the time evolution of the age pyramid. This dynamics is driven by age-dependent birth and death rates which are the average of microscopic birth and death rates over the stable characteristics distribution of the invariant swap pattern. In addition, we proved that it is not possible to keep

Figure 6.3: Swap rate, stable composition and equivalent death and birth rates for  $e(x) = \frac{1}{x}$



track of the population structure in terms of characteristics, even in the case of age-independent demographic rates, in other words that the corresponding sequence of measure-valued processes is not tight. This gives a natural and interesting example of a sequence of processes which does not converge. A set of computable invariant distributions for the swap patterns have also been given, namely a simple mixture kernel and a swap to the nearest neighbor mechanism. When considering reference age-dependent birth and death rates that reproduce real demographic data, numerical illustrations have been performed showing the equivalent birth and death rates in a proportional hazard setting. Next steps in this direction could concern (i) the study of the associated central limit theorem, and the link with the speed at which the swap pattern reaches its invariant measure and (ii) the assessment of the numerical gain when considering the equivalent birth and death rates compared to the

Figure 6.4: Swap rate, stable composition and equivalent death and birth rates for  $e(x) = x/40$

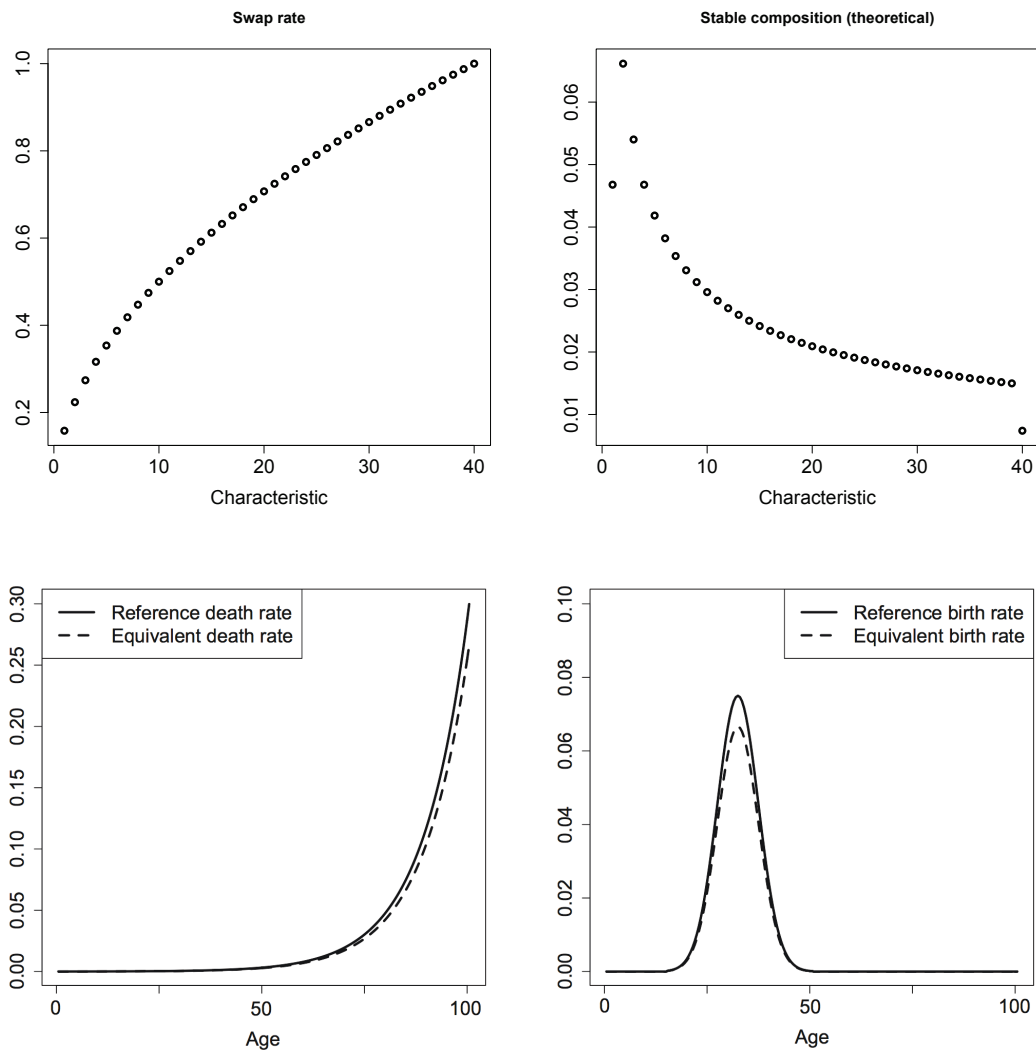


original model. These two aspects are linked in the assessment of the speed at which the original model reaches the average one. They are left for further research.

## Acknowledgements

The author thanks his supervisors Nicole El Karoui and Stéphane Loisel for enlightening discussions, as well as fruitful discussions on particular topics with Mathieu Rosenbaum and Nicolas Gilliers on non-tightness issues, Thibault Bourgeron on steady states for partial differential equations, and also Pierre-Antoine Corre for his help and Casimir Emako for several interesting remarks.

Figure 6.5: Swap rate, stable composition and equivalent death and birth rates for  $e(x) = \sqrt{x/40}$



# Bibliography

- Aalen, Odd O, Søren Johansen. 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 141–150. 118, 268
- Adamopoulos, L. 1975. Some counting and interval properties of the mutually-exciting processes. *Journal of Applied Probability* 78–86. 68, 211
- Aït-Sahalia, Y., Cacho-Diaz J., Laeven R.J.A. 2010. Modeling financial contagion using mutually exciting jump processes. Tech. rep., National Bureau of Economic Research. 69, 211
- Alai, D. H., S. Arnold(-Gaille), M. Sherris. 2015. Modelling Cause-of-Death Mortality and the Impact of Cause-Elimination. *Annals of Actuarial Science*, 9(1), 167-186 . 238, 242, 250
- Aldous, David. 1978. Stopping times and tightness. *The Annals of Probability* 335–340. 309
- Andersen, PK, Ø Borgan, RD Gill, N. Keiding. 1993. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York. 106, 107, 118, 131, 268, 287
- Arnold, S., M. Sherris. 2013. Forecasting mortality trends allowing for cause-of-death mortality dependence. *North American Actuarial Journal*, 17(4), 273-282 . 242
- Arnold, S., M. Sherris. 2015a. Causes-of-death mortality: What do we know on their dependence? *North American Actuarial Journal*, in press . 242, 250
- Arnold, S., M. Sherris. 2015b. International cause-specific mortality rates: New insights from a cointegration analysis. *Working paper, HEC Lausanne* . 242
- Arnold, Séverine, Alexandre Boumezoued, Héloïse Labit Hardy, Nicole El Karoui. 2015. Cause-of-death mortality: What can be learned from population dynamics? *HAL preprint Id: hal-01157900* . 5, 7, 34
- Auger, Pierre, JC Poggiale, E Sánchez. 2012. A review on spatial aggregation methods involving several time scales. *Ecological Complexity* **10** 12–25. 300
- Baddeley, A. 2007. Spatial point processes and their applications. *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004* 1–75. 170
- Bensusan, H. 2010. Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance vie. Ph.D. thesis, École Polytechnique. 36, 239, 300



- Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Bridging the gap from microsimulation practice to population models: a survey. *Work in progress* . 5, 7, 34, 212, 214, 223, 233, 239, 245, 247, 248, 268, 300, 301, 307
- Bezborodov, Viktor. 2015. Markov birth-and-death dynamics of populations. *arXiv preprint arXiv:1502.06783* . 181
- Biffis, Enrico. 2005. Affine processes for dynamic mortality and actuarial valuations. *Insurance: mathematics and economics* **37**(3) 443–468. 59, 192
- Billiard, Sylvain, Régis Ferrière, Sylvie Méléard, Viet Chi Tran. 2014. Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks. *Journal of Mathematical Biology* 1–32. 144, 300
- Billingsley, Patrick. 2009. *Convergence of probability measures*, vol. 493. Wiley-Interscience. 208, 311, 313, 317
- Blanchet, Didier, Sophie Buffeteau, Emmanuelle Crenner, Sylvie Le Minez. 2009. The destinie 2 microsimulation model: overview and illustrative results. *2 nd IMA conference, Ottawa*. 25, 36, 158, 166, 168
- Bloom, David E, David Canning. 2007. Commentary: The Preston Curve 30 years on: still sparking fires. *International Journal of Epidemiology* **36**(3) 498–499. 161
- Bolker, Benjamin, Stephen W Pacala. 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical population biology* **52**(3) 179–197. 49, 197
- Bonnet, C., C. Burricand, C. Colin, A. Flipo, P.R. Mahieu, B. Sédillot. 1999. Le modèle de microsimulation dynamique: Destinie. *Document de travail INSEE* . 240
- Boumezoued, A. 2015a. Macroscopic behavior of heterogenous populations with fast random life histories. *Working Paper* . 6, 7, 35, 268
- Boumezoued, A. 2015b. Population viewpoint on Hawkes processes. *HAL preprint Id: hal-01149752, To appear in Advances in Applied Probability 48.2 (June 2016)* . 5, 7, 34, 175
- Boumezoued, A., N. El Karoui, S. Loisel. 2015. Measuring mortality heterogeneity dynamics with interval-censored data. *Working Paper* . 5, 7, 34, 299
- Bovier, Anton, Shi-Dong Wang. 2013. Trait substitution trees on two time scales analysis. *arXiv preprint arXiv:1304.4640* . 144, 300
- Brémaud, P., L. Massoulié. 1996. Stability of nonlinear Hawkes processes. *The Annals of Probability* 1563–1588. 43, 179, 222
- Brémaud, P., L. Massoulié. 2002. Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability* 205–222. 67, 68, 69, 75, 210, 211, 224, 225, 236
- Bryk, Anthony S, Stephen W Raudenbush. 1992. *Hierarchical linear models: applications and data analysis methods..* Sage Publications, Inc. 162
- Cairns, A. J., D. Blake, K. Dowd, A. Kessler. 2014. Phantoms never die: Living with

- unreliable mortality data. Tech. rep., Herriot Watt University, Edinburgh. 18, 19, 20, 21, 163, 164, 244
- Cairns, A.J.G., D. Blake, K. Dowd. 2006. A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance* **73**(4) 687–718. 15, 16, 129, 158, 163, 284
- Cairns, Andrew JG, David Blake, Kevin Dowd. 2008. Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal* **2008**(2-3) 79–113. 25
- Cairns, Andrew JG, David Blake, Kevin Dowd, Guy D Coughlan, David Epstein, Alen Ong, Igor Balevich. 2009. A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal* **13**(1) 1–35. 55, 56, 158, 188
- Canning, David. 2011. The causes and consequences of demographic transition. *Population studies* **65**(3) 353–361. 159, 164
- Carriere, J. F. 1994. Dependent decrement theory. *Transactions of the Society of Actuaries* **46** 45–74. 89, 102, 238, 242
- Champagnat, Nicolas, Régis Ferrière, Sylvie Méléard. 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology* **69**(3) 297–321. 36, 49, 52, 144, 158, 197, 205, 239, 300, 301
- Chiang, C. L. 1968. Introduction to stochastic processes in biostatistics. *John Wiley and Sons, New York* . 238, 242
- Çınlar, E. 2011. *Probability and stochastics*, vol. 261. Springer. 42, 43, 170, 222
- Cléménçon, Stéphan, Viet Chi Tran, Hector de Arazoza. 2008. A stochastic sir model with contact-tracing: large population limits and statistical inference. *Journal of Biological Dynamics* **2**(4) 392–414. 52, 205
- Cohen, Joel E. 2003. Human population: the next half century. *science* **302**(5648) 1172–1175. 164
- Commenges, Daniel, Anne Gégout-Petit. 2007. Likelihood for generally coarsened observations from multistate or counting process models. *Scandinavian journal of statistics* **34**(2) 432–450. 120, 270
- Commenges, Daniel, Pierre Joly, Anne Gégout-Petit, Benoit Liqueur. 2007. Choice between semi-parametric estimators of Markov and non-Markov multi-state models from coarsened observations. *Scandinavian Journal of Statistics* **34**(1) 33–52. 120, 121, 124, 270, 271, 279
- Costa, Manon. 2015. A piecewise deterministic model for prey-predator communities. *arXiv preprint arXiv:1503.02492* . 144, 300
- Costa, Manon, Céline Hauzy, Nicolas Loeuille, Sylvie Méléard. 2015. Stochastic eco-evolutionary model of a prey-predator community. *Journal of mathematical biology* 1–50. 144, 300
- Cutler, David, Grant Miller. 2005. The role of public health improvements in health advances: the twentieth-century United States. *Demography* **42**(1) 1–22. 161

- Cutler, David M, Angus S Deaton, Adriana Lleras-Muney. 2006. The determinants of mortality. Tech. rep., National Bureau of Economic Research. 160
- Czado, Claudia, Florian Rudolph. 2002. Application of survival analysis methods to long-term care insurance. *Insurance: Mathematics and Economics* **31**(3) 395–413. 119, 132, 269, 287
- Da Fonseca, J., R. Zaatour. 2014. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets* **34**(6) 548–579. 69, 211
- Daley, DJ, D. Vere-Jones. 2003. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods of Probability and its Applications*. Springer, New York. 224
- Daley, DJ, D. Vere-Jones. 2008. An introduction to the theory of point processes. Volume II: General theory and structure. probability and its applications. 43, 222
- Dassios, A., H. Zhao. 2011. A dynamic contagion process. *Advances in applied probability* **43**(3) 814–846. 67, 68, 69, 73, 75, 78, 210, 211, 217, 225, 228
- Dassios, Angelos, Hongbiao Zhao, et al. 2013. Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability* **18**(62). 79
- Delattre, S., N. Fournier, M. Hoffmann. 2014. High dimensional Hawkes processes. *arXiv preprint arXiv:1403.5764* . 43, 179, 222, 223
- Delwarde, A., M. Denuit. 2006. *Construction de tables de mortalité périodiques et prospectives*. Economica. 18, 250
- Diaz, Belinda Aparicio, Thomas Fent, Alexia Prskawetz, Laura Bernardi. 2007. Projections of age-specific fertility rates through an agent-based model of social interaction. *Work session on demographic projections* 49. 165
- Dieckmann, Ulf, Richard Law. 2000. Relaxation projections and the method of moments. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity (U Dieckmann, R. Law, JAJ Metz, editors)*. Cambridge University Press, Cambridge 412–455. 49, 197
- Dimitrova, D. S., S. Haberman, V. K. Kaishev. 2013. Dependent competing risks: Cause elimination and its impact on survival. *Insurance: Mathematics and Economics* . 92, 102, 238, 242, 243
- Donnelly, PD. 2010. Explaining the Glasgow effect: could adverse childhood experiences play a role? *Public Health* **124**(9) 498–499. 169
- Doumic, Marie, Marc Hoffmann, Nathalie Krell, Lydia Robert, et al. 2015. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli* **21**(3) 1760–1799. 103
- Doumic, Marie, Marc Hoffmann, Patricia Reynaud-Bouret, Vincent Rivoirard. 2012. Non-parametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis* **50**(2) 925–950. 103

- Duée, M. 2005. La modélisation des comportements démographiques dans le modèle de microsimulation Destinée. *Document de travail INSEE/DESE G* . 240
- Elandt-Johnson, R. C. 1976. Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. *Scandinavian Actuarial Journal* (1) 37–51. 92, 241, 243
- Engen, S, O Bakke, A. Islam. 1998. Demographic and environmental stochasticity: concepts and definitions. *Biometrics* **54** 840–846. 25
- Errais, E., K. Giesecke, L. Goldberg. 2010. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics* **1**(1) 642–665. 69, 211
- Ferriere, R., V.C. Tran. 2009. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM: Proceedings*, vol. 27. 289–310. 36, 158, 239, 300, 301
- Fogel, Robert W. 1997. New findings on secular trends in nutrition and mortality: some implications for population theory. . 160
- Foucher, Yohann, Magali Giral, Jean-Paul Soulillou, Jean-Pierre Daures. 2007. A semi-Markov model for multistate and interval-censored data with multiple terminal events. application in renal transplantation. *Statistics in medicine* **26**(30) 5381–5393. 120, 121, 124, 270, 271, 279
- Fougère, Maxime, Marcel Mérette. 1999. Population ageing and economic growth in seven OECD countries. *Economic Modelling* **16**(3) 411–427. 239
- Fournier, N., S. Méléard. 2004. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Annals of applied probability* **14**(4) 1880–1919. 36, 50, 52, 144, 158, 179, 181, 197, 205, 207, 223, 233, 239, 247, 300, 301
- Fries, James F. 1980. Aging, natural death, and the compression of morbidity. *New England journal of medicine* **303**(3) 130–135. 18, 130, 162, 285
- Frydman, Halina. 1995. Nonparametric estimation of a Markov ‘illness-death’ process from interval-censored observations, with application to diabetes survival data. *Biometrika* **82**(4) 773–789. 120, 270
- Frydman, Halina, Michael Szarek. 2009. Nonparametric estimation in a Markov “illness–death” process from interval censored observations with missing intermediate transition status. *Biometrics* **65**(1) 143–151. 120, 270
- Garcia, Nancy L, Thomas G Kurtz. 2006. Spatial birth and death processes as solutions of stochastic equations. *Alea* **1** 281–303. 179, 224, 236
- Garcia, Nancy L, Thomas G Kurtz. 2008. Spatial point processes and the projection method. *In and Out of Equilibrium 2*. Springer, 271–298. 179
- Garcia, Nancy Lopes. 1995. Birth and death processes as projections of higher-dimensional Poisson processes. *Advances in applied probability* 911–930. 179
- Gaüzère, Franck, Daniel Commenges, Pascale Barberger-Gateau, Luc Letenneur, Jean-François Dartigues. 1999. Maladie et dépendance: description des évolutions par des modèles multi-états. *Population (french edition)* 205–222. 119, 269

- Gavrilov, LA, NS Gavrilova. 1991. The biology of life span: a quantitative approach. . 18, 130, 162, 285
- Giroi, F., G. King. 2006. Demographic forecasting. *Cambridge University Press, Cambridge* . 242
- Goldstein, Harvey. 2011. *Multilevel statistical models*, vol. 922. John Wiley & Sons. 162
- Gompertz, Benjamin. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London* 513–583. 11, 123, 163, 279
- Gray, Linsay, Alastair H Leyland. 2009. A multilevel analysis of diet and socio-economic status in Scotland: investigating the ‘Glasgow effect’. *Public health nutrition* **12**(09) 1351–1358. 169
- Grigelionis, B. 1971. The representation of integer-valued random measures as stochastic integrals over the Poisson measure. *Litovsk. Mat. Sb.* **11** 93–108. 43, 178, 222
- Guilmoto, C. Z. 2012. Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100. *Demography* **49**(1) 77–100. 52, 205, 244
- Gupta, Ankit, JAJ Metz, Viet Chi Tran. 2014. A new proof for the convergence of an individual based model to the trait substitution sequence. *Acta Applicandae Mathematicae* **131**(1) 1–27. 144, 149, 298, 300, 312
- Hardiman, S., N. Bercot, J.-P. Bouchaud. 2013. Critical reflexivity in financial markets: a Hawkes process analysis. *arXiv preprint arXiv:1302.1405* . 216
- Harris, T.E. 1963. The theory of branching processes. Tech. rep. 214
- Hautphenne, S., G. Latouche. 2012. The Markovian binary tree applied to demography. *Journal of mathematical biology* **64**(7) 1109–1135. 240
- Hawkes, A.G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**(1) 83–90. 26, 67, 68, 209, 210, 211
- Hawkes, A.G., D. Oakes. 1974. A cluster process representation of a self-exciting process. *Journal of Applied Probability* 493–503. 26, 68, 69, 210, 211, 213, 224
- Helms, Florian, Claudia Czado, Susanne Gschlößl. 2005. Calculation of ltc premiums based on direct estimates of transition probabilities. *ASTIN Bulletin* **35** 455–469. 119, 269
- HFD. 2014. The human fertility database. max planck institute for demographic research (germany) and vienna institute of demography (austria). URL [www.humanfertility.org](http://www.humanfertility.org). 21
- HMD. 2014. The human mortality database: French data. URL <http://www.mortality.org/cgi-bin/hmd/country.php?cntr=FRA&level=2>. 12, 163, 187
- Hock, Heinrich, David N Weil. 2012. On the dynamics of the age structure, dependency, and consumption. *Journal of population economics* **25**(3) 1019–1043. 239
- Hoffmann, Marc, Adélaïde Olivier. 2014. Nonparametric estimation of the division rate of an age dependent branching process. *arXiv preprint arXiv:1412.5936* . 103
- Hyndman, R. J., H. Booth. 2008. Stochastic population forecasts using functional data

- models for mortality, fertility and migration. *International Journal of Forecasting* **24**(3) 323–342. 240
- Iannelli, M., M. Martcheva, F. A. Milner. 2005. *Gender-structured population modeling: mathematical methods, numerics, and simulations*, vol. 31. Siam. 93, 239, 244, 245, 264
- INSEE. 2011. Sources et méthodes: Les indicateurs démographiques. URL [http://www.insee.fr/fr/methodes/sources/pdf/Indicateurs\\_demographiques.pdf](http://www.insee.fr/fr/methodes/sources/pdf/Indicateurs_demographiques.pdf). 251
- INSEE. 2014. Fécondité selon l'âge détaillé de la mère. URL [http://www.insee.fr/fr/themes/tableau.asp?reg\\_id=0&ref\\_id=bilandemo2](http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=bilandemo2).
- Jackson, Christopher H, et al. 2011. Multi-state models for panel data: the msm package for r. *Journal of Statistical Software* **38**(8) 1–29. 121
- Jacod, Jean. 1979. *Calcul stochastique et problemes de martingales*. Springer. 43
- Jensen, Jens Ledet, Jesper Møller. 1991. Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability* 445–461. 172
- Joffe, Anatole, Michel Métivier. 1986. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability* 20–65. 309
- Joly, Pierre, Daniel Commenges, Catherine Helmer, Luc Letenneur. 2002. A penalized likelihood approach for an illness–death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3**(3) 433–443. 118, 120, 121, 124, 268, 270, 271, 279
- Jourdain, Benjamin, Sylvie Méléard, Wojbor A Woyczynski. 2012. Lévy flights in evolutionary ecology. *Journal of mathematical biology* **65**(4) 677–707. 311, 314
- Jovanović, S., J. Hertz, S. Rotter. 2014. Cumulants of Hawkes point processes. *arXiv preprint arXiv:1409.5353* . 69, 211
- Kaishev, V. K., D. S. Dimitrova, S. Haberman. 2007. Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics* . 238, 242
- Kalbfleisch, JD, Jerald F Lawless. 1985. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**(392) 863–871. 120, 270
- Kang, Minhee, Stephen W Lagakos. 2007. Statistical methods for panel data from a semi-Markov process, with application to hpv. *Biostatistics* **8**(2) 252–264. 120, 270
- Keiding, Niels. 1990. Statistical inference in the lexis diagram. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **332**(1627) 487–509. 18
- Kendall, David G. 1949. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)* **11**(2) 230–282. 52, 205
- Kerstan, J. 1964. Teilprozesse Poissonscher prozesse. *Trans. Third Prague Conf. Informa-*

- tion Theory, *Statist. Decision Functions, Random Processes (Liblice, 1962)* 377–403. 43, 178, 222
- Keyfitz, N. 1968. Changing vital rates and age distributions. *Population Studies* **22**(2) 235–251. 100, 258
- Keyfitz, N. 1972. The mathematics of sex and marriage. *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, vol. 4. University of California Press Berkeley, 89–108. 239, 244
- Kuehn, Christian. 2015. Moment closure-a brief review. *arXiv preprint arXiv:1505.02190* . 49, 197
- Kurtz, TG. 1989. Stochastic processes as projections of Poisson random measures. special invited paper at ims meeting, washington, dc. *DC Unpublished* . 179
- Kurtz, Thomas G. 1992. Averaging for martingale problems and stochastic approximation. *Applied Stochastic Analysis*. Springer, 186–209. 149, 306, 312, 313
- Kwon, Hyuk-Sung, Bruce L Jones. 2006. The impact of the determinants of mortality on life insurance and annuities. *Insurance: Mathematics and Economics* **38**(2) 271–288. 119, 269
- Kwon, Hyuk-Sung, Bruce L Jones. 2008. Applications of a multi-state risk factor/mortality model in life insurance. *Insurance: Mathematics and Economics* **43**(3) 394–402. 119, 269
- Lee, R.D., L.R. Carter. 1992. Modeling and forecasting US mortality. *Journal of the American Statistical Association* **87**(419) 659–671. 158
- Levantesi, Susanna, Massimiliano Menzietti. 2012. Managing longevity and disability risks in life annuities with long term care. *Insurance: Mathematics and Economics* **50**(3) 391–401. 119, 269
- Lewis, P. A. W., Shedler. 1978. Simulation of nonhomogeneous Poisson processes by thinning. Tech. rep., Naval Postgraduate School. 43, 178, 222
- Li, J., C. O’Donoghue. 2013. A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* **6**(2) 3–55. 240
- Manton, K. G. 1986a. Past and future life expectancy increases at later ages: Their implications for the linkage of morbidity, disability, and mortality. *Journal of Gerontology*, **41**(5), 672-681 . 242
- Manton, K. G., E. Stallard, J. W. Vaupel. 1986b. Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, **81**(395), 635- 644 . 242
- Massoulié, L. 1998. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic processes and their applications* **75**(1) 1–30. 43, 47, 179, 222, 233
- Mayhew, Les, David Smith. 2013. A new method of projecting populations based on trends in life expectancy and survival. *Population studies* **67**(2) 157–170. 163

- McKendrick, A.G. 1926. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.* **54** 98–130. 21, 33, 143, 239, 245, 299
- McKeown, Thomas. 1976. The modern rise of population. *London, England: Edward Arnold* . 160
- Méléard, Sylvie, Sylvie Roelly. 1993. Sur les convergences étroite ou vague de processus à valeurs mesures. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* **317**(8) 785–788. 310
- Méléard, Sylvie, Viet Chi Tran. 2012. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications* **122**(1) 250–276. 45, 50, 144, 148, 149, 197, 298, 300, 305, 306, 308, 311, 312
- Metz, Johan AJ, Stefan AH Geritz, Géza Meszéna, Frans JA Jacobs, Joost S Van Heerwaarden, et al. 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *Stochastic and spatial structures of dynamical systems* **45** 183–231. 197
- Mirkin, Barry, Mary Beth Weinberger. 2001. The demography of population ageing. *Population Bulletin of the United Nations* **42**(43) 37–53. 239
- Morand, E., L. Toulemon, Pennek S., Baggio R., F. Billari. 2010. Demographic modelling: the state of the art. *SustainCity Working Paper, 2.1a, Ined, Paris.*. 36, 158, 165
- Oakes, D. 1975. The markovian self-exciting process. *Journal of Applied Probability* 69–77. 26, 69, 71, 211, 214
- Oelschläger, Karl. 1990. Limit theorems for age-structured populations. *The Annals of Probability* 290–318. 50, 197, 248
- Ogata, Yoshihiko. 1981. On Lewis' simulation method for point processes. *Information Theory, IEEE Transactions on* **27**(1) 23–31. 43, 178, 222
- Orcutt, Guy H. 1957. A new type of socio-economic system. *The Review of Economics and Statistics* **39**(2) 116–123. 36, 158, 165
- Plank, Michael J, Richard Law. 2014. Spatial point processes and moment dynamics in the life sciences: a parsimonious derivation and some extensions. *Bulletin of mathematical biology* **77**(4) 586–613. 197
- Plat, Richard. 2009. On stochastic mortality modeling. *Insurance: Mathematics and Economics* **45**(3) 393–404. 163
- Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, VT. Farewell, NE. Breslow. 1978. The analysis of failure times in the presence of competing risks. *Biometrics* 541–554. 238, 242
- Preston, Samuel H. 1975. The changing relation between mortality and level of economic development. *Population studies* **29**(2) 231–248. 161
- Putter, H, M Fiocco, RB Geskus. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* **26**(11) 2389–2430. 238, 242
- Rambaldi, M., P. Pennesi, F. Lillo. 2014. Modeling FX market activity around macroeco-



- conomic news: a Hawkes process approach. *arXiv preprint arXiv:1405.6047* . 67, 75, 210, 225
- Richards, SJ. 2008. Applying survival models to pensioner mortality data. *British Actuarial Journal* **14**(02) 257–303. 163, 164
- Roelly-Coppoletta, Sylvie. 1986. A criterion of convergence of measure-valued processes: application to measure branching processes. *Stochastics: An International Journal of Probability and Stochastic Processes* **17**(1-2) 43–65. 309
- Rooney, Cleo, Clare Griffiths, Lois Cook. 2002. The implementation of ICD-10 for cause of death coding-some preliminary results from the bridge coding study. *Health Statistics Quarterly* (13) 31–41. 102
- Rosen, M. 2006. Forecasting life expectancy and mortality in sweden, some comments on methodological problems and potential approaches. technical report 4. Tech. rep., Social Insurance Studies from the Swedish Social Insurance. 242
- Saichev, A., D. Sornette. 2011. Generating functions and stability study of multivariate self-excited epidemic processes. *The European Physical Journal B-Condensed Matter and Complex Systems* **83**(2) 271–282. 69, 211
- Schoen, R. 1983. Measuring the tightness of a marriage squeeze. *Demography* **20**(1) 61–78. 52, 205, 244
- Silverman, Eric, Jakub Bijak, Jason Noble. 2011. Feeding the beast: can computational demographic models free us from the tyranny of data? . 158, 163
- Spielauer, Martin. 2011. What is social science microsimulation? *Social Science Computer Review* **29**(1) 9–20. 165, 167
- Strehler, Bernard L, Albert S Mildvan. 1960. General theory of mortality and aging. *Science* **132**(3418) 14–21. 18, 129, 130, 162, 163, 284, 285
- Strulik, Holger, Sebastian Vollmer. 2013. Long-run trends of human aging and longevity. *Journal of Population Economics* **26**(4) 1303–1323. 18, 129, 130, 163, 284, 285
- Subramanian, SV, Kelvin Jones, Craig Duncan. 2003. *Multilevel methods for public health research*. Neighborhoods and health. New York: Oxford University Press. 162
- Tesfatsion, Leigh. 2002. Agent-based computational economics: Growing economies from the bottom up. *Artificial life* **8**(1) 55–82. 36, 158, 165
- Titman, Andrew C, Linda D Sharples. 2010. Semi-Markov models with Phase-Type sojourn distributions. *Biometrics* **66**(3) 742–752. 120, 270
- Touraine, Celia. 2013. Modèles illness-death pour données censurées par intervalle: application à l'étude de la démence. Ph.D. thesis, Bordeaux 2. 120, 121, 270
- Touraine, Céilia, Catherine Helmer, Pierre Joly. 2013. Predictions in an illness-death model. *Statistical methods in medical research* 0962280213489234. 121, 124, 270, 271, 279
- Tran, V.C. 2006. Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques . 50, 197, 200, 207, 214, 248, 304, 309

- Tran, V.C. 2008. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics* **12**(1) 345–386. 37, 50, 52, 197, 198, 205, 212, 214, 223, 233, 245, 300, 301, 307
- Tsai, S. P., E. S. Lee, R. J. Hardy. 1978. The effect of a reduction in leading causes of death: Potential gains in life expectancy. *American Journal of Public Health* . 238, 242
- Tsiatis, A. 1975. A non-identifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* . 88, 242
- Van Imhoff, Evert, Wendy Post. 1998. Microsimulation methods for population projection. *Population* **10**(1) 97–136. 167, 168, 196
- Vaupel, J. W., A. I. Yashin. 1983. The deviant dynamics of death in heterogeneous populations. technical report rr-83-001. Tech. rep., International Institute for Applied Systems Analysis (IIASA). 242
- Volinsky, Chris T, Adrian E Raftery. 2000. Bayesian information criterion for censored survival models. *Biometrics* **56**(1) 256–262. 283
- Von Foerster, H. 1959. *The Kinetics of Cellular Proliferation*. Grune & Stratton. 21, 33, 143, 239, 245, 300
- Von Neumann, John. 1951. Various techniques used in connection with random digits . 47
- Wang, Frank JS. 1977. A central limit theorem for age-and density-dependent population processes. *Stochastic Processes and their Applications* **5**(2) 173–193. 50, 197, 248
- Wei, Shaoceng. 2015. Multi-state models for interval censored data with competing risk . 120, 270
- Wheatley, S., V. Filimonov, D. Sornette. 2014. Estimation of the Hawkes process with renewal immigration using the EM algorithm. *Swiss Finance Institute Research Paper* (14-53). 67, 75, 210, 225
- WHO. 2014. The World Health Organization: the WHO Mortality Database. URL [http://www.who.int/healthinfo/statistics/mortality\\_rawdata/en/](http://www.who.int/healthinfo/statistics/mortality_rawdata/en/).
- Willekens, Frans. 2005. Biographic forecasting: bridging the micro-macro gap in population forecasting. *New Zealand population review* **31**(1) 77–124. 158, 166
- Willets, RC. 2004. *The cohort effect: insights and explanations*. Cambridge Univ Press. 29, 30, 55, 57, 164, 188, 189
- Wilmoth, J. R. 1995. Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* . 238, 242
- Wilmoth, John R, Shiro Horiuchi. 1999. Rectangularization revisited: Variability of age at death within human populations. *Demography* **36**(4) 475–495. 18, 130, 162, 285
- Zemyan, Stephen M. 2012. *The Classical Theory of Integral Equations*. Springer. 303, 319