



HAL
open science

Non Parametric Estimation of General Population Process Parameters.

Paulien Jeunesse

► **To cite this version:**

Paulien Jeunesse. Non Parametric Estimation of General Population Process Parameters.. Statistics [math.ST]. Université Paris sciences et lettres, 2019. English. NNT : 2019PSLED013 . tel-03222387

HAL Id: tel-03222387

<https://theses.hal.science/tel-03222387>

Submitted on 10 May 2021

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 de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Université Paris-Dauphine

Estimation non paramétrique du taux de mort dans un modèle de population générale : Théorie et applications.

École doctorale n°543

ÉCOLE DOCTORALE DE DAUPHINE

Spécialité SCIENCES

Soutenue par **Paulien Jeunesse**
le 8 janvier 2019

Dirigée par **Marc Hoffmann**

COMPOSITION DU JURY :

M. Marc Hoffmann
Professeur, Université Paris-Dauphine
Directeur de thèse

Mme Agathe Guilloux
Professeur, Université d'Évry Val d'Essonne
Rapporteur

M. Nicolas Champagnat
Directeur de recherche INRIA, Université de Lorraine
Rapporteur

Mme Eva Löcherbach
Professeur, Université de Cergy-Pontoise
Présidente du jury

M. Viet-Chi Tran
Maître de Conférences, Université de Lille
Membre du jury

M. Stéphane Mischler
Professeur, Université Paris-Dauphine
Membre du jury

M. Vincent Rivoirard
Professeur, Université Paris-Dauphine
Membre du jury

L'étude du taux de mortalité dans des modèles de population humaine ou en biologie est le coeur de ce travail. Cette thèse se situe à la frontière de la statistique des processus, de la statistique nonparamétrique et de l'analyse.

Dans une première partie, centrée sur une problématique actuarielle, un algorithme est proposé pour estimer les tables de mortalité, utiles en assurance. Cet algorithme se base sur un modèle déterministe de population. Ces nouvelles estimations améliorent les résultats actuels en prenant en compte la dynamique globale de la population. Ainsi les naissances sont incorporées dans le modèle pour calculer le taux de mort. De plus, ces estimations sont mises en lien avec les travaux précédents, assurant ainsi la continuité théorique de notre travail.

Dans une deuxième partie, nous nous intéressons à l'estimation du taux de mortalité dans un modèle stochastique de population. Cela nous pousse à utiliser des arguments propres à la statistique des processus et à la statistique nonparamétrique. On trouve alors des estimateurs non-paramétriques adaptatifs dans un cadre anisotrope pour la mortalité et la densité de population, ainsi que des inégalités de concentration non asymptotiques quantifiant la distance entre le modèle stochastique et le modèle déterministe limite utilisé dans la première partie. On montre que ces estimateurs restent optimaux dans un modèle où le taux de mort dépend d'interactions, comme dans le cas de la population logistique.

Dans une troisième partie, on considère la réalisation d'un test pour détecter la présence d'interactions dans le taux de mortalité. Ce test permet en réalité de juger de la dépendance temporelle de ce taux. Sous une hypothèse, on montre alors qu'il est possible de détecter la présence d'interactions. Un algorithme pratique est proposé pour réaliser ce test.

Mots-Clés

Statistique nonparamétrique, Dépendance, Statistique des processus aléatoires, Inégalités de concentration, Equation aux dérivées partielles, Adaptativité, Inégalité oracle, Sélection de fenêtre, Estimateurs à noyaux, vitesse de convergence minimax, test nonparamétrique.

In this thesis, we study the mortality rate in different population models to apply our results to demography or biology. The mathematical framework includes statistics of process, nonparametric estimations and analysis.

In a first part, an algorithm is proposed to estimate the mortality tables. This problematic comes from actuarial science and the aim is to apply our results in the insurance field. This algorithm is founded on a deterministic population model. The new estimates we gets improve the actual results. Its advantage is to take into account the global population dynamics. Thanks to that, births are used in our model to compute the mortality rate. Finally these estimations are linked with the precedent works. This is a point of great importance in the field of actuarial science.

In a second part, we are interested in the estimation of the mortality rate in a stochastic population model. We need to use the tools coming from nonparametric estimations and statistics of process to do so. Indeed, the mortality rate is a function of two parameters, the time and the age. We propose minimax optimal and adaptive estimators for the mortality and the population density. We also demonstrate some non asymptotic concentration inequalities. These inequalities quantify the deviation between the stochastic process and its deterministic limit we used in the first part. We prove that our estimators are still optimal in a model where the mortality is influenced by interactions. This is for example the case for the logistic population.

In a third part, we consider the testing problem to detect the existence of interactions. This test is in fact designed to detect the time dependance of the mortality rate. Under the assumption the time dependance in the mortality rate comes only from the interactions, we can detect the presence of interactions. Finally we propose an algorithm to do this test.

Keywords

Nonparametric statistics, Dependence, Statistics of process, Concentration inequalities, Partial differential equation, Adaptive estimation, Oracle inequality, Bandwidth selection, Kernel estimation, minimax rate of convergence, nonparametric testing.

REMERCIEMENTS

Je remercie tout d'abord Marc Hoffmann qui m'a encadré tout au long de ces trois années. Marc, ce travail n'aurait pu être ce qu'il est sans votre soutien et vos conseils. Vous avez été un excellent encadrant, à la fois sur le plan technique et surtout le plan humain. Vous m'avez poussé dans mes retranchements en maintenant une exigence vis à vis de mes multiples idées plus que nécessaires pour réussir à terminer cette thèse. Cela m'a permis d'acquérir une rigueur. Cet encadrement a aussi renforcé mon goût pour la statistique des processus et la statistique non paramétrique. Pour tout cela je vous remercie.

Je suis gré à Agathe Guilloux et Nicolas Champagnat d'avoir accepté d'assurer la charge de rapporter ma thèse. Je suis touché de l'intérêt que vous avez porté à mes travaux. Je vous remercie pour vos lectures attentives et vos remarques pertinentes qui m'ont permis d'améliorer mon manuscrit.

Je remercie Eva Löcherbach, Stéphane Mischler, Vincent Rivoirard et Viet-Chi Tran d'avoir accepté de prendre part au jury de ma thèse.

Le laboratoire du CEREMADE fût un environnement scientifique propice à cette thèse. J'ai pu y suivre des cours de recherche ainsi que des exposés enrichissants. Je remercie tous ses membres pour leur bienveillance à mon égard. Ce fût un plaisir d'enseigner les statistiques à Dauphine pendant trois années. Merci à ceux qui ont veillé pendant cette période à la bonne organisation des TDs. Je remercie également toutes les personnes qui ont accompagné mes différentes démarches administratives.

Sur un plan plus personnel, je souhaite remercier mes amis qui ont été plus que présents pour moi, durant et avant ces trois années, et avec qui j'ai passé des moments inoubliables, hors du monde des mathématiques. Je souhaite remercier ma famille, qui m'a toujours soutenu. Je tiens à remercier mon frère qui a nourri d'intéressantes réflexions pour cette thèse et autres.

Enfin, merci à Ruozhen pour tout.

TABLE DES MATIÈRES

Introduction	1
1 Cadre mathématique et statistique	1
1.1 Estimateurs à noyaux	1
1.2 Vitesse minimax	2
1.3 Estimateur adaptatif	3
1.4 Modèle de population	4
2 Première partie : Utilisation du modèle déterministe pour estimer le taux de mortalité dans un cadre discret.	7
2.1 Motivation	7
2.2 Résultats de la première contribution	9
3 Deuxième partie : Estimation nonparamétrique du taux de mort.	11
3.1 Résultats du chapitre 2	12
3.2 Résultats du chapitre 3	15
4 Troisième Partie : Test de présence d'interactions	17
5 Perspectives	19
5.1 Enrichissement dans le modèle discret	19
5.2 Enrichissement dans le modèle stochastique	20
5.3 Estimation de l'interaction	20
6 Composition de la thèse	21
1 A new inference strategy for general population mortality tables	22
1 Introduction	22
2 Model and inference strategy	24
2.1 Non-homogeneous birth-death dynamics	24
2.2 Observables in the Lexis diagram	24
2.3 Death rate inference	26
2.4 Main result	28
2.5 Proof of Proposition 1.2	30
2.6 Discussion	31
3 Numerical results	33
4 Concluding remarks	39

2	Nonparametric inference of age-structured models in a large population limit	41
1	Introduction	41
1.1	Setting	41
1.2	Motivation	42
1.3	Link with literature on death rate inference	43
1.4	Results and organisation of the paper	43
2	The microscopic model and its large population limit	45
2.1	Notation	45
2.2	Construction of the model	45
2.3	Stability of the model	47
3	Nonparametric estimation of g and μ	49
3.1	Kernel approximation	49
3.2	Construction of estimators of g and μ	50
3.3	Oracle inequalities	51
4	Adaptive estimation under anisotropic Hölder smoothness	53
4.1	The smoothness of the McKendrick Von Foester equation	53
4.2	Minimax lower bounds	54
4.3	Adaptive estimation under anisotropic Hölder smoothness	55
5	Numerical illustration	57
6	Proof of Theorem 2.6	60
6.1	A first stability result	60
6.2	Stability of the stochastic term	62
6.3	Proof of Theorem 2.6	68
6.4	Remaining proofs of Section 2	72
7	Proofs of Section 3 and 4	74
7.1	Proof of Theorem 2.10	74
7.2	Proof of Theorem 2.12	76
7.3	Proof of Theorem 2.17	81
7.4	Proof of Theorem 2.18	86
7.5	Proof of Theorem 2.19	86
8	Appendix	87
8.1	Proof of Proposition 2.24	87
8.2	Proof of Proposition 2.16	89
8.3	Further estimates on the McKendricks Von Voester equation	91
3	Nonparametric inference of age-structured models in a large population limit with interactions, immigration and characteristics.	92
1	Introduction	92
1.1	Motivation	92
1.2	Setting	93
1.3	The microscopic model and its large population limit	93
2	Stability of the Model	96
2.1	Definition	96
3	Nonparametric estimation of g and μ	99
3.1	Construction of estimators of g and μ	100
3.2	Oracle inequalities	100
4	Minimax estimation under anisotropic Hölder smoothness	101

4.1	The smoothness of the McKendrick-Von Forster equation	101
4.2	Minimax lower bounds	102
4.3	Adaptive estimation under anisotropic Hölder smoothness	103
5	Numerical illustration	104
6	Proof of Theorem 3.13	106
6.1	A first stability result	107
6.2	Stability of the stochastic term	114
6.3	Proof of the first part of Theorem 3.13	120
6.4	Proof of the second part of Theorem 3.13	120
7	Proof of Theorem 3.15 and 3.19	121
7.1	Majoration of the death term	121
7.2	Control of \mathcal{I}_T^N and $ \Delta_T^N $	121
7.3	Proof of the Theorem 3.19	124
7.4	Proof of the proposition 3.9	125
8	Remaining proofs	126
8.1	Proof of the property 3.39	126
8.2	Proof of the property 3.21	128
8.3	Proof of Theorem 3.25	129
4	Nonparametric test of time dependance of age-structured models in a large population limit	133
1	Introduction	133
1.1	Motivation	133
1.2	Setting	133
1.3	Model	134
2	Non parametric test	134
2.1	General setting	134
2.2	Upper bound	136
2.3	Lower bound	136
3	Discussion and numerical illustration	137
3.1	Implementation	137
3.2	Numerical result	139
3.3	Algorithm	141
4	Proof of theorem 4.4	141
4.1	To begin with the end	141
4.2	Proof of the proposition 4.11	142
4.3	Proof of property 4.15	146
5	Remaining proofs	148
5.1	Discussion around the lower bound	148
5.2	Proofs of the property 4.6	149

1 Cadre mathématique et statistique

Le but de cette thèse est l'estimation du taux de mortalité lorsqu'il dépend du temps. Il s'agit donc d'évaluer une fonction dépendant de deux paramètres, le temps et l'âge. Cette thèse s'inscrit dans la suite de la thèse d'Alexandre Boumezoued [8] et des travaux de Marc Hoffmann, Marie Doumic, Adélaïde Olivier et al. [22, 35, 35].

1.1 Estimateurs à noyaux

L'estimation de fonctions rentre dans le cadre de la statistique nonparamétrique. Dans cette thèse, nous nous sommes particulièrement intéressés aux estimateurs à noyaux tels que définis dans [57]. Une introduction aux méthodes nonparamétriques peut aussi se trouver dans [63]. Cette famille d'estimateurs est indexée par une fenêtre h .

Pour ces estimateurs, dans leur utilisation initiale, on cherche à estimer une densité réelle s grâce à un échantillon X_1, \dots, X_N indépendant et identiquement distribué, selon la densité s , de N observations. Il faut alors introduire une fonction K intégrable et telle que $\int K(u)du = 1$. On l'appelle noyau. La famille des estimateurs à noyau est alors

$$\hat{s}_N(x) = \frac{1}{N} \sum_{i=1}^N K_h(X_i - x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

Un noyau est dit d'ordre L si

- pour tout $j \leq L$, $u^j K(u)$ est intégrable
- pour tout $1 \leq j \leq L$, $\int u^j K(u)du = 0$
- $\int u^{L+1} K(u)du \neq 0$, sachant que $u^{L+1} K(u)$ est intégrable.

La notion d'ordre d'un noyau joue un rôle important dans le calcul de la vitesse minimax.

1.2 Vitesse minimax

On cherche à estimer une fonction s . Pour quantifier la vitesse d'estimation d'un estimateur \hat{s} vers $s : \mathbb{R}^d \rightarrow \mathbb{R}$ nous nous plaçons dans le cadre minimax pour lequel il est nécessaire d'avoir une fonction de perte $\rho(s, \hat{s})$. Il est possible de s'intéresser à des pertes L^p où $\rho(s, \hat{s})^p = \int_A |s(x) - \hat{s}(x)|^p dx$ avec $A \subset \mathbb{R}^d$. Dans notre étude, nous nous limitons à l'étude d'une perte ponctuelle, c'est-à-dire où $\rho(s, \hat{s}) = |s(x_0) - \hat{s}(x_0)|^2$. Le risque est alors $\mathbb{E}[\rho(s, \hat{s})]$.

Dans le cadre minimax, soit N une asymptotique, souvent le nombre d'observations, on peut alors définir la vitesse minimax d'estimation d'une fonction. On notera alors \hat{s}_N un estimateur de s . Si \mathcal{V} est une classe de fonctions à laquelle s est supposée appartenir, la quantité d'intérêt est

$$R_N(\mathcal{V}) = \inf_{\hat{s}_N} \sup_{s \in \mathcal{V}} \mathbb{E}[\rho(\hat{s}_N, s)]$$

où l'infimum est pris sur tous les estimateurs de s . Ce risque est appelé risque minimax et dépend seulement de la classe fonctionnelle à laquelle on suppose que la fonction à estimer appartient.

On dit qu'un estimateur \hat{s}_N atteint la vitesse minimax r_N s'il existe deux constantes $C_U > 0$ and $C_L > 0$ telles que

$$\sup_{s \in \mathcal{V}} \mathbb{E}[\rho(s, \hat{s})] \leq C_U r_N^2 \tag{1}$$

et

$$R_N(\mathcal{V}) \geq C_L r_N^2 \tag{2}$$

La condition (1) est la borne supérieure et la condition (2), la borne inférieure. Si les deux sont vérifiées avec la même vitesse, on obtient donc la vitesse minimax optimale qui n'est définie qu'à une constante près.

L'espace fonctionnel \mathcal{V} auquel appartient la fonction que l'on veut estimer est fondamental. La vitesse optimale r_N dépend de cet espace et sera notée $r_N(\mathcal{V})$. Dans cette thèse, nous allons travailler avec des espaces de régularité Hölder qui se prêtent bien à l'utilisation du risque ponctuel. Pour une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ on dit que $f \in \mathcal{H}^\alpha(x_0)$ si $\forall y \in \mathcal{U}_{x_0}$, \mathcal{U}_{x_0} voisinage de x_0 , on a

$$|f^{(n)}(y) - f^{(n)}(x)| \leq C|y - x|^{\{\alpha\}}$$

avec $\alpha = n + \{\alpha\}$. On peut alors étendre au cas bidimensionnel avec $f \in \mathcal{H}^{\alpha, \beta}(x_0, y_0)$ si $f_{y_0} = f(\cdot, y_0) \in \mathcal{H}^\beta(x_0)$ et $f_{x_0} = f(x_0, \cdot) \in \mathcal{H}^\alpha(y_0)$. L'extension à de plus grandes dimensions se fait aisément.

Nous renvoyons à [63, 27, 32] pour une définition plus rigoureuse et précise de ces espaces dans le cadre statistique, ainsi qu'une introduction à la statistique nonparamétrique et voir les différents risques et estimateurs existants.

Nous allons illustrer la vitesse minimax pour les espaces de Hölder et le risque ponctuel. Pour la comprendre, il est nécessaire d'avoir quelques notions supplémentaires. Dorénavant, nous indexerons par h et N l'estimateur \hat{s} . Le risque ponctuel $\mathbb{E}[|s(x_0) - \hat{s}_{h,N}(x_0)|^2]$ est tel que

$$\mathbb{E}[|s(x_0) - \hat{s}_{h,N}(x_0)|^2] \leq B_{N,h}^2 + V_{N,h}$$

où $B_{N,h}$ est le biais de l'estimateur et $V_{N,h}$ sa variance. Cette décomposition est typique des estimateurs à noyaux et de la statistique nonparamétrique en général. Pour obtenir la vitesse

optimale, il faut alors trouver la fenêtre h_N qui résout le problème de maximisation suivant $h_N = \operatorname{argmin}_{h \in [0,1]} B_{N,h}^2 + V_{N,h}$. Ce problème a une unique solution puisque la somme est convexe. La théorie de l'approximation nous assure que $B_{N,h}^2$ est croissant en h , et cette croissance dépend de \mathcal{V} . La variance est quant à elle décroissante en h et indépendante de \mathcal{V} .

Si $s \in \mathcal{H}^\alpha(x_0)$, on peut montrer qu'en choisissant un noyau d'ordre $L \geq \alpha + 1$ on obtient $B_{N,h}^2 \approx h^{2\alpha}$ et $V_{N,h} \approx \frac{1}{Nh}$, où \approx signifie égal à une constante près. On obtient alors une vitesse minimax de l'ordre de $N^{-\frac{\alpha}{2\alpha+1}}$.

Si $s \in \mathcal{H}^{\alpha,\beta}(x_0, y_0)$, en adaptant l'approche pour la rendre multidimensionnelle, et en notant $\frac{1}{\gamma} = \frac{1}{\alpha} + \frac{1}{\beta}$ ou encore $\gamma = \frac{\alpha\beta}{\alpha+\beta}$, on montre que la vitesse minimax est de l'ordre de $N^{-\frac{\gamma}{2\gamma+1}}$.

Dans ce cadre, la vitesse minimax dépend de la régularité de la fonction que l'on cherche à évaluer. En pratique, une telle connaissance n'est pas donnée au praticien et il est donc nécessaire de pouvoir avoir des estimateurs que l'on appelle adaptatifs. C'est-à-dire des estimateurs qui atteignent la vitesse minimax optimale sans supposer de régularité pour la fonction à estimer.

1.3 Estimateur adaptatif

Trouver un estimateur adaptatif consiste alors à avoir un estimateur atteignant la vitesse minimax optimale sur $\mathcal{H}^\beta(x_0)$ alors même qu'on ne connaît pas la régularité de la fonction que l'on cherche à estimer. On supposera simplement que la fonction que l'on cherche à estimer est dans un espace de régularité minimale $\mathcal{H}^\alpha(x_0)$, $\alpha < \beta$.

De manière plus formelle, cela consiste à supposer que la fonction à estimer $s \in \mathcal{V}_1$ et qu'on s'intéresse à $R_N(\mathcal{V}_2)$ où $\mathcal{V}_1 \subset \mathcal{V}_2$. Il faut alors trouver un estimateur \hat{s}_* de s tel que la condition (1) devienne alors en sachant seulement que s appartient à \mathcal{V}_2 .

$$\sup_{s \in \mathcal{V}_1} \mathbb{E}[\rho(s, \hat{s}_*)] \leq C_U r_N^2(\mathcal{V}_1)$$

On voit que cela est la meilleure borne que l'on puisse espérer puisque $r_N(\mathcal{V}_1) \leq r_N(\mathcal{V}_2)$. On cherche à estimer s qui appartient à \mathcal{V}_1 bien que ce soit là une inconnue du problème. Si s appartient uniquement à \mathcal{V}_2 , l'estimateur aura la vitesse $r_N(\mathcal{V}_2)$, c'est donc en ce sens qu'il est adaptatif. La vitesse de l'estimateur s'adapte à la régularité intrinsèque inconnue de la fonction à estimer. Dans notre exemple, on aurait ainsi $\mathcal{V}_1 = \mathcal{H}^\beta(x_0)$, $\mathcal{V}_2 = \mathcal{H}^\alpha(x_0)$. Dans le cas où on cherche à estimer une densité unidimensionnelle à partir d'un N -échantillon, on obtient les vitesses $r_N(\mathcal{V}_1) = N^{-\frac{\beta}{2\beta+1}}$ qui sont grandement inférieures à $r_N(\mathcal{V}_2) = N^{-\frac{\alpha}{2\alpha+1}}$.

Il existe une procédure pour construire un tel estimateur à partir de la famille des estimateurs à noyaux. Puisque la variance $V_{N,h}$ ne dépend pas de l'espace fonctionnel, il ne reste que le biais $B_{N,h}$ à calculer pour pouvoir trouver la meilleure fenêtre h .

Pour obtenir un estimateur adaptatif, tout dépend donc du biais qu'il faut pouvoir estimer avec précision. Or, trouver un estimateur du biais est chose non aisée en général. La méthode de Goldenschluger Lepski, introduite dans [28, 29], permet de s'en affranchir.

Cette méthode propose de comparer deux à deux les estimateurs. On introduit \mathcal{H}_N une grille sous ensemble de $[0, 1]^d$. Pour $h = (h_1, \dots, h_d)$ et $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_d)$, on écrit $h \leq \mathbf{h}$ si pour tout $i \leq d$, on a $h_i \leq \mathbf{h}_i$.

On introduit alors

$$A_{N,h} = \max_{h' \in \mathcal{H}_N, h' \leq h} \left\{ \rho(\hat{s}_{h,N}, \hat{s}_{h',N})^2 - V_{N,h} - V_{N,h'} \right\}_+$$

et

$$\hat{h}_N = \operatorname{argmin}_{h \in \mathcal{H}_N} A_{N,h} + V_{N,h}$$

L'estimateur adaptatif obtenu via la méthode de Goldenschluger Lepski est alors $\hat{s}_* = \hat{s}_{\hat{h}_N}$. Cette méthode fonctionne s'il est possible d'obtenir des inégalités oracles, à savoir si on peut montrer qu'il existe une constante C telle que

$$\mathbb{E}[\rho(\hat{s}_*, s)^2] \leq C \inf_{h \in \mathcal{H}_N} [B_{N,h}^2 + V_{N,h}] + \delta_N$$

avec δ_N tendant vers 0 assez vite. Pour obtenir ce type d'inégalité, il est nécessaire d'obtenir une inégalité de concentration sur la famille d'estimateurs. Plus précisément, on a avec notre fonction de risque ρ

$$A_{N,h} \leq 4 \sup_{h' \in \mathcal{H}_N} \{ \rho(\hat{s}_{h',N}, \mathbb{E}(\hat{s}_{h',N}))^2 - V_{N,h'} \}_+ + 4 \{ \rho(\hat{s}_{h,N}, \mathbb{E}(\hat{s}_{h,N}))^2 - V_{N,h} \}_+ + 4B_{N,h}^2$$

Il suffit alors de contrôler $\mathbb{E}[4 \sup_{h' \in \mathcal{H}_N} \{ \rho(\hat{s}_{h',N}, \mathbb{E}(\hat{s}_{h',N}))^2 - V_{N,h'} \}_+]$ ce qui est possible via une inégalité de concentration sur $\rho(\hat{s}_{h',N}, \mathbb{E}(\hat{s}_{h',N}))^2$.

En effet, une inégalité de concentration non asymptotique doit assurer qu'avec grande probabilité $\rho(\hat{s}_{h',N}, \mathbb{E}(\hat{s}_{h',N}))^2$ est proche de $V_{N,h'}$. Le lien entre ces deux quantités est simplement

$$\mathbb{E} [\rho(\hat{s}_{h',N}, \mathbb{E}(\hat{s}_{h',N}))^2] = V_{N,h'}$$

1.4 Modèle de population

Le besoin d'un modèle de population ayant une asymptotique liée à un modèle déterministe simple s'impose pour réaliser notre travail. Les modèles introduits dans [26, 17, 16] et développés dans [62] se prêtent parfaitement à notre étude. Ils sont par ailleurs étudiés dans [8, 2] pour ce type d'application. Il n'est donc pas nouveau d'y penser pour réaliser de tels travaux.

Modèle stochastique

On utilise les notations suivantes pour décrire les populations. Les populations sont vues comme des mesures ponctuelles sur \mathbb{R}_+ et la mesure $Z_t = \sum_{i=1}^{n_t} \delta_{a_i}(da)$ représente une population. La population est telle que chaque individu est régi par les comportements suivants :

- il donne naissance à un taux $b(t, a)$
- il meurt à un taux $\mu(t, a)$

On fixe le temps sur lequel on regarde la population et on s'intéresse à une asymptotique en grande population. Pour cela on va prendre N individus au départ et renormaliser par N . Ainsi chaque individu aura un poids de $1/N$. On note pour une population Z_s , τ_t la translation de tous les âges d'un temps t . On a $\tau_t Z_s = \sum_{i=1}^{n_s} \delta_{a_i+t}(da)$. On note aussi $a_i(Z_s)$ l'âge de l'individu i dans

la population Z_s . Pour tout $t \in [0, T]$ on peut définir le processus comme solution de l'équation stochastique suivante :

$$\begin{aligned} Z_t^N &= \tau_t Z_0^N + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq b(s, a_i(Z_s^N)), i \leq \langle N Z_s^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta) \\ &\quad - N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{a_i(Z_s^N) + t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_s^N)), i \leq \langle N Z_s^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(ds, di, d\vartheta) \end{aligned} \quad (3)$$

avec \mathcal{Q}_1 et \mathcal{Q}_2 deux processus de Poissons indépendants d'intensité $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$. Pour toute fonction f de l'âge on définit $\langle Z_t^N, f \rangle = \sum_{i=1}^{N_t} f(a_i(Z_t^N)) = \int_0^\infty f(a) Z_t^N(da)$. On note aussi \lesssim pour inférieur à une constante près.

Hypothèse 1. *On a*

- (i) *b et μ bornées,*
- (ii) *$\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$ presque sûrement et $Z_0^N \rightarrow \xi_0$ étroitement pour un $\xi_0 \in \mathcal{M}_+$ déterministe,*
- (iii) *$\xi_0(da) = g_0(a)da$ pour une certaine fonction g_0 bornée telle que $\int_0^\infty g_0(a)da < \infty$.*

Modèle déterministe

Sous cette hypothèse on obtient la convergence de Z_t^N vers $\xi_t(da) = g(t, a)da$. De plus la densité limite g vérifie une équation de transport:

$$\begin{cases} \partial_t g(t, a) + \partial_a g(t, a) + \mu(t, a)g(t, a) = 0 \\ g(0, a) = g_0(a), \quad g(t, 0) = \int_0^\infty b(t, a)g(t, a)da. \end{cases} \quad (4)$$

Cette équation est connue depuis longtemps par les démographes. Mise en avant par McKendrick dans [53] et Von Foerster dans [66], elle est étudiée depuis longtemps. De plus elle s'inscrit dans un cadre bien plus général. Elle est en effet une équation possible apparaissant en biologie. Nous invitons le lecteur à se reporter à [58, 67], pour avoir une introduction plus poussée sur ces sujets.

Enrichissement du modèle

Pour des raisons pratiques évidentes, ce modèle est trop simple. Bien qu'il constitue déjà un défi intéressant pour les questions statistiques comme nous le verrons, on peut d'ores-et-déjà introduire un modèle plus complexe ayant un intérêt, tant en biologie qu'en actuariat.

L'idée du modèle est d'incorporer des interactions entre les individus et des traits particuliers autre que l'âge. On comprend l'importance de ce type de modèle et la littérature abonde en ce sens, comme on peut le voir dans [41] où les interactions jouent un rôle fondamental ou encore dans le cas de la population avec interaction logistique [65]. Les traits sont eux aussi très importants et apparaissent souvent dans les modèles comme dans [15, 24]. Afin d'appliquer nos travaux en biologie, il est nécessaire de prendre cela en compte.

On introduit donc un espace de trait \mathcal{X} , comme par exemple la taille de l'individu ou son type, cet espace peut donc être discret ou continue. On introduit aussi une probabilité $P(dx)$ sur cet

espace et un noyau de transition $k_b : [0, T] \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$, telle que la probabilité de changer de caractéristiques en naissant à partir d'un individu ayant l'âge a et le trait x au temps t , soit $k_b(s, a, x, x')P(dx')$. On note aussi un noyau d'interaction $U : (\mathbb{R}_+ \times \mathcal{X})^2 \rightarrow \mathbb{R}$.

Dans ce cadre les populations Z_t^N sont des mesures sur $\mathbb{R}_+ \times \mathcal{X}$, on notera $Z_t^N(da, dx) = \sum_{i=1}^{N_t} \delta_{a_i, x_i}(da, dx)$ avec x_i le trait de l'individu i .

Le taux de mort s'écrit alors

$$\mu(s, a, x, Z_s^N) = \mu_0(s, a, x) + \int_0^\infty U(a, x, \alpha, y) Z_s^N(d\alpha, dy).$$

On peut alors réécrire l'équation (3), en notant simplement $a_i(Z_{s^-}^N) = a_i$ et $x_i(Z_{s^-}^N) = x_i$, respectivement l'âge et le trait de l'individu i dans la population $Z_{s^-}^N$, pour avoir

$$\begin{aligned} Z_t^N &= \tau_t Z_0^N \\ &+ N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X}} \delta_{t-s, x'}(da, dx) \mathbf{1}_{\{0 \leq \vartheta \leq b(s, a_i, x_i) k_b(s, a, x, x'), i \leq \langle N Z_{s^-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta, dx') \\ &- N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{a_i+t-s, x_i}(da, dx) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i, x_i, Z_{s^-}^N), i \leq \langle N Z_{s^-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(ds, di, d\vartheta) \end{aligned}$$

avec \mathcal{Q}_2 et \mathcal{Q}_1 deux mesures de Poissons indépendantes d'intensité respective $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$ et $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta P(dx')$ et $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$. L'hypothèse 1 est elle aussi modifiée pour devenir

Hypothèse 2. On a

(i) b, μ_0, k_b et U bornés,

(ii) $\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$ presque sûrement et $Z_0^N \rightarrow \xi_0$ étroitement pour un $\xi_0 \in \mathcal{M}_+$ déterministe,

(iii) $\xi_0(da, dx) = g_0(a, x)daP(dx)$ pour une certaine fonction g_0 bornée telle que $\int_0^\infty g_0(a)da < \infty$.

Modèle déterministe

Sous cette hypothèse on obtient de nouveau la convergence de Z_t^N vers $g(t, a, x)daP(dx)$. De plus la densité limite g vérifie une équation de transport:

$$\begin{cases} \partial_t g(t, a, x) + \partial_a g(t, a, x) + [\mu_0(t, a, x) + \int_0^\infty U(a, x, \alpha, y)g(t, \alpha, y)d\alpha P(dy)] g(t, a, x) = 0 \\ g(0, a, x) = g_0(a, x), \quad g(t, 0, x) = \int_0^\infty \int_{\mathcal{X}} k_b(t, a, x', x)b(t, a, x')g(t, a, x')daP(dx'). \end{cases}$$

Cette équation appartient elle aussi aux équations de biologie connues. Le cas où $U = 1$ et sans traits permet par exemple d'obtenir la population avec interaction logistique, grandement étudiée. Si on ne prend que deux traits, il est possible de retrouver une équation de type Lotka et Volterra, comme souligné dans [62].

2 Première partie : Utilisation du modèle déterministe pour estimer le taux de mortalité dans un cadre discret.

2.1 Motivation

Dans la lignée de la thèse d’Alexandre Boumezoued, nous nous sommes intéressés aux tables de mortalité. Les tables de mortalité sont fondamentales pour les études actuarielles. Elles donnent le taux de mortalité pour chaque classe d’âge et chaque période temporelle. L’histoire des tables de mortalité est riche et il est possible d’y trouver une introduction dans [21].

Les premiers démographes ont déjà posé le besoin de répondre à deux problèmes.

- (1) Le taux de mortalité dépend de l’âge et du temps.
- (2) Il est nécessaire d’avoir une compréhension globale de la dynamique de population pour trouver le taux de mortalité.

L’estimation dans le cadre discret du taux de mortalité pose néanmoins plusieurs problèmes. On ne dispose en effet que de données agrégées pour réaliser les estimations de mortalité. Les données sont souvent représentées dans un diagramme de Lexis, avec par exemple la figure 1. Dans cette figure l’abscisse correspond aux années civiles, ici cela va de 2008 à 2011, et l’ordonnée correspond aux âges des individus observés.

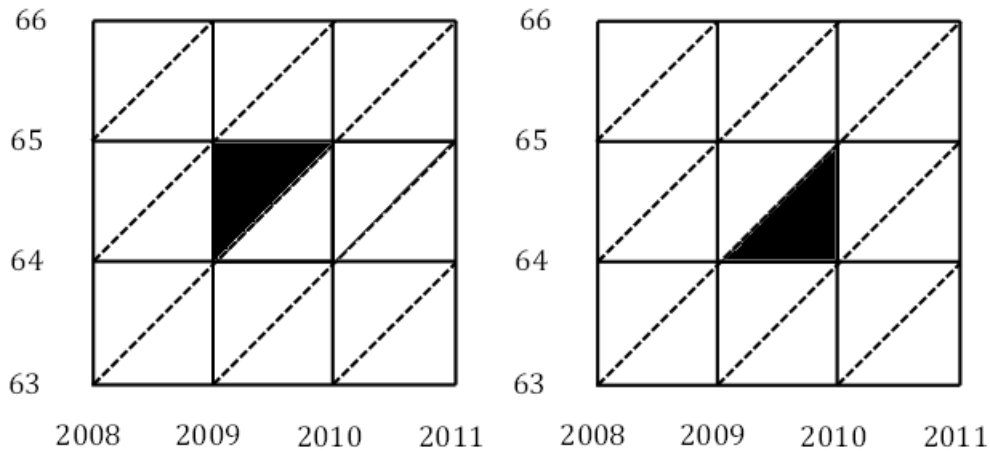


Figure 1: Diagramme de Lexis. Gauche : En noir, triangle supérieur associé à l’âge 64 et l’année 2009. Droite : En noir, triangle inférieur associé à l’âge 64 et l’année 2009.

Les données disponibles sont le nombre de personnes agrégées sur chaque bord des carrés du diagramme, c’est-à-dire la population ayant un certain âge pour une année donnée, et le nombre de morts dans chaque triangle. Ces données sont disponibles via l’HMD (Human Mortality Database) [34]. On peut donner un exemple pour la France avec la figure suivante.

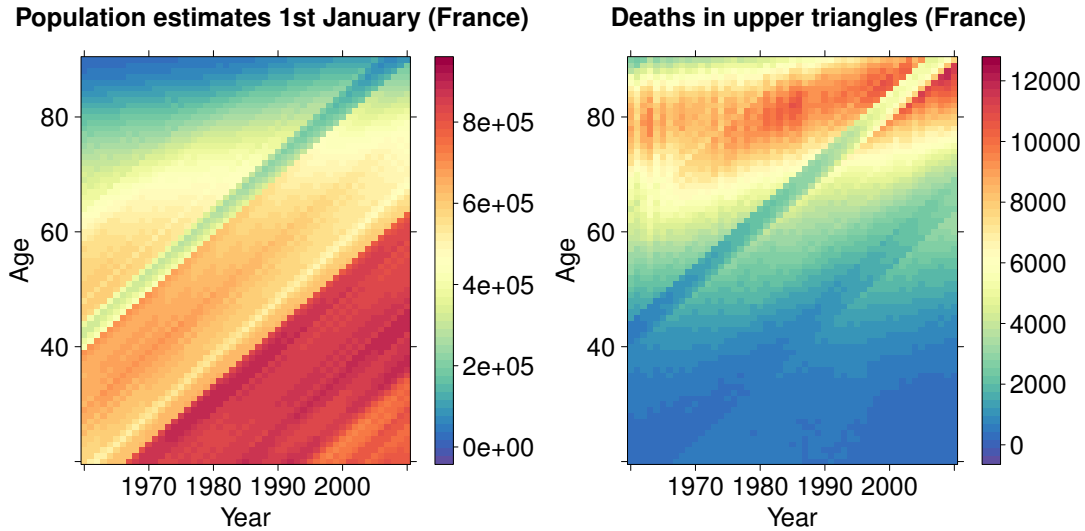


Figure 2: Gauche : Population france pour chaque âge et année. Droite : Nombre de morts dans les triangles supérieurs du diagramme de Lexis.

Afin de trouver le taux de mortalité dans ce cadre, plusieurs proxys existent. Néanmoins, les méthodes employées aboutissent souvent à un effet, dit l'effet cohorte que l'on peut voir sur la figure 3. Dans cette figure, le taux d'amélioration de la mortalité correspond à $\frac{\mu_U(x,t+1) - \mu_U(x,t)}{\mu_U(x,t)}$ avec $\mu_U(x,t)$ la mortalité, constante, dans le triangle supérieur à l'âge x et au temps t . Le taux d'amélioration est donc relatif. Un taux négatif correspond à une amélioration, au sens où le taux de mort diminue. Dans le cas contraire le taux de mort augmente.

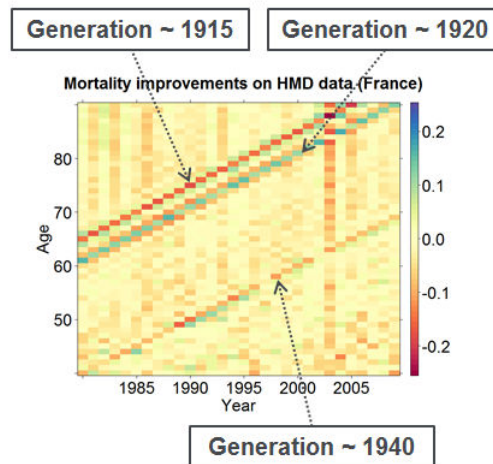


Figure 3: Taux d'amélioration de la mortalité. La génération 1915 a une mortalité plus faible.

A notre connaissance, les premières explications de cet effet ont été données par [60] en 2008. Il fut conjecturé que ces effets étaient dus à des chocs dans les naissances. Plus précisément, ces chocs correspondent à un nombre de naissances drastiquement différent pour certaines périodes historiques, telles que les guerres ou les épidémies. Cela a été confirmé en 2016 dans [14].

Plusieurs personnes ont essayé de résoudre ce problème typique via différentes approches [12, 40] avec en particulier [30] qui supprime l'effet cohorte. Mais à notre connaissance aucune approche n'a mis en oeuvre l'utilisation du modèle déterministe décrit précédemment permettant la prise en compte de la dynamique globale de population. Récemment, la méthodologie de l'HMD et L'HFD (Human fertility database)[33] a changé et tend à une prise en compte similaire à ce qui est fait dans la première contribution de cette thèse.

Les questions qu'il est nécessaire de se poser sont donc les suivantes:

1. Comment peut-on faire pour prendre en compte les naissances?
2. Quel est le lien avec les anciennes estimations?

La première est naturelle au vue de ce qui a été dit. La seconde vient du problème inhérent au métier d'actuaire pour lequel chaque nouvelle méthode doit s'inscrire dans un cadre historique.

2.2 Résultats de la première contribution

Les résultats qui suivent sont issus du chapitre 1 de la thèse. L'idée est d'utiliser le modèle déterministe 4 pour effectuer nos calculs dans lequel le taux de mort est supposé constant sur chaque triangle indexé par x, t pour l'âge x et le temps t . Comme on ne dispose que des morts sur chaque triangle, nous devons supposer le taux de mort constant sur chaque triangle. Il est en effet impossible d'avoir un taux de mort plus précis avec les données actuelles. On note $\mu_L(x, t)$ la valeur du taux de mort sur le triangle inférieur indexé par l'âge x et le temps t . De même on note $\mu_U(x, t)$ le taux de mort sur le triangle supérieur indexé par l'âge x et le temps t .

Les données sont donc $D_L(t, x)$ le nombre de morts sur le triangle inférieur commençant au temps t et à l'âge x . De la même manière on notera $D_U(t, x)$ le nombre de morts sur le triangle supérieur et $N(t, x)$ le nombre d'individus d'âge x entre t et $t + 1$. On suppose de plus que nous est donné $g(t, 0)$, c'est-à-dire le profil des naissances chaque année. Cela n'est vrai que sur une base mensuelle agrégée mais il est facile de discrétiser les résultats. On définit alors

$$L_y(\theta) = \frac{\int_0^1 g(y+v, 0) \exp(-\theta v) dv}{\int_0^1 g(y+v, 0) dv}.$$

On a alors le résultat suivant, voir l'algorithme 1 du premier chapitre, pour calculer le taux de mort.

Algorithme 3. Commencer à l'âge $x = 0$:

- (i) Résoudre l'équation suivante pour estimer le taux de mort $\mu_L(x, t)$ pour les triangles inférieurs sur toutes les années t possibles,

$$\exp(-\mu_L(t, x)) L_{t-x}(H(t, x) - \mu_L(t, x)) = \left(1 - \frac{D_L(t, x)}{N(t, x)}\right) L_{t-x}(H(t, x)),$$

(ii) Ensuite, résoudre l'équation

$$\begin{aligned} & L_{t-x-1}(H(t-1, x) - \mu_L(t-1, x)) \\ &= \left(1 + \frac{D_U(t, x)}{N(t, x+1)}\right) L_{t-x-1}(H(t-1, x) - \mu_L(t-1, x) + \mu_U(t, x)). \end{aligned}$$

pour inférer le taux de mort $\mu_U(t, x)$ sur le triangle supérieur pour toute année t disponible,
(ii) Calculer la valeur $H(t, x) = H(t-1, x) + \mu_U(t, x) - \mu_L(t-1, x)$ pour chaque t , définir $x \leftarrow x+1$ et retourner à l'étape (i) .

avec $H(0, y) = 0$.

Cela nous permet d'aboutir aux améliorations illustrées dans les figures 4 et 5.

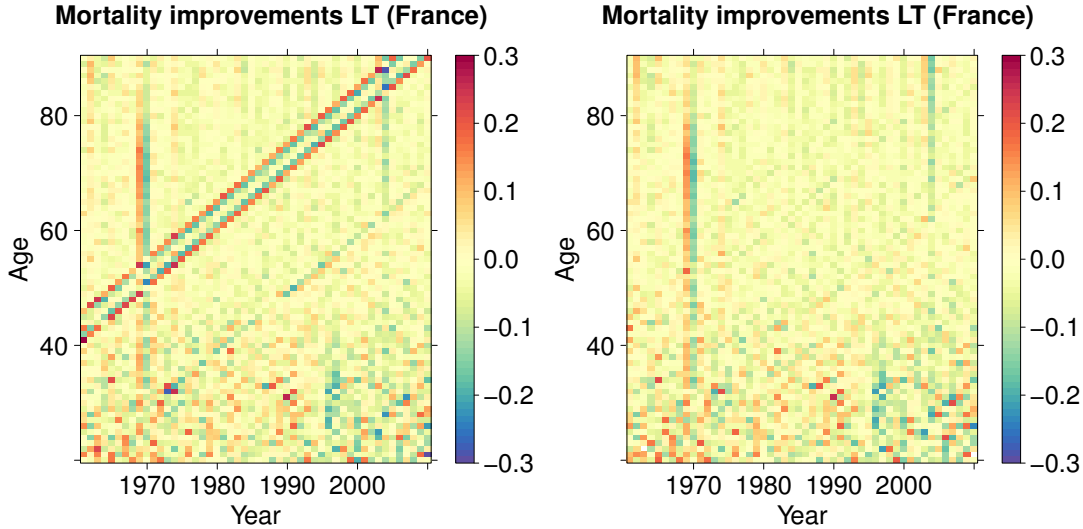


Figure 4: Amélioration de la mortalité dans les triangles inférieurs.

De plus il est possible de relier les estimés obtenus à ceux des méthodes précédentes. Nous ne détaillerons pas les calculs ici et invitons le lecteur à lire la contribution en lien avec cette section, à savoir le chapitre 1.

Nous avons grâce à cette contribution pu répondre aux deux questions que nous nous posions. Néanmoins, cette méthode ne peut être considérée que comme une approximation asymptotique en utilisant le modèle déterministe limite du modèle stochastique de population. La question qui se pose alors est de trouver un moyen d'estimer l'écart entre le modèle stochastique et sa limite pour cette application.

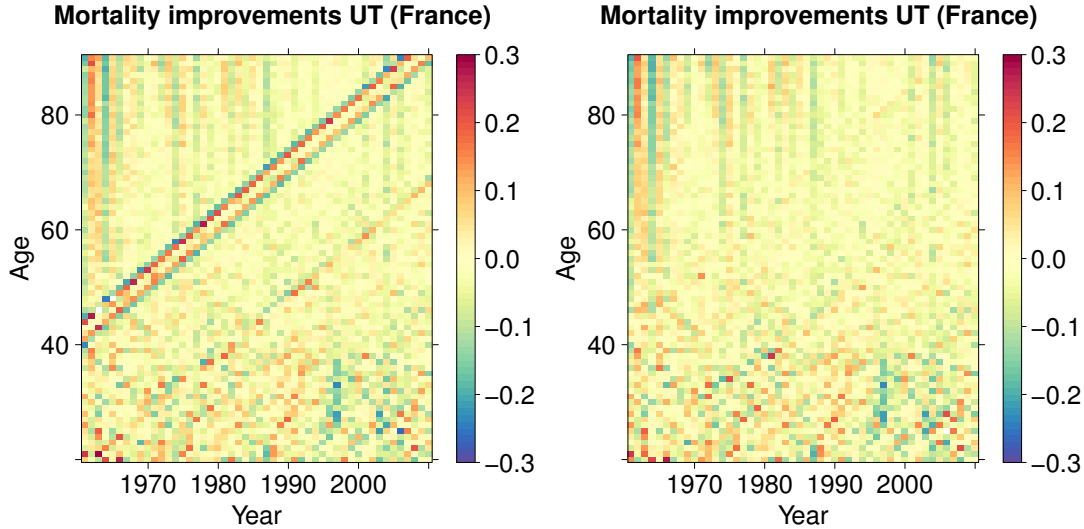


Figure 5: Amélioration de la mortalité dans les triangles supérieurs.

3 Deuxième partie : Estimation nonparamétrique du taux de mort.

Les chapitres 2 et 3 sont intimement liés. Ils sont issus d'une même problématique : comprendre l'écart entre le modèle stochastique et la limite. Pour cela nous nous sommes intéressés à un modèle plus théorique dans lequel nous nous sommes fixés l'objectif d'estimer le taux de mortalité, et la densité de population en se donnant uniquement le processus de population Z_t^N , $t \in [0, T]$ et T fixé.

On note par ailleurs qu'il est impossible de retrouver le taux de fertilité dans ce cadre puisque nous ne disposons pas de la généalogie. En ce sens, notre travail est différent des travaux sur les populations de cellules pour lesquelles le taux de mort et de naissance est le même. Une cellule meurt en donnant naissance à plusieurs cellules. Nous invitons le lecteur à voir la littérature sur ce sujet, voir [22, 35] ou [51], liste non exhaustive pour comprendre les différentes techniques mises en jeu. Néanmoins celles-ci diffèrent totalement de notre cadre, même si elles restent un point de comparaison intéressant et fondamental.

Par ailleurs, la littérature concernant le taux de mortalité du point de vue théorique comme nous l'envisageons est assez large et variée. C'est en effet une problématique assez ancienne apparaissant dans plusieurs domaines d'études appliquées, tels que la biologie, la médecine, l'analyse de survie pour des machines. Des auteurs de différents domaines se sont donc intéressés à des problèmes de ce type, voir [3], [20], [40], [52], [55], [19]. Ces papiers prennent en compte la dépendance temporelle du taux de mort.

Quelques articles se sont aussi intéressés à une résolution du problème via des outils propres aux équations différentielles. Ce type d'approche trouve une grande littérature en biologie, voir [58], [56].

A notre connaissance, il n'existe pas de résultat dans le cadre minimax pour ce problème précis. Aussi avons-nous à l'esprit les questions suivantes, typiques d'un problème d'estimation nonparamétrique.

- (1) Peut-on trouver un estimateur de μ , le taux de mortalité et de g , la densité de population, optimal au sens minimax?
- (2) Si oui, peut-on trouver un estimateur adaptatif de ces fonctions?
- (3) Peut-on quantifier les déviations entre le modèle stochastique et le modèle déterministe limite?

On rajoute la troisième question, en lien avec le premier chapitre. Il est important de noter que la question 3 trouve en partie sa réponse dans la thèse de Tran [62], puisque y sont démontrées des inégalités de grandes déviations en lien avec le processus Z_t^N . Néanmoins, notre objectif dans la question 3 est d'avoir des résultats non asymptotiques, bien qu'on pourrait se satisfaire de résultats asymptotiques pour une première réponse dans le cadre du chapitre 1.

Nous avons choisi de ne pas introduire de traits dans les résultats présentés dans cette introduction. Nous avons aussi choisi de commencer par traiter le cas où le taux de mort ne dépendait pas d'interactions. Par simplicité nous présenterons les résultats de la même manière dans l'introduction. Nous verrons que les résultats sont de nature légèrement différente, justifiant une présentation séparée.

Nous faisons l'hypothèse de régularité suivante sur les paramètres.

Hypothèse 4. $b \in \mathcal{H}^{\alpha,\beta}$, $\mu_0 \in \mathcal{H}^{\gamma,\delta}$, $U \in \mathcal{H}^{\rho,\eta}$ et $g_0 \in \mathcal{H}^\nu$. Avec $\min(\alpha, \beta, \nu) \geq \max(\gamma, \delta) + 1$ et $\delta - 1 \leq \gamma \leq \delta$.

On notera $\mu = \mu_0$ sans interactions et $\mu(t, a) = \mu_0(t, a) + \int_0^\infty U(a, \alpha) Z_t^N(d\alpha)$ s'il y a des interactions. Nous avons ajouté U dans cette hypothèse que nous utilisons pour le cas sans et avec interactions. Ce choix est dicté par la simplicité et lorsqu'il n'y a pas d'interactions, les contraintes sur U sont inutiles, U n'étant pas un paramètre du modèle.

3.1 Résultats du chapitre 2

On s'est intéressé dans un premier temps à un modèle sans interactions et sans traits. Bien que l'estimation de μ dans le modèle inhomogène en temps devienne un problème bidimensionnel, l'estimation de g reste un problème unidimensionnel. En effet, pour déterminer la densité de population à un temps donné, il suffit de connaître la population à ce temps précis. On a alors besoin d'estimer une fonction d'un paramètre, l'âge, pour pouvoir retrouver g .

Les deux théorèmes sont déduits des théorèmes 2.17, 2.18, 2.19 du chapitre 2. Dans la suite K , K^b et K^t sont des noyaux d'ordre assez grand pour assurer que la vitesse minimax optimale est atteignable. On peut trouver une définition de l'ordre d'un noyau dans [63].

Théorème 5. *Sous les hypothèses 1, 4, en supposant que la densité g est bornée inférieurement (hypothèse 2.11) et qu'on dispose d'un bon contrôle de la convergence de Z_0^N vers g_0 (hypothèse 2.5), il existe h_N , constructible par la méthode de Goldenschluger Lepski, tel que l'estimateur*

$$\hat{g}_{N, h_N}(s, a) = \int_0^\infty K_{h_N}(\alpha - a) Z_s^N(d\alpha)$$

converge vers g , la densité de population, à la vitesse optimale dans le cadre minimax.

3. DEUXIÈME PARTIE : ESTIMATION NONPARAMÉTRIQUE DU TAUX DE MORT.

Pour l'estimation du taux de mort, on définit tout d'abord $\Gamma^N(dt, da) = \sum_{i=1}^{D_T^N} \delta_{\tau_i, a_i}(dt, da)$, avec D_T^N le nombre de morts, τ_i, a_i les dates et âges des morts. On obtient un résultat similaire à la densité.

Théorème 6. *Sous les hypothèses 1, 4, en supposant que la densité g est bornée inférieurement (hypothèse 2.11) et qu'on dispose d'un bon contrôle de la convergence de Z_0^N vers g_0 (hypothèse 2.5), il existe $h_{1,N}$ et $h_{2,N}$, constructibles par la méthode de Goldenschluger Lepski, tel que l'estimateur*

$$\hat{\mu}_{N, h_{1,N}, h_{2,N}, h_N}(s, a)_{\varpi} = \frac{\int_0^T \int_0^\infty K_{h_{1,N}}^t(x-s) K_{h_{2,N}}^b((x-\alpha)-(s-a)) \Gamma^N(dx, d\alpha)}{\hat{g}_{N, h_N}(s, a)} \mathbf{1}_{\hat{g}_{N, h_N}(s, a) \geq \varpi}$$

converge vers le taux de mortalité μ à la vitesse optimale dans le cadre minimax.

Le choix de ϖ se fait en fonction de la borne inférieure de la densité g . C'est une constante calculable.

On note que μ est issu de la division de deux estimateurs dans le système de coordonnées obtenu par le temps t et la date de naissance $t - a$. Ce choix de coordonnées se justifie pour améliorer la régularité de la fonction à estimer en l'occurrence μg , ce qui améliore la vitesse d'estimation et nous permet d'obtenir la vitesse optimale.

De plus en utilisant la méthode de Goldenschluger Lepski, on trouve des estimateurs adaptatifs et optimaux à un facteur logarithme près. On peut illustrer cela en calculant les vitesses de nos estimateurs, comme montré sur les figures 6 et 7. Dans ces figures on trouve en abscisse le logarithme du nombre d'individus dans la population initiale, et en ordonnées le logarithme de l'erreur moyenne empirique d'estimation. En théorie ces courbes devraient être des droites avec une pente dépendant de la régularité de la fonction à estimer.

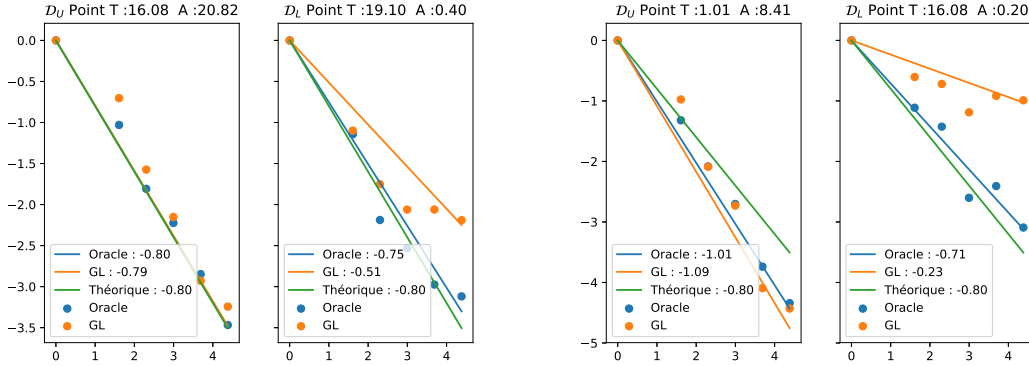


Figure 6: Vitesse en échelle logarithmique de l'estimation de la densité g sur 50 simulations. Les points d'estimation sont écrits dans les graphes.

3. DEUXIÈME PARTIE : ESTIMATION NONPARAMÉTRIQUE DU TAUX DE MORT.

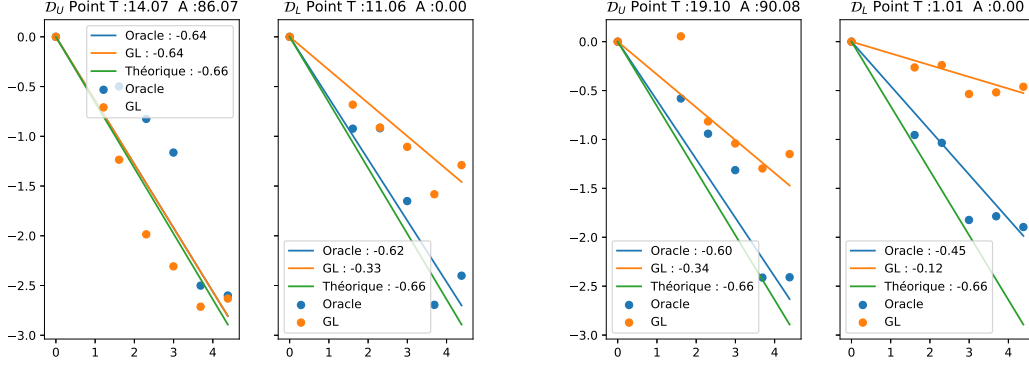


Figure 7: Vitesse en échelle logarithmique de l'estimation de la densité g sur 50 simulations. Les points d'estimation sont écrits dans les graphes.

Les deux figures précédentes nous permettent de vérifier que la vitesse d'estimation, calculée de manière empirique sur 50 simulations, est proche de la vitesse théorique. La courbe en bleu est la courbe oracle, c'est-à-dire la courbe pour laquelle on calcule le risque directement grâce à la fonction que l'on cherche à estimer. En temps normal, la courbe orange, obtenue à partir des données, est au-dessus de la courbe bleue, elle-même au-dessus de la courbe verte. On remarque pour le taux de mort que l'estimation a plus de difficultés à converger en certains points, la vitesse est plus faible que celle qu'on pourrait espérer. Ceci est la conséquence d'un manque d'information en ces points. Plus prosaïquement, il y a peu de gens et donc de morts en ces points. On remarque aussi certains points pour la densité de population où la convergence est plus rapide. Cela peut être dû à la forme particulière du taux de naissance que l'on a choisi, et à la faible dépendance temporelle du taux de mort dans cette simulation.

On introduit deux quantités.

$$\mathcal{W}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| \int_0^\infty w_2(t-a) f_t(a) (Z_t^N(da) - g(t,a)da) \right|$$

et

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| \int_0^t w_1(s) \int_0^\infty w_2(s-a) f_s(a) (Z_s^N(da) - g(s,a)da) ds \right|,$$

où w_1 and $w_2(\cdot)$ sont deux fonctions de poids bornées qui dépendent possiblement de N . On note $f_t(a) = f(t,a)$ pour f bornée. Implicitement on suppose que \mathcal{F} est assez régulier pour que ces variables aléatoires existent et soient mesurables. On note pour toute fonction $|w|_{1,\infty}^2 = |w|_1 |w|_\infty$, avec $|\cdot|_\infty$ et $|\cdot|_1$, la norme 1 et la norme infinie usuelle.

L'ensemble \mathcal{F} est muni de la norme infinie et on suppose que $\text{diam}(\mathcal{F}) \leq 1$. On suppose que l'entropie métrique de \mathcal{F} est finie, c'est-à-dire

$$\int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

Avec $\mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)$ le nombre de boules de taille ϵ pour la norme infinie, nécessaire pour recouvrir \mathcal{F} . Si \mathcal{F} est assez riche alors on a le théorème suivant, issu du théorème 2.6 du chapitre 2.

Théorème 7. *S'il existe une constante $C > 0$ telle que $\mathbb{P} [|w_2|_{1,\infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_0 \geq CN^{-1/2}(1+u)] \leq (e^u - 1) \wedge 1$ (hypothèse 2.5), si \mathcal{F} est assez riche (hypothèse 2.2), si l'entropie métrique de \mathcal{F} est finie et si l'hypothèse 1 est vérifiée, il existe deux constantes C_1 et C_2 , dépendant explicitement de quantités connues, telles que*

$$\mathbb{P} \left[|w_1|_{1,\infty}^{-1} |w_2|_{1,\infty}^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \geq C_1 N^{-1/2}(1+u) \right] \leq (e^u - 1) \wedge 1$$

et pour tout $t \in [0, T]$ on a

$$\mathbb{P} \left[|w_2|_{1,\infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_t \geq C_2 N^{-1/2}(1+u) \right] \leq (e^u - 1) \wedge 1$$

On a donc répondu en grande partie aux questions que l'on s'était posées dans ce cadre. Deux points sont améliorables. Le premier point vient du besoin du taux de mort de dépendre du temps. Si ce n'est pas le cas, nos simulations, mais aussi la borne inférieure, indiquent une vitesse en \sqrt{N} , bien meilleure que la vitesse minimax avec la régularité en âge que l'on obtient. Cela ne contredit donc pas l'hypothèse 4 mais nous indique qu'il serait possible de relaxer les conditions sur la régularité en temps du taux de mortalité tout en gardant un estimateur adaptatif optimal. Le second point concerne l'inégalité de concentration. Si celle-ci est suffisante pour nos besoins statistiques, elle n'est pas optimale. On peut se rendre compte de cela en prenant simplement un processus de population avec un taux de naissance constant et un taux de mort nul. On obtient alors un processus de Poisson usuel et l'inégalité de concentration que l'on a est sous optimal dans ce cadre.

3.2 Résultats du chapitre 3

Dans ce cadre, à savoir un modèle avec interactions et sans traits, on obtient des résultats similaires, bien que plus faibles pour les inégalités de concentration. On obtient là aussi l'optimalité de nos estimateurs et la possibilité d'avoir des estimateurs adaptatifs optimaux, à un facteur en puissance de logarithme près. Les deux premiers théorèmes sont une conséquence des théorèmes 3.25, 3.26 et 3.27 du chapitre 3.

Théorème 8. *Sous les hypothèses 2, 4, en supposant que la densité g est bornée inférieurement (hypothèse 3.18), qu'on dispose d'un bon contrôle de la convergence de Z_0^N vers g_0 (hypothèse 3.11) et que l'interaction est constante en dehors d'un compact (hypothèse 3.7), il existe h_N , constructible par la méthode de Goldenschluger Lepski, tel que l'estimateur*

$$\hat{g}_{N, h_N}(s, a) = \int_0^\infty K_{h_N}(\alpha - a) Z_s^N(d\alpha)$$

converge vers g , la densité de population, à la vitesse optimale dans le cadre minimax.

Pour l'estimation du taux de mort, on obtient un résultat similaire. De nouveau la constante ϖ est calculable en fonction de la borne inférieure sur la densité g .

Théorème 9. *Sous les hypothèses 2, 4, en supposant que la densité g est bornée inférieurement (hypothèse 3.18), qu'on dispose d'un bon contrôle de la convergence de Z_0^N vers g_0 (hypothèse 3.11) et que l'interaction est constante en dehors d'un compact (hypothèse 3.7), il existe $h_{1,N}$ et $h_{2,N}$, constructibles par la méthode de Goldenschluger Lepski, tel que l'estimateur*

$$\hat{\mu}_{N, h_{1,N}, h_{2,N}, h_N}(s, a)_\varpi = \frac{\int_0^T \int_0^\infty K_{h_{1,N}}^t(x-s) K_{h_{2,N}}^b((x-\alpha) - (s-a)) \Gamma^N(dx, d\alpha)}{\hat{g}_{N, h_N}(s, a)} \mathbf{1}_{\hat{g}_{N, h_N}(s, a) \geq \varpi}$$

converge vers le taux de mortalité μ à la vitesse optimale dans le cadre minimax.

On peut par ailleurs illustrer cela avec les graphes de vitesse.

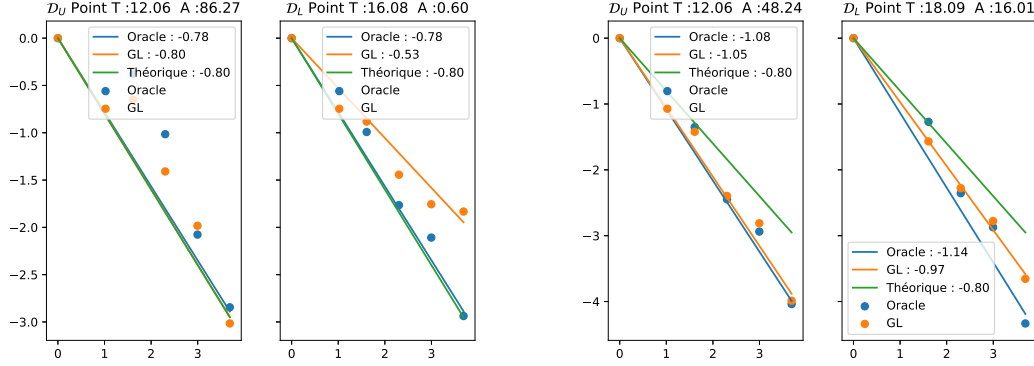


Figure 8: Vitesse en échelle logarithmique de l'estimation de la densité g sur 50 simulations. Les points d'estimation sont écrits dans les graphes.

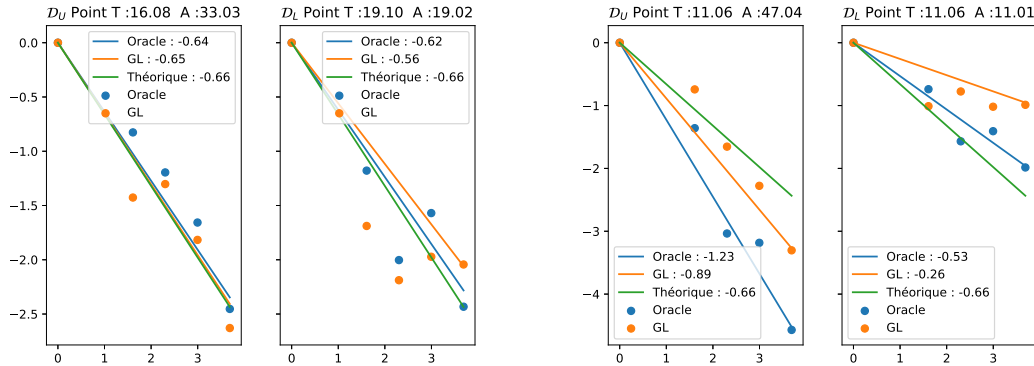


Figure 9: Vitesse en échelle logarithmique de l'estimation de la densité μg sur 50 simulations. Les points d'estimation sont écrits dans les graphes.

En revanche les inégalités de concentration s'expriment sous une forme plus faible en raison de la présence des interactions. On prend des hypothèses similaires pour \mathcal{F} et pour $|w_2|_{1,\infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_0$. Le théorème suivant est une conséquence du théorème 3.13 du chapitre 3.

Théorème 10. *On se place sous les hypothèses 3.1, 3.3, 3.4, 3.5, 3.11 et on suppose que l'entropie métrique de \mathcal{F} est finie. Pour toute fonction φ une fonction strictement positive telle que $\forall x, y > 0 \varphi(x + y) \lesssim \varphi(x) + \phi_\varphi(y)$ où ϕ_φ est une fonction connue dépendant de φ , croissante et positive telle que pour tout $C > 0 \phi_\varphi(Cx) \lesssim \phi_\varphi(x)$.*

Si $|w_2 w_3|_\infty \lesssim N^{1/2} |w_2 w_3|_1$, il existe un évènement \mathcal{B}_N avec $\mathbb{P}(\mathcal{B}_N^c) \lesssim e^{-\sqrt{N}}$ tel que

(i) il existe une variable aléatoire X^N telle que

$$\mathbb{P}\left[|w_1|_{1,\infty}^{-1}|w_2|_{1,\infty}^{-1}X^N \geq CN^{-1/2}(1+u)\right] \leq (e^u - 1) \wedge 1$$

telle que

$$\mathbb{E}[\varphi(\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T)] \lesssim \mathbb{E}[\varphi(\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T)|\mathcal{B}_N] + \mathbb{E}[\varphi(X^N)] + \mathbb{P}(B_N^c)\phi_\varphi(|w_1 w_2|_1)$$

et

$$\mathbb{P}(|w_1 w_2|_{1,\infty}^{-1}\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \geq C'N^{-1/2}(1+u)|\mathcal{B}_N) \leq (e^u - 1) \wedge 1.$$

(ii) pour tout $t \in [0, T]$ il existe une variable aléatoire X_t^N telle que

$$\mathbb{P}\left[|w_2|_{1,\infty}^{-1}X_t^N \geq C''N^{-1/2}(1+u)\right] \leq (e^u - 1) \wedge 1$$

telle que

$$\mathbb{E}[\varphi(\mathcal{W}_{w_2}^N(\mathcal{F})_t)] \lesssim \mathbb{E}[\varphi(\mathcal{W}_{w_2}^N(\mathcal{F})_t)|\mathcal{B}_N] + \mathbb{E}[\varphi(X_t^N)] + \mathbb{P}(B_N^c)\phi_\varphi(|w_2|_1)$$

et

$$\mathbb{P}(|w_2|_{1,\infty}^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T \geq C'''N^{-1/2}(1+u)|\mathcal{B}_N) \leq (e^u - 1) \wedge 1.$$

Nos résultats s'appliquent de manière plus générale au cas avec des traits. Nous ne l'avons pas écrit pour des raisons de simplicité.

Ce chapitre laisse deux questions importantes. Tout d'abord, pouvons-nous savoir si μ dépend du temps, ce qui est fondamental puisque le problème devient alors de nouveau unidimensionnel et donne une vitesse minimax différente dans ce cas. Ensuite, nous pouvons nous demander s'il est possible de tester la présence d'interactions, responsable d'une dépendance temporelle mais aussi d'une dégradation de la vitesse minimax.

4 Troisième Partie : Test de présence d'interactions

Etant capable d'estimer le taux de mort en présence d'interactions, une question naturelle est alors

(1) Peut-on tester la présence d'interactions?

Il faut pour répondre à cette question se rappeler qu'on ne dispose que d'une population et des trajectoires de vie de chaque individu. A notre connaissance ce problème n'a jamais été traité dans la littérature. Cela s'explique par le fait que les résultats sur le modèle qu'on utilise sont récents. D'autre part, à notre connaissance, cette question précise, à savoir la détection d'interactions dans le taux de mort dans une population structurée par âge, ne s'est pas encore posée. Bien qu'il existe une littérature abondante sur les tests d'interactions entre des coordonnées de vecteurs de variables aléatoires, notre problème est bien différent. Les données n'ont aucune structure d'indépendance.

Un test relativement simple serait de disposer de deux populations indépendantes avec les mêmes paramètres, à l'exception de la condition initiale. Dans ces conditions il suffirait de tester si les deux taux de mort estimés dans chacune des populations sont les mêmes. Si c'était le cas, il n'y aurait pas d'interactions, sinon il y en aurait. Dans notre cadre, avec une seule population, il

n'y a aucun moyen de créer deux populations indépendantes puisqu'il n'y a aucun moyen de savoir qui n'aurait pas dû mourir sans le phénomène d'interactions.

Il est par ailleurs possible de voir qu'avec une seule population on ne pourra jamais détecter la présence d'interactions. En effet, si cela était possible, on devrait pouvoir le faire dans le modèle déterministe. Or, avec une seule population, on ne peut avoir que $\mu(t, a) = \mu_0(t, a) + \int_0^\infty U(a, \alpha)g(t, \alpha)d\alpha$. Avec la donnée de g , U reste non identifiable tant que μ_0 dépend du temps. Si μ_0 est indépendant du temps U est alors identifiable et détectable.

Aussi le test que nous avons réalisé permet de vérifier si le taux de mortalité dépend du temps ou non. On pose $d_{\mathcal{F}}(\mu) = |\mu - m_\mu|_2$ avec $m_\mu(a) = \int_0^T \mu(s, a)g(s, a)ds / \int_0^T g(s, a)ds$. Donc $d_{\mathcal{F}} = 0$ si μ ne dépend pas du temps. On note $d_{\mathcal{F}}^N$ un estimateur précis de $d_{\mathcal{F}}$ et

$$\mathcal{F}(\rho_N) = \{f \in \mathcal{F} \mid d_{\mathcal{F}}(f) \geq \rho_N\}.$$

Nous invitons le lecteur à voir [37], [38] pour une définition plus précise des tests nonparamétriques dans le cadre minimax. On pourra aussi trouver dans [45] une problématique proche de celle qui nous intéresse dans cette thèse.

Pour un test ϕ_N on définit l'erreur globale comme la somme de l'erreur de première et seconde espèce.

$$R(C, \phi_N, \rho_N) = \sup_{\mu \in \mathcal{H}^\beta} \mathbb{P}_\mu(\phi_N = 1) + \sup_{\mu \in \mathcal{F}(C\rho_N)} \mathbb{P}_\mu(\phi_N = 0)$$

où \mathbb{P}_μ consiste en la probabilité pour le modèle paramétrisé par μ . Soit $0 < \alpha < 1$, la borne supérieure consiste à trouver un test ϕ_N^* et une constante C_\star tels que

$$\limsup_{N \rightarrow \infty} R(C, \phi_N^*, \rho_N^*) \leq \alpha$$

pour tout $C \geq C_\star$ et $\rho_N \geq \rho_N^*$.

Théorème 11. *Il existe une constante C_\star dépendant seulement des paramètres et $\phi_N^* = \mathbf{1}_{d_{\mathcal{F}}^N \geq \frac{C_\star}{2} \rho_N}$ telle que*

$$\limsup_{N \rightarrow \infty} R(C, \phi_N^*, \rho_N^*) = 0$$

pour tout $C \geq C_\star$ et $\rho_N \geq \rho_N^* = C_K \log(N) N^{-\frac{\min(\alpha, \beta, \delta, \rho)}{2 \min(\alpha, \beta, \delta, \rho) + 1}}$

On obtient ainsi la borne supérieure pour le test, voir le théorème 4.4 du chapitre 4. Notre approche pour détecter la présence d'interactions nécessite de faire une hypothèse supplémentaire.

Hypothèse 12. μ_0 est indépendant du temps, c'est-à-dire que

$$\mu(t, a) = \mu_0(a) + \int_0^\infty U(a, \alpha)Z_t^N(d\alpha)$$

Sous cette hypothèse la dépendance temporelle est équivalente à la détection d'interactions, ce qui répond en partie à la question. En pratique, on est capable de réaliser ce test via l'algorithme suivant, issu de la sous-section 3.3 du chapitre 4, où les τ_i, a_i correspondent aux temps et âges de morts dans la population. Les t_i sont une subdivision de $[0, T]$ telle que pour tout i , il existe i_0 tel que $\tau_i = t_{i_0}$.

Pour l'algorithme qui suit on définit D_T^N , le nombre de morts dans la population stochastique observée Z_t^N entre 0 et T . R_T^N est le cardinal d'une subdivision de l'intervalle $[0, T]$ contenant l'ensemble des temps de morts et assez grand pour assurer que les résultats aient un sens. La proposition 4.6 permet de mieux comprendre les contraintes sur R_T^N .

Algorithme 13.

- (1) Calculer pour tout t_i et a_k , $\theta_{i,k}^N = \hat{g}_*^N(t_i, a_k)$ l'estimateur adaptatif de g , et $\mu_{i,k}^N = \hat{\mu}_*^N(t_i, a_k)$ l'estimateur adaptatif de μ . Obtenir la variance $V_{i,k}^N$ de l'algorithme issue de la méthode de Goldenschluger Lepski.
- (2) Poser $V^N = 0.2 \max(V_{i,k}^N)$
- (3) Calculer $c_{\mathcal{F}}^N = \frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \mu_{i,k}^N \left[\mathbf{1}_{t_i=\tau_k} - \frac{1}{\int_0^T \int_0^\infty \mathbf{1}_{u=a_k} Z_s^N(du)ds} (t_{i+1} - t_i) \theta_{i,k}^N \right]$.
- (4) Retourner la valeur $\mathbf{1}_{c_{\mathcal{F}}^N \geq V^N}$

Cela nous donne les figures suivantes.

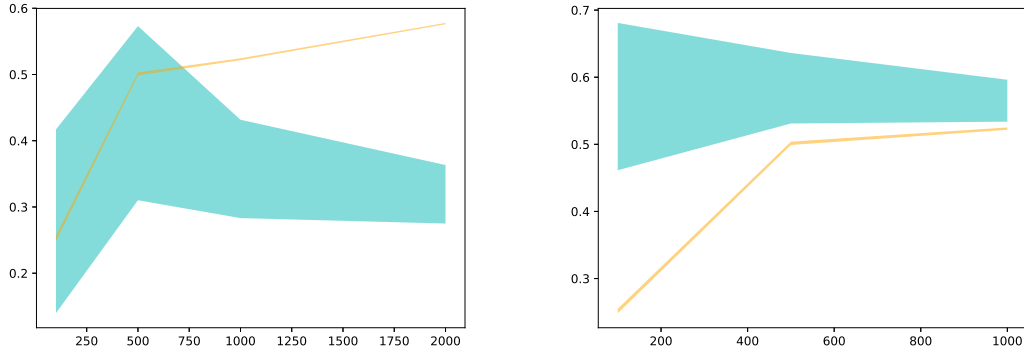


Figure 10: Variance en jaune, statistique de test en bleu. Dispersion à 95% réalisée sur 50 simulations. N en abscisse. Gauche: simulation sans interactions. Droite: simulation avec interactions.

5 Perspectives

5.1 Enrichissement dans le modèle discret

Estimation de l'erreur

Les inégalités de concentration obtenues dans la seconde partie de cette thèse peuvent être appliquées à une large classe d'espaces fonctionnels \mathcal{F} . En notant que l'on peut lier les données disponibles au processus Z_t^N comme étant des intégrales de ce processus sur des domaines particuliers. On peut lier nos inégalités de concentration aux estimations réalisées dans le premier chapitre.

En prenant pour ensemble \mathcal{F} les indicatrices des triangles dans le diagramme de Lexis, indexées par t, x dans $\{0, 1, \dots, T\} \times \{0, 1, \dots, A_{max}\}$, on peut montrer que l'entropie métrique de \mathcal{F} est finie (tout simplement parce que le nombre d'indicatrices est fini).

Ainsi en théorie nous devrions être capable de quantifier l'erreur entre notre estimation et le taux μ . Il nous reste cependant à comprendre comment enrichir l'algorithme pour calculer de front l'estimation et une estimation de la variance.

Test de dépendance temporelle

Une autre question intéressante serait de chercher à adapter notre test sur la dépendance temporelle au cadre discret. Cela a en effet un réel intérêt pratique puisqu'il est nécessaire de savoir si les tables de mortalités changent dans le temps parce que le taux sous jacent change ou simplement à cause de l'erreur d'estimation.

Il est assez clair que ce point ne pourra se faire qu'après avoir traité le point sur les estimations de l'erreur dans le modèle discret.

5.2 Enrichissement dans le modèle stochastique

Affiner les inégalités de concentration

Une vision attentive de la preuve des inégalités de concentration que l'on obtient nous invite à penser que nous pourrions obtenir des bornes plus fines. Plus précisément, l'article de François Bolley [7] nous invite à penser que nous devrions avoir une borne de type

$$e^{-(x+1)\ln(x+1)+x}$$

au lieu d'une borne exponentielle. Pour cela nous devons travailler avec d'autres normes d'Orlicz introduites dans le chapitre 2, et utiliser les résultats récents dans [68].

Ce raffinement permettrait d'avoir des facteurs logarithmes à une puissance plus faible dans les bornes supérieures des estimateurs adaptatifs.

Calibration du test

Pour le moment la constante 0.2 dans le test est empirique. Nous aimerions ajouter à l'algorithme un moyen de calibrer cette constante selon les données.

Cela nous demande une meilleure compréhension de la quantité V^N que nous définissons.

5.3 Estimation de l'interaction

Notre hypothèse sur μ dans le chapitre des tests nous permet en réalité de pousser plus loin. Nous sommes en effet en mesure d'estimer

$$\mu(t, a) = \mu_0(a) + \int_0^\infty U(a, \alpha)g(t, \alpha)d\alpha.$$

On peut aussi estimer g . On fixe a , on choisit t_0 quelconque dans $[0, T]$ et alors on obtient

$$\bar{\mu}(t, a) = \mu(t, a) - \mu(t_0, a) = \mu_0(a) - \mu_0(a) + \int_0^\infty U(a, \alpha)[g(t, \alpha) - g(t_0, \alpha)]d\alpha.$$

On voit alors qu'on retrouve un problème de Fredholm typique des problèmes inverses. Plus précisément on obtient le problème suivant, avec pour inconnue h , pour tout a

$$f(x) = \int_0^\infty K(s, x)h(s)ds$$

avec $f(x) = \bar{\mu}(x, a)$, $K(s, x) = g(x, s) - g(t_0, s)$ et $h(s) = U(a, s)$. Il est alors possible de retrouver U . Le problème vient du besoin d'estimer g pour obtenir le noyau dans l'équation intégrale puis $\bar{\mu}$.

Pour la résolution de ce problème inverse il serait intéressant d'essayer d'autres méthodes d'estimation nonparamétrique, telle que la méthode par projection.

6 Composition de la thèse

Cette thèse se compose de quatre chapitres dont la rédaction repose sur les travaux suivants :

- [*Chapitre I*] *A new inference strategy for general population mortality tables*, avec M. Hoffmann et A. Boumezoued, soumis, voir [10]
- [*Chapitre II*] *Nonparametric inference of age-structured models in a large population limit*, avec M. Hoffmann et A. Boumezoued, en préparation.
- [*Chapitre III*] *Nonparametric inference of age-structured models in a large population limit with interactions, immigration and characteristics*, en préparation.
- [*Chapitre IV*] *Nonparametric test of time dependance of age-structured models in a large population limit*, en préparation.

CHAPTER 1

A NEW INFERENCE STRATEGY FOR GENERAL POPULATION MORTALITY TABLES

1 Introduction

General population mortality tables are crucial inputs for actuarial studies as they provide estimates of mortality rates for several age classes at several periods in time. Since the publication of the first mortality tables (attributed to John Graunt in 1662), the mathematical problem of providing consistent statistical estimates of mortality has fascinated mathematicians - for a brief history the reader is referred to the well documented dedicated part of the introduction of [21]. Two centuries later, there was a huge development of graphical formalizations of life trajectories within a population by Lexis (1875) and his contemporaries. These first demographers showed that it is crucial to address simultaneously two components: (1) Consider the fact that the death rate depends on both age and time (non-homogeneous setting) and (2) Understand the mortality rate as an aggregate quantity which depends on an underlying population dynamics.

Recently, several papers and publications paid attention to data quality issues in the way we usually build mortality tables, especially in relation with the 'discrete time' nature of population estimates provided by national censuses. To our knowledge, the first insights have been suggested by [60]; his conjecture was focused on the 1919 birth cohort for England & Wales, for which he suggested that errors occurred in the computation of mortality rates due to shocks in the births series. The ONS methodology has then been studied by [14] in several directions, who confirmed the conjecture by [60] and proposed an approach to illustrate and correct mortality tables, applied to the data for England & Wales; the *Convexity Adjustment Ratio* introduced in their work has then been adapted by [9] who focused on the Human Mortality Database (HMD) - which provides mortality tables for more than 30 countries and regions worldwide - and showed that these anomalies are universal while using the 'population dynamics' point of view to properly define mortality estimates. To build new mortality tables for several countries, a link with the Human Fertility Database (HFD, the HMD counterpart for fertility) has been made to correct such errors in a systematic way.

However, all precedent contributions did not succeed to introduce a proper mathematical setting for computing mortality rates based on information extracted from censuses. In this paper, we aim

at performing a first step in this direction by deriving an inference strategy from a deterministic population dynamics model. The derivation of a consistent theory in the stochastic setting is in parallel provided in a companion theoretical paper, see [11].

The main difficulty in establishing a consistent theory to estimate mortality rates lies in points (1) and (2) mentioned above, which can be summarized as follows: inferring an age and time dependent mortality rate based on a population dynamics model. In the literature, we argue that each point is treated separately.

The inference of a time dependent death rate also depending on a time-dependent covariate (possibly age), which relates to point (1), has been addressed from a non-parametric perspective by [3], [20], [40], [52], [55], [12], [19]. From [40], *"One way of understanding the difficulties in establishing an Aalen theory in the Lexis diagram is that although the diagram is two-dimensional, all movements are in the same direction (slope 1) and in the fully non-parametric model the diagram disintegrates into a continuum of life lines of slope 1 with freely varying intensities across lines. The cumulation trick from Aalen's estimator (generalizing ordinary empirical distribution functions and Kaplan & Meier's (1958) non-parametric empirical distribution function from censored data) does not help us here."* This explains why data aggregation and smoothing is required to derive an estimate with two crossing dimensions, age and time.

On the other side, the inference of an age-dependent death rate in an homogeneous birth-death model (or similar) - point (2) - has been addressed by [18], [35]. To our knowledge, no statistical method deals with the usual problem faced by demographers related to the construction of a mortality table based on population estimates and death counts.

In this paper, we rely on a deterministic age-structured population model and derive exact formulas in the so-called Lexis diagram, allowing to build new and improved mortality estimates. The inference problem is summarized as follows:

- The death rate depends on both age and time and is to be estimated,
- The population evolves as an age-structured and time inhomogeneous birth-death process,
- The following observables are available in the Lexis diagram:
 - The number of individuals in each one-year age-class, assumed to be recorded at each beginning of year,
 - The number of deaths in annual Lexis triangles,
 - The number of births, available each month (or more generally at some intra-year frequency).

Note that the practical availability of annual population estimates as well as death counts in the Lexis triangle can be achieved according to the Human Mortality Database, whereas the Human Fertility Database is a public source providing in particular number of births by months for several countries. Such population, death and fertility data allows at this date the method proposed in this paper to be applied to around 10 countries. For other countries, the data (especially number of births by month) has to be reached by means of national institutes.

The paper is organized as follows. In Section 2, we present the non-homogeneous birth-death model and derive the inference strategy - the related interpretations and link with existing estimators is discussed in Subsection 2.6. In Section 3, we compute mortality tables according to our method and compare it to those obtained by the usual formulas. The paper ends with some concluding remarks in Section 4.

2 Model and inference strategy

2.1 Non-homogeneous birth-death dynamics

Let us denote by $\mu(a, t)$ the mortality rate at exact age $a \in \mathbb{R}_+ = [0, \infty)$ and exact time $t \in \mathbb{R}_+$, with an arbitrary time origin - let us also denote by $g(a, t)$ the population density at (a, t) , a non-negative real value. In its core definition, the death rate drives the number of living in a closed population. Formally, consider $g(0, \nu)$ the newborn at (exact) time ν (starting number in the cohort born at time ν), then the survivors at some age $a > 0$ in the cohort writes

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

Changing variables to represent $g(a, t)$, and differentiating by age and time, leads to the transport component of the so-called McKendrick-Von Foerster equation (see [53] and [66]):

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

with notation $\partial_a \equiv \partial/\partial a$. Clearly, at this stage, the population dynamics of $g(a, t)$ is not fully specified as the future path of $g(a, t)$ depends on the quantity $g(0, t - a)$. The McKendrick-Von Foerster specifies how births are given in the (asexual) population, based on a birth rate $b(a, t)$, as

$$\text{for each time } \nu > 0, \quad g(0, \nu) = \int_0^\infty g(a, \nu)b(a, \nu)da.$$

That is simply, the newborn at each time is given by the total number of birth from all parents alive at the same time.

2.2 Observables in the Lexis diagram

We work here in the Lexis diagram - that is we study lifelines in the time \times age coordinates. In an ideal demographic world, two kinds of population estimates are recorded in the one-year age \times time square:

- Population at exact time t , with age x at its last birthday:

$$P(x, t) = \int_x^{x+1} g(a, t)da. \quad (1.2)$$

- Individuals who attained exact age x during the year $[t, t + 1)$:

$$N(x, t) = \int_t^{t+1} g(x, s)ds.$$

An illustration of population estimates $P(x, t)$ for the French population extracted from the Human Mortality Database is given in Figure 1.1. This can be analysed in the light of a Lexis diagram in several directions. First, the diagonal effects appear clearly showing that generations (or cohorts) are not equally represented: as an example, the generations born between around 1915 and 1920 are less represented (World War I), whereas the generations born after around 1970 are highly represented (Baby Boom). In this work, the impact of the discrepancy between birth

patterns from one year to the next is of interest, as it introduces some bias in the classical formulas used in practice for death rate estimation.

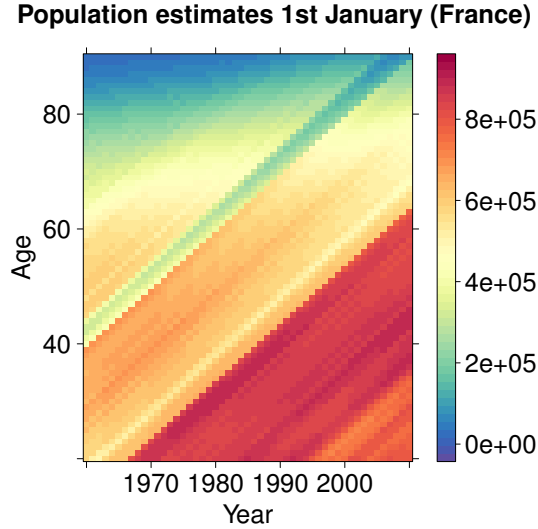


Figure 1.1: Population estimates for France by year for one-year age classes extracted from the Human Mortality Database

Also, the number of deaths are provided on the upper and lower triangles of the Lexis diagram, as defined below.

Definition 1.1. The upper (U) and lower (L) triangles for each age range x and observation year t are the age \times times sets defined by

$$T_U(x, t) = \{(a, s) : a \in [x, x + 1) \text{ and } s \in [t, t - x + a)\}, \quad (1.3)$$

and

$$T_L(x, t) = \{(a, s) : a \in [x, x + 1) \text{ and } s \in [t - x + a, t + 1)\}. \quad (1.4)$$

Based on this definition, the number of death in the Lexis triangles can be written

$$D_U(x, t) = \iint_{T_U(x, t)} \mu(a, s)g(a, s)dads \text{ and } D_L(x, t) = \iint_{T_L(x, t)} \mu(a, s)g(a, s)dads. \quad (1.5)$$

An illustration of death counts in the Lexis triangles (x, t) for the French population extracted from the Human Mortality Database is represented in Figure 1.2. Variations in number of deaths are closely linked to those of the underlying exposure (Figure 1.1) but also to the death rate itself, to be estimated.

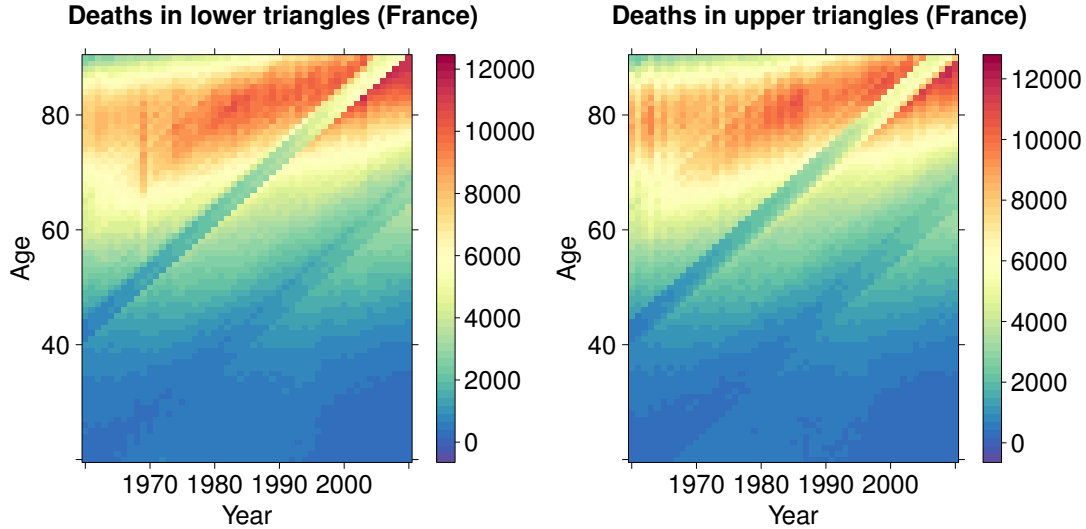


Figure 1.2: Death counts in Lexis triangles extracted from the Human Mortality Database

Assuming that the population is closed, the following fundamental relations appear (which can be proved by integration by parts):

$$\begin{aligned} N(x+1, t) &= P(x, t) - D_U(x, t), \\ P(x, t+1) &= N(x, t) - D_L(x, t). \end{aligned} \tag{1.6}$$

The assumption of closed-population is further discussed in Subsection 2.6.

In addition to population estimates and death counts, as analyzed by [14] and [9], we aim at including birth counts by month in the inference process - these can be extracted from the Human Fertility Database for a variety of countries. The dynamics of number of births by month in France is illustrated in Figure 1.3. The interpretation of this dynamics can be linked to that of Figures 1.1 (population estimates, see (1.2)) and 1.2 (death counts in Lexis triangles, as defined in (1.5)). Indeed, a similar information arises as the number of births are low in the period 1915-1920, which explains in particular the diagonal effect in Figure 1.1. Even more importantly, the dynamics at the monthly scale gives insight on what happens inside each year, then can be used to assess how the population is distributed inside a given age band. This is of great interest as the population distribution appears classically in the form of an 'exposure-to-risk', and more precisely the formulas we exhibit in order to estimate the death rate rely explicitly on the births distribution - as such, number of births by month are the key inputs for the inference strategy proposed here as it refines standard annual estimates. This is developed in the following.

2.3 Death rate inference

When two time-dependent dimensions are involved (here age and calendar time), the natural generalization of classical non-parametric estimates of the death rate is not direct (see again the discussion in [40]), therefore smoothing is required - see e.g. [52] and [55] for the analysis of such two dimensional kernel estimator based on continuous observation. Unfortunately, for building national mortality tables one does not observe continuously the living population (only possibly

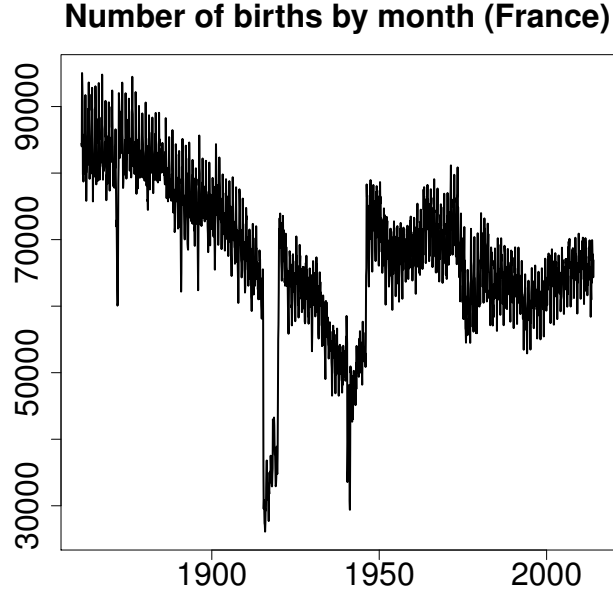


Figure 1.3: Number of birth by month extracted from the Human Fertility Database

the date of death through death certificates), therefore standard kernel smoothing techniques are neither applicable here. This leads to define some geometry on which the death rate is assumed to be piecewise constant, which allows to use aggregate information by year and age-class to derive (approximate) estimators.

In the classical demographic and actuarial practice, it is considered two versions of general population mortality tables: period and cohort. We propose here a brief discussion of these two versions and refer the reader to [9] for more details (and a study dedicated to period mortality tables). The two versions are illustrated in Figure 1.4.

- The period table provides death rate estimates based on the assumption that it is piecewise constant on squares in the Lexis diagram; each square (x, t) is equal to the region $T_U(x, t) \cup T_L(x, t)$, where the Lexis triangles T_U and T_L have been defined in Equations (1.3) and (1.4). The key advantage of period tables is that they provide an estimate of death rate by using information of a single year; the related drawback is that two generations (cohorts) are merged for a given death rate at (x, t) : the lifelines crossing the triangle $T_L(x, t)$ are born in year $t - x$, whereas those crossing $T_U(x, t)$ are born in year $t - x - 1$. This way, the period tables do not strictly reflect the mortality of single cohorts.
- The cohort table is based on the assumption that the death rate is constant on parallelograms $T_L(x, t) \cup T_U(x, t + 1)$, with the advantage that a given death rate at (x, t) relates to lifelines arising from a single cohort: that of people born in year $t - x$. However, the information provided by this death rate reflects conditions of the two consecutive years t and $t + 1$, as illustrated in Figure 1.4.

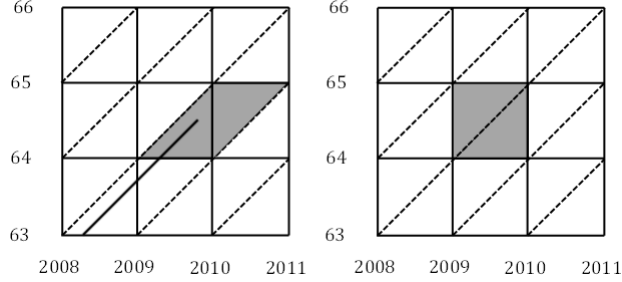


Figure 1.4: Population used (in grey) for the computation of the cohort death rate (left) and period death rate (right) for age 64 and year 2009.

Overall, period and cohort tables provide complementary information and their use is driven by the underlying objective. In this paper, we illustrate our method on the computation of triangle-based mortality tables, which generalize period and cohort mortality tables in a natural way as the death rate is assumed to be piecewise constant on Lexis triangles, instead of squares of parallelograms. This will allow us to draw analyses at a more granular scale compared to the two versions available in practice.

2.4 Main result

In the derivation of the inference formulas, we assume the death rate to be piecewise constant on Lexis triangles:

Assumption 1.1. *The death rate is piecewise constant on Lexis triangles, that is for each integer x and t ,*

$$\begin{aligned} \forall(a, s) \in T_L(x, t), \mu(a, s) &= \mu_L(x, t), \\ \forall(a, s) \in T_U(x, t), \mu(a, s) &= \mu_U(x, t). \end{aligned}$$

From the transport component described in Equation (1.1), for any upper or lower triangle which we denote T , and on which the death rate is constant equal to μ_T , it follows that:

$$\iint_T (\partial_a + \partial_s)g = - \iint_T \mu g = -\mu_T \iint_T g.$$

As the left hand side is the opposite of the number of deaths as introduced in Equation (1.5), it follows from the previous equation that the death rate can be written as the ratio

$$\mu_L(x, t) = \frac{D_L(x, t)}{E_L(x, t)} \text{ and } \mu_U(x, t) = \frac{D_U(x, t)}{E_U(x, t)},$$

where

$$E_L(x, t) = \iint_{T_L(x, t)} g(a, s)dad s \text{ and } E_U(x, t) = \iint_{T_U(x, t)} g(a, s)dad s$$

are the so-called 'exposures-to-risk' in the lower and upper triangle respectively.

Now, the number of deaths in Lexis triangles being observed (as provided by the Human Mortality Database), it remains to appropriately compute the exposure-to-risk. In the literature

dedicated to longevity studies, this quantity is approximated by annual observables, see e.g. [59] section 2.3.4 as well as references therein. The recent update of the Human Mortality Database methodology allowing to include monthly data is further discussed in Subsection 2.6. The standard annual approximation can be illustrated for period tables (see Subsection 2.3) for which the exposure-to-risk writes

$$E(x, t) = \int_t^{t+1} \int_x^{x+1} g(a, s) da ds = \int_t^{t+1} P(x, s) ds.$$

A possible approximation is therefore given by the trapezoid rule as

$$E(x, t) \approx \frac{1}{2} [P(x, t + 1) + P(x + 1, t)].$$

On the other hand, the exposure-to-risk can also be written as $E(x, t) = \int_x^{x+1} N(a, t) da$ and then approximated by $\frac{1}{2} [N(x, t) + N(x + 1, t)] = \frac{1}{2} [P(x, t) + P(x + 1, t)] + \frac{1}{2} [D_L(x, t) - D_U(x, t)]$, which leads to another possible approximation. Note that the Version 5 estimates of the Human Mortality Database rely on a demographic reasoning leading to an approximation in between the two previous ones - see the analysis in [9] for more details.

Overall, classical approximations have the advantage of being based on observables only, leading to a closed-form for the death rate estimate. The counterpart of this feature is that the validity of the underlying approximation can be put into question in periods in which the population curve (as $s \mapsto P(s, x)$) appears far from linear in a year.

We now detail the recursive and implicit scheme for computing death rate estimates, based on equations linking the death rate with the observables in the Lexis diagram introduced in subsection 2.2. Before stating the main result, we introduce two key quantities: first, the Laplace transform of the random variable 'date of birth in year y ', introduced as:

$$L_y(\theta) = \frac{\int_0^1 g(0, y + v) \exp(-\theta v) dv}{\int_0^1 g(0, y + v) dv},$$

and second, the cumulative gain in longevity at age x last birthday within the same cohort born in year $t - x$ (a diagonal in the Lexis diagram), that is between those born at exact time $t - x$ and those born at the end of the year $[t - x, t - x + 1)$, defined by:

$$H(x, t) = \sum_{y=0}^{x-1} \mu_U(y, t - x + y + 1) - \mu_L(y, t - x + y), \quad x \in \mathbb{N}^*. \quad (1.7)$$

The result at the core of the inference strategy is stated below:

Proposition 1.2. *Consider the transport Equation (1.1). Under Assumption 1.1, the following equalities hold:*

$$\exp(-\mu_L(x, t)) L_{t-x}(H(x, t) - \mu_L(x, t)) = \left(1 - \frac{D_L(x, t)}{N(x, t)}\right) L_{t-x}(H(x, t)), \quad (1.8)$$

and

$$\begin{aligned} & L_{t-x-1}(H(x, t - 1) - \mu_L(x, t - 1)) \\ &= \left(1 + \frac{D_U(x, t)}{N(x + 1, t)}\right) L_{t-x-1}(H(x, t - 1) - \mu_L(x, t - 1) + \mu_U(x, t)). \end{aligned} \quad (1.9)$$

The proof is detailed in the next part, along with a detailed discussion in Subsection 2.6. The resulting algorithm is described in Section 3.

2.5 Proof of Proposition 1.2

To prove (1.8), let us first focus on the exposure-to-risk in the lower triangle

$$E_L(x, t) = \int_t^{t+1} \int_x^{x+s-t} g(a, s) da ds.$$

According to the transport equation (1.1), the population density in the lower triangle can be expressed as

$$\begin{aligned} g(a, s) &= g(x, s - a + x) \exp\left(-\int_x^a \mu(u, s - a + u) du\right) \\ &= g(x, s - a + x) \exp(-(a - x)\mu_L(x, t)). \end{aligned}$$

where the last equality comes from the assumption of a piecewise constant death rate on Lexis triangles. By the change of variable $v \leftarrow s - a + x - t$, the exposure-to-risk can then be rewritten as

$$\begin{aligned} E_L(x, t) &= \int_t^{t+1} \int_x^{x+s-t} g(x, s - a + x) \exp(-(a - x)\mu_L(x, t)) da ds \\ &= \int_0^1 \int_{t+v}^{t+1} g(x, t + v) \exp(-(s - v - t)\mu_L(x, t)) ds dv. \end{aligned}$$

By straightforward computation, one finally gets the following expression for the exposure-to-risk in the lower triangle:

$$E_L(x, t) = \int_0^1 g(x, t + v) \frac{1 - \exp(-(v - 1)\mu_L(x, t))}{\mu_L(x, t)} dv. \quad (1.10)$$

Also note that $D_L(x, t) = \mu_L(x, t)E_L(x, t) = \int_0^1 g(x, t + v) (1 - \exp(-(v - 1)\mu_L(x, t))) dv$ and $N(x, t) = \int_0^1 g(x, t + v) dv$ so that

$$N(x, t) - D_L(x, t) = \int_0^1 g(x, t + v) \exp(-(v - 1)\mu_L(x, t)) dv.$$

Let us now derive the population density at exact age x , for any $v \in [0, 1)$,

$$\begin{aligned} g(x, t + v) &= g(0, t - x + v) \exp\left(-\int_0^x \mu(u, t - x + v + u) du\right) \\ &= g(0, t - x + v) \exp\left(-\sum_{y=0}^{x-1} \int_y^{y+1} \mu(u, t - x + v + u) du\right) \\ &= g(0, t - x + v) \exp\left(-\sum_{y=0}^{x-1} \int_y^{y+1-v} \mu(u, t - x + v + u) du - \sum_{y=0}^{x-1} \int_{y+1-v}^{y+1} \mu(u, t - x + v + u) du\right) \\ &= g(0, t - x + v) \exp\left(-(1 - v) \sum_{y=0}^{x-1} \mu_L(y, t - x + y) - v \sum_{y=0}^{x-1} \mu_U(y, t - x + y + 1)\right) \\ &= S(x, t)g(0, t - x + v) \exp(-vH(x, t)), \end{aligned} \quad (1.11)$$

where $S(x, t) = \exp\left(-\sum_{y=0}^{x-1} \mu_L(y, t-x+y)\right)$ is the survival function at age x for individuals which attained (exact) age x at (exact) time t , and where the cumulative death rate differential within the cohort $H(x, t)$ has been introduced in Equation (1.7). Let us now combine the previous result to get

$$N(x, t) - D_L(x, t) = S(x, t)e^{-\mu_L(x, t)} \int_0^1 g(0, t-x+v)e^{-v(H(x, t)-\mu_L(x, t))} dv,$$

and finally, let us apply some renormalization of the right hand side, first by $N(x, t)$ and second by $\int_0^1 g(0, t-x+v)dv$ to get the following formula, which reduces to Equation (1.8):

$$\left(1 - \frac{D_L(x, t)}{N(x, t)}\right) \frac{1}{\int_0^1 g(0, t-x+v)dv} = \frac{S(x, t)e^{-\mu_L(x, t)} \int_0^1 \tilde{g}(0, t-x+v)e^{-v(H(x, t)-\mu_L(x, t))} dv}{S(x, t) \int_0^1 \tilde{g}(0, t-x+v)e^{-vH(x, t)} dv}.$$

where $\tilde{g}(0, t-x+v) = \frac{g(0, t-x+v)}{\int_0^1 g(0, t-x+v)dv}$.

The proof of (1.9) follows similarly. Since $E_U(x, t) = \int_t^{t+1} \int_{x+s-t}^{x+1} g(a, s)dad s$ and $g(a, s) = g(x+1, s+x+1-a) \exp((x+1-a)\mu_U(x, t))$, then by changing variables, one gets $E_U(x, t) = \int_0^1 g(x+1, t+v) \frac{\exp(v\mu_U(x, t))-1}{\mu_U(x, t)} dv$, so that

$$N(x+1, t) + D_U(x, t) = \int_0^1 g(x+1, t+v) \exp(v\mu_U(x, t)) dv.$$

Then as $g(x+1, t+v) = g(0, t-x-1+v)S(x+1, t) \exp(-vH(x+1, t))$, one finally obtains

$$\left(1 + \frac{D_U(x, t)}{N(x+1, t)}\right) L_{t-x-1}(H(x+1, t)) = L_{t-x-1}(H(x+1, t) - \mu_U(x, t)),$$

which leads to the result, as the following equality is verified from the definition in Equation (1.7):

$$H(x+1, t) = H(x, t-1) + \mu_U(x, t) - \mu_L(x, t-1).$$

2.6 Discussion

Exposure-to-risk interpretation. The equality (1.10) can be interpreted as follows: for each individual attaining exact age x at time $t+v$, its contribution to the exposure-to-risk in the lower triangle is $\frac{1-\exp((v-1)\mu_L(x, t))}{\mu_L(x, t)}$, which depends on the unobserved death rate to be estimated. This contrasts with classical methods which compute approximations of the exposure-to-risk based on observables. At first order, assuming $\mu_L(x, t) \ll 1$, one recovers that $E_L(x, t) \approx \int_0^1 g(x, t+v)(1-v)dv$ and the related interpretation that the contribution of any individual which attained exact age x at time $t+v$ and living through the lower triangle is simply $1-v$ as it can be measured in the Lexis diagram.

Biased birthday density. The formula derived in (1.11) shows that the birthdays density at some age x is exponentially biased through $H(x, t)$ compared to the initial birthdays distribution (at age zero). This is true in general in the triangle model for the piecewise constant death rate, as well as in the period table for which the cumulative death rate difference matrix reduces to $H(x, t) = \sum_{y=0}^{x-1} \mu(y, t-x+y+1) - \mu(y, t-x+y)$ where $\mu(x, t)$ denotes the period death rate for

the square (x, t) . Moreover, as one expects in general some mortality improvement over the years, age being fixed, one may be interested in interpreting the case $H(x, t) < 0$ - in this situation, one sees that the initial birthdays distribution is distorted to the highest birthdays (youngest individuals) in the cohort as age goes. This demonstrates how even in a discrete time specification, individuals in the same cohort may experience different death rates over life (more precisely they pass through the same rates but do not 'spend the same time' in each triangle or square, so that the resulting survival functions are different). However, it is interesting to note that for the cohort table, which by definition assumes that $\mu(y, t - x + y + 1) = \mu(y, t - x + y)$, the H matrix vanishes, so that the initial birthdays distribution perfectly propagates towards highest ages. This justifies why period tables (as provided e.g. by the Human Mortality Database) are mainly concerned with abnormal isolated cohort effects, whereas cohort tables show a more reasonable structure - see the discussion in [9].

Closed population assumption. Due to the renormalization in the final result (1.8), the death rate relates to the closest annual population estimate; therefore, the assumption that the population is closed is only local in terms of population count, as the population estimate N may include population flow effects. Also, the assumption of a closed population implies here that the birthdays distribution at some age is obtained as a transformation of the initial birth distribution - to this extent the assumption applies globally in each cohort.

Link with estimates of the Human Mortality Database. Moreover, it is worth mentioning that at the time of writing, the Human Mortality Database released an update on February 2018, including in particular a revision of exposure calculation based on monthly birth counts. We now make the link with both the new Version 6 and the old Version 5 of the HMD Methods Protocol.

From (1.10), it can be shown by performing a first order expansion in $\mu_L(x, t)$ that

$$E_L(x, t) \approx E_L^{(1)}(x, t) - \mu_L(x, t)E_L^{(2)}(x, t),$$

where

$$E_L^{(1)}(x, t) := N(x, t) \left(1 + \frac{L'_{t-x}(H(x, t))}{L_{t-x}(H(x, t))} \right),$$

and

$$E_L^{(2)}(x, t) = \frac{1}{2}N(x, t) \left[1 + \frac{2L'_{t-x}(H(x, t)) + L''_{t-x}(H(x, t))}{L_{t-x}(H(x, t))} \right].$$

Let us denote by B_{t-x} the random variable with values in $[0, 1]$ that represents the time of birth in the year $t - x$, with mean $m_{t-x} := \mathbb{E}[B_{t-x}]$ and variance $\sigma_{t-x}^2 := \text{Var}(B_{t-x})$.

Under the assumption $H(x, t) = 0$, that is no mortality improvement between the youngest and oldest individuals within the same cohort, one can write

$$E(x, t) \approx N(x, t)(1 - m_{t-x}) - \frac{1}{2}\mu_L(x, t)N(x, t)((1 - m_{t-x})^2 + \sigma_{t-x}^2).$$

Note again that the assumption $H(x, t) = 0$ is not consistent with the piecewise constant death rate assumption on Lexis triangles, nor with the framework underlying the period tables.

Now, if one uses (1.6) and replaces $\mu_L(x, t) = \frac{D_L(x, t)}{E_L(x, t)}$ by its zero order approximation

$$\mu_L(x, t) \approx \frac{D_L(x, t)}{N(x, t)(1 - m_{t-x})},$$

one finally obtains the formula (51) displayed in the Version 6 in the HMD methods protocol:

$$E_L(x, t) \approx P(x, t + 1)(1 - m_{t-x}) + \frac{D_L(x, t)}{2(1 - m_{t-x})} ((1 - m_{t-x})^2 - \sigma_{t-x}^2).$$

Finally, if one assumes births to be uniformly distributed, then $m_{t-x} = \frac{1}{2}$ and $\sigma_{t-x}^2 = \frac{1}{12}$ so that the classical formula in Version 5 methods protocol is recovered (see Appendix E therein for the original derivation):

$$E_L(x, t) \approx \frac{1}{2}P(x, t + 1) + \frac{1}{6}D_L(x, t).$$

3 Numerical results

Based on Proposition 1.2, one can exhibit a recursive and implicit scheme for computing the death rates, as described below.

Algorithm 1. *For age x starting at zero:*

- (i) *Solve Equation (1.8) to estimate the death rate $\mu_L(x, t)$ for the lower triangles of any available year t ,*
- (ii) *Then based on the previous estimates, solve Equation (1.9) to infer the death rate $\mu_U(x, t)$ for the upper triangles of any available year t ,*
- (ii) *Compute the value for $H(x + 1, t) = H(x, t - 1) + \mu_U(x, t) - \mu_L(x, t - 1)$ for possible years t , let $x \leftarrow x + 1$ and go to step (i) .*

Remark 1. *Note that the method is past dependent - this is natural as any change in past death rates modify the future birthdays distribution in the cohort. This way, any revision of past mortality rates, which may occur in practice, requires the re-use of the methodology which will provide a new mortality table.*

In Figures 1.5 to 1.9, we depict the death rate estimates obtained with the method developed in this paper applied to French data sourced from the Human Mortality Database (annual population estimates, Figure 1.1 and number of deaths in Lexis triangles, Figure 1.2) and the Human Fertility Database (births by months, Figure 1.3). The number of births by month are used to approximate the Laplace transform of the birthdays distribution which is used in the inference process.

The results are compared with estimates as they would be classically computed based on annual observables (see [69] and [9] for further details):

$$\widehat{\mu}_L(x, t) = \frac{D_L(x, t)}{\frac{1}{2}N(x, t) - \frac{1}{3}D_L(x, t)} \text{ and } \widehat{\mu}_U(x, t) = \frac{D_U(x, t)}{\frac{1}{2}N(x + 1, t) + \frac{1}{3}D_U(x, t)}.$$

Each figure includes on the right the ratio between the new and the old estimate, which helps quantify the differences between both. First, the ratio is for several age classes close to one, which indicates that the new estimate does not differ much from the classical one, in other words that the classical approximation is valid. However, one sees strong deviations for specific ages in time, and this translates over time and ages, so that it appears that the anomalies belong to specific generations. To assess this specificity, we depict in Figure 1.10 mortality improvement rates separated between upper and lower triangles as

$$\frac{\mu_L(x, t + 1) - \mu_L(x, t)}{\mu_L(x, t)} \text{ and } \frac{\mu_U(x, t + 1) - \mu_U(x, t)}{\mu_U(x, t)}.$$

Clearly, the isolated cohort effects disappear in the new mortality tables: mainly the diagonals around 1915 and 1920, and to a lower extent those born around 1940; note that this indeed corresponds to the shocks in birth numbers as illustrated in Figure 1.3, which confirms from a mathematical perspective the previous contributions by [60], [14] and [9].

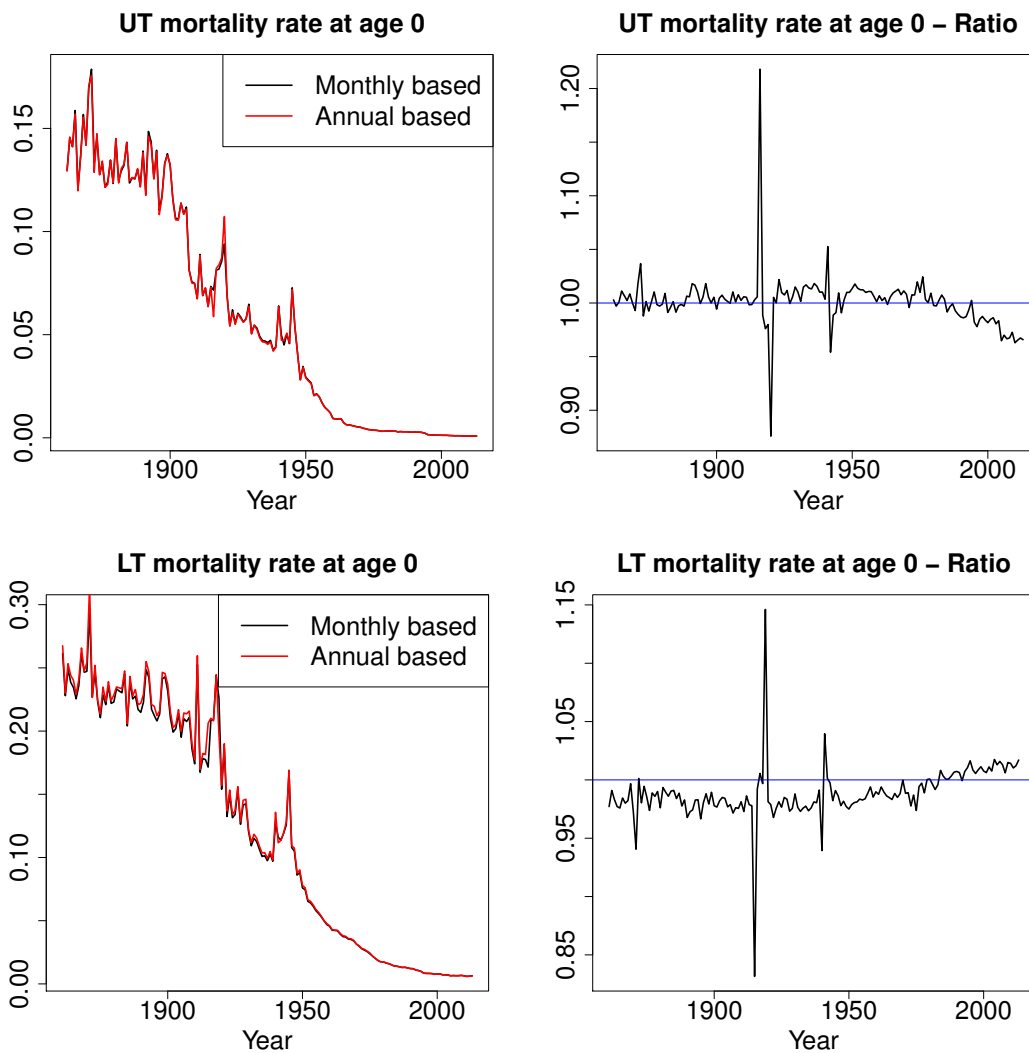


Figure 1.5: Left: death rates estimated based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: Upper triangle. Bottom: Lower triangle.

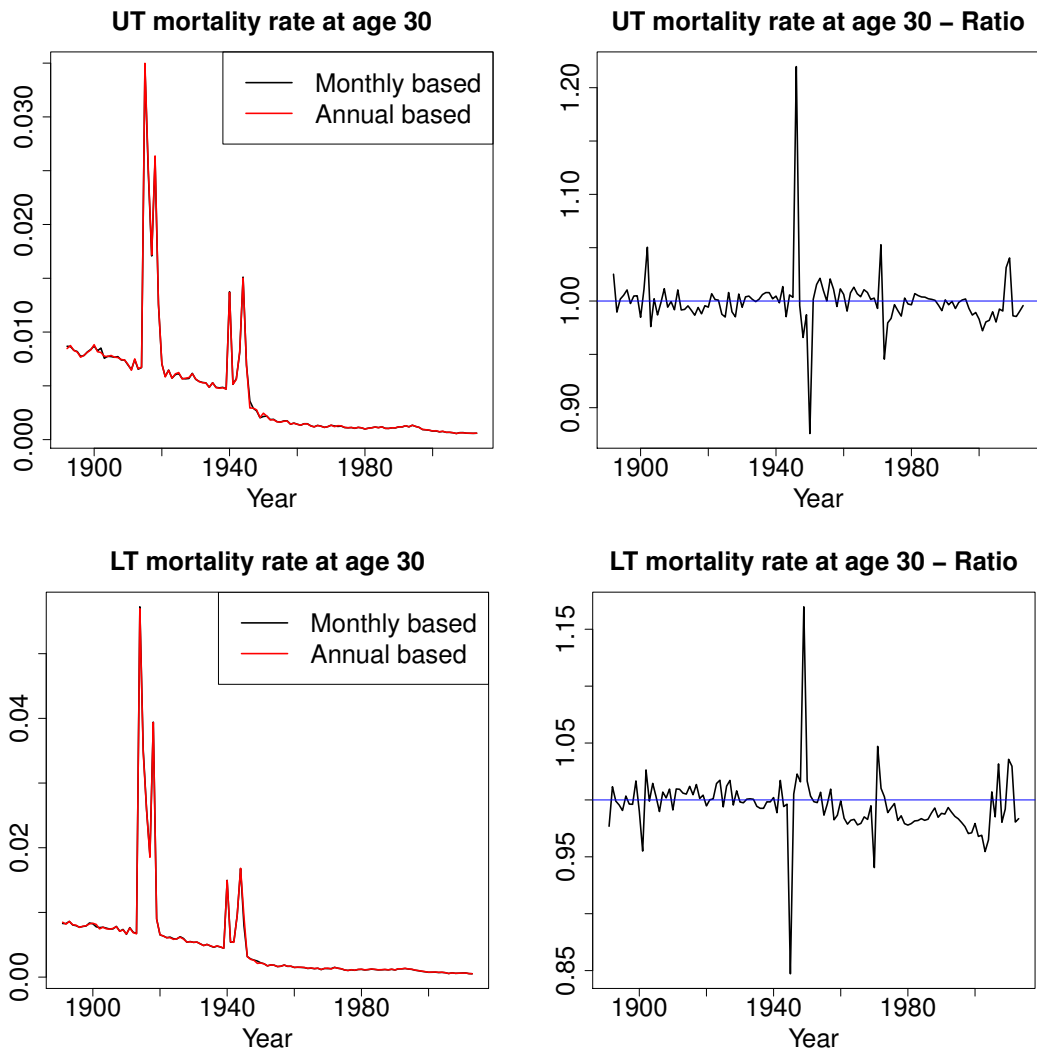


Figure 1.6: Left: death rates estimated based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: Upper triangle. Bottom: Lower triangle.

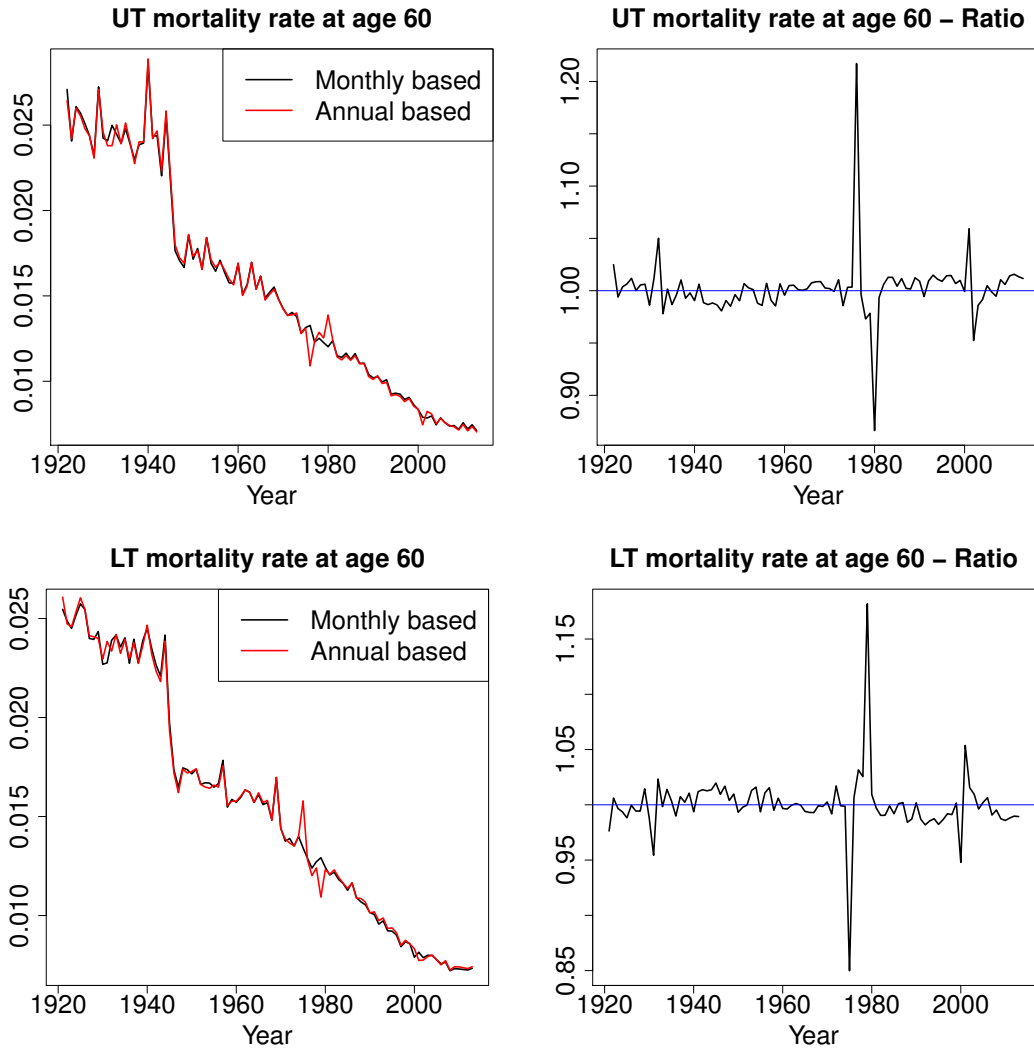


Figure 1.7: Left: death rates estimated based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: Upper triangle. Bottom: Lower triangle.

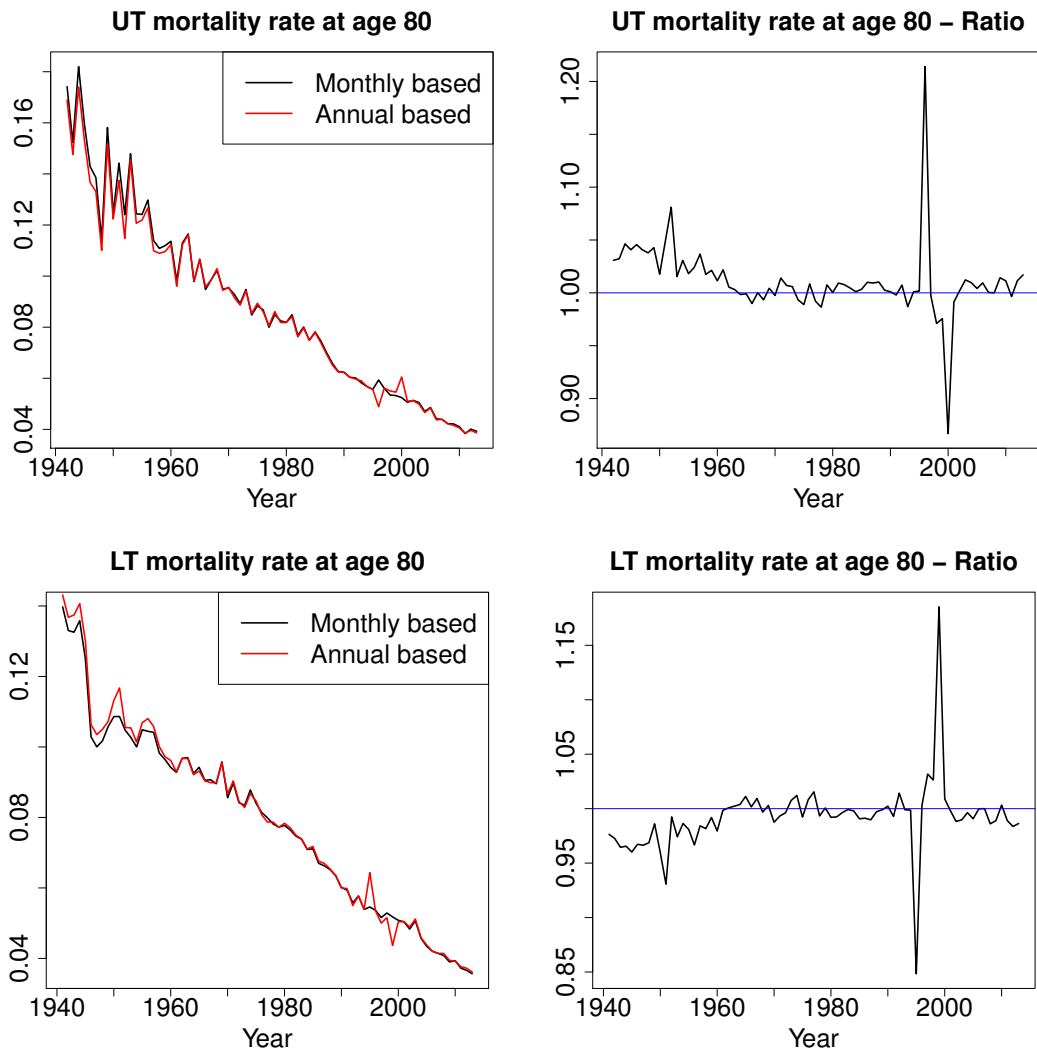


Figure 1.8: Left: death rates estimated based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: Upper triangle. Bottom: Lower triangle.

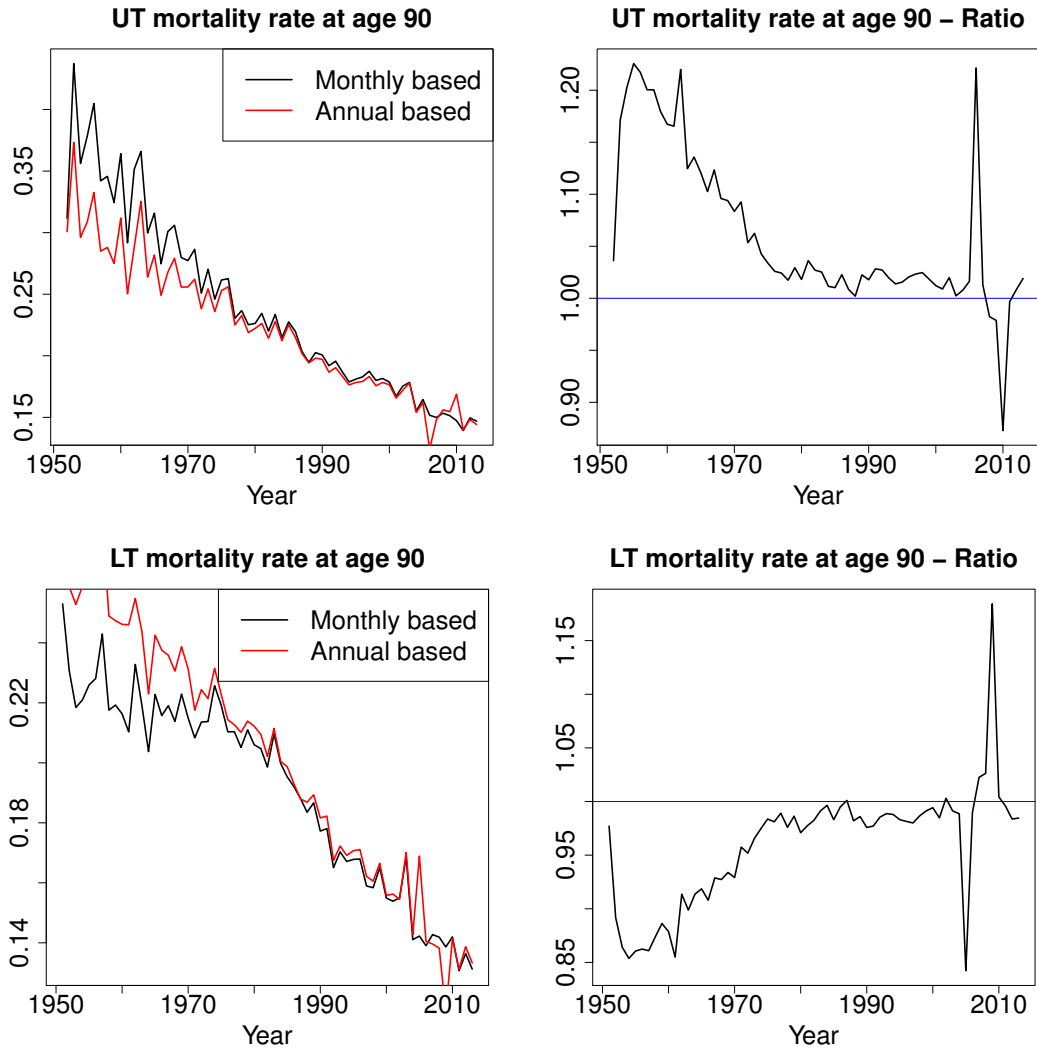


Figure 1.9: Left: death rates estimated based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: Upper triangle. Bottom: Lower triangle.

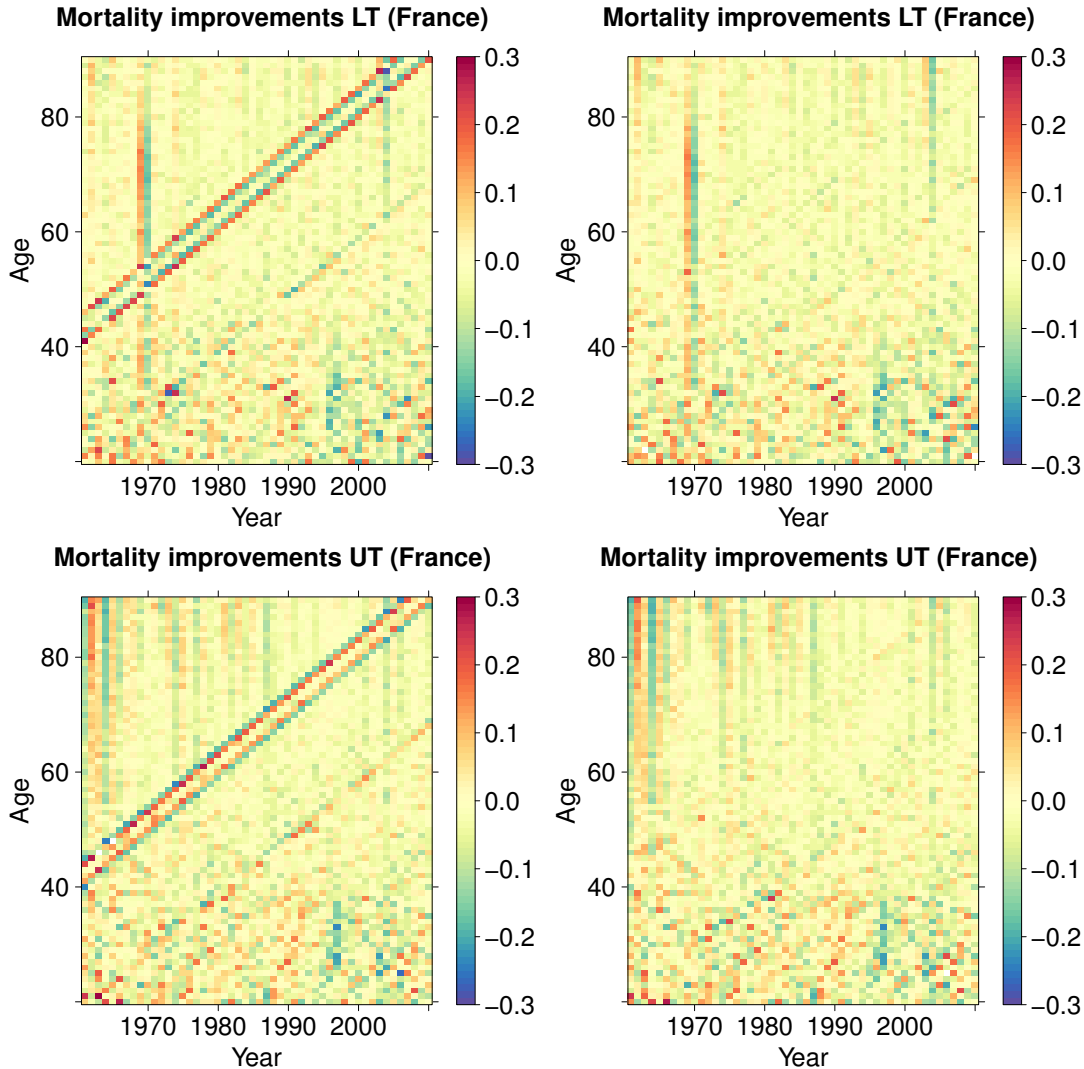


Figure 1.10: Left: mortality improvement rates based on the new inference method (in black), and compared to estimates using the standard method based on annual population records (in red). Right: ratio between new and old estimates. Top: improvement of mortality rates for the upper triangles. Bottom: improvements of mortality rates for the lower triangles.

4 Concluding remarks

In this paper, we proposed an inference strategy for general population mortality tables based on the derivation of formulas in the Lexis diagram, which relate the death rate with annual observables and the intra-year distribution of birthdays over ages. The method therefore uses monthly birth counts to refine classical mortality estimates. The new mortality tables show better features, including the fact that previous anomalies in the form of isolated cohort effects disappear, which

confirms from a mathematical perspective the previous contributions by [60], [14] and [9].

Several topics remain to be addressed to strengthen the methodology. First, it is of interest to account for population flows which may for several countries deform the closest population count, as well as distort the birthdays distribution over ages. Second, we emphasize that it is of importance to derive confidence intervals for the prediction, by going beyond the classical Poisson approximation to measure sampling risk. To this extent a stochastic population dynamics model is required, as well as the dedicated statistical framework.

CHAPTER 2

NONPARAMETRIC INFERENCE OF AGE-STRUCTURED MODELS IN A LARGE POPULATION LIMIT

1 Introduction

1.1 Setting

Suppose one wishes to recover a probability density g over the nonnegative real line $\mathbb{R}_+ = [0, \infty)$ from a N -sample a_1, \dots, a_N , where the a_i are not necessarily independent. If $Z^N = N^{-1} \sum_{i=1}^N \delta_{a_i}$ denotes the empirical distribution of the N -sample, designing a good statistical estimator of g usually requires a fine quantitative control of the fluctuations in the convergence

$$\int_{\mathbb{R}_+} \psi(a) Z^N(da) \rightarrow \int_0^\infty \psi(a) g(a) da \quad (2.1)$$

as N grows, for a large enough class of test functions ψ . Moreover, the performance of such a procedure depends on the smoothness properties of the function g , typically quantified by a smoothness parameter, like a (possibly fractional) number of derivatives in any reasonable sense and is usually unknown by the practitioner. For suitable ψ (possibly data-dependent), optimal estimators can be found provided good concentration inequalities are available for (2.1), following the broad guiding principle of Lepski's method [47, 28, 29] or other adaptive methods like model selection or wavelets, see for instance the comprehensive textbooks of Giné and Nickl [27] or Härdle *et al.* [32] or Tsybakov [63]. In this paper, we generalise the classical situation described above by adding a time variable. We investigate statistical inference of a time-evolving particle system governed by stochastic dynamics: for every $t \in [0, T]$, we observe the state of a population of (approximately) N particles, encoded by its empirical measure $Z^N = (Z_t^N)_{0 \leq t \leq T}$. Informally, Z^N is solution to a certain stochastic differential equation

$$\mathcal{H}_{b,\mu}^N(Z^N) = 0,$$

described in details in (2.7) below. The equation $\mathcal{H}_{b,\mu}^N$ is parametrised by two functional parameters b and μ and $Z_t^N(da)$ represents the state of a population structured in age $a \in \mathbb{R}_+$, alimanted by

a time-inhomogeneous fertility rate $b(t, a)$ and decimated by a mortality rate $\mu(t, a)$. Moreover, we are given an initial empirical age distribution Z_0^N at time $t = 0$. Under appropriate regularity conditions on b and μ and if Z_0^N is close enough to an initial limiting distribution g_0 , the following convergence

$$\int_0^T \int_{\mathbb{R}_+} \psi(t, a) Z_t^N(da) dt \rightarrow \int_0^T \int_0^\infty \psi(t, a) g(t, a) da dt \quad (2.2)$$

holds in a large population limit $N \rightarrow \infty$, where

$$\mathcal{H}_{b, \mu}(g) = 0, \quad (2.3)$$

and $\mathcal{H}_{b, \mu}$ is the McKendrick Von Foester transport equation [53, 66], detailed in (2.8) below and that reveals the interplay between the limiting solution g and the model parameter b and μ . This situation generalises (2.1) in a time-dependent framework.

Informally, our statistical problem takes the following form: estimate g or the parameters of the model b, μ from data Z^N in the limit $N \rightarrow \infty$. In this setting, it is crucial to understand: (i) the quantitative properties of the convergence (2.2) and in particular, how concentration inequalities can be obtained (with a view towards an adaptive estimation scheme in the idea of Lepski's principle) and (ii) what is the structure of (2.3) in terms of the smoothness properties of b, μ and g . In particular, the anisotropic smoothness of g viewed as a graph-manifold can benefit from $\mathcal{H}_{b, \mu}$ and lead to better approximation properties in certain directions along the characteristics of the transport.

1.2 Motivation

Of primary interest for us is human demography through the recent efforts and contributions for improving mortality estimates, see [14, 9, 10] among others and the references therein. In particular, the recent development of large human datasets like the Human Mortality Database (HMD) and Human Fertility Database (HFD) [34, 33] – in open access – allows one to process fertility and mortality data simultaneously, and subsequently addresses demographical issues such as the anomalies of cohort effects that have long fascinated demographers and actuaries [60, 14]. In this rejuvenated context, it becomes reasonable to study the estimation of population density or mortality rate in the enriched dynamical framework provided by birth-death particle systems that converge to the classical McKendrick Von Foester equation in a large population limit, and revisit classical studies like *e.g.* [40, 55] for statistical estimation of the death rate; see the detailed literature review in next section. In this setting, we consider the idealised model where we can observe the (renormalised) evolution of the state of the population Z_t^N continuously for $t \in [0, T]$, where $t = 0$ is the starting date for the observation of the population and $t = T$ a terminal time horizon, fixed once for all. We are interested in identifying or estimating the parameters of the model. Of major importance is the inhomogeneous death rate $\mu(t, a)$. In our framework, we cannot recover the birth rate since we are not given any genealogical input: mathematically, this simply expresses the lack of injectivity of the mapping $b \mapsto g$. Still, our observation enables us to identify the functions $(t, a) \mapsto g(t, a)$ and $(t, a) \mapsto \mu(t, a)$ in the limit $N \rightarrow \infty$ and establish a thorough nonparametric estimation program, in the methodology of adaptive minimax estimation.

1.3 Link with literature on death rate inference

The main difficulty in establishing a consistent theory to estimate mortality rates comes from two key points: (i) incorporate the fact that the death rate depends on both age and time (non-homogeneous setting) and (ii) use as observables the outcome of a stochastic population dynamics (birth-death process). In the literature, we argue that each point is treated separately. The inference of a time-dependent death rate also related to a time-dependent covariate (possibly age), which relates to the first point has been addressed from a nonparametric perspective by e.g. [3, 20, 40, 52, 55, 12, 19] and the references therein. From [40], *"One way of understanding the difficulties in establishing an Aalen theory in the Lexis diagram is that although the diagram is two-dimensional, all movements are in the same direction (slope 1) and in the fully non-parametric model the diagram disintegrates into a continuum of life lines of slope 1 with freely varying intensities across lines. The cumulation trick from Aalen's estimator (generalizing ordinary empirical distribution functions and Kaplan & Meier's (1958) nonparametric empirical distribution function from censored data) does not help us here."* On the other side, the inference of an age-dependent death rate in an homogeneous birth-death model (or similar) - oiuyr second point - has been addressed in [18, 22, 36] among others. To the best of our knowledge, no statistical method deals with the usual problem faced by demographers related to the inference of a time and age-dependent death rate table based on the observation of population dynamics. Note that in this paper, the observation of the population is assumed to be continuous over time, whereas in practice the information on population exposure is extracted from census (point observation); these practical considerations are discussed in a companion paper, see [10].

1.4 Results and organisation of the paper

In a first part of the paper, Section 2, we construct the microscopic model that describes the state of the population Z_t^N by means of a birth-death process characterised via a stochastic differential equation – given in (2.7) – driven by a random Poisson measure. We recall its convergence in a large population limit to the solution of the McKendrick Von Foester equation g based on classical results of [62, 26]. Our next step consists in quantifying the stability of the convergence $Z^N \rightarrow g$. To that end and anticipating the subsequent statistical analysis, we introduce two-distances:

$$\mathcal{W}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathbb{R}_+} w_2(t-a) f(t,a) (Z_t^N(da) - g(t,a) da)$$

and its integrated version

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathbb{R}_+} w_2(s-a) f(s,a) (Z_s^N(da) - g(s,a) da) ds,$$

where w_1 and w_2 are two bounded weight functions and \mathcal{F} a rich enough class of function with complexity measured in terms of entropy conditions. Note that formally $\mathcal{W}_{w_2}^N(\mathcal{F})_t$ is a degenerate version of $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t$. Taking $w_1 = w_2 = 1$ is reminiscent of the Wassertein-1 like distance if \mathcal{F} consists of 1-Lipschitz functions for instance. However, for the statistical analysis, we must be able to handle approximating kernels that do not have bounded Lipschitz norms, hence the presence of the weights w_1 and w_2 that can accomodate such kernels. The main result of this section, Theorem 2.6 states that under appropriate regularity conditions on b and μ , if $|w_2|_{1, \infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_0$ is of (small) order r_N , so are

$$|w_2|_{1, \infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_T \quad \text{and} \quad (|w_1|_{1, \infty} |w_2|_{1, \infty})^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T.$$

The rate of decay r_N possibly inflates by an order $N^{-1/2}$ and the result holds in terms of exponential decay of the fluctuation probabilities. The functional control $|\cdot|_{1,\infty} = (|\cdot|_1 |\cdot|_\infty)^{1/2}$ interpolates between L^1 and L^∞ -norms, and is sufficient to handle the behaviour of statistical kernels in an optimal way, since it can therefore be compared to the usual L^2 -norm that appears in variance terms. The concentration of $\mathcal{W}_{w_2}^N(\mathcal{F})_T$ expresses a kind of stability of the particle system from $t = 0$ to $t = T$, while the more intricate control of $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$ is crucial to control variance terms in bi-variate kernel estimators for the nonparametric estimation of $g(t, a)$ and $\mu(t, a)$. The proof relies on a combination of martingales techniques in the spirit of Tran [62], a central reference for the paper, combined with classical tools from concentration of processes indexed by functions under entropy controls, following for instance Ledoux-Talagrand [44].

In a second part, Section 3, we construct nonparametric estimators of $g(t, a)$ and $\mu(t, a)$ by means of kernel approximation: we consider estimators of the form

$$\widehat{g}_h^N(t, a) = K_h \star Z_t^N(a)$$

for $g(t, a)$, where \star denotes convolution and $K_h = h^{-1}K(h^{-1}\cdot)$, with $|K|_1 = 1$, is a kernel normalised in L^1 with bandwidth $h > 0$. It is noteworthy that for estimating the population density $g(t, a)$ at time t , the information Z_t^N is sufficient and we do not need the data $(Z_s^N, s \neq t)$. The situation is very different for estimating $\mu(t, a)$ the main parameter of interest. We construct a quotient estimator, inspired from a Nadaraya-Watson type procedure, and use

$$\widehat{\mu}_{h_1, h_2, h_3}^N(t, a) = \frac{(H_{h_1} \otimes K_{h_2} \circ \varphi) \star \Gamma^N(du, ds)}{\widehat{g}_{h_3}^N(t, a)} \quad (2.4)$$

where $\Gamma^N(du, ds)$ is the point process of the death occurrences in the population lifetime that can be extracted from Z^N and that converges to $\pi = \mu g$, see (2.18) in Section 3.2 for the details. In (2.4), we consider a bivariate kernel $H \otimes K$ with bandwidth (h_1, h_2) and $\varphi(t, a) = (t, t - a)$ is a certain change of coordinates that enables one to benefit from the smoothness along the characteristics of the transport. The choice of the bandwidths h_1, h_2, h_3 is chosen according to the data Z^N itself, in the spirit of Lepski's principle [28, 29]. In Theorems 2.10 and 2.12, we derive oracle inequalities that control the pointwise risk of $\widehat{g}_h^N(t, a)$ and $\widehat{\mu}_{h_1, h_2, h_3}^N(t, a)$ in terms of optimal balance between the error propagation of Theorem 2.6 and the linear approximation kernels.

Section 4 is devoted the adaptive estimation of g and μ for the pointwise risk under smoothness constraints. In a first part, we study the smoothness of g when b and μ belong to anisotropic Hölder spaces (and for simplicity, we assume that the initial condition g_0 is sufficiently smooth). Thanks to the relatively explicit form of the solution of the McKendrick Von Foester equation, we establish in Proposition 2.15 that when parametrised via φ , the function \widetilde{g} in the representation $g = \widetilde{g} \circ \varphi$ has explicitly quantifiable improved smoothness over g , suggesting to consider the approximation kernel $H_{h_1} \otimes K_{h_2} \circ \varphi$ for estimating π via the quotient estimator (2.4) that implicitly uses the representation of $\mu = \pi/g$. We establish in Theorem 2.17 minimax lower bounds for estimating $g(t, a)$ and $\mu(t, a)$ and prove in Theorems 2.18 and 2.19 that these bounds are optimal in some cases, thanks to the oracle inequalities established Theorems 2.10 and 2.12. In particular, we achieve minimax adaptation over anisotropic Hölder smoothness constraints, up to poly-logarithmic terms.

Sections 6 is devoted to the proof of the main concentration result of Theorem 2.6 and auxiliary stability results of Section 2. In Section 7, we give the proofs of the statistical results of 3 and 4. The Appendix Section 8 contains some useful technical and auxiliary results.

2 The microscopic model and its large population limit

2.1 Notation

\mathcal{M}_F denotes the set of finite point measures on $\mathbb{R}_+ = [0, \infty)$ and \mathcal{M}_{F+} the set of positive finite measures on \mathbb{R}_+ . Any $Z \in \mathcal{M}_F$ admits the representation $Z = \sum_{i=1}^n \delta_{a_i}$ for some ordered set $\{a_1, \dots, a_n\} \subset \mathbb{R}_+$. For a real-valued function f defined on \mathbb{R}_+ , we write

$$\langle Z, f \rangle = \int_{\mathbb{R}_+} f(a) Z(da) = \sum_{i=1}^n f(a_i).$$

In particular $n = \langle Z, \mathbf{1} \rangle$. For $Z = \sum_{i=1}^n \delta_{a_i} \in \mathcal{M}_F$, abusing notation slightly, we define the evaluation maps $a_i(Z) = a_i$ and for $t \geq 0$, the shift $\tau_t Z = \sum_{i=1}^n \delta_{a_i+t}$.

We fix once for all a terminal time $T > 0$ and $\mathcal{D} = [0, T] \times \mathbb{R}_+$. We work with the set of (measurable) functions

$$\mathcal{L}_{\mathcal{D}}^{\infty} = \{f : \mathcal{D} \rightarrow \mathbb{R}, \sup_{t,a} |f(t, a)| < \infty\},$$

implicitly continued on $\mathbb{R} \times \mathbb{R}$ by setting $f(t, a) = 0$ for $(t, a) \notin \mathcal{D}$ and also introduce

$$\mathcal{L}_{\mathcal{D}}^{\text{time}} = \{f : [0, T] \rightarrow \mathbb{R}, \sup_t |f(t)| < \infty\}, \quad \mathcal{L}_{\mathcal{D}}^{\text{age}} = \{f : \mathbb{R}_+ \rightarrow \mathbb{R}, \sup_a |f(a)| < \infty\},$$

with natural embeddings $\mathcal{L}_{\mathcal{D}}^{\text{time}} \subset \mathcal{L}_{\mathcal{D}}^{\infty}$ and also $\mathcal{L}_{\mathcal{D}}^{\text{age}} \subset \mathcal{L}_{\mathcal{D}}^{\infty}$ for appropriate arguments. For $p = 1, 2$, we set

$$|f|_p = \left(\int_{\mathcal{D}} |f(t, a)|^p dt da \right)^{1/p}, \quad |f|_{\infty} = \sup_{(t,a) \in \mathcal{D}} |f(t, a)|, \quad |f|_{1, \infty} = (|f|_1 |f|_{\infty})^{1/2}. \quad (2.5)$$

For $0 \leq s \leq 1$, we write

$$\mathcal{C}_{\mathcal{D}}^s = \{f \in \mathcal{L}_{\mathcal{D}}^{\infty}, |f(t, a) - f(t', a')| \leq c(|t - t'|^s + |a - a'|^s) \forall (t, a), (t', a') \in \mathcal{D} \text{ for some } c > 0\} \quad (2.6)$$

for the set of s -Hölder continuous functions on \mathcal{D} .

2.2 Construction of the model

The basic assumptions on the model are the following:

Assumption 2.1. *We have*

- (i) $b \in \mathcal{L}_{\mathcal{D}}^{\infty}$ and $\mu \in \mathcal{L}_{\mathcal{D}}^{\infty}$,
- (ii) $NZ_0^N \in \mathcal{M}_F$ is random and satisfies¹ $\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$ almost-surely; moreover $Z_0^N \rightarrow \xi_0$ narrowly, for some deterministic $\xi_0 \in \mathcal{M}_+$,
- (iii) $\xi_0(da) = g_0(a)da$ for some $g_0 \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ such that $\int_0^{\infty} g_0(a)da < \infty$.

¹where $a_N \lesssim b_N$ means $\sup_{N \geq 1} a_N b_N^{-1} < \infty$.

For $t \in [0, T]$, consider the equation

$$\begin{aligned} Z_t^N &= \tau_t Z_0^N + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq b(s, a_i(Z_{s^-}^N)), i \leq \langle N Z_{s^-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta) \\ &\quad - N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{a_i(Z_{s^-}^N) + t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_{s^-}^N)), i \leq \langle N Z_{s^-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(ds, di, d\vartheta) \end{aligned} \quad (2.7)$$

where \mathcal{Q}_i , $i = 1, 2$ are independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$ with intensity measure $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$. In this setting, the distribution Z_0^N describes the renormalised state of the population at time $t = 0$ and $N \langle Z_0^N, \mathbf{1} \rangle$ its size.

Under Assumption 2.1 (i), we have existence and (strong) uniqueness of a solution to (2.7) in $\mathbb{D}([0, T], \mathcal{M}_+)$, the Skorokhod space of càdlàg processes with values in \mathcal{M}_+ . Under Assumption 2.1 (i) and (ii)², we even have the narrow convergence of Z^N in $\mathbb{D}([0, T], \mathcal{M}_+)$ to a deterministic limit $\xi \in \mathcal{C}([0, T], \mathcal{M}_+)$, see *e.g.* [62, 25].

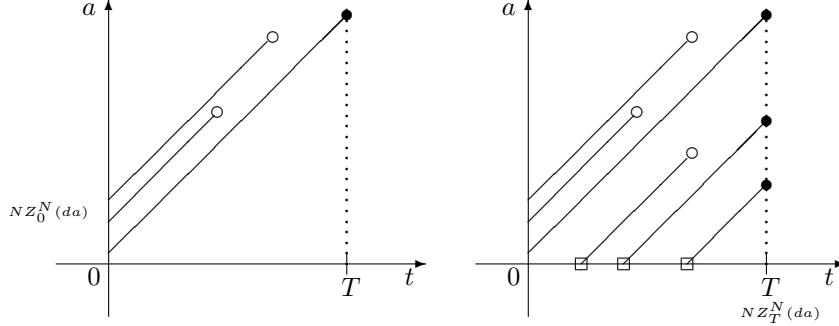


Figure 2.1: Sample path of $NZ_0^N(da)$ and its evolution without births (left), sample path of $(NZ_t^N(da))_{0 \leq t \leq T}$ (right).

Under Assumption 2.1 (iii), the limit $\xi = (\xi_t(da))_{0 \leq t \leq T}$ is smooth in the following sense: we have that $\xi_t(da) = g(t, a)da$ is a weak solution to

$$\begin{cases} \partial_t g(t, a) + \partial_a g(t, a) + \mu(t, a)g(t, a) = 0 \\ g(0, a) = g_0(a), \quad g(t, 0) = \int_0^\infty b(t, a)g(t, a)da. \end{cases} \quad (2.8)$$

The limit g is the solution of an inhomogeneous version of the McKendrick Von Foester transport equation (see [53, 66] and the comprehensive textbook of Perthame [58]). With the notation of Section 1.1, the equation $\mathcal{H}_{b, \mu}^N$ is given by (2.7) while $\mathcal{H}_{b, \mu}$ is given by (2.8).

²Actually, the condition of the almost-sure bound $\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$ can be relaxed to the significant weaker moment condition $\sup_{N \geq 1} \mathbb{E}[\langle Z_0^N, \mathbf{1} \rangle^{1+\epsilon}] < \infty$ for some $\epsilon > 0$.

2.3 Stability of the model

Preliminaries

The stability of $Z_t^N(da)$ relative to its limit $g(t, a)$ will be expressed in terms of weighted quantities of the form

$$\mathcal{W}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathbb{R}_+} w_2(t-a) f_t(a) (Z_t^N(da) - g(t, a) da)$$

and also

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathbb{R}_+} w_2(s-a) f_s(a) (Z_s^N(da) - g(s, a) da) ds,$$

where w_i , $i = 1, 2$ are two bounded weight functions (possibly taking negative values). For notational simplicity, we write $f_t(a) = f(t, a)$ for $f \in \mathcal{L}_D^\infty$ when no confusion is possible. Implicitly, we assume that \mathcal{F} is well-behaved in the sense that $\mathcal{W}_{w_2}^N(\mathcal{F})_t$ and $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t$ are measurables, as random variables on the ambient probability space over which Z^N is defined.

The structure of \mathcal{F}

We look for \mathcal{F} rich enough while having a controlled complexity, measured in terms of entropy. Let \mathfrak{s}_t , \mathfrak{t}_t and \mathfrak{u}_t be the operators on \mathcal{L}_D^∞ defined by

$$\mathfrak{s}_t(f) = ((s, a) \mapsto f(t, a+t)), \quad \mathfrak{t}_t(f) = ((s, a) \mapsto f(t, t-s)), \quad \mathfrak{u}_t(f) = ((s, a) \mapsto f(t, t+a-s)).$$

The minimal structure on \mathcal{F} we need is the following

Assumption 2.2. *We have $0, c_0, c_0 b, c_0 \mu \in \mathcal{F}$ for some constant $c_0 > 0$. Moreover, for every $t \in [0, T]$, the class \mathcal{F} is stable under the following operations:*

$$f \mapsto -f, \quad (f, g) \mapsto fg, \quad f \mapsto \mathfrak{s}_t(f), \quad f \mapsto \mathfrak{t}_t(f), \quad f \mapsto \mathfrak{u}_t(f). \quad (2.9)$$

Let $\text{diam}_{|\cdot|_\infty}(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} |f - g|_\infty$ and write $\mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)$ for the minimal number of ϵ -balls for the $|\cdot|_\infty$ -metric that are necessary to cover \mathcal{F} .

Proposition 2.3. *Let \mathcal{F} be the minimal set satisfying Assumption 2.2 for some $c_0 > 0$ such that $c_1 = c_0 \max(|b|_\infty, |\mu|_\infty) < 1$. If moreover $b, \mu \in \mathcal{C}_D^s$ for some $s > 0$ (\mathcal{C}^s is the set of Hölder continuous functions defined in (2.6)), then*

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty. \quad (2.10)$$

Concentration properties

Definition 2.4 (mild concentration). *A sequence of nonnegative random variables $(X^N)_{N \geq 1}$ has a mild concentration property of order $0 \leq r_N \rightarrow 0$ if for large enough N , we have*

$$\mathbb{P}(X^N \geq (1+u)r_N) \leq \frac{1}{e^u - 1} \quad \text{for every } u \geq 0.$$

Assumption 2.5. *The sequence*

$$|w_2|_{1, \infty}^{-1} \max_{h=1, w_2} \mathcal{W}_h^N(\mathcal{F})_0$$

has a mild concentration property of order r_N for some $0 \leq r_N \rightarrow 0$.

Theorem 2.6. *Work under Assumptions 2.1, 2.2 and 2.5. Assume moreover $\text{diam}_{|\cdot|_\infty}(\mathcal{F}) \leq 1$ and*

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

If w_2 has compact support with length support bounded in N by some $u > 0$ and satisfies an estimate of the form

$$|w_2|_\infty \lesssim \max(N^{1/2}, r_N^{-1})|w_2|_1, \quad (2.11)$$

then

$$(|w_1|_{1,\infty}|w_2|_{1,\infty})^{-1}\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_{1,\infty}^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T$$

share both a mild concentration property of order $C \max(r_N, N^{-1/2})$, for an explicitly computable $C = C(u, e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty, g_0, |w_1|_1, |w_2|_1) > 0$ continuous in its arguments. In particular, if $|w_i|_1$, $i = 1, 2$ is uniformly bounded in N , then C can be chosen independently of N .

Several remarks are in order: **1)** If the initial condition Z_0^N is close to its limit g_0 in $\mathcal{W}_{w_2}(\mathcal{F})_0$ -norm of order r_N , Theorem 2.6 states that the error inflates in $\mathcal{W}_{w_2}(\mathcal{F})_t$ -norm by a factor no worse than $N^{-1/2}$ for $t \in [0, T]$. In particular, whenever $r_N \lesssim N^{-1/2}$, the error propagation is stable. **2)** The order of magnitude of the error propagation is $\max(N^{-1/2}, r_N)$, as one could expect. As for the order in terms of w_1 or w_2 , the ideal order would be the integrated squared-error norm $|w_i|_2$ as a variance term in a central limit theorem for instance. Here, we obtain the slightly worse interpolation quantity $|w_i|_{1,\infty}$ which is always bigger than $|w_i|_2$. However, for statistical purposes, when w_i is replaced by a kernel $w_i = h_N^{-1}K(h_N^{-1}\cdot)$ for some kernel K such that $|K|_1 = 1$, the order is sharp, since in that case

$$|w_i|_{1,\infty} \approx h_N^{-1/2} \approx |w_i|_2$$

and moreover $|w_i|_1$ is uniformly bounded in N . The fact that we have here the correct order for dilating kernels is crucial for nonparametric estimation and is the main purpose (and difficulty) of Theorem 2.6. This seems to be a standard situation for nonparametric estimation in structured populations, where such effects are also met, see [22, 35, 5]. **3)** If w_2 is not compactly supported or if (2.11) does not hold, we still have that

$$(|w_1|_{1,\infty}|w_2|_\infty)^{-1}\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_\infty^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T$$

share both a mild concentration property of order $C \max(r_N, N^{-1/2})$, as explicitly obtained in the proof. However, such a result is not sufficient for nonparametric estimation: picking $w_2 = h_N^{-1}K(h_N^{-1}\cdot)$ yields $|w_2|_\infty \approx h_N^{-1}$ which is dramatically worse than the expected $h_N^{-1/2}$ in kernel estimation. **4)** The constant C also depends on the length of the support of w_2 , but that may be considered as fixed once for all for later statistical purposes. **5)** Assumption 2.5 implies the moment estimate

$$\mathbb{E}\left[\max_{h=1,w_2} \mathcal{W}_h^N(\mathcal{F})_0^p\right] \lesssim |w_2|_{1,\infty}^p r_N^p \quad \text{for every } p > 0. \quad (2.12)$$

In particular, if $Z_0^N(da)$ consists of a N -drawn of independent random variables with common distribution $g_0(a)da$ with $\int_0^\infty a^p g_0(a)da < \infty$, we have Assumption 2.5 if g_0 has subgaussian tails or the weaker (2.12) if $\int_0^\infty a^p g_0(a)da < \infty$ with r_N of order $N^{-1/2}$.

We end this section by giving a global stability result for the propagation of the error $Z_t^N(da) - g(t, a)da$, given a preliminary control on $Z_0^N(da) - g(0, a)da$, which relies on the techniques developed in Theorem 2.6, but with a weaker moment condition for the initial control of the particle system.

Proposition 2.7. *Work under Assumptions 2.1 and 2.2. If*

$$\mathbb{E}\left[\max_{k=1, w_2} \mathcal{W}_k^N(\mathcal{F})_0^p\right] \leq |w_2|_{1, \infty}^p r_N^p \quad (2.13)$$

for some $r_N \geq 0$ and $p \geq 1$, and if w_2 is compactly supported and satisfies an estimate of the form $|w_2|_\infty \lesssim \max(N^{1/2}, r_N^{-1})|w_2|_1$, then

$$\mathbb{E}\left[\mathcal{W}_{w_2}^N(\mathcal{F})_T^p\right] \lesssim |w_2|_{1, \infty}^p \max(N^{-p/2}, r_N^p) \quad (2.14)$$

and

$$\mathbb{E}\left[\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T^p\right] \lesssim (|w_1|_{1, \infty} |w_2|_{1, \infty})^p \max(N^{-p/2}, r_N^p). \quad (2.15)$$

3 Nonparametric estimation of g and μ

3.1 Kernel approximation

Definition 2.8. *A kernel K of (integer) order $\ell_0 \geq 0$ is a bounded function with compact support in \mathbb{R}_+ such that*

$$\int_{\mathbb{R}_+} \kappa^\ell K(\kappa) d\kappa = \mathbf{1}_{\{\ell=0\}}, \quad \text{for } \ell = 0, \dots, \ell_0.$$

For a bandwidth $h > 0$, we set $K_h(\kappa) = h^{-1}K(h^{-1}\kappa)$ so that $|K_h|_1 = |K|_1$. In order to approximate functions of $\mathcal{L}_{\mathcal{D}}^\infty$, we use bivariate kernels defined by

$$H \otimes K(t, a) = H(t)K(a) \quad \text{for } (t, a) \in \mathcal{D},$$

with $H \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$ and $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$. For a bivariate bandwidth $\mathbf{h} = (h_1, h_2)$ with $h_i > 0$, we set

$$(H \otimes K)_{\mathbf{h}}(t, a) = H_{h_1}(t)K_{h_2}(a).$$

and define the linear approximation by the convolution

$$(H \otimes K)_{\mathbf{h}} \star f(t, a) = \int_0^T \int_{\mathbb{R}_+} f(s, u) (H \otimes K)_{\mathbf{h}}(t - s, a - u) ds du. \quad (2.16)$$

We may also approximate f in another system of coordinates: if $\varphi : \mathcal{D} \rightarrow \mathcal{D}$ is invertible, reparametrise f via

$$f(t, a) = \tilde{f} \circ \varphi(t, a)$$

and define the φ -skewed linear approximation

$$(H \otimes K)_{\mathbf{h}} \circ \varphi \star f(t, a) = \int_0^T \int_{\mathbb{R}_+} f(s, u) ((H \otimes K)_{\mathbf{h}} \circ \varphi)(s - t, u - a) ds du$$

so that $((H \otimes K)_{\mathbf{h}} \circ \varphi) \star f(t, a) = (H \otimes K)_{\mathbf{h}} \star \tilde{f}(\varphi(t, a))$. The φ -skewed approximation potentially has better approximation properties for \tilde{f} in the vicinity of $\varphi(t, a)$ than f in the vicinity (t, a) , as it will become transparent in Section 4 below.

3.2 Construction of estimators of g and μ

Construction of an estimator of g

Let $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ be a kernel of order $\ell_0 \geq 0$. For $(t, a) \in \mathcal{D}$, we consider the family of estimators

$$\widehat{g}_h^N(t, a) = K_h \star Z_t^N(a) = \int_{\mathbb{R}_+} K_h(u - a) Z_t^N(du), \quad h > 0. \quad (2.17)$$

Remark 2.9. *At first glance, it may seem slightly surprising to build an estimator of the bivariate function $g(t, a)$ by means of (2.17) that uses data Z_t^N only and discards the observation $(Z_s^N, s \neq t)$. For instance, one may consider estimators of the form*

$$((H \otimes K)_h \circ \varphi) \star Z^N(t, a) = \int_0^T \int_{\mathbb{R}_+} ((H \otimes K)_h \circ \varphi)(s - t, u - a) Z_s^N(du)$$

Formally $\widehat{g}_h^N(t, a) = (H_{h_1=0} \otimes K_h) \star Z^N(t, a)$ without any specific change of coordinates and we will see that such a simple procedure already achieves minimax optimality, see Section 4.3 below.

Construction of the process of death occurrences

We first extract from the data $(Z_t^N(da))_{0 \leq t \leq T}$ the random measure

$$\Gamma^N(dt, da) = \sum_{i \geq 1} \delta_{(T_i, A_i)}(dt, da) \quad \text{on } [0, T] \times \mathbb{R}_+$$

associated with the successive times T_i of the death occurrences of the population during the observation period $[0, T]$, together with the corresponding ages A_i of the individuals that die at time T_i .

Remember that the evaluation mappings $a_i(Z_t^N)$ in the representation $Z_t^N = N^{-1} \sum_{i \geq 1} \delta_{a_i(Z_t^N)}$ are ordered:

$$a_1(Z_t^N) < a_2(Z_t^N) < \dots$$

and that $t \mapsto a_i(Z_t)$ is increasing with slope one unless a birth or a death occurs, in which case we have a non-negative or a negative jump. It follows that

$$\Gamma^N(dt, da) = \sum_{s > 0} \mathbf{1}_{\{i^* = \inf\{i \geq 1, \Delta a_i(Z_s^N) > 0\} < \infty\}} \delta_{(s, a_i^*(Z_{s-}^N))}(dt, da) \quad (2.18)$$

on $[0, T] \times \mathbb{R}_+$, where we set $\Delta a_i(Z_s^N) = a_i(Z_s^N) - a_i(Z_{s-}^N)$ and with the usual convention $\inf \emptyset = \infty$. This second representation in terms of the jump measure of the processes $a_i(Z_t^N)$ gives an explicit construction of $\Gamma^N(dt, da)$ as a function of $(Z_t^N(da), t \in [0, T])$.

Construction of an estimator of μ

Let $H \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$ and $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ be two kernels. For $(t, a) \in \mathcal{D}$ and $\varphi(t, a) = (t, t - a)$, consider the family

$$\widehat{\pi}_h^N(t, a) = \int_0^T \int_{\mathbb{R}_+} ((H \otimes K)_h \circ \varphi)(s - t, u - a) \Gamma^N(ds, du), \quad \mathbf{h} = (h_1, h_2) \quad \text{with } h_i > 0, \quad (2.19)$$

that estimate the function $\pi = \mu g$. An estimator of $\mu(t, a)$ is obtained by considering the ratio

$$\widehat{\mu}_{h, \mathbf{h}}^N(t, a)_\varpi = \frac{\widehat{\pi}_{\mathbf{h}}^N(t, a)}{\widehat{g}_h^N(t, a) \vee \varpi} \quad (2.20)$$

for some threshold $\varpi > 0$, and is thus specified by the bandwidths $h > 0$, $\mathbf{h} = (h_1, h_2)$ with $h_i > 0$ and $\varpi > 0$.

3.3 Oracle inequalities

Estimation of g , data-driven bandwidth

Pick a lattice \mathcal{G}_1^N included in $[N^{-1/2}, (\log N)^{-1}]$ and such that $\text{Card}(\mathcal{G}_1^N) \lesssim N$. The algorithm, based on the Lepski's principle as defined in the Goldenshluger-Lepski's method [28, 29] requires the family of linear estimators

$$\left(\widehat{g}_h^N(t, a), h \in \mathcal{G}_1^N \right)$$

defined in (2.17) and selects an appropriate bandwidth $h = \widehat{h}^N(t, a)$ from the data $(Z_i^N(da))_{0 \leq t \leq T}$. For $(t, a) \in \mathcal{D}$, writing $\{x\}_+ = \max(x, 0)$, define

$$A_h^N(t, a) = \max_{h' \leq h, h' \in \mathcal{G}_1^N} \left\{ \left(\widehat{g}_h^N(t, a) - \widehat{g}_{h'}^N(t, a) \right)^2 - \left(V_h^N + V_{h'}^N \right) \right\}_+,$$

where

$$V_h^N = \left(4(\log N) C^* N^{-1/2} |K_h|_{1, \infty} \right)^2 \quad (2.21)$$

and C^* is a (known) upper bound of the constant C of Theorem 2.6. (Remember that the constant C depends on the parameters of the model via $|b|_\infty, |\mu|_\infty$ and g_0 .) Let

$$\widehat{h}^N(t, a) \in \operatorname{argmin}_{h \in \mathcal{G}_1^N} \left(A_h^N(t, a) + V_h^N \right).$$

The data-driven Goldenshluger-Lepski estimator of $g(t, a)$ is defined as

$$\widehat{g}_\star^N(t, a) = \widehat{g}_{\widehat{h}^N(t, a)}^N(t, a). \quad (2.22)$$

Oracle estimate

We need some notation. Given a kernel K_h , the bias at scale h of g at point (t, a) is defined as

$$\mathcal{B}_h^N(g)(t, a) = \sup_{h' \leq h, h' \in \mathcal{G}_1^N} \left| \int_0^\infty K_{h'}(u - a) g(t, u) du - g(t, a) \right|. \quad (2.23)$$

We are ready to give our first estimation result for every $(t, a) \in \mathcal{D}_- = \mathcal{D} \setminus \{t = a\}$.

Theorem 2.10. *Work under Assumptions 2.1, 2.2 and 2.5 with $r_N \leq N^{-1/2}$ and some \mathcal{F} that satisfies $e(\mathcal{F}) < \infty$. For $(t, a) \in \mathcal{D}_-$, specify $\widehat{g}_\star^N(t, a)$ with a bounded and compactly supported kernel K . The following oracle inequality holds true*

$$\mathbb{E} \left[\left(\widehat{g}_\star^N(t, a) - g(t, a) \right)^2 \right] \lesssim \inf_{h \in \mathcal{G}_1^N} \left(\mathcal{B}_h^N(g)(t, a)^2 + V_h^N \right) + \delta_N$$

for large enough N , with $\delta_N = N^{-1}$ and up to a constant that depends on C^* and K .

Some remarks: **1)** The fact that we measure the performance of \widehat{g}_\star^N at point (t, a) in pointwise squared-error loss is inessential here. Other integrated norms like $|\cdot|_p$ would work as well, following the general proof of Lepski's principle [47, 28, 29]. However, if we need a fine control of the bias in terms of smoothness space, this is no longer true and is linked to the anisotropic and spatial inhomogeneous smoothness structure of the solution g . This will become transparent in Theorems 2.18 and 2.34 below. **2)** In (2.21), the choice of C^\star has to be set in principle prior to the data analysis and is of course difficult to calibrate. It depends on upper bounds on many quantities like $e(\mathcal{F})$ that appear in the constant of Theorem 2.6 or supremum of norms of the unknown parameters b and μ . Moreover, the explicit value C^\star obtained by tracking the constants in the computations of Section 6 is certainly too large. In practice, we need to inject some further prior knowledge and calibrate the threshold by some other method, possibly using data. Such approaches in the context of Lepski's principle have been developed lately in [42]. **3)** The proof relies on Theorem 2.6 which requires $e(\mathcal{F})$ to be finite. However, this requirement is not heavy, as soon as b and μ have a minimal global Hölder smoothness, as stems from Proposition 2.3.

Estimation of μ , data-driven bandwidth

Analogously to the bandwidth-selection method for estimation of g following Lepski's principle, we pick a discrete set $\mathcal{G}_2^N \subset [N^{-1/2}, (\log N)^{-1}]^2$ with cardinality $\text{Card}\mathcal{G}_2^N \lesssim N$. The construction is similar to that of $\widehat{g}_\star^N(t, a)$, given in addition the family of estimators

$$(\widehat{\pi}_{\mathbf{h}}^N(t, a), \mathbf{h} \in \mathcal{G}_2^N)$$

defined in (2.19). For $(t, a) \in \mathcal{D}$, let

$$\mathbf{A}_{\mathbf{h}}^N(t, a) = \max_{\mathbf{h}' \in \mathcal{G}_2^N} \{(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \widehat{\pi}_{\mathbf{h}'}^N(t, a))^2 - (\mathbf{V}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}'}^N)\}_+,$$

where

$$\mathbf{V}_{\mathbf{h}}^N = (4(\log N)C^\star N^{-1/2} |H_{h_1}|_{1, \infty} |K_{h_2}|_{1, \infty})^2 \quad (2.24)$$

and C^\star is a (known) upper bound of the constant C of Theorem 2.6. Let

$$\widehat{\mathbf{h}}^N(t, a) \in \text{argmin}_{\mathbf{h} \in \mathcal{G}_2^N} (\mathbf{A}_{\mathbf{h}}^N(t, a) + \mathbf{V}_{\mathbf{h}}^N(t, a)).$$

The data-driven Goldenshluger-Lepski estimator of $\mu(t, a)$ is defined as

$$\widehat{\mu}_\star^N(t, a)_\varpi = \mu_{\widehat{\mathbf{h}}^N(t, a), \widehat{\mathbf{h}}^N(t, a)}^N(t, a)_\varpi. \quad (2.25)$$

Oracle estimates

In order to estimate μ in squared-error loss consistently with the quotient estimator (2.25), we need a (local) lower bound assumption on $g(t, a)$. Let

$$\begin{aligned} \mathcal{D}_U &= \{(t, a) \in \mathcal{D}, a > t\}, \\ \mathcal{D}_L &= \{(t, a) \in \mathcal{D}, a < t\}, \end{aligned}$$

and $\mathcal{D}^- = \mathcal{D} \setminus \{t = a\}$ so that $\mathcal{D}^- = \mathcal{D}_L \cup \mathcal{D}_U$. A sufficient condition is given by the following

Assumption 2.11. For every $(t, a) \in \mathcal{D}^-$ there exists an open set $\mathcal{U}_{(t,a)}$ such that

$$\inf_{u \in \mathcal{U}_{(t,a)}} b(t-a, t-a+u)g_0(u) \geq \delta \text{ if } (t, a) \in \mathcal{D}_L \quad (2.26)$$

and

$$g_0(t-a) \geq \delta \text{ if } (t, a) \in \mathcal{D}_U, \quad (2.27)$$

for some $\delta > 0$.

We need some notation. For $\mathbf{h} = (h_1, h_2)$ and $\mathbf{h}' = (h'_1, h'_2)$ in \mathcal{G}_2^N , we say that $\mathbf{h} \leq \mathbf{h}'$ if $h_1 \leq h'_1$ and $h_2 \leq h'_2$ hold simultaneously. Given a bivariate kernel $H \otimes K$, the bias at scale \mathbf{h} of $\pi = \mu g$ at point (t, a) in the direction φ is defined as

$$\mathcal{B}_{\mathbf{h}}^N(\mu g)(t, a) = \sup_{\mathbf{h}' \leq \mathbf{h}, \mathbf{h}' \in \mathcal{G}_2^N} \left| \int_{\mathcal{D}} ((H \otimes K)_{\mathbf{h}'} \circ \varphi)(s-t, u-a)\pi(s, u) duds - \pi(t, a) \right|. \quad (2.28)$$

Theorem 2.12. Work under Assumptions 2.1, 2.2, 2.5 with $r_N \leq N^{-1/2}$ and some \mathcal{F} that satisfies $e(\mathcal{F}) < \infty$ together with Assumption 2.11. For $(t, a) \in \mathcal{D}^-$ specify $\hat{\mu}_{\star}^N(t, a)_{\varpi}$ with kernels H, K . The following oracle inequality holds true

$$\mathbb{E}[(\hat{\mu}_{\star}^N(t, a)_{\varpi} - \mu(t, a))^2] \lesssim \inf_{h \in \mathcal{G}_1^N} (\mathcal{B}_h^N(g)(t, a)^2 + \mathbf{V}_h^N) + \inf_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\mu g)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) + \delta_N$$

for large enough N and small enough $\varpi > 0$, with $\delta_N = N^{-1}$ and up to a constant that depends on C^* and the kernels H, K .

Some remarks: **1)** Similar to the case of Theorem 2.10, other loss functions can be chosen. **2)** We see that the performance of $\hat{\mu}_{\star}^N(t, a)_{\varpi}$ is similar to the worst performance of the estimation of the product $\pi = \mu g$ and the estimation of g , as is standard in the study of quotient estimator in the classical Nadaraya-Watson (NW) sense [4, 54]. However, the situation is quite different here than what is customary in standard nonparametric regression with NW: the estimation of $g(t, a)$ is actually equivalent to the estimation of a univariate function, while $\pi(t, a)$ is related to a genuinely bi-variate estimation problem that suffers from a dimensional effect. Therefore, there is good hope to obtain here an optimal procedure, as will become transparent under Hölder anisotropic smoothness scales in the subsequent minimax theorems 2.17 and 2.19 below. **3)** The same remark about the choice of C^* (and also the threshold ϖ) as in Theorem 2.10 above are valid in the context of the estimation of $\mu(t, a)$.

4 Adaptive estimation under anisotropic Hölder smoothness

4.1 The smoothness of the McKendrick Von Foester equation

Definition 2.13. Let $\alpha > 0$, $x_0 \in \mathbb{R}$ and \mathcal{U}_{x_0} be a neighbourhood of x_0 . We say that $f : \mathcal{U}_{x_0} \rightarrow \mathbb{R}$ belongs to $\mathcal{H}^{\alpha}(x_0)$ if³ for every $x, y \in \mathcal{U}_{x_0}$

$$|f^{(n)}(y) - f^{(n)}(x)| \leq C|y - x|^{\{\alpha\}} \quad (2.29)$$

having $\alpha = n + \{\alpha\}$ for a non-negative integer n and $0 < \{\alpha\} \leq 1$.

³The definition depends on \mathcal{U}_{x_0} , further omitted in the notation.

We obtain a semi-norm by setting $|f|_{\mathcal{H}^\alpha(x_0)} = \sup_{x \in \mathcal{U}_{x_0}} |f(x)| + C_{\mathcal{U}_{x_0}}(f)$, where $C_{\mathcal{U}_{x_0}}(f)$ is the smallest constant C for which (2.29) holds. The extension to multivariate functions is straightforward:

Definition 2.14. *The bivariate function f belongs to the anisotropic Hölder class $\mathcal{H}^{\alpha_1, \alpha_2}(x_0, y_0)$ if*

$$|f|_{\mathcal{H}^{\alpha_1, \alpha_2}(x_0, y_0)} = |f(\cdot, y_0)|_{\mathcal{H}^{\alpha_1}(x_0)} + |f(x_0, \cdot)|_{\mathcal{H}^{\alpha_2}(y_0)} < \infty.$$

We write $f \in \mathcal{H}$ if for every $(t, a) \in \mathcal{D}$, we have $f \in \mathcal{H}^{\sigma, \tau}(t, a)$.

Assumption 2.15. *For some $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$ and for every $(t, a) \in \mathcal{D}$, we have*

$$b \in \mathcal{H}^{\alpha, \beta}(t, a), \quad \mu \in \mathcal{H}^{\gamma, \delta}(t, a), \quad g_0 \in \mathcal{H}^\nu(a).$$

We give two results about the pointwise smoothness of the solution of the McKendrick Von Foester equation on $\mathcal{D}^- = \mathcal{D} \setminus \{t = a\}$, depending on the choice of coordinates. The smoothness of g differs on \mathcal{D}_U where only mortality affects the population and \mathcal{D}_L , where both mortality and birth come into play. Introduce also the change of coordinates $\varphi(t, a) = (t, t - a)$ that maps

$$\begin{aligned} \mathcal{D}_U &\rightarrow \varphi(\mathcal{D}_U) = \tilde{\mathcal{D}}_U = \{(t, a') \in \mathcal{D}, 0 \leq t \leq T, a' < 0\} \\ \mathcal{D}_L &\rightarrow \tilde{\mathcal{D}}_L = \varphi(\mathcal{D}_L) = \{(t, a') \in \mathcal{D}, 0 \leq t \leq T, 0 < a' < t\} \end{aligned}$$

onto smoothly. This defines in turn

$$\tilde{g} : \tilde{\mathcal{D}}_U \cup \tilde{\mathcal{D}}_L \rightarrow \mathbb{R}_+ \quad \text{via} \quad g(t, a) = \tilde{g} \circ \varphi(t, a).$$

Proposition 2.16. *Work under Assumptions 2.1, and 2.15.*

(i) *We have*

$$g \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta), \min(\alpha, \beta, \gamma+1, \delta)} \quad \text{on } \mathcal{D}_L \quad \text{and} \quad g \in \mathcal{H}^{\min(\gamma+1, \delta), \max(\gamma \wedge (\delta+1), \delta)} \quad \text{on } \mathcal{D}_U.$$

(ii) *We have the following improvement of the anisotropic smoothness when the parametrisation is given by \tilde{g} :*

$$\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\alpha, \beta, \gamma+1, \delta)} \quad \text{on } \tilde{\mathcal{D}}_L \quad \text{and} \quad \tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \max(\gamma \wedge (\delta+1), \delta)} \quad \text{on } \tilde{\mathcal{D}}_U.$$

The proof of Proposition 2.16 is relatively straightforward, given explicit representations of the solution g in terms of b, μ and g_0 , and is given in Appendix 8.2.

4.2 Minimax lower bounds

For $\alpha, \beta > 0$ and $L > 0$, we set

$$\mathcal{H}_L^{\alpha, \beta}(t, a) = \{f \in \mathcal{L}_\mathcal{D}^\infty, |f|_\infty + |f|_{\mathcal{H}^{\alpha, \beta}(t, a)} \leq L\},$$

where the semi-norm $|\cdot|_{\mathcal{H}^{\alpha, \beta}(t, a)}$ is defined after Definition 2.14. We also set, for $\epsilon > 0$,

$$\mathcal{L}_{\mathcal{D}, \epsilon}^\infty = \{f \in \mathcal{L}_\mathcal{D}^\infty, \inf_{(t, a) \in \mathcal{D}} f(t, a) \geq \epsilon\}$$

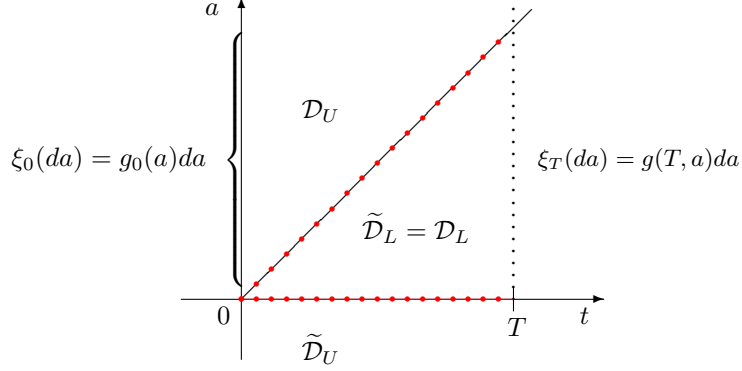


Figure 2.2: $\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\alpha, \beta, \gamma+1, \delta)}$ on $\tilde{\mathcal{D}}_L$ and $\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\gamma, \delta+1)}$ on $\tilde{\mathcal{D}}_U$.

Remember that under Assumption 2.1, any point (b, μ, g_0) with $b, \mu, g_0 \in \mathcal{L}_{\mathcal{D}}^{\infty}$ defines a unique solution g to the McKendrick Von Foester equation (2.8). Let

$$s_{\text{dens}}^- = \max(\gamma, \delta) \quad \text{and} \quad s_{\text{death}}^- = (\gamma^{-1} + \delta^{-1})^{-1}.$$

Under a non-degeneracy condition of the form $\mu \in \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty}$, we obtain the following minimax lower bound:

Theorem 2.17. *Work under Assumptions 2.1 and 2.11. Let $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$ and $L > 0$. For every $(t, a) \in \mathcal{D}^-$, we have*

$$\inf_F \sup_{b, \mu, g_0} \mathbb{E}[|F - g(t, a)|] \gtrsim N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)} \quad (2.30)$$

and

$$\inf_F \sup_{b, \mu, g_0} \mathbb{E}[|F - \mu(t, a)|] \gtrsim N^{-s_{\text{death}}^- / (2s_{\text{death}}^- + 1)}, \quad (2.31)$$

where the infimum is taken over all estimators and the supremum over

$$b \in \mathcal{H}_L^{\alpha, \beta}(t, a), \quad \mu \in \mathcal{H}_L^{\gamma, \delta}(t, a) \cap \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty} \quad \text{and} \quad g_0 \in \mathcal{H}_L^{\nu}(t, a).$$

Some remarks: **1)** As for the previous estimation results in Theorems 2.10 and 2.12, a glance at the proof shows that the lower bound actually holds for a wider class of loss functions, including loss in probability. We keep up to the statements (2.30) and (2.31) in expected pointwise absolute value for simplicity. **2)** If we take $\gamma = \delta$ for simplicity, we see that $s_{\text{dens}}^- = \gamma$ while $s_{\text{death}}^- = \gamma/2$. Therefore, although we are estimating bi-variate functions, the estimation difficulty for $g(t, a)$ is really that of a 1-dimensional function while the estimation of $\mu(t, a)$ remains that of a genuinely bivariate function. Heuristically, there is no information about the population density $g(t, a)$ captured by $(Z_s^N, s \neq t)$ while the estimation of the death rate $\mu(t, a)$ requires dynamical knowledge from the process $\Gamma^N(ds, du)$ for which a truly 2-dimensional information domain around (t, a) is required in order to identify $\mu(t, a)$.

4.3 Adaptive estimation under anisotropic Hölder smoothness

Our next result shows the performance of $g_{\star}^N(t, a)$ defined in (2.22) and gives optimal up to inessential logarithmic factors in some cases. Moreover, $g_{\star}^N(t, a)$ is nearly smoothness adaptive. More

precisely, let

$$s_{\text{dens}}^+(t, a) = \max(\gamma \wedge (\delta + 1), \delta) \mathbf{1}_{\mathcal{D}_U}(t, a) + \min(\alpha, \beta, \gamma + 1, \delta) \mathbf{1}_{\mathcal{D}_L}(t, a), \quad (2.32)$$

and note that $s_{\text{dens}}^+(t, a) \leq s_{\text{dens}}^-(t, a)$ always.

Theorem 2.18. *Work under Assumptions 2.1, 2.2, 2.5 with $r_N \leq N^{-1/2}$ and some \mathcal{F} that satisfies $e(\mathcal{F}) < \infty$, and Assumption 2.11. Specify $\widehat{g}_\star^N(t, a)$ with a compactly supported kernel of order $\ell_0 \geq 0$ and pick*

$$\mathcal{G}_1^N = (x_1^N < x_2^N < \dots < x_N^N)$$

a subdivision of $[N^{-1/2}, (\log N)^{-1}]$ with $\max_{1 \leq i \leq N-1} (x_{i+1}^N - x_i^N) \lesssim N^{-1}$ so that $\text{Card } \mathcal{G}_1^N \lesssim N$. For every $(t, a) \in \mathcal{D}_-$ and large enough N , we have

$$\sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{g}_\star^N(t, a) - g(t, a))^2])^{1/2} \lesssim \left(\frac{(\log N)^2}{N} \right)^{s_{\text{dens}}^+(t, a) \wedge \ell_0 / (2s_{\text{dens}}^+(t, a) \wedge \ell_0 + 1)}, \quad (2.33)$$

where the supremum is taken over $b \in \mathcal{H}_L^{\alpha, \beta}(t, a)$, $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$, $g_0 \in \mathcal{H}_L^\nu(t, a)$ with $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$ and $L > 0$.

Some remarks: **1)** Comparing with the minimax lower bound of Theorem 2.17, we see that both upper and lower bounds (2.30) and (2.33) agree on \mathcal{D}_U if $\delta \leq \gamma \leq \delta + 1$ and on \mathcal{D}_L if $\delta - 1 \leq \gamma \leq \delta$ (and if α and β are sufficiently large too), provided the order ℓ_0 of the kernel K is sufficiently large. The rates are tight up to an inessential logarithmic factor. We do not know about the optimality in g beyond this domain, but we see that the difficulty of the estimation of $g(t, a)$ is equivalent to the difficulty of the univariate function $a \mapsto g(t, a)$ for which the time variable t is simply a parameter: it suffices to piece together the estimators $\widehat{g}_\star^N(t, a)$ for every t in order to estimate the graph $(t, a) \mapsto g(t, a)$. **2)** While we already know that a logarithmic payment is unavoidable for a smoothness adaptive estimator (see the classical Lepski-Low phenomenon, [47, 50]) we do not know whether the order we find in the log term is correct (*i.e.* $(\log N)^2$ versus the classical $\log N$ payment). This stems from Theorem 2.6 and the mild concentration property as we define it, where exponential tail are obtained versus subgaussian tails, but this order seems genuinely linked to the Poissonian behaviour of the noise and it is not clear that we can extend our statistical result in order to remove the extra $\log N$ error-term in (2.33).

Similarly, $\mu_\star^N(t, a)$ defined in (2.25) also shares near optimality in some cases. Define

$$s_L(\alpha, \beta, \gamma, \delta) = (\min(\gamma, \delta)^{-1} + \min(\alpha, \beta, \gamma + 1, \delta)^{-1})^{-1},$$

$$s_U(\gamma, \delta) = (\min(\gamma, \delta)^{-1} + \delta^{-1})^{-1},$$

and

$$s_{\text{death}}^+(t, a) = s_U(\gamma, \delta) \mathbf{1}_{\mathcal{D}_U}(t, a) + s_L(\alpha, \beta, \gamma, \delta) \mathbf{1}_{\mathcal{D}_L}(t, a). \quad (2.34)$$

Note that $s_{\text{death}}^+(t, a) \leq s_{\text{death}}^-(t, a)$ always.

Theorem 2.19. *Work under Assumptions 2.1, 2.2, 2.5 with $r_N \leq N^{-1/2}$ and some \mathcal{F} that satisfies $e(\mathcal{F}) < \infty$, and Assumption 2.11. Specify $\mu_\star^N(t, a)$ with kernels H, K of order $\ell_0 \geq 0$ and pick $\mathcal{G}_2^N = \mathcal{G}_1^N \times \mathcal{G}_1^N$ so that $\text{Card } \mathcal{G}_2^N \lesssim N^2$. For every $(t, a) \in \mathcal{D}^-$ and large enough N , we have*

$$\sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{\mu}_\star^N(t, a) - \mu(t, a))^2])^{1/2} \lesssim \left(\frac{(\log N)^2}{N} \right)^{s_{\text{death}}^+(t, a) \wedge \ell_0 / (2s_{\text{death}}^+(t, a) \wedge \ell_0 + 1)}, \quad (2.35)$$

where the supremum is taken over $b \in \mathcal{H}_L^{\alpha,\beta}(t,a)$, $\mu \in \mathcal{H}_L^{\gamma,\delta}(t,a)$, $g_0 \in \mathcal{H}_L^\nu(t,a)$, with $\alpha, \beta, \gamma, \delta > 0$, $\nu \geq \max(\gamma, \delta) + 1$ and $L > 0$.

Some remarks: **1)** The same remark as **2)** after the statement of Theorem 2.18 holds here. **2)** The minimax optimality situation is somewhat clearer for estimating μ : we see that we have near optimality on \mathcal{D}_U as soon as $\gamma \leq \delta$, while the upper and lower bounds only agree if $\gamma \leq \delta \leq \gamma + 1$ on \mathcal{D}_L (and if α and β are sufficiently large too), provided the order ℓ_0 of the kernel K is sufficiently large. Thus situation is somewhat similar to the estimation of g on \mathcal{D}_U , see Theorem 2.18 above. **3)** The rate of estimation is triggered by the smoothness of $\pi = \mu g$ since the estimation of the quotient g will always be better, for

$$s_{\text{death}}^+(t,a) \leq s_{\text{dens}}^+(t,a) \text{ for every } (t,a) \in \mathcal{D}_-$$

always. However, in order to achieve optimality, we need to optimise the approximation property of π by looking at the smoothness of $\tilde{\pi} = \tilde{\mu}\tilde{g}$, with $\mu = \tilde{\mu} \circ \varphi$. This benefit is obtained thanks to Proposition 2.16 and is given in details in the proof. We would lose by a polynomial order in the rate of convergence given in (2.35) if we used a kernel of the form $(H \otimes K)_h$ instead of $(H \otimes K)_{h \circ \varphi}$ for the estimation of the numerator π in the representation $\mu = \pi/g$.

5 Numerical illustration

The simulations are realised with the parameters:

- (i) g_0 is the density of a gaussian random variable centered in 40 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t,a) = \mathbf{1}_{20 \leq a \leq 40}$. Such a birth rate is not Hölder, however we can prove similar result with such piecewise constant function.
- (iii) $\mu(t,a) = 0,04 \exp(0,0074a) \exp(-0,005t)$. We take a death rate very high in order to have enough deaths for the estimation of the death rate.

We consider the domain $[0, 20] \times [0, 120]$, which means $T = 20$ and $A_{max} = 120$.

We estimate g on the grid $\mathcal{T}^g = \{k1,005, 0 \leq k < 20\}$ and $\mathcal{A}^g = \{k0,2002, 0 \leq k < 600\}$. We estimate μ and μg on the grid $\mathcal{T}^\mu = \mathcal{T}^g$, and $\mathcal{A}^\mu = \{k1,0008, 0 \leq k < 120\}$.

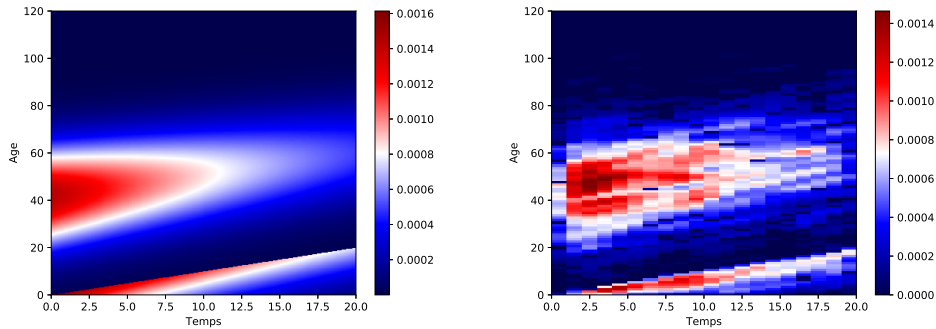


Figure 2.6: Left: true μg density. Right: Estimation with GoldenSchluger Lespki Method, for $N=4000$.

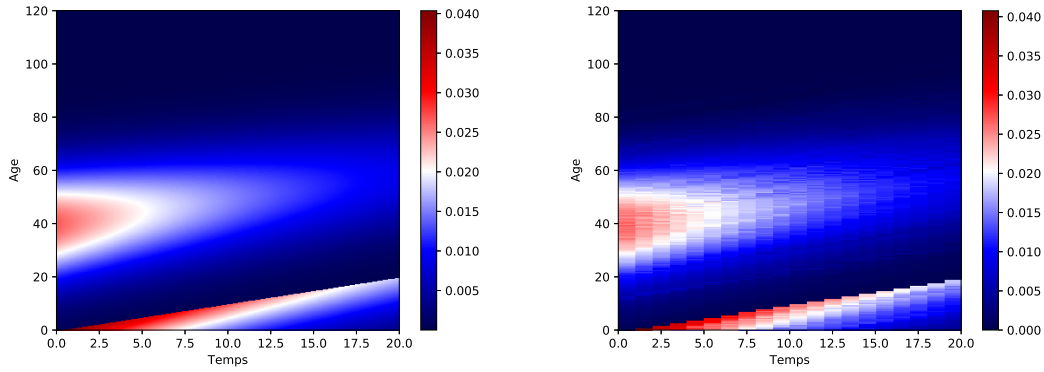


Figure 2.3: Left: true population density. Right: Estimation with GoldenSchluger Lespki Method, for $N = 4000$.

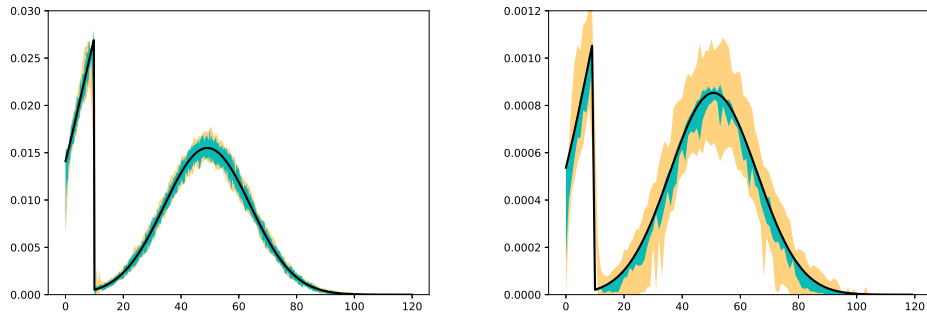


Figure 2.4: For $T = 10$, $N = 4000$, comparison between the true function (black) and the 95% confidence interval on 50 simulations. Oracle in blue, adaptive estimator in yellow. Left: Estimation of g , Right : Estimation of μg .

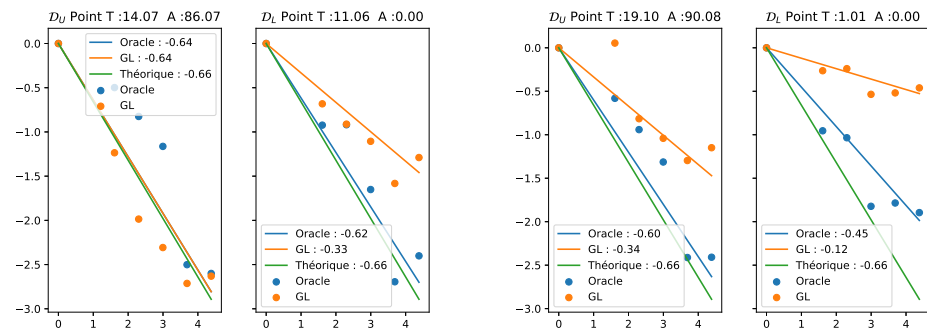


Figure 2.7: Log speed of the estimation of μg realized on 50 simulations. Points are written in the graphs.

5. NUMERICAL ILLUSTRATION

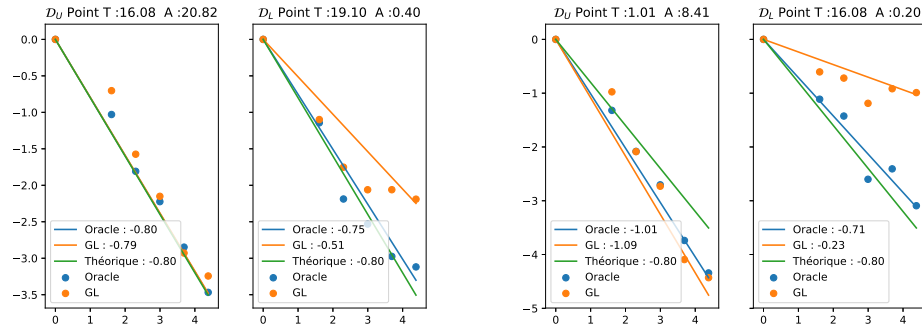


Figure 2.5: Log speed of the estimation of the population density g realized on 50 simulations. Points are written in the graphs.

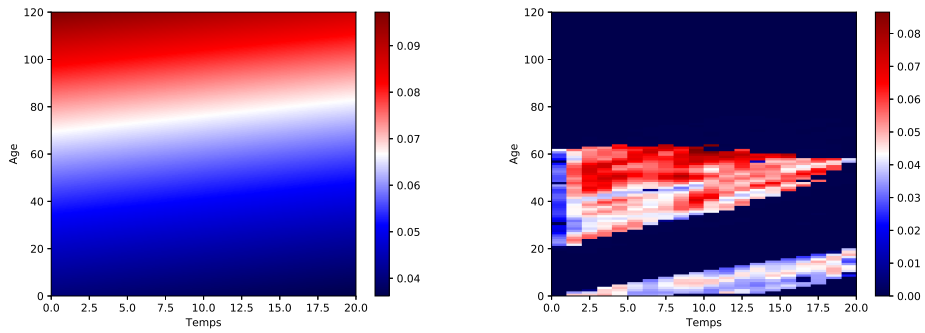


Figure 2.8: Left: true μ density. Right: Estimation with GoldenSchluger Lespki Method, for $N = 4000$, and a threshold of 0.01

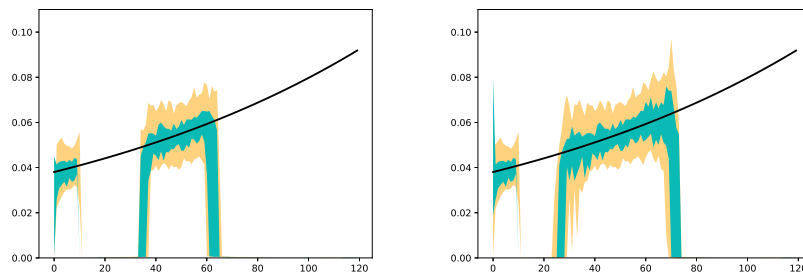


Figure 2.9: For $T = 10$, $N = 4000$, estimation of μ . The true value is in black, in blue the 95% interval confidence for the oracle estimate and in yellow the 95% interval confidence for the adaptive estimate, realized on 50 simulations. Left: threshold of 0.01. Right: threshold of 0.005

From the figure 2.8 the estimation seems very far from μ . This is true, however we can see from the figure 2.9 the threshold is very important.

6 Proof or Theorem 2.6

This section is devoted to the proof of the concentration properties of the model stated in Theorem 2.6. Recall that $w_1 \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$ and $w_2 \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ are two continuous weight functions. We introduce two fundamental processes for which we will establish concentration properties:

$$\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| \int_0^t w_1(s) M_s^N(w_2(s - \cdot) f_s) ds \right|,$$

where $M_t^N(f)$ is defined in (2.37) below and

$$\mathcal{M}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} |M_t^N(w_2(t - \cdot) f_t)|.$$

6.1 A first stability result

Proposition 2.20. *Work under Assumptions 2.1 and 2.2. Then $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$ is bounded above by*

$$|w_1|_1 \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} \int_0^T |w_1(t)| (\mathcal{W}_{w_2, 1}^N(\mathcal{F})_t + \mathcal{W}_{1, w_2}^N(\mathcal{F})_t) dt + \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T,$$

where c_0 is defined in Assumption 2.2.

Proof. By (2.7), the action hoeffy $\langle Z_t^N, f_t \rangle$ of $Z_t^N(da)$ for $f \in \mathcal{L}_{\mathcal{D}}^{\infty}$ can be written as

$$\begin{aligned} \langle Z_t^N, f_t \rangle &= \int_0^{\infty} f_t(t+a) Z_0^N(da) \\ &\quad + \int_0^t \int_0^{\infty} (b(s, a) f_t(t-s) - \mu(s, a) f_t(a+t-s)) Z_s^N(da) ds + M_t^N(f_t), \end{aligned} \quad (2.36)$$

with

$$M_t^N(f_t) = N^{-1} \int_0^t \int_{\{i \leq n_{s^-}^N\} \times \mathbb{R}_+} (f_t(t-s) \mathbf{1}_{\{\theta \leq b\}} - f_t(a_i(Z_{s^-}^N) + t-s) \mathbf{1}_{\{b \leq \theta \leq b+\mu\}}) \tilde{\mathcal{Q}}(ds, di, d\theta). \quad (2.37)$$

In the above formula, $n_t^N = N \langle Z_t^N, \mathbf{1} \rangle$ is the size of the population at time t , the functions b and μ in the indicators are evaluated at points $(s, a_i(Z_{s^-}^N))$ and $\tilde{\mathcal{Q}}(ds, di, d\theta) = \mathcal{Q}(ds, di, d\theta) - ds(\sum_{k \geq 1} \delta_k(di)) d\theta$ is the compensated measure of the Poisson measure \mathcal{Q} .

Apply now (2.36) to the test function $a \mapsto w_2(t-a) f_t(a)$ with $f \in \mathcal{F}$, subtract $g(t, a) da$ in the equation above, noting that $g(t, a)$ solves (2.8), set $\eta_t^N(da) = Z_t^N(da) - g(t, a) da$ and obtain

$$\begin{aligned} \int_{\mathbb{R}_+} w_2(t-a) f_t(a) \eta_t^N(da) &= \int_{\mathbb{R}_+} w_2(-a) f_t(t+a) \eta_0^N(da) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} (w_2(s) f_t(t-s) b(s, a) - w_2(s-a) f_t(a+t-s) \mu(s, a)) \eta_s^N(da) ds + M_t^N(w_2(t - \cdot) f_t). \end{aligned}$$

Multiplying each term by $\omega_1(t)$, integrating from 0 to T and taking absolute values, we also have

$$\left| \int_0^T w_1(t) \int_{\mathbb{R}_+} w_2(t-a) f_t(a) \eta_t^N(da) dt \right| \leq I + II + III + IV,$$

with

$$\begin{aligned} I &= \left| \int_0^T w_1(t) \int_{\mathbb{R}_+} w_2(-a) f_t(t+a) \eta_0^N(da) dt \right|, \\ II &= \left| \int_0^T w_1(t) \int_0^t \int_{\mathbb{R}_+} w_2(s) f_t(t-s) b(s,a) \eta_s^N(da) ds dt \right|, \\ III &= \left| \int_0^T w_1(t) \int_0^t \int_{\mathbb{R}_+} w_2(s-a) f_t(a+t-s) \mu(s,a) \eta_s^N(da) ds dt \right|, \\ IV &= \left| \int_0^T w_1(t) M_t^N(w_2(t-\cdot) f_t) dt \right|. \end{aligned}$$

By Assumption 2.2, we have $f_t(t+a) \in \mathcal{F}$ therefore $I \leq \mathcal{W}_{w_2}^N(\mathcal{F})_0$. Using that $c_0 f_t(t-s) b(s,a) \in \mathcal{F}$, we also have

$$\begin{aligned} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s) f_t(t-s) b(s,a) \eta_s^N(da) ds \right| &\leq c_0^{-1} \sup_{f \in \mathcal{F}} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s) \mathbf{1}(s-a) f_t(a) \eta_s^N(da) ds \right| \\ &= c_0^{-1} \mathcal{W}_{w_2,1}(\mathcal{F})_t, \end{aligned}$$

Therefore $II \leq c_0^{-1} \int_0^T |w_1(t)| \mathcal{W}_{w_2,1}(\mathcal{F})_t dt$. In the same way,

$$\begin{aligned} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s-a) f_t(a+t-s) \mu(s,a) \eta_s^N(da) ds \right| &\leq c_0^{-1} \sup_{f \in \mathcal{F}} \left| \int_0^t \int_{\mathbb{R}_+} \mathbf{1}(s) w_2(s-a) f_t(a) \eta_s^N(da) ds \right| \\ &= c_0^{-1} \mathcal{W}_{1,w_2}(\mathcal{F})_t \end{aligned}$$

and $III \leq c_0^{-1} \int_0^T |w_1(t)| \mathcal{W}_{1,w_2}(\mathcal{F})_t dt$ follows likewise. Finally,

$$|IV| \leq \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^N(w_2(t-\cdot) f_t) dt \right| = \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T.$$

Summing up the estimates, we obtain the conclusion noting that

$$\sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) \int_{\mathbb{R}_+} w_2(t-a) f_t(a) \eta_t^N(da) dt \right| = \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$$

since \mathcal{F} is stable under $f \mapsto -f$ by Assumption 2.2. \square

Proposition 2.21. *Work under Assumptions 2.1 and 2.2. We have*

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \lesssim |w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0 + \max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T,$$

where (k_1, k_2) and (l_1, \dots, l_4) range over permutations of $(1, w_2)$ and $(1, 1, w_1, w_2)$ respectively. The symbol \lesssim means inequality up to an explicitly computable constant depending on T and c_0 from Assumption 2.2.

Proof. Apply first Proposition 2.20 with $w_1 = 1$ and $w_2 = 1$ to obtain

$$\begin{aligned}\mathcal{W}_{1,1}^N(\mathcal{F})_T &\leq T\mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \int_0^T \mathcal{W}_{1,1}^N(\mathcal{F})_t dt + \mathcal{M}_{1,1}^N(\mathcal{F})_T \\ &\leq (T\mathcal{W}_1^N(\mathcal{F})_0 + \mathcal{M}_{1,1}^N(\mathcal{F})_T) e^{2c_0^{-1}T} \\ &= \mathcal{G}^{(1),N}(\mathcal{F})_T,\end{aligned}$$

say, by Grönwall lemma. Next, by Proposition 2.20 applied to $(w_2, 1)$, we obtain

$$\begin{aligned}\mathcal{W}_{w_2,1}^N(\mathcal{F})_T &\leq |w_2|_{L^1([0,T])} \mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \int_0^T |w_2(t)| \mathcal{W}_{1,1}^N(\mathcal{F})_t dt + \mathcal{M}_{w_2,1}^N(\mathcal{F})_T \\ &\leq |w_2|_{L^1([0,T])} (\mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \mathcal{G}^{(1),N}(\mathcal{F})_T) + \mathcal{M}_{w_2,1}^N(\mathcal{F})_T \\ &= \mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T,\end{aligned}$$

say. Apply now Proposition 2.20 with $(1, w_2)$ so that

$$\begin{aligned}\mathcal{W}_{1,w_2}^N(\mathcal{F})_T &\leq T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} \int_0^T (\mathcal{W}_{w_2,1}^N(\mathcal{F})_t + \mathcal{W}_{1,w_2}^N(\mathcal{F})_t) dt + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T \\ &\leq T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} T \mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \int_0^T \mathcal{W}_{1,w_2}^N(\mathcal{F})_t dt + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T \\ &\leq (T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + CT \mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T) e^{c_0^{-1}T} \\ &= \mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T\end{aligned}$$

say, by the previous estimate and Grönwall lemma again. By Proposition 2.20 and the two previous bounds, we infer that $\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$ is less than

$$|w_1|_{L^1([0,T])} \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} |w_1|_{L^1([0,T])} (\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T) + \mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T.$$

Expanding the estimates $\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T$ and $\mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T$ in terms of their appropriate arguments concludes the proof. \square

By Proposition 2.21, we see that the stability of the system is controlled by the initial approximation $\mathcal{W}_{w_2}^N(\mathcal{F})_0$ (including $w_2 = 1$) and the propagation of the stochastic term $\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$. We now turn to that latter term.

6.2 Stability of the stochastic term

For $f \in \mathcal{L}_D^{\text{age}}$, let

$$\widetilde{\mathcal{M}}_{w_1,w_2}^N(f)_t = \int_0^t w_1(s) M_s^N(w_2(s-\cdot)f) ds$$

and

$$\widetilde{\mathcal{M}}_{w_2}^N(f)_t = M_t^N(w_2(t-\cdot)f).$$

In particular, since \mathcal{F} is stable under $f \mapsto -f$, we have

$$\sup_{f \in \mathcal{F}} \widetilde{\mathcal{M}}_{w_1,w_2}^N(f)_T = \sup_{f \in \mathcal{F}} |\widetilde{\mathcal{M}}_{w_1,w_2}^N(f)_T| = \mathcal{M}_{w_2}^N(\mathcal{F})_T \quad (2.38)$$

and

$$\sup_{f \in \mathcal{F}} \widetilde{\mathcal{M}}_{w_2}^N(f)_T = \sup_{f \in \mathcal{F}} |\widetilde{\mathcal{M}}_{w_2}^N(f)_T| = \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

For $\kappa \geq 0$, consider the event

$$\mathcal{A}_\kappa^N = \left\{ \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \exp(|b|_\infty T)(1 + \kappa) \right\}, \quad (2.39)$$

and for $\lambda \geq 0$, set

$$\vartheta_{w_1, w_2}^N(f)_\lambda = 2NT|w_1|_\infty^{-1} \exp(|b|_\infty T)(|b|_\infty + |\mu|_\infty) \rho(N^{-1} \lambda |w_1 w_2|_\infty |f|_\infty),$$

where $\rho(x) = e^x - x - 1$.

Proposition 2.22. *Work under Assumptions 2.1. For large enough N , we have*

$$\int_0^\infty \mathbb{P}((\mathcal{A}_\kappa^N)^c) e^\kappa d\kappa \leq \frac{1}{2} \quad (2.40)$$

and for $\lambda \geq 0$,

$$\mathbb{E} \left[\exp(\lambda |\widetilde{\mathcal{M}}_{w_1, w_2}^N(f)_T - \widetilde{\mathcal{M}}_{w_1, w_2}^N(g)_T|) \mathbf{1}_{\mathcal{A}_\kappa^N} \right] \leq 2 \exp(|w_1|_1 (1 + \kappa) \vartheta_{w_1, w_2}^N(f - g)_\lambda). \quad (2.41)$$

Moreover, (2.41) remains true with $\widetilde{\mathcal{M}}_{w_2}^N(f)_T - \widetilde{\mathcal{M}}_{w_2}^N(g)_T$, replacing formally w_1 by 1 in the right-hand side of the inequality.

Proof. We first prove (2.40), namely

$$\int_0^\infty e^\kappa \mathbb{P} \left(\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle > \exp(|b|_\infty T)(1 + \kappa) \right) d\kappa \leq \frac{1}{2}.$$

Step 1) Consider the equation

$$\widetilde{Z}_t^N = \tau_t Z_0^N + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq |b|_\infty, i \leq N \langle \widetilde{Z}_{s-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta)$$

defined on the same probability space as $(Z_t(da))_{0 \leq t \leq T}$. Applying (2.36) with $b = |b|_\infty$, $\mu = 0$ and $f_t = 1$, we obtain

$$\langle \widetilde{Z}_t^N, \mathbf{1} \rangle = \langle \widetilde{Z}_0^N, \mathbf{1} \rangle + |b|_\infty \int_0^t \langle \widetilde{Z}_s^N, \mathbf{1} \rangle ds + M_t^N(\mathbf{1}),$$

and for every $\lambda \geq 0$, by Itô's formula:

$$\exp(\lambda \langle \widetilde{Z}_t^N, \mathbf{1} \rangle) = \exp(\lambda \langle \widetilde{Z}_0^N, \mathbf{1} \rangle) + N|b|_\infty (e^{\lambda/N} - 1) \int_0^t \langle \widetilde{Z}_s^N, \mathbf{1} \rangle \exp(\lambda \langle \widetilde{Z}_s^N, \mathbf{1} \rangle) ds + \xi_t,$$

where $(\xi_t)_{0 \leq t \leq T}$ is a local martingale. By localisation, one can prove that $\mathbb{E}[\xi_t] = 0$. Writing $f(t, \lambda) = \mathbb{E}[\exp(\lambda \langle \widetilde{Z}_t^N, \mathbf{1} \rangle)]$, it follows that

$$f(t, \lambda) = f(0, \lambda) + N|b|_\infty (e^{\lambda/N} - 1) \int_0^t \partial_\lambda f(s, \lambda) ds. \quad (2.42)$$

The solution of the transport equation (2.42) at time $t = T$ with initial condition $f(0, \lambda) = f_0(\lambda)$ is given by

$$f(T, \lambda) = f_0 \left(N \log \frac{e^{\lambda/N - |b|_\infty T}}{1 - (1 - e^{-|b|_\infty T})e^{\lambda/N}} \right) \leq \exp \left(qN \log \frac{e^{\lambda/N - |b|_\infty T}}{1 - (1 - e^{-|b|_\infty T})e^{\lambda/N}} \right),$$

where the last inequality stems from $f_0(\lambda) = \mathbb{E}[\exp(\lambda \langle \tilde{Z}_0^N, \mathbf{1} \rangle)] = \mathbb{E}[\exp(\lambda \langle Z_0^N, \mathbf{1} \rangle)] \leq e^{q\lambda}$ for some q by Assumption 2.1 (ii).

Step 2) With the notation $r = \exp(-|b|_\infty T)$, the usual Chernoff bound argument yields

$$\begin{aligned} \log \mathbb{P}(\langle \tilde{Z}_T^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa)) &\leq -\lambda r^{-1}(1 + \kappa) + qN \log \frac{r e^{\lambda/N}}{1 - (1-r)e^{\lambda/N}} \\ &\leq -Nr^{-1}(1 + \kappa) \log \left(\left(1 - \frac{rq}{\kappa+1}\right) \frac{1}{1-r} \right) + qN \log \frac{\kappa+1-rq}{1-r} \\ &\leq \log C_1 - C_2 N \kappa \end{aligned}$$

for the choice $\lambda = N \log \left(\left(1 - \frac{rq}{\kappa+1}\right) \frac{1}{1-r} \right)$ and for two constants $C_i = C_i(q, r) > 0$ that do not depend on N . Noting that by construction, $\sup_{t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \langle \tilde{Z}_T^N, \mathbf{1} \rangle$, we finally obtain

$$\begin{aligned} \int_0^\infty e^\kappa \mathbb{P} \left(\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa) \right) d\kappa &\leq \int_0^\infty e^\kappa \mathbb{P}(\langle \tilde{Z}_T^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa)) d\kappa \\ &\leq C_1 \int_0^\infty e^{(1-C_2 N)\kappa} d\kappa = \frac{C_1}{C_2 N - 1} \leq \frac{1}{2} \end{aligned}$$

for $N \geq (1 + 2C_1)/C_2$, and (2.40) is proved.

Step 3) We now turn to (2.41). For $t_0 \in [0, T]$ and $f \in \mathcal{L}_D^{\text{age}}$, define

$$B_{t,t_0}^N(f) = N \int_0^{t \wedge t_0} \int_0^\infty \left(b(s, a) \rho(N^{-1}f(t_0 - s)) + \mu(s, a) \rho(N^{-1}f(a + t_0 - s)) \right) Z_s^N(da) ds. \quad (2.43)$$

Lemma 2.23. *For every $t_0 \in [0, T]$ and $f, g \in \mathcal{L}_D^{\text{age}}$, there exists a nonnegative random variable $\Lambda_{t_0, t_0}^N(f - g)$ with $\mathbb{E}[\Lambda_{t_0, t_0}^N(f - g)] = 1$ such that*

$$\mathbb{E}[\exp(M_{t_0}^N(f) - M_{t_0}^N(g))] = \mathbb{E}[\Lambda_{t_0, t_0}^N(f - g) \exp B_{t_0, t_0}^N(f - g)].$$

Proof. Fix $t_0 \in [0, T]$ and for $f \in \mathcal{L}_D^{\text{age}}$, define the random process

$$\widetilde{M}_{t,t_0}^N(f) = N^{-1} \int_0^{t \wedge t_0} \int_{\{i \leq n_s^N\} \times \mathbb{R}_+} (f(t_0 - s) \mathbf{1}_{\{b \leq \theta\}} - f(a_i(Z_{s^-}^N) + t_0 - s) \mathbf{1}_{\{b \leq \theta \leq b + \mu\}}) \widetilde{\mathcal{Q}}(ds, di, d\theta),$$

obtained by keeping $t = t_0$ fixed in the integrand of $M_{t \wedge t_0}^N(f)$ defined in (2.37). By construction, $(\widetilde{M}_{t,t_0}^N(f))_{0 \leq t \leq T}$ is a martingale. In turn, a simple consequence of Itô's formula, see *e.g.* Tran [62] shows that the random process

$$t \mapsto \Lambda_{t,t_0}^N(f) = \exp(\widetilde{M}_{t,t_0}^N(f) - B_{t,t_0}^N(f))$$

is a martingale such that $\mathbb{E}[\Lambda_{t,t_0}^N(f)] = 1$. Noting that $M_{t_0}^N(f) = \widetilde{M}_{t_0,t_0}^N(f)$ at $t = t_0$, we also have

$$\begin{aligned}\mathbb{E}[\exp(M_{t_0}^N(f) - M_{t_0}^N(g))] &= \mathbb{E}[\exp(M_{t_0}^N(f - g))] \\ &= \mathbb{E}[\exp(\widetilde{M}_{t_0,t_0}^N(f - g))] \\ &= \mathbb{E}[\Lambda_{t_0,t_0}^N(f - g) \exp B_{t_0,t_0}^N(f - g)].\end{aligned}$$

□

Let $\lambda \geq 0$. We substitute $f - g$ by $a \mapsto \lambda w_1(t_0)w_2(t_0 - a)(f(a) - g(a))$ and look for an upper bound for

$$B_{t_0,t_0}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)).$$

Step 4) Observe first that $\rho(x) = e^x - x - 1$ implies that for any nonnegative function $\psi \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$, we have

$$\begin{aligned}\rho(N^{-1}\lambda\psi(a')(f(a) - g(a))) &\leq N^{-1}\psi(a')|f - g|_{\infty} \int_0^{\lambda} (\exp(\kappa N^{-1}|\psi|_{\infty}|f - g|_{\infty}) - 1) d\kappa \\ &= \frac{\psi(a')}{|\psi|_{\infty}} \rho(N^{-1}\lambda|\psi|_{\infty}|f - g|_{\infty}).\end{aligned}$$

Therefore, with $\psi(a') = w_1(t_0)w_2(t_0 - a')$ and $a' = t_0 - s$, we derive

$$\rho(N^{-1}\lambda w_1(t_0)w_2(s)(f(t_0 - s) - g(t_0 - s))) \leq \frac{w_1(t_0)w_2(s)}{|w_1w_2|_{\infty}} \rho(N^{-1}\lambda|w_1w_2|_{\infty}|f - g|_{\infty})$$

and

$$\rho(N^{-1}\lambda w_1(t_0)w_2(s - a)(f(a + t_0 - s) - g(a + t_0 - s))) \leq \frac{w_1(t_0)w_2(s - a)}{|w_1w_2|_{\infty}} \rho(N^{-1}\lambda|w_1w_2|_{\infty}|f - g|_{\infty})$$

with $a' = a + t_0 - s$ follows likewise. Plugging these two estimates in the definition (2.43) of B_{t_0,t_0}^N , we infer on $\mathcal{A}_{\kappa}^N = \{\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \exp(|b|_{\infty}T)(1 + \kappa)\}$ the chain of inequalities

$$\begin{aligned}& B_{t_0,t_0}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \\ & \leq N(|b|_{\infty} + |\mu|_{\infty}) \frac{w_1(t_0)}{|w_1w_2|_{\infty}} \rho(N^{-1}\lambda|w_1w_2|_{\infty}|f - g|_{\infty}) \int_0^{t_0} \int_0^{\infty} (w_2(s) + w_2(s - a)) Z_s^N(da) ds \\ & \leq N(|b|_{\infty} + |\mu|_{\infty}) \frac{w_1(t_0)}{|w_1|_{\infty}} \rho(N^{-1}\lambda|w_1w_2|_{\infty}|f - g|_{\infty}) 2T \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \\ & \leq N(|b|_{\infty} + |\mu|_{\infty}) \frac{w_1(t_0)}{|w_1|_{\infty}} \rho(N^{-1}\lambda|w_1w_2|_{\infty}|f - g|_{\infty}) \exp(|b|_{\infty}T)(1 + \kappa) 2T \\ & = w_1(t_0)(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda.\end{aligned}$$

We derive

$$\begin{aligned}& \exp(\lambda w_1(t_0)M_{t_0}^N(w_2(t_0 - \cdot)(f - g))) \mathbf{1}_{\mathcal{A}_{\kappa}^N} \\ & \leq \exp(w_1(t_0)(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda) \Lambda_{t_0,t_0}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \mathbf{1}_{\mathcal{A}_{\kappa}^N}.\end{aligned}\tag{2.44}$$

Step 5) For every integer $n \geq 1$ and $\lambda \geq 0$, $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$, define

$$\Delta_{w_1, w_2}^{N, n}(f - g)_\lambda = \exp\left(\lambda T n^{-1} \sum_{i=1}^n w_1(iT n^{-1}) M_{iT n^{-1}}^N(w_2(iT n^{-1} - \cdot)(f - g))\right).$$

Applying repeatedly (2.44) with $t_0 = iT n^{-1}$ and integrating with respect to $\mathbb{E}_{\mathcal{A}_\kappa^N}[\cdot] = \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \mathbb{E}[\cdot \mathbf{1}_{\mathcal{A}_\kappa^N}]$, we obtain

$$\begin{aligned} \mathbb{E}_{\mathcal{A}_\kappa^N}[\Delta_{w_1, w_2}^{N, n}(f - g)_\lambda] &\leq \exp\left(T n^{-1} \sum_{i=1}^n w_1(iT n^{-1})(1 + \kappa)\vartheta_{w_1, w_2}^N(f - g)_\lambda\right) \times \\ &\quad \times \mathbb{E}_{\mathcal{A}_\kappa^N}\left[\prod_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda T w_1(t_0)w_2(t_0 - \cdot)(f - g))^{1/n}\right] \\ &\leq \exp\left(T n^{-1} \sum_{i=1}^n w_1(iT n^{-1})(1 + \kappa)\vartheta_{w_1, w_2}^N(f - g)_\lambda\right) \mathbb{P}(\mathcal{A}_\kappa^N)^{-1}, \end{aligned}$$

where we used the fact that the geometric mean is controlled by the arithmetic mean:

$$\begin{aligned} \prod_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))^{1/n} &\leq \\ n^{-1} \sum_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) & \end{aligned}$$

and the fact that

$$\begin{aligned} &\mathbb{E}_{\mathcal{A}_\kappa^N}\left[\Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))\right] \\ &\leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \mathbb{E}\left[\Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))\right] = \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \end{aligned}$$

since $\Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))$ has expectation 1 by Lemma 2.23. Using

$$\liminf_{n \rightarrow \infty} \Delta_{w_1, w_2}^{N, n}(f - g)_\lambda = \exp\left(\lambda \int_0^T w_1(s) M_s^N(w_2(s - \cdot)(f - g)) ds\right)$$

by convergence of Riemann sums, letting $n \rightarrow \infty$, we obtain by Fatou lemma

$$\mathbb{E}_{\mathcal{A}_\kappa^N}\left[\exp\left(\lambda \int_0^T w_1(s) M_s^N(w_2(s - \cdot)(f - g)) ds\right)\right] \leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(|w_1|_1(1 + \kappa)\vartheta_{w_1, w_2}^N(f - g)_\lambda). \quad (2.45)$$

Noting that Lemma 2.23 also holds for $-M_s^N(f)$ and applying (2.45) to $-M_s^N(w_2(s - \cdot)(f - g))$, we infer

$$\mathbb{E}_{\mathcal{A}_\kappa^N}\left[\exp\left(\lambda \left|\int_0^T w_1(s) M_s^N(w_2(s - \cdot)(f - g)) ds\right|\right)\right] \leq 2\mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(|w_1|_1(1 + \kappa)\vartheta_{w_1, w_2}^N(f - g)_\lambda),$$

but since $\int_0^T w_1(s) M_s^N(w_2(s - \cdot)(f - g)) ds = \widetilde{\mathcal{M}}_{w_1, w_2}^N(f)_T - \widetilde{\mathcal{M}}_{w_1, w_2}^N(g)_T$, the estimate (2.41) is established.

Step 6) It remains to prove (2.41) for $\mathcal{M}_{w_2}^N(f-g)_T$. We first integrate (2.44) for $w_1 = 1$ at $t_0 = T$ so that $|w_1|_1 = T$ and proceed exactly as in Step 5) to obtain

$$\mathbb{E}_{\mathcal{A}_\kappa^N} \left[\exp \left(\lambda M_T^N (w_2(T - \cdot)(f - g)) \right) \right] \leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp \left(T(1 + \kappa) \vartheta_{1, w_2}^N (f - g)_\lambda \right).$$

Applying the same argument for $-M_T^N$, we also have

$$\mathbb{E}_{\mathcal{A}_\kappa^N} \left[\exp \left(\lambda |M_T^N (w_2(T - \cdot)(f - g))| \right) \right] \leq 2\mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp \left(T(1 + \kappa) \vartheta_{1, w_2}^N (f - g)_\lambda \right)$$

which is the desired result. \square

Proposition 2.22 is the main ingredient to obtain a concentration inequality for the processes $(\widetilde{\mathcal{M}}_{w_2}^N(f)_T)_{f \in \mathcal{F}}$ and $(\widetilde{\mathcal{M}}_{w_1, w_2}^N(f)_T)_{f \in \mathcal{F}}$, and in turn, a deviation bound for $\mathcal{M}_{w_1, w_2}(\mathcal{F})_T$ and $\mathcal{M}_{w_2}(\mathcal{F})_T$ thanks to (2.38). The proof is given in Section 6.4 below.

More precisely, consider the apparently more general situation where we have a real-valued process $\xi(f)_{f \in \mathcal{F}}$ indexed by some metric set (\mathcal{F}, d) and a family of events $\mathcal{A}(\kappa)_{\kappa > 0}$ satisfying the following properties:

$$\int_0^\infty \mathbb{P}(\mathcal{A}(\kappa)^c) e^\kappa d\kappa \leq \frac{1}{2}, \quad (2.46)$$

and

$$\mathbb{E} \left[\exp \left(\lambda |\xi(f) - \xi(g)| \right) \mathbf{1}_{\mathcal{A}(\kappa)} \right] \leq 2 \exp \left(c_1(1 + \kappa) \rho(c_2 d(f, g) \lambda) \right), \quad (2.47)$$

for every $\lambda \geq 0$ and some $c_1, c_2 > 0$.

Proposition 2.24. *Assume that $\xi(f)_{f \in \mathcal{F}}$ and $\mathcal{A}(\kappa)_{\kappa > 0}$ satisfy (2.46) and (2.47) and that $\xi(f_0) = 0$ for some $f_0 \in \mathcal{F}$. Then there exists a choice $\varpi = \varpi(c_1, c_2) > 0$ such that for every $u \geq 0$:*

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}} |\xi(f)| \geq 8 \left(u + \int_0^{\text{diam}_{\tilde{d}}(\mathcal{F})} \log(1 + \mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)) d\epsilon \right) \right) \leq (e^{u/\text{diam}_{\tilde{d}}(\mathcal{F})} - 1)^{-1},$$

where $\tilde{d} = \varpi d$, $\text{diam}_{\tilde{d}}(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} \tilde{d}(f, g)$ and $\mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)$ is the minimal number of balls of \tilde{d} -size $\epsilon > 0$ that are necessary to cover \mathcal{F} .

Remark 2.25. *We show in Remark 2.35 at the end of the proof of Proposition 2.24 that if $c_1 \geq 308$, we may pick $\varpi(c_1, c_2) = k\sqrt{c_1 c_2}$, with $k = 2\sqrt{77}$.*

The proof of Proposition 2.24 relies on standard concentration techniques and goes back to Dudley [23]. We use the classical textbook of Ledoux-Talagrand [44] and detail the computations in the Appendix section 8. Combining Proposition 2.22 and 2.24, we obtain the following

Theorem 2.26. *Work under Assumptions 2.1. Assume $\text{diam}_{|\cdot|_\infty} \mathcal{F} \leq 1$ and*

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

For large enough N , there exists an explicit choice of $C = C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) > 0$, given in the proof below, such that for every $u \geq 0$:

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq (1 + u)CN^{-1/2}|w_1|_{1, \infty}|w_2|_\infty) \leq (e^{ue(\mathcal{F})} - 1)^{-1} \quad (2.48)$$

and

$$\mathbb{P}(\mathcal{M}_{w_2}^N(\mathcal{F})_T \geq (1 + u)CN^{-1/2}|w_2|_\infty) \leq (e^{ue(\mathcal{F})} - 1)^{-1}. \quad (2.49)$$

Proof. We plan to apply Proposition 2.24 with $\xi(f) = \widetilde{\mathcal{M}}_{w_1, w_2}^N(f)_T$, having $\xi(f_0) = 0$ for $f_0 = 0$. We take $\mathcal{A}(\kappa) = \mathcal{A}_\kappa^N$ defined in (2.39) and notice that (2.46) is satisfied by (2.40). Also, we have (2.47) by (2.41) with

$$c_1 = 2NT|w_1|_1|w_1|_\infty^{-1} \exp(|b|_\infty T)(|b|_\infty + |\mu|_\infty) \quad \text{and} \quad c_2 = |w_1 w_2|_\infty N^{-1}, \quad (2.50)$$

for the metric $d(f, g) = |f - g|_\infty$. Setting $\widetilde{d} = \varpi d$ with ϖ taken from Proposition 2.24, we have $\text{diam}_{\widetilde{d}}(\mathcal{F}) = \varpi \text{diam}_d(\mathcal{F}) \leq \varpi$ by assumption and also $\mathcal{N}(\mathcal{F}, \widetilde{d}, \epsilon) \leq \mathcal{N}(\mathcal{F}, d, \epsilon/\varpi)$. It follows that

$$\int_0^{\text{diam}_{\widetilde{d}}(\mathcal{F})} \log(1 + \mathcal{N}(\mathcal{F}, \widetilde{d}, \epsilon)) d\epsilon \leq \int_0^\varpi \log(1 + \mathcal{N}(\mathcal{F}, d, \epsilon/\varpi)) d\epsilon = \varpi e(\mathcal{F}),$$

which is finite by assumption. Since $\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T = \sup_{f \in \mathcal{F}} |\xi(f)| = \sup_{f \in \mathcal{F}} \xi(f)$, remember (2.38), we may apply Proposition 2.24 and obtain, for every $u \geq 0$,

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8(\varpi e(\mathcal{F}) + u)) \leq (e^{u/\varpi} - 1)^{-1},$$

or equivalently

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8\varpi e(\mathcal{F})(1 + u)) \leq (e^{ue(\mathcal{F})} - 1)^{-1}.$$

By Remark 2.25 (see also Remark 2.35), we pick $\varpi = \varpi(c_1, c_2) = k\sqrt{c_1 c_2}$ with $k = 2\sqrt{77}$, assuming $c_1 \geq 308$ which is satisfied for sufficiently large N by (2.50). Using (2.50) again, it follows that

$$\begin{aligned} 8k\sqrt{c_1 c_2} e(\mathcal{F})(1 + u) &= (1 + u)8k e(\mathcal{F}) e^{\frac{1}{2}|b|_\infty T} \sqrt{2T} (|b|_\infty + |\mu|_\infty)^{1/2} N^{-1/2} (|w_1|_1 |w_1|_\infty)^{1/2} |w_2|_\infty \\ &= (1 + u)CN^{-1/2} |w_1|_{1, \infty} |w_2|_\infty, \end{aligned}$$

say, with

$$C = C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) = 8k e(\mathcal{F}) e^{\frac{1}{2}|b|_\infty T} \sqrt{2T} (|b|_\infty + |\mu|_\infty)^{1/2} \quad (2.51)$$

and (2.48) follows. The proof of (2.49) is obtained in the same way and is omitted. \square

Remark 2.27. (i) Up to inflating the constant C by a multiplicative factor $\max(1, e(\mathcal{F}))$, we see that Theorem 2.26 implies a mild concentration property for

$$(|w_1|_{1, \infty} |w_2|_\infty)^{-1} \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_\infty^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$$

with rate $CN^{-1/2}$.

(ii) The initial bound $|w_2|_{1, \infty}$ of Assumption 2.5 inflates to $|w_2|_\infty$ in (2.49). This defect actually has dramatic consequences when applied to subsequent statistical estimation: w_2 becomes a kernel depending on N that mimicks a Dirac mass which is not stable for the $|\cdot|_\infty$ metric. Improving on this estimates is actually the key difficulty in the proof of Theorem 2.6.

6.3 Proof of Theorem 2.6

The weakness of Theorem 2.26 lies in the use of Proposition 2.22, where the control (2.41) somehow needs to be improved. This improvement however uses the results of Theorem 2.26 that we are going to iterate.

Step 1) By Proposition 2.21, we have $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \lesssim I + II$, with

$$I = |w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0$$

and

$$II = \max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T.$$

Since $|w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0 \lesssim \max_{k=1, w_2} \mathcal{W}_k^N(\mathcal{F})_0$ up to a constant that only depends on T , $|w_1|$ and $|w_2|_{L^1([0, T])}$, we have by Assumption 2.5 that $(|w_1|_{1, \infty} |w_2|_\infty)^{-1} I$ has a mild concentration property (actually, we can even replace $|w_2|_\infty$ by $|w_2|_{1, \infty}$). Next, by Theorem 2.26, the mild concentration property also holds for

$$(|w_1|_{1, \infty} |w_2|_\infty)^{-1} \mathcal{W}_{k, l}^N(\mathcal{F})_T, \quad \text{with } (k, l) \in \{(1, 1), (w_2, 1), (1, w_2)\}$$

up to an appropriate change in the constants, and therefore it carries over to $(|w_1|_{1, \infty} |w_2|_\infty)^{-1} II$ since $\max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T \lesssim \sum_{(k, l)} \mathcal{W}_{k, l}^N(\mathcal{F})_T$ where the summation holds over $\{(1, 1), (w_2, 1), (1, w_2)\}$. In turn,

$$(|w_1|_{1, \infty} |w_2|_\infty)^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$$

has a mild concentration property of order $C' \max(r_N, N^{-1/2})$, for some $C' > 0$ that depends on c_0 of Assumption 2.2, T , $|w_1|_1$, $|w_2|_{L^1([0, T])}$ and the constant $C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty)$ of Theorem 2.26 defined in (2.51).

Step 2) We next carefully revisit Step 4) of the proof of Proposition 2.22. We have

$$\begin{aligned} & B_{t_0, t_0}^N(\lambda w_1(t_0) w_2(t_0 - \cdot)(f - g)) \\ & \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) \int_0^T \int_0^\infty (w_2(s) + w_2(s - a)) Z_s^N(da) ds. \end{aligned}$$

Adding and subtracting the limit $g(t, a) da$, we also have

$$\begin{aligned} & \int_0^T \int_0^\infty (w_2(s) + w_2(s - a)) Z_s^N(da) ds \\ & \leq \int_0^T \int_0^\infty (w_2(s) + w_2(s - a)) g(s, a) da ds + \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T + \mathcal{W}_{1, w_2}^N(\mathcal{F})_T \\ & \leq |w_2|_g + \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T + \mathcal{W}_{1, w_2}^N(\mathcal{F})_T, \end{aligned}$$

where, for $f \in \mathcal{F}_b^{\text{age}}$ we set

$$|f|_g = |f|_{L^1([0, T])} \sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da + T |f|_1 |g|_\infty \wedge |f|_\infty |g|_1. \quad (2.52)$$

This bound is tighter than the estimate $2|w_2|_\infty T \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle$ that we used in Step 4) of the proof of Proposition 2.22. Introduce now the family of events

$$\mathcal{B}_\kappa^N = \left\{ \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \leq 5C' \max(r_N, N^{-1/2}) |w_2|_{1, \infty} (1 + \kappa) \right\}, \quad \kappa > 0,$$

and

$$\mathcal{C}_\kappa^N = \left\{ \mathcal{W}_{1,w_2}^N(\mathcal{F})_T \leq 5C' \max(r_N, N^{-1/2}) \sqrt{T} |w_2|_\infty (1 + \kappa) \right\}, \quad \kappa > 0,$$

where C' is the constant of Step 1). On $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$, we now have

$$\begin{aligned} & B_{t_0, t_0}^N (\lambda w_1(t_0) w_2(t_0 - \cdot) (f - g)) \\ & \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) (|w_2|_g + C_{w_2}^N) (1 + \kappa) \\ & = w_1(t_0) (1 + \kappa) \tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda, \end{aligned}$$

say, with

$$C_{w_2}^N = 5C' \max(r_N, N^{-1/2}) (|w_2|_{1, \infty} + |w_2|_\infty \sqrt{T}) \quad (2.53)$$

and

$$\tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda = N(|b|_\infty + |\mu|_\infty) \frac{|w_2|_g + C_{w_2}^N}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty).$$

We thus have established that (2.41) of Proposition 2.22 holds with $\tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda$ instead of $\vartheta_{w_1, w_2}^N (f - g)_\lambda$ and $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$ instead of \mathcal{A}_κ^N .

Step 3) We now prove an analogous bound as (2.40) replacing \mathcal{A}_κ^N by $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$. Applying Theorem 2.26 with $(w_1, w_2) = (w_2, 1)$ up to an inflation of C by $\max(e(\mathcal{F}), 1)$ with the substitution $1 + u = 5(1 + \kappa)$, we obtain

$$\begin{aligned} \mathbb{P}((\mathcal{B}_\kappa^N)^c) &= \mathbb{P}(\mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \geq 5C' \max\{r_N, N^{-1/2}\} |w_2|_{1, \infty} (1 + \kappa)) \\ &= \mathbb{P}(\mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \geq (1 + u) C' \max(r_N, N^{-1/2}) |w_2|_{1, \infty}) \\ &\leq (\exp(u) - 1)^{-1} = (\exp(4 + \kappa + 4\kappa) - 1)^{-1} \leq e^{-5\kappa}. \end{aligned}$$

It follows that

$$\int_0^\infty \mathbb{P}((\mathcal{B}_\kappa^N)^c) e^\kappa d\kappa \leq \int_0^\infty e^{-4\kappa} d\kappa = \frac{1}{4}.$$

In the same way, applying Theorem 2.26 with $(w_1, w_2) = (1, w_2)$ and up to an inflating the constant C again, we obtain

$$\mathbb{P}((\mathcal{C}_\kappa^N)^c) = \mathbb{P}(\mathcal{W}_{1, w_2}^N(\mathcal{F})_T \geq 5C' \max(r_N, N^{-1/2}) |w_2|_\infty \sqrt{T} (1 + \kappa)) \leq e^{-5\kappa}$$

Hence $\int_0^\infty \mathbb{P}((\mathcal{C}_\kappa^N)^c) e^\kappa d\kappa \leq \frac{1}{4}$ follows likewise and (2.40) is proved with $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$ in place of \mathcal{A}_κ^N .

Step 4) We may now reproduce the proof of Theorem 2.26 with our new estimates from Step 2) : the estimate (2.50) now becomes

$$c'_1 = N(|b|_\infty + |\mu|_\infty) \frac{|w_1|_1}{|w_1 w_2|_\infty} (|w_2|_g + C_{w_2}^N) \quad \text{and} \quad c'_2 = c_2 = N^{-1} |w_1 w_2|_\infty,$$

and thanks to Step 3), we may apply in this new setting Proposition 2.24 to obtain

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8\varpi(c'_1, c'_2) e(\mathcal{F}) (1 + u)) \leq (e^{ue(\mathcal{F})} - 1)^{-1}.$$

Again, we may pick $\varpi = \varpi(c'_1, c'_2) = k\sqrt{c'_1}c'_2$ with $k = 2\sqrt{77}$, assuming $c_1 \geq 308$ which is true for N is large enough, and it follows that

$$\begin{aligned} & 8k\sqrt{c'_1}c'_2e(\mathcal{F}) \\ &= 8k e(\mathcal{F})(|b|_\infty + |\mu|_\infty)^{1/2}N^{-1/2}|w_1|_{1,\infty}|w_2|_\infty^{1/2}(|w_2|_g + C_{w_2}^N)^{1/2} \\ &\leq C''N^{-1/2}|w_1|_{1,\infty}|w_2|_\infty^{1/2}(|w_2|_g + C_{w_2}^N)^{1/2} \end{aligned}$$

say, with

$$C'' = C''(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) = 8k \max(1, e(\mathcal{F}))(|b|_\infty + |\mu|_\infty)^{1/2} \max(5C'\sqrt{T}, 1)^{1/2}.$$

For $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$, define now

$$[f]_{1,\infty}^{\varepsilon_N} = |f|_\infty^{1/2}(|f|_g + \varepsilon_N(|f|_{1,\infty} + |f|_\infty))^{1/2}.$$

We have proved that for $\varepsilon_N = \max(r_N, N^{-1/2})$, the sequence

$$(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$$

has a mild concentration property with rate $C''N^{-1/2}$. Applying the same argument as for Step 1) above, the mild concentration property carries over to

$$(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$$

with rate $C'' \max(r_N, N^{-1/2})$, possibly up to inflating the constant $C'' > 0$.

Step 5) We finally show that $[w_2]_{1,\infty}^{\varepsilon_N} \lesssim |w|_{1,\infty}$ up to a constant that only depends on $|b|_\infty, |\mu|_\infty, |g_0|_\infty$ and T , under the additional assumption that w_2 has compact support and $|w_2|_\infty \lesssim \varepsilon_N^{-1}|w_2|_1$. By definition of $|w_2|_g$ in (2.52), we have

$$|w_2|_g \lesssim |w_2|_1 \left(\sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da + |g|_\infty \right) \lesssim |w_2|_1$$

by the estimates of Lemma 2.37 in Appendix 8.3. Moreover, the compact support of w_2 implies $|w_2|_{1,\infty} \leq |w_2|_\infty |\text{supp}(w_2)|^{1/2} \lesssim |w_2|_\infty$. It follows that

$$[w_2]_{1,\infty}^{\varepsilon_N} \lesssim |w_2|^{1/2}(|w_2|_1 + \varepsilon_N|w_2|_\infty)^{1/2} \lesssim |w_2|_{1,\infty}.$$

Let us note that the constant may possibly depend on $|\text{supp}(w_2)|$ which is bounded above by \mathbf{u} by assumption.

Step 6) It remains to prove a mild concentration property for $([w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T$ with rate $C'' \max(r_N, N^{-1/2})$. The property holds for

$$([w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_2}^N(\mathcal{F})_T$$

with the same proof as for $(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$. We omit the details. Next, reproducing the beginning of the proof of Proposition 2.20 and applying (2.36) to the test function $a \mapsto w_2(t-a)f_t(a)$ with $f \in \mathcal{F}$, we obtain

$$\mathcal{W}_{w_2}^N(\mathcal{F})_T \leq \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1}(\mathcal{W}_{w_2,1}^N(\mathcal{F})_T + \mathcal{W}_{1,w_2}^N(\mathcal{F})_T) + \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

By Proposition 2.21, we further have

$$\mathcal{W}_{w_2,1}^N(\mathcal{F})_T \lesssim \mathcal{W}_1^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T$$

and

$$\mathcal{W}_{1,w_2}^N(\mathcal{F})_T \lesssim \max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T,$$

up to a constant that only depends on T , c_0 , $|w_1|_1$ and $|w_2|_{L^1([0,T])}$, therefore $\mathcal{W}_{w_2}^N(\mathcal{F})_T$ is of order

$$\max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T + \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

The mild concentration property of $(|w_2|_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$ and $(|w_1|_{1,\infty} [w_2]_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$ enables us to control the last two terms. The first term has the correct order by Assumption 2.5. The proof of Theorem 2.6 is complete.

6.4 Remaining proofs of Section 2

Proof of Proposition 2.7

We repeat the argument of Step 6) in the proof of Theorem 2.6 above. By Proposition 2.21, we have

$$\mathcal{W}_{w_2}^N(\mathcal{F})_T \lesssim \max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T$$

and thus

$$\mathbb{E}[\mathcal{W}_{w_2}^N(\mathcal{F})_T^p] \lesssim \mathbb{E}[\max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0^p] + \mathbb{E}[\max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T^p] + \mathbb{E}[\mathcal{M}_{w_2}^N(\mathcal{F})_T^p],$$

up to a constant that depends on p , T , c_0 , $|w_1|_1$ and $|w_2|_{L^1([0,T])}$. The first term is of order $|w_2|_{1,\infty}^p r_N$ by Assumption. For the two other terms we use the identity $\mathbb{E}[Z^p] = p \int_0^\infty x^{p-1} \mathbb{P}(Z \geq x) dx$ for a nonnegative random variable Z and conclude with the mild concentration property of $(|w_2|_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$ and $(|w_1|_{1,\infty} [w_2]_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$.

Proof of Proposition 2.3

Let \mathcal{F}_0 denote the minimal set that contains $0, c_0, c_0\mu, c_0b$ and that is stable under the operations defined in (2.9) except for the pointwise product $(f, g) \mapsto f \cdot g$. We also set, for $f \in \mathcal{L}_D^\infty$:

$$\mathcal{A}(f)_{t_1,t_2}^{(k,l)} = ((s, a) \mapsto f(t_1, t_2 + ka - ls))$$

with $t_1, t_2 \in [0, T]$ and $k, l = 0, 1$.

Step 1) We claim that

$$\mathcal{F}_0 \subseteq \left\{ 0, \pm c_0, \pm c_0\mu, \pm c_0b, \pm \mathcal{L}(c_0b)_{t_1,t_2}^{(k,l)}, \pm \mathcal{A}(c_0\mu)_{t_1,t_2}^{(k,l)}, \text{ for every } t_1, t_2 \in [0, T], k, l = 0, 1 \right\}. \quad (2.54)$$

Indeed, one can check the following stability properties:

$$\mathfrak{s}_t(\mathcal{A}(f)_{t_1,t_2}^{(k,l)})(s, a) = \mathcal{A}_{t_1,t_2}^{(k,l)}(t, t+a) = f(t_1, t_2 + kt + ka - lt) = \mathcal{A}_{t_1,t_2+kt-lt}^{(k,0)}(s, a),$$

$$\begin{aligned}\mathfrak{t}_t(\mathcal{A}(f)_{t_1, t_2}^{(k, l)})(s, a) &= \mathcal{A}_{t_1, t_2}^{(k, l)}(t, t - s) = f(t_1, t_2 - lt + ka - ks) = \mathcal{A}_{t_1, t_2 + kt - lt}^{(0, k)}(s, a), \\ \mathfrak{u}_t(\mathcal{A}(f)_{t_1, t_2}^{(k, l)})(s, a) &= \mathcal{A}_{t_1, t_2}^{(k, l)}(t, t - s + a) = f(t_1, t_2 + kt - ks + ka - lt) = \mathcal{A}_{t_1, t_2 + kt - lt}^{(k, k)}(s, a).\end{aligned}$$

This proves (2.54).

Step 2) We now prove that if $b, \mu \in \mathcal{C}^s$ for some $0 < s \leq 1$ with Hölder constant $L > 0$, then

$$\mathcal{N}(\mathcal{F}_0, |\cdot|_\infty, \epsilon) \lesssim \epsilon^{-2/s}, \quad (2.55)$$

up to a constant that only depends on s, T and L . Indeed, if $f \in \mathcal{C}^s$ with Hölder constant $L > 0$, we have

$$\begin{aligned}|\mathcal{L}(f)_{t_1, t_2}^{(k, l)} - \mathcal{L}(f)_{t'_1, t'_2}^{(k, l)}|_\infty &= \sup_{s, a} |f(t_1, t_2 + ka - ls) - f(t'_1, t'_2 + ka - ls)| \\ &\leq L(|t_1 - t'_1|^s + |t_2 - t'_2|^s),\end{aligned}$$

therefore, for fixed (k, l) and $f \in \mathcal{C}^s$, the ϵ -covering number of $\{\mathcal{L}(f)_{t_1, t_2}^{(k, l)}, t_1, t_2 \in [0, T]\}$ in $|\cdot|_\infty$ is the same as that of $[0, T]^2$ equipped with the metric $d((t_1, t_2) - (t'_1, t'_2)) = L(|t_1 - t'_1|^s + |t_2 - t'_2|^s)$. Since $\mathcal{N}([0, T], \epsilon, L|\cdot|^\gamma) = T \mathcal{N}([0, T], (\epsilon/L)^{1/\gamma}, |\cdot|) = TL^{1/\gamma} \epsilon^{-1/\gamma}$, we have that $\mathcal{N}([0, T]^2, \epsilon, d) \lesssim \epsilon^{-2/s}$ and (2.55) is established.

Step 3) We now consider the class $\mathcal{F}_0^{\text{prod}}$ that contains \mathcal{F}_0 and that is stable under the operation $(f, g) \mapsto fg$. Since $\mathfrak{s}_t(fg) = \mathfrak{s}_t(f)\mathfrak{s}_t(g)$, $\mathfrak{t}_t(fg) = \mathfrak{t}_t(f)\mathfrak{s}_t(g)$, $\mathfrak{u}_t(fg) = \mathfrak{u}_t(f)\mathfrak{s}_t(g)$, the class $\mathcal{F}_0^{\text{prod}}$ contains the minimal class \mathcal{F} .

Let $f = \prod_{\ell=1}^m f_\ell \in \mathcal{F}_0^{\text{prod}}$, with $f_\ell \in \mathcal{F}_0$. For every ℓ , we have $|f_\ell|_\infty \leq c_1 < 1$, with $c_1 = c_0 \max(b|_\infty, |\mu|_\infty) < 1$ by assumption. Therefore, if $m \geq \log \epsilon / \log c_1 = m(\epsilon)$, we have $|f|_\infty = |f - 0|_\infty \leq \epsilon$. Now, let g_i be $\mathcal{N}(\mathcal{F}_0, \epsilon m(\epsilon)^{-1}, |\cdot|_\infty)$ functions in \mathcal{F}_0 such that, for every $f \in \mathcal{F}_0$, there exists an index $i(f)$ such that $|f - g_{i(f)}|_\infty \leq \epsilon m(\epsilon)^{-1}$. If $m \leq \log \epsilon / \log c_0$, we have

$$\left| f - \prod_{\ell=1}^m g_{i(f_\ell)} \right|_\infty = \left| \prod_{\ell=1}^m f_\ell - \prod_{\ell=1}^m g_{i(f_\ell)} \right|_\infty \leq c_1^{m-1} m \epsilon m(\epsilon)^{-1} \leq \epsilon.$$

As a result, the family $\{0, \prod_{\ell=1}^k g_\ell, k = 1, \dots, m(\epsilon)\}$ is a family of centers of balls of radius at most ϵ that are sufficient to cover $\mathcal{F}_0^{\text{prod}}$. It follows that

$$\mathcal{N}(\mathcal{F}_0^{\text{prod}}, \epsilon, |\cdot|_\infty) \leq \mathcal{N}(\mathcal{F}_0, m(\epsilon)\epsilon, |\cdot|_\infty)^{m(\epsilon)+1} \lesssim (\epsilon m(\epsilon))^{-2m(\epsilon)/s}.$$

Step 4) We have established $\mathcal{F} \subseteq \mathcal{F}_0^{\text{prod}}$ and therefore

$$\begin{aligned}\epsilon(\mathcal{F}) &= \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon \leq \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}_0^{\text{prod}}, \epsilon, |\cdot|_\infty)) d\epsilon \\ &\lesssim \int_0^1 \log(\epsilon m(\epsilon))^{-2m(\epsilon)/s} d\epsilon \lesssim \int_0^1 (\log \epsilon)^2 d\epsilon < \infty.\end{aligned}$$

The proof of Proposition 2.3 is complete.

7 Proofs of Section 3 and 4

7.1 Proof of Theorem 2.10

Remember that the condition $r_N \leq N^{-1/2}$ is in force in this section.

Preliminaries

We first write a standard bias-variance decomposition in squared-error loss, based upon the stability result of Corollary 2.7.

Lemma 2.28. *Let $h \in \mathcal{G}_1^N$. If \widehat{g}_h^N is specified with a bounded and compactly supported kernel K , we have*

$$\mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \lesssim \mathcal{B}_h^N(g)(t, a)^2 + \mathbf{V}_h^N,$$

where $B_h(g)(t, a)$ and \mathbf{V}_h^N are defined in (2.28) and (2.21) respectively.

Proof. Write $\widehat{g}_h^N(t, a) - g(t, a) = I + II$, with

$$I = \int_0^\infty K_h(u - a)g(t, u)du - g(t, a)$$

and

$$II = \int_{\mathbb{R}_+} K_h(u - a)(Z_t^N(du) - g(t, u))du.$$

We have $I^2 \leq B_h(g)(t, a)^2$. For the stochastic term, we have

$$|II| \leq \mathcal{W}_{K_h(t-a-\cdot)}^N(\mathcal{F})_t \tag{2.56}$$

Moreover

$$|K_h(t - a - \cdot)|_\infty \leq |K_h(t - a - \cdot)|_\infty = h^{-1}|K|_\infty \lesssim |K_h(t - a - \cdot)|_1 N^{1/2} \tag{2.57}$$

as soon as $h^{-1} \lesssim N^{1/2}$ since $|K_h(t - a - \cdot)|_1 = |K|_1 = 1$. This condition is true for any $h \in \mathcal{G}_1^N$ using the fact that K is bounded and compactly supported. We may then apply Corollary 2.7 and obtain

$$\mathbb{E}[II^2] \lesssim |K_h(t - a - \cdot)|_{1, \infty}^2 N^{-1} \lesssim (C^* N^{-1/2} |K_h|_{1, \infty})^2 = \mathbf{V}_h^N.$$

□

Completion of proof of Theorem 2.10

We essentially repeat the main argument of the Goldenshluger-Lepski method (see *e.g.* [28, 29] for the pointwise risk) in a setting that we need to adapt to our context.

Step 1) For any $h \in \mathcal{G}_1^N$, forcing $\widehat{g}_h^N(t, a)$ in the risk decomposition and by definition of $A_h^N(t, a)$ and $\widehat{h}^N(t, a)$, we successively have

$$\begin{aligned}
& \mathbb{E}[(\widehat{g}_\star^N(t, a) - g(t, a))^2] \\
& \lesssim \mathbb{E}[(\widehat{g}_\star^N(t, a) - \widehat{g}_h^N(t, a))^2] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\
& \lesssim \mathbb{E}[\{(\widehat{g}_{\widehat{h}^N(t, a)}^N(t, a) - \widehat{g}_h^N(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{\widehat{h}^N(t, a)}^N\}_+ + \mathbf{V}_h^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\
& \lesssim \mathbb{E}[\mathbf{A}_{\max(\widehat{h}^N(t, a), h)}^N(t, a) + \mathbf{V}_h^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\
& \lesssim \mathbb{E}[\mathbf{A}_h^N(t, a)] + \mathbf{V}_h^N + \mathbb{E}[\mathbf{A}_{\widehat{h}^N(t, a)}^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\
& \lesssim \mathbb{E}[\mathbf{A}_h^N(t, a)] + \mathbf{V}_h^N + \mathcal{B}_h^N(g)(t, a)^2
\end{aligned}$$

where we applied Lemma 2.28 to obtain the last line.

Step 2) We first estimate $A_h^N(t, a)$. Write $g_h(t, a)$ for $\int_{\mathbb{R}_+} K_h(u - a)g(t, u)du$. For $h, h' \in \mathcal{G}_1^N$ with $h' \leq h$, since

$$\begin{aligned}
& (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 \\
& \leq 4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 + 4(g_h(t, a) - g(t, a))^2 + 4(g_{h'}(t, a) - g(t, a))^2 + 4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2,
\end{aligned}$$

we have

$$\begin{aligned}
& (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N \\
& \leq 8\mathcal{B}_h^N(g)(t, a)^2 + (4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N) + (4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N).
\end{aligned}$$

using $h' \leq h$ in order to bound $(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2$ by the bias at scale h . It follows that

$$\begin{aligned}
& (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N \\
& \leq 8\mathcal{B}_h^N(g)(t, a)^2 + 4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N + 4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N,
\end{aligned}$$

and taking maximum over $h' \leq h$, we obtain

$$\begin{aligned}
& \max_{h' \leq h} \{(\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N\}_+ \tag{2.58} \\
& \leq 8\mathcal{B}_h^N(g)(t, a)^2 + \{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N\}_+ + \max_{h' \leq h} \{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N\}_+.
\end{aligned}$$

Step 3) We estimate the expectation of the first stochastic term in the right-hand side of (2.58). Since $|\widehat{g}_h^N(t, a) - g_h(t, a)| \leq \mathcal{W}_{K_h(t-a-\cdot)}^N$, we successively have

$$\begin{aligned}
\mathbb{E}[\{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N\}_+] &= \int_0^\infty \mathbb{P}(4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N \geq \kappa) d\kappa \\
&= \int_0^\infty \mathbb{P}(|\widehat{g}_h^N(t, a) - g_h(t, a)| \geq \frac{1}{2}(\mathbf{V}_h^N + \kappa)^{1/2}) d\kappa \\
&\leq \int_0^\infty \mathbb{P}(\mathcal{W}_{K_h(t-a-\cdot)}^N \geq \frac{1}{2}(\mathbf{V}_h^N + \kappa)^{1/2}) d\kappa.
\end{aligned}$$

We may apply Theorem 2.6 with $w_2 = K_h(t - a - \cdot)$ since K is compactly supported and having (2.57) of Lemma 2.28 above. By the change of variable

$$\frac{1}{2}(\mathbb{V}_h^N + \kappa)^{1/2} = (1 + u)C''|K_h|_{1,\infty}N^{-1/2},$$

we then obtain

$$\begin{aligned} & \mathbb{E}[\{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbb{V}_h^N\}_+] \\ & \leq 8C''|K_h|_{1,\infty}N^{-1/2} \int_{\frac{1}{2C''}(\mathbb{V}_h^N)^{1/2}|K_h|_{1,\infty}^{-1}N^{1/2}-1}^{\infty} (1 + u) \min((e^u - 1)^{-1}, 1) du \\ & \lesssim \exp\left(-\frac{1}{2C''}(\mathbb{V}_h^N)^{1/2}|K_h|_{1,\infty}^{-1}N^{1/2}\right) \leq N^{-2} \end{aligned}$$

by definition of \mathbb{V}_h^N .

Step 4) For the second stochastic term, we use the rough estimate

$$\begin{aligned} \mathbb{E}[\max_{h' \leq h} \{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbb{V}_{h'}^N\}_+] & \leq \sum_{h' \leq h} \mathbb{E}[\{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbb{V}_{h'}^N\}_+] \\ & \lesssim \text{Card}(\mathcal{G}_1^N)N^{-2} \lesssim N^{-1} \end{aligned}$$

where we used Step 3) to bound each term $\mathbb{E}[\{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbb{V}_{h'}^N\}_+]$ independently of h together with $\text{Card}(\mathcal{G}_1^N) \lesssim N$. In conclusion, we have proved through Steps 2)-4) that $\mathbb{E}[\mathbf{A}_h^N(t, a)] \lesssim \delta_N$. Therefore, from Step 1), we conclude

$$\mathbb{E}[(\widehat{g}_*^N(t, a) - g(t, a))^2] \lesssim \mathcal{B}_h^N(g)(t, a)^2 + \mathbb{V}_h^N + \delta_N$$

for any $h \in \mathcal{G}_1^N$. The proof of Theorem 2.10 is complete.

7.2 Proof of Theorem 2.12

Preliminaries

We first study the behaviour of the process $\Gamma^N(dt, da)$ of death occurrences introduced in Section 3.2 and represented via (2.18).

Lemma 2.29. *With the notation of Section 2.2, we have*

$$\Gamma^N(dt, da) = N^{-1} \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} \delta_{a_i(Z_s^N)}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_s^N)), i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(dt, di, d\vartheta), \quad (2.59)$$

where \mathcal{Q}_2 is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$ with intensity $dt(\sum_{k \geq 1} \delta_k(di))d\vartheta$. Moreover, for nonnegative weights $w_1 \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$ and $w_2 \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$, we have

$$\left| \int_0^T \int_0^\infty w_1(s)w_2(s-u)(\Gamma^N(ds, du) - \mu(s, u)g(s, u)duds) \right| \leq \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T + |(\Delta_{w_1, w_2}^N)_T|, \quad (2.60)$$

where $t \mapsto (\Delta_{w_1, w_2}^N)_t$ is a square integrable martingale with predictable compensator

$$\langle \Delta_{w_1, w_2}^N \rangle_t = N^{-1} \int_0^t \int_0^\infty w_1(s)^2 w_2(s-u)^2 \mu(s, u) Z_s^N(du) ds. \quad (2.61)$$

Proof. The representation (2.59) is straightforward. We add and subtract in the left-hand side of (2.60) the term $\int_0^T \int_0^\infty w_1(s)w_2(s-u)\mu(s,u)Z_s^N(du)ds$ and obtain the desired inequality with

$$\begin{aligned} (\Delta_{w_1, w_2}^N)_t &= \int_0^t \int_0^\infty w_1(s)w_2(s-u)(\Gamma^N(ds, du) - \mu(s, u)Z_s^N(du)) \\ &= N^{-1} \int_0^t \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} w_1(s)w_2(s-u)\delta_{a_i(Z_{s-}^N)}(da)\mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_{s-}^N)), i \leq \langle NZ_{s-}^N, \mathbf{1} \rangle\}} \tilde{\mathcal{Q}}_2(ds, di, d\vartheta), \end{aligned}$$

where $\tilde{\mathcal{Q}}_2(ds, di, d\vartheta) = \mathcal{Q}_2(ds, di, d\vartheta) - ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$ is the associated compensated measure. Thus $(\Delta_{w_1, w_2}^N)_t$ is a martingale and (2.61) follows. \square

We next study the deviation of $(\Delta_{w_1, w_2}^N)_T$. Define

$$V_{w_1, w_2}^N = (4C^*(\log N)N^{-1/2}|w_1|_{1, \infty}|w_2|_{1, \infty})^2. \quad (2.62)$$

where C^* is the constant defined in (2.24) in Section 3.3. Let also

$$\chi_{w_1, w_2}^N = N^{-1}|w_1|_\infty|w_2|_\infty|\mu|_\infty$$

and

$$\xi_{w_1, w_2}^N = 16N^{-1}|\mu|_\infty|g|_\infty|w_1|_2^2|w_2|_2^2(V_{w_1, w_2}^N)^{-1/2}(\log N) \quad (2.63)$$

Lemma 2.30. *For $u > 2^{-6}V_{w_1, w_2}^N(\log N)^{-2}$, we have*

$$\mathbb{P}(|(\Delta_{w_1, w_2}^N)_T| \geq u^{1/2}) \leq 2 \exp\left(-\frac{u^{1/2}}{2(\chi_{w_1, w_2}^N + \xi_{w_1, w_2}^N)}\right) + 2\mathbb{P}(N^{-1}|\mu|_\infty \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T \geq \frac{1}{2}\xi_{w_1, w_2}^N u^{1/2}).$$

Proof. We plan to apply a classical deviation inequality for martingales (see *e.g.* Lemma 2.1 in van de Geer [64] or the classical textbook by Shorak and Wellner [61]), namely:

$$\mathbb{P}((\Delta_{w_1, w_2}^N)_T \geq v, \langle \Delta_{w_1, w_2}^N \rangle_T \leq w) \leq \exp\left(-\frac{v^2}{2(v\chi_{w_1, w_2}^N + w)}\right) \quad (2.64)$$

for every $v, w \geq 0$, where $\chi_{w_1, w_2}^N = N^{-1}|w_1|_\infty|w_2|_\infty|\mu|_\infty$ is an almost-sure bound of the size of the jumps of $(\Delta_{w_1, w_2}^N)_T$. With $v = u^{1/2}$ and $w = \xi_{w_1, w_2}^N u^{1/2}$, inequality (2.64) gives

$$\mathbb{P}(|(\Delta_{w_1, w_2}^N)_T| \geq u^{1/2}) \leq 2 \exp\left(-\frac{u^{1/2}}{2(\chi_{w_1, w_2}^N + \xi_{w_1, w_2}^N)}\right) + 2\mathbb{P}(\langle \Delta_{w_1, w_2}^N \rangle_T \geq \xi_{w_1, w_2}^N u^{1/2}).$$

Inserting the term $N^{-1} \int_0^t \int_{\mathbb{R}_+} w_1(s)^2 w_2(s-u)^2 g(s, u) duds$ in (2.61), we obtain

$$\langle \Delta_{w_1, w_2}^N \rangle_T \leq N^{-1}|\mu|_\infty(|w_1|_2^2|w_2|_2^2|g|_\infty + \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T),$$

therefore

$$\mathbb{P}(\langle \Delta_{w_1, w_2}^N \rangle_T \geq \xi_{w_1, w_2}^N u^{1/2}) \leq \mathbb{P}(N^{-1}|\mu|_\infty \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T \geq \frac{1}{2}\xi_{w_1, w_2}^N u^{1/2})$$

as soon as

$$N^{-1}|\mu|_\infty|g|_\infty|w_1|_2^2|w_2|_2^2 < \frac{1}{2}\xi_{w_1, w_2}^N u^{1/2}, \quad (2.65)$$

but by definition of ξ_{w_1, w_2}^N in (2.63), this condition is equivalent to $u > 2^{-6}V_{w_1, w_2}^N(\log N)^{-2}$. \square

Under Assumption 2.11, we have a uniform lower bound on $g(t, a)$.

Lemma 2.31. *Work under Work under Assumptions 2.1 and 2.11. Then, there exists $\epsilon > 0$ depending on $\delta(t, a)$ defined in (2.26) and (2.27) and $|\mu|_\infty$ and T such that $g(t, a) \geq \epsilon$.*

The proof uses an explicit representation of $g(t, a)$ established in Proposition 2.16 and is delayed until Appendix 8.3.

Completion of proof of Theorem 2.12

Let $(h, \mathbf{h}) \in \mathcal{G}_1^N \times \mathcal{G}_2^N$ and set $\pi(t, a) = \mu(t, a)g(t, a)$.

Step 1) We plan to use the following decomposition

$$\widehat{\mu}_{h, \mathbf{h}}^N(t, a)_\varpi - \mu(t, a) = I + II,$$

with

$$I = \frac{\pi(t, a)(g(t, a) - \widehat{g}_h^N(t, a) \vee \varpi)}{g(t, a)\widehat{g}_h^N(t, a) \vee \varpi}$$

and

$$II = \frac{(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a))g(t, a)}{g(t, a)\widehat{g}_h^N(t, a) \vee \varpi}.$$

First, we have

$$|I| \leq (\epsilon\varpi)^{-1}|\mu|_\infty|g|_\infty|g(t, a) - \widehat{g}_h^N(t, a) \vee \varpi| \leq (\epsilon\varpi)^{-1}|\mu|_\infty|g|_\infty|g(t, a) - \widehat{g}_h^N(t, a)|$$

thanks to Lemma 2.31 as soon as $\varpi \leq \epsilon \leq g(t, a)$. In the same way,

$$|II| \leq (\epsilon\varpi)^{-1}|g|_\infty|\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a)|$$

follows. Picking $h = \widehat{h}^N(t, a)$, $\mathbf{h} = \widehat{\mathbf{h}}^N(t, a)$ and taking square and expectation, we have thus established

$$\mathbb{E}[(\widehat{\mu}_{\star}^N(t, a)_\varpi - \mu(t, a))^2] \lesssim \mathbb{E}[(\widehat{g}_{\widehat{h}^N(t, a)}^N(t, a) - g(t, a))^2] + \mathbb{E}[(\widehat{\pi}_{\widehat{\mathbf{h}}^N(t, a)}^N(t, a) - \pi(t, a))^2] \quad (2.66)$$

as soon as $\varpi \leq \epsilon$. By Theorem 2.10, we already have the desired bound for the first term in the right-hand side of (2.66).

Step 2) We study the second term in the right-hand side of (2.66). For any $\mathbf{h} \in \mathcal{G}_2^N$, repeating Step 1) of the proof of Theorem 2.10, we have

$$\mathbb{E}[(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a))^2] \lesssim \mathbb{E}[\mathbf{A}_{\mathbf{h}}^N(t, a)] + \mathbf{V}_{\mathbf{h}}^N + \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2.$$

In order to estimate $\mathbb{E}[\mathbf{A}_{\mathbf{h}}^N(t, a)]$, we repeat Step 2) of the proof of Theorem 2.10 and obtain

$$\begin{aligned} & \max_{\mathbf{h}' \leq \mathbf{h}} \{ (\widehat{\pi}_{\mathbf{h}}^N(t, a) - \widehat{\pi}_{\mathbf{h}'}^N(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N - \mathbf{V}_{\mathbf{h}'}^N \}_+ \\ & \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2 + \{4(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi_{\mathbf{h}}(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N\}_+ + \max_{\mathbf{h}' \leq \mathbf{h}} \{4(\widehat{\pi}_{\mathbf{h}'}^N(t, a) - \pi_{\mathbf{h}'}(t, a))^2 - \mathbf{V}_{\mathbf{h}'}^N\}_+. \end{aligned} \quad (2.67)$$

Step 3) We estimate the expectation of the first stochastic term in the right-hand side of the last inequality. Using the same trick as in (2.56), we have by (2.60) that

$$\{4(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi_{\mathbf{h}}(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N\}_+ \lesssim I + II,$$

with

$$I = \{8\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N\}_+$$

and

$$II = \{8(\Delta_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N)_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N\}_+.$$

We bound each term separately. First, we have

$$\begin{aligned} \mathbb{E}[I] &= \int_0^\infty \mathbb{P}(8\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N \geq \kappa) d\kappa \\ &= \int_0^\infty \mathbb{P}(\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T \geq \frac{1}{2\sqrt{2}}(\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N + \kappa)^{1/2}) d\kappa \lesssim N^{-3} \end{aligned}$$

applying Theorem 2.6 with $w_1 = H_{h_1}(\cdot - t)$ and $w_2 = K_{h_2}(\cdot - (t - a))$ in the same way as Step 3) in the proof of Theorem 2.10. As for II , we have

$$\mathbb{E}[II] = \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \mathbb{P}(|\Delta_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N)_T| \geq \frac{1}{2\sqrt{2}}\kappa^{1/2}) d\kappa \quad (2.68)$$

and we plan to apply Lemma 2.30 with $w_1 = H_{h_1}(\cdot - t)$ and $w_2 = K_{h_2}(\cdot - (t - a))$. Setting $u = \frac{1}{8}\kappa$, the condition of Lemma 2.30 is fulfilled as soon as $\kappa > 8 \cdot 2^{-6} V_{H_{h_1}, K_{h_2}}^N (\log N)^{-2} = \frac{1}{8} V_{H_{h_1}, K_{h_2}}^N (\log N)^{-2}$ which is the case here since the integral in (2.68) above is taken for $\kappa \geq \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N = \frac{1}{2} V_{H_{h_1}, K_{h_2}}^N$. It follows that

$$\mathbb{E}[II] \leq III + IV,$$

with

$$III = 2 \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \exp\left(-\frac{\kappa^{1/2}}{4\sqrt{2}(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)}\right) d\kappa$$

and

$$IV = 2 \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \mathbb{P}(N^{-1}|\mu|_\infty \mathcal{W}_{(H_{h_1})^2, (K_{h_2})^2}(\mathcal{F})_T \geq \frac{1}{4\sqrt{2}}\xi_{H_{h_1}, K_{h_2}}^N \kappa^{1/2}) d\kappa.$$

First, we write

$$III = 4(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^2 \int_{v_N}^\infty \kappa e^{-\frac{\kappa}{4\sqrt{2}}} d\kappa$$

with

$$v_N = \frac{\sqrt{2}}{2}(\mathbf{V}_{\mathbf{h}}^N)^{1/2}(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^{-1}.$$

Note that

$$(V_{H_{h_1}, K_{h_2}}^N)^{1/2} = h_1^{-1/2} h_2^{-1/2} N^{-1/2} (\log N) 4C^* |H|_{1, \infty} |K|_{1, \infty}.$$

It follows that

$$\begin{aligned} &\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N \\ &= N^{-1} |H_{h_1}|_\infty |K_{h_2}|_\infty |\mu|_\infty + 16N^{-1} |\mu|_\infty |g|_\infty |H_{h_1}|_2^2 |K_{h_2}|_2^2 (V_{|H_{h_1}|, |K_{h_2}|}^N)^{-1/2} (\log N) \\ &= N^{-1} h_1^{-1} h_2^{-1} |\mu|_\infty |H|_\infty |K|_\infty + N^{-1/2} h_1^{-1/2} h_2^{-1/2} 4C^* |\mu|_\infty |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1, \infty} |K|_{1, \infty}}. \end{aligned}$$

By definition of \mathcal{G}_2^N we have $h_i \geq N^{-1/2}$ hence

$$(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^2 \leq (|\mu|_\infty |H|_\infty |K|_\infty + 4C^* |\mu|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}})^2$$

follows and the term in front of the integral in *III* is bounded. Moreover,

$$\begin{aligned} v_N &= \frac{\sqrt{2}}{2} \frac{(\log N) 4C^* |H|_{1,\infty} |K|_{1,\infty}}{N^{-1/2} h_1^{-1/2} h_2^{-1/2} |\mu|_\infty |H|_\infty |K|_\infty + 4C^* |\mu|_\infty |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}}} \\ &\geq \frac{\sqrt{2}}{2} \frac{C^* |H|_{1,\infty} |K|_{1,\infty}}{|\mu|_\infty (|H|_\infty |K|_\infty + 4C^* |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}})} (\log N) = C^{(3)} \log N \end{aligned}$$

say. Since

$$\int_{v_N}^\infty \kappa e^{-\frac{\kappa}{4\sqrt{2}}} d\kappa \lesssim v_N e^{-\frac{1}{4\sqrt{2}} v_N} \lesssim (\log N) N^{-C^{(3)}}$$

it suffices to check that $C^{(3)} > 2$ in order to have that *III* is smaller in order than N^{-2} and thus asymptotically negligible. We finally bound the term *IV*. Applying Theorem 2.6 with $(w_1, w_2) = (H_{h_1}^2, K_{h_2}^2)$, by the change of variable

$$\frac{N}{4\sqrt{2}|\mu|_\infty} \xi_{H_{h_1}, K_{h_2}}^N \kappa^{1/2} = (1+u) C'' |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty} N^{-1/2}$$

we obtain that *IV* is of order

$$y_N \int_{z_N}^\infty (1+u) \min((\exp(u)-1)^{-1}, 1) du,$$

with

$$y_N = ((\xi_{H_{h_1}, K_{h_2}}^N)^{-1} |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty} N^{-3/2})^2 \quad \text{and} \quad z_N = \frac{N^{3/2} \xi_{H_{h_1}, K_{h_2}}^N (V_{\mathbf{h}}^N)^{1/2}}{8C'' |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty}} - 1.$$

Straightforward computations show that $y_N \lesssim h_1^{-2} h_2^{-2} N^{-2} \lesssim 1$ by construction of \mathcal{G}_2^N . Finally

$$z_N = \frac{8|g|_\infty |H|_2 |K|_2}{C'' |H|_\infty |K|_\infty} \frac{\log N}{h_1^{1/2} h_2^{1/2} N^{-1/2}} - 1 \geq C^{(4)} \log N - 1.$$

say. One can check that $C^{(4)} > 2$ and we can therefore conclude that *IV* also has a negligible order.

Step 4) The control of the second term in the right-hand side of (2.67) is done in the same way as in Step 4) of the proof of Theorem 2.10 and only inflates the previous bound by a factor or order $\text{Card}(\mathcal{G}_2^N) \lesssim N^2$. In turn $\mathbb{E}[\mathbf{A}_{\mathbf{h}}^N(t, a)] \lesssim N^{-1}$ and we have established by Step 2) that for any $\mathbf{h} \in \mathcal{G}_2^N$,

$$\mathbb{E}[(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a))^2] \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2 + V_{\mathbf{h}}^N + \delta_N \quad (2.69)$$

holds true with $\delta_N \lesssim N^{-1}$. Putting together Step 1) and Theorem 2.10 completes the proof.

7.3 Proof of Theorem 2.17

Preliminaries

We let $(Z_t)_{0 \leq t \leq T}$ denote the canonical process on⁴ $\mathbb{D}([0, T], \mathcal{M}_{F+})$ endowed with the weak topology and equipped with its Borel sigma-field. If Υ is a probability measure on \mathcal{M}_{F+} and if $b, \mu \in \mathcal{L}_{\mathcal{D}}^{\infty}$, we write $\mathbb{P}_{b, \mu, \Upsilon}^N$ for the (necessarily unique) probability measure on $\mathbb{D}([0, T], \mathcal{M}_{F+})$ under which $(Z_t)_{0 \leq t \leq T}$ is a weak solution to (2.7) with $\mathcal{L}(Z_0) = \Upsilon$.

Proposition 2.32. *For $i = 1, 2$, let $b_i, \mu_i \in \mathcal{L}_{\mathcal{D}}^{\infty}$ such that $\text{supp}(b_2) \subset \text{supp}(b_1)$ and $\text{supp}(\mu_2) \subset \text{supp}(\mu_1)$. For any initial condition $\mathcal{L}(Z_0) = \Upsilon$, we have*

$$\|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV} \lesssim N^{1/2} (|b_1^{-1}b_2 - 1|_2 + |\mu_1^{-1}\mu_2 - 1|_2),$$

where $\|\cdot\|_{TV}$ denotes total variation distance, up to an explicitly computable constant that only depends on μ_1 and b_1 .

Proof. The proof is classical, and we only sketch it. Thanks to the Doléans-Dade exponential for semimartingales (see e.g. [39] or Löcherbach [49, 48] in the context of birth and death processes) and abbreviating $f(s, a_i(Z_s^-))$ by $f^i(s)$, we have

$$\begin{aligned} \frac{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N}{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} &= N^{-1} \int_0^T \int_0^{\infty} (b_2(s, a) - b_1(s, a) + \mu_2(s, a) - \mu_1(s, a)) Z_s^N(da) ds \\ &+ \int_0^T \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} \mathbf{1}_{\{i \leq \langle N Z_s^N, \mathbf{1} \rangle\}} \left(\mathbf{1}_{\{0 \leq \vartheta \leq b_1^i(s)\}} \log \frac{b_2^i(s)}{b_1^i(s)} + \mathbf{1}_{\{b_1^i(s) \leq \vartheta \leq \mu_1^i(s)\}} \log \frac{\mu_2^i(s)}{\mu_1^i(s)} \right) \mathcal{Q}_1(ds, di, d\vartheta), \end{aligned}$$

where \mathcal{Q}_1 is a Poisson random measures on $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$ with intensity $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$ under $d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N$. By Pinsker's inequality, it follows that

$$\begin{aligned} &\|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV}^2 \\ &\leq \frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[\log \frac{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N}{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N} \right] \\ &= -\frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[\log \frac{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N}{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \right] \\ &= \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[\int_0^T \int_0^{\infty} (b_2 - b_1 + \mu_2 - \mu_1 - b_1 \log \frac{b_2}{b_1} - \mu_1 \log \frac{\mu_2}{\mu_1})(s, a) Z_s^N(da) \right] \\ &= \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[\int_0^T \int_0^{\infty} (b_1 \varphi(b_1^{-1}b_2 - 1) + \mu_1 \varphi(\mu_1^{-1}\mu_2 - 1))(s, a) Z_s^N(da) \right], \end{aligned}$$

with $\varphi(x) = x - \log(1+x) \leq x^2$ for $x \geq 0$. Therefore

$$\begin{aligned} \|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV}^2 &\lesssim \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[\int_0^T \int_0^{\infty} ((b_1^{-1}b_2 - 1)^2 + (\mu_1^{-1}\mu_2 - 1)^2)(s, a) Z_s^N(da) \right] \\ &\lesssim N (|b_1^{-1}b_2 - 1|_2^2 + |\mu_1^{-1}\mu_2 - 1|_2^2) \end{aligned}$$

and Proposition 2.32 is proved. \square

⁴remember that \mathcal{M}_{F+} denotes the set of positive finite measures on \mathbb{R}_+

Representation of g in terms of (g_0, b, μ)

We need some notation. Let

$$L_{b,\mu}(t, a) = b(t, a) \exp\left(-\int_{t-a}^t \mu(s, s-t+a) ds\right) \text{ for } (t, a) \in \mathcal{D}_L,$$

$$M_{b,\mu,g_0}(t) = \int_0^\infty b(t, t+u) g_0(u) \exp\left(-\int_0^t \mu(s, u+s) ds\right) du, \text{ for } t \in [0, T],$$

and define $B_{b,\mu,g_0} : [0, T] \rightarrow \mathbb{R}_+$ as the solution to the integral equation

$$B_{b,\mu,g_0}(t) = M_{b,\mu,g_0}(t) + \int_0^t B_{b,\mu,g_0}(a) L_{b,\mu}(t, t-a) da \text{ for every } t \in [0, T]. \quad (2.70)$$

Note that Assumptions 2.1 and 2.15 ensure the existence and uniqueness of (2.70). Define next

$$g(t, a) = \begin{cases} g_0(a-t) \exp\left(-\int_0^t \mu(s, a-t+s) ds\right) & \text{on } \mathcal{D}_U \\ B_{b,\mu,g_0}(t-a) \exp\left(-\int_{t-a}^t \mu(s, a+s-t) ds\right) & \text{on } \mathcal{D}_L \end{cases} \quad (2.71)$$

and set for instance $g(t, a) = 0$ on $\{a = t\}$. One can check that g defined in (2.71) is a weak solution to the McKendricks Von Voester equation (2.8).

Completion of proof of Theorem 2.17

We follow a classical two-point lower bound argument using Le Cam's lemma: if \mathbb{P}_i , $i = 1, 2$ are two probability measures defined on the same probability space and $\Psi(\mathbb{P}_i) \in \mathbb{R}$ is a functional of \mathbb{P}_i , we have

$$\inf_F \max_{i=1,2} \mathbb{E}_{\mathbb{P}_i} [|F - \Psi(\mathbb{P}_i)|] \geq \frac{1}{2} |\Psi(\mathbb{P}_1) - \Psi(\mathbb{P}_2)| (1 - \|\mathbb{P}_1 - \mathbb{P}_2\|_{TV}), \quad (2.72)$$

where the infimum is taken over all estimators of $\Psi(\mathbb{P}_i)$, see *e.g.* [43] among many other references.

Step 1) To prove (2.30), we pick

$$g_0 \in \mathcal{H}_L^\nu(a), \quad b_0 \in \mathcal{H}_L^{\alpha,\beta}(t, a), \quad \mu_1 \in \mathcal{H}_L^{\gamma,\delta}(t, a) \cap \mathcal{L}_{\mathcal{D},\epsilon}^\infty$$

arbitrarily, together with a sequence Υ^N such that $N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$ almost-surely under Υ^N and $\Upsilon^N(da) \rightarrow g_0(da)$ weakly as $N \rightarrow \infty$. Next, define

$$\mu_2^N(s, u) = \mu_1(s, u) (1 + \psi_{t-a}^N(s, u)),$$

where

$$\psi_{t-a}^N(s, u) = cN^{-1/2} \tau_N^{1/2} \psi(\tau_N(s - u - (t - a))),$$

with $\tau_N = N^{1/(2s_{\text{death}}^- + 1)} = N^{1/(2 \max(\gamma, \delta) + 1)}$ and an infinitely many times differentiable nonnegative function ψ with compact support that satisfies $\psi(0) = 1$, $|\psi|_2^2 = 1$. Finally, pick $c > 0$ small enough so that the property

$$\mu_2^N \in \mathcal{H}_L^{\gamma,\delta}(t, a) \cap \mathcal{L}_{\mathcal{D},\epsilon}^\infty$$

holds, uniformly in N . This is possible since

$$|\psi_{t-a}^N(\cdot, t-a)|_{\mathcal{H}^\gamma(t)} = cN^{-1/2}\tau_N^{1/2+\gamma}|\psi|_{\mathcal{H}^\gamma(t)} \leq c|\psi|_{\mathcal{H}^\gamma(t)} \leq c|\psi|_{\mathcal{H}^\gamma(t,a)}$$

and

$$|\psi_{t-a}^N(t, \cdot)|_{\mathcal{H}^\delta(t-a)} = cN^{-1/2}\tau_N^{1/2+\delta}|\psi|_{\mathcal{H}^\delta(t-a)} \leq c|\psi|_{\mathcal{H}^\delta(t-a)} \leq c|\psi|_{\mathcal{H}^\delta(a)}.$$

By Proposition 2.32, we have

$$\|\mathbb{P}_{b_0, \mu_1, \Upsilon^N} - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}\|_{TV} \lesssim N^{1/2}|\mu_1^{-1}\mu_2^N - 1|_2 = N^{1/2}|\psi_{t-a}^N|_2 = c^{1/2}|\psi|_2^2 \leq \frac{1}{2} \quad (2.73)$$

say, for large enough N and sufficiently small c .

Step 2) Let $(t, a) \in \mathcal{D}_U$. We let

$$\Psi(\mathbb{P}_{b, \mu, \Upsilon}^N) = g(t, a) = \tilde{g}(t', a') = g_0(-a') \exp\left(-\int_0^{t'} \mu(s, s-a') ds\right)$$

by (2.71) above, with $(t', a') = (t, t-a) = \varphi(t, a)$. It follows that

$$\begin{aligned} & |\Psi(\mathbb{P}_{b_0, \mu_2^N, \Upsilon}^N) - \Psi(\mathbb{P}_{b_0, \mu_1, \Upsilon}^N)| \\ &= g_0(-a') \exp\left(-\int_0^{t'} \mu_1(s, s-a') ds\right) \left| \exp\left(-\int_0^{t'} \psi_{a'}^N(s, s-a') ds\right) - 1 \right| \\ &\geq g_0(-a') \exp\left(-(|\mu_1|_\infty + |\psi_{a'}^N|_\infty)t'\right) \int_0^{t'} \psi_{a'}^N(s, s-a') ds \\ &\geq \frac{1}{2} g_0(-a') \exp(-|\mu_1|_\infty t') cN^{-1/2}\tau_N^{1/2}\psi(0)t' \\ &\gtrsim N^{-s_{\text{dens}}^-/(2s_{\text{dens}}^-+1)} \end{aligned} \quad (2.74)$$

using $|e^{-x} - 1| \geq xe^{-x}$ for $x \geq 0$ and the fact that $e^{-t'|\psi_{a'}^N|_\infty} \geq \frac{1}{2}$ say, for sufficiently large N .

Step 3) Let $(t, a) \in \mathcal{D}_L$. We now have

$$\Psi(\mathbb{P}_{b, \mu, \Upsilon}^N) = g(t, a) = \tilde{g}(t', a') = B_{b, \mu, g_0}(a') \exp\left(-\int_{a'}^{t'} \mu(s, s-a') ds\right),$$

by (2.71) and where B_{b, μ, g_0} is defined in (2.70). It follows that

$$\begin{aligned} & |\Psi(\mathbb{P}_{b_0, \mu_2^N, \Upsilon}^N) - \Psi(\mathbb{P}_{b_0, \mu_1, \Upsilon}^N)| \\ &= |B_{b_0, \mu_2^N, g_0}(a') \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - B_{b_0, \mu_1, g_0}(a') \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right)| \\ &\geq |I| - |II|, \end{aligned}$$

with

$$I = B_{b_0, \mu_1, g_0}(a') \left(\exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right) \right)$$

and

$$II = (B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a')) \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right).$$

To bound I from below, we proceed as in Step 2). For simplicity, we assume moreover here that $b_0(t, a) = b_0$ is constant. We have $B_{b_0, \mu_1, g_0}(a') \geq M_{b_0, \mu_1, g_0}(t) \geq b_0 |g_0|_1 e^{-|\mu_1|_\infty a'}$ and in the same way as for (2.74) we can check that

$$\begin{aligned} & \left| \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right) \right| \\ & \geq \frac{1}{2} e^{-|\mu_1|_\infty (t'-a')} (t' - a') N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)} \end{aligned}$$

for large enough N hence

$$|I| \geq \frac{1}{2} b_0 |g_0|_1 e^{-|\mu_1|_\infty t'} (t' - a') c N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)}. \quad (2.75)$$

In order to bound II from above, we use the following technical facts that are checked in the same way as before: for every $(t, a) \in \mathcal{D}_L$, we have

$$\begin{aligned} |M_{b_0, \mu_2^N, g_0}(a') - M_{b_0, \mu_1, g_0}(a')| & \leq b_0 |\mu_1|_\infty T \int_0^\infty g_0(u) \psi_{(t, a)}^N(-u) du \\ & \leq b_0 |\mu_1|_\infty T |g_0|_\infty |\psi_{t-a}^N|_1 \\ & = b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 \ll N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)} \end{aligned} \quad (2.76)$$

and

$$|L_{b_0, \mu_2^N, g_0}(a', a' - a) - L_{b_0, \mu_1, g_0}(a', a' - a)| \leq b_0 |\mu_1|_\infty c N^{-1/2} \tau_N^{1/2} \psi(\tau_N(u - a'))(a' - u) \quad (2.77)$$

and since $B_{b_0, \mu, g_0}(t) \leq b_0 |g_0|_1 + b_0 \int_0^t B_{b_0, \mu, g_0}(s) ds$ for every $t \in [0, T]$, we infer

$$B_{b_0, \mu, g_0}(t) \leq b_0 |g_0|_1 e^{b_0 T} \quad (2.78)$$

by Grönwall lemma. It follows that

$$\begin{aligned} & B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a') = M_{b_0, \mu_2^N, g_0}(a') - M_{b_0, \mu_1, g_0}(a') \\ & + \int_0^{a'} B_{b_0, \mu_1, g_0}(a) (L_{b_0, \mu_2^N, g_0}(a', a' - a) - L_{b_0, \mu_1, g_0}(a', a' - a)) da \\ & + \int_0^{a'} L_{b_0, \mu_2^N, g_0}(a', a' - a) (B_{b_0, \mu_2^N, g_0}(a) - B_{b_0, \mu_1, g_0}(a)) da. \end{aligned}$$

Taking absolute values and using (2.76), (2.77) and (2.78), we derive

$$\begin{aligned} |B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a')| & \leq b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 \\ & + b_0^2 |g_0|_1 e^{b_0 T} |\mu_1|_\infty c N^{-1/2} \tau_N^{1/2} \int_0^{a'} \psi(\tau_N(u - a'))(a' - u) du \\ & + b_0 \int_0^{a'} |B_{b_0, \mu_2^N, g_0}(s) - B_{b_0, \mu_1, g_0}(s)| ds. \end{aligned}$$

Using $\tau_N^{1/2} \int_0^{a'} \psi(\tau_N(u - a'))(a' - u) du \leq \tau_N^{-1/2} T |\psi|_1$, we derive

$$|B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a')| \leq b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 (1 + b_0 e^{b_0 T}) e^{b_0 a'}$$

by Grönwall lemma again. We conclude

$$|II| \ll N^{-s_{\text{dens}}^-/(2s_{\text{dens}}^-+1)} \quad (2.79)$$

Comparing (2.75) and (2.79), we see that

$$|\Psi(\mathbb{P}_{b_0, \mu_2^N}^N, \Upsilon) - \Psi(\mathbb{P}_{b_0, \mu_1}^N, \Upsilon)| \gtrsim N^{-s_{\text{dens}}^-/(2s_{\text{dens}}^-+1)} \quad (2.80)$$

Step 4). Combining (2.74) or (2.80) with (2.72) and (2.73), we successively obtain

$$\begin{aligned} \sup_{b, \mu, g_0} \mathbb{E}_{\mathbb{P}_{b, \mu, \Upsilon^N}^N} [|F - g(t, a)|] &\geq \frac{1}{2} \max_{i=1,2} \mathbb{E}_{\mathbb{P}_{b_0, \mu_i, \Upsilon^N}^N} [|F - \Psi(\mathbb{P}_{b_0, \mu_i}^N, \Upsilon^N)|] \\ &\geq \frac{1}{4} |\Psi(\mathbb{P}_{b_0, \mu_1}^N, \Upsilon^N) - \Psi(\mathbb{P}_{b_0, \mu_2^N}^N, \Upsilon^N)| (1 - \|\mathbb{P}_{b_0, \mu_1, \Upsilon^N}^N - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}^N\|_{TV}) \\ &\gtrsim N^{-s_{\text{death}}^-/(2s_{\text{death}}^-+1)} \end{aligned}$$

and (2.30) follows.

Step 5) To prove (2.31), we proceed as in Step 1), considering now the perturbation

$$\mu_2^N(s, u) = \mu_1(s, u)(1 + \psi_{t,a}^N(s, u)),$$

with

$$\psi_{t,a}^N(u) = cN^{-1/2} \tau_N^{-1/2} \psi(\tau_N(s-t)) (\tilde{\tau}_N)^{1/2} \psi(\tilde{\tau}_N(u-a))$$

and $\tau_N^\delta = (\tilde{\tau}_N)^\gamma = N^{s(\gamma, \delta)/(2s(\gamma, \delta)+1)}$ and an infinitely many times differentiable function ψ with compact support that satisfies $\psi(0) = 1$, $|\psi|_2^2 = 1$. Finally, we pick $c > 0$ small enough so that the property

$$\mu_2^N \in \mathcal{H}_L^{\gamma, \delta}(t, a) \cap \mathcal{L}_{\mathcal{D}, \epsilon}^\infty$$

holds, uniformly in N . This is possible since

$$|\psi_{(t,a)}^N|_{\mathcal{H}^\gamma(t)} \leq cN^{-1/2} \tau_N^{1/2+\gamma} (\tilde{\tau}_N)^{1/2} |\psi|_{\mathcal{H}^\gamma(t)} \lesssim c$$

and

$$|\psi_{(t,a)}^N|_{\mathcal{H}^\delta(a)} \leq cN^{-1/2} \tau_N^{1/2} (\tilde{\tau}_N)^{1/2+\delta} |\psi|_{\mathcal{H}^\delta(a)} \lesssim c$$

likewise. Finally, we note that

$$|\mu_2^N(t, a) - \mu_1(t, a)| \geq |\mu_1(t, a) \psi_{t,a}^N(t, a)| \geq \epsilon c N^{-1/2} \tau_N^{1/2} (\tilde{\tau}_N)^{1/2} \gtrsim N^{-s_{\text{death}}^-/(2s_{\text{death}}^-+1)} \quad (2.81)$$

and

$$\|\mathbb{P}_{b_0, \mu_1, \Upsilon^N}^N - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}^N\|_{TV} \lesssim N^{1/2} |\mu_1^{-1} \mu_2^N - 1|_2 = N^{1/2} |\psi_{(t,a)}^N|_2 = c^{1/2} \leq \frac{1}{2} \quad (2.82)$$

say, for sufficiently small $c > 0$, by Proposition 2.32, which conditions are satisfied since μ_1 and μ_2^N are bounded below. The end of the proof is similar to that of Step 4) with $\Psi(\mathbb{P}_{b, \mu, \Upsilon^N}^N) = \mu(t, a)$ together with the bounds (2.81) and (2.82). Therefore (2.31) is proved and Theorem 2.17 follows.

7.4 Proof of Theorem 2.18

By (i) of Proposition 2.16 the smoothness assumptions on (b, μ, g) imply

$$u \mapsto g(t, u) \in \mathcal{H}_{L'}^{\min(\alpha, \beta, \gamma+1, \delta)}(a) \text{ for } (t, a) \in \mathcal{D}_L, \quad (2.83)$$

and

$$u \mapsto g(t, u) \in \mathcal{H}_{L'}^{\max(\gamma \wedge (\delta+1), \delta)}(a) \text{ for } (t, a) \in \mathcal{D}_U \quad (2.84)$$

for some L' that depends on L and the smoothness parameters only. For any $h \in \mathcal{G}_1^N$, by standard kernel approximation, see *e.g.* [63] the smoothness properties (2.83) and (2.84) together with the definition (2.32) of s_{dens}^+ imply

$$\left| \int_0^\infty K_h(u-a)g(t, u)du - g(t, a) \right| \lesssim h^{s_{\text{dens}}^+ \wedge \ell_0},$$

up to a constant that depends on K , s_{dens}^+ and L' only. It follows that

$$\mathcal{B}_h^N(g)(t, a)^2 \lesssim h^{2s_{\text{dens}}^+ \wedge \ell_0}.$$

We also have

$$\mathbb{V}_h^N \lesssim (\log N)^2 N^{-1} h^{-1}$$

up to a constant that depends on C'' of Theorem 2.6 and K . By Theorem 2.10, we conclude

$$\begin{aligned} \mathbb{E}[(\hat{g}_*^N(t, a) - g(t, a))^2] &\lesssim \min_{h \in \mathcal{G}_1^N} (h^{2s_{\text{dens}}^+ \wedge \ell_0} + (\log N)^2 N^{-1} h^{-1}) + \delta_N \\ &\lesssim \left(\frac{(\log N)^2}{N} \right)^{2s_{\text{dens}}^+ \wedge \ell_0 / (2s_{\text{dens}}^+ \wedge \ell_0 + 1)} \end{aligned}$$

using the definition of \mathcal{G}_1^N . Moreover, this estimate is uniform in (b, μ, g_0) . The proof of Theorem 2.18 is complete.

7.5 Proof of Theorem 2.19

Define $\tilde{\mu}$ via $\mu = \tilde{\mu} \circ \varphi$ and set $\tilde{\pi} = \tilde{\mu} \tilde{g}$.

Step 1) Write $\mu(t, a) = \tilde{\mu}(t', a') = \mu(t', t' - a')$ with $(t', a') = \varphi(t, a) = (t, t - a)$. The property $\mu \in \mathcal{H}_{L'}^{\gamma, \delta}(t, a)$ for every $(t, a) \in \mathcal{D}$ implies $\tilde{\mu} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \delta}(t', a')$ for every $(t', a') \in \varphi(\mathcal{D}) = \mathcal{D}$, for some other constant L' that depends on L . By (ii) of Proposition 2.16 it follows that

$$\tilde{\pi} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \min(\alpha, \beta, \gamma+1, \delta)}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_L$$

and

$$\tilde{\pi} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \delta}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_U.$$

Let $(t, a) \in \mathcal{D}_L$ so that $\varphi(t, a) \in \tilde{\mathcal{D}}_L$. By standard kernel approximation again, we infer

$$\begin{aligned} &\left| ((H \otimes K)_h \circ \varphi) \star \pi(t, a) - \pi(t, a) \right| \\ &= \left| (H \otimes K)_h \star \tilde{\pi}(\varphi(t, a)) - \tilde{\pi}(\varphi(t, a)) \right| \\ &= \left| \int_0^T \int_{\mathbb{R}_+} H_{h_1}(\varphi_1(t, a) - s) K_{h_2}(\varphi_2(t, a) - u) \tilde{\pi}(s, u) ds du - \tilde{\pi}(\varphi(t, a)) \right| \\ &\lesssim h_1^{\min(\gamma, \delta) \wedge \ell_0} + h_2^{\min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} \end{aligned}$$

up to a constant that depends on H, K, L' and the smoothness parameters only and where we have set $\varphi(t, a) = (\varphi_1(t, a), \varphi_2(t, a))$. Similarly, if $(t, a) \in \mathcal{D}_U$, we have

$$|((H \otimes K)_{\mathbf{h}} \circ \varphi) \star \pi(t, a) - \pi(t, a)| \lesssim h_1^{\min(\gamma, \delta) \wedge \ell_0} + h_2^{\delta \wedge \ell_0}.$$

It follows that

$$\mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2 \lesssim \begin{cases} h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} & \text{if } (t, a) \in \mathcal{D}_L \\ h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \delta \wedge \ell_0} & \text{if } (t, a) \in \mathcal{D}_U. \end{cases} \quad (2.85)$$

We also have

$$\mathbf{V}_{\mathbf{h}}^N \lesssim (\log N)^2 N^{-1} h_1^{-1} h_2^{-1} \quad (2.86)$$

up to a constant that depends on C'' of Theorem 2.6 and H, K .

Step 2) By Theorems 2.12 and 2.18, we have

$$\mathbb{E}[(\mu_{\star}^N(t, a)_{\varpi} - \mu(t, a))^2] \lesssim \left(\frac{(\log N)^2}{N}\right)^{2s_{\text{dens}}^+ \wedge \ell_0 / (2s_{\text{dens}}^+ \wedge \ell_0 + 1)} + \min_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\gamma)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) + \delta_N. \quad (2.87)$$

Moreover, by definition of s_L involved in (2.34), we have

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} + (\log N)^2 N^{-1} h_1^{-1} h_2^{-1}) \lesssim \left(\frac{(\log N)^2}{N}\right)^{2s_L \wedge \ell_0 / (2s_L \wedge \ell_0 + 1)}$$

and likewise, by definition of s_U involved in (2.34), we have

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\gamma, \delta) \wedge \ell_0} + (\log N)^2 N^{-1} h_1^{-1} h_2^{-1}) \lesssim \left(\frac{(\log N)^2}{N}\right)^{2s_U \wedge \ell_0 / (2s_U \wedge \ell_0 + 1)}.$$

Therefore, putting together (2.85) and (2.86) and using the definition of s_{death} in (2.34) we obtain

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\gamma)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) \lesssim \left(\frac{(\log N)^2}{N}\right)^{s_{\text{death}}^+(t, a) \wedge \ell_0 / (2s_{\text{death}}^+(t, a) \wedge \ell_0 + 1)}.$$

Since $s_{\text{dens}}^+ \geq s_{\text{death}}$, inequality (2.87) becomes

$$\mathbb{E}[(\mu_{\star}^N(t, a)_{\varpi} - \mu(t, a))^2] \lesssim \left(\frac{(\log N)^2}{N}\right)^{s_{\text{death}}(t, a) \wedge \ell_0 / (2s_{\text{death}}(t, a) \wedge \ell_0 + 1)} + \delta_N.$$

Since the estimate is uniform in (b, μ, g_0) and $\delta_N \lesssim N^{-1}$, this completes the proof of Theorem 2.19.

8 Appendix

8.1 Proof of Proposition 2.24

Preliminaries

For $x \geq 0$ and $q \geq 1$, define $\psi_q(x) = \exp(x^q) - 1$. Let also

$$\|\xi(f)\|_{\psi_q} = \inf \{c > 0, \mathbb{E}[\psi_q(c^{-1} \xi(f))] \leq 1\}$$

and

$$D = \text{diam}_d(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} d(f, g).$$

Proposition 2.33 (Theorem 11.2, Eq. (11.4) p. 302 in [44]). *In the setting of Proposition 2.24, if $\|\xi(f) - \xi(g)\|_{\psi_q} \leq d(f, g)$ and $E = \int_0^D \psi_q^{-1}(\mathcal{N}(\mathcal{F}, d, \epsilon)) d\epsilon < \infty$, then*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\xi(f)| \geq 8(E + u)\right) \leq \psi_q(u/D)^{-1},$$

provided $\xi(f_0) = 0$ for some $f_0 \in \mathcal{F}$.

We also recall the following bound based on a classical Chernoff bound argument, proof of which we omit. For $x \geq 0$, let $\tilde{\rho}(x) = (1 + x) \log(1 + x) - x$.

Lemma 2.34. *Let X be a non-negative random variable on some probability space equipped with a probability measure \mathbb{Q} . If, for some $k_1, k_2, k_3 > 0$, we have*

$$\mathbb{E}_{\mathbb{Q}}[e^{\lambda X}] \leq k_1 \exp(k_2 \rho(k_3 \lambda)) \text{ for every } \lambda \geq 0,$$

then, for every $u \geq 0$,

$$\mathbb{Q}(X \geq u) \leq k_1 \exp(-k_2 \tilde{\rho}(u/k_2 k_3)).$$

Proof of Proposition 2.24

Thanks to Proposition 2.33, all we need is an upper bound for $\|\xi(f) - \xi(g)\|_{\psi_1}$. Let $\kappa > 0$. We plan to apply Lemma 2.34 with $\mathbb{Q} = \mathbb{P}(\cdot | \mathcal{A}(\kappa))$, $X = |\xi(f) - \xi(g)|$, $k_1 = 2\mathbb{P}(\mathcal{A}(\kappa))$, $k_2 = c_1(1 + \kappa)$, $k_3 = c_2 d(f, g)$ and using (2.47). It follows that for every $u \geq 0$

$$\mathbb{P}(|\xi(f) - \xi(g)| \geq u) \leq 2 \exp(-c_1(1 + \kappa) \tilde{\rho}(u/c_1(1 + \kappa)c_2 d(f, g))) + \mathbb{P}(\mathcal{A}(\kappa)^c). \quad (2.88)$$

Now, let $c > 0$. We have

$$\begin{aligned} \mathbb{E}[\psi_1(c^{-1}|\xi(f) - \xi(g)|)] &= \mathbb{E}[\exp(c^{-1}|\xi(f) - \xi(g)|)] - 1 \\ &= \int_1^\infty \mathbb{P}(\exp(c^{-1}|\xi(f) - \xi(g)|) \geq \kappa) d\kappa \\ &= \int_0^\infty \mathbb{P}(|\xi(f) - \xi(g)| \geq c\kappa) e^\kappa d\kappa \\ &\leq 2 \int_0^\infty \exp(-c_1(1 + \kappa) \tilde{\rho}(c\kappa/c_1(1 + \kappa)c_2 d(f, g))) e^\kappa d\kappa + \frac{1}{2}, \end{aligned}$$

where we applied (2.88) with $u = c\kappa$ and used (2.46) for bounding the second term. It suffices then to pick $\varpi = \varpi(c_1, c_2) > 0$ such that

$$2 \int_0^\infty \exp(-c_1(1 + \kappa) \tilde{\rho}(\varpi(c_1, c_2)\kappa/c_1(1 + \kappa)c_2)) e^\kappa d\kappa \leq \frac{1}{2}. \quad (2.89)$$

Using (2.89) in the previous estimate with $c = \varpi d(f, g)$, we obtain

$$\mathbb{E}[\psi_1(\varpi d(f, g))^{-1} |\xi(f) - \xi(g)|] \leq 1$$

and therefore

$$\|\xi(f) - \xi(g)\|_{\psi_1} \leq \varpi d(f, g) = \tilde{d}(f, g),$$

say. We may then apply Proposition 2.33 with \tilde{d} instead of d and Proposition 2.24 follows.

Remark 2.35. In (2.89), we may choose $\varpi(c_1, c_2) = k\sqrt{c_1}c_2$ for some $k > 0$ that does not depend on c_1 nor c_2 . Indeed, since $\tilde{\rho}(x) \geq \frac{1}{4}x^2$ for $x \in [0, 1]$, given the ansatz $\varpi(c_1, c_2) = k\sqrt{c_1}c_2$ in (2.89), it suffices to show the existence of k satisfying $k \leq \sqrt{c_1}$ and

$$\int_0^\infty \exp\left(-\frac{k^2}{4} \frac{\kappa^2}{1+\kappa} + \kappa\right) d\kappa \leq \frac{1}{4}. \quad (2.90)$$

One can check that (2.90) holds for large enough k . A rough bound is $k = 2\sqrt{77}$, and therefore $c_1 \geq 308$ ensures the requirement $k \leq \sqrt{c_1}$.

8.2 Proof of Proposition 2.16

The behaviour of the solution $\xi_t(da) = g(t, a)da$ of the McKendricks Von Voester transport equation is studied in numerous textbooks, see *e.g.* [58]. The proof goes along a classical representation of g in terms of an auxiliary function solution to a certain renewal equation that enables one to study the pointwise smoothness of $(t, a) \mapsto g(t, a)$.

Preliminaries

We start with the following technical result, which is merely an observation:

Lemma 2.36. *If for some $\sigma, \tau > 0$ and for every $(t, a) \in \mathcal{D}$ we have $f \in \mathcal{H}^{\sigma, \tau}(t, a)$, then, for every $(t', a') \in \mathcal{D}$,*

$$(i) \ u \mapsto \int_0^u f(s, u) ds \in \mathcal{H}^{\min(\sigma+1, \tau)}(a'),$$

$$(ii) \ u \mapsto \int_0^{t'} f(s, u+s) ds \in \mathcal{H}^{\max(\sigma \wedge (\tau+1), \tau)}(a').$$

Proof. Property (i) is straightforward. To obtain (ii), we first write

$$G_{t'}(u) = \int_0^{t'} f(s, u+s) ds = \int_u^{u+t'} f(s-u, s) ds = \int_u^{u+t'} \tilde{f}(s, u) ds$$

with $\tilde{f}(s, u) = f(s-u, s)$, so that $\tilde{f} \in \mathcal{H}^{\min(\sigma, \tau), \sigma}(t' + a', a')$ for every $(t', a') \in \mathcal{D}$. Writing

$$\int_u^{u+t'} \tilde{f}(s, u) ds = \int_0^{u+t'} \tilde{f}(s, u) ds - \int_0^u \tilde{f}(s, u) ds$$

an applying (i), we obtain $u \mapsto \int_0^u \tilde{f}(s, u) ds \in \mathcal{H}^{\min(\min(\sigma, \tau)+1, \sigma)}(a') = \mathcal{H}^{\min(\sigma, \tau+1)}(a')$ for every $a' \in \mathbb{R}_+$. Similarly, $u \mapsto \int_u^{u+t'} \tilde{f}(s, u) ds \in \mathcal{H}^{\min(\sigma, \tau+1)}(a')$ therefore $G_{t'} \in \mathcal{H}^\tau(a') \in \mathcal{H}^{\min(\sigma, \tau+1)}$. But since $G_{t'} \in \mathcal{H}^\tau(a')$ trivially holds, we have in fact $G_{t'} \in \mathcal{H}^\tau(a') \cap \mathcal{H}^{\min(\sigma, \tau+1)}(a') = \mathcal{H}^{\max(\sigma \wedge (\tau+1), \tau)}(a')$. \square

Completion of proof of Proposition 2.16

For $\sigma, \tau > 0$, we write $f \in \mathcal{H}^{\sigma, \tau}$ if $f \in \mathcal{H}^{\sigma, \tau}(t, a)$ for every $(t, a) \in \mathcal{D}$.

Step 1) For fixed a , we have $(s, t) \mapsto \mu(s, a - t + s) \in \mathcal{H}^{\gamma \wedge \delta, \delta}$ hence by (i) of Lemma 2.36 we have $t \mapsto \int_0^t \mu(s, a - t + s) ds \in \mathcal{H}^{\min((\gamma \wedge \delta) + 1, \delta)} = \mathcal{H}^{\min(\gamma + 1, \delta)}$. For fixed t , we have $(s, a) \mapsto \mu(s, a - t) \in \mathcal{H}^{\gamma, \delta}$ hence $a \mapsto \int_0^t \mu(s, a - t + s) ds \in \mathcal{H}^{\min(\gamma, \delta + 1) \vee \delta}$ by (ii) of Lemma 2.36. It follows that

$$(t, a) \mapsto \exp\left(-\int_0^t \mu(s, a - t + s) ds\right) \in \mathcal{H}^{\min(\gamma + 1, \delta), \max(\gamma \wedge (\delta + 1), \delta)}.$$

Also $(t, a) \mapsto g_0(a - t) \in \mathcal{H}^{\nu, \nu} \subset \mathcal{H}^{\min(\gamma + 1, \delta), \max(\gamma \wedge (\delta + 1), \delta)}$ since $\nu \geq \max(\gamma, \delta) + 1$ hence the result on \mathcal{D}_U . In the same way, on \mathcal{D}_L , we have $t \mapsto \int_0^t \mu(s, a - t + s) ds \in \mathcal{H}^{\min(\gamma + 1, \delta)}$ and $t \mapsto \int_0^{t-a} \mu(s, a - t + s) ds \in \mathcal{H}^{\min(\gamma + 1, \delta)}$ by (i) of Lemma 2.36 hence

$$t \mapsto \int_{t-a}^t \mu(s, a - t + s) ds \in \mathcal{H}^{\min(\gamma + 1, \delta)}. \quad (2.91)$$

Moreover, $\int_{t-a}^t \mu(s, a + s - t) ds = -\int_0^a \mu(s + t, a + s) ds$ and $(s, a) \mapsto \mu(s, a + s - t) \in \mathcal{H}^{\gamma \wedge \delta, \delta}$ for fixed t , therefore

$$a \mapsto \int_{t-a}^t \mu(s, a - t + s) ds \in \mathcal{H}^{\min(\gamma + 1, \delta)} \quad (2.92)$$

by (i) of Lemma 2.36 likewise. Putting together (2.91) and (2.92), we conclude

$$(t, a) \mapsto \exp\left(-\int_{t-a}^t \mu(s, a - t + s) ds\right) \in \mathcal{H}^{\min(\gamma + 1, \delta), \min(\gamma + 1, \delta)}. \quad (2.93)$$

The property $b \in \mathcal{H}^{\alpha, \beta}$ together with (2.93) entail $L_{b, \mu} \in \mathcal{H}^{\min(\alpha, \gamma + 1, \delta), \min(\beta, \gamma + 1, \delta)}$ hence

$$(t, a) \mapsto L_{b, \mu}(t, t - a) \in \mathcal{H}^{\min(\alpha, \gamma + 1, \delta), \min(\alpha, \beta, \gamma + 1, \delta)}$$

and

$$t \mapsto \int_0^t B_{b, \mu, g_0}(a) L_{b, \mu}(t, t - a) da \in \mathcal{H}^{\min(\alpha, \beta, \gamma + 1, \delta)}, \quad (2.94)$$

follows by (i) of Lemma 2.36. Plainly,

$$t \mapsto M_{b, \mu, g_0}(t) \in \mathcal{H}^{\min(\alpha, \beta, \gamma + 1, \delta + 1)} \quad (2.95)$$

and putting together (2.94) and (2.95), we conclude

$$t \mapsto B_{b, \mu, g_0}(t) \in \mathcal{H}^{\min(\alpha, \beta, \gamma + 1, \delta)}. \quad (2.96)$$

hence $(t, a) \mapsto B_{b, \mu, g_0}(t - a) \in \mathcal{H}^{\min(\alpha, \beta, \gamma + 1, \delta), \min(\alpha, \beta, \gamma + 1, \delta)}$. The result of Proposition 2.16 (i) follows.

Step 2) Writing $(t', a') = \varphi(t, a) = (t, t - a)$, the representation (2.71) now becomes

$$g(t, a) = \tilde{g}(t', a') = \begin{cases} g_0(-a') \exp\left(-\int_0^{t'} \mu(s, s - a') ds\right) & \text{on } \tilde{\mathcal{D}}_U \\ B_{b, \mu, g_0}(a') \exp\left(-\int_{a'}^{t'} \mu(s, s - a') ds\right) & \text{on } \tilde{\mathcal{D}}_L. \end{cases} \quad (2.97)$$

On $\tilde{\mathcal{D}}_U$, we have $t' \mapsto \int_0^{t'} \mu(s, s - a') ds \in \mathcal{H}^{\min(\gamma, \delta)+1}$ and $a' \mapsto \int_0^{t'} \mu(s, s - a') ds \in \mathcal{H}^{\max(\gamma \wedge (\delta+1), \delta)}$ by (ii) of Lemma 2.36 for the second case, hence

$$(t', a') \mapsto \exp\left(-\int_0^{t'} \mu(s, s - a') ds\right) \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\gamma, \delta+1)}.$$

Since $(t', a') \mapsto g_0(-a') \in \mathcal{H}^{\infty, \nu}$ hence the result since $\nu \geq \delta$. Similarly, on $\tilde{\mathcal{D}}_L$, by (2.96), we have $(t', a') \mapsto B_{b, \mu, g_0}(a') \in \mathcal{H}^{\infty, \min(\alpha, \beta, \gamma+1, \delta)}$ and the same arguments as before yield

$$(t', a') \mapsto \int_{a'}^t \mu(s, s - a') ds \in \mathcal{H}^{\min(\gamma+1, \delta+1), \max(\gamma \wedge (\delta+1), \delta)}.$$

Combining these two properties gives the result on $\tilde{\mathcal{D}}_L$ and completes (ii) of Proposition 2.16.

8.3 Further estimates on the McKendricks Von Voester equation

The following result is a classical estimate of the renewal equation, see for instance [58].

Lemma 2.37 ([58], Theorem 2.2. in Chapter 2). *Work under Assumptions 2.1. We have*

$$\sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da \leq \int_0^\infty g_0(a) da e^{|b-\mu|_\infty T}$$

and

$$|g|_\infty \leq \max(|g_0|_\infty, |b|_\infty \sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da)$$

Proof of Lemma 2.31

On \mathcal{D}_U , by (2.71) in the proof of Proposition 2.16, we have

$$g(t, a) = g_0(a - t) \exp\left(-\int_0^t \mu(s, a - t + s) ds\right) \geq \delta(t, a) e^{-|\mu|_\infty T}$$

by (2.27) of Assumption 2.11. On \mathcal{D}_L , (2.71) yields the representation

$$g(t, a) = B_{b, \mu, g_0}(t - a) \exp\left(-\int_{t-a}^a \mu(s, a + s - t) ds\right) \geq B_{b, \mu, g_0}(t - a) e^{-|\mu|_\infty t}$$

and by (2.70), we further have

$$\begin{aligned} B_{b, \mu, g_0}(t - a) &\geq M_{b, \mu, g_0}(t - a) \\ &= \int_0^\infty b(t - a, t - a + u) g_0(u) \exp\left(-\int_0^{t-a} \mu(s, u + s) ds\right) du \\ &\geq \delta |\mathcal{U}(t, a)| e^{-|\mu|_\infty (t-a)} \end{aligned}$$

by (2.26) of Assumption 2.11. The proof of Lemma 2.31 is complete.

Acknowledgements We gratefully acknowledge insightful comments and discussions with Nicole El Karoui, O. Lepski and V. C. Tran.

CHAPTER 3

NONPARAMETRIC INFERENCE OF AGE-STRUCTURED MODELS IN A LARGE POPULATION LIMIT WITH INTERACTIONS, IMMIGRATION AND CHARACTERISTICS.

1 Introduction

1.1 Motivation

In the precedent chapter we are interested in the estimation of the mortality rate in a time dependent and age structured framework. However in biology, the models needed are more wide. This is for example the case in the well known logistic population [65], where the mortality rate depends on the state of the actual population. This belongs to a more general framework where there is the presence of interactions in the mortality rate. One model of this kind can be found in [41].

Another important point is the presence of characteristics, which can be for example the size of individual or their weight. This is for example the case of models where the quantity of interest is the size as in [24]. This is also of interest in actuarial science as highlighted in [8].

However the models change radically with interactions and characteristics. In a deterministic point of view, these models are widely studied as in [58, 67]. The stochastic model is also studied in term of probalistic point of view as in [62, 26, 17].

From a statistical point of view, there exists models taking into account the dependence between individuals thanks to covariate as in [19]. This is the case of the doubly stochastic Poisson process, like cox model, as we can find in [31]. There is also article which study the competition between the cause of death. This is a particular case of the characteristics case, since we can consider each cause of death as a marker of the individuals, as a characteristic. A good introduction to this subject can be found in [1]. We could also look into other fields to find some concentration inequalities, as in [7, 6].

Nevertheles, to our knowledge, there is no statistical studies of such a model with interactions and characteristics.

1.2 Setting

We investigate statistical inference of a stochastic time evolving density of a population alimeted by time inhomogeneous mortality, fertility, mutation and immigration in a large population limit.

The individuals will have characteristic (x, l) from a space $\mathcal{X} \times \mathcal{L}$ with measure $P(dx) \times Q(dl)$. \mathcal{X} is the space of the continuous characteristic and must be seen as a compact of \mathbb{R} . The multivariate case is not treated to keep simple notation. However this is possible to extend all our results to this case. \mathcal{L} is the space of the discrete characteristic with cardinal $L < \infty$.

The data are : (i) an initial condition Z_0^N such that $NZ_0^N \in \mathcal{M}_F$, the set of finite point measures on $\mathcal{U} = \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}$.

(ii) five functions

$$b, e, e_i : [0, T] \times \mathcal{U} \rightarrow \mathbb{R}_+$$

$$p : [0, T] \times \mathcal{U} \rightarrow [0, 1]$$

$$\mu : [0, T] \times \mathcal{U} \times \mathbb{R} \rightarrow \mathbb{R}_+$$

The fertility rate $b(s, a, x, l)$ accounts for the births produced by the individual having age a , at time s with characteristic (x, l) . $p(s, a, x, l)$ is the probability of mutation at birth. The swap or mutation rate, depending what is the characteristic, $e(s, a, x, l)$ accounts for the mutation (or the swap) of the individual having age a , at time s with characteristic (x, l) . e_i stands for the immigration rate. Finally the mortality rate $\mu(s, a, x, l, u)$ accounts for the death of an individual having age a , at time s with characteristic (x, l) and pressure u coming from the interactions.

(iii) three kernels

$$U : (\mathbb{R}_+ \times \mathcal{X} \times \mathcal{L})^2 \rightarrow \mathbb{R}$$

$$k_b, k_e : [0, T] \times \mathcal{U} \times \mathcal{X} \times \mathcal{L} \rightarrow \mathbb{R}_+$$

Where U is the interaction kernel. It accounts for the pressure coming from the other individuals in the population.

k_b and k_e are transition kernels. We suppose implicitly there exist such functions to define the kernels which are then $k_b(s, a, x, l, x', l')P(dx')Q(dl')$ and $k_e(s, a, x, l, x', l')P(dx')Q(dl')$.

1.3 The microscopic model and its large population limit

Any $Z \in \mathcal{M}_F$ admits the representation $Z = \sum_{i=1}^n \delta_{a_i, x_i, l_i}$ for some (ordered) $\{a_1, \dots, a_n\} \subset \mathbb{R}_+$. For real-valued function f defined on \mathcal{U} , we write

$$\langle Z, f \rangle = \int_{\mathcal{U}} f(a, x, l) Z(da, dx, dl) = \sum_{i=1}^n f(a_i, x_i, l_i).$$

In particular $n = \langle Z, \mathbf{1} \rangle$. In this setting, the distribution Z_0^N describes the renormalised state of the initial population and $N \langle Z_0^N, \mathbf{1} \rangle$ is the size of the population at time $t = 0$.

We work with the set of (measurable) functions

$$\mathcal{F}_b = \left\{ f : [0, T] \times \mathcal{U} \rightarrow \mathbb{R}, \sup_{t, a, x, l} |f(t, a, x, l)| < \infty \right\},$$

implicitly continued on $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathcal{L}$ by setting $f(t, a, x, l) = 0$ for $(t, a, x, l) \notin [0, T] \times \mathcal{U}$. We sometimes write $f_t(a, x, l) = f(t, a, x, l)$ when no confusion is possible and define the usual associated L^p -norms

$$|f|_p = \left(\int_0^T \int_{\mathcal{U}} |f(t, a, x, l)|^p dt da P(dx) Q(dl) \right)^{1/p} \quad \text{and} \quad |f|_{\infty} = \sup_{t, a, x, l} |f(t, a, x, l)|,$$

for $p \in [1, \infty]$, whenever well-defined. We will also use the interpolation quantity

$$|f|_{1, \infty} = (|f|_1 |f|_{\infty})^{1/2}. \quad (3.1)$$

The basic assumptions on the model are the following:

Assumption 3.1. *There exist a constant \mathcal{R} such that*

$$\max(|b|_{\infty}, |e|_{\infty}, |k_b|_{\infty}, |k_e|_{\infty}, |U|_{\infty}, |e_i|_{\infty}) \leq \mathcal{R}.$$

We also assume that

$$\forall (s, a, x, l) \in [0, T] \times \mathcal{U}, \quad |\mu(s, a, x, l, u_1) - \mu(s, a, x, l, u_2)| \leq \mathcal{R}|u_1 - u_2|$$

$|e_i|_1 < \infty$ and finally μ is at most linear in term of u , which means

$$|\mu(s, a, x, l)| \leq \mathcal{R}(1 + |u|)$$

Remark 3.2. *The assumption about the interaction seems to be strong. However we will often use $\mu(s, a, x, l, u) = \mu_r(s, a, x, l) + \eta u$. This is indeed the most natural way to introduce interactions. Remind that μ is a rate, so the add of new risk is additive.*

Assumption 3.3. *There exist $\alpha > 0$, such that $\sup_{N \geq 1} \mathbb{E}[\langle Z_0^N, \mathbf{1} \rangle^{1+\alpha}] < \infty$ and*

$$Z_0^N \rightarrow \xi_0 \quad \text{as } N \rightarrow \infty$$

narrowly, for some deterministic $\xi_0 \in \mathcal{M}_+$, the set positive finite measures on $\mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}$.

For $Z = \sum_{i=1}^n \delta_{a_i, x_i, l_i} \in \mathcal{M}_F$ we define the evaluation maps $a_i(Z) = a_i$, $x_i(Z) = x_i$, $l_i(Z) = l_i$ and for $t \geq 0$, the shift $\tau_t Z = \sum_{i=1}^n \delta_{a_i+t, x_i, l_i}$. Let Q_{μ} , Q_b , Q_e three independent Poisson random measures on $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}$ with intensity measures $ds (\sum_{k \geq 1} \delta_k(di)) d\vartheta P(dx) Q(dl)$. Let Q_i be an independent of the precedent ones Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}$ with intensity $ds da P(dx) Q(dl)$. We note $Z_t^N U(a_i, x_i, l_i) = \int_{\mathcal{U}} U(a_i, x_i, l_i, \alpha, y, v) Z_t^N(d\alpha, dy, dv) = \langle Z_t^N, U(a_i, x_i, l_i, \cdot, \cdot, \cdot) \rangle$.

Before writing the equation, we define

$$\begin{aligned} m_1(Z^N, i, s) &= p(s, a_i, x_i, l_i) b(s, a_i, x_i, l_i) \\ m_2(Z^N, i, s, x, l) &= (1 - p(s, a_i, x_i, l_i)) b(s, a_i, x_i, l_i) k_b(s, a_i, x_i, l_i, x, l) \\ m_3(Z^N, i, s, x, l) &= \mu(s, a_i, x_i, l_i, Z_s^N U(a_i, x_i, l_i)) \\ m_4(Z^N, i, s, x, l) &= e(s, a_i, x_i, l_i) k_e(s, a_i, x_i, l_i, x, l) \\ m_5(N, s, a, x, l) &= N e_i(s, a, x, l) \end{aligned}$$

For $t \in [0, T]$, write m_i when there is no confusion possible and consider the equation

$$\left\{ \begin{array}{l} Z_t^N = \tau_t Z_0^N \\ + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} \delta_{t-s, x_i, l_i} \mathbf{1}_{\{0 \leq \vartheta \leq m_1\}} \\ + \delta_{t-s, x, l} \mathbf{1}_{\{m_1 \leq \vartheta \leq m_2\}} Q_b(ds, di, d\vartheta, dx, dl) \\ - N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} \left[\delta_{a_i+t-s, x_i, l_i} \mathbf{1}_{\{0 \leq \vartheta \leq m_3\}} \right] Q_\mu(ds, di, d\vartheta, dx, dl) \\ + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} (\delta_{a_i+t-s, x, l} \\ - \delta_{a_i+t-s, x_i, l_i}) \mathbf{1}_{\{0 \leq \vartheta \leq m_4\}} Q_e(ds, di, d\vartheta, dx, dl) \\ + N^{-1} \int_0^t \int_{\mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \delta_{a+t-s, x, l} \mathbf{1}_{\{0 \leq \vartheta \leq m_5\}} Q_i(ds, da, dx, dl) \end{array} \right. \quad (3.2)$$

Under Assumption 3.1, we have existence and (strong) uniqueness of a solution to (3.1) in $\mathbb{D}([0, T], \mathcal{M}_+)$, the Skorokhod space of càdlàg processes with values in \mathcal{M}_+ .

Under Assumption 3.1 and 3.3, we even have the narrow convergence of Z^N in $\mathcal{D}([0, T], \mathcal{M}_+)$ to a deterministic limit $\xi \in \mathcal{C}([0, T], \mathcal{M}_+)$, see e.g. [62, 17, 26], or [8] for a proof with swap.

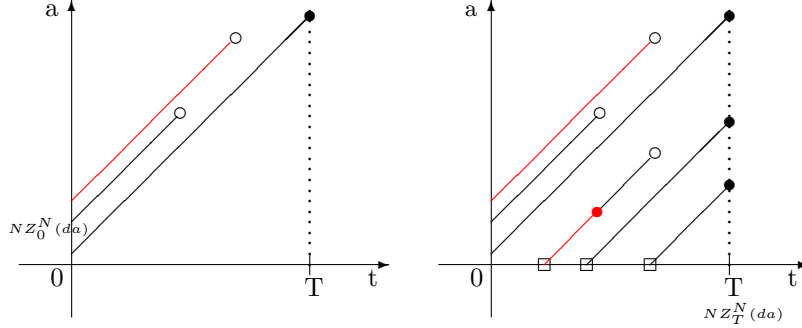


Figure 3.1: Sample path of $N Z_0^N(da)$ and its evolution without births (left), sample path of $(N Z_t^N(da))_{0 \leq t \leq T}$ (right). Black and Red lines symbolizes the characteristics.

Assumption 3.4. We have $\xi_0(da, dx, dl) = g_0(a, x, l) da P(dx) Q(dl)$ for some $g_0 \in \mathcal{F}_b$, with $|g_0|_1 < \infty$.

Under the Assumptions 3.1, 3.3 and 3.4, the limit $\xi = (\xi_t(da, dx, dl))_{0 \leq t \leq T}$ is smooth: for every $t \in [0, T]$ we have $\xi_t(da, dx, dl) = g(t, a, x, l) da P(dx) Q(dl)$, where

$$\left\{ \begin{array}{l} \partial_t g(t, a, x, l) + \partial_a g(t, a, x, l) \\ + [\mu(t, a, x, l, \int_{\mathcal{U}} U((v, y, \alpha), (l, x, a)) g(t, \alpha, y, v)) + e(t, a, x, l)] g(t, a, x, l) \\ = e_i(t, a, x, l) + \int_{\mathcal{X} \times \mathcal{L}} e(t, a, x', l') k_e(t, a, x', l', x, l) g(t, a, x', l') P(dx') Q(dl') \\ g(0, a, x, l) = g_0(a, x, l), \\ g(t, 0, x, l) = \int_0^\infty p(t, a, x, l) g(t, a, x, l) b(t, a, x, l) da \\ + \int_{\mathcal{U}} (1 - p(t, a, x', l')) g(t, a, x', l') b(t, a, x', l', x, l) P(dx') Q(dl') da \end{array} \right. \quad (3.3)$$

The limit $g(t, a, x, l)$ is the solution of a generalization of a special McKendrick Von Foester transport equation.

2 Stability of the Model

2.1 Definition

Let $w_1 : \mathbb{R} \rightarrow \mathbb{R}$, $w_2 : \mathbb{R} \rightarrow \mathbb{R}$, $w_3 : \mathbb{R} \rightarrow \mathbb{R}$ and $w_4 : \mathcal{L} \rightarrow \mathbb{R}_+$ be four given continuous weight functions. Assume that $|w_i|_\infty < \infty$, $|w_i|_1 < \infty$ and $|w_i|_2 < \infty$. This assumption is not needed for the probabilistic results but will be fulfilled in the statistical part. We do it now in order to simplify the bounds when those norm appears.

Let $\mathcal{F} \subset \mathcal{F}_b$ be such that $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$. Define $\eta_s^N(da, dx, dl) = Z_s^N(da, dx, dl) - g(s, a, x, l)daP(dx)Q(dl)$ and

$$\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathcal{U}} w_2(s-a)w_3(x)w_4(l)f_t(a, x, l)\eta_s^N(da, dx, dl)ds$$

and

$$\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathcal{U}} w_2(t-a)w_3(x)w_4(l)f_t(a, x, l)\eta_t^N(da, dx, dl).$$

Formally, we obtain $\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t$ from $\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t$ by letting $w_1 = \delta_t$.

Assumption 3.5. Define for any $f \in \mathcal{F}$

$$\phi_0(f, t)(s, a, x, l) = f(t, t+a, x, l)$$

$$\phi_1(f, t)(s, a, x, l) = p(s, a, x, l)b(s, a, x, l)f_t(t-s, x, l)$$

$$\phi_2(w_3, w_4, f, t)(s, a, x, l) =$$

$$\int_{\mathcal{X}} \int_{\mathcal{L}} (1-p(s, a, x, l))b(s, a, x, l)w_3(x')w_4(l')f_t(t-s, x', l')k_b(s, a, x, l, x', l')P(dx')Q(dl')$$

$$\phi_3(f, t)(s, a, x, l) = \mu(s, a, x, l, g_s U(a, x, l))f_t(a+t-s, x, l)$$

$$\phi_4(w_3, w_4, f, t)(s, a, x, l) =$$

$$\int_{\mathcal{X}} \int_{\mathcal{L}} e(s, a, x, l)k_e(s, a, x, l, x', l')w_3(x')w_4(l')f_t(a+t-s, x', l')P(dx')Q(dl')$$

$$\phi_5(f, t)(s, a, x, l) = e(s, a, x, l)f_t(a+t-s, x, l)$$

We have $0 \in \mathcal{F}$, all the constants $C \leq C_0 \in \mathcal{F}$ with $C_0 > 0$ such that for all $(t, a_0, x_0, l_0) \in [0, T] \times \mathcal{U}$

$$C_0^{-1}\phi_U(t, a_0, x_0, l_0)(s, a, x, l) = C_0^{-1}U(a_0 + s - t, x_0, l_0, a, x, l) \in \mathcal{F}$$

. And there exist $C(w_3, w_4) \leq C_1|w_3w_4|_1$, $C_1 \leq C_0$ such that \mathcal{F} is stable by the following operations

$$f \mapsto -f, \quad f \mapsto C_0^{-1}\phi_i(f, t), \quad \forall i \in \{0, 1, 3, 5\} \quad f \mapsto C(w_3, w_4)^{-1}\phi_i(w_3, w_4, f, t) \quad \forall i \in \{2, 4\}$$

$\forall t \in [0, T]$.

Remark 3.6. It is important to notice in the assumption that we assume $C(w_3, w_4) \leq C_0 |w_3 w_4|_1$. We will see $C(w_3, w_4)$ depends on $\int_{\mathcal{X} \times \mathcal{L}} w_3(x) w_4(l) Q(dx) P(dl)$.

Indeed suppose f and the function belongs to some Hölder space. See the definition 2.13 where here we note $\mathcal{H}^\alpha(L)$ for a function in \mathcal{H}^α , uniformly for all point, and L is the Hölder constant. Then

$$\phi_i(w_3, w_4, f, t)(s, a, x, l) \in \mathcal{H}^{a,b,c,d}(L(w_3, w_4))$$

With $L(w_3, w_4) \lesssim \int_{\mathcal{X}} w_3(x) P(dx) \int_{\mathcal{L}} w_4(l) Q(dl)$. And we know that if $\varphi \in \mathcal{H}^{a,b,c,d}(L)$ then $L^{-1}\varphi \in \mathcal{H}^{a,b,c,d}(1)$. The precedent assumption implies an implicit assumption about the measure $P(dx)$. The most simple is to assume that $P(dx)$ is absolutely continuous with respect to the Lebesgue measure.

Assumption 3.7. There exist $A \geq 0$ such that $U(\cdot, x, l, \alpha, y, b)$ is constant outside $[0, A]$.

Remark 3.8. The precedent assumption implies we can consider in the minimal entropy the support of the age is bounded above. Indeed if we come back to the proof we use the fact that $U(a+t-s, x, l, \alpha, y, b)$ is in \mathcal{F} up to a constant, but if this is a constant this is true. So the condition $C_0^{-1}U(a+t-s, x, l, \alpha, y, b) \in \mathcal{F}$ is only for $t, a, x, l \in [0, T] \times [0, A+T] \times \mathcal{X} \times \mathcal{L}$.

Let $\text{diam}_{|\cdot|_\infty}(\mathcal{F}) = \sup_{f,g \in \mathcal{F}} |f-g|_\infty$ and write $\mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)$ for the minimal number of ϵ -balls for the $|\cdot|_\infty$ -metric that are necessary to cover \mathcal{F} .

Proposition 3.9. Under the assumptions 3.5 and 3.7, the minimal space \mathcal{F}_0 satisfying those assumptions is such that

$$e(\mathcal{F}_0) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}_0, |\cdot|_\infty, \epsilon)) d\epsilon < \infty$$

Definition 3.10 (mild concentration). A sequence of nonnegative random variables $(X^N)_{N \geq 1}$ has a mild concentration property of order $0 \leq r_N \rightarrow 0$ if for large enough N , we have

$$\mathbb{P}(X^N \geq (1+u)r_N) \leq \frac{1}{e^u - 1} \text{ for every } u \geq 0.$$

Assumption 3.11. The sequence $|w_2 w_3|_{1,\infty}^{-1} |w_4|_\infty^{-1} \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0$ has the mild concentration property of order r_N for some $0 \leq r_N \rightarrow 0$ and there exist $q > 0$ and $p \geq 2$ such that

$$\mathbb{E}[\exp(\lambda \mathcal{W}_{1,1,1}^N(\mathcal{F})_0)] \leq e^{\max(\lambda^p, \lambda)q}.$$

Remark 3.12. Several remarks are in order :

1) Assumption 3.11 implies the moment estimate

$$\mathbb{E}[\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0^p] \leq 2^{p+3} p! r_N^p |w_2 w_3|_{1,\infty}^p |w_4|_\infty^p \text{ for every } p \geq 2, \quad (3.4)$$

since, noting $X^N = |w_2 w_3|_{1,\infty}^{-1} |w_4|_\infty^{-1} \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0$,

$$\begin{aligned} \mathbb{E}[\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0^p] &= p \int_0^\infty \kappa^{p-1} \mathbb{P}(\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0 \geq \kappa) d\kappa \\ &= |w_2 w_3|_{1,\infty}^p |w_4|_\infty^p r_N^p \int_{-1}^\infty (1+u)^{p-1} \mathbb{P}(X^N \geq (1+u)r_N) du \\ &\leq |w_2 w_3|_{1,\infty}^p |w_4|_\infty^p r_N^p \int_{-1}^\infty (1+u)^{p-1} \min\{(e^u - 1)^{-1}, 1\} du \\ &\lesssim 2^{p+3} p! |w_2 w_3|_{1,\infty}^p |w_4|_\infty^p r_N^p. \end{aligned} \quad (3.5)$$

In particular we have

$$\mathbb{E}[\mathcal{W}_{1,1,1}^N(\mathcal{F})_0^p] \leq 2^{p+3} p! r_N^p T^{p/2} = \frac{p!}{2} B_T^N (C_T^N)^{p-2}$$

with $B_T^N = 64Tr_N^2$ and $C_T^N = 2\sqrt{T}r_N$. And from the Berstein inequality proof we get

$$\mathbb{E}[\exp(\lambda W_{1,1,1}^N(\mathcal{F})_0)] \leq \exp(\lambda \mathbb{E}[W_{1,1,1}^N(\mathcal{F})_0]) + \exp\left(\frac{B_T^N \lambda^2}{2(1 - C_T^N \lambda)}\right) \quad \forall \lambda \in [0, 1/C_T^N]$$

which is close to the second part of the assumption but not sufficient in the proof.

- 2) This assumption is true if $r_N^{-1}|w_2 w_3|_{1,\infty}^{-1}|w_4|_{\infty}^{-1} \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0$ has a gaussian tail bound.
- 3) This assumption is true if Z_0^N is obtained thanks to a N -sample of a bounded random variable.

Theorem 3.13. Work under Assumptions 3.1, 3.3, 3.4, 3.5 and 3.11. Assume moreover

$$\text{diam}_{|\cdot|_{\infty}}(\mathcal{F}) \leq 1$$

and

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_{\infty}, \epsilon)) d\epsilon < \infty.$$

Note φ a function such that $\varphi > 0$, $\forall x, y > 0$ $\varphi(x+y) \lesssim \varphi(x) + \phi_{\varphi}(y)$ where ϕ_{φ} is a known function for φ , increasing and positive, such that for all $C > 0$ $\phi_{\varphi}(Cx) \lesssim \phi_{\varphi}(x)$.

If $|w_2 w_3|_{\infty} \lesssim N^{1/2}|w_2 w_3|_1$, there exist an event \mathcal{B}_N with $\mathbb{P}(\mathcal{B}_N^c) \lesssim e^{-\sqrt{N}}$ such that

- (i) there exist a random variable $\widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$ such that

$$|w_1 w_2 w_3|_{1,\infty}^{-1} |w_4|_{\infty}^{-1} \widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$$

has a mild concentration property of order $r_N \wedge N^{-1/2}$. We have the bound

$$\begin{aligned} & \mathbb{E}[\varphi(\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T)] \lesssim \\ & \mathbb{E}[\varphi(\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T) | \mathcal{B}_N] + \mathbb{E}[\varphi(\widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T)] + \mathbb{P}(\mathcal{B}_N^c) \phi_{\varphi}(|w_1 w_2 w_3 w_4|_1) \end{aligned}$$

and under $\mathbb{P}(\cdot | \mathcal{B}_N)$, $|w_1 w_2 w_3|_{1,\infty}^{-1} |w_4|_{\infty}^{-1} \mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$ has a mild concentration property of order $r_N \wedge N^{-1/2}$.

- (ii) there exist a random variable $\widetilde{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t$ such that for any t we have a mild concentration property of order $r_N \wedge N^{-1/2}$ for $|w_2 w_3|_{1,\infty}^{-1} |w_4|_{\infty}^{-1} \widetilde{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t$. We get

$$\begin{aligned} & \mathbb{E}[\varphi(\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_T)] \lesssim \\ & \mathbb{E}[\varphi(\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_T) | \mathcal{B}_N] + \mathbb{E}[\varphi(\widetilde{W}_{w_2, w_3, w_4}^N(\mathcal{F})_T)] + \mathbb{P}(\mathcal{B}_N^c) \phi_{\varphi}(|w_2 w_3 w_4|_1) \end{aligned}$$

and under $\mathbb{P}(\cdot | \mathcal{B}_N)$, $(|w_2 w_3|_{1,\infty}^{-1} |w_4|_{\infty}^{-1}) \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t$ has a mild concentration property of order $r_N \wedge N^{-1/2}$ for any t .

Remark 3.14. Two families of functions are important in this Theorem. **1)** The functions $\varphi_p : x \rightarrow x^p$ which obviously satisfy the assumptions with $\phi_{\varphi_p} = \varphi_p$. **2)** The functions of the form $\varphi_a : x \rightarrow \{x^2 - a\}_+$ which satisfy the assumptions with $\phi_{\varphi_a}(x) = a + 2x^2$, since $\{(x+y)^2 - a\}_+ \leq \{2x^2 - a\}_+ + 2y^2 \leq 2\{x^2 - a\}_+ + a + 2y^2$.

Now defining $\Gamma^N(ds, da, dx, dl)$ the point process of death in the same spirit of the precedent chapter, see the subsection 3.2 of the precedent chapter. We can define $\mathcal{D}_T^N(w_1, w_2, w_3, w_4)$ by

$$\int_0^T \int_{\mathcal{U}} w_1(s)w_2(s-a)w_3(x)w_4(l)\mathfrak{D}_s^N(da, dx, dl)$$

where $\mathfrak{D}_s^N(da, dx, dl) = \Gamma^N(ds, da, dx, dl) - \mu(s, a, x, l, g_s U(a, x, l))g(s, a, x, l)dsdaP(dx)Q(dl)$. Define $v_N = |w_1 w_2 w_3|_{1, \infty} |w_4|_{\infty} N^{-1/2}$,

Theorem 3.15. *Under the assumptions 3.1, 3.3, 3.4, 3.5, 3.7 and 3.11, we have the existence of a constant $C(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C$ such that*

$$\begin{aligned} \mathbb{P}(\mathcal{D}_T^N(w_1, w_2, w_3, w_4) \geq C v_N (1+u)) &\lesssim (e^u - 1)^{-1} \wedge 1 + (e^{\sqrt{1+u}-1} - 1)^{-1} \wedge 1 \\ &+ (e^{\sqrt{1+u}} - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\sqrt{1+u}}-1} - 1)^{-1} \wedge 1 \\ &+ 2 \exp\left(-\frac{(1+u)^2}{8C_2}\right) \mathbf{1}_{u \leq \frac{C_2}{C_1 v_N} - 1} \\ &+ 2 \exp\left(-\frac{\sqrt{u+1}}{v_N C_1}\right) \mathbf{1}_{u \geq \frac{C_2}{C_1 v_N} - 1} \end{aligned}$$

assuming $N^{1/2} v_N \rightarrow \infty$, $v_N \rightarrow 0$ and the existence of a constant \tilde{C} independent of N such that $|w_1 w_2 w_3|_{1, \infty}^2 = \tilde{C} |w_1 w_2 w_3|_{\infty}$. Those assumptions are fulfilled for our statistical purpose.

3 Nonparametric estimation of g and μ

We know from the precedent chapter we have to do estimation in another system of coordinates, which is $s, s-a, x, l$.

Since the last coordinate is discrete, we will just use for any l_0 in \mathcal{L} , the indicator function $\mathbf{1}_{l=l_0}$ as "kernel". All the precedent Theorem can be applied with such a weight function w_4 .

We will use multivariate kernels of the form $\mathcal{K}^{l_0} = K^1 \times K^2 \times K^3 \times \mathbf{1}_{l=l_0}$. \mathcal{K}^{l_0} is defined by

$$\mathcal{K}^{l_0}(t, a, x, l) = K^1(t)K^2(t-a)K^3(x)\mathbf{1}_{l=l_0} \text{ for } (t, a, x, l) \in [0, T] \times \mathcal{U}$$

For a multivariate bandwidth $\mathbf{h} = (h_1, h_2, h_3)$ with $h_i > 0$, we set

$$\mathcal{K}_h^{l_0}(t, a, x) = (h_1 h_2 h_3)^{-1} \mathcal{K}^{l_0}(h_1^{-1}t, h_2^{-1}a, h_3^{-1}x, l) = K_{h_1}^1(t)K_{h_2}^2(t-a)K_{h_3}^3(x)\mathbf{1}_{l=l_0}$$

For a function $f \in \mathcal{F}_b$ we define the linear approximation

$$\mathcal{K}_h^{l_0} \star f(t, a, x, l) = \int_0^T \int_{\mathbb{R}_+ \times \mathcal{X}} f(s, a, x, l_0) \mathcal{K}_h(s, a, x) ds da P(dx) Q(l_0)$$

From now we will write \mathcal{K}^{l_0} as \mathcal{K} taking an implicit l_0 .

3.1 Construction of estimators of g and μ

Construction of an estimator of g

We consider the family of estimators

$$\hat{g}_h^N(t, a, x, l) = K_h \star Z_t^N(a, x, l) = \int_{\mathcal{U}} K_{h_1}^1(u - a) K_{h_2}^2(y - x) \mathbf{1}_{v=l} Z_t^N(du, dy, dv)$$

, where K^1 and K^2 are two kernels of order $L \geq 0$.

Construction of an estimator of μ

We can define the process of death occurrences as before. For $(t, a, x, l) \in [0, T] \times \mathcal{U}$ consider the family of estimators

$$\hat{\gamma}_h^N(t, a, x, l_0) = \int_0^T \int_{\mathcal{U}} \mathcal{K}_h(s - t, u - a, y - x, l) \Gamma^N(ds, du, dy, dl), \quad \mathbf{h} = (h_1, h_2, h_3) \text{ with } h_i > 0. \quad (3.6)$$

An estimator of $\mu(t, a)$ is obtained by considering the ratio

$$\hat{\mu}_{h, \mathbf{h}}^N(t, a) = \begin{cases} \hat{g}_h^N(t, a)^{-1} \hat{\gamma}_h^N(t, a) & \text{if } \hat{g}_h^N(t, a) \neq 0 \\ 0 & \text{otherwise,} \end{cases} \quad (3.7)$$

and is specified by the bandwidths $\mathbf{h} = (h_1, h_2, h_3)$ with $h_i > 0$ and $\tilde{h} = (\tilde{h}_1, \tilde{h}_2)$ with $\tilde{h}_i > 0$.

3.2 Oracle inequalities

We define the quantity in the same spirit of the precedent chapter, see lemma 2.28. In our setting, the \mathcal{G}_N^1 will be of cardinal N^2 , since we have two dimensions, and we get bigger lattice \mathcal{G}_N^2 too. The estimator are the one obtained by the Goldenschluger Lepski Method.

Let

$$\mathcal{B}_h^N(g)(t, a) = \sup_{h'_1 \leq h_1, h'_2 \leq h_2, h' \in \mathcal{G}_1^N} \left| \int_0^\infty K_{h'_1}(u - a) K_{h'_2}(y - x) g(t, u, x, l_0) d\mu P(dy) - g(t, a, x, l_0) \right|. \quad (3.8)$$

and

$$V_h^N \approx (\ln(N) r_N \wedge N^{-1/2} |K_{h_1}^2 K_{h_2}^3|_{1, \infty})^2 \quad (3.9)$$

up to a constant depending only on $\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1$.

Theorem 3.16. *Work under Assumptions 3.1, 3.3, 3.4, 3.5, 3.11 and 3.7. For $(t, a, x, l) \in \mathcal{D}_U \cup \mathcal{D}_L$, specify $g_\star^N(t, a, x, l)$ with bounded and compactly supported kernels K^2 and K^3 . The following oracle inequality holds true*

$$\mathbb{E}[(\hat{g}_\star^N(t, a, x, l) - g(t, a, x, l))^2] \lesssim \inf_{h \in \mathcal{G}_1^N} (\mathcal{B}_h^N(g)(t, a, x, l)^2 + V_h^N) + \delta_N$$

for large enough N , with $\delta_N = N^{-1}$ and up to a constant that depends on $\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1$ and kernels.

Remark 3.17. *The proof of this Theorem is a direct consequence from the proof of the precedent chapter and the Theorem 3.13. We just have to take $\varphi(x) = \{x^2 - V_h^N\}_+$ as pointed out in the remark following the Theorem 3.13.*

Here we define

$$V_h^N = ((\ln N)^4 C^* \max(N^{-1/2}, r_N) |\mathcal{K}_{h_1}^1|_{1,\infty} |\mathcal{K}_{h_2}^2|_{1,\infty} |\mathcal{K}_{h_3}^3|_{1,\infty})^2$$

, for C^* a constant big enough, as we will see in the proof. We just note the power of the logarithm which is bigger than expected, we lost indeed some precision because of the interaction part.

Assumption 3.18. *For every $(t, a, x, l) \in \mathcal{D}_-$, the interior of the domain, there exists an open set $\mathcal{U}_{(t,a,x,l)}$ such that*

$$\inf_{(u \in \mathcal{U}_{(t,a,x,l)})} b(t-a, t-a+u, x, l) g_0(u) \geq \delta \text{ if } (t, a, x, l) \in \mathcal{D}_L \quad (3.10)$$

and

$$g_0(t-a, x, l) \geq \delta \text{ if } (t, a, x, l) \in \mathcal{D}_U, \quad (3.11)$$

for some $\delta > 0$.

We define $\widehat{\mu}_*^N(t, a, x, l)_\varpi$ in the same spirit than in the precedent chapter, see (2.25).

Theorem 3.19. *Work under Assumptions 3.1, 3.3, 3.4, 3.5, 3.11, 3.18 and 3.7. For $\mathcal{D}_U \cup \mathcal{D}_L$ specify $\widehat{\mu}_*^N(t, a, x, l)_\varpi$ with bounded and compactly supported kernels $\mathcal{K}^1, \mathcal{K}^2$ and \mathcal{K}^3 . The following oracle inequality holds true*

$$\begin{aligned} \mathbb{E}[(\widehat{\mu}_*^N(t, a, x, l)_\varpi - \mu(t, a, x, l))^2] &\lesssim \\ \inf_{h \in \mathcal{G}_1^N} (\mathcal{B}_h^N(g)(t, a, x, l)^2 + \mathbf{V}_h^N) &+ \inf_{h \in \mathcal{G}_2^N} (\mathcal{B}_h^N(\mu g)(t, a, x, l)^2 + \mathbf{V}_h^N) + \delta_N \end{aligned}$$

for large enough N and small enough $\varpi > 0$, with $\delta_N = N^{-1}$ and up to a constant that depends on $C^*, \mathcal{R}, C_0, T, |g_0|_1, |e_i|_1$ and the kernels K^i and \mathcal{K}^i .

4 Minimax estimation under anisotropic Hölder smoothness

4.1 The smoothness of the McKendrick-Von Forster equation

We will now only work without characteristic. Indeed our point is to focus on interaction. We use the former definition of Hölder space, see the definition 2.13.

Assumption 3.20. *We suppose that $b \in \mathcal{H}^{\alpha,\beta}$, $\mu \in \mathcal{H}^{\gamma,\delta,v}$, $g_0 \in \mathcal{H}^\nu$ and $U \in \mathcal{H}^{\rho,\eta}$, where $\nu > \alpha + 1, \beta + 1, \gamma + 1, \delta + 1, \rho + 1$.*

Proposition 3.21. *Under the assumptions 3.1, 3.3, 3.4 and 3.20 we have*

$$g \in \mathcal{H}^{\min(\alpha,\beta,\gamma+1,v,\delta,\rho), \min(\alpha,\beta,\gamma+1,v,\delta,\rho)} \text{ for } (t, a) \in \mathcal{D}_L$$

and

$$g \in \mathcal{H}^{\min(\alpha,\beta,\gamma+1,v,\delta,\rho), \min(\delta,\rho,v)} \text{ for } (t, a) \in \mathcal{D}_U$$

We also have

$$\tilde{g} \in \mathcal{H}^{\min(\alpha, \beta, \gamma, \delta, \nu, \rho)+1, \min(\alpha, \beta, \gamma+1, \nu, \delta, \rho)} \text{ for } (t, a) \in \widetilde{\mathcal{D}}_L$$

and

$$\tilde{g} \in \mathcal{H}^{\min(\alpha, \beta, \gamma, \nu, \delta, \rho)+1, \min(\delta, \rho, \nu)} \text{ for } (t, a) \in \widetilde{\mathcal{D}}_U$$

Proposition 3.22. *Under the assumptions 3.1, 3.3, 3.4 and 3.20 we have , noting $\bar{\mu}(t, a) = \mu(t, a, g_t U(a))$,*

$$\bar{\mu} \in \mathcal{H}^{\min(\alpha, \beta, \gamma, \nu, \delta, \rho), \min(\delta, \rho, \nu)} \text{ for } (t, a) \in \mathcal{D}_L$$

and

$$\bar{\mu} \in \mathcal{H}^{\min(\alpha, \beta, \gamma, \delta, \rho, \nu), \min(\delta, \rho, \nu)} \text{ for } (t, a) \in \mathcal{D}_U$$

Remark 3.23. *We can compare the proposition 3.21 to the proposition in the precedent chapter, see the proposition 2.16. It is important to note the change of variable let the dependence in α, β with respect to the time in the upper domain and in the lower domain after the change of variable. This is simply due to the fact $\bar{\mu}$ depends on g over the two domains and so we cannot avoid this loss of regularity.*

4.2 Minimax lower bounds

For $\alpha, \beta > 0$ and $L > 0$, we set

$$\mathcal{H}_L^{\alpha, \beta}(t, a) = \{f \in \mathcal{L}_{\mathcal{D}}^{\infty}, |f|_{\infty} + |f|_{\mathcal{H}^{\alpha, \beta}(t, a)} \leq L\},$$

where the semi-norm $|\cdot|_{\mathcal{H}^{\alpha, \beta}(t, a)}$ is defined after Definition 2.14. We also set, for $\epsilon > 0$,

$$\mathcal{L}_{\mathcal{D}, \epsilon}^{\infty} = \{f \in \mathcal{L}_{\mathcal{D}}^{\infty}, \inf_{(t, a) \in \mathcal{D}} f(t, a) \geq \epsilon\}$$

Remember that under Assumption 3.1, any point (b, μ, U, g_0) defines a unique solution g to the McKendrick Von Foester equation (3.3). We add the following assumption in order to get the lower bound.

Assumption 3.24. *There exists a function $\mathbf{m} \in \mathcal{H}^{\gamma, \delta}$ such that*

$$\mu(s, a, u) = \mathbf{m}(s, a) + u$$

With this assumption, we see that $\nu = \infty$ disappear from all the regularity. Let

$$s_{\text{dens}}^- = \min(\max(\gamma, \delta), \rho) \text{ and } s(\gamma, \delta, \rho) = \min(\rho, (\gamma^{-1} + \delta^{-1})^{-1}).$$

Under a non-degeneracy condition of the form $\bar{\mu} \in \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty}$, which is a condition only on \mathbf{m} if \mathfrak{k} is little enough, we obtain the following minimax lower bound:

Theorem 3.25. *Work under Assumptions 3.1, 3.3, 3.4, 3.20 and 3.24. Let $\alpha, \beta, \gamma, \delta, \rho > 0, \nu \geq \max(\alpha, \beta, \gamma, \delta) + 1$ and $L > 0$. For every $(t, a) \in \mathcal{D}^-$, we have*

$$\liminf_{N \rightarrow \infty} \inf_F \sup_{b, U, \mu, g_0} \mathbb{P}(N^{s_{\text{dens}}^-(t, a)/(2s_{\text{dens}}^-(t, a)+1)} |F - g(t, a)| \geq C) > 0 \quad (3.12)$$

and

$$\inf_F \sup_{b, U, \mu, g_0} \mathbb{E}[|F - \mu(t, a)|] \gtrsim N^{-s(\gamma, \delta, \rho)/(2s(\gamma, \delta, \rho)+1)}, \quad (3.13)$$

where the infimum is taken over all estimators and the supremum over

$$b \in \mathcal{H}_L^{\alpha, \beta}(t, a), \quad \mathbf{m} \in \mathcal{H}_L^{\gamma, \delta}(t, a) \cap \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty}, \quad U \in \mathcal{H}^{\rho, \eta} \text{ and } g_0 \in \mathcal{H}_L^{\nu}(t, a).$$

4.3 Adaptive estimation under anisotropic Hölder smoothness

Our next result give the upper bound for our estimator $g_\star^N(t, a)$ defined in the spirit of (2.22). Moreover, $g_\star^N(t, a)$ is nearly smoothness adaptive. More precisely, let

$$s_{\text{dens}}^+(t, a) = \min(\delta, \rho) \mathbf{1}_{\mathcal{D}_U}(t, a) + \min(\alpha, \beta, \gamma + 1, \delta, \rho) \mathbf{1}_{\mathcal{D}_L}(t, a), \quad (3.14)$$

and note that $s_{\text{dens}}^+(t, a) \leq s_{\text{dens}}^-(t, a)$ always.

Theorem 3.26. *Work under Assumptions 3.1, 3.3, 3.4, 3.20, 3.5, 3.7 and 3.24 with $r_N \leq N^{-1/2}$. Specify $g_\star^N(t, a)$ with a compactly supported kernel of order $\ell_0 \geq 0$ and pick*

$$\mathcal{G}_1^N = (x_1^N < x_2^N < \dots < x_N^N)$$

a subdivision of $[N^{-1/2}, (\log N)^{-1}]$ with $\max_{1 \leq i \leq N-1} (x_{i+1}^N - x_i^N) \lesssim N^{-1}$ so that $\text{Card } \mathcal{G}_1^N \lesssim N$. For every $(t, a) \in \mathcal{D}^-$ and large enough N , we have

$$\sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{g}_\star^N(t, a) - g(t, a))^2])^{1/2} \lesssim \left(\frac{(\log N)^2}{N} \right)^{s_{\text{dens}}^+(t, a) \wedge \ell_0 / (2s_{\text{dens}}^+(t, a) \wedge \ell_0 + 1)}, \quad (3.15)$$

where the supremum is taken over $b \in \mathcal{H}_L^{\alpha, \beta}(t, a)$, $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$, $g_0 \in \mathcal{H}_L^\nu(t, a)$ with $\alpha, \beta, \gamma, \delta > 0, \nu \geq \min(\gamma, \delta) + 1$ and $L > 0$.

Similarly, $\mu_\star^N(t, a)$ defined in the spirit of (2.25) also

shares near optimality on \mathcal{D}_U if $\delta \geq \gamma + 1$ and on \mathcal{D}_L provided b is regular enough. Define

$$s_L(\alpha, \beta, \gamma, \delta) = (\min(\alpha, \beta, \gamma + 1, \delta, \rho)^{-1} + \min(\alpha, \beta, \gamma, \delta, \rho)^{-1})^{-1},$$

$$s_U(\gamma, \delta) = (\min(\delta, \rho)^{-1} + \min(\alpha, \beta, \gamma, \delta, \rho)^{-1})^{-1},$$

let

$$s_{\text{death}}^+(t, a) = s_U(\gamma, \delta) \mathbf{1}_{\mathcal{D}_U}(t, a) + s_L(\alpha, \beta, \gamma, \delta) \mathbf{1}_{\mathcal{D}_L}(t, a). \quad (3.16)$$

and note that $s_{\text{death}}^+(t, a) \leq s_{\text{death}}^-$ always.

Theorem 3.27. *Work under Assumptions 3.1, 3.3, 3.4, 3.20, 3.5, 3.7, 3.24 and 3.18 with $r_N \leq N^{-1/2}$. Specify $\mu_\star^N(t, a)$ with kernels H, K of order $\ell_0 \geq 0$ and pick $\mathcal{G}_2^N = \mathcal{G}_1^N \times \mathcal{G}_1^N$ so that $\text{Card } \mathcal{G}_2^N \lesssim N^2$. For every $(t, a) \in \mathcal{D}^-$ and large enough N , we have*

$$\sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{\mu}_\star^N(t, a) - \mu(t, a))^2])^{1/2} \lesssim \left(\frac{(\log N)^4}{N} \right)^{s_{\text{death}}^+(t, a) \wedge \ell_0 / (2s_{\text{death}}^+(t, a) \wedge \ell_0 + 1)},$$

where the supremum is taken over $b \in \mathcal{H}_L^{\alpha, \beta}(t, a)$, $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$, $g_0 \in \mathcal{H}_L^\nu(t, a)$, with $\alpha, \beta, \gamma, \delta > 0, \nu \geq \min(\gamma, \delta) + 1$ and $L > 0$.

Remark 3.28. *Those Theorems are a direct consequence of the ones obtained in the subsection 3.2 and the proposition 3.21, see the subsection 7.4 and 7.5 of the precedent chapter for a detailed calculus.*

5 Numerical illustration

The simulation is realised with the parameters:

- (i) g_0 is the density of a gaussian random variable centered in 40 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t, a) = \mathbf{1}_{20 \leq a \leq 40}$. Such a birth rate is not Hölder, however we can prove similar result with such piecewise constant function.
- (iii) $\mu(t, a) = 0,004 \exp(0,0074a) \exp(-0,005t) + 0,1 * \int_0^\infty g(t, \alpha) d\alpha$. We take a death rate very high in order to have enough deaths for the estimation of the death rate.

We consider the domain $[0, 20] \times [0, 120]$, which means $T = 20$ and $A_{max} = 120$.

We estimate g on the grid $\mathcal{T}^g = \{k1,005, 0 \leq k < 20\}$ and $\mathcal{A}^g = \{k0,2002, 0 \leq k < 600\}$. We estimate μ and μg on the grid $\mathcal{T}^\mu = \mathcal{T}^g$, and $\mathcal{A}^\mu = \{k1,0008, 0 \leq k < 120\}$.

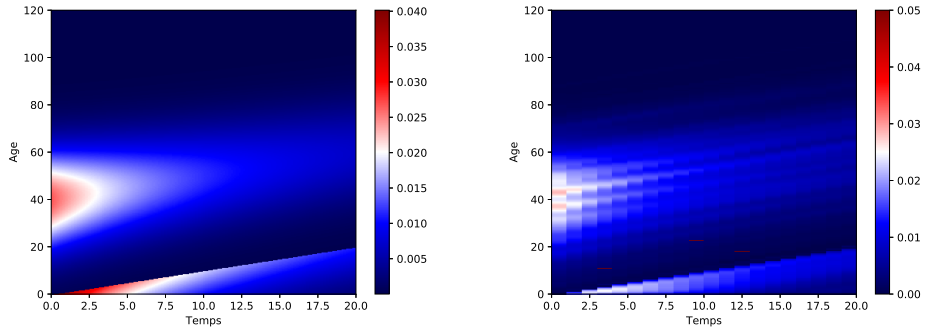


Figure 3.2: Left: true population density. Right: Estimation with GoldenSchluger Lespki Method, for $N = 4000$.

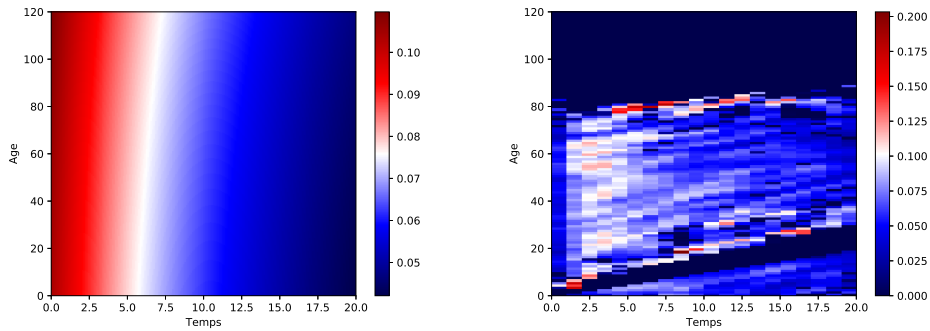


Figure 3.3: Left: true μ density. Right: Estimation with GoldenSchluger Lespki Method, for $N = 4000$, and a threshold of 0.001

5. NUMERICAL ILLUSTRATION

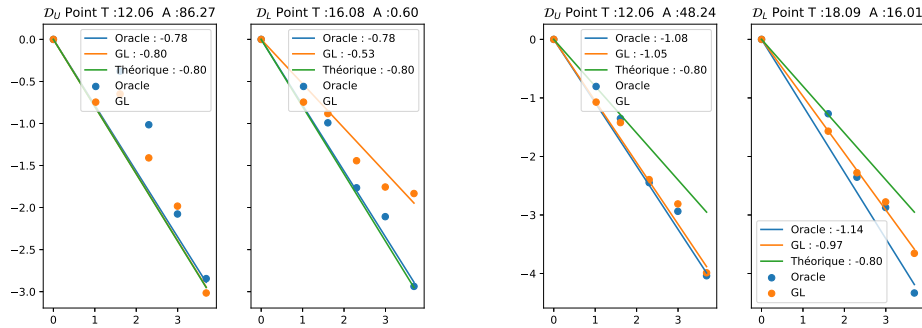


Figure 3.4: log speed of the estimation of the population density g realized on 50 simulations. Points are written in the graphs.

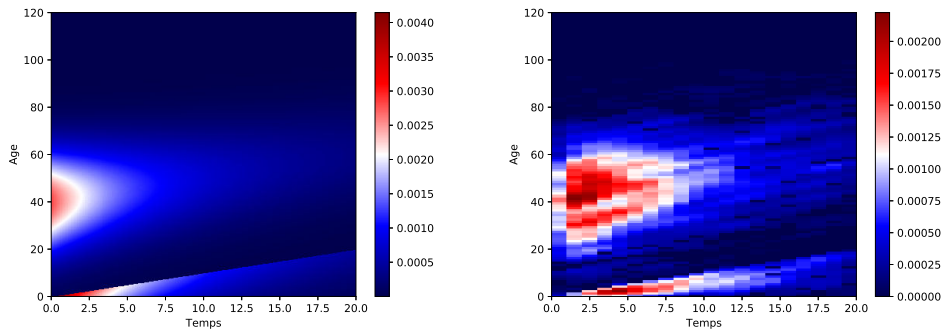


Figure 3.5: Left: true μg density. Right: Estimation with GoldenSchluger Lespki Method, for $N = 4000$.

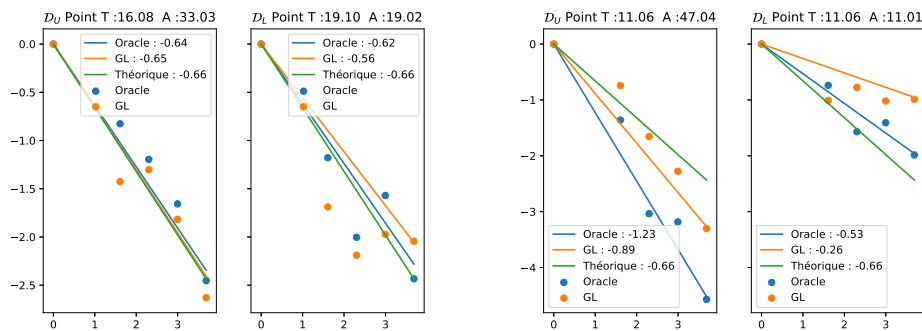


Figure 3.6: log speed of the estimation of μg realized on 50 simulations. Points are written in the graphs.

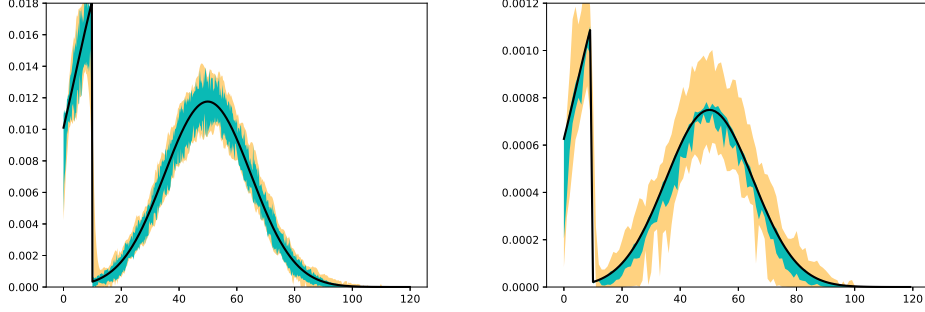


Figure 3.7: For $T = 10$, $N = 4000$, comparison between the true function (black) and the 95% confidence interval on 50 simulations. Oracle in blue, adaptive estimator in yellow. Left: Estimation of g , Right : Estimation of μg .

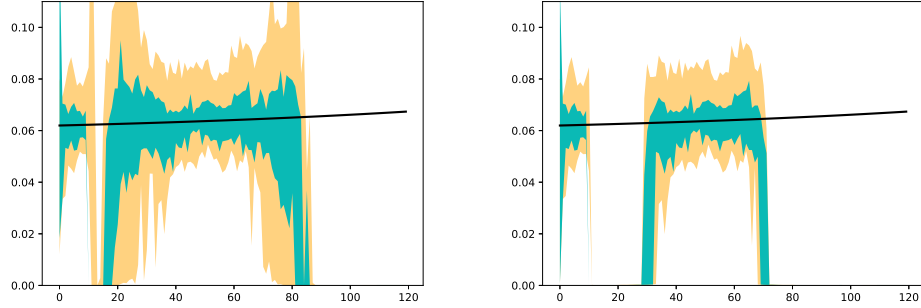


Figure 3.8: For $T = 10$, $N = 4000$, estimation of μ . The true value is in black, in blue the 95% interval confidence for the oracle estimate and in yellow the 95% interval confidence for the adaptative estimate, realized on 50 simulations. Left: threshold of 0.001. Right: threshold of 0.005

From the figure 3.3 the estimation seems very far from μ . This is true, however we can see from the figure 3.8 the threshold is very important.

6 Proof of Theorem 3.13

This section is devoted to the proof of the concentration properties of the model stated in Theorem 3.13. Recall that

$$\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathcal{U}} w_2(s-a) w_3(x) w_4(l) f_t(a, x, l) \eta_s^N(da, dx, dl) ds$$

and

$$\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathcal{U}} w_2(t-a) w_3(x) w_4(l) f_t(a, x, l) \eta_s^N(da, dx, dl).$$

Set also

$$\mathcal{M}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| \int_0^t w_1(s) M_s^N(w_2(s - \cdot) w_3 w_4 f_s) ds \right|$$

and

$$\mathcal{M}_{w_2, w_3, w_4}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} |M_t^N(w_2(t - \cdot) w_3 w_4 f_t)|$$

We will use the process \tilde{Z}_t^N created by coupling. In this coupling, people cannot die and when they swap they create a clone. We take \mathcal{R} for the swap and birth rate of this new population. It is not difficult to see that for any positive function we then have $\langle Z_t^N, f \rangle \leq \langle \tilde{Z}_t^N, f \rangle$. Let define for $\tilde{Z}_t^N, \tilde{\mathcal{W}}^N$ in the same way. We also define \mathbf{g} (to not have a confusion with \tilde{g} which is g with a change of variable) as the limit of \tilde{Z}^N . This limit exists and satisfy the equation

$$\begin{aligned} \langle \mathbf{g}, f_t \rangle &= \int_{\mathcal{U}} f_t(t + a, x, l) g_0(da, dx, dl) da P(dx) Q(dl) \\ &+ \int_0^t \int_{\mathcal{U}} (p(s, a, x, l) \mathcal{R} f_t(t - s, x, l) \mathbf{g}(s, a, x, l) da P(dx) Q(dl) ds \\ &+ \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} \mathcal{R} \tau_1(s, a, x, l, x', l') f_t(t - s, x', l') P(dx') Q(dl') \mathbf{g}(s, a, x, l) da P(dx) Q(dl) ds \\ &+ \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} \mathcal{R} k_e(s, a, x, l, x', l') f_t(a + t - s, x', l') P(dx') Q(dl') \mathbf{g}(s, a, x, l) da P(dx) Q(dl) ds \\ &+ \int_0^t \int_{\mathcal{U}} e_i(s, a, x, l) f(s, a, x, l) da P(dx) Q(dl) ds \end{aligned}$$

where

$$\tau_1(s, a, x, l, x', l') = (1 - p(s, a, x, l)) k_b(s, a, x, l, x', l').$$

We remark that $\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0$ is the same since the initial condition is the same.

6.1 A first stability result

Proposition 3.29. *We have*

$$\widetilde{\mathcal{W}}_{1,1,1}^N(\mathcal{F})_t \leq C_0 \mathcal{W}_{1,1,1}^N(\mathcal{F})_0 + 4C_0 \widetilde{\mathcal{W}}_{1,1,1}^N(\mathcal{F})_t + \widetilde{\mathcal{M}}_{1,1,1}^N(\mathcal{F})_t$$

for any $t \in [0, T]$.

The proof of this proposition will be easily deduced from the proof of the next proposition.

Remark 3.30. *As a corollary we see with $\tilde{n}_t = N \langle \tilde{Z}_t^N, 1 \rangle$ that*

$$N^{-1} \tilde{n}_t \leq N^{-1} \tilde{n}_T \leq |\mathbf{g}_T|_1 + \widetilde{\mathcal{W}}_{1,1,1}^N(\mathcal{F})_T \leq (|g_0|_1 + |e_i|_1) e^{2\mathcal{R}T} + \tilde{P}_T^N \quad (3.17)$$

where $\tilde{P}_T^N = C_0 \widetilde{\mathcal{W}}_{1,1,1}^N(\mathcal{F})_0 + 4C_0 \widetilde{\mathcal{W}}_{1,1,1}^N(\mathcal{F})_T + \widetilde{\mathcal{M}}_{1,1,1}^N(\mathcal{F})_T$

Proposition 3.31. *Work under Assumptions 3.1, 3.3, 3.4 and 3.5. Then $\mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$ is bounded above by*

$$C_0 |w_1|_1 \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0 + C_0 \left(\int_0^T |w_1(t)| \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_t dt + \int_0^T |w_1(t)| \mathcal{W}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t dt \right) +$$

$$\begin{aligned}
& C_0 |w_3 w_4|_1 \left(\int_0^T |w_1(t)| \mathcal{W}_{w_2,1,1,1}^N(\mathcal{F})_t dt + \int_0^T |w_1(t)| \mathcal{W}_{1,w_2,1,1}^N(\mathcal{F})_t dt \right) \\
& + C_0 \mathcal{R} \int_0^T \int_{\mathcal{U}} |w_1(t) w_2(t-a) w_3(x) w_4(l)| \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t \tilde{Z}_t^N(dx, da, dl) dt + \\
& \mathcal{M}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t
\end{aligned}$$

We also have in the case where $w_1 = w_2 = w_3 = w_4 = 1$,

$$\begin{aligned}
& TC_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + 4C_0 \int_0^T \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t dt \\
& + C_0 \mathcal{R} \left[\widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_0 + \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_T + \widetilde{\mathcal{M}}_{1,1,1,1}^N(\mathcal{F})_T \right] \int_0^T \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t dt + \\
& \mathcal{M}_{1,1,1,1}^N(\mathcal{F})_t
\end{aligned}$$

as soon as $\mathcal{F} \subset \{|f|_\infty \leq 1\}$.

Proof. By (3.2), the action $\langle Z_t^N, f_t \rangle$ of $Z_t^N(da, dx, dl)$ for $f \in \mathcal{F}_b$ can be written as

$$\begin{aligned}
\langle Z_t^N, f_t \rangle &= \int_{\mathcal{U}} f_t(t+a, x, l) Z_0^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} (p(s, a, x, l) b(s, a, x, l) f_t(t-s, x, l) Z_s^N(da, dx, l) \\
&+ \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} b(s, a, x, l) \tau_1(s, a, x, l, x', l') f_t(t-s, x', l') P(dx') Q(dl') Z_s^N(da, dx, dl) \\
&- \int_0^t \int_{\mathcal{U}} [\mu(s, a, x, l, Z_s^N U(a, x, l)) + e(s, a, x)] f_t(a+t-s, x, l) Z_s^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} e(s, a, x) k_e(s, a, x, l, x', l') f_t(a+t-s, x', l') P(dx') Q(dl') Z_s^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} e_i(s, a, x, l) f(s, a, x, l) da P(dx) Q(dl) ds \\
&+ [M_t^{N, birth}(f) + M_t^{N, death}(f) + M_t^{N, imm}(f) + M_t^{N, swap}(f)]
\end{aligned} \tag{3.18}$$

with

$$M_t^{N, death}(f_t) = N^{-1} \int_0^t \int_{\{i \leq n_s^N\} \times \mathcal{U}} \mathbf{1}_{\{m_3 \leq \theta\}} (-f_t(a_i + t-s, x_i, l_i)) \tilde{Q}_\mu(ds, di, d\theta, dx, dl). \tag{3.19}$$

We can easily find the other one by the same construction.

Let's remind we have for any $f \in \mathcal{F}$, as defining in the assumption 3.5,

$$\begin{aligned}
\phi_0(f, t)(s, a, x, l) &= f(t, t + a, x, l) \\
\phi_1(f, t)(s, a, x, l) &= p(s, a, x, l)b(s, a, x, l)f_t(t - s, x, l) \\
\phi_2(w_3, w_4, f, t)(s, a, x, l) \\
&= \int_{\mathcal{X}} \int_{\mathcal{L}} (1 - p(s, a, x, l))b(s, a, x, l)w_3(x')w_4(l')f_t(t_s, x', l')k_b(s, a, x, l, x', l')P(dx')Q(dl') \\
\phi_3(f, t)(s, a, x, l) &= \mu(s, a, x, l, g_s U(a, x, l))f_t(a + t - s, x, l) \\
\phi_4(w_3, w_4, f, t)(s, a, x, l) \\
&= \int_{\mathcal{X}} \int_{\mathcal{L}} e(s, a, x, l)k_e(s, a, x, l, x', l')w_3(x')w_4(l')f_t(a + t - s, x', l')P(dx')Q(dl') \\
\phi_5(f, t)(s, a, x, l) &= e(s, a, x, l)f_t(a + t - s, x, l)
\end{aligned}$$

We write ϕ_0 with a s but we have in mind this function is constant according to time. Apply now (3.18) to the test function $\varphi_t = (a, x, l) \mapsto w_2(t - a)w_3(x)w_4(l)f_t(a, x)$ with $f \in \mathcal{F}$ to obtain the equation

$$\begin{aligned}
\langle Z_t^N, \varphi_t \rangle &= \int_{\mathcal{U}} w_2(-a)w_3(x)w_4(l)\phi_0(f, t)(s, a, x, l)Z_0^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} w_2(s)w_3(x)w_4(l)\phi_1(f, t)(s, a, x, l)Z_s^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} \phi_2(w_3, w_4, f, t)(s, a, x, l)w_2(s)Z_s^N(da, dx, dl) \\
&- \int_0^t \int_{\mathcal{U}} \mu(s, a, x, l, Z_s^N U(a, x, l))w_2(s - a)w_3(x)w_4(l)f_t(a + t - s, x, l)Z_s^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} \phi_4(w_3, w_4, f, t)(s, a, x, l)w_2(s - a)Z_s^N(da, dx, dl) \\
&- \int_0^t \int_{\mathcal{U}} \phi_5(f, t)(s, a, x, l)w_2(s - a)w_3(x)w_4(l)Z_s^N(da, dx, dl) \\
&+ \int_0^t \int_{\mathcal{U}} e_i(s, a, x, l)w_2(s - a)w_3(x)w_4(l)f_t(a + t - s, x, l)daP(dx)Q(dl)ds \\
&+ [M_t^{N, birth}(\varphi_t) + M_t^{N, death}(\varphi_t) + M_t^{N, imm}(\varphi_t) + M_t^{N, swap}(\varphi_t)]
\end{aligned}$$

Subtracting $g(t, a, x, l)daP(dx)Q(dl)$ in the equation above and using that $g(t, a, x, l)$ solves (2.8), we also have

$$\begin{aligned}
\langle \eta_t^N, \varphi_t \rangle &= \int_{\mathcal{U}} w_2(-a)w_3(x)w_4(l)\phi_0(f,t)(s,a,x,l)\eta_0^N(da,dx,dl) \\
&+ \int_0^t \int_{\mathcal{U}} w_2(s)w_3(x)w_4(l)\phi_1(f,t)(s,a,x,l)\eta_s^N(da,dx,dl) \\
&+ \int_0^t \int_{\mathcal{U}} \phi_2(w_3,w_4,f,t)(s,a,x,l)w_2(s)\eta_s^N(da,dx,dl) \\
&- \int_0^t \int_{\mathcal{U}} [\mu(s,a,x,l,Z_s^N U(a,x,l))] w_2(s-a)w_3(x)w_4(l)f_t(a+t-s,x,l)Z_s^N(da,dx,dl) \\
&+ \int_0^t \int_{\mathcal{U}} \phi_3(f,t)(s,a,x,l)w_2(s-a)w_3(x)w_4(l)g(s,a,x,l)P(dx)Q(dl)dads \\
&+ \int_0^t \int_{\mathcal{U}} \phi_4(w_3,w_4,f,t)(s,a,x,l)w_2(s-a)\eta_s^N(da,dx,dl) \\
&- \int_0^t \int_{\mathcal{U}} \phi_5(f,t)(s,a,x,l)w_2(s-a)w_3(x)w_4(l)\eta_s^N(da,dx,dl) \\
&+ [M_t^{N,birth}(\varphi_t) + M_t^{N,death}(\varphi_t) + M_t^{N,imm}(\varphi_t) + M_t^{N,swap}(\varphi_t)]
\end{aligned}$$

Remark 3.32. *Immigration term disappear except in the martingale part. This comes from the fact this is a independant poisson process with the same limit and mean.*

We have to handle the term with interactions, for that we note

$$\Delta\mu(s,a,x,l) = \mu(s,a,x,l,g_s U(a,x,l)) - \mu(s,a,x,l,Z_s^N U(a,x,l))$$

and we do the following.

$$\begin{aligned}
&- \int_0^t \int_{\mathcal{U}} [\mu(s,a,x,l,Z_s^N U(a,x,l)) + e(s,a,x)] w_2(s-a)w_3(x)w_4(l)f_t(a+t-s,x,l)Z_s^N(da,dx,dl) \\
&+ \int_0^t \int_{\mathcal{U}} \mu(s,a,x,l,g_s U(a,x,l))w_2(s-a)w_3(x)w_4(l)f_t(a+t-s,x,l)g(s,a,x,l)P(dx)Q(dl)dads \\
&= - \int_0^t \int_{\mathcal{U}} \mu(s,a,x,l,g_s U(a,x,l))w_2(s-a)w_3(x)w_4(l)f_t(a+t-s,x,l)\eta_s^N(da,dx,dl)ds \\
&+ \int_0^t \int_{\mathcal{U}} \Delta\mu(s,a,x,l)w_2(s-a)w_3(x)w_4(l)f_t(a+t-s,x,l)Z_s^N(da,dx,dl)ds
\end{aligned}$$

Multiplying each term by $w_1(t)$ and integrating from 0 to T , we successively obtain

$$\begin{aligned}
&\int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} w_2(s)w_3(x)w_4(l)\phi_1(f,t)(s,a,x,l)\eta_s^N(da,dx,dl)dsdt \leq C_0 \int_0^T |w_1(t)|\mathcal{W}_{w_2,1,w_3,w_4}(\mathcal{F})_t dt, \\
&\int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} w_2(s)\phi_2(w_3,w_4,f,t)(s,a,x,l)\eta_s^N(da,dx,dl)dsdt \leq C_0|w_3w_4|_1 \int_0^T |w_1(t)|\mathcal{W}_{w_2,1,1,1}(\mathcal{F})_t dt \\
&\int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} \phi_5(f,t)(s,a,x,l)w_2(s-a)w_3(x)w_4(l)\eta_s^N(da,dx,dl)dsdt \leq C_0 \int_0^T |w_1(t)|\mathcal{W}_{1,w_2,w_3,w_4}(\mathcal{F})_t dt \\
&\int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} w_2(s-a)w_3(x)w_4(l)\phi_3(f,t)(s,a,x,l)\eta_s^N(dx,da,dl)dsdt \leq C_0 \int_0^T |w_1(t)|\mathcal{W}_{1,w_2,w_3,w_4}(\mathcal{F})_t dt
\end{aligned}$$

and

$$\int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} w_2(s-a) \phi_4(w_3, w_4, f, t)(s, a, x, l) \eta_s^N(da, dx, dl) ds dt \leq C_0 |w_3 w_4|_1 \int_0^T |w_1(t)| \mathcal{W}_{1, w_2, 1, 1}(\mathcal{F})_t dt$$

After those control, we have to take care of the last term which is

$$\int_0^t \int_{\mathcal{U}} \Delta \mu(s, a, x, l) w_2(s-a) w_3(x) w_4(l) f_t(a+t-s, x, l) Z_s^N(da, dx, dl) ds.$$

By assumption we know this is less or equal to

$$\int_0^T |w_1(t)| \int_0^t \int_{\mathcal{U}} \mathcal{R} \left| \int_{\mathcal{U}} U(a, x, l, \alpha, y, b) \eta_s^N(d\alpha, dy, db) \right| |w_2(s-a) w_3(x) w_4(l)| |f|_{\infty} Z_s^N(da, dx, dl) ds dt$$

Since we assume that $|f|_{\infty} \leq 1$ we just have to care about the term without f . We will use a coupling argument to continue. We can bound Z^N by \tilde{Z}^N . So we see that the preceding term is less than

$$\int_0^T |w_1(t)| \int_0^t \int_{\mathcal{U}} \mathcal{R} \left| \int_{\mathcal{U}} U(a, x, l, \alpha, y, b) \eta_s^N(d\alpha, dy, db) \right| |w_2(s-a) w_3(x) w_4(l)| \tilde{Z}_s^N(da, dx, dl) ds dt$$

Using transport, and the fact in the population \tilde{Z}^N there is no death we have a bound which is

$$\int_0^T |w_1(t)| \int_0^t \int_{\mathcal{U}} \mathcal{R} \left| \int_{\mathcal{U}} U(a+s-t, x, l, \alpha, y, b) \eta_s^N(d\alpha, dy, db) \right| |w_2(t-a) w_3(x) w_4(l)| \tilde{Z}_t^N(da, dx, dl) ds dt$$

Indeed, let $c(x, a)$ be a formal function. Then we have (since $x_i(s) = x_i(t)$)

$$\begin{aligned} \int_{\mathcal{U}} c(x, a) \tilde{Z}_s^N(dx, da) &= \sum_{i=1}^{\tilde{N}_s} c(x_i(s), a_i(s)) = \\ \sum_{i=1}^{\tilde{N}_s} c(x_i(t), a_i(t) + s - t) &\leq \sum_{i=1}^{\tilde{N}_t} c(x_i(t), a_i(t) + s - t) = \int_{\mathcal{U}} c(x, a + s - t) \tilde{Z}_t^N(dx, da) \end{aligned}$$

Now we can just use again the quantity we define to get the bound

$$\int_0^T \int_{\mathcal{U}} |w_1(t) w_2(t-a) w_3(x) w_4(l)| \mathcal{W}_{1, 1, 1, 1}(\mathcal{F})_t \tilde{Z}_t^N(da, dx, dl) ds dt$$

The result easily follows. \square

Remark 3.33. In a similar way of the proof of the proposition 3.31 we get that $\tilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$ is bounded by

$$\begin{cases} |w_1|_1 C_0 \mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0 + \\ C_0 \left(\int_0^T |w_1(t)| \mathcal{W}_{w_2, 1, w_3, w_4}^N(\mathcal{F})_t dt + \int_0^T |w_1(t)| \mathcal{W}_{1, w_2, w_3, w_4}^N(\mathcal{F})_t dt \right) + \\ C_0 |w_3 w_4|_1 \left(\int_0^T |w_1(t)| \mathcal{W}_{w_2, 1, 1, 1}^N(\mathcal{F})_t dt + \int_0^T |w_1(t)| \mathcal{W}_{1, w_2, 1, 1}^N(\mathcal{F})_t dt \right) + \\ \tilde{\mathcal{M}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_t \end{cases} \quad (3.20)$$

Proposition 3.34. *Work under Assumptions 3.1, 3.3, 3.4 and 3.5. We have*

$$\widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T \lesssim \widetilde{L}_1^N(w_1, w_2, w_3, w_4)_T + \widetilde{L}_2^N(w_1, w_2, w_3, w_4)_T$$

up to an explicitly computable constant depending on T and C_0 , as soon as $\mathcal{F} \subset \{|f|_\infty \leq 1\}$. With

$$\widetilde{L}_1^N(w_1, w_2, w_3, w_4)_T = |w_1|_{L^1[0, T]} \max |h_1^2|_{L^1[0, T]} |h_1^3 h_1^4|_1 \mathcal{W}_{h_2^2, h_2^3, h_2^4}^N(\mathcal{F})_0$$

and

$$\widetilde{L}_2^N(w_1, w_2, w_3, w_4)_T = \max |h_1^1 h_1^2|_{L^1[0, T]} |h_1^3 h_1^4|_1 \widetilde{\mathcal{M}}_{h_2^1, h_2^2, h_2^3, h_2^4}^N(\mathcal{F})_T$$

where each (h_1^i, h_2^i) equals $(1, w_i)$ or $(w_i, 1)$.

Proof. To control it we will extensively use the relation (3.20). We apply first with $w_1 = 1$, $w_2 = 1$, $w_3 = 1$ and $w_4 = 1$ to obtain

$$\begin{aligned} \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_T &\leq TC_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + 4C_0 \int_0^T \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_t dt + \widetilde{\mathcal{M}}_{1,1,1,1}^N(\mathcal{F})_T \\ &\leq (TC_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + \widetilde{\mathcal{M}}_{1,1,1,1}^N(\mathcal{F})_T) e^{4C_0 T} \\ &= \widetilde{\mathcal{G}}^{(1), N}(\mathcal{F})_T, \end{aligned}$$

say, by Grönwall lemma. Next, by relation (3.20) with $w_1 = w_2 = 1$, we obtain

$$\begin{aligned} \widetilde{\mathcal{W}}_{1,1,w_3,w_4}^N(\mathcal{F})_T &\leq TC_0 \mathcal{W}_{1,w_3,w_4}^N(\mathcal{F})_0 + 2C_0 \int_0^T \widetilde{\mathcal{W}}_{1,1,w_3,w_4}^N(\mathcal{F})_t dt \\ &\quad + \int_0^T \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_t 2C(w_3, w_4) dt + \widetilde{\mathcal{M}}_{1,1,w_3,w_4}^N(\mathcal{F})_T \\ &\leq (TC_0 \mathcal{W}_{1,w_3,w_4}^N(\mathcal{F})_0 + 2C_0 |w_3, w_4|_1 T \widetilde{\mathcal{G}}^{(1), N}(\mathcal{F})_T + \widetilde{\mathcal{M}}_{1,1,w_3,w_4}^N(\mathcal{F})_T) e^{2C_0 T} \\ &= \widetilde{\mathcal{G}}^{(2), N}_{w_3, w_4}(\mathcal{F})_T, \end{aligned}$$

say, thanks to $\widetilde{\mathcal{G}}^{(1), N}(\mathcal{F})_t$ increasing in t since $\widetilde{\mathcal{M}}_{1,1,1,1}^N(\mathcal{F})_t$ is increasing in t . Apply now relation (3.20) with $w_2 = w_3 = w_4 = 1$, we obtain

$$\begin{aligned} \widetilde{\mathcal{W}}_{w_1,1,1,1}^N(\mathcal{F})_T &\leq |w_1|_1 C_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + 4C_0 \int_0^T w_1(t) \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_t dt + \widetilde{\mathcal{M}}_{w_1,1,1,1}^N(\mathcal{F})_T \\ &\leq |w_1|_1 (C_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + 4C_0 \widetilde{\mathcal{G}}^{(1), N}(\mathcal{F})_T) + \widetilde{\mathcal{M}}_{w_1,1,1,1}^N(\mathcal{F})_T \\ &= \widetilde{\mathcal{G}}^{(3), N}_{w_1}(\mathcal{F})_T, \end{aligned}$$

say. Apply now relation (3.20) with $w_1 = w_3 = w_4 = 1$ so that

$$\begin{aligned} \widetilde{\mathcal{W}}_{1,w_2,1,1}^N(\mathcal{F})_T &\leq TC_0 \mathcal{W}_{w_2,1,1}^N(\mathcal{F})_0 + 2C_0 \int_0^T \widetilde{\mathcal{W}}_{w_2,1,1,1}^N(\mathcal{F})_t dt \\ &\quad + 2C_0 \int_0^T \widetilde{\mathcal{W}}_{1,w_2,1,1}^N(\mathcal{F})_t dt + \widetilde{\mathcal{M}}_{1,w_2,1,1}^N(\mathcal{F})_T \\ &\leq (TC_0 \mathcal{W}_{w_2,1,1}^N(\mathcal{F})_0 + 2C_0 T \widetilde{\mathcal{G}}^{(3), N}_{w_2}(\mathcal{F})_T + \widetilde{\mathcal{M}}_{1,w_2,1,1}^N(\mathcal{F})_T) e^{2C_0 T} \\ &= \widetilde{\mathcal{G}}^{(4), N}_{w_2}(\mathcal{F})_T, \end{aligned}$$

say. Apply now relation (3.20) with $w_2 = 1$ so that

$$\begin{aligned}
\widetilde{\mathcal{W}}_{w_1,1,w_3,w_4}^N(\mathcal{F})_T &\leq |w_1|_1 C_0 \mathcal{W}_{1,w_3,w_4}^N(\mathcal{F})_0 + 2C_0 \int_0^T w_1(t) \widetilde{\mathcal{W}}_{1,1,w_3,w_4}^N(\mathcal{F})_t dt \\
&\quad + 2C(w_3, w_4) \int_0^T w_1(t) \widetilde{\mathcal{W}}_{1,1,1,1}^N(\mathcal{F})_t dt + \widetilde{\mathcal{M}}_{w_1,1,w_3,w_4}^N(\mathcal{F})_T \\
&\leq (C_0 \mathcal{W}_{1,w_3,w_4}^N(\mathcal{F})_0 + 2C_0 \widetilde{\mathcal{G}}_{w_3,w_4}^{(2),N}(\mathcal{F})_T \\
&\quad + 2C_0 |w_3 w_4|_1 \widetilde{\mathcal{G}}_{w_3,w_4}^{(1),N}(\mathcal{F})_T) |w_1|_1 + \widetilde{\mathcal{M}}_{w_1,1,w_3,w_4}^N(\mathcal{F})_T \\
&= \widetilde{\mathcal{G}}_{w_1,w_3,w_4}^{(5),N}(\mathcal{F})_T,
\end{aligned}$$

say. Apply now relation (3.20) with $w_1 = 1$ so that

$$\begin{aligned}
\widetilde{\mathcal{W}}_{1,w_2,w_3,w_4}^N(\mathcal{F})_T &\leq TC_0 \mathcal{W}_{w_2,w_3,w_4}^N(\mathcal{F})_0 + C_0 \int_0^T (\widetilde{\mathcal{W}}_{w_2,1,w_3,w_4}^N(\mathcal{F})_t + \widetilde{\mathcal{W}}_{1,w_2,w_3,w_4}^N(\mathcal{F})_t) dt \\
&\quad + C(w_3, w_4) \int_0^T (\widetilde{\mathcal{W}}_{w_2,1,1,1}^N(\mathcal{F})_t + \widetilde{\mathcal{W}}_{1,w_2,1,1}^N(\mathcal{F})_t) dt + \widetilde{\mathcal{M}}_{1,w_2,w_3,w_4}^N(\mathcal{F})_T \\
&\leq (TC_0 \mathcal{W}_{w_2,w_3,w_4}^N(\mathcal{F})_0 + C_0 \widetilde{\mathcal{G}}_{w_2,w_3,w_4}^{(5),N}(\mathcal{F})_T \\
&\quad + C_0 |w_3 w_4|_1 \left[\widetilde{\mathcal{G}}_{w_2}^{(3),N}(\mathcal{F})_T + \widetilde{\mathcal{G}}_{w_2}^{(4),N}(\mathcal{F})_T \right] + \widetilde{\mathcal{M}}_{1,w_2,w_3,w_4}^N(\mathcal{F})_T) e^{C_0 T} \\
&= \widetilde{\mathcal{G}}_{w_2,w_3,w_4}^{(6),N}(\mathcal{F})_T,
\end{aligned}$$

say, by the previous estimate and Grönwall lemma again. By relation (3.20) and the six previous bounds, we infer that $\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T$ is less than

$$\begin{aligned}
&|w_1|_1 C_0 \mathcal{W}_{w_2,w_3,w_4}^N(\mathcal{F})_0 + C_0 |w_1|_1 (\widetilde{\mathcal{G}}_{w_2,w_3,w_4}^{(5),N}(\mathcal{F})_T + \widetilde{\mathcal{G}}_{w_2,w_3,w_4}^{(6),N}(\mathcal{F})_T) \\
&+ C_0 |w_3 w_4|_1 |w_1|_1 (\widetilde{\mathcal{G}}_{w_2}^{(3),N}(\mathcal{F})_T + \widetilde{\mathcal{G}}_{w_2}^{(4),N}(\mathcal{F})_T) + \widetilde{\mathcal{M}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T.
\end{aligned}$$

expanding the estimates $\mathcal{G}_{w_i}^{(i),N}(\mathcal{F})_T$ in terms of their appropriate arguments concludes the proof without interaction. \square

Remark 3.35. We can set $w_4 = \mathbf{1}_{l=l_0}$ in the precedent proof. This is easy to check by following carefully each step of the proof.

Proposition 3.36. Work under Assumptions 3.1, 3.3, 3.4 and 3.5. We have

$$\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \lesssim e^{\mathcal{R}C_0 T \tilde{P}_T^N} \left[L_1^N(w_1, w_2, w_3, w_4)_T + L_2^N(w_1, w_2, w_3, w_4)_T \right]$$

up to an explicitly computable constant depending on T , C_0 , \mathcal{R} , $|g_0|_1$ and $|e_i|_1$, as soon as $\mathcal{F} \subset \{|f|_\infty \leq 1\}$. With

$$L_1^N(w_1, w_2, w_3, w_4)_T = |w_1|_{L^1[0,T]} \max |h_1^2|_1 |h_1^3 h_1^4|_1 \mathcal{W}_{h_2^2, h_2^3, h_2^4}^N(\mathcal{F})_0$$

and

$$L_2^N(w_1, w_2, w_3, w_4)_T = \max |h_1^1|_{L^1[0,T]} |h_1^2|_1 |h_1^3 h_1^4|_1 \left[\mathcal{M}_{h_2^1, h_2^2, h_2^3, h_2^4}^N(\mathcal{F})_T + \widetilde{\mathcal{M}}_{h_2^1, h_2^2, h_2^3, h_2^4}^N(\mathcal{F})_T \right]$$

where each (h_1^i, h_2^i) equals $(1, w_i)$ or $(w_i, 1)$. $|w_2|_{\bar{1}}$ has to be understood as $\max(T, 1)$ when $w_2 = 1$ and $\max(|w_2|_{L^1[0,T]}, |w_2|_1)$ when we have a fixed w_2 .

Proof. We follow exactly the same step than the precedent ones. The only difference comes from the interaction term, remind

$$C_0 \mathcal{R} \int_0^T \int_{\mathcal{U}} |w_1(t)w_2(t-a)w_3(x)w_4(l)| \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t \tilde{Z}_t^N(dx, da, dl) dt$$

For the first step, where $w_1 = w_2 = w_3 = w_4 = 1$, we then have from the proposition 2.20 and (3.17) that

$$\begin{aligned} \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T &\leq TC_0 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 \\ &\quad + (4C_0 + C_0 \mathcal{R}[(|g_0|_1 + |e_i|_1)e^{2\mathcal{R}T} + \tilde{P}_T^N]) \int_0^T \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t dt + \mathcal{M}_{1,1,1,1}^N(\mathcal{F})_T \\ &\leq (T\mathcal{W}_{1,1,1,1}^N(\mathcal{F})_0 + \mathcal{M}_{1,1,1,1}^N(\mathcal{F})_T) e^{C_0(4+\mathcal{R}[(|g_0|_1 + |e_i|_1)e^{2\mathcal{R}T} + \tilde{P}_T^N])T} \\ &= \mathcal{G}^{(1),N}(\mathcal{F})_T, \end{aligned}$$

After this step we just bound the interaction term with the following

$$\begin{aligned} &\int_0^T \int_{\mathcal{U}} |w_1(t)w_2(t-a)w_3(x)w_4(l)| \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_t \tilde{Z}_t^N(dx, da, dl) dt \\ &\leq \int_0^T \int_{\mathcal{U}} |w_1(t)w_2(t-a)w_3(x)w_4(l)| \mathcal{G}^{(1),N}(\mathcal{F})_T \tilde{Z}_t^N(dx, da, dl) dt \\ &\quad \leq \mathcal{G}^{(1),N}(\mathcal{F})_T \left[|w_1 w_2 w_3 w_4|_{\mathfrak{g}} + \widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T \right] \end{aligned}$$

writing

$$|w_1 w_2 w_3 w_4|_{\mathfrak{g}} = \int_0^T \int_{\mathcal{U}} |w_1(t)w_2(t-a)w_3(x)w_4(l)| \mathfrak{g}(t, a, x, l) dt da P(dx) Q(dl)$$

where we can set each w_i to the value we want.

We have $|w_1 w_2 w_3 w_4|_{\mathfrak{g}} \leq (|g_0|_1 + |e_i|_1)e^{2\mathcal{R}T} |w_1|_{L^1[0,T]} |w_2|_{\bar{1}} |w_3 w_4|_1$.

Following the precedent proofs with those bounds leads us to the result. \square

By Proposition 3.36, we see that the stability of the system is controlled by the initial approximation $\mathcal{W}_{w_2, w_3, w_4}^N(\mathcal{F})_0$ and the propagation of the stochastic term $\mathcal{M}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$, but also the term without interaction $\widetilde{\mathcal{M}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$. We now turn to those latter terms.

6.2 Stability of the stochastic term

We will first cut $\mathcal{M}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T$ into four terms. We define

$$\mathcal{M}_{w_1, w_2, w_3, w_4}^{b,N}(\mathcal{F})_T = \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^{N, birth}(w_2(t-\cdot)w_3 w_4 f_t) dt \right|$$

$$\begin{aligned}\mathcal{M}_{w_1, w_2, w_3, w_4}^{\mu, N}(\mathcal{F})_T &= \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^{N, death}(w_2(t - \cdot) w_3 w_4 f_t) dt \right| \\ \mathcal{M}_{w_1, w_2, w_3, w_4}^{i, N}(\mathcal{F})_T &= \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^{N, imm}(w_2(t - \cdot) w_3 w_4 f_t) dt \right| \\ \mathcal{M}_{w_1, w_2, w_3, w_4}^{s, N}(\mathcal{F})_T &= \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^{N, swap}(w_2(t - \cdot) w_3 w_4 f_t) dt \right|\end{aligned}$$

and also without any confusion possible for each $c \in \{b, \mu, i, s\}$,

$$\mathcal{M}_{w_1, w_2, w_3, w_4}^{c, N}(f)_T = \left| \int_0^T w_1(t) M_t^{N, c}(w_2(t - \cdot) w_3 w_4 f_t) dt \right|$$

To remind we have

$$\begin{aligned}& M_t^{N, birth}(w_2(t - \cdot) w_3 w_4 f_t) \\ &= N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle N Z_{s-}^N, \mathbf{1} \rangle\}} [w_2(s) w_3(x_i) w_4(l_i) f_t(t - s, x_i, l_i) \mathbf{1}_{\{0 \leq \vartheta \leq m_1\}}] \\ &+ w_2(s) w_3(x) w_4(l) f_t(t - s, x, l) \mathbf{1}_{\{m_1 \leq \vartheta \leq m_2\}}] \widetilde{Q}_b(ds, di, d\vartheta, dx, dl) \\ & M_t^{N, death}(w_2(t - \cdot) w_3 w_4 f_t) \\ &= \frac{-1}{N} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle N Z_{s-}^N, \mathbf{1} \rangle, 0 \leq \vartheta \leq m_3\}} \\ &w_2(s - a_i) w_3(x_i) w_4(l_i) f_t(t + a_i - s, x_i, l_i) \widetilde{Q}_\mu(ds, di, d\vartheta, dx, dl) \\ & M_t^{N, imm}(w_2(t - \cdot) w_3 w_4 f_t) \\ &= N^{-1} \int_0^t \int_{\mathcal{U}} w_2(s - a) w_3(x) w_4(l) f_t(a + t - s, x, l) \mathbf{1}_{\{0 \leq \vartheta \leq m_5\}} \widetilde{Q}_i(ds, da, dx, dl) \\ & M_t^{N, swap}(w_2(t - \cdot) w_3 w_4 f_t) \\ &= N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle N Z_{s-}^N, \mathbf{1} \rangle\}} [(w_2(s - a_i) w_3(x) w_4(l) f_t(t + a_i - s, x, l) \\ &- w_2(s - a_i) w_3(x_i) w_4(l_i) f_t(a_i + t - s, x_i, l_i)) \mathbf{1}_{\{0 \leq \vartheta \leq m_4\}}] \widetilde{Q}_e(ds, di, d\vartheta, dx, dl)\end{aligned}$$

We will extensively use the following lemma

Lemma 3.37. *Let a martingale $M_t(f)$*

$$\frac{1}{N} \int_0^t \int_{\mathbb{N} \times [0, \infty) \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{i \leq N_{s-}} \mathbf{1}_{\vartheta < \tau(s, a_i, x_i, l_i, x', l', Z_s^N U(a_i, x_i, l_i))} f(s, a_i, x_i, l_i, x', l') \widetilde{Q}(ds, di, d\vartheta, dx', dl')$$

We can compute its exponential martingale $\Lambda(\lambda, t)$ associated with $\lambda M_t(f)$ to obtain

$$\Lambda(\lambda, t) = \exp(\lambda M_t(f) - B_t(\lambda, f))$$

with $B(\lambda, f)$ equals to

$$\int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} \tau(s, a, x, l, x', l', Z_s^N U(a, x, l)) N \rho \left(\lambda \frac{f(s, a, x, l, x', l')}{N} \right) P(dx') Q(dl') Z_s^N(da, dx, dl) ds$$

Remark 3.38. The Ito's formula shows that $\Lambda(\lambda, t)$ is a local martingale. By usual argument, and in the same idea of our control in the following of this chapter, one can show this is a true martingale.

Proof. We use Ito's Formula noting $\Lambda(\lambda, t)$ simply Λ_t

$$\begin{aligned} \Lambda_t &= \Lambda_0 + \int_0^t \Lambda_{s-} \lambda dM_s - \int_0^t \Lambda_s dB_s + \sum_{s \leq t} [(e^{\lambda M_s} - e^{\lambda M_{s-}}) e^{-B_t} - \lambda \Lambda_{s-} \Delta M_s] \\ &= \Lambda_0 + \int_0^t \Lambda_{s-} \lambda dM_s - \int_0^t \Lambda_s dB_s + \sum_{s \leq t} \Lambda_{s-} \rho(\lambda \Delta M_s) \end{aligned}$$

It implies the finite variation part is

$$\begin{aligned} & - \int_0^t \Lambda_s dB_s \\ & + \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} \tau(s, a, x, l, x', l', Z_s^N U(a, x, l)) N \rho \left(\lambda \frac{f(s, a, x, l, x', l')}{N} \right) P(dx') Q(dl') Z_s^N(da, dx, dl) ds \end{aligned}$$

Which ends the proof. \square

Step 1 First control of $\widetilde{\mathcal{M}}^N$.

We begin with the control of this quantity. We have for this one to follow only exactly the steps of the proof without interaction done in the chapter 2. One can see how to do it with the characteristics. Notations are more complicated but the ideas are the same.

In the step 3 of the precedent proof, see 6.2, for each quantity we control $B_{t_0, t_0}^{N, birth}(f)$ which is equal to

$$\begin{aligned} & \int_0^t \int_{\mathcal{U}} \mathcal{R} [p(s, a, x, l) N \rho(N^{-1} f(t_0, t_0 - s, x, l)) \\ & + \int_{\mathcal{X} \times \mathcal{L}} (1 - p(s, a, x, l)) k_b(s, a, x, l, x', l') N \rho(N^{-1} f(t_0, t_0 - s, x', l')) Q(dl') P(dx')] Z_s^N(da, dx, dl) ds \end{aligned}$$

by

$$\begin{aligned} & |b|_{\infty} N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} \left[\frac{f(t_0, t_0 - s, x, l)}{|f|_{\infty}} \right. \\ & \left. + \int_{\mathcal{X} \times \mathcal{L}} |k_b|_{\infty} \frac{f(t_0, t_0 - s, x', l')}{|f|_{\infty}} Q(dl') P(dx') \right] Z_s^N(da, dx, dl) ds \end{aligned}$$

We do the same for swap and immigration (we dont have death since we are with the process without death and swap out).

$$B_{t_0, t_0}^{N, imm}(f) = \int_0^t \int_{\mathcal{U}} e_i(s, a, x, l) N \rho(N^{-1} f(t_0, a + t_0 - s, x, l)) da ds P(dx) Q(dl)$$

which can be bounded by

$$|e_i|_{\infty} N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} \frac{f(t_0, a + t_0 - s, x, l)}{|f|_{\infty}} da ds P(dx) Q(dl)$$

Then we have $B_{t_0, t_0}^{N, swap}(f)$ equals to

$$\int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} e(s, a, x, l) k_e(s, a, x, l, x', l') N \rho(N^{-1} f(t_0, a + t_0 - s, x', l')) P(dx') Q(dl') Z_s^N(da, dx, dl)$$

which can be bounded by

$$|e|_{\infty} |k_e|_{\infty} N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} \int_{\mathcal{X} \times \mathcal{L}} \frac{f(t_0, a + t_0 - s, x', l')}{|f|_{\infty}} P(dx') Q(dl') Z_s^N(da, dx, dl)$$

Introducing $\mathcal{A}(\kappa) = \{N^{-1} \tilde{n}_T \leq C(1 + \kappa)\}$ where C is big enough to ensure that

$$\int_0^{\infty} \mathbb{P}(\mathcal{A}(\kappa)^c) e^{\kappa} d\kappa \leq \frac{1}{2}, \quad (3.21)$$

and

$$\mathbb{E}[\exp(\lambda |\xi(f) - \xi(g)|) \mathbf{1}_{\mathcal{A}(\kappa)}] \leq 2 \exp(c_1(1 + \kappa) \rho(c_2 d(f, g) \lambda)), \quad (3.22)$$

for every $\lambda \geq 0$ and some $c_1, c_2 > 0$.

Proposition 3.39. *Work under the assumption 3.11, then we have (3.21)*

And so we can control \tilde{n}_T with a great probability. We use again the same argument.

Proposition 3.40. *Assume that $(\xi(f))_{f \in \mathcal{F}}$ and $(\mathcal{A}(\kappa))_{\kappa > 0}$ satisfy (3.21) and (3.22) and that $\xi(f_0) = 0$ for some $f_0 \in \mathcal{F}$. Then there exists a choice $\varpi = \varpi(c_1, c_2) > 0$ such that for every $u \geq 0$:*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\xi(f)| \geq 8(u + \int_0^{\text{diam}_{\tilde{d}}(\mathcal{F})} \log(1 + \mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)) d\epsilon)\right) \leq (e^{u/\text{diam}_{\tilde{d}}(\mathcal{F})} - 1)^{-1},$$

where $\tilde{d} = \varpi d$, $\text{diam}_{\tilde{d}}(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} \tilde{d}(f, g)$ and $\mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)$ is the minimal number of balls of \tilde{d} -size $\epsilon > 0$ that are necessary to cover \mathcal{F} .

We see the only point is to get the dependence of c_1 and c_2 in term of N and w_i norm. We just have here after the first step

$$c_1 \approx \frac{N}{|w_1 w_2 w_3 w_4|_{\infty}} |w_1|_1 |w_2 w_3 w_4|_{\infty}$$

$$c_2 \approx \frac{|w_1 w_2 w_3 w_4|_\infty}{N}$$

Which in the end gives us the following mild concentration of order $N^{-1/2}$ for

$$\{|w_1|_{1,\infty} |w_2 w_3 w_4|_\infty\}^{-1} \widetilde{\mathcal{W}}^N_{w_1, w_2, w_3, w_4}(\mathcal{F})_T$$

Remark 3.41. w_4 is not very important to take care of. Indeed in the end we just have to check if we are able to take the indicator function in one trait l_0 . If one check the proof he will see everything can be written in a way to let w_4 be an indicator function.

Step 2 Final control of $\widetilde{\mathcal{M}}^N$.

We follow the path of the precedent chapter to obtain the control after iteration of $\widetilde{\mathcal{M}}^N$ and $\widetilde{\mathcal{W}}^N$.

This is easy to see the only problems will come from the swap and the birth. We just have to take care of the last terms and we can use again the quantity $\widetilde{\mathcal{W}}^N$ to bound those two terms. We want to bound

$$\int_0^t \int_{\mathcal{U}} [w_2(s) w_3(x) w_4(l) + \int_{\mathcal{X} \times \mathcal{L}} w_2(s) w_3(x') w_4(l') Q(dl') P(dx')] Z_s^N(da, dx, dl)$$

by

$$\widetilde{\mathcal{W}}^N_{w_2, 1, w_3, w_4}(\mathcal{F})_T + |w_3|_1 |w_4|_1 \widetilde{\mathcal{W}}^N_{w_2, 1, 1, 1}(\mathcal{F})_T + |w_2 \cdot 1 \cdot w_3 w_4|_{\mathfrak{g}} + |w_3 w_4|_1 |w_2 \cdot 1|_{\mathfrak{g}}$$

which is, with big probability, and up to a constant, thanks to the first step smaller than

$$|w_2 \cdot 1 \cdot w_3 w_4|_{\mathfrak{g}} + |w_3 w_4|_1 |w_2 \cdot 1|_{\mathfrak{g}} + |w_3|_\infty |w_4|_\infty \frac{|w_2|_{1,\infty}}{N^{1/2}} + |w_3|_\infty |w_4|_\infty \frac{|w_2|_\infty}{N^{1/2}}$$

We can do the same for the swap. We will write to make it short $|w_2 \cdot 1 \cdot w_3 w_4|_{\mathfrak{g}} + |w_3 w_4|_1 |w_2 \cdot 1|_{\mathfrak{g}} = |w_2 w_3 w_4|$. So we see that we have for the iteration, using the concentration from before and assuming we do as in the precedent chapter with c'_1 equals to

$$\frac{N}{|w_1 w_2 w_3 w_4|_\infty} |w_1|_1 \left[|w_2 w_3 w_4| + |w_3|_1 |w_4|_1 \frac{|w_2|_{1,\infty}}{N^{1/2}} + |w_3|_\infty |w_4|_\infty \frac{|w_2|_{1,\infty}}{N^{1/2}} + |w_3|_\infty |w_4|_\infty \frac{|w_2|_\infty}{N^{1/2}} \right]$$

and

$$c'_2 = c_2 \approx \frac{|w_1 w_2 w_3 w_4|_\infty}{N}$$

Since we assume that $|w_3 w_2|_\infty N^{-1/2} \lesssim |w_3 w_2|_1$, in the end we have the mild concentration of order $N^{-1/2}$ for

$$\{|w_1 w_2 w_3|_{1,\infty} |w_4|_\infty\}^{-1} \widetilde{\mathcal{M}}^N_{w_1, w_2, w_3, w_4}(\mathcal{F})_T$$

Then, in the same spirit of the precedent chapter we have a mild concentration of order $r_N \wedge N^{-1/2}$ for $\{|w_1 w_2 w_3|_{1,\infty} |w_4|_\infty\}^{-1} \widetilde{\mathcal{W}}^N_{w_1, w_2, w_3, w_4}(\mathcal{F})_T$, $\{|w_2 w_3|_{1,\infty} |w_4|_\infty\}^{-1} \widetilde{\mathcal{M}}^N_{w_2, w_3, w_4}(\mathcal{F})_T$ and $\{|w_2 w_3|_{1,\infty} |w_4|_\infty\}^{-1} \widetilde{\mathcal{W}}^N_{w_2, w_3, w_4}(\mathcal{F})_T$.

Step 3 Control of \mathcal{M}^N . Now we can easily control this quantity. Indeed the only complicated term will come with the term from death. This is the one bringing the interactions coming from μ .

To remind we have

$$B_{t_0, t_0}^{N, death}(f) = \int_0^t \int_{\mathcal{U}} \mu(s, a, x, l, Z_s^N U(a, x, l)) N \rho(N^{-1} f(t_0, t_0 - s, x, l)) Z_s^N(da, dx, dl)$$

Which can be bounded via the assumption 3.1

$$|\mu|_{\infty} (1 + |U|_{\infty} N^{-1} N_T) N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} \frac{f(t_0, t_0 - s, x, l)}{|f|_{\infty}} Z_s^N(da, dx, dl)$$

We use the fact we control well \tilde{Z}_t^N , and we get the bound

$$\mathcal{R}(1 + \mathcal{R} N^{-1} \tilde{n}_T) N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} \frac{f(t_0, t_0 - s, x, l)}{|f|_{\infty}} \tilde{Z}_s^N(da, dx, dl)$$

And if we put the term with w_1, w_2, w_3 et w_4 we get

$$\mathcal{R}(1 + \mathcal{R} N^{-1} \tilde{n}_T) |w_1 w_2 w_3 w_4|_{\infty}^{-1} N \rho(N^{-1} |f|_{\infty}) \int_0^t \int_{\mathcal{U}} w_1(t_0) w_2(t_0 - s) w_3(x) w_4(l) d(f_1, f_2) \tilde{Z}_s^N(da, dx, dl)$$

We can just use now that

$$\begin{aligned} & \int_0^t \int_{\mathcal{U}} [w_1(t_0) w_2(t_0 - s) w_3(x) w_4(l) d(f_1, f_2) \tilde{Z}_s^N(da, dx, dl) \\ & \lesssim |w_1(t_0)| |w_2 w_3 w_4|_{\mathfrak{g}} + |w_1(t_0)| \widetilde{\mathcal{W}}_{1, w_2, w_3, w_4}^N(\mathcal{F})_T \end{aligned}$$

and $\tilde{n}_T N^{-1} \lesssim 1 + \tilde{P}_T^N$.

Using the control we have of these quantities which are respectively

$$\mathbb{P}\left(\widetilde{\mathcal{W}}_{1, w_2, w_3, w_4}^N(\mathcal{F})_T \gtrsim |w_1 w_2 w_3|_{1, \infty} |w_4|_{\infty} N^{-1/2} (1 + u)\right) \leq \frac{1}{e^u - 1}$$

and

$$\mathbb{P}\left(\tilde{P}_T^N \gtrsim \frac{1 + u}{N^{1/2}}\right) \leq \frac{1}{e^u - 1}$$

This is then easy to see that we can define an event \mathcal{B}_N where $\tilde{P}_T^N \leq 2$ and with $\mathbb{P}(\mathcal{B}_N^c) \leq e^{-\sqrt{N}}$ for N big enough (for example $N \geq 16$ is sufficient). If we define $\mathbb{P}_{\mathcal{B}}$ the conditional probability with respect to the event \mathcal{B}_N , we can apply the proposition 3.40 with this probability. Indeed we have, defining $\mathcal{A}(\kappa) = \{\widetilde{\mathcal{W}}_{w_1, w_2, w_3, w_4}^N(\mathcal{F})_T \lesssim (1 + \kappa)\}$, that

$$\int_0^{\infty} \mathbb{P}_{\mathcal{B}}(\mathcal{A}(\kappa)^c) e^{\kappa} d\kappa \leq 1/2$$

For N great enough, since we can assume that $\mathbb{P}(\mathcal{B}_N) \leq 1/2$ and we have

$$\mathbb{P}_{\mathcal{B}}(\mathcal{A}(\kappa)^c) \leq \mathbb{P}(\mathcal{A}(\kappa)^c) / \mathbb{P}(\mathcal{B}_N) \leq 2\mathbb{P}(\mathcal{A}(\kappa)^c).$$

We also have the second condition

$$\mathbb{E}_{\mathcal{B}}[\exp(\lambda|\xi(f) - \xi(g)|)\mathbf{1}_{\mathcal{A}(\kappa)}] \leq 4\exp(c_1(1 + \kappa)\rho(c_2d(f, g)\lambda)), \quad (3.23)$$

With

$$c_1 \approx 2|w_1|_1 \left[|w_2w_3w_4|_{\mathfrak{g}} + |w_2w_3|_{1,\infty}|w_4|_{\infty}N^{-1/2} \right] |w_1w_2w_3w_4|_{\infty}^{-1}N$$

and

$$c_2 \approx \frac{|w_1w_2w_3w_4|_{\infty}}{N}$$

where \approx stands for equal up to a constant.

This will lead us to the mild concentration under $\mathbb{P}_{\mathcal{B}}$ of $|w_1w_2w_3|_{1,\infty}^{-1}|w_4|_{\infty}^{-1}\mathcal{M}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T$ with rate $N^{-1/2}$.

6.3 Proof of the first part of Theorem 3.13

The control under $\mathbb{P}_{\mathcal{B}}$ is easy to obtain. It just comes from the fact that $\mathbb{P}_{\mathcal{B}}(A) \leq 2\mathbb{P}(A)$ for N great enough. So the assumption on the initial condition will transfer under $\mathbb{P}_{\mathcal{B}}$.

Now we just have to go on the control of $\mathbb{E}[\varphi(\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)]$.

To do so we will just write

$$\mathbb{E}[\varphi(\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)] = \mathbb{E}_{\mathcal{B}}[\varphi(\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)]\mathbb{P}(\mathcal{B}_N) + \mathbb{P}(\mathcal{B}_N^c)\mathbb{E}[\varphi(\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)|\mathcal{B}_N^c]$$

So we have the first term of the Theorem. For the second term we just have to use that

$$\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \lesssim |w_1w_2w_3w_4|_{\mathfrak{g}} + \widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T + |w_1w_2w_3w_4|_{\mathfrak{g}}.$$

So we get

$$\varphi(\mathcal{W}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T) \lesssim \phi_{\varphi}(|w_1w_2w_3w_4|_{\mathfrak{g}}) + \varphi(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)$$

and

$$\mathbb{P}(\mathcal{B}_N^c)\mathbb{E}[\varphi(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)|\mathcal{B}_N^c] \leq \mathbb{E}[\varphi(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T)]$$

We conclude the proof with

$$|w_1w_2w_3w_4|_{\mathfrak{g}} \lesssim |w_1w_2w_3w_4|_1$$

6.4 Proof of the second part of Theorem 3.13

In the same spirit of the precedent chapter and what we have done in the proof, we can obtain the mild concentration at rate $N^{-1/2}$ of $|w_2w_3|_{1,\infty}^{-1}|w_4|_{\infty}^{-1}\widetilde{W}_{w_2,w_3,w_4}^N(\mathcal{F})_t$ for any t , and of $|w_2w_3|_{1,\infty}^{-1}|w_4|_{\infty}^{-1}W_{w_2,w_3,w_4}^N(\mathcal{F})_t$ for any t on \mathcal{B}_N .

And we use the bound

$$W_{w_2,w_3,w_4}^N(\mathcal{F})_t \lesssim |w_2w_3w_4|_1 + \widetilde{W}_{w_2,w_3,w_4}^N(\mathcal{F})_t$$

which concludes the proof.

7 Proof of Theorem 3.15 and 3.19

7.1 Majoration of the death term

Lemma 3.42.

$$\mathcal{D}_T^N(w_1, w_2, w_3, w_4) \leq |(\Delta_{w_1, w_2, w_3, w_4}^N)_T| + \mathcal{I}_T^N(w_1, w_2, w_3, w_4) \quad (3.24)$$

With $(\Delta_{w_1, w_2, w_3, w_4}^N)_T$ is a squared integrable martingale with predictable compensator $\langle \Delta_{w_1, w_2, w_3, w_4}^N \rangle_t$ equals to

$$N^{-1} \int_0^t \int_{\mathcal{U}} w_1(s)^2 w_2(s-a)^2 w_3(x)^2 w_4(l)^2 \mu(s, a, x, l, Z_{s-}^N U(a, x, l)) Z_{s-}^N (da, dx, dl) ds. \quad (3.25)$$

Moreover we have

$$\langle \Delta_{w_1, w_2, w_3, w_4}^N \rangle_T \lesssim N^{-1} [\mathcal{I}_T^N(w_1^2, w_2^2, w_3^2, w_4^2) + |w_1 w_2 w_3 w_4|_2^2] \quad (3.26)$$

The last term is

$$\begin{aligned} \mathcal{I}_T^N &= \left| \int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} \mu(s, a, x, l, Z_s^N U(a, x, l)) w_2(s-a) w_3(x) w_4(l) Z_s^N (da, dx, dl) ds dt \right. \\ &\quad \left. - \int_0^T w_1(t) \int_0^t \int_{\mathcal{U}} \mu(s, a, x, l, gU(a, x, l)) w_2(s-a) w_3(x) w_4(l) g(s, a, x, l) P(dx) Q(dl) dadts dt \right| \end{aligned}$$

Proof. We have the following representation in our model

$$\Gamma^N(dt, da) = N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle NZ_{s-}^N, \mathbf{1} \rangle, 0 \leq \vartheta \leq m_3\}} \delta_{a_i, x_i, l_i} (da, dx, dl) Q_\mu(ds, di, d\vartheta, dx, dl)$$

We add and subtract in the left-hand side of (3.24) the term

$$\int_0^T \int_{\mathcal{U}} w_1(s) w_2(s-a) w_3(x) w_4(l) \mu(s, a, x, l, Z_{s-}^N U(a, x, l)) Z_s^N (da, dx, dl) ds$$

and readily obtain the desired inequality with $(\Delta_{w_1, w_2, w_3, w_4}^N)_t$ equals to

$$\begin{aligned} &\int_0^t \int_{\mathcal{U}} w_1(s) w_2(s-a) w_3(x) w_4(l) (\Gamma^N(ds, da, dx, dl) - \mu(s, a, x, l, Z_{s-}^N U(a, x, l)) Z_s^N (da, dx, dl)) \\ &= N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle NZ_{s-}^N, \mathbf{1} \rangle, 0 \leq \vartheta \leq m_3\}} w_1(s) w_2(s-a_i) w_3(x_i) w_4(l_i) \tilde{Q}_\mu(ds, di, d\vartheta, dx, dl), \end{aligned}$$

where $\tilde{Q}_\mu(ds, di, d\vartheta, dx, dl) = Q_\mu(ds, di, d\vartheta, dx, dl) - ds \otimes di \otimes d\vartheta \otimes P(dx) \otimes Q(dl)$ is the associated compensated measure. Thus $(\Delta_{w_1, w_2, w_3, w_4}^N)_t$ is a martingale and (3.25) readily follows. The proof of (3.26) is easily obtained as done multiple times before. \square

7.2 Control of \mathcal{I}_T^N and $|\Delta_T^N|$

We write $v_N(w_1, w_2, w_3, w_4) = |w_1 w_2 w_3|_{1, \infty} |w_4|_\infty N^{-1/2}$. We assume that $|w_1 w_2 w_3|_\infty |w_4|_\infty N^{-1} \approx v_N^2$, which is true in our statistical purpose. We also work with the minimal set \mathcal{F} as defined in the proposition 3.9

Lemma 3.43. *There exists $C(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C$ such that*

$$\mathbb{P}(\mathcal{I}_T^N(w_1, w_2, w_3, w_4) \geq C v_N(w_1, w_2, w_3, w_4)(1+u)) \lesssim (e^u - 1)^{-1} \wedge 1 + (e^{\sqrt{1+u}-1} - 1)^{-1} \wedge 1$$

Proof. We have

$$\mathcal{I}_T^N \leq C_1[|w_1|_1 \mathcal{W}_{1,w_2,w_3,w_4}^N(\mathcal{F})_T + |w_1 w_2 w_3 w_4|_1 \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T + \widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T]$$

So

$$\begin{aligned} \mathbb{P}(\mathcal{I}_T^N \geq 3C_1[C' + C'' + C''']v_N(1+u)) &\leq \mathbb{P}(|w_1|_{1,\infty} \mathcal{W}_{1,w_2,w_3,w_4}^N(\mathcal{F})_T \geq C'v_N(1+u)) \\ &\quad + \mathbb{P}(|w_1 w_2 w_3|_{1,\infty} |w_4|_\infty \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T \geq C''v_N(1+u)) \\ &\quad + \mathbb{P}(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T \geq C'''v_N(1+u)) \\ &\lesssim (e^u - 1)^{-1} \wedge 1 \\ &\quad + (e^u - 1)^{-1} \wedge 1 \\ &\quad + \mathbb{P}(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \geq \sqrt{C'''v_N(1+u)}) \\ &\quad + \mathbb{P}(\mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T \geq \sqrt{C'''v_N(1+u)}) \end{aligned}$$

By union bound. Then for N great enough, we have $v_N \leq 1$ so $\sqrt{v_N} \geq v_N$, and $v_N \geq N^{-1/2}$

$$\begin{aligned} \mathbb{P}(\mathcal{I}_T^N \geq 3C_1[C' + C'' + C''']v_N(1+u)) &\lesssim (e^u - 1)^{-1} \wedge 1 \\ &\quad + \mathbb{P}(\widetilde{\mathcal{W}}_{w_1,w_2,w_3,w_4}^N(\mathcal{F})_T \geq v_N \sqrt{C'''(1+u)}) \\ &\quad + \mathbb{P}(\mathcal{W}_{1,1,1,1}^N(\mathcal{F})_T \geq N^{-1/2} \sqrt{C'''(1+u)}) \\ &\lesssim (e^u - 1)^{-1} \wedge 1 \\ &\quad + (e^{\sqrt{1+u}-1} - 1)^{-1} \wedge 1 \end{aligned}$$

□

Lemma 3.44. *There exist constants $C(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C$, $C_1(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C_1$ and $C_2(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C_2$ such that*

$$\begin{aligned} \mathbb{P}(|\Delta_T^N| \geq C v_N u) &\lesssim (e^{\sqrt{u}} - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\sqrt{u}}-1} - 1)^{-1} \wedge 1 + 2 \exp\left(-\frac{u^2}{8C_2}\right) \mathbf{1}_{u \leq \frac{C_2}{C_1 v_N}} \\ &\quad + 2 \exp\left(-\frac{\sqrt{u}}{v_N C_1}\right) \mathbf{1}_{u \geq \frac{C_2}{C_1 v_N}} \end{aligned}$$

for N great enough assuming $N v_N$ converge to infinity.

Proof. Let $B_{w_1,w_2,w_3,w_4}^N = B^N$ be the bound of the jumps of $\Delta_{w_1,w_2,w_3,w_4}^N$. We plan to apply a classical deviation inequality for martingales (see *e.g.* Lemma 2.1 in van der Geer [64] or the classical textbook Shorack and Wellner [61]), namely:

$$\mathbb{P}(|\Delta_T^N| \geq u, \langle \Delta^N \rangle_T \leq v, B^N \leq c) \leq 2 \exp\left(-\frac{v}{c^2} \tilde{\rho}\left(\frac{cu}{v}\right)\right) \quad (3.27)$$

Let $\mathcal{A}(\alpha, \beta) = \{\tilde{n}_T/N \leq C_1[1 + \frac{1+\alpha}{\sqrt{N}}]\}$, and $N^{-1}\mathcal{I}_T^N(w_1^2, w_2^2, w_3^2, w_4^2) \leq C_2 v_N^2(1 + \beta)N^{-1/2}$, with C_1, C_2 such that

$$\begin{aligned} \mathbb{P}(\mathcal{A}(\alpha, \beta)^c) &\leq \mathbb{P}(\tilde{n}_T/N \geq C_1[1 + \frac{1+\alpha}{\sqrt{N}}]) + \mathbb{P}(N^{-1}\mathcal{I}_T^N(w_1^2, w_2^2, w_3^2, w_4^2) \geq C_2 v_N^2(1 + \beta)N^{-1/2}) \\ &\lesssim (e^\alpha - 1)^{-1} \wedge 1 + (e^\beta - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\beta}-1} - 1)^{-1} \wedge 1 \end{aligned}$$

Under the event $\mathcal{A}(\alpha, \beta)$ we have

$$B^N \leq |w_1 w_2 w_3 w_4|_\infty N^{-1} C_1 [1 + (1 + \alpha)N^{-1/2}] \lesssim v_N^2 C_1 [1 + (1 + \alpha)N^{-1/2}]$$

and

$$\langle \Delta^N \rangle_T \leq v_N^2 C_2 [1 + (1 + \beta)N^{-1/2}]$$

And we get

$$\begin{aligned} \mathbb{P}(|\Delta_T^N| \geq v_N u) &\lesssim \mathbb{P}(|\Delta_T^N| \geq v_N u, \mathcal{A}(\alpha, \beta)) + (e^\alpha - 1)^{-1} \wedge 1 \\ &\quad + (e^\beta - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\beta}-1} - 1)^{-1} \wedge 1 \\ &\lesssim 2 \exp\left(-\frac{C_2 [1 + (1 + \beta)N^{-1/2}]}{v_N^2 C_1^2 [1 + (1 + \alpha)N^{-1/2}]^2} \tilde{\rho}\left(v_N u \frac{C_1 [1 + (1 + \alpha)N^{-1/2}]}{C_2 [1 + (1 + \beta)N^{-1/2}]}\right)\right) \\ &\quad + (e^\alpha - 1)^{-1} \wedge 1 + (e^\beta - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\beta}-1} - 1)^{-1} \wedge 1 \end{aligned}$$

Taking $\beta(u) = \alpha(u) = \sqrt{u}$ we get

$$\begin{aligned} &2 \exp\left(-\frac{C_2 [1 + (1 + \beta)N^{-1/2}]}{v_N^2 C_1^2 [1 + (1 + \alpha)N^{-1/2}]^2} \tilde{\rho}\left(v_N u \frac{C_1 [1 + (1 + \alpha)N^{-1/2}]}{C_2 [1 + (1 + \beta)N^{-1/2}]}\right)\right) \\ &= 2 \exp\left(-\frac{C_2}{v_N^2 C_1^2 [1 + (1 + \sqrt{u})N^{-1/2}]^2} \tilde{\rho}\left(v_N u \frac{C_1}{C_2}\right)\right) \\ &\leq 2 \exp\left(-\frac{u^2}{4C_2} \frac{1}{1 + (1 + \frac{\sqrt{C_1 v_N + \sqrt{C_2}}}{\sqrt{N C_1 v_N}})}\right) \mathbf{1}_{u \leq \frac{C_2}{C_1 v_N}} + 2 \exp\left(-\frac{\sqrt{u}}{v_N C_1}\right) \mathbf{1}_{u \geq \frac{C_2}{C_1 v_N}} \end{aligned}$$

□

combining the three precedent lemmas we get the Theorem. Indeed by the lemma 3.42, we have

$$\mathcal{D}_T^N(w_1, w_2, w_3, w_4) \leq |(\Delta_{w_1, w_2, w_3, w_4}^N)_T| + \mathcal{I}_T^N(w_1, w_2, w_3, w_4)$$

So by union bound

$$\begin{aligned} \mathbb{P}(\mathcal{D}_T^N(w_1, w_2, w_3, w_4) \geq 2C v_N(1 + u)) &\leq \mathbb{P}(|(\Delta_{w_1, w_2, w_3, w_4}^N)_T| \geq C v_N(1 + u)) \\ &\quad + \mathbb{P}(\mathcal{I}_T^N(w_1, w_2, w_3, w_4) \geq C v_N(1 + u)) \end{aligned}$$

We just adapt the lemma 3.44 to get

$$\begin{aligned} \mathbb{P}(|(\Delta_{w_1, w_2, w_3, w_4}^N)_T| \geq C v_N(1 + u)) &\lesssim (e^{\sqrt{1+u}} - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\sqrt{1+u}-1}} - 1)^{-1} \wedge 1 \\ &\quad + 2 \exp\left(-\frac{(1+u)^2}{8C_2}\right) \mathbf{1}_{u \leq \frac{C_2}{C_1 v_N} - 1} + 2 \exp\left(-\frac{\sqrt{u+1}}{v_N C_1}\right) \mathbf{1}_{u \geq \frac{C_2}{C_1 v_N} - 1} \end{aligned}$$

7.3 Proof of the Theorem 3.19

Let $A_N = A|w_1w_2w_3|_{1,\infty} \ln(N)^4 N^{-1/2}$ and $v_N = |w_1w_2w_3|_{1,\infty} N^{-1/2}$, such that $A_N/v_N = A \ln(N)^4$.

Lemma 3.45. *Let X be a random variable and $C(\mathcal{R}, C_0, T, |g_0|_1, |e_i|_1) = C$ a constant such that*

$$\begin{aligned} \mathbb{P}(X \geq Cv_N(1+u)) &\lesssim (e^u - 1)^{-1} \wedge 1 + (e^{\sqrt{1+u}-1} - 1)^{-1} \wedge 1 \\ &\quad + (e^{\sqrt{1+u}} - 1)^{-1} \wedge 1 + (e^{\sqrt{1+\sqrt{1+u}}-1} - 1)^{-1} \wedge 1 \\ &\quad + 2 \exp\left(-\frac{(1+u)^2}{8C_2}\right) \mathbf{1}_{u \leq \frac{c_2}{c_1 v_N} - 1} + 2 \exp\left(-\frac{\sqrt{u+1}}{v_N C_1}\right) \mathbf{1}_{u \geq \frac{c_2}{c_1 v_N} - 1} \end{aligned}$$

then $\mathbb{E}[\{(X^2) - A_N^2\}_+] \lesssim N^{-3}$ for a good choice of A .

Proof. Few remarks before doing the inequalities, $1/v_N \geq A_N/v_N \rightarrow \infty$. For any increasing function $f(u)$ we have $(e^{f(u)} - 1)^{-1} \leq 2e^{-f(u)}$ as long as $f(u) \geq \ln(2)$, which is the case for $f(u) = \sqrt{u}$ when $u \geq \ln(2)^2$ and for $f(u) = \sqrt{1+\sqrt{u}} - 1$ when $u \geq (1 + \ln(2))^2 - 1$. $A_N/v_N \geq 1$ for N great enough, so we get $\iota(A_N/v_N) \lesssim A_N/v_N$ where $\iota(x) = \sqrt{\sqrt{x} + 1}$. So we get the following bound for $\mathbb{E}[\{(X^2) - A_N^2\}_+]$ by copying the argument in the precedent chapter,

$$\begin{aligned} \mathbb{E}[\{(X^2) - A_N^2\}_+] &\lesssim \int_0^\infty \mathbb{P}(X \geq (A_N^2 + \kappa)^{1/2}) d\kappa \\ &\lesssim \int_{A_N/Cv_N-1}^\infty v_N^2 \mathbb{P}(X \geq Cv_N(1+u))(1+u) du \end{aligned}$$

In order to simplify the notation, we will write $A_N/Cv_N - 1 \geq A/C \ln(N)^4 - 1 = u_N - 1$ for N great enough, and we get

$$\begin{aligned} &\lesssim \int_{u_N-1}^\infty v_N^2 (1+u) e^{-u} du + \int_{u_N-1}^\infty v_N^2 (1+u) e^{-\sqrt{1+u}} du + \int_{u_N-1}^\infty v_N^2 (1+u) e^{-\sqrt{1+\sqrt{1+u}}} du \\ &\quad + \int_{u_N-1}^\infty v_N^2 2 \exp\left(-\frac{(1+u)^2}{8C_2}\right) (1+u) du + v_N^2 \int_{\frac{c_2}{c_1 v_N} - 1}^\infty 2 \exp\left(-\frac{\sqrt{u+1}}{v_N C_1}\right) (1+u) du \\ &\lesssim \int_{u_N}^\infty v_N^2 u e^{-u} du + \int_{u_N}^\infty v_N^2 u e^{-\sqrt{u}} u du + \int_{u_N}^\infty v_N^2 u e^{-\sqrt{1+\sqrt{u}}} du \\ &\quad + \int_{u_N}^\infty v_N^2 2 \exp\left(-\frac{u^2}{8C_2}\right) u du + v_N^2 \int_{\frac{c_2}{c_1 v_N}}^\infty 2 \exp\left(-\frac{\sqrt{u}}{v_N C_1}\right) u du \\ &\lesssim v_N^2 [-u e^{-u} - u]_{u_N}^\infty + v_N^2 \left[-2e^{-\sqrt{u}}(u^{3/2} + 3u + 6\sqrt{u} + 6)\right]_{u_N}^\infty \\ &\quad + v_N^2 \left[-4e^{1-\sqrt{1+\sqrt{u}}}(u^{3/2}(\iota(u) + 7) + 6u(7\iota(u) + 36) + 24\sqrt{u}(36\iota(u) + 115) + \mathfrak{C}_1 \iota(u) + \mathfrak{C}_2)\right]_{u_N}^\infty \\ &\quad + v_N^2 \exp\left(-\frac{u_N^2}{8C_2}\right) \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}[\{(X^2) - A_N^2\}_+] &\lesssim v_N^2 \exp\left(-\frac{u_N^2}{8C_2}\right) + v_N^{1/2} A_N^{3/2} e^{-\sqrt{u_N}} + A_N^{5/2} v_N^{-1/2} e^{-u_N^{1/4}} + A_N v_N e^{-u_N} \\
 &\lesssim v_N^2 e^{-\widetilde{C}_1 A^2 \ln(N)^8} + v_N^{1/2} A_N^{3/2} e^{-\sqrt{A} \widetilde{C}_2 \ln(N)^2} \\
 &\quad + A_N^{5/2} v_N^{-1/2} e^{-A^{1/4} \ln(N) \widetilde{C}_3} + A_N v_N e^{-A \ln(N)^4 \widetilde{C}_4}
 \end{aligned}$$

And taking A big enough we get the result. \square

Which leads to the result following carefully the proof of the precedent chapter, see the subsection 7.2. We just have to take care of one last point, which is if the minimal entropy is finite, but we know this is the case thank to the proposition 3.9.

7.4 Proof of the proposition 3.9

We assume here that $|w_3 w_4|_1$ is just a constant, depending on the kernels. We use the ϕ_i defined in the assumption 3.5.

The minimal set \mathcal{F} is the one where

1. The constants in $[0, C_0]$ are in \mathcal{F}
2. for all $(t, a_0, x_0, l_0) \in [0, T] \times \mathcal{U}$

$$C_0^{-1} \phi_U(t, a_0, x_0, l_0)(s, a, x, l) = C_0^{-1} U(a_0 + s - t, x_0, l_0, a, x, l) \in \mathcal{F}$$

3. \mathcal{F} is stable by $f \rightarrow -f$, and $f \rightarrow C_0^{-1} \phi_i(f, t, w_3, w_4)$ for all $t \in [0, T]$, $f \in \mathcal{F}$ and w_3, w_4 fixed.

Define \mathcal{F}_0 as the set of the functions being in \mathcal{F} without any composition with any ϕ_i . So

$$\mathcal{F}_0 = \{0, C_0^{-1}, \forall (t, a_0, x_0, l_0) \in [0, T] \times \mathcal{U}, C_0^{-1} \phi_U(t, a_0, x_0, l_0)(s, a, x, l)\}$$

Assume that γ is the minimal regularity for each coordinates for all the parameters functions, then we have

Proposition 3.46.

$$N(\mathcal{F}, \epsilon, \|\cdot\|_\infty) \lesssim \epsilon^{\log(\epsilon)/(\gamma \log(2))}$$

Proof. We will write $P_K(f)$ any composition of K $C_0^{-1} \phi_i(\cdot, t_j)$ starting from f , so implicitly there is sequence of t_j implied in the composition. We will also note $P_{\widetilde{K}}(f)$ when we take another sequence \widetilde{t}_j , but with the same sequence of ϕ_i . We first notice that we just need $K = -\log(\epsilon)/\log(2)$ compositions since $|P_M(f)|_\infty \leq C_0^{-M+1} L_0^{M+1} \leq 2^{-M+1}$, since $|f|_\infty \leq C_0^{-1} L_0$, where

$$L_0 = \max(1, \mathcal{R}, \mathcal{R}^2 C_w |w_3|_1 |w_4|_1, \mathcal{R}(1 + \mathcal{R}C(|g_0|, \mathcal{R}, T))).$$

And we take C_0 such that $L_0 \leq C_0/2$, which is always possible.

For any f_1, f_2, t_1 and t_2 we have

$$\begin{aligned}
 C_0^{-1} |\phi_i(f_1, t_1) - \phi_i(f_2, t_2)|_\infty &\leq C_0^{-1} |\phi_i(f_1, t_1) - \phi_i(f_2, t_1)|_\infty + C_0^{-1} |\phi_i(f_2, t_1) - \phi_i(f_2, t_2)|_\infty \\
 &\leq C_0^{-1} L_0 |f_1 - f_2|_\infty + C_0^{-1} L_0 C_\gamma |t_1 - t_2|^\gamma \\
 &\leq |f_1 - f_2|_\infty + |t_1 - t_2|^\gamma
 \end{aligned}$$

Where C_γ is a constant taking as the maximum of all the Hölder constant of each parameter function, depending also of the kernel. We take C_0 big enough to have $C_0^{-1} L_0 C_\gamma \leq 1/2$.

Lemma 3.47. *For any sequences t_j and \tilde{t}_j , $j \leq K$ we have $|P_K(f) - P_{\tilde{K}}(f)|_\infty \leq \sum_{j=1}^K |t_j - \tilde{t}_j|^\gamma$*

Proof. We do it by induction. For $K = 1$ we have $|C_0^{-1}\phi_i(f, t_1) - C_0^{-1}\phi_i(f, t_2)| \leq |t_1 - t_2|^\gamma$. Then we assume it is true for K .

$$\begin{aligned} |P_{K+1} - P_{\tilde{K}+1}|_\infty &= |\phi_i(P_K, t_{K+1}) - \phi_i(P_{\tilde{K}}, \tilde{t}_{K+1})|_\infty \\ &\leq |P_K - P_{\tilde{K}}|_\infty + |t_{K+1} - \tilde{t}_{K+1}|^\gamma \\ &\leq \sum_{j=1}^K |t_j - \tilde{t}_j|^\gamma + |t_{K+1} - \tilde{t}_{K+1}|^\gamma \end{aligned}$$

which ends the proof on the lemma. \square

Defining the distance d between two sequence of t_i as $d(t_i, \tilde{t}_i) = \sum_{j=1}^{\log(\epsilon)/\log(2)} |t_j - \tilde{t}_j|^\gamma$. As long as $d(t_i, \tilde{t}_i) \leq \epsilon$, we are sure that $|P_K(f) - P_{\tilde{K}}(f)|_\infty \leq \epsilon$.

So for each f we can define less than $5^{(-\log(\epsilon)/\log(2))} * T^{-\log(\epsilon)/(\log(2)\gamma)} \epsilon^{-\log(\epsilon)/(\log(2)\gamma)}$ compositions K_i to ensure there exist, for any composition K , a i_0 such that $|P_K(f) - P_{K_{i_0}}(f)|_\infty \leq \epsilon$. Indeed we have to choose a sequence in $[0, T]^{-\log(\epsilon)/\log(2)}$ such that $\sum_{j=1}^{-\log(\epsilon)/\log(2)} |t_j - \tilde{t}_j|^\gamma \leq \epsilon$. And each sequence leads us to $5^{(-\log(\epsilon)/\log(2))}$ composition possible.

To finish we just have to count the number of function $f \in \mathcal{F}_0$ we have to take at the beginning of each composition. But thanks to the assumption 3.7 we know that U is not constant only on a compact support for the age included in $[0, A]$. We also assume that \mathcal{X} is included in $[-X, X]$. So $\mathcal{N}(\mathcal{F}_0, |\cdot|_\infty, \epsilon) \lesssim (ATX)\epsilon^{-1/\gamma} + C_0/\epsilon$. C_0/ϵ comes from the fact we have all the constants $[0, C_0]$ in \mathcal{F} .

This leads us to the end of the proof. \square

8 Remaining proofs

8.1 Proof of the property 3.39

We have with

$$\begin{aligned} \tilde{m}_1(Z^N, i, s, x, l) &= p(s, a_i, x_i, l_i) \mathcal{R}k_b(s, a_i, x_i, l_i, x, l) \\ \tilde{m}_2(Z^N, i, s, x, l) &= (1 - p(s, a_i, x_i, l_i)) \mathcal{R}k_b(s, a_i, x_i, l_i) \\ \tilde{m}_4(Z^N, i, s, x, l) &= \mathcal{R}k_e(s, a_i, x_i, l_i, x, l) \\ \tilde{m}_5(N, s, a, x, l) &= Ne_i(s, a, x, l) \end{aligned}$$

$$\left\{ \begin{array}{l} \tilde{Z}_t^N = \tau_t \tilde{Z}_0^N \\ + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle N \tilde{Z}_s^N, \mathbf{1} \rangle\}} \delta_{t-s, x_i, l_i} \mathbf{1}_{\{0 \leq \vartheta \leq \tilde{m}_1\}} \\ + \delta_{t-s, x, l} \mathbf{1}_{\{\tilde{m}_1 \leq \vartheta \leq \tilde{m}_2\}} Q_b(ds, di, d\vartheta, dx, dl) \\ + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \mathbf{1}_{\{i \leq \langle N \tilde{Z}_s^N, \mathbf{1} \rangle\}} \delta_{a_i+t-s, x, l} \mathbf{1}_{\{0 \leq \vartheta \leq \tilde{m}_4\}} Q_e(ds, di, d\vartheta, dx, dl) \\ + N^{-1} \int_0^t \int_{\mathbb{R}_+ \times \mathcal{X} \times \mathcal{L}} \delta_{a+t-s, x, l} \mathbf{1}_{\{0 \leq \vartheta \leq \tilde{m}_5\}} Q_i(ds, da, dx, dl) \end{array} \right. \quad (3.28)$$

an adaptation of 3.2. Then we have by Ito's formula $\exp(\lambda\langle\tilde{Z}_t^N, \mathbf{1}\rangle)$ equals to

$$\begin{aligned} & \exp(\lambda\langle\tilde{Z}_0^N, \mathbf{1}\rangle) + \xi_t \\ & + N\mathcal{R}(e^{\frac{\lambda}{N}} - 1) \int_0^t \int_{\mathcal{U}} p(s, a, x, l) \exp(\lambda\langle\tilde{Z}_s^N, \mathbf{1}\rangle) \tilde{Z}_s^N(da, dx, dl) ds \\ & + N\mathcal{R}(e^{\frac{\lambda}{N}} - 1) \int_0^t \int_{\mathcal{U}} (1 - p(s, a, x, l)) \exp(\lambda\langle\tilde{Z}_s^N, \mathbf{1}\rangle) \tilde{Z}_s^N(da, dx, dl) ds \\ & + N\mathcal{R}(e^{\frac{\lambda}{N}} - 1) \int_0^t \int_{\mathcal{U}} \exp(\lambda\langle\tilde{Z}_s^N, \mathbf{1}\rangle) \tilde{Z}_s^N(da, dx, dl) ds \\ & + N(e^{\frac{\lambda}{N}} - 1) \int_0^t \int_{\mathcal{U}} e_i(s, a, x, l) \exp(\lambda\langle\tilde{Z}_s^N, \mathbf{1}\rangle) ds da P(dx) Q(dl) \end{aligned}$$

since

$$\int_{\mathcal{X} \times \mathcal{L}} k_e(s, a, x, l, x', l') P(dx') Q(dl') = \int_{\mathcal{X} \times \mathcal{L}} k_b(s, a, x, l, x', l') P(dx') Q(dl') = 1$$

and where $\mathbb{E}(\xi_t) = 0$. Introducing $f(t, \lambda) = \mathbb{E}[\exp(\lambda\langle\tilde{Z}_t^N, \mathbf{1}\rangle)]$, $I_s = \int_{\mathcal{U}} e_i(s, a, x, l) da P(dx) Q(dl)$ and $a_N(\lambda) = N(e^{\frac{\lambda}{N}} - 1)$, it follows that

$$f(t, \lambda) = f(0, \lambda) + 2\mathcal{R}a_N(\lambda) \int_0^t \partial_\lambda f(s, \lambda) ds + a_N(\lambda) \int_0^t I_s f(s, \lambda) ds$$

Now we solve this partial differential equation. First we write $\tilde{f}(x, y) = f(x + y, b_N(y))$ where $b_N(y) = -2\mathcal{R}a_N(y)$ and we have

$$\partial_y \tilde{f}(x, y) = a_N(b_N(y)) I_{x+y} \tilde{f}(x, y)$$

So $\tilde{f}(x, y) = K(x) \exp(\int_0^y a_N(b_N(s)) I_{x+s} ds)$. And now we can use the initial condition to get

$$K(x) = f_0(b_N(-x)) \exp\left(\int_{-x}^0 a_N(b_N(s)) I_{x+s} ds\right)$$

Now with μ_N^λ such that $b_N(\mu_N^\lambda) = \lambda$ we get

$$f(t, \lambda) = \tilde{f}(t - \mu_N^\lambda, \mu_N^\lambda) = f_0(b_N(\mu_N^\lambda - t)) \exp\left(\int_{\mu_N^\lambda - t}^{\mu_N^\lambda} a_N(b_N(s)) I_{t - \mu_N^\lambda + s} ds\right)$$

with $b_N(y) = N \ln\left(\frac{K_0 e^{2\mathcal{R}y}}{1 - K_0 + K_0 e^{2\mathcal{R}y}}\right)$, $a_N(b_N(y)) = \frac{K_0 - 1}{1 - K_0 + K_0 e^{2\mathcal{R}y}} N$ and $\mu_N^\lambda = \frac{1}{2\mathcal{R}} \ln\left(\frac{(1 - K_0)e^{\frac{\lambda}{N}}}{(1 - e^{\frac{\lambda}{N}})K_0}\right)$.

The solution of such a system, where $f(0, \lambda) = f_0(\lambda)$ is

$$f(t, \lambda) \leq f_0\left(N \ln\left(\frac{e^{-2\mathcal{R}t + \frac{\lambda}{N}}}{1 - e^{\frac{\lambda}{N}} [1 - e^{-2\mathcal{R}t}]}\right)\right) \exp\left(|e_i|_1 \left[\frac{e^{\frac{\lambda}{N}} - 1}{e^{\frac{\lambda}{N}} + 1} + \frac{e^{\frac{\lambda}{N}} - 1}{1 + e^{2\mathcal{R}t} e^{\frac{\lambda}{N}}}\right]\right)$$

Since $\left[\frac{e^{\frac{\lambda}{N}} - 1}{e^{\frac{\lambda}{N}} + 1} + \frac{e^{\frac{\lambda}{N}} - 1}{1 + e^{2\mathcal{R}t} e^{\frac{\lambda}{N}}}\right] \leq 2$ and $N_0/N \leq |g_0|_1 + C_0 \mathcal{W}_{1,1,1}^N(\mathcal{F})_0$, we have

$$f_0(\lambda) = \mathbb{E}[\lambda N_0/N] \leq e^{\lambda|g_0|_1} \exp(\max(\lambda, \lambda^p) C_0 q)$$

Since we will take $\lambda_N = N \log \left(\frac{1 - e^{-2\mathcal{R}T} e^{-1/N}}{1 - e^{-2\mathcal{R}T}} \right)$, we will have $N \ln \left(\frac{e^{-2\mathcal{R}t + \frac{\lambda_N}{N}}}{1 - e^{-\frac{\lambda_N}{N}} [1 - e^{-2\mathcal{R}t}]} \right) \leq 1$ if \mathcal{R} and N are great enough. So we can only use the case when $\max(\lambda^p, \lambda) = 1$ since $\lambda \leq 1$. And $f_0(\lambda) \leq e^{\tilde{q}\lambda}$, with $\tilde{q} \leq |g_0|_1 + q$, and we can conclude the proof as in the precedent chapter.

8.2 Proof of the property 3.21

We use two lemmas, the first one is the proposition of the precedent chapter, noting $\bar{\mu}(t, a) = \mu(t, a, gU(t, a))$ and assuming $\bar{\mu} \in \mathcal{H}^{l, m}$, we have

Lemma 3.48. *Work under Assumptions 3.1, 3.3, 3.4 and 3.20.*

(i) We have

$$g \in \mathcal{H}^{\min(\alpha, \beta, l+1, m), \min(\alpha, \beta, l+1, m)}(t, a) \text{ for } (t, a) \in \mathcal{D}_L$$

and

$$g \in \mathcal{H}^{\min(l+1, m), \max(m, \min(l, m+1))}(t, a) \text{ for } (t, a) \in \mathcal{D}_U.$$

(ii) The following improvement of the anisotropic smoothness holds when the parametrisation is given by \tilde{g} :

$$\tilde{g} \in \mathcal{H}^{\min(l+1, m+1), \min(\alpha, \beta, l+1, m)}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_L$$

and

$$\tilde{g} \in \mathcal{H}^{\min(l+1, m+1), \max(m, \min(l, m+1))}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_U.$$

The second lemma will let us do iteration to get the proposition.

Lemma 3.49. *Under the assumptions and $g \in \mathcal{H}^{\omega, \sigma}$ then $\bar{\mu} \in \mathcal{H}^{\min(\gamma, \omega, v), \min(\delta, \rho, v)}$*

Proof. $\mu \in \mathcal{H}^{\gamma, \delta, v}$, and $U \in \mathcal{H}^{\rho, \eta}$, so $f_U(t, a) = \int_0^\infty g(t, \alpha)U(a, \alpha)d\alpha \in \mathcal{H}^{\omega, \rho}$, if we assume that every derivative exists.

Two cases arise in the composition of function, the first one is when $v < 1$ and when $v \geq 1$. But we know that $v \geq 1$ according to the assumption 3.1. In this case the composition of function is just the minimum of regularity in each variable, which ends the proof. \square

We now begin with $\omega_0 = 0$ and $\sigma_0 = 0$, since we know that g is bounded. Taking now general ω_i and σ_i , from the lemma 3.49 we know that $\bar{\mu} \in \mathcal{H}^{\min(\gamma, \omega_i, v), \min(\delta, \rho, v)}$. From lemma 3.48 and $\bar{\mu} \in \mathcal{H}^{\min(\gamma, \omega_i, v), \min(\delta, \rho, v)}$ we get

$$g \in \mathcal{H}^{\min(\alpha, \beta, \min(\gamma, \omega_i, v)+1, \min(\delta, \rho, v)), \min(\alpha, \beta, \min(\gamma, \omega_i, v)+1, \min(\delta, \rho, v))}(t, a) \text{ for } (t, a) \in \mathcal{D}_L$$

and

$$g \in \mathcal{H}^{\min(\min(\gamma, \omega_i, v)+1, \min(\delta, \rho, v)), \max(\min(\delta, \rho, v), \min(\min(\gamma, \omega_i, v), \min(\delta, \rho, v)+1))}(t, a) \text{ for } (t, a) \in \mathcal{D}_U.$$

The following improvement of the anisotropic smoothness holds when the parametrisation is given by \tilde{g} :

$$\tilde{g} \in \mathcal{H}^{\min(\min(\gamma, \omega_i, v)+1, \min(\delta, \rho, v)+1), \min(\alpha, \beta, \min(\gamma, \omega_i, v)+1, \min(\delta, \rho, v))}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_L$$

and

$$\tilde{g} \in \mathcal{H}^{\min(\min(\gamma, \omega_i, v) + 1, \min(\delta, \rho, v) + 1), \max(\min(\delta, \rho, v), \min(\min(\gamma, \omega_i, v), \min(\delta, \rho, v) + 1))}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_U.$$

And we get the following iteration, since if we carefully do it one can see σ depends on the domain but ω has to be taken as the minimum in the regularity with respect to time,

$$\begin{aligned} \omega_{i+1} &= \min(\alpha, \beta, \min(\gamma, \omega_i, v) + 1, \min(\delta, \rho, v)) \wedge \min(\min(\gamma, \omega_i, v) + 1, \min(\delta, \rho, v)) \\ \sigma_{i+1} &= \min(\alpha, \beta, \min(\gamma, \omega_i, v) + 1, \min(\delta, \rho, v)) \text{ on } \mathcal{D}_L \end{aligned}$$

and

$$\sigma_{i+1} = \max(\min(\delta, \rho, v), \min(\min(\gamma, \omega_i, v), \min(\delta, \rho, v) + 1)) \text{ on } \mathcal{D}_U$$

And this is sufficient to conclude the proof in a formal way. We have

$$\begin{aligned} \omega &= \min(\alpha, \beta, \gamma + 1, v, \delta, \rho) \\ \sigma &= \min(\alpha, \beta, \gamma + 1, \omega + 1, v, \delta, \rho) = \min(\alpha, \beta, \gamma + 1, v, \delta, \rho) \text{ on } \mathcal{D}_L \end{aligned}$$

and

$$\sigma = \max(\min(\delta, \rho, v), \min(\alpha, \beta, \gamma, v, \delta, \rho)) = \min(\delta, \rho, v) \text{ on } \mathcal{D}_U$$

For \tilde{g} we get

$$\min(\min(\gamma, \omega, v) + 1, \min(\delta, \rho, v) + 1) = \min(\gamma + 1, \omega + 1, v + 1, \delta + 1, \rho + 1)$$

So on $\tilde{\mathcal{D}}_L$ we have

$$\min(\gamma + 1, \omega + 1, v + 1, \delta + 1, \rho + 1) = \min(\alpha, \beta, \gamma, \delta, v, \rho) + 1$$

and on $\tilde{\mathcal{D}}_U$ we have

$$\min(\gamma + 1, \omega + 1, v + 1, \delta + 1, \rho + 1) = \min(\alpha, \beta, \gamma, \delta, v, \rho) + 1$$

One last point is to ensure the lemma 3.49 is true. But since we are in Hölder space, we can ensure the existence of the derivative in a , since g is integrable. For the derivative in t we just have to think about the form of g and see we can bound any derivative of g by a integrable function, so in the worst case we could have only $f_U(t, a) \in \mathcal{H}^{\omega - \zeta, \rho}$ for any $\zeta < \omega$. But this does not change the result of the proof.

8.3 Proof of Theorem 3.25

The proof of the lower bound use the same argument than the precedent chapter, see subsection 7.3, coming from the usual argument for lower bound. Since we do only pointwise estimation we can rely on the test between two alternatives. The only difference comes from the interaction.

Lemma 3.50. *Let the model 1 be the one with parameter b_1, μ_1, U_1 and 2 the one with the parameters b_2, μ_2, U_2 , we get $\log\left(\frac{d\mathbb{P}_2}{d\mathbb{P}_1}\right)$ equals to*

$$\begin{aligned} & -N \int_0^t \int_0^\infty b_2(s, a) - b_1(s, a) + \mu_2(s, a, Z_s^N U(a)) - \mu_1(s, a, Z_s^N U(a)) Z_s^N(da) ds \\ & + \int_0^t \int_{\mathbb{N}^* \times [0, \infty)} \mathbf{1}_{i \leq \langle Z_{s-}^N, \mathbf{1} \rangle} \mathbf{1}_{0 \leq \vartheta < b_1(s, A_i(Z_{s-}^N))} \log\left(\frac{b_2(s, A_i(Z_{s-}^N))}{b_1(s, A_i(Z_{s-}^N))}\right) Q_b(ds, di, d\vartheta) \\ & + \int_0^t \int_{\mathbb{N}^* \times [0, \infty)} \mathbf{1}_{i \leq \langle Z_{s-}^N, \mathbf{1} \rangle} \mathbf{1}_{0 \leq \vartheta < \mu_1(s, A_i, Z_s^N U_1(A_i))} \log\left(\frac{\mu_2(s, A_i, Z_s^N U_2(A_i))}{\mu_1(s, A_i, Z_s^N U_1(A_i))}\right) Q_\mu(ds, di, d\vartheta) \end{aligned}$$

From this lemma we can easily compute a kullback leibler divergence. We assume that $b_1 = b_2$, then $D(\mathbb{P}_2 || \mathbb{P}_1)$ equals to

$$\begin{aligned} & -\mathbb{E}_{\mathbb{P}_2} \left[-\log\left(\frac{d\mathbb{P}_1}{d\mathbb{P}_2}\right) \right] \\ & N\mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \left\langle \mu_1(s, a, Z_s^N U_1(a)) - \mu_2(s, a, Z_s^N U_2(a)) - \mu_2 \log\left(\frac{\mu_1(s, a, Z_s^N U_1(a))}{\mu_2(s, a, Z_s^N U_2(a))}\right), Z_s^N \right\rangle ds \right] \\ & N\mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \left\langle \mu_2((\mu_1 - \mu_2)/\mu_2 - \log((\mu_1 - \mu_2)/\mu_2 + 1)), Z_s^N \right\rangle ds \right] \end{aligned}$$

We get with $\varphi(x) = x - \log(1+x) \leq x^2$ and the fact μ_2 is bounded under,

$$D(\mathbb{P}_2 || \mathbb{P}_1) \lesssim N\mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \int_0^\infty (\mu_1(s, a, Z_s^N U_1(a)) - \mu_2(s, a, Z_s^N U_2(a)))^2 Z_s^N(da) ds \right]$$

We now note $\alpha(s, a) = \mathbf{m}_1(s, a) - \mathbf{m}_2(s, a, u) = \psi_1(s)\psi_2(s-a)$. We take $\mathfrak{k}(U_1(a, \alpha) - U_2(a, \alpha)) = \psi_U(a, \alpha)$.

$$\begin{aligned} & \lesssim N\mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \int_0^\infty (\mathbf{m}_1(s, a) - \mathbf{m}_2(s, a))^2 \tilde{Z}_s^N(da) ds \right] + N\mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \int_0^\infty (Z_s^N \psi_U(a))^2 \tilde{Z}_s^N(da) ds \right] \\ & \lesssim N \int_0^T \int_0^\infty \psi_1^2(s)\psi_2^2(s-a) \mathfrak{g}(s, a) dad s + \mathbb{E}_{\mathbb{P}_2} \left[\widetilde{W}_{\psi_1^2, \psi_2^2}^N(\mathcal{F})_T \right] \\ & \quad + N|\psi_U|_2^2 + NC_{\psi_U}^2 \mathbb{E}_{\mathbb{P}_2} \left[\int_0^T \int_0^\infty \mathcal{W}_1^N(\tilde{\mathcal{F}})_s^2 \tilde{Z}_s^N(da) ds | \mathcal{B}_N \right] + Ne^{-\sqrt{N}} \end{aligned}$$

Where $\tilde{\mathcal{F}}$ is generated as \mathcal{F} with the only difference we have ψ_U in place of U .

Now we assume that $|\psi_2|_\infty^2 \leq N^{1/2}|\psi_2|_2^2$ and we have

$$D(\mathbb{P}_2 || \mathbb{P}_1) \lesssim N|\psi_1|_2^2 |\psi_2|_2^2 + N^{1/2}|\psi_1|_\infty |\psi_2|_\infty |\psi_1|_2 |\psi_2|_2 + N|\psi_U|_2^2 + C_{\psi_U}^2 + Ne^{-\sqrt{N}}$$

C_{ψ_U} is the maximum between the Hölder constant and the infinite norm. One can show that if $\psi_U(a, \alpha) = cN^{-1/2} \sqrt{\gamma_U} K(\gamma_U a, \alpha)$, then the Hölder constant is less than $cN^{-1/2} \gamma_U^{1/2+\rho}$, which is less a constant c by construction, and we have $C_{\psi_U}^2 \leq c^2 N^{-1} \max(\gamma_U, \gamma_U^{1+2\rho}) = c^2 N^{-1} \gamma_U^{1+2\rho}$ and which is less than c^2 by construction.

We begin with a lemma very important for the following. We continue to assume that $\mathbf{m}_1 - \mathbf{m}_2 = \psi(s, u)$ and $U_1 - U_2 = \psi_U$.

Lemma 3.51. *Let g_1 and g_2 be the solution of the equation (3.3) with $g_0, b, \mathbf{m}_1, U_1$ and $g_0, b, \mathbf{m}_1, U_2$ respectively. Then, we have with $\mathfrak{C} = 4\mathcal{R}T|g_0|_1 e^{2\mathcal{R}T} + |g_0|_1$*

(i) *In the case $\psi(s, u) = \psi(s - u)$, note $a_N = |g_0|_\infty |\psi|_1 + \max(|g_0|_\infty, \mathcal{R}|g_0|_1 e^{\mathcal{R}T}) |\psi_U|_1$, then*

$$\int_0^\infty |g_1(t, \alpha) - g_2(t, \alpha)| d\alpha \leq \mathfrak{C} a_N e^{\mathcal{R}T}$$

(ii) *In the case $\psi(s, u) = \psi_1(s)\psi_2(u)$, note $a_N = |\psi_1|_2 |\psi_2|_2 + \max(|g_0|_\infty, \mathcal{R}|g_0|_1 e^{\mathcal{R}T}) |\psi_U|_1$, then*

$$\int_0^\infty |g_1(t, \alpha) - g_2(t, \alpha)| d\alpha \leq \mathfrak{C} a_N e^{\mathcal{R}T}$$

Proof. We use the fact that g can be written explicitly in function of all the parameters. We also note $\int_0^\infty |g_1(t, \alpha) - g_2(t, \alpha)| d\alpha = Y_t^N$. We begin by the second point where $\psi(s, u) = \psi_1(s)\psi_2(u)$. From that we get on \mathcal{D}_U that $|g_1(t, a) - g_2(t, a)|$ is less than

$$\begin{aligned} &\leq g_0(a-t) \left[\int_0^t \mathbf{m}_1(s, s+a-t) - \mathbf{m}_2(s, s+a-t) ds \right. \\ &\quad \left. + \int_0^t \int_0^\infty g_1(t, \alpha) U_1(s+a-t, \alpha) - g_2(t, \alpha) U_2(s+a-t, \alpha) d\alpha ds \right] \\ &\leq g_0(a-t) \left[|\psi_1|_2 |\psi_2|_2 + \mathcal{R} \int_0^t Y_s^N ds + \int_0^t \max(|g_0|_\infty, \mathcal{R}|g_0|_1 e^{\mathcal{R}T}) |\psi_U(s+a-t, \alpha)| d\alpha ds \right] \\ &\leq g_0(a-t) \left[|\psi_1|_2 |\psi_2|_2 + \mathcal{R} \int_0^t Y_s^N ds + \max(|g_0|_\infty, \mathcal{R}|g_0|_1 e^{\mathcal{R}T}) |\psi_U|_1 \right] \\ &\leq g_0(a-t) \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right] \end{aligned}$$

And we get by integrating from t to ∞ $\int_t^\infty |g_1(t, a) - g_2(t, a)| da \leq |g_0|_1 \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right]$.

Now we work on \mathcal{D}_L . One can see the subsection 7.3 of the precedent chapter to get a better definition of what we write $B_1 = B_{b, \mu_1, g_0}$ and $B_2 = B_{b, \mu_2, g_0}$ here. This is not difficult to show, following carefully the proof of the precedent chapter that

$$\max(|B_1|_\infty, |B_2|_\infty) \leq \mathcal{R}|g_0|_1 e^{\mathcal{R}T}$$

and

$$\begin{aligned} |B_1(t) - B_2(t)| &\leq \left(\mathcal{R}|g_0|_1 \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right] + \mathcal{R}|g_0|_1 e^{\mathcal{R}T} \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right] \right) e^{\mathcal{R}T} \\ |B_1(t) - B_2(t)| &\leq 2\mathcal{R}|g_0|_1 e^{2\mathcal{R}T} \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right] \end{aligned}$$

by a similar argument than in the precedent chapter we finally have

$$\int_0^t |g_1(t, a) - g_2(t, a)| da \leq 2\mathcal{R}T|g_0|_1 e^{2\mathcal{R}T} \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right] + \mathcal{R}T|g_0|_1 e^{\mathcal{R}T} \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right]$$

$$\int_0^t |g_1(t, a) - g_2(t, a)| da \leq 4\mathcal{R}T|g_0|_1 e^{2\mathcal{R}T} \left[a_N + \mathcal{R} \int_0^t Y_s^N ds \right]$$

Finally we get, noting $\mathfrak{C} = 4\mathcal{R}T|g_0|_1 e^{2\mathcal{R}T} + |g_0|_1$,

$$Y_t^N \leq \mathfrak{C} a_N e^{\mathcal{R}t}$$

For the first term where $\psi(s, u) = \psi(s - u)$, we just have to be careful in $\int_0^t \mathbf{m}_1(s, s + a - t) - \mathbf{m}_2(s, s + a - t) ds = t\psi(t - a)$, and $\int_{t-a}^t \mathbf{m}_1(s, s - a + t) - \mathbf{m}_2(s, s - a + t) ds = a\psi(t - a)$.

Following the same proof than for the second term one can show that $|g_0|_1 |\psi_1|_2 |\psi_2|_2$ becomes $|g_0|_\infty |\psi|_1$. \square

Now we can work with specific function ψ and ψ_U to prove the lower bound for g . We take for a t_0, a_0 ,

$$\mathbf{m}_1 - \mathbf{m}_2 = N^{-1/2} \gamma_1^{1/2} K_1(\gamma_1(s - u - t_0 + a_0))$$

and $U_1(a, \alpha) - U_2(a, \alpha) = N^{-1/2} \gamma_U^{1/2} K_U(\gamma_U(a - a_0)) K(\alpha)$.

The Hölder condition implies that $\gamma_1 = N^{\frac{1}{2 \max\{\gamma, \delta+1\}}}$ and $\gamma_U = N^{\frac{1}{2\rho+1}}$. Following the proof of the precedent chapter and since Y_t^N is negligible with respect to $N^{-1/2} \gamma_1^{1/2}$ and $N^{-1/2} \gamma_U^{1/2}$ we get the minimum of the two regularity.

Now we can work with specific function ψ_1, ψ_2 and ψ_U to prove the lower bound for μ . We take for a t_0, a_0 ,

$$\mathbf{m}_1 - \mathbf{m}_2 = N^{-1/2} \gamma_1^{1/2} \gamma_2^{1/2} K_1(\gamma_1^N(s - t_0)) K_2(\gamma_2(u - a_0))$$

and $U_1(a, \alpha) - U_2(a, \alpha) = N^{-1/2} \gamma_U^{1/2} K_U(\gamma_U(a - a_0)) K(\alpha)$.

$$|\mu_1(t_0, a_0, (g_1)_{t_0} U_1(a_0)) - \mu_2(t_0, a_0, (g_2)_{t_0} U_2(a_0))| \geq |\psi_1(0) \psi_2(0) + N^{-1/2} \gamma_U^{1/2} K_U(0) \tilde{C}| - |Y_{t_0}^N \mathcal{R}|$$

We have from the Hölder condition, with $s = s(\gamma, \delta) = (\gamma^{-1} + \delta^{-1})^{-1}$ that $\gamma_U = N^{\frac{1}{2\rho+1}}$, $\gamma_1 = N^{\frac{\delta(\gamma+\delta)-1}{2s+1}}$ and $\gamma_2 = N^{\frac{\gamma(\gamma+\delta)-1}{2s+1}}$. So, with $s^- = \min(s, \rho)$,

$$|\psi_1(0) \psi_2(0) + N^{-1/2} \gamma_U^{1/2} K_U(0) \tilde{C}| \geq N^{\frac{-s^-}{2s^-+1}}$$

And we have $Y_{t_0}^N \leq \mathfrak{C} a_N e^{\mathcal{R}t}$ with $a_N = |\psi_1|_2 |\psi_2|_2 + \max(|g_0|_\infty, \mathcal{R}|g_0|_1 e^{\mathcal{R}T}) |\psi_U|_1$, and this is easy to see that $a_N N^{\frac{-s^-}{2s^-+1}} \rightarrow 0$, so for N great enough we have

$$|\mu_1(t_0, a_0, (g_1)_{t_0} U_1(a_0)) - \mu_2(t_0, a_0, (g_2)_{t_0} U_2(a_0))| \geq N^{\frac{-s^-}{2s^-+1}}$$

The end of the proof follow the usual argument, since we control the Kullback Leibler divergence in a good way, see the proposition 2.32.

CHAPTER 4

NONPARAMETRIC TEST OF TIME DEPENDANCE OF AGE-STRUCTURED MODELS IN A LARGE POPULATION LIMIT

1 Introduction

1.1 Motivation

In the precedent chapter we infer the mortality rate in presence of interactions. This is a question of interest in biology where the models with interactions come often. It can for example modeled the pressure coming from the lack of food in some environment.

We will use the nonparametric testing framework as introduced in Ingster [37, 38]. In our case we want to test if the mortality rate depends on time. This kind of testing has been studied. In [45, 46], the point is to test if the functions belongs to a parametric set of functions, but this can be generalized to any subset of the functionnal space the function originally belongs. This is also a topic studied in [13] with Hölder balls.

However, to our knowledge this problematic have not been done in the framework we work with during this thesis.

1.2 Setting

Remind the model and the data. We have :

- Z_0^N the initial population such that $NZ_0^N \in \mathcal{M}_F$
- Two functions

$$\begin{aligned} b &: [0, T] \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \\ \mu &: [0, T] \times \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}_+ \end{aligned}$$

The fertility rate $b(s, a)$ accounts for the births produced by the individual having age a at time s . The mortality rate $\mu(s, a, u)$ accounts for the death of an individual having age a at time s and pressure u coming from the interaction.

- A kernel

$$U : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$$

This interaction kernel bring a pressure of the form $\int_0^\infty g(s, \alpha)U(a, \alpha)d\alpha$ at each time s and age a .

1.3 Model

We refer to [62] to get a proper definition of the microscopic model we will use here.

In order to get a statistical test, we need to use a particular model. If we do no other assumptions it would be impossible to detect the presence of interaction. We do not know the minimal assumption to assume we can detect interaction. However we can try to give some intuition about it. Let's take a general function $\mu(s, a, u)$. We are able to find an estimator of

$$\mu \left(s, a, \int_0^\infty g(s, \alpha)U(a, \alpha)d\alpha \right) = \bar{\mu}(s, a)$$

We see that μ has interactions only if μ depends on g . So one can try to show the dependence between $\bar{\mu}$ and g . However if we do that directly we won't have any result. Indeed $\bar{\mu}$ even without interaction is linked with g , by the simple fact that we cannot estimate μ if there is noone. The problem is very difficult and impossible to solve without a further assumption. The time variation of μ can compensate the interaction term. We need an assumption which control this time dependence of μ .

To do so we have chosen a simple assumption where we assume we know the time dependence and this dependence has a particular structure. However we don't know any other way to proceed with the observations we have.

Assumption 4.1. *There exist a known function $\beta > 0$ such that*

$$\mu(s, a, u) = \beta(s)\mu_0(a, u)$$

In this setting if $\beta^{-1}(s)\bar{\mu}(s, a) = \mu_0(a, \int_0^\infty g(s, \alpha)U(a, \alpha)d\alpha)$ depends on time, then we have a interaction. Indeed the time dependence of $\mu_0(a, \int_0^\infty g(s, \alpha)U(a, \alpha)d\alpha)$ comes, only, from the interaction part. This gives us a way to detect the presence of interaction. We need to detect the time dependence of the death rate.

2 Non parametric test

2.1 General setting

We will work with $\bar{\gamma} = \min(\alpha, \beta, \rho, \delta)$ and $\bar{\delta} = \min(\alpha, \beta, \delta, \rho)$ which are the effective regularity of μ_0 in our setting under the assumption 3.20 of the precedent chapter. We take the worst regularity, one can check that with those regularity we have $g \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}$. Since we take the worst regularity on \mathcal{D}_L and \mathcal{D}_U we have $\bar{\gamma} = \bar{\delta}$. However we will keep two notation even if those constants are equal. We want to highlight the bidimensionnal part of the problem.

We define for any function $f \in \mathcal{F} = \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)$, with g the population density solution of (3.3), without immigration and characteristic,

$$d_{\mathcal{F}}(f) = \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty \varphi(s, a) [f(s, a) - m_f(a)] g(s, a) da ds \quad (4.1)$$

where $m_f(a) = \int_0^T g(s, a) f(s, a) ds / \int_0^T g(s, a) ds$.

It is easy to see that

$$d_{\mathcal{F}}(f) = \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty [\varphi(s, a) - m_\varphi(a)] f(s, a) g(s, a) da ds$$

From now, we note

$$|f|_{2,g}^2 = \int_0^T \int_0^\infty f^2(s, a) g(s, a) da ds$$

and

$$\mathcal{F}(\rho_N) = \{f \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L), d_{\mathcal{F}}(f) \geq \rho_N\}$$

We implicitly normalize the function φ such that their norm $|\varphi|_{2,g} = 1$ in the definition of $d_{\mathcal{F}}$. This is just technical to avoid some constant during the calculus. Now we can define our non parametric test, we will test $H_0 : \mu \in \mathcal{H}^{\bar{\delta}}(L)$ against the alternative $H_1 : \mu \in \mathcal{F}(\rho_N)$. Here $\mathcal{H}^{\bar{\delta}}(L)$ est l'espace de Hölder auquel appartient μ si ce taux ne dépend que de l'âge.

Our aim is to find ρ_N , the minimax rate of the test. We will also define the candidate for the test statistic which will be

$$d_{\mathcal{F}}^N = \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty [\varphi(s, a) - m_\varphi^N(a)] \beta^{-1}(s) \Gamma^N(da, ds)$$

where $m_\varphi^N(a) = \mathfrak{V}^N(a) \int_0^T \varphi(s, a) \hat{g}_*^N(s, a) ds$ and Γ^N is the point process associated with the deaths. $\mathfrak{V}^N(a)$ is just $1 / \int_0^T \int_0^\infty \mathbf{1}_{u=a} Z_s^N(du) ds$. We remark that m^N is always defined since we integrate against Γ^N and we evaluate in the time of death of an individual, so we have at least one individual at each time and $\int_0^T \int_0^\infty \mathbf{1}_{u=a} Z_s^N(du) ds \geq 1/N$. We will work with the following assumption in order to ensure a control of \mathfrak{V} .

Assumption 4.2. $g_0 = 0$ for $a \geq A_{\max}$.

Remark 4.3.

0) The assumption 4.2 is there to ensure the compactity of the set where we work. With this assumption we have $g \neq 0$ only on $[0, T] \times [0, A_{\max} + T]$. It can be release. This will be done in a future work.

1) From now, we stop to write β . One can easily check than everything will be the same with $\beta = 1$. Indeed according to the proof of Theorem 2.19 we have $\Gamma^N(da, ds) \approx \bar{\mu}(s, a) g(s, a) da ds$, so

$$\begin{aligned} \beta^{-1}(s) \Gamma^N(da, ds) &\approx \beta^{-1}(s) \bar{\mu}(s, a) g(s, a) \\ &\approx \mu_0 \left(a, \int_0^\infty g(s, \alpha) U(a, \alpha) d\alpha \right) g(s, a) \end{aligned}$$

2) If $\mu \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}$ then $\frac{\mu(s,a) - m_\mu(a)}{|\mu - m_\mu|_{2,g}} \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}$. Since we have $m_\mu \in \mathcal{H}^{\bar{\delta}}$ according to the proposition 3.21. So we just have

$$d_{\mathcal{H}^{\bar{\gamma}, \bar{\delta}}}(\mu) \approx |\mu - m_\mu|_{2,g} = d(\mu)$$

and $d(\mu) = 0$ if and only if $\mu(s, a) = m_\mu(a)$ where $g(s, a) \neq 0$. There is only equality up to a constant since the Hölder constant can change.

3) From now, and till the end of the chapter, we will assume we work implicitly under the assumptions 3.1, 3.3, 3.4, 3.20, 3.5, 3.7, 3.24 and 3.18. With our choice of \mathcal{F} we can ensure that $d(\mu) = 0$ is equivalent to μ does not depend on time, where $g \neq 0$. We can also be sure the assumption 3.5 is satisfied. The assumption 3.24 is satisfied since this is our setting. Also to be sure that $e(\mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)) < \infty$ we take $\bar{\delta} \geq 2$

4) Since we work under the assumption 3.18 we don't need to take care about the definition of m_f and m_f^N since g is bounded below. To be rigorous we should carry the lower bound of g into the definition of the test. However, in application we do not use the theoretical definition.

2.2 Upper bound

For a test ϕ_N we can define the global error as the sum of the first and second kind error.

$$R(C, \phi_N, \rho_N) = \sup_{\mu \in \mathcal{H}^\beta} \mathbb{P}_\mu(\phi_N = 1) + \sup_{\mu \in \mathcal{F}(C\rho_N)} \mathbb{P}_\mu(\phi_N = 0)$$

where \mathbb{P}_μ stands for the probability when we have the mortality rate μ in the model.

Taking $0 < \alpha < 1$ the upper bound consists to find a test ϕ_N^* and a constant C_* such that

$$\limsup_{N \rightarrow \infty} R(C, \phi_N^*, \rho_N^*) \leq \alpha$$

for all $C \geq C_*$ and $\rho_N \geq \rho_N^*$ where ρ_N^* will be the minimax rate for the test.

Theorem 4.4. *There exists a constant C_* depending only on \mathcal{R} , C_0 , $|g_0|_1$ and T such that with $\phi_N^* = \mathbf{1}_{d_{\mathcal{F}} \geq \frac{C_*}{2} \rho_N}$, we have*

$$\limsup_{N \rightarrow \infty} R(C, \phi_N^*, \rho_N^*) = 0$$

for all $C \geq C_*$ and $\rho_N \geq \rho_N^* = C_K \log(N) N^{-\frac{\min(\alpha, \beta, \delta, \rho)}{2 \min(\alpha, \beta, \delta, \rho) + 1}}$

In this theorem, the constant \mathcal{R} , C_0 , $|g_0|_1$ and T comes from the assumption of the precedent chapter. The constant C_K comes from the choice of the kernel K which is done to estimate the population density g .

2.3 Lower bound

Taking $0 < \alpha < 1$ the lower bound consist to find a constant C^* such that

$$\liminf_{N \rightarrow \infty} \inf_{\phi_N} R(C, \phi_N, \rho_N) \geq \alpha$$

for all $C \leq C^*$ and $\rho_N \leq \rho_N^*$ where ρ_N^* will be the minimax rate for the test.

One typical way to get this kind of bound is to consider several alternative, here $\mu_i \in \mathcal{F}(C\rho_N)$ for $1 \leq i \leq M$ against one $\mu_0 \in \mathcal{H}^\beta$. Then we have

$$R(C, \phi_N, \rho_N) \geq \mathbb{P}_{\mu_0}(\phi_N = 1) + \frac{1}{M} \sum_{i=1}^M \mathbb{P}_{\mu_i}(\phi_N = 0)$$

and noting $\mathbb{P}_{\mu_k} = \mathbb{P}_k$, $\xi_k = \frac{d\mathbb{P}_k}{d\mathbb{P}_0}|_T$ and $Z = \frac{1}{M} \sum_{i=1}^M \xi_k$, we get

$$R(C, \phi_N, \rho_N) \geq (1 - \delta) [1 - \mathbb{P}_0(|Z - 1| \leq \delta)]$$

for all $\delta > 0$. To prove the lower bound we need to prove that $\mathbb{P}_0(|Z - 1| \leq \delta) \xrightarrow{N \rightarrow \infty} 0$ which is possible only when M depends on N and if we control nicely Z .

Remark 4.5. *We can expect from the precedent chapter and what we do in the upper bound for the test than the lower bound will be the one of the estimation of the population density. So we expect to have a lower bound of order $N^{-\frac{\min(\delta, \rho)}{2 \min(\delta, \rho) + 1}}$.*

3 Discussion and numerical illustration

3.1 Implementation

Write (τ_i, a_i) the support of $\Gamma^N(ds, da)$. Remind $\Gamma^N(ds, da)$ is the point measure of death, so each (τ_i, a_i) is the time and age of one death. We assume there are D_T^N deaths between 0 and T . So we can write

$$\Gamma^N(ds, da) = \sum_{i=1}^{D_T^N} \delta_{\tau_i, a_i}(ds, da)$$

Then we have for all φ

$$\int_0^T \int_0^\infty [\varphi(s, a) - m_\varphi^N(a)] \Gamma^N(ds, da) = \frac{1}{N} \sum_{k=1}^{D_T^N} [\varphi(\tau_k, a_k) - m_\varphi^N(a_k)]$$

Now we will simply approximate the integral with a Riemann sum. Let $(t_i)_{i \leq R_T^N}$ be the subdivision of $[0, T]$ such that each time of death is in the subdivision and $\sup_i |t_{i+1} - t_i| \leq L_N$. This implies that $R_T^N \geq \frac{T}{L_N}$. With this subdivision we get

$$m_\varphi^N(a_k) = \mathfrak{W}_k^N \int_0^T \hat{g}_\star^N(s, a_k) \varphi(s, a_k) ds \approx \mathfrak{W}_k^N \sum_{i=1}^{R_T^N} (t_{i+1} - t_i) \varphi(t_i, a_k) \hat{g}_\star^N(t_i, a_k)$$

Where $\mathfrak{W}_k^N = \mathfrak{W}^N(a_k)$, which is fully computable since

$$\mathfrak{W}^N(a_k) = \int_0^T \int_0^\infty \mathbf{1}_{u=a} Z_s^N(du) ds$$

and $\int_0^T \int_0^\infty \mathbf{1}_{u=a} Z_s^N(du) ds$ is just the number of individuals getting the age a between 0 and T divided by N . Noting $\theta_{i,k}^N = \hat{g}_\star^N(t_i, a_k)$ and $\varphi_{i,k} = \varphi(t_i, a_k)$ we get

$$\int_0^T \int_0^\infty [\varphi(s, a) - m_\varphi^N(a)] \Gamma^N(ds, da) \approx \frac{1}{N} \sum_{k=1}^{D_T^N} [\varphi(\tau_k, a_k) - \mathfrak{W}_k^N \sum_{i=1}^{R_T^N} (t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N]$$

And now since $\{\tau_k\} \subset \{t_i\}$ we have

$$\frac{1}{N} \sum_{k=1}^{D_T^N} [\varphi(\tau_k, a_k) - \mathfrak{W}_k^N \sum_{i=1}^{R_T^N} (t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N] = \frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \varphi_{i,k} [\mathbf{1}_{t_i=\tau_k} - \mathfrak{W}_k^N (t_{i+1} - t_i) \theta_{i,k}^N]$$

We note $w_{i,k}^N = [\mathbf{1}_{t_i=\tau_k} - \mathfrak{A}_k^N(t_{i+1} - t_i)\theta_{i,k}^N] \frac{1}{N}$ and w^N the associated vector. We also define $D^N \mathcal{F} = \{(\varphi_{i,k})_{i \leq R_T^N, k \leq D_T^N} \text{ s.t. } \varphi \in \mathcal{F}\}$ we have

$$d_{\mathcal{F}}^N \approx \sup_{\varphi \in D^N \mathcal{F}} \langle \varphi, w^N \rangle = c_{\mathcal{F}}^N$$

The following property gives a true bound to the precedent approximation.

Proposition 4.6. *If $h_N = \min(h_N(t_i, a_k))$, where $h_N(t_i, a_k)$ is the bandwidth obtained in the Goldenschluger Lepski method for the point t_i, a_k , we have*

$$|d_{\mathcal{F}}^N - c_{\mathcal{F}}^N| \lesssim N^{-1} \tilde{n}_T [N^{-1} \tilde{n}_T h_N^{-2} L_N^2 + L_N^{1+\delta} h_N^{-1}]$$

where \tilde{n}_T is a upper bound of the number of individuals between 0 and T in the population.

Corollary 4.7. *If $L_N = \max(h_N(\rho_N^*)^{1/2}, (h_N \rho_N^*)^{1/(1+\delta)}) = (h_N \rho_N^*)^{1/(1+\delta)}$ then*

$$|d_{\mathcal{F}}^N - c_{\mathcal{F}}^N| \lesssim \frac{\tilde{n}_T^2}{N^2} \rho_N^*$$

Remark 4.8. *From the corollary 4.7, we see that $c_{\mathcal{F}}^N$ can converge to $d_{\mathcal{F}}$ as fast as $d_{\mathcal{F}}^N$. The typical case is where $\rho_N^* \approx N^{-2/5}$, which means we take a kernel of order 2 and the regularity are all greater than 2, see the remark 4.3. We have with $L_N = N^{-1/5}$ and $R_T^N \geq TN^{1/5}$ the same speed of convergence for the two quantities. So in our practical implementation with $N = 1000$ and $T = 20$ we need $R_T^N \geq 80$ in all our simulations. This is done simply by considering all the death event since $D_T^N \geq 80$ in our simulations. In fact R_T^N will always be of the order of N since it has to be bigger than the number of deaths.*

We know now that without considering the restriction $\varphi \in D^N \mathcal{F}$, $\sup \langle \varphi, w^N \rangle$ is obtained for $\varphi = \alpha^N w^N$ with some α^N . Moreover we know that, asymptotically, the maximum of $\sup \langle \varphi, w^N \rangle$ exists. It is when $\varphi = \mu$. Indeed $c_{\mathcal{F}}^N$ converges to $|\mu - m_{\mu}|_{2,g}$ by proposition 4.6. w^N converges to 0, since

$$\begin{aligned} |w^N|_2^2 &= \frac{1}{N^2} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} [\mathbf{1}_{t_i=\tau_k} - \mathfrak{A}_k^N(t_{i+1} - t_i)\theta_{i,k}^N]^2 \\ &= \frac{1}{N^2} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \mathbf{1}_{t_i=\tau_k} - 2\mathbf{1}_{t_i=\tau_k} \mathfrak{A}_k^N(t_{i+1} - t_i)\theta_{i,k}^N + (\mathfrak{A}_k^N(t_{i+1} - t_i)\theta_{i,k}^N)^2. \end{aligned}$$

Our only way to get something computable is so to calculate the estimate of μ by the preceding chapter. Since the minimax speed of convergence, as shown in the theorem 3.27, of this estimator is below the one of $c_{\mathcal{F}}^N$, this is not the optimal way. Let $\mu_{i,k}^N$ be the estimate of μ taken in each point t_i, a_k , we get the following quantity for our test

$$\frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \mu_{i,k}^N [\mathbf{1}_{t_i=\tau_k} - \mathfrak{A}_k^N(t_{i+1} - t_i)\theta_{i,k}^N]$$

Remark 4.9.

1) *According to the proposition of this chapter, there exist a more efficient way to do the practical test. However we don't know yet how to do it without computing the estimate of μ which is computationally very expensive.*

2) *We can compute only for $t_i = \tau_k$ since asymptotically, and even for our example with $N=1000$, we have $D_T^N \geq T/L_N$. This limits the number of estimation we have to do.*

3.2 Numerical result

We have done four tests. In order to compute $C^* \rho_N^*$ we have taken the maximum of all the variance obtained in the Goldenschluger Lespki algorithm. This is the best candidate we can find without further computation. We test in those four example if we find the time dependence of the death rate. As said in the introduction of this chapter, we always consider that $\beta(s)$ is identical to 1.

The simulation 1 has be done with the parameters:

- (i) g_0 is the density of a gaussian random variable centered in 40 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t, a) = \mathbf{1}_{20 \leq a \leq 40}$. Such a birth rate is not Hölder, however we can prove similar result with such piecewise constant function.
- (iii) $\mu(t, a) = 0,004 \exp(0,0074a) \exp(-0,005t)$.

The simulation 2 has be done with the parameters:

- (i) g_0 is the density of a gaussian random variable centered in 40 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t, a) = \mathbf{1}_{20 \leq a \leq 40}$. Such a birth rate is not Hölder, however we can prove similar result with such piecewise constant function.
- (iii) $\mu(t, a) = 0,004 \exp(0,0074a)(1 + 5t)$. We take a dependence in time since our test is here to know if μ depends on time.

The simulation 3 has be done with the parameters:

- (i) g_0 is the sum of a density of a gaussian random variable centered in 20 with a variance of 15^2 conditioned to be between 0 and 120 and a density of a gaussian random variable centered in 70 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t, a) = 0,03 \exp\left(-\frac{(30-a)^2}{2*20^2}\right) \exp\left(-\frac{(5-t)^2}{2*2^2}\right)$.
- (iii) $\mu(t, a) = 0,004 \exp(0,0074a) \exp(-0,005t)$.

The simulation 4 has be done with the parameters:

- (i) g_0 is the sum of a density of a gaussian random variable centered in 20 with a variance of 15^2 conditioned to be between 0 and 120 and a density of a gaussian random variable centered in 70 with a variance of 15^2 conditioned to be between 0 and 120.
- (ii) $b(t, a) = 0,03 \exp\left(-\frac{(30-a)^2}{2*20^2}\right) \exp\left(-\frac{(5-t)^2}{2*2^2}\right)$.
- (iii) $\mu(t, a) = 0,004 \exp(0,0074a) \exp(-0,005t) + 0,01 \int_0^\infty g(t, \alpha) d\alpha$.

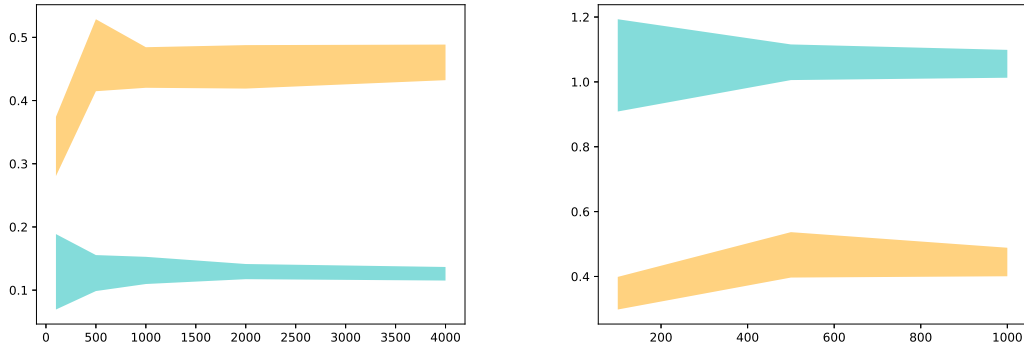


Figure 4.1: Variance in yellow and test statistic in blue. N in absciss. 95% dispersion realized on 50 simulations. Left: simulation 1. Right: simulation 2.

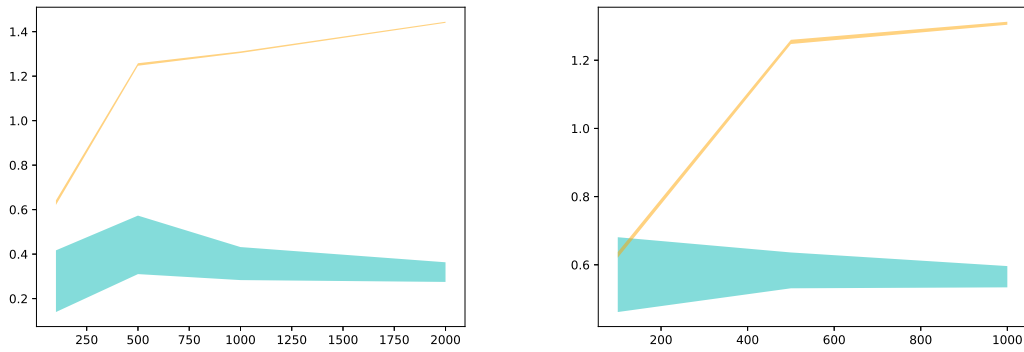


Figure 4.2: Variance in yellow and test statistic in blue. N in absciss. 95% dispersion realized on 50 simulations. Left: simulation 3. Right: simulation 4.

The results in the figure 4.1, are the one we expect. For the figure 4.2, we have a problem about the simulation 4. However we see the test statistic, compare to the one in simulation 3 is greater, around 0.6 for the simulation 4, to compare to 0.4 for the simulation 3. Also the variance is greater in simulation 3 compared to the test statistic than in the simulation 4. So we think the only problem here is the calibration of the constant for the test. So with a new calibration for the variance we get the figure 4.3. This calibration is simply the constant which multiply the maximum of the goldenchluger lepsi variance we obtained. Between the figure 4.2 and 4.3 we just divide this constant by 3.

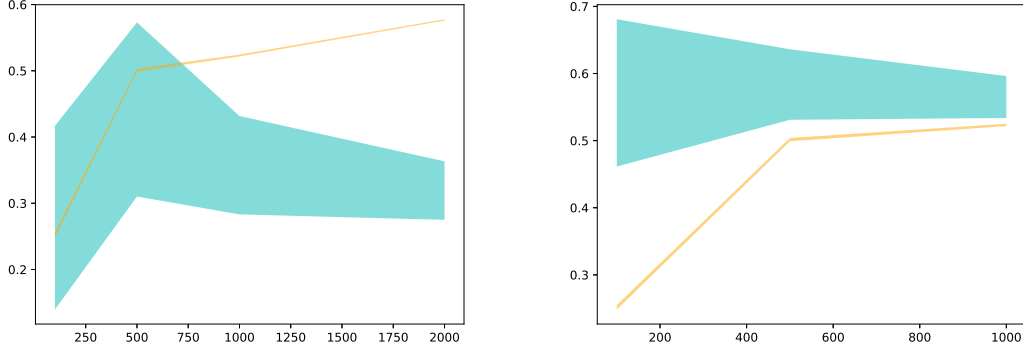


Figure 4.3: New calibration constant for the simulation 3 and 4. Variance in yellow and test statistic in blue. N in absciss. 95% dispersion realized on 50 simulations. Left: simulation 3. Right: simulation 4.

3.3 Algorithm

To finish this section we propose an algorithm to compute the test. We assume we have the deaths (τ_i, a_i) .

- (1) Compute for all t_i and a_k , $\theta_{i,k}^N$ thanks to the Goldenschluger Lepski Method. Get the variance $V_{i,k}^N$ obtained in the Goldenschluger Lepski Algorithm.
- (2) Set $V^N = 0.2 \max(V_{i,k}^N)$
- (3) Compute $c_{\mathcal{F}}^N = \frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \mu_{i,k}^N \left[\mathbf{1}_{t_i=\tau_k} - \mathfrak{V}_k^N(t_{i+1} - t_i) \theta_{i,k}^N \right]$.
- (4) Return the value of $\mathbf{1}_{c_{\mathcal{F}}^N \geq V^N}$

Remark 4.10. *The value 0.2 for the calibration constant has no other justification than a empirical one. This is the constant which gives us the best results in all our simulations. A good way to find a theoretical justification for this constant would be to investigate further the link between the variance obtained in the Goldenschluger Lepski algorithm and the true variance which appeared in the concentration inequality.*

We can just be sure that $\max(V_{i,k}^N) \geq \rho_N^$ and they have the same order in N . So there exist a constant $C \leq 1$ such that $\rho_N^* = CV^N$. 0.2 comes from our empirical test.*

4 Proof of theorem 4.4

4.1 To begin with the end

We note $\rho_N^* = \log(N) |w_1 w_2|_{1,\infty} N^{-1/2}$, where w_i are the kernels with the bandwith chosen to complied with a point on \mathcal{D}_L . We have to take the worth speed of convergence since we depends on the population density on all the domain. In our setting we have $\gamma = \infty$ and $\nu = \infty$. According

to the precedent chapter $\rho_N^* = \log(N)|w_1 w_2|_{1,\infty} N^{-1/2} = C_K \log(N) N^{-\frac{\min(\alpha, \beta, \delta, \rho)}{2 \min(\alpha, \beta, \delta, \rho) + 1}}$, where C_K is a constant depending only of the kernels chosen. We begin by a property which is sufficient to finish the proof.

Proposition 4.11. *There exist a constant C depending on \mathcal{R} , C_0 , T , $|g_0|_1$ and the kernels such that*

$$\mathbb{P}_\mu(|d_{\mathcal{F}}^N - d_{\mathcal{F}}| \geq C\rho_N^*) \lesssim \delta_N$$

with $\delta_N \xrightarrow{N \rightarrow \infty} 0$.

Now we finish the proof of the theorem before turning into the proof of the proposition 4.11. We use the test $\phi_N^* = \mathbf{1}_{d_{\mathcal{F}}^N \geq C\rho_N^*}$.

If we take $\mu \in H^\beta(L)$, this mean we are in the alternative H_0 . Then $d_{\mathcal{F}} = 0$ and we have

$$\mathbb{P}_\mu(\phi_N^* = 1) = \mathbb{P}_\mu(d_{\mathcal{F}}^N \geq C\rho_N^*) \leq \mathbb{P}_\mu(|d_{\mathcal{F}}^N| \geq C\rho_N^*) \leq \delta_N$$

Applying proposition 4.11 since $|d_{\mathcal{F}}^N| = |d_{\mathcal{F}}^N - d_{\mathcal{F}}|$.

And now taking $\mu \in \mathcal{F}(C'\rho_N)$ for $\rho_N \geq \rho_N^*$. We have

$$\mathbb{P}_\mu(\phi_N^* = 0) = \mathbb{P}_\mu(d_{\mathcal{F}}^N \leq C\rho_N^*)$$

Since $d_{\mathcal{F}} \geq C'\rho_N \geq C'\rho_N^*$ and we can have $d_{\mathcal{F}}^N \leq C\rho_N^*$ only if at least $|d_{\mathcal{F}}^N - d_{\mathcal{F}}| \geq (C' - C)\rho_N^*$ and we get

$$\mathbb{P}_\mu(d_{\mathcal{F}}^N \leq C\rho_N^*) \leq \mathbb{P}_\mu(|d_{\mathcal{F}}^N - d_{\mathcal{F}}| \geq (C' - C)\rho_N^*) \leq \delta_N$$

If $C' \geq 2C$, so $C_* = 2C$ and in the end we have

$$\limsup_{N \rightarrow \infty} R(C', \phi_N^*, \rho_N) \leq \limsup_{N \rightarrow \infty} R(C_*, \phi_N^*, \rho_N^*) \lesssim \limsup_{N \rightarrow \infty} \delta_N = 0$$

Now we just have to do the proof of the proposition 4.11.

4.2 Proof of the proposition 4.11

During this proof we begin fix a $\mu \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)$ during all the proof.

Lemma 4.12. *Let ψ_1 , ψ_2 and X a set such that $\sup_{x \in X} \psi_1(x) = \sup_{x \in X} |\psi_1(x)|$ and $\sup_{x \in X} \psi_2(x) = \sup_{x \in X} |\psi_2(x)|$, then*

$$|\sup_{x \in X} \psi_1(x) - \sup_{x \in X} \psi_2(x)| \leq \sup_{x \in X} |\psi_2(x) - \psi_1(x)|$$

Proof. The proof of this lemma comes from the fact that

$$\sup_{x \in X} \psi_1(x) - \sup_{x \in X} \psi_2(x) \leq \sup_{x \in X} |\psi_1(x) - \psi_2(x)| + \underbrace{\sup_{x \in X} |\psi_2(x)| - \sup_{x \in X} \psi_2(x)}_{=0}$$

And this is the same for $\sup_{x \in X} \psi_1(x) - \sup_{x \in X} \psi_2(x)$ which concludes the proof of the lemma. \square

For the next lemma we remind that we work under all the assumptions, especially the one assuming that $g_0 = 0$ for $a \geq A_{\max}$.

Lemma 4.13. $\forall a \leq A_{\max} + T$, there exists a constant C_1 such that

$$\mathbb{P} \left(\left| \mathfrak{Y}^N(a) - \frac{1}{\int_0^T g(s, a) ds} \right| \gtrsim \frac{1+u}{\sqrt{N}} \right) \leq e^{-C_1 \sqrt{N}} + \frac{1}{e^u - 1}$$

Proof. We assume the space for the age is compact. We can add the function $a \rightarrow \mathbf{1}_{a=a_0}$, pour tout $a_0 \leq A_{\max} + T$ to all the function which must be in the minimal fonctionnal space \mathcal{F} as defined in the proposition 3.9.

From this we get that $X^N = \left| \int_0^T \int_0^\infty \mathbf{1}_{u=a} Z_s^N(du) ds - \int_0^T g(s, a) ds \right|$ have a mild concentration of order $N^{-1/2}$. From this we deduce easily the result for $\mathfrak{Y}^N(a)$. \square

Let $X = \mathcal{F}$,

$$\psi_1(\varphi) = \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi(a)) \mu(s, a) g(s, a) dads$$

and

$$\psi_2(\varphi) = \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi^N(a)) \Gamma^N(ds, da)$$

Since \mathcal{F} satisfy the assumption 3.5 we have for any $\varphi \in \mathcal{F}$, $-\varphi \in \mathcal{F}$.

So $\sup_{\varphi \in \mathcal{F}} |\psi_i(\varphi)| = \sup_{\varphi \in \mathcal{F}} \psi_i(\varphi)$, since $\psi_i(-\varphi) = -\psi_i(\varphi)$. Indeed we can take a fast proof for ψ_1 and the idea is the same for ψ_2 .

Let $\epsilon > 0$, and φ_ϵ such that $|\psi_1(\varphi_\epsilon)| \geq \sup_{\varphi \in \mathcal{F}} |\psi_1(\varphi)| - \epsilon$. Then we can assume that $|\psi_1(\varphi_\epsilon)| = \psi_1(\varphi_\epsilon)$. if not, we just have to take $-\varphi_\epsilon$. It implies that $\sup_{\varphi \in \mathcal{F}} \psi_1(\varphi) \geq \sup_{\varphi \in \mathcal{F}} |\psi_1(\varphi)|$ and obviously we have $\sup_{\varphi \in \mathcal{F}} \psi_1(\varphi) \leq \sup_{\varphi \in \mathcal{F}} |\psi_1(\varphi)|$. Which gives $\sup_{\varphi \in \mathcal{F}} |\psi_1(\varphi)| = \sup_{\varphi \in \mathcal{F}} \psi_1(\varphi)$. We can apply lemma 4.12 to get

$$\begin{aligned} |d_{\mathcal{F}}^N - d_{\mathcal{F}}| &= \left| \sup_{\varphi \in \mathcal{F}} \psi_2(\varphi) - \sup_{\varphi \in \mathcal{F}} \psi_1(\varphi) \right| \\ &\leq \sup_{\varphi \in \mathcal{F}} |\psi_2(\varphi) - \psi_1(\varphi)| \\ &\leq \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi(a)) \mu(s, a) g(s, a) dads \\ &\quad - \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi^N(a)) \Gamma^N(da, ds) \\ &\leq T_1 + T_2 \end{aligned}$$

where

$$T_1 = \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi(a)) [\Gamma^N(ds, da) - g(s, a) \mu(s, a) dads]$$

and

$$T_2 = \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty (m_\varphi^N(a) - m_\varphi(a)) \Gamma^N(ds, da)$$

First we take care of T_1 . If we have $\varphi(s, a) - m_\varphi(a) \in \mathcal{F}$, which is the case when we work with $\mathcal{F} = \mathcal{H}^{\bar{\gamma}, \bar{\delta}}$, then

$$\begin{aligned} T_1 &\leq \sup_{\tilde{\varphi} \in \mathcal{F}} \int_0^T \int_0^\infty \tilde{\varphi}(s, a) [\Gamma^N(da, ds) - g(s, a)\mu(s, a)dads] \\ &\leq \mathcal{D}_{1,1}^N(\mathcal{F})_T \end{aligned}$$

with

$$\mathcal{D}_{w_1, w_2}^N(\mathcal{F})_T = \sup_{\varphi \in \mathcal{F}} \left| \int_0^T \int_0^\infty \varphi(s, a) w_1(s) w_2(s-a) [\Gamma^N(da, ds) - g(s, a)\mu(s, a)dads] \right|$$

Remark 4.14. *This is where the remark 4.3 is fully needed to ensure the simplicity of the proof. Since we take $\mathcal{F} = \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)$, we have $\varphi(s, a) - m_\varphi(a) \in \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)$ up to a constant. One could relax the choice of $\mathcal{F} = \mathcal{H}^{\bar{\gamma}, \bar{\delta}}(L)$, however this would prove another upper bound since we would no longer have $d(\mu) \approx |\mu - m_\mu|_{2,g}$ in this case.*

Proposition 4.15. *We have, noting $v_N = |w_1 w_2|_{1, \infty} N^{-1/2}$, there exist \mathfrak{C}_1 such that*

$$\mathbb{P}(\mathcal{D}_{w_1, w_2}^N(\mathcal{F})_T \geq \mathfrak{C}_1 v_N (1+u)) \lesssim (e^u - 1)^{-1} \wedge 1 + e^{-\sqrt{N}}$$

up to a constant depending only on \mathcal{R} , C_0 , T , and $|g_0|_1$, where these constants are defined in the precedent chapter.

So we have

$$\mathbb{P}(T_1 \geq \mathfrak{C}_1 \rho_N^*) \leq \mathbb{P}(\mathcal{D}_{1,1}^N(\mathcal{F})_T \geq \mathfrak{C}_1 \rho_N^*) \leq \mathbb{P}(\mathcal{D}_{1,1}^N(\mathcal{F})_T \geq \tilde{C} \ln(N) N^{-1/2}) \lesssim 1/N + e^{-\sqrt{N}} \lesssim 1/N$$

Now we can turn into T_2 . We need to cut according to \mathfrak{Y}^N . More precisely we have $T_2 \leq T_{2,1} + T_{2,2}$ where

$$\begin{aligned} T_{2,1} &= \sup_{\varphi \in \mathcal{F}} \int_0^\infty (m_\varphi^N(a) - \mathfrak{Y}^N(a) \int_0^T g(s, a) \varphi(s, a) \Gamma^N(ds, da)) \\ T_{2,2} &= \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty g(s, a) \varphi(s, a) (\mathfrak{Y}^N(a) - \frac{1}{\int_0^T g(s, a) ds}) \Gamma^N(ds, da) \end{aligned}$$

These two quantities can be bounded up to a constant by a quantity of the form

$$\sup_{\varphi \in \mathcal{F}} \frac{1}{N} D_T^N \sup_{k \leq D_T^N} Y_k^N$$

with $Y_k^N = \int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx$ for $T_{2,1}$ and $Y_k^N = \left| \mathfrak{Y}^N(a_k) - \frac{1}{\int_0^T g(s, a_k) ds} \right|$ for $T_{2,2}$.

We show the result for bounding $T_{2,1}$.

$$\begin{aligned}
 & \sup_{\varphi \in \mathcal{F}} \int_0^\infty (m_\varphi^N(a) - \mathfrak{Y}^N(a) \int_0^T g(s, a) \varphi(s, a)) \Gamma^N(ds, da) \\
 &= \sup_{\varphi \in \mathcal{F}} \int_0^T \int_0^\infty \mathfrak{Y}^N(a) \left[\int_0^T (\hat{g}_*^N(x, a) - g(x, a)) \varphi(x, a) dx \right] \Gamma^N(ds, da) \\
 &\leq \int_0^T \int_0^\infty \left[\int_0^T (\hat{g}_*^N(x, a) - g(x, a)) dx \right] \Gamma^N(ds, da) \\
 &\lesssim \frac{1}{N} D_T^N \sup_{k \leq D_T^N} \int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx
 \end{aligned}$$

From the theorem 3.26 we know how to control $\mathbb{E} [|\hat{g}_*^N(x, a_k) - g(x, a_k)|^p | B_N]$, and we have the following lemma, easy consequence of the Bernstein inequality.

Lemma 4.16. *Let $Y_s > 0$ such that for all $s \in [0, T]$, $\mathbb{E}[|Y_s|^p] \leq p! B C^{p-2}$, then*

$$\mathbb{P} \left(\left| \int_0^T |Y_s| - \mathbb{E}(|Y_s|) ds \right| \geq \epsilon \right) \leq 2 \exp \left(\frac{-\epsilon^2}{2(B_T + C_T \epsilon)} \right)$$

where $B_T = 4TB$ and $C_T = 2TC$.

We have $\mathbb{E} [|\hat{g}_*^N(x, a_k) - g(x, a_k)|^p | B_N] \leq M p! v_N^p$. So we can apply the lemma with $B = M v_N^2$ and $C = v_N$ and we have

$$\begin{aligned}
 & \mathbb{P} \left(\int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq \widetilde{C}_1 T v_N (1 + u) | B_N \right) \\
 &= \mathbb{P} \left(\int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq \int_0^T \mathbb{E} [|\hat{g}_*^N(x, a_k) - g(x, a_k)|] dx + \epsilon | B_N \right) \\
 &\leq 2 \exp \left(\frac{-\epsilon^2}{4T v_N (2M v_N + \epsilon)} \right)
 \end{aligned}$$

where $\epsilon = \widetilde{C}_1 T v_N (1 + u) - \int_0^T \mathbb{E} [|\hat{g}_*^N(x, a_k) - g(x, a_k)|] dx \geq T v_N u$.

Since $x \mapsto \exp \left(\frac{-x^2}{4T v_N (2M v_N + x)} \right)$ is decreasing for $x \geq 0$ we have

$$\begin{aligned}
 \mathbb{P} \left(\int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq \widetilde{C}_1 T v_N (1 + u) | B_N \right) &\leq 2 \exp \left(-\frac{T^2 v_N^2 u^2}{4T v_N^2 (2M + Tu)} \right) \\
 &\leq \exp \left(-\frac{Tu^2}{4(2M + u)} \right)
 \end{aligned}$$

So we have

$$\mathbb{P} \left(\int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq AT \widetilde{C}_1 \rho_N^* | B_N \right) \lesssim \exp \left(-\frac{A^2 T \ln(N)^2}{4(2M + A \ln(N))} \right) \lesssim N^{-AT}$$

for N great enough, and we take $A = 2/T$. We can write with $K = 2\widetilde{C}_2\widetilde{C}_1$

$$\begin{aligned} \mathbb{P}(T_2 \geq K\rho_N^*) &\leq \mathbb{P}\left(\frac{1}{N}D_T^N \sup_{k \leq D_T^N} \int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq K\rho_N^* | D_T^N \leq \widetilde{C}_2 N, B_N\right) \\ &\quad + \mathbb{P}(D_T^N \geq \widetilde{C}_2 N) + \mathbb{P}(B_N^c) \\ &\leq \widetilde{C}_2 N \sup_{k \leq \widetilde{C}_2 N} \mathbb{P}\left(\int_0^T |\hat{g}_*^N(x, a_k) - g(x, a_k)| dx \geq 2\widetilde{C}_1 \rho_N^* | B_N\right) + 1/N + e^{-\sqrt{N}} \lesssim N^{-1} \end{aligned}$$

This conclude the part for $T_{2,1}$. For $T_{2,2}$ we just need to use the lemma 4.13. This conclude the proof with $\delta_N = N^{-1}$.

4.3 Proof of property 4.15

Lemma 4.17.

$$\mathcal{D}_{w_1, w_2}^N(\mathcal{F})_T \leq \sup_{\varphi \in \mathcal{F}} \mathcal{I}_T^N(w_1, w_2)(\varphi) + \sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \quad (4.2)$$

Where $\Delta_{w_1, w_2}^N(\varphi)$ is a squared integrable martingale with predictable compensator $\langle \Delta_{w_1, w_2}^N(\varphi) \rangle_t$ equals to

$$N^{-1} \int_0^t \int_{\mathcal{U}} w_1(s)^2 w_2(s-a)^2 \varphi(s, a)^2 \mu(s, a, Z_{s-}^N U(a)) Z_{s-}^N(da) ds. \quad (4.3)$$

Moreover we have

$$\sup_{\varphi \in \mathcal{F}} \mathcal{I}_T^N(w_1, w_2)(\varphi) \lesssim |w_1|_1 \mathcal{W}_{1, w_2}^N(\mathcal{F})_T + |w_1 w_2|_1 \mathcal{W}_{1, 1}^N(\mathcal{F})_T + \widetilde{\mathcal{W}}_{w_1, w_2}^N(\mathcal{F})_T \mathcal{W}_{1, 1}^N(\mathcal{F})_T \quad (4.4)$$

Proof. Comes from the proof of the lemma 3.42. \square

We next study the deviation of $\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T|$.

Lemma 4.18. For any $\varphi_1, \varphi_2 \in \mathcal{F}$ we have on a certain event $\mathcal{A}_N \cap \mathcal{B}_N$

$$|\Delta_{w_1, w_2}^N(\varphi_1) - \Delta_{w_1, w_2}^N(\varphi_2)|_\psi \lesssim v_N |\varphi_1 - \varphi_2|_\infty$$

Where $|\cdot|_\psi$ stands for the orlicz norm with $\psi(x) = e^x - 1$ and $\mathbb{P}(\mathcal{A}_N^c \cap \mathcal{B}_N) \lesssim e^{-\sqrt{N}}$.

With this lemma we can then conclude applying the result on metric entropy, as found in [44], we get a mild concentration on $\mathcal{A}_N \cap \mathcal{B}_N$ for $\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)|$ with rate v_N .

Proof. Let $B_{w_1, w_2}(\varphi)$ be the bound of the jumps of $\Delta_{w_1, w_2}^N(\varphi)$. Now we work on \mathcal{B}_N .

We plan to apply a classical deviation inequality for martingales (see *e.g.* Lemma 2.1 in van der Geer [64] or the classical textbook Shorak and Wellner [61]), namely:

$$\mathbb{P}(|\Delta_{w_1, w_2}^N(\varphi)_T| \geq u, \langle \Delta_{w_1, w_2}^N(\varphi) \rangle_T \leq v, B_{w_1, w_2}(\varphi) \leq c) \leq 2 \exp\left(-\frac{v}{c^2} \tilde{\rho}\left(\frac{cu}{v}\right)\right) \quad (4.5)$$

Since $\Delta_{w_1, w_2}^N(\varphi_1 - \varphi_2) = \Delta_{w_1, w_2}^N(\varphi_1) - \Delta_{w_1, w_2}^N(\varphi_2)$, we will just work with a generic φ during the proof.

We have

$$B_{w_1, w_2}(\varphi) \lesssim N^{-1} |w_1|_\infty |w_2|_\infty \mathcal{R} \left[1 + (1 + \tilde{P}_T^N) \right] (1 + \tilde{P}_T^N) |\varphi|_\infty$$

And we have from the lemma 3.42

$$\langle \Delta_{w_1, w_2}^N(\varphi) \rangle_T \lesssim N^{-1} [\mathcal{I}_T^N(w_1^2, w_2^2) + |w_1 w_2|_2^2].$$

We take \mathcal{A}_N the event $\{\mathcal{W}_{1,1}^N(\mathcal{F})_T \lesssim 1\}$. We also define $\mathcal{A}(\kappa)$ as an event where $\mathcal{W}_{1, w_2}^N(\mathcal{F})_T \lesssim (1 + \kappa) |w_2|_2 N^{-1/2}$ and $\widetilde{\mathcal{W}}_{w_1, w_2}^N(\mathcal{F})_T \lesssim (1 + \kappa) |w_1 w_2|_2 N^{-1/2}$, up to a good constant such that (4.6) is true. And $\tilde{P}_T^N \lesssim 1$ since we are on \mathcal{B}_N . We have

$$B_{w_1, w_2}(\varphi) \lesssim N^{-1} |w_1|_\infty |w_2|_\infty |\varphi|_\infty \lesssim v_N |\varphi|_\infty, \quad \langle \Delta_{w_1, w_2}^N \rangle_T \lesssim (1 + \kappa) v_N^2 |\varphi|_\infty^2$$

We can impose to find an event $\mathcal{A}(\kappa)$ such that

$$\int_0^\infty \mathbb{P}(\mathcal{A}(\kappa)^c) e^\kappa d\kappa \leq 1/2 \quad (4.6)$$

as in the proof of Theorem 3.13 for example. Then we have a constant C,

$$\mathbb{P}(|\Delta_{w_1, w_2}^N(\varphi)_T| \geq u, \mathcal{A}(\kappa), \mathcal{B}_N, \mathcal{A}_N) \leq 2 \exp\left(-C_1 \frac{(1 + \kappa)}{v_N^2} \tilde{\rho}\left(C_2 \frac{u}{(1 + \kappa) |\varphi|_\infty}\right)\right)$$

And we can calculate $\mathbb{E}[\psi_1(c^{-1} |\Delta_{w_1, w_2}^N(\varphi_1)_T - \Delta_{w_1, w_2}^N(\varphi_2)_T|) \mathbf{1}_{B_N}]$ equals to

$$\begin{aligned} &= \mathbb{E}[\exp(c^{-1} |\Delta_{w_1, w_2}^N(\varphi_1)_T - \Delta_{w_1, w_2}^N(\varphi_2)_T|) \mathbf{1}_{B_N}] - 1 \\ &= \int_1^\infty \mathbb{P}(\exp(c^{-1} |\Delta_{w_1, w_2}^N(\varphi_1)_T - \Delta_{w_1, w_2}^N(\varphi_2)_T|) \geq \kappa, \mathcal{B}_N, \mathcal{A}_N) d\kappa \\ &= \int_0^\infty \mathbb{P}(|\Delta_{w_1, w_2}^N(\varphi_1)_T - \Delta_{w_1, w_2}^N(\varphi_2)_T| \geq c\kappa, \mathcal{B}_N, \mathcal{A}_N) e^\kappa d\kappa \\ &\leq 2 \int_0^\infty \exp\left(-C_1 \frac{(1 + \kappa)}{v_N^2} \tilde{\rho}\left(C_2 \frac{c\kappa}{|\varphi|_\infty (1 + \kappa)}\right)\right) e^\kappa d\kappa + \frac{1}{2}, \end{aligned}$$

Then we just have to take $c = \tilde{C} |\varphi|_\infty v_N$, to conclude. \square

From this we have

$$\begin{aligned} \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \geq C v_N (1 + u)) &\leq \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \geq C v_N (1 + u), B_N, \mathcal{A}_N) \\ &\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \geq C v_N (1 + u), B_N^c, \mathcal{A}_N) \\ &\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \geq C v_N (1 + u), B_N, \mathcal{A}_N^c) \\ &\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\Delta_{w_1, w_2}^N(\varphi)_T| \geq C v_N (1 + u), B_N^c, \mathcal{A}_N^c) \\ &\leq (e^u - 1)^{-1} \wedge 1 + 2\mathbb{P}(B_N^c) + \mathbb{P}(\mathcal{A}_N^c, B_N) \\ &\lesssim (e^u - 1)^{-1} \wedge 1 + e^{-\sqrt{N}} \end{aligned}$$

$$\begin{aligned}
\mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\mathcal{I}_{w_1, w_2}^N(\varphi)_T| \geq Cv_N(1+u)) &\leq \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\mathcal{I}_{w_1, w_2}^N(\varphi)_T| \geq Cv_N(1+u), B_N, \mathcal{A}_N) \\
&\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\mathcal{I}_{w_1, w_2}^N(\varphi)_T| \geq Cv_N(1+u), B_N^c, \mathcal{A}_N) \\
&\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\mathcal{I}_{w_1, w_2}^N(\varphi)_T| \geq Cv_N(1+u), B_N, \mathcal{A}_N^c) \\
&\quad + \mathbb{P}(\sup_{\varphi \in \mathcal{F}} |\mathcal{I}_{w_1, w_2}^N(\varphi)_T| \geq Cv_N(1+u), B_N^c, \mathcal{A}_N^c) \\
&\leq (e^u - 1)^{-1} \wedge 1 + 2\mathbb{P}(B_N^c) + \mathbb{P}(\mathcal{A}_N^c, B_N) \\
&\lesssim (e^u - 1)^{-1} \wedge 1 + e^{-\sqrt{N}}
\end{aligned}$$

which gives us the result by union bound.

5 Remaining proofs

5.1 Discussion around the lower bound

Even without any proof we have some result about the lower bound. First of all we take $\mu_0 = 1$ and we can take $\mu_k = 1 + \nu_k$. We have then the following lemma.

Lemma 4.19. \mathbb{P}_k is absolutely continuous with respect to \mathbb{P}_0 and we have

$$\begin{aligned}
\ln(\xi_k) = \ln\left(\frac{d\mathbb{P}_k}{d\mathbb{P}_0}\Big|_T\right) &= N \int_0^T \int_0^\infty -\nu_k(s, a) + \ln(1 + \nu_k(s, a)) Z_s^N(da) ds \\
&\quad + \int_0^T \int_{\mathbb{N} \times \mathbb{R}_+} \mathbf{1}_{i \leq N(Z_{s-}^N, 1), \vartheta \leq 1} \log(1 + \nu_k(s, a_i)) \tilde{Q}_\mu(ds, di, d\vartheta)
\end{aligned}$$

This lemma is just an adaptation of the lemma 3.50. We can from this lemma get a control of $\ln(\xi_k)$. We have

$$|\ln(\xi_k)| \leq Y_k + |M_k(T)|$$

where $Y_k = N|\nu_k|_{2,g}^2 + N\widetilde{W}_{\nu_k^2}^N(\mathcal{F})_T$ and $M_k(T)$ is a martingale such that

$$\langle M_k \rangle_T \leq Y_k$$

Lemma 4.20.

$$\mathbb{P}(|\ln(\xi_k)| \geq 2C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2) \leq$$

Proof.

$$\begin{aligned}
&\mathbb{P}(|\ln(\xi_k)| \geq 2C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2) \\
&\leq \mathbb{P}(N\widetilde{W}_{\nu_k^2}^N(\mathcal{F})_T \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)) \\
&\quad + \mathbb{P}(|M_k(T)| \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)) \\
&\lesssim (e^u - 1)^{-1} \wedge 1 \\
&\quad + e^{-C'|\nu_k|_2|\nu_k|_\infty^{-1} N^{1/2}(1+u)\bar{\rho}} \left(\frac{|\nu_k|_\infty C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)}{C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2} \right)
\end{aligned}$$

Since $|\Delta M_k| \leq |\nu_k|_\infty$

$$\begin{aligned}
 & \mathbb{P}(|M_k(T)| \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)) \leq \\
 & \mathbb{P}(|M_k(T)| \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u), \langle M_k \rangle_T \leq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2) \\
 & + \mathbb{P}(Y_k \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2) \\
 & \lesssim \exp\left(-C'|\nu_k|_2|\nu_k|_\infty^{-1} N^{1/2}(1+u)\tilde{\rho}\left(\frac{|\nu_k|_\infty C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)}{C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u) + N|\nu_k|_{2,g}^2}\right)\right) + (e^u - 1)^{-1} \wedge 1
 \end{aligned}$$

$$\mathbb{P}(N\widetilde{W}_{\nu_k}^N(\mathcal{F})_T \geq C'|\nu_k|_2|\nu_k|_\infty N^{1/2}(1+u)) \lesssim (e^u - 1)^{-1} \wedge 1 \quad \square$$

To end the lower bound we just need to control $\mathbb{E}[\frac{1}{M^2}(\sum_{k=1}^M \xi_k - 1)^2]$ since in this case we would have with $Z = \frac{1}{M} \sum_{k=1}^M \xi_k$

$$\mathbb{P}(|Z - 1| \leq \delta) \leq \mathbb{E}[\frac{1}{M^2}(\sum_{k=1}^M \xi_k - 1)^2] \delta^{-2}$$

However we are only able to control with our precedent control $\mathbb{E}[\ln(\xi_k)^2]$ which is not sufficient to conclude.

5.2 Proofs of the property 4.6

We use the lemma 4.12 to get that

$$|d_{\mathcal{F}}^N - c_{\mathcal{F}}^N| \leq \sup_{\varphi \in \mathcal{F}} |d^N(\varphi) - c^N(\varphi)|$$

where

$$d^N(\varphi) = \int_0^T \int_0^\infty (\varphi(s, a) - m_\varphi^N(a)) \Gamma^N(ds, da)$$

and

$$c^N(\varphi) = \int_0^T \int_0^\infty \varphi(s, a) \Gamma^N(ds, da) - \frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} \mathfrak{B}_k^N(t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N$$

So we have since \mathfrak{W}_k^N is bounded uniformly in k by the lemma 4.13 and the fact that g is bounded below.

$$\begin{aligned}
& |d^N(\varphi) - c^N(\varphi)| \\
& \leq \mathfrak{W}_k^N \left| \frac{1}{N} \sum_{k=1}^{D_T^N} \sum_{i=1}^{R_T^N} (t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N - \frac{1}{N} \sum_{k=1}^{D_T^N} \int_0^T \hat{g}_\star^N(s, a_k) \varphi(s, a_k) ds \right| \\
& \lesssim \frac{1}{N} \sum_{k=1}^{D_T^N} \left| \sum_{i=1}^{R_T^N} (t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N - \int_0^T \hat{g}_\star^N(s, a_k) \varphi(s, a_k) ds \right| \\
& \lesssim \frac{1}{N} D_T^N \sup_{k \leq D_T^N} \sum_{i=1}^{R_T^N} \left| (t_{i+1} - t_i) \varphi_{i,k} \theta_{i,k}^N - \int_{t_i}^{t_{i+1}} \hat{g}_\star^N(s, a_k) \varphi(s, a_k) ds \right| \\
& \lesssim \frac{1}{N} D_T^N \sup_{k \leq D_T^N} \sup_{i \leq R_T^N} \int_{t_i}^{t_{i+1}} |\varphi_{i,k} \theta_{i,k}^N - \hat{g}_\star^N(s, a_k) \varphi(s, a_k)| ds \\
& \lesssim \frac{1}{N} D_T^N \sup_{k \leq D_T^N} \sup_{i \leq R_T^N} \int_{t_i}^{t_{i+1}} |\varphi_{i,k} \theta_{i,k}^N - \hat{g}_\star^N(s, a_k) \varphi(s, a_k)| ds \\
& \lesssim \frac{1}{N} D_T^N \sup_{k \leq D_T^N} \sup_{i \leq R_T^N} \int_{t_i}^{t_{i+1}} |\theta_{i,k}^N - \hat{g}_\star^N(s, a_k)| ds + \int_{t_i}^{t_{i+1}} |\hat{g}_\star^N(s, a_k) (\varphi_{i,k} - \varphi(s, a_k))| ds
\end{aligned}$$

We get, since K is lipschitz and the lipschitz constant of K_h is h^{-2}

$$\sup_{i \leq R_T^N} \int_{t_i}^{t_{i+1}} |\theta_{i,k}^N - \hat{g}_\star^N(s, a_k)| ds \lesssim \frac{\tilde{n}_T}{N} |t_{i+1} - t_i|^2 \frac{1}{h_N^2} \lesssim \frac{\tilde{n}_T}{N} \frac{L_N^2}{h_N^2}$$

and

$$\sup_{i \leq R_T^N} \int_{t_i}^{t_{i+1}} |\hat{g}_\star^N(s, a_k) (\varphi_{i,k} - \varphi(s, a_k))| ds \lesssim \frac{1}{h_N} |t_{i+1} - t_i|^{1+\bar{\delta}} \lesssim L_N^{1+\bar{\delta}} h_N^{-1}$$

Which leads to the end of the proof.

BIBLIOGRAPHY

- [1] Séverine Arnold, Alexandre Boumezoued, H eloise Labit Hardy, and Nicole El Karoui. Cause-of-death mortality: What can be learned from population dynamics? 2015.
- [2] Harry Bensusan. *Risques de taux et de long evit e: Mod elisation dynamique et Applications aux produits d eriv es et   l'assurance-vie*. PhD thesis, Ecole Polytechnique X, 2010.
- [3] Rudolf Beran. Nonparametric regression with randomly censored survival data. Technical report, Technical Report, Univ. California, Berkeley, 1981.
- [4] Herman J. Bierens. *Topics in advanced econometrics*. Cambridge University Press, Cambridge, 1994. Estimation, testing, and specification of cross-section and time series models.
- [5] S. Val ere Bitseki Penda, Marc Hoffmann, and Ad elai de Olivier. Adaptive estimation for bifurcating Markov chains. *Bernoulli*, 23(4B):3598–3637, 2017.
- [6] SG Bobkov, Friedrich G otze, et al. Concentration of empirical distribution functions with applications to non-iid models. *Bernoulli*, 16(4):1385–1414, 2010.
- [7] Fran ois Bolley, Arnaud Guillin, and C edric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.
- [8] Alexandre Boumezoued. *Approches micro-macro des dynamiques de populations h et rog enes structur ees par  ge. Application aux processus auto-excitants et   la d emographie*. PhD thesis, Paris 6, 2016.
- [9] Alexandre Boumezoued. Improving HMD mortality estimates with HFD fertility data. *To appear in the North American Actuarial Journal*, 2016.
- [10] Alexandre Boumezoued, Marc Hoffmann, and Paulien Jeunesse. A new inference strategy for general population mortality tables. 2018.
- [11] Alexandre Boumezoued, Marc Hoffmann, and Paulien Jeunesse. Statistical inference for an in-homogeneous age-structured population process. *Forthcoming*, 2018.

-
- [12] Elodie Brunel, Fabienne Comte, and Agathe Guilloux. Estimation strategies for censored lifetimes with a lexis-diagram type model. *Scandinavian Journal of Statistics*, 35(3):557–576, 2008.
- [13] Adam D Bull and Richard Nickl. Adaptive confidence sets in l2. *Probability Theory and Related Fields*, 156(3-4):889–919, 2013.
- [14] Andrew JG Cairns, David Blake, Kevin Dowd, and Amy R Kessler. Phantoms never die: living with unreliable population data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):975–1005, 2016.
- [15] Àngel Calsina and journal=Ecological Modelling volume=133 number=1-2 pages=33–43 year=2000 publisher=Elsevier Cuadrado, Sílvia. A model for the adaptive dynamics of the maturation age.
- [16] Nicolas Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic processes and their applications*, 116(8):1127–1160, 2006.
- [17] Nicolas Champagnat, Régis Ferrière, and Sylvie Méléard. Individual-based probabilistic models of adaptive evolution and various scaling approximations. In *Seminar on Stochastic Analysis, Random Fields and Applications V*, pages 75–113. Springer, 2007.
- [18] Stéphan Clémentçon, Viet Chi Tran, and Hector De Arazoza. A stochastic SIR model with contact-tracing: large population limits and statistical inference. *Journal of Biological Dynamics*, 2(4):392–414, 2008.
- [19] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 47, pages 1171–1196. Institut Henri Poincaré, 2011.
- [20] Dorota M Dabrowska. Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*, pages 181–197, 1987.
- [21] DJ Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods of Probability and its Applications*. Springer, New York, 2003.
- [22] Marie Doumic, Marc Hoffmann, Nathalie Krell, and Lydia Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3):1760–1799, 2015.
- [23] R. M. Dudley. Universal Donsker classes and metric entropy. *Ann. Probab.*, 15(4):1306–1326, 1987.
- [24] Bruno Ernande, Ulf Dieckmann, and Mikko Heino. Adaptive changes in harvested populations: plasticity and evolution of age and size at maturation. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1537):415–423, 2004.
- [25] Nicolas Fournier and Sylvie Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.

- [26] Nicolas Fournier, Sylvie Méléard, et al. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, 14(4):1880–1919, 2004.
- [27] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.
- [28] Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008.
- [29] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011.
- [30] Quentin Guibert, Olivier Lopez, and Pierrick Piette. Forecasting mortality rate improvements with a high-dimensional var. 2017.
- [31] Agathe Guilloux, Sarah Lemler, and Marie-Luce Taupin. Adaptive kernel estimation of the baseline function in the cox model with high-dimensional covariates. *Journal of Multivariate Analysis*, 148:141–159, 2016.
- [32] Wolfgang Härdle, Axel Werwatz, Marlene Müller, and Stefan Sperlich. Nonparametric density estimation. In *Nonparametric and Semiparametric Models*, pages 39–83. Springer, 2004.
- [33] HFD. The human fertility database. max planck institute for demographic research (germany) and vienna institute of demography (austria).
- [34] HMD. The human mortality database.
- [35] Marc Hoffmann and Adélaï de Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Process. Appl.*, 126(5):1433–1471, 2016.
- [36] Marc Hoffmann and Adélaïde Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and their Applications*, 126(5):1433–1471, 2016.
- [37] Yuri Ingster and Irina A Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2012.
- [38] Yuri Izmailovich Ingster. On the minimax nonparametric detection of signals in white gaussian noise. *Problemy Peredachi Informatsii*, 18(2):61–73, 1982.
- [39] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [40] Niels Keiding. Statistical inference in the lexis diagram. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 332(1627):487–509, 1990.
- [41] Eva Kisdi. Evolutionary branching under asymmetric competition. *Journal of Theoretical Biology*, 197(2):149–162, 1999.

-
- [42] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2):298–335, 2017.
- [43] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
- [44] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [45] Oleg V Lepski, Vladimir G Spokoiny, et al. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.
- [46] OV Lepski and AB Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117(1):17–48, 2000.
- [47] O. V. Lepskiĭ. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.
- [48] E. Löcherbach. Likelihood ratio processes for Markovian particle systems with killing and jumps. *Stat. Inference Stoch. Process.*, 5(2):153–177, 2002.
- [49] Eva Löcherbach. LAN and LAMN for systems of interacting diffusions with branching and immigration. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(1):59–90, 2002.
- [50] Mark G. Low. Nonexistence of an adaptive estimator for the value of an unknown probability density. *Ann. Statist.*, 20(1):598–602, 1992.
- [51] Aline Marguet. *Branching processes for structured populations and estimators for cell division*. PhD thesis, Université Paris-Saclay, 2017.
- [52] Ian W McKeague and Klaus J Utikal. Inference for a nonlinear counting process regression model. *The Annals of Statistics*, pages 1172–1187, 1990.
- [53] A.G. McKendrick. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.*, 54:98–130, 1926.
- [54] È. A. Nadaraja. On a regression estimate. *Teor. Veroyatnost. i Primenen.*, 9:157–159, 1964.
- [55] Jens P Nielsen and Oliver B Linton. Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, pages 1735–1748, 1995.
- [56] Khashayar Pakdaman, Benoît Perthame, and Delphine Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1):55, 2009.
- [57] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.
- [58] Benoît Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.
- [59] Ermanno Pitacco, Michel Denuit, and Steven Haberman. *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, 2009.

- [60] SJ Richards. Detecting year-of-birth mortality patterns with limited data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):279–298, 2008.
- [61] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Reprint of the 1986 original [MR0838963].
- [62] Viet Chi Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Probab. Stat.*, 12:345–386, 2008.
- [63] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [64] Sara van de Geer. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.*, 23(5):1779–1801, 1995.
- [65] Pierre-François Verhulst. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, 10:113–126, 1838.
- [66] H. Von Foerster. *The Kinetics of Cellular Proliferation*. Grune & Stratton, 1959.
- [67] Glenn F Webb and Graham Webb. *Theory of nonlinear age-dependent population dynamics*. CRC Press, 1985.
- [68] Jon A Wellner. The bennett-orlicz norm. *Sankhya A*, 79(2):355–383, 2017.
- [69] John R Wilmoth, Kirill Andreev, Dmitri Jdanov, Dana A Gleis, C Boe, M Bubenheim, D Philipov, V Shkolnikov, and P Vachon. Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock*. URL: <http://mortality.org> [version 31/05/2007], 9:10–11, 2007.

Résumé

L'étude du taux de mortalité dans des modèles de population humaine ou en biologie est le coeur de ce travail. Cette thèse se situe à la frontière de la statistique des processus, de la statistique nonparamétrique et de l'analyse.

Dans une première partie, centrée sur une problématique actuarielle, un algorithme est proposé pour estimer les tables de mortalité, utiles en assurance. Cet algorithme se base sur un modèle déterministe de population. Ces nouvelles estimations améliorent les résultats actuels en prenant en compte la dynamique globale de la population. Ainsi les naissances sont incorporées dans le modèle pour calculer le taux de mort. De plus, ces estimations sont mises en lien avec les travaux précédents, assurant ainsi la continuité théorique de notre travail.

Dans une deuxième partie, nous nous intéressons à l'estimation du taux de mortalité dans un modèle stochastique de population. Cela nous pousse à utiliser des arguments propres à la statistique des processus et à la statistique nonparamétrique. On trouve alors des estimateurs nonparamétriques adaptatifs dans un cadre anisotrope pour la mortalité et la densité de population, ainsi que des inégalités de concentration non asymptotiques quantifiant la distance entre le modèle stochastique et le modèle déterministe limite utilisé dans la première partie. On montre que ces estimateurs restent optimaux dans un modèle où le taux de mort dépend d'interactions, comme dans le cas de la population logistique.

Dans une troisième partie, on considère la réalisation d'un test pour détecter la présence d'interactions dans le taux de mortalité. Ce test permet en réalité de juger de la dépendance temporelle de ce taux. Sous une hypothèse, on montre alors qu'il est possible de détecter la présence d'interactions. Un algorithme pratique est proposé pour réaliser ce test.

Mots Clés

Statistique nonparamétrique, Dépendance, Statistique des processus aléatoires, Inégalités de concentration, Equation aux dérivées partielles, Adaptativité, Inégalité oracle, Sélection de fenêtre, Estimateurs à noyaux, vitesse de convergence minimax, test nonparamétrique.

Abstract

In this thesis, we study the mortality rate in different population models to apply our results to demography or biology. The mathematical framework includes statistics of process, nonparametric estimations and analysis.

In a first part, an algorithm is proposed to estimate the mortality tables. This problematic comes from actuarial science and the aim is to apply our results in the insurance field. This algorithm is founded on a deterministic population model. The new estimates we gets improve the actual results. Its advantage is to take into account the global population dynamics. Thanks to that, births are used in our model to compute the mortality rate. Finally these estimations are linked with the precedent works. This is a point of great importance in the field of actuarial science.

In a second part, we are interested in the estimation of the mortality rate in a stochastic population model. We need to use the tools coming from nonparametric estimations and statistics of process to do so. Indeed, the mortality rate is a function of two parameters, the time and the age. We propose minimax optimal and adaptive estimators for the mortality and the population density. We also demonstrate some non asymptotics concentration inequalities. These inequalities quantify the deviation between the stochastic process and its deterministic limit we used in the first part. We prove that our estimators are still optimal in a model where the mortality is influenced by interactions. This is for example the case for the logistic population.

In a third part, we consider the testing problem to detect the existence of interactions. This test is in fact designed to detect the time dependence of the mortality rate. Under the assumption the time dependence in the mortality rate comes only from the interactions, we can detect the presence of interactions. Finally we propose an algorithm to do this test.

Keywords

Nonparametric statistics, Dependance, Statistics of process, Concentration inequalities, Partial differential equation, Adaptive estimation, Oracle inequality, Bandwidth selection, Kernel estimation, minimax rate of convergence, nonparametric testing.