NNT : 201XSACLX073

THÉSE DE DOCTORAT
DE L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À L'ÉCOLE POLYTECHNIQUE

École doctorale n°574
École doctorale de mathématiques Hadamard
Spécialité de doctorat: Mathématiques appliquées

par

ALINE MARGUET

Processus de branchement pour des populations structurées et estimateurs pour la division cellulaire

Thèse soutenue à l'École Polytechnique, le 27 novembre 2017 devant un jury composé de

M.	VINCENT BANSAYE	École Polytechnique	(Directeur de thèse)
M.	JEAN BERTOIN	Universität Zürich	(Président du jury)
Mme.	MARIE DOUMIC	Université Pierre et Marie Curie	(Examinatrice)
Mme.	ANNE GÉGOUT-PETIT	Université de Lorraine	(Rapporteur)
M.	MARC HOFFMANN	Université Paris Dauphine	(Directeur de thèse)
Mme.	SYLVIE MÉLÉARD	École Polytechnique	(Examinatrice)

Après avis des rapporteurs :

Mme.	ANNE GÉGOUT-PETIT	Université de Lorraine
M.	ANDREAS KYPRIANOU	University of Bath

Remerciements

Mes premiers remerciements vont naturellement à Marc et Vincent. D'abord, pour la confiance que vous m'avez accordée en acceptant d'encadrer ma thèse. Ensuite, pour votre soutien continu tout au long de ses trois années sans lequel rien de tout ceci n'aurait été possible. Merci pour votre disponibilité, malgré vos emplois du temps que je sais chargés. Merci de m'avoir appris tant de choses et surtout, merci de m'avoir communiqué votre enthousiasme pour ce métier.

Merci à Anne Gégout-Petit et à Andreas Kyprianou d'avoir accepté de rapporter ma thèse ; j'en suis très honorée. Je suis très heureuse que Jean Bertoin, Marie Doumic et Sylvie Méléard aient accepté de participer à mon jury de thèse. Je remercie particulièrement Sylvie, pour son aide durant ces trois années ainsi que pour l'énergie qu'elle emploie à faire vivre la chaire MMB. J'ai beaucoup appris grâce aux journées organisées par la chaire, de par leur qualité scientifique, mais aussi par les rencontres que j'ai pu y faire. Je remercie pour cela également les membres de la chaire.

J'ai eu la chance de passer ces trois années au sein du CMAP. Merci à toute l'équipe administrative, pour leur aide précieuse dans les nombreuses démarches que j'ai eu à effectuer durant la thèse. Merci aux nombreux (trop nombreux pour en faire la liste ...) (ex-)doctorants du CMAP, et particulièrement aux (ex-)occupants du bureau 2009, ainsi qu'aux abonnés de la pause de 16h, avec qui j'ai partagé mes joies, mes coups de blues et qui ont partagés avec moi leurs gâteaux. Je garderai un très bon souvenir de ces trois années, et c'est en partie grâce à vous.

Enfin, merci à mes amis et à ma famille pour leur soutien. Et merci à Timothée, pour tout.

Contents

Introduction en français

Sommaire

1 Croissance, vieillissement et division cellulaire

La dynamique microscopique des populations de cellules est un phénomène complexe lors duquel de nombreux facteurs rentrent en jeu. L'étude de ces populations au niveau des individus a permis la découverte et la compréhension de nombreux phénomènes, tel que le transfert de gènes entre les espèces. De plus, de nombreuses techniques [BSJ04], comme la fluorescence ou la cytométrie, permettent d'observer les populations de cellules et de récupérer de nombreuses données. L'exploitation de ces données nécessite alors la compréhension des biais liés à la procédure d'échantillonnage dans la population ainsi que la modélisation de la dynamique des population de cellules au niveau microscopique. La construction d'un tel modèle requiert la prise en compte de certaines caractéristiques qui jouent un rôle clé dans les mécanismes cellulaires. Ces populations sont donc dîtes structurées car le cycle de vie et la dynamique de chaque individu dépend de caractéristiques. Nous donnons ici quelques exemples de dépendances entre les différents processus intervenant dans la dynamique de populations de cellules.

Croissance et division cellulaire

Un des mécanismes majeur intervenant dans la dynamique des populations de cellules est la division cellulaire. Il est associé à la croissance cellulaire. Les techniques modernes de suivi de population de cellules à l'aide de microfluides [WRP$^+$10] permettent de suivre la dynamique d'un grand nombre de cellules et ainsi de récupérer une quantité importante de données. Elles consistent à piéger les cellules dans des cavités de petite taille. Lorsque la cellule se divise, la cellule fille se retrouve hors du creux. Elle est alors évacuée par le flux d'une solution qui circule au-dessus des cavités. À l'aide de ces techniques microfluidiques, Wang et al. [WRP$^+$10] ont mis en évidence la stabilité du mécanisme de croissance pour la bactérie E. coli. Une question se pose alors : comment les cellules contrôlent-elles leur taille et l'homéostasie, i.e. le maintien à une valeur bénéfique pour l'individu, de la taille ? Différents modèles ont été proposés pour expliquer le contrôle de la taille par les bactéries : le modèle "timer" dans lequel la durée de vie de chaque individu est fixée ou encore le modèle "sizer" où la taille à la division est fixée. Récemment, les travaux de Taheri-Araghi et al. [TABS$^+$15] ont permis de valider expérimentalement, pour les bactéries E. coli et Bacillus subtilis, le modèle "adder". Dans ce modèle, la taille de chaque individu augmente d'une valeur fixée entre deux divisions. Par ailleurs, les modèles "sizer" et "timer" ont été invalidés par ces mêmes expériences. Plus récemment, le modèle "adder" a été validé pour la levure Saccharomyces cerevisiae [SRA16].

Ces études ont mis en évidence un comportement moyen pour le mécanisme de croissance de cellules mais aussi une certaine variabilité phénotypique au niveau individuel. Celle-ci est en général bénéfique pour les populations de cellules puisqu'elle permet par exemple une plus grande résistance aux perturbations extérieures. Pour un aperçu des différents travaux portant sur les causes de cette hétérogénéité, nous renvoyons le lecteur à [SA02] pour le cas de Saccharomyces cerevisiae et [Ave06] pour un cadre plus général. En particulier, pour le cas de la levure, les taux de croissances et les tailles des individus sont très hétérogènes au sein d'une même population. Les différences de tailles sont dues à la division asymétrique de la levure (voir Figure 1.1) car les cellules filles sont plus petites que leur ancêtre direct.

Vieillissement cellulaire

Un autre mécanisme largement étudié pour les populations de cellules est le vieillissement cellulaire. Pour S. cerevisiae, l'âge se compte habituellement en nombre de fois où la cellule s'est divisée. Dans une population de levures, la part la plus importante est constituée

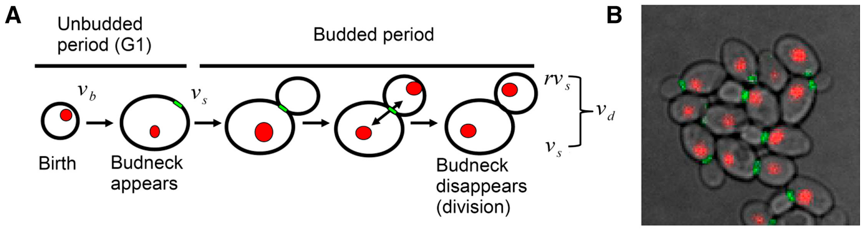

FIGURE 1.1 : Image issue de [SRA16]. A. Cycle de croissance et division de la levure S. cerevisiae. B. Microcolonie de S. cerevisiae. Un marqueur fluorescent colore en rouge le noyau des cellules et en vert l'anneau séparant le bourgeon de la cellule mère, indiquant ainsi une division.

d'individus "jeunes", qui ne se sont jamais divisés. Ceci impacte la fitness de la population car un certain nombre de caractères physiologiques des cellules sont liés à l'âge, comme la taille ou la durée d'un cycle avant division. Kale et al. [KJ96] ont étudié la résistance de S. cerevisiae à l'exposition aux UV en fonction de l'âge des individus. Ils ont observés que la résistance aux UV augmente dans un premier temps avec l'âge, atteint son maximum pour les cellules ayant subis 8 divisions puis décroît avec l'âge. Ainsi, cette augmentation de la résistance n'est pas due à l'augmentation de la taille des cellules avec l'âge car les cellules les plus âgées sont les plus vulnérables à l'exposition aux UV. Elle peut par contre s'expliquer par une régulation dépendante de l'âge de la production des protéines chargées de réparer les dégâts causés par les UV. La résistance aux stress extérieurs dépend donc de l'âge de la cellule. L'âge est donc un facteur important à considérer lors de la modélisation de populations de cellules. Pour plus de détails sur les différents changements au niveau individuel dus au vieillissement cellulaire et leur rôle éventuel dans le vieillissement cellulaire, nous renvoyons le lecteur à [DLJB14].

Les premières études sur le vieillissement cellulaire chez S. cerevisiae ont mis en évidence un phénomène de rajeunissement : le dernier bourgeon d'une cellule avant sa mort est encore capable de se diviser [Joh66, Mül71]. De plus, les cellules filles issues des dernières divisions ont une durée de vie en terme de reproduction moins longue que les cellules issues des premiers bourgeons. Ceci plaide pour l'existence de "facteurs de vieillissement" vérifiant les conditions suivantes [HG08] :

- ils s'accumulent avec l'âge,
- ils sont répartis de manière asymétrique à la division entre la cellule mère et la cellule fille,
- leur absence ou leur réduction entraîne une augmentation de la durée de vie de la cellule,
- l'augmentation du nombre de facteurs de vieillissement entraîne une diminution de la durée de vie de la cellule.

L'ADN ribosomique extrachromosomique (ERC en anglais) satisfait ces conditions. Il est l'un des candidats potentiels pour expliquer une partie du phénomène de vieillissement chez la levure. Les progrès récents dans la collecte de données, notamment grâce aux techniques microfuidiques [CCK17], laissent espérer de nouvelles avancées sur la compréhension du viellissement cellulaire. En effet, le suivi individuel d'une cellule tout au long de sa vie permet de récupérer des données sur la durée de vie des cellules, la dynamique du temps entre deux divisions, mais également sur la morphologie des cellules [LVH$^+$12, XZZ$^+$12]. Les effets du vieillissement cellulaire et ses causes peuvent donc être étudiés plus en détails.

Le vieillissement chez S. cerevisiae semble donc être étroitement lié à la division asymétrique des cellules. Ce phénomène a également été observé pour une bactérie avec division asymétrique par Ackermann et al. [ASJ03]. Pour la bactérie E. coli, la division est symétrique, ce qui rend l'étude expérimentale du vieillissement cellulaire chez cette bactérie plus compliquée. Stewart et al. [SMPT05] ont suivi la croissance de colonies de cellules par fluorescence pendant neuf générations, récupérant ainsi les données physiologiques de chaque cellule. Lors de la division cellulaire chez E. coli, la cellule se coupe en deux parties de tailles égales. Ainsi, chaque cellule fille hérite d'un pôle nouveau, correspondant au pôle issu du milieu de la cellule mère, et d'un pôle plus ou moins âgé. En particulier, la cellule fille qui hérite du pôle anciennement nouveau de la cellule mère serait moins âgée que la cellule qui hérite du pôle qui existait déjà plusieurs divisions auparavant. Les expériences de Stewart et al. [SMPT05] révèlent que les cellules possédant le pôle âgé présentent notamment un taux de croissance plus faible et une probabilité de mort plus importante que les cellules héritant du pôle nouveau. Les résultats de cette étude plaident en faveur de l'existence d'un phénomène de vieillissement cellulaire chez E. coli dû à une asymétrie fonctionnelle comme la localisation dans les pôles de la cellule de certains composants.

Ainsi, l'étude de modèles individus-centrés pour des populations de cellules est motivée par de nombreuses questions. De plus, les nombreuses dépendances entre les différents mécanismes inhérents à la dynamique des populations de cellules nécessitent une approche mathématique rigoureuse du problème.

2 Processus de branchement et populations structurées

Nous nous intéressons dans cette thèse à la dynamique de population sans interactions structurées par un trait. Nous présentons dans cette partie les outils mathématiques pour la modélisation de telles dynamiques.

2.1 Processus de branchement

Les processus de branchements sont des modèles individus-centrés sans interactions basés sur une description probabiliste de la vie des individus. Les domaines d'application de ces processus sont variés : ils peuvent servir à modéliser les populations de végétaux ou d'animaux mais aussi les populations de cellules ou la dynamique de polymères. La présentation qui suit est inspirée de [AN72].

Processus de Galton-Watson multitype

Historiquement, les premiers processus de branchement ont été introduits indépendamment par Bienaymé en 1845 puis par Galton et Watson en 1873 pour étudier la "survie" des patronymes nobles. Ces processus permettent donc d'aborder des questions d'extinction de population mais également d'étudier la stabilisation de caractéristiques spécifiques dans une population telles que la taille ou le taux de croissance.

Pour étudier la dynamique de populations structurées, on associe à chaque individu une caractéristique, souvent appelée trait. Celle-ci peut par exemple correspondre à un allèle d'un gène dont on veut étudier l'évolution dans la population ou à l'état d'un système de régulation des gènes dans le cas de l'étude de l'effet d'un switch moléculaire. Le processus de Galton-Watson associé est appelé processus de branchement multitype. Le nombre de descendants et la transmission du trait d'un individu dépendent alors de son trait.

Soit $p \geqslant 0$. On note $(\mathbf{e_i}, 1 \leqslant i \leqslant p)$ la base canonique de $\mathbb{R}^p$.

Définition. *Un processus de Galton-Watson p-multitype est une chaîne de Markov*

$$(\mathbf{Z}_n = (Z_{n,1}, \ldots, Z_{n,p}), n \geqslant 0)$$

à valeurs $\mathbb{N}^p$ de noyau associé :

$$\mathbb{P}\left(\mathbf{Z}_1 = \mathbf{j} \middle| \mathbf{Z}_0 = \mathbf{e_i}\right) = p_i(\mathbf{j}), \quad \mathbf{j} \in \mathbb{N}^p,\ 1 \leqslant i \leqslant p,$$

et vérifiant pour tout $n \in \mathbb{N}$:

$$\mathbf{Z}_{n+1} = \sum_{i=1}^{p} \sum_{k=0}^{Z_{n,i}} \xi_{k,i},$$

où $(\xi_{k,i}, 1 \leqslant i \leqslant p, k \geqslant 0)$ sont des variables aléatoires indépendantes à valeurs dans $\mathbb{N}^p$ dont la distribution est donnée par :

$$\mathbb{P}\left(\xi_{k,i} = \mathbf{j}\right) = p_i(\mathbf{j}), \quad \forall k \geqslant 0.$$

La i-ième coordonnée de $\mathbf{Z}_n$, notée $Z_{n,i}$, correspond au nombre d'individus à la génération n de type i dans la population. On désigne par $Z_{n,j}^{(i)}$ le nombre de descendant de type j à la génération n d'un individu de type i.

Critère d'extinction. Pour les processus de Galton-Watson avec un seul type, l'extinction de la population survient si le nombre moyen de descendants par individu est inférieur ou égal à 1. Pour une étude plus détaillée des probabilités d'extinction pour les processus de Galton-Watson, nous renvoyons le lecteur à [AN72], Section I.A.5. Dans le cas multitype, le critère d'extinction dépend de la valeur propre maximale de la matrice M dont les coefficients :

$$M_{i,j} = \mathbb{E}\left[Z_{1,j}^{(i)}\right], \quad 1 \leqslant i, j \leqslant p,$$

correspondent au nombre moyen de descendants de type j à la génération 1 d'un individu de type i. On suppose que le processus n'est pas dégénéré i.e. qu'il existe $1 \leqslant i, j \leqslant p$ tels que $M_{i,j} \neq 1$. D'après le théorème de Perron-Frobenius [Ser02], si M est une matrice strictement positive, elle admet une valeur propre maximale $\rho > 0$ associée à un vecteur propre à droite strictement positif $\mathbf{u}$ et un vecteur propre à gauche $\mathbf{v}$. On désigne par $\mathbf{q}$ le vecteur dont la i-ème coordonnée correspond à la probabilité d'extinction d'un processus de Galton-Watson multitype issu d'un individu de type i. On a alors le résultat suivant :

Théorème. *Avec les notations précédentes :*

- *Si $\rho \leqslant 1$, $\mathbf{q} = \mathbf{1}$ et le processus s'éteint presque sûrement.*
- *Si $\rho > 1$, $\mathbf{q} < \mathbf{1}$ et $\mathbf{q}$ est l'unique solution dans $\mathbb{R}^p$ de $\mathbf{f}(\mathbf{q}) = \mathbf{q}$, où $\mathbf{f}$ désigne la fonction génératrice de $\mathbf{Z}$.*

La preuve de ce résultat figure dans [Har63], Section II.7. sous l'hypothèse plus générale suivante : il existe un entier $N > 0$ tel que M^N soit strictement positive. Le processus est dit sous-critique lorsque $\rho < 1$, sur-critique si $\rho > 1$ et critique dans le cas où $\rho = 1$.

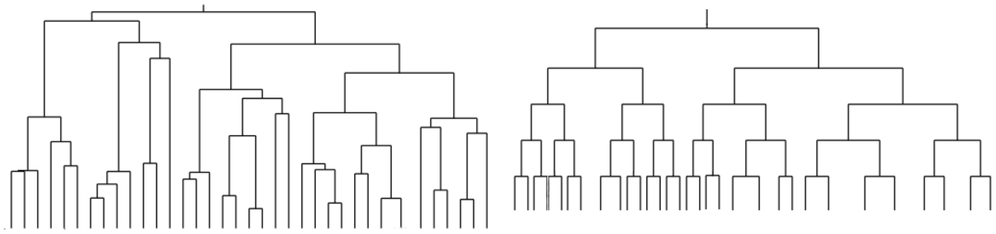

Figure 2.1 : À gauche : Généalogie descendante d'une population avec générations chevauchantes. À droite : Généalogie descendante d'une population avec reproduction saisonnière.

Comportement asymptotique. Les premiers travaux portant sur l'étude du comportement asymptotique des processus de Galton-Watson dans le cas sur-critique sont ceux de Harris (1963), de Kesten et Stigum (1966) et d'Athreya (1970). Introduisons tout d'abord la condition suivante :

$$\mathbb{E}\left[Z_{1,j}^{(i)} \log Z_{1,j}^{(i)}\right] < \infty, \text{ pour tout } 1 \leqslant i, j \leqslant p \tag{$\ast$}$$

On a alors le théorème suivant, dû à Kesten et Stigum :

Théorème. *On suppose que le processus n'est pas dégénéré et que* M *est strictement positive. Alors, si* $\rho > 1$*, on a :*

$$\lim_{n \to +\infty} \frac{\mathbf{Z}_n}{\rho^n} = \mathbf{v} W, \; p.s.,$$

où W *est une variable aléatoire positive. De plus,* $\mathbb{P}(W > 0) > 0 \Leftrightarrow (\ast)$ *et si* $(\ast)$ *est vérifiée, on a pour tout* $1 \leqslant i \leqslant p$ *:*

- $\mathbb{E}\left[W \middle| \mathbf{Z}_0 = \mathbf{e}_i\right] = u_i,$
- $\mathbb{P}(W = 0 | \mathbf{Z}_0 = \mathbf{e}_i) = q_i,$

avec $\mathbf{e}_i$ *le* i*-ème vecteur de base de* $\mathbb{R}^p$*.*

La preuve initiale de ce théorème est donnée dans [KS66]. Les proportions asymptotiques des individus de chaque type dans la population sont donc déterministes, données par le vecteur propre à gauche de M. En temps long, le seul aléa provient des fluctuations de la taille totale de la population. Une autre preuve de ce théorème a été proposée par Lyons, Pemantle et Peres [LPP95] basé sur la notion d'arbre biaisé par la taille détaillée en Section 2.2.

Dans les cas critique ou sous-critique, des résultats de convergence du processus conditionné à la non-extinction ont été prouvés par Joffe et Spitzer. Nous renvoyons le lecteur aux Sections V.4 et V.5 de [AN72] pour le détail des résultats.

Processus de Markov branchants en temps continu et dimension finie

Pour l'étude des dynamiques de populations au sein desquelles la reproduction n'est pas saisonnière telles que les populations de cellules, le modèle de Galton-Watson décrit précédemment n'est pas adéquat. Les processus de Markov branchants en temps continu sont plus adaptés à la modélisation de populations dont le générations se chevauchent (voir Figure 2.1). Les premiers travaux sur la dynamique des populations en temps continu à l'aide de processus de branchement sont dus à Yule et portaient sur des processus de naissance et mort. Nous nous intéressons dans cette section aux processus avec un nombre fini de types, pouvant modéliser par exemple un état phénotypique pour l'étude d'un switch dans une population de cellules. La définition d'un processus de Markov branchant en temps continu est analogue à la définition en temps discret.

Définition. *Un processus de Markov branchant p-multitype en temps continu* $(\mathbf{Z}(t), t \geqslant 0)$ *est un processus de Markov à valeurs dans* $\mathbb{N}^p$ *vérifiant pour tout* $t, s \geqslant 0$ *:*

$$\mathbf{Z}(t+s) = \sum_{j=1}^{p} \sum_{k=1}^{Z_j(s)} \mathbf{Z}^{k,(j)}(t),$$

où $Z_j(s)$ *est le* j*-ième coordonnée de* $\mathbf{Z}(s)$ *et* $(\mathbf{Z}^{k,(j)}(t), t \geqslant 0)$ *sont des processus indépendants conditionnellement à* $\mathbf{Z}(s)$ *de loi identique à celle de* $(\mathbf{Z}(t), t \geqslant 0)$ *issu d'un individu de type* j*.*

Les caractéristiques infinitésimales du processus de Markov branchant en temps continu sont :

- les paramètres des lois exponentielles qui déterminent les durées de vie des individus en fonction de leur type,
- les distributions $\mathbf{p}^{(i)} = \left(p_1^{(i)}, \ldots, p_p^{(i)}\right)$ sur $\mathbb{N}^p$ du nombre de descendants de chaque type d'un individu de type i, pour $1 \leqslant i \leqslant p$.

En fonction de ces caractéristiques, les équations de Kolmogorov décrivent la dynamique infinitésimale du processus. Nous renvoyons le lecteur à [AN72] Section V.7. pour le détail de ces équations dans le cas d'un processus de Markov branchant p-multitype.

L'analogue de la matrice M dans le cas continu est la matrice $(M(t), t \geqslant 0)$ dont les coefficients sont donnés pour tout $1 \leqslant i, j \leqslant p$ par :

$$M_{i,j}(t) = \mathbb{E}\left[Z_j^{(i)}(t)\right],$$

où $Z_j^{(i)}(t)$ correspond au nombre de descendants de type j au temps t dans une population issu d'un individu de type i. On suppose qu'il existe un temps $t_0 > 0$ tel que pour tout $1 \leqslant i, j \leqslant p$, $M_{i,j}(t_0) > 0$. Alors, d'après le théorème de Perron-Frobenius, il existe une valeur propre maximale $\rho(t_0) > 0$ pour $M(t_0)$. De plus, on a pour tout $t, s \geqslant 0$:

$$\frac{M(t+s) - M(t)}{s} = \frac{M(t)(M(s) - I)}{s} \xrightarrow[s \to 0]{} M(t)\frac{d}{ds}M(s)|_{s=0}.$$

Il existe donc une matrice A telle que pour tout $t \geqslant 0$:

$$M(t) = \exp(At).$$

Alors, A et $M(t)$ ont les mêmes vecteurs propres et si on désigne par $\lambda_1, \ldots, \lambda_p$ les valeurs propres complexes de A, les valeurs propres de $M(t)$ sont données par $\exp(\lambda_1 t), \ldots, \exp(\lambda_p t)$. La valeur propre λ_1 de partie réelle maximale de A est réelle. Elle est l'analogue de ρ dans l'étude en temps discret et est appelée paramètre de Malthus. Elle correspond à la vitesse exponentielle de croissance de la population. Les résultats de convergence du processus sont alors similaires à ceux du temps discret. On a en particulier le théorème suivant :

Théorème. *Si le processus* $\mathbf{Z}$ *est non dégénéré et s'il existe* $t_0 \geqslant 0$ *tel que* $M_{i,j}(t_0) > 0$ *pour tout* $1 \leqslant i, j \leqslant p$*, alors on a :*

$$\mathbf{Z}(t)e^{-\lambda_1 t} \xrightarrow[t \to +\infty]{} \mathbf{v}W \quad p.s.$$

où $\mathbf{v}$ *est le vecteur propre à gauche (strictement positif) associé à la valeur propre* λ_1 *de* A *et* W *est une variable aléatoire positive. De plus, on a équivalence entre les propositions suivantes :*

- *il existe* $1 \leqslant k \leqslant p$ *tel que* $\mathbb{P}(W > 0 \big| \mathbf{Z}(0) = \mathbf{e_k}) > 0$*,*
- *pour tout* $1 \leqslant i, j \leqslant p$*,* $\mathbb{E}\left[\xi_{i,j} \log \xi_{i,j}\right] < \infty$*, avec* $(\xi_{i,1}, \ldots, \xi_{i,p})$ *distribué selon* $\mathbf{p^{(i)}}$*.*

Le poids exponentiel correspond à la croissance de la population et les répartitions des types à la limite sont à nouveau donnés un vecteur propre à gauche.

Processus âge-dépendants

Les processus âge-dépendant ou processus de Bellman-Harris constituent une catégorie particulière des processus de branchement. Ils permettent de modéliser des populations où la durée de vie des individus n'est pas distribuée selon une loi exponentielle. Dans ce cas, le processus qui compte le nombre d'individus en vie au temps t n'est plus un processus de Markov et les résultats des parties précédentes ne s'appliquent pas. Dans un souci de clarté, nous nous limitons ici à la description des processus de Bellman-Harris avec un seul type. Nous renvoyons le lecteur à [Har63] Section VI.28 pour une généralisation au cas multitype.

On désigne par $(Z(t), t \geqslant 0)$ le nombre d'individus en vie dans la population au temps t. Les durées de vie des individus sont des variables aléatoires indépendantes identiquement distribuées de fonction de répartition G. À sa mort, chaque individu est remplacé par un nombre aléatoire de descendants tiré suivant la distribution $(p_k, k \geqslant 0)$. L'objet mathématique au cœur de l'étude des processus âge-dépendants est encore la fonction génératrice :

$$F(s,t) = \sum_{k=0}^{\infty} \mathbb{P}(Z(t) = k) s^k.$$

Cette fonction vérifie un équation intégrale qui résume la dynamique du processus. En effet, si on note l_0 la durée de vie de l'individu initial on a pour tout $t \geqslant 0$ et $|s| \leqslant 1$:

$$F(s,t) = \mathbb{E}\left(s^{Z(t)}\right) = s\mathbb{P}\left(l_0 > t\right) + \mathbb{E}\left(s^{Z(t)} \mathbf{1}_{l_0 \leqslant t}\right).$$

De plus, si on note ν_0 la variable aléatoire correspondant au nombre de descendants du premier individu, on obtient :

$$\mathbb{E}\left(s^{Z(t)} \mathbf{1}_{l_0 \leqslant t}\right) = \int_0^t \mathbb{E}\left(s^{\sum_{k=1}^{\nu_0} Z_k(t-u)}\right) dG(u),$$

où Z_k sont des processus de Bellman-Harris indépendants. Finalement, on a :

$$F(s,t) = s(1 - G(t)) + \int_0^t h(F(s, t-u)) dG(u),$$

où h désigne la fonction génératrice associée à la loi de reproduction $(p_k, k \geqslant 0)$. De cette équation, on déduit une équation de renouvellement pour le nombre moyen d'individu $M(t)$ dans la population au temps t :

$$M(t) = 1 - G(t) + m \int_0^t M(t-u) dG(u),$$

où $m = h'(1)$ est le nombre moyen de descendants d'un individu. Si on suppose de plus que G n'est pas une distribution sur un réseau (voir Définition VI.17.1 dans [Har63]), le comportement asymptotique de M est alors donné par le théorème suivant :

Théorème. *Supposons que* $m > 1$. *Soit* $\alpha > 0$ *telle que :*

$$m \int_0^\infty e^{-\alpha t} dG(t) = 1.$$

Alors,

$$M(t) \sim \frac{m-1}{\alpha m^2 \int_0^\infty t e^{-\alpha t} dG(t)} e^{\alpha t}, t \to \infty.$$

En particulier, la croissance est exponentielle à vitesse α où α est le paramètre de Malthus. Une expression asymptotique du second moment peut également être obtenue [Har63].

Les résultats obtenus dans cette thèse portent sur une généralisation des processus de Bellmann-Harris au cas où la durée de vie ne dépend pas seulement de l'âge de l'individu mais d'un trait évoluant suivant une certaine dynamique. Notons que dans ce cas, le formalisme des fonctions génératrices n'est plus adapté au problème. D'autre part, la transmission du trait joue un rôle crucial dans l'étude de ces processus. En particulier, le phénomène de renouvellement est propre au processus structuré en âge, ceci car chaque individu nait à l'âge 0.

2.2 Épine et arbre biaisé par la taille

Les travaux de Harris, Kesten-Stigum et Athreya sur les processus de Galton-Watson multitypes sur-critiques mentionnés dans la section précédente ont permis d'exhiber la répartition des types dans la population en temps long à travers le vecteur propre à gauche de la matrice de reproduction M. La compréhension des mécanismes qui produisent cette répartition est cruciale pour répondre aux questions liées au vieillissement cellulaire détaillées en Section 1.

Les arbres biaisés par la taille constituent une première étape dans la compréhension des biais liés à l'échantillonnage. Ces objets ont été considérés par Kallenberg [Kal77], Chauvin et Rouault [CR88] puis Chauvin, Rouault et Wakolbinger [CRW91]. Lyons, Pemantle et Peres [LPP95] s'appuient sur les arbres biaisés par la taille pour proposer une nouvelle preuve du théorème de Kesten-Stigum. À chaque nœud de l'arbre, on associe un label grâce aux notations d'Ulam, Harris et Neveu. Soit

$$\mathcal{U} = \bigcup_{n \in \mathbb{N}} (\mathbb{N}^*)^n,$$

l'ensemble des labels. La racine de l'arbre est notée $\emptyset$. Lorsqu'un individu $u \in \mathcal{U}$ meurt, ses K descendants ont pour labels $u1, \ldots, uK$. Ces notations permettent en particulier de conserver la généalogie d'un individu grâce à son label.

On introduit un nouveau type d'arbre aléatoire : les arbres de Galton-Watson biaisés par la taille. Pour cela, on définit la distribution biaisée par la taille $(\widehat{p}_k, k \geqslant 0)$ par :

$$\widehat{p}_k = \frac{k p_k}{m}, \quad k \geqslant 0,$$

où m est la moyenne de la loi de reproduction $(p_k, k \geqslant 0)$. Les grandes fratries sont ainsi favorisées. On construit alors l'arbre de Galton-Watson biaisé par la taille de la manière suivante :

- on commence avec un individu v_0,
- il a un nombre de descendants distribué suivant $\widehat{p}$,

- on choisit l'un de ses descendants uniformément au hasard, disons v_1.
- les sous-arbres issus des autres descendants sont des arbres de Galton-Watson,
- on répète les mêmes étapes avec v_1 au lieu de v_0.

La suite de labels $(v_k, k \geqslant 0)$ forme l'épine. Soit a un arbre de Galton-Watson de longueur n et v un individu à la génération n dans a. On a :

$$\mathbb{P}\left(\left(\widehat{\mathcal{A}}_n, \xi\right) = (a, v)\right) = \frac{1}{m^n}\mathbb{P}\left(\mathcal{A}_n = a\right). \tag{1}$$

avec $\mathcal{A}_n$ et $\widehat{\mathcal{A}}_n$ les variables aléatoires correspondant respectivement à un arbre de Galton-Watson à la génération n et à un arbre de Galton-Watson biaisé à la génération n et ξ la variable aléatoire correspondant à l'épine, c'est-à-dire au choix d'une lignée distinguée. Remarquons en particulier que la mesure ainsi définie sur les arbres de Galton-Watson jusqu'à la génération n avec une lignée distinctive n'est pas une mesure de probabilité car sa masse est de m^{-n}. On note Z_n le nombre d'individus à la génération n dans a. On a alors l'égalité suivante permettant de relier la loi des arbres de Galton-Watson sous la mesure biaisée à celle sous la mesure non-biaisée :

$$\mathbb{P}\left(\widehat{\mathcal{A}}_n = a\right) = \frac{Z_n}{m^n}\mathbb{P}\left(\mathcal{A}_n = a\right). \tag{2}$$

Comme $W_n := Z_n m^{-n}$ est une martingale, il est naturel d'effectuer le changement de mesure correspondant et de considérer la mesure associée au processus biaisé. À l'aide de cette nouvelle mesure, Lyons, Peres et Pemantle [LPP95] et Lyons [Lyo97] proposent notamment une preuve alternative au théorème de Kesten-Stigum.

Cette décomposition à l'aide de l'épine s'étend aux processus de Galton-Watson multitypes en considérant une loi de reproduction pondérée par le vecteur propre à gauche de la matrice de reproduction [KLPP97, BK04, Ath00]. Le cas des processus âge-dépendants a été traité par Olofsson [Olo98, Olo09]. Des généralisations en temps continu ont été proposés par exemple par Georgii et Baake [GB03] pour les processus branchants multitypes afin d'étudier les caractéristiques ancestrales des individus "typiques" dans la population. Enfin, la notion d'épine a été développée et utilisée pour l'étude de divers processus. Nous renvoyons le lecteur aux travaux de Kyprianou [Kyp04] pour l'étude du comportement asymptotique et de l'unicité des propagations d'onde de l'équation KPP. Pour les processus de fragmentation, l'utilisation de l'épine, également appelée fragment marqué, a été dévelopée par Bertoin et Rouault [BR05] et par Harris, Knobloch et Kyprianou [HKK10].

Hardy et Harris [HH09] ont généralisé la construction de l'épine au cas de traits infinidimensionels, notamment pour l'étude du mouvement brownien branchant. Tout comme Lyons, Pemantle et Peres [LPP95], ils considèrent l'ensemble $\widetilde{\mathcal{T}}$ des arbres de Galton-Watson (en temps continu) avec une lignée distinctive, à savoir l'épine. La suite des labels des individus formant l'épine est notée ξ et le label de l'individu correspondant à l'épine au temps t est noté ξ_t. L'ensemble $\widetilde{\mathcal{T}}$ est muni de la mesure de probabilité P canonique pour le processus de Markov branchant $(Z_t, t \geqslant 0)$ dont les caractéristiques sont les suivantes :

- la dynamique de chaque individu dans la population est donnée par une copie indépendante d'un processus de Markov $(X_t, t \geqslant 0)$ sur un espace mesurable $\mathcal{X}$. On note X_t^u le trait de u à l'instant t.
- un individu u meurt au temps t au taux $B(X_t^u)$ où $B : \mathcal{X} \to (0, \infty)$ est une fonction mesurable. On note β_u le temps de mort de u.

- lorsqu'un individu u meurt à la position x, il est remplacé par $A_u + 1$ individus à cette même position, avec A_u une copie indépendante de $A(x)$ et $\mathbb{P}(A(x) = k) = p_k(x)$ pour tout $k \geqslant 0$. On note $m(x) = \sum_{k \geqslant 0} kp_k(x)$ la moyenne de $A(x)$.
- les individus évoluent ensuite indépendamment les uns des autres.

Ce modèle très général est très proche de celui considéré dans les Chapitres 1 et 2. Il permet en particulier de modéliser des populations structurées spatialement. Notons que dans ce modèle les individus ne peuvent pas mourir sans laisser de descendants et que le branchement est local c'est-à-dire que les descendants naissent à l'endroit où meurt leur parent.

On définit plusieurs filtrations sur l'ensemble $\widetilde{\mathcal{T}}$ des arbres marqués avec épine qui permettent de séparer les différentes informations dont on dispose. Nous donnons ici une définition informelle de ces filtrations et renvoyons le lecteur à [HH09] pour leur définition rigoureuse.

- $\mathcal{F}_t$ contient toutes les informations sur le processus de branchement jusqu'au temps t mais ne connait rien de l'épine,
- $\widetilde{\mathcal{F}}_t$ est construite à partir de $\mathcal{F}_t$ en ajoutant les informations sur l'épine : elle contient donc toutes les informations sur le processus de branchement et sur l'épine jusqu'au temps t,
- $\mathcal{G}_t$ est la filtration qui correspond au mouvement de l'épine. Elle ne contient pas d'information sur la lignée à laquelle l'épine correspond dans l'arbre.

Pour $x \in \mathcal{X}$, on définit P^x la mesure de probabilité sur $(\widetilde{\mathcal{T}}, \mathcal{F}_\infty)$ telle que $\left(\widetilde{\mathcal{T}}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t \geqslant 0}, P^{\cdot}\right)$ soit le modèle canonique pour le processus de branchement Z avec

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{t \geqslant 0} \mathcal{F}_t\right).$$

Comme précédemment, Hardy et Harris définissent une mesure de probabilité $\widetilde{P}^x$ à partir de P^x pour laquelle l'épine correspond à une lignée dans l'arbre. Pour toute fonction $\widetilde{\mathcal{F}}_t$-mesurable f, on pose

$$\int_{\widetilde{\mathcal{T}}} f d\widetilde{P}^x = \int_{\widetilde{\mathcal{T}}} \sum_{u \in V_t} f_u \prod_{v \leqslant u} \frac{1}{1 + A_v} dP^x,$$

avec

$$f = \sum_{u \in V_t} f_u \mathbf{1}_{\{\xi_t = u\}},$$

où V_t désigne l'ensemble des individus en vie au temps t et pour tout $u \in V_t$, f_u est un fonction mesurable par rapport à $\mathcal{F}_t$. La différence par rapport aux travaux de Lyons et al. [LPP95, Lyo97] est que cette mesure est une mesure de probabilité. Sous cette mesure, la dynamique de l'épine entre les sauts suit le processus de Markov $(X_t, t \geqslant 0)$ et à la mort du représentant de l'épine, l'un de ses descendants est choisi uniformément au hasard pour le remplacer. Ils considèrent ensuite un changement de mesure à l'aide d'une martingale. Il permet de prendre en compte les biais présents dans l'arbre, dus notamment à la croissance de la population, de la même manière que Lyons, Pemantle et Peres [LPP95] (voir (2)). Une des nouveautés apportées par les travaux de Hardy et Harris est la généralisation de la martingale considérée pour le changement de mesure. Soit $\zeta(t)$ une $\mathcal{G}_t$-martingale qui ne dépend que de l'épine et soit $\widetilde{\zeta}(t)$ la martingale donnée par :

$$\widetilde{\zeta}(t) = \prod_{u<\xi_t} \frac{1+A_u}{1+m\left(\xi_{S_u}\right)} \prod_{v<\xi_t} \left(1+m\left(\xi_{S_v}\right)\right) e^{-\int_0^t B(\xi_s)m(\xi_s)ds} \zeta(t).$$

Le changement de mesure correspondant à celui considéré dans le cas discret (2) est alors donné dans le cas continu par :

$$\left.\frac{d\widehat{P}}{d\widetilde{P}}\right|_{\widetilde{\mathcal{F}}_t} = \frac{\widetilde{\zeta}(t)}{\widetilde{\zeta}(0)}.$$

Ainsi, $\widehat{P}$ est une nouvelle mesure de probabilités sur $\left(\widetilde{\mathcal{T}}, \widetilde{\mathcal{F}}_\infty\right)$. Le comportement de l'épine sous la mesure $\widehat{P}$ est alors l'analogue du comportement de l'épine lors de la construction d'un arbre de Galton-Watson biaisé par la taille en ce qui concerne la reproduction. Des biais supplémentaires apparaissent à cause du chevauchement des générations et de la dynamique des individus entre les sauts. Le premier produit dans l'expression de la martingale $\widetilde{\zeta}$ permet d'introduire le biais par la taille et le second entraine une accélération du temps : les durées de vie des individus le long de l'épine sont plus courtes car le taux de division $B(\cdot)$ est remplacé par $(1+m(\cdot))B(\cdot)$. Pour une description précise du comportement de l'épine sous $\widehat{P}$, nous renvoyons le lecteur à [HH09]. Les auteurs donnent également l'analogue de la décomposition en épine initialement prouvée dans [Lyo97] qui s'avère utile notamment pour l'étude du comportement asymptotique de martingales additives.

Le cadre spectral est propice à l'utilisation des outils liés à l'épine car il permet d'exhiber des martingales simples. Considérons un processus de Markov branchant où la dynamique des individus suit un processus de Markov $(X_t, t \geqslant 0)$ de générateur $\mathcal{G}$, où le branchement est binaire, local et a lieu au taux B. S'il existe une fonction ϕ et $\lambda \in \mathbb{R}$ tels que $(\mathcal{G}+B)\phi = \lambda\phi$, alors,

$$W_t := \langle \phi, Z_t \rangle e^{-\lambda t},$$

est une $\sigma(Z_t, t \geqslant 0)$-martingale. L'existence d'éléments propres permet donc notamment de simplifier l'étude asymptotique du processus de branchement Z en considérant son comportement contre une fonction propre. Pour plus de détails sur ces techniques, nous renvoyons le lecteur aux travaux d'Engländer, Harris et Kyprianou [EHK10] ou de Cloez [Clo17].

2.3 Représentation par des processus à valeurs mesures et équation de croissance-fragmentation

Dans les modèles que nous considérons, chaque individu est caractérisé par un vecteur de traits phénotypiques évoluant dans $\mathcal{X} \subset \mathbb{R}^d$. Le processus à valeurs mesures associé à la population permet alors de suivre la dynamique de l'ensemble de la population tout en considérant les dynamiques individuelles. On définit $\mathcal{M}_F(\mathcal{X})$ l'ensemble des mesures positives finies sur $\mathcal{X}$ et $\mathcal{M}_P(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ le sous-ensemble des mesures ponctuelles sur $\mathcal{X}$:

$$\mathcal{M}_P(\mathcal{X}) = \left\{ \sum_{i=1}^n \delta_{x_i}, n \geqslant 0, x_1, \ldots, x_n \in \mathcal{X} \right\},$$

où δ_x représente la masse de Dirac au point x. À l'aide de ce formalisme, nous représentons la dynamique microscopique de la population par le processus Z défini par :

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

où V_t représente la population au temps t et X_t^u le trait de l'individu u au temps t.

Les processus à valeurs mesures ont été largement utilisés pour l'étude de populations structurées. Nous mentionnons ici le travail de Fournier et Méléard [FM04] sur la dynamique d'une population statique (par exemple des plantes) structurée spatialement avec de la compétition. De nombreuses études de population structurées reposent sur les techniques initiées dans [FM04], comme par exemple les travaux de Tran sur les processus structurés en âge [Tra06].

La représentation par processus à valeurs mesures permet de définir le processus de branchement structuré comme l'unique solution d'une équation différentielle stochastique (voir Chapitre 1, Théorème 1.2.2). Considérons l'exemple d'une population structurée en taille, au sein de laquelle les individus croissent exponentiellement à taux $a > 0$ et se divisent à taux B en deux parties égales. Le processus à valeurs mesures associé $(Z_t, t \geqslant 0)$ est solution de l'équation suivante :

$$\begin{aligned} \langle Z_t, f \rangle = & \langle Z_0, f \rangle + \int_0^t \int_0^\infty axf'(x) Z_s(dx) ds && (3) \\ & + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+} \mathbf{1}_{\{u \in V_s,\ \theta \leqslant B(X_s^u)\}} \left(2f\left(\frac{X_s^u}{2} \right) - f(X_s^u) \right) M(ds, du, d\theta), && (4) \end{aligned}$$

pour toute fonction mesurable $f : \mathbb{R}_+ \to \mathbb{R}$ et tout $t \geqslant 0$, où $M(ds, du, d\theta)$ est une mesure ponctuelle de Poisson sur $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ d'intensité $ds \otimes du \otimes d\theta$. La première intégrale traduit la croissance exponentielle des individus entre les divisions. La deuxième intégrale correspond aux divisions qui arrivent à taux B et lors desquelles un individu de trait x est remplacé par 2 individus de trait $x/2$. Ces équations décrivent la dynamique du processus et permettent l'utilisation des outils du calcul stochastique. On peut ainsi calculer diverses quantités associées à la population comme par exemple la variance du nombre d'individus dans la population (voir Chapitre 2, Lemme 2.4.6).

Un objet crucial pour l'étude de la dynamique du trait d'un individu typique détaillée dans le Chapitre 1 est le semi-groupe du premier moment donnée pour tout $s \leqslant t$, $x \in \mathcal{X}$ et toute fonction mesurable $f : \mathcal{X} \to \mathbb{R}$ par :

$$R_{s,t} f(x) = \mathbb{E}\left[\sum_{u \in V_t} f(X_t^u) \middle| Z_s = \delta_x \right]. \tag{5}$$

Il s'agit d'un semi-groupe non-conservatif. Dans le cas de l'exemple précédent, il vérifie l'équation suivante :

$$R_{s,t} f(x) = f(x) + \int_s^t \int_0^\infty axf'(x) R_{s,r}(dx) dr + \int_s^t \int_0^\infty B(x) \left(2f\left(\frac{x}{2} \right) - f(x) \right) R_{s,r}(dx) dr,$$

pour toute fonction mesurable $f : \mathbb{R}_+ \to \mathbb{R}$ et tout $0 \leqslant s \leqslant t$. On peut définir le générateur $\mathcal{A}$ associé au semi-groupe de premier moment et dont l'expression dans le cas de l'exemple précédent est donnée par :

$$\mathcal{A} f(x) = axf'(x) + B(x) \left(2f\left(\frac{x}{2} \right) - f(x) \right).$$

Tout comme pour l'équation différentielle stochastique (3), on peut lire la dynamique des individus sur le générateur : le premier terme correspond à la croissance exponentielle des individus et le second correspond au mécanisme de division.

Pour l'étude asymptotique des processus de branchement effectuée dans le Chapitre 2, l'utilisation des processus à valeurs mesures permet également de simplifier les résultats de convergence car le processus ainsi que sa limite appartiennent à l'ensemble des mesures ponctuelles.

Ces processus à valeurs mesures sont la version microscopique de modèles macroscopiques. Ainsi, de nombreux travaux portent sur le lien entre les différentes échelles de modèles [FM04, BT11, Clo17]. Dans le cas des processus de branchement, la limite grande population de ces processus est déterministe, donnée par le semi-groupe de premier moment défini précédemment (5), ceci grâce à la propriété de branchement. L'étude en grande population des processus considérés dans cette thèse se ramène donc à l'étude des équations croissance-fragmentation associées.

Notons $n(t,dx)$ la distribution vérifiant pour tout fonction mesurable $f : \mathcal{X} \to \mathbb{R}$:

$$\langle n(t,\cdot), f\rangle = R_{0,t}f(x) = \mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\right], \quad \text{and } n(0,dy) = n^{\text{in}}(dy).$$

On considère le cas où $n(t,\cdot)$ est absolument continue par rapport à la mesure de Lebesgue, ce qui est vérifié si la mesure initiale n^{in} est elle-même à densité. La fonction $n(t,x)$ correspond à la densité d'individus de trait x dans la population à l'instant t. En passant à l'espérance dans l'équation différentielle stochastique (3), on obtient l'équation de croissance-fragmentation suivante d'inconnue n :

$$\partial_t n(t,x) + a\partial_x(xn(t,x)) + B(x)n(t,x) = 4B(2x)n(t,2x).$$

Plus généralement, on appelle équation de croissance-fragmentation une équation d'inconnue n de la forme :

$$\partial_t n(t,x) + \partial_x(v(x)n(t,x)) + B(x)n(t,x) = \int_0^1 B\left(\frac{x}{z}\right) n\left(t,\frac{x}{z}\right)\frac{\rho(dz)}{z}, \tag{6}$$

avec conditions initiales $v(0)n(t,0) = 0$ pour tout $t > 0$ et $n(0,x) = n^{\text{in}}(x)$ pour tout $x \geqslant 0$ et où v représente le taux de croissance des individus, B le taux de division et ρ le noyau de fragmentation.

Les équations de croissance-fragmentation apparaissent dans de nombreux domaines comme l'étude de croissance de populations bactériennes mais aussi l'étude des polymères ou encore le protocole de contrôle de transmission (ou TCP). Ces équations initialement étudiées d'un point de vue déterministe et donc macroscopique correspondent aux limites en grande population des processus de branchement. Les équations de croissance-fragmentation ont fait l'objet de nombreux travaux [Per07], portant notamment sur l'existence de solutions et l'analyse du comportement asymptotique de ces solutions. L'analyse spectrale du semi-groupe associé est la clé du problème. En particulier, dans le cas d'un opérateur compact, le théorème de Krein-Rutmann assure l'existence d'une valeur propre associée à un vecteur propre positif [Du06]. En 2016, Mischler et Scher [MS16] ont développé et utilisé cette théorie, pour montrer la convergence exponentielle des solutions vers des profils asymptotiques pour des taux de fragmentation dans une certaine classe.

L'étude asymptotique des solutions repose sur l'existence de solutions aux problèmes aux valeurs propres, direct et dual, associé à (6) d'inconnus (λ, N, ϕ) donnés par :

$$\begin{cases} \lambda N(x) + (v(x)N(x))' + B(x)N(x) = \int_0^1 B\left(\frac{x}{z}\right) N\left(\frac{x}{z}\right)\frac{\rho(dz)}{z}, \\ N \geqslant 0,\ \int_0^\infty N(x)dx = 1, \end{cases} \tag{7}$$

et

$$\begin{cases} \lambda\psi(x) - v(x)\psi'(x) + B(x)\psi(x) = B(x)\int_0^1 \psi(zx)\rho(dz), \\ \psi \geqslant 0, \int_0^\infty N(x)\psi(x)dx = 1. \end{cases}$$

On a alors l'équation de conservation suivante [MS16] :

$$\int_0^\infty n(t,x)\psi(x)dx = e^{\lambda t}\int_0^\infty \psi(x)n^{\mathrm{in}}(x)dx.$$

La première valeur propre λ est le paramètre de Malthus, mentionnée précédemment dans la Section 2.1. Elle correspond au taux de croissance exponentielle de la population. Nous renvoyons le lecteur à [Oli17] pour une étude de l'impact de la variabilité des caractéristiques individuelles dans une population sur le paramètre de Malthus. Asymptotiquement, les solutions de l'équation (6) vérifient lorsque t tend vers l'infini :

$$n(t,x) \sim e^{\lambda t}\left\langle n^{\mathrm{in}}, \psi\right\rangle N(x),$$

pour une certaine norme pondérée [MMP05]. Ceci nous indique qu'en temps long, le profil en taille de la population est donné par N et ne dépend pas de la condition initiale. Nous retrouvons ce phénomène "d'oubli" de la condition initiale pour les modèles probabilistes dans la Section 2.5.

L'existence de solutions aux problèmes aux valeurs propres pour des équations de croissance-fragmentation a été largement étudiée. Citons par exemple les travaux de Diekmann, Heijmans et Thieme [DHT84], Perthame et Ryzhik [PR05], Michel [Mic06]. Perthame et Ryzhik [PR05] ont également montré la convergence à vitesse exponentielle de la solution de l'équation de croissance-fragmentation vers le profil stationnaire dans le cas d'une division binaire et d'une croissance linéaire. Pour des travaux portant sur la vitesse convergence dans des cadres plus généraux, nous renvoyons le lecteur à [MMP05, BG17].

Récemment, Bouguet [Bou16], Bertoin et Watson [BW17] ont étudié ces équations de croissance-fragmentation en adoptant un point de vue probabiliste. Cette nouvelle approche leur permet en particulier d'exhiber des estimées fines relatives au comportement asymptotique des solutions des équations de croissance-fragmentation, à l'aide de critères de type Foster-Lyapunov (voir Section 2.5). Ces travaux sont ainsi en lien avec l'approche par ergodicité des lignées ancestrales proposée dans le Chapitre 2.

2.4 Formule Many-to-One

L'une des formules clés apportées par l'épine est la formule Many-to-One. Elle permet de résumer la dynamique de l'ensemble de la population grâce à la dynamique de l'épine. On considère le processus de Markov branchant $(Z_t, t \geqslant 0)$ dont les caractéristiques sont les suivantes :

- la dynamique de chaque individu dans la population est donnée par une copie indépendante d'un processus de Markov $(X_t, t \geqslant 0)$ sur un espace mesurable $\mathcal{X}$ de générateur infinitésimal $\mathcal{G}$. On note X_t^u le trait de u à l'instant t.

- un individu u meurt au temps t au taux $B(X_t^u)$ où $B : \mathcal{X} \to (0,\infty)$ est une fonction mesurable. On note β_u le temps de mort de u.

- lorsqu'un individu u de trait x meurt, il est remplacé par A_u individus, avec A_u une copie indépendante de $A(x)$ et $\mathbb{P}(A(x) = k) = p_k(x)$ pour tout $k \geqslant 0$. On note $m(x) = \sum_{k\geqslant 0} kp_k(x)$ la moyenne de $A(x)$.

- le trait à la naissance des k descendants d'un individu de trait x sont distribués selon le noyau $P^k(x,dy)$. Pour $1 \leqslant i \leqslant k$, on note $P_i^k(x,dy) = P^k\left(x, \mathcal{X}^{i-1} \times dy \times \mathcal{X}^{k-i}\right)$ la $i$ème marginale de P^k.
- les individus évoluent ensuite indépendamment les uns des autres.

Ce processus, plus général que celui présenté dans la Section 2.2, est l'objet d'étude de cette thèse. À partir du semi-groupe de premier moment $(R_{s,t}, s \leqslant t)$ introduit en (5) et de son générateur associé, on obtient la formule Many-to-One classique, encore appelée formule de Feynman-Kac [DM04] :

$$R_{s,t}f(x) = \mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\big|Z_s = \delta_x\right] = \mathbb{E}\left[f(Y_t)e^{\int_0^t B(Y_s)(m(Y_s)-1)ds}\big|Y_s = x\right], \quad (8)$$

où $(Y_t, t \geqslant 0)$ est un processus de Markov de générateur infinitésimal donnée pour toute fonction test $f : \mathcal{X} \to \mathbb{R}$ par :

$$\mathcal{A}f(x) = \mathcal{G}f(x) + B(x)m(x)\sum_{k\geqslant 0}\frac{kp_k(x)}{m(x)}\frac{1}{k}\sum_{i=1}^{k}\left(f(y) - f(x)\right)P_i^k(x,dy), \quad x \in \mathcal{X}.$$

Le terme exponentiel dans le membre de droite de (8) correspond à la croissance de la population. Le processus Y correspond au trait d'un individu "typique" dans la population. Dans le cas d'un branchement local, i.e. $P^k(x,dy) = k\delta_x$ pour tout $k \in \mathbb{N}$, on obtient $\mathcal{A}f(x) = \mathcal{G}f(x)$. La dynamique de Y est donc identique à celle de X et correspond au processus de l'épine introduit dans la Section 2.2. Hardy et Harris proposent dans [HH09] une généralisation de cette formule qui prend en compte l'ensemble de la trajectoire des individus dans le cas d'un branchement local. Bansaye et al. [BDMT11] étendent ces résultats au cas d'un branchement non-local à taux constant.

Pour répondre aux problématiques introduites en Section 1, une étape clé est la caractérisation de la dynamique du trait d'un individu échantillonné dans la population afin d'appréhender les biais dus à l'échantillonnage. Or le membre de droite dans l'équation (8) correspond à la trajectoire d'un individu typique pénalisé par la taille de la population. Cet objet n'est pas mathématiquement facile à appréhender ni à utiliser pour comprendre et exhiber la dynamique du trait d'un individu typique dans la population. Une première étape consiste à séparer la masse totale de la population de la dynamique du trait. Pour $f : \mathcal{X} \to \mathbb{R}$ et $x \in \mathcal{X}$, on considère donc l'opérateur renormalisé suivant :

$$P_{s,t}f(x) = \frac{R_{s,t}f(x)}{m(x,s,t)}, \quad \text{avec } m(x,s,t) = R_{s,t}\mathbf{1}(x) = \mathbb{E}\left(\#V_t\right), \quad \text{pour tout } 0 \leqslant s \leqslant t. \quad (9)$$

Ainsi $m(x,s,t)$ correspond à la masse totale de la population au temps t issue au temps s d'un individu de trait x. La famille $(P_{s,t}, t \geqslant s)$ est constituée d'opérateurs conservatifs mais ils ne possèdent pas la propriété de semi-groupe. Une deuxième étape consiste donc à considérer l'opérateur suivant :

$$P_{r,s}^{(t)}f(x) = \frac{R_{r,s}(fm(\cdot,s,t))}{m(x,r,t)}, \quad \text{avec } 0 \leqslant r \leqslant s \leqslant t.$$

On obtient alors une famille de semi-groupes $(P_{r,s}^{(t)}, r \leqslant s \leqslant t)$ conservatifs, pour tout $t \geqslant 0$, et $r \leqslant t$. Dans le cas d'un processus de branchement en temps discret, cette idée a été développée par Bansaye [Ban15]. Une renormalisation similaire à l'aide d'éléments propres associés au générateur du semi-groupe de premier moment a également été proposée par

Cloez [Clo17] pour une étude asymptotique du processus de branchement. On retrouve également une renormalisation similaire dans les travaux de Del Moral et Villemonais [DMV16] pour l'étude des diffusions tuées.

Pour tout $t \geqslant 0$, on peut associer au semi-groupe conservatif $(P_{r,s}^{(t)}, r \leqslant s \leqslant t)$ un processus de Markov. Grâce à la renormalisation effectuée précédemment, ce processus correspond au trait d'un individu "typique" dans la population. Cette famille de semi-groupes est indexée par le temps d'échantillonnage t et est inhomogène en temps à cause de la dépendance du nombre total d'individus dans la population au temps t en la condition initiale et plus précisément de la dépendance en le trait du taux de division B. L'étude de ce semi-groupe et du processus de Markov associé est cruciale pour la compréhension de la dynamique des populations structurées considérées dans cette thèse. L'objectif principal du Chapitre 1 est la caractérisation de ce processus, appelé processus auxiliaire, par son générateur infinitésimal. Ceci nous permet d'exhiber les différents biais liés à l'échantillonnage. Le Chapitre 2 porte ensuite sur le comportement asymptotique du trait d'un individu typique le long de ses lignées ancestrales. Les résultats de ce chapitre s'appuient sur des techniques d'ergodicité pour les processus de Markov inhomogènes en temps appliquées au processus auxiliaire et sur des calculs le long des généalogies.

Il existe des généralisations des formules Many-to-One qui permettent de caractériser la loi du trait de plusieurs individus dans la population au temps t. Des telles formules ont été proposées dans le cas d'un taux de division constant par Bansaye et al. [BDMT11], dans le cas d'un branchement local avec un taux de division non constant par Harris et Roberts [HR17] et dans un cadre plus général avec des outils spectraux par Cloez [Clo17]. Ces formules s'avèrent notamment utiles pour l'étude des corrélations entre les individus. Nous établissons dans le Chapitre 1 une formule Many-to-One pour les paires ainsi que pour tout l'arbre.

2.5 Ergodicité et vitesse de convergence pour un processus de Markov

Pour étudier le comportement asymptotique de populations structurées, un outil clé est la notion d'ergodicité. Celle-ci apparaît notamment dans le Chapitre 3 pour l'estimation statistique des paramètres d'un modèle de population structurée ainsi que dans le Chapitre 2 pour montrer la convergence du processus décrivant la dynamique du trait d'un individu typique dans la population. L'ergodicité d'un processus comprend plusieurs propriétés : la convergence de ce processus, uniforme ou non en la condition initiale et éventuellement à vitesse exponentielle. Nous présentons dans un premier temps un aperçu des résultats existants concernant l'ergodicité des chaînes de Markov en temps discret.

Temps discret. On considère $(X_n, n \geqslant 0)$ une chaîne de Markov à valeurs dans un espace $\mathcal{X}$ localement compact, séparable et métrisable de matrice de transition P. On désigne par $\mathcal{B}(\mathcal{X})$ la tribu borélienne associée à $\mathcal{X}$. La condition de Doeblin (Hypothèse A ci-dessous) est une hypothèse classique pour obtenir l'ergodicité de la chaîne.

Hypothèse A. *Il existe une mesure de probabilité φ sur $\mathcal{B}(\mathcal{X})$, $\varepsilon < 1$, $\delta > 0$ et $m \in \mathbb{N}$ tels que :*

$$\varphi(A) > \varepsilon \Rightarrow \inf_{x \in \mathcal{X}} P^m(x, A) > \delta.$$

Cette condition assure l'ergodicité uniforme et exponentielle des chaînes de Markov apériodiques.

Théorème ([MT09, Théorème 16.0.2])**.** *Si X est une chaîne de Markov apériodique vérifiant l'hypothèse A, alors on a*

$$\sup_{x\in\mathcal{X}} \|P^n(x,\cdot)-\pi\| \xrightarrow[n\to+\infty]{} 0,$$

et la convergence a lieu à vitesse exponentielle.

Ce résultat d'ergodicité est un résultat fort car il est uniforme en la condition initiale. Dans le cadre du modèle de population structurée étudié au Chapitre 3, les hypothèses considérées ne nous permettent pas de montrer la convergence uniforme en la condition initiale de la chaîne. D'autres hypothèses, moins contraignantes que la condition de Doeblin, entraînent l'ergodicité (non-uniforme) de la chaîne.

Hypothèse B. *Il existe une fonction $V:\mathcal{X}\to\mathbb{R}_+$, $c<1$ et $d>0$ tels que :*

$$PV(x)\leqslant cV(x)+d,$$

pour tout $x\in\mathcal{X}$.

Cette hypothèse, appelée condition de drift ou de Foster-Lyapunov, assure que le processus est ramené vers l'ensemble $\{x: V(x)<K\}$ pour un certain $K>0$ et permet le contrôle de la durée des excursions en dehors de cet ensemble [MT09, Théorème 11.3.4].

Hypothèse C. *Il existe $\alpha\in(0,1)$ et une mesure de probabilité ν telles que :*

$$\inf_{x\in\mathcal{C}} P(x,\cdot)\geqslant\alpha\nu(\cdot),$$

avec $\mathcal{C}=\{x: V(x)<R\}$ et $R>2d(1-c)^{-1}$ avec c,d définis dans l'hypothèse B.

Cette hypothèse est l'analogue localisé de la condition de Doeblin. Elle assure une certaine uniformité du processus en la condition initiale. Ces deux hypothèses sont suffisantes pour montrer l'ergodicité de la chaine de Markov.

Théorème 1 ([HM11, Théorème 1.2])**.** *Sous les Hypothèses B et C, la chaîne de Markov de transition P admet une unique mesure invariante π et il existe $C>0$ et $\gamma\in(0,1)$ tels que :*

$$\|P^nf-\pi f\|\leqslant C\gamma^n\|f-\pi f\|,$$

pour toute fonction mesurable $f:\mathcal{X}\to\mathbb{R}_+$ telle que $\|f\|<\infty$, avec $\pi f=\int f(x)\pi(dx)$ et $\|f\|=\sup_{x\in\mathcal{X}}|f(x)|\,(1+V(x))^{-1}$.

Nous utilisons notamment ce résultat d'ergodicité dans le Chapitre 3 pour l'estimation des paramètres du modèle de population structurée. Pour une étude détaillée de l'ergodicité pour des chaînes de Markov en temps discret, nous renvoyons le lecteur à [MT09].

Temps continu. Les hypothèses à vérifier pour montrer l'ergodicité d'un processus de Markov en temps continu sont similaires à celles du cas discret. Soit $(X_t, t\geqslant 0)$ un processus de Markov à valeur dans un espace $\mathcal{X}$ localement compact, séparable et mesurable. Pour tout $x\in\mathcal{X}$ et $A\in\mathcal{B}(\mathcal{X})$, on désigne par :

$$P_t(x,A):=\mathbb{P}_x(X_t\in A),$$

le semi-groupe du processus X. On note $\mathcal{A}$ le générateur infinitésimal de X et $\mathcal{D}(\mathcal{A})$ son domaine.

L'existence de mesures invariantes pour un processus de Markov en temps continu est liée à l'existence de mesures invariantes pour son analogue en temps discret [Hai10].

Proposition. *Soit* $P := P_T$ *pour un certain* $T > 0$*. Alors, si* π *est une mesure invariante pour* P*, la mesure* $\widehat{\pi}$ *définie par :*

$$\widehat{\pi}(A) = \frac{1}{T} \int_0^T P_t \pi(A) dt,$$

est une mesure invariante pour le semi-groupe P_t*.*

Dans le cas des processus de Markov en temps continu, l'hypothèse de type Foster-Lyapunov s'énonce en terme de générateur.

Hypothèse D. *Il existe une fonction mesurable* $V : \mathcal{X} \to [0, \infty)$ *et* $c, K > 0$ *tels que :*

$$\mathcal{A}V(x) \leqslant -cV(x) + K, \quad \forall x \in \mathcal{X}.$$

Il existe diverses formulations équivalentes de cette hypothèse. Nous renvoyons le lecteur à [DMT95] pour plus de détails. L'équivalent de l'Hypothèse C est parfois énoncée en terme de petite set [MT93a, MT93b]. Nous considérons également l'hypothèse suivante :

Hypothèse E. *Pour tout* $t > 0$*, il existe* $\alpha_t \in (0, 1)$ *et une mesure de probabilité* ν_t *sur* $\mathcal{X}$ *tels que :*

$$\inf_{x \in B(R,V)} P_t(x, \cdot) \geqslant \alpha_t \nu_t(\cdot),$$

avec $B(R, V) = \{x \in \mathcal{X} : V(x) \leqslant R\}$ *pour un certain* $R > \frac{2K}{c}$ *avec* c, K *définis dans l'Hypothèse D.*

Ces hypothèses assurent que le processus de Markov possède la propriété "d'oubli" de la condition initiale.

Théorème 2. *Sous les Hypothèses D et E, il existe* $C > 0$ *et* $\overline{c} > 0$ *tels que :*

$$|P_t\varphi(x) - P_t\varphi(y)| \leqslant Ce^{-\overline{c}t} d(x, y),$$

pour toute fonction φ *telle que* $\|\varphi\|_\infty \leqslant 1$*, où* d *désigne une distance sur* $\mathcal{X}$*.*

Nous renvoyons le lecteur à [Hai10] pour plus de détails sur le sujet ainsi que pour des conditions assurant la convergence sous-géométrique du processus de Markov. Dans le Chapitre 2, nous adaptons les Hypothèses D et E au cas inhomogène en temps. On obtient alors la condition d'oubli de la condition initiale pour le processus auxiliaire, ce qui nous permet de montrer la convergence de la mesure empirique le long des lignées ancestrales.

3 Estimation statistique dans des populations structurées

Le nombre croissant de données sur les populations de cellules, grâce notamment aux techniques microfluidiques, ouvre la voie à l'estimation statistique. L'un des enjeux est la compréhension du mécanisme de division cellulaire et plus particulièrement, la dépendance du taux de division des cellules en certaines caractéristiques. Nous considérons ici plusieurs modèles qui ont été étudiés afin de répondre à ces questions.

3.1 Étude du vieillissement chez E. coli

Les premiers travaux de modélisation et d'étude statistiques portant sur l'étude de Stewart et al. [SMPT05] sont ceux de Guyon [Guy07]. Il modélise une population de cellules qui présentent une division symétrique telles que la bactérie E. coli. à l'aide de chaînes bifurcantes. Le formalisme qu'il développe est notamment utilisé dans le Chapitre 3. Après une étude générale du modèle individu-centré, notamment sur son comportement asymptotique, Guyon propose un modèle auto-régressif spécifique pour l'étude du vieillissement chez E. coli. Soit $(X_n, n \geqslant 0)$ une chaîne de Markov vérifiant pour tout $n \geqslant 0$:

$$\left\{ \begin{array}{lcl} X_{2n} & = & \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} & = & \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{array} \right.$$

avec $\alpha_0, \alpha_1 \in (-1, 1)$, $\beta_0, \beta_1 \in \mathbb{R}$ et $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geqslant 1)$ une suite des variable gaussiennes bivariées centrées indépendantes et identiquement distribuées de matrice de covariance Γ. La suite $(X_n, n \geqslant 0)$ correspond aux taux de croissance des individus dans la population, $2n$ et $2n + 1$ désignent les cellules filles de parent n héritant respectivement du nouveau pôle (partie de membrane créee lors de la division précédente) et du pôle âgé (voir Section 1). Guyon propose alors de tester l'égalité des couples de paramètres (α_0, β_0) et (α_1, β_1) à partir de l'estimation des paramètres par maximum de vraisemblance. Les résultats du test statistique appliqué aux données de Stewart et al. [SMPT05] indiquent une différence significative entre les paramètres de transmission du taux de croissance pour les deux types de cellules filles, mettant ainsi en lumière une disymétrie dans la transmission du taux de croissance chez E. coli. Ceci soutient donc la thèse d'un phénomène de vieillissement pour cette bactérie. Citons également les travaux de Delmas et Marsalle [DM10] généralisant l'étude de Guyon au cas d'une chaîne de Markov bifurcante sur un arbre de Galton-Watson, autorisant ainsi certaines cellules à mourir sans se diviser.

3.2 Estimation du taux de division

Outre l'existence d'un phénomène de vieillissement cellulaire chez E. coli, une des questions que pose la dynamique des populations de cellules est le fonctionnement du mécanisme de division cellulaire. De nombreuses études statistiques s'intéressent ainsi à l'estimation du taux de division. Le but est de comprendre quelles caractéristiques sont à l'origine du déclenchement de la division cellulaire. Nous renvoyons le lecteur aux travaux de Robert et al. [RHK+14] pour une étude de la dépendance du taux de division en l'âge ou en la taille grâce à l'estimation non-paramétrique du taux de division à partir des données de Stewart et al. [SMPT05] et Wang et al. [WRP+10]. Nous présentons maintenant les différentes techniques statistiques qui ont été mises en œuvre pour rendre possible de telles études.

Les premiers travaux portant sur la reconstruction du taux de division [PZ07, DPZ09] reposent sur une approche déterministe du problème. À l'aide des résultats de convergence exponentielle des solutions des équations de croissance-fragmentation vers leur profil stable (voir Section 2.3), les auteurs se ramènent à la résolution du problème inverse suivant : comment reconstruire le taux de division B, solution de l'équation (7) de la Section 2.3, à partir du couple (λ, N), où λ est le paramètre de Malthus et N le profil asymptotique de la population ? Une des difficultés provient alors du fait que la valeur N_ε observée est bruitée de telle sorte que

$$\|N_\varepsilon - N\| < \varepsilon.$$

Or sans hypothèse de régularité sur la dérivée de N_ε, le problème de reconstruction de B est mal posé. Une solution consiste alors à "régulariser" l'équation (7) en considérant un problème approché bien posé et à contrôler la différence en norme entre la solution au problème approché et la solution au problème initial.

Une des limitations de cette approche est que l'on suppose connues certaines propriétés de N, notamment de régularités, qui semblent difficiles à vérifier à partir des données expérimentales. Doumic et al. [DHRBR12] ont proposé une étude statistique du problème pour faire disparaître cette limitation. Ainsi, ils proposent un estimateur à noyau $\widehat{N}$ de N à partir de mesures individuelles sur des cellules choisies au hasard dans la population à un instant donné. Les techniques de [DPZ09] leur permettent de construire un estimateur de B en utilisant $\widehat{N}$ au lieu de N_ε qui converge au taux $n^{-s/(2s+3)}$ vers B, où s est un coefficient traduisant la régularité imposée à B.

Les modèles individu-centrés permettent d'obtenir une meilleure vitesse de convergence grâce la prise en compte de la structure des généalogies. Les travaux de Doumic et al. [DHKR15] s'appuient ainsi sur une modélisation probabiliste de la population de cellules pour construire un estimateur du taux de division. Plus précisément, ils considèrent une population structurée par la taille où chaque individu u croît exponentiellement à un taux τ_u. On note ξ_u la taille à la naissance de u. Ainsi, si $\xi_u(t)$ désigne la taille de u à l'instant t, on a :

$$\xi_u(t) = \xi_u e^{\tau_u t}.$$

Un individu de taille x se divise au taux $B(x)$ et à la division, il est remplacé par deux individus de taille $x/2$. Le taux de croissance de chaque cellule fille est distribué de manière indépendante suivant le noyau $\rho(v,\cdot)$, où v est le taux de croissance de la cellule mère. Un objet clé dans l'étude du processus de Markov branchants est le noyau de transition associé à la distribution du couple (ξ_u, τ_u) donné par :

$$P_B\left(\left(\xi_{u^-}, \tau_{u^-}\right), \left(\xi_u, \tau_u\right)\right) = \frac{B(2\xi_u)}{\xi_u \tau_{u^-}} \mathbf{1}_{\left\{\xi_u \geqslant \xi_{u^-}/2\right\}} \exp\left(-\int_{\xi_{u^-}/2}^{\xi_u} \frac{B(2s)}{s\tau_{u^-}} ds\right) \rho\left(\tau_{u^-}, \tau_u\right),$$

où u^- désigne l'ancêtre de u. Un estimateur du taux de division B est alors obtenu à l'aide de la formule suivante :

$$B(x) = \frac{x}{2} \frac{\nu_B(x/2)}{\mathbb{E}_{\nu_B}\left[\frac{1}{\tau_{u^-}} \mathbf{1}_{\left\{\xi_{u^-} \leqslant x,\ \xi_u \geqslant x/2\right\}}\right]},$$

où ν_B désigne la mesure invariante associée au noyau P_B. Doumic et al. [DHKR15] construisent ainsi un estimateur à noyau pour le taux de division et montrent la convergence de cet estimateur en norme $\mathbb{L}^2$ à taux $\log(n) n^{-s/(2s+1)}$, où s est un paramètre correspondant à la régularité de B. Le taux de convergence est ainsi meilleur que celui de [DHRBR12]. Nous renvoyons le lecteur aux travaux de Bitseki-Penda et al. [BPHO17] pour une amélioration de la vitesse de convergence de l'estimateur du taux de division grâce à des inégalités de déviation de type Bernstein.

Dans les travaux cités précédemment, il est supposé que l'on dispose de données généalogiques sur la population de cellules, c'est-à-dire que l'on dispose des caractéristiques de tous les individus dans la population jusqu'à la génération n. Cependant, pour les populations de cellules, les générations sont chevauchantes (voir Figure 2.1) et les individus en vie à un instant T ne sont pas tous de la même génération. Ainsi, si les données que l'on considère sont composées des traits des individus qui ont vécu jusqu'à l'instant T, les procédures d'estimation précédentes ne peuvent pas être utilisées car les données sont biaisées : les individus que se reproduisent moins vite sont moins représentés dans la population arrêtée au temps T que dans la population jusqu'à la génération n où tous les individus sont recensés, peu importe la longueur de leur cycle de vie. Hoffmann et Olivier [HO16] ont proposé un estimateur du taux de division dans le cas de telles données pour un modèle structuré en âge. Ils considèrent un processus de Bellmann-Harris où chaque individu a une durée de vie distribuée suivant $f(x)dx$. À sa mort, un individu est remplacé par un nombre aléatoire de descendants distribué suivant $(p_k, k \geqslant 2)$ ayant chacun un durée de vie également distribuée suivant $f(x)dx$. Le taux de division B est alors obtenu grâce à la formule suivante :

$$B(x) = \frac{f(x)}{1 - \int_0^x f(y)dy}, \quad x \geqslant 0.$$

À partir de cette représentation et de la convergence de mesures empiriques, obtenues par des techniques d'épine, ils proposent un estimateur à noyau pour B et prouvent la convergence en probabilité de cet estimateur à vitesse exponentielle vers le taux de division. L'une des spécificités du modèle en âge est que les caractéristiques des cellules à la naissance sont égales : au début de leur existence, les individus sont supposés être d'âge nul. Cette propriété n'est plus vérifiée pour les modèles en taille, ce qui complique le problème de l'estimation des paramètres.

3.3 Estimation du noyau de fragmentation

Comme expliqué dans la Section 1, la manière dont les cellules se divisent (par exemple en bourgeonnant ou en se divisant en deux parties égales) joue un rôle clé dans le mécanisme du vieillissement cellulaire. Ainsi, l'estimation statistique du noyau de fragmentation est un objectif important pour la compréhension de ce phénomène.

Hoang [Hoa15] a proposé un estimateur du noyau de fragmentation dans le cadre d'une population structurée par la taille. Dans son modèle, chaque individu a un certain degré de toxicité qui augmente linéairement avec le temps à taux $\alpha > 0$. Cette modélisation du vieillissement de la cellule par l'accumulation d'une toxicité est inspirée des travaux de Stewart et al. [SMPT05]. Les cellules se divisent à un taux constant et lorsqu'une cellule se divise, sa toxicité à la division est répartie entre les deux cellules filles suivant la fraction aléatoire $\Gamma \in (0, 1)$ de densité h. Hoang propose alors un estimateur adaptatif de h pour lequel il prouve une inégalité oracle. Cet estimateur est construit à partir des données dans la population sur $[0, T]$. Pour la construction d'un estimateur dans le cas d'un taux de division non constant et de données composées des cellules en vie à un instant T donné, nous renvoyons le lecteur à [Hoa16].

4 Résumé des résultats de la thèse

Cette thèse est constituée de trois chapitres. Les deux premiers chapitres sont consacrés à l'étude probabiliste des processus de Markov structurés branchants. Le troisième chapitre est consacré à l'étude statistique de ces modèles.

On considère un processus de Markov branchant à valeurs mesures $(Z_t, t \geqslant 0)$ défini de la manière suivante :

- chaque individu u dans la population a un trait $(X_t^u, t \geqslant 0)$ dont la dynamique suit un processus de Markov à valeurs dans un espace mesurable $\mathcal{X} \subset \mathbb{R}^d$ de générateur $\mathcal{G}$ et de domaine associé $\mathcal{D}(\mathcal{G})$,
- l'individu u meurt au temps t au taux $B(X_t^u)$,
- à sa mort, un individu de trait x est remplacé par un nombre aléatoire de descendants distribué selon $(p_k(x), k \geqslant 0)$,

- le trait à la naissance du j-ième descendant parmi k est distribué suivant $P_j^{(k)}(x, \cdot)$, pour tout $1 \leqslant j \leqslant k$.

L'objet d'étude est alors le processus à valeurs mesures Z dont une représentation est donnée par :

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

où V_t désigne l'ensemble des individus en vie au temps t. Dans le premier chapitre, on s'intéresse à la caractérisation du trait le long de sa lignée ancestrale d'un individu échantillonné dans la population à un instant t. Nous exhibons ainsi les biais liés à l'échantillonnage dans le cas d'un taux de division dépendant du trait. Le Chapitre 2 porte sur le comportement asymptotique de la mesure empirique associée au processus de branchement. Son étude repose sur l'ergodicité du processus introduit dans le Chapitre 1. Enfin, le troisième chapitre porte sur l'estimation des paramètres dans une population branchante structurée par une diffusion. Il s'agit donc d'un cas particulier des modèles étudiés dans les deux premiers chapitres.

4.1 Chapitre 1 : échantillonnage uniforme dans une population structurée branchante

Ce chapitre est constitué de l'article "Uniform sampling in a structured branching population" [Mar16], accepté pour publication dans Bernoulli en juillet 2017.

Dans ce chapitre, on étudie l'échantillonnage d'individus dans une population structurée branchante. La question qui nous intéresse est la suivante : si l'on choisit uniformément au hasard un individu en vie dans la population à un instant donné t, que peut-on dire de ses caractéristiques et de celles de ses ancêtres le long de sa lignée ancestrale ?

Le premier résultat porte sur l'existence et l'unicité du processus de Markov structuré branchant Z précédemment décrit. Les hypothèses sous lesquelles nous montrons ce résultat sont relativement générales. En particulier, le taux de division n'est pas supposé borné mais majoré par un polynôme afin d'assurer la non-explosion du processus en temps fini. Le Théorème 1.2.2 du Chapitre 1 assure que le processus à valeurs mesures Z est l'unique solution forte d'une équation différentielle stochastique avec sauts.

Dans une deuxième partie, on étudie le processus auxiliaire associé au semi-groupe $(P_{r,s}^{(t)}, r \leqslant s \leqslant t)$ défini précédemment par (9). Le résultat principal porte sur la caractérisation du trait le long de sa lignée ancestrale d'un individu échantillonné dans la population.

Théorème (Théorème 1.3.1). *Sous des hypothèses assurant la non-explosion de la population en temps fini, on a pour toute fonction mesurable positive F sur l'ensemble des processus càdlàg (continus à droite et limités à gauche), pour tout $t \geqslant 0$ et $x \in \mathcal{X}$,*

$$\mathbb{E}\left[\sum_{u \in V_t} F\left(X_s^u, s \leqslant t\right)\right] = m(x,0,t)\mathbb{E}\left[F\left(Y_s^{(t)}, s \leqslant t\right)\right], \tag{10}$$

où $\left(Y_s^{(t)}, s \leqslant t\right)$ est un processus de Markov inhomogène de générateur $\left(\mathcal{A}_s^{(t)}, s \leqslant t\right)$ vérifiant :

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_s^{(t)} f(x) + \widehat{B}_s^{(t)}(x) \int_{\mathcal{X}} \left(f\left(y\right) - f\left(x\right)\right) \widehat{P}_s^{(t)}\left(x, dy\right),$$

avec :

$$\widehat{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}\left(m(\cdot,s,t)f\right)(x) - f(x)\,\mathcal{G}\left(m(\cdot,s,t)\right)(x)}{m(x,s,t)},$$

$$\widehat{B}_s^{(t)}(x) = B(x)\int_{\mathcal{X}} \frac{m(y,s,t)}{m(x,s,t)} m(x,dy),$$

$$\widehat{P}_s^{(t)}(x,dy) = m(y,s,t)m(x,dy)\left(\int_{\mathcal{X}} m(y,s,t)m(x,dy)\right)^{-1},$$

et :

$$m(x,A) := \sum_{k\geqslant 0} p_k(x)\sum_{j=1}^{k} P_j^{(k)}(x,A),$$

désigne le nombre moyen de descendants d'un individu de trait x dont le trait est dans A.

On retrouve ici le fait que l'inhomogénéité en temps est due à la dépendance en le trait du nombre moyen d'individus dans la population. De plus, les différents biais dus à la croissance du nombre d'individus sont explicites. Tout d'abord, le processus auxiliaire saute plus souvent que n'importe quel individu dans la population lorsqu'un saut lui permet d'atteindre un trait plus efficace en terme de nombre de descendants. Cette augmentation du taux de division a également été observée par exemple dans [CR88, LPP95, HH09, BDMT11]. De la même manière, la loi de reproduction favorise les traits à la naissance efficaces en terme de nombre de descendants. Enfin, la dynamique du trait entre les divisions est également biaisée à cause de la dépendance en le trait du taux de division et du caractère stochastique de cette évolution. En effet, dans le cas d'une dynamique déterministe pour le trait, on a $\widehat{\mathcal{G}}_s^{(t)} f(x) = \mathcal{G}f(x)$. Nous montrons également d'autres formules Many-to-One qui permettent la caractérisation du trait d'un couple d'individus ou de l'ensemble des individus qui ont vécu jusqu'à l'instant t. Ces formules s'avèreront notamment utiles pour l'étude du comportement asymptotique de la mesure empirique dans le Chapitre 2.

Le processus auxiliaire est donc un candidat pour caractériser la loi du trait d'un individu échantillonné dans la population au temps t le long de sa lignée ancestrale. Cependant, ce processus prend en compte les biais liés à la croissance de la population. Or ces biais n'apparaissent pas instantanément dans la population. C'est pourquoi la loi du trait d'un individu échantillonné dans la population au temps t est caractérisée par le processus auxiliaire uniquement pour des temps suffisamment grands. Nous montrons dans le Chapitre 1 un résultat en grande population, ce qui nous permet de mimer le temps long. On désigne par $X^{U(t),\nu}$ le trait d'un individu échantillonné uniformément dans le population du temps t, de distribution initiale ν. Si $\nu_n = \sum_{i=1}^n \delta_{X_i}$ avec X_i des variables aléatoires indépendantes et identiquement distribuées selon ν, on montre la convergence en loi suivante :

Théorème (Théorème 1.4.1)**.** *Sous des hypothèses assurant la non-explosion de la population en temps fini, la suite $\left(X_{[0,t]}^{U(t),\nu_n}, n\geqslant 0\right)$ converge en loi quand $n\to+\infty$ dans l'espace de Skorokhod vers $Y_{[0,t]}^{(t),\pi_t}$ où*

$$\pi_t(dx) = \frac{m(x,0,t)\nu(dx)}{\int_{\mathcal{X}} m(y,0,t)\nu(dy)},$$

est la loi initiale de $Y_{[0,t]}^{(t),\pi_t}$.

Ainsi, le processus auxiliaire est le bon objet pour caractériser la loi d'un individu échantillonné dans la population au temps t issue d'une grande population initiale.

Enfin, nous illustrons les résultats précédents par des exemples concrets de modèles de populations de cellules. Dans le cas d'une adaptation du modèle structuré en taille étudié dans [DHKR15] où chaque individu croît exponentiellement à un taux $a > 0$ et se divise à un taux $B(x) = x$, le nombre moyen d'individus dans la population est donné par :

$$m(x, s, t) = 1 + \frac{x}{a}\left(e^{a(t-s)} - 1\right), \quad x \in \mathbb{R}_+, 0 \leqslant s \leqslant t.$$

Ainsi, le générateur du processus auxiliaire est donné par :

$$\mathcal{A}_s^{(t)} f(x) = axf'(x) + x\left(1 + \frac{1}{1 + \frac{x}{a}(e^{a(t-s)} - 1)}\right)\left(f\left(\frac{x}{2}\right) - f(x)\right).$$

Dans ce cas, le seul biais est l'accélération du temps i.e. le taux de division est plus important pour le processus auxiliaire que pour n'importe quel individu dans la population. Dans la Section 1.4.2 du Chapitre 1, nous proposons une généralisation de cet exemple à une famille plus générale de taux de division, éventuellement dépendant du temps. Nous détaillons aussi le cas d'une population dont le trait croît linéairement ainsi que le cas de la prolifération de parasites dans une population de cellules également étudié dans [BT11].

4.2 Chapitre 2 : loi des grands nombres pour des processus de Markov branchants à l'aide de l'ergodicité des lignées ancestrales

Ce chapitre est composée d'une prépublication intitulée " A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages" [Mar17].

Cette partie porte sur le comportement en temps long de la mesure empirique associée au processus de branchement Z donnée par :

$$\frac{1}{N_t} \sum_{u \in V_t} \delta_{X_t^u},$$

où N_t désigne le cardinal de V_t. Grâce à l'étude des généalogies et à l'ergodicité du processus auxiliaire Y le long de sa lignée ancestrale, on montre la convergence $\mathbb{L}^2$ de la mesure empirique vers le distribution asymptotique de Y. Ainsi, la mesure aléatoire caractérisant le trait d'un individu échantillonné dans la population est asymptotiquement déterministe et est donnée par la distribution de Y. Notre approche repose sur l'utilisation du générateur du processus auxiliaire exhibé dans le chapitre précédent.

Pour alléger les notations, on considère dans ce chapitre le cas particulier d'une division binaire : un individu de trait x est remplacé à sa mort par deux individus dont les traits sont distribués suivant le noyau de probabilité $P(x, \cdot)$. Les résultats de ce chapitre peuvent facilement être généralisés au cas d'un processus avec un nombre aléatoire de descendants.

L'étude du comportement asymptotique de la mesure empirique a déjà été réalisée dans de nombreux cas. Asmussen et Hering [AH76] ont montré la convergence de la mesure empirique pour des processus de branchement généraux dans les cas discret et continu grâce à une hypothèse de décomposition du semi-groupe du premier moment. Ils traitent en particulier du cas des diffusions branchantes. Nous renvoyons le lecteur à [EHK10] pour une généralisation de ces résultats sur les diffusions branchantes. Athreya et Kang ont montré la convergence de la mesure empirique pour une processus de branchement en temps discret dans les cas d'un espace de trait discret [AK98a] ou continu [AK98b]. Nous citons également les travaux de Guyon [Guy07] sur les chaînes de Markov bifurcantes et leur généralisation au cas d'un arbre de Galton-Watson par Delmas et Marsalle [DM10]. Pour des résultats en environnement changeant, nous renvoyons le lecteur à [BH15, Ban15]. En temps continu et espace de traits discret, Georgii et Baake [GB03] ont montré une loi des grands nombres portant sur le type des individus le long d'une lignée ancestrale. Leurs résultats reposent sur les techniques d'épines développées par Lyons et al. [LPP95] et détaillées en Section 2.2. Harris et Robert [HR14] utilisent la construction de l'épine de [HH09] pour montrer la convergence de la mesure empirique dans le cas d'un processus de Markov branchant avec branchement local. Pour des théorèmes centraux limites, nous renvoyons le lecteur aux travaux de Guyon [Guy07] et Ren et al. [RSZ14].

L'une des techniques pour montrer la convergence de la mesure empirique repose sur l'existence d'éléments propres pour le générateur infinitésimal du semi-groupe de premier moment, [LPP95, GB03, EHK10]. Une autre technique permet de s'affranchir du cadre spectral. Elle consiste à considérer une approche de la question basée sur le processus auxiliaire i.e. la caractérisation de la loi du trait d'un individu choisi uniformément dans la population. Sous des hypothèses d'ergodicité du processus auxiliaire, on peut montrer la convergence de la mesure empirique vers la distribution limite de ce processus. Cette technique a été développée en temps discret notamment par Athreya et Kang [AK98b] et Guyon [Guy07], ainsi qu'en temps continu par Bansaye et al. [BDMT11] dans le cas neutre et par Cloez [Clo17] dans le cas non-neutre en combinant les deux approches. Nous montrons dans le Chapitre 2 la convergence de la mesure empirique grâce à l'ergodicité du processus auxiliaire dans le cas non-neutre en s'affranchissant du cadre spectral.

Comme vu précédemment, dans le cas d'un taux de division non constant, le processus auxiliaire est inhomogène en temps et ce même si la dynamique du trait elle-même est homogène en temps. Nous avons donc adapté le travail de Hairer et Mattingly [HM11] pour montrer l'ergodicité d'un processus de Markov inhomogène en temps.

Hypothèse F. *On suppose que :*

1. *il existe une fonction* $V : \mathcal{X} \to \mathbb{R}_+$ *et* $c, d > 0$ *tels que pour tout* $x \in \mathcal{X}$, $t \geqslant 0$ *et* $s \leqslant t$ *:*
$$\mathcal{A}_s^{(t)} V(x) \leqslant -cV(x) + d,$$
2. *pour tout* $0 < r < s$, *il existe* $\alpha_{s-r} \in (0,1)$ *et une mesure de probabilité* $\nu_{r,s}$ *sur* $\mathcal{X}$ *tels que pour tout* $t \geqslant s$ *:*
$$\inf_{x \in B(R,V)} P_{r,s}^{(t)}(x, \cdot) \geqslant \alpha_{s-r} \nu_{r,s}(\cdot),$$
avec $B(R,V) = \{x \in \mathcal{X} : V(x) \leqslant R\}$ *pour un certain* $R > \frac{2d}{c}$ *avec* c, d *définis dans le premier point.*

Cette hypothèse rassemble les deux conditions classiques, détaillées dans la Section 2.5, qui permettent de montrer l'ergodicité exponentielle d'un processus (voir Théorème 1). La première est une condition de type Foster-Lyapunov et le seconde une condition de minoration. Sous ces deux conditions, on montre un résultat d'oubli de la condition initiale pour les trajectoires du processus auxiliaire.

Proposition (Proposition 2.3.1)**.** *Sous l'Hypothèse F, il existe* $\overline{c} > 0$ *tel que pour tout* $x, y \in \mathcal{X}$, $T > 0$, *pour toute fonction mesurable bornée* $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$ *et tout* $0 \leqslant r \leqslant t$,
$$\left| P_{r,t,T}F(x) - P_{r,t,T}F(y) \right| \leqslant C e^{-\overline{c}(t-r)} d(x,y) \left\| F \right\|_\infty ,$$
où d *est une distance sur* $\mathcal{X}$, C *est une constant positive et*
$$P_{r,t,T}F(x) := \mathbb{E}\left[F\left(Y_{t+s}^{(t+T)}, s \leqslant T \right) | Y_r^{(t+T)} = x \right]. \tag{11}$$

Ces résultats permettent de contrôler la trajectoire du processus inhomogène en temps considéré sur un intervalle de taille T.

Nous montrons un résultat de type loi des grands nombres grâce au résultat d'ergodicité du processus auxiliaire précédemment énoncé.

Théorème (Proposition 2.3.3). *Sous des hypothèses assurant la non-explosion de la population en temps fini et sous l'Hypothèse F, pour tout* $x_0, x_1 \in \mathcal{X}$ *et* $T > 0$*, on a :*

$$\left(\frac{\sum_{u \in V_{t+T}} F\left(X^u_{t+s}, s \leqslant T\right)}{N_{t+T}} - \mathbb{E}_{x_1}\left[F\left(Y^{(t+T)}_{t+s}, s \leqslant T\right)\right] \right) \xrightarrow[t \to +\infty]{} 0,$$

où la convergence a lieu dans $\mathbb{L}^2$ *avec pour mesure initiale* δ_{x_0}.

Ainsi, la dynamique du processus auxiliaire correspond asymptotiquement à la dynamique du trait le long des lignées ancestrales d'un individu échantillonné.

Enfin, nous illustrons ce résultat par l'étude détaillée du cas d'une population structurée en taille où chaque individu croît exponentiellement à taux $a > 0$, se divise à taux $B(t, x) = x\varphi(t)$ au temps t si sa taille à l'instant t est x et donne naissance à deux individus de taille θx et $(1-\theta)x$, où θ est distribué uniformément sur $[\varepsilon, 1-\varepsilon]$, pour un certain $\varepsilon > 0$. Ce modèle est similaire à celui étudié dans [DHKR15] et dans les exemples du Chapitre 1. Nous montrons dans ce cas l'ergodicité exponentielle des trajectoires du processus auxiliaire ainsi que la convergence de la mesure empirique du processus de branchement.

4.3 Chapitre 3 : estimation statistique dans une population branchante structurée par une diffusion

Ce chapitre est constitué d'un travail en collaboration avec Marc Hoffmann. Il est consacré à l'estimation des différents paramètres qui régissent la division cellulaire. Le processus de branchement étudié est un cas particulier du processus Z décrit précédemment. On suppose que la dynamique du trait suit une diffusion sur $\mathbb{R}$:

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x \in \mathbb{R},$$

où $(W_t, t \geqslant 0)$ est un mouvement brownien standard. Chaque individu meurt à un taux $B(\cdot)$ qui dépend de son trait et laisse une descendance formée de deux individus. Les traits à la naissance des descendants d'un individu de trait x sont θx et $(1-\theta)x$, où θ est une variable aléatoire sur $[0, 1]$ distribuée suivant un noyau κ symétrique par rapport à 1/2. Contrairement aux deux chapitres précédents, on s'intéresse ici à la chaîne de Markov formée des traits des individus à la naissance, adoptant ainsi un point de vue généalogique plutôt que chronologique.

On utilise les notations de Ulam-Harris-Neveu introduites en Section 2.2. On note $\mathbb{G}_n \subset \mathcal{U}$ l'ensemble des individus qui composent la génération n et

$$\mathbb{T}_n = \bigcup_{m=0}^{n} \mathbb{G}_n,$$

l'ensemble des individus dans la population jusqu'à la génération n. Pour $u \in \mathcal{U}$, on note u^- l'unique ancêtre de u à la génération précédente. On s'intéresse alors au problème de l'estimation des différents paramètres du modèle à partir de données observées suivantes :

$$\mathbb{X}^n := (X_u, u \in \mathbb{T}_n),$$

où X_u désigne le trait à la naissance de l'individu u.

L'obtention de telles observations semble réaliste d'un point de vue expérimental, à l'aide notamment des techniques microfluidiques ou de fluorescence détaillées dans la Section 1.

La chaîne de Markov $(X_u, u \in \mathcal{U})$ est une chaîne de Markov bifurcante [Guy07, DM10] sur $\mathcal{X}$ de noyau de transition $\mathcal{P} : \mathcal{X} \to \mathcal{X} \times \mathcal{X}$. Le noyau $\mathcal{P}$ traduit la manière dont le trait se transmet aux deux descendants.

On considère le processus de l'individu étiqueté noté Y. On a pour tout $m \in \mathbb{N}$:

$$Y_m = Y_{\varnothing\epsilon_1\ldots\epsilon_m},$$

avec $(\epsilon_k, k \geqslant 0)$ indépendants et identiquement distribués suivant une loi de Bernoulli de paramètre 1/2. Comme l'évolution du temps est généalogique, la valeur de la chaîne ainsi construite au rang n correspond au choix uniforme d'un individu parmi les individus en vie à la génération n. On note :

$$\mathcal{Q} = \frac{\mathcal{P}_1 + \mathcal{P}_2}{2},$$

le noyau de transition associé à Y, où $\mathcal{P}_1$ et $\mathcal{P}_2$ sont les deux marginales du noyau de transition $\mathcal{P}$. Dans notre cas, on a $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{Q}$.

Comme détaillé précédemment dans la Section 3, l'estimation statistique dans des modèles de population structurées porte exclusivement à notre connaissance sur des cas où la dynamique du trait des individus est déterministe. Dans le cas d'une diffusion branchante, Höpfner et al. [HHL02] ont proposé un estimateur à noyau du taux de mort B. Leur approche utilise de manière cruciale le fait que le trait des descendants à la naissance est identique au trait de leur ancêtre à sa mort ce qui n'est pas le cas dans notre modèle.

Sous des hypothèses de régularité, on montre que $\mathcal{Q}(x, dy) = \mathcal{Q}_B(x, dy) = q_B(x, y)dy$ avec

$$q_B(x, y) = \int_0^1 B\left(\frac{y}{z}\right) \sigma\left(\frac{y}{z}\right)^{-2} \mathbb{E}\left[\int_0^{+\infty} e^{-\int_0^t B(\phi_x(s))ds} dL_t^{\frac{y}{z}}(\phi_x)\right] \kappa(z) \frac{dz}{z},$$

où $L_t^y(\phi_x)$ désigne le temps local au temps t en y de la semi-martingale $(\phi_x(t), t \geqslant 0)$. L'intégrale par rapport au temps local résulte du fait que la durée de vie des individus dépend de la dynamique stochastique du trait. Ainsi, la probabilité de passer de x à y en une génération dépend du temps moyen passé par la diffusion $(\phi_x(t), t \geqslant 0)$ dans l'intervalle $[y, +\infty)$. L'une des difficultés provient du fait que la densité du noyau de transition q_B dépend de la moyenne de l'ensemble de la trajectoire du trait $(\phi_x(t), t \geqslant 0)$ et non de la trajectoire à un instant donné t.

Estimation non-paramétrique. On s'intéresse dans un premier temps au comportement asymptotique de la chaîne Y. Sous des conditions appropriées sur les paramètres régissant la dynamique du flot stochastique, on montre que les Hypothèses B et C de la Section 2.5 sont vérifiées par le noyau $\mathcal{Q}$, entraînant ainsi l'ergodicité du processus.

Théorème (Théorème 3.2.2)**.** *Sous des hypothèses appropriées sur la dynamique du flot, $\mathcal{Q}$ admet une mesure invariante ν. De plus, il existe $C > 0$ et $\rho \in (0, 1)$ tels que pour tout $x \in \mathbb{R}$ et $m \in \mathbb{N}$:*

$$|\mathcal{Q}^m \varphi(x) - \nu(\varphi)| \leqslant C\rho^m (1 + V(x)) \|\varphi - \nu(\varphi)\|_\infty .$$

pour toute fonction mesurable $\varphi : \mathbb{R} \to \mathbb{R}$ telle que

$$\sup_{x \in \mathbb{R}} \frac{\varphi(x)}{1 + V(x)} < \infty,$$

avec $V(x) = x^2$.

Ainsi l'ergodicité n'est pas uniforme mais dépend de la condition initiale. Nous verrons ultérieurement que si le trait évolue sur un espace compact, on obtient l'ergodicité uniforme du processus en vérifiant la condition de Doeblin A de la Section 2.5.

Ce résultat d'ergodicité nous permet d'étudier le comportement asymptotique de la mesure empirique donnée par :

$$\mathcal{M}_n(\psi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \psi(X_{u^-}, X_u),$$

pour toute fonction test $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, où $\mathbb{T}_n^\star = \mathbb{T}_n \backslash \{\varnothing\}$. Elle converge lorsque n tend vers l'infini vers $\nu(\mathcal{Q}\psi)$. L'estimation non-paramétrique de q et ν s'effectue à l'aide de noyaux d'approximation.

Définition. *On appelle noyau d'ordre k toute fonction $G : \mathcal{X} \to \mathbb{R}$ à support compact satisfaisant $\int_{\mathcal{X}} x^l G(x) dx = \mathbf{1}_{\{l=0\}}$ pour tout $l = 0, \ldots, k$.*

Afin d'obtenir une bonne vitesse de convergence pour des estimateurs à noyau de ν et $\mathcal{Q}$, il est nécessaire de contrôler la dépendance en ψ de la vitesse de convergence de la mesure empirique. Plus précisément, si l'on considère un noyau $G_h := h^{-1}G(\cdot/h)$, où h est un paramètre de lissage, on a :

$$|G_h|_1 \lesssim 1, \quad |G_h|_2^2 \lesssim h^{-1}, \quad |G_h|_\infty^2 \lesssim h^{-2}.$$

La vitesse de convergence de $\mathbb{E}_\mu\left[\mathcal{M}_n(G_h)^2\right]$ doit avoir le bon ordre en h, à savoir h^{-1}, pour obtenir une bonne vitesse de convergence pour les estimateurs à noyaux construits à partir de G_h. Nous renvoyons le lecteur à la preuve du Théorème 3.3.3 dans le Chapitre 3 pour plus de détails.

Pour toute fonction $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, on pose $\psi^\star(x) = \sup_{y \in \mathcal{X}} |\psi(x,y)|$, $\psi_\star(y) = \sup_{x \in \mathcal{X}} |\psi(x,y)|$ et

$$|\psi|_{\wedge 1} = \int_{\mathcal{X}\times\mathcal{X}} |\psi(x,y)| dx dy \bigwedge \int_{\mathcal{X}} \psi_\star(y) dy.$$

Pour toute mesure positive μ sur $\mathcal{X}$, on pose également :

$$|\psi|_\rho = \int_{\mathcal{X}\times\mathcal{X}} |\psi(x,y)| \rho(dx) dy + |\psi|_{\wedge 1}.$$

Une analyse fine de la variance de la mesure empirique le long des généalogies nous permet d'obtenir la vitesse de convergence désirée avec un contrôle sur ψ de l'ordre de $|\psi|_2^2$.

Théorème (Théorème 3.2.3). *Soit μ une mesure de probabilité sur $\mathbb{R}$ telle que $\mu(V^2) < \infty$. Soit $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ une fonction bornée telle que $\psi_\star$ soit à support compact. Alors, sous les hypothèses assurant l'ergodicité de la chaîne Y et si $\rho \leqslant \frac{1}{2}$, on a pour tout $n \in \mathbb{N}$:*

$$\mathbb{E}_\mu\big[\big(\mathcal{M}_n(\psi) - \nu(\mathcal{Q}\psi)\big)^2\big] \lesssim |\mathbb{T}_n|^{-1}\big(|\psi^2|_{\mu+\nu} + |\psi^\star\psi|_\mu + \big(1 + \mu(V^2)^{1/2}\big)|\psi_\star|_1 |\psi|_\nu\big),$$

où le symbole $\lesssim$ signifie à une constante explicite près dépendante de $\mathcal{Q}$ et du support de $\psi_\star$.

Soit G un noyau d'ordre k. On considère le noyau G_h, où h est un paramètre de lissage, défini par $G_h(x) = h^{-1}G(h^{-1}x)$ pour tout $x \in \mathcal{X}$. Pour tout $x_0 \in \mathcal{X}$, on a :

$$\mathcal{M}_n(G_h(\cdot - x_0)) \xrightarrow[n \to +\infty]{\mathbb{P}} \int_{\mathcal{X}} G_h(x - x_0) \nu(x) dx.$$

Un estimateur de ν est donc donné par :

$$\widehat{\nu}(x_0) = \mathcal{M}_n(G_h(\cdot - x_0)). \tag{12}$$

De la même manière, si l'on considère $G^{\otimes 2}_{h_1,h_2}(x,y) = h_1^{-1}h_2^{-1}G(h_1^{-1}x)G(h_2^{-1}x)$, on a pour tout $x_0, y_0 \in \mathcal{X}^2$:

$$\mathcal{M}_n\left(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} \int_{\mathcal{X}\times\mathcal{X}} G^{\otimes 2}_{h_1,h_2}(x - x_0, y - y_0)q(x,y)\nu(x)dx.$$

On obtient donc l'estimateur suivant de la densité q :

$$\widehat{q}_n(x_0,y_0) = \frac{\mathcal{M}_n\left(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\right)}{\mathcal{M}_n(G_h(\cdot - x_0)) \vee \varpi}, \tag{13}$$

où le dénominateur permet de corriger le poids ν apparaissant à la limite et $\varpi > 0$ est un seuil assurant la bonne définition de l'estimateur. Les estimateurs à noyau ainsi définis sont consistants.

Théorème (Théorème 3.3.3). *Soit $\alpha, \beta > 0$. On considère $\widehat{\nu}_n(x_0)$ défini en (12) par un noyau d'ordre $k > \max(\alpha,\beta)$ et un paramètre de lissage $h = |\mathbb{T}_n|^{-1/(2\beta+1)}$ et $\widehat{q}_n(x_0,y_0)$ défini par (13) à l'aide du même noyau et avec $h_1 = |\mathbb{T}_n|^{-s(\alpha,\beta)/(\alpha\wedge\beta)(2s(\alpha,\beta)+1)}$, $h_2 = |\mathbb{T}_n|^{-s(\alpha,\beta)/\beta(2s(\alpha,\beta)+1)}$ et $\varpi_n \to 0$. Alors, sous des hypothèses assurant l'ergodicité de la chaîne Y, on a*

$$\left(\mathbb{E}\left[\left(\widehat{\nu}_n(x_0) - \nu(x_0)\right)^2\right]\right)^{1/2} \lesssim |\mathbb{T}_n|^{-\beta/(2\beta+1)},$$

et

$$\left(\mathbb{E}\left[\left(\widehat{q}_n(x_0,y_0) - q(x_0,y_0)\right)^2\right]\right)^{1/2} \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)},$$

avec $s(\alpha,\beta)^{-1} = (\alpha\wedge\beta)^{-1} + \beta^{-1}$. De plus, les majorations sont uniformes en $\mathcal{Q}$ pour $\mathcal{Q}$ dans une certaine classe de régularité de Hölder dépendant de α et β.

Estimation paramétrique du taux de division. On s'intéresse dans un second temps à l'estimation paramétrique du taux de division B par maximum de vraisemblance. Ce problème est complexe et nous avons donc besoin de considérer des hypothèses supplémentaires. La principale difficulté provient du fait que le résultat obtenu précédemment sur la vitesse de convergence de la mesure empirique s'applique dans le cas de fonctions ψ bornées. Or la fonction $\psi(x,y) = \log(q(x,y))$ ne l'est pas forcément. Afin de contourner cette difficulté, nous considérons le cas des processus de Markov branchants où la dynamique du trait suit une diffusion réfléchie sur le compact $[0,L]$, pour un certain $L > 0$. La dynamique du trait entre les divisions est alors donnée par l'équation différentielle stochastique suivante :

$$\begin{cases} dX_t = r(X_t)dt + \sigma(X_t)dW_t + d\ell_t, \\ \ell_t = \int_0^t \left(\mathbf{1}_{X_s=0} + \mathbf{1}_{X_s=L}\right) d\ell_s, \end{cases}$$

où $(W_t, t \geqslant 0)$ est un mouvement brownien standard et la deuxième équation traduit les réflexions au bord du domaine. Dans ce cas, la densité de transition q de la chaîne Y est légèrement modifiée. Son support est notamment restreint à cause des réflexions.

Les preuves des résultats obtenus dans le cas d'une diffusion non-réfléchie s'adaptent aisément au cas avec réflexion. La compacité du domaine nous permet de montrer l'ergodicité uniforme du processus le long d'une lignée. De plus, la convergence de la mesure empirique à vitesse $\mathbb{T}_n^{-1}$ est toujours vérifiée.

On suppose à présent que le taux de division B appartient à la classe suivante :

$$\mathcal{B} = \left\{B : \mathcal{X} \to \mathbb{R}, B(x) = B_0(\vartheta, x), x \in \mathcal{X}, \vartheta \in \Theta\right\},$$

où $x \mapsto B_0(\vartheta, x)$ est connu au paramètre ϑ près et $\Theta \subset \mathbb{R}_+^d$ est un ensemble compact de paramètres qui caractérisent le taux de division. On montre l'identifiabilité du modèle pour une classe ordonnée de taux de division i.e. que l'on peut comparer tous les éléments de la classe deux à deux (voir Définition 3.3.5 et Proposition 3.3.6).

Pour souligner la dépendance en ϑ, on introduit un indice dans les notations précédemment définies. On suppose que l'on dispose d'observations $(X_u, u \in \mathbb{T}_n)$ du processus jusqu'à la génération n obtenues pour un certain noyau $\mathcal{P}_\vartheta$. On considère alors l'estimateur du maximum de vraisemblance (EMV) donné par :

$$\widehat{\vartheta}_n \in \operatorname*{argmax}_{\vartheta \in \Theta} \left\{ \frac{1}{\mathbb{T}_n} \sum_{u \in \mathbb{T}_n} \log \left(q_\vartheta \left(X_{u^-}, X_u \right) \right) \right\}.$$

La dynamique choisie pour le trait nous permet de borner le noyau de transition et de montrer la consistance de l'EMV.

Théorème (Théorème 3.3.7). *Sous des hypothèses de régularité du taux de division, $\widehat{\vartheta}_n$ converge en probabilité vers ϑ lorsque n tend vers l'infini.*

Enfin, nous montrons la normalité asymptotique de l'estimateur. On définit $\Psi(\vartheta)$ la matrice d'information de Fisher dont les coefficients sont donnés pour tout $1 \leqslant i, j \leqslant d$ par :

$$\Psi(\vartheta)_{i,j} = \nu_\vartheta \mathcal{Q}(\vartheta) \left(\frac{\partial_{\vartheta_i} q_\vartheta \partial_{\vartheta_j} q_\vartheta}{q_\vartheta^2} \right).$$

Théorème (Théorème 3.3.9). *Sous des hypothèses de régularité et si $\Psi(\vartheta)$ est inversible, pour tout ϑ dans l'intérieur de Θ, on a :*

$$\mathbb{T}_n^{1/2} \left(\widehat{\vartheta}_n - \vartheta \right) \to \mathcal{N} \left(0, \Psi(\vartheta)^{-1} \right),$$

en loi lorsque n tend vers l'infini, où $\mathcal{N}\left(0, \Psi(\vartheta)^{-1}\right)$ désigne la loi normale d-dimensionnelle de moyenne 0 *et de matrice de covariance l'inverse de la matrice de Fisher $\psi(\vartheta)$.*

Malgré sa normalité asymptotique, l'implémentation numérique de l'estimateur est un problème complexe que nous comptons aborder de manière approfondie. Notons que nous n'avons pas trouvé d'autre contraste pour l'estimation de $x \mapsto B(x)$, en particulier, nous ne sommes pour l'instant pas en mesure de l'estimer de manière non-paramétrique.

Introduction in English

Content

1 Growth, aging and cell division

The microscopic dynamic of cell populations is a complex phenomenon depending on many factors. The study of these populations at an individual level led to the discovery of new phenomena such as the horizontal gene transfer. Moreover, there is a large number of techniques [BSJ04], such as fluorescence or cytometry, allowing to observe cell populations and to gather a large amount of data. The analysis of those data relies on the understanding of the bias appearing in the process of sampling and on the modeling of cell population dynamics at an individual level. The construction of such a model requires to take into account the characteristics that play a key role in the mechanisms of cell division. Such cell populations are said to be structured populations because the life cycle of each cell depends on individual characteristics. We now give some examples of dependencies between the different processes that play a role in the dynamic of cell populations.

Growth and cell division

One of the most important mechanism in the dynamic of cell populations is cell division. It is linked with the growth of the cell. Recent techniques for the monitoring of cell populations using microfluidics [WRP$^+$10] allow to follow the dynamic of a large number of cells and to obtain a large amount of data. The cells are trapped in small cavities and when a cell divides, the daughter cell is pushed outside the hole. Then, it is flushed out by the constant flow of a solution over the holes. Using these microfluidics techniques, Wang et al. [WRP$^+$10] highlighted the stability of the growth mechanism for the bacteria E.coli. It raises the following question: how do cells control their size and the homeostasie i.e. the property of a system to remain at a beneficial value, of their size? Several models have been suggested to explain this phenomenon: the "timer" model or the "sizer" model in which respectively the lifetime or the size at division of each individual is fixed. Recently, Taheri-Araghi et al. [TABS$^+$15] fitted experimentally the "adder" model for the bacteria E. coli and Bacillus subtilis. In this model, a fixed value is added to the size of each individual between two divisions. Besides, the models "sizer" and "timer" did not fit the data of [TABS$^+$15]. Moreover, experimental validation of the "adder" model has also been done for the yeast Saccharomyces cerevisiae [SRA16].

Those studies highlight an average behavior for the growth of cells. But there is also a phenotypical variability at an individual level. It is in general beneficial for the cell population because it increases the resistance of the population to external perturbations. For an overview of the different works on the reasons for this heterogeneity, we refer the reader to [SA02] for the case of Saccharomyces cerevisiae and to [Ave06] for a more general framework. In particular, for the case of the yeast, the growth rates and the sizes of individuals are very heterogeneous among a population. The size differences are due to the asymmetric division of the yeast, (see Figure 1) because the buds are smaller than their mother cell.

Cellular aging

An other mechanism, which is widely studied for cell populations, is cellular aging. For S. cerevisiae, the age usually corresponds to the number of times a cell divided. In a yeast population, most individuals are young i.e. they have not divided yet. It has an impact on the fitness of the cell population because some physiological characteristics are linked with the age of the cell, such as its size or the duration of a cell cycle. Kale and Jazwinski [KJ96] studied the correlation between the resistance of S. cerevisiae to UV exposition and the age of the cells. They observed that UV resistance increases with age, is maximum for

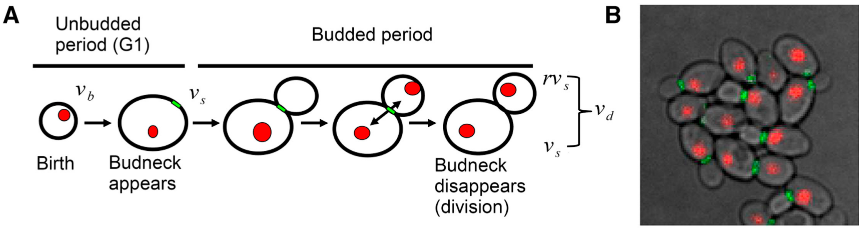

Figure 1.1: Picture from [SRA16]. A. Budding yeast cell cycle. B. Growing microcolonies of S. cerevisiae. The red and green fluorescent markers correspond respectively to the nucleus and to the bud neck ring which indicates cell division.

cells which divided 8 times and then, decreases with age. Therefore, the increase in UV resistance is not due to the growth of the cells because older cells are more vulnerable to UV exposition. However, it can be explained by an internal regulation, depending on the age of the cell, for the production of proteins responsible for the reparation of damages caused by UV exposition. Then, resistance to external stresses depends on the age of the cell. Aging is therefore an important characteristic to consider for the modeling of cell population dynamics.

The first studies on aging for S. cerevisiae revealed a rejuvenation phenomenon: the last bud of a cell before its death is still capable of dividing [Joh66, Mül71]. Moreover, the daughter cells coming from those last divisions have a smaller reproductive lifetime than the cells coming from the first buds. It suggests the existence of "aging factors" satisfying the following conditions [HG08]:

- they accumulate with age,
- at division, they are asymmetrically distributed between the bud and the mother cell,
- their absence or diminution lead to an increase of the cell lifetime,
- their augmentation leads to a decrease of the cell lifetime.

The extrachromosomal rDNA circles (ERC) fulfill those conditions. Then, it is a potential candidate for explaining a part of the phenomenon of aging for the yeast. For more details on the different changes at an individual level due to aging and their eventual role in aging, we refer the reader to [DLJB14]. Recent progresses in the collection of data, especially by means of microfluidics [CCK17], augur well for the understanding of the aging mechanisms. Indeed, by following a cell during its entire life, you can collect data on the lifetime of individuals, on the dynamic of cells between two divisions but also on the morphology of cells [LVH$^+$12, XZZ$^+$12]. The effects and causes of aging can therefore be studied in more details.

Aging for S. cerevisiae seems to be closely related to the asymmetrical division of cells. It has also been observed for a bacteria with asymmetrical division by Ackermann et al. [ASJ03]. For the bacteria E. coli, division is symmetrical so that the experimental study of cellular aging is more complicated. Stewart et al. [SMPT05] followed the growth of cells colonies using fluorescence for nine generations. They gathered the physiological data of each cell. At division, the bacterium E. coli split in two equal parts. Each daughter cell inherits a new pole, corresponding to the pole coming from the middle of the mother cell, and an other older pole. In particular, the daughter cell which inherits the pole previously new for the mother cell should be younger than the daughter cell which inherits the pole that already existed some divisions before. The experiment of Stewart et al. [SMPT05] revealed that cells which inherit the older pole have a smaller growth rate and a more important death probability than cells which inherit the new pole. The results of this study suggest that their is an aging phenomenon for the bacterium E. coli due to a functional asymmetry such as the localization in the poles of the cell of important components.

Finally, the study of individual-based models for cell populations is motivated by numerous questions. Moreover, the various dependencies between the different mechanisms involved in the cell population dynamics require a rigorous mathematical approach of the problem.

2 Branching processes and structured population

In this thesis, we are interested in the dynamic of structured populations without interactions. In this section, we present some of the mathematical tools for the modeling of such dynamics.

2.1 Spine and size biased tree

The pioneering works of Harris, Kesten-Stigum and Athreya on multitype supercritical Galton-Watson processes revealed the asymptotic repartition of types in the population which is given by the left eigenvector of the reproduction matrix. The question of understanding the mechanisms leading to such a distribution is crucial to answer the problematic linked with cellular aging detailed in Section 1.

The notion of size-biased tree is at the heart of this question of the understanding of sampling bias. It has been considered by Kallenberg [Kal77], Chauvin and Rouault [CR88] then Chauvin, Rouault and Wakolbinger [CRW91]. Lyons, Pemantle and Peres [LPP95] suggested a new proof of the Kesten-Stigum theorem based on size-biased trees. To each node in the tree, we associate a label using the Ulam-Harris-Neveu notation. Let

$$\mathcal{U} = \bigcup_{n \in \mathbb{N}} (\mathbb{N}^*)^n ,$$

be the set of all possible labels. The first individual is denoted by $\varnothing$. We an individual $u \in \mathcal{U}$ dies, its K descendants are labeled $u1, \ldots, uK$. Those notation allow in particular to keep track of the genealogy of an individual through its label. We introduce a new type of random trees: the size-biased Galton-Watson trees. We define the size-biased distribution $(\widehat{p}_k, k \geqslant 0)$ by

$$\widehat{p}_k = \frac{kp_k}{m}, \quad \text{for } k \geqslant 0,$$

where m is the mean of the reproduction law $(p_k, k \geqslant 0)$. This law favors big families with many siblings. Then, we construct the size-biased Galton-Watson tree as follows:

- we start with a single individual v_0,
- it has a random number of descendants distributed according to $\widehat{p}$,
- we pick randomly one of these descendants, say v_1,
- the subtrees arising from the other descendants are standard Galton-Watson trees,
- we repeat the same steps by replacing v_0 with v_1.

The sequence of labels $(v_k, k \geqslant 0)$ constitutes the spine. Let t be a Galton-Watson tree of length n and v an individual belonging to the nth generation in t. We have:

$$\mathbb{P}\left(\left(\widehat{\mathcal{A}}_n, \xi\right) = (t, v)\right) = \frac{1}{m^n} \mathbb{P}\left(\mathcal{A}_n = t\right),$$

with $\mathcal{A}_n$ and $\widehat{\mathcal{A}}_n$ the random variables corresponding respectively to a Galton-Watson tree at generation n and to a size-biased Galton-Watson tree at generation n and ξ the random variable corresponding to the spine. Let us mention that, as defined, the measure on the space of Galton-Watson trees until generation n with a distinguished line of descent is not a probability measure because its total mass is equal to m^{-n}. Let Z_n be the number of individuals in generation n in t. Then, we have the following result linking the law of Galton-Watson trees with the size-biased reproduction law to the law of Galton-Watson trees with the standard reproduction law:

$$\mathbb{P}\left(\widehat{\mathcal{A}}_n = t\right) = \frac{Z_n}{m^n}\mathbb{P}\left(\mathcal{A}_n = t\right). \tag{1}$$

As $W_n := Z_n m^{-n}$ is a martingale, it is natural to consider the associated changed measure, which corresponds to the law of the size-biased process. Using this new measure, Lyons, Peres and Pemantle [LPP95] and Lyons [Lyo97] give an alternative proof of the Kesten-Stigum theorem.

This decomposition of the tree using the spine extends to multitype Galton-Watson processes [KLPP97, BK04, Ath00] using the eigenelements of the reproduction matrix and to age-dependant processes [Olo98, Olo09]. Generalizations to continuous time have been done by Georgii and Baake [GB03] for multitype branching Markov processes for the study of the ancestral types of typical individuals in the population. Finally, spine techniques have been developed and used for the study of various other processes. We refer the reader to the work of Kyprianou [Kyp04] for the study of the asymptotic behavior and of the uniqueness of wave propagation for the KPP equation. For fragmentation processes, the use of the spine, also refered to as tagged-fragment, has been developed by Bertoin et al. [BR05] and Harris et al. [HKK10].

Hardy and Harris [HH09] generalized the construction of the spine to the case of infinite-dimensional traits, in particular for the study of the branching Brownian motion. As Lyons, Pemantle and Peres [LPP95], they consider the set $\widetilde{\mathcal{T}}$ of Galton-Watson trees (in continuous time) with a distinguish line of descent, although known as the spine. The sequence of labels of the individuals constituting the spine is still denoted by ξ and the label of the individual corresponding to the spine at time t is denoted by ξ_t. The set $\widetilde{\mathcal{T}}$ is equipped with the canonical probability measure P for the branching process $(Z_t, t \geqslant 0)$ which characteristics are the following:

- the dynamic of each individual in the population is an independent copy of a Markov process $(X_t, t \geqslant 0)$ on a measurable space $\mathcal{X}$, we denote by X_t^u the trait of u at time t,
- an individual u dies at time t at rate $B(X_t^u)$ where $B : \mathcal{X} \to (0, \infty)$ is a measurable function,
- when an individual u dies at position x, it is replaced by $A_u + 1$ individuals at the same position, where A_u is an independent copy of $A(x)$ and $\mathbb{P}(A(x) = k) = p_k(x)$ for all $k \geqslant 0$, we denote by $m(x) = \sum_{k\geqslant 0} k p_k(x)$ the mean of $A(x)$,
- then, the individuals evolve independently from each other following the same dynamic.

This very general model is similar to the one we study in this thesis. It allows in particular to model spatially structured populations. Note that in this model an individual can note die without leaving any descendant and that the branching mechanism is local, i.e. the trait at birth of the descendants is the same as the trait at death of their parent. We now define several filtrations on the set $\widetilde{\mathcal{T}}$ of labeled trees that allow to separate the different available informations. We give here an informal description of those filtrations and we refer the reader to [HH09] for their rigorous definition.

- $\mathcal{F}_t$ contains all the information on the branching process until time t but do not know anything about the spine.
- $\widetilde{\mathcal{F}}_t$ is constructed from $\mathcal{F}_t$ by adding the information on the spine: it contains all the information on the branching process and the spine until time t.
- $\mathcal{G}_t$ is the filtration corresponding to the spine dynamic. It does not contain any information on the lineage of the spine in the tree.

For $x \in \mathcal{X}$, we define P^x the probability measure on $(\widetilde{\mathcal{T}}, \mathcal{F}_\infty)$ such that $\left(\widetilde{\mathcal{T}}, \mathcal{F}_\infty, (\mathcal{F}_t)_{t\geqslant 0}, P^{\cdot}\right)$ is the canonical model for the branching process Z with

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{t\geqslant 0} \mathcal{F}_t\right).$$

As before, Hardy and Harris define a probability measure $\widetilde{P}^x$ from P^x for which the spine corresponds a lineage in the tree. For every $\widetilde{\mathcal{F}}_t$-measurable function f, let

$$\int_{\widetilde{\mathcal{T}}} f d\widetilde{P}^x = \int_{\widetilde{\mathcal{T}}} \sum_{u\in V_t} f_u \prod_{v\leqslant u} \frac{1}{1+A_v} dP^x,$$

with

$$f = \sum_{u\in V_t} f_u \mathbf{1}_{\{\xi_t = u\}},$$

where V_t denote the set of all individual alive at time t and for all $u \in V_t$, f_u is a measurable function with respect to $\mathcal{F}_t$. Contrary to the measure exhibited in the work of Lyons et al. [LPP95, Lyo97], this measure is a probability measure. Under this measure, the spine moves as the Markov process $(X_t, t \geqslant 0)$ and at death, it is replaced by a uniformly chosen individual among its descendants. We consider a change of measure by means of a martingale. It allows to take into account the bias in the tree due to the growth of the population, in the same way as Lyons, Pemantle and Peres [LPP95] (see (1)). One of the novelty appearing in the work of Hardy and Harris is that one can choose any martingale for the change of measure. Let $\zeta(t)$ be a $\mathcal{G}_t$-martingale depending only on the spine and let $\widetilde{\zeta}(t)$ be the martingale given by:

$$\widetilde{\zeta}(t) = \prod_{u<\xi_t} \frac{1+A_u}{1+m\left(\xi_{S_u}\right)} \prod_{v<\xi_t} \left(1+m\left(\xi_{S_v}\right)\right) e^{-\int_0^t B(\xi_s)m(\xi_s)ds} \zeta(t).$$

The corresponding change of measure is given by:

$$\left.\frac{d\widehat{P}}{d\widetilde{P}}\right|_{\widetilde{\mathcal{F}}_t} = \frac{\widetilde{\zeta}(t)}{\widetilde{\zeta}(0)}.$$

It is the analogous of (1) in continuous time. Thus, $\widehat{P}$ is a new probability measure on $\left(\widetilde{\mathcal{T}}, \widetilde{\mathcal{F}}_\infty\right)$. The behavior of the spine under the measure $\widehat{P}$ is identical to the behavior of the spine previously described in discrete time zas far as the reproduction is concerned. Supplementary bias appear because of the overlapping of generations and because of the stochastic dynamic of individuals between jumps. The first product in the expression of the martingale $\widetilde{\zeta}$ corresponds to the introduction of a size bias and the second product corresponds to a time acceleration. Indeed, the lifetimes of individuals belonging to the spine are shorter because the division rate $B(\cdot)$ is replaced by $(1 + m(\cdot))B(\cdot)$. For a precise description of the spine behavior under $\widehat{P}$, we refer the reader to [HH09]. They also give the analogous of the spine decomposition initially proved in [Lyo97] which is especially useful for the asymptotic study of additive martingales.

The spectral framework is appropriate for the use of spine techniques because it allows to exhibit simple martingales. Consider a branching Markov process where the dynamic of individuals follows a Markov process $(X_t, t \geqslant 0)$ with generator $\mathcal{G}$, with binary local branching at rate B. If there exist a function ϕ and $\lambda \in \mathbb{R}$ so that $(\mathcal{G} + B)\phi = \lambda\phi$, then,

$$W_t := \langle \phi, Z_t \rangle e^{-\lambda t},$$

is a $\sigma(Z_t, t \geqslant 0)$-martingale. The existence of eigenelements simplifies the asymptotic study of the branching process by considering the dynamic penalized by the eigenvector. For more details on these techniques, we refer the reader to [EHK10, Clo17].

2.2 On the use of measure-valued processes and growth-fragmentation equations

In the models considered in this thesis, each individual is characterized by a vector of phenotypical traits belonging to $\mathcal{X} \subset \mathbb{R}^d$. The use of measure-valued processes enables us to follow the dynamic of the whole population but also individual dynamics. Let $\mathcal{M}_F(\mathcal{X})$ be the set of all finite positive measures on $\mathcal{X}$ and $\mathcal{M}_P(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ the subset of punctual measure on $\mathcal{X}$:

$$\mathcal{M}_P(\mathcal{X}) = \left\{ \sum_{i=1}^{n} \delta_{x_i}, n \geqslant 0, x_1, \ldots, x_n \in \mathcal{X} \right\},$$

where δ_x denotes the Dirac mass at point x. Using this formalism, we represent the microscopic dynamic of the population by the process Z defined by:

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

where V_t represents the population at time t and X_t^u represents the trait of u at time t.

Measure-valued processes have widely been used for the study of structured populations. We mention here the work of Fournier and Méléard [FM04] one the dynamic of a spatially structured fixed population (plants, for example) with interactions. Many studies of structured population are based on techniques initiated in [FM04], such as the work of Tran on age-structured processes [Tra06].

The use of measure-valued processes enables the definition of the structured branching process as the unique solution of a stochastic differential equation (see Chapter 1, Theorem 1.2.2). Let us consider the example of a size-structured population where each individual grows exponentially at rate $a > 0$ and divides at rate B in two equal parts. The associated measure-valued process $(Z_t, t \geqslant 0)$ is then solution of the following equation:

$$\begin{aligned} \langle Z_t, f \rangle = & \langle Z_0, f \rangle + \int_0^t \int_0^\infty a x f'(x) Z_s(dx) ds \\ & + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+} \mathbf{1}_{\{u \in V_s,\ \theta \leqslant B(X_s^u)\}} \left(2f\left(\frac{X_s^u}{2}\right) - f(X_s^u) \right) M(ds, du, d\theta), \end{aligned} \tag{2}$$

for all measurable functions$f : \mathbb{R}_+ \to \mathbb{R}$ and all $t \geqslant 0$, where $M(ds, du, d\theta)$ is a punctual Poisson measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ with intensity $ds \otimes du \otimes d\theta$. The first integral corresponds to the exponential growth of individuals between divisions whereas the second integral corresponds to division events happening at rate B during which an individual with trait x is replaced by 2 individuals with trait $x/2$. These equations describe the dynamic of the process and enable the use of tools from stochastic calculus. Therefore, we can calculate various quantities associated with the process such as the variance of the number of individuals in the population (see Chapter 2, Lemma 2.4.6).

A key object for the analysis of the dynamic of the trait of a typical individual studied in Chapter 1 is the first moment semi-group. It is given for all $s \leqslant t$, $x \in \mathcal{X}$ and all measurable functions $f : \mathcal{X} \to \mathbb{R}$ by:

$$R_{s,t}f(x) = \mathbb{E}\left[\sum_{u \in V_t} f(X_t^u) \big| Z_s = \delta_x\right]. \tag{3}$$

This semi-group is non-conservative. In the case of the previous example of a size-structured population, it solves the following equation:

$$R_{s,t}f(x) = f(x) + \int_s^t \int_0^\infty axf'(x)R_{s,r}(dx)dr + \int_s^t \int_0^\infty B(x)\left(2f\left(\frac{x}{2}\right) - f(x)\right)R_{s,r}(dx)dr,$$

for all measurable functions $f : \mathbb{R}_+ \to \mathbb{R}$ and all $0 \leqslant s \leqslant t$. We can define the generator $\mathcal{A}$ associated with the first moment semi-group. In the case of the previous example, it is given by:

$$\mathcal{A}f(x) = axf'(x) + B(x)\left(2f\left(\frac{x}{2}\right) - f(x)\right).$$

As on the stochastic differential equation (2), we can read the dynamic of the individuals on the generator: the first term corresponds to the exponential growth and the second term to the mechanism of division.

Measure-valued processes are the microscopic version of macroscopic models. Thus, there are many works on the link between the different scales of models [FM04, BT11, Clo17]. In the case of branching processes, the large population limit is deterministic, given by the first-moment semi-group previously defined in (3). This result holds by means of the branching property. Thus, the study in large population of the processes considered in this thesis reduces to the study of the associated growth-fragmentation equation.

Let $n(t, dx)$ be the distribution such that for all measurable functions $f : \mathcal{X} \to \mathbb{R}$:

$$\langle n(t,\cdot), f\rangle = R_{0,t}f(x) = \mathbb{E}\left[\sum_{u \in V_t} f(X_t^u)\right], \quad \text{and } n(0, dy) = n^{\text{in}}(dy).$$

We consider the case where $n(t, dx) = n(t, x)dx$, which holds in particular if the initial measure n^{in} is absolutely continuous with respect to the Lebesgue measure. The function $n(t, x)$ corresponds to the density of individuals with trait x in the population at time t. Taking the expectation in the stochastic differential equation (2), we get the following growth-fragmentation equation, with unknown n:

$$\partial_t n(t,x) + a\partial_x(xn(t,x)) + B(x)n(t,x) = 4B(2x)n(t,2x).$$

More generally, a growth-fragmentation equation with unknown n is an equation of the form:

$$\partial_t n(t,x) + \partial_x(v(x)n(t,x)) + B(x)n(t,x) = \int_0^1 B\left(\frac{x}{z}\right) n\left(t, \frac{x}{z}\right) \frac{\rho(dz)}{z}, \tag{4}$$

with initial conditions $v(0)n(t,0) = 0$ for all $t > 0$ and $n(0,x) = n^{\text{in}}(x)$ for all $x \geqslant 0$. Thus, v corresponds to the growth rate of individuals, B to the division rate and ρ to the fragmentation kernel.

Growth-fragmentation equations appear in various domains such as the study of growth of bacterial population but also the study of polymers or the transmission control protocol (TCP). These equations, first studied from a deterministic and macroscopic point of view, are the large population limit of branching processes. Numerous works focus on growth-fragmentation equations [Per07], especially on the existence of solutions and on the study of the asymptotic behavior of these solutions. The spectral analysis of the associated semi-group is the key of the problem. In 2016, Mischler and Scher [MS16] developed and used this theory to prove the exponential convergence of solutions to their asymptotic profile for a certain class of fragmentation rates.

The asymptotic study of solutions of growth-fragmentation equations relies on the existence of solutions to the associated eigenvalue problems, direct and dual, with unknown (λ, N, ϕ) given by:

$$\begin{cases} \lambda N(x) + (v(x)N(x))' + B(x)N(x) = \int_0^1 B\left(\frac{x}{z}\right) N\left(\frac{x}{z}\right) \frac{\rho(dz)}{z}, \\ N \geqslant 0,\ \int_0^\infty N(x)dx = 1, \end{cases} \tag{5}$$

and

$$\begin{cases} \lambda\psi(x) - v(x)\psi'(x) + B(x)\psi(x) = B(x)\int_0^1 \psi(zx)\rho(dz), \\ \psi \geqslant 0,\ \int_0^\infty N(x)\psi(x)dx = 1. \end{cases}$$

Then, we have the following conservation law [MS16]:

$$\int_0^\infty n(t,x)\psi(x)dx = e^{\lambda t}\int_0^\infty \psi(x)n^{\text{in}}(x)dx.$$

The first eigenvalue λ is the Malthus parameter. It corresponds to the exponential growth rate of the population. We refer the reader to [Oli17] for a study on the impact of variability of individual characteristics on the Malthus parameter. Asymptotically, the solutions of equation (4) satisfy as t tends to infinity:

$$n(t,x) \sim e^{\lambda t}\left\langle n^{\text{in}}, \psi\right\rangle N(x),$$

for a weighted norm [MMP05]. This indicates that the asymptotic size profile of the population is given by N and do not depend on the initial condition. We will find again this phenomenon of the process forgetting its initial condition for probabilistic models in Section 2.4.

The existence of solutions to the eigenvalue problems associated with growth-fragmentation equation has been widely studied. We mention the works of Diekmann et al. [DHT84], Perthame and Ryzhik [PR05], Michel [Mic06]. Perthame and Ryzhik [PR05] also proved the exponential convergence of the solutions to growth-fragmentation equations towards the stationary profile in the case of a binary division and linear growth. We refer the reader to [MMP05, BG17] for works on the speed of convergence in more general frameworks.

Recently, Bouguet [Bou16], Bertoin and Watson [BW17] studied growth-fragmentation equations adopting a probabilistic point of view. By means of this new approach, they exhibited precise estimates for the asymptotic behavior the solutions to growth-fragmentation equations. Their works rely on Foster-Lyapunov criteria (see Section 2.4) and are therefore linked with the approach followed in this thesis.

2.3 Many-to-One formula

One of the key formulas provided by spine techniques is the so-called Many-to-One formula. It resumes the dynamic of the whole population (many) by the dynamic of the spine (one). We consider the branching Markov process $(Z_t, t \geqslant 0)$ which characteristics are the following:

- the dynamic of each individual in the population is given by an independent copy of a Markov process $(X_t, t \geqslant 0)$ on a measurable space $\mathcal{X}$ with generator $\mathcal{G}$, and we recall that X_t^u corresponds to the trait of u at time t.
- an individual u dies at time t at rate $B(X_t^u)$ where $B : \mathcal{X} \to (0, \infty)$ is a measurable function,
- when an individual u dies with trait x, it is replaced by A_u individuals, with A_u being an independent copy of $A(x)$, $\mathbb{P}(A(x) = k) = p_k(x)$ for all $k \geqslant 0$, and we denote by $m(x) = \sum_{k \geqslant 0} k p_k(x)$ the mean of $A(x)$,
- the trait at birth of the k descendants of an individual with trait x are distributed according to $P^k(x, dy)$ and for $1 \leqslant i \leqslant k$, we denote by $P_i^k(x, dy)$ the ith marginal of P^k,
- then, the individuals evolve independently from each other following the same dynamic.

This process is more general than the one previously described in Section 2.1. It is the object of study of Chapters 1 and 2. From the first moment semi-group and its associated generator, we obtain the classical Many-to-One formula, also known as Feynman-Kac formula [DM04]:

$$R_{s,t} f(x) = \mathbb{E}\left[\sum_{u \in V_t} f(X_t^u) \big| Z_s = \delta_x\right] = \mathbb{E}\left[f(Y_t) e^{\int_0^t B(Y_s)(m(Y_s)-1)ds} \big| Y_s = x\right], \qquad (6)$$

where $(Y_t, t \geqslant 0)$ is a Markov process with infinitesimal generator given for all test functions $f : \mathcal{X} \to \mathbb{R}$ by:

$$\mathcal{A} f(x) = \mathcal{G} f(x) + B(x) m(x) \sum_{k \geqslant 0} \frac{k p_k(x)}{m(x)} \frac{1}{k} \sum_{i=1}^{k} \left(f(y) - f(x)\right) P_i^k(x, dy), \quad \text{for } x \in \mathcal{X}.$$

The exponential term on the right-hand side of (6) corresponds to the growth of the population. The process Y corresponds to the trait of a "typical" individual in the population. In the case of a local branching, i.e. $P^k(x, dy) = k\delta_x$ for all $k \in \mathbb{N}$, we have $\mathcal{A} f(x) = \mathcal{G} f(x)$. The dynamic of Y is then the same as the dynamic of X and corresponds to the spine process introduced in Section 2.1. Hardy and Harris give in [HH09] a generalization of this formula which takes into account the whole dynamic of individuals in the case of a local branching. Bansaye et al. [BDMT11] extended these results to the case of a process with non-local branching and constant division rate.

To answer the questions introduced in Section 1, a key step is the characterization of the dynamic of a sampled individual in the population, in order to understand the bias due to the sampling procedure. The right-hand side of equation (6) corresponds to the trajectory of a typical individual penalized by the size of the population. This object is not mathematically easy to understand or to use for the characterization of the dynamic of a typical individual, which we aim at doing in this thesis. Then, a first step consists in separating the total mass of the population from the dynamic of the trait. We consider the following normalized operator:

$$P_{s,t}f(x) = \frac{R_{s,t}f(x)}{m(x,s,t)}, \tag{7}$$

where

$$m(x,s,t) = R_{s,t}\mathbf{1}(x) = \mathbb{E}\left(\#V_t\right),$$

corresponds to the total mass of the population at time t descended from an individual with trait x at time s. The family $(P_{s,t}, t \geqslant s)$ is composed of conservative operators without the semi-group property. A second step consists in defining the following operator:

$$P^{(t)}_{r,s}f(x) = \frac{R_{r,s}(fm(\cdot,s,t))}{m(x,r,t)}.$$

Thus, we obtain a family $(P^{(t)}_{r,s}, r \leqslant s \leqslant t)$ of conservative semi-groups, for $t \geqslant 0$, and $r \leqslant t$. In the case of discrete branching processes, this idea has been developed by Bansaye [Ban15]. A similar approach with the definition of a new family of semi-groups using eigenelement of the first moment semi-group has been suggested by Cloez [Clo17] for the asymptotic study of branching processes. Del Moral and Villemonais [DMV16] used also a similar normalization for the study of killed diffusions.

For all $t \geqslant 0$, we consider the time-inhomogeneous Markov process associated with the conservative semi-group $(P^{(t)}_{r,s}, r \leqslant s \leqslant t)$. By means of the preceding normalization, this process corresponds to the dynamic of the trait of a typical individual in the population. This family of semi-groups is indexed by the sampling time t. The time inhomogeneity comes from the dependence of the size of the population on the initial condition and more precisely, from the dependence of the division rate B on the trait. The study of this time-inhomogeneous process is the heart of this thesis. The main goal of Chapter 1 is the characterization of this process, named auxiliary process, by its infinitesimal generator. Thus, we exhibit the different bias due to the sampling procedure. In Chapter 2, we study the asymptotic behavior of the trait of a typical individual along its ancestral lineage. The results of this chapter rely on the ergodicity techniques for time-inhomogeneous Markov processes applied to the auxiliary process and on computations along genealogies.

2.4 Ergodicity and speed of convergence for a Markov process

For the study of the asymptotic behavior of structured populations, ergodic techniques are very useful. We apply them in Chapter 2, to prove the convergence of the process describing the dynamic of the trait of a typical individual in the population and in Chapter 3 for the statistical reconstruction of the parameters of a structured population. There are several "degrees" of ergodicity for a process: the convergence can be uniform in the initial condition or not, and the speed of convergence can be exponential. We first present some existing results on the ergodicity of discrete time Markov chains.

Discrete time. Let $(X_n, n \geqslant 0)$ be a Markov chain with values in a locally compact, separable metric space $\mathcal{X}$ with transition P. Let $\mathcal{B}(\mathcal{X})$ denote the Borel σ-algebra of $\mathcal{X}$. Doeblin's condition (Assumption A below) is a classical assumption to test to obtain the ergodicity of the Markov chain.

Assumption A. *There exists a probability measure* φ *on* $\mathcal{B}(\mathcal{X})$, $\varepsilon < 1$, $\delta > 0$ *and* $m \in \mathbb{N}$ *such that:*

$$\varphi(A) > \varepsilon \Rightarrow \inf_{x \in \mathcal{X}} P^m(x, A) > \delta.$$

Under this condition, if the Markov chain X is aperiodic, it is uniformly exponentially ergodic.

Theorem ([MT09, Theorem 16.0.2]). *If* X *is an aperiodic Markov chain satisfying Assumption A, we have:*

$$\sup_{x \in \mathcal{X}} \| P^n(x, \cdot) - \pi \| \xrightarrow[n \to +\infty]{} 0,$$

and the convergence is exponentially fast.

This ergodic result is strong because it is uniform in the initial condition. For the structured population model studied in Chapter 3, assumptions that we consider are not sufficient for the proof of the uniform convergence. Under less demanding assumptions, we have the (non-uniform) ergodicity of the chain.

Assumption B. *There exists a function* $V : \mathcal{X} \to \mathbb{R}_+$, $c < 1$ *and* $d > 0$ *such that:*

$$PV(x) \leqslant cV(x) + d,$$

for all $x \in \mathcal{X}$.

This assumption, known as drift condition or Foster-Lyapunov condition, ensures that the process gets back to the set $\{x : V(x) < K\}$ for some $K > 0$ and enables to control the length of excursions outside this set [MT09, Theorem 11.3.4].

Assumption C. *There exist* $\alpha \in (0, 1)$ *and a probability measure* ν *such that:*

$$\inf_{x \in \mathcal{C}} P(x, \cdot) \geqslant \alpha \nu(\cdot),$$

with $\mathcal{C} = \{x : V(x) < R\}$ *and* $R > 2d(1 - c)^{-1}$ *where* c, d *are defined in Assumption B.*

This assumption is the local analogous of Doeblin's condition. It ensures some kind of uniform behavior in the initial condition for the process. Those two conditions are sufficient to obtain the ergodicity of the chain.

Theorem 1 ([HM11, Theorem 1.2]). *Under Assumptions B and C, the Markov chain with transition* P *admits a unique invariant measure* π *and there exist* $C > 0$ *and* $\gamma \in (0, 1)$ *such that:*

$$\| P^n f - \pi f \| \leqslant C \gamma^n \| f - \pi f \|,$$

for all measurable functionsf $: \mathcal{X} \to \mathbb{R}_+$ *such that* $\|f\| < \infty$, *where* $\pi f = \int f(x) \pi(dx)$ *and* $\|f\| = \sup_{x \in \mathcal{X}} |f(x)| \, (1 + V(x))^{-1}$.

For the detailed study of ergodicity for Markov chain in discrete time, we refer the reader to [MT09].

Continuous time. The conditions under which we can prove the ergodicity of a Markov process in continuous time are similar to the one in discrete time. Let $(X_t, t \geqslant 0)$ be a Markov process with values in a locally compact, separable metric space $\mathcal{X}$. For all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$, we denote by:

$$P_t(x, A) := \mathbb{P}_x(X_t \in A),$$

the semi-group of X. Let $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ be the infinitesimal generator of X and its associated domain.

The existence of invariant measure for a continuous-time Markov process is linked to the existence of invariant measures for its analogous in discrete time [Hai10].

Proposition. *Let $P := P_T$ for some $T > 0$. Then, if π is an invariant measure for P, the measure $\widehat{\pi}$ defined by:*

$$\widehat{\pi}(A) = \frac{1}{T} \int_0^T P_t \pi(A) dt,$$

is an invariant measure for the semi-group P_t.

In the case of continuous-time Markov process, the Foster-Lyapunov condition concerns the generator of X.

Assumption D. *There exists a measurable function $V : \mathcal{X} \to [0, \infty)$ and $c, K > 0$ such that:*

$$\mathcal{A}V(x) \leqslant -cV(x) + K, \quad \forall x \in \mathcal{X}.$$

There exists several equivalent formulation of this hypothesis. We refer the reader to [DMT95] for more details. The continuous-time equivalent of Assumption C uis usually written in terms of petite sets. This approach is developed in [MT93a, MT93b]. Here, we consider an other formalism.

Assumption E. *For all $t > 0$, there exists $\alpha_t \in (0, 1)$ and a probability measure ν_t on $\mathcal{X}$ such that:*

$$\inf_{x \in B(R,V)} P_t(x, \cdot) \geqslant \alpha_t \nu_t(\cdot),$$

where $B(R, V) = \{x \in \mathcal{X} : V(x) \leqslant R\}$ for some $R > \frac{2K}{c}$ where c, K are defined in Assumption D.

Those conditions ensure that the Markov process "forget" its initial condition.

Theorem 2. *Under Assumptions D and E, there exist $C > 0$ and $\overline{c} > 0$ such that:*

$$|P_t \varphi(x) - P_t \varphi(y)| \leqslant C e^{-\overline{c}t} d(x, y),$$

for all functions φ such that $\|\varphi\|_\infty \leqslant 1$, where d is a distance on $\mathcal{X}$.

We refer the reader to [Hai10] for more details on the subject and for conditions ensuring the sub-geometric convergence of the Markov process. In Chapter 2, we adapt Assumptions D and E to the case of time-inhomogeneous Markov processes. Therefore, we prove that the auxiliary process "forget" its initial condition, enabling the proof of the convergence of the empirical measure along ancestral lineages.

3 Statistical estimation for structured populations

The growing amount of data on cell populations, by means of microfluidic techniques, paves the way to statistical estimation. A major issue is the understanding of the mechanisms of cell division and more specifically, of the dependence of the division rate to some characteristics. We present here several models that have been considered to answer those questions.

3.1 Studies on aging for E. Coli

The first work of modeling and statistical estimation focusing on the study of Stewart et al. [SMPT05] is the work of Guyon [Guy07]. He models a cell population with symmetrical division such as the bacterium E. coli with bifurcating Markov chains. After a general study of the individual-based model, especially on the asymptotic behavior of the empirical measure, Guyon introduce a specific auto-regressive model for the study of aging of E. coli. Let $(X_n, n \geqslant 0)$ be a Markov chain satisfying for all $n \geqslant 0$:

$$\begin{cases} X_{2n} & = & \alpha_0 X_n + \beta_0 + \varepsilon_{2n}, \\ X_{2n+1} & = & \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases}$$

where $\alpha_0, \alpha_1 \in (-1, 1)$, $\beta_0, \beta_1 \in \mathbb{R}$ and $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geqslant 1)$ is a sequence of independent and identically distributed gaussian bivariate centered random variables with covariance matrix Γ. The sequence $(X_n, n \geqslant 0)$ corresponds to the growth rates of individuals in the population, $2n$ and $2n+1$ correspond to the daughter cells of the individual n inheriting respectively of the new pole and the old one (see Section 1). Guyon tests the couples of parameters (α_0, β_0) and (α_1, β_1) for equality by means of the maximum likelihood estimation of the parameters. The results of the statistical testing applied to the data of Stewart et al. [SMPT05] show a significant difference between the transmission parameters for the growth rate for the two daughter cells. Thus, there is an asymmetry in the transmission of the growth rate for the bacterium E. coli. This phenomenon indicates an aging phenomenon for this bacterium. For a generalization of the work of Guyon to a bifurcating Markov chain on a Galton-Watson tree, we refer the reader to [DM10].

3.2 Estimation of the division rate

Besides the existence of an aging phenomenon for E. Coli, a key question for the understanding of cell population dynamics is the functioning of the mechanism of cell division. Thus, many statistical works focus on the estimation of the division rate. The goal is to decide which characteristics trigger cell division. We refer the reader to the work of Robert et al. [RHK$^+$14] for the study of the dependence of the division rate on age or size by means of nonparametric estimation of the division rate using the data of Stewart et al. [SMPT05] and Wang et al. [WRP$^+$10]. We now present different statistical techniques developed to answer those questions.

The first works on the estimation of the division rate [PZ07, DPZ09] rely on a deterministic approach of the question. Using the exponential convergence of the solutions of growth-fragmentation equations to their stationary profile (see Section 2.2), the problem is reduced to the resolution of the following inverse problem: how to reconstruct the division rate B, solution of equation (5) of Section 2.2, from the couple (λ, N), where λ is the Malthus parameter and N the asymptotic profile of the population? One of the difficulties comes from the fact that the observed value N_ε is noisy so that

$$\|N_\varepsilon - N\| < \varepsilon.$$

Without any regularity assumption on the derivative of N_ε, the problem of reconstruction of B is ill-posed. A solution consists in "regularizing" equation (5) and looking at an well-posed approximate problem. Then, we need to control the norm difference between the solution to the approximate problem and the solution to the initial problem.

A limitation of this approach is that some properties of N are supposed to be known, such as regularity properties, and they seem hard to verify in practice from experimental data. Doumic et al. [DHRBR12] developed a statistical study of the problem to tackle this question. They gave a kernel estimator $\widehat{N}$ of N from individual cell measurements randomly picked in the population at a given time. By means of techniques from [DPZ09], they constructed an estimator of B using $\widehat{N}$ instead of N_ε which converges at rate $n^{-s/(2s+3)}$ to B, where s is a regularity coefficient corresponding to the imposed regularity on B.

Using individual-based models, we can obtain a better speed of convergence for the estimator by taking into account the structure of genealogies. In the work of Doumic et al. [DHKR15], they construct an estimator of the division rate from a probabilistic model for the cell population. More precisely, they consider a size-structured population where each individual u grows exponentially at rate τ_u. We denote by ξ_u the size at birth of u. Then, if $\xi_u(t)$ denotes the size u at time t, we have:

$$\xi_u(t) = \xi_u e^{\tau_u t}.$$

An individual of size x divides at rate $B(x)$ and at division, it is replaced by two individuals with size $x/2$. The growth rate of each daughter cell is distributed independently according to the kernel $\rho(v, \cdot)$, where v is the growth rate of their ancestor. A key object for the statistical study of the branching Markov process is the transition kernel associated with the couple (ξ_u, τ_u) given by:

$$P_B\left(\left(\xi_{u^-}, \tau_{u^-}\right), \left(\xi_u, \tau_u\right)\right) = \frac{B(2\xi_u)}{\xi_u \tau_{u^-}} \mathbf{1}_{\{\xi_u \geqslant \xi_{u^-}/2\}} \exp\left(-\int_{\xi_{u^-}/2}^{\xi_u} \frac{B(2s)}{s\tau_{u^-}} ds\right) \rho\left(\tau_{u^-}, \tau_u\right),$$

where u^- corresponds to the ancestor of u. An estimator of the division rate B can be constructed from the following formula:

$$B(x) = \frac{x}{2} \frac{\nu_B(x/2)}{\mathbb{E}_{\nu_B}\left[\frac{1}{\tau_{u^-}} \mathbf{1}_{\{\xi_{u^-} \leqslant x,\ \xi_u \geqslant x/2\}}\right]},$$

where ν_B denotes the invariante measure associated with the kernel P_B. Doumic et al. [DHKR15] give an estimator for the division kernel and prove the $\mathbb{L}^2$-convergence of this estimator at rate $\log(n) n^{-s/(2s+1)}$, where s is again a regularity parameter for B. The rate of convergence is then better than the one obtained in [DHRBR12]. We refer the reader to the work of Bitseki-Penda et al. [BPHO17] for an improvement of the speed of convergence of the estimator by means of Bernstein like deviation inequalities.

In the aforementioned works, the data are assumed to be genealogical, meaning that we have available the characteristics of all individuals in the population until the nth generation. However, for cell populations, the generations overlap (see Figure 2.1 in the French introduction) and all individuals alive at a given time T do not belong to the same generation. Thus, if the considered data are composed of traits of all individuals in the tree until time T, the estimation procedures detailed above cannot be used because the data are biased: individuals that reproduce slower are less represented in the population stopped at time T than in the population at generation n where all individuals are inventoried, whatever their lifetime. Hoffmann and Olivier [HO16] gave an estimator of the division rate for such data in the case of an age-structured model. They considered a Bellmann-Harris process where each individual lives for a time distributed as $f(x)dx$. At death, an individual is replaced by a random number of descendants distributed as $(p_k, k \geqslant 2)$, each of them also having a lifetime distributed as $f(x)dx$. The division rate B is then reconstructed from the following formula:

$$B(x) = \frac{f(x)}{1 - \int_0^x f(y)dy}, \quad \text{for } x \geqslant 0.$$

From this representation and convergence results for empirical measures obtained by means of spinal techniques, they gave a kernel estimator of B and prove the exponential convergence in probability of this estimator towards the division rate. One of the specificities of the age-structured model is that the characteristic of the individuals at birth are all equal. This is not the case any more for size-structured models so that the estimation of parameters is more complex for those models.

3.3 Estimation of the fragmentation kernel

As explained in Section 1, the way cells divide (by bugging or dividing in two equal parts) plays a key role in the mechanism of cell aging. Then, the estimation of the fragmentation kernel is also an important step in the understanding of this phenomenon.

Hoang [Hoa15] gave an estimator of the fragmentation kernel in the case of a size-structured population. In his model, each individual has a certain degree of toxicity that grows linearly with time at rate $\alpha > 0$. This modeling of cellular aging by the accumulation of toxicity is inspired from the work of Stewart et al. [SMPT05]. Cells divide at a constant rate and when a cell divides, its toxicity is distributed to its two daughter cells according to the random fraction $\Gamma \in (0,1)$ with probability density function h. Hoang gave an adaptative estimator of h for which he proved an oracle inequality. The construction of this estimator relies on data of the population on $[0,T]$. For the construction of an estimator of the fragmentation kernel in the case of a varying division rate and data of the population at a given time T, we refer the reader to [Hoa16].

4 Overview of the thesis

This thesis consists of three chapters. The first two chapters are dedicated to the probabilistic study of branching structured Markov processes. The third chapter consists of the statistical study of these models.

We consider a measure-valued branching Markov process $(Z_t, t \geqslant 0)$ which characteristics are the following:

- each individual u in the population has a trait $(X_t^u, t \geqslant 0)$ which dynamic follows a Markov process with associated generator and domain $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$,
- u dies at time t at rate $B(X_t^u)$,
- at death, an individual with trait x is replaced by a random number of descendants distributed according to $(p_k(x), k \geqslant 0)$,
- the trait at birth of the jth descendant among k is distributed according to $P_j^{(k)}(x, \cdot)$, for all $1 \leqslant j \leqslant k$.

The object of study is the measure-valued process Z represented by:

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

where V_t corresponds to the individuals alive at time t. In the first chapter, we are interested in the characterization of the trait along its ancestral lineage of a uniformly sampled individual in the population at time t. We exhibit the bias due to the uniform sampling in the case of a division rate depending on the trait. Chapter 2 focus on the asymptotic behavior of the empirical measure associated with the branching process. This asymptotic study relies on the ergodicity of the auxiliary process introduced in the first chapter. Finally, in the third chapter, we are interested in the statistical estimation of the parameters of a branching population structured by a diffusion. It is a particular case of the model studied in the first two chapters.

4.1 Chapter 1: uniform sampling in a branching structured population

This chapter consists of the article "Uniform sampling in a structured branching population" [Mar16], accepted for publication in Bernoulli in July 2017.

In this chapter, we study the uniform sampling of individuals in a structured branching population. The problematic is the following: if you pick uniformly at random an individual in the population at time t, what are its characteristics and the characteristics of its ancestors along its ancestral lineage?

The first result is the existence and uniqueness of the process Z previously described. We prove this result under general assumptions. In particular, we do not assume that the division rate is bounded by a constant but that it is bounded by a polynomial in order to ensure the non-explosion of the population process in finite time. Theorem 1.2.2 of Chapter 1 ensures that the measure-valued process Z is the unique strong solution of a stochastic differential equation with jumps.

Then, we study the penalized process associated with the semi-group $(P_{r,s}^{(t)}, r \leqslant s \leqslant t)$ previously defined in (7). The main result of this chapter consists of the characterization by the infinitesimal generator of this process of the trait along its ancestral lineage of a uniformly sampled individual in the population.

Theorem (Theorem 1.3.1)**.** *Under assumptions ensuring the non-explosion of the branching process in finite time, for all measurable positive functions F on the space of càdlàg processes (right continuous with left limits), and for all $t \geqslant 0$ and $x \in \mathcal{X}$,*

$$\mathbb{E}\left[\sum_{u \in V_t} F\left(X_s^u, s \leqslant t\right)\right] = m(x,0,t)\mathbb{E}\left[F\left(Y_s^{(t)}, s \leqslant t\right)\right], \tag{8}$$

where $\left(Y_s^{(t)}, s \leqslant t\right)$ is a time-inhomogeneous Markov process with generator $\left(\mathcal{A}_s^{(t)}, s \leqslant t\right)$ satisfying:

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_s^{(t)} f(x) + \widehat{B}_s^{(t)}(x) \int_{\mathcal{X}} \left(f\left(y\right) - f\left(x\right)\right) \widehat{P}_s^{(t)}\left(x, dy\right),$$

with:

$$\widehat{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}\left(m(\cdot,s,t) f\right)(x) - f\left(x\right) \mathcal{G}\left(m(\cdot,s,t)\right)(x)}{m(x,s,t)},$$

$$\widehat{B}_s^{(t)}(x) = B(x) \int_{\mathcal{X}} \frac{m(y,s,t)}{m(x,s,t)} m(x,dy),$$

$$\widehat{P}_s^{(t)}(x,dy) = m(y,s,t) m(x,dy) \left(\int_{\mathcal{X}} m(y,s,t) m(x,dy)\right)^{-1},$$

and:

$$m(x,A) := \sum_{k\geqslant 0} p_k(x) \sum_{j=1}^{k} P_j^{(k)}(x,A),$$

corresponds to the mean number of descendants whose trait is in A of an individual with trait x at death.

We see that the time-inhomogeneity of the auxiliary process comes from the dependence of the mean number of individuals in the population on the trait. Moreover, the bias corresponding to the growth of the population are explicit. This auxiliary process jumps more often than any individual in the population, if jumping is beneficial in term of number of descendants. This modification of the division rate also appears in [CR88, LPP95, HH09, BDMT11]. The reproduction law is biased in the same way. Finally, the dynamic of the trait between divisions is also biased because of the dependence of the division rate on the trait and of the stochastic behavior of this dynamic. In particular, in the case of a deterministic dynamic, we have $\widehat{\mathcal{G}}_s^{(t)} f(x) = \mathcal{G} f(x)$. We also prove other Many-to-One formulas that give the characterization, in terms of auxiliary process, of the trait of a couple of individuals or of all individuals until time t. These formulas are especially useful for the study of the asymptotic behavior of the empirical measure addressed in Chapter 2.

The auxiliary process is a natural candidate for the characterization of the law along its ancestral lineage of a uniformly sampled individual in the population. However, this is a penalized process which take into account the bias due to the growth of the population, and these bias do not appear instantaneously at the beginning of the process in the population. Then, the law of the trait of a uniformly sampled individual in the population at time t is characterized by the auxiliary process only for large times. In Chapter 1, we consider the trait of a uniformly sampled individual when the size of the initial population goes to infinity in order to mimic the asymptotic behavior of the population. The long time behavior is addressed in Chapter 2. Let $X^{U(t),\nu}$ be the trait of a uniformly sampled individual in the population at time t with initial distribution ν. If $\nu_n = \sum_{i=1}^{n} \delta_{X_i}$, where the X_i's are independent identically distributed as ν, we prove the convergence in law of $X^{U(t),\nu_n}$ as n goes to infinity.

Theorem (Theorem 1.4.1). *Under assumptions ensuring the non-explosion of the branching process in finite time, the sequence $\left(X_{[0,t]}^{U(t),\nu_n}, n \geqslant 0\right)$ converges in law in the space of Skorokhod towards $Y_{[0,t]}^{(t),\pi_t}$ where*

$$\pi_t(dx) = \frac{m(x,0,t)\nu(dx)}{\int m(y,0,t)\nu(dy)}$$

is the initial law of $Y_{[0,t]}^{(t),\pi_t}$.

Then, the auxiliary process characterizes the law of a uniformly sampled individual in the population, when the initial population goes to infinity.

Finally, we illustrate the preceding results by concret examples of models for cell populations. We consider a simplification of the size-structured model studied in [DHKR15] where each individual growth exponentially at rate $a > 0$ and dies at rate $B(x) = x$. The mean number of individuals in the population is given by:

$$m(x,s,t) = 1 + \frac{x}{a}\left(e^{a(t-s)} - 1\right),$$

for $x \in \mathbb{R}_+$, and $0 \leqslant s \leqslant t$. The generator of the auxiliary process is then given by:

$$\mathcal{A}_s^{(t)} f(x) = axf'(x) + x\left(1 + \frac{1}{1 + \frac{x}{a}(e^{a(t-s)} - 1)}\right)\left(f\left(\frac{x}{2}\right) - f(x)\right).$$

In this example, the only bias is the time acceleration i.e. the division rate is higher than the division rate of any individual in the population. In Section 1.4.2 of Chapter 1, we give a generalization of this example to a more general family of division rates, possibly time-dependent. We also detail the case of a size-structured population where each individual grows lineary and the case of parasites proliferation also studied in [BT11].

4.2 Chapter 2: law of large numbers for branching Markov processes by the ergodicity of ancestral lineages

This chapter consists of a preprint entitled " A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages" [Mar17].

This chapter addresses the question of the asymptotic behavior of the empirical measure associated with the branching process Z given by:

$$\frac{1}{N_t} \sum_{u \in V_t} \delta_{X_t^u},$$

where N_t is the cardinality of V_t. By means of the study of the genealogies and the ergodicity of the auxiliary process Y along its ancestral lineage, we prove the $\mathbb{L}^2$ convergence of the empirical measure towards the asymptotic distribution of Y. Then, the random measure characterizing the trait of a uniformly sampled individual in the population is asymptotically deterministic and corresponds to the limiting distribution of Y. Our approach is based on the use of the infinitesimal generator of the auxiliary process given in the previous chapter.

To shorten notation, we consider in this chapter only the case of a binary division. Then, when he died, an individual with trait x is replaced by two individuals which traits at birth are given by the probability kernel $P(x, \cdot)$. The results of this chapter can easily be extended to the case of a process with a random number of descendants.

The study of the asymptotic behavior of the empirical measure has already been considered in numerous cases. Asmussen and Hering [AH76] proved the convergence of the empirical measure for branching processes in discrete or continuous time by means of an assumption ensuring a useful decomposition of first-moment semi-group. In particular, they address the case of branching diffusions, also studied in [EHK10]. Athreya and Kang proved the convergence of the empirical measure for a branching process in discrete time with a discrete trait [AK98a] or continuous trait [AK98b]. We also mention the work of Guyon [Guy07] on bifurcating Markov chains and the generalization of those results to the case of a Galton-Watson tree by Delmas and Marsalle [DM10]. For results in random environment, we refer the reader to [BH15, Ban15]. In continuous time and for a continuous trait, Georgii and Baake [GB03] proved a law of large numbers on the type of individuals along an ancestral lineage. Their results rely on spine techniques developped by Lyons et al. [LPP95] and detailed in Section 2.1. Harris and Robert [HR14] use the spine construction of [HH09] to prove the convergence of the empirical measure in the case of a branching Markov processes with local branching. For central limit theorems, we refer the reader to [Guy07, RSZ14].

One way of proving the convergence of the empirical measure relies on the existence of eigenelements for the infinitesimal generator of the first moment semi-group [LPP95, GB03, EHK10]. We use an other approach that does not require the use of a spectral framework. It is based on the use of the auxiliary process defined in Chapter 1. Using the ergodic behavior of this penalized process, we prove the convergence of the empirical measure towards the limiting distribution of the auxiliary process. This technique has already been developped in discrete time by Athreya and Kang [AK98b] and Guyon [Guy07], and in continuous time by Bansaye et al. [BDMT11] in the neutral case, by Cloez [Clo17] in a general framework combining the two approaches. In Chapter 2, we prove the convergence of the empirical measure by means of the ergodicity of the auxiliary process in a general framework without the use of eigenelements.

As detailed in Chapter 1, in the case of a non-constant division rate, the auxiliary process is time-inhomogeneous, even if the dynamic of the trait itself is time-homogeneous. We adapt the work of Hairer and Mattingly [HM11] to the case of a time-inhomogeneous Markov process to prove the ergodicity of the auxiliary process.

Assumption F. *We assume that*

1. *there exists a function* $V : \mathcal{X} \to \mathbb{R}_+$ *and* $c, d > 0$ *such that for all* $x \in \mathcal{X}$*,* $t \geqslant 0$ *and* $s \leqslant t$*:*
$$\mathcal{A}_s^{(t)} V(x) \leqslant -cV(x) + d,$$
2. *for all* $0 < r < s$*, there exists* $\alpha_{s-r} \in (0,1)$ *and a probability measure* $\nu_{r,s}$ *on* $\mathcal{X}$ *such that for all* $t \geqslant s$*:*
$$\inf_{x \in B(R,V)} P_{r,s}^{(t)}(x, \cdot) \geqslant \alpha_{s-r} \nu_{r,s}(\cdot),$$
with $B(R,V) = \{x \in \mathcal{X} : V(x) \leqslant R\}$ *for a given* $R > \frac{2d}{c}$ *with* c, d *defined in the first point.*

This assumption gathers the two classical conditions used to prove the ergodicity of a process (see Section 2.4): a Foster-Lyapunov condition and the second a minorization condition. Under these assumptions, we prove the ergodicity of the trajectory of the auxiliary process.

Proposition (Proposition 2.3.1)**.** *Under Assumption F, there exists* $\overline{c} > 0$ *such that for all* $x, y \in \mathcal{X}$*,* $T > 0$*, for all measurable bounded functions* $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$ *and all* $0 \leqslant r \leqslant t$*,*

$$\left| P_{r,t,T} F(x) - P_{r,t,T} F(y) \right| \leqslant C e^{-\overline{c}(t-r)} d(x,y) \left\| F \right\|_\infty,$$

where d *is a distance on* $\mathcal{X}$*,* C *is a positive constant and*

$$P_{r,t,T} F(x) := \mathbb{E}\left[F\left(Y_{t+s}^{(t+T)}, s \leqslant T \right) \middle| Y_r^{(t+T)} = x \right]. \tag{9}$$

This result allows to control the entire trajectory of the time-inhomogeneous process on an interval.

In the same spirit of [BDMT11] but with a non-constant division rate, we prove a law of large numbers result by means of the previous proposition.

Theorem (Proposition 2.3.3)**.** *Under assumptions ensuring the non-explosion of the branching process in finite time and under assumption F, for all* $x_0, x_1 \in \mathcal{X}$ *and* $T > 0$*, we have:*

$$\left(\frac{\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T \right)}{N_{t+T}} - \mathbb{E}_{x_1}\left[F\left(Y_{t+s}^{(t+T)}, s \leqslant T \right) \right] \right) \xrightarrow[t \to +\infty]{} 0,$$

where the convergence holds in $\mathbb{L}^2$ *with initial measure* δ_{x_0}*.*

Then, the dynamic of the auxiliary process corresponds asymptotically to the dynamic of the trait along ancestral lineages of a uniformly sampled individual.

We illustrate this result with the detailed study of a size-structured population where each individual grows exponentially at rate $a > 0$, divides at rate $B(t,x) = x\varphi(t)$ at time t if its size at time t is x and give birth to two individuals with sizes θx and $(1-\theta)x$, where θ is uniformly distributed on $[\varepsilon, 1-\varepsilon]$, where $\varepsilon > 0$. This model has already been studied in [DHKR15] and in Chapter 1. We prove the exponential ergodicity of the trajectory of the auxiliary process associated with this model and the convergence of the empirical measure of the branching process.

4.3 Chapter 3: statistical estimation for branching Markov processes

This chapter consists of a work in collaboration with Marc Hoffmann. It is dedicated to the statistical estimation of the different parameters that come into play in the dynamic of a cell population. We study a particular case of the branching process Z previously described. We assume that the dynamic of the trait of each individual follows a diffusion on $\mathbb{R}$:

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x \in \mathbb{R},$$

where $(W_t, t \geqslant 0)$ is a standard Brownian motion. Each individual dies at rate $B(\cdot)$ depending on its trait and is replaced by two individuals. The traits at birth of the descendants of an individual with trait x are given by θx and $(1-\theta)x$, where θ is a random variable on $[0,1]$ distributed according to a probability density κ symmetrical with respect to 1/2. Contrary to the two previous chapters, we consider here the Markov chain in discrete time given by the trait at birth of individuals in the population. Hence, we adopt a genealogical point of view rather than chronological.

We use again the Ulam-Harris-Neveu notation introduced in Section 2.1. We denote by $\mathbb{G}_n \subset \mathcal{U}$ the set of individuals belonging to the nth generation and by

$$\mathbb{T}_n = \bigcup_{m=0}^{n} \mathbb{G}_n,$$

the set of all individuals in the population until the nth generation. For all $u \in \mathcal{U}$, we denote by u^- the parent of u. We are interested in the reconstruction of the different parameters of the model from observed data given by

$$\mathbb{X}^n := (X_u, u \in \mathbb{T}_n),$$

where X_u denotes the trait at birth of u.

From an experimental point of view, obtaining such data is realistic for example by means of recent microfluidic techniques detailed in Section 1.

The Markov chain $(X_u, u \in \mathcal{U})$ is a bifurcating Markov chain [Guy07, DM10] on $\mathcal{X}$ with transition $\mathcal{P} : \mathcal{X} \to \mathcal{X} \times \mathcal{X}$. The transition $\mathcal{P}$ describes the transmission of the trait to the two descendants.

We consider the process of the tagged-branch denoted by Y. It corresponds to the uniform choice of a branch in the genealogical tree. More precisely, we have for all $m \in \mathbb{N}$:

$$Y_m = Y_{\varnothing\epsilon_1\ldots\epsilon_m},$$

where $(\epsilon_k, k \geqslant 0)$ are independent Bernoulli random variables with parameter 1/2. As the time dynamic is genealogical, the value of Y_m corresponds to the uniform choice of an individual among the individuals in generation m. We denote by:

$$\mathcal{Q} = \frac{\mathcal{P}_1 + \mathcal{P}_2}{2},$$

the transition associated with the Markov chain Y, where $\mathcal{P}_1$ and $\mathcal{P}_2$ are the marginals of the transition $\mathcal{P}$. Note that in our case, $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{Q}$.

As detailed in Section 3, the statistical estimation in structured population models focus essentially on cases where the dynamic of the trait is deterministic. In the case of a branching diffusion, Höpfner et al. [HHL02] gave a kernel estimator of the death rate B. Their approach relies on the fact that the trait at birth of the individuals is identical to the trait at death of its ancestor, which is not the case in our model.

Under regularity assumptions, we prove that $\mathcal{Q}(x, dy) = \mathcal{Q}_B(x, dy) = q_B(x, y)dy$ with

$$q_B(x,y) = \int_0^1 B\left(\frac{y}{z}\right)\sigma\left(\frac{y}{z}\right)^{-2}\mathbb{E}\left[\int_0^{+\infty} e^{-\int_0^t B(\phi_x(s))ds}\, dL_t^{\frac{y}{z}}(\phi_x)\right]\kappa(z)\frac{dz}{z},$$

where $L_t^y(\phi_x)$ denotes the local time at location y and time t of the semi-martingale $(\phi_x(t), t \geqslant 0)$. The integral with respect to the local time results from the fact that the lifetime of individuals depends on the stochastic dynamic of the trait. Therefore, the probability to jump from x to y in one generation depends on the average time spent by the diffusion $(\phi_x(t), t \geqslant 0)$ in the interval $[y, +\infty)$. In particular, one of the difficulties of the model comes from the fact that the density q_B depends one the average of the whole trajectory of the trait $(\phi_x(t), t \geqslant 0)$ rather than on the trajectory at a given time t.

Nonparametric estimation. First, we focus on the asymptotic behavior of the Markov chain Y. Under appropriate conditions on the dynamic of the diffusion, we prove that Assumptions B and C of Section 2.4 are satisfied by the transition $\mathcal{Q}$, so that the Markov chain Y is ergodic.

Theorem (Theorem 3.2.2)**.** *Under appropriate assumptions on the dynamic of the flow,* $\mathcal{Q}$ *admits an invariant measure* ν*. Moreover, there exists* $C > 0$ *and* $\rho \in (0, 1)$ *such that for all* $x \in \mathbb{R}$ *and* $m \in \mathbb{N}$*:*

$$|\mathcal{Q}^m\varphi(x) - \nu(\varphi)| \leqslant C\rho^m(1 + V(x))\,\|\varphi - \nu(\varphi)\|_\infty\,.$$

for all measurable functions $\varphi : \mathbb{R} \to \mathbb{R}$ *such that*

$$\sup_{x\in\mathbb{R}} \frac{\varphi(x)}{1 + V(x)} < \infty,$$

with $V(x) = x^2$*.*

Therefore, we do not prove the uniform ergodicity of Y but a weaker property that depends on the initial condition. In the case of a reflected diffusion, the trait takes its values in a compact and we prove the uniform ergodicity of the chain Y by checking Doeblin's condition A of Section 2.4.

Using this ergodicity result, we can study the asymptotic behavior of the empirical measure given by:

$$\mathcal{M}_n(\psi) = \frac{1}{|\mathbb{T}_n^\star|}\sum_{u\in\mathbb{T}_n^\star} \psi(X_{u^-}, X_u),$$

for all test functions $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, where $\mathbb{T}_n^\star = \mathbb{T}_n\backslash\{\varnothing\}$. It converges as n tends to infinity towards $\nu(\mathcal{Q}\psi)$. Our estimates for the density q and the invariant measure ν rely on approximation kernels.

Definition. *A function* $G : \mathcal{X} \to \mathbb{R}$ *is a kernel of order* k *if it is compactly supported and satisfies* $\int_{\mathcal{X}} x^l G(x)dx = \mathbf{1}_{\{l=0\}}$ *for* $l = 0, \ldots, k$*.*

For the convergence of kernel estimators of ν and $\mathcal{Q}$, we need to control the dependence in ψ of the speed of convergence of the empirical measure. More precisely, if we consider the kernel $G_h := h^{-1}G(\cdot/h)$, where h is a bandwidth parameter, we have:

$$|G_h|_1 \lesssim 1, \quad |G_h|_2^2 \lesssim h^{-1}, \quad |G_h|_\infty^2 \lesssim h^{-2}.$$

The speed of convergence of $\mathbb{E}_\mu\left[\mathcal{M}_n(G_h)^2\right]$ must have the right order in h, namely h^{-1}, to obtain a good speed of convergence for the kernel estimators. We refer the reader to the proof of Theorem 3.3.3 in Chapter 3 for more details.

For all functions $\psi : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$, we set $\psi^\star(x) = \sup_{y\in\mathcal{X}} |\psi(x,y)|$, $\psi_\star(y) = \sup_{x\in\mathcal{X}} |\psi(x,y)|$ and

$$|\psi|_{\wedge 1} = \int_{\mathcal{X}\times\mathcal{X}} |\psi(x,y)|dxdy \bigwedge \int_{\mathcal{X}} \sup_{x\in\mathcal{X}} |\psi(x,y)|dy.$$

For all probability measures μ on $\mathcal{X}$, we set:

$$|\psi|_\rho = \int_{\mathcal{X}\times\mathcal{X}} |\psi(x,y)|\rho(dx)dy + |\psi|_{\wedge 1}.$$

A sharp analysis of the variance of the empirical measure along genealogies allows us to obtain the desired rate of convergence with a control on ψ of ordrer $|\psi|_2^2$.

Theorem (Theorem 3.2.3)**.** *Let μ be a probability measure on $\mathbb{R}$ such that $\mu(V^2) < \infty$. Let $\psi : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a bounded function such that $\psi_\star$ is compactly supported. Then, under assumptions ensuring the ergodicity of the Markov chain Y and if $\rho \leqslant \frac{1}{2}$, we have for all $n \in \mathbb{N}$:*

$$\mathbb{E}_\mu\big[\big(\mathcal{M}_n(\psi) - \nu(\mathcal{Q}\psi)\big)^2\big] \lesssim |\mathbb{T}_n|^{-1}\big(|\psi^2|_{\mu+\nu} + |\psi^\star\psi|_\mu + \big(1+\mu(V^2)^{1/2}\big)|\psi_\star|_1|\psi|_\nu\big),$$

where the symbol $\lesssim$ means up to a constant depending only on $\mathcal{Q}$ and supp(ψ).

For a given kernel G, we consider the kernel G_h, whith bandwidth parameter h, defined by $G_h(x) = h^{-1}G(h^{-1}x)$ for $x \in \mathcal{X}$. For all $x_0 \in \mathcal{X}$,

$$\mathcal{M}_n(G_h(\cdot - x_0)) \xrightarrow[n\to+\infty]{\mathbb{P}} \int_{\mathcal{X}} G_h(x-x_0)\nu(x)dx.$$

A kernel estimator for ν is then given by:

$$\widehat{\nu}(x_0) = \mathcal{M}_n(G_h(\cdot - x_0)). \tag{10}$$

In the same way, if we consider $G^{\otimes 2}_{h_1,h_2}(x,y) = h_1^{-1}h_2^{-1}G(h_1^{-1}x)G(h_2^{-1}x)$, we have for all $x_0, y_0 \in \mathcal{X}^2$:

$$\mathcal{M}_n\left(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} \int_{\mathcal{X}\times\mathcal{X}} G^{\otimes 2}_{h_1,h_2}(x-x_0, y-y_0)q(x,y)\nu(x)dx.$$

An kernel estimator of the density q is then given by:

$$\widehat{q}_n(x_0,y_0) = \frac{\mathcal{M}_n\left(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\right)}{\mathcal{M}_n(G_h(\cdot - x_0)) \vee \varpi}, \tag{11}$$

where the denominator balances the superfluous weight ν in the numerator and $\varpi > 0$ is a threshold ensuring the good definition of the quotient. These estimators are consistent.

Theorem (Theorem 3.3.3)**.** *Let* $\alpha, \beta > 0$*. Specify* $\widehat{\nu}_n(x_0)$ *defined in* (10) *by a kernel of order* k *and a bandwidth parameter* $h = |\mathbb{T}_n|^{-1/(2\beta+1)}$ *and* $\widehat{q}_n(x_0, y_0)$ *defined in* (11) *by the same kernel and bandwidth parameters* $h_1 = |\mathbb{T}_n|^{-s(\alpha,\beta)/(\alpha\wedge\beta)(2s(\alpha,\beta)+1)}$*,* $h_2 = |\mathbb{T}_n|^{-s(\alpha,\beta)/\beta(2s(\alpha,\beta)+1)}$ *and* $\varpi_n \to 0$*. Then, under assumptions ensuring the ergodicity of the Markov chain* Y*, we have*

$$\Big(\mathbb{E}\big[\big(\widehat{\nu}_n(x_0) - \nu(x_0)\big)^2\big]\Big)^{1/2} \lesssim |\mathbb{T}_n|^{-\beta/(2\beta+1)},$$

and

$$\Big(\mathbb{E}\big[\big(\widehat{q}_n(x_0, y_0) - q(x_0, y_0)\big)^2\big]\Big)^{1/2} \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)},$$

with $s(\alpha,\beta)^{-1} = (\alpha\wedge\beta)^{-1} + \beta^{-1}$*. Moreover, these bounds are uniform in* $\mathcal{Q}$ *for* $\mathcal{Q}$ *in a given Hölder class depending on* α *and* β*.*

Parametric estimation of the division rate. In this part, we focus on the statistical reconstruction of the division rate B by means of maximum likelihood estimation. This question is complex and we need to consider extra assumptions. The main difficulty comes from the fact that the previous result on the speed of convergence of the empirical measure only apply to bounded functions ψ. However the function $\psi(x, y) = \log(q(x, y))$ is not necessarily bounded. To circumvent this difficulty, we consider that case of branching Markov processes with a trait dynamic following a reflected diffusion on a compact set $[0, L]$, for some $L > 0$. The dynamic of the trait is then given by the following stochastic differential equation:

$$\begin{cases} dX_t = r(X_t)dt + \sigma(X_t)dW_t + d\ell_t, \\ \ell_t = \int_0^t \left(\mathbf{1}_{X_s=0} + \mathbf{1}_{X_s=L}\right) d\ell_s, \end{cases}$$

where $(W_t, t \geqslant 0)$ is a standard Brownian motion. The second equation corresponds to the reflections at the edge of the domain. The density function q of the Markov chain Y is slightly modified. In particular, its support is limited because of the reflections.

The proofs of the results obtained in the case of a standard diffusion easily adapt to the case with reflections. The compactness of the domain allow us to prove the uniform ergodicity of the process Y and the empirical measure still converges at speed $\mathbb{T}_n^{-1}$.

We now assume that the division rate belongs to the following parametric class of functions:

$$\mathcal{B} = \{B : \mathcal{X} \to \mathbb{R}, B(x) = B_0(\vartheta, x), x \in \mathcal{X}, \vartheta \in \Theta\},$$

where $x \mapsto B_0(\vartheta, x)$ is known up to the parameter ϑ and $\Theta \subset \mathbb{R}_+^d$ is a compact set of parameters characterizing the division rate. We prove the identifiability of the model for a certain class of ordered division rates (see Definition 3.3.5 and Proposition 3.3.6).

In order to emphasize the dependence in ϑ of the different quantities, we add subscripts to previously introduced notation. We assume that we have observations $(X_u, u \in \mathbb{T}_n)$ of the process until generation n drawn according to a certain law $\mathcal{P}_\vartheta$. We consider the maximum likelihood estimator (MLE) given by:

$$\widehat{\vartheta}_n \in \underset{\vartheta\in\Theta}{\operatorname{argmax}} \left\{ \frac{1}{\mathbb{T}_n} \sum_{u\in\mathbb{T}_n} \log\left(q_\vartheta\left(X_{u^-}, X_u\right)\right) \right\}.$$

The dynamic that we chose for the trait allows us to bound the density of the transition kernel and thus to prove the consistence of the MLE.

Theorem (Theorem 3.3.7). *Under regularity assumptions for the division rate,* $\widehat{\vartheta}_n$ *converges in probability towards* ϑ *when* n *tends to infinity.*

Finally, we prove the asymptotic normality of the MLE. Let $\Psi(\vartheta)$ be the Fisher information at point $\vartheta \in \Theta$ as the $d \times d$ matrix with entries:

$$\Psi(\vartheta)_{i,j} = \nu_\vartheta \mathcal{Q}(\vartheta) \left(\frac{\partial_{\vartheta_i} q_\vartheta \partial_{\vartheta_j} q_\vartheta}{q_\vartheta^2} \right),$$

for $1 \leqslant i, j \leqslant d$.

Theorem (Theorem 3.3.9). *Under regularity assumptions and if* $\Psi(\vartheta)$ *is nonsingular, for all* ϑ *in the interior of* Θ*, we have:*

$$\mathbb{T}_n^{1/2} \left(\widehat{\vartheta}_n - \vartheta \right) \to \mathcal{N} \left(0, \Psi(\vartheta)^{-1} \right),$$

in law as n *tends to infinity, where* $\mathcal{N}\left(0, \Psi(\vartheta)^{-1}\right)$ *denote the* d*-dimensional Gaussian distribution with mean* 0 *and covariance the inverse of the Fisher matrix* $\psi(\vartheta)$*.*

The numerical implementation of the maximimum likelihood estimator $\widehat{\vartheta}_n$ is a challenging problem that we plan to address fully. Let us mention that for now, we did not find an other constrat than the MLE for the estimation of $x \mapsto B(x)$, in particular, we do not know how to estimate it nonparametrically.

Chapter 1

Uniform sampling in a structured branching population

Abstract

We are interested in the dynamics of a structured branching population where the trait of each individual moves according to a Markov process. The rate of division of each individual is a function of its trait and when a branching event occurs, the trait of the descendants at birth depends on the trait of the mother and on the number of descendants. In this chapter, we explicitly describe the penalized Markov process, named auxiliary process, corresponding to the dynamics of the trait along the spine by giving its associated infinitesimal generator. We prove a Many-to-One formula and a Many-to-One formula for forks. Furthermore, we prove that this auxiliary process characterizes exactly the process of the trait of a uniformly sampled individual in the large population approximation. We detail three examples of growth-fragmentation models: the linear growth model, the exponential growth model and the parasite infection model..

Content

1.1 Introduction

The characterization of the sampling of individuals in a population is a key issue for branching processes with several motivations in statistics and biology. We refer to the work of Durrett [Dur78] and references therein for the study of the genealogy of a branching Markov process and the study of the degree of relationship between k individuals chosen randomly at time t in the population. In particular, he analyzed the asymptotics of the so-called reduced branching process $N_t(s)$ defined as the number of individuals alive at time s which have offspring alive at time t. An approximation of this process by a pure birth process is given in [O'C95]. The question of finding the coalescing time of individuals in a Galton-Watson tree is addressed in [Zub76] and the coalescent structure of continuous-time Galton-Watson trees is studied in [HJR17]. We refer to [Ath12, LP13] for more results on this question and to [Hon11] for results concerning the Bellman-Harris branching process. The pedigree of a typical individual in a supercritical branching process has also been investigated asymptotically for multi-type branching processes with a finite number of types in [GB03], with i.i.d life-times in [AAI11] and with an age-structure in [NJ84]. The characterization of the sampling is the key to obtain asymptotic results on the branching process [KLPP97, BDMT11, Clo17] and to infer the parameters of the model [Guy07, DHKR15, HO16].

In this chapter, we consider a continuous-time structured branching Markov process where the trait of each individual moves according to a Markov process and influences the branching events. The purpose of this chapter is to characterize the trait of a typical individual uniformly sampled from the population at time t and its associated ancestral lineage. In particular, we exhibit the bias due to the structure of the population and to the sampling. We also describe the traits of a uniformly sampled couple in the current population. Therefore, we provide new applications in a non-neutral framework for cell division (Section 1.2.2), even for models in a varying environment.

We now describe informally the process, while its rigorous construction and characterization as a càdlàg measure-valued process under Assumptions A and B are detailed in Section 1.2. We assume that individuals behave independently and that for each individual u:

- its trait $(X_t^u)_{t\geqslant 0}$ evolves as an $\mathcal{X}$-valued Markov process with infinitesimal generator $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$, where $\mathcal{X} \subset (\mathbb{R}_+)^d$ is a measurable space for some $d \geqslant 1$,
- it dies at time t at rate $B(X_t^u)$,
- at its death, an individual with trait x is replaced by $k \in \mathbb{N}$ individuals with probability $p_k(x)$ and $m(x) = \sum_{k\geqslant 1} kp_k(x)$,
- the trait of the jth child among k is distributed as $P_j^{(k)}(x, \cdot)$ for all $1 \leqslant j \leqslant k$.

We use the notion of spine, which is a distinguished line of descent in the branching process, and Many-to-One formulas, which have been developed from the notion of size-biased tree, considered by Kallenberg [Kal77], Chauvin and Rouault [CR88], Chauvin, Rouault and Wakolbinger [CRW91] with a Palm measure approach and Lyons, Peres and Pemantle [LPP95]. For general results on branching processes using these techniques, including the spinal decomposition, we refer to [KLPP97, Ath00] for discrete-time models and to [GB03, HH09, Clo17] for continuous-time branching processes. Non-locality at branching has also been investigated using the spine for continuous-state branching processes in [KP16] and for superdiffusions in [CRS17]. These previous works ensure in particular that if we denote by V_t the set of individuals alive at time t and by N_t its cardinal, we have the well-known Many-to-One formula:

$$\mathbb{E}\left[\sum_{u\in V_t} f\left(X_t^u\right)\right] = \mathbb{E}\left[f\left(Y_t\right)e^{\int_0^t B(Y_s)(m(Y_s)-1)ds}\right], \tag{1.1}$$

where f is a non-negative measurable function and $(Y_t)_{t\geqslant 0}$ follows the dynamics of a tagged-particle i.e. the same dynamics of all the particles between jumps and at a jump, the unique daughter particle is chosen uniformly at random among all the daughter particles. This formula can be seen as a Feynman-Kac formula [DM04] with a weight on the right-hand side relying on the whole ancestral lineage of current individuals which corresponds to the growth of the population. In this case, under spectral assumptions, the asymptotic behavior of the number of individuals has been well studied in [LPP95, KLPP97, Ath00, GB03, BK04]. We also refer to the work of Bansaye and al. [BDMT11] for law of large numbers theorems using Many-to-One formulas.

On the right-hand side of (1.1) appears a Markov process with penalized (or rewarded) trajectories which describes the dynamics of the trait of a typical individual. This corresponds to a time-inhomogeneous Markov process $Y^{(t)}$, indexed by $t\geqslant 0$, for which we provide the following formula for every non-negative measurable function F on the space of càdlàg processes:

$$\mathbb{E}\left[\sum_{u\in V_t} F\left(X_s^u, s\leqslant t\right)\right] = m(x,0,t)\mathbb{E}\left[F\left(Y_s^{(t)}, s\leqslant t\right)\right], \tag{1.2}$$

where for $x\in\mathcal{X}$ and $0\leqslant s\leqslant t$,

$$m(x,s,t) := \mathbb{E}\left[N_t\middle|Z_s=\delta_x\right], \tag{1.3}$$

and

$$Z_t = \sum_{u\in V_t}\delta_{X_t^u},$$

is the empirical measure of the process. We explicit the generator $\left(\mathcal{A}_s^{(t)}\right)_{s\leqslant t}$ of this auxiliary process: for all well-chosen functions f, $x\in\mathcal{X}$ and $s<t$, we have:

$$\mathcal{A}_s^{(t)}f(x) = \widehat{\mathcal{G}}_s^{(t)}f(x) + \widehat{B}_s^{(t)}(x)\int_{\mathcal{X}}(f(y)-f(x))\widehat{P}_s^{(t)}(x,dy),$$

where

$$\widehat{\mathcal{G}}_s^{(t)}f(x) = \frac{\mathcal{G}\left(m(\cdot,s,t)f\right)(x) - f\left(x\right)\mathcal{G}\left(m(\cdot,s,t)\right)(x)}{m(x,s,t)},$$

$$\widehat{B}_s^{(t)}(x) = B(x)\int_{\mathcal{X}}\frac{m(y,s,t)}{m(x,s,t)}m(x,dy),$$

$$\widehat{P}_s^{(t)}(x,dy) = m(y,s,t)m(x,dy)\left(\int_{\mathcal{X}}m(y,s,t)m(x,dy)\right)^{-1},$$

and

$$m(x,A) := \sum_{k\geqslant 0}p_k(x)\sum_{j=1}^{k}P_j^{(k)}(x,A),$$

denotes the expected number of children with trait in the Borel set A of an individual of trait x.

Moreover, we give some very simple and interesting examples where we can find the expression of the generator of the auxiliary process: we detail three models for the dynamics of a cell population (see Section 1.2.2).

The Many-to-One formula (1.2) splits the behavior of the entire population into a term characterizing the growth of the population and a term characterizing the dynamics of the trait. This separation in two terms is the key to the study of the ancestral trait of a uniformly sampled individual. Indeed, we prove in Theorem 1.4.1, that the auxiliary process describes the ancestral lineage of a sampled individual in a branching population at a fixed time when the initial population is large. More precisely, if we denote by $X^{U(t),\nu}$ the trait of a uniformly sampled individual from a population at time t with initial distribution ν and if $\nu_n = \sum_{i=1}^{n} \delta_{X_i}$ where X_i are i.i.d. random variables with law ν, under some assumptions, we prove the following convergence in law:

$$X_{[0,t]}^{U(t),\nu_n} \underset{n\to+\infty}{\longrightarrow} Y_{[0,t]}^{(t),\pi_t}, \text{ where } \pi_t(dx) = \frac{\mathbb{E}_x(N_t)\nu(dx)}{\int \mathbb{E}_x(N_t)\nu(dx)}, \tag{1.4}$$

and $Y^{(t),\pi_t}$ denotes the auxiliary process with initial condition distributed as π_t. This result shows that the auxiliary process is the appropriate tool for the study of the trait along the ancestral lineage of a sampling. We notice in particular that the dependence on the trait of the average number of individuals in the population plays a crucial part in the creation of a bias.

Finally, we refer the reader to Chapter 2 for results on the asymptotic behavior of the process of a sampling. In particular, under some assumptions ensuring the ergodicity of the auxiliary process, a law of large number for the empirical distribution of ancestral trajectories is proven. The asymptotic behavior of the process of a sampling has already been studied in [BDMT11] in the case of a constant division rate and in [Clo17] in a spectral framework.

Outline. Section 1.2 is devoted to the rigorous construction of our process. In Section 1.2.1, we first describe in detail the model and in Theorem 1.2.2, we prove the existence and uniqueness of the branching process. Then, in Section 1.2.2, we introduce our three examples of cell division models: the size-structured model with linear or exponential growth and the parasite infection model. In Section 1.3, we detail the properties of the Markov process along the spine. In particular, in Theorem 1.3.1, we prove the Many-to-One formula which describes the dynamics of a typical individual in the population. Finally, we give two other Many-to-One formulas, one for the dynamics of the whole tree in Proposition 1.3.5 and an other one for the dynamics of a couple of traits in Proposition 1.3.6. Section 1.4 concerns the ancestral lineage of a uniform sampling at a fixed time in a large population. More precisely, in Theorem 1.4.1, we prove the convergence (1.4). In Section 1.4.2, we give explicitly the dynamics of the auxiliary process for our three examples of cell population models. Finally, in Section 1.5, we give some useful comments on the model and some other examples.

Notation. We use the classical Ulam-Harris-Neveu notation to identify each individual. Let

$$\mathcal{U} = \bigcup_{n\in\mathbb{N}} (\mathbb{N}^*)^n .$$

The first individual is labeled by $\varnothing$. When an individual $u \in \mathcal{U}$ dies, its K descendants are labeled $u1, \dots, uK$. If u is an ancestor of v, we write $u \leqslant v$.

1.2 Definition and existence of the structured branching process

First, we introduce some useful notations and objects to characterize the branching process. Henceforth, we work on a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

Dynamics of the trait. Let $\mathcal{X} = \mathcal{Y} \times \mathbb{R}_+$ where $\mathcal{Y} \subset (\mathbb{R}_+)^d$ is a measurable space for some $d \geqslant 1$. It is the state space of the Markov process describing the trait of the individuals. The second component, with values in $\mathbb{R}_+$, is a time component. We assume that $(A_t, t \geqslant 0)$ is a strongly continuous contraction semi-group with associated infinitesimal generator $\mathcal{G} : \mathcal{D}(\mathcal{G}) \subset \mathcal{C}_b(\mathcal{X}) \to \mathcal{C}_b(\mathcal{X})$, where $\mathcal{C}_b(\mathcal{X})$ denotes the space of continuous bounded functions from $\mathcal{X}$ to $\mathbb{R}$.

Then, according to Theorem 4.4.1 in [EK86], there is a unique solution to the martingale problem associated with $(\mathcal{G}, \mathcal{D}(\mathcal{G})$, denoted by $(X_t, t \geqslant 0)$. It is an $\mathcal{X}$-valued càdlàg strong Markov process. For all $0 \leqslant s \leqslant t$, $x \in \mathcal{X}$, we denote by $\Phi(x, s, t)$ the corresponding stochastic flow i.e. $(\Phi(x, s, t), t \geqslant s)$ is the unique solution of the martingale problem associated with $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$ satisfying $\Phi(x, s, s) = x$. We have the following properties:

- for all $f \in \mathcal{D}(\mathcal{G})$, $0 \leqslant s \leqslant t$ and $x \in \mathcal{X}$:

  $$f\left(\Phi(x, s, t)\right) - f(x) - \int_s^t \mathcal{G}f\left(\Phi(x, s, r)\right) dr, \tag{1.5}$$

  is a $\sigma(X_t, t \geqslant 0)$-martingale where $\sigma(X_t, t \geqslant 0)$ is the natural filtration associated with X.

- for each $0 \leqslant s \leqslant t$, $\Phi(\cdot, s, t)$ is a measurable map from $\mathcal{X}$ to $\mathcal{X}$,

- for each $0 \leqslant r \leqslant s \leqslant t$ and all $x \in \mathcal{X}$, $\Phi\left(\Phi(x, r, s), s, t\right) = \Phi(x, r, t)$, almost surely.

We refer the reader to [Kun97] for more properties on stochastic flows.

Remark 1.2.1. According to the Hille-Yoshida theorem (see [EK86, Theorem 1.2.6]), $\mathcal{D}(\mathcal{G})$ is dense in $\mathcal{C}_b(\mathcal{X})$ for the topology of uniform convergence.

Division events. An individual with trait x dies at an instantaneous rate $B(x)$, where B is a continuous function from $\mathcal{X}$ to $\mathbb{R}_+$. It is replaced by $A_u(x)$ children, where $A_u(x)$ is a $\mathbb{N}$-valued random variable with distribution $(p_k(x), k \geqslant 0)$. For convenience, we assume that $p_1(x) \equiv 0$ for all $x \in \mathcal{X}$. The trait at birth of the jth descendant among k is given by the random variable $F_j^{(k)}(x, \theta)$, where $\left(F_j^{(k)}(\cdot, \cdot), j \leqslant k, k \in \mathbb{N}\right)$ is a family of measurable functions from $\mathcal{X} \times [0, 1]$ to $\mathcal{X}$ and θ is a uniform random variable on $[0, 1]$. For all $k \in \mathbb{N}$, let $P^{(k)}(x, \cdot)$ be the probability measure on $\mathcal{X}^k$ corresponding to the trait distribution at birth of the k descendants of an individual with trait x. We denote by $P_j^{(k)}(x, \cdot)$ the jth marginal distribution of $P^{(k)}$ for all $k \in \mathbb{N}$ and $j \leqslant k$ i.e. for all Borel sets $A \subset \mathcal{X}$, we have $P_j^{(k)}(x, A) = P^{(k)}\left(x, \mathcal{X}^{j-1} \times A \times \mathcal{X}^{k-j}\right)$.

We denote by $\mathcal{M}_P(\mathcal{X})$ the set of point measures on $\mathcal{X}$. Following Fournier and Méléard [FM04], we work in $\mathbb{D}\left(\mathbb{R}_+, \mathcal{M}_P\left(\mathcal{X}\right)\right)$, the state of càdlàg measure-valued processes. For any $\bar{Z} \in \mathbb{D}\left(\mathbb{R}_+, \mathcal{M}_P\left(\mathcal{U} \times \mathcal{X}\right)\right)$, we write $\bar{Z}_t(du, dx) = \sum_{u \in V_t} \delta_{(u, X_t^u)}(du, dx)$ and:

$$Z_t(dx) = \sum_{u \in V_t} \delta_{X_t^u}(dx), \ t \geqslant 0,$$

the marginal measure of $\bar{Z}_t(du, dx)$ on $\mathcal{X}$, where V_t represents the set of individuals alive at time t. We set $N_t = \# V_t$. Moreover, for any process $\bar{Z} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{U} \times \mathcal{X}))$, we define recursively the associated sequence of jump times by

$$T_0(\bar{Z}) = 0 \text{ and } T_{k+1}(\bar{Z}) = \inf \left\{ t > T_k(\bar{Z}),\ N_t \neq N_{T_k(\bar{Z})} \right\},$$

with the standard convention that $\inf\{\varnothing\} = +\infty$.

In order to ensure the non-explosion in finite time of such a process, we need to consider two sets of hypotheses. The first one controls what happens regarding divisions (in term of rate of division and of mass creation).

Assumption A. *We consider the following assumptions:*

1. *There exist $b_1, b_2 \geqslant 0$ and $\gamma \geqslant 1$ such that for all $x \in \mathcal{X}$,*
$$B(x) \leqslant b_1 |x|^\gamma + b_2.$$
2. *For all $t \geqslant 0$, there exists $\ell(t) \in \mathbb{R}_+$, increasing in t, such that for all $x = (y, t) \in \mathcal{X}$, $k \in \mathbb{N}$ and $\theta \in [0, 1]$:*
$$\sum_{i=1}^{k} F_i^{(k)}(x, \theta) \leqslant x \vee \ell(t), \text{ componentwise.}$$
3. *There exists $\overline{m} \geqslant 0$ such that for all $x \in \mathcal{X}$,*
$$m(x) = \sum_k k p_k(x) \leqslant \overline{m}.$$
4. *For all $x \in \mathcal{X}$ and $s \geqslant 0$, we have:*
$$\lim_{t \to +\infty} \int_s^t B(\Phi(x, s, r))\, dr = +\infty, \text{ almost surely.}$$

The first point controls the life-times of individuals via the division rate. In particular, if $\gamma = 0$, B is bounded and the non-explosion in finite time of the number of individuals in the previously defined process is obvious. In more general framework, we have to consider the other point of Assumption A in order to prove the non-explosion in finite time. The second point of Assumption A means that we consider a fragmentation process with a possibility of mass creation at division when the mass is small enough. In particular, clones are allowed in the case of bounded traits and bounded number of descendants and any finite type branching structured process can be considered. The dependence in t of the threshold ℓ allows us to consider models in a varying environment. The last point of Assumption A ensures that each individual divides after a certain time.

We make a second assumption to control the behavior of traits between divisions.

Assumption B. *There exist $c_1, c_2 \geqslant 0$ such that for all $x \in \mathcal{X}$:*

$$\mathcal{G} h_\gamma(x) \leqslant c_1 h_\gamma(x) + c_2,$$

where γ is defined in Assumption A and for $x \in (\mathbb{R}_+)^d$, $h_\gamma(x) = |x|^\gamma = \left(\sum_{i=1}^d x_i\right)^\gamma$.

Assumptions A(1) and B are linked via the parameter γ which controls the balance between the growth of the population and the dynamics of the trait.

1.2.1 Existence and uniqueness of the structured branching process

We now prove the strong existence and uniqueness of the structured branching process. Let $E = \mathcal{U} \times \mathbb{R}_+ \times [0,1] \times [0,1]$ and $M\left(ds, du, dz, dl, d\theta\right)$ be a Poisson point measure on $\mathbb{R}_+ \times E$ with intensity $ds \otimes n(du) \otimes dz \otimes dl \otimes d\theta$, where $n(du)$ denotes the counting measure on $\mathcal{U}$. Let $\left(\Phi^u\right)_{u\in\mathcal{U}}$ be a family of independent stochastic flows satisfying (1.5) describing the individual-based dynamics. We assume that M and $\left(\Phi^u\right)_{u\in\mathcal{U}}$ are independent. We denote by $\mathcal{F}_t$ the filtration generated by the Poisson point measure M and the family of stochastic flows $(\Phi^u(x,s,t), u \in \mathcal{U}, x \in \mathcal{X}, s \leqslant t)$ up to time t.

For all $x \in \mathcal{X}$, there exists a function $G(x,\cdot) : [0,1] \to \mathbb{N}$ such that:

$$G(x,l) \overset{d}{=} \left(p_k(x),\ k \in \mathbb{N}\right),$$

where l is a uniform random variable on $[0,1]$. This formalism will prove useful in the use of Poisson point measure to describe the jumps in the measure-valued branching process. For convenience, for all $x \in \mathcal{X}$ and θ, l uniform random variables on $[0,1]$, we write:

$$F_i(x,l,\theta) = F_i^{(G(x,l))}(x,\theta).$$

We denote by

$$\bar{\mathcal{D}}(\mathcal{G}) := \left\{f : \mathcal{U} \times \mathcal{X} \to \mathbb{R} \text{ such that } f(u,\cdot) \in \mathcal{D}(\mathcal{G})\ \forall u \in \mathcal{U}\right\}.$$

For all $0 \leqslant s \leqslant t$, $f \in \bar{\mathcal{D}}(\mathcal{G})$, $x \in \mathcal{X}$ and $u \in \mathcal{U}$, we consider the $\mathcal{F}_t$-martingale $\left(M^{f,u}_{s,t}(x), t \geqslant s\right)$ defined by:

$$\begin{aligned} M^{f,u}_{s,t}(x) :=& f(u,t,\Phi^u(x,s,t)) - f(u,s,x) \\ &- \int_s^t \left(\mathcal{G}f(u,r,\Phi^u(x,s,r)) + \partial_r f(u,r,\Phi^u(x,s,r))\right) dr. \end{aligned} \tag{1.6}$$

Theorem 1.2.2. *Under Assumptions A(1-3) and B, there exists a strongly unique $\mathcal{F}_t$-adapted càdlàg process taking values in $\mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ such that for all $f \in \bar{\mathcal{D}}(\mathcal{G})$ and $t \geqslant 0$:*

$$\begin{aligned} \langle \bar{Z}_t, f\rangle = f\left(\varnothing, 0, x_0\right) &+ \int_0^t \int_{\mathcal{U}\times\mathcal{X}} \left(\mathcal{G}f(u,s,x) + \partial_s f\left(u,s,x\right)\right) \bar{Z}_s\left(du,dx\right) ds + M^f_{0,t}(x) \\ &+ \int_0^t \int_E \mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant B\left(X^u_{s-}\right)\right\}} \left(\sum_{i=1}^{G(X^u_s,l)} f\left(u,s,F_i\left(X^u_s,l,\theta\right)\right) - f\left(u,s,X^u_{s-}\right)\right) \\ &\qquad M\left(ds,du,dz,dl,d\theta\right), \end{aligned} \tag{1.7}$$

where for all $s \geqslant 0$ and $t \geqslant s$:

$$M^f_{s,t}(x) = \sum_{k\geqslant 1} \mathbf{1}_{s\leqslant T_{k-1}(\bar{Z})<t} \sum_{u\in V_{T_{k-1}(\bar{Z})}} M^{f,u}_{T_{k-1}(\bar{Z}),T_k(\bar{Z})\wedge t}\left(X^u_{T_{k-1}(\bar{Z})}\right),$$

is a $\mathcal{F}_t$-martingale.

The existence and uniqueness of such measure-valued process has first been studied by Fournier and Méléard [FM04]. We also refer to [Tra06, BM15] for different extensions and to [BT11] for the case of branching processes. Here, we obtain the non-explosion of the branching process in finite time under quite general assumptions (no bounded branching rate, random number of offspring, random transmission of the trait).

The proof of this theorem is split into four lemmas. First, in Lemma 1.2.3, we prove the existence of a $\mathcal{F}_t$-adapted càdlàg measure-valued process $\bar{Z}$ solution of (1.7) for all $t \in [0, T_k(\bar{Z}))$ and all $k \in \mathbb{N}$. Then, in Lemma 1.2.4, we prove that $\left(M^f_{s,t}, t \geqslant 0\right)$ is a $\mathcal{F}_t$-martingale. Next, in Lemma 1.2.5, we prove the uniqueness of the increasing sequence $(T_k(\bar{Z}),\ k \geqslant 0)$ corresponding to the jump times of a solution $\bar{Z}$ to (1.7) and the uniqueness of a $\mathcal{F}_t$-adapted càdlàg solution to (1.7) for $t \in [0, T_k(\bar{Z}))$ for all $k \in \mathbb{N}$. Finally, in Lemma 1.2.6, we prove that the sequence of jump times tends to infinity resulting in the existence and uniqueness of the process on $\mathbb{R}_+$.

Lemma 1.2.3. *There exists a $\mathcal{F}_t$-adapted càdlàg measure-valued process $(\bar{Z}_t,\ t \geqslant 0) \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ which is solution of (1.7) for all $f \in \bar{\mathcal{D}}(\mathcal{G})$ and for all $t \in [0, T_k(\bar{Z}))$, $k \in \mathbb{N}$.*

Proof. See Section 1.6.1 in the appendix. □

The existence of such processes has already been studied in [BT11] in the case of a trait following a Feller diffusion. From Lemma 1.2.3, we deduce the existence of a càdlàg measure-valued process $Z \in \mathcal{M}_P(\mathcal{X})$ solution of (1.7) which is given by the projection of the solution $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ on the second coordinate.

Lemma 1.2.4. *Let $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ be a solution of (1.7) whose construction is given in the previous lemma. Let $k \in \mathbb{N}$. For all $0 \leqslant s \leqslant t \leqslant T_k(\bar{Z})$, $x \in \mathcal{X}$ and $f \in \bar{\mathcal{D}}(\mathcal{G})$,*

$$M^f_{s,t}(x) = \sum_{k \geqslant 1} \mathbf{1}_{\{s \leqslant T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}}} M^{f,u}_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}\left(X^u_{T_{k-1}(\bar{Z})}\right),$$

is an $\mathcal{F}_t$-martingale.

Proof. Let $k \in \mathbb{N}$ and $0 \leqslant s \leqslant t \leqslant T_k(\bar{Z})$. Let $f \in \bar{\mathcal{D}}(\mathcal{G})$ and $x \in \mathcal{X}$. Then, for all $s \leqslant r \leqslant t$, we have:

$$\begin{aligned}
&\mathbb{E}\left[M^f_{s,t}(x) | \mathcal{F}_r\right] - M^f_{s,r}(x) \\
=&\mathbb{E}\left[\sum_{k \geqslant 1} \mathbf{1}_{\{r \leqslant T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}(\bar{Z})}} M^{f,u}_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}\left(X^u_{T_{k-1}(\bar{Z})}\right) | \mathcal{F}_r\right] \\
=&\mathbb{E}\left[\sum_{k \geqslant 1} \mathbf{1}_{\{r \leqslant T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}(\bar{Z})}} \mathbb{E}\left[M^{f,u}_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}\left(X^u_{T_{k-1}(\bar{Z})}\right) | \mathcal{F}_{T_{k-1}(\bar{Z})}\right] | \mathcal{F}_r\right] = 0,
\end{aligned}$$

because $\left(M^{f,u}_{s,t}(x), t \geqslant s\right)$ is a $\mathcal{F}_t$-martingale. □

Next, we prove the uniqueness of the sequence of jump times $(T_k(\bar{Z}),\ k \geqslant 0)$ associated with a solution $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ to (1.7) and the uniqueness of the solution on $[0, T_k(\bar{Z}))$, for all $k \in \mathbb{N}$.

Lemma 1.2.5. *The increasing sequence $(T_k(\bar{Z}),\ k \geqslant 0)$ corresponding to the jump times of a solution $\bar{Z}$ to (1.7) is strongly unique. Moreover, the strong uniqueness of a $\mathcal{F}_t$-adapted càdlàg measure-valued solution to (1.7) holds, for $t \in [0, T_k(\bar{Z}))$ and for all $k \in \mathbb{N}$.*

Proof. See Section 1.6.2 in the appendix. □

Lemma 1.2.6. *Under Assumptions A(1-3) and B, the strongly unique sequence of jump times of a solution $\bar{Z}$ to (1.7) tends to infinity as k tends to infinity, almost surely.*

Proof. Let $T > 0$. To shorten notation, we write T_k instead of $T_k(\bar{Z})$. We prove that almost surely there is no accumulation of jumps on $[0, T]$ of the solution of (1.7) previously constructed on $[0, T_k[$, for all $k \in \mathbb{N}$. Let $k \in \mathbb{N}$ and $\left(\bar{Z}_t, t \leqslant T_k\right)$ be the solution of (1.7) up to the kth division time. Using equation (1.7) applied to the constant function equal to 1, we have for all $t \leqslant T_k \wedge T$:

$$\begin{aligned}\mathbb{E}_{\delta_x}\left(N_t\right) &= 1 + \int_0^t \mathbb{E}_{\delta_x}\left(\sum_{u \in V_s} B\left(X_s^u\right)\left(m(X_s^u) - 1\right)\right) ds \\ &\leqslant 1 + \overline{m} b_1 \int_0^t \mathbb{E}_{\delta_x}\left(\sum_{u \in V_s} |X_s^u|^\gamma\right) ds + \overline{m} b_2 \int_0^t \mathbb{E}_{\delta_x}\left(N_s\right) ds, \qquad (1.8)\end{aligned}$$

where the inequality comes from Assumption A(1) and A(3). Recalling that $h_\gamma(x) = \left(\sum_{i=1}^d x_i\right)^\gamma$, for $x \in (\mathbb{R}_+)^d$, we have using (1.7):

$$\begin{aligned}\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} |X_t^u|^\gamma\right] &= |x|^\gamma + \int_0^t \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} \mathcal{G} h_\gamma(X_s^u)\right] ds \\ &+ \int_0^t \int_{[0,1]} \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} B(X_s^u) \sum_{k \geqslant 0} p_k(X_s^u)\left(\sum_{j=1}^k \left|F_j^{(k)}(X_s^u, \theta)\right|^\gamma - |X_s^u|^\gamma\right)\right] d\theta ds.\end{aligned}$$

Next, using Assumption A(2) and B, we get that:

$$\begin{aligned}\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} |X_t^u|^\gamma\right] \leqslant |x|^\gamma + \int_0^t & \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} \left(c_1 |X_s^u|^\gamma + c_2\right)\right] \\ &+ \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} B(X_s^u) \ell(s) \mathbf{1}_{\{|X_s^u|^\gamma \leqslant \ell(s)\}}\right] ds.\end{aligned}$$

Finally, using Assumption A(1) and the fact that $t \mapsto \ell(t)$ is increasing, we get:

$$\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} |X_t^u|^\gamma\right] \leqslant |x|^\gamma + c_1 \int_0^t \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} |X_s^u|^\gamma\right] ds + \left(c_2 + \left(b_1 \ell(t) + b_2\right) \ell(t)\right) \int_0^t \mathbb{E}_{\delta_x}\left[N_s\right] ds.$$

Adding this inequality to (1.8) we obtain for all $t \leqslant T_k \wedge T$:

$$\mathbb{E}_{\delta_x}\left[N_t\right] + \mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} |X_t^u|^\gamma\right] \leqslant 1 + |x|^\gamma + A(T) \int_0^t \left(\mathbb{E}_{\delta_x}\left[N_s\right] + \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} |X_s^u|^\gamma\right]\right) ds,$$

where $A(T) = c_1 + c_2 + b_1 \ell(T)^2 + b_2 \ell(T) + (b_1 + b_2)\overline{m}$. According to Grönwall Lemma, we obtain for all $t \leqslant T_k \wedge T$:

$$\mathbb{E}_{\delta_x}\left[N_t\right] + \mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} |X_t^u|^\gamma\right] \leqslant \left(1 + |x|^\gamma\right) e^{A(T)t} < \infty.$$

Finally, the average number of individuals in the population at time t is bounded for t in compact sets and there is no explosion of the population in finite time. □

Before moving to the next section , we introduce the first-moment semi-group $(R_{s,t},\ t \geqslant s)$ associated with the branching process: for all $s \geqslant 0$, $t \geqslant s$ and $x \in \mathcal{X}$, let

$$R_{s,t} f(x) = \mathbb{E}\left[\sum_{u \in V_t} f\left(X_t^u\right) \middle| Z_s = \delta_x\right], \qquad (1.9)$$

where f is a measurable function. Applying equation (1.7) to $f \equiv 1$, we obtain by taking the expectation :

$$R_{s,t}\mathbf{1}(x) = m(x,s,t) = 1 + \int_s^t \mathbb{E}\left[\sum_{u\in V_r} B(X_r^u)(m(X_r^u)-1)\big|Z_s = \delta_x\right] dr. \tag{1.10}$$

In particular, if $B \equiv b$ and $m(x) = m$ for all $x \in \mathcal{X}$, we obtain $m(x,s,t) = e^{b(m-1)(t-s)}$.

1.2.2 Some growth-fragmentation models for cell population dynamics

In this section, we consider growth-fragmentation processes: at division, the trait of the ancestor is shared between the children and the number of individuals in the population increases. Moreover, we focus on models where the trait moves according to a diffusion given by its associated generator of following form:

$$\mathcal{G}f(x) = r(x)f'(x) + \sigma^2(x)f''(x),$$

where r and σ are measurable functions. This class covers several dynamics for the trait. Here, we present three of them. In particular, we give an explicit formula for the average number of individuals in the population at time t. We first give a useful equation concerning models with such a dynamics. For all $s \geqslant 0$, $t \geqslant s$ and $x \in \mathcal{X}$, applying (1.7) to $f(x) = x$ and taking the expectation, we obtain:

$$R_{s,t}\mathbf{Id}(x) = x + \int_s^t R_{s,u}r(x)du, \tag{1.11}$$

where $(R_{s,t})_{t\geqslant s}$ is defined in (1.9).

Linear growth model

We consider here a size-structured model. More precisely, the size of each cell grows linearly at a rate $a > 0$ supposed to be identical for each cell and divisions occur at rate $B(x) = \alpha x$, $\alpha > 0$. At fission, the cell splits into two daughter cells of size $\frac{x}{2}$, when x denotes the size of the mother at splitting. Deciding whether the cells' growth follows a linear or an exponential dynamics has fueled a large debate in the literature (see [Coo06] and references therein). The linear growth model has been considered for example in [DMZ10] for the calibration of a deterministic growth-fragmentation model from experimental data and in [Hoa15] for the estimation of the division rate.

Using the previous notation, the process $(X_t, t \geqslant 0)$ describing the size of a cell starting from x_0 is given by:

$$X_t = x_0 + at,$$

and the associated generator is given for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ by:

$$\mathcal{G}f(x) = af'(x).$$

Then, the branching process $(Z_t, t \geqslant 0)$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned}\langle Z_t, f\rangle =& \langle Z_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} af'(x)Z_s(dx)ds \\ &+ \int_0^t \int_{\mathcal{U}\times\mathbb{R}_+} \mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant \alpha X_{s-}^u\right\}} \left(2f\left(\frac{X_{s-}^u}{2}\right) - f\left(X_{s-}^u\right)\right) M(ds,du,dz),\end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ with intensity $ds \otimes n(du) \otimes dz$. The first integral corresponds to the dynamics of the population between two divisions. The integral with respect to the Poisson point measure represents to the jump part of the process and the indicator function corresponds to the fact that an individual u jumps at time s if it is in the population at time s^- and if the division rate at X^u_{s-} is large enough. In this case u, it is removed from the population and two descendants with trait $X^u_{s-}/2$ appear.

The validity of Assumptions A and B is trivial for this model with $\gamma = 1$. Let us compute the average number of individuals in the population at time t. For all $s \leqslant t$ and $x \in \mathbb{R}$, we have using (1.10):

$$m(x,s,t) = 1 + \alpha \int_s^t \mathbb{E}\left(\sum_{u \in V_r} X^u_r | Z_s = \delta_x \right) dr. \tag{1.12}$$

Combining (1.11) and (1.12), we obtain:

$$m(x,s,t) = 1 + \alpha \int_s^t \left(x + a \int_s^r m(x,s,\tau) d\tau \right) dr,$$

and for all $x \in \mathcal{X}$ and $s \geqslant 0$, $m(x,s,\cdot)$ is the solution of the following Cauchy problem with unknown f:

$$\begin{cases} f''(t) = a\alpha f(t), \\ f(s) = 1, \; f'(s) = \alpha x. \end{cases}$$

with explicit solution given by:

$$m(x,s,t) = \frac{1}{2}\left(e^{\overline{a}(t-s)} + e^{-\overline{a}(t-s)} \right) + \frac{x}{2}\sqrt{\frac{\alpha}{a}} \left(e^{\overline{a}(t-s)} - e^{-\overline{a}(t-s)} \right),$$

where $\overline{a} = \sqrt{a\alpha}$. The population size is exponential in time as in the neutral case.

Exponential growth model in a varying environment

We assume here that the growth of the cells is exponential at rate a. This exponential growth model has been studied in [DHKR15] in the case of a specific growth rate for each individual in order to infer the division rate of the population. Here, we assume that the division rate is a function of time, mimicking a varying environment. More precisely, we set $B(x,t) = \alpha(t)x$, with α a positive function. The generator for the dynamics of the size is given for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ by:

$$\mathcal{G} f(x) = axf'(x).$$

We still assume that the branching is binary and that the size of the descendants at birth are both $x/2$ if x is the size of the mother at splitting. Then, the branching process $(Z_t, t \geqslant 0)$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle =& \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} axf'(x) Z_s(dx) ds \\ &+ \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+} \mathbf{1}_{\left\{ u \in V_{s-}, \; z \leqslant \alpha(s) X^u_{s-} \right\}} \left(2f\left(\frac{X^u_{s-}}{2} \right) - f\left(X^u_{s-} \right) \right) M(ds, du, dz), \end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ with intensity $ds \otimes n(du) \otimes dz$. Moreover, we have using (1.11) with $r(x) = ax$:

$$\mathbb{E}\left(\sum_{u \in V_t} X^u_t | Z_s = \delta_x \right) = xe^{a(t-s)}.$$

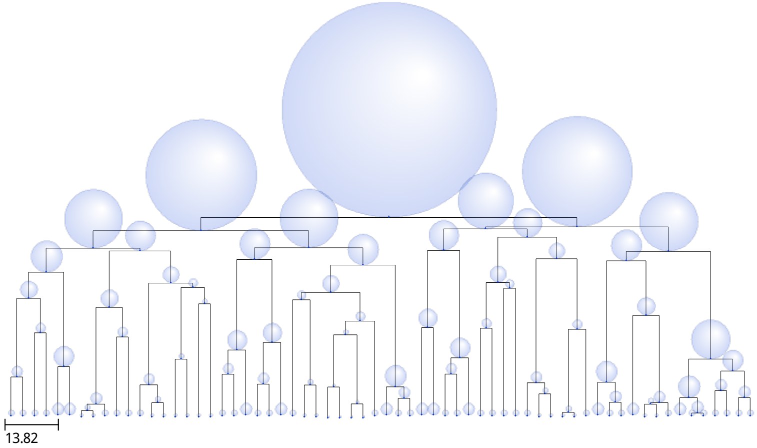

Figure 1.2.1: Descending genealogy from an individual with size 1 until time $T = 50$ of a size-structured population. Each cell grows exponentially at rate 0.01 and divide at rate $B(x) = x$. The scale refers to the length of the branches.

Combining this with equation (1.10), we obtain:

$$m(x, s, t) = 1 + x \int_s^t \alpha(r) e^{a(r-s)} dr.$$

In particular, if $\alpha(r) \equiv \alpha$ with α a positive constant, we obtain:

$$m(x, s, t) = 1 + \frac{\alpha x}{a} \left(e^{a(t-s)} - 1 \right).$$

The growth is again exponentially fast in time. On Figure 1.2.1, we see a numerically simulated example of such a size-structured population. The diameter of each circle corresponds to the size of the individual at division and the length of the branch represents the lifetime of each individual. In particular, we notice the link between the lifetime and the size of each individual: the bigger the cell, the shorter its lifetime. We also observe that the sizes of the cells at birth decrease as a function of time. This is due to the fact that the growth is not fast enough to compensate for the divisions.

Parasite infection model

This model is a continuous version of Kimmel's multilevel model for plasmids [Kim97] which has already been studied in the case of a constant or monotone division rate by Bansaye and Tran in [BT11]. It models the proliferation of a parasite infection in a cell population. More precisely, we assume here that the trait $(X_t, t \geqslant 0)$ is a Markov process describing the quantity of parasites in each cell which evolves as a Feller diffusion process:

$$X_t = X_0 + \int_0^t g X_s ds + \int_0^t \sqrt{2\sigma^2 X_s} dB_s,$$

where $(B_s)_{s \geqslant 0}$ is standard Brownian motion and $g, \sigma > 0$ are some fixed parameters. The generator for the dynamics of the quantity of parasites is given for any function $f \in \mathcal{C}^2(\mathbb{R}_+)$ by:

$$\mathcal{G} f(x) = g x f'(x) + \sigma^2 x f''(x).$$

We assume here that a cell with a quantity x of parasites will potentially divide at a rate $B(x) = \alpha x + \beta$, $\alpha, \beta > 0$ into two daughter cells with a quantity δx and $(1-\delta)x$ of parasites respectively, where δ is a random variable with uniform distribution on $[0,1]$. We need β to be strictly positive so that even cells without any parasites divide after some time. The branching process $(Z_t, t \geqslant 0)$ is then solution of the following equation, for any function $f \in \mathcal{C}^2(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned}\langle Z_t, f\rangle = \langle Z_0, f\rangle &+ \int_0^t \int_{\mathbb{R}_+} \left(gxf'(x) + \sigma^2 x f''(x)\right) Z_s(dx)ds + \int_0^t \sum_{u \in V_s} \sqrt{2\sigma^2 X_s^u} f'(X_s^u) dB_s^u \\ &+ \int_0^t \int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]} \mathbf{1}_{\left\{u \in V_{s-},\ z \leqslant \alpha X_{s-}^u + \beta\right\}} \left(f\left(\delta X_{s-}^u\right) + f\left((1-\delta)X_{s-}^u\right) - f\left(X_{s-}^u\right)\right) \\ &\qquad M(ds, du, dz, d\delta),\end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times [0,1]$ with intensity $ds \otimes n(du) \otimes dz \otimes d\delta$ and $(B_s^u, s \geqslant 0)_{u \in \mathcal{U}}$ is a family of standard Brownian motions. In particular, the generator corresponding to first moment semi-group is given for any function $f \in \mathcal{C}^2(\mathbb{R})$ and $x \in \mathcal{X}$ by:

$$\mathcal{F}_{\text{inf}} f(x) = gxf'(x) + \sigma^2 x f''(x) + (\alpha x + \beta)\left(\int_0^1 \left[f(\delta x) + f((1-\delta)x)\right] d\delta - f(x)\right).$$

Therefore, we notice that if (V, λ) are eigenelements of $\mathcal{F}_{\text{inf}}$, we have $\mathcal{F}_{\text{inf}} V(0) = \beta V(0)$ so that $V(0) = 0$ if $\lambda \neq \beta$ and we cannot apply usual techniques using eigenelements.

Let us compute the average number of individuals in the population after time t. We have using (1.10):

$$m(x,s,t) = 1 + \alpha \int_s^t \mathbb{E}\left[\sum_{u \in V_r} X_r^u \middle| Z_s = \delta_x\right] dr + \beta \int_s^t m(x,s,r)dr.$$

Again, using (1.11), we obtain:

$$\mathbb{E}\left[\sum_{u \in V_r} X_r^u \middle| Z_s = \delta_x\right] = xe^{g(r-s)}.$$

Then, combining the two previous equations, we get after differentiation:

$$\partial_t m(x,s,t) = \alpha x e^{g(t-s)} + \beta m(x,s,t),$$

and finally:

$$m(x,s,t) = \frac{\alpha x}{g-\beta} e^{g(t-s)} + \left(1 - \frac{\alpha x}{g-\beta}\right) e^{\beta(t-s)},$$

if $g \neq \beta$ and:

$$m(x,s,t) = (1 + \alpha(t-s))\, xe^{\beta(t-s)},$$

if $g = \beta$. In the three examples above, the mean number of individuals in the population is an affine function of the trait of the initial individual. However, this is not the rule. For example, Cloez developed in [Clo17](Corollary 6.1.) the case of a dynamics of the trait following an Ornstein-Uhlenbeck process where the dependence in x is not affine.

For other examples and comments, including a link with the integro-differential model, we refer to Section 1.5.

1.3 The trait of sampled individuals at a fixed time : Many-to-One formulas

In order to characterize the trait of a uniformly sampled individual, the spinal approach ([CR88],[LPP95]), consists in following a "typical" individual in the population whose behavior summarizes the behavior of the entire population. Biggins [Big77] used this approach for the study of branching random walks extending Kingman results [Kin75]. The spinal approach has then been extended to various frameworks ([HW96],[KLPP97],[HH09]). In particular, Georgii and Baake [GB03] used spine techniques in a spectral framework to describe the asymptotic distribution of the trait of a uniformly sampled individual in the population and its ancestral lineage in the case of a finite set of possible trait.

In this section, we specify the generator of the process describing the trait along the spine. The existence of our auxiliary process does not rely on the existence of spectral elements for the mean operator of the branching process.

With a slight abuse of notation, for all $u \in V_t$ and $s < t$, we denote by X_s^u the trait of the unique ancestor living at time s of u.

1.3.1 The auxiliary process

Let us define

$$\mathcal{D}(\mathcal{A}) = \left\{ f \in \mathcal{D}(\mathcal{G}) \text{ s.t. } m(\cdot, s, t) f \in \mathcal{D}(\mathcal{G}) \ \forall t \geqslant 0,\ \forall s \leqslant t \right\}.$$

From now on, we assume that for all $x \in \mathcal{X}$, $t \geqslant 0$ and $s \leqslant t$, $m(x, s, t) \neq 0$.

We now recall the operator and functions needed for the definition of the auxiliary process, and introduce additional notations. For all $f \in \mathcal{D}(\mathcal{A})$, $x \in \mathcal{X}$ and $s < t$, we write:

$$\widehat{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}\left(m(\cdot, s, t) f\right)(x) - f(x)\, \mathcal{G}\left(m(\cdot, s, t)\right)(x)}{m(x, s, t)}, \tag{1.13}$$

$$\widehat{B}_s^{(t)}(x) = B(x) \Lambda(x, s, t), \tag{1.14}$$

$$\widehat{P}_s^{(t)}(x, dy) = \Lambda^{-1}(x, s, t) \frac{m(y, s, t)}{m(x, s, t)} m(x, dy), \tag{1.15}$$

where:

$$\Lambda(x, s, t) = \int_{\mathcal{X}} \frac{m(y, s, t)}{m(x, s, t)} m(x, dy).$$

In order to prove a Many-to-One formula, we need to consider the following assumption:

Assumption C. *There exists a function C such that for all $j \leqslant k$, $j, k \in \mathbb{N}$ and $0 \leqslant s \leqslant t$, we have:*

$$\sup_{x \in \mathcal{X}} \sup_{s \in [0,t]} \int_{\mathcal{X}} \frac{m(y, s, t)}{m(x, s, t)} P_j^{(k)}(x, dy) \leqslant C(t),\ \forall t \geqslant 0.$$

Assumption D. *For all $t \geqslant 0$ we have:*

- *for all $x \in \mathcal{X}$, $s \mapsto m(x, s, t)$ is differentiable on $[0, t]$ and its derivative is continuous on $[0, t]$,*
- *for all $x \in \mathcal{X}$, $f \in \mathcal{D}(\mathcal{A})$, $s \mapsto \mathcal{G}(m(\cdot, s, t) f)(x)$ is continuous,*

- *$\mathcal{D}(\mathcal{A})$ is dense in $\mathcal{C}(\mathcal{X})$ for the topology of uniform convergence.*

The last point of this assumption allows us to extend our formulas to all measurable functions with respect to the Skorokhod topology using a monotone class argument. Moreover, combining Lemma 1.3.4 and Remark 1.2.1, this assumption is in particular satisfied if $\mathcal{D}(\mathcal{G})$ is stable by product.

Theorem 1.3.1. *Under Assumptions A(1-3), B, C and D, for all $t \geqslant 0$, for all $x_0 \in \mathcal{X}$ and for all non-negative measurable functions $F : \mathbb{D}([0,t], \mathcal{X}) \to \mathbb{R}_+$ we have:*

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{u \in V_t} F\left(X_s^u, s \leqslant t\right)\right] = m(x_0, 0, t)\mathbb{E}_{x_0}\left[F\left(Y_s^{(t)}, s \leqslant t\right)\right], \tag{1.16}$$

where $\left(Y_s^{(t)}, s \leqslant t\right)$ is a time-inhomogeneous Markov process whose law is characterized by its associated infinitesimal generators $\left(\mathcal{A}_s^{(t)}\right)_{s \leqslant t}$ given for $f \in \mathcal{D}(\mathcal{A})$ and $x \in \mathcal{X}$ by:

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_s^{(t)} f(x) + \widehat{B}_s^{(t)}(x) \int_{\mathcal{X}} \left(f\left(y\right) - f\left(x\right)\right) \widehat{P}_s^{(t)}\left(x, dy\right). \tag{1.17}$$

Formula (1.16) has a natural interpretation in terms of semi-groups. If f is a non-negative measurable function, we set for any $0 \leqslant r \leqslant s \leqslant t$ and any $x \in \mathcal{X}$:

$$P_{r,s}^{(t)} f(x) := \frac{\mathbb{E}\left[\sum_{u \in V_t} f\left(X_s^u\right) | Z_r = \delta_x\right]}{m(x, r, t)} = \mathbb{E}\left[f\left(Y_s^{(t)}\right) | Y_r^{(t)} = x\right]. \tag{1.18}$$

In other words, $\left(P_{r,s}^{(t)}\right)_{r \leqslant s \leqslant t}$ is a conservative (non-homogeneous) semi-group and the auxiliary process $Y^{(t)}$ is its time-inhomogeneous associated Markov process corresponding to the right-hand side of (1.1). We can exhibit this process using a change of probability measure. Indeed, by Feynman-Kac's formula, we have

$$P_{r,s}^{(t)} f(x) = m(x, r, t)^{-1} \mathbb{E}\left[e^{\int_r^s B(X_v)(m(X_v)-1)dv} m(X_s, s, t) f(X_s) \big| X_r = x\right],$$

where the Markov process $(X_s, r \leqslant s \leqslant t)$ corresponds to dynamics of the tagged-particle which infinitesimal generator $\mathcal{M}$ is given by

$$\mathcal{M} f(x) = \mathcal{G} f(x) + B(x) m(x) \sum_{k \geqslant 0} \frac{k p_k(x)}{m(x)} \frac{1}{k} \sum_{i=1}^{k} \int_{\mathcal{X}} \left(f(y) - f(x)\right) P_i^{(k)}(x, dy).$$

Then, the change of probability measure given by the $\sigma(X_l, l \leqslant s)$-martingale

$$M_s^{(t)} := \frac{e^{\int_r^s B(X_s)(m(X_s)-1)ds} m(X_s, s, t)}{m(x, r, t)}, \quad \text{for } r \leqslant s \leqslant t$$

exhibits the probability measure corresponding to the auxiliary process.

Before proving Theorem 1.3.1, we give some links between our approach and previous works on this subject. In the neutral case, i.e. B and $(p_k)_{k \in \mathbb{N}}$ constants, the auxiliary process coincides with the one in [BDMT11] i.e. for all $f \in \mathcal{D}(\mathcal{G})$ and $x \in \mathcal{X}$, the infinitesimal generator of the auxiliary process is given by:

$$\mathcal{A} f(x) = \mathcal{G} f(x) + B m \sum_{k \geqslant 0} \widehat{p}_k \left(\frac{1}{k} \sum_{j=1}^{k} \int_{\mathcal{X}} \left(f(y) - f(x)\right) P_j^{(k)}(x, dy)\right),$$

where $\widehat{p}_k = kp_k m^{-1}$ denote the biased reproduction law. The dynamics of this auxiliary process heavily depends on the comparison between $m(x,s,t)$ and $m(y,s,t)$, for $x,y \in \mathcal{X}$. It emphasizes several bias due to growth of the population. First, the auxiliary process jumps more than the original process, if jumping is beneficial in terms of number of descendants. This phenomenon of time-acceleration also appears for examples in [CR88, LPP95, HH09]. Moreover, the reproduction law favors the creation of a large number of descendant as in [BDMT11] and the non-neutrality favors individuals with an "efficient" trait at birth in terms of number of descendants. Finally, a new bias appears on the dynamics of the trait because of the combination of the random evolution of the trait and non-neutrality. Indeed, if the dynamics of the trait is deterministic, we have $\widehat{\mathcal{G}}_s^{(t)} f(x) = \mathcal{G} f(x)$.

The auxiliary process has been guessed through a discretization of the model using the expression of the auxiliary process in [Ban15]. However, the proof of Theorem 1.3.1 does not rely on a discretization argument but on the uniqueness of the solution to the integro-differential equation (1.20). The proof is decomposed in four parts: first, in Lemma 1.3.2, we prove that the integro-differential equation (1.20) admits a unique solution which corresponds to the semi-group of the auxiliary process defined in (1.18). Afterwards, in Lemma 1.3.3, we prove that the infinitesimal generator of this auxiliary process verifies (1.17). Then, we prove Theorem 1.3.1 for any function such that $F(x) = f_1(x_{t_1}) \ldots f_k(x_{t_k})$, $x \in \mathbb{D}([0,t],\mathcal{X})$, by induction on $k \in \mathbb{N}$. Finally, we extend the set of functions for which (1.16) is satisfied using a monotone class argument.

Let $t \geqslant 0$. We define the following family of semi-groups for $f \in \mathcal{D}(\mathcal{A})$:

$$Q_{s,r}^{(t)} f(x) = \frac{A_{r-s}(m(\cdot,r,t)f)(x)}{m(x,s,t)}, \quad s \leqslant r \leqslant t.$$

We also define:

$$\widetilde{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}(m(\cdot,s,t)f)(x) + f(x)\partial_s m(x,s,t)}{m(x,s,t)}. \tag{1.19}$$

Lemma 1.3.2. *Let $t \geqslant 0$. Under Assumptions A(1-3), B, C and D, for all $x_0 \in \mathcal{X}$ and $t_0 \leqslant t$, the family of probability measures $\left(P_{t_0,s}^{(t)}(x_0,\cdot), t_0 \leqslant s \leqslant t\right)$ is the unique solution of the following equation with unknown $(\mu_{t_0,s}(x_0,\cdot),\ t_0 \leqslant s \leqslant t)$:*

$$\begin{aligned}\mu_{t_0,s}\left(x_0,f\right) =& f\left(t_0,x_0\right) + \int_{t_0}^{s}\int_{\mathcal{X}} \left(\widetilde{\mathcal{G}}_r^{(t)} f(r,x) + \partial_r f(r,x)\right) \mu_{t_0,r}\left(x_0,dx\right) dr \\ &+ \int_{t_0}^{s}\int_{\mathcal{X}} \left[\widehat{B}_r^{(t)}(x)\int_{\mathcal{X}} f\left(r,y\right) \widehat{P}_r^{(t)}(x,dy) - B(x) f\left(r,x\right)\right] \mu_{t_0,r}\left(x_0,dx\right) dr,\end{aligned} \tag{1.20}$$

for all functions f such that, $f(s,\cdot) \in \mathcal{D}(\mathcal{A})$ for all $s \geqslant 0$ and $s \mapsto f(s,x)$ is continuously differentiable for all $x \in \mathcal{X}$.

Proof. Let $t \geqslant 0$ and let f be as in the statement of the lemma. The proof falls naturally into two parts. We first prove that $\left(P_{t_0,s}^{(t)}(x_0,\cdot), t_0 \leqslant s \leqslant t\right)$ is a solution of (1.20). We show that for all $t_0 \leqslant s \leqslant t,\ x_0 \in \mathcal{X}$,

$$m(x_0,t_0,t) P_{t_0,s}^{(t)} f(t_0,x_0) = \mathbb{E}\left(\langle Z_s, f(s,\cdot) m(\cdot,s,t)\rangle \middle| Z_{t_0} = \delta_{x_0}\right).$$

Indeed, from (1.18), the left-hand side of the above equation is equal to:

$$\begin{aligned}
m(x_0,t_0,t)P^{(t)}_{t_0,s}f(t_0,x_0) &= \mathbb{E}\left[\sum_{v\in V_s}\sum_{\substack{u\in V_t\\ u\geqslant v}} f\left(s,X_s^v\right)\big|Z_{t_0}=\delta_{x_0}\right]\\
&= \mathbb{E}\left[\sum_{v\in V_s} f\left(s,X_s^v\right)\mathbb{E}_{\delta_{x_0}}\left(\sum_{\substack{u\in V_t\\ u\geqslant v}} 1\Big|\mathcal{F}_s\right)\big|Z_{t_0}=\delta_{x_0}\right]\\
&= \mathbb{E}\left[\sum_{v\in V_s} f\left(s,X_s^v\right)\mathbb{E}\left(\sum_{u\in V_t} 1\Big|Z_s=\delta_{X_s^u}\right)\right]\\
&= \mathbb{E}\left[\sum_{v\in V_s} f\left(s,X_s^v\right)m\left(X_s^v,s,t\right)\big|Z_{t_0}=\delta_{x_0}\right].
\end{aligned}$$

Then, applying (1.7) to the function $f(s,\cdot)m(\cdot,s,t)$ and taking the expectation, we obtain:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{v\in V_s} f\left(s,X_s^v\right)m\left(X_s^v,s,t\right)\big|Z_{t_0}=\delta_{x_0}\right] = m(x_0,t_0,t)f\left(t_0,x_0\right)\\
&+\int_{t_0}^s\int_{\mathcal{X}}\left(\mathcal{G}\left(f(r,\cdot)m(\cdot,r,t)\right)(x)+f(r,x)\partial_r m(x,r,t)+\partial_r f(r,x)m(x,r,t)\right)R_{t_0,r}(x_0,dx)dr\\
&+\int_{t_0}^s\int_{\mathcal{X}}B(x)\left(\sum_{k\geqslant 0}p_k(x)\sum_{j=1}^k\int_{\mathcal{X}}f\left(r,y\right)m\left(y,r,t\right)P_j^{(k)}\left(x,dy\right)-f\left(r,x\right)m\left(x,r,t\right)\right)\\
&\qquad R_{t_0,r}(x_0,dx)dr. \qquad (1.21)
\end{aligned}$$

Finally, dividing by $m(x_0,t_0,t)$, we obtain that $\left(P^{(t)}_{t_0,s}(x_0,\cdot),t_0\leqslant s\leqslant t\right)$ is a solution of (1.20).

We now prove the uniqueness of a solution to (1.20). Without loss of generality, we assume that $t_0=0$. This part of the proof is adapted from [BT11]. Let $\left(\gamma^1_{s,t},s\leqslant t\right)$ and $\left(\gamma^2_{s,t},s\leqslant t\right)$ be two solutions of equation (1.20). Let us recall that the total variation norm is given for all measures γ^1,γ^2 on $\mathcal{X}$ with finite mass by:

$$\left\|\gamma^1-\gamma^2\right\|_{TV}=\sup_{\phi\in\mathcal{C}_b(\mathcal{X},\mathbb{R}),\|\phi\|_\infty\leqslant 1}\left|\gamma^1(\phi)-\gamma^2\left(\phi\right)\right|,$$

where $\mathcal{C}_b\left(\mathcal{X},\mathbb{R}\right)$ denotes the set of continuous bounded functions from $\mathcal{X}$ to $\mathbb{R}$. The idea is to find a function which cancels the first integral in (1.20). Let $x\in\mathcal{X}$, $t\geqslant 0$ and $r\leqslant t$. We begin by computing the differential of $\left(Q^{(t)}_{s,r}f(x),\ s\leqslant r\leqslant t\right)$ with respect to s. First, $s\mapsto A_{r-s}(m(\cdot,r,t)f)(x)$ is differentiable because $x\mapsto m(x,r,t)f(x)\in\mathcal{D}(\mathcal{G})$ and according to the backward equation, its derivative is $s\mapsto\mathcal{G}(A_{r-s}(m(\cdot,r,t)f))(x)=A_{r-s}(\mathcal{G}(m(\cdot,r,t)f))(x)$. Furthermore, $s\mapsto m(x,s,t)^{-1}$ is differentiable because $s\mapsto m(x,s,t)$ is differentiable according to the first point of Assumption D and because $m(x,s,t)\neq 0$ for all $x\in\mathcal{X}$, $t\geqslant 0$

and $s \leqslant t$. Then, we have for all $s \geqslant 0$ and $r \geqslant s$:

$$\begin{aligned}
\partial_s Q_{s,r}^{(t)} f(x) &= \frac{\partial_s A_{r-s}\left(m(\cdot,r,t)f\right)(x)}{m(x,s,t)} - \frac{\partial_s m(x,s,t)}{m(x,s,t)^2} A_{r-s}\left(m(\cdot,r,t)f\right)(x) \\
&= -\frac{\mathcal{G}\left(A_{r-s}\left(m(\cdot,r,t)f\right)\right)(x)}{m(x,s,t)} - \frac{\partial_s m(x,s,t)}{m(x,s,t)} \frac{A_{r-s}\left(m(\cdot,r,t)f\right)(x)}{m(x,s,t)} \\
&= -\left(\frac{\mathcal{G}\left(m(\cdot,r,t)Q_{s,r}^{(t)}f\right)(x)}{m(x,s,t)} + \frac{\partial_s m(x,s,t)}{m(x,s,t)} Q_{s,r}^{(t)} f(x)\right).
\end{aligned}$$

Therefore, for all $s \leqslant t$, $f \in \mathcal{D}(\mathcal{A})$, we have:

$$\partial_s Q_{s,r}^{(t)} f(x) = -\widetilde{\mathcal{G}}_s^{(t)} Q_{s,r}^{(t)} f(x). \tag{1.22}$$

Let $f \in \mathcal{D}(\mathcal{A})$ be such that $\|f\|_\infty \leqslant 1$. Let $\tau_n(x) = \inf\{t \geqslant 0, X_t \notin \mathcal{B}(x,n)\}$ where $\mathcal{B}(x,n)$ is the $\mathcal{X}$-ball of radius n and centered in x. Let us define for all $x \in \mathcal{X}$, $s \leqslant r \leqslant t$ and $n \in \mathbb{N}$:

$$Q_{s,r}^{(t),n} f(x) = \frac{\mathbb{E}_x\left[m(X_{r\wedge\tau_n(x)-s}, r\wedge\tau_n(x), t) f(X_{r\wedge\tau_n(x)-s})\right]}{m(x,s,t)}.$$

We still have $\partial_s Q_{s,r}^{(t),n} f(x) = -\widetilde{\mathcal{G}}_s^{(t)} Q_{s,r}^{(t),n} f(x)$. Moreover, for all $s \leqslant r \leqslant t$ and all $x \in \mathcal{X}$, we have:

$$\begin{aligned}
\left|Q_{s,r}^{(t),n} f(x)\right| &\leqslant \frac{\mathbb{E}_x\left[m(X_{r\wedge\tau_n(x)-s}, r\wedge\tau_n(x), t)\right]}{m(x,s,t)} \\
&\leqslant \frac{\mathbb{E}_x\left[m(X_{r\wedge\tau_n(x)-s}, r\wedge\tau_n(x), t)\right]}{\mathbb{E}\left[\mathbf{1}_{\Omega_{r\wedge\tau_n(x)}} m(X^{\varnothing}_{r\wedge\tau_n(x)}, r\wedge\tau_n(x), t) \middle| Z_s = \delta_x\right]},
\end{aligned}$$

where $\Omega_r = \left\{T_1(\bar{Z}) > r\right\}$. Conditioning with respect to $\sigma\left(X_s, s \leqslant r\wedge\tau_n(x)\right)$ on the denominator we obtain:

$$\begin{aligned}
\left|Q_{s,r}^{(t),n} f(x)\right| &\leqslant \frac{\mathbb{E}_x\left[m(X_{r\wedge\tau_n(x)-s}, r\wedge\tau_n(x), t)\right]}{\mathbb{E}\left[\exp\left(-\int_0^{r\wedge\tau_n(x)} B(X_u)du\right) m(X^{\varnothing}_{r\wedge\tau_n(x)}, r\wedge\tau_n(x), t) \middle| X_s^{\varnothing} = x\right]} \\
&\leqslant \frac{\mathbb{E}\left[m(X_{r\wedge\tau_n(x)}, r\wedge\tau_n(x), t) \middle| X_s = x\right]}{\exp\left(-r\overline{B}_n(x)\right)\mathbb{E}\left[m(X^{\varnothing}_{r\wedge\tau_n(x)}, r\wedge\tau_n(x), t) \middle| X_s^{\varnothing} = x\right]} \leqslant e^{r\overline{B}_n(x)},
\end{aligned} \tag{1.23}$$

where $\overline{B}_n(x) = \sup_{y\in\mathcal{B}(x,n)} B(y)$.

Let $T_n = \inf\left\{s \leqslant t, \gamma_{s,t}^1\left(x_0, \mathcal{B}(x_0,n)^c\right) + \gamma_{s,t}^2\left(x_0, \mathcal{B}(x_0,n)^c\right) > 0\right\}$ where $\mathcal{B}(x_0,n)^c$ is the complementary of the $\mathcal{X}$-ball of radius n and centered in x_0 with the convention that $\inf\varnothing = +\infty$. Then, using that $\left(\gamma_{s,t}^i, s \leqslant t\right)$, for $i = 1, 2$, are solutions of (1.20), we have for all $s \leqslant r \leqslant t$:

$$\begin{aligned}
&\langle \gamma^i_{s\wedge T_n,t}(x_0,\cdot), Q^{(t),n}_{s\wedge T_n,r} f\rangle = Q_{0,r}^{(t),n} f(x_0) \\
&\quad + \int_0^{s\wedge T_n} \int_{\mathcal{X}} \left[\widehat{B}_u^{(t)}(x) \int_{\mathcal{X}} Q_{u,r}^{(t),n} f(y) \widehat{P}_u^{(t)}(x,dy) - B(x) Q_{u,r}^{(t),n} f(x)\right] \gamma_{u,t}^i(x_0,dx)\,du.
\end{aligned}$$

Using (1.23), we get: we have:

$$\begin{aligned}
&\left|\gamma^1_{s\wedge T_n,t}\left(x_0, Q^{(t),n}_{s\wedge T_n,r}f\right) - \gamma^2_{s\wedge T_n,t}\left(x_0, Q^{(t),n}_{s\wedge T_n,r}f\right)\right| \\
&= \left|\int_0^{s\wedge T_n}\int_{\mathcal{X}}\left[\widehat{B}^{(t)}_u(x)\int_{\mathcal{X}} Q^{(t),n}_{u,r}f(y)\,\widehat{P}^{(t)}_u(x,dy) - B(x)Q^{(t),n}_{u,r}f(x)\right]\left(\gamma^1_{u,t}-\gamma^2_{u,t}\right)(x_0,dx)\,du\right| \\
&= \left|\int_0^{s\wedge T_n}\int_{\mathcal{X}} B(x)\left[\int_{\mathcal{X}} Q^{(t),n}_{u,r}f(y)\,\frac{m(y,u,t)}{m(x,u,t)}m(x,dy) - Q^{(t),n}_{u,r}f(x)\right]\left(\gamma^1_{u,t}-\gamma^2_{u,t}\right)(x_0,dx)\,du\right| \\
&\leqslant (C(t)\overline{m}+1)e^{r\overline{B}_{r(n,x_0)}(x_0)}\overline{B}_n(x_0)\int_0^{s\wedge T_n}\left\|\gamma^1_{u,t}-\gamma^2_{u,t}\right\|_{TV}du \\
&\leqslant (C(t)\overline{m}+1)e^{r\overline{B}_{r(n,x_0)}(x_0)}\overline{B}_n(x_0)\int_0^{s}\left\|\gamma^1_{u\wedge T_n,t}-\gamma^2_{u\wedge T_n,t}\right\|_{TV}du,
\end{aligned}$$

where $r(n,x_0) = 2n + |x_0| + \ell(s)^{1/\gamma}$ and $C(t)$ is defined in Assumption C. Then Grönwall's lemma implies that $\left\|\gamma^1_{s\wedge T_n,t}-\gamma^2_{s\wedge T_n,t}\right\|_{TV} = 0$. Taking the limit as n tends to $+\infty$, we obtain $\left\|\gamma^1_{s,t}-\gamma^2_{s,t}\right\|_{TV} = 0$ and the uniqueness of the solution to (1.20). ☐

Lemma 1.3.3. *Let $t \geqslant 0$. Under Assumption D, the generator of the semi-group of the auxiliary process $\left(P^{(t)}_{r,s},\ r\leqslant s\leqslant t\right)$ defined in* (1.18) *is $\left(\mathcal{A}^{(t)}_s,\ s\leqslant t\right)$ defined on $\mathcal{D}(\mathcal{A})$.*

For the proof of this Lemma, we need a preliminary result which proof is given in Section 1.6.3 in the appendix.

Lemma 1.3.4. *For all $t \geqslant 0$ and $s \leqslant t$, $m(\cdot,s,t)\in\mathcal{D}(\mathcal{G})$.*

We can now prove Lemma 1.3.3.

Proof. Let $t \geqslant 0$ and $f\in\mathcal{D}(\mathcal{A})$. If we take the expectation of (1.7) and differentiate with respect to t, we get that for all functions g such that, $g(s,\cdot)\in\mathcal{D}(\mathcal{A})$:

$$\partial_t R_{s,t}g(x,s) = \mathcal{G}g(x,s) + \partial_s g(x,s) + B(x)\left(\int_{\mathcal{X}} g(y,s)m(x,dy) - g(x,s)\right) := \mathcal{R}g(x,s),$$

for all $x\in\mathcal{X}$ and $s\leqslant t$, because $t\mapsto\mathbb{E}\left[\langle Z_t,f\rangle\right]$ is continuous whenever f is continuous. Next, according to Assumption D, we have the following first order Taylor expansion: for all $x\in\mathcal{X}$, $r<t$ and $h>0$,

$$\begin{aligned}
P^{(t)}_{r,r+h}f(x) &= \frac{R_{r,r+h}(m(\cdot,r+h,t)f)(x)}{m(x,r,t)} \\
&= f(x) + \frac{\mathcal{R}(m(\cdot,r,t)f)(x)}{m(x,r,t)}h + \frac{\partial_r m(x,r,t)f(x)}{m(x,r,t)}h + o(h).
\end{aligned}$$

Then

$$\lim_{h\to 0}\frac{P^{(t)}_{r,r+h}f(x) - f(x)}{h} = \frac{\mathcal{R}(m(\cdot,r,t)f)(x)}{m(x,r,t)} + \frac{\partial_r m(x,r,t)f(x)}{m(x,r,t)},$$

and we obtain:

$$\mathcal{A}^{(t)}_r f(x) = \widetilde{\mathcal{G}}^{(t)}_r f(x) + \widehat{B}^{(t)}_r(x)\left[\int_{\mathcal{X}} f(y)\,\widehat{P}^{(t)}_r(x,dy) - B(x)f(x)\right],\quad r\leqslant t. \tag{1.24}$$

However, taking $f \equiv 1$ in (1.21), which is possible according to Lemma 1.3.4, and differentiating with respect to s yields:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{u\in V_s} \left(\mathcal{G}\left(m(\cdot,s,t)\right)\left(X_s^u\right)+\partial_s m(x,s,t)\right)\middle| Z_{t_0}=\delta_{x_0}\right]\\
&= -\mathbb{E}\left[B\left(X_s^u\right)\left(\sum_{k\geqslant 0} p_k\left(X_s^u\right)\sum_{j=1}^{k}\int_{\mathcal{X}} m\left(y,s,t\right)P_j^{(k)}\left(X_s^u,dy\right)-m\left(X_s^u,s,t\right)\right)\middle| Z_{t_0}=\delta_{x_0}\right].
\end{aligned}$$

Then, for $s=t_0$ we get :

$$\begin{aligned}
\partial_s m(x,t_0,t) = -\mathcal{G}\left(m(\cdot,t_0,t)\right)(x)+B\left(x\right)m(x,t_0,t)\\
- B\left(x\right)\sum_{k\geqslant 0} p_k\left(x\right)\sum_{j=1}^{k}\int_{\mathcal{X}} m(y,t_0,t)P_j^{(k)}\left(x,dy\right),
\end{aligned} \tag{1.25}$$

for all $t_0 \geqslant 0$. Combining (1.24), (1.25), we obtain formula (1.17) for the generator of the auxiliary process. ☐

Proof of Theorem 1.3.1. We prove the result by induction on $k \in \mathbb{N}$ for any separable function $F = f_1 \dots f_k$ with $f_i \in \mathcal{D}(\mathcal{A})$ for all $i = 1\dots k$. We consider the following proposition denoted by $\mathcal{H}_k$: for all $0 < s_1 \leqslant s_2 \leqslant \dots \leqslant s_k \leqslant t$, for all $x_0 \in \mathcal{X}$ and $f_1,\dots,f_k \in \mathcal{D}(\mathcal{A})$:

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_t} f_1\left(X_{s_1}^u\right)\dots f_k\left(X_{s_k}^u\right)\right] = m\left(x_0,0,t\right)\mathbb{E}_{x_0}\left[f_1\left(Y_{s_1}^{(t)}\right)\dots f_n\left(Y_{s_k}^{(t)}\right)\right].$$

First, $\mathcal{H}_1$ holds by (1.18). Assuming that $\mathcal{H}_{k-1}$ is true for some $k>1$, we now prove $\mathcal{H}_k$. Let $0 < s_1 \leqslant s_2 \leqslant \dots \leqslant s_k \leqslant t$ and $f_1,\dots,f_k$ be measurable non-negative functions such that $f_i \in \mathcal{D}(\mathcal{A})$ for all $1\leqslant i\leqslant k$. We have using the Markov property:

$$\begin{aligned}
&\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_t} f_1\left(X_{s_1}^u\right)\dots f_k\left(X_{s_k}^u\right)\right]\\
&= \mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_{s_{k-1}}} f_1\left(X_{s_1}^u\right)\dots f_{k-1}\left(X_{s_{k-1}}^u\right)\mathbb{E}\left[\sum_{\substack{v\in V_t\\ v\geqslant u}} f_k\left(X_{s_k}^v\right)\middle|\mathcal{F}_{s_{k-1}}\right]\right]\\
&= \mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_{s_{k-1}}} f_1\left(X_{s_1}^u\right)\dots f_{k-1}\left(X_{s_{k-1}}^u\right)\mathbb{E}\left[\sum_{v\in V_t} f_k\left(X_{s_k}^v\right)\middle| Z_{s_{k-1}}=\delta_{X_{s_{k-1}}^u}\right]\right].
\end{aligned}$$

We can now use the result proved in the case $k=1$ and the last term on the right hand side is equal to:

$$\begin{aligned}
&\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_{s_{k-1}}} f_1\left(X_{s_1}^u\right)\dots f_{k-1}\left(X_{s_{k-1}}^u\right)m\left(X_{s_{k-1}}^u,s_{k-1},t\right)\mathbb{E}\left[f_k\left(Y_{s_k}^{(t)}\right)\middle|Y_{s_{k-1}}^{(t)}=X_{s_{k-1}}^u\right]\right]\\
&= \mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_t} f_1\left(X_{s_1}^u\right)\dots f_{k-1}\left(X_{s_{k-1}}^u\right)\mathbb{E}\left[f_k\left(Y_{s_k}^{(t)}\right)\middle|Y_{s_{k-1}}^{(t)}=X_{s_{k-1}}^u\right]\right]\\
&= m(x_0,0,t)\mathbb{E}_{x_0}\left[f_1\left(Y_{s_1}^{(t)}\right)\dots f_{k-1}\left(Y_{s_{k-1}}^{(t)}\right)\mathbb{E}\left[f_k\left(Y_{s_k}^{(t)}\right)\middle|Y_{s_{k-1}}^{(t)}\right]\right],
\end{aligned}$$

where the last equality is obtained using the induction hypothesis.

Finally, using Assumption D and a monotone-class argument, we extend the result to all measurable functions with respect to the Skorokhod topology (see details in Appendix 1.6.4). ☐

We now develop two other Many-to-One formulas: one to characterize the trait of the individuals over the whole tree and the other to characterize the trait of a couple of individuals.

1.3.2 A Many-to-One formula for the whole tree

We denote by:

$$\mathcal{T} = \bigcup_{s \geqslant 0} V_s \subset \mathcal{U},$$

the set of all individuals in the population. Let us recall that for $u \in \mathcal{T}$, $\alpha(u)$ and $\beta(u)$ are random variables representing respectively the time of birth and death of u.

Proposition 1.3.5. *Under Assumptions A,B, C and D, for all $x_0 \in \mathcal{X}$ and for any non-negative measurable function $F : \mathbb{D}(\mathbb{R}_+, \mathcal{X}) \times \mathbb{R}_+ \to \mathbb{R}_+$, we have:*

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{u \in \mathcal{T}} F\left(X^u_{[0,\beta(u))}, \beta(u)\right)\right] = \int_0^{+\infty} m(x_0,0,s)\mathbb{E}_{x_0}\left[F\left(Y^{(s)}_{[0,s)}, s\right) B\left(Y^{(s)}_s\right)\right] ds. \tag{1.26}$$

The left-hand side of (1.26) describes the dynamics of the trait of all individuals that were in the population. The right-hand side is the equivalent in terms of auxiliary process. Then, according to this result, the sum of the contributions of all individuals in the population is equal to the average of the auxiliary process with respect to the mean number of individuals in the population. The weight B in the right-hand side comes from the density of the lifetimes. The terms might be infinite.

Proof. We follow [Clo17] (Lemma 3.8) and provide a proof for the whole trajectories. First, we recall that for any $u \in \mathcal{T}$ and any Borel set $A \subset \mathbb{R}_+$:

$$\mathbb{P}\left(\beta(u) \in A \,\middle|\, (X^u_s)_{s\geqslant 0}, \alpha(u)\right) = \int_A B\left(X^u_t\right) \exp\left(-\int_{\alpha(u)}^t B\left(X^u_s\right) ds\right) dt.$$

Then, for all non-negative measurable functions $f : \mathbb{D}(\mathbb{R}_+, \mathcal{X}) \to \mathbb{R}_+$, we have:

$$\begin{aligned}&\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{\beta(u)} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right] \\ =&\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \left(\int_{\alpha(u)}^{\tau} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right) B\left(X^u_\tau\right) \exp\left(-\int_{\alpha(u)}^{\tau} B\left(X^u_r\right) dr\right) d\tau\right].\end{aligned}$$

Next, using Fubini's Theorem, we obtain that the right-hand side of the above equation is:

$$\begin{aligned}&\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \left(\int_s^{+\infty} B\left(X^u_\tau\right) \exp\left(-\int_{\alpha(u)}^{\tau} B\left(X^u_r\right) dr\right) d\tau\right) F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right] \\ =&\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \exp\left(-\int_{\alpha(u)}^{s} B\left(X^u_r\right) dr\right) F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right] \\ =&\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} F\left(X^u_{[0,\beta(u))}, \beta(u)\right)\right],\end{aligned} \tag{1.27}$$

where the first equality is comes from of Assumption A(4). But $\{\alpha(u) \leqslant s < \beta(u), u \in \mathcal{T}\} = \{u \in V_s\}$, then:

$$\mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}}\int_{\alpha(u)}^{\beta(u)} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right] = \mathbb{E}_{\delta_{x_0}}\left[\int_0^{+\infty} \mathbf{1}_{\{u\in V_s\}} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right]. \tag{1.28}$$

Finally combining (1.27) and (1.28) we get:

$$\begin{aligned}
\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in\mathcal{T}} f\left(X^u_{[0,\beta(u))}, \beta(u)\right)\right] &= \sum_{u\in\mathcal{U}} \mathbb{E}_{\delta_{x_0}}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} F\left(X^u_{[0,\beta(u))}, \beta(u)\right)\right] \\
&= \sum_{u\in\mathcal{U}} \mathbb{E}_{\delta_{x_0}}\left[\int_0^{+\infty} \mathbf{1}_{\{u\in V_s\}} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right) ds\right] \\
&= \int_0^{+\infty} \mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_s} F\left(X^u_{[0,s)}, s\right) B\left(X^u_s\right)\right] ds \\
&= \int_0^{+\infty} m(x_0, 0, s)\mathbb{E}_{x_0}\left[F\left(Y^{(s)}_{[0,s)}, s\right) B\left(Y^{(s)}_s\right)\right] ds,
\end{aligned}$$

where the last equality comes from the Many-to-One formula (1.16). □

1.3.3 Many-to-One formulas for forks

In this section, we characterize the law of a couple of lineage coming from two individuals alive at time t. For former results on the subject, we refer to [BDMT11] for such formulas in the neutral case and to [HR17] for many-to-few-formulas on weighted k-fold sums over particles in the case of local branching. We aim at characterizing the dynamics of the trait of a couple of individual along the spine using our auxiliary process. Those formulas have already proved useful to control the variance of estimators [HO16].

For any two functions f, g, defined respectively on two intervals I_f, I_g, for any $[a, b) \subset I_f$, $[c, d) \subset I_g$, we define the concatenation $[f_{[a,b)}, g_{[c,d)}]$ by:

$$[f_{[a,b)}, g_{[c,d)}](t) = \begin{cases} f(t) & \text{, if } t \in [a, b), \\ g(t + c - b) & \text{, if } t \in [b, b + (d - c)). \end{cases}$$

Proposition 1.3.6. *Under Assumptions A, B, C and D, for any $t \geqslant 0$, $x_0 \in \mathcal{X}$ and for any non-negative measurable function $F : \mathbb{D}([0, t], \mathcal{X})^2 \to \mathbb{R}_+$:*

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{\substack{u,v\in V_t \\ u\neq v}} F\left(X^u_{[0,t]}, X^v_{[0,t]}\right)\right] = \int_0^t m(x_0, 0, s)\mathbb{E}_{x_0}\left[B\left(Y^{(s)}_s\right) J_{s,t} F\left(Y^{(s)}_{[0,s]}\right)\right] ds, \tag{1.29}$$

where for $(x_r, r \leqslant s) \in \mathbb{D}([0, s], \mathcal{X})$:

$$J_{s,t}F(x) = \sum_{a\neq b\in\mathbb{N}} \sum_{k\geqslant\max(a,b)} p_k(x_s) \int_0^1 m\left(F^{(k)}_a(x_s, \theta), s, t\right) m\left(F^{(k)}_b(x_s, \theta), s, t\right) H_{s,t}F\left(x, F^{(k)}_a(x_s, \theta), F^{(k)}_b(x_s, \theta)\right) d\theta,$$

and for all $s \leqslant t$, $(x_s, s \leqslant t) \in \mathbb{D}([0, t], \mathcal{X})$ and $y_1, y_2 \in \mathcal{X}$:

$$H_{s,t}F(x, y_1, y_2) = \mathbb{E}\left[F\left([x_{[0,s)}; Y^{(t),1}_{[s,t]}], [x_{[0,s)}; Y^{(t),2}_{[s,t]}]\right) \mid \left(Y^{(t),1}_s, Y^{(t),2}_s\right) = (y_1, y_2)\right],$$

and $(Y^{(t),1}_s,\ s \leqslant t), (Y^{(t),2}_s,\ s \leqslant t)$ are two independent copies of $(Y^{(t)}_s,\ s \leqslant t)$.

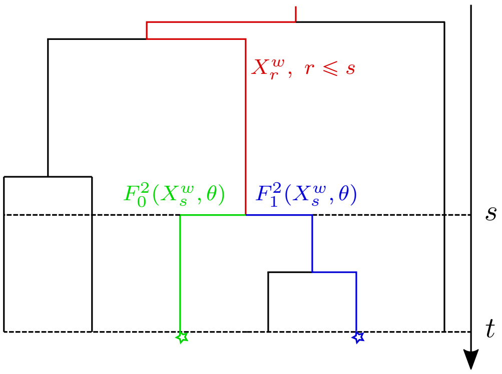

Figure 1.3.1: Forks.

According to this proposition, the sum of the contributions of each couple in the population at time t corresponds to an integral of a product of a count term and a term characterizing the dynamics of the trait the couple. The integral is over all possible death times $s \in [0,t]$ for the most recent common ancestor w of u and v, where $u \neq v \in V_t$. For example, in the case of Figure 1.3.1, if we pick the green star and the blue star, the lineage of their most recent common ancestor w is coloured in red. For the count term, $m(x_0, 0, s)$ corresponds to the choice w among the individuals in the population at time s and $m\left(F_a^{(k)}(x_s,\theta), s, t\right) m\left(F_b^{(k)}(x_s,\theta), s, t\right)$ corresponds to the choice of u and v among the descendants of w. In the example, with our choice of w, there is only one choice for u and two for v. Before s, the traits along the ancestral lineage of u and v are identical. After the death of w, the dynamics of the trait of the ancestor of u and the ancestor of v become independent conditionally to the trait of w at death. This explains the term $H_{s,t}F\left(x, F_a^{(k)}(x_s,\theta), F_b^{(k)}(x_s,\theta)\right)$ above where x, $(Y_s^{(t),1},\ s \leqslant t)$ and $(Y_s^{(t),2},\ s \leqslant t)$ describe the dynamics of the trait along the red, green and blue path respectively.

This formula is similar to the Many-to-Two formula proved in [HR17] but as in the Many-to-One formula (1.16), the count terms are separated from the terms corresponding to the dynamics of the trait of a "typical" individual contrary to the formula in [HR17]. This decomposition is useful for the study of the asymptotic behavior of the branching process. We refer the reader to Chapter 2 for an example of use of this formula to prove a law of large numbers.

Proof. Let $t \geqslant 0$ and $x_0 \in \mathcal{X}$. First we prove (1.29) for $F(x,y) = f_1(x)f_2(y)$, where $f_i : \mathbb{D}([0,t],\mathcal{X}) \to \mathbb{R}_+$ are non-negative measurable functions for $i = 1,2$. Let us denote by A the left-hand side of (1.29). We explicit the most recent common ancestor w of two individuals u, v living at time t and we obtain:

$$\begin{aligned}
A &= \mathbb{E}_{\delta_{x_0}}\left[\sum_{w\in\mathcal{U}}\sum_{a_1\neq a_2\in\mathbb{N}}\sum_{\tilde{u}_1,\tilde{u}_2\in\mathcal{T}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\mathbf{1}_{\{wa_1\tilde{u}_1\in V_t,\ wa_2\tilde{u}_2\in V_t\}}\prod_{i=1,2} f_i\left(\left[X^w_{[0,\beta(w))}; X^{wa_i\tilde{u}_i}_{[\beta(w),t]}\right]\right)\right]\\
&= \mathbb{E}_{\delta_{x_0}}\left[\sum_{\substack{wa_1\neq wa_2\in\mathcal{T}\\ a_1,a_2\in\mathbb{N}}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\mathbb{E}\left[\prod_{i=1,2}\sum_{\substack{u_i\in V_t\\ u_i\geqslant wa_i}} f_i\left(\left[X^w_{[0,\beta(w))}; X^{u_i}_{[\beta(w),t]}\right]\right)\Big|\mathcal{F}_{\beta(w)}\right]\right].
\end{aligned}$$

Then, applying successively the branching property and the Markov property, we have:

$$\begin{aligned}A =&\mathbb{E}_{\delta_{x_0}}\left[\sum_{\substack{wa_1\neq wa_2\in\mathcal{T}\\ a_1,a_2\in\mathbb{N}}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\prod_{i=1,2}\mathbb{E}\left[\sum_{\substack{u_i\in V_t\\ u_i\geqslant wa_i}} f_i\left(\left[X^w_{[0,\beta(w))};X^{u_i}_{[\beta(w),t]}\right]\right)\Big|X^{wa_i}_{[0,\beta(w)]}\right]\right]\\ =&\mathbb{E}_{\delta_{x_0}}\left[\sum_{\substack{wa_1\neq wa_2\in\mathcal{T}\\ a_1,a_2\in\mathbb{N}}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\prod_{i=1,2}\mathbb{E}\left[\sum_{\substack{u_i\in V_t\\ u_i\geqslant wa_i}} f_i\left(\left[\widetilde{x};X^{u_i}_{[\beta(w),t]}\right]\right)\Big|X^{wa_i}_{\beta(w)}\right]_{\widetilde{x}=X^w_{[0,\beta(w))}}\right].\end{aligned}$$

Next, using the Many-to-One formula (1.16) we obtain:

$$\begin{aligned}A = \mathbb{E}_{\delta_{x_0}}\Bigg[&\sum_{\substack{wa_1\neq wa_2\in\mathcal{T}\\ a_1,a_2\in\mathbb{N}}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\prod_{i=1,2} m\left(X^{wa_i}_{\beta(w)},\beta(w),t\right)\\ &\mathbb{E}\left[f_i\left(\left[\widetilde{x};Y^{(t)}_{[\beta(w),t]}\right]\right)\Big|Y^{(t)}_{\beta(w)}=X^{wa_i}_{\beta(w)}\right]_{\widetilde{x}=X^w_{[0,\beta(w))}}\Bigg].\end{aligned}$$

Expliciting the distribution of the trait at birth of wa and wb yields

$$\begin{aligned}A = \mathbb{E}_{\delta_{x_0}}\Bigg[&\sum_{\substack{wa_1\neq wa_2\in\mathcal{T}\\ a_1,a_2\in\mathbb{N}}}\mathbf{1}_{\{t\geqslant\beta(w)\}}\sum_{k\geqslant\max(a_1,a_2)} p_k\left(X^w_{\beta(w)}\right)\int_0^1\prod_{i=1,2} m\left(F^{(k)}_{a_i}\left(X^w_{\beta(w)},\theta\right),\beta(w),t\right)\\ &\times\mathbb{E}\left[f_i\left(\left[\widetilde{x};Y^{(t)}_{[\beta(w),t]}\right]\right)\Big|Y^{(t)}_{\beta(w)}=F^{(k)}_{a_i}\left(X^w_{\beta(w)^-},\theta\right)\right]_{\widetilde{x}=X^w_{[0,\beta(w))}}d\theta\Bigg].\end{aligned}$$

Applying the Many-to-One formula over the whole tree (1.26) yields:

$$\begin{aligned}A =&\int_0^t m(x_0,0,s)\mathbb{E}_x\Bigg[B\left(Y^{(s)}_s\right)\sum_{a_1\neq a_2\in\mathbb{N}}\sum_{k\geqslant\max(a_1,a_2)} p_k\left(Y^{(s)}_s\right)\\ &\int_0^1\prod_{i=1,2} m\left(F^{(k)}_{a_i}\left(Y^{(s)}_s,\theta\right),s,t\right)\mathbb{E}\left[f_i\left(\left[\widetilde{x};Y^{(t)}_{[s,t]}\right]\right)\Big|Y^{(t)}_s=F^{(k)}_{a_i}\left(Y^{(s)}_s,\theta\right)\right]_{|\widetilde{x}=Y^{(s)}_{[0,s)}}d\theta\Bigg]ds\\ =&\int_0^t m(x_0,0,s)\mathbb{E}_{x_0}\left[B\left(Y^{(s)}_s\right)J_{s,t}\left(f_1\otimes f_2\right)\left(Y^{(s)}_{[0,s)}\right)\right]ds,\end{aligned}$$

where $f_1\otimes f_2(x) = f_1(x)f_2(x)$. Finally, we obtain (1.29) using a monotone class argument.

$\square$

Let us explicit a particular case of formula (1.29). We define:

$$J_2(f,g)(x) = \sum_{a\neq b}\sum_{k\geqslant\max(a,b)} p_k\left(x\right)\int_0^1 f\left(F^{(k)}_a\left(x,\theta\right)\right)g\left(F^{(k)}_b\left(x,\theta\right)\right)d\theta. \tag{1.30}$$

J_2 represents the average trait at birth of two uniformly chosen children from an individual of type x. For simplicity of notation, we write $J_2f(x)$ instead of $J_2(f,f)(x)$. Let us recall that:

$$P^{(t)}_{r,s}f(x) = \mathbb{E}\left[f\left(Y^{(t)}_s\right)\Big|Y^{(t)}_r = x\right].$$

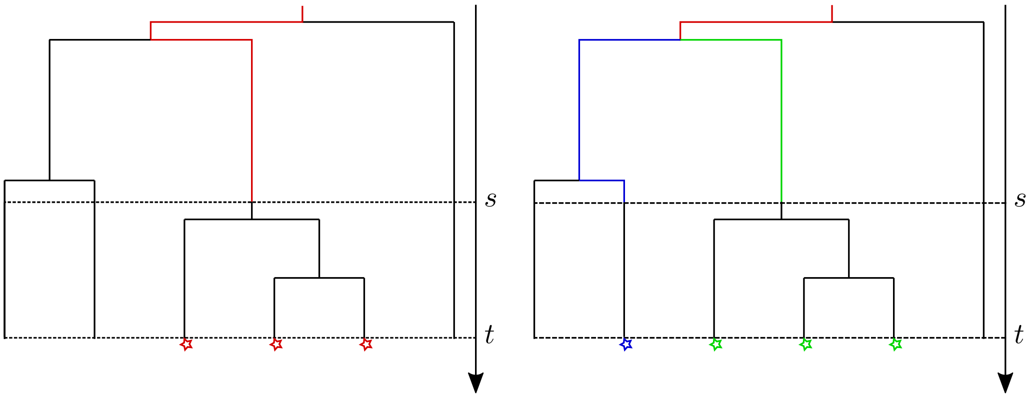

Figure 1.3.2: Trees and forks.

Corollary 1.3.7. *Under Assumptions A,B, C and D, for any non-negative measurable functions f_t, g_t from $\mathcal{X}\times\mathbb{R}^+$ to $\mathbb{R}$ and any $x_0\in\mathcal{X}$ we have for $s\leqslant t$:*

$$\begin{aligned}\mathbb{E}_{\delta_{x_0}}\left[\sum_{\substack{u,v\in V_t\\ u\neq v}} f_t\left(X_s^u\right) g_t\left(X_s^v\right)\right] &= \int_s^t m(x_0,0,r)\mathbb{E}_{x_0}\left[f_t\otimes g_t\left(Y_s^{(r)}\right) B\otimes J_2 m\left(\cdot,r,t\right)\left(Y_r^{(r)}\right)\right]dr\\ &+\int_0^s m(x_0,0,r)\mathbb{E}_{x_0}\left[B\otimes J_2\left(m(\cdot,r,t)P_{r,s}^{(t)}f_t, m(\cdot,r,t)P_{r,s}^{(t)}g_t\right)\left(Y_r^{(r)}\right)\right]dr. \quad (1.31)\end{aligned}$$

The first integral corresponds to the couple of individuals alive at time t whose most recent common ancestor died after time s. It is the case for example on Figure 1.3.2 if you pick two red stars on the tree on the left-hand side. The product $m(x_0,0,r)J_2(m(\cdot,r,t)(y)$, with $y\in\mathcal{X}$ corresponds to the average number of such couple at time t whose most recent common ancestor died at time r with $s\leqslant r\leqslant t$. The second integral corresponds to couples $(u,v)\in V_t$ of individuals whose most recent common ancestor w died before s. It is the case on Figure 1.3.2 if you pick one blue star and one green star on the tree on the right-hand side. In this case, unlike in the previous one, the value of the trait of the individuals at time s is not the same. The dynamics of the trait along the blue lineage or along the green lineage are independent conditionally to the trait of their common ancestor at death. This explains the terms $P_{r,s}^{(t)}f_t$ and $P_{r,s}^{(t)}g_t$ that appear in the second integral. As before, the remaining terms depending on the average number of individuals in the population is a count term.

1.4 Ancestral lineage of a uniform sampling at a fixed time with a large initial population

The Many-to-One formula (1.16) gives us the law of the trait of a uniformly sampled individual in an "average" population. But the characterization of the law of the trait of a uniformly sampled individual in the effective population is more complex because the number of individuals alive at time t is stochastic and depends on the dynamics of the trait of individuals. As the auxiliary process takes into account the bias in the population due to the number of individuals, it characterizes the law of a uniformly sampled individual only when the bias are in place i.e. when there is a certain amount of individuals. Indeed, the dynamics of the first individual in the population is not biased. That is why we now look at the ancestral lineage of a uniform sampling in a large population.

1.4.1 Convergence of the sampling process on a fixed time interval

It only makes sense to speak of a uniformly sampled individual at time t if the population does not become extinct before time t. For all $t \geqslant 0$, let $\Omega_t = \{N_t > 0\}$ denote the event of survival of the population. Let $\nu \in \mathcal{M}_P(\mathcal{X})$ be such that:

$$\mathbb{P}_\nu(\Omega_t) > 0. \tag{1.32}$$

We set

$$\nu_n := \sum_{i=1}^{n} \delta_{X_i}, \tag{1.33}$$

where X_i are i.i.d. random variables with distribution ν. For $t \geqslant 0$, we denote by $U(t)$ the random variable with uniform distribution on V_t conditionally on Ω_t and by $\left(X_s^{U(t)},\ s \leqslant t\right)$ the process describing the trait of a sampling along its ancestral lineage. If X is a stochastic process, we denote by X^ν the process with initial distribution $\nu \in \mathcal{M}_P(\mathcal{X})$. In particular, for all $t \geqslant 0$, $Y^{(t),\nu}$ corresponds to the auxiliary process with $Y_0^{(t)} \sim \nu$. For all $0 \leqslant s \leqslant t$,

$$m(\nu, s, t) = \mathbb{E}\left(N_t | Z_s = \nu\right),$$

denote the average number of individuals in the population after time t starting from a population distributed as ν at time s.

Theorem 1.4.1. *Under Assumptions A(1-3),B, C and D, for any $t \geqslant 0$, the sequence $\left(X_{[0,t]}^{U(t),\nu_n}\right)_{n \geqslant 0}$ converges in law in $\mathbb{D}([0,t],\mathcal{X})$ to $Y_{[0,t]}^{(t),\pi_t}$ where:*

$$\pi_t(dx) = \frac{m(x,0,t)\nu(dx)}{m(\nu,0,t)}.$$

Proof. Let $t \geqslant 0$. Let $(X_i)_{1 \leqslant i \leqslant n}$ be i.i.d random variables with distribution ν and $\nu_n = \sum_{i=1}^{n} \delta_{X_i}$. Let $F : \mathbb{D}([0,t],\mathcal{X}) \to \mathbb{R}_+$ be a bounded measurable function. First, we notice that:

$$\frac{1}{n} N_t^{\nu_n} = \frac{1}{n} \sum_{i=1}^{n} N_t^{(i)}, \tag{1.34}$$

where $N_t^{(i)}$ are independent copies of N_t with initial distribution δ_{X_i}. According to the law of large numbers, (1.34) converges almost surely as n tends to infinity to $m(\nu,0,t) = \int_{\mathcal{X}} m(x,0,t)\nu(dx)$. Next, let $\Omega_t(\nu_n) = \{N_t^{\nu_n} > 0\}$. $(\Omega_t(\nu_n))_{n \geqslant 0}$ is a increasing sequence. According to (1.32), there exists $0 < \varepsilon(t) \leqslant 1$ such that:

$$\mathbb{P}(\Omega_t(\nu_n)^C) \leqslant (1-\varepsilon(t))^n \underset{n \to +\infty}{\longrightarrow} 0,$$

so that:

$$\mathbf{1}_{\{\Omega_t(\nu_n)^C\}} \underset{n \to +\infty}{\longrightarrow} 0, \text{ almost surely.}$$

We have:

$$\mathbb{E}\left[F\left(X_{[0,t]}^{U(t),\nu_n}\right)\right] = \mathbb{E}\left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{1}{N_t^{\nu_n}} \sum_{u \in V_t^{\nu_n}} F\left(X_{[0,t]}^{u}\right)\right] \mathbb{P}\left(\Omega_t(\nu_n)\right)^{-1}.$$

Let $V_t^{(i)}, i = 1 \dots n$ be independent identically distributed populations at time t coming from an individual with trait $X_i \sim \nu$ at 0. Then:

$$\mathbb{E}\left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{1}{N_t^{\nu_n}} \sum_{u \in V_t^{\nu_n}} F\left(X_{[0,t]}^u\right)\right] = \mathbb{E}\left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{n}{N_t^{\nu_n}} \frac{1}{n} \sum_{i=1}^{n} \sum_{u \in V_t^{(i)}} F\left(X_{[0,t]}^u\right)\right]. \quad (1.35)$$

According to the strong law of large numbers,

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{u \in V_t^{(i)}} F\left(X_{[0,t]}^u\right) \underset{n \to +\infty}{\longrightarrow} \mathbb{E}_\nu\left[\sum_{u \in V_t} F\left(X_{[0,t]}^u\right)\right], \text{ almost surely.}$$

Taking the limit in (1.35) as n tends to infinity, we have by dominated convergence because F is bounded:

$$\mathbb{E}\left[F\left(X_{[0,t]}^{U(t),\nu_n}\right)\right] \underset{n \to +\infty}{\longrightarrow} \frac{1}{m(\nu, 0, t)} \int_{\mathcal{X}} \mathbb{E}_x\left[\sum_{u \in V_t} F\left(X_{[0,t]}^u\right)\right] \nu(dx),$$

because $\mathbb{P}\left(\Omega_t(\nu_n)\right) \to 1$ as n tends to infinity. Finally, applying the Many-to-one formula (1.16), we obtain:

$$\mathbb{E}\left[F\left(X_{[0,t]}^{U(t),\nu_n}\right)\right] \underset{n \to +\infty}{\longrightarrow} \frac{\int_{\mathcal{X}} m(x, 0, t) \mathbb{E}_x\left[F\left(Y_{[0,t]}^{(t)}\right)\right] \nu(dx)}{m(\nu, 0, t)}.$$

□

Remark 1.4.2. If we start with n individuals with the same trait x, we obtain:

$$\mathbb{E}\left[F\left(X_{[0,t]}^{U(t),\nu_n}\right)\right] \underset{n \to +\infty}{\longrightarrow} \mathbb{E}_x\left[F\left(Y_{[0,t]}^{(t)}\right)\right].$$

Therefore, the auxiliary process describes exactly the dynamics of the trait of a uniformly sampled individual in the large population limit, if all the starting individuals have the same trait. If the initial individuals have different traits at the beginning, the large population approximation of a uniformly sampled individual is a linear combination of the auxiliary process.

Remark 1.4.3. One can easily generalizes this results to a k-tuple of individuals uniformly picked at time t. But if you start with a population of size n and you pick k individuals uniformly at random at time t, when n tends to infinity, the probability that those k individuals comes from the same initial individual is zero. Then, the trajectories of their traits are independent and we get for example in the case $k = 2$, for any $f, g : \mathbb{D}\left([0,t], \mathcal{X}\right) \to \mathbb{R}_+$ bounded measurable functions:

$$\mathbb{E}\left[f\left(X_{[0,t]}^{U_1(t),\nu_n}\right) g\left(X_{[0,t]}^{U_2(t),\nu_n}\right)\right] \underset{n \to +\infty}{\longrightarrow} \mathbb{E}_x\left[f\left(Y_{[0,t]}^{(t),1}\right) g\left(Y_{[0,t]}^{(t),2}\right)\right],$$

where $U_1(t), U_2(t)$ are independent random variables with uniform distribution on V_t and the processes $\left(Y_s^{(t),1}, s \leqslant t\right), \left(Y_s^{(t),2}, s \leqslant t\right)$ are i.i.d. distributed as $\left(Y_s^{(t)}, s \leqslant t\right)$.

Remark 1.4.4. An other way of characterizing the trait of a uniformly sampled individual via the auxiliary process is to look at the long time behavior of the process instead of looking at the large population behavior. This is the subject of Chapter 2.

1.4.2 The trait of a uniformly sampled individual for growth-fragmentation models

The auxiliary process is a good way of getting simulated random variables corresponding to the trait of a uniformly sampled individual with large initial population or to the trait of a uniformly sampled individual for large times (see Chapter 2). Indeed, it is much more quicker to simulate one trajectory of the auxiliary process rather than the dynamics of an entire population. In this section, we detail the auxiliary process for our three examples introduced in Section 1.2.2.

Linear growth model

For the linear growth model with binary division (Section 1.2.2), Assumption C is satisfied for $C \equiv 1$ and the large population limit of the ancestral process of a sampling grows linearly between two jumps and jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = \alpha x \left(1 + \frac{1 + e^{2\sqrt{\alpha}(t-s)}}{1 - x\sqrt{\frac{\alpha}{a}} + e^{2\sqrt{\alpha}(t-s)} \left(1 + x\sqrt{\frac{\alpha}{a}} \right)} \right).$$

At a jump, there is a unique descendant with trait $\frac{x}{2}$ if x is the trait of its parent at the splitting time. In particular, the rate of division of the limiting process is bigger than the rate of division in a cell line for the original process. It means that in the large population limit, a typical individual has overcome more division than any individual.

Exponential growth model in a varying environment

For the exponential growth model in a varying environment with binary division (Section 1.2.2), Assumption C is satisfied for $C \equiv 1$ and the associated auxiliary process grows exponentially between two jumps and jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = (\alpha(s)x + \beta) \left(1 + \frac{1}{1 + x\int_s^t \alpha(r) e^{(a-\beta)(r-s)} dr} \right).$$

The rate of division of the limiting process is again bigger than the division rate of any individual. At a jump, there is a unique descendant with trait $\frac{x}{2}$ if x is the trait of its parent at the splitting time.

This example is a good illustration of the fact that the large population limit of the size of a uniformly sampled individual does not correspond to the size of a tagged cell, i.e. the size along a lineage where at each division, you choose randomly one daughter cell. In fact, as the division rate of the auxiliary process is larger than B, the number of divisions along the lineage of a uniformly sampled individual is bigger than the number of divisions along the lineage of tagged cell, resulting in a difference on the size of the individuals. However, the distribution of the number of divisions along the lineage of a uniformly sampled individual coincides with the one for the auxiliary process. On Figure 1.4.1, we can see those distributions: the two first distributions, corresponding to the distribution of the number of divisions along the lineage of a uniformly sampled individual and of the auxiliary process, are centered on a bigger number of divisions than the third distribution corresponding to a tagged cell.

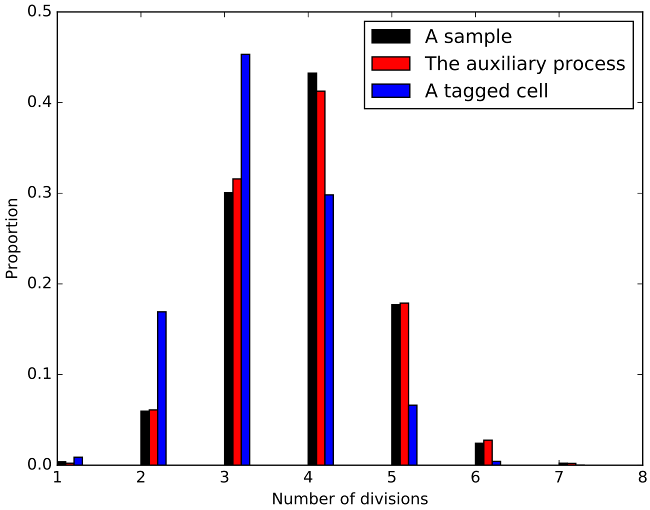

Figure 1.4.1: Distribution of the number of divisions in the lineage of a uniformly sampled individual (black bars), of the auxiliary process (red bars) and of a tagged cell (blue bars). For each case, we used 5000 realizations of each process until time $t = 50$ with parameters $a = 0.1$ and $x_0 = 1$. The distribution of the number of divisions almost coincides for the auxiliary process and a sampled individual. However, the distribution of the number of divisions for a tagged cell is different from the two previous ones. Indeed, it is more likely to sample an individual whose ancestors divided many times, that is why, the distributions of the number of divisions for the auxiliary process and for a uniformly sampled individual are centered on bigger values than the distribution of the number of divisions for a tagged cell.

Parasite infection model

For this cell division model with parasite infection, Assumption C is satisfied for $C \equiv 1$ and the auxiliary process evolves as a Feller diffusion with infinitesimal generator:

$$\mathcal{F}_s^{(t)} f(x) = \left(gx + 2\sigma^2 \frac{\alpha x \left(e^{g(t-s)} - e^{\beta(t-s)}\right)}{\alpha x \left(e^{g(t-s)} - e^{\beta(t-s)}\right) + (g-\beta)e^{\beta(t-s)}} \right) f'(x) + \sigma^2 x f''(x),$$

so that the drift of the limit of the process of the ancestral trait of a sampling is bigger than the original drift in the population. Then, the limiting process jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = (\alpha x + \beta) \left(1 + \frac{1}{1 + \frac{\alpha x}{g-\beta} \left(e^{(g-\beta)(t-s)} - 1\right)} \right).$$

Thus, the division rate of the limiting process is also bigger than the rate of division in a cell line for the original process.
The trait of the newborn cell is distributed according to the following probability law:

$$\widehat{P}_s^{(t)}(x, dy) = \mathbf{1}_{\{0 \leqslant y \leqslant x\}} \frac{2(g-\beta) + 2\alpha y \left(e^{(g-\beta)(t-s)} - 1\right)}{2(g-\beta) + \alpha x \left(e^{(g-\beta)(t-s)} - 1\right)} \frac{dy}{x}.$$

In fact, because cells divide faster when they have more parasites inside them, it is a good strategy, in order to have a lot of descendants in a long time scale, to choose to give a lot of parasites to your daughter cell. Moreover, the evolution of the trait is biased: the drift in the Feller diffusion is more important for the auxiliary process because a cell with more parasites divides faster so that it produces more descendants.

1.5 Further comments and examples

We can apply the results of this work to various models and we choose to detail in this chapter only three of them based on biological and computational considerations. However, we review in this section some other interesting models.

1.5.1 The age-structured population model

In this model, the quantity of interest is the age of each individual which grows linearly. The life-time of each individual is a random variable with cumulative distribution function G. Such models have been first introduced by Bellman and Harris in [BH52] and have recently been studied in order to infer the division rate [HO16]. Let $B : \mathbb{R}_+ \to \mathbb{R}$ be the rate of division of each cell defined via:

$$G(t) = 1 - \exp\left(-\int_0^t B(s)ds\right).$$

The branching process $(Z_t)_{t\geqslant 0}$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned}\langle Z_t, f\rangle =& \langle Z_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} f'(x) Z_s(dx)ds \\ &+ \int_0^t \int_{\mathcal{U}\times\mathbb{R}_+\times\mathbb{N}} \mathbf{1}_{\left\{u\in V_{s-},\ \theta\leqslant B\left(X_{s-}^u\right)\right\}} \left(kf\left(0\right) - f\left(X_{s-}^u\right)\right) M(ds, du, d\theta, dk),\end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times \mathbb{N}$ with intensity $ds \otimes n(du) \otimes d\theta \otimes p(dk)$, where p denotes the distribution of the number of descendants.

In order to get information on the average number of individuals in the population at time t, we follow Harris in [Har63] (Chap. 6) and we obtain:

$$m(0,0,t) = 1 - G(t) + m\int_0^t m(0,0,t-u)dG(u),$$

where m is the average number of descendants at division. Using the life-time distribution conditioned to be greater than x for the first individual we have:

$$m(x,0,t) = \frac{m(0,0,t+x) - m\int_0^x m(0,0,t+x-u)dG(u)}{1-G(x)}. \tag{1.36}$$

We cannot find an explicit solution to this renewal equation except in the case of an exponentially distributed life-time but we know the asymptotic behavior of a solution (see [Har63]). In particular, if G is non-lattice and $m > 1$, let α be the positive root of:

$$m\int_0^\infty e^{-\alpha t}dG(t) = 1.$$

Then,

$$m(0,0,t) \underset{t\to+\infty}{\sim} c(\alpha, m) n_1 e^{\alpha t},$$

where

$$c(\alpha, m) = \frac{m-1}{\alpha m^2 \int_0^\infty t e^{-\alpha t} dG(t)},$$

and n_1 is explicitly given in [Har63] Theorem 17.1 and the rate of division of the auxiliary process is given for large t by:

$$\widehat{B}_s^{(t)} \sim B(x)\frac{e^{-\alpha x}(1-G(x))}{1 - m\int_0^x e^{-\alpha u}dG(u)}.$$

1.5.2 Multi-type branching process and switching

An example of phenomenon that we would like to understand using a model on a finite state space is the phenotypic switching, i.e. the capacity to achieve multiple internal states in response to a single set of external inputs. Examples of studies of switching can be found in [OTL$^+$04] or [LK10]. For an asymptotic characterization of the ancestral lineage of a typical individual for models with a trait on a finite state space, we refer to [GB03]. We assume here that an individual can be in state 0 or 1 which is constant during its lifetime. An individual in state $x = 0, 1$ divide at rate $B(x) = b_x$ and at division, it is replaced by 2 individuals. We denote by p the probability of switching at birth. We assume that this probability does not depend on the trait. Therefore, the trait only affects the lifetime of individuals. We obtain for the generator of the first moment semi-group for any function f taking values in $\{0, 1\}$ and any $x \in \{0, 1\}$:

$$\mathcal{F}_{\text{switch}} f(x) = B(x) \left(2f(x)(1-p) + 2f(\overline{x})p - f(x) \right),$$

where $\overline{x} = 1 - x$. Moreover we have:

$$\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} B(X_t^u) \right] = (b_1 - b_0)\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} X_t^u \right] + b_0 \mathbb{E}_{\delta_x}\left[N_t \right].$$

and after calculations, we obtain:

$$\mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} X_t^u \right] = x + (b_1(1-2p) - 2pb_0) \int_0^t \mathbb{E}_{\delta_x}\left[\sum_{u \in V_s} X_s^u \right] ds + 2pb_0 \int_0^t \mathbb{E}_{\delta_x}\left[N_s \right] ds.$$

Then if we write:

$$\mu(t) = \mathbb{E}_{\delta_x}\left[N_t \right], \ \nu(t) = \mathbb{E}_{\delta_x}\left[\sum_{u \in V_t} X_t^u \right], \ \forall t \geqslant 0,$$

we obtain:

$$\begin{pmatrix} \partial_t \mu \\ \partial_t \nu \end{pmatrix} = \begin{pmatrix} b_0 & b_1 - b_0 \\ 2pb_0 & b_1(1-2p) - 2pb_0 \end{pmatrix} \begin{pmatrix} \mu \\ \nu \end{pmatrix}.$$

For example, for $p = 0.5$, writing $\gamma = \frac{b_0}{b_1}$, we have:

$$m(1, s, t) = m(0, s, t) + \left[e^{\sqrt{b_0 b_1}(t-s)} - e^{-\sqrt{b_0 b_1}(t-s)} \right] \frac{1}{2\sqrt{\gamma}}(1 - \gamma).$$

In particular, the transition kernel of the auxiliary process is given by:

$$\widehat{P}_s^{(t)}(x, dy) = \frac{m(x, s, t)\delta_x(dy) + m(\overline{x}, s, t)\delta_{\overline{x}}(dy)}{m(x, s, t) + m(\overline{x}, s, t)},$$

so that if $\gamma > 1$, i.e. $b_0 > b_1$, the auxiliary process switches more from 1 to 0 at a jump because $m(0, s, t) > m(1, s, t)$.

1.5.3 Markovian jump processes for the dynamics of the trait

The dynamics of some characteristics of a cell are non-continuous and thus cannot be described by a diffusion type process. For example, this is the case for the dynamics of populations inside a cell such as plasmids or extra-chromosomal DNA. Then, an other generalization of Kimmel's multilevel model for plasmids [Kim97] is the following: we assume that the trait of each individual evolves as a birth and death process with birth rate $\lambda > 0$ and death rate $\mu > 0$. We assume here that $\lambda - \mu > 0$. The generator of the process corresponding to the dynamics of the trait is then given for any measurable function $f : \mathbb{N} \to \mathbb{R}_+$ and any $x \in \mathbb{N}$ by:

$$\mathcal{G}f(x) = \lambda(f(x+1) - f(x)) + \mu(f(x-1) - f(x)).$$

We assume that a cell with x plasmids divides at a rate $B(x)$ and that at division, the plasmids are randomly allocated to one of the two daughter cell. The branching process $(Z_t)_{t\geqslant 0}$ is solution of the following equation, for any measurable function $f : \mathbb{N} \to \mathbb{R}_+$ and any $x \in \mathcal{X}$:

$$\begin{aligned}\langle Z_t, f\rangle = \langle Z_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} \sum_{u\in V_s} \big[&\mathbf{1}_{\{\theta \leqslant \lambda X_s^u\}} \left(f(X_s^u + 1) - f(X_s^u)\right) \\ &+\mathbf{1}_{\{\lambda X_s^u \leqslant \theta \leqslant (\lambda+\mu) X_s^u\}} \left(f(X_s^u - 1) - f(X_s^u)\right)\big] Q^u(ds, d\theta) \\ +\int_0^t \int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]} \mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant B\left(X_{s-}^u\right)\right\}} &\left(f\left(\delta X_{s-}^u\right) + f\left((1-\delta)X_{s-}^u\right) - f\left(X_{s-}^u\right)\right) M(ds, du, dz, d\delta),\end{aligned}$$

where $(Q^u)_{u\in\mathcal{U}}$ is a family of Poisson point measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $ds \otimes d\theta$ and M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times [0,1]$ with intensity $ds \otimes n(du) \otimes dz \otimes d\delta$.

For example, for the division rate $B(x) = x$, we obtain for the average number of individuals in the population after a time t:

$$m(x, s, t) = 1 + \frac{x}{\lambda - \mu}\left(e^{(\lambda-\mu)t} - 1\right).$$

In particular, the motion of the auxiliary process between jumps is given by the following generator:

$$\begin{aligned}\widehat{\mathcal{G}}_s^{(t)} f(x) =& \lambda\left[1 + \frac{e^{(\lambda-\mu)(t-s)} - 1}{\lambda - \mu + x\left(e^{(\lambda-\mu)(t-s)} - 1\right)}\right](f(x+1) - f(x)) \\ &+ \mu\left[1 - \frac{e^{(\lambda-\mu)(t-s)} - 1}{\lambda - \mu + x\left(e^{(\lambda-\mu)(t-s)} - 1\right)}\right](f(x-1) - f(x)).\end{aligned}$$

The birth rate of the plasmid population for the auxiliary process is bigger than λ and the death rate is smaller than μ. This can be explained again by the fact that cells with a lot of plasmids divides more so that they are more represented at sampling.

1.5.4 Link with the integro-differential model

The study of the average process associated with the measure-valued branching process Z is interesting in the sense that it characterizes the macroscopic evolution of the population. For a more detailed study of this link see for example [CCF16]. The following result is a corollary of Theorem 1.2.2 of Section 1.2. We recall that for all $s \geqslant 0$, $t \geqslant s$ and $x \in \mathcal{X}$,

$$R_{s,t} f(x) = \mathbb{E}\left[\sum_{u\in V_t} f\left(X_t^u\right) \middle| Z_s = \delta_x\right],$$

where f is a measurable function.

Corollary 1.5.1. *Let $f \in D(\mathcal{G})$, $s \geqslant 0$ and $x_0 \in \mathcal{X}$. Under Assumptions A(1-3) and B, the measure $\left(R_{s,t}(x_0,\cdot)\right)_{t\geqslant 0}$ is the unique solution to the following integro-differential equation:*

$$
\begin{aligned}
R_{s,t}f(s,x_0) = f(s,x_0) &+ \int_s^t \int_{\mathcal{X}} \left(\mathcal{G}f(r,x) + \partial_r f(r,x)\right) R_{s,r}(x_0,dx)ds \\
&+ \int_s^t \int_{\mathcal{X}} B(x) \left[\sum_{k\geqslant 0} p_k(x) \sum_{j=1}^{k} \int_{\mathcal{X}} f(r,y) P_j^{(k)}(x,dy) - f(r,x)\right] R_{s,r}(x_0,dx)ds,
\end{aligned}
\tag{1.37}
$$

where $\left(R_{s,t}\right)_{t\geqslant s}$ is defined in (1.9).

One can prove this result taking the expectation in (1.7) and using the same arguments as in the proof of Corollary 2.4 in [Clo17].

Let $n(t,\cdot) := R_{0,t}(x_0,\cdot)$. Equation (1.37) can be rewritten as:

$$
\begin{cases}
\partial_t n(t,x) = \mathcal{G}^T n(t,x) + \sum_{k\geqslant 0}\sum_{j=1}^{k} K_j^{(k)}\left(Bp_k n(t,\cdot)\right) - B(x)n(t,x), \\
n(0,x)dx = \delta_{x_0}(dx).
\end{cases}
$$

where $\mathcal{G}^T$ is the adjoint operator of $\mathcal{G}$ and $K_j^{(k)}$ is the adjoint operator of $f \mapsto \int_{\mathcal{X}} f(y) P_j^{(k)}(x,dy)$ as in [Clo17].

For example, in the case of the cell division model with exponential growth introduced in Section 1.2.2, we obtain in a weak sense:

$$
\partial_t n(t,x) + \partial_x\left(axn(t,x)\right) = 4B(2x)n(t,2x) - B(x)n(t,x).
$$

This is a classical growth-fragmentation equation as the one studied in [MS16] or [CCF16]. The solutions of the associated eigenvalue problem permit in particular to quantify the asymptotic global growth rate of the population.

1.6 Appendix

1.6.1 Proof of Lemma 1.2.3

We give a recursive construction of the solution to (1.7). For all $u \in \mathcal{U}$, we denote the birth time and the death time of u respectively by $\alpha(u)$ and $\beta(u)$. Let $x_0 \in \mathcal{X}$ be given. We construct a structured population $Y^k = \left(\bar{Z}^k, (X_s^u,\ s \geqslant T_k(\bar{Z}^k),\ u \in V_{T_k(\bar{Z}^k)})\right)$, where $\bar{Z}^k \in \mathbb{D}\left(\mathbb{R}_+, \mathcal{M}_P\left(\mathcal{U}\times\mathcal{X}\right)\right)$ is such that $T_{k+1} = +\infty$. We set $\alpha(\varnothing) = 0$, $X_0^{\varnothing} = x_0$, $V_0 = \{\varnothing\}$ and $\bar{Z}_t^0 \equiv \delta_{(\varnothing,x_0)}$ for all $t \geqslant 0$, so that:

$$
Y^0 = \left(\bar{Z}^0, \left(\Phi^{\varnothing}(x_0,0,t), t \geqslant 0\right)\right).
$$

Let $k \geqslant 1$. We now construct Y^{k+1}. For all $u \in V_{T_k(\bar{Z}^k)}$ such that $\alpha(u) = T_k(\bar{Z}^k)$, we set $X_t^u = \Phi^u(X_{\alpha(u)}^u, \alpha(u), t)$, for all $t \geqslant \alpha(u)$. For all $u \in V_{T_k(\bar{Z}^k)}$, let:

$$
\beta(u) = \inf\left\{t > \alpha(u),\ \int_{\alpha(u)}^{t}\int_{\mathbb{R}_+} \mathbf{1}_{\left\{z \leqslant B(X_{s^-}^u)\right\}} M(ds,\{u\},dz,[0,1],[0,1]) > 0\right\}.
$$

Let $T = \inf\left\{\beta(u),\ u \in V_{T_k(\bar{Z}^k)}\right\}$. Let $(T, U_{k+1}, \theta_{k+1}, L_{k+1}, A_{k+1})$ be the unique quintuplet of random variables such that $M\left(\{T\},\{U_{k+1}\},\{\theta_{k+1}\},\{L_{k+1}\},\{A_{k+1}\}\right) = 1$. Let:

$$
V_T = V_{T^-}\backslash\{U_{k+1}\}\bigcup\left\{U_{k+1}1,\dots,U_{k+1}G(U_{k+1},T,L_{k+1})\right\},
$$

and for all $i = 1, \dots, G\left(X_T^{U_{k+1}}, L_{k+1}\right)$, we set $\alpha(U_{k+1}i) = T$ and:

$$X_{\alpha(U_{k+1}i)}^{U_{k+1}i} = F_i\left(X_T^{U_{k+1}}, L_{k+1}, A_{k+1}\right).$$

We set

$$\begin{aligned}
\bar{Z}_t^{k+1} &= \bar{Z}_t^k, \text{ for all } t \in [0, T_k(\bar{Z}^k)],\\
\bar{Z}_t^{k+1} &= \sum_{u \in V_{T_k(\bar{Z}^k)}} \delta_{(u, X_t^u)}, \text{ for all } t \in [T_k(\bar{Z}^k), T[,\\
\bar{Z}_t^{k+1} &= \sum_{u \in V_T} \delta_{\left(u, X_T^u\right)}, \text{ for all } t \geqslant T.
\end{aligned}$$

Finally, we set $Y^{k+1} = (\bar{Z}^{k+1}, (X_s^u,\ s \geqslant T, u \in V_T))$ so that $T_{k+1}(\bar{Z}^{k+1}) = T$.

Let $\bar{Z}$ be the measure-valued branching process on $\mathbb{R}_+$ satisfying for all $k \in \mathbb{N}$ and all $t \geqslant 0$:

$$\bar{Z}_{t \wedge T_k(\bar{Z}^k)} = \bar{Z}_t^k.$$

Therefore, $T_k(\bar{Z}) = T_k(\bar{Z}^k)$ for all $k \in \mathbb{N}$. To shorten notation, we write T_k instead of $T_k(\bar{Z})$ until the end of the proof.

Let $f \in \mathcal{D}(\mathcal{G})$. We now prove by induction the following property:

$$\mathcal{H}_k : \left\{\forall t \in [T_k, T_{k+1}),\ \langle \bar{Z}_t, f\rangle \text{ is a solution to (1.7).}\right\}$$

First, $\mathcal{H}_0$ is obviously true. Assume that $\mathcal{H}_{k-1}$ is true. Let $t \in [T_k, T_{k+1})$. We recall that U_k denotes the individual who dies at time T_k. We denote by:

$$V_{t,1} = V_{T_{k-1}} \backslash \{U_k\},\ V_{t,2} = \{u \in V_t | \alpha(u) = T_k\},$$

the set of all individuals born strictly before T_k except U_k and the descendants of U_k, respectively. We have:

$$\sum_{u \in V_t} f(u, t, X_t^u) = \sum_{u \in V_{t,1}} f(u, t, X_t^u) + \sum_{u \in V_{t,2}} f(u, t, X_t^u),$$

and

$$f(u, t, X_t^u) = f\left(u, t, \Phi^u\left(X_{T_{k-1}}, T_{k-1}, t\right)\right).$$

As none of the individuals in $V_{T_{k-1}} \backslash \{U_k\}$ divides on $[T_{k-1}, t]$, we obtain using (1.6):

$$\begin{aligned}
f(u, t, X_t^u) = f\left(u, T_{k-1}, X_{T_{k-1}}^u\right) &+ \int_{T_{k-1}}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds\\
&+ M_{T_{k-1}, t}^{f,u}\left(X_{T_{k-1}}^u\right).
\end{aligned}$$

Then, we split both the integral term and the martingale in two terms to separate the behavior of the population before T_k and after T_k. We add and subtract the contribution corresponding to U_k to get a sum over all individuals alive at time T_{k-1}:

$$\begin{aligned}
\sum_{u \in V_{t,1}} f(u, t, X_t^u) = &\sum_{u \in V_{T_{k-1}} \backslash \{U_k\}} \left[\int_{T_k}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_k, t}^{f,u}\left(X_{T_k}^u\right)\right]\\
&- f\left(U_k, T_k^-, X_{T_k^-}^{U_k}\right) + \sum_{u \in V_{T_{k-1}}} \Big[f\left(u, T_{k-1}, X_{T_{k-1}}^u\right)\\
&+ \int_{T_{k-1}}^{T_k} (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_{k-1}, T_k}^{f,u}\left(X_{T_{k-1}}^u\right) \Big].
\end{aligned} \tag{1.38}$$

Using the induction hypothesis, we have:

$$\begin{aligned}\sum_{u\in V_{T_{k-1}}} f\left(u,T_{k-1},X^u_{T_{k-1}}\right) &= f\left(u,0,x_0\right)\\ &\quad+\int_0^{T_{k-1}}\int_{\mathcal{U}\times\mathcal{X}}\left(\mathcal{G}f(u,s,x)+\partial_s f\left(u,s,x\right)\right)\bar{Z}_s\left(dudx\right)ds+M^f_{0,T_{k-1}}(x_0)\\ &\quad+\int_0^{T_{k-1}}\int_E \mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant B\left(X^u_{s-}\right)\right\}}\left(\sum_{i=1}^{G(X^u_s,l)} f\left(u,s,F_i\left(X^u_s,l,\theta\right)\right)-f\left(u,s,X^u_{s-}\right)\right)\\ &\qquad M\left(ds,du,dz,dl,d\theta\right).\end{aligned}\tag{1.39}$$

Moreover, for all $s\in[T_{k-1},T_k[$, $V_s=V_{T_{k-1}}$, so that we have:

$$\sum_{u\in V_{T_{k-1}}}\int_{T_{k-1}}^{T_k}\left(\mathcal{G}f(u,s,X^u_s)+\partial_s f\left(u,s,X^u_s\right)\right)ds=\int_{T_{k-1}}^{T_k}\sum_{u\in V_s}\left(\mathcal{G}f(u,s,X^u_s)+\partial_s f\left(u,s,X^u_s\right)\right)ds.\tag{1.40}$$

Finally, we obtain combining (1.38), (1.39) and (1.40):

$$\begin{aligned}\sum_{u\in V_{t,1}} f\left(u,t,X^u_t\right) &= f\left(\varnothing,0,x_0\right)+\int_0^{T_k}\int_{\mathcal{U}\times\mathcal{X}}\left(\mathcal{G}f(u,s,x)+\partial_s f\left(u,s,x\right)\right)\bar{Z}_s\left(dudx\right)ds\\ &\quad+M^f_{0,T_{k-1}}(x_0)+\sum_{u\in V_{T_{k-1}}} M^{f,u}_{T_{k-1},T_k}\left(X^u_{T_{k-1}}\right)\\ &\quad+\int_0^{T_{k-1}}\int_E \mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant B\left(X^u_{s-}\right)\right\}}\left(\sum_{i=1}^{G(X^u_s,l)} f\left(u,s,F_i\left(X^u_s,l,\theta\right)\right)-f\left(u,s,X^u_{s-}\right)\right)\\ &\qquad M\left(ds,du,dz,dl,d\theta\right)\\ -f\left(U_k,T_k^-,X^{U_k}_{T_k^-}\right)&+\sum_{u\in V_{T_{k-1}}\setminus\{U_k\}}\left[\int_{T_k}^t\left(\mathcal{G}f(u,s,X^u_s)+\partial_s f\left(u,s,X^u_s\right)\right)ds+M^{f,u}_{T_k,t}\left(X^u_{T_k}\right)\right].\end{aligned}\tag{1.41}$$

Next, using again (1.6), we have:

$$\sum_{u\in V_{t,2}} f\left(u,t,X^u_t\right)=\sum_{u\in V_{t,2}}\left[f(u,T_k,X^u_{T_k})+\int_{T_k}^t\left(\mathcal{G}f(u,s,X^u_s)+\partial_s f\left(u,s,X^u_s\right)\right)ds+M^{f,u}_{T_k,t}\left(X^u_{T_k}\right)\right].\tag{1.42}$$

Moreover, by definition of $V_{t,2}$:

$$\sum_{u\in V_{t,2}} f(u,T_k,X^u_{T_k})=\int_{T_{k-1}}^t\int_E\mathbf{1}_{\left\{u\in V_{s-},\ z\leqslant B(X^u_s)\right\}}\sum_{i=1}^{G(X^u_s,l)} f(u,s,F_i(X^u_s,l,\theta))M(ds,du,dz,dl,d\theta).\tag{1.43}$$

Adding the martingale terms of (1.41) and (1.42), we obtain:

$$\begin{aligned}&M^f_{0,T_{k-1}}(x_0)+\sum_{u\in V_{T_{k-1}}} M^{f,u}_{T_{k-1},T_k}\left(X^u_{T_{k-1}}\right)+\sum_{u\in V_{T_{k-1}}\setminus\{U_k\}} M^{f,u}_{T_{k-1},t}\left(X^u_{T_{k-1}}\right)\\ &\quad+\sum_{u\in V_{T_k},\alpha(u)=T_k} M^{f,u}_{T_k,t}\left(X^u_{T_k}\right)=M^f_{0,T_k}(x_0).\end{aligned}\tag{1.44}$$

Finally, we obtain the result combining (1.42),(1.43) and (1.44).

1.6.2 Proof of Lemma 1.2.5

Let $\bar{Z}^{(1)}$ and $\bar{Z}^{(2)}$ be two solutions of (1.7) associated with the previously defined family of flows and Poisson point measure. For all $k \in \mathbb{N}$, we write $T_k^{(i)} = T_k(\bar{Z}^{(i)})$, $i = 1, 2$. We assume that $\bar{Z}_0^{(1)} = \bar{Z}_0^{(2)} = \delta_x$, for some $x \in \mathcal{X}$. We have $T_0^{(1)} = T_0^{(2)} = 0$. We prove by induction on $k \in \mathbb{N}$ the following proposition:

$$\mathcal{H}_k : \ T_{k+1}^{(1)} = T_{k+1}^{(2)} \text{ and } \forall t \in [T_k^{(1)}, T_{k+1}^{(1)}),\ \forall f \in \bar{\mathcal{D}}(\mathcal{G}),\ \langle \bar{Z}_t^{(1)}, f\rangle = \langle \bar{Z}_t^{(2)}, f\rangle.$$

First, $\mathcal{H}_0$ is true because:

$$T_1^{(1)} = T_1^{(2)} = \inf\left\{ t > 0,\ \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leqslant B(\Phi^{\varnothing}(x,0,s))\}} M(ds, \{\varnothing\}, dz, [0,1], [0,1]) > 0 \right\}$$

and for all $t \in [0, T_1)$, $\bar{Z}_t^{(1)} = \bar{Z}_t^{(2)} = \delta_{\Phi^{\varnothing}(x,0,t)}$. Let us assume that $\mathcal{H}_{k-1}$ is true. We first prove the second point of $\mathcal{H}_k$. By (1.7), we have for $i = 1, 2$:

$$\langle \bar{Z}^{(i)}_{T_k^{(1)}}, f\rangle = \langle \bar{Z}^{(i)}_{T_{k-1}^{(1)}}, f\rangle + I_1 + I_2,$$

where

$$I_1 = \int_{T_{k-1}^{(1)}}^{T_k^{(1)}} \int_{\mathcal{U}\times\mathcal{X}} \left(\mathcal{G}f(u,s,x) + \partial_s f\,(u,s,x)\right) \bar{Z}_s^{(i)}\,(du,dx)\,ds + M^{f,(i)}_{T_{k-1}^{(1)}, T_k^{(1)}}(x)$$

$$I_2 = \int_{T_{k-1}^{(1)}}^{T_k^{(1)}} \int_E \mathbf{1}_{\left\{u \in V_{s^-}^{(i)},\ z \leqslant B\left(X_{s^-}^{u,(i)}\right)\right\}} \Bigg(\sum_{i=1}^{G(X_s^u, l)} f\big(u, s, F_i\big(X_s^u, l, \theta\big)\big) - f\Big(u, s, X_{s^-}^{u,(i)}\Big)\Bigg)$$
$$M\left(ds, du, dz, dl, d\theta\right).$$

As the jump integral I_2 depends only on the process strictly before $T_k^{(1)}$, we obtain, using the induction hypothesis, that $\langle \bar{Z}^{(1)}_{T_k^{(1)}}, f\rangle = \langle \bar{Z}^{(2)}_{T_k^{(1)}}, f\rangle$. The evolution of the trait for $t \in \left[T_k^{(1)}, T_{k+1}^{(1)} \wedge T_{k+1}^{(2)}\right)$ only depends on the family of flows given at the beginning and which are the same for both solutions. Hence, it remains to prove that $T_{k+1}^{(1)} = T_{k+1}^{(2)}$. And it is the case because this jump time only depend on the state of the population at $T_k^{(1)}$, on the flows and on the Poisson point measure M. Finally, for all $t \in \left[T_k^{(1)}, T_{k+1}^{(1)}\right)$, we have: $\langle \bar{Z}_t^{(1)}, f\rangle = \langle \bar{Z}_t^{(2)}, f\rangle$.

Moreover, the measure-valued process is entirely characterized by $\left\{\langle \bar{Z}_t, f\rangle, f \in \bar{\mathcal{D}}(\mathcal{G})\right\}$ according to Remark 1.2.1. Therefore, there is a unique càdlàg measure-valued strong solution to (1.7) up to the kth jump time for all $k \in \mathbb{N}$.

1.6.3 Proof of Lemma 1.3.4

Let $r, s < t$ and $x \in \mathcal{X}$. We consider the event $\Omega_r = \{\text{no division before } r\}$. Then:

$$\frac{1}{r}\left(\mathbb{E}(m(X_r, s, t) \big| X_0 = x) - m(x, s, t)\right) = A(x, r, s, t) + B(x, r, s, t) + C(x, r, s, t),$$

where

$$A(x, r, s, t) = \frac{1}{r}\left(m(X_r, s, t) \big| X_0 = x) - \mathbb{E}(< Z_r, m(\cdot, s, t) f > \mathbf{1}_{\Omega_r} \big| Z_0 = \delta_x)\right),$$

$$B(x, r, s, t) = \frac{1}{r}\mathbb{E}(< Z_r, m(\cdot, s, t) > (\mathbf{1}_{\Omega_r} - 1) \big| Z_0 = \delta_x),$$

$$C(x, r, s, t) = \frac{1}{r}\left(\mathbb{E}(< Z_r, m(\cdot, s, t) > \big| Z_0 = \delta_x) - m(x, s, t)\right).$$

First,

$$A(x,r,s,t) = \frac{1}{r}\mathbb{E}_x\left(m(X_r,s,t)\mathbf{1}_{\Omega_r^C}\right).$$

Conditioning with respect to $\sigma(X_u, u \leqslant r-s)$ we obtain:

$$A(x,r,s,t) = \frac{1}{r}\mathbb{E}_x\left(m(X_r,s,t)\left(1-e^{-\int_0^r B(X_u)du}\right)\right) \xrightarrow[r\to 0]{} m(x,s,t)B(x).$$

Next, we have:

$$B(x,r,s,t) = -\frac{1}{r}\mathbb{E}(< Z_r, m(\cdot,s,t) > \mathbf{1}_{\Omega_r^C}\big|Z_0=\delta_x).$$

Then, let us denote T_1 the random variable corresponding to the lifetime of the first individual. We have using the Markov property and the branching property:

$$\begin{aligned}
B(x,r,s,t) &= -\frac{1}{r}\quad\mathbb{E}\left(\sum_{u\in V_r} m(X_r^u,s,t)\mathbf{1}_{T_1<r}\big|Z_0=\delta_x\right)\\
&= -\frac{1}{r}\quad\mathbb{E}\left(\mathbf{1}_{T_1<r}\sum_{k\geqslant 0} p_k\left(X_{T_1}^{\varnothing}\right)\right.\\
&\qquad\left.\sum_{j=0}^{k}\int_0^1 \mathbb{E}\left(\sum_{u\in V_r} m(X_r^u,s,t)\big|X_{T_1}^u = F_j^{(k)}(X_{T_1}^{\varnothing},\theta)\right)P_j^{(k)}(X_{T_1}^{\varnothing},d\theta)\big|Z_0=\delta_x\right).
\end{aligned}$$

Next, exhibiting the distribution of T_1 we obtain:

$$\begin{aligned}
B(x,r,s,t) = -\frac{1}{r}\int_0^r \mathbb{E}&\left(B\left(X_v^{\varnothing}\right)e^{-\int_0^v B\left(X_l^{\varnothing}\right)dl}\sum_{k\geqslant 0} p_k\left(X_v^{\varnothing}\right)\right.\\
&\left.\sum_{j=0}^{k}\int_0^1 \mathbb{E}\left(\sum_{u\in V_r} m(X_r^u,s,t)\big|X_v^u = F_j^{(k)}(X_v^{\varnothing},\theta)\right)P_j^{(k)}(X_v^{\varnothing},d\theta)\big|Z_0=\delta_x\right).
\end{aligned}$$

Finally:

$$B(x,r,s,t) \xrightarrow[r\to 0]{} -B(x)\sum_{k\geqslant 0} p_k(x)\sum_{j=1}^{k}\int_0^1 m\left(F_j^{(k)}(x,\theta),s,t\right)P_j^{(k)}(x,d\theta).$$

For the last term we have:

$$\begin{aligned}
C(x,r,s,t) &= \frac{1}{r}\left(\mathbb{E}(< Z_r, m(\cdot,s,t) > \big|Z_0=\delta_x) - m(x,s,t)\right)\\
&= \frac{1}{r}\left(\mathbb{E}\left(\sum_{u\in V_r} m(X_r^u,s,t)\big|Z_0=\delta_x\right) - m(x,s,t)\right)\\
&= \frac{1}{r}\left(\mathbb{E}\left(\sum_{u\in V_r}\mathbb{E}\left(\sum_{u\in V_{t-s}} 1\big|Z_0=\delta_{X_r^u}\right)\big|Z_0=\delta_x\right) - m(x,s,t)\right)\\
&= \frac{1}{r}\left(m(x,s,t+r)-m(x,s,t)\right) \xrightarrow[r\to 0]{} -\partial_t m(x,s,t)
\end{aligned}$$

Finally,

$$\lim_{r\to 0}\frac{1}{r}\left(A_r(m(\cdot,s,t))(x) - m(x,s,t)\right),$$

exists and $m(\cdot,s,t) \in \mathcal{D}(\mathcal{G})$.

1.6.4 Details of the proof of Theorem 1.3.1

We detail here the use of the monotone-class theorem in the proof of Theorem 1.3.1.

Using Remark 1.2.1, (1.16) is satisfied for any function of the form $F = \mathbf{1}_{B_1} \dots \mathbf{1}_{B_n}$, where B_i are Borel sets, for $i = 1 \dots n$, for all $n \in \mathbb{N}$. Let us define:

$$\mathcal{H} = \{F : \mathbb{D}([0,t], \mathcal{X}) \to \mathbb{R}_+ \text{ bounded and measurable such that (1.16) is satisfied for } F\},$$

and:

$$I = \left\{ \bigcap_{i=1}^{n} \{x \in \mathbb{D}([0,t], \mathcal{X}),\ x(s_i) \in B_i\}, n \in \mathbb{N},\ s_1, \dots, s_n \in \mathbb{R}_+,\ B_1, \dots, B_n \text{ Borel sets} \right\}.$$

First, I is a π-system and $\sigma(I) = \mathcal{D}$ where $\mathcal{D}$ is the Borel σ-field associated with the skorokhod topology on $\mathbb{D}([0,t], \mathcal{X})$ ([Bil13] Theorem 12.5). Then, applying the monotone-class theorem ([Wil91], Theorem 3.14.), we obtain that $\mathcal{H}$ contains all measurable functions with respect to the Skorokhod topology.

Chapter 2

A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages

Abstract

We are interested in the dynamics of a structured branching population where the trait of each individual moves according to a Markov process. The rate of division of each individual is a function of its trait and when a branching event occurs, the trait of a descendant at birth depends on the trait of the mother. We prove a law of large numbers for the empirical distribution of ancestral trajectories. It ensures that the empirical measure converges to the mean value of the spine which is a time-inhomogeneous Markov process describing the trait of a typical individual along its ancestral lineage. Our approach relies on ergodicity arguments for this time-inhomogeneous Markov process. We apply this technique on the example of a size-structured population with exponential growth in varying environment.

Content

2.1 Introduction

We are interested in the asymptotic behavior of a continuous-time structured branching Markov process. Each individual in the population is characterized by a trait which follows a Markovian dynamics and which influences the branching events. This trait may describe the position of an individual, its size, the number of parasites inside a cell, etc. The purpose of this chapter is to prove a law of large number i.e. the convergence of the empirical measure to a deterministic limit.

The law of large numbers has already been proved in many different cases. For the convergence in discrete time of the proportions of individuals with a certain type in the population, we refer to [AK98a, AK98b] with respectively a discrete or continuous set of types. The generalisation of the law of large numbers to general branching Markov processes has been obtained by Asmussen and Hering in [AH76] in both discrete and continuous time. Their proof relies on a specific decomposition of the first moment semi-group which apply to the case of branching diffusions. In the context of cellular aging, Guyon [Guy07] proved the convergence of the empirical measure for bifurcating Markov chains using the ergodicity of the spine. A generalization of those results to binary Galton-Watson processes can be found in [DM10]. For results in varying environment, we mention [BH15, Ban15]. In continuous-time, we refer to [GB03] for asymptotic results in the case of a finite number of types, to [HR14] for a strong law of large numbers in the case of local branching and to [RSZ14] for central limit theorems. The specific case of branching diffusions, popularized by Asmussen and Hering [AH76], is adressed in [EHK10]. We also mention [EW06, Eng09] for the study of the case of superdiffusions. For nonlocal branching results in continuous-time, we refer to [BT11] for the study of the proportion of infected cells in a population, to [BDMT11] for the case of a general Markov branching process with a constant division rate and to [Clo17] for the convergence of an empirical measure in the general case. Some of those results relies on spectral theory. Here, we will follow an other approach which requires no use of eigenelements as in [BDMT11] or [Guy07]. In particular, it can be applied to time-inhomogeneous dynamics.

The question of the asymptotic behavior of structured branching processes appears in many different situations and in particular in the modeling of cell population dynamics. In this context, the law of large numbers is a key result for the construction of an estimating procedure for the parameters of the model. We refer to [HO16] for the estimation of the division rate in the case of an age-structured population.

In this chapter, we prove the convergence of the empirical measure for a class of general branching Markov processes, using spinal techniques. More precisely, we use the characterization of the trait along a typical ancestral lineage introduced in Chapter 1. We adapt the techniques of [HM11] and we prove that under classical conditions [MT09], the semi-group of the auxiliary process, which is a time-inhomogeneous Markov process, is ergodic. Using this property, we prove a law of large numbers for the empirical distribution of ancestral trajectories. We also apply this technique to an example in varying environment where the law of large numbers result holds.

We describe briefly the branching process $(Z_t, t \geqslant 0)$ and we refer to Chapter 1 for its rigorous construction. We assume that individuals behave independently and that for each individual u in the population:

- its trait $(X_t^u, t \geqslant 0)$ follows a Markov process on $\mathcal{X}$ with infinitesimal generator and domain $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$,
- it dies at time t at rate $B(t, X_t^u)$ and is replaced by 2 individuals,
- the trait of the two children are both distributed according to $P(x, \cdot)$.

We focus on the empirical measure which describes the current state of the population:

$$\frac{1}{N_t} \sum_{u \in V_t} \delta_{X_t^u}, \; t \geqslant 0,$$

where V_t denotes the set of individuals alive at time t and N_t its cardinal. A crucial quantity for the study of this probability measure is the first moment semi-group applied to the constant function equal to 1:

$$m(x, s, t) := \mathbb{E}\left[N_t \middle| Z_s = \delta_x\right].$$

It is the mean number of individuals in the population at time t starting at time s with a single individual with trait $x \in \mathcal{X}$. In fact, the behavior of the empirical measure is linked with the behavior of a uniformly chosen individual in the population and the mean number of individuals in the population. More precisely, we have the following result, referred to as a Many-to-One formula (Theorem 1.3.1): for all non-negative measurable functions F on the space of càdlàg processes, for all $0 \leqslant s \leqslant t$ and $x_0 \in \mathcal{X}$,

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{u \in V_t} F\left(X_s^u, s \leqslant t\right)\right] = m(x_0, 0, t)\mathbb{E}_{x_0}\left[F\left(Y_s^{(t)}, s \leqslant t\right)\right], \tag{2.1}$$

where $\left(Y_s^{(t)}, s \leqslant t\right)$ is a time-inhomogeneous Markov process, called the auxiliary process, whose infinitesimal generators $\left(\mathcal{A}_s^{(t)}, s \leqslant t\right)$ are given for all suitable functions f and $x \in \mathcal{X}$ by:

$$\begin{aligned}\mathcal{A}_s^{(t)} f(x) = & \frac{\mathcal{G}(m(\cdot, s, t)f)(x) - f(x)\mathcal{G}(m(\cdot, s, t))(x)}{m(x, s, t)} \\ & + 2B(s, x) \int_{\mathcal{X}} (f(y) - f(x)) \frac{m(y, s, t)}{m(x, s, t)} P(x, dy).\end{aligned} \tag{2.2}$$

It corresponds to the trait of a typical individual in the population (see Chapter 1). The auxiliary process and its asymptotic behavior are the keys to obtain the main result of this chapter which is the following law of large numbers for the empirical distribution of ancestral trajectories:

$$\left(\frac{\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T\right)}{N_{t+T}} - \mathbb{E}_{x_1}\left[F\left(Y_{t+s}^{(t+T)}, s \leqslant T\right)\right]\right) \xrightarrow[t \to +\infty]{} 0, \text{ in } \mathbb{L}_2(\delta_{x_0}),$$

for all $x_0, x_1 \in \mathcal{X}$ and $T > 0$, where the $\mathbb{L}_2(\delta_{x_0})$-convergence is the $\mathbb{L}_2$-convergence with initial measure δ_{x_0}.

This result ensures that the behavior of the whole population becomes deterministic asymptotically and that this behavior is given by the limit behavior of the auxiliary process. This weak law of large numbers gives information on the ancestral lineages in the population. To establish this result, we prove in particular that under the classical drift and minorization conditions [MT09] adapted to the time-inhomogeneous case, the auxiliary process is ergodic in the sense that there exists $\overline{c} > 0$ such that for all $x, y \in \mathcal{X}$, $T > 0$, for all bounded measurable functions $F : \mathbb{D}([0, T], \mathcal{X}) \to \mathbb{R}$ and all $0 \leqslant r \leqslant t$, we have:

$$\left|P_{r,t,T}F(x) - P_{r,t,T}F(y)\right| \leqslant Ce^{-\overline{c}(t-r)} d(x, y) \left\|F\right\|_{\infty},$$

where d is a distance on $\mathcal{X}$, C is a positive constant and

$$P_{r,t,T}F(x) := \mathbb{E}\left[F\left(Y^{(t+T)}_{t+s}, s \leqslant T\right)|Y^{(t+T)}_{r} = x\right]. \tag{2.3}$$

We also apply our method to study a size-structured population with a division rate that depends both on the trait and the time. Spectral techniques fall apart in this case because of the time dependence of the division rate whereas our method works. We prove the law of large for the distribution of ancestral trajectories in this special case. In particular, we exhibit a Lyapunov function for the time-inhomogeneous auxiliary process associated with this population dynamics and we establish the minorization condition.

Outline In Section 2.2, we detail the structured branching process and the assumptions considered for its existence and uniqueness. Then, in Section 2.3, we study the asymptotic behavior of the empirical measure: first, in Section 2.3.1, we give our result on the ergodicity of the auxiliary process, then, in Section 2.3.2, we state the law of large numbers for the empirical distribution of ancestral trajectories for the structured branching process. Section 2.3.3 is dedicated to the proofs. Finally, in Section 2.4, we apply the techniques developed in the previous sections to study the asymptotic behavior of a size-structured population in a fluctuating environment.

Notation. We use the classical Ulam-Harris-Neveu notation to identify each individual. Let

$$\mathcal{U} = \bigcup_{n\in\mathbb{N}} \{0,1\}^n .$$

The first individual is labeled by $\varnothing$. When an individual $u \in \mathcal{U}$ dies, its descendants are labeled by $u0, u1$. If u is an ancestor of v, we write $u \leqslant v$. With a slight abuse of notation, for all $u \in V_t$ and $s < t$, we denote by X^u_s the trait of the unique ancestor living at time s of u.

We also introduce the following notation for the time-inhomogeneous auxiliary process: for all measurable functions f, we set

$$\mathbb{E}_x\left(f\left(Y^{(t)}_s\right)\right) := \mathbb{E}\left(f\left(Y^{(t)}_s\right)|Y^{(t)}_0 = x\right),$$

for all $x \in \mathcal{X}$, $0 \leqslant s \leqslant t$.

2.2 The structured branching process

First, we introduce some useful notations and objects to characterize the branching process. Henceforth, we work on a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

Dynamics of the trait. Let $\mathcal{X} \subset (\mathbb{R}_+)^d$ be a measurable space for some $d \geqslant 1$. It is the state space of the Markov process describing the trait of the individuals. Let $\mathcal{G} : \mathcal{D}(\mathcal{G}) \subset \mathcal{C}_b(\mathcal{X}) \to \mathcal{C}_b(\mathcal{X})$ be the infinitesimal generator associated with a strongly continuous contraction semi-group where $\mathcal{C}_b(\mathcal{X})$ denotes the continuous bounded functions on $\mathcal{X}$. Then, $(X_t, t \geqslant 0)$ is the unique $\mathcal{X}$-valued càdlàg strong Markov process solution of the martingale problem associated with $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$ (Theorems 4.4.1 and 4.4.2 in [EK86]). We denote by $(X^x_t, t \geqslant 0)$ the corresponding process starting from $x \in \mathcal{X}$.

Division events. An individual with trait x at time t dies at an instantaneous rate $B(t,x)$, where B is a continuous function from $\mathbb{R}_+ \times \mathcal{X}$ to $\mathbb{R}_+$. It is replaced by two children. Their trait at birth are distributed according to the probability measure $P(x,\cdot)$ on $\mathcal{X}^2$. The probability measures corresponding to the marginal distributions are supposed to be equal. By a slight abuse of notation, we will also denote them by P.

Remark 2.2.1. For the sake of clarity, we consider only binary division but the model can easily be extended to a random number of descendants as in Chapter 1. The choice of equal marginal distribution for the traits at birth simplifies calculation but is not mandatory.

In order to ensure the non-explosion in finite time of such a process, we need to consider the following hypotheses.

Assumption A. *We suppose that:*

1. *there exist* $b_1, b_2 : \mathbb{R}_+ \to \mathbb{R}_+^*$ *continuous and* $\gamma \geqslant 1$ *such that for all* $(t,x) \in \mathbb{R}_+ \times \mathcal{X}$,
$$B(t,x) \leqslant b_1(t)\,|x|^\gamma + b_2(t),$$
2. *for all* $x \in \mathcal{X}$:
$$\lim_{t\to+\infty} \int_0^t B(s, X_s^x)ds = +\infty, \text{ almost surely,}$$
3. *for all* $x \in \mathcal{X}$:
$$\int_{\mathcal{X}^2} (y_1 + y_2) P(x, dy_1 dy_2) \leqslant x, \text{ componentwise,}$$
4. *there exist* $c_1, c_2 \geqslant 0$ *such that for all* $x \in \mathcal{X}$:
$$\mathcal{G}h_\gamma(x) \leqslant c_1 h_\gamma(x) + c_2,$$
where γ *is defined in the first item and for* $x \in (\mathbb{R}_+)^d$, $h_\gamma(x) = |x|^\gamma = \left(\sum_{i=1}^d x_i\right)^\gamma$.

Under Assumption A, we have the strong existence and uniqueness of the structured branching process Z in the state of càdlàg measure-valued processes, where for all $t \geqslant 0$:

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},\ t \geqslant 0.$$

We refer to Theorem 1.2.2 for more details and to [FM04] for the study of càdlàg measure-valued processes.

For the existence of the auxiliary process $Y^{(t)}$ with infinitesimal generators given by (2.2), we need to consider additional assumptions on the mean number of individuals in the population at a given time. Let us define

$$\mathcal{D}(\mathcal{A}) = \left\{ f \in \mathcal{D}(\mathcal{G}) \text{ s.t. } m(\cdot,s,t)f \in \mathcal{D}(\mathcal{G})\ \forall t \geqslant 0,\ \forall s \leqslant t \right\},$$

the domain of the infinitesimal generator of the auxiliary process.

Assumption B. *We suppose that for all* $t \geqslant 0$:

- *for all* $x \in \mathcal{X}$, $s \mapsto m(x,s,t) \in \mathcal{C}^1([0,t])$,
- *for all* $x \in \mathcal{X}$, $f \in \mathcal{D}(\mathcal{A})$, $s \mapsto \mathcal{G}(m(\cdot,s,t)f)(x)$ *is continuous.*

This assumption allows us to derive the expression of the generator of the auxiliary process (Lemma 1.3.3). It is in particular satisfied in the example developed in Section 2.4 and in the examples of Chapter 1.

Assumption C. *For all $t \geqslant 0$:*

$$\sup_{x\in\mathcal{X}}\sup_{s\leqslant t}\int_{\mathcal{X}}\frac{m(y,s,t)}{m(x,s,t)}P(x,dy) < +\infty,$$

This assumption tells us that we control uniformly in x the benefit or the penalty of a division. In the general case, the control of the ratio $m(y,s,t)(m(x,s,t))^{-1}$ seems difficult to obtain. We refer to Chapter 1 or to Section 2.4 for examples where this assumption is satisfied.

2.3 Asymptotic behaviour of the structured branching process

The purpose of this section is to prove the law of large numbers result. We show that asymptotically, the behavior of the whole population corresponds to the mean behavior of the auxiliary process introduced in Chapter 1. The ergodicity of this process is the key for the proof of the law of large numbers. We notice that the ergodicity of the auxiliary process is also required for the proof of the convergence of the empirical measure in [Guy07, BDMT11, Clo17].

In Subsection 2.3.1, we prove the ergodicity of the auxiliary process. Then, in Subsection 2.3.2, we state the main theorem of this chapter which is the convergence in $\mathbb{L}_2$-norm of the difference between the empirical measure and the mean value of the auxiliary process towards zero as time goes to infinity.

2.3.1 Ergodicity of the auxiliary process

For all $t \geqslant 0$, we denote by $\left(P_{r,s}^{(t)}, r \leqslant s \leqslant t\right)$ the semi-group of the auxiliary process defined in (2.2) by its infinitesimal generators. Then, for all $x \in \mathcal{X}$, $r \leqslant s \leqslant t$, and $f : \mathcal{X} \to \mathbb{R}$, we have:

$$P_{r,s}^{(t)}f(x) = \mathbb{E}\left[f\left(Y_s^{(t)}\right)|Y_r^{(t)} = x\right].$$

The next assumption gather two classical hypothesis to obtain the ergodicity of a process (see [MT09]). We adapt them to the time-inhomogeneous case.

Assumption D. *We suppose that:*

1. *there exists a function $V : \mathcal{X} \to \mathbb{R}_+$ and $c, d > 0$ such that for all $x \in \mathcal{X}$, $t \geqslant 0$ and $s \leqslant t$:*
$$\mathcal{A}_s^{(t)}V(x) \leqslant -cV(x) + d,$$
2. *for all $0 < r < s$, there exists $\alpha_{s-r} \in (0,1)$ and a probability measure $\nu_{r,s}$ on $\mathcal{X}$ such that for all $t \geqslant s$:*
$$\inf_{x\in B(R,V)} P_{r,s}^{(t)}(x,\cdot) \geqslant \alpha_{s-r}\nu_{r,s}(\cdot),$$
with $B(R,V) = \{x \in \mathcal{X} : V(x) \leqslant R\}$ for some $R > \frac{2d}{c}$ where c, d are defined in the first point.

Adapting directly [HM11], we prove that the semi-group of the auxiliary process is a contraction operator for a well-chosen norm. For all $\beta > 0$, we define the following metric on $\mathcal{X}$:

$$d_\beta(x,y) = \begin{cases} 0 & x = y \\ 2 + \beta V(x) + \beta V(y) & x \neq y. \end{cases}$$

We can now state the result on the ergodic behavior of the trajectories of auxiliary process.

Proposition 2.3.1. *Let $T > 0$. Under Assumptions A-D, there exists $\bar{c} > 0$ and $\beta > 0$ such that for all $x, y \in \mathcal{X}$, for all bounded measurable functions $F : \mathbb{D}([0,T],\mathcal{X}) \to \mathbb{R}$ and all $0 \leqslant r \leqslant t$, we have:*

$$|P_{r,t,T}F(x) - P_{r,t,T}F(y)| \leqslant Ce^{-\bar{c}(t-r)} \|F\|_\infty d_\beta(x,y). \tag{2.4}$$

where $C > 0$ is a positive constant.

In the case of a division rate independent of time, the auxiliary process is still time-inhomogeneous but we obtain the convergence of the trajectories of the auxiliary process.

Proposition 2.3.2. *Let $T \geqslant 0$. Assume that $B(t,x) \equiv B(x)$ for all $t \geqslant 0$ and $x \in \mathcal{X}$. Then, under Assumptions A-D, there exists a probability measure Π on the Borel σ-field of $\mathbb{D}([0,T],\mathcal{X})$ endowed with the Skorokhod distance such that for all bounded measurable functions $F : \mathbb{D}([0,T],\mathcal{X}) \to \mathbb{R}$ and for all $x \in \mathcal{X}$:*

$$|P_{0,t,T}F(x) - \Pi(F)| \leqslant Ce^{-\bar{c}t} \|F\|_\infty \left(2 + 2\beta V(x) + \beta\frac{d}{c}\right).$$

This convergence is different from classical ergodicity results because $(P_{0,t}^{(t)}, t \geqslant 0)$ is not a semi-group.

2.3.2 A law of large numbers

Before stating the law of large numbers, we need to consider a final set of assumptions. For $x, y \in \mathcal{X}$ and $s > 0$, let

$$\varphi_s(x,y) = \sup_{t \geqslant s} \frac{m(x,0,s)m(y,s,t)}{m(x,0,t)}. \tag{2.5}$$

It quantifies the benefit, in term of number of individuals at time t, of "changing" the trait of the entire population at time s by the trait y. For all $x \in \mathcal{X}$, we define:

$$c(x) = \liminf_{t \to \infty} \frac{\log(m(x,0,t))}{t}, \tag{2.6}$$

which corresponds to the growth rate of the total population. In particular, if the division rate is constant $B \equiv b$, we have that $c(x) \equiv b(m-1)$. Moreover, in our case, $c(x) > 0$ for all $x \in \mathcal{X}$.

Using the same notations as in Chapter 1, we set for all measurable functions $f : \mathcal{X} \to \mathbb{R}$ and for all $x \in \mathcal{X}$:

$$Jf(x) = 2\int_{\mathcal{X}} f(y_0) f(y_1) P(x, dy_0 dy_1). \tag{2.7}$$

It represents the trait at birth of the descendants of an individual.

Assumption E. *There exists $\alpha_1, D_1 \geqslant 0$ such that $\alpha_1 < c(x)$ for all $x \in \mathcal{X}$ and for all $t > 0$:*

$$\mathbb{E}_x \left[B\left(t, Y_t^{(t)}\right) J\left(V\varphi_t\left(x, \cdot\right)\right)\left(Y_t^{(t)}\right)\right] \leqslant D_1 e^{\alpha_1 t},$$

where V is defined in Assumption D.

This is a technical assumption. In particular, if φ_t, B, V are upper bounded by polynomials and if we can control the moments of the measure m, the first point of Assumption E amounts to bound the moments of the auxiliary process. We refer the reader to Lemma 2.4.5 in the last section of this chapter for the verification of this hypothesis in an example.

We first state a slightly less strong result than the law of large numbers.

Theorem 2.3.3. *Let $T > 0$. Under Assumptions A-E, we have for all bounded measurable functions $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$, for all $x_0, x_1 \in \mathcal{X}$:*

$$\mathbb{E}_{\delta_{x_0}} \left[\left(\sum_{u \in V_{t+T}} \frac{F\left(X_{t+s}^u, s \leqslant T\right) - P_{0,t,T} F\left(x_1\right)}{m(x_0, 0, t+T)} \right)^2 \right] \xrightarrow[t \to \infty]{} 0. \tag{2.8}$$

Moreover, the rate of convergence is lower-bounded by:

$$v(t) = \exp\left(\min\left(\overline{c}, \frac{c(x_0) - \alpha_1}{2}\right) t\right),$$

where $\overline{c}$ is defined below in (2.15).

As in [Guy07, BDMT11], we could generalize this result to unbounded functions F satisfying specific conditions such as $P_{0,t}^{(t)} F \leqslant e^{bt}$ for some $b < c(x)$. The rate of convergence of the empirical measure depends both on the growth rate of the population and on the rate that governs the exponential ergodicity for the auxiliary process. The same type of rate of convergence appeared in [HO16], Theorem 3, in the case of an age structured population.

In order to derive the law of large number from the previous result, we need to control the variance of the number of individuals in the population.

Assumption F. *For all $x \in \mathcal{X}$,*

$$\sup_{t \geqslant 0} \mathbb{E}_{\delta_x} \left(\left(\frac{N_t}{m(x, 0, t)} \right)^2 \right) < \infty.$$

The meaning of this assumption is that the number of individuals at time t in the population is of the same order as the expected number of individuals in the population at time t. We can now state the law of large numbers.

Corollary 2.3.4. *Let $T > 0$. Under Assumptions A-F, for all bounded measurable functions $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$, for all $x_0, x_1 \in \mathcal{X}$, we have:*

$$\frac{\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T\right)}{N_{t+T}} - P_{0,t,T} F(x_1) \xrightarrow[t \to +\infty]{} 0, \text{ in } \mathbb{L}_2(\delta_{x_0}).$$

Remark 2.3.5. It is possible to extend this convergence to population processes allowing death events i.e. if $p_0 \not\equiv 0$. In this case, one should check that $c(x)$ defined in (2.6) is strictly positive in order to use the same proof for the convergence and the convergence is only valid on the survival event $\{N_t > 0\}$.

Remark 2.3.6. We are not able to give the rate of convergence in this case because we did not prove the convergence of $(N_t m(x,t)^{-1}, t \geqslant 0)$, for $x \in \mathcal{X}$.

In the case of a division rate that does not depend on time, even if the auxiliary process is still time-inhomogeneous, we have its convergence when the time goes to infinity. We obtain the following result.

Corollary 2.3.7. *Let $T > 0$. Under Assumptions A-F, if $B(t,x) \equiv B(x)$ for all $t \geqslant 0$ and $x \in \mathcal{X}$, there exists a probability measure Π on the Borel σ-field of $\mathbb{D}([0,T],\mathcal{X})$ endowed with the Skorokhod distance such that:*

$$\frac{\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T\right)}{N_{t+T}} \xrightarrow[t \to +\infty]{} \Pi(F), \text{ in } \mathbb{L}_2(\delta_{x_0}).$$

Therefore, the empirical measure of ancestral trajectories converges toward the limit of the auxiliary process.

2.3.3 Proofs

We first give a useful inequality. Combining the first point of Assumption D and Dynkin's formula applied to $x \mapsto e^{ct}V(x)$ where c, V are defined in Assumption D, we have:

$$P_{r,s}^{(t)} V(x) \leqslant e^{-c(s-r)} V(x) + \frac{d}{c}\left(1 - e^{-c(s-r)}\right). \tag{2.9}$$

We will use this inequality in the two following subsections.

Proof of Proposition 2.3.1

This is adapted from [HM11]. We consider the semi-norm on measurable functions from $\mathcal{X}$ into $\mathbb{R}$ defined by:

$$|||f|||_\beta = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_\beta(x,y)}.$$

We also introduce the following weighted norm:

$$\|f\|_\beta = \sup_x \frac{|f(x)|}{1 + \beta V(x)}.$$

Step 1. Let $0 \leqslant r \leqslant s \leqslant t$ and $f : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function. First, we prove that for all $\Delta > 0$, there exists $\overline{\alpha}_\Delta \in (0,1)$ and $\beta_\Delta > 0$ such that for all $r > 0$ and all $t \geqslant r + \Delta$:

$$|||P_{r,r+\Delta}^{(t)} f|||_{\beta_\Delta} \leqslant \overline{\alpha}_\Delta |||f|||_{\beta_\Delta}. \tag{2.10}$$

Let $\beta > 0$ that will be specified later. Fix $R > \frac{2d}{c}$ and $f : \mathcal{X} \to \mathbb{R}$ such that $|||f|||_\beta \leqslant 1$. Using Lemma 2.1 in [HM11], we can assume without loss of generality that $\|f\|_\beta \leqslant 1$. To obtain (2.10), it is sufficient to prove that for all $x, y \in \mathcal{X}$, there exists $\overline{\alpha}_\Delta \in (0,1)$ and $\beta_\Delta > 0$ such that:

$$\left| P_{r,r+\Delta}^{(t)} f(x) - P_{r,r+\Delta}^{(t)} f(y) \right| \leqslant \overline{\alpha}_\Delta d_{\beta_\Delta}(x,y).$$

If $x = y$, the claim is true. Let $x \neq y \in \mathcal{X}$. We assume first that x and y are such that:

$$V(x) + V(y) \geqslant R.$$

Then, we have:

$$\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| \leqslant 2 + \beta P^{(t)}_{r,r+\Delta}V(x) + \beta P^{(t)}_{r,r+\Delta}V(y),$$

because $\|f\|_\beta \leqslant 1$. Next, using (2.9) we obtain:

$$\begin{aligned}\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| &\leqslant 2 + \beta e^{-c\Delta}\left(V(x)+V(y)\right) + 2\beta\frac{d}{c}\left(1-e^{-c\Delta}\right)\\ &\leqslant 2 + \beta e^{-c\Delta}\left(V(x)+V(y)\right) + 2\beta\frac{d}{Rc}(V(x)+V(y))\left(1-e^{-c\Delta}\right).\end{aligned}$$

Let $\gamma^0_\Delta = e^{-c\Delta} + \frac{2d}{Rc}\left(1-e^{-c\Delta}\right)$. We have $\gamma^0_\Delta < 1$. Then:

$$\begin{aligned}\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| &\leqslant 2 + \beta\gamma^0_\Delta(V(x)+V(y))\\ &\leqslant \left(\frac{2+\gamma^0_\Delta\beta(V(x)+V(y))}{2+\beta(V(x)+V(y))}\right)(2+\beta V(x)+\beta V(y))\\ &\leqslant \gamma^1_\Delta d_\beta(x,y), \qquad (2.11)\end{aligned}$$

where

$$\gamma^1_\Delta = \frac{2+\beta R\gamma^0_\Delta}{2+\beta R} < 1.$$

Assume now that x and y are such that:

$$V(x)+V(y) < R.$$

Let us consider the following linear operator:

$$\widetilde{P}^{(t)}_{r,r+\Delta} = \frac{1}{1-\alpha_\Delta}P^{(t)}_{r,r+\Delta} - \frac{\alpha_\Delta}{1-\alpha_\Delta}\nu_{r,r+\Delta}.$$

We have:

$$\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| = (1-\alpha_\Delta)|\widetilde{P}^{(t)}_{r,r+\Delta}f(x) - \widetilde{P}^{(t)}_{r,r+\Delta}f(y)|.$$

According to the second point of Assumption D, $\widetilde{P}^{(t)}_{r,r+\Delta}f(x) \geqslant 0$ for all $f \geqslant 0$ and $x \in B(R,V)$. Then:

$$\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| \leqslant (1-\alpha_\Delta)\left(\widetilde{P}^{(t)}_{r,r+\Delta}f(x) + \widetilde{P}^{(t)}_{r,r+\Delta}f(y)\right).$$

Next, using that $\|f\|_\beta \leqslant 1$ and that $\widetilde{P}^{(t)}_{r,r+\Delta}V(x) \leqslant \frac{1}{1-\alpha_\Delta}P^{(t)}_{r,r+\Delta}V(x)$, we get:

$$\begin{aligned}\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| &\leqslant 2(1-\alpha_\Delta) + \beta\left(P^{(t)}_{r,r+\Delta}V(x) + P^{(t)}_{r,r+\Delta}V(y)\right)\\ &\leqslant 2\left(1-\alpha_\Delta+\beta\frac{d}{c}\left(1-e^{-c\Delta}\right)\right) + \beta e^{-c\Delta}(V(x)+V(y)),\end{aligned}$$

where the second inequality comes from (2.9). Let $\alpha^0_\Delta \in (0, \frac{2d}{Rc}\alpha_\Delta)$. Then, if we fix $\beta = \beta_\Delta := cd^{-1}\alpha^0_\Delta$, we have:

$$\begin{aligned}\left|P^{(t)}_{r,r+\Delta}f(x) - P^{(t)}_{r,r+\Delta}f(y)\right| &\leqslant 2\left(1-\alpha_\Delta+\alpha^0_\Delta\right) + \beta_\Delta e^{-c\Delta}(V(x)+V(y))\\ &\leqslant \gamma^2_\Delta d_{\beta_\Delta}(x,y), \qquad (2.12)\end{aligned}$$

where

$$\gamma^2_\Delta = e^{-c\Delta} \vee (1-(\alpha_\Delta-\alpha^0_\Delta)).$$

Finally, combining (2.11) and (2.12) and noticing that $\gamma^1_\Delta > \gamma^2_\Delta$, yields the result with $\overline{\alpha}_\Delta = \gamma^1_\Delta$.

Step 2. We now prove (2.4). Conditioning with respect to $\sigma\left(Y_u^{(t+T)}, r \leqslant u \leqslant t\right)$ and using the Markov property, we obtain:

$$P_{r,t,T}F(x) - P_{r,t,T}F(y) = \int_{\mathcal{X}} P_{t,t,T}F(z)\left(P_{r,t}^{(t+T)}(x,dz) - P_{r,t}^{(t+T)}(y,dz)\right). \tag{2.13}$$

For all $z \in \mathcal{X}$, we set $g(z) = P_{t,t,T}F(z)$. Let $\Delta > 0$ and let $l(r,t) \in \mathbb{N}, \varepsilon_{r,t} \geqslant 0$ be such that $t - r = l(r,t)\Delta + \varepsilon_{r,t}$ and $\varepsilon_{r,t} < \Delta$. Using (2.10), we have:

$$\begin{aligned}\left|P_{r,t}^{(t+T)}g(x) - P_{r,t}^{(t+T)}g(y)\right| &= \left|P_{r,r+\Delta}^{(t+T)}P_{r+\Delta,t}^{(t+T)}g(x) - P_{r,r+\Delta}^{(t+T)}P_{r+\Delta,t}^{(t+T)}g(y)\right| \\ &\leqslant \overline{\alpha}_{\Delta} d_{\beta_{\Delta}}(x,y) \|| P_{r+\Delta,t}^{(t+T)} g \||_{\beta_{\Delta}} \\ &\leqslant \left(\overline{\alpha}_{\Delta}\right)^{l(r,t)} d_{\beta}(x,y)\left\|g\right\|_{\infty},\end{aligned}$$

where $\beta = cd^{-1}$. Finally, we obtain:

$$\left|P_{r,t}^{(t+T)}g(x) - P_{r,t}^{(t+T)}g(y)\right| \leqslant Ce^{-\overline{c}(t-r)}d_{\beta}(x,y)\left\|g\right\|_{\infty}, \tag{2.14}$$

where $C := 1 + \frac{cR}{2d}$ and

$$\overline{c} := \sup_{\Delta > 0} \log(\overline{\alpha}_{\Delta}^{-1})\Delta^{-1}. \tag{2.15}$$

In particular, $\overline{c} < c$ because $\overline{\alpha}_{\Delta} > e^{-c\Delta}$. Finally, combining (2.13) and (2.14), and using that $\left\|P_{t,t+T}^{(t+T)}F\right\|_{\infty} \leqslant \|F\|_{\infty}$, we get the result.

Proof of Proposition 2.3.2

Let $F : \mathbb{D}([0,T],\mathcal{X}) \to \mathbb{R}$ be a bounded measurable function. We have for all $t, r \geqslant 0$:

$$P_{0,t+r,T}F(x) = \mathbb{E}_x\left[F\left(Y_{t+r+s}^{(t+r+T)}, s \leqslant T\right)\right] = \mathbb{E}_x\left[\mathbb{E}\left[F\left(Y_{t+r+s}^{(t+r+T)}, s \leqslant T\right)|\mathcal{F}_r^{(t+r+T)}\right]\right].$$

Using the Markov property, we have:

$$P_{0,t+r,T}F(x) = \int_{\mathcal{X}} P_{r,t+r,T}F(y)P_{0,r}^{(t+r+T)}(x,dy).$$

Since B does not depend on time, we have $m(y,r,t+r+T) = m(y,0,t+T)$. Then, using the Many-to-One formula (2.1) and the Markov property, we get:

$$\begin{aligned}P_{0,t+r,T}F(x) &= \int_{\mathcal{X}} \frac{\mathbb{E}\left[\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T\right)|Z_0 = \delta_y\right]}{m(y,0,t+T)} P_{0,r}^{(t+r+T)}(x,dy) \\ &= \int_{\mathcal{X}} P_{0,t,T}F(y)P_{0,r}^{(t+r+T)}(x,dy).\end{aligned}$$

Next,

$$|P_{0,t+r,T}F(x) - P_{0,t,T}F(x)| \leqslant \int_{\mathcal{X}} |P_{0,t,T}F(y) - P_{0,t,T}F(x)|\, P_{0,r}^{(t+r+T)}(x,dy).$$

Then, according to (2.4), there exist $\overline{c} > 0$, $\beta > 0$ and a constant $C > 0$ such that:

$$\begin{aligned}|P_{0,t+r,T}F(x) - P_{0,t,T}F(x)| &\leqslant Ce^{-\overline{c}t}\left\|F\right\|_{\infty} \int_{\mathcal{X}} \left(2 + \beta V(y) + \beta V(x)\right) P_{0,r}^{(t+r+T)}(x,dy) \\ &\leqslant Ce^{-\overline{c}t}\left\|F\right\|_{\infty}\left(2 + 2\beta V(x) + \beta\frac{d}{c}\right) \xrightarrow[r,t \to +\infty]{} 0,\end{aligned}$$

where the last inequality comes from (2.9). Finally, $(P_{0,t,T}F(x), t \geqslant 0)$ has a limit as $t \to +\infty$ and this limit is independent of x by (2.4).

Proof of Theorem 2.3.3, Corollary 2.3.4 and Corollary 2.3.7

Let $F : \mathbb{D}([0,T],\mathcal{X}) \to \mathbb{R}_+$ be a bounded measurable function. For all $x \in \mathbb{D}([0,t+T],\mathcal{X})$, $x_1 \in \mathcal{X}$, we define the following function:

$$\phi_{t,T}(x_1,(x_s, s \leqslant t+T)) = F\left(x_{t+s}, s \leqslant T\right) - P_{0,t,T}F(x_1).$$

Proof of Theorem 2.3.3. We have:

$$\mathbb{E}_{\delta_{x_0}}\left[\left(\sum_{u\in V_{t+T}} \frac{\phi_{t,T}(x_1,(X_s^u, s\leqslant t+T))}{m(x_0,0,t+T)}\right)^2\right] = A(t,T) + B(t,T),$$

where

$$A(t,T) = m(x_0,0,t+T)^{-2}\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\in V_{t+T}} \phi_{t,T}(x_1,(X_s^u, s\leqslant t+T))^2\right],$$

$$B(t,T) = m(x_0,0,t+T)^{-2}\mathbb{E}_{\delta_{x_0}}\left[\sum_{u\neq v\in V_{t+T}} \phi_{t,T}(x_1,(X_s^u, s\leqslant t+T))\phi_{t,T}(x_1,(X_s^v, s\leqslant t+T))\right].$$

For the first term, we have:

$$A(t,T) \leqslant 4e^{-c(x_0)(t+T)}\left\|F\right\|_\infty^2 \xrightarrow[t\to+\infty]{} 0.$$

For the second term, using the Many-to-One formula for forks (Proposition 1.3.6), we have:

$$m(x_0,0,t+T)^2B(t,T) = \int_0^{t+T} m(x_0,0,s)\mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right) J_{s,t+T}\phi_{t,T}(x_1,\cdot)\left(Y_r^{(s)}, r\leqslant s\right)\right]ds,$$

where for $x \in \mathbb{D}([0,s],\mathcal{X})$:

$$\begin{aligned} J_{s,t+T}\phi_{t,T}(x_1,\cdot)\left(x\right) = 2\int_{\mathcal{X}^2} & m\left(y_0,s,t+T\right)\mathbb{E}\left[\phi_{t,T}\left(x_1,\left(\widetilde{Y}_r^{(t+T)}, r\leqslant t+T\right)\right)|Y_s^{(t+T)} = y_0\right] \\ & m\left(y_1,s,t+T\right)\mathbb{E}\left[\phi_{t,T}\left(x_1,\left(\widetilde{Y}_r^{(t+T)}, r\leqslant t+T\right)\right)|Y_s^{(t+T)} = y_1\right]P(x_s,dy_0dy_1),\end{aligned}$$

where:

$$\widetilde{Y}_r^{(t+T)} = \begin{cases} x_r & \text{if } r < s, \\ Y_r^{(t+T)} & \text{if } s\leqslant r\leqslant t+T.\end{cases}$$

We split the integral into two parts:

$$B(t,T) = I_1 + I_2,$$

where:

$$I_1 = m(x_0,0,t+T)^{-2}\int_t^{t+T} m(x_0,0,s)\mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right) J_{s,t+T}\phi_{t,T}(x_1,\cdot)\left(Y_r^{(s)}, r\leqslant s\right)\right]ds,$$

$$I_2 = m(x_0,0,t+T)^{-2}\int_0^{t} m(x_0,0,s)\mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right) J_{s,t+T}\phi_{t,T}(x_1,\cdot)\left(Y_r^{(s)}, r\leqslant s\right)\right]ds.$$

For the first integral, we have:

$$\begin{aligned}
I_1 &\leqslant 4\left\|F\right\|_\infty^2 \int_t^{t+T} m(x_0,0,s)^{-1}\mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right)J\varphi_s\left(x_0,\cdot\right)\left(Y_s^{(s)}\right)\right]ds \\
&\leqslant 4\left\|F\right\|_\infty^2 \int_t^{t+T} e^{-c(x_0)s}D_1 e^{\alpha_1 s}ds \\
&\leqslant 4\left\|F\right\|_\infty^2 \frac{D_1}{c(x_0)-\alpha_1}e^{(\alpha_1-c(x_0))t} \xrightarrow[t\to+\infty]{} 0,
\end{aligned}$$

where the second inequality comes from Assumption E. Therefore, we only have to deal with the remaining integral I_2. First, we notice that for any $0 \leqslant s \leqslant t$ and $0 \leqslant r \leqslant T$,

$$\widetilde{Y}_{t+r}^{(t+T)} = Y_{t+r}^{(t+T)}.$$

Therefore, we get:

$$\phi_{t,T}\left(x_1,\left(\widetilde{Y}_r^{(t+T)}, r \leqslant t+T\right)\right) = \phi_{t,T}\left(x_1,\left(Y_r^{(t+T)}, r \leqslant t+T\right)\right).$$

Next, Assumption E yields:

$$\begin{aligned}
I_2 \leqslant& \int_0^t m(x_0,0,s)^{-1} \\
&\times \mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right)J\left(\varphi_s(x_0,\cdot)\mathbb{E}\left(\phi_{t,T}\left(x_1,\left(Y_r^{(t+T)}, r \leqslant t+T\right)\right)|Y_s^{(t+T)}=\cdot\right)\right)\left(Y_s^{(s)}\right)\right]ds.
\end{aligned}$$

Moreover, for any $y \in \mathcal{X}$ and $s \leqslant t$, we have:

$$\mathbb{E}\left(\phi_{t,T}\left(x_1,\left(Y_r^{(t+T)}, r \leqslant t+T\right)\right)\middle|Y_s^{(t+T)}=y\right) = P_{s,t,T}F(y) - P_{0,t,T}F(x_1).$$

According to Proposition 2.3.1, there exists $\overline{c} > 0$, $\beta > 0$ and $C > 0$ such that:

$$\left|P_{s,t,T}F(y) - P_{0,t,T}F(x_1)\right| \leqslant Ce^{-\overline{c}(t-s)}\left\|P_{t,t+T}^{(t+T)}F\right\|_\infty \int_{\mathcal{X}} d_\beta(y,x_2)P_{0,s}^{(t+T)}(x_1,dx_2).$$

Finally:

$$\left|P_{s,t,T}F(y) - P_{0,t,T}F(x_1)\right| \leqslant Ce^{-\overline{c}(t-s)}\left\|F\right\|_\infty\left(2+\beta V(y)+\beta V(x_1)+\beta\frac{d}{c}\right).$$

Then we have:

$$\begin{aligned}
I_2 \leqslant& C\left\|F\right\|_\infty^2 \int_0^t e^{-2\overline{c}(t-s)}m(x_0,0,s)^{-1} \\
&\times \mathbb{E}_{x_0}\left[B\left(Y_s^{(s)}\right)J\left(\varphi_s(x_0,\cdot)\left(2+\beta V(\cdot)+\beta V(x_1)+\beta\frac{d}{c}\right)\right)\left(Y_s^{(s)}\right)\right]ds.
\end{aligned}$$

Next, using Assumption E we obtain:

$$I_2 \leqslant C\left\|F\right\|_\infty^2\left(2+\beta+\beta V(x_1)+\beta\frac{d}{c}\right)\int_0^t e^{-2\overline{c}(t-s)}e^{(\alpha_1-c(x_0))s}ds,$$

where $C > 0$ denotes a positive constant which can vary from line to line. Then:

$$\begin{aligned}
I_2 &\leqslant C\left\|F\right\|_\infty^2 e^{-2\overline{c}t}\int_0^t e^{(\alpha_1-c(x_0)+2\overline{c})s}ds \\
&\leqslant C\frac{\left\|F\right\|_\infty^2}{\alpha_1-c(x_0)+2\overline{c}}e^{-2\overline{c}t}\left(e^{(\alpha_1-c(x_0)+2\overline{c})t}-1\right) \\
&\leqslant C\frac{\left\|F\right\|_\infty^2}{\alpha_1-c(x_0)+2\overline{c}}\left(e^{(\alpha_1-c(x_0))t}-e^{-2\overline{c}t}\right) \\
&\leqslant C\left\|F\right\|_\infty^2 e^{-\min(2\overline{c},c(x_0)-\alpha_1)t}.
\end{aligned}$$

Finally, we obtain:

$$A(t,T) + B(t,T) \leqslant C \|F\|_\infty^2 e^{-\min(2\overline{c}, c(x_0)-\alpha_1)t}$$

where C is a constant depending on $x_0, \beta, V(x_1), c, d, c(x_0), \alpha_1, R$. □

We now prove Corollary 2.3.4.

Proof of Corollary 2.3.4. Let $T > 0$, $\varepsilon > 0$, $x_0 \in \mathcal{X}$ and let $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$ be a bounded measurable function. Let $\delta > 0$. We have:

$$\begin{aligned}
&\mathbb{E}_{\delta_{x_0}}\left[\left(\frac{\sum_{u \in V_{t+T}} \phi_{t,T}\left(x_1, (X_s^u, s \leqslant t+T)\right)}{N_{t+T}}\right)^2\right] \\
&\leqslant \delta^2 \mathbb{E}_{\delta_{x_0}}\left[\left(\frac{\sum_{u \in V_{t+T}} \phi_{t,T}\left(x_1, (X_s^u, s \leqslant t+T)\right)}{m(x_0, 0, t+T)}\right)^2\right] + 4\|F\|_\infty^2 \mathbb{P}_{\delta_{x_0}}\left(\frac{N_t}{m(x_0,0,t+T)} \leqslant \delta^{-1}\right).
\end{aligned}$$

According to Paley-Zygmund inequality and Assumption F, we have:

$$\begin{aligned}
\mathbb{P}_{\delta_{x_0}}\left(N_t \leqslant \delta^{-1} m(x_0,0,t+T)\right) &\leqslant 1 - (1-\delta^{-1})^2 \mathbb{E}_{\delta_{x_0}}\left[\left(\frac{N_{t+T}}{m(x_0,0,t+T)}\right)^2\right]^{-1} \\
&\leqslant 1 - \frac{(1-\delta^{-1})^2}{g(x_0)}, \qquad (2.16)
\end{aligned}$$

where $g : \mathcal{X} \to \mathbb{R}_+$ is such that for all $x_0 \in \mathcal{X}$, $\mathbb{E}_{\delta_{x_0}}\left[N_{t+T}^2 m(x_0,0,t+T)^{-2}\right] \leqslant g(x_0)$. Finally, we can fix δ such that, combining (2.16) and Theorem 2.3.3, for t large enough, we have:

$$\mathbb{E}_{\delta_{x_0}}\left[\left(\frac{\sum_{u \in V_{t+T}} \phi_{t,T}\left(x_1, (X_s^u, s \leqslant t+T)\right)}{N_{t+T}}\right)^2\right] \leqslant \varepsilon.$$

□

Corollary 2.3.7 is a direct consequence of Corollary 2.3.2 and 2.3.4.

2.4 Asymptotic behavior of a time-inhomogeneous dynamics: application of ergodicity techniques

In the study of population dynamics, time-inhomogeneity typically appears in fluctuating environment. This effect can be modeled by a division rate that changes over time. In this section, we show how our method via the ergodicity of the auxiliary process applies to such models.

We consider a size-structured cell population in a fluctuating environment: each cell grows exponentially at rate $a > 0$ and division occurs at time t at rate $B(t,x) = x\varphi(t)$, if x is the size of the cell at time t. We assume that $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and that there exist $\varphi_1, \varphi_2 > 0$ such that for all $t \in \mathbb{R}_+$:

$$\varphi_1 \leqslant \varphi(t) \leqslant \varphi_2.$$

This choice of a polynomial dependence in the size for the division rate B is classical in the study of growth-fragmentation [MS16]. The originality comes from the function φ which model a changing environment.

At division, the cell splits into two daughter cells of size θx and $(1-\theta)x$, with $\theta \sim \mathcal{U}([\varepsilon, 1-\varepsilon])$ for some $0 < \varepsilon < \frac{1}{2}$ and x the size of the cell at division. Then, the process that we consider is a Piecewise Deterministic Markov Process (PDMP) with jump rate B and transition density function P given by:

$$P(x,y) = \begin{cases} \frac{1}{1-2\varepsilon} & \text{if } \varepsilon x \leqslant y \leqslant (1-\varepsilon)x, \\ 0 & \text{otherwise.} \end{cases}$$

Let us first make some comments on the choice of the model. The function φ is lower bounded to ensure that each cell effectively divides after some time. The upper bound is convenient for the calculations. An interesting example is $B(t,x) = x(\alpha + \beta \sin(t))$, with $\alpha - \beta > 0$ for the modeling of the growth of a cell population in a periodic environment. Finally, we consider a uniform law on $[\varepsilon, 1-\varepsilon]$ for the kernel at division but the next lemmas can easily be extend to a more general kernel.

Following the same calculations as in Section 1.2.2, we have:

$$m(x,s,t) = 1 + x\int_s^t \varphi(r)e^{a(r-s)}dr, \ \forall x \in \mathbb{R}_+.$$

Moreover, in this case, the infinitesimal generator of the auxiliary process is given by:

$$\mathcal{A}_s^{(t)} f(x) = axf'(x) + 2x\varphi(s)\int_{\varepsilon x}^{(1-\varepsilon)x} (f(y) - f(x))\frac{m(y,s,t)}{m(x,s,t)}\frac{dy}{(1-2\varepsilon)x}, \tag{2.17}$$

for all $f : \mathbb{R}_+ \to \mathbb{R}$ continuously differentiable, $s, t \in \mathbb{R}_+$ such that $s < t$ and $x \in \mathbb{R}_+$. Then, the division rate of the auxiliary process is given by:

$$\widehat{B}_s^{(t)}(x) = 2x\varphi(s)\frac{m(\frac{x}{2}, s, t)}{m(x,s,t)},$$

and the transition kernel for the trait at birth is given by:

$$\widehat{P}_s^{(t)}(x,dy) = \widehat{P}_s^{(t)}(x,y)dy = \frac{m(y,s,t)}{x(1-2\varepsilon)m(\frac{x}{2},s,t)}\mathbf{1}_{\{\varepsilon x \leqslant y \leqslant (1-\varepsilon)x\}}dy.$$

Finally, we set:

$$\phi(s,t) = \int_s^t \varphi(r)e^{a(r-s)}dr.$$

We have the following result on the asymptotic behavior of the measure-valued branching process:

Theorem 2.4.1. *Let $T > 0$. For all bounded measurable functions $F : \mathbb{D}([0,T], \mathcal{X}) \to \mathbb{R}$, for all $x_0, x_1 \in \mathcal{X}$, we have:*

$$\frac{\sum_{u \in V_{t+T}} F\left(X_{t+s}^u, s \leqslant T\right)}{N_{t+T}} - \mathbb{E}_{x_1}\left[F\left(Y_{t+s}^{(t+T)}, s \leqslant T\right)\right] \xrightarrow[t \to +\infty]{} 0, \text{ in } \mathbb{L}_2(\delta_{x_0}). \tag{2.18}$$

The proof of Theorem 2.4.1 is detailed in several lemmas. First, in Lemma 2.4.2, we exhibit a Lyapunov function and a probability measure which ensure that Assumption D is satisfied. Next, in Lemma 2.4.3, we prove that the moments of the auxiliary process are bounded. Finally, in Lemmas 2.4.5 and 2.4.6, we prove that Assumptions E and F are satisfied.

Let $V(x) = \frac{1}{x} + x$ for $x \in \mathbb{R}_+^*$.

Lemma 2.4.2. *We have the following:*

1. *There exists $d(\varepsilon) > 0$ such that for all $0 \leqslant s \leqslant t$ and $x \in \mathbb{R}_+^*$ we have:*

$$\mathcal{A}_s^{(t)} V(x) \leqslant -aV(x) + d(\varepsilon).$$

2. *For all $R > 2d(\varepsilon)a^{-1}$, for all $r < s \leqslant t$, there exists $\alpha_{s-r} > 0$ such that for all Borel set A of $\mathbb{R}_+$:*

$$\inf_{x \in B(R,V)} \mathbb{P}\left(Y_s^{(t)} \in A \middle| Y_r^{(t)} = x\right) \geqslant \alpha_{s-r} \nu_{r,s}(A).$$

Proof. We first prove that V is a Lyapunov function. Let us compute $\mathcal{A}_s^{(t)} V_1(x)$ where $V_1(x) = x$. We have for $x \in \mathbb{R}_+$:

$$\begin{aligned}\mathcal{A}_s^{(t)} V_1(x) &= ax + \frac{2}{1-2\varepsilon}\varphi(s) \int_{\varepsilon x}^{(1-\varepsilon)x} (y-x) \frac{1+y\phi(s,t)}{1+x\phi(s,t)} dy \\ &= ax - \varphi(s)x^2 + \frac{2}{3}\varphi(s)(\varepsilon^2 - \varepsilon + 1)x^2 \left(1 - \frac{1}{1+x\phi(s,t)}\right).\end{aligned}$$

Then, we obtain

$$\mathcal{A}_s^{(t)} V_1(x) \leqslant a\left(1 - \frac{1}{3a}\varphi(r)(1+2\varepsilon-2\varepsilon^2)x\right)x \leqslant -ax + \frac{3a^2}{\varphi_1(1+2\varepsilon-2\varepsilon^2)}.$$

Next, let $V_2(x) = \frac{1}{x}$. We have:

$$\mathcal{A}_s^{(t)} V_2(x) = -\frac{a}{x} + \frac{2}{1-2\varepsilon}\varphi(s) \int_{\varepsilon x}^{(1-\varepsilon)x} \left(\frac{1}{y} - \frac{1}{x}\right) \frac{1+y\phi(s,t)}{1+x\phi(s,t)} dy.$$

Using that for all $x \geqslant 0$ and $y \in [\varepsilon x, (1-\varepsilon)x]$, $1 + y\phi(s,t) \leqslant 1 + x\phi(s,t)$, we get:

$$\mathcal{A}_s^{(t)} V_2(x) \leqslant -\frac{a}{x} + 2\varphi(s)C(\varepsilon),$$

where $C(\varepsilon) = \frac{1}{1-2\varepsilon}\left[\log\left(\frac{1-\varepsilon}{\varepsilon}\right) - (1-2\varepsilon)\right]$. Noticing that $C(\varepsilon) > 0$ because $\varepsilon > \frac{1}{2}$ yields:

$$\mathcal{A}_s^{(t)} V_2(x) \leqslant -aV_2(x) + 2\varphi_2 C(\varepsilon). \tag{2.19}$$

Finally:

$$\mathcal{A}_s^{(t)} V(x) \leqslant -aV(x) + d(\varepsilon),$$

where

$$d(\varepsilon) = 2\varphi_2 C(\varepsilon) + \frac{3a^2}{\varphi_1(1+2\varepsilon-2\varepsilon^2)}.$$

Next, we prove the minorization condition. Let us specify the shape of the subset $B(R,V)$ of $\mathbb{R}_+$ that we will consider. For all $R > 2d(\varepsilon)a^{-1}$, we have:

$$B(R,V) = \{x \in \mathbb{R}_+,\ V(x) < R\} = \{x_1(R) < x < x_2(R)\}, \tag{2.20}$$

where:

$$x_1(R) = \frac{R - \sqrt{R^2-4}}{2},\ x_2(R) = \frac{R + \sqrt{R^2-4}}{2}.$$

Now, we prove the second point. Let $R > 2d(\varepsilon)a^{-1}$, $x \in B(R,V)$ and let A be a Borel set. Let $n \in \mathbb{N}$ be such that

$$\left(\frac{1-\varepsilon}{\varepsilon}\right)^{n-1} > \frac{x_2(R)}{x_1(R)}. \tag{2.21}$$

Let $0 \leqslant r < s \leqslant t$. Considering the case where the auxiliary process jumped exactly n times between r and s, we have:

$$P_{r,s}^{(t)}(x,A) \geqslant \mathbb{E}\left[\mathbf{1}_{\left\{Y_s^{(t)}\in A\right\}}\mathbf{1}_{\{r\leqslant\tau_1\leqslant s\}}\mathbf{1}_{\{\tau_1\leqslant\tau_2\leqslant s\}}\ldots\mathbf{1}_{\{\tau_{n-1}\leqslant\tau_n\leqslant s\}}\mathbf{1}_{\{\tau_{n+1}\geqslant s\}}\middle|Y_r^{(t)}=x\right],$$

where τ_i denotes the time of the ith jump of the auxiliary process, $i = 1,\ldots,n$. Let us denote by $\mathcal{F}_s^{(t)}$ the filtration generated by the auxiliary process $(Y_s^{(t)}, s \leqslant t)$ up to time s. Conditioning with respect to $\mathcal{F}_{\tau_1}^{(t)}$ and using the strong Markov property and the fact that between two jumps, the growth of the auxiliary process is exponential at rate a, we get:

$$\begin{aligned}P_{r,s}^{(t)}(x,A) \geqslant \mathbb{E}\Bigg[\mathbf{1}_{\{r\leqslant\tau_1\leqslant s\}}\int_{J_{r,\tau_1}(x)}\mathbb{E}\left[\mathbf{1}_{\left\{Y_s^{(t)}\in A\right\}}\mathbf{1}_{\{\tau_1\leqslant\tau_2\leqslant s\}}\ldots\mathbf{1}_{\{\tau_{n-1}\leqslant\tau_n\leqslant s\}}\mathbf{1}_{\{\tau_{n+1}\geqslant s\}}\middle|Y_{\tau_1}^{(t)}=y_1\right]&\\ \times\widehat{P}_{\tau_1}^{(t)}\left(xe^{a(\tau_1-r)},y_1\right)dy_1\Big|Y_r^{(t)}=x\Bigg]&.\end{aligned}$$

where for all $r \leqslant s \leqslant t$ and $x \in \mathcal{X}$, $J_{r,s}(x) = \left[\varepsilon x e^{a(s-r)}; (1-\varepsilon)xe^{a(s-r)}\right]$. Introducing the probability density of the first division time τ_1 yields

$$\begin{aligned}P_{r,s}^{(t)}(x,A) \geqslant \int_r^s g_r^{(t)}(x,t_1)\int_{J_{r,t_1}(x)}\mathbb{E}\left[\mathbf{1}_{\{t_1\leqslant\tau_2\leqslant s\}}\ldots\mathbf{1}_{\{\tau_{n-1}\leqslant\tau_n\leqslant s\}}\mathbf{1}_{\{\tau_{n+1}\geqslant s\}}\mathbf{1}_{\left\{Y_s^{(t)}\in A\right\}}\middle|Y_{t_1}^{(t)}=y_1\right]&\\ \times\widehat{P}_{t_1}^{(t)}\left(xe^{a(t_1-r)},y_1\right)dy_1&,\end{aligned}$$

where for all $r \leqslant s \leqslant t$ and $x \in \mathcal{X}$,

$$g_r^{(t)}(x,s) = \widehat{B}_s^{(t)}\left(xe^{a(s-r)}\right)\exp\left(-\int_r^s\widehat{B}_u^{(t)}\left(xe^{a(u-r)}\right)du\right).$$

Using the same argument iteratively, we get

$$\begin{aligned}P_{r,s}^{(t)}(x,A) \geqslant \int_{E_0}g_r^{(t)}(x,t_1)\int_{E_1}g_{t_1}^{(t)}(y_1,t_2)\ldots\int_{E_{n-1}}g_{t_{n-1}}^{(t)}(y_{n-1},t_n)e^{-\int_{t_n}^s\widehat{B}_u^{(t)}\left(y_ne^{a(u-t_n)}\right)du}&\\ \times\mathbf{1}_{\left\{y_ne^{a(s-t_n)}\in A\right\}}\prod_{i=0}^{n-1}\widehat{P}_{t_{i+1}}^{(t)}\left(y_ie^{a(t_{i+1}-t_i)},dy_{i+1}\right)dt_n\times\ldots\times dt_1&,\end{aligned}$$

where $y_0 = x$ and $t_0 = r$ and $E_i = [t_i, s] \times J_{t_i,t_{i+1}}(y_i)$, for $i = 0,\ldots,n-1$. Next, since $x \mapsto \widehat{B}_s^{(t)}(x)$ is increasing, we have:

$$\begin{aligned}\prod_{i=0}^{n}\exp\left(-\int_{t_i}^{t_{i+1}}\widehat{B}_u^{(t)}\left(y_ie^{a(u-t_i)}\right)du\right) &\geqslant \exp\left(-\int_r^s\widehat{B}_u^{(t)}\left(xe^{a(u-r)}\right)du\right)\\ &\geqslant e^{-2\varphi_2a^{-1}x_2(R)\left(e^{a(s-r)}-1\right)},\end{aligned}$$

where $t_{n+1} = s$. Noticing that:

$$\widehat{B}_s^{(t)}(x) \geqslant x\varphi_1,\ \widehat{P}_s^{(t)}(x,y) \geqslant \frac{2\varepsilon}{x(1-2\varepsilon)},$$

yields

$$P^{(t)}_{r,s}(x,A) \geqslant C_{r,s} \int_{\mathcal{E}_{n-2}} \left(\int_{t_{n-1}}^{s} \left(\int_{J_{t_{n-1},t_n}(y_{n-1})} \mathbf{1}_{\left\{y_n e^{a(s-t_n)} \in A\right\}} dy_n \right) dt_n \right) dy_{n-1} dt_{n-1} \dots dy_1 dt_1,$$

where $\mathcal{E}_{n-2} = E_0 \times \ldots \times E_{n-2}$ and

$$C_{r,s} = \exp\left(-2\varphi_2 \frac{x_2(R)}{a} \left(e^{a(s-r)} - 1 \right) \right) \left(\frac{2\varphi_1 \varepsilon}{1 - 2\varepsilon} \right)^n .$$

Applying the substitution $z = y_n e^{a(s-t_n)}$, we get:

$$P^{(t)}_{r,s}(x,A) \geqslant C_{r,s} I^{(n)}_{r,s}(x,A),$$

where:

$$I^{(n)}_{r,s}(x,A) = \frac{1}{a} \int_{\mathcal{E}_{n-2}} \left(1 - e^{-a(s-t_{n-1})} \right) \left(\int_{J_{t_{n-1},s}(y_{n-1})} \mathbf{1}_{\{z \in A\}} dz \right) dy_{n-1} dt_{n-1} \dots dy_1 dt_1.$$

Let $0 < \delta_1 < 1 < \delta_2$ be such that

$$\left(\frac{\delta_1}{\delta_2} \right)^{n-1} \geqslant \frac{\varepsilon}{1-\varepsilon}. \tag{2.22}$$

We prove the following proposition by induction for $n \geqslant 1$: there exists $C > 0$ depending on ε, n, δ_1, δ_2 and a such that

$$I^{(n)}_{r,s}(x,A) \geqslant C \left(1 - e^{-a(s-r)} \right)^n x^{n-1} \int_{\delta_2^{n-1} \varepsilon^n x e^{a(s-r)}}^{\delta_1^{n-1} (1-\varepsilon)^n x e^{a(s-r)}} \mathbf{1}_{z \in A} dz.$$

The verification for $n = 1$ is straightforward. We assume now that the proposition is satisfied for $n - 1$, for some $n \in \mathbb{N}$. Then, there exists $C > 0$ such that

$$\begin{aligned} I^{(n)}_{r,s}(x,A) &= \int_r^s \int_{J_{r,t_1}(x)} I^{(n-1)}_{t_1,s}(y_1, A) dy_1 dt_1 \\ &\geqslant C \int_r^s \int_{J_{r,t_1}(x)} y_1^{n-2} \left(1 - e^{-a(s-t_1)} \right)^{n-1} \int_{\delta_2^{n-2} \varepsilon^{n-1} y_1 e^{a(s-t_1)}}^{\delta_1^{n-2} (1-\varepsilon)^{n-1} y_1 e^{a(s-t_1)}} \mathbf{1}_{\{z \in A\}} dz dy_1 dt_1. \end{aligned}$$

Switching the integrals and using that $y_1 > \varepsilon x e^{a(t_1 - r)}$, we get:

$$I^{(n)}_{r,s}(x,A) \geqslant C \int_r^s \left(1 - e^{-a(s-t_1)} \right)^{n-1} x^{n-2} e^{a(n-2)(t_1 - r)} \left(I_1 + I_2 + I_3 \right) dt_1,$$

where

$$\begin{aligned} I_1 &= \int_{\delta_2^{n-2} \varepsilon^n x e^{a(s-r)}}^{\delta_2^{n-2} (1-\varepsilon) \varepsilon^{n-1} x e^{a(s-r)}} \mathbf{1}_{\{z \in A\}} dz \left(\frac{z}{\delta_2^{n-2} \varepsilon^{n-1}} e^{-a(s-t_1)} - \varepsilon x e^{a(t_1 - r)} \right), \\ I_2 &= \int_{\delta_2^{n-2} (1-\varepsilon) \varepsilon^{n-1} x e^{a(s-r)}}^{\delta_1^{n-2} (1-\varepsilon)^{n-1} \varepsilon x e^{a(s-r)}} \mathbf{1}_{\{z \in A\}} dz (1 - 2\varepsilon) x e^{a(t_1 - r)}, \\ I_3 &= \int_{\delta_1^{n-2} (1-\varepsilon)^{n-1} \varepsilon x e^{a(s-r)}}^{\delta_1^{n-2} (1-\varepsilon)^n x e^{a(s-r)}} \mathbf{1}_{\{z \in A\}} dz \left((1-\varepsilon) x e^{a(t_1 - r)} - \frac{z}{\delta_1^{n-2} (1-\varepsilon)^{n-1}} e^{-a(s-t_1)} \right). \end{aligned}$$

Next, using that $\delta_2\varepsilon \leqslant (1-\varepsilon)$ and $\delta_1(1-\varepsilon)\geqslant \varepsilon$, we obtain,

$$\begin{aligned}I_1 &\geqslant \int_{\delta_2^{n-1}\varepsilon^n x e^{a(s-r)}}^{\delta_2^{n-2}(1-\varepsilon)\varepsilon^{n-1}xe^{a(s-r)}} \mathbf{1}_{\{z\in A\}}dz(\delta_2-1)\varepsilon x e^{a(t_1-r)},\\ I_3 &\geqslant \int_{\delta_1^{n-2}(1-\varepsilon)^{n-1}\varepsilon x e^{a(s-r)}}^{\delta_1^{n-1}(1-\varepsilon)^n x e^{a(s-r)}} \mathbf{1}_{\{z\in A\}}dz(1-\delta_1)(1-\varepsilon)xe^{a(t_1-r)}.\end{aligned}$$

Therefore, gathering the three integrals, we get

$$I_{r,s}^{(n)}(x,A)\geqslant C\left(1-e^{-a(s-r)}\right)^n x^{n-1}\int_{\delta_2^{n-1}\varepsilon^n x e^{a(s-r)}}^{\delta_1^{n-1}(1-\varepsilon)^n x e^{a(s-r)}} \mathbf{1}_{\{z\in A\}}dz$$

where the constant C varies from line to line and the proposition holds at stage n. Finally, we have

$$P_{r,s}^{(t)}(x,A)\geqslant \alpha_{s-r}\nu_{r,s}(A),$$

where

$$\begin{aligned}\alpha_{s-r} =&C\left(1-e^{-a(s-r)}\right)^n x_1(R)^{n-1}e^{a(s-r)}\left(\delta_1^{n-1}(1-\varepsilon)^n x_1(R)-\delta_2^{n-1}\varepsilon^n x_2(R)\right)C_{r,s}\\ \nu_{r,s}(A)=&\frac{1}{e^{a(s-r)}\left(\delta_1^{n-1}(1-\varepsilon)^n x_1(R)-\delta_2^{n-1}\varepsilon^n x_2(R)\right)}\int_{\delta_2^{n-1}\varepsilon^n x e^{a(s-r)}}^{\delta_1^{n-1}(1-\varepsilon)^n x e^{a(s-r)}} \mathbf{1}_{\{z\in A\}}dz,\end{aligned}$$

and

$$\delta_1^{n-1}(1-\varepsilon)^n x_1(R)-\delta_2^{n-1}\varepsilon^n x_2(R)>0$$

according to (2.21) and (2.22). □

Next, we check that the first point of Assumption E is satisfied. We first prove that the moments of the auxiliary process are bounded. For all $p\in\mathbb{N}^*$, $0\leqslant s\leqslant t$ and $x\geqslant 0$, we denote by:

$$f_p^{(t)}(x,s)=\mathbb{E}_x\left[\left(Y_s^{(t)}\right)^p\right].$$

Lemma 2.4.3. *For all $p\in\mathbb{N}^*\bigcup\{-1\}$ and $x\geqslant 0$, we have:*

$$\sup_{t\geqslant 0}\sup_{s\leqslant t}\mathbb{E}_x\left[\left(Y_s^{(t)}\right)^p\right]<+\infty.$$

Remark 2.4.4. The moments that we need to control in order to check Assumption E depend on the function V. The shape of the Lyapunov function $V(x)=x+x^{-1}$ was convenient for the proof of the second point of Lemma 2.4.2. Indeed, the proof relies on the fact that $B(R,V)$ is lower bounded by a positive real number. This is the case because of the term x^{-1} in V. Because of this term, we need to control the first harmonic moment of the auxiliary process.

Proof. Let $p\in\mathbb{N}^*$ be a positive integer. We have using (2.17) and Dynkin's formula:

$$\begin{aligned}f_p^{(t)}(x,s)=&x^p+ap\int_0^s f_p^{(t)}(x,r)dr\\ &+2\int_0^s\mathbb{E}_x\left[\varphi(r)\int_{\varepsilon Y_r^{(t)}}^{(1-\varepsilon)Y_r^{(t)}}\left(y^p-\left(Y_r^{(t)}\right)^p\right)\frac{1+y\phi(r,t)}{1+Y_r^{(t)}\phi(r,t)}dy\right]\frac{dr}{1-2\varepsilon}.\end{aligned}$$

By differentiation with respect to s of the last equality we have:

$$\partial_s f_p^{(t)}(x,s) = apf_p^{(t)}(x,s) + 2\mathbb{E}_x\left[\varphi(s)\int_{\varepsilon Y_s^{(t)}}^{(1-\varepsilon)Y_s^{(t)}}\left(y^p - \left(Y_s^{(t)}\right)^p\right)\frac{1+y\phi(s,t)}{1+Y_s^{(t)}\phi(s,t)}\frac{dy}{1-2\varepsilon}\right].$$

Next, we notice that for $\varepsilon x \leqslant y \leqslant (1-\varepsilon)x$, we have:

$$\frac{m(y,s,t)}{m(x,s,t)} \geqslant \frac{1+\varepsilon x\phi(s,t)}{1+x\phi(s,t)} \geqslant \varepsilon.$$

Then

$$\begin{aligned}\partial_s f_p^{(t)}(x,s) &\leqslant apf_p^{(t)}(x,s) + 2\varepsilon\mathbb{E}_x\left[\varphi(s)\int_{\varepsilon Y_s^{(t)}}^{(1-\varepsilon)Y_s^{(t)}}\left(y^p - \left(Y_s^{(t)}\right)^p\right)\frac{dy}{1-2\varepsilon}\right]\\ &\leqslant apf_p^{(t)}(x,s) - C(\varepsilon)f_{p+1}^{(t)}(x,s)\end{aligned}$$

where $C(\varepsilon) := \frac{2\varepsilon}{1-2\varepsilon}\varphi_1\left(1-2\varepsilon-\frac{(1-\varepsilon)^{p+1}-\varepsilon^{p+1}}{p+1}\right)$. Moreover, $C(\varepsilon) > 0$ because $\varepsilon < \frac{1}{2}$. Applying Jensen inequality, we have $f_{p+1}^{(t)}(s) \geqslant f_p^{(t)}(s)^{1+1/p}$. Finally, we obtain the following differential inequality:

$$\partial_s f_p^{(t)}(x,s) \leqslant F\left(f_p^{(t)}(x,s)\right),$$

where $F(x) = apx - C(\varepsilon)x^{1+1/p}$ for all $x \geqslant 0$. We notice that there exists $x_0 > 0$ such that $F > 0$ on $(0, x_0)$ and $F < 0$ on $(x_0, +\infty)$. Then, any solution to the equation $y' = F(y)$ is bounded by $y(0) \vee x_0$ and so is $f_p^{(t)}(x,\cdot)$.

Next, we prove that the first harmonic moment of the auxiliary process is bounded. Let us recall that $V_2(x) = 1/x$. Let $x \in \mathcal{X}$ and $0 \leqslant s \leqslant t$. According to Kolmogorov's forward equation, we have:

$$\partial_s P_{0,s}^{(t)}V_2(x) = P_{0,s}^{(t)}\mathcal{A}_s^{(t)}V_2(x).$$

Using (2.19) we get:

$$\partial_s P_{0,s}^{(t)}V_2(x) \leqslant -aP_{0,s}^{(t)}V_2(x) + 2\varphi_2 C(\varepsilon).$$

Finally, using Grönwall's inequality, we obtain:

$$P_{0,s}^{(t)}V_2(x) \leqslant \left(\frac{1}{x} - \frac{2\varphi_2 C(\varepsilon)}{a}\right)e^{-as} + \frac{2\varphi_2 C(\varepsilon)}{a}.$$

□

Lemma 2.4.5. *For all $x \geqslant 0$, we have:*

$$\sup_{t\geqslant 0}\mathbb{E}_x\left[B\left(t, Y_t^{(t)}\right)J\left(V\varphi_t(x,\cdot)\right)\left(Y_t^{(t)}\right)\right] < +\infty.$$

Proof. First, we have for all $x \in \mathbb{R}_+$ and all $s, t \in \mathbb{R}_+$ with $s \leqslant t$:

$$1 + \frac{x}{a}\varphi_1(e^{a(t-s)}-1) \leqslant m(x,s,t) \leqslant 1 + \frac{x}{a}\varphi_2(e^{a(t-s)}-1).$$

Then, for all $x, y \in \mathcal{X}$, we obtain:

$$\begin{aligned}\varphi_t(x,y) = \sup_{r \geqslant t} \frac{m(x,0,t)m(y,t,r)}{m(x,0,r)} &\leqslant \sup_{r\geqslant t} \frac{\left(1+\frac{x}{a}\varphi_2 e^{at}\right)\left(1+\frac{y}{a}\varphi_2 e^{a(r-t)}\right)}{1+\frac{x}{a}\varphi_1\left(e^{ar}-1\right)} \\ &\leqslant \frac{\left(1+\frac{x}{a}\varphi_2\right)\left(1+\frac{y}{a}\varphi_2\right)}{\frac{x}{a}\varphi_1 \wedge 1}.\end{aligned}$$

Next, for all $\theta \in (0,1)$, we have:

$$\varphi_t\left(x,\theta y\right)\varphi_t\left(x,(1-\theta)y\right) \leqslant \left(\varphi_t(x,y)\right)^2 \leqslant A_1(x)A_2(y)$$

where

$$A_1(x) = \left(\frac{x}{a}\varphi_1 \wedge 1\right)^{-2}\left(1+\frac{x}{a}\varphi_2\right)^2, \ A_2(y) = \left(1+\frac{y}{a}\varphi_2\right)^2.$$

Moreover, for $\theta \in [\varepsilon, 1-\varepsilon]$ and for all $x \in \mathcal{X}$, $V(\theta x)V((1-\theta)x) \leqslant (\varepsilon x)^{-2} + x^2 + 2\varepsilon^{-1}$. Then:

$$\begin{aligned} J\left(V\varphi_t(x,\cdot)\right)(y) &\leqslant 2\int_{\varepsilon}^{1-\varepsilon} V\left(\theta y\right)V\left((1-\theta)y\right)\varphi_t\left(x,\theta y\right)\varphi_t\left(x,(1-\theta)y\right)\frac{d\theta}{1-2\varepsilon} \\ &\leqslant \left((\varepsilon y)^{-2}+y^2+2\varepsilon^{-1}\right)A_1(x)A_2(y) \leqslant A_1(x)\sum_{k=0}^{6} C_k(\varepsilon)y^{k-2},\end{aligned}$$

where for all $k = 0\ldots 6$ $C_k(\varepsilon)$ are constants depending on $x, a, \varepsilon, \varphi_2$. Then, we get:

$$\mathbb{E}_x\left[B\left(t,Y_t^{(t)}\right)J\left(V\varphi_t(x,\cdot)\right)\left(Y_t^{(t)}\right)\right] \leqslant \varphi_2 A_1(x)\sum_{k=0}^{6} C_k(\varepsilon)\mathbb{E}_x\left[\left(Y_t^{(t)}\right)^{k-1}\right] < \infty,$$

according to Lemma 2.4.3. □

Last, we verify that Assumption F is satisfied.

Lemma 2.4.6. *For all $t \geqslant 0$, $x \in \mathcal{X}$, we have:*

$$\mathbb{E}_{\delta_{x_0}}\left[\left(\frac{N_t}{m(x_0,0,t)}\right)^2\right] \leqslant \frac{a^2+\varphi_2 x\left(a+2\varphi_2 x\right)+\varphi_2^2x^2}{\left(\min(a,\varphi_1 x)\right)^2}.$$

Proof. According to Itô formula, we have for all $x \in \mathcal{X}$ and $t \geqslant 0$:

$$\mathbb{E}_{\delta_x}\left[N_t^2\right] = 1 + x\int_0^t \varphi(s)e^{as}\left(2\mathbb{E}_{\delta_x}\left[N_s\right]+1\right)ds.$$

After some calculations, we obtain:

$$\mathbb{E}_{\delta_x}\left[N_t^2\right] \leqslant \frac{e^{2at}}{a^2}\left(a^2+\varphi_2 x\left(a+2\varphi_2 x\right)+\varphi_2^2 x^2\right)$$

Moreover, we have:

$$m(x,0,t)^2 \geqslant e^{2at}\left(e^{-at}+\frac{x}{a}\varphi_1(1-e^{-at})\right)^2 \geqslant e^{2at}\left(\min\left(1,\frac{x}{a}\varphi_1\right)\right)^2,$$

and the result follows. □

Chapter 3

Statistical estimation in a randomly structured branching population

Abstract

We consider a binary branching process structured by a stochastic trait that evolves according to a diffusion process that triggers the branching events, in the spirit of Kimmel's model of cell division with parasite infection. Based on the observation of the trait at birth of the first n generations of the process, we construct nonparametric estimator of the transition of the associated bifurcating chain and study the parametric estimation of the branching rate. In the limit $n \to \infty$, we obtain asymptotic efficiency in the parametric case and minimax optimality in the nonparametric case.

Content

3.1 Introduction

3.1.1 Motivation

The statistics of branching Markov processes has been in constant progress over the last decade, in particular thanks to a thorough study of structured populations, with a strong input from evolutionary or cell division modelling in mathematical biology (see for instance the textbooks [M1́6, Per07] and the references therein). Several models have been considered, with data processed either in discrete or continuous time, and an accompanying statistical inference program has progressively emerged, in which one addresses the problem of the identification and uncertainty quantification of critical parameters like branching rates, modeled as functions of biological traits like age, size and so on.

In many cases, this approach is linked to certain piecewise deterministic Markov models or bifurcating Markov chains (BMC) in discrete time. These models are well understood from a probabilist point of view (in discrete time Guyon [Guy07], Bitseki-Penda et al. [BPDG14, BPEBG17], in continuous time Bansaye and Méléard [BM15], Bansaye et al. [BDMT11] or more recently Marguet [Mar16] for a general approach). Moreover, a fairly complete picture of their statistical structure has been established over the last few years in the group of Doumic and collaborators [BPHO17, DHRBR12, DHKR15, HO16], see also Bitseki-Penda and Olivier [BPO17], de Saporta et al. [dSGPM12, dSGPM14], Azaïs et al. [ADGP14] or recently Bitseki-Penda and Roche [BPR17]. In these models, the traits of a population between branching events like cell division evolve through time according to a dynamical system. The next logical step is to replace this deterministic evolution by a random flow, that allows one to account for traits like that may have their own random evolution according to exogeneous input. A paradigmatic example is given by Kimmel's model (see Kimmel [Kim97] and Bansaye [Ban08]) where the trait is given by a density of parasites within a cell that evolve according to a diffusion process. The statistical analysis of such models is the topic of the present chapter.

We consider a population model with binary division triggered by a trait $x \in \mathcal{X}$ where $\mathcal{X} \subseteq \mathbb{R}$ is an open (possibly unbounded) interval. The trait $\phi_x(t)$ of each individual evolves according to

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \;\; \phi_x(0) = x \tag{3.1}$$

where $r, \sigma : \mathcal{X} \to \mathcal{X}$ are regular functions and $(W_t)_{t \geqslant 0}$ is a standard Brownian motion. Each individual with trait x at time t dies according to a division rate $x \mapsto B(x)$. (An individual with trait $\phi_x(t)$ at time t dies with probability $B(\phi_x(t))dt$ during the interval $[t, t+dt]$). At division, a particle is replaced by two new individuals with trait at birth given respectively by θx and $(1-\theta)x$ where θ is drawn according to $\kappa(x)dx$ for some symmetric probability density function $\kappa(x)$ on $[0,1]$.

The model is described by the traits of the population, formally given as a Markov process

$$X = (X_1(t), X_2(t), \ldots), \;\; t \geqslant 0 \tag{3.2}$$

with values in $\bigcup_{k \geqslant 1} \mathcal{X}^k$, where the $X_i(t)$ denote the (ordered) traits of the living particles at time t. Its distribution is entirely determined by an initial condition at $t = 0$ and by the parameters (r, σ, B, κ).

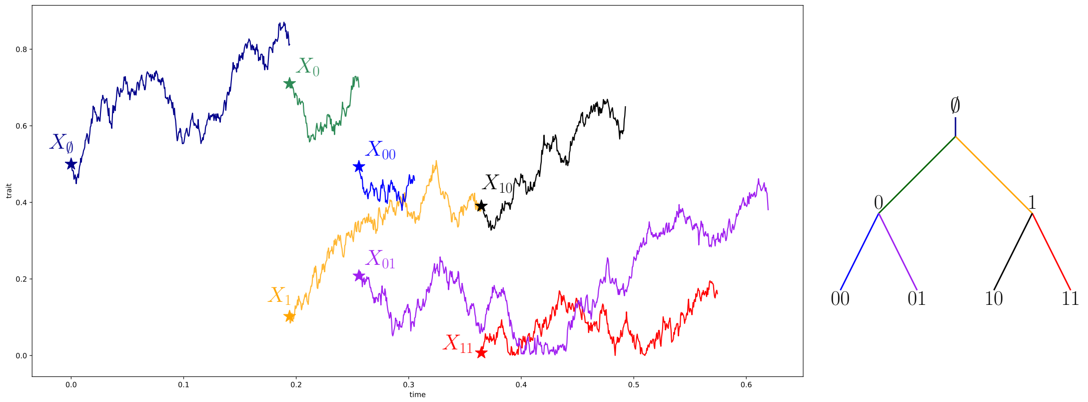

Figure 3.1.1: Example of a trajectory and its associated genealogy.

3.1.2 Statistical setting by reduction to a bifurcating Markov chain model

We assume we have data at branching events (i.e. at cell division) and we wish to make inference on the parameters of the model. Using the Ulam-Harris-Neveu notation, for $m \geqslant 0$, let $\mathbb{G}_m = \{0,1\}^m$ (with $\mathbb{G}_0 = \{\varnothing\}$) and introduce the infinite genealogical tree

$$\mathbb{T} = \bigcup_{m \in \mathbb{N}} \mathbb{G}_m.$$

For $u \in \mathbb{G}_m$, set $|u| = m$ and define the concatenation $u0 = (u,0) \in \mathbb{G}_{m+1}$ and $u1 = (u,1) \in \mathbb{G}_{m+1}$. For $n \geqslant 0$, let $\mathbb{T}_n = \bigcup_{m=0}^{n} \mathbb{G}_m$ denote the genealogical tree up to the n-th generation. We denote by X_u the trait at birth of an individual $u \in \mathbb{T}$. From the branching events, we assume that we observe

$$\mathbb{X}^n = (X_u)_{u \in \mathbb{T}_n},$$

so that we have $2^{n+1} - 1$ random variables with value in $\mathcal{X}$. Asymptotics are taken as n grows to infinity. An example of trajectory is represented on Figure 3.1.1 with the associated genealogy.

There are several objects of interest that we may try to infer from the data $\mathbb{X}^n$. First, one may notice that the Markov structure of X in (3.2) turns $(X_u, u \in \mathbb{T})$ into a *bifurcating Markov chain* according to the terminology introduced by Delmas and Guyon [Guy07]. A bifurcating Markov chain is specified by **1)** a measurable state space, here $\mathcal{X}$ (endowed with its Borel sigma-field) with a Markov kernel $\mathcal{P}$ from $\mathcal{X}$ to $\mathcal{X} \times \mathcal{X}$ and **2)** a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \geqslant 0}, \mathbb{P}\big)$. Following Guyon, [Guy07], Definition 2, we have the

Definition 3.1.1. *A bifurcating Markov chain (BMC) is a family* $(X_u)_{u \in \mathbb{T}}$ *of random variables with value in* $\mathcal{X}$ *such that* X_u *is* $\mathcal{F}_{|u|}$*-measurable for every* $u \in \mathbb{T}$ *and*

$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_m} \psi_u(X_u, X_{u0}, X_{u1}) \,\Big|\, \mathcal{F}_m\Big] = \prod_{u \in \mathbb{G}_m} \mathcal{P}\psi_u(X_u) \tag{3.3}$$

for every $m \geqslant 0$ *and any family of (bounded) measurable functions* $(\psi_u)_{u \in \mathbb{G}_m}$*, where* $\mathcal{P}\psi(x) = \int_{\mathcal{X} \times \mathcal{X}} \psi(x, y_1, y_2)\mathcal{P}(x, dy_1 dy_2)$ *denotes the action of* $\mathcal{P}$ *on* ψ*.*

The distribution of $(X_u)_{u \in \mathbb{T}}$ is thus entirely determined by $\mathcal{P}$ and an initial distribution for $X_\varnothing$. A key role for understanding the asymptotic behavior of the bifurcating Markov chain is the so-called *tagged-branch* chain, that consists in picking a lineage at random in the population $(X_u)_{u\in\mathbb{T}}$: it is a Markov chain with value in $\mathcal{X}$ defined by $Y_0 = X_\varnothing$ and for $m \geqslant 1$:

$$Y_m = X_{\varnothing\epsilon_1\ldots\epsilon_m},$$

where $(\epsilon_m)_{m\geqslant 1}$ is a sequence of independent Bernoulli random variables with parameter 1/2, independent of $(X_u)_{u\in\mathbb{T}}$, with transition

$$\mathcal{Q} = (\mathcal{P}_1 + \mathcal{P}_2)/2 \tag{3.4}$$

obtained from the marginal transitions of $\mathcal{P}$:

$$\mathcal{P}_1(x, dy) = \int_{y_2\in\mathcal{X}} \mathcal{P}(x, dy\,dy_2) \ \text{ and } \ \mathcal{P}_2(x, dy) = \int_{y_1\in\mathcal{X}} \mathcal{P}(x, dy_1 dy).$$

Guyon proves in [Guy07] that if $(Y_m)_{m\geqslant 0}$ is ergodic with invariant measure $\nu(dx)$ on $\mathcal{X}$, then a convergence of the type

$$\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} \psi(X_u, X_{u0}, X_{u1}) \to \int_{\mathcal{X}} \mathcal{P}\psi(x)\nu(dx) \tag{3.5}$$

holds as $n \to \infty$ for appropriate test functions g, almost surely and appended with appropriate central limit theorems (Theorem 19 in [Guy07]).

3.1.3 Main results

In this context, there are several quantities that can be inferred from the data $\mathbb{X}_n$ as n grows and that are important in order to understand the dynamics of $(X_u)_{u\in\mathbb{T}}$. Under suitable assumptions on the stochastic flow (3.1), the transition $\mathcal{Q}$ admits an invariant measure ν and we have fast convergence of the tagged-chain $(Y_m)_{m\geqslant 1}$ to equilibrium. This enables us to construct in a first part nonparametric estimators of ν and $\mathcal{Q}$ with an optimal rate of convergence. Notice that since the fragmentation kernel $\kappa(x)$ that distributes the trait of the parent between its two children is symmetric, we have $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{Q}$. This reveals the structure of the underlying BMC.

However, estimators of ν and $\mathcal{Q}$ do not give us any insight about the parameters (r, σ, B, κ) of the model. In a second part, we investigate the inference of the division rate $x \mapsto B(x)$ as a function of the trait $x \in \mathcal{X}$ when the other parameters r, σ and $\kappa(x)$ are known. This seemingly stringent assumption is necessary given the observation scheme $\mathbb{X}^n$. If extraneous data were available, estimators of the parameters r, σ and κ could be obtained in a relatively straightforward manner:

i) As soon as a discretisation of the values of the flow are available, standard techniques about inference in ergodic diffusions can be applied to recover $x \mapsto r(x)$ and $x \mapsto \sigma(x)^2$, see for instance [Hof99, Kut04].

ii) The fact that an individual u distributed its traits to its offspring in a conservative way enables one to recover the fraction θ_u distributed among the children. Indeed the individual u born at b_u with lifespan at $d_u - b_u$ has trait $\phi_{X_u}(d_u - b_u)$ at its time of death. It follows that its children have trait at birth given by

$$X_{u0} = \theta_u \phi_{X_u}(d_u - b_u), \ \ X_{u1} = (1-\theta_u)\phi_{X_u}(d_u - b_u),$$

where the θ_u are drawn independently from the distribution $\kappa(x)dx$ and therefore, the relationship $\frac{X_{u0}}{X_{u_1}} = \frac{\theta_u}{1-\theta_u}$ identifies θ_u. In turn, the estimation of $x \mapsto \kappa(x)$ reduces to a standard density estimation problem from data $(\theta_u)_{u\in\mathbb{T}_n}$, see for instance [Hoa15].

The identification and estimation of the branching rate $x \mapsto B(x)$ from data $\mathbb{X}^n$ is more delicate and is the topic of the second part of this chapter. Under minimal regularity assumptions developed in Section 3.2 below, it is not difficult to obtain an explicit representation of the transition $\mathcal{Q}(x,dy) = \mathcal{Q}_B(x,dy) = q_B(x,y)dy$ that reads

$$q_B(x,y) = \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]dz, \tag{3.6}$$

where $L_t^y(\phi_x)$ denotes the local time at t in y of the semimartingale $(\phi_x(t))_{t\geqslant 0}$. Assuming (r,σ,κ) known (or identified by extraneous observation schemes) we study the estimation of $x \mapsto B(x)$ when B belongs to a parametric class of functions $\{B_\vartheta, \vartheta \in \Theta\}$ for some regular subset of the Euclidean space $\mathbb{R}^d$. Under a certain ordering property (Definition 3.3.5 in Section 3.3.2 below) that ensures the identifiability of the model and suitable standard regularity properties, we realize a standard maximum likelihood estimation of B thanks to (3.6) by maximising the contrast

$$\vartheta \mapsto \prod_{u\in\mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u), \;\; \vartheta \in \Theta,$$

(with $\mathbb{T}_n^\star = \mathbb{T}_n \backslash \mathbb{G}_0$) and we prove that it achieves asymptotic efficiency and discuss its practical implementation. It is noteworthy that for the parametric estimation of B, there is no straightforward contrast minimization procedure (at least we could not find any) whereas $q_B(x,y)$ is explicit. The fairly intricate dependence of B in the representation (3.6) makes however the whole scheme relatively delicate, both mathematically and numerically.

Clearly, other observation schemes are relevant in the context of cell division modelling. For instance, one could consider a (large) time $T > 0$ and observe the branching process X_t defined in (3.2) for every $t \in [0,T]$. This entails the possibility to extract the times (T_u) at which branching events occur, like e.g. in [HO16]. However, the continuous time setting is drastically different and introduce the additional difficulty of bias sampling, an issue we avoid in the present context. Alternatively, one could consider the augmented statistical experiment where one observes $(X_u, T_u)_{u\in\mathbb{T}_n}$, but the underlying mathematical structure is presumably not simpler. Our results show in particular that for the parametric estimation of the branching rate B, although the times at which branching event occur are statistically informative, their observation is not necessary to obtain optimal rates of convergence as soon as (r,σ,κ) are known.

3.1.4 Organization of the chapter

Section 3.2.1 is devoted to the construction of the stochastic model, our assumptions and the accompanying statistical experiments. In particular, we have a nice structure enough so that explicit representations of $\mathcal{P}$ and $\mathcal{Q}$ are available (Proposition 3.2.1). We give a first result on the geometric ergodicity of the model via an explicit Lyapunov function in Theorem 3.2.2 and derive in Theorem 3.2.3 a rate of convergence for the variance of empirical measures of the data $\mathbb{X}^n = (X_u)_{u\in\mathbb{T}_n}$ against test functions $\varphi(X_u)$ or $\psi(X_{u^-}, X_u)$ with a sharp control in terms of adequate norms for φ, ψ that do not follow from the standard application of the geometric ergodicity of Theorem 3.2.2. This is crucial for the subsequent applications to the nonparametric estimation of $\mathcal{Q}$ and its invariant measure ν that are given in Theorem 3.3.3 of Section 3.3.1. Section 3.3.2 is devoted to the parametric estimation of the branching rate, where an asymptotically efficient result is proved for a maximum likelihood estimator in Theorem 3.3.9. It is based on a relatively sharp study of the transition $\mathcal{Q}$, thanks to local time properties of the stochastic flow that triggers the branching events.

3.2 A cell division model structured by a stochastic flow

The proofs are postponed to Section 3.4 and an Appendix Section 3.5 contains useful auxiliary results.

3.2.1 Assumptions and well-posedness of the stochastic model

Dynamics of the traits

Remember that $\mathcal{X} \subseteq \mathbb{R}$ is an open, possibly unbounded interval. The flow is specified by $r, \sigma : \mathcal{X} \to \mathcal{X}$ which are measurable and that satisfy the following assumption:

Assumption 1. *For some* $r_1, \sigma_1, \sigma_2 > 0$, *we have* $|r(x)| \leqslant r_1(1 + |x|)$ *and* $\sigma_1 \leqslant \sigma(x) \leqslant \sigma_2$, *for every* $x \in \mathcal{X}$. *Moreover, for some* $r_2 > 0$, *we have* $\operatorname{sgn}(x)r(x) < 0$ *for* $|x| \geqslant r_2$ *(with* $\operatorname{sgn}(x) = \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x\leqslant 0\}}$*).*

Under Assumption 1, there is a unique strong solution to (3.1) (for instance [Øks03], Theorem 5.2.1.). We denote by $(\Phi_x(t), t \geqslant 0)$ the unique solution to (3.1) with initial condition $x \in \mathcal{X}$. In particular, $(\Phi_x(t), t \geqslant 0)$ is a strong Markov process and is ergodic (cf. [Kut04], Theorem 1.16.). Note that when $\mathcal{X}$ is bounded, the drift condition $\operatorname{sgn}(x)r(x) < 0$ for large enough x can be dropped.

Division events.

An individual with trait x dies at an instantaneous rate $x \mapsto B(x)$, where $B : \mathcal{X} \to [0, \infty)$ satisfies the following condition:

Assumption 2. *The function* $x \mapsto B(x)$ *is continuous. Moreover, for some* $b_1, b_2 > 0$ *and* $\gamma \geqslant 0$, *we have* $b_1 \leqslant B(x) \leqslant b_2 |x|^\gamma + b_1$ *for every* $x \in \mathcal{X}$.

Under Assumptions 1 and 2, the process X in (3.2) is well defined and the size of the population does note explode in finite time almost-surely, see for instance Marguet [Mar16]. Note that the lower bounds for σ and B are not needed for the well-posedness of X but rather for later statistical purposes.

Fragmentation of the trait at division

Finally, we make an additional set of assumptions on the fragmentation distribution $\kappa(z)dz$ that ensures in particular the non-degeneracy of the process.

Assumption 3. *We have*

$\kappa(z) = \kappa(1 - z)$ *for almost every* $z \in [0, 1]$,

$\operatorname{supp}(\kappa) \subset [\varepsilon, 1 - \varepsilon]$ *for some* $0 < \varepsilon < 1/2$,

$\inf_{z\in[\varepsilon,1-\varepsilon]} \kappa(z) \geqslant \delta$.

The symmetry condition of Assumption 3 on the fragmentation distribution $\kappa(z)dz$ ensures that the law of the trait at birth is identical for both children. The other parts of Assumption 3 are slightly technical and may presumably be relaxed.

Representations of $\mathcal{P}$ and $\mathcal{Q}$

Under Assumptions 1, 2 and 3, we obtain closed-form formulas for the transition $\mathcal{P}$ defined via (3.3) and the mean or marginal transition $\mathcal{Q}$ of the BMC $(X_u)_{u\in\mathbb{T}}$, see (3.4) that also gives the transition probability of the discrete Markov chain with value in $\mathcal{X}$ corresponding to the trait at birth along an ancestral lineage. These representations are crucial for the subsequent analysis of the variance of the estimators of $\mathcal{P}$ and of the invariant measure ν.

Proposition 3.2.1. *Work under Assumptions 1, 2 and 3. For every $x, y, y_1, y_2 \in \mathcal{X}$, we have*

$$\mathcal{P}(x, dy_1dy_2) = p(x, y_1, y_2)dy_1dy_2 \text{ and } \mathcal{Q}(x, dy) = q(x, y)dy,$$

with

$$p(x, y_1, y_2) = \frac{\kappa\big(y_1/(y_1+y_2)\big)}{y_1+y_2} B\,(y_1+y_2)\,\sigma(y_1+y_2)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^{y_1+y_2}(\phi_x)\Big] \tag{3.7}$$

and

$$q(x, y) = \int_0^1 \frac{\kappa(z)}{z}B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^{y/z}(\phi_x)\Big]dz, \tag{3.8}$$

where $L_t^y(\phi_x)$ denotes the local time at t in y of the semimartingale $(\phi_x(t))_{t\geqslant 0}$.

Note that by Assumption 3, we have $p(x, y_1, y_2) = p(x, y_2, y_1)$ for every $x, y_1, y_2 \in \mathcal{X}$ and also $\int_{\mathcal{X}} p(x, y_1, y)dy_1 = \int_{\mathcal{X}} p(x, y, y_2)dy_2 = q(x, y)$ for every $x, y \in \mathcal{X}$.

3.2.2 Convergence of empirical measures

Convergence of $\mathcal{Q}$ to equilibrium

Assumptions 1,2 and 3 imply a drift condition for the Lyapunov function $V(x) = x^2$ on $\mathcal{X}$ and a minorization condition over a small set so that in turn $\mathcal{Q}$ is geometrically ergodic.

Let $\mathbb{Q} = \mathbb{Q}(r_i, b_i, \sigma_i, z_i, \gamma, \varepsilon, \delta, i = 1, 2)$ be the class of all transitions $\mathcal{Q} = \mathcal{Q}(r, \sigma, B, \kappa)$ defined over $\mathcal{X}$ that satisfy Assumptions 1,2 and 3 with appropriate constants. An invariant probability measure for $\mathcal{Q}$ is a probability ν on $\mathcal{X}$ such that $\nu\mathcal{Q} = \nu$, where $\nu\mathcal{Q}(dy) = \int_{x\in\mathcal{X}} \nu(dx)\mathcal{Q}(x, dy)$. Define

$$\mathcal{Q}^r(x, dy) = \int Q(x, dz)Q^{r-1}(z, dy) \text{ with } \mathcal{Q}^0(x, dy) = \delta_x(dy)$$

for the r-th iteration of $\mathcal{Q}$. For $\varphi : \mathcal{X} \to \mathbb{R}$, we set

$$|\varphi|_V = \sup_{x\in\mathcal{X}} \frac{|\varphi(x)|}{1+V(x)}$$

and write $\nu(\varphi) = \int_{\mathcal{X}} \varphi(x)\nu(dx)$ when no confusion is possible.

Theorem 3.2.2 (Convergence to equilibrium)**.** *Work under Assumptions 1, 2 and 3. Then any $\mathcal{Q} \in \mathbb{Q}$ admits an invariant probability distribution ν. Moreover, for $V(x) = x^2$, there exist $C = C(\mathcal{Q}) > 0$ and $\rho = \rho(\mathcal{Q}) \in (0, 1)$ such that for every $m \geqslant 1$, the bound*

$$\big|\mathcal{Q}^m\varphi - \nu(\varphi)\big|_V \leqslant C\rho^m\big|\varphi - \nu(\varphi)\big|_V$$

holds as soon as $|\varphi|_V < \infty$, $\sup_{\mathcal{Q}\in\mathbb{Q}} C(\mathcal{Q}) < \infty$ and $\sup_{\mathcal{Q}\in\mathbb{Q}} \rho(\mathcal{Q}) < 1$.

In particular, we have $|\mathcal{Q}^m\varphi(x) - \nu(\varphi)| \leqslant C\rho^m(1 + V(x))|\varphi - \nu(\varphi)|_\infty$ for every $x \in \mathcal{X}$.

Control of empirical variances

Theorem 3.2.2 is the key ingredient in order to control the rate of convergence of empirical means of the form

$$\mathcal{M}_n(\psi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \psi(X_{u^-}, X_u) \tag{3.9}$$

that converge to $\nu\mathcal{Q}(\psi)$ for any reasonable test function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. (If $\varphi : \mathcal{X} \to \mathbb{R}$ we set $\mathcal{M}_n(\varphi) = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \varphi(X_u) \to \nu(\varphi)$ and we have a formal correspondence between the two expressions by writing $\psi(x, y) = \varphi(y)$ as a function of the second variable.) In order to derive nonparametric estimators of ν and $\mathcal{Q}$ by means of kernel functions ψ that shall depend on n, we need sharp estimates in terms of ψ.

We shall restrict our study to transitions $\mathcal{Q} \in \mathbb{Q}$ for which the geometric rate of convergence to equilibrium $\rho = \rho(\mathcal{Q})$ given in Theorem 3.2.2 satisfies $\rho(\mathcal{Q}) \leqslant 1/2$. Let $\mathbb{Q}_{1/2} \subset \mathbb{Q}$ denote the set of such transitions.

We need some notation. We denote by $|\cdot|_1$ the usual L^1-norm w.r.t. the Lebesgue measure on $\mathcal{X} \times \mathcal{X}$. For a function $\psi = \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ we set $\psi^\star(x) = \sup_{y \in \mathcal{X}} |\psi(x, y)|$ and $\psi_\star(y) = \sup_{x \in \mathcal{X}} |\psi(x, y)|$ and define

$$|\psi|_{\wedge 1} = \int_{\mathcal{X} \times \mathcal{X}} |\psi(x, y)| dx dy \bigwedge \int_{\mathcal{X}} \psi_\star(y) dy.$$

Note in particular that when $\psi(x, y) = \varphi(y)$ is a function of y only, we may have that $|\psi_\star|_1 = \int_{\mathcal{X}} |\varphi(y)| dy$ is finite while ψ is not integrable on $\mathcal{X} \times \mathcal{X}$ as a function of two variables. For a positive measure ρ on $\mathcal{X}$, let also

$$|\psi|_\rho = \int_{\mathcal{X} \times \mathcal{X}} |\psi(x, y)| \rho(dx) dy + |\psi|_{\wedge 1}.$$

We write $\mathbb{P}_\mu$ for the law of $(X_u)_{u \in \mathbb{T}}$ with initial distribution μ for $X_\varnothing$. Remember that $V(x) = x^2$ from Theorem 3.2.2.

Theorem 3.2.3. *Work under Assumptions 1, 2 and 3. Let μ be a probability measure on $\mathcal{X}$ such that $\mu(V^2) < \infty$. Let $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a bounded function such that $\psi_\star$ is compactly supported. The following estimate holds true:*

$$\mathbb{E}_\mu\big[\big(\mathcal{M}_n(\psi) - \nu(\mathcal{Q}\psi)\big)^2\big] \lesssim |\mathbb{T}_n|^{-1}\big(|\psi^2|_{\mu+\nu} + |\psi^\star \psi|_\mu + \big(1 + \mu(V^2)^{1/2}\big)|\psi_\star|_1 |\psi|_\nu\big),$$

where the symbol $\lesssim$ means up to an explicitly computable constant that depends on $\mathcal{Q}$ and on supp$(\psi_\star)$ only. Moreover, the estimate is uniform in $\mathcal{Q} \in \mathbb{Q}_{1/2}$.

Several remarks are in order: **1)** We have a sharp order in terms of the test functions ψ, that behave no worse than $\int_{\mathcal{X}^2} \psi^2$ under minimal regularity on ν which is satisfied, see Lemma 3.4.5 below (and of course μ, although this restriction could easily be relaxed). This behaviour is the one expected for instance in the IID case and is crucial for the subsequent statistical application of Theorem 3.3.3 where the functions ψ will be kernel depending on n. **2)** Theorem 3.2.3 can be derived with other more competitive tools in the uniform geometric ergodic case, see e.g. [BPHO17, BPR17], a case we do not recover here since $\mathcal{X}$ is not necessarily bounded. **3)** Theorem 3.2.3 has an analog in [DHKR15] for piecewise deterministic growth-fragmentation models, but our proof is somewhat simpler here and sharper (we do not pay the superfluous logarithmic term in [DHKR15]).

3.3 Statistical estimation

3.3.1 Nonparametric estimation of $\mathcal{Q}$ and ν

Under Assumptions 1, 2 and 3, any $\mathcal{Q}(x, dy) = q(x, y)dy$ admits an invariant probability measure $\nu(dx) = \nu(x)dx$, the regularity of $\nu(x)$ being inherited from that of $\mathcal{Q}$ via $\nu(x) = \int_{\mathcal{X}} q(z, x)\nu(dz)$.

Fix $(x_0, y_0) \in \mathcal{X} \times \mathcal{X}$. We are interested in constructing estimators of $q(x_0, y_0)$ and $\nu(x_0)$ from the observation $\mathbb{X}^n$ when both functions satisfy some Hölder regularity properties in the vicinity of (x_0, y_0). To that end, we need approximating kernels.

Definition 3.3.1. *A function $G : \mathcal{X} \to \mathbb{R}$ is a kernel of order k if it is compactly supported and satisfies $\int_{\mathcal{X}} x^\ell G(x)dx = \mathbf{1}_{\{\ell=0\}}$ for $\ell = 0, \ldots, k$.*

The construction and numerical tractability of approximating kernels is documented in numerous textbooks, see for instance Tsybakov [Tsy04]. For bandwidth parameters $h, h_1, h_2 > 0$, we set

$$G_h(y) = h^{-1}G(h^{-1}y)$$

and

$$G^{\otimes 2}_{h_1,h_2}(x, y) = h_1^{-1}h_2^{-1}G(h_1^{-1}x)G(h_1^{-1}y)$$

and obtain approximations of $\nu(y_0)$ and $q(x_0, y_0)$ by setting

$$G_h \star \nu(y_0) = \int_{\mathcal{X}} G_h(y - y_0)\nu(y)dy$$

and

$$G_h \star q(x_0, y_0) = \int_{\mathcal{X}\times\mathcal{X}} G^{\otimes 2}_{h_1,h_2}(x - x_0, y - y_0)\nu(x)q(x, y)dxdy.$$

The convergence of $\mathcal{M}_n(\varphi)$ to $\nu(\varphi)$ suggests to pick $\varphi = G_h(\cdot - x_0)$ that converges to $G_h \star \nu(x_0)$ and use it as a proxy of $\nu(x_0)$. We obtain the estimator

$$\widehat{\nu}_n(x_0) = \mathcal{M}_n\big(G_h(\cdot - x_0)\big),$$

specified by the choice of $h > 0$ and the kernel G. Likewise, an estimator of $q(x_0, y_0)$ is obtained by considering the quotient estimator with numerator $\psi = G_{h_1,h_2}(\cdot - x_0, \cdot - y_0)$ that converges to $G_h \star \nu(x_0)q(x_0, y_0)$ and denominator $\widehat{\nu}_{n,h}(x_0)$ in order to balance the superfluous weight $\nu(x_0)$ in the numerator. We obtain the estimator

$$\widehat{q}_n(x_0, y_0) = \frac{\mathcal{M}_n\big(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\big)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi},$$

specified by the choice of $h, h_1, h_2 > 0$, a threshold $\varpi > 0$ and the kernel G. In order to quantify the kernel approximation, we introduce anisotropic Hölder classes. For $\alpha > 0$, we write $\alpha = \lfloor \alpha \rfloor + \{\alpha\}$ with $\lfloor \alpha \rfloor$ an integer and $0 < \{\alpha\} \leqslant 1$.

Definition 3.3.2. *Let $\alpha, \beta > 0$ and $\mathcal{V}_{x_0}$ and $\mathcal{V}(y_0)$ be bounded neighborhoods of x_0 and y_0.*

i) The function $\varphi : \mathcal{V}_{x_0} \to \mathbb{R}$ belongs to the Hölder class $\mathcal{H}^\alpha(x_0)$ if

$$|\varphi^{(\lfloor \alpha \rfloor)}(y) - \varphi^{(\lfloor \alpha \rfloor)}(x)| \leqslant C|y - x|^{\{\alpha\}} \text{ for every } x, y \in \mathcal{V}_{x_0}. \tag{3.10}$$

ii) The function $\psi : \mathcal{V}_{x_0} \times \mathcal{V}_{y_0} \to \mathbb{R}$ *belongs to the anisotropic Hölder class* $\mathcal{H}^{\alpha,\beta}(x_0, y_0)$ *if*

$$x \mapsto \psi(x, y_0) \in \mathcal{H}^{\alpha}(x_0) \text{ and } y \mapsto \psi(x_0, y) \in \mathcal{H}^{\beta}(y_0)$$

hold simultaneously.

We obtain a semi-norm on $\mathcal{H}^{\alpha}(x_0)$ by setting $|\varphi|_{\mathcal{H}^{\alpha}(x_0)} = \sup_{x \in \mathcal{V}_{x_0}} |\varphi(x)| + c_{\alpha}(\varphi)$, where $c_{\alpha}(\varphi)$ is the smallest constant for which (3.10) holds. Likewise, we equip $\mathcal{H}^{\alpha,\beta}(x_0, y_0)$ with the semi-norm $|\psi|_{\mathcal{H}^{\alpha,\beta}(x_0,y_0)} = |\psi(\cdot, y_0)|_{\mathcal{H}^{\alpha}(x_0)} + |\psi(x_0, \cdot)|_{\mathcal{H}^{\beta}(y_0)}$. The space $\mathcal{H}^{\alpha,\beta}(x_0, y_0)$ is appended with (semi) Hölder balls

$$\mathcal{H}^{\alpha,\beta}(x_0, y_0)(R) = \big\{\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, |\psi|_{\mathcal{H}^{\alpha,\beta}(x_0,y_0)} \leqslant R\big\}, \quad R > 0.$$

We are ready to state our convergence result over transitions $\mathcal{Q}$ that belong to

$$\mathbb{Q}^{\alpha,\beta}_{1/2}(R) = \mathbb{Q}_{1/2} \cap \mathcal{H}^{\alpha,\beta}(x_0, y_0)(R), \quad R > 0,$$

with a slight abuse of notation.

Theorem 3.3.3. *Work under Assumptions 1,2, 3. Assume that the initial distribution* μ *is absolutely continuous w.r.t. the Lebesgue measure with a locally bounded density function and satisfies* $\mu(V^2) < \infty$.

Let $\alpha, \beta > 0$. *Specify* $\widehat{\nu}_n(y_0)$ *by a kernel of order* $k > \max\{\alpha, \beta\}$, $h = |\mathbb{T}_n|^{-1/(2\beta+1)}$ *and* $\widehat{q}_n(x_0, y_0)$ *with the same kernel,* $h_1 = |\mathbb{T}_n|^{-s(\alpha,\beta)/(\alpha\wedge\beta)(2s(\alpha,\beta)+1)}$, $h_2 = |\mathbb{T}_n|^{-s(\alpha,\beta)/\beta(2s(\alpha,\beta)+1)}$ *and* $\varpi_n \to 0$. *Then, for every* $R > 0$

$$\sup_{\mathcal{Q} \in \mathbb{Q}^{\alpha,\beta}_{1/2}(R)} \big(\mathbb{E}\big[\big(\widehat{\nu}_n(y_0) - \nu(y_0)\big)^2\big]\big)^{1/2} \lesssim |\mathbb{T}_n|^{-\beta/(2\beta+1)}$$

and

$$\sup_{\mathcal{Q} \in \mathbb{Q}^{\alpha,\beta}_{1/2}(R)} \big(\mathbb{E}\big[\big(\widehat{q}_n(x_0, y_0) - q(x_0, y_0)\big)^2\big]\big)^{1/2} \lesssim \varpi_n^{-1} |\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$$

hold true, where $s(\alpha, \beta)^{-1} = (\alpha \wedge \beta)^{-1} + \beta^{-1}$ *is the effective anisotropic smoothness associated with* (α, β).

Several remarks are in order: **1)** We obtain an optimal result in the minimax sense for estimating $\nu(y_0)$ and in the case $\beta \geqslant \alpha$ for estimating $q(x_0, y_0)$. This stems from the fact that the representation $\nu(x) = \int_{\mathcal{X}} \nu(y) q(y, x) dy$ henceforth $q \in \mathcal{H}^{\alpha,\beta}$ implies that $\nu \in \mathcal{H}^{\beta}$. In turn, the numerator of $\widehat{q}_n(x_0, y_0)$ is based on the estimation of the function $\nu(x)q(x, y) \in \mathcal{H}^{\alpha\wedge\beta,\beta}$. **2)** In the estimation of $q(x_0, y_0)$, we have a superfluous term ϖ_n^{-1} in the error that can be taken arbitrarily small, and that comes from the denominator of the estimator. It can be removed, however at a significant technical cost. Alternatively, one can get rid of it by weakening the error loss: it is not difficult to prove

$$\big(\mathbb{E}\big[\big(\widehat{q}_n(x_0, y_0) - q(x_0, y_0)\big)^p\big]\big)^{1/p} \lesssim |\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)} \text{ for every } 0 < p < 2,$$

and the result of course also holds in probability. **3)** The assumption that μ is absolutely continuous can also be removed. **4)** Finally, a slightly annoying fact is that the estimators $\widehat{\nu}_n(x_0)$ and $\widehat{q}_n(x_0, y_0)$ require the knowledge of (α, β) to be tuned optimally, and this is not reasonable in practice. It is possible to tune our estimators in practice by cross-validation in the same spirit as in [HO16], but an adaptive estimation theory still needs to be established. This lies beyond the scope of the chapter, and requires concentration inequalities, a result we do not have here, due to the fact that the model is not uniformly geometrically ergodic (otherwise, we could apply the same strategy as in [BPHO17, BPR17]).

3.3.2 Parametric estimation of the division rate

In order to conduct inference on the division rate $x \mapsto B(x)$, we need more stringent assumptions on the model so that we can apply the results of Theorem 3.2.3. The main difficulty lies in the fact that we need to apply Theorem 3.2.3 to test functions of the form $\psi(x, y) = \log q(x, y)$ when applied to the loglikelihood of the data, and that these functions are possibly unbounded.

A stochastic trait model as a diffusion on a compact with reflection at the boundary

We circumvent this difficulty by assuming that the trait $\phi_x(t)$ of each individual evolves in a bounded interval with reflections at the boundary and with no loss of generality, we take $\mathcal{X} = [0, L]$ for some $L > 0$. The dynamics of the traits now follows

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t + d\ell_t, \tag{3.11}$$

where the solution $(\ell_t)_{t\geqslant 0}$ to $\ell_t = \int_0^t \big(\mathbf{1}_{\{X_s=0\}} + \mathbf{1}_{\{X_s=L\}}\big)\, d\ell_s$ accounts for the reflection at the boundary and $(W_t)_{t\geqslant 0}$ is a standard Brownian motion. Under Assumption 1 (that reduces here to the boundedness of r, σ and the ellipticity of σ) there exists a unique strong solution to (3.11), see for instance Theorem 4.1. in [Tan79].

A slight modification of Proposition (3.2.1) gives the following explicit formulas for the transitions $\mathcal{P}$ and $\mathcal{Q}$. Remember that by Assumption 3, we have $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$. Define

$$\mathcal{D} = \big\{0 < y_1 \leqslant \varepsilon L,\ \tfrac{\varepsilon}{1-\varepsilon} y_1 \leqslant y_2 \leqslant \tfrac{1-\varepsilon}{\varepsilon} y_1\big\} \cup \big\{\varepsilon L \leqslant y_1 \leqslant (1-\varepsilon)L,\ \tfrac{\varepsilon}{1-\varepsilon} y_1 \leqslant y_2 \leqslant \tfrac{L-y_1}{y_1}\big\}.$$

Then the explicit formula for $p(x, y_1, y_2)$ given in (3.7) remains unchanged provided $(x, y_1, y_2) \in \mathcal{X} \times \mathcal{D}$ and it vanishes outside of $\mathcal{X} \times \mathcal{D}$. For $q(x, y)$, the formula (3.8) now becomes

$$q(x, y) = \int_{y/L}^{1} \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\left[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\right] dz, \tag{3.12}$$

for $(x, y) \in \mathcal{X} \times \{0 \leqslant y \leqslant (1-\varepsilon)L\}$ and 0 otherwise.

Adapting the proof of Theorem 3.2.2 to the case of a diffusion living on a compact interval (formally replacing $[-w, w]$ by $[0, L]$ in the proof of Proposition 3.4.2 below) one easily checks that Theorem 3.2.2 remains valid in this setting (applying for instance Theorem 4.3.16 in [CMR05]). In turn, Theorem 3.2.3 also holds true in the case of a reflected diffusion. For parametric estimation, the control on the variance of $\mathcal{M}_n(\psi)$ is less demanding and we will simply need the following

Corollary 3.3.4. *Work under Assumptions 1,2, 3 in the case of a reflected diffusion on* $[0, L]$ *for the evolution of the trait* $(\phi_x(t),\ t \geqslant 0)$. *Let* $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. *Then, for any probability measure* μ, *we have*

$$\sup_{\mathcal{Q}\in\mathbb{Q}_{1/2}} \mathbb{E}_\mu\big[(\mathcal{M}_n(\psi) - \nu(\psi))^2\big] \lesssim |\mathbb{T}_n|^{-1} \sup_{x,y} \psi(x, y)^2.$$

Maximum likelihood estimation

From now on, we fix a triplet $(r_0, \sigma_0, \kappa_0)$ and we let the division rate $x \mapsto B(x)$ belong to a parametric class

$$\mathcal{B} = \big\{B : \mathcal{X} \to \mathbb{R}, B(x) = B_0(\vartheta, x), x \in \mathcal{X}, \vartheta \in \Theta\big\},$$

where $x \mapsto B_0(x, \vartheta)$ is known up to the parameter $\vartheta \in \Theta$, and $\Theta \subset \mathbb{R}^d$ for some $d \geqslant 1$ is a compact subset of the Euclidean space. In this setting, the model is entirely characterized by ϑ which is our parameter of interest. A first minimal stability requirement of the parametric model is the following

Assumption 4. *We have* $\mathbb{Q}(\mathcal{B}) = \{\mathcal{Q} = \mathcal{Q}(r_0, \sigma_0, B, \kappa), B \in \mathcal{B}\} \subset \mathbb{Q}_{1/2}$.

A second minimal requirement is the identifiability of the class $\mathcal{B}$, namely the fact that the map

$$B \mapsto \mathcal{Q}(r_0, \sigma_0, B, \kappa_0)$$

from $\mathcal{B}$ to $\mathbb{Q}$ is injective. This is satisfied in particular if $\mathcal{B}$ satisfies a certain orderliness property.

Definition 3.3.5. *A class $\mathcal{B}$ of functions from $\mathcal{X} \to [0, \infty)$ is orderly if $\varphi_1, \varphi_2 \in \mathcal{B}$ implies either $\varphi_1(x) \leqslant \varphi_2(x)$ for every $x \in \mathcal{X}$ or $\varphi_2(x) \leqslant \varphi_1(x)$ for every $x \in \mathcal{X}$.*

Proposition 3.3.6. *Let $\mathcal{B}$ be orderly in the sense of Definition 3.3.5 and $\mathcal{Q}(\mathcal{B}) \subset \mathbb{Q}$ for some $(r_0, \sigma_0, \kappa_0)$. Then $B \mapsto \mathcal{Q}(r_0, \sigma_0, B, \kappa_0)$ is injective.*

We further stress the dependence on ϑ by introducing a subscript in the notation whenever relevant. We formally obtain a statistical experiment

$$\mathcal{E}^n = \big\{\mathbb{P}^n_\vartheta, \vartheta \in \Theta\big\}$$

by letting $\mathbb{P}^n_\vartheta$ denote the law of $\mathbb{X}^n = (X_u, u \in \mathbb{T}_n)$ under $\mathcal{P}_\vartheta$ with initial condition $X_\varnothing$ distributed according to ν_ϑ on the product space $\mathcal{X}^{|\mathbb{T}_n|}$ endowed with its Borel sigma-field. The experiment $\mathcal{E}^n$ is dominated by the Lebesgue measure on $\mathcal{X}^{|\mathbb{T}_n|}$ and we obtain a likelihood function by setting

$$\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big) = \prod_{u \in \mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u), \tag{3.13}$$

where X_{u^-} denotes the trait of the parent of u. Taking any maximizer of (3.13) we obtain a maximum likelihood estimator

$$\widehat{\vartheta}_n \in \underset{\vartheta \in \Theta}{\operatorname{argmax}}\, \mathcal{L}_n\big(\vartheta, \mathbb{X}^n\big)$$

provided a maximizer exists.

Convergence results and asymptotic efficiency

We first have an existence and consistency result of $\widehat{\vartheta}_n$ under the following non-degeneracy assumption that strengthens Assumption 2.

Assumption 5. *The function $B_0 : \Theta \times \mathcal{X} \to [0, \infty)$ is continuous and for some positive b_3, b_4, we have*

$$0 < b_3 \leqslant \inf_{\vartheta, x} B_0(\vartheta, x) \leqslant \sup_{\vartheta, x} B_0(\vartheta, x) \leqslant b_4$$

Moreover, the class $\mathcal{B} = \big\{B_0(\vartheta, \cdot), \vartheta \in \Theta\big\}$ is orderly in the sense of Definition 3.3.5.

Theorem 3.3.7. *Work under Assumptions 1, 3, 4 and 5. Then, for every $\vartheta \in \Theta$, $\widehat{\vartheta}_n$ converges to ϑ in probability as $n \to \infty$.*

Our next result gives an explicit rate of convergence and asymptotic normality for $\widehat{\vartheta}_n$. We need further regularity assumptions.

Assumption 6. *The set Θ has non empty interior and, for every $x \in \mathcal{X}$ the map $\vartheta \mapsto B_0(\vartheta, x)$ is three times continuously differentiable. Moreover, for every $1 \leqslant i, j, k \leqslant d$:*

$$\sup_{\vartheta,x} |\partial_{\vartheta_i} B_0(\vartheta, x)| + \sup_{\vartheta,x} |\partial^2_{\vartheta_i \vartheta_j} B_0(\vartheta, x)| + \sup_{\vartheta,x} |\partial^3_{\vartheta_i \vartheta_j \vartheta_k} B_0(\vartheta, x)| < \infty.$$

Introduce the Fisher information $\Psi(\vartheta)$ at point $\vartheta \in \Theta$ as the $d \times d$-matrix with entries:

$$\Psi(\vartheta)_{i,j} = \nu_\vartheta \mathcal{Q}_\vartheta\Big(\frac{\partial_{\vartheta_i} q_\vartheta \, \partial_{\vartheta_j} q_\vartheta}{q_\vartheta^2}\Big) = \int_{\mathcal{X}\times\mathcal{X}} \frac{\partial_{\vartheta_i} q_\vartheta(x,y) \, \partial_{\vartheta_j} q_\vartheta(x,y)}{q_\vartheta(x,y)^2} \nu_\vartheta(x) q_\vartheta(x,y) dx dy,$$

for $1 \leqslant i, j \leqslant d$.

Assumption 7. *For every ϑ in the interior of Θ, the matrix $\Psi(\vartheta)$ is nonsingular.*

Although standard in regular parametric estimation, Assumption 7 is not obviously satisfied even if we have the explicit formula (3.12), for $q_\vartheta(x, y)$, due to its relatively intricate form. We can however easily show that it is satisfied in the special case of a trait evolving as a reflected diffusion with constant drift. More general parametrizations are presumably possible, adapting the proof delayed until Appendix 3.5.2.

Proposition 3.3.8. *Assume $d = 1$, $B_0(\vartheta, x) = \vartheta$ for every $x \in \mathcal{X}$, with $\Theta = [\vartheta_1, \vartheta_2] \subset (0, \infty)$, $r(x) = r_1 < 0$ and $\sigma(x) = \sigma_0 > 0$ for every $x \in \mathcal{X}$. Let $\kappa(z) = (1 - 2\varepsilon)^{-1}$ for every $z \in [\varepsilon, 1 - \varepsilon]$. There exists an explicit open interval $\mathcal{I} \subset (0, 1/2)$ such that Assumption 7 is satisfied as soon as $\varepsilon \in \mathcal{I}$.*

We are ready to state our final result on asymptotic normality of $\widehat{\vartheta}_n$.

Theorem 3.3.9. *Work under Assumptions 1, 3, 4, 5, 6 and 7. For every ϑ in the interior of Θ, we have*

$$\mathbb{T}_n^{1/2}(\widehat{\vartheta}_n - \vartheta) \to \mathcal{N}\big(0, \Psi(\vartheta)^{-1}\big)$$

in distribution as $n \to \infty$, where $\mathcal{N}(0, \Psi(\vartheta)^{-1})$ denote the d-dimensional Gaussian distribution with mean 0 and covariance the inverse of the Fisher matrix $\Psi(\vartheta)$.

Several remarks are in order: **1)** Although asymptotically optimal, the practical implementation of $\widehat{\vartheta}_n$ is a challenging question that we plan to address in a systematic way. **2)** As for classical estimation in diffusion processes (see e.g. [DCFZ86, GCJ93]), the assumptions of Theorem 3.3.9, especially Assumption 7 are standard. However, the fact that they hold true in the simple case of Proposition 3.3.8 and a glance at the proof is an indication that they are certainly true in wider generality. **3)** We could not find another contrast estimator than the MLE, and in particular, we do not know how to estimate $x \mapsto B(x)$ nonparametrically. One formal possibility is to let $d = d_n \to \infty$ as $n \to \infty$, and we could certainly obtain an optimal result provided $\|\Psi(\vartheta)^{-1}\| \lesssim d_n$ in operator norm. Such a result necessarily depends on the regularity of parametrization $\vartheta \mapsto B_0(\vartheta, \cdot)$ via the parametrization $\vartheta \mapsto \partial_{\vartheta_i} q_\vartheta$. Such a study seems out of reach with the tools developed here.

3.4 Proofs

3.4.1 Proof of Proposition 3.2.1

We first prove (3.8). By the occupation times formula, a change of variable and Fubini's theorem, for any bounded $\varphi : \mathcal{X} \to \mathbb{R}$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}
\mathcal{Q}\varphi(x) &= \mathbb{E}\big[\varphi(X_u)\big|X_{u^-} = x\big] \\
&= \mathbb{E}\Big[\int_0^1 \kappa(z) \int_0^\infty \varphi\big(z\phi_x(t)\big)B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}dtdz\Big] \\
&= \mathbb{E}\Big[\int_0^1 \frac{\kappa(z)}{z} \int_{\mathbb{R}} \varphi(y)B(y/z) \int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}\frac{dL_t^{y/z}(\phi_x)}{\sigma(y/z)^2}dydz\Big] \\
&= \int_{\mathbb{R}} \varphi(y) \int_0^1 \frac{\kappa(z)}{z}B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^{y/z}(\phi_x)dz\Big]dy,
\end{aligned}$$

where $(L_t^y(\phi_x))_{t \geqslant 0}$ is the local time of ϕ_x at $y \in \mathcal{X}$. Since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, the above integrals are well defined and (3.8) is established. We turn to (3.7). By (3.3), for any bounded $\psi : \mathcal{X}^3 \to \mathbb{R}$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}
\mathcal{P}\psi(x) &= \mathbb{E}\Big[\int_0^1 \kappa(z) \int_0^\infty \psi\big(x, z\phi_x(t), (1-z)\phi_x(t)\big)B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}dtdz\Big] \\
&= \int_{\mathbb{R}} \int_0^1 \psi\big(x, y_1, \tfrac{1-z}{z}y_1\big)\frac{B\big(y_1/z\big)}{\sigma(y_1/z)^2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^{y_1/z}(\phi_x)\Big]\frac{\kappa(z)}{z}dzdy_1 \\
&= \int_{\mathcal{D}} \psi\big(x, y_1, y_2\big)\frac{\kappa\big(y_1/(y_1+y_2)\big)}{y_1+y_2}\frac{B\left(y_1+y_2\right)}{\sigma(y_1+y_2)^2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^{y_1+y_2}(\phi_x)\Big]
\end{aligned}$$

where we set $y_2 = \frac{1-z}{z}y_1$ in order to obtain the last line. The integral is taken over the domain

$$\mathcal{D} = \big\{(y_1, y_2) \in \mathbb{R}^2,\ \tfrac{\varepsilon}{1-\varepsilon}y_1 \leqslant y_2 \leqslant \tfrac{1-\varepsilon}{\varepsilon}y_1\big\} \subset \mathrm{supp}\big((y_1, y_2) \mapsto \kappa(z_1+z_2)\big)$$

therefore the above integral is well defined and the representation (3.7) is proved.

3.4.2 Proof of Theorem 3.2.2

The proof goes along a classical path: we establish a drift and a minorisation condition in Proposition 3.4.1 and 3.4.2 below, and then apply for instance Theorem 1.2. in [HM11], see also the references therein.

Proposition 3.4.1 (Drift condition). *Let* $V(x) = x^2$. *Work under Assumptions 1,2 and 3. There exist explicitly computable* $0 < v_1 = v_1(\varepsilon) < 1$ *and* $v_2 = v_2(\varepsilon, r_1, r_2, \sigma_1, \sigma_2, b_1) > 0$ *such that*

$$\mathcal{Q}V(x) \leqslant v_1V(x) + v_2.$$

Proposition 3.4.2 (Minorisation condition). *Work under Assumption 1, 2 and 3. For large enough* $w > 0$, *there exists* $\alpha \in (0,1)$ *and a probability measure* μ *on* $\mathcal{X}$ *such that*

$$\inf_{\{x, |x| \leqslant w\}} \mathcal{Q}(x, \mathcal{A}) \geqslant \alpha\mu(\mathcal{A})$$

for every Borel set $\mathcal{A} \subset \mathcal{X}$.

Proof of Proposition 3.4.1. Fix $x \in \mathcal{X}$ and let $m(\kappa) = \int_0^1 z^2\kappa(z)dz$. By Itô formula, we obtain the decomposition

$$\mathcal{QV}(x) = \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\Big[z^2\phi_x(t)^2 B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dtdz = m(\kappa)(I + II + III + IV),$$

where

$$\begin{aligned}
I &= x^2 \int_0^\infty \mathbb{E}\Big[B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dt,\\
II &= 2\int_0^\infty \mathbb{E}\Big[\int_0^t \phi_x(u)r(\phi_x(u))du\, B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dt,\\
III &= \int_0^\infty \mathbb{E}\Big[\int_0^t \sigma\big(\phi_x(u)\big)^2 du\, B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dt,\\
IV &= 2\int_0^\infty \mathbb{E}\Big[\int_0^t \phi_x(u)\sigma\big(\phi_x(u)\big)dW_u\, B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dt.
\end{aligned}$$

First, note that $\int_0^\infty B(\phi_x(t))e^{-\int_0^t B(\phi_x(s))ds}dt = 1$, therefore $I = x^2$ by Fubini's theorem. We turn to II. By Fubini's theorem again:

$$\begin{aligned}
II &= 2\,\mathbb{E}\Big[\int_0^\infty \phi_x(u)r\big(\phi_x(u)\big)\Big(\int_u^\infty B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}dt\Big)du\Big]\\
&= 2\,\mathbb{E}\Big[\int_0^\infty \phi_x(u)r\big(\phi_x(u)\big)e^{-\int_0^u B(\phi_x(s))ds}du\Big]\\
&= 2\int_{\mathbb{R}} zr(z)\sigma(z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^z(\phi_x)\Big]dz.
\end{aligned}$$

by the occupation times formula. By Assumption 1 we have $zr(z) < 0$ for $|z| \geqslant r_2$, therefore:

$$\begin{aligned}
II &\leqslant 2\int_{-r_2}^{r_2} zr(z)\sigma(z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^z(\phi_x)\Big]dz\\
&\leqslant 2r_1r_2(1+r_1)\sigma_1^{-2}\int_{-r_2}^{r_2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds}dL_t^z(\phi_x)\Big]dz\\
&\leqslant 2r_1r_2(1+r_1)\sigma_1^{-2}\int_0^\infty e^{-b_1t}\mathbb{P}\big(-t_2 \leqslant \phi_x(t) \leqslant r_2\big)dt \leqslant 2r_1r_2(1+r_1)\sigma_1^{-2}b_1^{-1}.
\end{aligned}$$

using successively Assumption 1, 2 and the occupation times formula. For the term III, we have

$$\begin{aligned}
III &= \int_0^\infty \mathbb{E}\Big[\sigma\big(\phi_x(u)\big)^2\int_u^\infty B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]du\\
&= \int_0^\infty \mathbb{E}\Big[\sigma(\phi_x(u))^2 e^{-\int_0^u B(\phi_x(s))ds}\Big]du
\end{aligned}$$

by Fubini again and this last quantity is less than $\sigma_2^2\int_0^\infty e^{-b_1t}dt = \sigma_2^2b_1^{-1}$ by Assumption 1 and 2. Similarly for the term IV, we have

$$\begin{aligned}
IV &= 2\,\mathbb{E}\Big[\int_0^\infty \sigma\big(\phi_x(u)\big)\int_u^\infty B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}dtdW_u\Big]\\
&= 2\,\mathbb{E}\Big[\int_0^\infty \sigma\big(\phi_x(u)\big)e^{-\int_0^u B(\phi_x(s))ds}dW_u\Big]
\end{aligned}$$

and this last quantity vanishes. Putting the estimates for I, II, III and IV together, we conclude

$$\mathcal{Q}V(x) \leqslant m(\kappa)x^2 + m(\kappa)(2r_1r_2(1+r_1)\sigma_1^{-2} + \sigma_2^2)b_1^{-1}.$$

Since $\operatorname{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, we have $m(\kappa) \leqslant (1-\varepsilon)^2 < 1$ and this completes the proof with $v_1 = m(\kappa)$ and $v_2 = m(\kappa)(2r_1r_2(1+r_1)\sigma_1^{-2} + \sigma_2^2)b_1^{-1}$.

☐

Proof of Proposition 3.4.2. Step 1). Let $x \in [-w, w]$ and $\mathcal{A} \subset \mathcal{X}$ be a Borel set. Applying Assumption 3, introducing the event $\mathcal{W}(\phi_x(t)) = \{2w \leqslant \phi_x(t) \leqslant 2w\sqrt{(1-\varepsilon)/\varepsilon}\}$, applying Fubini's theorem and a change of variable, we successively obtain

$$\begin{aligned}
\mathcal{Q}(x,\mathcal{A}) &= \int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\{z\phi_x(t)\in\mathcal{A}\}}B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\Big]dtdz\\
&\geqslant b_1\int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))}\mathbf{1}_{\{z\phi_x(t)\in\mathcal{A}\}}e^{-\int_0^t B\big(\phi_x(s)\big)ds}\Big]dtdz\\
&\geqslant b_1\int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))}e^{-\int_0^t B(\phi_x(s))ds}\int_0^{\phi_x(t)}\mathbf{1}_{\{y\in\mathcal{A}\}}\kappa\big(y/\phi_x(t)\big)\phi_x(t)^{-1}dy\Big]dt\\
&= \int_{\mathbb{R}} \mathbf{1}_{\{y\in A\}}f(x,y)dy,
\end{aligned}$$

with $f(x,y) = b_1\int_0^\infty \mathbb{E}\big[\mathbf{1}_{\mathcal{W}(\phi_x(t))}\mathbf{1}_{\{y\leqslant\phi_x(t)\}}e^{-\int_0^t B(\phi_x(s))ds}\kappa\big(y/\phi_x(t)\big)\phi_x(t)^{-1}\big]dt$.

Step 2). We now prove that f is bounded below independently of x. By Assumption 3 and the definition of $\mathcal{W}(\phi_x(t))$, we have

$$\begin{aligned}
f(x,y) &\geqslant \delta b_1\int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))}\mathbf{1}_{\{\varepsilon\phi_x(t)\leqslant y\leqslant(1-\varepsilon)\phi_x(t)\}}e^{-\int_0^t B(\phi_x(s))ds}\phi_x(t)^{-1}\Big]dt\\
&\geqslant \frac{\delta b_1}{2w\sqrt{(1-\varepsilon)/\varepsilon}}\mathbf{1}_{\big\{2w\sqrt{\varepsilon(1-\varepsilon)}\leqslant y\leqslant 2w(1-\varepsilon)\big\}}h(x,y),
\end{aligned}$$

with $h(x,y) = \int_0^\infty \mathbb{E}\big[\mathbf{1}_{\mathcal{W}(\phi_x(t))}e^{-\int_0^t B(\phi_x(s))ds}\big]dt$. Let $\Delta = (1+\sqrt{(1-\varepsilon)/\varepsilon})w$ denote the mid-point of the interval $[2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]$. Let also $T_y^x = \inf\{t > 0, \phi_x(t) \geqslant y\}$ denote the exit time of the interval $(\inf\mathcal{X}, y)$ by $(\phi_x(t))_{t\geqslant 0}$. It follows that

$$\begin{aligned}
h(x,y) &\geqslant \mathbb{E}\Big[\int_{T_\Delta^x}^\infty \mathbf{1}_{\mathcal{W}(\phi_x(t))}e^{-\int_0^t B(\phi_x(s))ds}dt\Big]\\
&= \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))}e^{-\int_0^{t+T_\Delta^x} B(\phi_x(s))ds}\Big]dt\\
&= \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))}e^{-\int_0^{T_\Delta^x} B(\phi_x(s))ds}e^{-\int_0^t B\big(\phi_x(s+T_\Delta^x)\big)ds}\Big]dt\\
&\geqslant \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))}e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)}e^{-\int_0^t B\big(\phi_x(s+T_\Delta^x)\big)ds}\Big]dt,
\end{aligned}$$

by Assumption 2 and because $\phi_x(s) \leqslant \Delta$ for every $s \leqslant T_\Delta^x$. Applying the strong Markov property, we further obtain

$$h(x,y) \geqslant \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_\Delta(t))}e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)}e^{-\int_0^t B(\phi_\Delta(s))ds}\Big]dt \tag{3.14}$$

since $\phi_x(T_\Delta^x) = \Delta$ for $x \leqslant w < \Delta$. Introduce next

$$T_{\mathcal{W}}^{\Delta} = \inf\{t \geqslant 0,\ \phi_\Delta(t) \notin [2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]\},$$

i.e. the exit time of $[2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]$ by $(\phi_\Delta(t))_{t\geqslant 0}$. By (3.14) and Assumption 2 again, it follows that

$$\begin{aligned} h(x,y) &\geqslant \mathbb{E}\Big[\int_0^{T_{\mathcal{W}}^{\Delta}} \mathbf{1}_{\mathcal{W}(\phi_\Delta(t))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-\int_0^t (b_2|\phi_x(s)|^\gamma+b_1)ds} dt\Big] \\ &\geqslant \mathbb{E}\left[\int_0^{T_{\mathcal{W}}^{\Delta}} \mathbf{1}_{\mathcal{W}(\phi_\Delta(t))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-v_3 t} dt\right], \end{aligned}$$

using that $\phi_x(t) \leqslant 2w\sqrt{(1-\varepsilon)/\varepsilon}$ for $t \leqslant T_{\mathcal{W}}^{\Delta}$ and where $v_3 = b_2(2w\sqrt{(1-\varepsilon)/\varepsilon})^\gamma + b_1$. Since $\Delta > -w$, the event $\{T_\Delta^x \leqslant T_\Delta^{-w}\}$ holds almost-surely for every $x \in [-w, w]$ and therefore

$$h(x,y) \geqslant \mathbb{E}\Big[\int_0^{T_{\mathcal{W}}^{\Delta}} e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)} e^{-tv_3} dt\Big] \geqslant v_3^{-1}\mathbb{E}\big[e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)}\big]\mathbb{E}\big[1-e^{-T_{\mathcal{W}}^{\Delta}v_3}\big]. \tag{3.15}$$

by the independence of $T_{\mathcal{W}}^{\Delta}$ and T_Δ^{-w}. Furthermore, for every $a, x \in \mathcal{X}$ with $x < a$, we have

$$\mathbb{P}\left(T_a^x < \infty\right) = \lim_{u \to -\infty} \frac{s(u) - s(x)}{s(u) - s(a)},$$

where

$$s(x) = \int_{\inf \mathcal{X}}^x \exp(-2\int_{\inf \mathcal{X}}^y r(z)\sigma(z)^{-2}dz)dy,$$

is the scale function associated to $(\phi_x(t))_{t\geqslant 0}$. By the classical Feller classification of scalar diffusions (see e.g. Revuz and Yor [RY99]), we have the equivalence $\mathbb{P}\left(T_a^x < \infty\right) = 0$ if only if $\int_{\inf \mathcal{X}} s(x)dx = 0$ but that latter property contradicts Assumption 1. Therefore, there exist $w_1, \delta_1 > 0$ such that $\mathbb{P}(T_\Delta^{-w} \leqslant w_1) > \delta_1$. It follows that

$$\mathbb{E}\big[e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)}\big] \geqslant \mathbb{E}\big[e^{-w_1(b_2\Delta^\gamma+b_1)}\mathbf{1}_{\{T_\Delta^{-w}\leqslant w_1\}}\big] \geqslant \delta_1 e^{-w_1(b_2\Delta^\gamma+b_1)} \tag{3.16}$$

and since $T_{\mathcal{W}}^{\Delta} > 0$ almost surely, there exists $\delta_2 > 0$, independent of x, such that

$$\mathbb{E}\big[1-e^{-T_{\mathcal{W}}^{\Delta}v_3}\big] > \delta_2. \tag{3.17}$$

Back to (3.15), putting together (3.16) and (3.17), we obtain

$$h(x,y) \geqslant v_3^{-1}\delta_1 e^{-w_1(b_2\Delta^\gamma+b_1)}\delta_2$$

and eventually

$$f(x,y) \geqslant \frac{\delta b_1\delta_1\delta_2}{2w\sqrt{(1-\varepsilon)/\varepsilon}v_3}\mathbf{1}_{\left\{2w\sqrt{\varepsilon(1-\varepsilon)}\leqslant y\leqslant 2w(1-\varepsilon)\right\}} e^{-w_1(b_2\Delta^\gamma+b_1)}.$$

Step 3). Define the probability measure $\mu(dy) = f(y)dy$ on $\mathcal{X}$ by

$$f(y) = \frac{1}{2w\big(1-\varepsilon-\sqrt{\varepsilon(1-\varepsilon)}\big)}\mathbf{1}_{\left\{2w\sqrt{\varepsilon(1-\varepsilon)}\leqslant y\leqslant 2w(1-\varepsilon)\right\}},$$

and let $\alpha = (1-\varepsilon-\sqrt{\varepsilon(1-\varepsilon)})\frac{b_1\delta\delta_1\delta_2}{\sqrt{(1-\varepsilon)/\varepsilon}v_3}e^{-w_1(b_2\Delta^\gamma+b_1)}$. We may assume that $0 < \alpha < 1$ (the lower bound remains valid if we replace δ by $\delta' < \delta$ for instance) and we thus have established

$$\mathcal{Q}(x, \mathcal{A}) \geqslant \alpha\mu(\mathcal{A}),$$

for an arbitrary Borel set $\mathcal{A} \subset \mathcal{X}$. The proof of Proposition 3.4.2 is complete. □

3.4.3 Proof of Theorem 3.2.3

Preparations

We first state a useful estimate on the local time of $L_t^y(\phi_x)$ as $t \to \infty$. Its proof is delayed until Appendix 3.5.1.

Lemma 3.4.3. *Work under Assumption 1. For every compact $\mathcal{K} \subset \mathcal{X}$ and for every $t \geqslant 0$, we have*

$$\sup_{x\in\mathcal{X}, y\in\mathcal{K}} \mathbb{E}\left[L_t^y\left(\phi_x\right)\right] \lesssim 1 + t^{3/2},$$

up to a constant that only depends on r_1, r_2 and σ_2. In particular, for every $c > 0$, the function

$$y \mapsto \int_0^\infty e^{-ct} \sup_{x\in\mathcal{X}} \mathbb{E}\left[L_t^y\left(\phi_x\right)\right] dt$$

is well-defined and locally bounded, uniformly over $\mathbb{Q}$.

Lemma 3.4.3 enables us to obtain estimates on the action of $\mathcal{P}$ and $\mathcal{Q}$ on functions $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with nice behaviours that will prove essential for obtaining Theorem 3.2.3. We set

$$\mathcal{Q}\psi(x) = \int_{\mathcal{X}} \psi(x,y)q(x,y)dy, \quad \mathcal{P}(\psi\otimes\psi)(x) = \int_{\mathcal{X}\times\mathcal{X}} \psi(x,y_1)\psi(x,y_2)p(x,y_1,y_2)dy_1dy_2,$$

where $p(x,y_1,y_2)$ and $q(x,y)$ are given in Proposition 3.2.1.

Lemma 3.4.4. *Work under Assumptions 1, 2 and 3. Let $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be bounded and such that $\psi_\star$ has compact support. There exists a constant $c_{\mathrm{supp}(\psi_\star)}$ depending on $\mathrm{supp}(\psi_\star)$ (and $\mathcal{Q}$) such that*

$$|\mathcal{Q}\psi(x)| \leqslant c_{\mathrm{supp}(\psi_\star)} \int_{\mathcal{X}} |\psi(x,y)|dy \tag{3.18}$$

and

$$|\mathcal{P}(\psi\otimes\psi)(x)| \leqslant c_{\mathrm{supp}(\psi_\star)} \psi^\star(x) \int_{\mathcal{X}} |\psi(x,y)|dy.$$

Note in particular that (3.18) implies in turn the estimate

$$|\mathcal{Q}^2\psi(x)| \leqslant c^2_{\mathrm{supp}(\psi_\star)} |\psi|_1 \wedge c_{\mathrm{supp}(\psi_\star)}| \int_{\mathcal{X}} \psi_\star(y)|dy \lesssim |\psi|_{\wedge 1}.$$

Proof. By Assumption 2, we have

$$\begin{aligned}
|\mathcal{Q}\psi(x)| &= \Big|\int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\big[\psi(x, z\phi_x(t))B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\big]dtdz\Big| \\
&\leqslant \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\big[\big|\psi(x, z\phi_x(t))\big|\big(b_2\left|\phi_x(t)\right|^\gamma + b_1\big)e^{-b_1 t}\big]dtdz \\
&\leqslant \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\Big[\big|\psi(x, z\phi_x(t))\big|\big(b_2\left|\phi_x(t)\right|^\gamma + b_1\big)\int_t^\infty b_1 e^{-b_1 s}ds\Big]dtdz.
\end{aligned}$$

Next, by Fubini's theorem and the occupation times formula, we derive

$$\begin{aligned}|\mathcal{Q}\psi(x)| &\leqslant \int_0^1 \kappa(z)\int_0^\infty b_1 e^{-b_1 s}\int_0^s \mathbb{E}\big[|\psi(x,z\phi_x(t))|(b_2|\phi_x(t)|^\gamma+b_1)\big]dtdsdz\\ &\leqslant \int_0^1 \kappa(z)\int_0^\infty b_1 e^{-b_1 s}\int_{\mathbb{R}} |\psi(x,zy)|(b_2|y|^\gamma+b_1)\sigma(y)^{-2}\mathbb{E}[L_s^y(\phi_x)]dydsdz\\ &= \int_0^1 \kappa(z)z^{-1}\int_0^\infty b_1 e^{-b_1 s}\int_{\mathbb{R}} |\psi(x,y)|(b_2|y/z|^\gamma+b_1)\sigma(y/z)^{-2}\mathbb{E}[L_s^{y/z}(\phi_x)]dydsdz\\ &\leqslant c_{\mathrm{supp}(\psi_\star)}\int_{\mathbb{R}}|\psi(x,y)|dy,\end{aligned}$$

and (3.18) is proved with

$$c_{\mathrm{supp}(\psi_\star)} = \sup_{y\in\mathrm{supp}(\psi_\star),x\in\mathcal{X},z\in[\varepsilon,1-\varepsilon]} b_1(b_2|y/z|^\gamma+b_1)\sigma(y/z)^{-2}z^{-1}\int_0^\infty e^{-b_1 s}\mathbb{E}[L_s^{y/z}(\phi_x)]ds.$$

This last quantity is finite by Lemma 3.4.3. For the second estimate, we have

$$\begin{aligned}|\mathcal{P}(\psi\otimes\psi)(x)| &\leqslant \int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\big[\big|\psi\big(x,z\phi_x(t)\big)\psi\big(x,(1-z)\phi_x(t)\big)\big|B(\phi_x(t))e^{-\int_0^t B(\phi_x(s))ds}\big]dtdz\\ &\leqslant \big|\mathcal{Q}\psi(x)\big|\sup_y\big|\psi(x,y)\big|\end{aligned}$$

and we conclude by applying (3.18). ☐

Completion of proof of Theorem 3.2.3

Without loss of generality, we may assume that $\nu(\mathcal{Q}\psi)=0$, the general case being obtained by considering the function $\psi(x,y)-\nu(\mathcal{Q}\psi)$. Of course, the compact support property is lost by adding a constant and one has to be careful when revisiting the estimates of *Step 1)* to *Step 4)* below. They exhibit additional error terms that all have the right order using Lemma 3.4.4 and the fact that $\mathcal{P}\mathbf{1}=\mathcal{Q}\mathbf{1}=\mathbf{1}$.

Step 1). We start with a standard preliminary decomposition, see for instance [BPDG14, BPEBG17], expanding the sum in $u\in\mathbb{T}_n^\star$. We have

$$\begin{aligned}\mathbb{E}[\mathcal{M}_n(\psi)^2] &= |\mathbb{T}_n^\star|^{-2}\mathbb{E}_\mu\big[\big(\sum_{m=1}^n\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\big)^2\big]\\ &\leqslant |\mathbb{T}_n^\star|^{-2}\Big(\sum_{m=1}^n\Big(\mathbb{E}_\mu\big[\big(\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\big)^2\big]\Big)^{1/2}\Big)^2\end{aligned}$$

by triangle inequality. Thus Theorem 3.2.3 amounts to control

$$\mathbb{E}_\mu\big[\big(\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\big)^2\big] = I_m + II_m,$$

with

$$\begin{aligned}I_m &= \mathbb{E}_\mu\big[\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)^2\big],\\ II_m &= \mathbb{E}_\mu\big[\sum_{u,v\in\mathbb{G}_m,u\neq v}\psi(X_{u^-},X_u)\psi(X_{v^-},X_v)\big].\end{aligned}$$

Step 2). The control of the term I_m is straightforward: by Lemma 3.4.4 we have

$$\begin{aligned} I_1 &= 2\mu(\mathcal{Q}\psi^2) \leqslant 2c_{\mathrm{supp}(\psi_\star)} \int_{\mathcal{X}\times\mathcal{X}} \psi(x,y)^2 \mu(dx)dy \ \text{ for } \ m=1, \\ I_m &= 2^m \mu(\mathcal{Q}^m \psi^2) \leqslant 2^m \big(c^2_{\mathrm{supp}(\psi_\star)} |\psi^2|_1 \wedge c_{\mathrm{supp}(\psi_\star)} |\psi_\star^2|_1\big) \ \text{ for } \ m \geqslant 2, \end{aligned}$$

therefore $I_m \lesssim 2^m |\psi^2|_\mu$ holds for every $m \geqslant 1$. In the case $\nu(Q\psi) \neq 0$, we replace $|\psi^2|_\mu$ by $|\psi^2|_{\mu+\nu}$.

Step 3). We further decompose the main term $II_m = III_m + IV_m$, having

$$\begin{aligned} III_m &= \mathbb{E}_\mu\Big[\sum_{w\in\mathbb{G}_{m-1}} \psi(X_w, X_{w0})\psi(X_w, X_{w1})\Big], \\ IV_m &= \mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v}\ \sum_{i,j=0,1} \psi(X_u, X_{ui})\psi(X_v, X_{vj})\Big]. \end{aligned}$$

The control of III_m is straightforward:

$$III_m = \mathbb{E}_\mu\Big[\sum_{w\in\mathbb{G}_{m-1}} \mathcal{P}(\psi\otimes\psi)(X_w)\Big] = 2^{m-1}\mu\big(\mathcal{Q}^{m-1}\mathcal{P}(\psi\otimes\psi)\big).$$

In the same way as for the term I_m, by Lemma 3.4.4, one readily checks that $|III_m| \lesssim 2^{m-1}|\psi^\star\psi|_\mu$.

Step 4). We now turn to the main term IV_m. Writing here $u\wedge v$ for the most common recent ancestor of u and v, conditioning w.r.t. $\mathcal{F}_{|u\wedge v|+1}$ and using the conditional independence of (X_u, X_{ui}) and (X_v, X_{vj}) given $\mathcal{F}_{|u\wedge v|+1}$ thanks to the BMC property (3.3), we successively obtain

$$\begin{aligned} IV_m &= \mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v}\ \sum_{i,j=0,1} \mathbb{E}_\mu\left[\psi(X_u, X_{ui})\psi(X_v, X_{vj})\big|\mathcal{F}_{|u\wedge v|+1}\right]\Big] \\ &= \mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v}\ \sum_{i,j=0,1} \mathbb{E}_\mu\left[\psi(X_u, X_{ui})\big|\mathcal{F}_{|u\wedge v|+1}\right]\mathbb{E}_\mu\left[\psi(X_v, X_{vj})\big|\mathcal{F}_{|u\wedge v|+1}\right]\Big] \\ &= \mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v}\ \sum_{i,j=0,1} \mathbb{E}_\mu\left[\mathcal{Q}\psi(X_u)\big|\mathcal{F}_{|u\wedge v|+1}\right]\mathbb{E}_\mu\left[\mathcal{Q}\psi(X_v)\big|\mathcal{F}_{|u\wedge v|+1}\right]\Big] \\ &= 4\,\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v} \mathcal{Q}^{|u|-|u\wedge v|}\psi(X_{u^*})\mathcal{Q}^{|u|-|u\wedge v|}\psi(X_{v^*})\Big], \end{aligned}$$

where u^* (respectively v^*) is the descendant of $u\wedge v$ which is an ancestor of u (respectively v). Conditioning further w.r.t. $\mathcal{F}_{|u\wedge v|}$ we obtain

$$\begin{aligned} IV_m &= 4\,\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v} \mathcal{P}\big(\mathcal{Q}^{|u|-|u\wedge v|}\psi\otimes\mathcal{Q}^{|u|-|u\wedge v|}\psi\big)(X_{u\wedge v})\Big] \\ &= 4\sum_{u,v\in\mathbb{G}_{m-1}, u\neq v} \mu\big(\mathcal{Q}^{|u\wedge v|}\mathcal{P}(\mathcal{Q}^{|u|-|u\wedge v|}\psi\otimes\mathcal{Q}^{|u|-|u\wedge v|}\psi)\big) \\ &= 4\sum_{l=1}^{m-1}\ \sum_{u,v\in\mathbb{G}_{m-1}, |u\wedge v|=m-l-1} \mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi\otimes\mathcal{Q}^l\psi)\big), \end{aligned}$$

obtaining the last term by rearranging the sum $u,v\in\mathbb{G}_{m-1}$ that expands over indices $|u|-|u\wedge v| = m-1-|u\wedge v|$ that vary from 1 to $m-1$. Notice also that for $l=1,\ldots,m-1$,

$$\big|\{u\neq v\in\mathbb{G}_{m-1}, |u\wedge v| = m-l-1\}\big| = |\mathbb{G}_{m-l-1}| \times 2^l \times 2^{l-1},$$

where $|\mathbb{G}_{m-l-1}|$ is the number of common ancestors, 2^l is the number of choices for u (the first descendant in generation $m-1$ of the ancestor from generation $m-l-1$) and 2^{l-1} is the number choices of v (the second descendant, satisfying $u \wedge v \in \mathbb{G}_{m-1}$). We finally obtain

$$IV_m = 4 \sum_{l=1}^{m-1} 2^{m+l-2} \mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi \otimes \mathcal{Q}^l\psi)\big).$$

By Lemma 3.4.4 and Theorem 3.2.2 one easily obtains the following estimates:

$$|\mu\left(\mathcal{Q}^{m-2}\mathcal{P}\left(\mathcal{Q}\psi \otimes \mathcal{Q}\psi\right)\right)| \lesssim |\psi_\star|_1 |\psi|_{\wedge 1} \ \text{ for } \ l = 1,$$

and for $l \geqslant 2$:

$$|\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi \otimes \mathcal{Q}^l\psi)\big)| \lesssim |\psi|_{\wedge 1}^2 \ \wedge \ \rho^{2(l-1)} |\psi_\star|_1^2 \, \mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}((1+V)\otimes(1+V))\big).$$

In the case $\nu(Q\psi) \neq 0$, we replace $|\psi|_{\wedge 1}$ by $|\psi|_\nu$. We claim that

$$\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}((1+V)\otimes(1+V))\big) \lesssim 1 + \mu(V^2) \tag{3.19}$$

and postpone the proof of (3.19) to Step 6 below. It follows that for any $\ell \geqslant 1$:

$$\begin{aligned}
|IV_m| &\lesssim 2^m |\psi_\star|_1 |\psi|_{\wedge 1} + 2^m \sum_{l=1}^{\infty} 2^\ell \big(|\psi|_{\wedge 1}^2 \ \wedge \ \rho^{2(l-1)} |\psi_\star|_1^2 \big(1 + \mu(V^2)\big)\big) \\
&\leqslant 2^m |\psi_\star|_1 |\psi|_{\wedge 1} + 2^m \Big(|\psi|_{\wedge 1}^2 \sum_{l=1}^{\ell} 2^l + |\psi_\star|_1^2 \big(1+\mu(V^2)\big) \sum_{l=\ell+1}^{\infty} 2^{-l}(2\rho)^{2l}\Big) \\
&\leqslant 2^m |\psi_\star|_1 |\psi|_{\wedge 1} + 2^m \inf_{\ell \geqslant 1} \Big(|\psi|_{\wedge 1}^2 2^\ell + |\psi_\star|_1^2 \big(1+\mu(V^2)\big) 2^{-\ell}\Big) \\
&\lesssim 2^m \big(1 + \mu(V)^{1/2}\big) |\psi_\star|_1 |\psi|_{\wedge 1}.
\end{aligned}$$

Step 5). Putting together the estimates obtained for I_m in Step 2, III_m in Step 3 and IV_m in Step 5, and recalling $II_m = III_m + IV_m$ we eventually derive:

$$\mathbb{E}_\mu\big[\big(\sum_{u \in \mathbb{G}_m} \psi(X_{u^-}, X_u)\big)^2\big] \lesssim 2^m \big(|\psi^2|_\mu + |\psi^\star \psi|_\mu + \big(1 + \mu(V)^{1/2}\big)|\psi_\star|_1 |\psi|_{\wedge 1}\big).$$

In the case $\nu(Q\psi) \neq 0$, we replace $|\psi^2|_\mu$ by $|\psi^2|_{\mu+\nu}$ and $|\psi|_{\wedge 1}$ by $|\psi|_\nu$ as follows from Step 2 and 4. Taking square root, summing in $1 \leqslant m \leqslant n$, taking square again and normalizing by $|\mathbb{T}_n^\star|^{-2}$ we obtain Theorem 3.2.3.

Step 6). It remains to establish (3.19). We only sketch the argument which is similar to the proof of Proposition 3.4.1. First, one easily obtains

$$\mathcal{P}((1+V)\otimes(1+V)) \lesssim 1 + \mathcal{Q}V(x) + QV^2(x),$$

and it follows that

$$\begin{aligned}
\mu(\mathcal{Q}^{m-l-1}\mathcal{P}\big((1+V)\otimes(1+V)\big)\big) &\lesssim 1 + \mu(\mathcal{Q}^{m-l}V) + \mu(\mathcal{Q}^{m-l}V^2) \\
&\lesssim 1 + \mu(V) + \mu(\mathcal{Q}^{m-l}V^2)
\end{aligned}$$

by Proposition 3.4.1. Applying Itô's formula and using Assumptions 1 and 2 on can check that

$$\mathcal{Q}V^2(x) \lesssim 1 + V(x)^2 + \mathcal{Q}V(x) \lesssim 1 + V(x)^2$$

by Proposition 3.4.1 again. We obtain (3.19) by integrating w.r.t. μ.

3.4.4 Proof of Theorem 3.3.3

Preparations

We first establish local estimates on the invariant density ν.

Lemma 3.4.5. *Work under Assumptions 1, 2 and 3. Let $\mathcal{Q} \in \mathbb{Q}$ and let ν be the associated invariant density of Theorem 3.2.2. Let $x_0 \in \mathcal{X}$. There exist positive constants $c_i = c_i(x_0, \mathcal{Q})$ and a bounded neighborhood $\mathcal{V}_{x_0}$ with non-empty interior such that*

$$0 < c_1 \leqslant \inf_{x \in \mathcal{V}_{x_0}} \nu(x) \leqslant \sup_{x \in \mathcal{V}_{x_0}} \nu(x) \leqslant c_2.$$

Moreover, $0 < \inf_{\mathcal{Q} \in \mathbb{Q}} c_1(x_0, \mathcal{Q}) \leqslant \sup_{\mathcal{Q} \in \mathbb{Q}} c_2(x_0, \mathcal{Q}) < \infty$.

Proof. Let $\mathcal{V}_{x_0} = [a, b] \subset \mathcal{X}$ be a bounded neighborhood of x_0 and

$$\mathcal{V}_{x_0}^{\varepsilon} = \big[a/(1-\varepsilon) \wedge a/\varepsilon,\ b/\varepsilon \vee b/(1-\varepsilon)\big].$$

By Proposition 3.2.1, using Assumptions 1, 2 and 3, for every $x \in \mathcal{V}_{x_0}$, we successively obtain

$$\begin{aligned}
\nu(x) &= \int_{\mathcal{X}} \nu(y) q(y, x) dy \\
&= \int_{\mathcal{X}} \nu(y) \int_0^1 \frac{\kappa(z)}{z} B(x/z) \sigma(x/z)^{-2} \mathbb{E}\Big[\int_0^{\infty} e^{-\int_0^t B(\phi_y(s))ds} dL_t^{x/z}(\phi_y)\Big] dz dy \\
&\leqslant (b_2 |x|^{\gamma} + b_1) \varepsilon^{-(1+\gamma)} \sigma_1^{-2} \int_{\mathcal{X}} \nu(y) \sup_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\Big[\int_0^{\infty} e^{-b_1 t} dL_t^x(\phi_y)\Big] dy \\
&= (b_2 |x|^{\gamma} + b_1) \varepsilon^{-(1+\gamma)} \sigma_1^{-2} b_1 \int_{\mathcal{X}} \nu(y) \int_0^{\infty} e^{-b_1 t} \sup_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\big[L_t^x(\phi_y)\big] dt dy.
\end{aligned}$$

By Lemma 3.4.3, we have $\sup_{x \in \varepsilon \mathcal{V}_{x_0}} \mathbb{E}\big[L_t^x(\phi_y)\big] \lesssim 1 + t^{3/2}$ uniformly over $\mathbb{Q}$ and the first part of the lemma easily follows. For the second part of the lemma, we have

$$\begin{aligned}
\nu(x) &\geqslant b_1 \sigma_2^{-2} \int_{\mathcal{X}} \nu(y) \inf_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\Big[\int_0^{\infty} e^{-b_2 \int_0^t |\phi_y(s)|^{\gamma} ds - b_1 t} dL_t^x(\phi_y)\Big] dy \\
&\geqslant b_1 \sigma_2^{-2} \int_{[-N,N] \cap \mathcal{X}} \nu(y) \inf_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\Big[\int_0^{T} e^{-(b_2 M^{\gamma} + b_1)t} dL_t^x(\phi_y) \mathbf{1}_{\{\sup_{t \leqslant T} |\phi_y(t)| \leqslant M\}}\Big] dy \\
&\geqslant b_1 \sigma_2^{-2} e^{-(b_2 M^{\gamma} + b_1)T} \int_{[-N,N] \cap \mathcal{X}} \nu(y) \inf_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\big[L_T^x(\phi_y) \mathbf{1}_{\{\sup_{t \leqslant T} |\phi_y(t)| \leqslant M\}}\big] dy
\end{aligned}$$

for arbitrary constants $M, N, T > 0$. Since $\mathbb{E}\big[L_T^x(\phi_y) \mathbf{1}_{\{\sup_{t \leqslant T} |\phi_y(t)| \leqslant M\}}\big] \uparrow \mathbb{E}\big[L_T^x(\phi_y)\big]$ uniformly in $(x, y) \in \mathcal{V}_{x_0} \times [-N, N]$ as M grows, pick M large enough so that for every $y \in [-N, N] \cap \mathcal{X}$, we have

$$\inf_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\big[L_T^x(\phi_y) \mathbf{1}_{\{\sup_{t \leqslant T} |\phi_y(t)| \leqslant M\}}\big] \geqslant \tfrac{1}{2} \inf_{x \in \mathcal{V}_{x_0}^{\varepsilon}} \mathbb{E}\big[L_T^x(\phi_y)\big].$$

Next, we use the fact that Assumption 1 implies that the law of the random variable $\phi_y(t)$ admits a density $\rho_t(y, x)$ w.r.t. the Lebesgue measure and that this density is bounded away from zero on compact sets in (y, x), see for instance [DCFZ86, GCJ93]. In turn $\mathbb{E}\big[L_T^x(\phi_y)\big] = \int_0^T \rho_t(y, x) dt \geqslant \tau_T > 0$ for some τ_T depending also on M and N and we infer

$$\nu(x) \geqslant \frac{\tau_T}{2} b_1 \sigma_2^{-2} e^{-(b_2 M^{\gamma} + b_1)T} \int_{[-N,N] \cap \mathcal{X}} \nu(y) dy$$

and we obtain the result by taking N sufficiently large. The proof is complete. ☐

Completion of proof of Theorem 3.3.3

Step 1). Write $\widehat{\nu}_n(y_0) - \nu(y_0) = I + II$, with

$$I = \frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} G_h(X_u - y_0) - \nu(\mathcal{Q}G_h(\cdot - y_0)) \ \text{ and } \ II = G_h \star \nu(y_0) - \nu(y_0).$$

We plan to apply Theorem 3.2.3 to I with $\psi(x,y) = \varphi(y) = G_h(y-y_0)$. By Lemma 3.4.5, ν is locally bounded and we easily check that

$$|\psi^2|_{\mu+\nu} \lesssim \int_{\mathcal{X}} |G_h(y-y_0)^2|dy = h^{-1}\int_{\mathcal{X}} G(y)^2 dy \lesssim h^{-1},$$

$$|\psi^\star\psi|_\mu \lesssim \sup_y |G_h(y)| \int_{\mathcal{X}} |G_h(y-y_0)|dy \lesssim h^{-1},$$

and

$$|\psi_\star|_1|\psi|_{\wedge 1} \lesssim \Big(\int_{\mathcal{X}} |G_h(y-y_0)|dy\Big)^2 \lesssim 1.$$

Therefore, by Theorem 3.2.3, we have $\mathbb{E}_\mu[I^2] \lesssim |\mathbb{T}_n|^{-1}h^{-1}$ and this term is of order $|\mathbb{T}_n|^{-2\beta/(2\beta+1)}$ from the choice of h. For the term II, Lemma 3.4.5 and the representation $\nu(x) = \int_{\mathcal{X}} \nu(y)q(y,x)dy$ show that $\nu \in \mathcal{H}^\beta(y_0)$ as soon as $q \in \mathcal{H}^{\alpha,\beta}(x_0,y_0)$. Then, by classical kernel approximation (see e.g. the textbook by Tsybakov [Tsy04]) we have that $II^2 \lesssim h^{2\beta}$ since the order k of the kernel G satisfies $k > \beta$, and thus II^2 the same order as I^2 from the choice of h.

Step 2). For the estimation of $q(x_0,y_0)$, write

$$\widehat{q}_n(x_0,y_0) - q(x_0,y_0) = I + II,$$

with

$$I = \frac{\mathcal{M}_n\big(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\big) - \nu(x_0)q(x_0,y_0)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n},$$

and

$$II = \frac{q(x_0,y_0)\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n\big)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n}.$$

We have $\big|I\big| \leqslant III + IV$, with

$$III = \varpi_n^{-1}\big|\mathcal{M}_n\big(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\big) - G^{\otimes 2}_{h_1,h_2} \star \nu(x_0)q(x_0,y_0)\big|$$

and

$$IV = \varpi_n^{-1}\big|G^{\otimes 2}_{h_1,h_2} \star \nu(x_0)q(x_0,y_0) - \nu(x_0)q(x_0,y_0)\big|.$$

We plan to apply Theorem 3.2.3 to bound III with $\psi(x,y) = G_{h_1,h_2}(x-x_0, y-y_0)$. Using Lemma 3.4.5 and the fact that μ is absolutely continuous, we have $|\psi|_{\mu+\nu} \lesssim |\psi|_1$. It readily follows that

$$|\psi^2|_{\mu+\nu} \lesssim |G_{h_1}(\cdot - x_0)^2|_1 |G_{h_2}(\cdot - y_0)^2|_1 \lesssim h_1^{-1}h_2^{-1},$$

$$\begin{aligned}|\psi^\star\psi|_\mu &\lesssim |G_{h_1}(\cdot - x_0)\sup_y |G_{h_2}(y-y_0)| G_{h_1}(\cdot - x_0)G_{h_2}(\cdot - y_0)|_1 \\ &= |G_{h_1}(\cdot - x_0)^2|_1 \sup_y |G_{h_2}(y-y_0)| |G_{h_2}(\cdot - y_0)|_1 \lesssim h_1^{-1}h_2^{-1},\end{aligned}$$

and

$$|\psi_\star|_1|\psi|_1 \lesssim \sup_x |G_{h_1}(x-x_0)||G_{h_2}(\cdot-y_0)|_1^2|G_{h_1}(\cdot-x_0)|_1 \lesssim h_1^{-1}.$$

We conclude

$$\mathbb{E}_\mu\big[III^2\big] \lesssim \varpi_n^{-2}|\mathbb{T}_n|^{-1}h_1^{-1}h_2^{-1},$$

and this term has order $\varpi_n^{-2}|\mathbb{T}_n|^{-2s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$ from the choice of h_1 and h_2. By kernel approximation and the fact that G has order $k > \max\{\alpha,\beta\}$, noting that $(x,y)\mapsto \mu(x)q(x,y)\in\mathcal{H}^{\alpha\wedge\beta,\beta}$, we have

$$|IV| \lesssim h_1^{\alpha\wedge\beta}+h_2^{\beta} \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$$

from the choice of h_1, h_2.

We turn to the term II. We plan to use

$$\big(\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\vee\varpi_n\big)^2 \lesssim \big(\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\big)^2+\mathbf{1}_{\left\{\mathcal{M}_n(G_h(\cdot-x_0))<\varpi_n\right\}}.$$

Pick n large enough so that $0<\varpi_n\leqslant\tau(x_0)=\frac{1}{2}\inf_{\mathcal{Q}\in\mathbb{Q},x\in\mathcal{V}_{x_0}}\nu(x)$, a choice which is possible by Lemma 3.4.5. Since $\{\mathcal{M}_n(G_h(\cdot-x_0))<\varpi_n\}\subset\{\mathcal{M}_n(G_h(\cdot-x_0))-\nu(x_0)<-\tau(x_0)\}$, we further infer

$$\begin{aligned}&\mathbb{E}_\mu\big[\big(\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\vee\varpi_n\big)^2\big]\\ \leqslant&\,\mathbb{E}_\mu\big[\big(\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\big)^2\big]+\mathbb{P}_\mu\big(\big|\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\big|\geqslant\tau(x_0)\big)\\ \lesssim&\,\mathbb{E}_\mu\big[\big(\nu(x_0)-\mathcal{M}_n\big(G_h(\cdot-x_0)\big)\big)^2\big].\end{aligned}$$

Applying *Step 1)* of the proof, we derive

$$\mathbb{E}_\mu\big[II^2\big]\lesssim\varpi_n^{-2}|\mathbb{T}_n|^{-2\beta/(2\beta+1)}$$

and this term has negligible order. The proof of Theorem 3.3.3 is complete.

3.4.5 Proof of Proposition 3.3.6

Let $s(x)=\int_0^x\exp\Big(-2\int_0^y\frac{r(z)}{\sigma(z)^2}dz\Big)$ and $m(x)=\frac{2}{\sigma(x)^2s'(x)}$. Consider the infinitesimal generator $\mathcal{L}$ associated to the diffusion process (3.11), written in its divergence form

$$\mathcal{L}f(x)=\frac{1}{m(x)}\frac{d}{dx}\Big(\frac{1}{s(x)}\frac{d}{dx}f(x)\Big),\ \ f\in\mathcal{D}(\mathcal{L}),$$

with domain $\mathcal{D}(\mathcal{L})$ densely defined on twice continuously differentiable functions f satisfying the boundary condition $f'(0)=f'(L)=0$. By Itô formula and Fubini's theorem, for $f\in\mathcal{D}(\mathcal{L})$, we have

$$\begin{aligned}\int_{\mathcal{X}}f(y)q(x,y)dy&=-\int_0^1\kappa(z)\mathbb{E}\Big[\int_0^\infty f\left(z\phi_x(t)\right)\frac{d}{dt}e^{-\int_0^t B(\phi_x(s))ds}dt\Big]dz\\&=\int_0^1\kappa(z)f_z(x)dz+\int_0^1\kappa(z)\mathbb{E}\Big[\int_0^\infty\mathcal{L}f_z\left(\phi_x(t)\right)e^{-\int_0^t B(\phi_x(s))ds}dt\Big]dz,\end{aligned}$$

where we set $f_z(x) = f(zx)$ for $z \in [\varepsilon, 1-\varepsilon]$ since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$ by Assumption 3. Pick $f(x) = \int_0^x \exp\big(2\varepsilon^{-1}\int_0^y e^{2\int_0^{u\varepsilon^{-1}} \frac{|r(v)|}{\sigma(v)^2}dv} s(u)^{-1}du\big)dy$, and note that

$$f''(zx) = \frac{2}{\varepsilon s(zx)} \exp\Big(2\int_0^{zx\varepsilon^{-1}} \frac{|r(v)|}{\sigma(v)^2}dv\Big) f'(zx).$$

It follows that for $z \geqslant \varepsilon$ and every $x \in \mathcal{X}$, we have

$$\begin{aligned}\mathcal{L}f_z(x) &= \frac{z}{m(x)} \frac{zs(x)f''(zx) - s'(x)f'(zx)}{s^2(x)} \\ &= \frac{zf'(zx)}{m(x)s(x)^2}\Big(\frac{2zs(x)}{\varepsilon s(zx)} e^{2\int_0^{zx\varepsilon^{-1}} \frac{|r(v)|}{\sigma(v)^2}dv} - e^{-2\int_0^x \frac{r(v)}{\sigma(v)^2}dv}\Big) > 0.\end{aligned}$$

Now let $B_1, B_2 : \mathcal{X} \to [0,\infty)$ be two functions in an orderly class $\mathcal{B}$ according to Definition 3.3.5 and write q_{B_1} and q_{B_2} for the associated transition densities. With no loss of generality, we may (and will) assume that $B_1(x) \leqslant B_2(x)$ for every $x \in \mathcal{X}$. Assume that $q_{B_1} = q_{B_2}$. Since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, we have

$$\begin{aligned}&\int_{\mathcal{X}} f(y)\big(q_{B_1}(x,y) - q_{B_2}(x,y)\big)dy \\ &= \int_\varepsilon^{1-\varepsilon} \kappa(z)\mathbb{E}\Big[\int_0^\infty \mathcal{L}f_z(\phi_x(t))\big(e^{-\int_0^t B_1(\phi_x(s))ds} - e^{-\int_0^t B_2(\phi_x(s))ds}\big)\Big]dtdz = 0.\end{aligned}$$

Our choice of f and the property $B_1 \leqslant B_2$ implies that the integrand is non-negative. It follows that

$$\kappa(z)\big(e^{-\int_0^t B_1(\phi_x(s))ds} - e^{-\int_0^t B_2(\phi_x(s))ds}\big) = 0$$

$dzdt \otimes \mathbb{P}$-a.s. Picking z such that $\kappa(z) > 0$, we obtain $\int_0^t B_1(\phi_x(s))ds = \int_0^t B_2(\phi_x(s))ds$ $\mathbb{P}$-a.s. for every $t \geqslant 0$ by continuity of the integrand in t. By the occupation times formula, it follows that $\int_X \big(B_1(y) - B_2(y)\big)L_t^y(\phi_x)dy = 0$, almost-surely, and by the ordering property, $B_1(y) = B_2(y)$ for every y such that $L_t^y(\phi_x) > 0$, i.e. for $y \in [\inf_{0\leqslant s\leqslant t}\phi_x(s), \sup_{0\leqslant s\leqslant t}\phi_x(s)] \to \mathcal{X}$ as $t \to \infty$. The proof of Proposition 3.3.6 is complete.

3.4.6 Proof of Theorem 3.3.7

Preparation for the proof

We first establish uniform bounds for $q_\vartheta(x,y)$. Remember that in the reflected case, we have $\mathcal{X} = [0, L]$ and $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$ under Assumption 3.

Lemma 3.4.6. *Work under Assumptions 1, 3 and 5. For sufficiently small $\eta > 0$, we have:*

$$0 < \inf_{x\in\mathcal{X}, y\in\mathcal{X}_\eta, \vartheta\in\Theta} q_\vartheta(x,y) \leqslant \sup_{x,y\in\mathcal{X},\vartheta\in\Theta} q_\vartheta(x,y) < \infty,$$

where $\mathcal{X}_\eta = [0, (1-\varepsilon)L - \eta]$.

Proof. The proof is close to that of Lemma 3.4.5. Let $x \in \mathcal{X}$ and $y \in \mathcal{X}_\eta$. We have

$$\begin{aligned}\inf_{\vartheta\in\Theta} q_\vartheta(x,y) &\geqslant b_3\sigma_2^{-2}\int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z}\mathbb{E}\Big[\int_0^\infty e^{-b_4 t}dL_t^{y/z}(\phi_x)\Big]dz \\ &\geqslant b_3\sigma_2^{-2}b_4\int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z}\int_0^\infty e^{-b_4 t}\mathbb{E}\big[L_t^{y/z}(\phi_x)\big]dtdz \\ &\geqslant (1-\varepsilon)^{-1}b_3\sigma_2^{-2}b_4\frac{\eta}{L}\int_0^\infty e^{-b_4 t}\inf_{x,y\in\mathcal{X}}\mathbb{E}\big[L_t^y(\phi_x)\big]dt\end{aligned}$$

According to [Cat92], Section 5, proof of Lemma 5.37, the law of $\phi_x(t)$ is absolutely continuous with density with $y \mapsto \rho_t(x,y)$ that can be taken continuous and that satisfies $\inf_{x,y\in\mathcal{X}} \rho_t(x,y) > 0$ for every $t > 0$. Therefore

$$\inf_{x,y\in\mathcal{X}} \mathbb{E}\big[L_t^y(\phi_x)\big] = \inf_{x,y\in\mathcal{X}} \int_0^t \rho_s(x,y)ds \geqslant \tau_t > 0$$

and the result follows. The upper bound readily follows from

$$\begin{aligned}\sup_{\vartheta\in\Theta} q_\vartheta(x,y) &\leqslant b_4\sigma_1^{-2}\int_\varepsilon^{1-\varepsilon}\frac{\kappa(z)}{z}dz \sup_{x,y\in\mathcal{X}}\mathbb{E}\big[\int_0^\infty e^{-b_3t}dL_t^y(\phi_x)\big] \\ &\leqslant \varepsilon^{-1}b_4\sigma_1^{-2}\sup_{x,y\in\mathcal{X}}\mathbb{E}\big[\int_0^\infty e^{-b_3t}dL_t^y(\phi_x)\big] \\ &= \varepsilon^{-1}b_4\sigma_1^{-2}b_3\sup_{x,y\in\mathcal{X}}\int_0^\infty e^{-b_3t}\mathbb{E}\big[L_t^y(\phi_x)\big]dt\end{aligned}$$

which is finite by Lemma 3.4.3. □

Completion of proof of Theorem 3.3.7

This proof is classical (see for instance van der Vaart [vdV98] Theorem 5.14). We nevertheless give it for self-containedness. For $a \in \Theta$, let

$$\mathcal{M}(a,\vartheta) = \int_{\mathcal{X}}\nu_\vartheta(dx)\int_{\mathcal{X}}\log q_a(x,y)q_\vartheta(x,y)dy.$$

First, $a \mapsto \mathcal{M}(a,\vartheta)$ has a unique maximum at $a = \vartheta$, as stems from the inequality $\log(x) \leqslant 2(\sqrt{x}-1)$ for $x \geqslant 0$. Indeed

$$\begin{aligned}\mathcal{M}(a,\vartheta)-\mathcal{M}(\vartheta,\vartheta) &= \int_{\mathcal{X}}\nu_\vartheta(dx)\int_{\mathcal{X}}\log\frac{q_a(x,y)}{q_\vartheta(x,y)}q_\vartheta(x,y)dy \\ &\leqslant \int_{\mathcal{X}}\nu_\vartheta(dx)\big(\int_{\mathcal{X}}2\sqrt{q_a(x,y)}\sqrt{q_\vartheta(x,y)}dy-2\big) \\ &\leqslant -\int_{\mathcal{X}}\nu_\vartheta(dx)\int_{\mathcal{X}}\Big(\sqrt{q_a(x,y)}-\sqrt{q_\vartheta(x,y)}\Big)^2dy \leqslant 0.\end{aligned}$$

Next, writing $m_{\mathcal{U}}(x,y) = \sup_{a\in\mathcal{U}}\log q_a(x,y)$, we prove that for every $a \neq \vartheta \in \Theta$, there exists a neighborhood $\mathcal{U}_a$ of a such that:

$$\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_a}\big) < \nu_\vartheta\big(\mathcal{Q}_\vartheta\log q_\vartheta\big) = \mathcal{M}(\vartheta,\vartheta). \tag{3.20}$$

Pick a decreasing sequence of open balls $(\mathcal{U}_\ell(a))_{\ell\geqslant1}$ around a with vanishing diameters. For every $x,y \in \mathcal{X}$ we have $m_{\mathcal{U}_\ell(a)}(x,y) \downarrow \log q_a(x,y)$ by continuity of $a \mapsto \log q_a(x,y)$ thanks to the continuity of B_0 according to Assumption 5. By Lemma 3.4.6, we also have $\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}}\big) < \infty$ for any $\mathcal{U} \subset \Theta$ therefore

$$\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_l(a)}\big) \downarrow \nu_\vartheta\big(\mathcal{Q}_\vartheta\log q_a\big) = \mathcal{M}(a,\vartheta) \leqslant \mathcal{M}(\vartheta,\vartheta)$$

by monotone convergence with equality only if $a = \vartheta$, and this proves the existence of $\mathcal{U}_a$ such that (3.20) holds. We are now ready to prove the consistency result. For $\eta' > 0$, the compact ball

$$\mathcal{C}_{\eta'}(\vartheta) = \{a \in \Theta, |a-\vartheta| \geqslant \eta'\}$$

can be covered by finitely many open neighbourhoods $\mathcal{U}_{a_1}, \dots \mathcal{U}_{a_p}$ with $a_i \in \mathcal{C}_{\eta'}(\vartheta)$ and such that (3.20) holds for every $\mathcal{U}_{a_i}$. For $\eta > 0$, let

$$m_{\mathcal{U}}^{(\eta)}(x,y) = \sup_{a\in\mathcal{U}} \log q_a(x,y) \mathbf{1}_{\{q_a(x,y)\geqslant \eta\}}.$$

Abbreviating $\mathcal{L}_n(a, (X_u)_{u\in\mathbb{T}_n})$ by $\mathcal{L}_n(a)$, it follows that

$$\begin{aligned} |\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \log\mathcal{L}_n(a) &\leqslant \max_{1\leqslant i\leqslant p} |\mathbb{T}_n^\star|^{-1} \sum_{u\in\mathbb{T}_n^\star} m_{\mathcal{U}_{a_i}}^{(\eta)}(X_{u^-}, X_u) \\ &\to \max_{1\leqslant i\leqslant p} \nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_{a_i}}\big) < \mathcal{M}(\vartheta,\vartheta) \end{aligned} \tag{3.21}$$

in probability as $n \to \infty$ and letting $\eta \to 0$, as stems from Corollary 3.3.4 and the fact that $\sup_{x,y\in\mathcal{X}_\eta} m_{\mathcal{U}_{a_i}}(x,y) < \infty$ by Lemma 3.4.6. Finally, if $\widehat{\vartheta}_n \in \mathcal{C}_{\eta'}(\vartheta)$, then, by definition of $\widehat{\vartheta}_n$, we have

$$|\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \log\mathcal{L}_n(a) \geqslant |\mathbb{T}_n^\star|^{-1} \log\mathcal{L}_n(\widehat{\vartheta}_n) \geqslant |\mathbb{T}_n^\star|^{-1}\log\mathcal{L}_n(\vartheta) > \mathcal{M}(\vartheta,\vartheta) - \epsilon_n,$$

where $\epsilon_n \to 0$ in probability, as follows from Corollary 3.3.4. We conclude the proof by noticing that

$$\big\{\widehat{\vartheta}_n \in \mathcal{C}_{\eta'}(\vartheta)\big\} \subset \big\{|\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \mathcal{L}_n(a) \geqslant \mathcal{M}(\vartheta,\vartheta) - \epsilon_n\big\}$$

and the fact that the probability of this last event converges to 0 by (3.21) as $n \to \infty$.

3.4.7 Proof of Theorem 3.3.9

Preparation for the proof

We start by proving some useful estimates on the gradient and Hessian of $\log q_\vartheta$. Let

$$\Gamma_\vartheta = \nabla_\vartheta \log q_\vartheta = \left(\partial_{\vartheta_1}\log q_\vartheta, \dots, \partial_{\vartheta_d}\log q_\vartheta\right), \quad \Gamma_{\vartheta,i} = \partial_{\vartheta_i}\log q_\vartheta, \quad 1 \leqslant i \leqslant d.$$

Lemma 3.4.7. *Work under Assumptions 1, 3, 5 and 6. For every $1 \leqslant i,j \leqslant d$ and $\eta > 0$, we have*

$$\sup_{x\in\mathcal{X}, y\in\mathcal{X}_\eta, \vartheta\in\Theta} |\Gamma_{\vartheta,i}(x,y)| < \infty, \quad \sup_{x\in\mathcal{X}, y\in\mathcal{X}_\eta, \vartheta\in\Theta} |\partial_\vartheta\Gamma_\vartheta(x,y)_{i,j}| < \infty, \quad \sup_{x\in\mathcal{X}, y\in\mathcal{X}_\eta, \vartheta\in\Theta} \big\|\partial_\vartheta^2\Gamma_\vartheta(x,y)\big\| < \infty$$

where $\|\cdot\|$ corresponds to the operator norm for the Hessian $\partial^2_\vartheta\Gamma_\vartheta(x,y)$.

Proof. According to Lemma 3.4.6, since

$$\Gamma_\vartheta(x,y) = \frac{\partial_\vartheta q_\vartheta(x,y)}{q_\vartheta(x,y)}$$

componentwise, it suffices to show $|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim 1$ in order to establish the first bound. We have:

$$\begin{aligned} \partial_{\vartheta_i} q_\vartheta(x,y) &= \int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z} \partial_{\vartheta_i} B_0(\vartheta, y/z)\sigma(y/z)^{-2}\,\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big] dz \\ &- \int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z} B_0(\vartheta,y/z)\sigma(y/z)^{-2}\,\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} \int_0^t \partial_{\vartheta_i} B_0(\vartheta,\phi_x(s))ds\, dL_t^{y/z}(\phi_x)\Big]. \end{aligned}$$

By Assumption 1, 5 and 6 we readily obtain

$$|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim \int_\varepsilon^{1-\varepsilon} \mathbb{E}\Big[\int_0^\infty (1+t) e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]\kappa(z)dz$$

By Assumption 6 and integration by part, we have

$$\begin{aligned}\mathbb{E}\Big[\int_0^\infty (1+t) e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big] &\leqslant \mathbb{E}\Big[\int_0^\infty (1+t)e^{-b_3 t} dL_t^{y/z}(\phi_x)\Big] \\ &= \int_0^\infty (1-b_3(1+t))e^{-b_3 t}\mathbb{E}[L_t^{y/z}(\phi_x)]dt.\end{aligned}$$

This last term is bounded by Lemma 3.4.3 and $|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim 1$ follows. We turn to the second bound: clearly, for $1 \leqslant i,j \leqslant d$

$$\partial_\vartheta \Gamma_\vartheta(x,y)_{i,j} = \frac{\partial^2_{\vartheta_i\vartheta_j} q_\vartheta(x,y) q_\vartheta(x,y) - \partial_{\vartheta_i} q_\vartheta(x,y)\partial_{\vartheta_j} q_\vartheta(x,y)}{q_\vartheta(x,y)^2}$$

and thanks to Lemma 3.4.6 and the first bound, we only need to show $|\partial^2_{\vartheta_i\vartheta_j} q_\vartheta(x,y)| \lesssim 1$ in order to obtain the second bound. Define

$$\omega_t(y,z,\vartheta) = \mathbf{1}_{\{\varepsilon\vee yL^{-1},1-\varepsilon\}}(z)\frac{\kappa(z)}{z\sigma(y/z)^2}\exp(-\int_0^t B_0(\vartheta,\phi_x(s))ds).$$

We have

$$\begin{aligned}\partial^2_{\vartheta_i\theta_j} q_\vartheta(x,y) &= \int_0^1 \partial^2_{\vartheta_i\vartheta_j} B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta) dL_t^{y/z}(\phi_x)\Big]dz \\ &+ \int_0^1 B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial_{\vartheta_i} B_0(\vartheta,\phi_x(s))ds\int_0^t \partial_{\vartheta_j} B_0(\vartheta,\phi_x(s))ds\, dL_t^{y/z}(\phi_x)\Big]dz \\ &- \sum_{(\ell,\ell')=(i,j),(j,i)}\int_0^1 \partial_{\theta_\ell} B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial_{\theta_{\ell'}} B_0(\vartheta,\phi_x(s))dsdL_t^{y/z}(\phi_x)\Big]dz \\ &+ \int_0^1 B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial^2_{\vartheta_i\vartheta_j} B_0(\vartheta,\phi_x(s))dsdL_t^{y/z}(\phi_x)\Big]dz\end{aligned}$$

and we proceed in the same way as for the first estimate, using repeatedly Assumption 1, 5 and 6. The proof of the third bound is analogous. □

Completion of proof of Theorem 3.3.9

This proof is classical (see for instance van der Vaart [vdV98] Theorem 5.41). We nevertheless give it for self-containedness. By definition of $\widehat{\vartheta}_n$ and a Taylor expansion around ϑ, we have

$$\begin{aligned}0 &= \sum_{u\in\mathbb{T}_n^\star}\Gamma_{\widehat{\vartheta}_n}(X_{u^-},X_u) \\ &= \sum_{u\in\mathbb{T}_n^\star}\Big(\Gamma_\vartheta(X_{u^-},X_u) + \partial_\vartheta\Gamma_\vartheta(X_{u^-},X_u)(\widehat{\vartheta}_n-\vartheta) + (\widehat{\vartheta}_n-\vartheta)^T\partial^2_\vartheta\Gamma_{\widetilde{\vartheta}_n}(X_{u^-},X_u)(\widehat{\vartheta}_n-\vartheta)\Big),\end{aligned}$$

for some $\widetilde{\vartheta}_n$ on the segment line between ϑ and $\widehat{\vartheta}_n$. Rearranging the sum and introducing the normalization $|\mathbb{T}_n^\star|^{1/2}$, we derive

$$\begin{aligned}&\big(|\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}\partial_\vartheta\Gamma_\vartheta(X_{u^-},X_u) + |\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}(\widehat{\vartheta}_n-\vartheta)^T\partial^2_\vartheta\Gamma_{\widetilde{\vartheta}_n}(X_{u^-},X_u)\big)|\mathbb{T}_n^\star|^{1/2}(\widehat{\vartheta}_n-\vartheta) \\ &= -\,|\mathbb{T}_n^\star|^{-1/2}\sum_{u\in\mathbb{T}_n^\star}\Gamma_\vartheta(X_{u^-},X_u).\end{aligned} \tag{3.22}$$

We plan to apply the central limit theorem for bifurcating Markov chain proved by Guyon, see [Guy07] Corollary 24 on the right-hand side. By Lemma 3.4.6 and 3.4.7 we have that $\mathcal{Q}_\vartheta(\Gamma_{\vartheta,i}\Gamma_{\vartheta,j})$ and $\mathcal{Q}_\vartheta(\Gamma_{\vartheta,i}\Gamma_{\vartheta,j}\Gamma_{\vartheta,k}\Gamma_{\vartheta,l})$ are bounded functions on $\mathcal{X}$ for all $1 \leqslant i,j,k,l \leqslant d$. Moreover, we have $\nu_\vartheta(\mathcal{Q}_\vartheta\Gamma_{\vartheta,i}) = 0$. Therefore

$$|\mathbb{T}_n^\star|^{-1/2} \sum_{u\in\mathbb{T}_n^\star} \Gamma_\vartheta(X_{u^-}, X_u) \to \mathcal{N}\big(0, \Psi(\vartheta)\big) \tag{3.23}$$

in distribution as $n \to \infty$, where $\Psi(\vartheta)$ is the Fisher information matrix defined after Assumption 6. Next, since $\partial_\vartheta\Gamma_\vartheta$ is bounded by Lemma 3.4.7, we have

$$|\mathbb{T}_n^\star|^{-1} \sum_{u\in\mathbb{T}_n^\star} \partial_\vartheta\Gamma_\vartheta(X_{u^-}, X_u) \to \Psi(\vartheta) \tag{3.24}$$

in probability as $n \to \infty$. Moreover, by Lemma 3.4.7, we have: $\sup_{x,y\in\mathcal{X},\vartheta\in\Theta} \big\|\partial_\vartheta^2\Gamma_\vartheta(x,y)\big\| < \infty$ and since $\widehat{\vartheta}_n - \vartheta$ converges to 0 by Theorem 3.3.7, it follows that

$$|\mathbb{T}_n^\star|^{-1} \sum_{u\in\mathbb{T}_n^\star} (\widehat{\vartheta}_n - \vartheta)^T \partial_\theta^2\Gamma_{\widetilde{\vartheta}_n}(X_{u^-}, X_u) \to 0 \tag{3.25}$$

in probability as $n \to \infty$ tends to infinity. Combining (3.23), (3.24) and (3.25) in (3.22) we finally obtain

$$\Psi(\vartheta)|\mathbb{T}_n^\star|^{1/2}(\widehat{\vartheta}_n - \vartheta) \to \mathcal{N}\big(0, \Psi(\vartheta)\big)$$

in distribution as $n \to \infty$. We conclude thanks to the invertibility of $\Psi(\vartheta)$ granted by Assumption 7.

3.5 Appendix

3.5.1 Proof of Lemma 3.4.3

Step 1). Fix $\delta > 0$ and let $\mathcal{K}_\delta = \{y \in \mathcal{X}, \inf_{z\in\mathcal{K}} |y - z| \leqslant \delta\}$ denote the δ-enlargement of $\mathcal{K}$. For $x \in \mathcal{X}$, let

$$\tau_x = \inf\{t \geqslant 0, \phi_x(t) \in \mathcal{K}_\delta\}, \ \ \inf\varnothing = \infty,$$

and

$$\phi_x^{\mathcal{K}_\delta}(t) = \begin{cases} \phi_{\sup\mathcal{K}_\delta}\big((t-\tau_x)_+\big) & \text{if} \quad x > \sup\mathcal{K}_\delta \\ \phi_x(t) & \text{if} \quad x \in \mathcal{K}_\delta \\ \phi_{\inf\mathcal{K}_\delta}\big((t-\tau_x)_+\big) & \text{if} \quad x < \inf\mathcal{K}_\delta. \end{cases}$$

For every $y \in \mathcal{K}$, we have $L_t^y(\phi_x) = L_t^y(\phi_x^{\mathcal{K}_\delta})$, and by Itô-Tanaka's formula, it follows that

$$L_t^y(\phi_x) = L_t^y(\phi_x^{\mathcal{K}_\delta}) = \big|\phi_x^{\mathcal{K}_\delta}(t) - y\big| - |\phi_x^{\mathcal{K}_\delta}(0) - y| - \int_0^t \operatorname{sgn}\big(\phi_x^{\mathcal{K}_\delta}(s) - y\big)d\phi_x^{\mathcal{K}_\delta}(s).$$

Assume first that $x > \sup\mathcal{K}_\delta$. Observing that $L_t^y(\phi_x) = 0$ on $\{\tau_x \geqslant t\}$, and that $d\phi_x^{\mathcal{K}_\delta}(s)$ vanishes on $[0, \tau_x)$ on $\{\tau_x < t\}$, we readily have

$$\begin{aligned} L_t^y(\phi_x) &= \big|\phi_x^{\mathcal{K}_\delta}(t) - y\big| - |\sup\mathcal{K}_\delta - y| - \int_{\tau_x\wedge t}^t \operatorname{sgn}\big(\phi_x^{\mathcal{K}_\delta}(s) - y\big)d\phi_x^{\mathcal{K}_\delta}(s) \\ &= \big|\phi_{\sup\mathcal{K}_\delta}\big((t-\tau_x)_+\big) - y\big| - |\sup\mathcal{K}_\delta - y| - \int_0^{(t-\tau_x)_+} \operatorname{sgn}\big(\phi_{\sup\mathcal{K}_\delta}(s) - y\big)d\phi_{\sup\mathcal{K}_\delta}(s). \end{aligned} \tag{3.26}$$

We plan to bound each term separately.

Step 2). By Itô's formula, $(\phi_{\sup \mathcal{K}_\delta}(t)-y)^2 = (\sup \mathcal{K}_\delta - y)^2 + I + II$, with

$$\begin{aligned} I &= \int_0^t \big(2(\phi_{\sup \mathcal{K}_\delta}(s)-y)r(\phi_{\sup \mathcal{K}_\delta}(s)) + \sigma(\phi_{\sup \mathcal{K}_\delta}(s))^2\big)ds, \\ II &= 2\int_0^t (\phi_{\sup \mathcal{K}_\delta}(s)-y)\sigma(\phi_{\sup \mathcal{K}_\delta}(s))dW_s. \end{aligned}$$

First,

$$I \leqslant 2\sigma_1^{-2}\int_{\mathbb{R}} (z-y)r(z)L_t^z(\phi_{\sup \mathcal{K}_\delta})dz + t\sigma_2^2$$

by the occupation times formula and Assumption 1. Introduce $|y|_{r_2} = |y| \vee r_2$, where r_2 is defined in Assumption 1. Since $z-y>0$ and $r(z)<0$ for $z>|y|_{r_2}$, we have $\int_{|y|_{r_2}}^{\infty}(z-y)r(z)L_t^z(\phi_x)dz < 0$. Similarly $\int_{-\infty}^{-|y|_{r_2}}(z-y)r(z)L_t^z(\phi_x)dz < 0$. It follows that

$$\begin{aligned} \int_{\mathbb{R}} (z-y)r(z)L_t^z(\phi_x)dz &\leqslant \int_{-|y|_{r_2}}^{|y|_{r_2}} (z-y)r(z)L_t^z(\phi_x)dz \\ &\leqslant r_1\int_{-|y|_{r_2}}^{|y|_{r_2}} |z-y|(1+|z|)L_t^z(\phi_x)dz \\ &\leqslant r_1(|y|_{r_2}-y)(1+|y|_{r_2})\int_{\mathbb{R}} L_t^z(\phi_x)dz \\ &\leqslant r_1(|y|_{r_2}-y)(1+|y|_{r_2})t, \end{aligned}$$

therefore

$$I \leqslant 2\sigma_1^{-2}r_1(|y|_{r_2}-y)(1+|y|_{r_2})t + \sigma_2^2 t = t\kappa(y)$$

say. Since $\mathbb{E}[II]=0$, we derive by Cauchy-Schwarz's inequality

$$\mathbb{E}\big[|\phi_{\sup \mathcal{K}_\delta}(t)-y|\big] \leqslant \sqrt{(\sup \mathcal{K}_\delta - y)^2 + t\kappa(y)}. \tag{3.27}$$

Step 3). We are ready to control each term of (3.26). We have

$$\begin{aligned} &\mathbb{E}\big[|\phi_{\sup \mathcal{K}_\delta}\big((t-\tau_x)_+\big)-y|\big] \\ &\leqslant |\sup \mathcal{K}_\delta - y| + \mathbb{E}\Big[\int_0^{(t-\tau_x)_+} \big|r\big(\phi_{\sup \mathcal{K}_\delta}(s)\big)\big|ds\Big] + \mathbb{E}\Big[\Big|\int_0^{(t-\tau_x)_+} \sigma\big(\phi_{\sup \mathcal{K}_\delta}(s)\big)dW_s\Big|\Big] \\ &\leqslant |\sup \mathcal{K}_\delta - y| + r_1\mathbb{E}\Big[\int_0^t (1+|\phi_{\sup \mathcal{K}_\delta}(s)|)ds\Big] + \mathbb{E}\Big[\sup_{u\leqslant t}\Big(\int_0^u \sigma\big(\phi_{\sup \mathcal{K}_\delta}(s)\big)dW_s\Big)^2\Big]^{1/2} \\ &\leqslant |\sup \mathcal{K}_\delta - y| + r_1 t + r_1\mathbb{E}\Big[\int_0^t |\phi_{\sup \mathcal{K}_\delta}(s)|ds\Big] + \sqrt{2}\sigma_2 t \\ &\leqslant |\sup \mathcal{K}_\delta - y| + r_1 t + r_1\int_0^t \sqrt{(\sup \mathcal{K}_\delta)^2 + s\kappa(0)}ds + \sqrt{2}\sigma_2 t \\ &\lesssim 1 + t^{3/2}, \end{aligned}$$

where we successively applied Assumption 1, Doob's inequality and (3.27). In the same way

$$\begin{aligned} &\Big| -\int_0^{(t-\tau_x)_+} \operatorname{sgn}\big(\phi_{\sup \mathcal{K}_\delta}(s)-y\big)d\phi_{\sup \mathcal{K}_\delta}(s)\Big| \\ \leqslant & r_1\int_0^t (1+|\phi_{\sup \mathcal{K}_\delta}(s)|)ds + \sup_{u\leqslant t}\Big|\int_0^u \operatorname{sgn}(\phi_{\sup \mathcal{K}_\delta}(s)-y)\sigma\big(\phi_{\sup \mathcal{K}_\delta}(s)\big)dW_s\Big|. \end{aligned}$$

Taking expectation and using the foregoing arguments, this last quantity is also of order $1 + t^{3/2}$ and Lemma 3.4.3 is proved for $x > \sup \mathcal{K}_\delta$.

Step 4). If $x < \inf \mathcal{K}_\delta$, we apply the same arguments, replacing $|\sup \mathcal{K}_\delta|$ by $|\inf \mathcal{K}_\delta|$ with obvious changes. Likewise if $x \in \mathcal{K}_\delta$ we may replace $|\sup \mathcal{K}_\delta|$ by $\max\{|\sup \mathcal{K}_\delta|, |\inf \mathcal{K}_\delta|\}$.

3.5.2 Proof of Proposition 3.3.8

Remember that

$$\Psi(\vartheta) = \nu_\vartheta\Big(\mathcal{Q}_\vartheta\Big(\frac{\partial_\vartheta q_\vartheta}{q_\vartheta}\Big)^2\Big) = \int_{\mathcal{X}} \nu_\vartheta(dx) \int_{\mathcal{X}} \frac{\big(\partial_\vartheta q_\vartheta(x,y)\big)^2}{q_\vartheta(x,y)} dy.$$

If $\mathcal{A} \subset \mathcal{X}$ is a Borel set with $\mathrm{Leb}(\mathcal{A}) > 0$, we have

$$\nu_\vartheta(\mathcal{A}) = \int_{\mathcal{X}\times\mathcal{X}} \mathbf{1}_{\mathcal{A}}(y) q_\vartheta(x,y) \nu_\vartheta(dx) dy \geqslant \inf_{x,y} q_\vartheta(x,y)\, \mathrm{Leb}(\mathcal{A}) > 0$$

since $\inf_{x,y\in\mathcal{X}} q_\vartheta(x,y) > 0$ by Lemma 3.4.6. By continuity of $y \mapsto \partial_\vartheta q_\vartheta(x,y)$ on $[0,L]$, it suffices then to show the existence $x, y \in \mathcal{X}$ such that $\partial_\vartheta q_\vartheta(x,y) > 0$. For $x, y \in \mathcal{X}$, we have

$$\begin{aligned}
\partial_\vartheta q_\vartheta(x,y) &= \int_{\varepsilon \vee y L^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z} \sigma^{-2} \mathbb{E}\Big[\int_0^\infty (1-\vartheta t) e^{-\vartheta t} dL_t^{y/z}(\phi_x)\Big] dz \\
&= \int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} \kappa(y/u) \sigma^{-2} \mathbb{E}\Big[\int_0^\infty (1-\vartheta t) e^{-\vartheta t} dL_t^{u}(\phi_x)\Big] \frac{du}{u} \\
&= \mathbb{E}\Big[\int_0^\infty \kappa\big(y/\phi_x(t)\big)(1-\vartheta t) e^{-\vartheta t} \mathbf{1}_{\{y(1-\varepsilon)^{-1} \leqslant \phi_x(t) \leqslant y\varepsilon^{-1}\}} \frac{dt}{\phi_x(t)}\Big] \\
&= \frac{1}{1-2\varepsilon} \int_0^\infty (1-\vartheta t) e^{-\vartheta t} \mathbb{E}\Big[\mathbf{1}_{\{y(1-\varepsilon)^{-1} \leqslant \phi_x(t) \leqslant y\varepsilon^{-1}\}} \frac{1}{\phi_x(t)}\Big] dt
\end{aligned}$$

by the change of variable $u = yz^{-1}$, the occupation times formula, and the specific form of κ. For $t \geqslant 0$, define

$$A_t(x,y) = \mathbb{E}\Big[\mathbf{1}_{\{y(1-\varepsilon)^{-1} \leqslant \phi_x(t) \leqslant y\varepsilon^{-1}\}} \frac{1}{\phi_x(t)}\Big] = \int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} \rho_t(x,z) \frac{dz}{z},$$

for which a closed-form formula is known, see for instance [Lin05], Section 4.1, given by

$$\rho_t(x,z) = \frac{2r_1 e^{2r_1 x}}{e^{2r_1 L} - 1} + \frac{2}{L} e^{r_1(z-x)} \sum_{n=1}^{\infty} \frac{e^{-a(n)t/2}}{a(n)} g(n,x) g(n,z),$$

with

$$g(n,x) = \frac{\pi n}{L} \cos\Big(x\frac{\pi n}{L}\Big) + r_1 \sin\Big(x\frac{\pi n}{L}\Big), \quad \text{and} \quad a(n) = r_1^2 + \pi^2 n^2 / L^2.$$

It follows that

$$A_t(x,y) = \frac{2r_1 e^{2r_1 x}}{e^{2r_1 L} - 1} \log\Big(\frac{1-\varepsilon}{\varepsilon}\Big) + \frac{2}{L} e^{-r_1 x} \sum_{n=1}^{\infty} \frac{e^{-a(n)t/2}}{a(n)} g(n,x) \mathcal{I}(n,y)$$

with $\mathcal{I}(n,y) = \int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} e^{r_1 z} g(n,z)\frac{dz}{z}$, and therefore

$$\begin{aligned}
\partial_\vartheta q_\vartheta(x,y) &= \frac{1}{1-2\varepsilon}\int_0^\infty (1-\vartheta t)e^{-\vartheta t}A_t(x,y)dt \\
&= \frac{1}{1-2\varepsilon}\frac{2}{L}e^{-r_1 x}\sum_{n=1}^\infty \Big(\int_0^\infty (1-\vartheta t)e^{-\vartheta t}\frac{e^{-a(n)t/2}}{a(n)}dt\Big) g(n,x)\mathcal{I}(n,y) \\
&= \frac{1}{(1-2\varepsilon)L}\int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} e^{-r_1(x-z)}\sum_{n=1}^\infty \frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2}\frac{dz}{z}.
\end{aligned}$$

Let $x \in [0,L]$ be such that $g(n,x) \neq 0$ for every $n \geqslant 1$. Since $x \mapsto g(n,x)$ is continuous on $[0,L]$, there exists $0 < \varepsilon_n < \frac{1}{2}$ such that $g(n,x)g(n,z) > 0$ for all $z \in \mathcal{J}(\varepsilon_n, x) = [2\varepsilon_n x, 2(1-\varepsilon_n)x]$. Let $N > 0$ be such that for all $z \in \mathcal{J}(\varepsilon_1, x)$:

$$|R_N(x,z)| = \Big|\sum_{n=N+1}^\infty \frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2}\Big| < \frac{g(1,x)g(1,z)}{(\vartheta + a(1)/2)^2},$$

which exists because by normal convergence of the above series. Then, for every $z \in \mathcal{J}(\max\{\varepsilon_n, 1 \leqslant n \leqslant N\}, x)$ we have

$$\sum_{n=1}^\infty \frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2} = \sum_{n=1}^N \frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2} + R_N(x,z) > \frac{g(1,x)g(1,z)}{(\vartheta + a(n)/2)^2} - |R_N(x,z)| > 0.$$

Finally, for $\varepsilon > \max\{\varepsilon_n, 1 \leqslant n \leqslant N\}$, picking $y = 2\varepsilon(1-\varepsilon)x$ yields $[y(1-\varepsilon)^{-1}, y\varepsilon^{-1}] = \mathcal{J}(\varepsilon, x) \subset \mathcal{J}(\max\{\varepsilon_n, 1 \leqslant n \leqslant N\}, x)$ so that $\partial_\vartheta q_\theta(x,y) > 0$.

Bibliography

[AAI11] K. B. Athreya, S. R. Athreya, and S. K. Iyer. *Supercritical age-dependent branching Markov processes and their scaling limits.* Bernoulli, 17(1):138–154, 2011.

[ADGP14] R. Azaïs, F. Dufour, and A. Gégout-Petit. *Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes.* Scandinavian Journal of Statistics. Theory and Applications, 41(4):950–969, 2014.

[AH76] S. Asmussen and H. Hering. *Strong limit theorems for general supercritical branching processes with applications to branching diffusions.* Probability Theory and Related Fields, 36(3):195–212, 1976.

[AK98a] K. Athreya and H.-J. Kang. *Some limit theorems for positive recurrent branching Markov chains: I.* Advances in Applied Probability, 30:693–710, 1998.

[AK98b] K. Athreya and H.-J. Kang. *Some limit theorems for positive recurrent branching Markov chains: II.* Advances in Applied Probability, 30:711–722, 1998.

[AN72] K. B. Athreya and P. E. Ney. *Branching Processes* (Springer-Verlag, New York-Heidelberg), 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[ASJ03] M. Ackermann, S. C. Stearns, and U. Jenal. *Senescence in a Bacterium with asymmetric division.* Science, 300(5627):1920–1920, 2003.

[Ath00] K. B. Athreya. *Change of measures for Markov chains and the $L \log L$ theorem for branching processes.* Bernoulli, 323–338, 2000.

[Ath12] K. B. Athreya. *Coalescence in critical and subcritical Galton-Watson branching processes.* Journal of Applied Probability, 49(3):627–638, 2012.

[Ave06] S. V. Avery. *Microbial cell individuality and the underlying sources of heterogeneity.* Nature reviews. Microbiology, 4(8):577, 2006.

[Ban08] V. Bansaye. *Proliferating parasites in dividing cells: Kimmel's branching model revisited.* The Annals of Applied Probability, 18(3):967–996, 2008.

[Ban15] V. Bansaye. *Ancestral lineages and limit theorems for branching Markov chains.* To appear in Journal of Theoretical Probability, 2015. Available at http://www.cmapx.polytechnique.fr/ bansaye/recherche.html.

[BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle, and V. C. Tran. *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees.* The Annals of Applied Probability, 21(6):2263–2314, 2011.

[BG17] E. Bernard and P. Gabriel. *Asymptotic behavior of the growth-fragmentation equation with bounded fragmentation rate.* Journal of Functional Analysis, 272(8):3455–3485, 2017.

[BH52] R. Bellman and T. Harris. *On age-dependent binary branching processes.* Annals of Mathematics, 280–295, 1952.

[BH15] V. Bansaye and C. Huang. *Weak law of large numbers for some Markov chains along non homogeneous genealogies.* ESAIM: PS, 19:307–326, 2015.

[Big77] J. D. Biggins. *Martingale convergence in the branching random walk.* Journal of Applied Probability, 25–37, 1977.

[Bil13] P. Billingsley. *Convergence of Probability Measures* (John Wiley & Sons), 2013.

[BK04] J. D. Biggins and A. E. Kyprianou. *Measure change in multitype branching.* Advances in Applied Probability, 544–581, 2004.

[BM15] V. Bansaye and S. Méléard. *Stochastic models for structured populations*, vol. 1 of *Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems* (Springer), 2015. Scaling limits and long time behavior.

[Bou16] F. Bouguet. *A probabilistic look at conservative growth-fragmentation equations.* arXiv:1609.02414, 2016.

[BPDG14] S. V. Bitseki Penda, H. Djellout, and A. Guillin. *Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application.* The Annals of Applied Probability, 24(1):235–291, 2014.

[BPEBG17] S. V. Bitseki Penda, M. Escobar-Bach, and A. Guillin. *Transportation and concentration inequalities for bifurcating Markov chains.* Bernoulli, 23(4B):3213–3242, 2017.

[BPHO17] S. V. Bitseki Penda, M. Hoffmann, and A. Olivier. *Adaptive estimation for bifurcating Markov chains.* Bernoulli, 23(4B):3598–3637, 2017.

[BPO17] S. V. Bitseki Penda and A. Olivier. *Autoregressive functions estimation in nonlinear bifurcating autoregressive models.* Statistical Inference for Stochastic Processes, 20(2):179–210, 2017.

[BPR17] S. V. Biteski Penda and A. Roche. *Local bandwidth selection for kernel density estimation in bifurcating Markov chain model.* arXiv:1706.07034, 2017.

[BR05] J. Bertoin and A. Rouault. *Discretization methods for homogeneous fragmentations.* Journal of the London Mathematical Society, 72(1):91–109, 2005.

[BSJ04] B. F. Brehm-Stecher and E. A. Johnson. *Single-cell microbiology: tools, technologies, and applications.* Microbiology and molecular biology reviews, 68(3):538–559, 2004.

[BT11] V. Bansaye and V. C. Tran. *Branching Feller diffusion for cell division with parasite infection.* Aléa, 8(81-127):241–242, 2011.

[BW17] J. Bertoin and A. Watson. *A probabilistic approach to spectral analysis of growth-fragmentation equations.* arXiv:1701.05061, 2017.

[Cat92] P. Cattiaux. *Stochastic calculus and degenerate boundary value problems.* Ann. Inst. Fourier (Grenoble), 42(3):541–624, 1992.

[CCF16] F. Campillo, N. Champagnat, and C. Fritsch. *Links between deterministic and stochastic approaches for invasion in growth-fragmentation-death models.* Journal of Mathematical Biology, 1–41, 2016.

[CCK17] K. L. Chen, M. M. Crane, and M. Kaeberlein. *Microfluidic technologies for yeast replicative lifespan studies.* Mechanisms of ageing and development, 161:262–269, 2017.

[Clo17] B. Cloez. *Limit theorems for some branching measure-valued processes.* Advances in Applied Probability, 49(2):549–580, 2017.

[CMR05] O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models.* Springer Series in Statistics (Springer, New York), 2005.

[Coo06] S. Cooper. *Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research.* Theoretical Biology and Medical Modelling, 3(1):10, 2006.

[CR88] B. Chauvin and A. Rouault. *KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees.* Probability theory and related fields, 80(2):299–314, 1988.

[CRS17] Z.-Q. Chen, Y.-X. Ren, and R. Song. *$L\log L$ criterion for a class of multitype superdiffusions with non-local branching mechanisms.* arXiv:1708.08219, 2017.

[CRW91] B. Chauvin, A. Rouault, and A. Wakolbinger. *Growing conditioned trees.* Stochastic Processes and their Applications, 39(1):117–130, 1991.

[DCFZ86] D. Dacunha-Castelle and D. Florens-Zmirou. *Estimation of the coefficients of a diffusion from discrete observations.* Stochastics, 19(4):263–284, 1986.

[DHKR15] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree.* Bernoulli, 21(3):1760–1799, 2015.

[DHRBR12] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. *Nonparametric estimation of the division rate of a size-structured population.* SIAM Journal on Numerical Analysis, 50(2):925–950, 2012.

[DHT84] O. Diekmann, H. Heijmans, and H. R. Thieme. *On the stability of the cell size distribution.* Journal of Mathematical Biology, 19(2):227–248, 1984.

[DLJB14] A. Denoth Lippuner, T. Julou, and Y. Barral. *Budding yeast as a model organism to study the effects of age.* FEMS Microbiology Reviews, 38(2):300–325, 2014.

[DM04] P. Del Moral. *Feynman-Kac Formulae* (Springer), 2004.

[DM10] J.-F. Delmas and L. Marsalle. *Detection of cellular aging in a Galton-Watson process.* Stochastic Processes and their Applications, 120(12):2495–2519, 2010.

[DMT95] D. Down, S. P. Meyn, and R. L. Tweedie. *Exponential and uniform ergodicity of Markov processes.* The Annals of Probability, 1671–1691, 1995.

[DMV16] P. Del Moral and D. Villemonais. *Exponential mixing properties for time inhomogeneous diffusion processes with killing.* To appear in Bernoulli, 2016.

[DMZ10] M. Doumic, P. Maia, and J. P. Zubelli. *On the calibration of a size-structured population model from experimental data.* Acta Biotheoretica, 58(4):405–413, 2010.

[DPZ09] M. Doumic, B. Perthame, and J. P. Zubelli. *Numerical solution of an inverse problem in size-structured population dynamics.* Inverse Problems, 25(4):045008, 2009.

[dSGPM12] B. de Saporta, A. Gégout-Petit, and L. Marsalle. *Asymmetry tests for bifurcating auto-regressive processes with missing data.* Statistics & Probability Letters, 82(7):1439–1444, 2012.

[dSGPM14] B. de Saporta, A. Gégout-Petit, and L. Marsalle. *Random coefficients bifurcating autoregressive processes.* ESAIM. Probability and Statistics, 18:365–399, 2014.

[Du06] Y. Du. *Order structure and topological methods in nonlinear partial differential equations*, vol. 2 (World Scientific), 2006.

[Dur78] R. Durrett. *The genealogy of critical branching processes.* Stochastic Processes and their Applications, 8(1):101–116, 1978.

[EHK10] J. Engländer, S. Harris, and A. Kyprianou. *Strong Law of Large Numbers for branching diffusions.* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 46(1):279–298, 2010.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov Processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics (John Wiley & Sons, Inc., New York), 1986. Characterization and convergence.

[Eng09] J. Engländer. *Law of large numbers for superdiffusions: The non-ergodic case.* Ann. Inst. H. Poincaré Probab. Statist., 45(1):1–6, 2009.

[EW06] J. Engländer and A. Winter. *Law of large numbers for a class of superdiffusions.* Ann. Inst. H. Poincaré Probab. Statist., 42(2):171 – 185, 2006.

[FM04] N. Fournier and S. Méléard. *A microscopic probabilistic description of a locally regulated population and macroscopic approximations.* The Annals of Applied Probability, 1880–1919, 2004.

[GB03] H.-O. Georgii and E. Baake. *Supercritical multitype branching processes: the ancestral types of typical individuals.* Advances in Applied Probability, 1090–1110, 2003.

[GCJ93] V. Genon-Catalot and J. Jacod. *On the estimation of the diffusion coefficient for multi-dimensional diffusion processes.* Annales de l'Institut Henri Poincaré. Probabilités et Statistiques, 29(1):119–151, 1993.

[Guy07] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging.* The Annals of Applied Probability, 17(5/6):1538–1569, 2007.

[Hai10] M. Hairer. *Convergence of Markov processes.* Lecture notes, 2010.

[Har63] T. E. Harris. *The theory of branching processes.* Die Grundlehren der Mathematischen Wissenschaften, Bd. 119 (Springer-Verlag), 1963.

[HG08] K. A. Henderson and D. E. Gottschling. *A mother's sacrifice: what is she keeping for herself?* Current opinion in cell biology, 20(6):723–728, 2008.

[HH09] R. Hardy and S. C. Harris. *A spine approach to branching diffusions with applications to L_p-convergence of martingales.* In *Séminaire de Probabilités XLII*, 281–330 (Springer), 2009.

[HHL02] R. Höpfner, M. Hoffmann, and E. Löcherbach. *Non-parametric Estimation of the Death Rate in Branching Diffusions.* Scandinavian journal of statistics, 29(4):665–692, 2002.

[HJR17] S. Harris, S. Johnston, and M. Roberts. *The coalescent structure of continuous-time Galton-Watson trees.* arXiv:1707.07993, 2017.

[HKK10] S. C. Harris, R. Knobloch, and A. E. Kyprianou. *Strong law of large numbers for fragmentation processes.* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 46(1):119–134, 2010.

[HM11] M. Hairer and J. C. Mattingly. *Yet another look at Harris' ergodic theorem for Markov chains.* In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, vol. 63 of *Progr. Probab.*, 109–117 (Birkhäuser/Springer Basel AG, Basel), 2011.

[HO16] M. Hoffmann and A. Olivier. *Nonparametric estimation of the division rate of an age dependent branching process.* Stochastic Processes and their Applications, 126(5):1433–1471, 2016.

[Hoa15] V. H. Hoang. *Estimating the division kernel of a size-structured population.* arXiv:1509.02872, 2015.

[Hoa16] V. H. Hoang. *Adaptive estimation for inverse problems with applications to cell divisions.* Ph.D. thesis, Université de Lille 1–Sciences et Technologies, 2016.

[Hof99] M. Hoffmann. *Adaptive estimation in diffusion processes.* Stochastic Processes and their Applications, 79(1):135–163, 1999.

[Hon11] J. J. Hong. *Coalescence in Bellman-Harris and multi-type branching processes.* Graduate Theses and Dissertations, 2011.

[HR14] S. Harris and M. Roberts. *A strong law of large numbers for branching processes: almost sure spine events.* Electronic Communications in Probability, 19:1–6, 2014.

[HR17] S. Harris and M. Roberts. *The many-to-few lemma and multiple spines.* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 53(1):226–242, 2017.

[HW96] S. C. Harris and D. Williams. *Large deviations and martingales for a typed branching diffusion, 1.* Astérisque, 236:133–154, 1996.

[Joh66] J. Johnston. *Reproductive capacity and mode of death of yeast cells.* Antonie Van Leeuwenhoek, 32(1):94–98, 1966.

[Kal77] O. Kallenberg. *Stability of critical cluster fields.* Mathematische Nachrichten, 77(1):7–43, 1977.

[Kim97] M. Kimmel. *Quasistationarity in a branching model of division-within-division.* In *Classical and modern branching processes (Minneapolis, MN, 1994)*, vol. 84 of *IMA Vol. Math. Appl.*, 157–164 (Springer, New York), 1997.

[Kin75] J. F. C. Kingman. *The first birth problem for an age-dependent branching process.* The Annals of Probability, 790–801, 1975.

[KJ96] S. P. Kale and S. M. Jazwinski. *Differential response to UV stress and DNA damage during the yeast replicative life span.* Genesis, 18(2):154–160, 1996.

[KLPP97] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. *A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes.* In *Classical and modern branching processes (Minneapolis, MN, 1994)*, vol. 84 of *IMA Vol. Math. Appl.*, 181–185 (Springer, New York), 1997.

[KP16] A. Kyprianou and S. Palau. *Extinction properties of multi-type continuous-state branching processes.* arXiv:1604.04129, 2016.

[KS66] H. Kesten and B. P. Stigum. *A limit theorem for multidimensional Galton-Watson processes.* The Annals of Mathematical Statistics, 37(5):1211–1223, 1966.

[Kun97] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*, vol. 24 of *Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge), 1997. Reprint of the 1990 original.

[Kut04] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes.* Springer Series in Statistics (Springer-Verlag London, Ltd., London), 2004.

[Kyp04] A. E. Kyprianou. *Travelling wave solutions to the KPP equation: alternatives to Simon Harris' probabilistic analysis.* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 40(1):53–72, 2004.

[Lin05] V. Linetsky. *On the transition densities for reflected diffusions.* Advances in Applied Probability, 37(2):435–460, 2005.

[LK10] S. Leibler and E. Kussell. *Individual histories and selection in heterogeneous populations.* Proceedings of the National Academy of Sciences, 107(29):13183–13188, 2010.

[LP13] A. Lambert and L. Popovic. *The coalescent point process of branching trees.* The Annals of Applied Probability, 23(1):99–144, 2013.

[LPP95] R. Lyons, R. Pemantle, and Y. Peres. *Conceptual Proofs of* $L\log L$ *Criteria for Mean Behavior of Branching Processes.* The Annals of Probability, 23(3):1125–1138, 1995.

[LVH$^+$12] S. S. Lee, I. A. Vizcarra, D. H. Huberts, L. P. Lee, and M. Heinemann. *Whole lifespan microscopic observation of budding yeast aging through a microfluidic dissection platform.* Proceedings of the National Academy of Sciences, 109(13):4916–4920, 2012.

[Lyo97] R. Lyons. *A simple path to Biggins' martingale convergence for branching random walk.* In *Classical and modern branching processes*, 217–221 (Springer), 1997.

[MÍ16] S. Méléard. *Modèles aléatoires en écologie et évolution. (French) [Random models in ecology and evolution].* Mathématiques et Applications (Berlin) [Mathematics and Applications] (Springer-Verlag, Berlin), 2016.

[Mar16] A. Marguet. *Uniform sampling in a structured branching population.* arXiv:1609.05678, to appear in Bernoulli, 2016.

[Mar17] A. Marguet. *A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages.* arXiv:1707.07993, submitted, 2017.

[Mic06] P. Michel. *Existence of a solution to the cell division eigenproblem.* Mathematical Models and Methods in Applied Sciences, 16(supp01):1125–1153, 2006.

[MMP05] P. Michel, S. Mischler, and B. Perthame. *General relative entropy inequality: an illustration on growth models.* Journal de mathématiques pures et appliquées, 84(9):1235–1260, 2005.

[MS16] S. Mischler and J. Scher. *Spectral analysis of semigroups and growth-fragmentation equations.* In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis* (Elsevier), 2016.

[MT93a] S. P. Meyn and R. L. Tweedie. *Stability of Markovian processes II: Continuous-time processes and sampled chains.* Advances in Applied Probability, 25(3):487–517, 1993.

[MT93b] S. P. Meyn and R. L. Tweedie. *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes.* Advances in Applied Probability, 518–548, 1993.

[MT09] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability* (Cambridge University Press, Cambridge), second ed., 2009.

[Mül71] I. Müller. *Experiments on ageing in single cells of Saccharomyces cerevisiae.* Archives of Microbiology, 77(1):20–25, 1971.

[NJ84] O. Nerman and P. Jagers. *The stable doubly infinite pedigree process of supercritical branching populations.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 65(3):445–460, 1984.

[O'C95] N. O'Connell. *The genealogy of branching processes and the age of our most recent common ancestor.* Advances in Applied Probability, 27(2):418–442, 1995.

[Øks03] B. Øksendal. *Stochastic Differential Equations.* Universitext (Springer-Verlag, Berlin), sixth ed., 2003. An introduction with applications.

[Oli17] A. Olivier. *How does variability in cell aging and growth rates influence the Malthus parameter?* Kinetic and Related Models, 10(2):481–512, 2017.

[Olo98] P. Olofsson. *The $x \log x$ condition for general branching processes.* Journal of applied probability, 35(3):537–544, 1998.

[Olo09] P. Olofsson. *Size-biased branching population measures and the multi-type $x \log x$ condition.* Bernoulli, 15(4):1287–1304, 2009.

[OTL$^+$04] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden. *Multistability in the lactose utilization network of Escherichia coli.* Nature, 427(6976):737–740, 2004.

[Per07] B. Perthame. *Transport Equations in Biology.* Frontiers in Mathematics (Birkhäuser Verlag, Basel), 2007.

[PR05] B. Perthame and L. Ryzhik. *Exponential decay for the fragmentation or cell-division equation.* Journal of Differential equations, 210(1):155–177, 2005.

[PZ07] B. Perthame and J. P. Zubelli. *On the inverse problem for a size-structured population model.* Inverse Problems, 23(3):1037, 2007.

[RHK$^+$14] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, and M. Doumic. *Division in Escherichia coli is triggered by a size-sensing rather than a timing mechanism.* BMC biology, 12(1):17, 2014.

[RSZ14] Y.-X. Ren, R. Song, and R. Zhang. *Central limit theorems for supercritical branching Markov processes.* Journal of Functional Analysis, 266(3):1716–1756, 2014.

[RY99] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* (Springer-Verlag, Berlin), third ed., 1999.

[SA02] E. Sumner and S. V. Avery. *Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast Saccharomyces cerevisiae.* Microbiology, 148(2):345–351, 2002.

[Ser02] D. Serre. *Matrices: Theory and Applications.* Graduate texts in mathematics (Springer), 2002.

[SMPT05] E. J. Stewart, R. Madden, G. Paul, and F. Taddei. *Aging and death in an organism that reproduces by morphologically symmetric division.* PLoS biology, 3(2):e45, 2005.

[SRA16] I. Soifer, L. Robert, and A. Amir. *Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy.* Current Biology, 26(3):356–361, 2016.

[TABS$^+$15] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. *Cell-size control and homeostasis in bacteria.* Current Biology, 25(3):385–391, 2015.

[Tan79] H. Tanaka. *Stochastic Differential Equations with Reflecting boundary condition in convex regions.* Stochastic Processes: Selected Papers of Hiroshi Tanaka, 9:157, 1979.

[Tra06] V. C. Tran. *Stochastic particle models for problems of adaptive evolution and for the approximations of statistical solutions.* Thesis, Université de Nanterre - Paris X, 2006.

[Tsy04] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*, vol. 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]* (Springer-Verlag, Berlin), 2004.

[vdV98] A. W. van der Vaart. *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press), 1998.

[Wil91] D. Williams. *Probability with martingales* (Cambridge university press), 1991.

[WRP$^+$10] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun. *Robust growth of Escherichia coli.* Current biology, 20(12):1099–1103, 2010.

[XZZ$^+$12] Z. Xie, Y. Zhang, K. Zou, O. Brandman, C. Luo, Q. Ouyang, and H. Li. *Molecular phenotyping of aging in single yeast cells using a novel microfluidic device.* Aging cell, 11(4):599–606, 2012.

[Zub76] A. M. Zubkov. *Limiting distributions of the distance to the closest common ancestor.* Theory of Probability and Its Applications, 20(3):602–612, 1976.

université
PARIS-SACLAY

École doctorale
de mathématiques
Hadamard (EDMH)

Titre : Processus de branchement pour des populations structurées et estimateurs pour la division cellulaire

Mots clefs : Processus de branchement, Formule Many-to-One, Loi des grands nombres, Ergodicité, Estimateur à noyau, Estimateur du maximum de vraisemblance

Résumé : Cette thèse porte sur l'étude probabiliste et statistique de populations sans interactions structurées par un trait. Elle est motivée par la compréhension des mécanismes de division et de vieillissement cellulaire. On modélise la dynamique de ces populations à l'aide d'un processus de Markov branchant à valeurs mesures. Chaque individu dans la population est caractérisé par un trait (l'âge, la taille, etc...) dont la dynamique au cours du temps suit un processus de Markov. Ce trait détermine le cycle de vie de chaque individu : sa durée de vie, son nombre de descendants et le trait à la naissance de ses descendants. Dans un premier temps, on s'intéresse à la question de l'échantillonnage uniforme dans la population. Nous décrivons le processus pénalisé, appelé processus auxiliaire, qui correspond au trait d'un individu "typique" dans la population en donnant son générateur infinitésimal. Dans un second temps, nous nous intéressons au comportement asymptotique de la mesure empirique associée au processus de branchement. Sous des hypothèses assurant l'ergodicité du processus auxiliaire, nous montrons que celui-ci correspond asymptotiquement au trait le long de sa lignée ancestrale d'un individu échantillonné uniformément dans la population. Enfin, à partir de données composées des traits à la naissance des individus dans l'arbre jusqu'à une génération donnée, nous proposons des estimateurs à noyau de la densité de transition de la chaine correspondant au trait le long d'une lignée ainsi que de sa mesure invariante. De plus, dans le cas d'une diffusion réfléchie sur un compact, nous estimons par maximum de vraisemblance le taux de division du processus. Nous montrons la consistance de cet estimateur ainsi que sa normalité asymptotique.

Title : Branching processes for structured populations and estimators for cell division

Keywords : Branching Markov processes, Many-to-One formulas, Law of large numbers, Ergodicity, Kernel estimator, Maximum likelihood estimator

Abstract : We study structured populations without interactions from a probabilistic and a statistical point of view. The underlying motivation of this work is the understanding of cell division mechanisms and of cell aging. We use the formalism of branching measure-valued Markov processes. In our model, each individual is characterized by a trait (age, size, etc...) which moves according to a Markov process. The rate of division of each individual is a function of its trait and when a branching event occurs, the trait of the descendants at birth depends on the trait of the mother and on the number of descendants. First, we study the trait of a uniformly sampled individual in the population. We explicitly describe the penalized Markov process, named auxiliary process, corresponding to the dynamic of the trait of a "typical" individual by giving its associated infinitesimal generator. Then, we study the asymptotic behavior of the empirical measure associated with the branching process. Under assumptions assuring the ergodicity of the auxiliary process, we prove that this process corresponds asymptotically to the trait along its ancestral lineage of a uniformly sampled individual in the population. Finally, we address the problem of parameter estimation in the case of a branching process structured by a diffusion. We consider data composed of the trait at birth of all individuals in the population until a given generation. We give kernel estimators for the transition density and the invariant measure of the chain corresponding to the trait of an individual along a lineage. Moreover, in the case of a reflected diffusion on a compact set, we use maximum likelihood estimation to reconstruct the division rate. We prove consistency and asymptotic normality for this estimator.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France