Aix-Marseille Université

ED 184 Mathématiques et Informatique

INSTITUT DE MATHEMATIQUE DE MARSEILLE (I2M) /UMR 7373

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Mathématiques

Aurélien VELLERET

Mesures quasi-stationnaires et applications à la modélisation de l'évolution biologique

Quasi-stationary distributions and application to models of biological evolution

Soutenue le 09/07/2020 devant le jury composé de :

M.	Nicolas CHAMPAGNAT	DR - Inria Nancy - Grand Est	Rapporteur
Mme.	Sylvie MELEARD	PR - Centre de Math. Appl.	Examinateur
M.	Vincent BANSAYE	PR - Centre de Math. Appl.	Examinateur
M.	François HAMEL	PR - Instit. de Math. de Marseille	Examinateur
M.	Etienne PARDOUX	PR - Instit. de Math. de Marseille	Directeur de thèse
M.	Michael KOPP	PR - Instit. de Math. de Marseille	Codirecteur de thèse

Après avis des rapporteurs :

M.	Nicolas CHAMPAGNAT	DR - Inria Nancy - Grand Est
M.	Michel BENAïM	PR - Instit. de Math. de Neuchâtel

Numéro national de thèse/suffixe local : 2020AIXM0226/010ED184

Résumé

Je décris le comportement en temps long de plusieurs processus qui illustrent les mécanismes de sélection naturelle. Sous certaines conditions que je détaille, les effets de sélection peuvent être interprétés comme un conditionnement qui biaise la dynamique d'un processus aléatoire "neutre". Ce processus évolue sur un espace $\mathcal{X}$, nommé espace des traits et potentiellement très général, notamment continu et non borné. On peut ainsi caractériser aussi bien la dynamique du profil complet de la population que celle du profil d'un individu (ou d'un sous-groupe) choisi uniformément dans la population. On voit naturellement apparaître dans ces modèles des transitions brutales d'états qui rendent l'analyse plus délicate que les cas traités classiquement. Cette thèse s'attaque à ce problème à la fois d'un point de vue général et sur une variété de processus spécifiques.

Chacun des chapitres qui suivent constitue l'objet d'un article en voie de révision ou sur le point d'être soumis. Le dernier fait exception, puisque sa publication est validée et qu'il doit sortir dans le volume 68 de ESAIM Proceedings and Surveys. Ce dernier chapitre est un peu particulier au sens où il présente une interprétation personnelle des résultats de différents articles récents. Son fil directeur est la session dédiée aux applications en biologie des processus stochastiques lors de la conférence Modélisation Aléatoire et Statistiques à laquelle j'ai participé en août 2018. Je suis le seul auteur de ces travaux, mis à part pour le 5e chapitre, qui est une collaboration avec M. Mariani et mon directeur E. Pardoux. Ce dernier a néanmoins eu une part très importante dans la révision de tous ces travaux, et je lui en suis très reconnaissant.

Après ce bref résumé, je reviendrai plus en détails sur les conclusions de ces chapitres en section 0.2, après la présentation des résultats de la littérature donnée en section 0.1. J'y présenterai d'abord les concepts élémentaires sur lesquels mes résultats se basent. Après la description des propriétés de processus Markov et la description en terme de semi-groupe, j'introduis assez vite les notions de mesures quasi-stationnaires et de capacité de survie. Cela amène à celle de Q-processus, obtenu via la transformée de Doob du processus initial. Avant d'introduire les résultats de convergence, il me semble alors important de revenir aux notions clés qui amènent aux théorèmes ergodiques. Les aspects de pondération de l'espace, d'irréductibilité et les conditions de Lyapunov sont utiles à garder en mémoire quoique je ne les utilise pas directement dans mes articles (pour l'instant !). Je les détaille donc en Section 0.1.2, d'autant qu'ils aideront à la discussion sur les résultats voisins de ceux de mon premier article. On verra alors en Section 0.1.3 comment ces notions s'étendent au cadre dit non-conservatif. La

Section 0.1.4 donne un aperçu sur les résultats généraux pour établir l'existence de QSD et les propriétés de convergence associées. En guise d'ouverture, on esquissera en Section 0.1.5 le fait que mes conditions, comme celles de plusieurs autres articles récents, permettent en outre de quantifier les fluctuations de la mesure empirique autour de sa référence quasi-ergodique en terme de Grandes Déviations. La question de l'estimation numérique des QSD est enfin introduite en Section 0.1.6.

La première partie de cette thèse, à savoir les deux premiers chapitres, consiste à mettre en place des conditions permettant d'obtenir des résultats analogues à la stationnarité et à l'ergodicité lorsqu'on prend en compte un tel conditionnement. Le premier chapitre introduit un premier jeu d'hypothèses, avec une extension développée au deuxième chapitre. Dans chacun de ces chapitres, je présente des exemples d'applications, qui illustrent la puissance de mes jeux d'hypothèses.

Mes critères sont facilement vérifiés, comme je le détaille dans mes applications, tirées de la modélisation de la sélection naturelle en écologie théorique. L'étude de la première famille de processus vise à mieux comprendre sous quelles conditions peut se faire l'adaptation d'une population à un environnement changeant. Ainsi, le processus considéré est un couplage de deux composantes, démographique et évolutive. Le conditionnement se fait sur la survie de la population, avec un effet implicite de sélection pris en compte lors des événements d'invasion par une sous-population porteuse d'une mutation. Plusieurs exemples qui ont illustré les deux premiers chapitres sont tirés de cette veine et introduisent des éléments essentiels pour la preuve du modèle couplé traité au chapitre 3.

Une deuxième famille de modèle compare des effets de sélection dans une population structurée en groupes en distinguant une compétition intra-groupes et entre groupes. L'objet du Chapitre 4 est ainsi d'exploiter les techniques de quasi-stationnarité pour décrire la dynamique interne d'un groupe choisi uniformément dans la population. Cela conduit à traiter la quasi-stationnarité au-delà du cas irréductible. Le Chapitre 5 porte sur la contre-sélection des mutations délétères dans une population asexuée, et donne encore une autre motivation à la quasi-stationnarité en tant qu'étude d'un équilibre méta-stable. On y verra un processus bien particulier pour lequel il est possible justifier une convergence en variation totale vers une unique distribution quasi-stationnaire quand bien même il évolue de manière diffusive en dimension infinie.

Je profite enfin de cette occasion pour ouvrir un plus large panorama sur ces modèles à mi-chemin entre écologie et évolution avec une discussion détaillée des exposés de la session dédiée lors des Journées de Modélisation Aléatoires et Stochastiques d'août 2018. Les limites d'échelles en grandes populations et temps longs y jouent un grand rôle pour éclairer les simulations de type individus-centrées et justifier l'analyse des modèles limites.

Mots clés : distribution quasi-stationnaire, écologie, évolution, méta-stabilité, modèle de l'optimum mobile, sélection de groupe, équations de réaction-diffusion, cliquet de

Muller, processus de Markov en temps et espace continus, équations différentielles stochastiques, processus à sauts, limites en temps longs, limites d'échelles, récurrence de Harris, couplages, équations aux valeurs propres, grandes déviations, processus de branchement

Abstract

I describe the long time behavior of several random processes that illustrate the mechanisms of natural selection. Under certain conditions that I will specify, selective effects can be interpreted as a conditioning that skews the dynamics of a "neutral" random process. This process evolves on a space $\mathcal{X}$, called the trait space and potentially very general, especially continuous and unbounded. We can thus characterize both the dynamics of the complete population profile and the profile of an individual (or a sub-group) selected uniformly across the population. In these models we see naturally occurring abrupt transitions that make the analysis more delicate than for more classical conservative Markovian processes. This thesis tackles this problem both from a general point of view and on a variety of specific processes.

Each of the following chapters constitutes the subject matter of an article under review or about to be submitted. The latter is an exception, since its publication is validated. It should be released in volume 68 of ESAIM Proceedings and Surveys. This last chapter is a bit special in the sense that it presents a personal interpretation of various results from several recent articles. Its guiding principle is the session dedicated to applications in biology of stochastic processes at the conference "Modélisation Aléatoire et Statistiques" in which I took part in August 2018. I am the sole author of these works, except for the fifth chapter, which is a collaboration with M. Mariani and my director E. Pardoux. The latter nonetheless contributed much in the revision of all these works, and for that, I'm very grateful to him.

The first part of this thesis, namely the first two chapters, is to put in place conditions that allow to obtain results similar to stationarity and ergodicity when such conditioning is taken into account. The first chapter introduces a first set of hypotheses, with an extension developed in the second chapter. In each of these chapters, I present examples of applications, that illustrate the power of my sets of assumptions.

My criteria are easily verified, as I detail in my applications, which come from the modeling of natural selection in theoretical ecology. The study of the first family of processes aims at better understanding under which conditions can a population be able to adapt to a changing environment. The process under consideration is a coupling of two components, the first one representing the demography and the second one the characters of the individuals. In this model, we need a conditioning on the event that the population has survived. An implicit effect of selection is also taken into account, in the way we express successive events of fixation of a new mutation in the population. Several examples that have illustrated the first two chapters are

derived from this motivation and introduce essential elements for the proof of the coupled model discussed in Chapter 3.

A second family of models compares selective effects in a population structured in groups distinguishing between intra-group and inter-group competition. The purpose of Chapter 4 is thus to exploit quasi-stationarity techniques to describe the internal dynamics of a group uniformly sampled across the population. This leads to the consider results of convergence to quasi-stationary distribution for a non-irreducible process. Chapter 5 deals with on the counter-selection of deleterious mutations in an asexual population. It provides another motivation for looking at quasi-stationarity in terms of metastability. A very specific process will be considered for which it is possible to justify a convergence in total variation towards a unique quasi-stationary distribution even though the process evolves in a diffusive manner in some infinite dimensional space.

Finally, I take advantage of this opportunity to open up a wider panorama on these models halfway between ecology and evolution. I present a detailed discussion of the presentations from the dedicated session during the conference Modélisation Aléatoires et Stochastiques that took place in August 2018. Scaling limits in large populations and long time scales play an important role in enlightening individual-based simulations and justify the analysis of limiting summary models.

Keywords: quasi-stationary distribution, ecology, evolution, meta-stability, model of the mobile optimum, group selection, reaction-diffusion equations, Muller's ratchet, Markov process in continuous time and continuous space, stochastic differential equations, jump processes, long time limits, scaling limits, Harris recurrence, couplings, eigenvalue equations, branching processes, large deviations

Remerciements

Si je me suis déclaré seul auteur de la plupart des travaux qui constituent cette thèse, je n'oublie pas à quel point je dois à ceux qui m'ont encadré ou qui simplement ont accompagné une part de mon cheminement jusqu'à aujourd'hui. Ce travail a manifestement été rien moins que solitaire.

Cette période a clairement été pour moi comme un parcours initiatique, bien au-delà de la dimension purement académique. Vu la longueur de cette thèse, je me permets de m'épancher assez librement. J'ai tardé jusqu'à bien tard la rédaction de cette section. Sans doute paradoxalement car il me semblait devoir dire trop de choses pour pouvoir mener cette thèse à son accomplissement. Ils sont bien trop nombreux ceux que j'aimerais mentionner comme jalons de ma vie. Il est néanmoins plus que l'heure de s'y mettre ! Car aussi bien ma culture chrétienne de famille que ma passion plus récente pour la pensée japonaise et les échos des dernières recherches en neurosciences, tout m'amène à sanctifier cette partie. La reconnaissance que l'on doit à autrui est le fondement de la reconnaissance de soi et des autres !

Ma reconnaissance va ici en tout premier lieu à Etienne, qui a su me pousser à voir toujours au-delà de ce que je croyais possible ! Son assurance à s'engager dans la vie sous toutes ses formes force le respect. Je viserai à être un mathématicien aussi exemplaire. Je pense bien sûr à la passion qu'il témoigne pour ses très nombreuses recherches à l'interface des probabilité et de la modélisation du vivant. Mais je pense aussi à sa présence toujours disponible, qui amène à de fréquentes discussions, simples et directes. Je pense aussi à l'engagement qu'il sait susciter dans ses nombreuses collaborations, qui l'ont amené aux quatre coins du monde. Le cycle de conférences internationales qu'il a initié au CIRM en février dernier a été une exceptionnelle opportunité pour moi. Quand bien même cela m'a en partie détourné de l'avancée de ma thèse, je ne regrette pas le moins du monde d'y avoir autant été présent. Je suis enfin admiratif de ses initiatives pour donner accès au meilleur niveau de probabilité à des jeunes de toute l'Afrique francophone. C'est à mes yeux une immense richesse que d'avoir côtoyé ceux venus de très loin qui ont pourtant été parmi mes plus proches camarades de thèse : Ténan, Ibrahima, Alphonse, Boris, Brice, Elma, nous sommes tous les héritiers d'Etienne, parmi une grande famille de grands chercheurs ! A l'image de mon directeur, je garderai le plus longtemps possible ce goût pour les randonnées en pleine nature, l'art sous toutes ses formes et la grande liberté d'esprit.

Je tacherai de préserver aussi dans mes travaux mathématiques la plus grande rigueur alliée au soin constant de rendre ces résultats les plus accessibles possibles. Même si j'avais déjà acquis par ma formation le soucis d'une grande précision, il a bien fallu m'y reprendre pour ne rien laisser noircir la pureté de cette belle architecture de travail. Quant à la clarté des expressions qui en guident la visite, je réalise à quel point elle ne coule pas de source. Car ce n'est pas le tout que de faire l'architecture la plus solide, la plus économe ou la plus élégante. Il faut aussi la faire vivre et dévoiler ses beaux recoins cachés au maximum de monde. Cela exige, outre une parfaite compréhension et encore une grande précision, d'abord un travail toujours recommencé pour améliorer la portée signifiante de ses mots.

Si mon éloge sera sans doute moins vibrant pour Michael, c'est sans nul doute que je me suis trop laissé abreuver à la source d'Etienne. Si la séparation géographique entre mon bureau de Château-Gombert et le sien à Saint-Charles a créé une barrière plus tenace que je ne pouvais penser, c'est toujours avec beaucoup de générosité qu'il a répondu à mes demandes de conseils. Resté attentif à ma progression avec Etienne, il m'a je crois laissé tranquillement tracer mon chemin même trop abstrait et trop peu concret à son goût. Finalement, j'aurai toujours dû repousser le travail précis d'analyse de mes simulations et l'écriture d'un article plus dédié à une audience d'écologues théoriciens. Les séminaires de maths-bio qu'il a organisé avec Etienne ont néanmoins toujours été sources d'inspiration. J'aurai souhaité que plus de personnes découvrent les passionnants chercheurs qu'il a accueilli. Je me suis aussi avec lui confronté à enseigner les statistiques à ces malheureux étudiants qui peinent à appréhender ce monde de l'abstraction mathématique. Je sais son implication même lors de cette période difficile de confinement.

Je remercie bien sûr les membres de mon jury qui ont accepté de faire l'évaluation de ce travail de thèse. C'est un grand honneur pour moi que de réunir cette assemblée prestigieuse.

Je ne veux pas non plus oublier de remercier mes rapporteurs, Michel et Nicolas. J'aurai bien souhaité leur donner de meilleures conditions pour bien relire mon imposant manuscrit. Suite à des changements de plan de dernière heure, j'ai néanmoins accumulé un retard qui a comprimé le temps que je leur accordais pour la relecture. Les conditions sanitaires récentes n'ont aussi pas aidé. Je leur suis donc très reconnaissant d'avoir aussi précisément analysé la teneur, la qualité et l'originalité de ce travail. J'ai grandement apprécié les corrections qui m'ont été transmises et par lesquelles j'achève ce manuscrit.

J'en profite pour remercier tous ceux qui viendront assister à ma soutenance, pour en faire un jour des plus mémorables.

Mais je n'en serais déjà pas là sans le dévouement et la qualité des professeurs qui ont jalonné ma formation. Deux d'entre eux parmi les plus marquants me font d'ailleurs l'honneur de faire partie de mon jury. J'espère que cette thèse rendra bien hommage à ce que je leur dois, d'abord comme étudiant et maintenant comme jeune collègue en apprentissage. Ils m'ont ouvert l'une des arches les plus ouvragées vers un

monde fantastique : avec des trésors d'abstraction, nous tentons d'esquisser à notre niveau rien moins que la dynamique du monde.

Cette reconnaissance envers mes professeurs a depuis longtemps eu une saveur particulière vis-à-vis de cette discipline que sont les mathématiques. Je me souviens de ce professeur de 5e qui a affirmé une fois en substance : "Voyez cet enfant qui n'a presque plus besoin de moi. Il est déjà presque capable d'apprendre directement des livres de maths." Et bien qu'il m'en coûte de perdre cet encadrement constant, c'est désormais bien devenu la direction que je poursuis. Je me rappelle aussi de ma professeur de 4e et 3e : Mme Motchidlover. Petite, âgée et frêle, on ne lui aurait guère préjugé tant d'assurance que ce qu'elle manifestait en nous enseignant. Alors qu'en raison d'un accident, je me suis retrouvé incapable pendant plus d'un mois de venir en cours, sa venue comme professeur particulier m'a vite désarçonné. C'est que je ne pouvais plus tirer partie de ma compréhension plus rapide que la moyenne pour prendre cela à la légère ! Mais j'ai découvert ainsi à quel point j'aimais me confronter à ces défis. Et il a bien fallu cela lorsque j'ai franchi les échelons de plus en plus techniques et ardus qui m'ont mené du Lycée Louis-le-Grand à l'entrée de l'ENS. Tous mes professeurs de mathématiques m'ont alors laissé une impression extra-ordinaire, avec une mention toute particulière pour M. Alarcon, Mme. Biolley et M. Randé. Une fois à l'ENS, c'est un autre univers qui s'est ouvert devant moi. Dans cette dimension de recherche, cela a été pour moi une grande ouverture sur plus de liberté et l'entretien d'une passion toujours plus affirmée. J'ai alors pu renouer avec l'apprentissage de la biologie et de l'évolution. Car j'en avais été marqué en particulier par la préparation en terminale du Concours Général des Lycées dans cette matière. Cela m'a orienté vers ce M2 d'Orsay de Maths pour les Sciences du Vivant et à ces superbes rencontres. Avec en particulier, pour n'en citer que quelques uns parmi ces grands noms, Régis Ferrière, Bertrand Maury, Amandine Veber et Denis Thieffry puis Christophe Giraud, Stéphane Robin, Nicolas Curien sans oublier bien sûr Sylvie et Vincent.

Le panorama s'est toujours plus élargi durant cette thèse avec des rencontres jusqu'à l'autre bout du monde. Je pense déjà à Jonathan Mattignly, résidant en Caroline du Nord aux USA, et Mauro Mariani, désormais bien installé à Moscou, dont la rencontre s'est poursuivie en la rédaction des chapitres 4 et 5 de ma thèse. Mais j'ai aussi une pensée d'amitié très sincère à Kim Daehong. Lorsque je l'ai sollicité pour prendre part au Summer Program de l'été 2019 organisé par la JSPS, il m'a accueilli à bras ouverts au Japon. Cela a été l'occasion d'un beau dépaysement, malgré ma familiarité d'avec la culture japonaise. L'un de mes prochains articles sur les grandes déviations de processus à sauts devra bien rendre honneur à la confiance qu'il m'a accordé. Ce devait constituer l'un des chapitres de cette thèse, mais la comparaison d'avec les résultats d'un article plus ancien ([KM03]) m'en a fait repousser la finalisation. Mais ce n'est que partie remise pour présenter des résultats encore plus éclatants !

Et je n'oublie pas tous les chercheurs que j'ai rencontré lors des nombreuses confé-

rences auxquelles j'ai participé. Parmi elles, je compte tout particulièrement les rencontres de la Chaire Modélisation Mathématique et Biodiversité dont Sylvie est la marraine et l'active gardienne. Je rappelle aussi ce mois thématique organisé au CIRM où a été manifeste la richesse et la diversité des approches mathématiques pour l'étude du vivant. Je suis aussi très reconnaissant aux organisateurs des différentes écoles d'été auxquelles j'ai participé à Turku, Güntzburg et Aussois, ainsi qu'à ceux qui m'ont invité à parler en séminaire, jusqu'au Kansai.

Parmi toutes ces rencontres, je suis persuadé qu'il y aura l'occasion de très belles collaborations. Depuis qu'il a son poste sur Avignon, j'ai déjà eu la chance de pouvoir interagir fréquemment avec Raphael. Vu la proximité de nos recherches et de nos centres d'intérêts allié à une belle complicité, je suis sûr qu'on aura de multiples occasions de retravailler ensemble !

Merci à ceux que j'ai rencontré en vue d'organiser mon post-doctorat l'année prochaine et qui ont bien pris le temps de considérer mon projet. Je cite à ce titre Anton Bovier, Anton Wakolbinger, Cornelia Pokalyuk, Michel Benaïm, Marie Doumic, Andreas Kyprianou et Philippe Robert.

Mais c'est aussi évidemment localement au sein de ce laboratoire de l'I2M que les rencontres ont été les plus faciles. Merci à Pierre Picco et à Florence, d'avoir accepter de constituer mon comité de suivi de thèse. Je n'ai pas tant que cela eu besoin de vous solliciter, mais c'est toujours rassurant de savoir à qui parler s'il y avait un soucis.

Je remercie aussi Fabienne et Alexandre avec qui je me suis toujours senti en confiance et stimulé de poursuivre mes avancées. Je pense aussi à Erwan, Charles et Remy, qui se sont occupé d'organiser ces séminaires de probabilité si instructifs. Merci à Gilles Didier, qui est si humble malgré la richesse de ses travaux. Merci à Eric que je retrouvais à ces séminaires de maths-bio et à tous ceux qui m'ont accueilli pour les séminaires de statistiques. J'ai eu une relation particulière avec Pierre Mathieu, qui m'a donné cours et avec qui je partage ma passion du Japon. Tes conseils me sont précieux, aussi bien en recherche que pour découvrir ce merveilleux pays.

Pour me soutenir sur le cours de statistiques destiné aux élèves du nouveau Master de sciences cognitives, merci à Bruno, Marieke, Alexis, Julie et les autres. Pour mes premiers débuts dans l'enseignement, merci à Marie-Thérèse et à Kacem, puis à Laurent Pézard, Marina Talet et enfin Christophe Bernard, Pierre Mangeol et Véronique Paban. J'ai aussi beaucoup d'admiration pour Pierre Pudlo dans la manière dont il sait enseigner les statistiques et que j'ai tenté de reprendre à ma façon.

Merci à tous les élèves que j'ai pu encadrer. Ils n'ont sans doute pas eu l'approche la plus simple, surtout quand j'ai eu la main sur un cours de statistiques. Mais nous avons eu des échanges pleins de vie et de rebondissements. Outre les connaissances pratiques, j'espère avoir su leur transmettre l'idée de cette passion qui m'anime pour cette discipline.

Une thèse, c'est aussi tout un ensemble de questions administratives et pratiques. Je suis heureux d'avoir pu compter sur Marie-Christine, Evelyne et Valérie pour la logistique et l'organisation d'événements. La préparation de la Journée des Doctorants en février 2019 avec Mélodie a été pour moi une immense joie. Je dois aussi à ce titre remercier l'implication de Fabienne et le soutien de Nadia et de Sonia, avec toute la direction de l'Ecole Doctorale 184. Ils souhaitent tous voir pérenniser ce type d'événements qui nous rassemblent, nous doctorants, en une communauté solidaire. En tant que représentant des doctorants (à vrai dire un bien piètre délégué), j'ai pu voir de l'intérieur à quel point ils sont attentifs à ce que nous ayons les meilleures conditions. C'est d'autant plus le cas en cette période difficile à vivre du confinement, quand trop jeune on manque de repères. J'espère que mes successeurs sauront profiter de toutes ces occasions qui s'ouvrent à eux s'ils savent en nourrir l'ambition.

Les Journées de Pi sont à cet égard exemplaires de la qualité et du rayonnement que l'on peut manifester. Guillaume, Annamaria, Elena, Joel, ils avaient très clairement un sacré talent d'organisation. Mais qu'est-ce que ça avait de spectaculaire d'avoir ce cycle de spectacles entraîné depuis Marseille jusqu'à Paris et Lyon !

Mes remerciements se tournent ensuite naturellement vers ceux avec qui j'ai partagé la plus de moments dans le quotidien de mes travaux au bureau. Nous avons tous bien œuvré à ce que ce bureau soit l'un des plus actifs ! A vrai dire, ce lieu m'a fait un peu peur la première fois, à l'idée d'un espace sanctuarisé pour le travail personnel. Mais c'est très vite devenu pour moi un second lieu de vie et je ne peux que m'attrister que ma compagne n'ait pas cette chance d'avoir un bureau pour son travail (les remerciements que je lui réserve attendent leur meilleure place finale !). Les échanges et la convivialité n'a pas entravé notre ardeur au travail, bien au contraire ! Et j'étais libre bien sûr de laisser vadrouiller mes pensées lors de mes virées dans l'espace boisé en aplomb du laboratoire.

Ténan, même s'il est plus âgé que moi et a dès le départ eu plus de maturité, est bien celui dont le parcours de thèse aura suivi le mien au plus proche. Je l'ai précédé de très peu avec mon stage de M2 et peut-être me considère-t-il tout de même comme son aîné. J'ai perçu avec quelle forte amitié il se réjouissais des avancées de ma recherche et je peux lui dire que c'est réciproque. Ténan, j'espère que tu pourras t'installer comme un grand chercheur dans ta Côte d'Ivoire natale. Il sera bon que tu retrouves ta proche famille dont tu as dû supporter si longtemps l'éloignement. Bon nombres de nos discussions sont restées sérieuses, tout en maintenant entre nous une grande complicité et un bel entrain.

Ibrahima était néanmoins imbattable question enthousiasme. J'en garde la belle image du grand frère toujours prêt à me prendre dans ses bras pour fêter nos retrouvailles. J'en garde aussi la fierté que tu avais d'enseigner au plus grand nombre et de porter haut les couleurs de ton pays, le Sénégal.

Avec Alphonse, la grande confiance que nous avons maintenant l'un pour l'autre ne s'est pas faite en un jour. Ces premiers jours toujours intimidants dans un autre univers sont désormais passés et je me réjouis de pouvoir échanger au sujet des textes qu'il m'a transmis et dont cette fin de thèse a retardé la lecture. Plus encore que tous les autres étudiants qu'Etienne a su faire inviter d'Afrique, j'ai été marqué par son ardeur au travail. J'ai souvent eu cette certitude qu'en venant à mon bureau, je le retrouverai forcément à ta place. Ce que l'on peut dire plus difficilement de moi, même si je trouvais toujours des techniques pour me mettre au travail à la dérobée. Ma capacité de travail est un savant mélange entre excitation, recherche de stabilité et d'accomplissement. L'envie d'aller faire autre chose n'a pas été si rare. Mais l'idée de ce qu'il est de mon engagement d'accomplir tient aussi de ce que vous parvenez à faire en partant de conditions bien moins favorables.

Yannis, cette fois, je me permets de dire ce que j'ai vite pensé à notre première rencontre, alors que je voulais me remettre au travail : décidément, qu'est-ce qu'il parle pour un doctorant de mathématiques ! Je m'y suis habitué et cela est resté raisonnable au sein du bureau (quoique). J'ai toujours apprécié la teneur de nos échanges et la confiance d'avoir quelqu'un sur qui compter. Nous avons finalement partagé beaucoup, notamment sur nos joies et nos frustrations comme doctorants et comme enseignants. Je ressens bien combien l'avancée de sa thèse est ardue et technique et lui souhaite le meilleur des courages pour en venir à bout !

Enfin, Elma, Guilio, Pierre, c'est avec joie que je vous retrouvais même si cela n'a pas autant duré. Mes pensées vont aussi à François Ged, venu pour son stage de M2 et qui aurait bien voulu poursuivre ici sa thèse. Nous avons eu très vite de très bons contacts. J'ai appris que sa thèse avec Jean Bertoin s'est bien passée, ce qui n'est pas pour m'étonner.

Boris aussi est quelquefois venu occuper ce bureau, même si historiquement le sien était plutôt à côté. C'est d'abord avec les rencontres vers Aussois que nous avons eu plus directement l'occasion de discuter. Son sujet de thèse, entre la théorie des grandes déviations et la modélisation de la Malaria, lui a manifestement donné bien du fil à retordre. Je me suis inquiété de ne plus le voir et j'aurai souhaité qu'il relâche un peu plus son fardeau avec nous. Mais il peut désormais être fier de ses accomplissements. Sans oublier de nous faire profiter des plus beaux airs d'opéra !

Et puis il y a la multitude des doctorants que j'ai côtoyé pendant toutes ces années, et avec qui j'ai très souvent partagé mes repas. Je suis hélas certain d'en oublier quelques uns, mais je sais que voir son nom fait déjà quelque chose. Merci à toi, Mélodie, pour ton engagement sincère comme représentante des doctorante et pour ton soutien dans ce que nous avons accompli avant d'être rattrapés par cette fin de thèse. Merci à Marina, Adrien, Florian, Pierre, Federico, Timothy, d'avoir répondu présent pour incarner cette Journée des Doctorants. Merci à Cécile, Marie-Eve, Guillaume, Damien, de s'être investis dans l'organisation des Séminaires des Doctorants. Merci à Olivier, devenu notre support informatique après la fin de sa thèse. Merci aussi à Jordan, Benoît, Valentin, Paolo, Dyonysis, Valdimiro, Rémi, qui m'ont précédé et m'ont accueilli avec joie. Merci encore à Irène, Arnaud, Marina, Luc, mais aussi Lydia,

Andrea, Vincent, Quentin, Antoine, avec qui j'ai partagé déjà plusieurs années. Merci toujours aux plus nouveaux : Clément, Anaïs, Enguerrand, Mingmin, Juan, Emily, Bastien, Octave, Youssouf. Je pense aussi aux post-doc : Alexandre, Sergey, Romain, qui nous ont rejoins.

Lors de la préparation de cette thèse, il y a par ailleurs toute une vie plus quotidienne, faite de plaisirs simples et de curiosité. Je suis donc pleinement reconnaissant à Patrick, pour son engagement comme chef de chœur à nous faire profiter des musiques les plus exotiques ; à Olivier pour son enseignement du Shorinji, qui sous couvert de luttes au corps parle de moralité, de respect et de coopération ; à Kaori qui m'a enseigné le japonais et la culture du pays avec autant de bienveillance. Sans oublier tous ceux, trop nombreux pour les citer, avec qui j'ai partagé les joies de ces apprentissages et de leur mise en pratique. Le yoga est enfin devenu ma nouvelle passion et j'enjoins tout un chacun à s'y mettre.

Merci à mes colocataires, Thibaut Desjardin, Thibaut Rozet, avec qui j'ai partagé de si bons moments et qui ont bien voulu m'entendre longtemps détailler mes travaux. Vous êtes devenus des amis irremplaçables. Ce qui ne m'empêche pas de garder l'amitié et le soutien de beaucoup d'autres, rencontrés au fil de toutes ses années et dont je reprends trop peu le contact. Jonathan, Diane, Raphael, Adrien, Manuel, Jules, Kevin, Pierre, Thomas, Edgard, Sophia, Olivier, François, etc... nombreux sont ceux à qui je repense avec joie.

Je ne peux pas terminer ces remerciements sans en adresser parmi les plus chaleureux à mes parents, et plus généralement à ma famille. Je me suis nourri à partir de leur soutien sans faille et de la grande liberté de choix qu'ils m'ont encouragé à développer. Si leur passion pour les livres, les langues et l'histoire antique est un peu trop dévorante, bénéficier d'une telle atmosphère de tolérance, d'ouverture et de culture a beaucoup contribué à faire de moi ce que je suis. Cette engouement pour de longs voyages culturels a aussi largement contribué à attiser ma curiosité. Cela a aussi forgé ma capacité de concentration avec l'enchaînement des musées d'arts et d'histoire.

Enfin, en tant que deuxième partie de moi-même, il y a Iris. Elle aura vécu cette thèse dès son début au plus près malgré sa compréhension difficile du sujet. C'est avec grande patience et une grande tendresse qu'elle m'a entendu relater ces espoirs de conquêtes qui me menaient au loin. De nombreuses frustrations sont encore venues entraver notre chemin commun, mais nous les affrontons avec panache. Faisant elle-même une thèse de philosophie, elle se rend bien compte elle-même de la difficulté de rédiger un document aussi long et aussi accompli. On s'est ainsi encouragé mutuellement à mener à leur terme ces grandes œuvres que nous voulons voir faire date ! Cette thèse fait déjà plus de pages qu'il n'y a de jours dans l'année et j'aimerai qu'elle puisse avoir la richesse d'une année avec elle. Le choix d'une carrière académique pour nous deux

n'annonce pas les augures les plus favorables pour pouvoir être ensembles au jour le jour. Je ferai néanmoins tout mon possible pour aussi réaliser ce rêve. Iris, sache bien que les moments que nous passons ensembles sont de purs instants de bonheur et de délice. A mieux savoir les retrouver et les pérenniser, je pourrai toujours mieux m'accomplir en tant qu'homme.

Table des matières

Table des figures

0 Introduction

0.1 Introduction à la quasi-stationnarité

0.1.1 Notations préliminaires

0.1.1.1 Processus de Markov homogène avec extinction et variable de biais multiplicative

Vu les applications qui vont nous occuper, nous avons choisi de privilégier les processus de Markov X homogène en temps continu, avec une extinction au temps τ_∂ et potentiellement une variable de biais au temps t donnée par la fonctionnelle multiplicative Z_t (non-nulle), comme explicité ci-après. L'espace d'état sur lequel évolue ce processus est noté $\mathcal{X}$, que l'on supposera par commodité Polonais (i.e. métrique, séparable et complet). Ce dernier est muni de la tribu Borélienne notée $\mathcal{B}$, ce qui définit les fonctions mesurables sur cet espace. La propriété de Markov est énoncée par rapport à une filtration $(\mathcal{F}_t)_{t\geq 0}$, qui peut être étendue au-delà de la filtration standard $(\mathcal{F}_t^X)_{t\geq 0} = \sigma(X_s\ s \leq t)$, et pour laquelle le processus $(Z_t)_{t\geq 0}$ est aussi adapté. L'espace de probabilité est noté Ω et pour tout $x \in \mathcal{X}$ je noterai $\mathbb{P}_x$ la loi du processus $(X_t)_{t\geq 0}$ partant de la condition initiale $x \in \mathcal{X}$. Pour des conditions initiales plus générales en terme d'une mesure positive μ sur $\mathcal{X}$, je noterai : $\mathbb{P}_\mu := \int_{\mathcal{X}} \mu(dx)\mathbb{P}_x$, de telle manière que $\mathbb{P}_\mu(X_0 \in dx) = \mu(dx)$.

Les notations avec tilde feront référence à la propriété de Markov (cf ci-dessous) : elles désignent les fonctionnelles d'une copie $\widetilde{X}$ du processus avec condition initiale X_T et indépendante de X conditionnellement à X_T, pour un certain temps d'arrêt T donné.

Definition 1. *Propriété de Markov forte avec extinction : pour tout $s > 0$, tout temps d'arrêt T, toute fonction mesurable positive g et toute condition initiale $x \in \mathcal{X}$:*

$$\mathbb{E}_x[g(X_{T+s}) \ ;\ T + s < \tau_\partial \,\big|\, \mathcal{F}_T] = \mathbb{E}_{X_T}[g(\widetilde{X}_s)]\ , \qquad \mathbb{P}_x \text{ p.s. sur } \{T < \tau_\partial\}, \tag{0.1.1}$$

où $\{T < \tau_\partial\}$ correspond bien à T fini sur l'événement $\{\tau_\partial = \infty\}$.

Une représentation trajectorielle de cette propriété peut être donnée dès lors qu'on se ramène à un processus absorbé, comme on le verra en Section 0.1.3.4. On renvoie notamment aux conditions énoncées au Lemme 0.1.3.3.

Conditions sur $(Z_t)_{t\geq 0}$: En définissant Z comme fonctionnelle multiplicative du processus X vérifiant la propriété de Markov forte, on suppose que Z n'atteint pas 0 (sans quoi il faudrait changer le temps d'arrêt τ_∂) et que pour tout $s > 0$, tout temps d'arrêt T, toute fonction mesurable positive g et toute condition initiale $x \in \mathcal{X}$:

$$\mathbb{E}_x[g(X_{T+s})Z_{t+s} \;;\; T+s < \tau_\partial \,\big|\, \mathcal{F}_T] = Z_T \mathbb{E}_{X_T}[g(\widetilde{X}_s)\widetilde{Z}_s]\;, \quad \mathbb{P}_x \text{ p.s. sur } \{T < \tau_\partial\},$$

On requiert des propriétés très voisines entre celles portées par l'extinction et par un tel biais. Si cette pondération est majorée, on peut d'ailleurs décrire le semi-groupe associé au système via le semi-groupe d'un processus avec pondération uniforme, en changeant le taux d'extinction. Dans le cas contraire, il faudra certes la prendre en compte mais les raisonnements sur les processus avec extinction mais sans biais doivent s'étendre assez directement dans le cas avec biais.

Même si les techniques que j'utilise reposent de manière très conséquente sur la propriété de Markov forte, pour les propositions qui suivent la propriété de Markov faible suffit. Par rapport aux définitions précédente, dans la version affaiblie seule les choix de T comme temps déterministe sont considérés.

0.1.1.2 Description en terme de semi-groupe

Les conditions énoncées en début de Section sur le processus se traduisent par sa description en terme d'une suite d'opérateurs, notée $(P_t)_{t\geq 0}$. Ceux-ci sont définis de la manière suivante par leur action "à gauche" sur les fonctions mesurables f positives sur $\mathcal{X}$ par :

$$P_t f(x) := \mathbb{E}_x(f(X_t) \times Z_t \;;\; t < \tau_\partial), \tag{0.1.2}$$

où je rappelle que Z définit une fonctionnelle multiplicative adaptée à $(\mathcal{F}_t^X)$. Les conditions de Markov faible évoquées plus haut sont exactement ce qui garantit que cette suite d'opérateurs vérifie la relation de semi-groupe :

$$\forall\, s,t \geq 0, \quad P_s.P_t.f = P_t.P_s.f = P_{t+s}f. \tag{0.1.3}$$

Ce semi-groupe agit aussi à droite sur les mesures positives μ sur $\mathcal{X}$ via :

$$\mu P_t(A) := \int_{\mathcal{X}} \mu(dx)\mathbb{E}_x(Z_t \;;\; X_t \in A\;,\; t < \tau_\partial). \tag{0.1.4}$$

Comme pour tout opérateur linéaire, ces deux actions sont associatives, i.e. $\int_{\mathcal{X}} (\mu.P_t)(dx) f(x) = \int_{\mathcal{X}} \mu(dy)(P_t.f)(y)$. Cela m'a motivé à utiliser les notations suivantes de dualité, en reprenant les notations précédentes :

$$\langle \mu \,\big|\, f \rangle = \int_{\mathcal{X}} \mu(dx) f(x),$$
$$\langle \mu.P_t \,\big|\, P_s.f \rangle = \langle \mu.P_{t+s} \,\big|\, f \rangle = \langle \mu \,\big|\, P_{t+s}.f \rangle := \langle \mu \,\big|\, P_{t+s} \,\big|\, f \rangle.$$

J'utiliserai ces différentes formulations selon ce sur quoi je veux mettre l'accent.

Je dirai qu'un semi-groupe $(Q_t)_{t \geq 0}$ est conservatif si pour tout $x \in \mathcal{X}$ et $t \geq 0$, $Q_t(x, \mathcal{X}) = 1$. Dans cette thèse, je réserve la notation $(Q_t)_{t \geq 0}$ aux semi-groupes conservatifs, et note $(P_t)_{t \geq 0}$ les semi-groupes plus généraux, a priori non-conservatifs. De manière analogue, le semi-groupe $(P_t)_{t \geq 0}$ sera dit sous-conservatif si pour tout $x \in \mathcal{X}$ et $t \geq 0$, $P_t(x, \mathcal{X}) \leq 1$. Par extension, on emploiera aussi ces notions pour le processus biaisé avec extinction. Des définitions analogues pourraient être employées avec $\mathcal{X}$ pondéré par une fonction V pour définir la V-conservativité (respectivement la V-sous-conservativité) en considérant $\langle \delta_x.P_t \,\big|\, V \rangle$ au lieu de $P_t(x, \mathcal{X})$, cf Section 0.1.2.1.

0.1.1.3 Notations pour les espaces d'action du semi-groupe

Les notations pour les ensembles de fonctions $\mathcal{F}_+(\mathcal{X})$, $\mathcal{F}_b(\mathcal{X})$, $C_b(\mathcal{X})$, $C^k(\mathcal{X})$ et $C_b^k(\mathcal{X})$, pour $k \geq 1$, désigneront respectivement les fonctions mesurables positives, mesurables bornées, continues bornées, dérivables jusqu'à l'ordre k et dérivables jusqu'à l'ordre k avec des dérivées bornées jusqu'à cet ordre. Les notations pour les ensembles de mesures $\mathcal{M}_+(\mathcal{X})$, $\mathcal{M}_1(\mathcal{X})$ et $\mathcal{M}(\mathcal{X})$ désigneront resp. les mesures positives finies, les mesures de probabilités et les mesures signées sur $\mathcal{X}$. D'autres notations seront introduites en Section 0.1.2.2 sur des questions d'irréductibilité et en Section 0.1.2.3 pour prendre en compte une pondération de $\mathcal{X}$ par une fonction de Lyapunov.

0.1.1.4 Mesure quasi-stationnaire et capacité de survie

On dit que α est une mesure quasi-stationnaire pour le semi-groupe $(P_t)_{t \geq 0}$ avec λ pour taux de croissance si pour tout $t \geq 0$:

$$\alpha P_t = e^{\lambda\, t}\, \alpha, \text{ ce qui implique en particulier } \forall\, t \geq 0,\ \mathbb{E}_\alpha(Z_t \ ;\ t < \tau_\partial) = e^{\lambda\, t} \tag{0.1.5}$$

- cf e.g. le Théorème 2.2 de [CMS13]. Cette notion généralise au cadre non-conservatif la propriété de stationnarité du cadre conservatif, avec $\lambda = 0$. Par habitude (bien partagée !), je reprendrai le sigle anglais de QSD pour les mesures quasi-stationnaires.

Dans un cadre sous-conservatif, en général avec juste de l'extinction mais pas de biais, $-\lambda \geq 0$ est parfois dénommée comme le taux d'extinction. Le nom d'exposant de persistance est aussi employé, surtout avec des résultats de convergence vers la capacité de survie, définie comme suit (voir aussi la Section 0.1.3.2).

On dit qu'une fonction h est une capacité de survie associée à α si elle est positive, $\langle\alpha\,\big|\,h\rangle = 1$ et si pour tout $t \geq 0$:

$$\forall\, t \geq 0,\ P_t\, h = e^{\lambda\, t} h. \tag{0.1.6}$$

Cela implique qu'elle appartient au domaine du générateur infinitésimal $\mathcal{L}$ et satisfait : $\mathcal{L}h = \lambda h$. Dans le cas d'un processus de Markov absorbé, cela signifie aussi que $(h(X_t)\, Z_t\, e^{-\lambda t}\mathbf{1}_{\{t<\tau_\partial\}})_{t\geq 0}$ est une martingale locale.

On a alors la propriété d'unicité suivante, facile à vérifier : si h_1 et h_2 sont deux capacités de survie bornées associée à α, alors elles coïncident sur

$$\left\{x \in \mathcal{X} \,\middle|\, \left\|\delta_x.e^{-\lambda t} P_t(dy) - \langle \delta_x e^{-\lambda t} P_t \,\big|\, \mathbf{1}\rangle \times \alpha(dy)\right\|_{TV} \to 0 \text{ as } t \to \infty\right\}.$$

Ce résultat a son équivalent pour un espace pondéré par une fonction, comme énoncé en Section 0.1.2.1.

0.1.1.5 Transformée de Doob et Q-processus

Avoir une telle capacité de survie permet de définir un autre processus de Markov via ce qu'on appelle classiquement la transformée de Doob ou la h-transformée du processus, définie comme suit pour $t \geq 0$:

$$Q_t := e^{-\lambda t} B[1/h].P_t.B[h]\ , \quad \text{avec } B[g].f(x) := g(x)\, f(x), \quad x \in \mathcal{X}, f, g \in \mathcal{F}_{>0}(\mathcal{X}) \tag{0.1.7}$$

$$\text{au sens où } Q_t(x, dy) := e^{-\lambda t} \frac{h(y)}{h(x)} P_t(x, dy), \quad x \in \mathcal{X}.$$

Une propriété essentielle de ce processus, outre qu'il est clairement de Markov, est le fait qu'il est conservatif sous réserve de condition d'intégrabilité. Cela se vérifie aisément via la propriété de martingale locale de $(h(X_t)\, Z_t\, e^{-\lambda t}\, \mathbf{1}_{\{t<\tau_\partial\}})_{t\geq 0}$. On remarque alors immédiatement que $\beta = \alpha B[h]$ définit une mesure stationnaire du Q-processus. Cela se traduit par la propriété de quasi-ergodicité suivante sur le processus biaisé avec extinction : pour toute fonction $f \in \mathcal{F}_+(\mathcal{X})$:

$$\mathbb{E}_\alpha[((1/t) \times \int_0^t f(X_s)\, ds) \times h(X_t)\, Z_t\ ;\ t < \tau_\partial] = \langle \alpha \,\big|\, h \times f\rangle = \langle \beta \,\big|\, f\rangle. \tag{0.1.8}$$

Cette propriété justifie l'appellation de β en tant que mesure quasi-ergodique du processus biaisé avec extinction. Cela sera d'autant plus justifié avec les convergences donnés en Section 0.1.3.2, comme cela est aussi décrit dans [HZZ19].

0.1.2 Critères pour une convergence exponentielle vers une unique mesure stationnaire

Avant-propos

La littérature portant sur l'ergodicité de processus de Markov et sur leur extension à des cadres non-conservatifs (cf paragraphe suivant) est extrêmement vaste. La bibliographie collectée par Pollett dans [Po15] est à ce titre tout à fait impressionnante ! Heureusement, ce sujet a aussi fait l'objet de plusieurs articles de synthèse. Citons en particulier la présentation générale dans [CMS13], celle consacrée aux espaces discrets dans [DP13] ou plus spécifiquement sur les modèles de dynamique de population dans [MV12].

J'ai dû orienter mes recherches dans la littérature pour me concentrer sur les approches constructives et assez générales, qui vont au-delà d'un espace d'état discret ou en temps discret. Il est probable que de nombreux résultats dont je n'ai pas connaissance, exploitant par exemple des méthodes basées sur la réversibilité ou des techniques perturbatives, apporteraient des éclaircissements supplémentaires et je n'ai aucune prétention d'exhaustivité.

Il se trouve d'ailleurs que cette section 0.1.2 reprend largement les résultats énoncés dans l'article [KM03], que j'ai en réalité découverts tardivement. Ce dernier porte a priori sur l'établissement de propriété de Grandes Déviations sur la mesure empirique d'un processus de Markov conservatif. Il n'y est nullement mentionné la quasi-stationnarité, qui est néanmoins abordée à travers la notion d'ergodicité multiplicative. Je me suis particulièrement intéressé à cet article lorsque j'ai voulu obtenir de telles propriétés de Grandes Déviations en conservant le conditionnement qui a fondé mon analyse de la quasi-stationnarité. J'ai ainsi réalisé la très forte connexion entre ces méthodes de Grandes Déviations et celles pour établir la quasi-stationnarité. La thèse de G. Ferré est à ce titre une bonne introduction, cf [Fe20]. C'est en cherchant à appliquer les résultats des deux premiers chapitres à ce cadre de Grandes Déviations que j'ai pu juger l'importance des résultats énoncés dans [KM03], ne serait-ce qu'en termes de quasi-stationnarité.

Il me semble donc essentiel de replacer mes résultats dans le cadre plus général énoncé dans [KM03], où s'inscrivent aussi plusieurs travaux récents. Cela permet en effet d'exploiter ces résultats précédents, notamment vis-à-vis de leurs conséquences en terme de Grandes Déviations. Il était de mon projet d'inclure un tel chapitre sur ce sujet dans ma thèse (sans doute avant de finaliser le chapitre 5) mais le temps m'a de toute façon manqué pour relier correctement l'analyse initiée dans [KM03], [KM05] puis dans les travaux de G. Ferré [Fe20] avec mes résultats préliminaires.

0.1.2.1 Espaces pondérés par une fonction

De manière générale, on ne s'attend pas à ce que la convergence, vers une QSD notamment, soit uniforme en la condition initiale. On introduit donc des fonctions, dites de Lyapunov, pour contrôler le temps de parcours nécessaire avant que le mélange puisse vraiment s'effectuer avec les lois stationnaires. Les résultats de convergence peuvent alors être énoncés plus directement en introduisant une pondération de l'espace d'état $\mathcal{X}$ via la fonction de Lyapunov. J'introduis dans ce paragraphe une présentation générale de ce type de pondération.

On considère donc une fonction $V : \mathcal{X} \mapsto [0, \infty]$, éventuellement infinie sauf au moins en un point. L'ensemble suivant est ainsi non-vide :

$$S_V := \{x \in \mathcal{X} \,\big|\, V(x) < \infty\}. \tag{0.1.9}$$

En pratique, on s'attend à ce que $S_V = \mathcal{X}$ ou au moins à ce que $\psi(S_V^c) = 0$ pour une mesure d'irréductibilité ψ (cf. le paragraphe suivant sur cette notion).

Soit L_∞^V l'espace vectoriel des fonctions mesurables $f : \mathcal{X} \mapsto \mathbb{R}$ pour lesquelles le ratio f/V définit une fonction bornée sur $\mathcal{X}$. La norme naturelle sur cet espace est dénotée $\|.\|_V$ et définie par :

$$\|f\|_V := \sup_{x \in \mathcal{X}} |f(x)|/V(x).$$

On s'intéresse aussi à normer les noyaux P sur $\mathcal{X}$ de la forme $(P(x, dy))_{x \in \mathcal{X}}$. La norme supremum associée à L_∞^V est ainsi naturellement définie comme suit :

$$|||P|||_V := \sup \left\{ \frac{\langle \delta_x \,\big|\, Pf \rangle}{\|f\|_V} \,\Big|\, x \in \mathcal{X} \,,\; f \in L_\infty^V \,,\; \|f\|_V > 0 \right\}.$$

En vue de relier nos résultats au cadre proposé dans un article récent sur ces questions [BCGM19], il est aussi intéressant de se placer dans le cadre d'opérateurs agissant sur les mesures. Car, comme elle est employée dans [BCGM19] que je reprends, la norme $|||.|||_V$ s'interprète aussi comme la norme d'opérateur sur l'espace vectoriel des mesures signées $\mathcal{M}(V)$ sur $\mathcal{X}$ qui intègrent V. Considérons d'abord l'espace $\mathcal{M}_+(V)$ des mesures positives qui intègrent V et est ainsi relié à l'espace $\mathcal{M}_+(\mathcal{X})$ des mesures positives finies via l'isomorphisme : $B[V] : \mathcal{M}_+(V) \to \mathcal{M}_+(\mathcal{X})$, d'inverse $B[1/V]$, défini comme suit pour toute mesure $\mu \in \mathcal{M}_+(V)$:

$$\mu B[V](dx) := V(x)\mu(dx). \tag{0.1.10}$$

Cet isomorphisme s'étend naturellement aux mesures signées, et on déduit de la décomposition de Hahn-Jordan des mesures signées finies que pour toute mesure $\mu \in \mathcal{M}(V)$, il existe un unique couple de mesures (μ_+, μ_-) dans $\mathcal{M}_+(V)$ mutuellement singulières et telles que $\mu = \mu_+ - \mu_-$. $\mathcal{M}(V)$ est alors naturellement muni de la norme suivante par isomorphisme :

$$\|\mu\|_{M(V)} := \|\mu B[V]\|_{TV} = \langle \mu_+ \Big| V \rangle + \langle \mu_- \Big| V \rangle. \tag{0.1.11}$$

Si on suppose de plus que V est uniformément minorée par 1, $\mu(\mathcal{D})$ est toujours finie pour $\mu \in \mathcal{M}(V)$ et $\mathcal{D}$ mesurable de $\mathcal{X}$, et bornée par $\|\mu\|_{M(V)}$. On définit aussi $\mathcal{M}_1(V)$ par l'isomorphisme sur les mesures de probabilités, i.e. $\mu \in M_1(V)$ si μ est positive et satisfait $\langle \mu \Big| V \rangle = 1$.

0.1.2.2 ψ-irréductibilité

On considère ici un semi-groupe $(P_t)_{t\geq 0}$ qui satisfait pour un certain $\theta_* \in \mathbb{R}$ en tant qu'automorphisme de L^V_∞ (ou de $\mathcal{M}(V)$, ce qui serait équivalent) :

$$\sup_{t\geq 0} e^{-\theta_* t} \|P_t\|_V < \infty. \tag{0.1.12}$$

Pour tout $\theta > \theta_*$, on peut ainsi définir le noyau résolvant, comme opérateur à gauche sur L^V_∞ et à droite sur $\mathcal{M}(V)$, par :

$$R_\theta := \int_0^\infty \theta e^{-\theta t} P_t dt,$$

avec la convention que $R := R_1$ (sous réserve que $\theta_* < 1$). Le semi-groupe est appelé ψ-irréductible pour une certaine mesure positive ψ sur $\mathcal{X}$ sous la condition suivante : il existe $\theta > \theta_*$ tel que pour toute fonction $s : \mathcal{X} \mapsto \mathbb{R}_+$ qui satisfait $\langle \psi \Big| s \rangle > 0$:

$$\langle \delta_x . R_\theta \Big| s \rangle > 0, \quad x \in \mathcal{X}.$$

Sans restriction, on suppose aussi que ψ est maximale au sens où toute autre mesure ψ' pour laquelle le processus est ψ'-irréductible est elle-même absolument continue par rapport à ψ [MT93]. L'ensemble des fonctions $s : \mathcal{X} \mapsto \mathbb{R}_+$ qui satisfont $\langle \psi \Big| s \rangle > 0$ est noté $\mathcal{F}_\psi(\mathcal{X})$, et ces fonctions sont appelées ψ-positives.

Une autre propriété essentielle est celle d'apériodicité. Elle est satisfaite dès lors que pour toute fonction $s \in \mathcal{F}_\psi(\mathcal{X})$ et $x \in \mathcal{X}$, $\langle \delta_x P_t \Big| s \rangle > 0$ pour tout t suffisamment grand.

Via l'irréductibilité, on peut s'intéresser à l'espace d'état accessible, qui est donné par le support de ψ. Un ensemble mesurable E est ainsi appelé plein si $\psi(E^c) = 0$. Il est appelé absorbant si pour tout $x \in E$, $\delta_x R(E^c) = 0$. Le lecteur pourra trouver à la Proposition 4.2.3 de [MT93] la preuve que pour tout processus de Markov ψ-irreductible, un ensemble absorbant non vide est forcément plein.

Une fonction $s \in \mathcal{F}_\psi(\mathcal{X})$ et une mesure positive ψ sur $\mathcal{X}$ sont nommées "small" (le terme "petite" ayant un autre usage) si pour un certain $\theta > \theta_*$:

$$R_\theta \geq |s\rangle \otimes \langle\zeta| \text{ au sens où } \delta_x R_\theta(A) \geq s(x)\zeta(A), \quad x \in \mathcal{X}, A \in \mathcal{B}. \tag{0.1.13}$$

En adaptant la preuve de la Proposition 5.5.5 de [MT93], on voit qu'on peut toujours trouver pour chaque processus X ψ-irréductible (conservatif) un $\theta > 0$ et un couple (s, ζ) qui satisfait (0.1.13) et pour lequel s est strictement positive sur $\mathcal{X}$ et ζ équivalente à toute mesure maximale d'irréductibilité ψ. Si une fonction "small" s est strictement minorée sur un ensemble mesurable E, alors E est aussi nommé "small". L'ensemble de ces fonctions "small" est dénoté $\mathcal{F}_\psi^\circ(\mathcal{X})$, tandis que $\mathcal{M}_+^\circ(\mathcal{X})$ dénote l'ensemble des mesure positives "small". Ces deux ensembles sont des cônes positifs, stables par addition.

Ces conditions d'irréductibilité et d'apériodicité sont essentielles pour les semi-groupes conservatifs. Même si la situation est plus complexe pour les semi-groupes non-conservatifs (cf discussion en Section 0.1.3.5), ces notions jouent tout de même un rôle clé pour la compréhension. Il est utile de préciser que restreindre l'espace d'état via un temps d'extinction τ'_∂ (antérieur à τ_∂) peut amener à exhiber une mesure d'irréductibilité ψ' singulière avec ψ. La survie associée à cette dernière est potentiellement moins bonne que celle de ψ'. On peut donc se demander s'il n'y a pas une restriction de l'espace d'état dont mesure d'irréductibilité aurait le taux de survie optimal.

Pour l'exemple, considérons un modèle source-puits : seule la transition de la source vers le puits a lieu, à un taux positif, avec un taux d'extinction ρ_1 depuis la source S et $\rho_0 > \rho_1$ depuis le puits P. Alors, $\psi = \delta_P$ est irréductible et maximale avec τ_∂. Si on définit τ'_∂ comme le temps de sortie de la source, il n'est pas difficile de voir que $\psi' = \delta_S$ est irréductible et maximale avec τ'_∂. Quand bien même $\tau'_\partial \leq \tau_\partial$ p.s., ψ' est associée à un taux d'extinction ρ_1 plus faible que ρ_0 auquel est associé ψ. Et il existe une mesure quasi-stationnaire qui charge les deux états et dont le taux d'extinction est exactement ρ_1.

0.1.2.3 Conditions de Lyapunov

Les conditions de Lyapunov que l'on considère prennent l'une des deux formes suivantes, tirées de [MT93]. La première permet de garantir l'ergodicité de manière générale, sans précision sur la vitesse de convergence, contrairement à la seconde où celle-ci sera nécessairement exponentielle.

(LG) $\mathcal{L}V \leq -f + bs\,, \qquad R \geq |s\rangle \otimes \langle\zeta|,$
où $f : \mathcal{X} \mapsto [1, \infty)$, $b \in (0, \infty)$, $s : \mathcal{X} \mapsto (0, 1]$ et $\zeta \in \mathcal{M}_1(\mathcal{X})$.

(LE) $\mathcal{L}V \leq -\delta V + bs\,, \qquad R \geq |s\rangle \otimes \langle\zeta|,$
où $V : \mathcal{X} \mapsto [1, \infty]$, $\delta, b \in (0, \infty)$, $s : \mathcal{X} \mapsto (0, 1]$ et $\zeta \in \mathcal{M}_1(\mathcal{X})$.

Il est à noter que (LE) implique (LG). Ce cadre de convergence à vitesse exponentielle est particulièrement justifié en vue d'étendre ces résultats à des semi-groupes non-conservatifs pour lesquels un déclin ou une croissance exponentielle est attendue. Un processus de Markov est ainsi appelé exponentiellement ergodique avec V pour fonction de Lyapunov s'il est ψ-irréductible, apériodique et s'il satisfait (LE) (avec cette fonction V).

0.1.2.4 Théorèmes ergodiques

Le théorème qui va suivre énonce que (LG) implique l'existence et l'unicité de la mesure invariante sur $\mathcal{X}$ pour le semi-groupe Q_t, notée π. Etant donné une telle mesure π, on définit aussi le noyau $\Pi := \mathbf{1} \otimes \pi$, i.e. $\langle \delta_x \Pi h \rangle = \langle \pi \,\big|\, h \rangle$ pour $x \in \mathcal{X}$ et $h \in L_\infty$, vers lequel doit tendre Q_t. Sous réserve que $\langle \pi \,\big|\, V \rangle < \infty$, Π définit un automorphisme de L_∞^V et il est élémentaire que $\|\Pi\|_V = \langle \pi \,\big|\, V \rangle$.

Theorem 0.1.1 ([KM03])**.** *Soit X un processus de Markov (conservatif) ψ-irréductible et apériodique. Pour toute fonction $f : \mathcal{X} \mapsto [1, \infty)$, les assertions suivantes sont équivalentes :*

(i) X est récurrent positif de probabilité invariante π avec $\langle \pi \,\big|\, f \rangle < \infty$.

(ii) Il existe un ensemble "small" $\mathcal{D}$ tel que :

$$\sup_{x \in \mathcal{D}} \mathbb{E}_x [\int_{[0,\tau_\mathcal{D})} f(X_t) dt] < \infty \,, \qquad \text{où } \tau_\mathcal{D} := \inf\{t \geq 1 \;;\; X_t \in \mathcal{D}\}.$$

(iii) La condition (LG) est valide avec cette fonction f pour une certaine fonction V.

Si ces conditions sont vérifiées, l'ensemble S_V défini en (0.1.9) avec V donné par (LG) est absorbant et plein. En outre, pour tout $x \in S_V$:

$$\sup_{g : |g| \leq f} \{ |\langle \delta_x \,\big|\, Q_t - \Pi \,\big|\, g \rangle| \} \to 0 \text{ as } t \to \infty.$$

La version exponentielle de ce théorème est la suivante :

Theorem 0.1.2 ([KM03])**.** *Soit X un processus de Markov (conservatif) ψ-irréductible et apériodique. Pour toute fonction $V : \mathcal{X} \mapsto [1, \infty)$, les assertions suivantes sont équivalentes :*

(i) Q_t converge vers Π en norme $\|.\|_V$ pour une certaine mesure de probabilité π, qui est alors nécessairement l'unique mesure invariante.

(ii) Il existe un ensemble "small" $\mathcal{D}$ et $\rho > 0$ tel que :

$$\sup_{x \in \mathcal{D}} \mathbb{E}_x \exp[\rho \tau_\mathcal{D}] < \infty \,, \qquad \text{où } \tau_C := \inf\{t \geq 1 \;;\; X_t \in \mathcal{D}\}.$$

(iii) La condition (LE) est valide avec cette fonction V.

Si ces conditions sont vérifiées, l'ensemble S_V défini en (0.1.9) avec V donné par (LG) est absorbant et plein. En outre, il existe $C, \gamma > 0$ tels que pour tout $t \geq 0$:

$$\|Q_t - \Pi\|_V \leq Ce^{-\gamma t}.$$

0.1.3 Semi-groupe de Markov non-conservatifs et mesures quasi-stationnaires

0.1.3.1 Valeur propre généralisée

En vue de relier nos résultats à ceux de [BCGM19] et [KM03], il me semble pertinent d'établir la nature des résultats de convergence attendus pour des semi-groupes définis de manière assez générale sur L_∞^V et $\mathcal{M}_+(V)$. Dans les résultats de ma thèse, ils seront toujours énoncés pour V constante à 1, sauf en ce qui concerne le Q-processus.

Dans la suite, nous ferons l'hypothèse (BV) suivante sur le semi-groupe $(P_t)_{t\geq 0}$: $(P_t)_{t\geq 0}$ est ψ-irréductible et apériodique et il existe $T > 0$ tel que la norme $\|P_t\|_V$ est uniformément bornée pour tout $t \leq T$. Il n'est pas difficile de voir par récurrence que cette propriété s'étend pour tout $T > 0$ et permet de borner la croissance asymptotique en t de $\|P_t\|_V$.

Theorem 0.1.3 ([KM03]). *Pour tout semi-groupe $(P_t)_{t\geq 0}$ satisfaisant (BV) il existe un unique $\lambda \in [-\infty, \infty)$ tel que pour tout $x \in \mathcal{X}$ et $s \in \mathcal{F}_+^\circ(\mathcal{X})$:*

$$\int_0^\infty e^{-\widetilde{\lambda} t} \langle \delta_x P_t \,\big|\, s\rangle dt \begin{cases} = \infty \text{ pour tout } x \in \mathcal{X}, \widetilde{\lambda} < \lambda, \\ < \infty \text{ pour presque tout } x \in \mathcal{X}[\psi], \widetilde{\lambda} > \lambda. \end{cases}$$

Dans le cas $\widetilde{\lambda} = \lambda$, on peut faire la distinction suivante, qui ne dépend pas du choix du couple (ζ, s) qui satisfait (0.1.13) :

— *$(P_t)_{t\geq 0}$ est λ-transient si $\int_0^\infty e^{-\lambda t}\langle \zeta \,\big|\, P_t s\rangle dt < \infty$*

— *$(P_t)_{t\geq 0}$ est λ-récurrent si $\int_0^\infty e^{-\lambda t}\langle \zeta \,\big|\, P_t s\rangle dt = \infty$*

Cette valeur λ est alors nommée la valeur propre principale généralisée, ce qui se justifie par le théorème qui suit. L'existence d'une fonction propre minimale est ainsi rendue équivalente à la λ-récurrence du semi-groupe $(P_t)_{t\geq 0}$.

Le Théorème 0.1.3 est tiré du Théorème 3.1 de [KM03], dont les auteurs s'appuient sur [Nu84],en particulier le Théorème 3.2 de ce dernier, et l'adaptent au cas du temps continu. La preuve repose sur le noyau résolvant R_θ, qui est bien défini d'après l'hypothèse (BV) pour θ assez grand et qui satisfait pour une certaine couple (ζ, s) "small" :

$$R_\theta \geq |s\rangle \otimes \langle \zeta|.$$

Cette approche se révèle à nouveau féconde pour obtenir une fonction propre pour le semi-groupe. Il s'agit d'abord de déterminer le rayon de convergence r_θ associé à R_θ, en remarquant que $\langle \zeta \,\big|\, (R_\theta)^k \,\big|\, s \rangle$ définit pour $k \geq 1$ une suite sur-multiplicative. La propriété suivante va ainsi caractériser r_θ :

$$\sum_{k=0}^{\infty} r^k \langle \delta_x \,\big|\, (R_\theta)^k \,\big|\, s \rangle \begin{cases} = \infty \text{ pour tout } x \in \mathcal{X}, r > r_\theta, \\ < \infty \text{ pour } \psi\text{-presque tout } x \in \mathcal{X}, r < r_\theta. \end{cases} \tag{0.1.14}$$

On peut alors définir le candidat suivant :

$$\check{h} := \sum_{k=0}^{\infty} (r_\theta)^k \, R_\theta . (R_\theta - |s\rangle \otimes \langle \zeta |)^k . s. \tag{0.1.15}$$

Nous sommes maintenant en mesure d'énoncer le théorème suivant, qui détermine bien λ comme valeur propre principale dans le cas λ-récurrent et spécifie le cas transient.

Theorem 0.1.4 ([KM03]). *Supposons que $(P_t)_{t\geq 0}$ satisfait à l'hypothèse (BV) et rappelons que λ est la valeur propre principale généralisée définie au Théorème 0.1.3. Alors la fonction $\check{h}$ est finie ψ-presque sûrement et :*

(i) Si $(P_t)_{t\geq 0}$ est λ-récurrent, alors $\check{h}$ satisfait pour tout $t \geq 0$: $P_t \check{h} = e^{\lambda t} \check{h}$.

(ii) Si $(P_t)_{t\geq 0}$ est λ-transient, alors quelle que soit la fonction $s \in \mathcal{F}_+^\circ(\mathcal{X})$, il existe $\delta > 0$ tel que $P_1 \check{h} = e^\lambda \check{h} - \delta s$.

Ainsi dans le cas transient, il existe une solution $\bar{h}$ à l'inégalité :

$$P_1 \bar{h} \leq e^\lambda \bar{h}, \tag{0.1.16}$$

où $\bar{h}$ est ψ-presque sûrement finie avec une inégalité stricte dès lors que $\bar{h}(x) < \infty$.

(iii) La solution de (0.1.15) est minimale, donc essentiellement unique malgré l'arbitraire dans le choix de ζ et s, au sens suivant. Si $\bar{h} : \mathcal{X} \mapsto (0, \infty)$ satisfait l'inégalité (0.1.16), alors il existe $c > 0$ tel que pour tout $x \in \mathcal{X}$, $\bar{h}(x) \geq c\check{h}(x)$. Si $(P_t)_{t\geq 0}$ est en outre λ-récurrent, alors on peut choisir un tel $c > 0$ tel que ψ-p.s. $\bar{h} = c\check{h}$.

Le Théorème 0.1.4 reprend essentiellement l'énoncé du Théorème 3.3 de [KM03], lui-même déduit du Théorème 5.1 de [Nu84].

0.1.3.2 Quasi-ergodicité exponentielle

Nous proposons la définition suivante en tant que généralisation de la notion d'ergodicité exponentielle au cadre non-conservatif. Il est à noter que l'unicité ne sera requise que dans l'espace de mesures $\mathcal{M}_1(V)$, ce qui n'exclut pas l'existence d'autres QSD dans $\mathcal{M}_1(\mathcal{X})$.

Un semi-groupe homogène P_t qui agit à gauche sur L^V_∞ et à droite sur $\mathcal{M}_+(V)$ est dit exponentiellement quasi-ergodique s'il existe $\lambda \in \mathbb{R}$, $C, \gamma > 0$, $h \in L^V_\infty$ positive sur $\mathcal{X}$, et $\alpha \in \mathcal{M}_1(V)$ tels que la propriété suivante est vérifiée

$$\|e^{-\lambda t} P_t - |h\rangle \otimes \langle\alpha|\|_V \leq C\, e^{-\gamma t}, \tag{0.1.17}$$

où $|h\rangle \otimes \langle\alpha|$ est l'opérateur de projection, défini à gauche sur L^V_∞ (dont nous prenons un représentant f) et à droite sur $\mathcal{M}(V)$ (avec μ pour représentant) par :

$$((|h\rangle \otimes \langle\alpha|).f)(x) := \langle\alpha|f\rangle \times h(x) \;;\; (\mu.(|h\rangle \otimes \langle\alpha|))(dx) := \langle\mu|h\rangle\alpha(dx) \tag{0.1.18}$$
$$\text{de sorte que } \mu.(|h\rangle \otimes \langle\alpha|).f := \langle\mu|h\rangle \times \langle\alpha|f\rangle.$$

Ce résultat condensé (0.1.17) a les implications suivantes en vue d'interpréter α et h :

(UQS) α est l'unique mesure quasi-stationnaire de $\mathcal{M}(V)$, avec λ pour taux de croissance.

(UCS) h est une capacité de survie dans L^V_∞, avec λ comme taux de croissance. Elle est associée à α au sens où $\langle\alpha \,\big|\, h\rangle = 1$.
Plus précisément, il y a unicité de cette capacité de survie selon la propriété suivante. Sous l'hypothèse qu'une fonction positive $\widetilde{h}$ de L^V_∞ est telle que pour un certain $\widetilde{\lambda} \in \mathbb{R}$, $\mathcal{L}\,\widetilde{h} = \widetilde{\lambda}\,\widetilde{h}$ est vérifié, alors $\widetilde{h}$ est nécessairement proportionnelle à h et $\widetilde{\lambda} = \lambda$.

(0.1.17) est en réalité équivalente à la conjonction de convergence pour les deux limites, de la mesure renormalisée et de sa normalisation, pour lesquelles on a recours aux définitions suivantes :

$$\mu A_t(dx) := \frac{\mu P_t(dx)}{\langle \mu P_t V\rangle}, \quad \mu \in \mathcal{M}(V), \tag{0.1.19}$$

$$h_t(x) \;:= e^{-\lambda\, t}\langle\delta_x.P_t.V\rangle = \frac{\langle\delta_x.P_t.V\rangle}{\langle\alpha.P_t.V\rangle}, \quad x \in \mathcal{X}. \tag{0.1.20}$$

(CH) $h \in L^V_\infty$ positive sur $\mathcal{X}$, et $(h_t)_{t\geq 0}$ converge à taux exponentiel vers h en norme $\|.\|_V$.

(CA) $\alpha \in \mathcal{M}_1(V)$ et il existe $C, \gamma > 0$ tel que pour tout t assez grand pour vérifier $C\exp(-\gamma t) < \langle\mu \,\big|\, h\rangle/2$:

$$\|\mu A_t - \alpha\|_V \leq \frac{4C}{\langle\mu \,\big|\, h\rangle} e^{-\gamma t}. \tag{0.1.21}$$

On voit que (CA) implique que α est la mesure quasi-limite pour toute condition initiale dans $\mathcal{M}_1(V)$, au sens où pour tout ensemble $\mathcal{D}$ mesurable de S_V, on a :

$$\lim_{t\to\infty} \mu A_t(\mathcal{D}) = \alpha(\mathcal{D}).$$

Proposition 0.1.3.1. (0.1.17) *est équivalente à* $(CH)+(CA)$ *si on suppose en outre que* $\|P_t\|_V$ *est bornée sur tout intervalle de temps fini. Sous ces conditions,* (UQS) *et* (UCS) *sont vérifiées.*

La preuve de cette équivalence est élémentaire dans le sens réciproque et assez facile dans le sens direct en évaluant d'abord l'opérateur sur la fonction V. Les détails sont laissés au lecteur. Il est bien connu que (CA) implique (UQS) : via la preuve du Lemme 7.2 dans [CC+09], on déduit le fait que α est une QSD du fait que c'est une quasi-limite. L'unicité vient de ce que $\mu_* A_t = \mu_*$ pour toute QSD $\mu_* \in \mathcal{M}(V)$. Il est à noter que d'autres QSD peuvent exister hors de $\mathcal{M}(V)$. Déduire de (CH) que h est une capacité de survie associée à λ n'est pas beaucoup plus difficile à obtenir, et je renvoie à la Section 1.5.4 du Chapitre 1 pour plus de détails. Que $\langle\alpha \,\big|\, h\rangle = 1$ découle immédiatement de la convergence de h_t vers h en notant que $\langle\alpha \,\big|\, h_t\rangle = 1$ par définition de h_t. $\square$

0.1.3.3 Transformée de Doob et lien avec l'ergodicité classique

A tout semi-groupe $(P_t)_{t\geq 0}$ non-conservatif exponentiellement quasi-ergodique est naturellement associé un semi-groupe $(Q_t)_{t\geq 0}$ conservatif qui est exponentiellement ergodique via le Q-processus, au vu de la Proposition suivante.

Proposition 0.1.3.2. *Si* $(P_t)_{t\geq 0}$ *satisfait* (0.1.17), *alors avec des notations identiques,* $(Q_t)_{t\geq 0}$ *satisfait :*

$$\|Q_t - |\mathbf{1}\rangle \otimes \langle \alpha B[h]| \|_{V/h} \leq C e^{-\gamma t}. \tag{0.1.22}$$

Réciproquement, si h *est une capacité de survie du semi-groupe* $(P_t)_{t\geq 0}$ *et si le semi-groupe associé* $(Q_t)_{t\geq 0}$ *satisfait :*

$$\|Q_t - |\mathbf{1}\rangle \otimes \langle \pi| \|_{V} \leq C e^{-\gamma t}, \tag{0.1.23}$$

pour une certaine mesure de probabilité π *et une certaine fonction de Lyapunov* V *positive (non-nécessairement minorée par* 1*), alors* $(P_t)_{t\geq 0}$ *satisfait :*

$$\|P_t - |h\rangle \otimes \langle \pi B[1/h]| \|_{V\times h} \leq C e^{-\gamma t}. \tag{0.1.24}$$

La vérification de cette proposition est en réalité tout à fait élémentaire. Ce résultat a tout particulièrement été exploité dans [FRS19], [DM15] ou [KM03] pour déduire la convergence vers (0.1.17) depuis des résultats plus classiques d'ergodicité pour les semi-groupes conservatifs. Les méthodes qui déduisent les propriétés du semi-groupe conservatif depuis sa transformée de Doob sont génériquement regroupées sous le nom de R-théorie. D'après le théorème 0.1.2, on voit par exemple qu'une convergence en norme $\|.\|_V$ pour (P_t) avec $t \to \infty$ (y compris en variation totale) est nécessairement exponentielle, sous réserve que V/h est uniformément minorée. Cette approche nécessite néanmoins de pouvoir construire au préalable la capacité de survie. Vu le Théorème 0.1.4, cela peut se faire en justifiant la λ-récurrence.

0.1.3.4 Interprétation en terme d'absorption d'un processus neutre

Un des cadres naturels où on voit apparaître de tels semi-groupes non-conservatifs concerne le cas d'un processus de Markov fort restreint à ne pas avoir atteint un certain état absorbant. De manière générale, il peut être intéressant de représenter de tels semi-groupes sous cette forme. Après avoir précisé ce que l'on entend par cette représentation, et ce qu'elle permet, nous verrons dans quels cadre nous pouvons nous y ramener.

Ce cadre est celui d'un processus de Markov fort homogène en temps continu et défini sur l'espace d'état $\mathcal{X} \cap \{\partial\}$ via la donnée de $(\Omega, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (X_t)_{t\in\mathbb{R}_+}, (P_x)_{x\in\mathcal{X}\cap\{\partial\}})$. On appellera ∂ le "cimetière", et c'est un élément absorbant pour $\mathcal{X}$. En posant son temps d'atteinte $\tau_\partial := \inf\{t \geq 0 : X_t = \partial\}$, cela signifie que $X_t = \partial$ pour tout $t \geq \tau_\partial$. Sommairement, on désignera simplement le processus par X et pour un tel processus conservatif, la propriété de Markov forte peut s'énoncer comme classiquement de la façon suivante :

Propriété de Markov forte pour un processus conservatif : Pour tout temps d'arrêt T, toute fonction mesurable positive Φ du processus et toute condition initiale $x \in \mathcal{X}$:

$$\mathbb{E}_x(\Phi[(X_{T+s})_{s\geq 0}] \,\big|\, \mathcal{F}_T) = \mathbb{E}_{X_T}(\Phi[(\widetilde{X}_s)_{s\geq 0}])\,, \qquad \mathbb{P}_x \text{ p.s. sur } \{T < \infty\}.$$

τ_∂ est aussi appelé le temps d'extinction de $\mathcal{X}$ au sens où X restreint comme processus à valeur dans $\mathcal{X}$ satisfait à la propriété de Markov forte énoncée à la Définition 1.

Cette notion de processus avec extinction peut s'étendre assez naturellement pour des processus avec plusieurs états absorbants ou avec d'autres critères d'extinction. La propriété de Markov forte énoncée à la Définition 1 permet de faire le lien entre ces représentations. Il est à noter que l'on peut vouloir rajouter de l'information par rapport à la filtration standard pour définir l'événement d'extinction. Cela peut en particulier permettre de prendre en considération un taux d'extinction dépendant des valeurs de X_t.

Lemme 0.1.3.3. *Soit $(\Omega, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (X_t)_{t\in\mathbb{R}_+}, (\mathbb{P}_x)_{x\in\mathcal{X}})$ un processus de Markov fort (conservatif) avec $\mathcal{X}$ Polonais. Supposons que τ_∂ est un temps d'arrêt pour $(\mathcal{F}_t)$, fini p.s. et tel que la condition énoncée en Définition 1 est satisfaite. Définissons $Y_t := X_t \mathbf{1}_{\{t<\tau_\partial\}} + \partial \mathbf{1}_{\{\tau_\partial \leq t\}}$, i.e. plus généralement $(\Omega, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (Y_t)_{t\in\mathbb{R}_+}, (\mathbb{P}_x)_{x\in\mathcal{X}\cap\{\partial\}})$, avec sous $\mathbb{P}_\partial$: $Y_t \equiv \partial$ pour tout $t \geq 0$. Alors, Y est un processus de Markov fort* *conservatif. Pour ce processus Y, ∂ est le seul état absorbant, dont τ_∂ est le temps d'atteinte.*

Il va de soi que l'on peut aussi définir un tel Y en partant d'un processus de Markov fort dont τ_∂ est le temps d'extinction.

Preuve : Sous ces hypothèses, avec l'extension de g sur $\mathcal{X} \cup \{\partial\}$ par $g(\partial) = 0$:

$$\begin{aligned}
\mathbb{E}_\mu(g(Y_{T+s}) \,\big|\, \mathcal{F}_T) &= \mathbb{E}_\mu(g(X_{T+s}) \;;\; T+s < \tau_\partial \,\big|\, \mathcal{F}_T) \\
&= \mathbf{1}_{\{T<\tau_\partial\}} \mathbb{E}_{X_T}(g(\widetilde{X}_s) \;;\; s < \widetilde{\tau_\partial}) \qquad \text{par (0.1.1)}, \\
&= \mathbf{1}_{\{T<\tau_\partial\}} \mathbb{E}_{Y_T}(g(\widetilde{Y}_s) \;;\; s < \widetilde{\tau_\partial}).
\end{aligned}$$

Pour $\mathbf{1}_{\{\partial\}}$, qui vaut 1 sur ∂ et 0 sinon :

$$\begin{aligned}
\mathbb{E}_\mu(\mathbf{1}_{\{\partial\}}(Y_{T+s}) \,\big|\, \mathcal{F}_T) &= 1 - \mathbb{P}_\mu(T+s < \tau_\partial \,\big|\, \mathcal{F}_T) \\
&= 1 - \mathbf{1}_{\{T<\tau_\partial\}} \mathbb{P}_{Y_T}(s < \widetilde{\tau_\partial}) \qquad \text{par (0.1.1)}, \\
&= \mathbf{1}_{\{\tau_\partial \le T\}} + \mathbf{1}_{\{T<\tau_\partial\}} \mathbb{P}_{Y_T}(\widetilde{\tau_\partial} \le s) = \mathbb{E}_{Y_T}(\mathbf{1}_{\{\partial\}}(\widetilde{Y}_s)).
\end{aligned}$$

Par linéarité et en itérant la condition de Markov, il est ensuite élémentaire de conclure que Y satisfait à la propriété de Markov forte plus générale. Il est clair aussi que τ_∂ est le temps d'atteinte de ∂, qui est absorbant. Puisque τ_∂ est fini p.s., ∂ est bien le seul état absorbant.

□

Cette approche peut donc être employée très largement pour des temps de sortie de domaines assez généraux, avec des applications évidentes en terme de métastabilité. Comparer différents choix d'événements d'extinction pour un même processus peut aussi être informatif. Une des questions naturelles à se poser est par exemple de savoir dans quelle mesure le choix précis des frontières d'un tel domaine affecte la dynamique en temps long du processus non absorbé. On peut ainsi empiriquement avoir une distinction assez nette du bassin d'attraction même pour un processus stochastique diffusif qui visite tout le domaine !

Cette représentation en terme d'un processus avec biais et absorption n'apparaît pas toujours aussi naturellement que pour ces questions de métastabilité ou les modèles avec extinction explicite (taille de population qui tombe à 0, plus aucun individu, etc.). Il est important de savoir la reconnaître sur des équations d'évolution, comme dans le cadre du Chapitre 4.

Imaginons qu'un résultat de convergence amène à étudier un processus $(\mu_t)_{t\ge 0}$ à valeur dans $\mathcal{M}_+(\mathcal{X})$, supposé uniquement défini comme solution pour tout $f \in \mathcal{D}(\mathcal{L})$ de l'équation :

$$\partial_t \langle \mu_t \,\big|\, f \rangle = \langle \mu_t \,\big|\, \mathcal{L} f \rangle + \langle \mu_t \,\big|\, r \times f \rangle, \tag{0.1.25}$$

où r est une fonction mesurable et où $\mathcal{L}$ est le générateur infinitésimal d'un processus de Markov, avec pour domaine $\mathcal{D}(\mathcal{L})$. Il est alors très naturel de se demander si on peut pas identifier μ_t comme $\mu_0 P_t$, où P_t est défini pour toute fonction mesurable positive par :

$$\langle \mu P_t \,\big|\, f \rangle = \mathbb{E}_\mu(f(X_t) \exp[\textstyle\int_0^t r(X_s)\, ds] \; ; \; t < \tau_\partial). \tag{0.1.26}$$

On appellerait ceci la représentation de Feynman-Kac de la solution μ_t. De manière générale, cette identification n'est pas claire sans des résultats de régularité supplémentaire sur μ_t, qui doivent découler de conditions de régularité et de bornes sur les coefficients de $\mathcal{L}$ et r. C'est d'autant plus délicat s'il y a gérer des effets de bords.

Il est plus classique, et plus facile, d'identifier $P_t u_0$ comme l'unique solution de l'équation

$$\partial_t u_t(x) = \mathcal{L} u_t(x) + r \times u_t(x). \tag{0.1.27}$$

Le lien entre la représentation de

On retrouve de telles expressions de Feynman-Kac notamment dans le cadre du comportement en espérance de processus de branchement, sous l'autre nom de formule many-to-one, cf e.g. [Cl17], [KA06], [KL+97]. L'introduction de la thèse d'A. Marguet aborde aussi ce sujet, cf Section 2.4, en vue de son article [Al19]. On la voit aussi dans les modèles de croissance-fragmentation, cf e.g. [Be19]. On peut l'interpréter en décrivant X comme la dynamique spontanée d'un individu (ou d'un polymère) biaisée par son taux de reproduction pour tenir compte d'un échantillonnage uniforme au temps t dans la population. Ce taux de reproduction ne prend en compte que la différence entre le taux de multiplication et le taux de disparition. Une même interprétation sous-tend les équations de réaction-diffusion non-locales qui sont linéaires, comme mentionné au Chapitre 2. Elles sont notamment étudiées pour des enjeux d'invasion par une nouvelle population, cf e.g [CH20], [CDM08], ou de persistance d'une population mise en danger par un changement d'environnement, cf e.g. [BCV16], [BHR08].

De manière analogue, l'équation :

$$\partial_t \langle \mu_t \,\big|\, f \rangle = \langle \mu_t \,\big|\, \mathcal{L} f \rangle + \langle \mu_t \,\big|\, r \times f \rangle - \langle \mu_t \,\big|\, r \rangle \times \langle \mu_t \,\big|\, f \rangle, \tag{0.1.28}$$

permet d'identifier μ_t comme μA_t, comme défini en (0.1.19) avec $V \equiv 1$. On la retrouve notamment dans les limites d'échelles de population où la taille de cette population est contrainte à rester constante.

Il n'est pas difficile de voir que l'équation caractéristique de μA_t plus généralement avec V dans le domaine de $\mathcal{L}$ est :

$$\partial_t \langle \mu_t \,\big|\, f \rangle = \langle \mu_t \,\big|\, \mathcal{L} f \rangle + \langle \mu_t \,\big|\, r \times f \rangle - \langle \mu_t \,\big|\, \mathcal{L} V + r \times V \rangle \times \langle \mu_t \,\big|\, f \rangle. \tag{0.1.29}$$

Jusqu'à présent, je ne crois pas avoir été confronté à ce type d'équation.

De même, l'équation :

$$\partial_t f_t = \mathcal{L} f_t + r \times f_t, \tag{0.1.30}$$

permet d'identifier f_t comme $P_t f_0$.

0.1.3.5 La question de l'irréductibilité

Les résultats généraux que l'on présente ici sont certes énoncés dans des cas irréductibles, mais leur portée n'est pas limitée à ce cadre. Comme on le verra au Chapitre 4, ils peuvent permettre de traiter des processus pour lesquels il existe des domaines quasi-absorbants au sens suivant. Un domaine $\mathcal{D}$ est quasi-absorbant si pour tout $t \in [\tau_{\mathcal{D}}, \tau_\partial)$, $X_t \in \mathcal{D}$ où $\tau_{\mathcal{D}}$ désigne le temps d'atteinte de $\mathcal{D}$ par $\mathcal{X}$. L'existence d'un tel état quasi-absorbant n'empêche pas le processus d'être irréductible au sens de la Section 0.1.2.2, mais pourrait constituer une difficulté pour établir la dynamique en temps long.

Il peut être particulièrement utile de distinguer plusieurs événements d'extinction qui correspondent à l'absorption par différents domaines quasi-absorbants. On peut voir la dynamique du processus comme la descente d'une cascade qu'il ne pourrait plus remonter mais où il pourrait naviguer dans les bassins intermédiaires un temps indéfini. Être en dehors d'un domaine quasi-absorbant $\mathcal{D}$ peut alors correspondre à un niveau hiérarchique supérieur. Notons que l'irréductibilité définit en réalité que le niveau le plus bas est unique. C'est en définitive une information très limitée si on ne peut pas garantir que les niveaux supérieurs ne se vident pas progressivement avec $t \to \infty$ pour la loi renormalisée du processus.

Ces difficultés ne sont en réalité pas réservées aux domaines quasi-absorbants et devrait plutôt être interprétés en termes de mesures supports. Essayons d'éclaircir ce que l'on peut déduire de l'irréductibilité avec déjà ce premier lemme.

Lemme 0.1.3.4. *Supposons que le processus absorbé X est ψ-irréductible, avec ψ comme énoncé en Section 0.1.2.2. Si α est une mesure quasi-limite pour une condition initiale absolument continue par rapport à ψ, alors α est elle-même absolument continue par rapport à ψ.*

Preuve : On vérifie d'abord que la loi de X_t est absolument continue par rapport à ψ, si sa condition initial μ l'est. La première étape est de voir que X est aussi irréductible par rapport à ψP_t. On vérifie ensuite que l'absolue continuité de μ par rapport à ψ entraîne celle de μP_t par rapport à ψP_t puis par rapport à ψ par maximalité de ψ. Le résultat est a fortiori vrai pour la loi de X_t conditionnée à $\{t < \tau_\partial\}$. Si une partie de la masse de α était portée par un domaine S singulier pour ψ, elle ne pourrait pas être approchée par la masse associée à $\mu A_t(S)$.

Regardons plus en détail la justification énoncée plus haut : pour un choix quelconque de $A \in \mathcal{B}$ tel que $\psi P_t(A) > 0$, définissions $A_n := \{x \in \mathcal{X} \,\big|\, \delta_x P_t(A) \geq 1/n\}$ avec $n \geq 1$. Puisque $\psi P_t(A) > 0$, il doit exister un n tel que $\psi(A_n) > 0$. Maintenant, pour tout mesure de probabilité μ sur $\mathcal{X}$, il existe $s > 0$ par irréductibilité de ψ tel que $\mu P_s(A_n) > 0$. On en déduit par la propriété de Markov que $\mu P_{t+s}(A) \geq \mu P_s(A_n)/n > 0$. Comme A et μ sont arbitraires, X est aussi irréductible par rapport à ψP_t. Un raisonnement tout à fait analogue permet de montrer l'absolue continuité de μP_t par rapport à ψP_t. Si $\psi(S) = 0$, on déduit que $\mu A_t(S) \equiv 0$ donc que $\alpha(S) = 0$. Les arguments plus hauts sont donc bien suffisants pour conclure le Lemme 0.1.3.4. $\square$

Plus généralement, décomposons le semi-groupe partant de la condition initiale μ entre sa partie absolue continue par rapport à ψ et sa partie singulière ν_t :

$$\mu P_t(dx) := r_t^\mu(x)\,\psi(dx) + \nu_t^\mu(dx).$$

Par maximalité de ψ, on déduit du Lemme que $\nu_t^\mu \equiv 0$ implique que pour tout $s \geq t$ $\nu_s^\mu \equiv 0$. Certes $\nu_s^\mu(\mathcal{X})$ est décroissante, ce qui induit des contraintes sur ν_s^μ lorsque le processus est conservatif. Mais si $\mu P_t(\mathcal{X})$ tend déjà vers 0, cela ne nous apprend pas grand chose. Et il est facile de construire des exemples pour lesquels $\nu_t^\mu(\mathcal{X})$ est équivalent à $\mu P_t(\mathcal{X})$ en temps long ! L'exemple archétypal est celui dit de "source-puits" à deux états, où l'état quasi-absorbant (le puits) a un taux d'extinction plus grand que l'autre état (la source). Une analyse élémentaire permet alors de conclure qu'il y a en réalité deux QSD. La première ne charge que le puits tandis que l'autre est portée par les deux états, avec le taux d'extinction de la source. Donner un critère permettant de garantir qu'en temps long, $\nu_t^\mu(\mathcal{X})$ est négligeable devant $\mu P_t(\mathcal{X})$ est le coeur de la réflexion portée par le Chapitre 2.

Ces enjeux arrivent très naturellement dans le cas de processus avec des sauts sur un espace d'état non discret (aussi en temps discret).

0.1.4 Principaux résultats d'existence et de convergence vers les mesures quasi-stationnaires

La question qui va nous intéresser dans la suite est de décrire les propriétés de convergence que l'on pourra obtenir sur ces états de quasi-équilibre. Nous discutons dans ce paragraphe des résultats auparavant connus et de ceux voisins des miens qui viennent d'être soumis.

0.1.4.1 Le cadre avantageux d'une décomposition spectrale

Comme par exemple dans le Théorème 3.2 de [CCM16], on peut être en mesure de diagonaliser le générateur infinitésimal et d'identifier la capacité de survie comme vecteur propre associé à la valeur propre principale λ. La QSD est alors la valeur propre de l'adjoint $\mathcal{L}^*$, facilement déduite a priori. A partir du trou spectral défini comme l'écart à la seconde valeur propre $\lambda_2 - \lambda$, on peut alors donner une estimée de la vitesse de convergence. Cette diagonalisation peut s'effectuer dès lors que l'on peut associer $\mathcal{L}$ à un opérateur auto-adjoint compact. Comme dans [KKT16], on peut généraliser ces approches dans le cadre de processus biaisés par une fonctionnelle multiplicative du processus.

La principale limitation des théories de décomposition spectrale est donc qu'elles requièrent une propriété de réversibilité pour définir un noyau auto-adjoint. Certes, pour des processus de diffusion uni-dimensionnels, cette condition peut souvent être vérifiée ; et plus généralement elle peut être facilement déduite de conditions faciles à valider comme celle de bilan détaillé ("detailed balance"). Il arrive aussi que les processus qui satisfont cette condition soient désignés comme étant "à l'équilibre" par opposition aux processus non-réversibles déclarés "hors d'équilibre", cf notamment la description détaillée de [TH13]. Cela peut justifier l'intérêt qui a été porté à ces processus réversibles. C'est néanmoins une condition très restrictive, surtout lorsque des sauts interviennent et/ou que l'espace d'état est de dimension au moins 2, comme on peut le voir dans l'appendice A de [CCM17].

0.1.4.2 Estimées de compacité, point fixe et méthodes de renouvellement

Un bon nombre de résultats de la littérature déduisent l'existence d'une distribution quasi-stationnaire comme limite d'une suite définie itérativement en modifiant légèrement les propriétés du générateur infinitésimal. Il faut alors une estimée de compacité sur l'opérateur qui définit l'itération suivante à partir de la précédente pour justifier l'existence d'une sous-suite convergente.

C'est en définitive la base du Théorème de Perron-Frobenius et de sa généralisation via le Théorème de Krein-Rutman. Citons à ce titre le travail récent en [AMZ20], qui traite les chaînes de Markov auto-régressives, avec des techniques d'approximation pour se ramener à des cadres d'espaces compacts.

Dans [CM+11], une méthode du point fixe de Tychonov est employée sur l'opérateur : $\mu \mapsto \mu P / \langle \mu.P.\mathbf{1} \rangle$. Les conditions présentées pour obtenir un ensemble convexe stable me semblent annoncer les hypothèses de [CV17c] et [BCGM19].

Dans les techniques dites de renouvellement, comme présentées par exemple dans [FK+95], on remplace l'extinction par un saut défini selon une loi indépendante du point d'extinction. On est ainsi ramené à l'étude d'un régime sans extinction, pour lequel l'existence d'une mesure stationnaire peut potentiellement être établi sans trop de difficultés. On s'attend à ce que, partant d'une mesure à l'itération k suffisamment proche d'une quasi-stationnaire, l'effet de ce nouveau saut soit très similaire à l'effet de la renormalisation sur la quasi-stationnaire. L'itération $k + 1$ donnée par la mesure stationnaire associée a donc de bonnes chances d'être plus proche encore de cette mesure quasi-stationnaire. En tout cas, les seuls points fixes de cet algorithme sont exactement les QSD. On espère donc pouvoir justifier, ou au moins valider par simulation, la convergence vers une mesure au moins le long de sous-suites.

Dans [BP10] aussi, les auteurs établissent de cette manière, pour un processus en espace discret, l'existence et l'unicité de la QSD et une borne explicite sur l'écart en variation totale entre la QSD et la première itération de l'algorithme décrit au paragraphe précédent. Cela requiert néanmoins une borne inférieure uniforme en la condition initiale sur la probabilité d'atteindre un état de référence avant ∂ et ne peut donc s'appliquer en espace continu avec un bord absorbant.

0.1.4.3 R-théorie et l'étude du Q-processus

Comme mentionné en Section 0.1.3.3, il est tout à fait raisonnable de chercher à justifier la convergence vers la mesure quasi-stationnaire à partir de la convergence du Q-processus vers sa mesure stationnaire. Ce résultat a tout particulièrement été exploité dans [FRS19], [DM15], [KM03], et plus anciennement en temps et espaces discrets dans [FK+96] et [SV96]. Les principales difficultés de cette approche sont a priori les suivantes :

(i) Il faut être en mesure de justifier au préalable que le semi-groupe de départ est bien λ-récurrent selon les notations du Théorème 0.1.3, de manière à définir une capacité de survie (à constante près) et ainsi la transformée de Doob associée.

(ii) Il faut pouvoir maîtriser l'ergodicité du Q-processus malgré la difficulté pour obtenir des informations précises sur cette capacité de survie, alors qu'elle joue potentiellement un rôle dans la contraction de Lyapunov.

Il semble que cette approche soit en revanche particulièrement efficace pour des processus facilement décrits par leur fonction génératrice, notamment les processus de type Galton-Watson.

0.1.4.4 Critères récents de convergence exponentielle

Les conditions que je vais présenter ont beaucoup de similitudes avec les jeux de conditions suivants. Excepté pour le premier [CV16], dont je m'inspire très directement, les autres approches ont été rendues publiques à la même période que mon premier article. Je ne présenterai en détail, outre mes conditions données en Section 0.2.2, que celles du premier chapitre, et celles de [BCGM19] qui se rattachent le plus à la notion de quasi-ergodicité exponentielle définie en Section 0.1.3.2. Etant donné la discussion précédente sur les conditions de Lyapunov, qui constituent la principale nouveauté des articles [CV17c], [FRS19] et [BCGM19], je vais plus me focaliser ici sur les résultats de convergence que semblent permettre chaque jeu d'hypothèses. Je renvoie aussi à la présentation donnée par W. Oçafrain dans l'introduction de sa thèse.

Dans [CV16], N. Champagnat et D. Villemonais ont établi des conditions directement inspirées de la condition de Doeblin (surtout celle notée "mélange") qu'ils montrent équivalentes à la quasi-ergodicité exponentielle à taux uniforme et en variation totale. Les conditions sont exprimées en ces termes (avec "U" pour uniforme) :
Il existe une mesure de probabilité $\zeta \in \mathcal{M}_1(\mathcal{X})$, $t, c > 0$ tels que :

(U1) **Mélange** :$\forall\, x \in \mathcal{X}, \quad \mathbb{P}_x(X_t \in dy \,\big|\, t < \tau_\partial) \geq c\,\zeta,$

(U2) **Comparaison asymptotique de survie**

$$\sup_{x\in\mathcal{X},t\geq 0} \frac{\mathbb{P}_x(t<\tau_\partial)}{\mathbb{P}_\zeta(t<\tau_\partial)} < \infty.$$

On a alors le théorème suivant :

Theorem 0.1.5 ([CV16]). *Le conjonction* $(U1)$ *et* $(U2)$ *est équivalente à la quasi-ergodicité exponentielle au sens où, avec les notations de la section 0.1.3.2,* (CH) *a lieu en norme* $\|.\|_\infty$ *et* (CA) *est remplacée par :*

$$\|\mu A_t - \alpha\|_{TV} \leq Ce^{-\gamma t}.$$

Les processus qui satisfont ces conditions de Champagnat-Villemonais sont "moralement" des processus apériodiques (à noter qu'une T-périodicité serait contradictoire avec (U1)), ayant la propriété de revenir rapidement sur les compacts (si ce n'est pas déjà leur espace d'état). Cette dernière propriété est notamment vérifiée par les processus descendant de l'infini dans le cas uni-dimensionnel, pour lesquels nous renvoyons à l'étude donnée en [BCMM19]. Des conditions spécifiques apparentées à celles de fonction de Lyapunov ont été aussi introduites dans [CV17b] pour traiter ces propriétés de descentes de l'infini dans des cadres potentiellement multi-dimensionnels.

Les conditions que je propose sont motivées par la constatation que l'unicité de la QSD α n'est pas incompatible avec une convergence non uniforme comme en (CA). On retrouve cet effet pour des processus tout à fait élémentaires dès lors que l'espace n'est pas compact mais que le processus est confiné par son extinction. La difficulté est la suivante : pour tout $t, \epsilon > 0$, on peut trouver une condition initiale μ telle que

$$\|\mu A_t - \alpha\|_{TV} \geq \epsilon.$$

J'ai donc cherché à affaiblir les critères $(U1)$ et $(U2)$ en des versions locales (typiquement sur un compact) sous la condition d'obtenir un bon contrôle du temps de retour à ce domaine local. L'irréductibilité est aussi encore assurée dans la généralisation de $(U1)$. J'en suis néanmoins resté au cadre de la variation totale.

Il apparaît vu le preprint [BCGM19] que V. Bansaye, B. Cloez, P. Gabriel et A. Marguet ont réussi pendant ce temps à définir des conditions nécessaires et suffisantes pour la quasi-ergodicité exponentielle telle que généralement définie en Section 0.1.3.2. Pour en simplifier la présentation, quelques notations sont utiles. Pour deux fonctions $f, g : \mathcal{D} \mapsto \mathbb{R}$, la notation $f \lesssim g$ signifie qu'il existe une constante $C > 0$ telle que $f \leq Cg$ sur $\mathcal{D}$. $f \sim g$ si $f \lesssim g$ et $g \lesssim f$. Par extension, $PV \lesssim V$ sur $[0,T]\times\mathcal{X}$ signifie qu'il existe $C > 0$ tel pour tout $(t,x) \in [0,T]\times\mathcal{X}$, $\langle \delta_x P_t V\rangle \leq CV(x)$, et de même pour $\gtrsim$.

Voici donc les conditions qu'ils proposent (où L fait référence à Lyapunov) :

Jeu d'hypothèses (L) sur le couple (V, S) de fonctions : Il existe $\tau, T > 0$, $R_S > R_V > 0$, $R_G \geq 0$, $(c, d) \in (0, 1]^2$, $K \subset \mathcal{X}$ et $\zeta \in \mathcal{M}_1(\mathcal{X})$ de support inclus dans K tels que :

$(L0)$ $S \leq V$ sur $\mathcal{X}$ et $V \sim S$ sur K ; $PV \lesssim V$ et $PS \gtrsim S$ sur $[0, T] \times \mathcal{X}$,

$(L1)$ **Mélange** Pour tout $x \in K$ et fonction positive $f \in L_\infty^{V/S}$:

$$\langle \delta_x . P_\tau \,\big|\, f \times S \rangle \geq c\, \langle \delta_x . P_\tau \,\big|\, S \rangle \times \langle \zeta \,\big|\, f \rangle, \tag{0.1.31}$$

$(L2)$ **Propriété de contraction de Lyapunov** $P_\tau V \leq R_V V + R_G \mathbf{1}_K S$,

$(L3)$ **Estimée de survie** $P_\tau S \geq R_S S$,

$(L4)$ **Comparaison asymptotique de survie**

$$\sup_{x \in K, n \in \mathbb{Z}_+} \langle \zeta \,\big|\, \frac{P_{n\tau} S}{S} \rangle^{-1} \times \langle \delta_x \,\big|\, \frac{P_{n\tau} S}{S} \rangle < \infty. \tag{0.1.32}$$

Selon le Theorème 1.1 de [BCGM19], si $(P_t)_{t \geq 0}$ satisfait au jeu d'hypothèse (**L**), alors il est exponentiellement quasi-ergodique de fonction de Lyapunov V. Réciproquement, s'il est exponentiellement quasi-ergodique de fonction de Lyapunov V, alors le jeu d'hypothèse (**L**) est vérifié pour le couple $(V, h/\|h\|_V)$.

Pour les processus en temps continu, des conditions de Lyapunov définies à partir du générateur sont données pour remplacer $(L2-3)$, de manière à être plus commodes à vérifier. Pour comparer avec mon jeu d'hypothèses, il est intéressant de remarquer la connexion forte qui existe entre la conditions de Lyapunov et les estimées exponentielles de temps de retour, pour laquelle on renvoie au Lemme 3.6 de [CV17c] suivant :

Lemme 0.1.4.1 ([CV17c]**).** *Supposons qu'il existe $\rho, t > 0$ tels que :*

$$\forall\, x \in_c X,\ \mathbb{E}_x(\exp[\rho(\tau_K \wedge \tau_\partial)]) < \infty\ , \qquad \sup_{x \in K} \mathbb{E}_x(\mathbb{E}_{X_t} \exp[-\rho(\widetilde{\tau}_K \wedge \widetilde{\tau_\partial})]) < \infty,$$

alors la fonction $V(x) := \mathbb{E}_x(\exp[-\rho(\tau_K \wedge \tau_\partial]))$ satisfait pour un certain $c > 0$:

$$\forall\, x \in \mathcal{X}, \quad \mathbb{E}_x(V(X_t) \,\big|\, t < \tau_K \wedge \tau_\partial) \leq V(x) e^{-\rho t}$$
$$\forall\, s \in [0, t],\ \forall\, x \in K, \quad \mathbb{E}_x(V(X_s) \,\big|\, s < \tau_\partial) \leq c.$$

Dans les exemples qui m'ont occupé, j'ai privilégié ces estimées de retour définies par domaine, plutôt que la définition de fonction de Lyapunov qui s'appuie sur une description infinitésimale du processus.

D'autres résultats qui affaiblissent les conditions $(U1)$ et $(U2)$ ont été proposés dans une direction similaire aux conditions (L) à la fois dans [CV17c] et dans [FRS19]. Les preuves sont à chaque fois différentes et les conclusions aussi. On déduit du Théorème 2.1 ou 3.5 de [CV17c] une convergence analogue à (CA) avec une norme d'opérateur de $(\mathcal{M}(V), \|.\|_{\mathcal{M}(V)})$ dans $(\mathcal{M}(\mathcal{X}), \|.\|_{TV})$, et exactement la convergence (CH) dans L^V_∞ (ainsi que pour $L^{V^p}_\infty$ pour $p \in [1-\epsilon, 1]$). La convergence analogue à (CA) fait intervenir une autre fonction que h, mais la linéarité de l'opérateur $e^{\lambda t} P_t$ devrait permettre de se ramener au cas avec h comme je l'ai remarqué pour mes résultats.

On déduit du Théorème 1 de [FRS19] une convergence pour tout $\mu \in \mathcal{M}(V)$ de μA_t vers α en norme $\|.\|_{\mathcal{M}(V)}$. La dépendance en μ semble néanmoins ne pas pouvoir être explicitée facilement. Par ailleurs, en reprenant le formalisme de (LE) en Section 0.1.2.3, les conditions de Lyapunov de cet article sont demandées pour tout $\delta > 0$, avec $s = \mathbf{1}_{\{K\}}$, et ce pour une même fonction de Lyapunov V quitte à prendre K et b assez grands. De telles conditions sont exclues pour des processus à sauts purs dont les sauts arrivent à taux fini.

0.1.4.5 Extensions de mes résultats attendues comme possible

Il devrait être assez clair que tous les résultats dans la suite s'étendent au cas du temps discret. Il est très probable que l'on pourrait assez facilement généraliser nos techniques à des cas non-homogènes en temps comme ils sont considérés en [CV18b], [BCG19], [DV16]. Dans sa thèse, W. Oçafrain s'est aussi particulièrement occupé de ces questions pour gérer le cas de frontières absorbantes mobiles, [Oc18], [Oc19], On attend alors des estimées de contraction sur $\|\mu_1 A_t - \mu_2 A_t\|_V$, qui n'auraient pas de raison particulière d'être exponentielle. Le cadre général d'un espace pondéré par une fonction V n'est en pratique pas abordé dans mes preuves, même si je m'attends à ce que l'adaptation à ce cadre ne soit pas trop difficile. De même, nous supposons toujours que la variable de biais multiplicative peut être bornée uniformément, ce que l'on pourrait vouloir relâcher.

Pour le cadre totalement général d'un semi-groupe non-conservatif, je dois mentionner que mon approche se fonde sur des contrôles de temps d'arrêt pour lesquels la représentation par un processus biaisé est essentielle.

0.1.5 Lien avec les Grandes déviations de la mesure empirique

Une introduction beaucoup plus complète sur ce lien avec les Grandes Déviations est donnée au début de la thèse de G. Ferré [Fe20], dont fait partie [FRS19]. Je ne compte pas rentrer dans trop de détails et renvoie à celle-ci. J'indique néanmoins une direction de recherche qui s'est ouverte plus largement avec ma découverte des résultats de [KM03].

Au vu de ces derniers, les méthodes permettant d'établir la quasi-ergodicité exponentielle de processus biaisés jouent un grand rôle pour évaluer les grandes déviations sur $\langle S_t \,\big|\, G \rangle := (1/t) \int_0^t G(X_s)\, ds$ lorsque t tend vers l'infini, où G est une fonction test au minimum mesurable et a priori dans L^V_∞. Dans l'article [KM03], ces déviations sont évaluées pour un processus de Markov conservatif, mais il apparaît que la généralisation à des processus biaisés ne devrait pas trop poser de problème. On s'intéresse donc généralement à la distribution de $\langle S_t \,\big|\, G \rangle$ conditionnellement à $\{t < \tau_\partial\}$ (ou avec la renormalisation si Z_t intervient). D'après (0.1.8), la limite attendue est alors $\langle \beta \,\big|\, G \rangle$ et il n'est pas très difficile de montrer que la convergence a bien lieu en espérance, avec le conditionnement bien sûr. On pourrait aussi prouver sans trop de mal que la variance de cette distribution tend asymptotiquement vers 0, et même qu'elle doit être d'ordre $1/t$, indépendante de la condition initiale à cet ordre.

Le résultat que l'on cherche typiquement à obtenir est le suivant :
pour tout δ dans un ensemble à préciser et tel que $\delta > \langle \beta \,\big|\, G \rangle$:

$$\forall\, \mu \in \mathcal{M}(V), \quad (1/t) \times \log \mathbb{E}_\mu(Z_t \;;\; \langle S_t \,\big|\, G \rangle \geq \delta \,,\; t < \tau_\partial) \to \Lambda(\delta), \tag{0.1.33}$$

où $\Lambda(\delta) < \lambda$ est attendue comme fonction strictement décroissante de δ. On aurait un résultat analogue pour $\delta < \langle \beta \,\big|\, G \rangle$ en inversant le sens de l'inégalité. D'après (CH), on en déduirait un déclin à taux $-(\lambda - \Lambda(\delta))$ de la probabilité de déviation renormalisée.

L'idée est de transformer la loi du processus via une variable de biais de la forme $Z^\theta_G(t) := \exp[\theta\, t\, \langle S_t \,\big|\, G \rangle]$. Elle est déjà présente dans les premiers travaux de Donsker et Varadhan [DV75]-[DV83] puis de Gärtner et Ellis [Gä77], [E84]. Dans cette expression, $\theta \in \mathbb{R}$ est à ajuster en fonction de la déviation que l'on cherche à obtenir. Choisir $\theta > 0$ va en effet permettre de donner plus de poids (y compris relatif) aux grandes réalisations de $\langle S_t \,\big|\, G \rangle$. La remarque importante est le fait que $Z^\theta_G(t) = \exp[\theta \int_0^t G(X_s)\, ds]$ est un processus adapté et multiplicatif. On est donc ramené à l'étude du processus avec un biais supplémentaire, avec le semi-groupe suivant :

$$P^\theta_t(x, dy) := \mathbb{E}_x[Z^\theta_G(t) \times Z_t \;;\; X_t \in dy \,,\; t < \tau_\partial], \qquad x \in \mathcal{X}, t \geq 0. \tag{0.1.34}$$

Si ce semi-groupe est exponentiellement quasi-ergodique, de mesure quasi-ergodique β^θ, on s'attend alors à ce que la loi de $\langle S_t \,\big|\, G \rangle$ soit très proche en temps long de $\langle \beta^\theta \,\big|\, G \rangle$, avec renormalisation.

On peut alors espérer obtenir (0.1.33) pour les valeurs de δ de la forme $\langle \beta^\theta \,\big|\, G \rangle$, où le semi-groupe P^θ est exponentiellement quasi-ergodique. Remarquons que pour toute fonction G bornée non nulle, il est facile de se ramener au cas où $\|G\|_\infty = 1$ et $\langle \pi \,\big|\, G \rangle = 0$ avec une translation et une dilatation de G. On sait en effet en ajuster l'effet sur $P_t\theta$.

Le Théorème 4.16 de [KM03], donné dans un cadre conservatif, est à ce titre particulièrement important. Il généralise la notation de P^θ avec θ complexe, ce que l'on notera plutôt P^η. On en déduit alors le générateur associé $\mathcal{L}^\eta := \mathcal{L} + \eta\, B[G]$. Son spectre $\mathcal{S}^\eta$ est défini comme l'ensemble des valeurs $z \in \mathbb{C}$ telles que $[zI - \mathcal{L}^\eta]^{-1}$ n'existe pas comme opérateur linéaire et borné. L^V_∞ et $\mathcal{M}(V)$ sont généralisées aisément au cas complexe. Dans la suite, $Re(z)$ désigne la partie réelle de z.

Theorem 0.1.6. *[**Théorème d'extension de l'ergodicité multiplicative**[KM03]] Supposons que le processus de Markov X est exponentiellement ergodique de fonction de Lyapunov $V : \mathcal{X} \mapsto [1, \infty)$ tel que $\langle \pi \,\big|\, V^2 \rangle < \infty$. Supposons en outre que G admet une variance asymptotique positive :*

$$\forall\, x \in \mathcal{X}, \quad \sigma^2_G = \lim_{t\to\infty} Var_x(\langle S_t \,\big|\, G\rangle / \sqrt{t}) > 0.$$

On se ramène aussi au cas où G est bornée par 1 et satisfait $\langle \pi \,\big|\, G\rangle = 0$. Alors, il existe $\theta^\vee, \omega^\vee, \epsilon > 0$ tels que pour tout $\eta = \theta + i\omega \in \mathbb{C}$ avec $|\theta| \leq \theta^\vee$ et $|\omega| \leq \omega^\vee$, il existe une unique valeur maximale isolée $\Lambda(\eta)$ dans $\mathcal{S}^\eta$, i.e. telle que :

$$Re(\Lambda(\eta)) = \max\{Re(z) \,\big|\, z \in \mathcal{S}^\eta\}\ , \qquad \mathcal{S}^\eta \cap \{z \,\big|\, Re(z) \geq Re(\Lambda(\eta)) - \epsilon\} = \{\Lambda(\eta)\}.$$

De plus, pour tout η de cette forme, il existe un unique couple de mesure propre $\alpha^\eta \in \mathcal{M}(V)$ et de fonction propre $h^\eta \in L^V_\infty$ associé à $\Lambda(\eta)$ sous les conditions suivantes :

(i) $\langle \alpha^\eta \,\big|\, \mathbf{1}\rangle = \langle \alpha^\eta \,\big|\, h^\eta\rangle = 1$,

(ii) $\alpha^\eta.P^\eta_t = \exp[\Lambda(\eta)t]\alpha^\eta$, $P^\eta_t.h^\eta = \exp[\Lambda(\eta)t]h^\eta$,

(iii) il existe $C, \gamma > 0$ tels que :

$$\left\| \exp[-\Lambda(\eta)t]\, P^\eta_t - |h^\eta\rangle \otimes \langle \alpha^\eta| \right\|_V \leq C\, e^{-\gamma t}. \tag{0.1.35}$$

La condition $\sigma^2_G > 0$ est aussi ce qui permet d'assurer que : $\theta \mapsto \langle \beta^\theta \,\big|\, G\rangle$ est strictement croissante. Sous réserve de savoir le montrer, on est alors en mesure d'avoir beaucoup plus précis que (0.1.33) avec même l'équivalent asymptotique en adaptant les Théorèmes 6.3 et 6.5 de [KM03] aux cas du temps continu. Cette condition sur la variance ne semble pas poser tant de problème que cela dans [KM03]. Le cas $\sigma_G = 0$ est exclu en vertu des résultats de la Proposition 2.4 (de [KM03]) qui montrent que ce cas impose que $\langle S_t \,\big|\, G\rangle$ doit pouvoir s'exprimer en fonctions de la condition initiales X_0 et finale X_t.

Il ne m'apparaît encore pas clair de savoir si tout se généralise bien au cadre non-conservatif. Notamment, si on sait que P^{θ_*} est exponentiellement quasi-ergodique, on pourrait effectuer a priori la transformée de Doob via h^{θ_*} pour se ramener au cas conservatif. On voudrait ainsi appliquer le Théorème 0.1.6 pour obtenir une régularité en θ au voisinage de θ_* des propriétés de quasi-ergodicité exponentielle de P^θ. La fonction de Lyapunov qui apparaît alors est V/h^{θ_*}, pour laquelle $\langle \beta^{\theta_*} \,\big|\, (V/h^{\theta_*})^2 \rangle$ n'est pas clair. Même si on affaiblit cette fonction de Lyapunov comme conseillé dans [KM03], se pose alors le problème de savoir revenir au cadre de départ et du choix des pondérations d'espace.

A terme, mon objectif sera aussi de traiter les Grandes Déviations de la mesure empirique de sauts : $\langle J_t \,\big|\, F \rangle := (1/t) \sum_{s \leq t} F(X_{s-}, X_s)$ comme cela a pu être fait dans [KKT16] pour des modèles réversibles. Je me suis en effet intéressé à quantifier les perturbations du profil de mutations permettant l'adaptation d'une population. C'était avec cet objectif que j'ai effectué un stage au Japon avec K. Daehong l'été dernier. C'est aussi dans ce cadre de sauts que mes techniques sont les plus nouvelles.

0.1.6 Estimation numérique

Je me suis intéressé aussi à observer par les simulations comment le choix des paramètres affecte les profils des QSD, mesures quasi-ergodiques et capacités de survie associées. J'ai aussi cherché à estimer numériquement le taux d'extinction λ, et à préciser si une convergence exponentielle en variation totale avec un taux de convergence γ' est pertinente. J'ai aussi regardé A priori ce taux observé γ' est plus rapide que le γ obtenu dans les preuves.

Pour toutes ces questions, il est important de savoir évaluer numériquement la dynamique de μA_t en contrôlant notamment l'effet de non-linéarité. Dans les modèles que j'ai pu étudier numériquement, je me suis principalement basé sur l'estimation de la fonction de densité. Cela correspond à faire une approche par volumes finis en discrétisant l'espace d'état en petits volumes dont on précise les transitions à chaque pas de temps. Inclure un taux d'extinction dans ce cadre ne pose alors pas de grosses difficultés. Si la discrétisation temporelle est assez fine, il suffit de renormaliser la distribution par un facteur multiplicatif. Par simplicité, les schémas utilisés sont explicites.

Pour des modèles en grandes dimensions (par exemple celui du chapitre 5 !), cette approche par densité montre ses limites car elle est trop gourmande en temps de calcul. Il faut alors avoir recours à l'estimation via les réalisations du processus lui-même. Il faut alors savoir comment poursuivre l'estimation une fois que le processus (ou une particule qui le représente) est amené à s'éteindre.

Une méthode classique est de faire évoluer en parallèle toute une population dont chaque particule se déplace indépendamment selon le processus. En espérant que la loi empirique sur la population est proche de la QSD, relancer les particules selon cette loi empirique devrait permettre à l'algorithme de se stabiliser à proximité de la QSD. Cette approche est au final assez analogue à la méthode de renouvellement présentée en Section 0.1.4.2. L'analyse de cette approche et de la précision de l'approximation constitue une bonne partie de la thèse de D. Villemonais et de ses travaux postérieurs, et en particulier [Vi11], [Vi13], [Vi14]. W. Oçafrain s'est intéressé avec lui à ces questions en [OV17] pour empêcher que des sauts simultanés ne viennent bloquer l'algorithme.

Une autre approche consiste à ne simuler qu'un processus à la fois et à utiliser la mesure empirique en temps comme approximation de la QSD lors de l'extinction. Cette approche initiée en temps discret et espace fini dans [AFP88] a donné suite à plusieurs extensions : [BCP18] pour le cas compact, [MV20] pour des espaces non-compacts avec un taux d'extinction borné, [BCV20] pour des diffusions avec extinction au bord d'un domaine borné.

0.2 Les résultats de la thèse

Cette thèse se décompose en six chapitres. A part le cinquième chapitre que je viens juste d'achever et qui est une collaboration avec M. Mariani et mon directeur E. Pardoux, et le dernier chapitre qui vient d'être accepté, ils constituent chacun un article soumis ou en passe de l'être. Un travail en préparation, avec une version ArXiv [Ve20], porte sur l'extension des résultats de Grandes Déviations dans le cadre d'un conditionnement à la survie et en incluant la mesure empirique des sauts. Cependant, il n'est pas encore mature. Le lien avec les techniques de [KM03] et [KM05], doit être approfondi, ainsi qu'avec les résultats de [Fe20], sachant que je n'ai découvert que récemment l'impact conséquent de ces travaux.

De même, les résultats de simulations ont jusqu'à présent été plutôt exploratoires et c'est un travail en préparation que de les exploiter dans des articles. J'en donnerai néanmoins un aperçu en complément des troisième et quatrième chapitres, que l'on pourra trouver en fin de manuscrit, juste avant la bibliographie.

0.2.1 Un modèle pour quantifier le rôle des mutation pour l'adaptation à un changement environnemental

La motivation originelle de cette thèse est le traitement de ce modèle d'adaptation qui prend la suite du travail de thèse d'Elma Nassar. Celui-ci s'est lui-même effectué sous l'encadrement de mes directeurs Etienne Pardoux et Michael Kopp et ceux-ci m'ont proposé en début de thèse d'étendre l'approche précédente pour tenir compte du risque d'extinction. Il a ainsi été question dès le début d'inclure un couplage entre le processus évolutif et un processus adaptatif et de caractériser la stabilité du processus associé face à un changement environnemental. Cela explique notamment que plusieurs exemples des deux premiers chapitres sont en réalité des versions simplifiées du modèle décrit en plus grand détail au troisième chapitre.

La famille de modèles étudiées dans le cadre de cette question de l'adaptation à un changement environnemental se base sur un processus évolutif avec la forme suivante :

$$X_t = x - \int_0^t V_s\, ds + \int_0^t \Sigma_s \cdot dB_s + \int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} w\ \mathbf{1}_{\{u\le U_s(w)\}}\, M(ds, dw, du),$$

où B est un Mouvement Brownien et M un Processus Ponctuel de Poisson (un PPP) sur $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+$ d'intensité $ds\,\nu(dw)\,du$, avec ν une mesure positive sur $\mathbb{R}^d$ telle que $\int_{\mathbb{R}^d} 1 \wedge \|w\|\nu(dw) < \infty$. V_s et Σ_s dépendent a priori de X_s, U_s de X_{s-}, avec potentiellement une dépendance envers un autre processus couplé N_s qui décrit la taille de la population. Via le terme $U_s(w)$, on explicite le taux auquel une mutation d'effet additif w est censé pouvoir envahir la population au temps t. V_s et Σ_s reprennent à la fois la variation du référentiel d'adaptation via les changements environnementaux et l'accumulation de mutations dans une limite où elles seraient infinitésimales. V_s en décrit alors la partie à variations bornées tandis que la partie martingale locale a jusqu'à présent été supposée adaptée à un Mouvement Brownien de référence. Dans ces modèles dits d'optimum mobile, X permet de décrire non pas tant les caractères intrinsèques des individus que plutôt la relation qui existe entre ces caractères et l'adaptation à leur environnement. Cette adaptation peut se mesurer ou bien explicitement en terme du taux de croissance d'un modèle de taille de population qui serait couplé à X, ou bien implicitement en terme d'un taux d'extinction sur X et sur l'expression du taux de sauts. Ici, on suppose que cette relation d'adaptation est homogène en temps tout en incluant un effet environnemental sur X via le terme $(V_s)_{s\ge 0}$. Cela signifie que le changement d'environnement se traduit uniquement par une translation dans l'espace des caractères de la relation qui relie ces caractères à l'adaptation.

0.2.1.1 Un processus purement diffusif

D'une manière générale, le processus peut être décrit par l'équation suivante :

$$(S) \begin{cases} dN_t = (r(X_t) - c\ N_t)\ N_t\ dt + \sigma\ \sqrt{N_t}\ dB_t^N \\ dX_t = b(X_t, N_t)\ dt + \theta(X_t, N_t)\ dB_t^X \end{cases}$$

avec condition initiale (n, x), B^N et B^X deux Mouvements Browniens standards indépendants, $c, \sigma > 0$, et r, b, θ étant des fonctions localement Hölder-continues. On requiert aussi que θ soit localement elliptique, au sens suivant : pour tout compact set K de $\mathbb{R}^{d+1}$, il existe $\bar{\theta} > 0$ tel que pour tout $(n, x) \in K$ et $\xi \in \mathbb{R}^d$: $\sum_{i,j} \theta_{i,j}(n, x)\xi_i\xi_j \ge \bar{\theta}\, |\xi|^2$.

On peut aussi se permettre d'introduire des catastrophes, qui arrivent de manière imprévisible au taux $\rho_c(x, n)$, et conduisent à l'extinction complète de la population. Vu les résultats du Chapitre 2, on serait capable d'obtenir la quasi-ergodicité exponentielle en introduisant des événements similaires conduisant simplement à une réduction de la taille de population. Nous laissons au lecteur intéressé le soin de voir comment tous ces éléments de preuve se combinent.

Theorem 0.2.1. *Sous les conditions d'ellipticité et de régularité Hölder mentionnées plus haut, et sous l'hypothèse que* $\limsup_{\|x\|\to\infty} r(x) = -\infty$*, le processus* (N, X) *avec extinction au temps* τ_∂ *est exponentiellement quasi-ergodique. La capacité de survie associée est minorée par une constante positive sur tout compact de* $\mathbb{R}_+ \times \mathbb{R}^d$*.*

Je ne précise pas de fonction de pondération V ce qui signifie qu'elle est constante égale à 1. Vu le Lemme 0.1.3.4 et les estimées déduites de l'inégalité de Harnack (cf le paragraphe suivant), il ne devrait pas être difficile de prouver que la QSD a une densité pour la mesure de Lebesgue. Il est probable que sous des conditions raisonnables, on puisse montrer que cette densité et la capacité de survie sont régulières. Pour l'instant, je ne me suis cependant que peu impliqué sur ces questions et ne sais prouver a priori que des minorations. Ces remarques sont aussi valables pour les résultats suivants.

Pour ce théorème, outre la condition de confinement $\limsup_{\|x\|\to\infty} r(x) = -\infty$, il suffit en réalité que le processus (N, X) satisfasse l'hypothèse (H) suivante (que l'on pourra résumer comme l'inégalité de Harnack dans les discussions). Cette constatation avait initialement été faite par Champagnat et Villemonais en Section 4 de [CV17b] et nous la reprenons.

Nous disons qu'un processus (Y_t) à valeurs dans $\mathcal{Y} \subset \mathbb{R}^d$ et de générateur infinitésimal $\mathcal{L}$ (qui inclut potentiellement un taux d'extinction ρ_e) satisfait l'hypothèse (H) sous la condition suivante :

Considérons deux sous-ensembles $\mathcal{D}, \mathcal{D}' \subset \mathcal{Y}$ *connexes par arcs, ouverts et relativement compacts, avec la relation d'inclusion* $\overline{\mathcal{D}} \subset \mathcal{D}'$*, de bords* C^∞*, et tels que pour tout point* $y \in \partial\mathcal{D}'$*, il existe une boule fermée* $C \in \mathbb{R}^d$ *(d'intérieur non vide) telle que* $C \cap \overline{\mathcal{D}'} = \{y\}$*. Alors, pour tout* $0 < t_1 < t_2$ *et fonctions de contrainte* C^2 *et positive :* $u_{\partial\mathcal{D}'} : (\{0\} \times \mathcal{D}') \cup ([0, t_2] \times \partial\mathcal{D}') \to [0, \infty)$*, il existe une unique solution* $u(t, x) \in C^{1,2}((0, t_2) \times \mathcal{D}') \cap C^0([0, t_2] \times \overline{\mathcal{D}'})$ *au problème :*

$$\begin{aligned} &\partial_t u(t, y) = \mathcal{L}u(t, y) && \text{sur } [0, t_2] \times \mathcal{D}' \,; \\ &u(t, y) = u_{\partial\mathcal{D}'}(y) && \text{sur } (\{0\} \times \mathcal{D}') \cup ([0, t_2] \times \partial\mathcal{D}'). \end{aligned}$$

Cette solution est positive sur $int(\mathcal{D}')$ *et satisfait, pour un certain* $C = C(\mathcal{L}, t_1, t_2, \mathcal{D}, \mathcal{D}') > 0$ *indépendant de* $u_{\partial\mathcal{D}'}$ *:*

$$\inf_{y\in\mathcal{D}} u(t_2, y) \geq C \, \sup_{y\in\mathcal{D}} u(t_1, y).$$

Remarques : (i) La preuve serait encore valable si t_2 devait dépendre de $\mathcal{D}$ et $\mathcal{D}'$.

(ii) Si le taux d'extinction est borné sur $\mathcal{D}'$ mais pas suffisamment régulier pour garantir l'existence ou la régularité d'une telle solution u, il faudrait prouver l'Hypothèse (H) pour le générateur $\mathcal{L} + \rho_e^k$ avec une famille $(\rho_e^k)_{k\geq 1}$ d'approximations régulières du taux d'extinction, en justifiant un contrôle uniforme de la constante C sur cette famille.

(iii) Avec les notations qui définissent le processus (X_t, N_t), l'Hypothèse (H) doit être justifiée pour le générateur $\mathcal{L}$ défini de la manière suivante pour toute fonction $f \in \mathbb{C}_b^2(\mathbb{R}_+ \times \mathbb{R}^d)$:

$$\begin{aligned}\mathcal{L}f(n,x) := [r(x) - c\,n]\,\partial_n f(x,n) + b(x,n)\,\partial_x f(x,n) \\ + n\,\sigma^2/2 \times \Delta_n f(x,n) + \theta^2(x,n)/2 \times \Delta_x f(x,n).\end{aligned}$$

L'Hypothèse (H) est déduite de résultats classiques de la littérature sous les conditions citées plus haut que r, b et θ soient Hölder-continues et l'opérateur $\theta \times \Delta_x$ localement elliptique.

0.2.1.2 Un processus évolutif à sauts purs confiné par le taux d'extinction

Dans le chapitre 2, nous considérons cette fois la dynamique

$$X_t = x - v\,t\,e_1 + \int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} w\,\mathbf{1}_{\left\{u \le g(X_{s-},w)\right\}}\,M(ds,dw,du). \qquad (0.2.36)$$

Ici, M est un PPP sur $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$, d'intensité $\pi(ds,dw,du) = ds\;dw\;du$. $g(x,w)$ décrit le taux de fixation dans une population dont les individus sont de type x d'une sous-population mutante de type $x + w$. Le changement environnemental est supposé se traduire ici par une translation à vitesse constante $v > 0$ du profil de la valeur sélective des individus (la "fitness", qu'elle soit mesurée au travers du taux d'extinction de la population ou du taux de croissance des individus). Cette translation se fait selon le sens d'accroissement de la première coordonnée suivant e_1. De plus, l'extinction peut advenir au temps τ_∂ de manière imprévisible au taux $\rho_e(x), x \in \mathbb{R}^d$, ce que l'on peut définir de la manière suivante :

$$\tau_\partial := \inf\{t \ge 0\ ;\ \textstyle\int_0^t \int_0^{\rho_c(X_s,N_s)} M^E(ds,du) \ge 1\},$$

où M^E est un PPP, d'intensité $ds\;du$, indépendant de M.

— **Jeu d'hypothèse** (PD) **:** (pour Piecewise-Deterministic)

(PD^1) ρ_e est localement bornée et $\lim_{\|x\|\to\infty} \rho_e(x) = +\infty$.
Aussi, l'explosion induit l'extinction au sens où : $\tau_\partial \le \sup_{\{\ell \ge 1\}} T_{\mathcal{D}_\ell}$.

h est mesurable et pour tout compact $K \subset \mathbb{R}^d$, il existe :

(PD^2) une borne supérieure $\rho_J^\vee$ sur le taux de saut :
$\forall\, x \in K, \quad 0 < \rho_J(x) := \int_{\mathbb{R}^d} g(x,w)\,dw \le \rho_J^\vee$.

(PD^3) une borne inférieure $g_\wedge > 0$ sur les sauts susceptibles de compenser la dérive, avec pour certains $0 < \delta S < S$:
$\forall\, x \in K,\ \forall\, w \in B(S\,e_1, \delta S), \quad g(x,w) \ge g_\wedge$.

(PD^4) *une estimée de tension sur le profil des sauts :* Pour tout $\epsilon > 0$, il existe $w_\vee$ tel que :
$\forall\, x \in K, \quad \int_{\mathbb{R}^d} g(x,w)/\rho_J(x)\, \mathbf{1}_{\{\|w\| \geq w_\vee\}}\, dw \leq \epsilon.$

(PD^5) une borne supérieure $g_\vee$ sur la densité de chaque saut :
$\forall\, x \in K,\ \forall\, w \in \mathbb{R}^d, \quad g(x,w) \leq g_\vee\, \rho_J(x).$

Theorem 0.2.2. *Considérons le processus X donné par l'équation (0.2.36) sous le jeu d'hypothèses* (**PD**)*. Alors, le processus X avec extinction au temps τ_∂ est exponentiellement quasi-ergodique et sa capacité de survie est minorée par une constante positive sur tout compact de $\mathbb{R}^d$.*

0.2.1.3 Un couplage entre les dynamiques démographiques et évolutives à sauts

Dans le cadre des résultats du chapitre 3, le processus qui décrit l'évolution conjointe de la taille de population et de l'écart phénotypique est donné comme suit :

$$\begin{cases} X_t = x - v\, t\, \mathbf{e_1} + \displaystyle\int_{[0,t]\times\mathbb{R}^d\times(\mathbb{R}_+)^2} w\; \mathbf{1}_{\left\{u_f \leq f_0(N_s)\right\}} \mathbf{1}_{\{u_g \leq g(X_{s-},w)\}} M(ds, dw, du_f, du_g) \\ N_t = n + \displaystyle\int_0^t \Big(r(X_s)\; N_s - \gamma_0 \times (N_s)^2\Big)\, ds + \sigma \int_0^t \sqrt{N_s}\, dB_s, \end{cases}$$

où N_t désigne la taille de population et X_t l'écart phénotypique. B désigne ici un Mouvement Brownien standard et M un Processus Ponctuel de Poisson sur $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+$ d'intensité $ds\, \nu(dw)\, du$, avec ν une mesure positive sur $\mathbb{R}^d$ telle que $\int_{\mathbb{R}^d} 1 \wedge \|w\| \nu(dw) < \infty$

2 jeux d'hypothèses ont été considérés pour ce système, selon que l'on inclut ou non la possibilité d'invasion par des mutations délétères. Concrètement, on s'attend à ce que celles-ci adviennent, mais soient très fortement contre-sélectionnées pour de grandes populations en état d'équilibre démographique. Par notre étude, nous montrons à la fois que de tels événements de fixation ne posent pas de problème pour cette convergence (en fait ils y contribuent au contraire), et qu'ils ne sont pas nécessaires non plus à cette convergence.

On regroupe ici les hypothèses que nous exploitons.

[$H1$] $f \in \mathcal{C}^0\Big(\mathbb{R}_+^*,\, \mathbb{R}_+\Big)$.

[$H2$] r est localement Lipschitz-continue sur $\mathbb{R}^d$.

[$H3$] $g \in \mathcal{C}^0\Big(\mathbb{R}^d \times \mathbb{R}^d,\, \mathbb{R}_+\Big)$ et est bornée sur tout $K \times \mathbb{R}^d$, où K est un compact de $\mathbb{R}^d$.
(1 est la borne attendue du fait de l'interprétation biologique de g, mais on peut étendre le modèle à des cas où g est seulement proportionnelle à la probabilité de fixation, cf. Corollaire 3.2.2.2)

[$H4$] $\quad \nu(\mathbb{R}^d) < \infty.$
Sous l'hypothèse [$H4$], sur tout intervalle de temps fini, seulement un nombre fini de mutations peuvent advenir.

[$H5$] $\quad \forall\, y > 0, \quad f(y) > 0$
Il est assez naturel d'imaginer que $f(0) = 0$ et $f(y) \to \infty$ lorsque $y \to \infty$, mais nous n'allons pas en avoir besoin ici.

[$H6$] Il existe $S, \nu_\wedge > 0$ et $0 < \delta S < S$ tel que $\nu(dw) \geq \nu_\wedge \, \mathbf{1}_{B(S+\delta S) \setminus B(S-\delta S)}\, dw$, où $B(R)$, pour $R > 0$, est la boule de rayon R centrée à l'origine.

[$H7$] $r(x) \to -\infty$, as $\|x\| \to \infty$.

[$H8$] g est positive.
En particulier, même les mutations délétères sont en mesure de se fixer dans la population.

[$H9$] Cette hypothèse énonce que les mutations délétères ne peuvent envahir la population, tandis que toutes les mutations avantageuses ont une probabilité positive de se fixer :

$$\begin{aligned} \forall\, x, w \in \mathbb{R}^d, \quad \|x + w\| < \|x\| \qquad &\Rightarrow \quad g(x,w) > 0 \\ \|x + w\| \geq \|x\| \qquad &\Rightarrow \quad g(x,w) = 0. \end{aligned}$$

[$H10$] $\nu(dw) << dw$ et la densité $g(x,w)\, \nu(w)$ (pour un saut de x vers $x + w$), qui caractérise la loi de l'effet additif des sauts par rapport à la mesure de Lebesgue, satisfait :

$$\forall\, x_\vee > 0, \quad \sup \left\{ \frac{g(x,w)\, \nu(w)}{\int_{\mathbb{R}^d} g(x,w')\, \nu(w')\, dw'} \; ; \; \|x\| \leq x_\vee,\, w \in \mathbb{R}^d \right\} < \infty.$$

Cette dernière hypothèse est requise uniquement pour le cas d'un processus évolutif multidimensionnel, i.e. $d \geq 2$.

Proposition 0.2.1.1. *Supposons que* [$H1$-4] *est vérifiée. Alors, pour toute condition initiale* $(x,y) \in \mathbb{R}^d \times \mathbb{R}_+^*$, *il existe une unique solution forte* $(X_t, Y_t)_{t \geq 0}$ *solution de* (S) *pour tout* $t < \tau_\partial$, *et telle que* $X_t = Y_t = 0$ *pour* $t \geq \tau_\partial$, *où* $\tau_\partial := \sup_{\{n \geq 1\}} T_Y^n \wedge \sup_{\{n \geq 1\}} T_X^n$,
$T_Y^n := \inf\{t \geq 0,\ Y_t \leq 1/n\}$, $\qquad T_X^n := \inf\{t \geq 0,\ \|X_t\| \geq n\}$.

Theorem 0.2.3. *Supposons le jeu d'hypothèse* [$H1$-7]. *Supposons en outre que* [$H8$] *ou* [$H9$] *est vérifiée. Si* $d \geq 2$, [$H10$] *est aussi demandée. Alors, le processus* (X, Y) *avec extinction au temps* τ_∂ *est exponentiellement quasi-ergodique.*

La difficulté de ce modèle tient à ce qu'il combine une composante de type déterministe par morceaux à une composante diffusive, ce qui n'avait pas été fait auparavant. De plus, le domaine est non borné avec extinction au bord et descente de l'infini selon une direction seulement.

0.2.1.4 Extensions attendues

J'ai eu l'occasion de lancer par ailleurs plusieurs projets de réflexion sur comment étendre ces résultats, que je suis encore loin d'avoir finalisés. J'aurai aimé par exemple pouvoir déduire des résultats analogues de stationnarité conditionnelles lorsque l'environnement est aléatoire avec une perte de mémoire de ses variations, tout en faisant partie du conditionnement. La thèse a pris une autre direction vu l'intérêt pratique et la généralité des critères introduits pour la preuve de la quasi-stationnarité. De même, j'aborderai peu dans cette thèse les simulations numériques, mes réflexions n'étant pas assez mûries et détaillées numériquement pour faire l'objet d'un article. J'en donnerai néanmoins un aperçu en complément des troisième et quatrième chapitres, que l'on pourra trouver en fin de manuscrit, juste avant la bibliographie.

Ces simulations m'ont déjà permis d'observer concrètement ces profils de QSD et de visualiser une distinction spécifique de régime entre le cas d'un changement environnemental soutenable et celui d'une adaptation "au pas de course". J'espère aussi sous peu compléter ces observations en analysant la manière dont la distribution empirique de mutations converge vers son asymptotique quasi-ergodique. En particulier, je voudrais voir si les convergences de type Grande Déviation sont valides et estimer les fonctions de taux associées. J'ai des premiers résultats extrêmement encourageants autant du point de vue théorique que pour ce qui est des simulations.

0.2.2 Critères pour établir la convergence exponentielle vers une unique QSD

0.2.2.1 Rappel des objectifs

Chacun des deux premiers chapitres reprend très largement un article en cours de révision. Ces travaux reprennent les techniques récemment développées par N. Champagnat et D. Villemonais, et notamment celui présentant les critères $(U1)$ et $(U2)$ donnés en Section 0.1.4.4. Ils ont d'ailleurs beaucoup publié sur le sujet depuis cet article fondateur [CV16] de 2016. De même, on déduit de mes hypothèses une convergence à vitesse exponentielle en variation totale des lois marginales sous conditionnement vers une unique distribution quasi-stationnaire. Outre le fait que j'explicite la preuve présentée dans [CV16] directement comme une minoration de la mesure conditionnelle, l'originalité de ce premier chapitre concerne le traitement de domaines transitoires et de la dépendance en la condition initiale. Pour les modèles d'adaptation qui m'ont intéressé, il est en effet fréquent qu'une partie de l'espace d'état ne soit pas viable car très voisine du bord d'extinction. Il se peut aussi qu'une partie de l'espace d'état reste à jamais inaccessible une fois quittée. Le fait que le processus doit bien s'échapper de ces zones peu viables avec un conditionnement à la survie n'entraîne cependant pas que les lois conditionnelles doivent se localiser en temps borné là où se concentre la QSD. On s'attend en effet à la difficulté potentielle suivante : pour tout $t, \epsilon > 0$, il existe une condition initiale μ telle que, avec A_t défini en (0.1.19) pour $V \equiv 1$:

$$\|\mu A_t - \alpha\|_{TV} \geq \epsilon.$$

Il faut aussi pouvoir gérer le fait que le domaine transitoire est potentiellement décomposé en plusieurs composantes pour lesquelles les justifications de ce caractère transitoire sont différentes. J'ai réussi néanmoins à m'occuper de ces problèmes avec un contrôle de moment exponentiel, en traitant séparément chaque partie de ce domaine transitoire et les circulations entre elles. Faire cela aurait sans doute été beaucoup plus délicat si je m'étais basé sur la définition de fonctions de Lyapunov adaptées, même si ces techniques se sont beaucoup popularisées lors de l'écriture de cette thèse. Par ailleurs, ce traitement permet de simplifier largement les problématiques d'effets de bords, comme on pourra le voir par exemple au chapitre 5.

On pourra aussi noter que j'adapte la preuve du Chapitre 1 à un cadre non irréductible pour démontrer les Propositions 2.2.2 et 2.5.1 du Chapitre 4. Je ne suis pas certain que cela aurait été aussi direct avec les autres résultats récents ([CV17c] ou [BCGM19]). Car j'exploite alors le contrôle explicite de la minoration de la loi conditionnée (sans me ramener au cas linéaire).

Nous détaillons ici les conclusions des deux premiers chapitres et renvoyons à ceux-ci la présentation des exemples illustratifs et des preuves, autres que ceux donnés en Section 0.2.1 précédente.

0.2.2.2 Les hypothèses

$(A0)$: **"Recouvrement de $\mathcal{X}$"**
Il existe une suite $(\mathcal{D}_\ell)_{\ell \geq 1}$ de sous-ensembles fermés de $\mathcal{X}$ (dont l'intérieur est noté $int(\mathcal{D}_\ell)$) tels que :

$$\forall\, \ell \geq 1,\ \mathcal{D}_\ell \subset int(\mathcal{D}_{\ell+1}) \quad \text{et} \quad \bigcup\nolimits_{\ell \geq 1} \mathcal{D}_\ell = \mathcal{X}. \tag{A0}$$

Soit $\mathbf{D}$ l'ensemble des sous-ensembles compacts de $\mathcal{X}$ qui sont inclus dans $\mathcal{D}_\ell$ à partir d'un certain rang.

Pour les définitions suivantes, on rappelle les notations suivantes pour le temps de sortie et le temps de première entrée d'un ensemble mesurable $\mathcal{D}$:

$$T_\mathcal{D} := \inf\left\{t \geq 0\ ;\ X_t \notin \mathcal{D}\right\}, \qquad \tau_\mathcal{D} := \inf\left\{t \geq 0\ ;\ X_t \in \mathcal{D}\right\}.$$

$(A1)$: **"Mélange"**
Il existe une mesure de probabilité $\zeta \in \mathcal{M}_1(\mathcal{X})$ telle que, pour tout $\ell \geq 1$, il existe $L > \ell$ et $c, t > 0$ tels que :

$$\forall\, x \in \mathcal{D}_\ell, \qquad \mathbb{P}_x\left[X_t \in dx\ ;\ t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\, \zeta(dx).$$

Remarque : *On pourra noter que* $(A1)$ *implique que* ζ *est une mesure d'irréductibilité, cf Section 0.1.2.2.*

On requiert pour la condition suivante une estimation du taux de survie défini comme suit :

$$\rho_S := \sup\left\{\rho \geq 0 \,\middle|\, \sup_{L\geq 1} \liminf_{t>0} e^{\rho t}\, \mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) = 0\right\} \vee 0. \tag{0.2.37}$$

$(A2)$: **"Echappée du domaine transitoire "**
Pour un certain $\rho > \rho_S$ *et* $E \in \mathbf{D}$ *:*

$$e_{\mathcal{T}} \ := \sup_{\{x\in\mathcal{X}\}} \mathbb{E}_x\left(\exp\left[\rho\left(\tau_\partial \wedge \tau_E\right)\right]\right) < \infty.$$

$(A3)$: **"Comparaison asymptotique de survie"**
Pour le $E \in \mathbf{D}$ défini en $(A2)$ et une mesure $\zeta \in \mathcal{M}_1(\mathcal{X})$ qui satisfait $(A1)$:

$$\limsup_{t\to\infty} \sup_{x\in E} \frac{\mathbb{P}_x(t < \tau_\partial)}{\mathbb{P}_\zeta(t < \tau_\partial)} < \infty.$$

Cette dernière hypothèse est potentiellement difficile à obtenir directement. Il peut être plus commode de d'exploiter l'hypothèse suivante, qui est l'objet essentiel de l'analyse proposée au Chapitre 2 :

$(A3_F)$: **"Absorption avec échecs"**
Etant donné $\zeta \in \mathcal{M}_1(\mathcal{X})$ *qui satisfait* $(A1)$, $\rho > 0$ *et* $E \in \mathbf{D}$ *défini en* $(A2)$, *pour tout* $\epsilon \in (0,\, 1)$, *on peut trouver* $t, c > 0$ *tels que pour tout* $x \in E$ *il existe un temps d'arrêt* U_A *tel que :*

$$\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\} \qquad \text{et} \qquad \mathbb{P}_x(U_A = \infty,\, t < \tau_\partial) \leq \epsilon\, \exp(-\rho\, t),$$

tandis que pour un certain temps d'arrêt V, *qui dépend a priori de* x *:*

$$\mathbb{P}_x\Big(X(U_A) \in dx' \ ; \ U_A < \tau_\partial\Big) \leq c\, \mathbb{P}_\zeta\Big(X(V) \in dx' \ ; \ V < \tau_\partial\Big).$$

En outre, on requiert qu'il existe un temps d'arrêt U_A^∞ *qui étend* U_A *au sens suivant :*
- $U_A^\infty := U_A$ *sur l'événement* $\{\tau_\partial \wedge U_A < \tau_E^t\}$, *où* $\tau_E^t := \inf\{s \geq t : X_s \in E\}$.
- *Sur l'événement* $\{\tau_E^t \leq \tau_\partial \wedge U_A\}$ *(qui par construction est identique à* $\{\tau_E^t < \tau_\partial\} \cap \{U_A = \infty\}$*) et conditionnellement à* $\mathcal{F}_{\tau_E^t}$, *la loi de* $U_A^\infty - \tau_E^t$ *coïncide avec celle de* $\widetilde{U}_A$ *pour une réalisation* $\widetilde{X}$ *du processus de Markov* $(X_t, t \geq 0)$ *avec pour condition initiale* $\widetilde{X}_0 := X(\tau_E^t)$ *et indépendant de* X *conditionnellement à* $X(\tau_E^t)$. *C'est-à-dire qu'on demande une propriété de Markov forte pour la loi de* $\widetilde{U}_A^\infty$ *au temps d'arrêt* τ_E^t.

Comme on le verra au Lemme 3.0.1 du Chapitre 1 et à la Remarque 2.4.1 du Chapitre 2, la valeur de ρ_S et les couples (ρ, E) pour lesquels l'hypothèse $(A3_F)$ (resp.

$(A3)$) est vérifiée ne dépendent pas du choix spécifique de ζ du moment que cette mesure satisfait à l'hypothèse $(A1)$. Par ailleurs, $(A1)$ implique une borne supérieure sur ρ_S. C'est très utile avec $(A3)$ si on peut prouver $(A2)$ pour tout $\rho > 0$, ou si on peut prouver conjointement $(A2)$ et $(A3_F)$ pour tout $\rho > 0$.

Nous dirons que le jeu d'hypothèse $(\mathbf{A})$ est vérifié sous les conditions suivantes : "L'hypothèse $(A0)$ est vérifiée, de même que $(A1)$ pour une certaine mesure $\zeta \in \mathcal{M}_1(\mathcal{X})$. En outre, il existe $\rho > \rho_S$ et $E \in \mathbf{D}$ tels que l'hypothèse $(A2)$ et soit $(A3)$ soit $(A3_F)$ sont vérifiées."

0.2.2.3 Le théorème central de cette thèse

Des Théorèmes 2.1-4 du chapitre 2, nous tirons le théorème central de cette thèse, à savoir :

Theorem 0.2.4. *Si le jeu d'hypothèse* $(\mathbf{A})$ *est vérifié, alors le processus* X *avec extinction au temps* τ_∂ *est exponentiellement quasi-ergodique au sens défini en Section 0.1.3.2.*

Il en est de même pour tout $L \geq 1$ *assez grand pour le processus* X *restreint à* $\mathcal{D}_L$ *via l'extinction au temps* $\tau_\partial \wedge T_{\mathcal{D}_L}$ *et les constantes impliquées dans ces convergences peuvent être choisies indépendantes de* L.

La preuve donnée au Chapitre 2 complète celle plus détaillée du Chapitre 1, pour laquelle l'hypothèse $(A3_F)$ n'est pas considérée. Le résultat de convergence est aussi amélioré avec cette fois-ci les convergences (CH) et (CA) qui explicitent la dépendance en la condition initiale. Pour autant, les avancées présentées au Chapitre 2 concernent essentiellement cette hypothèse $(A3_F)$ et son usage.

Plusieurs applications à des modèles simples sont données dans ces 2 premiers chapitres, notamment les modèles d'adaptation déjà mentionnés en Section 0.2.1. Au Chapitre 1, on traite un modèle de naissance et mort confiné par l'extinction en montrant que même pour des modèles aussi simples, la convergence n'est pas attendue comme uniforme en la condition initiale. Au Chapitre 2, les processus à sauts purs sont aussi à l'honneur, d'autant que leur traitement est souvent plus compliqué en réalité. Il faut en effet pouvoir justifier, comme énoncé en Section 0.1.3.5, que la mesure singulière à ψ, à savoir $\nu_t^\mu(dx)$, tend exponentiellement vers 0. Il est alors parfois nécessaire de procéder par étapes selon les directions imposées sur les sauts.

0.2.3 Modèle de sélection à deux niveaux

Au chapitre 4, nous considérons une population structurée en un grand nombre de groupes de grandes tailles. L'enjeu de ce chapitre est de préciser le conflit qui peut s'opérer entre des effets de sélection naturelle au sein des groupes et entre les groupes. Une question naturelle est notamment d'étudier la dynamique du niveau de coopération entre individus : sous quelles conditions les individus prêts à sacrifier une part de leur efficacité reproductive pour le bénéfice du groupe dans leur entier pourront-ils prospérer ?

Dans cette optique, nous mettons en conflit deux types d'individus : les types D (pour "Defectors") sont toujours supposés meilleurs à l'intérieur des groupes tandis que les types C (pour "Cooperators") apportent un avantage aux groupes qui en possèdent (sans que la relation soit nécessairement linéaire).

Il y a une vaste littérature sur la notion de sélection de groupe à laquelle ce travail se rattache. Il y est souvent établi que ces effets de groupes peuvent être assimilés à des bénéfices individuels en se plaçant dans le cadre de la parenté génétique : le bénéfice apporté aux autres est ainsi pondéré par la probabilité que cet autre possède ce même gène altruiste, elle-même donnée par le degré de parenté ("kinship") entre le donneur et le receveur. Il y a néanmoins deux hypothèses très forte dans les modèles qui suivent : les migrations d'individus entre les groupes ne sont pas permises ; lorsqu'un groupe se divise, les deux groupes qui en résultent ont exactement la même composition que le groupe qui s'est divisé.

La limite d'échelle que nous considérons est spécifiée comme le processus déterministe $(\mu_t)_{t\geq 0}$ à valeur dans les mesures de probabilité sur $[0,1]$ et solution de l'équation suivante :

$$\begin{aligned} &\partial_t \left\langle \mu_t \,\middle|\, f \right\rangle = \left\langle \mu_t \,\middle|\, \mathcal{L}_{WF} f \right\rangle - \left\langle \mu_t \,\middle|\, (\rho - \left\langle \mu_t \,\middle|\, \rho \right\rangle) f \right\rangle , \qquad \mu_0 = \mu_0, \\ &\text{où } \mathcal{L}_{WF} f(x) := x\,(1-x) \left[\gamma\, \partial^2_{xx} f(x) - s\, \partial_x f(x)\right], \end{aligned} \tag{0.2.38}$$

$\mu_t([0,x])$, pour $x \in (0,1)$, représente alors au temps t la proportion de groupes possédants une proportion de type C inférieur à x. Les paramètres cette équation sont les paramètres intra-groupes $s > 0$ pour la sélection et $\gamma > 0$ pour les fluctuations génétiques uniformes, et entre-groupe avec la fonction ρ qui décrit l'intensité de la sélection.

Une partie des fluctuations dans la reproduction des groupes sont décrites comme négligeables dans cette limite d'échelle, mais pourrait naturellement être incluses. C'est bien plutôt dû ici au fait de rechercher un modèle que l'on sait correctement analyser que véritablement motivé biologiquement.

Plus précisément, les événements de reproduction des groupes peuvent abstraitement être séparés en deux classes : la première classe regroupe la majorité de ces événements et se caractérise par le fait que le taux auquel les événements de cette classe se produisent sur un groupe donné ne dépend pas de la composition de ce groupe, avec un taux de l'ordre du nombre de groupes. On mentionnera ses effets en tant que "fluctuations uniformes" au niveau des groupes. Ce sont ces effets de fluctuations dont la description a été simplifiée comme négligeable. La seconde regroupe un nombre bien plus faible d'événements mais dont le taux d'arrivée dépend de la composition du groupe. On mentionnera sa contribution en tant qu'effets de sélection au niveau des groupes. La même distinction peut se faire sur les événements intra-groupes. Comme il est détaillé dans ce chapitre, les fluctuations uniformes intra-groupes sont cruciales pour donner toute sa force aux effets de sélection entre groupes : je préfère donc éviter d'en parler comme des effets "neutres" !

Ce modèle est tout d'abord introduit et justifié par un modèle individu-centré d'une population subdivisée en groupes, dont les tailles de groupes et le nombre de ceux-ci sont finis et fixés. Nous reprenons pour ce faire le formalisme introduit par Luo et Mattingly dans [Luo13] puis [LM15], avec la limite d'échelle que nous décrivons à mi-chemin entre les deux limites d'échelles obtenues dans [LM15]. Les individus se reproduisent à l'identique, de même que les groupes, avec des générations chevauchantes. On adopte alors l'approche dite de Moran, pour laquelle les naissances arrivant à un certain taux exponentiel amènent à choisir un individu (respectivement un groupe) uniformément dans la population, qui doit laisser sa place au nouveau venu. Il s'agit du seul événement de mortalité inclus dans le modèle, et les effets de sélection sont issus des variations dans le taux de natalité en fonction du type d'individu (resp. de groupe).

Avec l'idée de pouvoir appliquer nos résultats de convergence vers des mesures quasi-stationnaires, nous commençons par établir que μ_t peut s'écrire de la forme $\mu_0 A_t$ selon le formalisme introduit en Section 0.1.3.2. Pour ce faire, on définit d'abord la dynamique aléatoire interne au groupe, dont le générateur infinitésimal est $\mathcal{L}_{WF}$ (cf. Section 0.1.1.2), à savoir :

$$dX_t := -s\, X_t\,(1 - X_t)\, dt + \sqrt{2\gamma\, X_t\,(1 - X_t)}\, dB_t. \tag{0.2.39}$$

On introduit alors la pénalisation suivante, de type Feynman-Kac :

$$Z_t := \exp[-\textstyle\int_0^t \rho(X_s)\, ds]. \tag{0.2.40}$$

Cette valeur représente le taux d'extinction supplémentaire ressenti jusqu'au temps t par un groupe dont l'évolution est donnée par X en comparaison d'une valeur minorante de référence. Le choix naturel pour définir cette valeur de référence est de considérer le taux d'extinction le plus bas à proportion fixée d'individus de type C. La Proposition 1.2.1 du Chapitre 4 que l'on reprend ici fait le lien entre μ_t et la pénalisation de X par Z :

Proposition 0.2.3.1. *Soit pour tout $t \geq 0$ la mesure de probabilité μ_t définie par :*

$$\langle \mu_t \,\big|\, f \rangle := \mathbb{E}\left[f(X_t)\, Z_t\right] / \,\mathbb{E}\left[Z_t\right], \forall\, f \in \mathcal{C}([0,1]), \;. \tag{0.2.41}$$

$(\mu_t)_{t\geq 0}$ est l'unique solution de l'équation (0.2.38).

La deuxième étape est alors d'interpréter Z_t comme $\mathbb{P}(t < \tau_\partial \big| X)$. Pour ce faire, on introduit une variable exponentielle T_∂ indépendante de X et on définit le taux d'extinction via la formule suivante :

$$\tau_\partial := \inf \{t \geq 0 \; ; \; -\ln(Z_t) \geq T_\partial\} . \tag{0.2.42}$$

D'après la Proposition 1.2.2 du Chapitre 4, $(\mu_0 A_t)_{t\geq 0}$, comme définie en (0.1.19) avec $V \equiv 1$, est bien l'unique solution de l'équation (0.2.38). Les points d'équilibre pour la dynamique donnée par l'équation (0.2.38) coïncident donc exactement avec les mesures quasi-stationnaires associées au temps d'extinction τ_∂. Une QSD μ_* sera alors déclarée stable si pour toute mesure de probabilité μ_0 dont la variation totale est assez proche de μ_*, $\mu_0 A_t$ converge vers μ_* en variation totale.

Il faut particulièrement noter que le semi-groupe $(P_t)_{t\geq 0}$ n'est pas ψ-irréductible : Puisque X est absorbée aussi bien en 0 qu'en 1, pour tout $t \geq 0$, $\delta_0 P_t$ (resp. $\delta_1 P_t$) est supporté par $\{0\}$ (resp. $\{1\}$). Sachant que ces supports sont disjoints, aucune mesure d'irréductibilité ψ ne peut faire le pont entre les deux. Il est à noter au passage que l'on obtient immédiatement deux QSD, donc deux points d'équilibre, à savoir δ_0 et δ_1, avec des taux d'extinction resp. $\rho(1)$ et $\rho(0)$, a priori distincts.

Il faut donc d'abord traiter séparément les domaines irréductibles $\{0\}, \{1\}$ et $(0, 1)$ et gérer au cas par cas la façon dont ces effets se combinent. On est donc amené à considérer aussi le taux d'extinction $\tau_{0,1,\partial} := \tau_\partial \wedge \tau_0 \wedge \tau_1$, où τ_0 (resp. τ_1) désigne le temps d'atteinte de 0 (resp. 1). La Proposition 2.0.1 du Chapitre 4, que l'on replace ici dans le contexte de la Section 0.1.3.2, permet d'introduire l'unique QSD et l'unique capacité de survie des groupes polymorphes, avec le taux d'extinction associé ρ_α :

Proposition 0.2.3.2. *Le processus X avec extinction au temps $\tau_{0,1,\partial}$ est exponentiellement quasi-ergodique avec $\rho_\alpha > 0$ comme taux d'extinction et pour tout $\ell \geq 1$, il existe une constante positive qui minore sa capacité de survie sur $[1/\ell, 1 - 1/\ell]$. Par ailleurs, la convergence vers la QSD se fait uniformément en la condition initiale.*

Je montre alors aux Sections 4.2.1-7 que le comportement asymptotique de μ_t dépend uniquement de la relation entre ρ_α, $\rho(0)$ et $\rho(1)$. En supposant arbitrairement $\rho(1) \leq \rho(0)$, tous les cas possibles sont détaillés avec les temps de convergences associés (pour l'exhaustivité, on permet à s d'être négatif, même si le cas $s > 0$ est plus naturel). J'explicite aussi en Section 4.2.8 plusieurs résultats asymptotiques sur la comparaison entre ρ_α, $\rho(0)$ et $\rho(1)$ lorsque les paramètres s, γ et r tendent vers 0 ou au contraire vers l'infini.

De manière succincte, toute QSD stable associée à τ_∂ a nécessairement pour taux d'extinction $\rho(1) \wedge \rho_\alpha \leq \rho(0)$. Elle est unique excepté dans le cas dégénéré où $\rho(1) = \rho(0) \leq \rho_\alpha$: les QSD sont les combinaisons linéaires entre δ_0 et δ_1 et aucune n'est stable selon la définition plus haut. Dès lors que $\rho(1) < \rho(0) \wedge \rho_\alpha$, δ_1 est l'unique QSD stable associée à τ_∂ et δ_0 l'unique QSD instable (avec elle-même pour seul bassin d'attraction). Au contraire, si $\rho_\alpha < \rho(1) \leq \rho(0)$, alors l'unique QSD stable associée à τ_∂ coïncide avec celle de taux d'extinction $\tau_{0,1,\partial}$ sur le domaine $(0,1)$ (au facteur de renormalisation près). La capacité de survie est aussi proportionnelle à celle donnée par $\tau_{0,1,\partial}$, en particulier nulle sur 0 et 1. Le Q-processus est alors identique pour τ_∂ et $\tau_{0,1,\partial}$ (pour les conditions initiales dans $(0,1)$).

En Section 3, je discute de ce que l'on peut en déduire pour le modèle en population finie, ainsi que de ce qui risque de les différencier. Je m'appuie pour ce faire sur les simulations que j'ai réalisées du modèle déterministe limite (avec conditionnement). Cette discussion est encore assez générale, au sens où je n'ai pas su trouver le temps de proprement coder les simulations individus-centrées et de les comparer aux simulations du modèle limite. Je laisse ce travail pour un prochain article, en espérant aussi inclure le modèle limite qui prend aussi en compte les fluctuations uniformes au niveau des groupes. Les questions qui orientent cette discussion sont les suivantes :

— L'analyse asymptotique et les propriétés de stabilités de ces distributions quasi-stationnaires donnent-elles une vision pertinente des domaines d'attractions ?

— Ce type de convergence à taux exponentiel est-il représentatif de la dynamique observée ?

— Les fluctuations intra-groupes jouent-elles bien un rôle important pour permettre aux effets de sélection entre groupes de modérer ceux intra-groupes ?

Une perspective évolutive est aussi considérée, dans une limite de rares mutations. On imagine alors qu'un type résident domine jusqu'à se voir remplacer par un type mutant après de nombreuses tentatives infructueuses d'invasion. En supposant que les mutations qui sont générées ont autant de chances de favoriser la coopération que l'égoïsme, on se demande alors quelles types de mutations font effectivement évoluer le système. On se base pour se faire sur les propriétés attendues de stabilité du processus à taille de population finie autour des QSD stables.

Les conclusions de cette discussion peuvent être résumées ainsi :

• Dans tous les cas où δ_1 est prouvé comme stable, on s'attend à ce que, avec une grande probabilité, les groupes purement constitués de types C finissent par complètement envahir la population. Cela se vérifie pourvu que la condition initiale possède suffisamment de groupes hautement coopératifs pour que la diffusion dans ceux-ci génère rapidement une proportion non-négligeable de purs groupes C.

• Néanmoins, le comportement limite détaillé dans [LM15] sans fluctuations démographiques intra-groupes montre un cas où le individus de type D envahissent rapidement même une population avec une proportion importante de type C. On retrouve cet effet lorsque les fluctuations démographiques intra-groupes sont faibles : une très grande proportion des groupes deviennent visiblement de plus en plus dominés par les individus D. Le fait que les purs groupes D sont ceux qui se reproduisent le moins bien ne suffit pas à empêcher la fixation d'un grand nombre de groupes comme purs groupes D. Introduire de rares mutations d'individus D en individus C n'y changerait sans doute pas grand chose.

• Le cas $\rho(1) < \rho_\alpha < \rho(0)$ caractérise le fait qu'une sous-population de groupes polymorphes est mieux capable de se maintenir que les purs groupes D. Notamment, cela signifie que la domination par les individus D serait tout de même empêchée, même si on perturbe le modèle en introduisant de rares événements de migration entre groupes ou de mutations d'individus C en individus D. Ces conclusions de stabilité apparaissent bien sur les simulations si tant est que la QSD polymorphe α est rapidement approchée.

• De même, dans le cas $\rho_\alpha < \rho(1) \leq \rho(0)$, l'état polymorphe est a priori très stable. On observe bien que μ_t converge vers une unique limite dès lors que la proportion de groupes polymorphes n'a pas été violemment amputée au départ. Notamment, la proportion de purs groupes C et de purs groupes D se stabilisent.

• Lorsqu'on se restreint à ne considérer que les groupes polymorphes (i.e. en regardant $\mu A_t^{0,1}$), il existe une unique mesure quasi-stationnaire α, avec une convergence à vitesse exponentielle vers celle-ci. On pourrait donc s'attendre à ce que celle-ci soit rapidement approchée par $\mu A_t^{0,1}$. Sur les simulations, il n'est cependant pas clair que le domaine d'attraction de α soit très étendu au-delà du voisinage de l'état purement coopératif, en particulier pour de faibles fluctuations intra-groupes. La convergence vers celle-ci et son influence n'a potentiellement pas d'effet significatif si la proportion de groupe majoritairement de type C est trop faible. Un attracteur transitoire est alors susceptible de se manifester pour $\mu A_t^{0,1}$. La distribution de ce dernier est alors a priori largement dominée par les individus de type D. On peut l'interpréter en pratique comme une mesure quasi-stationnaire alternative, avec un taux d'extinction en général plus fort que $\rho(0)$. Cet effet est visible lorsque les transitions depuis un groupe majoritairement de type D jusqu'à majoritairement de type C sont trop coûteuses comme déviation du processus X pour être compensées rapidement par l'écart entre les taux d'extinction. De manière presque symétrique, la "vraie" QSD est dominée alors par les individus de type C, dont le maintien garantit sa stabilité. Les transitions du voisinage de la "pseudo"-QSD à la "vraie" QSD sont suffisamment longues pour que l'attraction vers la pseudo-QSD soit bien visible.

• En termes évolutifs, pour des effets de sélection très forts par rapport aux fluctuations intra-groupes, la situation du statu-quo est privilégiée : la probabilité de fixation d'une mutation qui défavorise son porteur sur un niveau de sélection n'est que très peu compensée par l'avantage qu'elle pourrait donner sur un autre niveau.

• Toujours en termes évolutifs, la coopération est privilégiée lorsqu'on accroît le niveau de fluctuations génétiques intra-groupes (i.e. la parenté entre individus d'un même groupe). Les mutations qui défavorisent le groupe de ceux qui les portent ont une probabilité de fixation très diminuée par rapport au cas inverse : la relation de comparaison entre la force de ces effets dans le cadre de sélection faible semble perdre ici sa pertinence quand les effets se conjuguent.

0.2.4 Cliquet de Muller

Le chapitre 5 aborde la question de la méta-stabilité dans le modèle dit du cliquet de Muller. Ce modèle a été introduit dans le but de quantifier l'efficacité avec laquelle la sélection naturelle permet d'empêcher via un mode de reproduction purement asexué la fixation et l'accumulation de mutations délétères. Contrairement au cas sexué, les mutations délétères ne peuvent alors être purgées qu'avec l'extinction des lignées qui en sont porteuses. On est par ce modèle en mesure de préciser le bénéfice d'avoir un mode de reproduction sexuée ou en tout cas avec une transmission dite horizontale entre individus de lignées différentes.

Pour simplifier, toutes les mutations sont délétères et indiscernables et la taille de population est fixée. Le taux de reproduction moyen d'un individu est ainsi donné comme fonction uniquement du nombre des mutations qu'il porte, que l'on compare au taux moyen de reproduction des autres individus. A cause de la variabilité génétique uniforme, il est inéluctable d'observer la disparition irréversible de la sous-population des individus de phénotype optimal, i.e. sans mutation. Pour une population capable de survivre avec un taux de croissance raisonnable, il apparaît a priori plus réaliste de supposer que de tels événements de disparition sont relativement distants. On s'intéresse plus particulièrement au cas où la distribution des mutations au sein de la population se stabilise entre chacune de ces disparitions, avec à chaque fois une translation d'une mutation en plus. Cela justifie ce nom de "cliquet" donné à ce modèle en référence à ces roues dentées qui ne peuvent aller que dans un sens avec des transitions ponctuelles entre équilibre successifs. Les clics font donc ainsi référence à l'extinction de cette sous-population optimale et ces quasi-équilibres sont l'objet de l'étude du Chapitre 5.

Le modèle principal d'étude dans ce chapitre est le système d'EDS suivant, de type Flemming-Viot et introduit dans [EPW09] et plus largement présenté dans [PSW12]0 en incluant des mutations compensatoires. Pour $i \in [\![0, d]\!]$, avec pour objectif le cas $d = \infty$, $X_i(t)$ spécifie la proportion d'individus de la population porteurs d'exactement i mutations délétères au temps t (avec la convention $X_{-1} \equiv 0$). Ces proportions satisfont :

$$\begin{aligned} \forall\, i \le d, \quad dX_i(t) &= \alpha(M_1(t) - i)\, X_i(t)\, dt + \lambda(X_{i-1}(t) - \mathbf{1}_{\{i<d\}} X_i(t))\, dt \\ &\quad + \sqrt{X_i(t)}\, dW_t^i - X_i(t)\, dW_t \end{aligned} \tag{$S^{(d)}$}$$
$$\text{où } W_t := \textstyle\sum_{j=0}^d \int_0^t \sqrt{X_j(s)} dW_s^j, \quad M_1(t) := \sum_{i=0}^d i\, X_i(t),$$

où $(W^i)_{i\ge 0}$ est une familles de Mouvement Browniens mutuellement indépendants. On considère alors comme extinction le clic défini plus haut, i.e. ici :

$$\tau_\partial := \inf\{t \ge 0 \ ; \ X_0(t) = 0\}. \tag{0.1}$$

Pour $d \in [\![1, \infty[\![$, le processus X est considéré sur

$$\mathcal{X}_d := \{(x_k)_{k \in [\![0,d]\!]} \in [0,1]^{d+1} \ ; \ \sum_k x_k = 1\},$$

tandis que pour $d = \infty$, l'espace d'état est réduit à :

$$\mathcal{X}_\infty := \{(x_k)_{k \in \mathbb{Z}_+} \in [0,1]^\infty \ ; \ \sum_k x_k = 1 \ , \ \sum_k k^6 x_k < \infty\}.$$

Theorem 0.2.5. *Que d soit fini ou infini, le processus X avec extinction au temps τ_∂ est exponentiellement quasi-ergodique. Pour tout $\ell \geq 3$, la capacité de survie est uniformément minorée sur $\mathcal{D}_\ell := \{x \in \mathcal{X}_d \,\big|\, x_0 \in (1/\ell, 1 - 1/\ell), \sum_{i=0}^d i^3 x_i \leq \ell\}$.*

Les paramètres impliqués dans ces résultats de convergence peuvent être choisis uniformément en $d \in [\![1, \infty]\!]$. Pour d et $L \in \mathbb{N}$ assez grands, ils peuvent être choisis uniformément pour l'extinction au temps $\tau_\partial \wedge T_{\mathcal{D}_L}$ où $T_{\mathcal{D}_L}$ est le temps de sortie de $\mathcal{D}_L$.

La borne $1 - 1/\ell$ au voisinage de 1 pourrait vraisemblablement être remplacée par 1 avec une analyse a posteriori. Notons $\mathcal{D}_\ell^1 := \{x \in \mathcal{X}_d \,\big|\, x_0 \in (1 - 1/\ell, 1], \sum_{i=0}^d i^3 x_i \leq \ell\}$. Une analyse du bord "$x_0 = 1$" doit permettre de prouver qu'il existe $t, c > 0$ et $\ell' \geq 1$ tel que pour tout $x \in \mathcal{D}_\ell^1$, $\mathbb{P}_x(X_t \in \mathcal{D}_{\ell'}, t < \tau_\partial) \geq c$. On en déduirait via $h(x) := e^{-\lambda t} \langle \delta_x P_t \,\big|\, h \rangle$ que η est aussi uniformément minorée sur $\mathcal{D}_\ell^1$. Je laisse ces détails au lecteur intéressé.

J'introduis au préalable un modèle en génération discrète et taille de population finie dont la limite d'échelle est le processus ci-dessus. Cela permet à la fois de donner plus de sens aux paramètres de ce modèle et de fournir un cadre simplifié, néanmoins nouvellement traité, pour l'étude de la quasi-stationnarité. Par simplicité, les transitions du modèle en population finie se font par générations discrètes, et on spécifie la manière dont les parents sont choisis avec un biais selon le nombre de mutations délétères. De manière générale, on s'attend à ce que l'étude du modèle limite donné par le système $(S^{(\infty)})$ fournisse néanmoins une description plus générale qui est moins dépendante des détails du modèle individu-centré considéré.

L'effet de la contre-sélection des porteurs de mutations délétères se traduit via le terme proportionnel à α dans le terme de dérive. Puisque la taille de population reste fixe, le taux de croissance des individus est globalement translaté pour être de 0 en moyenne sur toute la population. Nous supposons par simplicité que chaque mutation délétère induit un effet de sélection à la fois additif et identique. Le taux de croissance des individus porteurs de i mutations est donc proportionnel à la différence entre i et le nombre moyen de mutations, i.e. $M_1(t)$. L'advenue de nouvelles mutations est représentée par le terme proportionnel à λ dans le terme de dérive. λ correspond au taux auquel les individus porteurs de i mutations en reçoivent une additionnelle et deviennent ainsi des individus porteurs de $i+1$ mutations (sans qu'il y ait de dépendance en i sur ce taux). Finalement, le choix quasi-uniforme du parent pour chaque enfant de la génération suivante donne lieu au dernier terme de martingale. Pour simplifier nos notations, nous nous plaçons dans l'échelle de temps pour laquelle ce terme n'a pas de coefficient. Cela correspond à une accélération du temps $t \mapsto t'/N$, où t' est comptée en nombre de générations et N la taille de population. Dans un cadre plus général, l'accélération du temps serait du type $t \mapsto t'/N_e$, où N_e est la taille de population efficace, et où t' est comptée par rapport à l'échelle de temps de reproduction des individus.

Comme énoncé au Théorème 2.3 du Chapitre 5, le processus X avec $d = \infty$ et τ_∂ pour temps d'extinction est exponentiellement quasi-ergodique selon les notations de la Section 0.1.3.2. Ce théorème est précédé d'autres résultats similaires, qui établissent un tel résultat dans plusieurs cadres simplifiés et dont les preuves éclairent celle du Théorème 2.3. Tout d'abord, le cadre en population finie et borne finie $d \geq 1$ sur le nombre de mutations est une application très directe du Théorème de Perron Frobenius. La preuve est facilement déduite de la preuve du Théorème 2.1 et donc seulement esquissée. Le théorème 2.1 traite du modèle en population finie où les individus pourraient théoriquement avoir un nombre non borné de mutations. Le coeur de la preuve consiste alors à justifier qu'il est très exceptionnel de trouver la population avec un individu porteur d'un trop grand nombre de mutations. Avec la preuve du théorème 2.2, on traite ensuite le cas de la solution du système $(S^{(d)})$ pour d fini. Comparé aux résultats récents de quasi-stationnarité sur les diffusions, la difficulté principale dans ce cadre vient des effets de bords singuliers. Les bords du domaine sont irréguliers (avec des "coins" de toute dimension plus petite que d), avec une jonction particulière entre bord absorbant et bord répulsif. La preuve du Théorème 2.3 nécessite de gérer toutes ces difficultés, avec un niveau supplémentaire de difficulté. Il faut user d'astuce pour rendre négligeable l'influence de ces individus porteurs d'un grand nombre de mutations, qui ne disparaissent jamais vraiment. Pour y arriver, nous avons eu besoin d'un contrôle des moments sur la distribution des mutations dans la population : que ceux-ci restent grands est exceptionnel et il est rare de les voir réaugmenter trop vite. Ces contrôles nous permettent d'ailleurs d'obtenir un résultat de tension uniforme sur le paramètre d sur tous les moments en moyenne selon la distribution quasi-stationnaire de dimension d. Ce résultat est énoncé à la Proposition 2.3.1.

0.2.5 Différentes échelles de temps pour les modèles individus-centrés

Le dernier chapitre a un statut particulier dans cette thèse, au sens où il n'y est pas question de preuves de résultats nouveaux. C'est en fait une présentation personnelle d'un certain nombre de résultats récents de probabilités appliquées à la biologie, avec pour fil directeur les présentations données à ce sujet lors des Journées MAS en août 2018. La propriété suivante m'a paru rassembler les modèles présentés. Les auteurs des articles en question dérivent de modèles stochastiques individus-centrées la description asymptotique en grande taille de population d'une loi ou d'un processus, dans une échelle de temps longue par rapport à la durée de vie des individus. J'ai mis l'accent sur la variété de telles limites et des échelles de temps associées, en tentant de préciser comment ces effets étaient susceptibles de se combiner.

Tout d'abord, les travaux de M. Hoffman et A. Marguet portent sur l'estimation statistique de traits de population, i.e. ce qui caractérise les lois d'évolution des individus, à partir de l'histoire de vie de ces individus, cf [HM19]. On peut en déduire sous quelles échelles de temps l'aléa individuel est suffisamment intégré pour que la description en terme de population prenne vraiment sens. C'est à ces échelles que peuvent se faire ressentir les effets de sélection capable de discriminer des sous-populations selon les traits qui les différencient.

Je reprends alors les résultats de ce Chapitre 4, que j'avais présenté lors de cette conférence. Cela me permet de justifier que des conflits peuvent arriver entre différents effets de sélection naturelle. Dans le cas de tels conflits, les effets de sélection naturelle peuvent potentiellement dépendre de l'échelle de temps considéré, quand bien même elles seraient suffisantes pour distinguer les traits des individus.

Les autres modèles portent sur la trajectoire évolutive de la population, avec l'idée que les mutations sont relativement peu fréquentes et donc limitantes. Dans l'article de N. Champagnat et B. Henry [CH20], deux cadres sont présentés où les techniques de Grandes Déviations et d'optimisation apparaissent. Dans le premier, les mutations ne sont pas distinguées et sont représentées via un noyau de type Laplacien. Néanmoins, dans l'échelle de temps d'évolution considérée, cette variabilité interne à la population apparaît très réduite par rapport à la dynamique sélective qu'elle induit. Le coût en terme de probabilité pour voir apparaître une telle dynamique est exactement compensé par l'avantage sélectif qu'elle procure en terme de reproduction des individus. Dans le deuxième cadre, les traits d'individus possibles sont distingués et les transitions par mutation entre ces types sont très rares. De même, une dynamique quasi-déterministe apparaît lorsque l'avantage reproductif est capable de compenser l'exceptionnalité de ces transitions d'états.

Enfin, dans la continuité de l'exposé de V.C. Tran, nous abordons l'échelle de temps la plus longue pour laquelle la dynamique écologique de remplacement de populations par des sous-populations mutantes est instantanée. Dans ces modèles, on observe alors une succession de sauts évolutifs, où le nouveau type domine nécessairement le précédent. On dira alors que sa fitness est meilleure que le type résident dans une population résidente. Ce type de modèle a d'abord été introduit dans [MG+96], puis justifié comme limite d'échelle de modèles individus-centrés dans [Ch06]. Les intervalles de temps entre ces événements brutaux sont aléatoires et donnés via un processus de Poisson par l'arrivée imprévisible de la première mutation favorable. Pour simplifier l'analyse, ces relations de domination sont transitives dans ces modèles et la coexistence de différents types n'est pas autorisée. Les principes devraient néanmoins se justifier hors de ce cadre. On peut alors en parallèle intégrer à ce modèle la dynamique de propagation d'autres traits, tant qu'ils sont considérées neutres (pour la reproduction des individus), cf [BF⁺15].

Dans l'échelle de temps encore plus longue où de tels événements se succèdent en nombre, on peut alors inférer une dynamique déterministe. Sa vitesse est alors proportionnelle à la variation infinitésimale de fitness et à l'effet quadratique des mutations pondéré par leur taux d'apparition, sans oublier évidemment le taux de mutation et la taille de population résidente à l'équilibre. Si cette relation finale est simple et agréable pour une modélisation évolutive, sa justification propre en terme de processus individus centrés est néanmoins particulièrement exigeante. Notamment, on doit justifier en même temps que les mutations sont assez faibles pour que la fitness progresse de manière indistinguable mais aussi assez fortes pour que les événements de remplacement lors des mutations puissent être considérés ponctuels et bien distingués. A ces difficultés doivent se combiner les contrôles sur les fluctuations exceptionnelles de taille de population hors de leur équilibre, ce qui rend la preuve fournie en [BBC17] tout à fait impressionnante. On doit pour autant garder en mémoire que ces difficultés sont autant de limitations de l'interprétation déterministe proposée.

1 Unique quasi-stationary distribution, with a possibly stabilizing extinction

This chapter is taken from the preprint with the same name whose ArXiv reference can be found at the end of the bibliography (here is the link for the pdf version : [Chapter 1]).

Abstract

We establish sufficient conditions for exponential convergence to a unique quasi-stationary distribution in the total variation norm. These conditions also ensure the existence and exponential ergodicity of the Q-process, the process conditionned upon never being absorbed. The technique relies on a coupling procedure that is related to Harris recurrence (for Markov Chains). It applies to general continuous-time and continuous-space Markov processes. The main novelty is that we modulate each coupling step depending both on a final horizon of time (for survival) and on the initial distribution. By this way, we could notably include in the convergence a dependency on the initial condition. As an illustration, we consider a continuous-time birth-death process with catastrophes and a diffusion process describing a (localized) population adapting to its environment.

Keywords : quasi-stationary distribution ; survival capacity ; Q-process ; Harris recurrence ; birth-and-death process ; diffusion

1.1 Introduction

1.1.1 Presentation

Given a continuous-time and continuous-space Markov process with an absorbing state, we are interested in this work in the long time behavior of the process conditionally on not being absorbed (not being "extinct"). The proof can be adapted quasi verbatim to discrete time processes.

More precisely, our first concern is on the marginal —at time t— conditioned on not being extinct —also at time t— (the **MCNE** in short). We wish to highlight key conditions on the process such that these MCNE converge as $t \to \infty$ to a unique distribution α. This limiting distribution is called the **quasi-stationary distribution** (the **QSD**) —cf Subsections 1.1.3 and 1.2.2, or chapter 2 in [CMS13] for more details on this notion. The techniques we use allow us to establish not only the existence and uniqueness of the QSD, but also the exponential convergence in total variation norm, cf Theorem 1.2.1.

In addition, we deduce, under the same conditions, the existence of a specific eigenfunction h of the infinitesimal generator, with the same eigenvalue as the QSD. As time goes to infinity, the renormalizing factor at time t behaves asymptotically as $h\ \exp[-\lambda t]$, cf (1.2.4) and Theorem 1.2.2. This convergence motivates the name **survival capacity** that we give to h (sometimes described as the "reproductive value" in ecological models). Again, the convergence is exponential, but not uniform over the state space in our case. Moreover, we deduce the existence of the **Q-process**. Its marginal at time t is given by the limit (as $T \to \infty$) of the marginal of the original process at time t conditioned on not being extinct at time T, cf Theorem 1.2.3. Thus, it is often described as the process conditioned to never be absorbed. Finally, we deduce for the Q-process the existence and uniqueness of its stationary distribution β together with a property related to exponential ergodicity.

To deduce these results, our aim is to combine a large degree of generality with conditions as easy to verify as possible. A specificity of our approach is that it allows to deduce a coupling procedure depending on the initial condition that ensures a contraction in total variation towards the limiting distribution. It is only for commodity that we have restricted the analysis to cases where there is a unique QSD. One can find in Chapter 4 an application to group selection models where our procedure of proof is exploited to deduce the convergence to some QSD in a specific basin of attraction.

Although there is already a vast literature on QSDs (see notably the impressive bibliography collected by Pollett [Po15]), the approach we follow seems to have been explored only in the very recent years. For a review on the results that were previously obtained, we refer notably to general surveys as in [CMS13], in [DP13] or more specifically for population dynamics in [MV12]. We see already in these surveys how essential is the role played by the spectral theory. The spectral theory is very effective both to relate the QSD and the survival capacity to the first eigenvector of a diagonalizable operator and to identify the convergence rate as the gap between the first and the second eigenvalues (cf e.g. [CCM16]). The principal drawback of the spectral theory is that it usually relies on reversibility. Certainly, for 1 dimensional processes, this condition of time-symmetry is quite easily satisfied ; while, more generally, it can be deduced from conditions easy to verify (detailed balance notably). This may explain why reversibility is so extensively studied. Yet, it is a very restrictive condition for higher dimensions, as it is well explained in the appendix A of [CCM17].

Alternative methods are usually much less effective. In [CM+11], the authors prove the existence of the QSD via a Tychonov fixed point theorem. Another proof for the existence of the QSD is presented in [FK+95] for Markov Chains on $\mathbb{Z}_+$, based on compactness arguments and renewal techniques. In [BP10], the authors prove, under quite stringent conditions, the existence and uniqueness of the QSD and propose estimations of this QSD up to some computable time, again with renewal arguments. The authors of [DM15] relate the speed of convergence to QSD to the one of a related Doob's transform towards its stationary distribution. Yet the conditions of the last two papers seem to apply essentially to discrete-space processes, or at least when the extinction is in some sense uniformly bounded. The existence of the QSD and the survival capacity has also been related, at least for discrete time and discrete space, to the notion of R-positivity [SV96]. This is especially useful when the process is easily described by generating functions (in particular for Galton-Watson processes) but seems quite an abstract criterion otherwise. Still, it provides the main principle of focusing on the exponential rate of extinction, which is at the core of our study.

We exploit the idea, first exploited in [CV16], to rely on a more constructive method in the form of a strong regeneration condition, analogous to Harris' recurrence (what we can see maybe a bit more clearly in the present work). At the foundation of our proof is clearly the characterization given in [CV16] of the uniform exponential convergence to a unique QSD. As we can see in the applications we present (cf Section 1.4) lack of reversibility is not at all an issue for our proofs. The hope with these techniques is also to include easily more complexity on the stochastic models, (for instance time inhomogeneity) while relying on the same method with uniform in time estimates (cf [CV18b], [BCG19], [DV16]).

Our result of exponential convergence to a unique QSD applies in the general context of possibly non-reversible continuous-time and continuous-space Markov processes, with a multiplicative constant being possibly non-uniform over the initial conditions, for which little is known. To our knowledge, such results have only been obtained as we were working on the current article, in [CV17c], [FRS19] and [BCGM19], with seemingly the same foundation. Note that the proofs in [CV17c] and [BCGM19] concern the convergence towards a unique Yaglom distribution, which may not be unique as a QSD, for any initial conditions with a light enough tail. Also, it is proved in [BCGM19] that their conditions are not only sufficient, but also necessary. This is a good omen for our set of assumptions, seeing how closely their assumptions are related to ours.

These approaches are reminiscent of older techniques involving, in the case of discrete-time and discrete-space processes, the notion of "R-positive-recurrence" (cf e.g. [FK+96], [SV96]). Yet, the results in [CV17c], [FRS19], [BCGM19] and in the present paper are apparently the first to ensure such exponential convergence to the QSD (with maybe poor yet explicit rates) as well as to the stationary distribution for the Q-process. More precisely, the proofs in [FRS19] and [BCGM19] rely on the definition of Lyapunov functions (and two associated exponential rates) that we can connect respectively to our assumption $(A2)$ and our condition on ρ_S. In practice, it does not seem so clear to us how to find such Lyapunov functions especially when one wishes to combine simple bounds on different parts of the space. So we believe that our assumptions are easier to verify in many examples (cf e.g. our second application).

Besides, the techniques exploited in [CV17c], [FRS19] and [BCGM19] are quite different from ours. Whereas they deal with uniform bounds on a weighted norm, our proofs are much more constructive and rely on a control on entry times of core sets thanks to the competition between different behaviors. It extends to models where the uniqueness of the QSD does not hold due to transitivity conditions, as one can observe in the applications we have in Chapter 4. In particular, our work offers a new constructive perspective even for the results in [CV16] (cf Subsection 1.5.3) since the coupling steps which we introduce apply directly to the MCNE (and not to their linearized versions). Our arguments are in fact not so far from those presented in [FK$^+$96], adapted for the case of continuous-time and continuous-space processes. Our approach provides opportunities for extensions, quasi verbatim for the discrete-time case, probably not difficult for time non-homogeneous sub-Markov processes and for cases where there are several QSD, possibly with sufficiently heavy tail. Surely, in such cases without uniqueness, using Lyapunov criteria as in [CV17c] is certainly very efficient. Given the link between our assumption $(A2)$ and these Lyapunov criteria (cf Subsection 1.2.4), our approach could then be easily reformulated in terms of such Lyapunov criteria.

Like in [CV17c] and [FRS19], our proof requires a seemingly abstract assumption, $(A3)$, to compare the asymptotic of survival. It might be rather difficult to establish for continuous-time and continuous-space processes including jumps, that are not expected to be strong Feller (so that the arguments in [FRS19] do not apply). Contrary to the case of classical diffusive processes, it indeed cannot be inferred as easily from a version of Harnack inequality. In Chapter 2, we provide a very efficient and more easily verified condition which ensures $(A3)$, given the other assumptions. Since this condition is technical and may appear too abstract without the illustration of various examples, we encourage any interested reader to look at Chapters 3 and 5, besides the simple illustrations given in Chapter 2.

The remainder of Section 1.1 is organized as follows. Subsection 1.1.2 describes our general notations ; Subsection 1.1.3 presents our specific setup of a Markov process with extinction ; and Subsection 1.1.4 the decomposition of the state space on which we base our assumptions. Subsection 1.2.1 presents the main set of conditions which we show to be sufficient for the exponential convergence to the QSD. Subsection 1.2.2 states the three main theorems of the present paper, dealing respectively with the QSD, the survival capacity and the Q-process. The conditions that we present are then certainly numerous ; yet we believe that they are quite convenient to deal with in practice, except maybe for $(A3)$, for which we can only give a few hints in the present work (cf Subsection 1.2.3 and 1.4.2). In Subsection 1.2.4, we compare the specific form of convergence that we propose to what can be found in the literature. In Subsection 1.2.5, we discuss the different assumptions that we introduce. We then explain why we can simply consider the first from our two sets of assumptions in order to complete the proofs of the three theorems. These proofs are finally given in Section 1.5. In Section 1.4, we present two applications of our general theorems. Theses results seem to be new, but concern toy-models. We hope that they will help the reader get insight on our approach. The application of our theorems to more significant biological models is intended for following work.

1.1.2 Elementary notations

In the following, the notation $k \geq 1$ has generally to be understood as $k \in \mathbb{N}$ while $t \geq 0$ (resp. $c > 0$) should be understood as $t \in \mathbb{R}_+ := [0, \infty)$ (resp. $c \in \mathbb{R}_+^*$ $:= (0, \infty)$). In this context (with $m \leq n$), we denote classical sets of integers by : $\mathbb{Z}_+ := \{0, 1, 2...\}$, $\mathbb{N} := \{1, 2, 3...\}$, $[\![m, n]\!] := \{m,\ m+1, ..., n-1,\ n\}$, where the notation $:=$ makes explicit that we define some notation by this equality. For maxima and minima, we usually denote : $s \vee t := \max\{s, t\}$, $s \wedge t := \min\{s, t\}$. Accordingly, for a function ψ, $\psi^\wedge$ (resp. $\psi^\vee$) will usually be used for a lower-bound (resp. for an upper-bound) of ψ.

Let $\Big(\Omega; (\mathcal{F}_t)_{t\geq 0}; (X_t)_{t\geq 0}; (P_t)_{t\geq 0}; (\mathbb{P}_x)_{x\in\mathcal{X}\cup\partial}\Big)$ be a time homogeneous strong Markov process with cadlag paths on some Polish space $\mathcal{X} \cup \{\partial\}$ [[RW00], Definition III.1.1], where $(\mathcal{X}; \mathcal{B})$ is a measurable space and $\partial \notin \mathcal{X}$. We also assume that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous and complete. We recall that $\mathbb{P}_x(X_0 = x) = 1$, P_t is the transition function of the process satisfying the usual measurability assumptions and Chapman-Kolmogorov equation. The first entry time (resp. the first exit time) of $\mathcal{D}$, for some domain $\mathcal{D} \subset \mathcal{X}$, will generally be denoted by $\tau_{\mathcal{D}}$ (resp. by $T_{\mathcal{D}}$). While dealing with the Markov property between different stopping times, we wish to clearly indicate with our notation that we introduce a copy of X (ie with the same semigroup (P_t)) whose dependency upon X is limited to its initial condition. This copy (and the associated stopping times) is then denoted with a tilde ($\widetilde{X}$, $\widetilde{\tau_\partial}$, $\widetilde{T}_{\mathcal{D}}$ etc.). In the notation $\mathbb{P}_{X_{\tau_E}}(t - \tau_E < \widetilde{\tau_\partial})$ for instance, τ_E and X_{τ_E} refer to the initial process X while $\widetilde{\tau_\partial}$ refers to the copy $\widetilde{X}$.

1.1.3 The stochastic process with absorption

We consider a strong Markov processes absorbed at ∂ : the cemetery. More precisely, we assume that $X_s = \partial$ implies $X_t = \partial$ for all $t \geq s$ and that the extinction epoch : $\tau_\partial := \inf\{t \geq 0\ ;\ X_t = \partial\}$ is a stopping time. Thus, the family $(P_t)_{t\geq 0}$ defines a non-conservative semigroup of operators on the set $\mathcal{B}_+(\mathcal{X})$ (resp. $\mathcal{B}_b(\mathcal{X})$) of positive (resp. bounded) $(\mathcal{X}, \mathcal{B})$-measurable functions. For any probability measure μ on $\mathcal{X}$, that is $\mu \in \mathcal{M}_1(\mathcal{X})$, and $f \in \mathcal{B}_+(\mathcal{X})$ (or $f \in \mathcal{B}_b(\mathcal{X})$) we use the notations : $\mathbb{P}_\mu(.) := \int_{\mathcal{X}} \mathbb{P}_x(.)\, \mu(dx)$, $\langle \mu \,\big|\, f \rangle := \int_{\mathcal{X}} f(x)\, \mu(dx)$.

We denote by $\mathbb{E}_x$ (resp. $\mathbb{E}_\mu$) the expectation corresponding to $\mathbb{P}_x$ (resp. $\mathbb{P}_\mu$).

$$\begin{aligned} \mu P_t(dy) &:= \mathbb{P}_\mu(X_t \in dy), & \langle \mu P_t \,\big|\, f \rangle &= \langle \mu \,\big|\, P_t f \rangle = \mathbb{E}_\mu[f(X_t)], \\ \mu A_t(dy) &:= \mathbb{P}_\mu(X_t \in dy \,\big|\, t < \tau_\partial), & \langle \mu A_t \,\big|\, f \rangle &= \mathbb{E}_\mu[f(X_t) \,\big|\, t < \tau_\partial], \end{aligned}$$

μA_t is what we called the MCNE (at time t, with initial distribution μ).

In this setting, the family $(P_t)_{t\geq 0}$ (resp. $(A_t)_{t\geq 0}$) defines a linear but non-conservative semigroup (resp. a conservative but non-linear semigroup) of operators on $\mathcal{M}_1(\mathcal{X})$ endowed with the total variation norm : $\|\mu\|_{TV} := \sup\{|\mu(A)| \;;\; A \in \mathcal{B}\}$ for $\mu \in \mathcal{M}(\mathcal{X})$.

A probability measure α is said to be the *quasi-limiting distribution* of an initial condition μ if : $\forall B \in \mathcal{B}$, $\lim_{t\to\infty} \mathbb{P}_\mu(X_t \in B \,\big|\, t < \tau_\partial) := \lim_{t\to\infty} \mu A_t(B) = \alpha(B)$.

It is now classical (cf e.g. Proposition 1 in [MV12]) that α is then a quasi-stationary distribution or QSD, in the sense that : $\forall t \geq 0$, $\alpha A_t(dy) = \alpha(dy)$.

Our first purpose will be to prove that the assumptions in Subsection 1.2.1 provide sufficient conditions for the existence of a unique quasi-limiting distribution α, independent of the initial condition.

1.1.4 Specification on the state space

In the following Theorems, we will always assume the following decomposition of $\mathcal{X}$:

Assumption 0. *: **"Exhaustion of $\mathcal{X}$"** There exists a sequence $(\mathcal{D}_\ell)_{\ell\geq 1}$ of closed subsets of $\mathcal{X}$ such that :*

$$\forall n \geq 1,\; \mathcal{D}_\ell \subset \mathcal{D}_{\ell+1}^\circ \quad \textit{and} \quad \bigcup_{\ell\geq 1} \mathcal{D}_\ell = \mathcal{X}. \tag{A0}$$

This sequence will serve as a reference for the following statements. For instance, we will have control on the process through the fact that the initial distribution belongs to some set of the form :

$$\mathcal{M}_{\ell,\xi} := \{\mu \in \mathcal{M}_1(\mathcal{X}) \;;\; \mu(\mathcal{D}_\ell) \geq \xi\}, \qquad \text{with } \xi \in (0,1). \tag{1.1.1}$$

Note that for any $\xi > 0$: $\mathcal{M}_1(\mathcal{X}) = \bigcup_{\ell\geq 1} \mathcal{M}_{\ell,\xi}$. Let also :

$$\mathbf{D} := \{\mathcal{D} \;;\; \mathcal{D} \text{ is closed and there exists } \ell \geq 1 \text{ such that } \mathcal{D} \subset \mathcal{D}_\ell\}. \tag{1.1.2}$$

1.2 Exponential convergence to the QSD

1.2.1 Hypotheses

We recall that for any set $\mathcal{D}$, we defined the first exit and entry times as :

$$T_{\mathcal{D}} := \inf \{t \geq 0 \;;\; X_t \notin \mathcal{D}\}\,, \qquad \tau_{\mathcal{D}} := \inf \{t \geq 0 \;;\; X_t \in \mathcal{D}\}\,.$$

Assumption 1. *: **"Mixing property"***
There exists some probability measure $\zeta \in \mathcal{M}_1(\mathcal{X})$ such that, for any $\ell \geq 1$, there exists $L \geq \ell$, $c, t > 0$ such that :

$$\forall\, x \in \mathcal{D}_\ell, \qquad \mathbb{P}_x\left[X_t \in dx \;;\; t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\, \zeta(dx). \tag{A1}$$

Assumption 2. *: **"Escape from the Transitory domain"***
For given $\rho > 0$ and $E \in \mathbf{D}$:

$$e_{\mathcal{T}} := \sup_{x \in \mathcal{X}} \mathbb{E}_x\left(\exp\left[\rho\left(\tau_\partial \wedge \tau_E\right)\right]\right) < \infty. \tag{A2}$$

The order ρ in the previous exponential moment is required to be larger than the following **"survival estimate"** that involves the measure ζ in $(A1)$:

$$\rho_S := \sup \left\{ \rho \geq 0 \,\middle|\, \sup_{L \geq 1} \liminf_{t > 0}\; e^{\rho t}\, \mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) = 0 \right\} \vee 0. \tag{1.2.1}$$

Assumption 3. *: **"Asymptotic comparison of survival"** For a given $E \in \mathbf{D}$ and $\zeta \in \mathcal{M}_1(\mathcal{X})$:*

$$\limsup_{t \to \infty} \sup_{x \in E} \frac{\mathbb{P}_x(t < \tau_\partial)}{\mathbb{P}_\zeta(t < \tau_\partial)} < \infty \tag{A3}$$

We say that assumption $(\mathbf{A})$ holds, whenever :
"$(A1)$ holds for some $\zeta \in \mathcal{M}_1(\mathcal{X})$ and a sequence $(\mathcal{D}_\ell)$ that satisfies $(A0)$. Moreover, there exist $E \in \mathbf{D}$ such that $(A2)$, holds with some $\rho > \rho_S$ as well as $(A3)$."

As we shall see in Subsection 1.3, $(A1)$ implies that $\rho_S < \infty$. In order to ensure assumption $(\mathbf{A})$, we may not need to estimate precisely ρ_S : it is possible (depending on the process) that $(A2)$ is satisfied for any potential value for $\rho > 0$ (where E is likely to depend on ρ). Moreover, ρ_S as well as assumption $(A3)$ actually do not depend of the choice of ζ satisfying $(A1)$.

1.2.2 Main Theorems : the simplest set of assumptions

Theorem 1.2.1. *Assume that assumption* $(\mathbf{A})$ *holds. Then, there exists a unique QSD* α. *Moreover, we have exponential convergence to* α *of the MCNE's at a given rate* $\gamma > 0$. *More precisely, for any pair* $\ell \geq 1$ *and* $\xi \in (0,1)$, *there exists* $C = C(\ell, \xi) > 0$ *such that :*

$$\forall\, t > 0,\ \forall\, \mu \in \mathcal{M}_{\ell,\xi}, \quad \|\, \mathbb{P}_\mu\, [\, X_t \in dx \mid t < \tau_\partial] - \alpha(dx)\, \|_{TV} \leq C\, e^{-\gamma\, t}. \tag{1.2.2}$$

It is classical (cf e.g. Theorem 2.2 in [CMS13]) that, as a QSD, α is associated to some extinction rate λ :

$$\forall\, t \geq 0, \quad \mathbb{P}_\alpha(t < \tau_\partial) = e^{-\lambda\, t}, \text{ so that } \alpha P_t = e^{-\lambda\, t}\, \alpha \tag{1.2.3}$$

$$\text{Let :} \quad h_t(x) \ := e^{\lambda\, t}\mathbb{P}_x(t < \tau_\partial). \tag{1.2.4}$$

Theorem 1.2.2. *Again under assumption* $(\mathbf{A})$, *we have exponential convergence in the supremum norm of* $(h_t)_{t\geq 0}$ *to a limit* h, *with the rate* γ *deduced from* (1.2.2). *The function* h, *which describes the "survival capacity" of the initial condition* μ, *has a positive lower-bound on any* $\mathcal{D}_\ell$, *an upper-bound on* $\mathcal{X}$ *and vanishes on* ∂. *It also belongs to the domain of the infinitesimal generator* $\mathcal{L}$, *associated with the semi-group* $(P_t)_{t\geq 0}$ *on* $(B(\mathcal{X} \cup \{\partial\}); \|.\|_\infty)$, *and :*

$$\mathcal{L}\, h = -\lambda\, h, \qquad \textit{so} \quad \forall\, t \geq 0,\ P_t\, h = e^{-\lambda\, t} h. \tag{1.2.5}$$

Remark : Like in [CV16], it is also not difficult to show that there is no eigenvalue of $\mathcal{L}$ between 0 and $-\lambda$, and that h is the unique eigenvector associated to $-\lambda$.

Theorem 1.2.3. *Under again assumption* $(\mathbf{A})$, *we have :*
(i) **Existence of the Q-process :**
There exists a family $(\mathbb{Q}_x)_{x\in\mathcal{X}}$ *of probability measures on* Ω *defined by :*

$$\lim_{t\to\infty} \mathbb{P}_x(\Lambda_s \,\big|\, t < \tau_\partial) = \mathbb{Q}_x(\Lambda_s), \tag{1.2.6}$$

for all $\mathcal{F}_s$*-measurable set* Λ_s. *The process* $(\Omega; (\mathcal{F}_t)_{t\geq 0}; (X_t)_{t\geq 0}; (\mathbb{Q}_x)_{x\in\mathcal{X}})$ *is an* $\mathcal{X}$*-valued homogeneous strong Markov process.*

(ii) Transition kernel :
The transition kernel of the Markov process X *under* $(Q_x)_{x\in\mathcal{X}}$ *is given by :*

$$q(x;t;dy) = e^{\lambda\, t}\, \frac{h(y)}{h(x)}\, p(x;t;dy), \tag{1.2.7}$$

where $p(x;t;dy)$ *is the transition kernel of the Markov process* X *under* $(P_x)_{x\in\mathcal{X}}$. *In other words, for all* $\psi \in \mathcal{B}_b(\mathcal{X})$ *and* $t \geq 0$, $\quad \langle \delta_x\, Q_t \,\big|\, \psi\rangle = e^{\lambda\, t}\, \langle \delta_x\, P_t \,\big|\, h \times \psi\rangle \,/\, h(x)$, *where $(Q_t)_{t\geq 0}$ is the semi-group of X under $\mathbb{Q}$.*

(iii) Exponential ergodicity :
There is a unique invariant distribution of X under $\mathbb{Q}$, defined by :

$$\beta(dx) := h(x)\,\alpha(dx).$$

Moreover, there exists $\gamma > 0$ and $C = C(\ell, \xi)$ such that :

$$\forall\, t > 0,\ \forall\, \mu \in \mathcal{M}_{\ell,\xi}, \quad \|\mathbb{Q}_{\mu B[h]}(X_t \in dx) - \beta(dx)\|_{\frac{1}{h}} \leq C\, e^{-\gamma\, t}. \tag{1.2.8}$$

$$\text{where } \|\mu\|_{\frac{1}{h}} := \left\|\frac{\mu(dx)}{h(x)}\right\|_{TV} \geq \frac{\|\mu(dx)\|_{TV}}{\|h(x)\|_\infty}, \quad \mu B[h](dx) := h(x)\,\mu(dx)\,/\langle \mu \,\big|\, h\rangle,$$
$$\mathbb{Q}_\mu(dw) \ := \textstyle\int_{\mathcal{X}} \mu(dx)\,\mathbb{Q}_x(dw).$$

1.2.3 How to verify $(A3)$?

For discrete space, it is quite natural to deduce $(A3)$ from the fact that there exists t such that : $\inf_{x\in E} \mathbb{P}_\zeta(X_t = x) > 0$. We can thus couple some trajectories starting from ζ and passing in x at time t to the set of all trajectories starting from x. From this we can infer a lower-bound of the asymptotic survival ability of the former (starting from ζ) compared to the latter (starting from x). For an illustration, this coupling is exploited in the birth and death process in Subsection 1.4.1.

For continuous space however, the process starting from ζ will never hit precisely x. We need to wait a bit for the process starting from x to diffuse before the association we expect can be ensured. Although it appears quite more complicated, our argument is very similar. In cases where the Harnack inequality holds (notably pure diffusive processes, cf Subsection 1.4.2.2), one is usually able to prove :

$$\forall\, x \in E, \qquad \mathbb{P}_x\Big(X_t \in dx\ ;\ t < \tau_\partial\Big) \leq c\,\mathbb{P}_\zeta\Big(X_{t_\alpha} \in dx\ ;\ t_\alpha < \tau_\partial\Big),$$

where $t, t_\alpha, c > 0$ are independent of x. Like for the discrete space case, we then deduce $(A3)$ from the Markov property and an additional control of the survival on finite time-interval.

In a much general setting, and especially when jumps are involved in the process, the situations might get much tricky when one wishes to look for a similar coupling of trajectories. The issue is notably on exceptional behavior along which we have poor controls (no jump for a long time, too many jumps, too large etc.). We propose a criterion to handle these issues in Chapter 2. Several illustrations are presented there, and one can also look at Chapter 3 for a more developed application of our criteria (the mixing is due to a diffusive behavior on one component, and due to the jumps on the other).

1.2.4 Comparison with the literature

To our knowledge, dependencies on the initial condition in results like (1.2.2) can only be found in the very new results of [CV17c], [FRS19] and [BCGM19]. It is however quite natural for the models we have in mind, where extinction plays a stabilizing role, preventing transient dynamics. In the perspective of natural selection, we expect to observe the prevalence of trajectories leading to and gravitating around some basin of attraction, notably compared to those dragged away in deadlier regions. Although the burden of mal-adaptation may seem light in the short run, if it is too hard for the process to escape from less adapted areas, one can presume that the process cannot have been there for long. In particular, the trajectories starting from favorable initial conditions may outcompete what remains of the distribution, so that it becomes the leading part in the convergence to the QSD.

Upper-bounds of the form :

$$\| \mathbb{P}_\mu [X_t \in dx \mid t < \tau_\partial] - \alpha(dx) \|_{TV} \le C(\mu)\, e^{-\gamma\, t}. \tag{1.2.9}$$

assume generally $C(\mu) = \langle \mu \,\big|\, W \rangle$ in the case of stationary distribution (for processes without extinction). The use of such Lyapunov functions W has been thoroughly studied in [MT93] in the case of Markov Chains, or in e.g. [DMT95], [CGZ10], [CG16] for continuous time processes. The condition on W may relate to different probabilistic bound on the first entry time τ_E ; yet, including extinction, exponential moments appear compulsory (all the more since λ, the limiting rate of extinction, is not precisely known). When considering extinction, we lose also the property of linearity over the initial condition. This explains why upper-bounds like $\langle \mu \,\big|\, W \rangle$ are not so general and why we focus on general initial distributions and not only Dirac Masses.

For the case where extinction occurs, other expressions of $C(\mu)$ have been presented in two preprints ([CV17c] and [FRS19]) while we were working on this article. In [CV17c], (1.2.9) is obtained with a non-linear dependency of the form $C(\mu) = C\, \langle \mu \,\big|\, \psi_1 \rangle \,/\, \langle \mu \,\big|\, \psi_2 \rangle$. As can be seen in next Chapter 2, the dependency we introduce implies (1.2.9) with $C(\mu) = C / \langle \mu \,\big|\, h \rangle$, for some $C > 0$. So the former extends our result by including the more classical dependency through a Lyapunov function. In [FRS19] and [BCGM19], the convergence is stated in a weighted norm involving a weight function W (resp.V) related to the previous ψ_1. The dependency $C(\mu)$ stated in their analog of (1.2.9) is implicitly related to both $\langle \mu \,\big|\, h \rangle$ and $\langle \mu \,\big|\, W \rangle$ (their function h plays the same role as our). Note however that our condition on the survival ($\rho > \rho_S$) seems easier to interpret than its analog in [CV17c], [FRS19] and [BCGM19]. A dependency on ψ_1 (or on W) is neglected in our article : $(A2)$ ensures in a way that we can find some upper-bounded ψ_1 (we refer e.g. to Lemma 3.6 in [CV17c]). It would probably not be too hard to introduce, as long as one has control on the increase of such $W(X_t)$ when the process jumps at this time t.

1.2.5 Remarks on the Assumptions and the results

Remarks 1.2.5.1. *Since X is right-continuous and the filtration is both right-continuous and complete, the first entry time of any Borel set is a stopping time, cf. Theorem 52 in* [DM11]*, or more recently Theorem 2.4 in* [Ba10]*. It means in particular that the first exit time $T_{\mathcal{D}_\ell}$ and the first entry time $\tau_{\mathcal{D}_\ell}$ are stopping times (for any $\ell \geq 1$ and any initial condition). The result extends in fact to any iterated combination of the kind "next entry time of $\mathcal{D}_\ell$ after the first exit time of $\mathcal{D}_L$ following the first entry time of $\mathcal{D}_\ell$". For this, we shall use that there is a positive gap between each of the three random times (say $\tau_0 < T_1 < \tau_1$) involved, and that for any t, : $(s, \omega) \mapsto \mathbf{1}_{\{\tau_0(\omega) < s \leq t\}}$ has left continuous paths (and similarly with T_1 instead of τ_0 and possibly so on by induction).*

Remarks 1.2.5.2. *This property on first entry times is the main reason for us to assume $\mathcal{X}$ Polish. The space topology is not much exploited. As we can see in the illustrations of Chapter 2 (exploiting this result), it is not required for the process to be strong Feller : for jump processes, there may exist bounded measurable function f such that $P_t f$ is discontinuous. h itself might not be continuous, notably for discontinuous jump rate.*

Remarks 1.2.5.3. $(A1)$ *imposes a weak form of irreducibility condition, with this reference measure ζ, and a coherence in time to prevent periodicity.*

It may happen that there exists absorbing domains $\mathcal{D}_A$. (whose escape can only happen at τ_∂). Any MCNE with initial condition $x \in \mathcal{D}_A$ is necessarily supported in $\mathcal{D}_A$. Any ζ that satisfies $(A1)$ is thus also supported in $\mathcal{D}_A$. Moreover, if these MCNE converge to a unique QSD as in our result, this QSD is necessarily supported on $\mathcal{D}_A$ as well.

Remarks 1.2.5.4. *Assumption $(A1)$ is a stronger version of Doeblin's condition that appears for the convergence of Markov Chains without extinction. It also implies that any border of extinction shall be approached by the sequence $\mathcal{D}_\ell$ while $\ell \to \infty$, but never from inside any $\mathcal{D}_\ell$, since by Lemma 1.3.0.3 :*

$$\forall\, \ell \geq 1,\ \forall\, t > 0, \quad \inf\{\mathbb{P}_x(t < \tau_\partial)\ ;\ x \in \mathcal{D}_\ell\}\ > 0$$

Remarks 1.2.5.5. *For pure jump processes, one can generally choose $\mathcal{D}_L := \mathcal{D}_\ell$. For other processes, one often needs "a bit of space" between $\mathcal{D}_\ell$ and $\mathcal{D}_L^c$ to obtain a lower bound uniform in $x \in \mathcal{D}_\ell$ over trajectories from x to ζ staying inside $\mathcal{D}_L$.*

Remarks 1.2.5.6. *The definition of ρ_S is made to ensure that for any $\rho > \rho_S$, there exists $t_S, c > 0$ and $L \geq 1$ such that for any $t \geq t_S$:*

$$\mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) \geq c\, \exp[-\rho t].$$

In this expression, the left hand side is uniformly increasing with L. Almost sure extinction is not at all needed for our proof, which is why we defined ρ_S such that it equals 0 if for some L, there is a uniform lower-bound on $\mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L})$. It includes the case where there is no extinction, or only in some "transitory domain".

Remarks 1.2.5.7. *To understand (1.2.8), it is worth noticing that, considering some general initial condition in the left-hand side of (1.2.6), we obtain for the Q-process a biased initial condition :*

$$\forall \mu \in \mathcal{M}_1(\mathcal{X}), \qquad \lim_{t\to\infty}\mathbb{P}_\mu(\Lambda_s \,\big|\, t < \tau_\partial) = \mathbb{Q}_{\mu B[h]}(\Lambda_s). \tag{1.2.10}$$

To deduce (1.2.8) from (1.2.2), we reformulate (1.2.7) in terms of $B[h]$, P_t, A_t and Q_t :

$$\forall t \geq 0,\ \forall \mu \in \mathcal{M}_1(\mathcal{X}), \qquad (\mu B[h]) Q_t = (\mu P_t)\, B[h] = (\mu A_t)\, B[h]. \tag{1.2.11}$$

Originally, we intended to adapt the proof of Theorem 1.2.1 on the marginal at time t conditioned on survival at time $t+T$ to deduce a control uniform in T. This approach is effective but leads to a weaker result where $\|.\|_{\frac{1}{h}}$ is replaced by $\|.\|_{TV}$. The convergence of the MCNE is here more informative, because h is bounded.

1.3 Several implications of $(A1)$

Lemma 1.3.0.1. *Assume that $(A1)$ holds for two probability measures ζ^1 and ζ^2. Then, the associated values for ρ_S coincide. Moreover, the sets E for which assumption $(A3)$ holds are the same for both measures.*

"From mixing to regeneration, then survival" : $(A1)$ trivially implies, for any ℓ such that $\zeta(\mathcal{D}_\ell) > 0$ the following **regeneration estimate** :

There exists $t, c > 0$, $L \geq \ell$ such that, with $\mathcal{D}_S := \mathcal{D}_\ell \subset \mathcal{D}_L$:

$$\forall x \in \mathcal{D}_S, \qquad \mathbb{P}_x(X_t \in \mathcal{D}_S\ ;\ t < \tau_\partial \wedge T_{\mathcal{D}_L}) \geq c. \tag{1.3.1}$$

Lemma 1.3.0.2. *Assume that (1.3.1) holds (for t, $c > 0$, $\mathcal{D}_S \subset \mathcal{D}_L$ and $\zeta(\mathcal{D}_S) > 0$). Then, $\rho_S \leq -\frac{1}{t}\ln(c)$.*

In particular, we deduce that $(A1)$ implies $\rho_S < \infty$.

Lemma 1.3.0.3. *$(A1)$ is equivalent to the apparently stronger version (with the same ζ) :*
For any $\ell \geq 1$, and $t_\vee > 0$, there exists $L \geq \ell$, $t \geq t_\vee$ and $c > 0$ such that :

$$\forall x \in \mathcal{D}_\ell, \qquad \mathbb{P}_x\left[X_t \in dx\ ;\ t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\,\zeta(dx). \tag{$\overline{A1}$}$$

1.3.1 Proof of Lemma 1.3.0.1

Assume that ρ_S^1 is associated to a first choice of ζ^1 satisfying $(A1)$, and consider another choice ζ^2. By $(A0)$, there exists $\ell \geq 1$ such that $\zeta^2(\mathcal{D}_\ell) \geq 1/2$. By $(A1)$ applied to ζ^1, for some $c_J, t_J > 0$ and $L > \ell$:

$$\mathbb{P}_{\zeta^2}(X_{t_J} \in dx \ ; \ t_J < \tau_\partial \wedge T_{\mathcal{D}_L}) \geq c_J \zeta^1(dx). \tag{1.3.2}$$

By definition of ρ_S^1, for any $\rho > \rho_S^1$, there exists $c_S, t_S > 0$ and $L' \geq L$ such that :

$$\forall\, t \geq t_S, \ \mathbb{P}_{\zeta^1}(t < \tau_\partial \wedge T_{\mathcal{D}_{L'}}) \geq c_S \exp[-\rho t] \tag{1.3.3}$$

By combining (1.3.2), (1.3.3) and the Markov property, we deduce :

$$\limsup_{t>0} \frac{\exp[-\rho(t+t_J)]}{\mathbb{P}_{\zeta^2}(t+t_J < \tau_\partial \wedge T_{\mathcal{D}'_L})} \leq \Big(c_J\, c_S\, \exp[\rho t_J]\Big)^{-1} < \infty.$$

By optimizing in ρ, we deduce $\rho_S^2 \leq \rho_S^1$ and the equality by symmetry.

Concerning assumption $(A3)$, (1.3.2) and the Markov property imply that for any $t \geq 0$ and $x \in \mathcal{X}$:

$$\begin{aligned} &\mathbb{P}_{\zeta^2}(t+t_J < \tau_\partial) \geq c_J \mathbb{P}_{\zeta^1}(t < \tau_\partial) \\ \text{Thus} \qquad &\frac{\mathbb{P}_x(t+t_J < \tau_\partial)}{\mathbb{P}_{\zeta^2}(t+t_J < \tau_\partial)} \leq (c_J)^{-1} \frac{\mathbb{P}_x(t < \tau_\partial)}{\mathbb{P}_{\zeta^1}(t < \tau_\partial)}. \end{aligned}$$

If assumption $(A3)$ holds for ζ^1 and E, it thus holds also for ζ^2 and the same E. $\square$

1.3.2 Proof of Lemma 1.3.0.2

Assume (1.3.1). Let $x \in \mathcal{D}_S$, $\rho \ := -\frac{1}{t_{RG}} \ln(c_{RG})$, $T_L := \inf\{t \geq 0,\ X_t \notin \mathcal{D}_L\}$. By induction over $k \in \mathbb{N}$ and the Markov property :
$\forall\, k \geq 1, \qquad \inf_{x \in \mathcal{D}_S} \mathbb{P}_x(k\, t_{RG} < T_L) \geq \exp(-\rho\, k\, t_{RG})$.
Thus, for a general value of $t > 0$:

$$\begin{aligned} \inf_{x\in\mathcal{D}_S} \mathbb{P}_x(t < T_L) &\geq \inf_{x\in\mathcal{D}_S} \mathbb{P}_x\left(\left\lceil \tfrac{t}{t_{RG}} \right\rceil t_{RG} < T_L\right) \geq \exp\left(-\rho \left\lceil \frac{t}{t_{RG}} \right\rceil t_{RG}\right) \\ &\geq \exp(-\rho\,(t + t_{RG})) = c_S\, e^{-\rho\, t} \qquad \text{with } c_S \ := \exp(-\rho\, t_{RG}) = c_{RG}. \qquad \square \end{aligned}$$

1.3.3 Proof of Lemma 1.3.0.3 :

Let $\ell \geq L_S$ for which we apply $(A1)$. By induction with the Markov property, it is quite straightforward to extend the property $(A1)$ on $\mathcal{D}_\ell$ with the same L_M,

$t^{(k)} := k \times t$, $c^{(k)} := c \times (c\,\zeta(\mathcal{D}_S))^{k-1}$. Then, for any $t_\veebar > 0$, we only need to apply this extension for some $k \geq 1$ such that $t^{(k)} \geq t_\veebar$. On the other hand, $(\overline{A1})$ clearly implies $(A1)$ (take $t_\veebar = 0$), so that we have indeed proved $(A1) \Leftrightarrow (\overline{A1})$. $\square$

1.4 Two models to which our results apply

1.4.1 Birth-and-death process with catastrophes

We choose to illustrate our result with this example for its clear simplicity. In this birth and death process, the population can get extinct punctually at any time during what we call a catastrophe, and which happens at a rate depending on number of alive individuals. Otherwise, the process gets extinct when there is only a unique individual that ends up dying. To ensure uniqueness, we will impose that the catastrophe rate is large enough when the population size is large. Biologically, we could imagine that the population is under the threat of some voracious predators, but can stay hidden as long as the population size is not too large.

In fact, one has now quite a complete description of birth-and-death processes. It is proved in [MSV14] that there exists a unique QSD for one dimensional birth and death processes if and only if (1.2.9) holds with a uniform constant $C(\mu) = C > 0$. This equivalence is probably due to the fact that in these models, extinction can only occur once the process is inside some given compact set (i.e. once it has descended from infinity), as suggested in Theorem 19 in [DP13]. Like in [CV16] and as we will do, the authors of [DP13] include direct extinction from any state of the birth-and-death process (what is called a "catastrophe"). Theorem 19 in [DP13] states that the behavior of the process is the same if catastrophe only happens in a compact set. In Theorem 4.1 of [CV16], the authors prove that, for a bounded catastrophe rate, there is descent from infinity (see notably [BCMM19]) iff (1.2.9) holds with a uniform constant $C(\mu) = C$. This does not exclude however that (1.2.9) could hold without descent from infinity, which we prove with our technique.

1.4.1.1 Description of the process

X, the population size, is a time-homogeneous Markov Chain on $\mathbb{Z}_+$ where $\partial = 0$ is the absorbing state and $\mathcal{X} = \mathbb{N}$. Given $X_0 = n \geq 1$, there is a death with rate $d_n > 0$ (leading to $X = n - 1$), a birth with rate $b_n > 0$ (leading to $X = n + 1$) and a catastrophe with rate $c_n \geq 0$ (leading to $X = 0$). Since c_1 and d_1 play the same role (the transition is from $X = 1$ to $X = \partial$), we assume w.l.o.g. $c_1 = 0$. Actually, $d_1 > 0$ is not required in the following statements.

Theorem 1.4.1. *Assume that : for some $n \geq 1$ (thus for all n) $\mathbb{P}_n(\tau_\partial < \infty) = 1$*

$$\text{and} \qquad \liminf_{n\to\infty} c_n > \inf_{k\geq 1}(b_k + d_k + c_k). \tag{1.4.1}$$

Then, the conclusions of Theorem 1.2.1, 1.2.2 and 1.2.3 hold.

At least for some of the models, the speed of convergence towards the QSD cannot be uniformly bounded over all initial conditions, since :

Proposition 1.4.1.1. *We can define some postive values for $(b_n, d_n, c_n)_{n\geq 1}$ such that (1.4.1) holds and for which, whatever large the time $t > 0$, and whatever small the similarity threshold $\epsilon \in (0, 1)$, we can still find some initial condition $x \in \mathcal{X}$ such that :*

$$\left\|\mathbb{P}_x\left(X_t \in dy \,\middle|\, t < \tau_\partial\right) - \mathbb{P}_1\left(X_t \in dy \,\middle|\, t < \tau_\partial\right)\right\|_{TV} \geq 1 - \epsilon.$$

The proof of Theorem 1.4.1 and Proposition 1.4.1.1 are achieved resp. in Subsection 1.4.1.2-3.

Remarks 1.4.1.1. *The term $\inf_k(b_k + d_k + c_k)$ comes from a survival estimate of the simplest form : the process reaches some position k (as optimal as we need) on which to stay up to (large) time t. We are a priori very far from a necessary and sufficient condition : it seems hardly possible to infer generically the level of catastrophe rate that affects the process as it evolves at large values. On the other hand, the regeneration estimate could be improved by looking at the QSD restricted to $[\![1, k]\!]$ (by extinction when exiting the interval). Again, explicit values can hardly be obtained except for very specific models.*

Remarks 1.4.1.2. *By the condition $\mathbb{P}_n(\tau_\partial < \infty) = 1$, we mean that the process is non-explosive. With : $T_\infty := \lim_{n\to\infty} \inf\{t \geq 0 \,;\, X_t \geq n\}$, our condition means that "for some $n \geq 1$ (thus for all), $\mathbb{P}_n(T_\infty = \infty) = 1$", which is satisfied as soon as the sequence b_n/n is upper-bounded. We refer to Theorem 5.5.2 in [Mé16] for a more general condition (deduced from the case without catastrophe).*

Remarks 1.4.1.3. *Considering $\bar{\tau_\partial} := \inf\{t \geq 0 \,;\, X_t = 0\} \wedge T_\infty$ as the extinction epoch, our theorem extends to the case $T_\infty < \infty$. It also extends to models where catastrophes do not entirely exterminate the population. Assume for instance that after a catastrophe, from a population of size larger than some $K \geq 1$, only K individuals are to survive. We can keep the extinction for population of size initially lower than K, but it's not very significant here. Then (A2) can easily be adapted with $K \in E = [\![1, \ell_E]\!]$. The proof of the other assumptions remains the same.*

Remarks 1.4.1.4. *The alternative conditions given in [CV17c] seem also very efficient to obtain Theorem 1.4.1. Since ρ_S is finite, this will certainly not be the case for the ones in [FRS19].*

1.4.1.2 Proof of Theorem 1.4.1

By (1.4.1), let $k, \ell_E \geq 1$ and $\rho_E > 0$ be such that :

$$\begin{aligned} 0 < \widetilde{\rho}_S \ &:= b_k + d_k + c_k \ < \ \rho_E \ < \ \inf_{\{n \geq \ell_E+1\}} c_n := \widetilde{\rho}_E \\ \zeta \ &:= \delta_k, \quad \mathcal{D}_S \ := \{k\}\,, \quad E \ := [\![1, \ell_E]\!] \end{aligned} \tag{1.4.2}$$

Let $\mathcal{D}_\ell = [\![1, \ell \vee k]\!]$, for $\ell \geq 1$. in the following, we ensure $(\mathbf{A})$ (where $\zeta(\mathcal{D}_S) > 0$ is obvious). First, $(A0)$ is obvious.

Proof of $(A1)$ **and** $(A3)$ Let $n \geq k$. Consider

$$\partial^n \ := \{0\} \cup [\![n+1, \infty[\![, \qquad \tau_\partial^n \ := \inf\{t \geq 0 \ ; \ X_t \in \partial^n\}\,.$$

Then the process $Y_t \ := X_t\, \mathbf{1}_{\{t < \tau_\partial^n\}}$ is a Markov Chain on the finite space $[\![0, n]\!]$, absorbed at $\partial = 0$. Since $\forall\, \ell \geq 1,\ d_\ell > 0, b_\ell > 0$, this Chain (Y_t) is irreducible and it is elementary to prove that :

$$\forall\, t_Y > 0,\ \exists\, c_Y > 0, \quad \forall\, i, j \in [\![1, n]\!], \quad \mathbb{P}_i(Y_{t_Y} = j) \geq c_Y \tag{1.4.3}$$

With $j := k$ and $n := \ell_M \vee k$, (1.4.3) clearly implies $(A1)$ (with parameters $\ell = \ell_M, L = L_M, c = c_M, t = t_M$). We can indeed choose $\zeta := \delta_k$, $L_M := n$, $t_M = 1$ (arbitrary), and c_M the value of c_Y associated to the choice of $t_Y = t_M$.

With $i := k$ and $\ell = \ell_E$, (1.4.3) and the Markov property imply $(A3)$ for any E, because :

$$\begin{aligned} \forall\, t > 0,\ \forall\, j \in [\![1, \ell_E]\!], \quad \mathbb{P}_j(t < \tau_\partial) &\leq (1/c_Y) \times \mathbb{P}_k(X_{t_Y} = j \ ; \ t + t_Y < \tau_\partial) \\ &\leq (1/c_Y) \times \mathbb{P}_k(t < \tau_\partial) \end{aligned}$$

Proof of $(A2)$ By (1.4.2), we can upper-bound $\tau_\partial \wedge \tau_E$ by an exponential variable with rate $\widetilde{\rho}_E$ (thanks to the lower-bound of the extinction rate). Thus :

$$\forall\, t > 0,\ \forall\, n \geq \ell_E + 1, \quad \mathbb{P}_n(t < \tau_E \wedge \tau_\partial) \leq \exp(-\widetilde{\rho}_E\, t) \tag{1.4.4}$$

It is classical -by Fubini Theorem, and the integral expression of the exponential- to relate the exponential moment with the repartition function by :

$$\mathbb{E}_n\left(\exp\left[\rho_E\left(\tau_E \wedge \tau_\partial\right)\right]\right) = 1 + \rho_E \int_0^\infty \exp[\rho_E\, t]\, \mathbb{P}_n(t < \tau_E \wedge \tau_\partial)\, dt \tag{1.4.5}$$

By (1.4.4) and (1.4.5), we conclude :

$$\begin{aligned} \forall\, n \geq \ell_E + 1, \quad \mathbb{E}_n\left(\exp\left[\rho_E\left(\tau_E \wedge \tau_\partial\right)\right]\right) &\leq 1 + \rho_E \int_0^\infty \exp[-(\widetilde{\rho}_E - \rho_E)\, t]\, dt \\ &= 1 + \ \{\rho_E \ / \ (\widetilde{\rho}_E - \rho_E)\} < \infty \end{aligned}$$

Proof that for any $k \geq 1$ $b_k + d_k + c_k \geq \rho_S$ Immediately, by (1.4.2) :

$$\forall\, t \geq 0, \qquad \mathbb{P}_k(X_t = k\ ;\ t < \tau_\partial \wedge T_{\mathcal{D}_k}) \geq \mathbb{P}_k(\forall\, s \leq t,\ X_s = k) = \exp(-\widetilde{\rho}_S\, t)$$

1.4.1.3 Proof of Proposition 1.4.1.1

We consider one of the simplest choice, which is to take b_n, d_n linear in n (the classical Malthus' growth model, without competition) and c_n constant for $n \geq 2$. We can then choose arbitrarily :

$$\begin{aligned} b_1, d_1, \bar{b}, \bar{d} \in (0,\infty)^5, \quad c_2 > (b_1 + d_1), \\ \text{with } c_1 = 0, \qquad \forall\, n \geq 2, \quad b_n := \bar{b}\, n,\ d_n := \bar{d}\, n,\ c_n := c_2. \end{aligned} \tag{1.4.6}$$

(1.4.1) is clearly satisfied. There is no explosion for this model, so that extinction happens a.s. (ref Remark (iv) of Subsection 1.4.1.1).

We shall only need to consider transitions between values of the form $2^n, n \geq 2$. Let :

$$T_n := \inf\left\{t \geq 0\ ;\ X_t \leq 2^{n-1}\, or\, X_t \geq 2^{n+1}\right\}, \tag{1.4.7}$$

$$\tau_n := \inf\left\{t \geq 0\ ;\ X_t \leq 2^n\right\}. \tag{1.4.8}$$

We use the following lemma, whose proof is deferred after the one of Proposition 1.4.1.1 :

Lemma 1.4.1.2. *For some $u > 0$, it holds :* $\lim_{n\to\infty} \mathbb{P}_{2^n}(T_n \leq u) = 0.$

For given $t, \epsilon > 0$, let $K := \lfloor t/u \rfloor + 1$ and $N \geq 1$ (by Lemma 1.4.1.2) such that :

$$\mathbb{P}_1(X_t \leq 2^N \,\big|\, t < \tau_\partial) \geq 1 - \epsilon/2, \tag{1.4.9}$$

$$\forall\, n \geq N, \quad \mathbb{P}_{2^n}(T_n \leq u) \leq \epsilon \times e^{-(c_2 - d_1)t}/(4\,K). \tag{1.4.10}$$

With initial condition $x := 2^{N+K+1}$, in order that X reaches 2^N before time $t \leq K\,u$, he must at least once have got from 2^{N+k} to 2^{N+k-1} during a time-interval less than u, for some $1 \leq k \leq K+1$. With the Markov property, this implies, with (1.4.7), (1.4.8), (1.4.10)and the fact that the extinction rate is always lower-bounded by d_1 :

$$\begin{aligned} &\mathbb{P}_x(\tau_N \leq t) \leq \textstyle\sum_{\{k \leq K+1\}} \mathbb{P}_{2^{N+k}}(T_{N+k} \leq u) \leq \epsilon\, e^{-(c_2-d_1)t}/2. \\ &\mathbb{P}_x(\tau_N \leq t\ ;\ t < \tau_\partial) \leq e^{-d_1\, t}\, \mathbb{P}_x(\tau_N \leq t) \leq \epsilon\, e^{-c_2\, t}/2. \end{aligned}$$

Since the extinction rate is upper-bounded by c_2 : $\mathbb{P}_x(t < \tau_\partial) \geq e^{-c_2\, t}$.

This implies $\mathbb{P}_x(\tau_N \leq t \,\big|\, t < \tau_\partial) \leq \epsilon/2$. Therefore, with also (1.4.9) :

$$\begin{aligned}\|\delta_1 A_t - \delta_x A_t\|_{TV} &\geq \mathbb{P}_1(X_t \leq 2^N \,\big|\, t < \tau_\partial) - \mathbb{P}_x(X_t \leq 2^N \,\big|\, t < \tau_\partial)\\ &\geq 1 - \epsilon/2 - \epsilon/2 \geq 1 - \epsilon. \qquad \square\end{aligned}$$

Proof of Lemma 1.4.1.2 :
With initial condition 2^n, we can decompose X as a semi-martingale, up to time $t \wedge T_n$:

$$\forall\, t > 0, \qquad X_{t\wedge T_n} := 2^n + \textstyle\int_0^{t\wedge T_n} (\bar{b} - \bar{d})\, X_s\, ds + M_{t\wedge T_n}, \tag{1.4.11}$$

where $(M_{t\wedge T_n})_t$ is a martingale with bounded quadratic variation, with (1.4.7) :

$$< M >_{t\wedge T_n} = \textstyle\int_0^{t\wedge T_n} (\bar{b} + \bar{d})\, X_s\, ds \leq (\bar{b} + \bar{d})\, 2^{n+1}\, t. \tag{1.4.12}$$

Let $\quad u := (8\,|\bar{b} - \bar{d}| \vee 1)^{-1} \quad$ so that, by (1.4.7), a.s. :

$$\forall\, t \leq u, \qquad \left| \textstyle\int_0^{t\wedge T_n} (\bar{b} - \bar{d})\, X_s\, ds \right| \leq |\bar{b} - \bar{d}|\, 2^{n+1}\, u \leq 2^{n-2}. \tag{1.4.13}$$

$$\begin{aligned}\mathbb{P}_{2^n}(T_n \leq u) &\leq \mathbb{P}_{2^n}\left(\sup_{\{t\leq u\}} M_{t\wedge T_n} \geq 2^{n-2}\right) && \text{by (1.4.11) and (1.4.13)}\\ &\leq 2^{-(2\,n-4)}\, \mathbb{E}_{2^n}(< M >_{u\wedge T_n}) && \text{by Doob's inequality}\\ &\leq 2^{-(2\,n-4)}\, (\bar{b} + \bar{d})\, 2^{n+1}\, u && \text{by (1.4.12)}\\ &= \frac{4\,(\bar{b} + \bar{d})}{|\bar{b} - \bar{d}| \vee 1}\, 2^{-n} \underset{n\to\infty}{\longrightarrow} 0 && \text{with the definition of } u \qquad \square\end{aligned}$$

1.4.2 Adaptation of a population to its environment : application to a diffusion process

In this illustration, the notion of being in a mal-adapted region is quite intuitive and the criteria for the exponential convergence to a unique QSD rather natural. Again, the general proof for this illustrative example is unclear without our techniques, except maybe with those of [CV17c]. Yet, in this case, it is presumably quite technical to find a proper Lyapunov function (although our argument proves in fact that they exist). In fact, our control is deduced from local bounds ensuring both a rapid escape from the associated local domains together with sufficiently low transition rates between these domains.

1.4.2.1 Presentation of the model

We consider a simple coupled process describing the eco-evolutive dynamics of a population. We model the population size by a logistic Feller diffusion $(N_t)_{t\geq 0}$ where the growth rate $(r(X_t))_{t\geq 0}$ is changing randomly. Namely, the adaptation of the population and the change of the environment are assumed to act on a hidden process (X_t) in $\mathbb{R}^d$, from which the growth rate is deduced. For simplicity, we will assume that X_t evolves as a continuous Markov process driven by some Brownian Motion and a drift (possibly depending on N and X). For very low values of $r(X_t)$, it is expected that the population shall vanish very quickly. It would thus not change much of the result to introduce an absorbing boundary at some threshold of mal-adaptation. Yet, we want our result to be independent of any such truncation of the trait space and say that this large extinction is sufficient in itself to bound the mal-adaptation, while highlighting that the initial condition indeed matters here.

In a general setting, the process can be described as :

$$(S)\begin{cases} dN_t = (r(X_t) - c\ N_t)\ N_t\ dt + \sigma\ \sqrt{N_t}\ dB_t^N \\ dX_t = b(X_t, N_t)\ dt + \theta(X_t, N_t)\ dB_t^X \end{cases}$$

with initial conditions (n, x), B^N and B^X two independent Brownian Motions, $c, \sigma > 0$, and r, b, θ being locally Hölder continuous functions. We also require that θ is locally elliptic, in the following sense : for any compact set K of $\mathbb{R}^{d+1}$, there exists $\bar{\theta} > 0$ such that for any $(n, x) \in K$ and $\xi \in \mathbb{R}^d$: $\sum_{i,j} \theta_{i,j}(n, x)\xi_i\xi_j \geq \bar{\theta}\, |\xi|^2$.

It is also not much more costly to introduce catastrophes, arising at rate $\rho_c(x, n)$, leading to the complete extinction of the population. Partial extinction of the population (with jumps on the population size), are however quite more technical to deal with (because the Harnack inequality is not as obvious). In Chapter 2, where the focus is on processes with jumps, we shall present techniques that makes it much more manageable.

The main issue for this model is to specify the conditions for $(A2)$ to hold, either for any ρ, or with some ρ sufficiently large. We discuss these aspects in Subsection 1.4.2.3 for a precise control and we show in Subsection 1.4.2.4 a way to prove $(A2)$, efficient at least in a strong case where it holds for any ρ, i.e. for the following Theorem 1.4.2. For diffusions like this, $(A1)$ and $(A3)$ may be deduced quite roughly thanks to the Harnack inequalities, as presented in the next subsection. In order to satisfy assumption $(A3)$, there is no additional restriction on the set E, so that the only requirement on E is for $(A2)$.

1.4.2.2 Harnack inequalities for $(A1)$ and $(A3)$

In the following, we say that a process (Y_t) on $\mathcal{Y} \subset \mathbb{R}^d$ with generator $\mathcal{L}$ (including possibly an extinction rate ρ_c) satisfies Assumption (H) if :

Consider any path-connected open relatively compact sets $\mathcal{D}, \mathcal{D}' \subset \mathcal{Y}$, such that $\overline{\mathcal{D}} \subset \mathcal{D}'$, with C^∞ boundaries, and such that for any point $x \in \partial\mathcal{D}'$, there exists a closed ball $C \in \mathbb{R}^d$ such that $C \cap \overline{\mathcal{D}'} = \{x\}$. For any $0 < t_1 < t_2$ and non-negative C^2 constraints : $u_{\partial\mathcal{D}'} : (\{0\} \times \mathcal{D}') \cup ([0, t_2] \times \partial\mathcal{D}') \to [0, \infty)$, there exists a unique solution $u(t, x) \in C^{1,2}((0, t_2) \times \mathcal{D}') \cap C^0([0, t_2] \times \overline{\mathcal{D}'})$ to the problem :

$$\begin{aligned} &\partial_t u(t,x) = \mathcal{L}u(t,x) && \text{on } [0,t_2] \times \mathcal{D}' \,; \\ &u(t,x) = u_{\partial\mathcal{D}'}(x) && \text{on } (\{0\} \times \mathcal{D}') \cup ([0,t_2] \times \partial\mathcal{D}'). \end{aligned}$$

It is non-negative on $int(\mathcal{D}')$ and it satisfies, for some $C = C(\mathcal{L}, t_1, t_2, \mathcal{D}, \mathcal{D}') > 0$ independent of $u_{\partial\mathcal{D}'}$:

$$\inf_{x \in \mathcal{D}} u(t_2, x) \geq C \, \sup_{x \in \mathcal{D}} u(t_1, x).$$

Remarks 1.4.2.1. (i) *The proof would still hold if t_2 happens to depend on $\mathcal{D}$ and $\mathcal{D}'$.* (ii) *Even if the extinction rate is bounded on $\mathcal{D}'$ it might not be regular enough to guarantee the existence or regularity of such a solution u. It may however be possible to deal with such an issue : we would try to prove the Hypothesis (H) for the generator $\mathcal{L} + \rho_e^k$ with a family $(\rho_e^k)_{k\geq 1}$ of regular approximations of the extinction rate, justifying a uniform control of the constant C on this family.*

Theorem 1.4.2. *Assume that Assumption (H) holds, and $\limsup_{\|x\|\to\infty} r(x) = -\infty$. Then, all the results of Subsection 1.2.2 hold, and we have in particular exponential convergence in total variation of the MCNE to the unique QSD.*

In particular for (X_t, N_t), Assumption (H) shall hold for the generator :

$$\begin{aligned} \mathcal{L}f(x,n) := [r(x) - c\,n]\,\partial_n f(x,n) + b(x,n)\,\partial_x f(x,n) \\ + n\,\sigma^2/2 \times \Delta_n f(x,n) + \theta^2(x,n)/2 \times \Delta_x f(x,n). \end{aligned}$$

The proof for the existence and uniqueness of the solution u for such second-order partial operator with Hölder coefficients and elliptic diffusion coefficient can be found for instance in Corollary 2, Section 4, Chapter 3 of [Fr08]. It also ensures that the solution has two continuous x-derivatives and one continuous t-derivative. The fact it is non-negative is then a consequence of the Maximum principle (cf. e.g. Theorem 1, Chapter 2 in [Fr08]). Finally, the comparison comes from the parabolic Harnack inequality, exploiting the regularity of u and the fact it is non-negative. For its proof, we refer to Theorem 1.1 in [KS80]. The Harnack inequality on the open and path-connected set $\mathcal{D}$ is not too difficult to deduce from the local Harnack inequalities these authors provide.

One essentially need to cover $\bar{\mathcal{D}}$ by a finite number of balls included in $\mathcal{D}'$ on which the local inequalities can be applied. For any points $x, y \in \bar{\mathcal{D}}$, we can then construct a path between them for which the number of visited such balls is uniformly bounded.

The interval $[t_1, t_2]$ shall then be split into as many time-intervals and the local Harnack inequalities applied recursively to conclude that Assumption (H) holds.

Assumption (H) shall hold more generally, notably under the Hörmander condition instead of the condition of ellipticity (cf e.g. [PP04]). Lots of articles are dedicated to prove such estimates under various conditions.

Assumption (H) **with** $Y_t = (X_t, N_t)$ **implies** $(A1)$
We define for $(\mathcal{D}_\ell)$ a sequence of strictly increasing compact and path-connected sets with C^∞ boundaries whose union is $\mathcal{Y} := \mathbb{R}^d \times \mathbb{R}_+^*$. For some C^∞ function f with support in $\mathcal{D}_\ell = \mathcal{D}$, we apply Assumption (H) with $u_{\partial\mathcal{D}'}(t, y) := f(x)$ on $\{0\} \times \mathcal{D}_\ell$ and $u_{\partial\mathcal{D}'}(t, y) := 0$ on $\mathbb{R}_+^* \times \mathcal{D}_{\ell+1}$ (with $\mathcal{D}' := \mathcal{D}_{\ell+1}$). The solution u we obtain is identified thanks to Itô formula as : $u(t, y) := \mathbb{E}_y\left(f(Y_t) \;;\; t < \tau_\partial^{\ell+1}\right)$, with an additional extinction when the process exits $\mathcal{D}_{\ell+1}$. Applying Harnack inequalities implies thus that for any $y \in \mathcal{D}_\ell$, and some reference $y_0 \in \mathcal{D}_1$:

$$\mathbb{E}_y\left(f(Y_{t_2}) \;;\; t_2 < \tau_\partial^{\ell+1}\right) \geq c\, \mathbb{E}_{y_0}\left(f(Y_{t_1}) \;;\; t_1 < \tau_\partial^1\right).$$

As soon as $\mathbb{P}_{y_0}\left(Y_t \in \mathcal{D}_1 \;;\; t < \tau_\partial^1\right) > 0$, we can obtain a probability measure ζ, independent of ℓ, such that (since c does not depend on f) :

$$\forall\, y \in \mathcal{D}_\ell, \quad \mathbb{E}_y\left(Y_t \in dy \;;\; t < \tau_\partial^{\ell+1}\right) \geq c\, \zeta(dy).$$

Assumption (H) **with** $Y_t = (X_t, N_t)$ **implies** $(A3)$
The proof of $(A3)$ is a bit similar but much more technical because the reference measure is now in the upper-bound, so that we can no longer neglect trajectories exiting $\mathcal{D}'$. W.l.o.g., we consider E to be of the form $\mathcal{D}_\ell$ for ℓ sufficiently large. Since the support of ζ is included in $\mathcal{D}_1$, we wish to prove that there exists $c > 0$ such that for any $x \in \mathcal{D}_1$ and $y \in E$:

$$\mathbb{P}_y\left(X_t \in dx \;;\; t < \tau_\partial\right) \leq c\, \mathbb{P}_x\left(X_{t_\alpha} \in dx \;;\; t_\alpha < \tau_\partial\right), \tag{1.4.14}$$

where we can choose here $0 < t_\alpha < t$ arbitrary (c depending on this choice). (1.4.14) directly implies $(A3)$ with the functions $f_s(x) = \mathbb{P}_x(s - t_\alpha < \tau_\partial)$, and the Markov property.

In the step 4 of the proof given in Section 4 of [CV17b], N. Champagnat and D. Villemonais used a trick to obtain results such as (1.4.14). Their idea is to apply the parabolic Harnack inequality on some regular and compact domain $\mathcal{R}$ such that $E \subset \mathcal{R} \subset \mathcal{X}$ and $d(E, \partial\mathcal{R}) > 0$ while approximating the function :

$$u(t, x) \;:= \mathbb{E}_x\left(f(X_t) \;;\; t < \tau_\partial\right), \quad \text{with } \; t \geq t_{\mathcal{R}},\; x \in \mathcal{R}$$

defined for some $f \in C^\infty(\mathcal{X})$ and any choice of $0 < t_{\mathcal{R}} < t_\alpha$. Although we can prove (as they do) that u is continuous, it is a priori not regular enough to apply Harnack inequality directly. Thus, we approximate it on the parabolic boundary $[t_{\mathcal{R}}, \infty) \times \partial\mathcal{R} \bigcup \{t_{\mathcal{R}}\} \times \mathcal{R}$ by the family $(U_k)_{k\geq 1}$ of smooth ($\mathcal{C}_+^\infty$ w.l.o.g.) functions. We then deduce approximations of u in $[t_{\mathcal{R}}, \infty) \times \mathcal{R}$ by (smooth) solutions of :

$$\begin{aligned} &\partial_t u_k(t,x) - \mathcal{L} u_k(t,x) = 0, && t \geq t_{\mathcal{R}},\ x \in \mathcal{R}^\circ \\ &u_k(t,x) = U_k(t,x), && t \geq t_{\mathcal{R}},\ x \in \partial\mathcal{R}, \quad \text{or } t = t_{\mathcal{R}},\ x \in \mathcal{R}. \end{aligned}$$

By Assumption (H), the constant involved in the Harnack inequality does not depend on the values on the boundary. Thus, it applies with the same constant for the whole family of approximations u_k. We refer to the proof in [CV17b] to state that the Harnack inequality then extends to the approximated function u, where the regularity of $u \in C^{1,2}$ is required to apply the Itô formula on the process $u(t-s, X_s)$. Thus, (1.4.14) indeed holds (where we could have chosen any $x \in E$). □

1.4.2.3 Definition of the transitory domain

Depending on the process, one may have a rather precise upper-bound on ρ_S through a regeneration estimate of the form (1.3.1). For simplicity, assume here for instance that the demographical dynamics (on N) is quicker than the spatial/evolutionary dynamics (on X). Thus, while X_t stays around x, the demographical dynamics is concentrated around some value $n_x := r(x)/c$. In ecology, this value is usually called the "carrying capacity" (associated to the state x). Then, it seems rather natural to make $\mathcal{D}_S$ surround some portion of the graph $(x, n_x)_{x\in\mathbb{R}^d}$. Also, $\mathcal{D}_S$ is a priori chosen to ensure a large value of the growth rate $r(x)$ (thus large carrying capacity and low extinction rate) and possibly low values of $b(x, n_x)$ and $\sigma(x, n_x)$. Thus, the "natural" extinction is kept low and the process won't escape too quickly this area. The exact definition of L is not expected to change much our upper-bound of ρ_S.

Depending on such upper-bound $\widetilde{\rho}_S$ of ρ_S, one may extend the reasonning behind Theorem 1.4.2 to hold provided that $\limsup_{\|x\|\to\infty} r(x)$ is sufficiently low.

1.4.2.4 Escape from the transitory domain

A purpose of this section is to demonstrate how to prove $(A2)$ when, depending on the position in the transitory domain, there are various reasons for a quick escape. To combine several local estimates, dealing with suprema in the initial condition of exponential moments appear much more convenient than Lyapunov estimates, see Appendix A. Moreover, these exponential moments can be naturally deduced from probabilities of retention (or transfer) in the transitory domain for a finite given time, see Appendix B, C or D.

Decomposition of the transitory domain

In our choice for E, with three parameters $n_0 < n_\infty < n_E$ to be fixed, its complementary $\mathcal{T} = \mathcal{X} \setminus E$ is made up of 3 subdomains : "$y = \infty$", "$y = 0$", and "$\|x\| = \infty$", according to figure 1.1. They are formally defined as follow :

— $\mathcal{T}_\infty^N := \Big\{\mathbb{R}^d \setminus B(0\,,\,n_E)\Big\} \times (n_\infty, \infty) \cup B(0\,,\,n_E) \times (n_E, \infty)$ ("$y = \infty$")

— $\mathcal{T}_0 := B(0\,,\,n_E) \times [0, n_0]$ ("$y = 0$")

— $\mathcal{T}_\infty^X := \Big\{\mathbb{R}^d \setminus B(0\,,\,n_E)\Big\} \times (n_0, n_E]$ ("$\|x\| = \infty$")

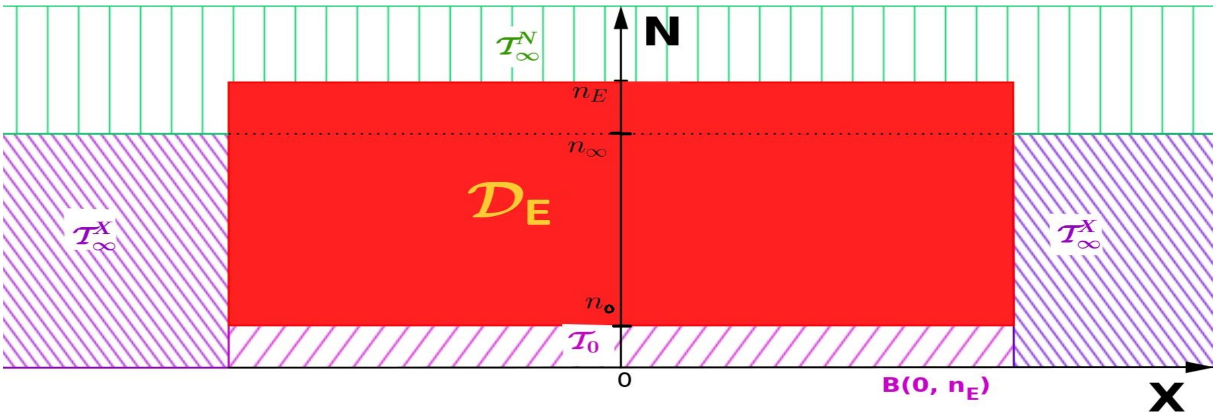


FIGURE 1.1 – subdomains for $(A2)$

Essentially, we will need to choose n_∞ sufficiently large to have the property of descent from infinity for $\mathcal{T}_\infty^N$; $n_E > n_\infty$ sufficiently large to have a growth rate so low that the population cannot maintain itself in $\mathcal{T}_\infty^X$; n_0 sufficiently small to prove that the population can hardly survive afterwards. Thus, the process will escape each region with an exponential moment. Yet, we also need to prove that the process will not circulate between the different transitory areas. Therefore, we will set these areas such that at least some of the transitions (those associated with an increase of the population size) happens with so low probability that Theorem 1.1 holds true.

For each of these domains, we define the following exponential moments that we shall relate by specific inequalities. Let $t_h > 0$ (a threshold needed to ensure the boundedness) and $\widehat{\tau}_E := \tau_E \wedge \tau_\partial \wedge t_h$ (remember that τ_E is the hitting time of E) :

— $\mathcal{E}_\infty^N := \sup\{\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)]\ ;\ (x,n) \in \mathcal{T}_\infty^N\}$,

— $\mathcal{E}_\infty^X := \sup\{\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)]\ ;\ (x,n) \in \mathcal{T}_\infty^X\}$,

— $\mathcal{E}_0 := \sup\{\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)]\ ;\ (x,n) \in \mathcal{T}_0\}$.

Implicitly, we assume ρ to be given. Then, $\mathcal{E}_\infty^N$, $\mathcal{E}_\infty^X$ and $\mathcal{E}_0$ can be seen as functions of n_0, n_∞ and n_E that need to be specified (while the dependency in t_h shall be negligible as $t_h \to \infty$).

A set of inequalities associating the local bounds

The local exponential moments that we introduce are related thanks to the three following propositions, obtained from local bounds mentioned in the following three lemmas. We refer to Appendices A, B, C and D to see first how to deduce $(A2)$ from the three propositions that follow, and then respectively for the (technical) proofs of the propositions (including the lemmas) :

Proposition 1.4.2.1. *Given any $\rho > 0$, we can define $n_\infty > 0$ and $C \geq 1$ such that, whatever $n_E > n_\infty$ and $t_h > 0$:* $\quad \mathcal{E}_\infty^N \leq C\, \left(1 + \mathcal{E}_\infty^X\right)$.

Proposition 1.4.2.2. *Given any ρ, ϵ, $n_\infty > 0$, we have, for some $C \geq 1$ (in fact independent of any parameters), and any n_E sufficiently large and $t_h > 0$:* $\mathcal{E}_\infty^X \leq C\, (1 + \mathcal{E}_0) + \epsilon\, \mathcal{E}_\infty^N$.

Proposition 1.4.2.3. *Given any ρ, ϵ, $n_\infty > 0$, we have, for some $C \geq 1$ and any n_E sufficiently large and $t_h > 0$:* $\quad \mathcal{E}_0 \leq C + \epsilon\, \left(\mathcal{E}_\infty^N + \mathcal{E}_\infty^X\right)$.

The associated elementary bounds on finite time

The main ingredient for these propositions are simple comparison properties that are specific to each of the transitory domain. By focusing on each of the domains separately (with the transitions between them), we can highly simplify our control on the dependency of the processes. Specific autonomous one-dimensional processes indeed acts as upper-bound for each of the domains. The values of (X_t) do not affect these auxilliary processes, but only the regions on which they act as upper-bounds.

Propositions 1.4.2.1 and 1.4.2.2 are deduced from the estimates given in Lemmas 1.4.2.4 and 1.4.2.5 on autonomous processes of the form :

$$N_t^D := n + \int_0^t (r - c\ N_s^D)\ N_s^D\ ds + \int_0^t \sigma\ \sqrt{N_s^D}\ dB_s. \tag{1.4.15}$$

Propositions 1.4.2.1 relies on the property of descent from infinity valid for any value of r :

Lemma 1.4.2.4. *Let N^D be the solution of (1.4.15), for some $r \in \mathbb{R}$ and $c > 0$, with n the initial condition. Then, for any $t, \epsilon > 0$ there exists $n_\infty > 0$ such that :*

$$\sup_{n>0} \mathbb{P}_n(t < \tau_\downarrow^D) \leq \epsilon \qquad \textit{with } \tau_\downarrow^D := \inf\left\{s \geq 0,\ N_s^D \leq n_\infty\right\}.$$

Proposition 1.4.2.2 relies on the strong negativity on the drift term :

Lemma 1.4.2.5. *Considering any $c, t > 0$, with $\tau_\partial^D := \inf\left\{t \geq 0,\ N_t^D = 0\right\}$:*

$$\sup_{n>0}\ \mathbb{P}_n\left(t < \tau_\partial^D\right) \underset{r \to -\infty}{\longrightarrow} 0.$$

Moreover, for any $n, \epsilon > 0$, there exists n_c such that, for any r sufficiently low, with $T_\infty^D := \inf\left\{t \geq 0,\, N_t^D \geq n_c\right\}$: $\quad \mathbb{P}_n\left(T_\infty^D \leq t\right) + \mathbb{P}_n\left(N_t^D \geq n\right) \leq \epsilon.$

Finally, Proposition 1.4.2.3 relies on an upper-bound given as a Continuous State Branching Process, for which the extinction rate is much more explicit. It is clearly as strong as needed for sufficiently small initial condition.

We recall that the complete proofs (of $(A2)$ from the propositions and of the propositions themselves) are deferred to respectively Appendices A, B, C and D. With this and the results of Subsection 1.4.2.2, we have concluded the proof of Theorem 1.4.2.

1.5 Proof of Theorems 1.2.1-3

In Subsection 1.5.3, we present the general principles of our coupling that concludes the proof of Theorem 1.2.1. These principles would alone end the proof in the context of the Assumption (A) in [CV16]). Yet, with our more general assumptions, these principles require the results of the two previous subsections. First, we prove in Subsection 1.5.1 that the MCNE will keep in the long run some mass on some specific set $\mathcal{D}_\circ$ (which is weaker but related in some sense to the tension of the laws) ; then we prove in Subsection 1.5.2 that $(A3)$ holds in fact for $\mathcal{X}$ instead of just E, cf Remark 3. At the end of Subsection 1.5.3, the proof of Theorem 1.2.1 is then complete. The following Subsection 1.5.4 and 1.5.5 then prove respectively Theorem 1.2.2 and 1.2.3.

1.5.1 Stabilization of the process in the long run

The main purpose of this section is to prove :

Theorem 1.5.1. *Assume that* $(\mathbf{A})$ *holds. Then, there exists* $\mathcal{M}_\circ = \mathcal{M}_{\ell_\circ, \xi_\circ}$ *(with* $\ell_\circ \geq 1,\ \xi_\circ > 0$*) such that for any* $\ell \geq 1$ *and* $\xi \in (0,1)$*, there exists* $t_\circ = t_\circ(\ell, \xi) > 0$ *such that :*

$$\forall\, \mu \in \mathcal{M}_{\ell,\xi},\ \forall\, t \geq t_\circ, \qquad \mu A_t \in \mathcal{M}_\circ. \tag{1.5.1}$$

Remarks 1.5.1.1. *Assumption* $(A3)$ *is not involved in the proof of Theorem 1.5.1. This will be important in Chapter 2 since we will exploit this Theorem to provide an alternative criterion to* $(A3)$*.*

Proof of Theorem 1.5.1 :
According to $(A2)$, let $\ell_E \geq 1$ and $\rho_E > \rho_S$ be such that :

$$\begin{aligned} \text{with } E &:= \mathcal{D}_{\ell_E},\ \tau_E^1 := \inf\left\{t \geq 0\ ;\ X_t \in E\right\}, \\ e_{\mathcal{T}} &:= \sup\nolimits_x \mathbb{E}_x\left(\exp\left[\rho_E\left(\tau_E^1 \wedge \tau_\partial\right)\right]\right) < \infty \end{aligned} \tag{1.5.2}$$

From (1.2.1), i.e. the definition of ρ_S, there exists $\widetilde{\rho}_S \in (\rho_S, \rho_E)$, $c_S > 0$ and $L \geq 1$, such that :

$$\forall\, t \geq 0,\ \mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) \geq c_S\ \exp(-\widetilde{\rho}_S\, t) \tag{1.5.3}$$

We then apply $(A1)$ with $\ell = \ell_E$ to state that there exists $L_C \geq L \vee \ell_E$, $t_C, c_C > 0$ such that, with $\mathcal{D}_C := \mathcal{D}_{L_C}$:

$$\forall\, x \in E, \quad \mathbb{P}_x\Big[X_{t_C} \in dy\,,\ t_C < \tau_\partial \wedge T_{\mathcal{D}_C}\Big] \geq c_C\, \zeta(dy) \tag{1.5.4}$$

W.l.o.g., we can replace $\mathcal{D}_L$ by $\mathcal{D}_C$ in (1.5.3).

We then define by induction over $i \in \mathbb{N}$:

$$\begin{aligned} &T_C^i := \inf\Big\{t \geq \tau_E^i\ ;\ X_t \notin \mathcal{D}_C\Big\}\,,\ T_C^0 := 0,\\ &\tau_E^{i+1} := \inf\Big\{t \geq T_C^i\ ;\ X_t \in E\Big\} \end{aligned}$$

In order to conclude the proof of Theorem 1.5.1, we need the following three Lemmas :

Lemma 1.5.1.1. ***First entry in*** E ***:*** *Assume that (1.5.2), $(A1)$ and (1.5.3) hold. Then, for any ℓ, ξ, there exists $C_E = C_E(\ell,\, \xi) > 0$ such that :*

$$\forall\, t_h > 0,\ \forall\, \mu \in \mathcal{M}_{\ell,\,\xi}, \quad \mathbb{P}_\mu(t_h \leq \tau_E^1 \,\big|\, t_h < \tau_\partial) \leq C_E\, e^{-(\rho_E - \widetilde{\rho}_S)\, t_h}.$$

Lemma 1.5.1.2. ***Containment of the process after*** T_C^i ***:***
Suppose that (1.5.3) and $(A1)$ hold. Then, there exists $\ell_\circ \geq L_C$, and $c_\circ > 0$ such that :

$$\forall\, x \in \mathcal{D}_C,\ \forall\, t > 0, \qquad \mathbb{P}_x\left(t < T_C^1 \wedge T_{\mathcal{D}_{\ell_\circ}} \wedge \tau_\partial\right) \geq c_\circ\, \exp[-\widetilde{\rho}_S\, t].$$

Lemma 1.5.1.3. ***Last exit from*** $\mathcal{D}_C$ ***:***
Suppose that $(A0)$, $(A1)$, (1.5.2), (1.5.3) hold (with $E \subset \mathcal{D}_C$) and $\rho_E > \widetilde{\rho}_S$. Then, there exists $C_L > 0$, such that for any $\mu \in \mathcal{M}_1(\mathcal{X})$ with $t_h > t > 0$:

$$\mathbb{P}_\mu\left(T_C^{I(t_h)} \leq t_h - t\,,\ t_h \leq \tau_E^{I(t_h)+1}\,,\ \tau_E^1 < t_h \,\middle|\, t_h < \tau_\partial\right) \leq C_L\, e^{-(\rho_E - \widetilde{\rho}_S)\, t},$$
$$\textit{with } I(t_h)\ := \max\Big\{i \geq 0\ ;\ T_C^i \leq t_h\Big\}\ (< \infty\ a.s.).$$

The proofs of these Lemmas are deferred, in the order of occurrence, after the proof that they imply Theorem 1.5.1.

1.5.1.1 Proof that Lemmas 1.5.1.1-3 imply Theorem 1.5.1

With Lemma 1.5.1.1 and 1.5.1.3, we obtain an upper-bound (with high probability) on how much time the process may have spent outside $\mathcal{D}_C$. Thus, we can associate most of trajectories ending outside $\mathcal{D}_C$ to others ending inside $\mathcal{D}_C$. From this association, we deduce a lower-bound on the probability to see the process in $\mathcal{D}_C$.

Let us first define $\mathcal{D}_\circ$ according to Lemma 1.5.1.2. In the following, we will define : $\mathcal{M}_\circ := \{\mu \in \mathcal{M}_1(\mathcal{X})\ ;\ \mu(\mathcal{D}_\circ) \geq \xi_\circ\}$ for a well-chosen $\xi_\circ$. Thanks to Lemma 1.5.1.3, we choose some $t > 0$ sufficiently large to ensure : $\forall\, t_h > t,\ \forall\, \mu \in \mathcal{M}_1(\mathcal{X})$,

$$\mathbb{P}_\mu\left(T_C^{I(t_h)} \leq t_h - t\ ,\ t_h \leq \tau_E^{I(t_h)+1}\ ,\ \tau_E^1 < t_h \,\middle|\, t_h < \tau_\partial\right) \leq {}^1\!/_4\ . \tag{1.5.5}$$

Let $\ell \geq 1$, $\xi \in (0,1)$. Thanks to Lemmas 1.5.1.1, we know that for some $t_\circ \geq t > 0$:

$$\forall\, t_h \geq t_\circ,\ \forall\, \mu \in \mathcal{M}_{\ell,\xi}, \quad \mathbb{P}_\mu\left(t_h \leq \tau_E^1 \,\middle|\, t_h < \tau_\partial\right) \leq {}^1\!/_4\ . \tag{1.5.6}$$

Let $\mu \in \mathcal{M}_{\ell,\xi}$. Let us first assume that :

$$\mathbb{P}_\mu\left(\tau_E^{I(t_h)+1} \leq t_h \,\middle|\, t_h < \tau_\partial\right) \geq {}^1\!/_4\ .$$

By definition of $I(t_h)$, on the event $\left\{\tau_E^{I(t_h)+1} \leq t_h\right\} \bigcap \{t_h < \tau_\partial\}$, we know that the process stays in $\mathcal{D}_C$ in the time-interval $[\tau_E^{I(t_h)+1}, t_h]$. In particular :

$$\mu A_{t_h}(\mathcal{D}_\circ) \geq \mu A_{t_h}(\mathcal{D}_C) \geq \mathbb{P}_\mu\left(\tau_E^{I(t_h)+1} \leq t_h \,\middle|\, t_h < \tau_\partial\right) \geq {}^1\!/_4\ . \tag{1.5.7}$$

where we recall that $\ell_\circ \geq \ell_C$ by Lemma 1.5.1.2.

Now that this case has been easily treated, we consider the complementary :

$$\mathbb{P}_\mu\left(\tau_E^{I(t_h)+1} \leq t_h \,\middle|\, t_h < \tau_\partial\right) < {}^1\!/_4\ .$$

Thus, by (1.5.5) and (1.5.6) : $\mathbb{P}_\mu\left(t_h - t < T_C^{I(t_h)}\ ,\ \tau_E^1 \leq t_h \,\middle|\, t_h < \tau_\partial\right) \geq {}^1\!/_4$. By defining the stopping time : $\tau_C := \inf\{s \geq t_h - t\ ;\ X_s \in \mathcal{D}_C\}$, we deduce :

$$\mathbb{P}_\mu\left(t_h - t \leq \tau_C < t_h \,\middle|\, t_h < \tau_\partial\right) \geq {}^1\!/_4\ . \tag{1.5.8}$$

By the Markov property, then Lemma 1.5.1.2 :

$$\begin{aligned}
&\mathbb{P}_\mu \left(X_{t_h} \in \mathcal{D}_\circ \,,\; t_h - t \le \tau_C < t_h \,,\; t_h < \tau_\partial\right) \\
&\quad \ge \mathbb{E}_\mu \left[\mathbb{P}_{X_{\tau_C}} \left(\widetilde{X}_{t_h - \tau_C} \in \mathcal{D}_\circ \,,\; t_h - \tau_C < \widetilde{\tau_\partial}\right) \,;\; t_h - t \le \tau_C < t_h \wedge \tau_\partial\right] \\
&\quad \ge c_\circ \exp[-\widetilde{\rho}_S t]\, \mathbb{P}_\mu \left[t_h - t \le \tau_C < t_h \wedge \tau_\partial\right] \\
&\quad \ge c_\circ \exp[-\widetilde{\rho}_S t]\, \mathbb{P}_\mu \left[t_h - t \le \tau_C < t_h \,\middle|\, t_h < \tau_\partial\right] \times \mathbb{P}_\mu \left[t_h < \tau_\partial\right].
\end{aligned}$$

So (1.5.8) indeed implies $\mu A_{t_h} (\mathcal{D}_\circ) \ge \xi_\circ$ with $\xi_\circ := c_\circ\, e^{-\widetilde{\rho}_S\, t}/4$.

Now, with $\mathcal{M}_\circ := \{\mu \in \mathcal{M}_1(\mathcal{X}) \,;\; \mu(\mathcal{D}_\circ) \ge \xi_\circ\}$ ($\xi_\circ$ given by the previous formula does not depend on ℓ, ξ or μ), we indeed prove (1.5.1) (we recall (1.5.7) for the first case). □

1.5.1.2 Proof of Lemma 1.5.1.1

By (1.5.2) and the Markov inequality :

$$\forall\, \mu,\; \forall\, t_h > 0, \quad \mathbb{P}_\mu \left(t_h \le \tau_E^1 \wedge \tau_\partial\right) \le e_{\mathcal{T}}\, e^{-\rho_E\, t_h}. \tag{1.5.9}$$

Let $\ell \ge 1$, $\xi \in (0,1)$. We apply $(A1)$, (1.5.3) and the Markov property to obtain that there exists $c = \xi\, c_S\, c(\ell) > 0$ such that : $\forall\, \mu \in \mathcal{M}_{\ell,\xi}$, $\forall\, t_h > 0$,

$$\mathbb{P}_\mu \left(t_h < \tau_\partial\right) \ge c\, e^{-\widetilde{\rho}_S\, t_h}. \tag{1.5.10}$$

Thus, by (1.5.9), (1.5.10), with : $C_E := e_{\mathcal{T}}/c > 0$,

$$\forall\, \mu \in \mathcal{M}_{\ell,\xi},\; \forall\, t_h > 0, \quad \mathbb{P}_\mu \left(t_h \le \tau_E^1 \,\middle|\, t_h < \tau_\partial\right) \le C_E \exp[-(\rho_E - \widetilde{\rho}_S)\, t_h]. \qquad \square$$

1.5.1.3 Proof of Lemma 1.5.1.2

Thanks to $(A1)$, applied with $\ell = L_C$, there exists some $\mathcal{D}_\circ$, $t_E, c_E > 0$ such that :

$$\forall\, x \in \mathcal{D}_C, \quad \mathbb{P}_x \left(\tau_E^1 \le t_E \wedge T_{\mathcal{D}_\circ} \wedge \tau_\partial\right) \ge c_E \tag{1.5.11}$$

Recalling (1.5.4), we deduce that conditionally on $\mathcal{F}_{\tau_E^1}$ and on the event $\{\tau_E^1 \le t_E \wedge T_{\mathcal{D}_\circ}\}$:

$$\mathbb{P}_{X_{\tau_E^1}} \left(\widetilde{X}_{t_C} \in dx \,,\; t_C < \widetilde{T_C^1} \wedge \widetilde{T}_{\mathcal{D}_\circ} \wedge \widetilde{\tau_\partial}\right) \ge c_C\, \zeta(dx). \tag{1.5.12}$$

By combining (1.5.11), (1.5.12), (1.5.3) and the Markov property, we deduce that :

$$\mathbb{P}_x\left(t_h < T_C^1 \wedge T_{\mathcal{D}_\circ} \wedge \tau_\partial\right) \geq c_\circ \exp[-\widetilde{\rho}_S\, t_h],$$
$$\text{with } c_\circ := c_E\, c_C\, c_S \exp[\widetilde{\rho}_S\,(t_E + t_C)] > 0. \qquad \square$$

1.5.1.4 Proof of Lemma 1.5.1.3

The idea is to use that it is extremely rare to observe that the process is still alive after experiencing an excursion outside E for a long time (and still be there). Indeed, compared to trajectories that stay inside E (in particular those reaching quickly $\mathcal{D}_S$ and not leaving $\mathcal{D}_L$) the probabilities of the associated events vanish with a larger rate : $\rho_E > \widetilde{\rho}_S$. We would like to initiate the comparison just before $T_C^{I(t_h)}$, where the process exits E for the last time before t_h. Yet, it is not a stopping time, with the jump possibly unpredictable, so that the proof gets somewhat more technical.

Let us first prove that $I(t_h) < \infty$. Since X has cadlag paths, we would have on the event $\{I(t_h) = \infty\}$: $\sup_j T_C^i = \sup_j \tau_E^i = T < t_h$ with $X_{T-} \in E \cap \overline{\mathcal{X} \setminus \mathcal{D}_C}$. Yet, by $(A0)$, this set is empty, so that a.s. $I(t_h) < \infty$. Then, exploiting a discretization of time in time-intervals of length t_L to be fixed later :

$$\begin{aligned}
P \ &:= \mathbb{P}_\mu\left(T_C^{I(t_h)} \leq t_h - t \,,\ t_h \leq \tau_E^{I(t_h)+1} \wedge \tau_\partial \,,\ \tau_E^1 < t_h\right)\\
&= \textstyle\sum_{\{i\geq 1\}} \mathbb{P}_\mu\left(T_C^i \leq t_h - t \,,\ t_h \leq \tau_E^{i+1} \wedge \tau_\partial\right)\\
&\leq \textstyle\sum_{\{i\geq 1\}} \sum_{\{k\geq 0\}} \mathbf{1}_{\{k\, t_L \leq t_h - t\}} \mathbb{P}_\mu\left(T_C^i \in (k\, t_L, (k+1)t_L] \,,\ t_h \leq \tau_E^{i+1} \wedge \tau_\partial\right)\\
&= \textstyle\sum_{\{i\geq 1\}} \sum_{\{k\geq 0\}} \mathbf{1}_{\{k\, t_L \leq t_h - t\}} \mathbb{E}_\mu\Big[\mathbb{P}_{X_{(k+1)t_L}}\left(t_h - (k+1)t_L \leq \widetilde{\tau}_E^1 \wedge \widetilde{\tau_\partial}\right) ;\\
&\qquad\qquad T_C^i \in (k\, t_L, (k+1)t_L] \,,\ (k+1)t_L \leq \tau_E^{i+1} \wedge \tau_\partial\Big],
\end{aligned}$$

where we used the Markov property. Exploiting (1.5.2) :

$$\begin{aligned}
P \leq e_T \textstyle\sum_{\{k\geq 0\}} \mathbf{1}_{\{k\, t_L \leq t_h - t\}} \exp[-\rho_E\,(t_h - (k+1)t_L)]&\\
\times \textstyle\sum_{\{i\geq 1\}} \mathbb{P}_\mu\left[T_C^i \in (k\, t_L, (k+1)t_L] \,,\ (k+1)t_L \leq \tau_E^{i+1} \wedge \tau_\partial\right].&
\end{aligned} \tag{1.5.13}$$

The trick is to observe that, by definitions of $\tau_E^i < T_C^i$, one shall have $X_s \in \mathcal{D}_C$ for any $s \in [\tau_E^i, T_C^i)$, in particular on some vicinity to the left of T_C^i. Defining for $k \geq 0$:

$$\tau_C^k := \inf\left\{s \geq k t_L : X_s \in \mathcal{D}_C\right\},$$

we see that the events $\{T_C^i \in (k\, t_L, (k+1)t_L]\} \cap \left\{(k+1)t_L \leq \tau_E^{i+1} \wedge \tau_\partial\right\}$ are disjoint (for k fixed) and included in the event $\left\{\tau_C^k < (k+1)t_L \wedge \tau_\partial\right\}$. On the other hand, exploiting the Markov property together with Lemma 1.5.1.2 :

$$\mathbb{P}_\mu\left[t_h < \tau_\partial\right] \geq c_\circ \exp[-\widetilde{\rho}_S(t_h - k\, t_L)]\, \mathbb{P}_\mu\left[\tau_C^k < (k+1)t_L \wedge \tau_\partial\right].$$

Coming back to (1.5.13), we deduce :

$$P \leq \frac{e_T\, e^{\rho_E\, t_L}}{c_\circ} \mathbb{P}_\mu\left[t_h < \tau_\partial\right] \times \textstyle\sum_{\{k\geq 0\}} \mathbf{1}_{\{k\, t_L \leq t_h - t\}} \exp[-(\rho_E - \widetilde{\rho}_S) \times (t_h - k t_L)]$$

The sum over k is upper-bounded by :

$$\exp[-(\rho_E - \widetilde{\rho}_S)t] \times \textstyle\sum_{\{\ell\geq 0\}} \exp[-\ell(\rho_E - \widetilde{\rho}_S)t_L] \leq \frac{e^{-(\rho_E - \widetilde{\rho}_S)t}}{1 - e^{-(\rho_E - \widetilde{\rho}_S)t_L}}$$

This concludes the proof of the Lemma, with : $C_L := \dfrac{e_T\, e^{\rho_E\, t_L}}{c_\circ(1 - e^{-(\rho_E - \widetilde{\rho}_S)t_L})}$. The choice of t_L is free, so that we can fix it to optimize this constant. □

1.5.2 Persistence

1.5.2.1 Theorem 1.5.2

For the proof of the following Theorem 1.5.2, we need the following Corollary of Theorem 1.5.1 :

Corollary 1.5.2.1. ***"Stability"*** *:*
Under assumption $(\mathbf{A})$*, there exists* $t_S, c'_S > 0$ *and* $\widetilde{\rho}_S \in (\rho_S, \rho_E)$ *such that :*

$$\forall\, u \geq 0,\ \forall\, t \geq u + t_S, \qquad \mathbb{P}_\zeta(t - u < \tau_\partial) \leq c'_S\, e^{\widetilde{\rho}_S u}\, \mathbb{P}_\zeta\left(t < \tau_\partial\right). \tag{1.5.14}$$

Theorem 1.5.2. *Assume that there exists* $\rho_E > \widetilde{\rho_S}$*,* $\mathcal{D}_S \subset \mathcal{X}$*,* $E \subset \mathcal{X}$ *and* $\zeta \in \mathcal{M}_1(\mathcal{X})$ *such that (A3), (A2), (1.5.3) and (1.5.14) hold. Then, there exists* $t_P, c_P > 0$ *such that :*

$$\forall\, x \in \mathcal{X},\ \forall\, t \geq t_P, \qquad \mathbb{P}_x\left(t < \tau_\partial\right) \leq c_P\, \mathbb{P}_\zeta\left(t < \tau_\partial\right). \tag{1.5.15}$$

1.5.2.2 Proof of Corollary 1.5.2.1 :

By (1.5.1) and (A1), there exists $c > 0$ such that for any v sufficiently large : $\zeta A_v \geq c\,\zeta$. with the Markov property, it implies for any $u \geq 0$:

$$\mathbb{P}_\zeta(v + u < \tau_\partial) \geq c\,\mathbb{P}_\zeta(v < \tau_\partial)\,\mathbb{P}_\zeta(u < \tau_\partial).$$

Exploiting (1.5.3) with $t = u$, we deduce Corollary 1.5.2.1 with $v = t - u$ and $c'_S = (c\, c_S)^{-1}$.

□

1.5.2.3 Proof of Theorem 1.5.2

From $(A3)$, there exists $t_A, c_A > 0$ such that :

$$\forall\, t \geq t_A,\ \forall\, x \in E, \quad \mathbb{P}_x(t < \tau_\partial) \leq c_A\, \mathbb{P}_\zeta(t < \tau_\partial) \tag{1.5.16}$$

This proof is very close to the one in [CV17b] (p13 :"Step 2 : Proof of $(A1)$"), except that, in (1.5.14), $t - u$ shall be larger than some value, and similarly for t in (1.5.16). To compare the notations, our $e_\mathcal{T}$, c_A and c_S refer resp. to their M, C_m and $4/c_1\, D_m\, D_{n_1}$. Thus, we won't detail it much and refer to [CV17b].

Let $\zeta \in \mathcal{M}_1(\mathcal{X})$, $t \geq t_P\ := t_S \vee t_A$ and $x \in \mathcal{X}$.

$$\mathbb{P}_x(t < \tau_\partial) \leq c_A\, \mathbb{E}_x\left[\mathbb{P}_\zeta(t - \tau_E < \widetilde{\tau_\partial})\ ;\ \tau_E < (t - t_P) \wedge \tau_\partial\right] + \mathbb{P}_x(t - t_P \leq \tau_E \wedge \tau_\partial) \tag{1.5.17}$$

thanks to property $(A3)$, since $t - \tau_E \geq t_P \geq t_A$ on $\{\tau_E < (t - t_P) \wedge \tau_\partial\}$.

By $(A2)$ (with the Markov inequality) and Corollary 1.5.2.1, with $u = t_A$ for the first term of (1.5.17) and $u = t - t_S$ for the second : $\forall t \geq 0$, $\forall x \in \mathcal{X}$,

$$\mathbb{P}_x(t < \tau_\partial) \leq \left(c_A + e^{\widetilde{\rho_S}(t_P - t_S)} / \,\mathbb{P}_\zeta\left(t_S < \tau_\partial\right)\right) \times c'_S\, e_\mathcal{T}\ \mathbb{P}_\zeta(t < \tau_\partial) \qquad \square$$

1.5.3 Coupling procedure : proof of Theorem 1.2.1

1.5.3.1 Definition of the uncoupled part

With a given set of parameters t_D, c_D, t_P, $c_P > 0$ (cf following subsection) we define for $t_h > t_P$:

$$J(t_h)\ := \lfloor (t_h - t_P)/t_D \rfloor\,. \tag{1.5.18}$$

For $t \geq 0$, $\mu \in \mathcal{M}_1(\mathcal{X})$, $t_h > t_P$, and $k \in \mathbb{N}$, let :

$$\begin{aligned} a(k,t) = a^{t_h}_\mu(k,t)\ &:= \mathbf{1}_{\{k \leq J(t_h),\, k\, t_D \leq t\}} \times {}^{c_D}\!/_{c_P}\ \left(1 - {}^{c_D}\!/_{c_P}\ \right)^{k-1} \\ &\qquad \times \frac{\mathbb{P}_\mu(t_h < \tau_\partial)}{\mathbb{P}_\mu(t < \tau_\partial)} \times \frac{\mathbb{P}_\zeta(t - k\, t_D < \tau_\partial)}{\mathbb{P}_\zeta(t_h - k\, t_D < \tau_\partial)}. \end{aligned} \tag{1.5.19}$$

Remarks 1.5.3.1. *As we can see in the proof of Lemma 1.5.3.7, $a(k,t)$ corresponds to the mass associated with the k-th step of coupling, considered at time t with the constraint that it must represent a fixed proportion of μA_{t_h} (at time t_h). We refer to Figure 2 for a presentation of the coupling architecture.*

Let $\quad r_j\ := 1 - \sum_{\{k \leq j\}} a(k, j\, t_D)$. Under the condition $r_j > 0$, that we will prove to be true by induction over $j \leq J(t_h)$, we define :

$$\nu_j(dx) := ({}^1\!/r_j) \times \left[\mu A_{j\, t_D}(dx) - \textstyle\sum_{\{k \leq j\}} a(k, j\, t_D)\, \zeta A_{(j-k)\, t_D}(dx)\right], \tag{1.5.20}$$

with the convention $\nu_0 := \mu$. Remark that this definition ensures $\nu_j(\mathcal{X}) = 1$.

Remarks 1.5.3.2. *ν_j shall correspond to the marginal of the process conditioned of not being already coupled at time $j\, t_D$. We normalize what remains of $\mu A_{j\, t_D}$ when we subtract the contribution of each coupling step (only those up to the j-th will contribute to the sum). The main difficulty will be to prove that, under suitable conditions, ν_j is indeed a positive measure, thus a probability measure. In Figure 2, the associated coupling procedure is presented for the more general case where we compare two initial conditions in some $\mathcal{M}_{\ell,\xi}$ rather than already in $\mathcal{M}_R$.*

1.5.3.2 Definition of the constants involved

For clarity, we denote by t_h (for horizon of time) the time t that appears in Theorem 1.2.1. During this coupling procedure, it will stay fixed, and won't appear in the other sections. The constants $c_P, t_P > 0$ come from Theorem 1.5.2, while $c_D, t_D > 0$ come from this corollary of Theorem 1.5.1 :

Proposition 1.5.3.1. ***"Coupling and Renewal"***
Suppose that (A1) holds and (1.5.1) also for some $\mathcal{M}_\circ := \mathcal{M}_{\ell_\circ, \xi_\circ}$. Then, with $\ell_R := \ell_\circ$, $\xi_R := \xi_\circ/2$, $\mathcal{M}_R := \mathcal{M}_{\ell_R, \xi_R}$, $\mathcal{D}_R := \mathcal{D}_{\ell_R}$, there exits $c_D \in (0,1)$ and $t_D \geq t_\circ(\ell_R, \xi_R)$ such that :

$$\forall\, \mu \in \mathcal{M}_R, \qquad \mu A_{t_D}(dx) \geq c_D\, \zeta(dx) \text{ and } \frac{\mu A_{t_D}(\mathcal{D}_R) - c_D}{1 - c_D} \geq \xi_R \tag{1.5.21}$$

Remarks 1.5.3.3. *The subscript D refers to "Doeblin's" condition, since we will likewise iteratively couple a proportion at most c_D of the distribution. The properties (1.5.21) and (1.5.15) make us able to prove the induction : $\nu_j \in \mathcal{M}_R \Rightarrow \nu_{j+1} \in \mathcal{M}_R$.*

Proof of Proposition 1.5.3.1 We apply (1.5.1) with $\ell = \ell_R$ and $\xi = \xi_R$.
Thus, with $t_R := t_\circ(\ell_R, \xi_R)$:

$$\forall\, \mu \in \mathcal{M}_R,\ \forall\, t \geq t_R, \qquad \mu A_t \in \mathcal{M}_\circ, \text{ i.e. } \mu A_t(\mathcal{D}_R) \geq 2\, \xi_R \tag{1.5.22}$$

We can then define $c_M \in (0, 1)$, $t_M \geq t_R$ thanks to $(A1)$, cf Subsection 1.3.3, such that :

$$\forall\, x \in \mathcal{D}_R,\ \mathbb{P}_x\left[X_{t_M} \in dx\ ;\ t_M < \tau_\partial\right] \geq c_M\, \zeta(dx). \tag{1.5.23}$$

We can then choose $c_D := c_M\, \xi_R \in (0,1)$, $t_D := t_M \geq t_R$ (for the statement of the

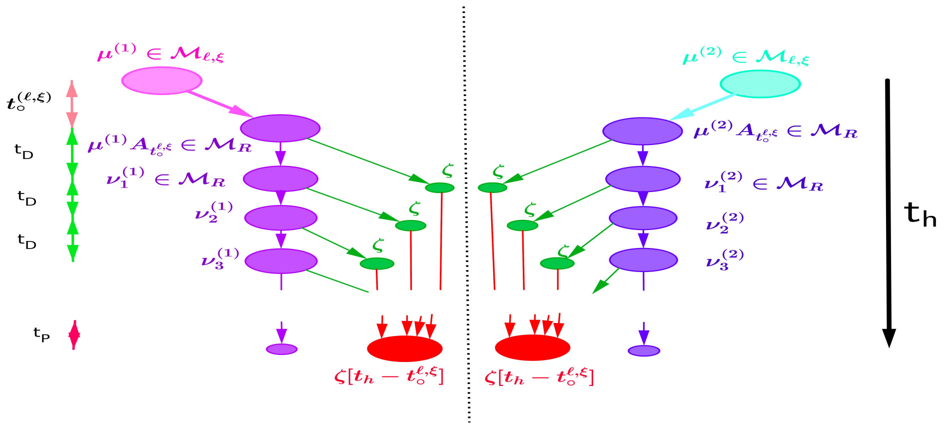


FIGURE 1.2 – Illustration of the coupling procedure

The figure illustrates the coupling procedure on two initial conditions $\mu^{(1)}$ and $\mu^{(2)}$. We can observe by symmetry how the MCNE are progressively decomposed with time descending along the vertical axis. By construction, the middle red parts (at time t_h) are common for both initial conditions (both its distribution $\zeta[t_h - t_\circ^{\ell,\xi}]$ and the amount of mass).
The procedure shall extend to cases where the QSD is not unique provided that $\mu^{(1)}$ and $\mu^{(2)}$ are in the same basin of attraction. This procedure already deals with an inhomogeneity in time due to the conditioning for survival at time t_h, so that the adaptation to inhomogeneous in time processes are likely to be easy.
Since both initial conditions belong to the same $\mathcal{M}_{\ell,\xi}$, the time $t_\circ = t_\circ^{\ell,\xi}$ needed to reach $\mathcal{M}_R$ can be chosen uniformly. Then, after every time-interval of size t_D and as long as time $t_h - t_P$ is not reached, we shall exploit property (1.5.21). We split the "remaining MCNE" (the ν_j after j splitting steps) in order to extract a component whose contribution to the MCNE at time t_h is explicit. This contribution (in the expression of $\zeta[t_h - t_\circ^{\ell,\xi}]$) is proportional to $\zeta A_{t_h - t}$ for a splitting at time t. For this contribution at time t_h to be fixed, note that the contribution to the MCNE at time t has to depend both on the remaining time $t_h - t$ and the specific value of ν_j.

---

proposition), and observe that :

$$\begin{aligned}
\forall\, \mu \in \mathcal{M}_R,\ \mu A_{t_M}(dx) &\geq \mu(\mathcal{D}_R)\, c_M\, \zeta(dx) && \text{by (1.5.23)}\\
&\geq \xi_R\, c_M\, \zeta(dx) = c_D\, \zeta(dx) && \text{because } \mu \in \mathcal{M}_R
\end{aligned}$$

$$\text{By (1.5.22)}, \qquad \frac{\mu A_{t_D}(\mathcal{D}_R) - c_D}{1 - c_D} \geq \frac{2\xi_R - c_D}{1 - c_D} = \frac{2 - c_M}{1 - \xi_R\, c_M}\xi_R \geq \xi_R,$$

where we used $1 \geq (1 - \xi_R)\, c_M$ (of course $c_M \in (0,1)$ and $\xi_R > 0$). $\square$

Remark : Our choice for ξ_R *and* c_D *is done for simplicity and can certainly be improved regarding the convergence rate* γ*. What we require is rather :* $\xi_R \leq \frac{c_M}{c_D} \wedge \frac{\xi_\circ - c_D}{1 - c_D}$.

1.5.3.3 Lower-bound on the marginals

At time t_h, for any initial condition $\mu \in \mathcal{M}_R$, the MCNE shall be lower-bounded by :

$$\zeta[t_h](dx) := \textstyle\sum_{\{j \leq J(t_h)\}} \left({}^{c_D}\!/_{c_P} \right) \times \left(1 - {}^{c_D}\!/_{c_P} \right)^{j-1} \zeta A_{t_h - j\, t_D}(dx) \geq 0 \tag{1.5.24}$$

Remark : The definition of $(\zeta[t])_{t \geq 0}$ implicitly depends on c_D, t_D, c_P and t_P, but not on μ, ℓ or ξ.

The proof of Theorem 1.2.1 will be completed thanks to Theorems 1.5.1, 1.5.2 and :

Proposition 1.5.3.2. *Suppose* (1.5.1)*,* (A1) *and* (1.5.15) *hold, with* c_D *and* t_D *chosen according to Proposition 1.5.3.1,* c_P*,* t_P *according to* (1.5.15)*. Then, to any pair* $\ell \geq 0$ *and* $\xi \in (0,1)$*, we can associate a time* $t_\circ = t_\circ(n, \xi) > 0$ *such that :*

$$\forall\, \mu \in \mathcal{M}_{\ell,\xi},\ \forall\, t_{h,2} \geq t_{h,1} \geq t_\circ, \qquad \mu A_{t_{h,2}} \geq \zeta[t_{h,1} - t_\circ] \tag{1.5.25}$$

Proof that (1.5.25) implies (1.2.2) Thanks to (1.5.25) :

$$\begin{aligned}
&\forall\, \ell \in \mathbb{N},\ \forall\, \xi \in (0,1),\ \forall\, (\mu_1,\, \mu_2) \in (\mathcal{M}_{\ell,\xi})^2,\ \forall\, t_{h,2} \geq t_{h,1} \geq t_\circ, \\
&\qquad \mu_1 A_{t_{h,1}} \geq \zeta[t_{h,1} - t_\circ] \quad \text{and} \quad \mu_2 A_{t_{h,2}} \geq \zeta[t_{h,1} - t_\circ] \\
&\text{thus } \|\mu_2 A_{t_{h,2}} - \mu_1 A_{t_{h,1}}\|_{TV} \leq \|\mu_2 A_{t_{h,2}} - \zeta[t_{h,1} - t_\circ]\|_{TV} + \|\mu_1 A_{t_{h,1}} - \zeta[t_{h,1} - t_\circ]\|_{TV} \\
&\qquad\qquad \leq 2 \times [1 - \zeta[t_{h,1} - t_\circ](\mathcal{X})].
\end{aligned} \tag{1.5.26}$$

$$\begin{aligned}
1 - \zeta[t_h](\mathcal{X}) &= 1 - \textstyle\sum_{\{j \leq J(t_h)\}} \left({}^{c_D}\!/_{c_P} \right) \left[1 - {}^{c_D}\!/_{c_P} \right]^{j-1} && \text{by (1.5.24)} \\
&= \left[1 - {}^{c_D}\!/_{c_P} \right]^{J(t_h)} \leq \exp[-\gamma\,(t_h - t_P - t_D)] && \text{by (1.5.18)} \\
&\quad \text{with } \gamma := -\frac{1}{t_D} \ln\left[1 - \frac{c_D}{c_P}\right]
\end{aligned} \tag{1.5.27}$$

Finally, with (1.5.26), (1.5.27) and $C = C(\ell,\, \xi) := 2\, \exp[\gamma\,(t_P + t_D + t_\circ(\ell, \xi))]$:

$$\|\mu_2 A_{t_{h,2}} - \mu_1 A_{t_{h,1}}\|_{TV} \leq C\, e^{-\gamma\, t_{h,1}}. \tag{1.5.28}$$

This states that for any $\mu \in \mathcal{M}_{\ell,\xi}$, $(\mu A_{t_h})_{\{t_h \geq 0\}}$ is a Cauchy-sequence for the total variation distance. Thus, it converges for this distance to some distribution $\alpha^{\ell,\xi}$. By letting $t_{h,2}$ go to infinity in the last inequality, we further get an exponential rate of convergence γ, independent of ℓ and ξ, and in particular on μ.

Since immediately : $\forall\, \ell \leq \ell',\ \forall\, \xi \geq \xi' > 0, \quad \mathcal{M}_{\ell,\xi} \subset \mathcal{M}_{\ell',\xi'}$, we deduce $\alpha^{\ell,\xi} = \alpha^{\ell\vee\ell',\xi\wedge\xi'} = \alpha^{\ell',\xi'}$, which means (since $\mathcal{M}_1(\mathcal{X}) = \bigcup_{(\ell,\xi)} \mathcal{M}_{\ell,\xi}$) that a unique distribution α is the attractor. In particular, there cannot be a QSD different from α.
For any initial condition μ : $\lim_{t\to\infty} \mathbb{P}_\mu(X_t \in dy \,\big|\, t < \tau_\partial) = \alpha(dy)$, where the convergence holds in the weak topology (ie α is a *quasi-limiting distribution*). One can then easily adapt the proof of Lemma 7.2 in [CC+09] to deduce that α is effectively a QSD and $\forall\, t \geq 0,\ \mathbb{P}_\alpha(t < \tau_\partial) = e^{-\lambda t}$. By letting $t_{h,2} \to \infty$ in (1.5.28), with $\mu_2 = \mu_1 = \mu \in \mathcal{M}_{\ell,\xi}$:

$$\| \, \mathbb{P}_\mu \left[X_t \in dx \mid t < \tau_\partial \right] - \alpha(dx) \, \|_{TV} \leq C(\ell,\xi)\, e^{-\gamma\, t}$$

This ends the proof of (1.2.2) given (1.5.25). □

Proof of Proposition 1.5.3.2 :
In order to achieve the induction "$\nu_j \in \mathcal{M}_R$ implies $\nu_{j+1} \in \mathcal{M}_R$" we ensure iteratively :

$$\mathbb{P}_{\nu_j}\left(t_h - j\, t_D < \tau_\partial\right) \leq c_P\, \mathbb{P}_{\nu_j}\left(t_D < \tau_\partial\right) \mathbb{P}_\zeta\left(t_h - [j+1]\, t_D < \tau_\partial\right), \tag{1.5.29}$$
$$\text{and } \nu_j A_{t_D} \geq c_D\, \zeta, \tag{1.5.30}$$

that come respectively from (1.5.15) and (1.5.21).
In practice, to conclude the proof of Theorem 1.5.1, we need :

Lemma 1.5.3.3. *Assume that* (1.5.15) *holds (cf Theorem 1.5.2). If moreover, for some* $j \leq J(t_h) - 1$, $r_j > 0, \nu_j \in \mathcal{M}_1(\mathcal{X})$, *then* (1.5.29) *holds.*

Lemma 1.5.3.4. *Assume* $r_j > 0$, $\nu_j \in \mathcal{M}_1(\mathcal{X})$, *and that* (1.5.29), (1.5.30) *hold.*
Then : $r_{j+1} > 0, \quad \nu_{j+1} \in \mathcal{M}_1(\mathcal{X})$.
Moreover, for any measurable subset $\mathcal{D}$ *of* $\mathcal{X}$ *:* $\quad \nu_{j+1}(\mathcal{D}) \geq (\nu_j A_{t_D}(\mathcal{D}) - c_D)/(1 - c_D)$.

Lemma 1.5.3.5. *Assume that :* $r_{J(t_h)} > 0$, $\nu_{J(t_h)} \in \mathcal{M}_1(\mathcal{X})$. *Then* $\mu A_{t_h} \geq \zeta[t_h]$.

Proof of Proposition 1.5.3.2 with Lemmas 1.5.3.3, 1.5.3.4 and 1.5.3.5 Let us first assume that $\mu \in \mathcal{M}_R$, where we use Proposition 1.5.3.1 together with $(A1)$ and (1.5.1) to define $\mathcal{M}_R$ such that (1.5.21) holds.
Then, by induction over $j \leq J(t_h)$, we state (I_j) : $r_j > 0$ and $\nu_j \in \mathcal{M}_R$. We initialize at $j = 0$, with $r_0 = 1$ and $\nu_0 := \mu \in \mathcal{M}_R$ by hypothesis.
Assume (I_j) for some $j \leq J(t_h) - 1$. Then, by (I_j) and (1.5.21), (1.5.30) holds. (1.5.15) implies (1.5.29), which implies with (I_j), (1.5.30) and Lemma 1.5.3.4 that : $r_{j+1} > 0$ and $\nu_{j+1} \in \mathcal{M}_1(\mathcal{X})$. Also thanks to Lemma 1.5.3.4 and again (1.5.21) we prove finally :
$\nu_{j+1}(\mathcal{D}_R) \geq \frac{1}{1-c_D}\left(\nu_j A_{t_D}(\mathcal{D}_R) - c_D\right) \geq \xi_R$. Therefore, (I_{j+1}) holds.

By induction, we get $(I_{J(t_h)})$ thus $r_{J(t_h)} > 0$ and $\nu_{J(t_h)} \in \mathcal{M}_R \subset \mathcal{M}_1(\mathcal{X})$. By Lemma 1.5.3.5, this indeed concludes the proof that :

$$\forall \mu \in \mathcal{M}_R,\ \forall t_h > 0, \qquad \mu A_{t_h} \geq \zeta[t_h] \tag{1.5.31}$$

For general initial conditions, recall that in Proposition 1.5.3.1, we constructed $\mathcal{M}_R$ such that $\mathcal{M}_\circ \subset \mathcal{M}_R$. Thus, (1.5.1) holds with $\mathcal{M}_R$ instead of $\mathcal{M}_\circ$.
Since moreover $t_{h,2} \geq t_{h,1}$: $\quad \mu A_{t_\circ + t_{h,2} - t_{h,1}} \in \mathcal{M}_R$.
Since $\mu A_{t_{h,2}} = \mu A_{t_\circ + t_{h,2} - t_{h,1}} A_{t_{h,1} - t_\circ}$, we finally deduce from (1.5.31) :

$$\mu A_{t_{h,2}} \geq \zeta[t_{h,1} - t_\circ].$$

□

Proof of Lemma 1.5.3.3 : $j \leq J(t_h) - 1$ means notably $t_h - [j+1]\, t_D \geq t_P$, thus :

$$\begin{aligned}
\mathbb{P}_{\nu_j}(t_h - j\, t_D < \tau_\partial) &= \mathbb{E}_{\nu_j}\left[\mathbb{P}_{X_{t_D}}(t_h - [j+1]\, t_D < \tau_\partial)\ ;\ t_D < \tau_\partial\right] \\
&\leq c_P\, \mathbb{P}_\zeta(t_h - [j+1]\, t_D < \tau_\partial) \times \mathbb{P}_{\nu_j}[t_D < \tau_\partial] \qquad \text{by (1.5.15)}
\end{aligned}$$
□

Proof of Lemma 1.5.3.5 : By (1.5.20), i.e. the definition of $\nu_{J(t_h)}$, and since (A_t) is a semigroup :

$$\begin{aligned}
\mu A_{t_h} &= \frac{\mathbb{P}_\mu(J(t_h)\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)}\, \mu A_{J(t_h)\, t_D} P_{t_h - J(t_h)\, t_D} \\
&\geq \frac{\mathbb{P}_\mu(J(t_h)\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \left[\textstyle\sum_{\{k \leq J(t_h)\}} a(k, J(t_h)\, t_D)\, \zeta A_{[J(t_h)-k]\, t_D}\ P_{t_h - J(t_h)\, t_D}\right] \\
&\geq \sum_{k \leq J(t_h)} \frac{\mathbb{P}_\mu(J(t_h)\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)}\, a(k, J(t_h)\, t_D)\, \frac{\mathbb{P}_\zeta(t_h - k\, t_D < \tau_\partial)}{\mathbb{P}_\zeta([J(t_h) - k]\, t_D < \tau_\partial)} \zeta A_{t_h - k\, t_D}.
\end{aligned}$$

Finally, by (1.5.19) and (1.5.24) : $\qquad \mu A_{t_h} \geq \zeta[t_h]$. □

Proof of Lemma 1.5.3.4 : For the proof of this lemma, we need the following elementary lemma, whose proof is postponed after we deduce Lemma 1.5.3.4 :

Lemma 1.5.3.6. *Assume that for some $j \leq J(t_h) - 1$, $r_j > 0$, $\nu_j \in \mathcal{M}_1(\mathcal{X})$ and (1.5.29) holds. Then :*

$$r_{j+1} > 0 \qquad \textit{and}\ \exists\, 0 < c_j \leq c_D, \quad \nu_{j+1}(dx) = (\nu_j A_{t_D}(dx) - c_j\, \zeta(dx)) \,/\, (1 - c_j).$$

Thanks to Lemma 1.5.3.6 together with (1.5.30) : $\nu_j \geq 0$ thus $\nu_j \in \mathcal{M}_1(\mathcal{X})$.

Moreover, for any measurable set $\mathcal{D}$:

$$\begin{aligned}\nu_{j+1}(\mathcal{D}) &\geq (\nu_j A_{t_D}(\mathcal{D}) - c_j) \,/(1-c_j) = 1 - (1-\nu_j A_{t_D}(\mathcal{D}))\,/\,(1-c_j)\\ &\geq (\nu_j A_{t_D}(\mathcal{D}) - c_D)\,/\,(1-c_D)\end{aligned}$$

since $\zeta(\mathcal{D}) \vee \nu_j A_{t_D}(\mathcal{D}) \leq 1,\; c_j \leq c_D$ and : $c \to 1-(1-\nu_j A_{t_D}(\mathcal{D}))\,/\,(1-c)$ is decreasing. □

Proof of Lemma 1.5.3.6 : First of all, we need to relate $1-\sum_{k\geq 1} a(k, j\,t_D)$ to the repartition of mass at time t_h, which is done in the proof of the following lemma, whose proof (similar to the next paragraph, yet much simpler) is postponed in Appendix E :

Lemma 1.5.3.7. *Assume that for some $j \leq J(t_h) - 1$: $\quad r_j > 0,\; \nu_j \in \mathcal{M}_1(\mathcal{X})$.*

$$\textit{Then } r_j = \left[1 - \frac{c_D}{c_P}\right]^j \times \frac{\mathbb{P}_\mu(t_h < \tau_\partial)}{\mathbb{P}_\mu(j\,t_D < \tau_\partial)} \times \frac{1}{\mathbb{P}_{\nu_j}(t_h - j\,t_D < \tau_\partial)}.$$

Proof of Lemma 1.5.3.6 with Lemma 1.5.3.7 :
By the definition of ν_j, cf (1.5.20) : $\mu A_{j\,t_D} = \sum_{k=1}^{j} a(k, j\,t_D)\,\zeta A_{(j-k)\,t_D} + r_j\,\nu_j$.

$$\begin{aligned}\mu A_{[j+1]\,t_D} &= \frac{\mathbb{P}_\mu(j\,t_D < \tau_\partial)}{\mathbb{P}_\mu([j+1]\,t_D < \tau_\partial)}\,\mu A_{j\,t_D}\cdot P_{t_D}\\ &= \textstyle\sum_{\{k\leq j\}} a(k, j\,t_D) \times \dfrac{\mathbb{P}_\mu(j\,t_D < \tau_\partial) \times \mathbb{P}_\zeta([j+1-k]\,t_D < \tau_\partial)}{\mathbb{P}_\mu([j+1]\,t_D < \tau_\partial) \times \mathbb{P}_\zeta([j-k]\,t_D < \tau_\partial)}\zeta A_{[j+1-k]\,t_D}\\ &\qquad + \ell_j\,\nu_j A_{t_D}, \end{aligned} \tag{1.5.32}$$

$$\text{where } \ell_j := r_j \times \frac{\mathbb{P}_\mu(j\,t_D < \tau_\partial)}{\mathbb{P}_\mu([j+1]\,t_D < \tau_\partial)} \times \mathbb{P}_{\nu_j}(t_D < \tau_\partial). \tag{1.5.33}$$

By (1.5.19), i.e. the definition of $a(k, j\,t_D)$:

$$a(k, j\,t_D) \times \frac{\mathbb{P}_\mu(j\,t_D < \tau_\partial)}{\mathbb{P}_\mu([j+1]\,t_D < \tau_\partial)} \times \frac{\mathbb{P}_\zeta([j+1-k]\,t_D < \tau_\partial)}{\mathbb{P}_\zeta([j-k]\,t_D < \tau_\partial)} = a(k, [j+1]\,t_D).$$

$$\text{Thus} \quad 1 = \textstyle\sum_{\{k\leq j\}} a(k, [j+1]\,t_D) + \ell_j \quad \text{i.e.} \quad r_{j+1} = \ell_j - a(j+1, [j+1]\,t_D) \tag{1.5.34}$$

by evaluating (1.5.32) on $\mathcal{X}$ and the definition of $r_{j+1}(cf.above(1.5.20))$.

By (1.5.33), (1.5.19) and by Lemma 1.5.3.7 :

$$\begin{aligned}
c_j :=& a(j+1, [j+1]\, t_D)/\ell_j \qquad (1.5.35)\\
=& \left[1 - \frac{c_D}{c_P}\right]^{-j} \times \frac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \times \frac{\mathbb{P}_{\nu_j}(t_h - j\, t_D < \tau_\partial)}{\mathbb{P}_{\nu_j}(t_D < \tau_\partial)} \times \frac{\mathbb{P}_\mu([j+1]\, t_D < \tau_\partial)}{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}\\
&\times \frac{c_D}{c_P} \left(1 - \frac{c_D}{c_P}\right)^j \times \frac{\mathbb{P}_\mu(t_h < \tau_\partial)}{\mathbb{P}_\mu([j+1]\, t_D < \tau_\partial)} \times \frac{1}{\mathbb{P}_\zeta(t_h - [j+1]\, t_D < \tau_\partial)}\\
=& \frac{c_D}{c_P} \frac{\mathbb{P}_{\nu_j}(t_h - j\, t_D < \tau_\partial)}{\mathbb{P}_{\nu_j}(t_D < \tau_\partial) \times \mathbb{P}_\zeta(t_h - [j+1]\, t_D < \tau_\partial)}.
\end{aligned}$$

Thanks to (1.5.29) : • $0 < c_j \leq c_D$.
Since $c_D < 1$, using (1.5.34) and (1.5.35) : • $r_{j+1} = \ell_j\ (1 - c_j) > 0$.
Finally, by (1.5.20), i.e. the definition of ν_{j+1}, (1.5.35) and (1.5.32) :

$$\begin{aligned}
\nu_{j+1} &= (1/r_{j+1}) \times \left[\mu A_{[j+1]\, t_D} - \textstyle\sum_{\{k \leq j+1\}} a(k, (j+1)\, t_D)\, \zeta A_{(j+1-k)\, t_D}\right]\\
&= (\nu_j A_{t_D} - c_j\, \zeta) \times \ell_j\, / r_{j+1}
\end{aligned}$$

• $\nu_{j+1} = (\nu_j A_{t_D} - c_j\, \zeta)\, / (1 - c_j)$. □

The proof of Theorem 1.2.1 is now complete (up to Appendix D).

1.5.4 Proof of Theorem 1.2.2 :

Step 1 : proof of the uniform convergence to h :

Considering the conclusions of Theorem 1.5.2, it is easily seen that we can replace ζ by α (by any probability measure in fact). To achieve this, we only need to apply $(A1)$ and adjust the value for c_P : $c'_P := c_P\, e^{-\lambda\, t_M}\, / \,(\alpha(\mathcal{D}_{\ell_M})\, c_M)$, where t_M, c_M are given by $(A1)$ for initial condition in $\mathcal{D}_{\ell_M}$. This can be translated in term of a uniform bound on h by :

$$\|h_\bullet\|_\infty := \sup_{\{t \geq 0\}} \|h_t\|_\infty \leq c'_P \vee e^{\lambda\, t_P} < \infty. \qquad (1.5.36)$$

Like in the proof of Proposition 2.3 in [CV16], we deduce that, for any $s, t > 0$, $\mu \in \mathcal{M}_{\ell,\, \xi}$:

$$|\langle \mu | h_t \rangle - \langle \mu | h_{t+s} \rangle| = \langle \mu | h_t \rangle\, |\langle \alpha - \mu | h_s \rangle| \leq \|h_\bullet\|_\infty^2\, C(\ell, \xi)\, e^{-\gamma\, t}. \qquad (1.5.37)$$

The constant C can actually be taken independently of ℓ, ξ. Indeed, because the previous expression is linear in μ and $\langle \alpha \,\big|\, h_t \rangle \equiv 1$:

$$|\langle \mu \,\big|\, h_t - h_{t+s} \rangle| = 2 |\langle \bar{\mu} \,\big|\, h_t - h_{t+s} \rangle|, \qquad \text{where } \bar{\mu} := (\mu + \alpha)/2.$$

By choosing ℓ sufficiently large to ensure $\xi := \alpha(\mathcal{D}_{\ell_M})/2 > 0$, we deduce that for any $\mu \in \mathcal{M}_1(\mathcal{X})$, $\bar{\mu} \in \mathcal{M}_{\ell,\xi}$. The inequality (1.5.37) is thus uniform in $\mu \in \mathcal{M}_1(\mathcal{X})$, so that (h_t) defines a Cauchy sequence for the uniform norm. We deduce that h_t converges to some unique function h, whose norm is also bounded by $\|h_\bullet\|_\infty$. □

Step 2 : Characterization of the survival capacity h

The rest of the proof is directly taken from [CV16].

As the punctual limit of (h_t), and since for any $t \geq 0$, h_t vanishes on ∂, this also hold for h. With the uniform bound (1.5.36), we deduce that h is also bounded.

As stated in the beginning of this Subsection 1.5.4, we can replace ζ by any probability measure μ in (1.5.15) (with specific values for $c_P(\mu) = c_P(\ell)/\xi, t_P(\mu) = t_P(\ell) > 0$). In particular, for $\mu = \delta_x$, with $x \in \mathcal{D}_\ell$:

$$\forall\, t \geq t_P(\ell), \quad \mathbb{P}_\alpha(t < \tau_\partial) \leq c_P(\ell)\, \mathbb{P}_x(t < \tau_\partial) \quad \text{thus } \forall\, t \geq t_P(\ell), \quad h_t(x) \geq c_P(\ell) > 0.$$

This proves that h has a positive lower-bound on any $\mathcal{D}_\ell$.

By the Markov property and (1.2.3) :

$$\forall\, h > 0, \quad P_h h(x) = \lim_{t\to\infty} \frac{\mathbb{E}_x\left[\mathbb{P}_{X_h}(t < \tau_\partial)\right]}{\mathbb{P}_\alpha(t < \tau_\partial)} = e^{-\lambda\, h} \lim_{t\to\infty} \frac{\mathbb{P}_x(t + h < \tau_\partial)}{\mathbb{P}_\alpha(t + h < \tau_\partial)} = e^{-\lambda\, h}\, h(x).$$

From this and (1.5.36), we immediately deduce that h is in the domain of $\mathcal{L}$ and $\mathcal{L}\, h = -\lambda\, h$.

□

1.5.5 Proof of Theorem 1.2.3 :

Except for (iii), for which we will prove (1.2.8), and for the uniqueness of the stationary distribution, the proof is almost the same as in [CV16].

Step 1 : Proof that the Q-process is well-defined and characterization

Let Λ_s be a $\mathcal{F}_s$-measurable set and $\mu \in \mathcal{M}_1(\mathcal{X})$. By the Markov property :

$$\mathbb{P}_\mu(\Lambda_s \,\big|\, t < \tau_\partial) = \mathbb{E}_\mu\left[e^{\lambda s}\, h_{t-s}(X_s)/\langle \mu \,\big|\, h_t\rangle \;;\; s < \tau_\partial \;,\; \Lambda_s\right]$$

By Theorem 1.2.2, the random variable $M_s^t := \mathbf{1}_{\{s<\tau_\partial\}}\, e^{\lambda s}\, h_{t-s}(X_s) \;/\; \langle \mu \,\big|\, h_t\rangle$, (where $t \geq s$) converges a.s. to :

$$M_s := \mathbf{1}_{\{s<\tau_\partial\}}\, e^{\lambda s}\, h(X_s) \;/\; \langle \mu \,\big|\, h\rangle, \tag{1.5.38}$$

where $\langle\mu \,\big|\, h\rangle > 0$ because h is positive on $\mathcal{X}$. For t sufficiently large (a priori depending on μ), we deduce from (1.5.36) and the convergence of $\langle\mu \,\big|\, \eta_t\rangle$ to $\langle\mu \,\big|\, \eta\rangle$:

$$0 \leq M_s^t \leq 2\, e^{\lambda s}\, \|h_\bullet\|_\infty \;/\; \langle\mu \,\big|\, h\rangle. \tag{1.5.39}$$

Thus, by the dominated convergence Theorem, we obtain that $\quad E_\mu(M_s) = 1.$

By the penalisation's theorem of Roynette, Vallois and Yor (cf Theorem 2.1 in [RVY06]) these two conditions imply that M is a martingale under P_μ and that $\mathbb{P}_\mu(\Lambda_s \,\big|\, t < \tau_\partial)$ converges to $E_\mu(M_s \;;\; \Lambda_s)$ for all $\Lambda_s \in \mathcal{F}_s$ when $t \to \infty$. In particular for $\mu = \delta_x$, this means that $\mathbb{Q}_x$ is well defined and :

$$\frac{d\mathbb{Q}_x}{d\mathbb{P}_x}\Big|_{\mathcal{F}_s} = \mathbf{1}_{\{s<\tau_\partial\}}\, e^{\lambda s}\, \frac{h(X_s)}{h(x)}. \tag{1.5.40}$$

(1.5.40) implies directly (1.2.7). Concerning (1.2.11) :

$$\begin{aligned}
\mu B[h] Q_t(dy) &= \int \frac{h(x)}{\langle\mu \,\big|\, h\rangle} \mu(dx)\, \frac{h(y)}{h(x)}\, e^{\lambda\, t}\, p(x;t;dy) \\
&= \frac{h(y)}{\langle\mu \,\big|\, P_t\, h\rangle}\, \mu P_t(dy) = \mu P_t B[h] \qquad\quad \text{by (1.2.5)} \\
&= \frac{h(y)\, \mathbb{P}_\mu(t<\tau_\partial)}{\langle\mu\, P_t \,\big|\, h\rangle} \times \frac{\mu P_t(dy)}{\mathbb{P}_\mu(t<\tau_\partial)} = \frac{h(y)}{\langle\mu\, A_t \,\big|\, h\rangle}\, \mu A_t(dy) = \mu A_t B[h]
\end{aligned}$$

For a more general convergence, with μ as initial condition and $\Lambda_s \in \mathcal{F}_s$, we deduce :

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{P}_\mu(\Lambda_s \,\big|\, t<\tau_\partial) &= E_\mu\left(e^{\lambda s}\, h(X_s) \,/\, \langle\mu \,\big|\, h\rangle \;;\; s<\tau_\partial, \Lambda_s\right) \\
&= \int_{\mathcal{X}} \mu(dx) \frac{h(x)}{\langle\mu \,\big|\, h\rangle} E_x\left(e^{\lambda s}\, \frac{h(X_s)}{h(x)} \;;\; s<\tau_\partial, \Lambda_s\right) \\
&= \int_{\mathcal{X}} \mu(dx) \frac{h(x)}{\langle\mu \,\big|\, h\rangle} \mathbb{Q}_x(\Lambda_s) = \mathbb{Q}_{\mu B[h]}(\Lambda_s).
\end{aligned}$$

by (1.5.40) and the definition of $\mu B[h]$ in (1.2.10).

Moreover, the convergence holds in fact in total variation over $\mathcal{F}_s$, as we prove it now. By the previous calculations, (1.5.39) and (1.5.36), for any $\epsilon > 0$:

$$\begin{aligned}
\left\| \mathbb{P}_\mu(dw \,\middle|\, t < \tau_\partial) - \mathbb{Q}_{\mu B[h]}(dw) \right\|_{TV,\mathcal{F}_s} &\leq \mathbb{E}_\mu |M_s^t - M_s| \\
&\leq 4\, e^{\lambda s}\, \frac{\|h_\bullet\|_\infty}{\langle \mu \,|\, h\rangle}\, \mathbb{P}_\mu(|M_s^t - M_s| \geq \epsilon) + \epsilon \\
\text{so } \limsup_{t\to\infty} \left\| \mathbb{P}_\mu(dw \,\middle|\, t < \tau_\partial) - \mathbb{Q}_{\mu B[h]}(dw) \right\|_{TV,\mathcal{F}_s} &\leq \epsilon,
\end{aligned}$$

By letting $\epsilon \to 0$, we conclude :

$$\forall\, s \in \mathbb{R}_+, \qquad \left\| \mathbb{P}_\mu(dw \,\middle|\, t < \tau_\partial) - \mathbb{Q}_{\mu B[h]}(dw) \right\|_{TV,\mathcal{F}_s} \xrightarrow[t\to\infty]{} 0 \tag{1.5.41}$$

For the proof that X defines a strong Markov process under $(\mathbb{Q}_x)_{x\in\mathcal{X}}$, we refer again to the proof in [CV16].

Step 2 : The invariant distribution for X under $\mathbb{Q}$

For all $t \geq 0$ and $f \in \mathcal{B}_b(\mathcal{X})$, with (1.5.40) :

$$\begin{aligned}
\langle \beta \,|\, Q_t\, f\rangle &= \langle \alpha \,|\, h \times Q_t\, f\rangle = e^{\lambda\, t}\, \langle \alpha \,|\, P_t\, (h \times f)\rangle \\
&= \langle \alpha \,|\, h \times f\rangle = \langle \beta \,|\, f\rangle
\end{aligned} \tag{1.5.42}$$

where we used (1.2.3). We prove the uniqueness with the next subsection.

Step 3 : Proof of (1.2.8)

Exploiting (1.2.11), we deduce from our definitions :

$$\begin{aligned}
\|(\mu B[h])Q_t - \beta\|_{\frac{1}{h}} &= \left\| \frac{\mu A_t}{\langle \mu A_t \,|\, h\rangle} - \alpha \right\|_{TV} \\
&\leq \langle \mu A_t \,|\, h\rangle^{-1} \times [\|\mu A_t - \alpha\|_{TV} + |\langle \mu A_t \,|\, h\rangle - 1|]
\end{aligned} \tag{1.5.43}$$

To ensure a lower-bound on $\langle \mu A_t \,|\, h\rangle$, we exploit (1.2.5) and write :

$$\langle \mu A_t \,|\, h\rangle = \frac{e^{\lambda t} \langle \mu P_t \,|\, h\rangle}{\langle \mu \,|\, h_t\rangle} = \frac{\langle \mu \,|\, h\rangle}{\langle \mu \,|\, h_t\rangle}.$$

We already know that h_t is uniformly upper-bounded and h has a lower-bound on any $\mathcal{D}_\ell$. Since $|\langle \mu A_t \,|\, h\rangle - 1| = |\langle \mu A_t - \alpha \,|\, h\rangle| \leq \|\mu A_t - \alpha\|_{TV}\ \|h_\bullet\|_\infty$, and exploiting

(1.5.43) and (1.2.2), we conclude that there exists $C' = C'(\ell, \xi) > 0$ such that :

$$\forall t > 0,\ \forall \mu \in \mathcal{M}_{\ell,\xi}, \quad \|\mathbb{Q}_{\mu B[h]}(X_t \in dx) - \beta(dx)\|_{\frac{1}{h}} \le C'\, e^{-\gamma\, t}. \qquad \square$$

Step 4 : Convergence with initial condition for the Q-process

When μ_Q is the initial condition of the Q-process, it is in general not possible to interpret it as $\mu B[h]$. Indeed, we should expect in this case $\mu(dx)$ to be proportional to $h(x)^{-1}\, \mu_Q(dx)$, which may not be integrable. Thus, the convergence to β might in general not be exponential.

However, it is exponential for measures with support in any of the $\mathcal{D}_\ell$, in particular Dirac masses. Indeed, we have a lower-bound of h : $\quad h^{(\ell)} := \inf\{h_x\ ;\ x \in \mathcal{D}_\ell\}$, which is positive because of $(A1)$ and (1.5.15). Thus, if $\mu_Q \in \mathcal{M}_1(\mathcal{X})$ has support on $\mathcal{D}_\ell$, $\langle \mu_Q \,\big|\, 1/h\rangle \le 1/h^{(\ell)} < \infty$, so :

$$\mu_Q = \mu B[h], \qquad \text{with } \mu(dx) := \mu_Q B(1/h) := \mu_Q(dx) \,/\, (h(x) \times \langle \mu_Q \,\big|\, 1/h\rangle).$$

Now, μ has the same support as μ_Q, thus $\mu(\mathcal{D}_\ell) = 1$, i.e. $\mu \in \mathcal{M}_{\ell,1}$. By (1.2.8) :

$$\|\mu_Q Q_t - \beta\|_{TV} = \|\mu B[h] Q_t - \beta\|_{TV} \le C(\ell, 1)\, e^{-\gamma\, t}.$$

More generally, since the Q-process is linear with its initial condition, and by $(A0)$, the property of uniqueness of the stationary distribution β holds.

Besides, to have exponential convergence, it suffices that : $\langle \mu_Q \,\big|\, 1/h\rangle < \infty$. It can be deduced from $\sum_{\ell \ge 1} \mu_Q(\mathcal{D}_\ell \setminus \mathcal{D}_{\ell-1}) \,/ h^{(\ell)} < \infty$ (note that one has lower-bounds of $h^{(\ell)}$). In any case, the convergence still holds in total variation. $\square$

Appendices :

Appendix A : Combine all the inequalities to prove $(A2)$

We shall first prove that an upper-bound of the global supremum can be deduced from the upper-bounds in Propositions 1.4.2.1-3. This shall hold at least for ϵ^X and ϵ^0 sufficiently small, which is obtained with n_E sufficiently large. The constraints on ϵ^X and ϵ^0 are mentioned while we handle the inequalities. We prove next that we can indeed find suitable choices of ϵ^X, ϵ^0, C^X_∞, C^N_∞ and C_0 for the upper-bounds in Propositions 1.4.2.1-3 to hold with these constraints.

t_h is introduced to make sure that $\mathcal{E}^X_\infty \vee \mathcal{E}^N_\infty \vee \mathcal{E}_0 < \infty$ $(\le \exp[\rho\, t_h])$. It is needed to justify the following inequalities, but this specific upper-bound plays no role.

By the upper-bounds in Propositions 1.4.2.2 and 1.4.2.1 :

$$\mathcal{E}_\infty^X \leq C_\infty^X\,(1+\mathcal{E}_0) + \epsilon^X\,C_\infty^N (1+\mathcal{E}_\infty^X)$$
$$(1-\epsilon^X\,C_\infty^N)\,\mathcal{E}_\infty^X \leq C_\infty^X + \epsilon^X\,C_\infty^N + C_\infty^X\,\mathcal{E}_0$$

So we assume in the following that : $\epsilon^X \leq (2\,C_\infty^N)^{-1}$, and we recall that $C_\infty^N \wedge C_\infty^X \geq 1$, which, combined with the upper-bounds of Proposition 1.4.2.3, yields :

$$\mathcal{E}_\infty^X \leq 3\,C_\infty^X + 2\,C_\infty^X\,\mathcal{E}_0, \qquad \mathcal{E}_\infty^N \leq 4\,C_\infty^N\,C_\infty^X + 2\,C_\infty^N\,C_\infty^X\,\mathcal{E}_0$$
$$\mathcal{E}_0 \leq C_0 + 7\,\epsilon^0\,C_\infty^N\,C_\infty^X + 4\,\epsilon^0\,C_\infty^N\,C_\infty^X \mathcal{E}_0$$
$$\text{thus}\ \left(1 - 4\,\epsilon^0 C_\infty^N\,C_\infty^X\right)\,\mathcal{E}_0 \leq C_0 + 7\,\epsilon^0 C_\infty^N\,C_\infty^X$$

In the following, we assume that : $\quad \epsilon^0 \leq \left(8\,C_\infty^N\,C_\infty^X\right)^{-1}.$
We recall also that $C_\infty^N \geq 1$ and $C_0 \geq 1$, so that :

$$\mathcal{E}_0 \leq 4\,C_0\,, \qquad \mathcal{E}_\infty^X \leq 11\,C_\infty^X\,C_0\,, \qquad \mathcal{E}_\infty^N \leq 12\,C_\infty^N\,C_\infty^X\,C_0.$$

Finally, provided : $\quad \epsilon^X \leq (2\,C_\infty^N)^{-1}, \qquad \epsilon^0 \leq \left(8\,C_\infty^N\,C_\infty^X\right)^{-1},$
conditions which we can satisfy and restrict the choices of n_∞ and $n_E > n_\infty$, we deduce :

$$sup_{(x,n)}\left\{\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)]\right\} \leq 12\,C_\infty^N\,C_\infty^X\,C_0 < \infty \tag{1.6.1}$$

More precisely, for any ρ, we obtain from Proposition 1.4.2.1 the constants y_∞ and C_∞^N which gives us a value for ϵ^X. We then deduce, thanks to Proposition 1.4.2.2, some value for n_E^X and C_∞^X. The value of ϵ^0 can then be fixed, so that we can choose, according to Proposition 1.4.2.3, some value $n_E^0 > 0$ and C_0. To make the inequalities of Propositions 1.4.2.2 and 1.4.2.3 hold, we can just take $n_E := n_E^X \vee n_E^0$. Taking the limit in (1.6.1) as $t_h \to \infty$ (recall that $\widehat{\tau}_E := \tau_E \wedge \tau_\partial \wedge t_h$) concludes the proof of $(A2)$.
□

Appendix B : Descent from infinity, proof of Proposition 1.4.2.1 :

Lemma 1.4.2.4 implies Proposition 1.4.2.1 :

We obtain by induction and the Markov property : $\quad \forall\, n > 0, \quad \mathbb{P}_n(k\,t < \tau_\downarrow^D) \leq \epsilon^k.$
Thus, by choosing ϵ sufficiently small (for any given value of $t > 0$), we ensure :

$$C_\infty^N := \sup_{\{n>0\}}\left\{\mathbb{E}_n[\exp(\rho\tau_\downarrow^D)]\right\} < +\infty.$$

A fortiori with $T_\downarrow := \inf\{t, N_t \leq n_\infty\} \wedge \tau_E \leq \tau_\downarrow^D$, $\sup_{(x,n)}\left\{\mathbb{E}_{(x,n)}[\exp(\rho\,T_\downarrow)]\right\} \leq C_\infty^N < \infty.$

At time $T_\downarrow$, the process is either in E or in $\mathcal{T}_\infty^X$. Thus :

$$\begin{aligned}\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)] \le \mathbb{E}_{(x,n)}[\exp(\rho\,T_\downarrow)\,;\ (x,n)_{T_\downarrow} \in E] \\ + \mathbb{E}_{(x,n)}\Big[\exp(\rho\,T_\downarrow)\mathbb{E}_{(X,N)_{T_\downarrow}}[\exp(\rho\,\widehat{\tau}_E)]\,;\ (X,N)_{T_\downarrow} \in \mathcal{T}_\infty^X\Big],\end{aligned}$$

with the Markov property and the fact that $(\widehat{\tau}_E - \mathcal{T}_\infty^X)_+ \le t_h$ on the event $\{(X,N)_{T_\downarrow} \in \mathcal{T}_\infty^X\}$.

$$\text{Therefore :} \qquad \mathcal{E}_\infty^N \le C_\infty^N\,(1+\mathcal{E}_\infty^X) \qquad \square$$

Proof of Lemma 1.4.2.4 :
The proof of this Lemma relies mainly on the same arguments as in [BP12], part 6, related to the descent from infinity. Let $Z_t := \sigma/2 \times \sqrt{N_t^D}$. It is solution to the following EDS :

$$Z_t := z + \int_0^t \psi\,(Z_s)\ ds + B_t,\ \text{where } \psi(z) := -\frac{1}{2\,z} + \frac{r\,z}{2} - c\,z^3. \tag{1.6.2}$$

As long as Z is very large and $|B|$ not exceptionally large, the leading term $-c\,Z_t^3$ indeed makes the process comes down in finite time. Let $V := Z - B$. It is the solution of the ODE :

$$\frac{dV_t}{dt} = -\frac{1}{2(V_t+B_t)} + \frac{r\,(V_t+B_t)}{2} - c\,(V_t+B_t)^3, \tag{1.6.3}$$

$$\text{Let } z_2 \ge z_1 := \sup\left\{z>0,\ \left|-\frac{1}{2z}+\frac{rz}{2}\right| \ge \frac{cz^3}{2}\right\}, \tag{1.6.4}$$

$$T_B := \inf\{t>0, B_t \notin [-z_2\,,\ 2\,z_2]\}, \qquad T_V := \inf\{t>0, V_t < 2z_2\}, \tag{1.6.5}$$

where we consider w.l.o.g. an initial condition z strictly bigger than $2\,z_2$, so that T_V is positive a.s. Then, as in [BP12], we get on the time interval $[0, T_B \wedge T_V]$:

$$\begin{gathered}B_t \ge -z_2 \ge -V_t/2,\ \text{implying}\ \ V_t+B_t \ge V_t/2 \text{ and } V_t+B_t \ge z_2,\\ \left|-\frac{1}{2(V_t+B_t)} + \frac{r\,(V_t+B_t)}{2}\right| \le \frac{c}{2}(V_t+B_t)^3,\\ \frac{d}{dt}\left[(V_t)^{-2}\right] = \frac{-2}{(V_t)^3}\frac{dV}{dt} \ge 2\times\left(c-\frac{c}{2}\right)\times\left(\frac{V_t+B_t}{V_t}\right)^3 \ge \frac{c}{8},\\ \text{thus } V_t^{(-2)} - z^{(-2)} \ge c\,t/8 \quad \text{and in particular} \quad V_t \le \sqrt{8/(c\times t)}.\end{gathered}$$

Thus, $\{t \le T_B\} \subset \{T_V \le t\} \cup \left\{V_t \le \sqrt{8/(c\times t)}\right\}$.
By (1.6.5), let z_2 be sufficiently big to ensure : $\quad \mathbb{P}(T_B < t) \le \epsilon$.

Then, denote : $\quad z_\infty := \left(\sqrt{8/(c\times t)} + 2\, z_2\right) \vee (4\, z_2)$.

We deduce that, on the event $\{t \le T_B\}$, either $Z_t \le z_\infty$ or $T_V \le t$ while $Z_{T_V} \le 4 z_2 \le z_\infty$. In any case, $\tau_\downarrow^D \le t$. Hence : $\quad \forall\, z > 0, \quad \mathbb{P}_z(t < \tau_\downarrow^D) \le \epsilon$. $\square$

Appendix C : Mal-adaptation too large, proof of Proposition 1.4.2.2 :

Lemma 1.4.2.5 implies Proposition 1.4.2.2 Let ρ, ϵ, $n_\infty > 0$ ($c > 0$ is the same as for the definition of Z). For simplicity, we choose $t := \log(2)/\rho > 0$ (i.e. $\exp[\rho\, t] = 2$), and assume w.l.o.g. $t < t_h$. We choose $r_\vee \in \mathbb{R}$ according to Lemma 1.4.2.5 such that :

$$\begin{aligned}
&\forall\, n > 0,\ \forall\, r \le r_\vee, \quad \mathbb{P}_n\left(t < \tau_\partial^D\right) \le e^{-\rho\, t}/2 = 1/4,\\
&\forall\, r \le r_\vee, \quad \mathbb{P}_{n_\infty}\left(T_\infty^D \le t\right) + \mathbb{P}_{n_\infty}\left(N_t^D \ge n_\infty\right) \le \epsilon/4.
\end{aligned}$$

Since $\limsup_{\|x\|\to\infty} r(x) = -\infty$, with n_E chosen sufficiently large : $\forall\, x \notin B(0, n_E),\ r(x) \le r_\vee$.

Let (X, N) with initial condition $(x, n) \in \mathcal{T}_\infty^X$. In the following, we denote :

$$\begin{aligned}
&T_\infty^N := \inf\left\{t \ge 0,\ N_t \ge n_c\right\}, \quad \tau_0 := \inf\left\{t > 0,\ (X,\, N)_t \in \mathcal{T}_0\right\},\\
&T := t \wedge T_\infty^N \wedge \tau_0 \wedge \tau_E \wedge \tau_\partial.
\end{aligned} \tag{1.6.6}$$

Since, on the event $\{T = t\}$, either $N_t \ge n_\infty$ or $(X, Y)_t \in \mathcal{T}_\infty^X$:

$$\begin{aligned}
\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)] &= \mathbb{E}_{(x,n)}[\exp(\rho\, T)\,;\ T = \widehat{\tau}_E] + \mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)\,;\ T = \tau_0]\\
&\quad + \mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)\,;\ T = t] + \mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)\,;\ T = T_\infty^N]\\
&\le \exp(\rho\, t)\ (1 + \mathcal{E}_0) + \exp(\rho\, t)\, \mathbb{P}_{(x,n)}[T = t]\mathcal{E}_\infty^X\\
&\quad + \exp(\rho\, t)\ \left(\mathbb{P}_{(x,n)}[T = T_\infty^N] + \mathbb{P}_{(x,n)}[N_t \ge n_\infty,\ T = t]\right)\ \mathcal{E}_\infty^N,
\end{aligned}$$

by the Markov property. Now, by (1.6.6), N^D is an upper-bound of N before T. Thus, by our definitions of $t, n_E, r_\vee$:

$$\mathbb{E}_{(x,n)}[\exp(\rho\,\widehat{\tau}_E)] \le 2\ (1 + \mathcal{E}_0) + \mathcal{E}_\infty^X/2 + \epsilon\,\mathcal{E}_\infty^N/2.$$

Taking the supremum over $(x, n) \in \mathcal{T}_\infty^X$ in the last inequality yields :
$\mathcal{E}_\infty^X \le 4\ (1 + \mathcal{E}_0) + \epsilon\,\mathcal{E}_\infty^N$. $\square$

Proof of Lemma 1.4.2.5 : We recall our definition of Z and ψ in (1.6.2). In the following, we condider ψ as a function of r, thus the notation $\psi_r(z) := -1/(2\, z) + (rz)/2 - c\, z^3$.

Step 1 : $\sup_{z>0} \psi_r(z) \underset{r\to-\infty}{\longrightarrow} -\infty$

Let $A > 0$, $z_A := \frac{2}{A}$ and $r_\vee := -A^2$. Then :

$$\begin{aligned} &\forall\, z \leq z_A,\ \forall\, r \leq 0, && \psi_r(z) \leq -1/(2\, z_A) = -A,\\ &\forall\, z \geq z_A,\ \forall\, r \leq r_\vee \leq 0, && \psi_r(z) \leq r_\vee\, z_A = -2A. \end{aligned}$$

□

Step 2 : bound on $Z^A_t := z - A\,t + B_t$ for A large :
Let $\epsilon, t_D > 0$. We can choose $\Delta z > 0$ such that, with $N \sim \mathcal{N}(0,1)$:

$$\mathbb{P}\left(\sup_{\{t \leq t_D\}} B_t \geq \Delta z\right) = 2\,\mathbb{P}\left(N \geq \Delta z \,/\, \sqrt{t_D}\right) \leq \epsilon. \tag{1.6.7}$$

Then, we can choose $A > 0$ (sufficiently big) such that :

$$\mathbb{P}\left(B_{t_D} \geq A\, t_D\right) = \mathbb{P}\left(N \geq A\,\sqrt{t_D}\right) \leq\ \epsilon.$$

We also choose $r_\vee$ thanks to step 1 such that :

$$\forall\, r \geq r_\vee, \quad \sup_{z>0} \psi_r(z) \leq -A \leq 0.$$

We now assume that the initial condition of Z satisfies $Z \leq z_\infty$ ($z_\infty = \sigma\sqrt{n_\infty}/2$).
For any $z_E \geq z_\infty + \Delta z$ and $r \leq r_\vee$, we deduce :

$$\begin{aligned} &\sup_{\{t \leq t_D\}} Z_t \leq z_\infty + \sup_{\{t \leq t_D\}} B_t\\ &\mathbb{P}\left(\sup_{\{t \leq t_D\}} Z_t \geq z_E\right) \leq \mathbb{P}\left(\sup_{\{t \leq t_D\}} B_t \geq \Delta z\right) \leq \epsilon \quad \text{by (1.6.7)}\\ &\mathbb{P}\left(Z_{t_D} \geq z_\infty\right) \leq \mathbb{P}\left(B_{t_D} \geq A\, t_D\right) \leq \epsilon \quad \text{by our choice of A and } r_\vee.\\ &\text{Thus } \mathbb{P}\left(T^D_\infty \leq t_D\right) \leq \mathbb{P}\left(Z_{t_D} \geq z_\infty\right) + \mathbb{P}\left(\sup_{\{t \leq t_D\}} Z_t \geq z_E\right) \leq 2\,\epsilon, \end{aligned}$$

with $n_E = (2\, z_E/\sigma)^2$. It proves the second claim of the Lemma (up to a change of ϵ by $\epsilon/2$).

Step 3 : descent from infinity and extinction
Now, we need to assume $c > 0$. Let again $\epsilon, t_D > 0$. Thanks to Lemma 1.4.2.4 (for $r = 0$ since $\mathbb{P}(t_D < \tau_\downarrow)$ is decreasing with r) we choose $z_\downarrow > 0$ such that, with $\tau_\downarrow := \inf\{t \geq 0,\ Z_t \leq z_\downarrow\}$:

$$\forall\, r \leq 0,\ \forall\, z > 0, \quad \mathbb{P}_{z_\infty}(t_D < \tau_\downarrow) \leq \epsilon \tag{1.6.8}$$

Like in the previous step, we choose $A > 0$ such that :

$$\mathbb{P}\left(B_{t_D} \geq A\, t_D - z_\downarrow\right) \leq \epsilon.$$

Again, we choose $r_\vee$ thanks to step 1 such that : $\quad \forall\, r \leq r_\vee, \quad \sup_{z>0} \psi_r(z) \leq -A \leq 0.$

Then, with $r \leq r_\vee$, on the event $\{\tau_\downarrow \leq t_D\}$, conditionally on $Z_{\tau_\downarrow}$:

$$\mathbb{P}_{N^D_{\tau_\downarrow}}\left(2\,t_D - \tau_\downarrow < \widetilde{\tau_\partial}^D\right) \leq \mathbb{P}_{Z_{\tau_\downarrow}}\left(\widetilde{Z}_{t_D} > 0\right) \leq \mathbb{P}\left(z_\downarrow - A\,t_D + B_{t_D} > 0\right) \leq \epsilon, \tag{1.6.9}$$

by our choices of A and $r_\vee$. Finally, by the Markov property, for any $z > 0$:

$$\begin{aligned}\mathbb{P}_{z_\infty}\left(2\,t_D < \tau_\partial^D\right) &\leq \mathbb{P}_{z_\infty}\left(t_D < \tau_\downarrow\right) + \mathbb{E}_{z_\infty}\left[\mathbb{P}_{Z_{\tau_\downarrow}}\left(2\,t_D - \tau_\downarrow < \widetilde{\tau_\partial^D}\right)\;;\;\tau_\downarrow \leq t_D\right]\\ &\leq 2\,\epsilon \qquad \text{with (1.6.8), (1.6.9)}\end{aligned}$$

which proves the first claim of the Lemma (replace ϵ by $\epsilon/2$ in the proof and take $t_D = t/2$).

$\square$

Appendix D : Too few individuals, proof of Proposition 1.4.2.3 :

For $(x, n) \in \mathcal{T}_0$, with n_E sufficiently large, we would like to say that mortality is so strong in this area that it overcomes an exponential growth at rate ρ. In order to get an estimate of mortality in $\mathcal{T}_0$, we will use some coupling with branching processes and consider the process after a time $t_D = 1$ (arbitrary). In practice, we prove that for any $\rho, \epsilon' > 0$, there exists $C' \geq 1$ such that for any n_E sufficiently large :

$$\mathcal{E}_0 \leq C' + \epsilon'\left(\mathcal{E}^N_\infty + \mathcal{E}^X_\infty + \mathcal{E}_0\right). \tag{1.6.10}$$

By taking $\epsilon' = (\epsilon \wedge 1)/2$, $C = 2\,C'$, it clearly implies Proposition 1.4.2.3.

The equation $N^U_t = n_0 + \int_0^t r_+ N^U_s ds + \sigma \int_0^t \sqrt{N^U_s}\, dB^N_s$ defines an upper-bound of N on $[0, t_D]$ as soon as $n \leq n_0$, which is notably a branching process. The survival of (X, N) beyond t_D clearly implies the survival of N^U beyond t_D. Let us define ρ_0 by the relation : $\mathbb{P}_{n_0}\left(t_D < \tau^U_\partial\right) =: \exp(-\rho_0\; t_D)$. For a branching process like N^U, it is classical that : $\rho_0 \to \infty$ as $n_0 \to 0$. Indeed, with $u(t, \lambda)$ the Laplace exponent of N^U (cf e.g. [Pa16] Subsection 4.2, notably Lemma 5) : $\mathbb{P}_{n_0}\left(\tau^U_\partial \leq t_D\right) = \exp[-n_0 \lim_{\lambda \to \infty} u(t_D, \lambda)] \to 1$, as $n_0 \to 0$.

So we can impose that $\rho_0 > \rho$, and even that $\exp(-(\rho_0 - \rho)\,t_D)$ is sufficiently small to make transitions from $\mathcal{T}_0$ to $\mathcal{T}_0$, $\mathcal{T}_\infty^N$ or $\mathcal{T}_\infty^X$ of little incidence.

$$\begin{aligned}
\mathbb{E}_{(x,n)}[\exp(\rho\widehat{\tau}_E)] \le \mathbb{E}_{(x,n)}\Big[\exp(\rho\widehat{\tau}_E)\ ;\ \widehat{\tau}_E < t_D\Big]& \\
+ \mathbb{E}_{(x,n)}\Big[\exp(\rho\widehat{\tau}_E)\ ;\ (x,n)_{t_D} \in \mathcal{T}_0 \cup \mathcal{T}_\infty^N \cup \mathcal{T}_\infty^X\Big]& \\
\le \exp[\rho\, t_D] + \exp(\rho\, t_D)\ (\mathcal{E}_0 + \mathcal{E}_\infty^N + \mathcal{E}_\infty^X)\ \mathbb{P}_{(x,n)}(t_D < \tau_\partial)& \\
\le C' + \epsilon'\ (\mathcal{E}_0 + \mathcal{E}_\infty^N + \mathcal{E}_\infty^X), \qquad \text{where } C' := \exp[\rho\, t_D]& \\
\text{and } \epsilon' := \exp(-(\rho_0 - \rho)\, t_D) \to 0 \text{ as } n_0 \to 0.&
\end{aligned}$$

□

Appendix E : Proof of Lemma 1.5.3.7

Like in the proof of Lemma 1.5.3.6 with Lemma 1.5.3.7 :

$$\begin{aligned}
\mu A_{t_h} &= \frac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)}\ \mu A_{j\, t_D} \cdot \mathbb{P}_{t_h - j\, t_D} \\
&= \sum_{k=1}^{j} a(k, j\, t_D) \times \frac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \times \frac{\mathbb{P}_\zeta(t_h - k\, t_D < \tau_\partial)}{\mathbb{P}_\zeta([j-k]\, t_D < \tau_\partial)} \zeta A_{t_h - k\, t_D} \\
&\quad + r_j \times \frac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \times \mathbb{P}_{\nu_j}(t_h - j\, t_D < \tau_\partial)\ \nu_j A_{t_h - j\, t_D} \qquad (1.6.11)
\end{aligned}$$

Yet, by (1.5.19) : $\quad a(k, j\, t_D) \times \dfrac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \times \dfrac{\mathbb{P}_\zeta(t_h - k\, t_D < \tau_\partial)}{\mathbb{P}_\zeta([j-k]\, t_D < \tau_\partial)} = \dfrac{c_D}{c_P}\ \left(1 - \dfrac{c_D}{c_P}\right)^{k-1}$,
so that we obtain, by evaluating the measures in (1.6.11) on $\mathcal{X}$:

$$r_j \times \frac{\mathbb{P}_\mu(j\, t_D < \tau_\partial)}{\mathbb{P}_\mu(t_h < \tau_\partial)} \times \mathbb{P}_{\nu_j}(t_h - j\, t_D < \tau_\partial) = 1 - \sum_{k=1}^{j} \frac{c_D}{c_P}\ \left(1 - \frac{c_D}{c_P}\right)^{k-1} = \left(1 - {}^{c_D}/_{c_P}\ \right)^j$$

□

2 Exponential quasi-ergodicity for processes with discontinuous trajectories

This chapter is taken from the third version of the preprint with the same name whose ArXiv reference can be found at the end of the bibliography (here is the link for the pdf version : [Chapter 2]). After the last rejection by a prestigious journal of probability, I've come to realize that a publication to a journal of analysis could be better suited.

Adapting the proof for the criteria I introduce is presumably not considered so original or difficult. The interest of its contribution will hopefully emerge from the following publications where it plays a crucial role. On the contrary, the applications described in this chapter may already generate more interest to researcher studying persistence of populations and traveling waves with an analytic viewpoint. Namely, the applications introduced in the paper concern processes close to the ones these researchers have been looking at. Thus, I plan to change the organization of the paper to adapt to this different audience. The focus will be on the description of solutions to non-linear reaction-diffusion equations with a linear reaction term. Additional results on existence, uniqueness and regularity of such solutions will then be required (and would probably be presented in an additional preprint).

Abstract :

This paper establishes exponential convergence to a unique quasi-stationary distribution in the total variation norm for a very general class of strong Markov processes. Specifically, we can treat non-reversible processes with discontinuous trajectories, which seems to be a substantial breakthrough. Considering jumps driven by Poisson Point Processes in two different applications, we intend to illustrate the potential of these results and motivate our criteria. Our set of conditions is expected to be much easier to verify than an implied property which is crucial in our proof, namely a comparison of asymptotic extinction rates between different initial conditions.

2.1 Introduction

2.1.1 Presentation

This work is concerned with the long time behavior of a general class of strong Markov processes, conditionally upon the fact that the process has not been absorbed in some "cemetery state" (is not "extinct"). We are first concerned with the marginal distribution (at time t) conditioned upon the fact that extinction has not occurred (also at time t). For simplicity, we will call this distribution the **MCNE** at time t.

We present a new set of key conditions on the process such that the MCNE converges as $t \to \infty$ to a unique distribution α. This limiting distribution is called the **quasi-stationary distribution** (the **QSD**) (cf Subsections 2.1.3 and 2.2.2, or chapter 2 in [CMS13] for more details on this notion). The current paper extends the results that we proposed in Chapter 1, with the purpose of providing conditions as general and as easy to verify as possible. As in Chapter 1, we obtain not only the existence and uniqueness of the QSD, but also the exponential convergence in total variation distance.

This result implies that the probability that the process goes extinct after time t decays asymptotically at a given exponential rate λ. Compensating for this extinction rate, we define the relative extinction probability $h_t(x)$ at time t, starting from state x (cf. Theorem 2.2.2). We prove that this sequence of functions of x converges in the supremum norm to the unique positive eigenfunction h of the infinitesimal generator, as $t \to \infty$. As for the QSD, its eigenvalue is exactly the opposite of the extinction rate. This convergence motivates the name **survival capacity** which we give to h. It is connected to the notion of "reproductive value" in ecological models, and indicates likewise the most favorable traits for having a large progeny.

Moreover, we obtain the existence of the **Q-process**. Its marginal at time t is given by the limit (as $T \to \infty$) of the marginal of the original process at time t conditioned on not being extinct at time T, cf. Theorem 2.2.4. Thus, it can be described as the process conditioned to becoming extinct only in a far future. We also deduce the existence and uniqueness of its stationary distribution β.

Compared to Chapter 1, the convergence results towards the QSD α and towards β are more precise regarding the dependence upon the initial condition and now correspond to the notion of exponential quasi-ergodicity given in the introduction, cf Subsection 0.1.3.2. The multiplicative factor that delays the exponential decay (at a uniform rate) can be expressed simply in terms of the survival capacity of this initial condition. Moreover, we can take into account in this factor that the initial condition is close to α in total variation for the first case, and close to β in a weighted norm for the second.

In the continuation of Chapter 1, our motivation is to obtain uniqueness as well as quantitative results for the speed of convergence, while relying on conditions that are as directly related to the result as possible. The approach we follow appears to be well suited for this purpose. It extends the principles of Harris' recurrence, with the following additional condition on the asymptotic extinction rate :

$$\limsup_{t\to\infty} \sup_{x\in\mathcal{X}} \frac{\mathbb{P}_x(t<\tau_\partial)}{\mathbb{P}_\zeta(t<\tau_\partial)} < \infty, \tag{2.1.1}$$

where ζ is a reference measure related to small sets and $\mathcal{X}$ the whole state space. This approach has been first exploited by Champagnat and Villemonais in [CV16] and several developments have been proposed in the aftermath. In [CV16], the two highlighted conditions (including (2.1.1)) are proved to be not only sufficient for the uniform exponential convergence towards a unique QSD, but actually necessary. This asymptotic estimate is in fact crucial in any of the current proofs extending the results presented in [CV16]. Notably, we refer to [CV18a], [CV17], [CV17b] [CV18b], [BCG19] and [DV16]. It is not only implied by the conditions, but also a major step in the proofs.

In the extension of [CV16] that we have presented in Chapter 1, we could also simplify the proof of the estimate (2.1.1), by restricting the initial conditions x that one needs to consider. The local version of (2.1.1) replacing $\mathcal{X}$ by some set $\mathcal{D}_E$, typically a compact set, is easier to verify. It extends into (2.1.1) provided that the first hitting time of $\mathcal{D}_E$ has moments of a sufficiently large order. Similarly, such a local estimate seems to be sufficient in the recent results of [CV17c] and [BCGM19], at the expense of defining a proper Lyapunov function. Referring for instance to [CG16], there are strong links between the conditions on the Lyapunov function and specific moments of first entry time. We believe that the argument in Chapter 1 give the proper intuition to understand the results of [CV17c] and [BCGM19] on how to extend local estimates of (2.1.1). With our new assumption, we generalize in some sense this principle that some components of the trajectories (started from a test initial condition x) may be considered transitory. The waiting time for specific jumps to occur is an example of such transitory periods.

Methods for verifying the estimate (2.1.1) (or some local version) have been so far developed in the case of discrete state space on one side, and exploiting Harnack's inequality in the case of diffusion processes on the other side. Notably, an approach for treating the case of jumps in continuous space is clearly lacking. The main purpose of the current paper is to fill this gap.

Irreducibility is the major issue that we want to tackle in the current paper. Notably, a jump rate too small in specific directions creates difficulties for establishing our results. One has not only to deal with absorbing sets, but more generally with absorbing measures : these are measures ν such that the MCNE at any time t has a density w.r.t. ν as soon as the initial condition has a density w.r.t. ν. Given that we are looking for convergence results in total variation, the QSD necessarily has a density with respect to any absorbing measure. In the general class of processes that we cover, we want to allow the existence of such absorbing measures ν without requiring that any MCNE at positive time has a density w.r.t. ν. It means notably that we do not rely on the strong Feller property as in [FRS19] : there are typically bounded functions f for which, whatever t, $P_t f$ is not continuous (the semi-group P_t will be specified in subsection 2.1.3).

The difficulty occurs even for very elementary processes. For simplicity, let us consider a pure jump processes with a bounded jump rate and such that the law of the jump sizes has a density w.r.t. the Lebesgue measure. Although the Lebesgue measure is clearly absorbing, it is generally unclear, due to the conditioning, that, for Dirac initial condition, the mass of the MCNE at the starting point vanishes for t large enough. The decay of such a mass as $t \to \infty$ has however to be ensured for the convergence in total variation towards a unique QSD. As pointed out in [BCL17] or in [BCGL17], one may otherwise observe even a concentration of the MCNE to a Dirac mass as $t \to \infty$. Punctual persistence effects is only the simplest problematic case. So the difficulty is even more acute when the jumps do not have a density.

The current paper does not provide a way to prove such a decay, but specify the type of decay that we require to prevent such persistence effects. Such irreducibility issues have specifically to be addressed for processes whose trajectories involve discontinuities, notably jumps or a change in direction as in the Zig-Zag process (for which stationary properties have been recently obtained [BRZ19]). Similar difficulties can also arise for processes in discrete time.

General surveys like [CMS13], [DP13] or more specifically for population dynamics [MV12] give an overview on the huge literature dedicated to QSD (for which Pollett has collected quite an impressive bibliography, cf. [Po15]). Reversible processes represent a very specific class for which very strong results can be obtained, using the spectral decomposition of the operator. When jumps in continuous space are involved, such property is generally not to be expected. They are even more exceptional when the state space is multidimensional : looking at the Appendix A of [CCM17], one can see how strong would be the constraints on the jump rates.

We focus in the following discussion on the general context of continuous-time and continuous-space Markov processes, without any assumption of reversibility. Then, the most elementary argument to ensure the existence of a QSD has to do with the compactness of the semi-group, as in the Krein-Rutman theorem. In [CM+11], the authors manage to characterize any QSD as a fixed point of the normalized semi-group (on a properly chosen measure space), and exploit Tychonov's theorem. Another classical technique, called renewal, aims at approximating some QSD by stationary distributions of a sequence of processes without extinction (cf e.g. [FK+95], where specific jumps are introduced to replace the extinction events). To our knowledge, very few papers have looked at continuous-time and continuous-space processes that are not diffusions. It is worth mentioning [ANT80], which applies to pure jump processes, and extends the techniques of "R-theory" to the continuous-time setting (cf e.g. [Tw74a] or [Tw74b] for the R-theory). General conditions are presented to ensure the existence of a QSD, which need not be easy to verify, as it is usually the case in R-theory. Note in conclusion that all these techniques do not provide any insight into the issue of uniqueness (except possibly [ANT80]) nor into the speed of convergence.

Up to now, the only existing strategy for obtaining uniqueness and the speed of convergence is the one initiated in [CV16] and notably extended in [CV17c] [FRS19], [BCGM19] and in our Chapter 1, for cases where the convergence is not uniform over the initial condition. The aim of such techniques is specifically to establish the exponential convergence in total variation to the QSD. Note that in [CV17c], [FRS19] and [BCGM19], the authors deduce the convergence from any initial law with a sufficiently light tail to a specific QSD. Alternative QSDs with much heavier tails are expected to exist in such context.

In Section 2.3, we present two applications of our general theorems. We have chosen elementary models of PDMP and pure jump processes for the analysis to be simple and to exemplify our approach. To our knowledge, they have not been treated yet. Quantitative convergence results are even more unlikely to be known. The closest results are possibly in the very recent [BCGL17], where their model corresponds to our second application restricted to the case of dimension one. We hope that they will help the reader to get insight on our criterion and its adaptability. Applications of our theorems to more detailed biological models are intended to be done in future work. We refer notably to Chapter 3, where we couple a diffusive process specifying the population size to a piecewise deterministic process specifying the adaptation of the population. Chapter 5 provides another application for the analysis of metastability between the clicks of the "Muller ratchet", that was somewhat unexpected. In this model of an asexual population with constant size, it may be observed that natural selection effectively prevent the accumulation of deleterious mutations. This criterion makes it possible to deal with individuals carrying too large a number of mutations, especially in the diffusive large population limit where they never completely disappear. But already, the applications of this paper are related to the solutions of specific non-local reaction-diffusion equations, see Subsection 2.3.1.4.

More generally, these techniques provide conditions ensuring the existence and uniqueness of the positive eigenvector of general linear non-local reaction diffusion equations (see notably Subsection 2.3.1.4 for some partial results and Subsection 2.3.1.1 for the related conditions). The long-time behavior of structured branching processes can also be captured by such results, thanks to the many-to-one formula (see [HH09]). Likewise, we expect direct applications for the existence and uniqueness of equilibrium to growth-fragmentation models (as in [MMP05]).

Another work in preparation (with a preprint already in ArXiv, cf [Ve20]) is devoted to the use of these results to deduce Large Deviation estimates on the empirical measures of the process X. We are then looking at ensuring the limit and estimating the following quantity :

$$\lim_{t\to\infty} \frac{-1}{t} \log \mathbb{P}_\mu \left[\Psi_t \le \gamma \,\middle|\, t < \tau_\partial\right], \qquad \text{where } \Psi_t := \tfrac{1}{t}[\textstyle\int_0^t G(X_s)\,ds + \sum_{s\le t} F(X_{s-}, X_s)],$$

where G, F are bounded measurable, the latter vanishing on the diagonal.

Our proofs would be easy to adapt to processes in discrete-time, and our techniques seem to generalize naturally to time-inhomogeneous processes, cf. [CV18b], [BCG19], [DV16], and probably to semi-Markov processes, i.e. pure jump processes for which the waiting time between jumps is not necessarily exponential.

The paper is organized as follows. Subsection 2.1.2 describes our general notations ; Subsection 2.1.3 presents our specific setup of a Markov process with extinction ; and Subsection 2.1.4 the decomposition of the state space on which we base our assumptions. Subsection 2.2.1 presents the main set of conditions which we show to be sufficient for the exponential convergence to the QSD. Subsection 2.2.2 states the three main theorems of this paper, that extend those of Chapter 1, dealing respectively with the QSD, the survival capacity and the Q-process. Finally, in Section 2.4, we detail our proofs.

2.1.2 Elementary notations

In this paper, $k \ge 1$ means $k \in \mathbb{N}$ while $t \ge 0$ (resp. $c > 0$) should be understood as $t \in \mathbb{R}_+ := [0, \infty)$ (resp. $c \in \mathbb{R}_+^* := (0, \infty)$). We denote classical sets of integers by : $\mathbb{Z}_+ := \{0, 1, 2...\}$, $\mathbb{N} := \{1, 2, 3...\}$, $[\![m, n]\!] := \{m, m+1, ..., n-1, n\}$ (for $m \le n$), where the notation $:=$ makes explicit that we define some notation by this equality. For maxima and minima, we usually denote : $s \vee t := \max\{s, t\}$, $s \wedge t := \min\{s, t\}$.

Let $\Big(\Omega; (\mathcal{F}_t)_{t\ge0}; (X_t)_{t\ge0}; (P_t)_{t\ge0}; (\mathbb{P}_x)_{x\in\mathcal{X}\cup\partial}\Big)$ be a time homogeneous strong Markov process with cadlag paths on some Polish space $\mathcal{X} \cup \{\partial\}$ [see [RW00], Definition III.1.1], where $\partial \notin \mathcal{X}$ is an isolated point and the topology on cX is denoted by $\mathcal{B}$. We also assume that the filtration $(\mathcal{F}_t)_{t\ge0}$ is right-continuous and complete. We recall that $\mathbb{P}_x(X_0 = x) = 1$, P_t is the semi-group of the process satisfying the usual measurability assumptions and the Chapman-Kolmogorov equation. The hitting time (resp. the exit time out) of $\mathcal{D}$, for some domain $\mathcal{D} \subset \mathcal{X}$, will generally be denoted by $\tau_\mathcal{D}$ (resp. by $T_\mathcal{D}$). While dealing with the Markov property between different stopping times, we wish to clearly indicate with our notation that we introduce a copy of X (ie a process with the same semigroup P_t) independent of X given its initial condition. This copy (and the associated stopping times) is then denoted with a tilde ($\widetilde{X}$, $\widetilde{\tau_\partial}$, $\widetilde{T}_\mathcal{D}$ etc.). In the notation $\mathbb{P}_{X(\tau_E)}(t - \tau_E < \widetilde{\tau_\partial})$ for instance, τ_E and $X(\tau_E)$ refer to the initial process X while $\widetilde{\tau_\partial}$ refers to the copy $\widetilde{X}$.

2.1.3 The stochastic process with absorption

We consider a strong Markov processes absorbed at ∂ : the cemetery. More precisely, we assume that $X_s = \partial$ implies $X_t = \partial$ for all $t \geq s$. This implies that the extinction time : $\tau_\partial \;:= \inf\{t \geq 0 \;;\; X_t = \partial\}$ is a stopping time. Thus, the family $(P_t)_{t\geq 0}$ defines a non-conservative semigroup of operators on the set $\mathcal{B}_+(\mathcal{X})$ (resp. $\mathcal{B}_b(\mathcal{X})$) of positive (resp. bounded) $(\mathcal{X}, \mathcal{B})$-measurable real-valued functions. For any probability measure μ on $\mathcal{X}$, that is $\mu \in \mathcal{M}_1(\mathcal{X})$, and $f \in \mathcal{B}_+(\mathcal{X})$ (or $f \in \mathcal{B}_b(\mathcal{X})$) we use the notations : $\mathbb{P}_\mu(.) \;:= \int_\mathcal{X} P_x(.)\, \mu(dx)$, $\langle \mu \,\big|\, f\rangle \;:= \int_\mathcal{X} f(x)\, \mu(dx)$.

We denote by $\mathbb{E}_x$ (resp. $\mathbb{E}_\mu$) the expectation according to $\mathbb{P}_x$ (resp. $\mathbb{P}_\mu$).

$$\mu P_t(dy) \;:= \mathbb{P}_\mu(X_t \in dy), \qquad \langle \mu P_t \,\big|\, f\rangle \;=\; \langle \mu \,\big|\, P_t f\rangle \;= \mathbb{E}_\mu[f(X_t)],$$
$$\mu A_t(dy) \;:= \mathbb{P}_\mu(X_t \in dy \,\big|\, t < \tau_\partial), \qquad \langle \mu A_t \,\big|\, f\rangle \;= \mathbb{E}_\mu[f(X_t) \,\big|\, t < \tau_\partial],$$

μA_t is what we called the MCNE at time t, with initial distribution μ.

In this setting, the family $(P_t)_{t\geq 0}$ (resp. $(A_t)_{t\geq 0}$) defines a linear but non-conservative semigroup (resp. a conservative but non-linear semigroup) of operators on $\mathcal{M}_1(\mathcal{X})$ endowed with the total variation norm :

$\|\mu\|_{TV} \;:= \sup\{|\mu(A)| \;;\; A \in \mathcal{B}\}$ for $\mu \in \mathcal{M}(\mathcal{X})$.

A probability measure α is said to be the *quasi-limiting distribution* of an initial condition μ if :

$\forall\, B \in \mathcal{B}, \quad \lim_{t\to\infty} \mathbb{P}_\mu(X_t \in B \,\big|\, t < \tau_\partial) := \lim_{t\to\infty} \mu A_t(B) = \alpha(B).$

It is now classical (cf e.g. Proposition 1 in [MV12]) that α is then a quasi-stationary distribution or QSD, in the sense that : $\forall\, t \geq 0, \quad \alpha A_t(dy) = \alpha(dy)$ and that for any such QSD, there exists an extinction rate λ such that : $\forall\, t \geq 0,\ \Pr_\alpha(t < \tau_\partial) = \exp[-\lambda\, t]$.

Our first purpose will be to prove that the assumptions in Subsection 2.2.1 provide sufficient conditions for the existence of a unique quasi-limiting distribution α, independent of the initial condition.

2.1.4 Specification on the state space

In the following Theorems, we will always assume :

$(A0)$: **"Exhaustion of $\mathcal{X}$"**

There exists a sequence $(\mathcal{D}_\ell)_{\ell\geq 1}$ of closed subsets of $\mathcal{X}$ such that (with $int(\mathcal{D})$ the interior of $\mathcal{D}$) :

$$\forall\, \ell \geq 1,\ \mathcal{D}_\ell \subset int(\mathcal{D}_{\ell+1}) \quad \text{and} \quad \textstyle\bigcup_{\ell\geq 1} \mathcal{D}_\ell = \mathcal{X}. \tag{A0}$$

This sequence will serve as a reference for the following statements. In particular, we will have control on the process through the fact that the initial distribution belongs to some set of the form :

$$\mathcal{M}_{\ell,\xi} \;:= \{\mu \in \mathcal{M}_1(\mathcal{X}) \;;\; \mu(\mathcal{D}_n) \geq \xi\}, \qquad \text{with } \xi \in (0,1]. \tag{2.1.2}$$

$$\text{Note that :} \qquad \forall \xi \in (0,1], \;, \quad \mathcal{M}_1(\mathcal{X}) = \bigcup_{\ell \geq 1} \mathcal{M}_{\ell,\xi}.$$

$$\text{Let also : } \mathbf{D} \;:= \{\mathcal{D} \;;\; \mathcal{D} \text{ is compact and } \exists \ell \geq 1,\; \mathcal{D} \subset \mathcal{D}_\ell\}. \tag{2.1.3}$$

2.2 Exponential convergence to the QSD

2.2.1 Hypotheses

Like in Chapter 1, our results rely on a set (**AF**) of assumptions. The label of the assumption is chosen to be consistent with the ones in Chapter 1. We detail first each basic assumption and then define (**AF**) in terms of those.

2.2.1.1 The assumptions common with Chapter 1

We recall the following definitions for the exit and first entry times of any set $\mathcal{D}$:

$$T_{\mathcal{D}} \;:= \inf\{t \geq 0 \;;\; X_t \notin \mathcal{D}\}, \qquad \tau_{\mathcal{D}} \;:= \inf\{t \geq 0 \;;\; X_t \in \mathcal{D}\}.$$

$(A1)$: **"Mixing property"**
There exists a probability measure $\zeta \in \mathcal{M}_1(\mathcal{X})$ such that, for any $\ell \geq 1$, there exists $L > \ell$ and $c, t > 0$ such that :

$$\forall x \in \mathcal{D}_\ell, \qquad \mathbb{P}_x[X_t \in dx \;;\; t < \tau_\partial \wedge T_{\mathcal{D}_L}] \geq c\,\zeta(dx).$$

$(A2)$: **"Escape from the Transitory domain"**
For a given $\rho > 0$ and $E \in \mathbf{D}$:

$$\sup_{\{x\in\mathcal{X}\}} \mathbb{E}_x(\exp[\rho(\tau_\partial \wedge \tau_E)]) < \infty.$$

ρ in the previous moment is required to be strictly larger than the following **"survival estimate"** :

$$\rho_S := \sup\left\{\gamma \geq 0 \,\middle|\, \sup_{L\geq 1} \liminf_{t>0}\; e^{\gamma t}\, \mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) = 0\right\} \vee 0. \tag{2.2.1}$$

2.2.1.2 A coupling approach including failures to ensure (2.1.1)

We now state the new condition that replace $(A3)$. As can be seen from our proof, its purpose is to imply $(A3)$, see Subsection 2.4.1.1. We also defer the discussion on this assumption to Subsection 2.2.3, where we will explain the origin of its title.

$(A3_F)$: **"Absorption with failures"**
Given $\zeta \in \mathcal{M}_1(\mathcal{X})$, $\rho > \rho_S$ and $E \in \mathbf{D}$, for any $\epsilon \in (0,1)$, there exist $t, c > 0$ such that for any $x \in E$ there exists a stopping time U_A such that :

$$\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\} \quad \text{and} \quad \mathbb{P}_x(U_A = \infty,\, t < \tau_\partial) \leq \epsilon\, \exp(-\rho\, t),$$

while for some stopping time V :

$$\mathbb{P}_x\Big(X(U_A) \in dx' \;;\; U_A < \tau_\partial\Big) \leq c\,\mathbb{P}_\zeta\Big(X(V) \in dx' \;;\; V < \tau_\partial\Big).$$

We further require that there exists a stopping time U_A^∞ extending U_A in the following sense :
- *$U_A^\infty := U_A$ on the event $\{\tau_\partial \wedge U_A < \tau_E^t\}$, where $\tau_E^t := \inf\{s \geq t : X_s \in E\}$.*
- *On the event $\{\tau_E^t \leq \tau_\partial \wedge U_A\}$ and conditionally on $\mathcal{F}_{\tau_E^t}$, the law of $U_A^\infty - \tau_E^t$ coincides with the one of $\widetilde{U}_A^\infty$ for a realization $\widetilde{X}$ of the Markov process $(X_t, t \geq 0)$ with initial condition $\widetilde{X}_0 := X(\tau_E^t)$ and independent of X conditionally on $X(\tau_E^t)$.*

2.2.1.3 The general sets of Assumptions

We say that assumption $(\mathbf{AF})$ holds, whenever :
"$(A1)$ holds for some $\zeta \in \mathcal{M}_1(\mathcal{X})$ and a sequence $(\mathcal{D}_\ell)$ satisfying $(A0)$. Moreover, there exist $\rho > \rho_S$ and $E \in \mathbf{D}$ such that assumptions $(A2)$ and $(A3_F)$ hold."

Remark : Almost sure extinction is not at all needed for our proof (which would in fact include the case where there is no extinction, or only in some "transitory domain").

2.2.2 Main Theorems

Theorem 2.2.1. *Assume that $(\mathbf{AF})$ holds. Then, there exists a unique QSD α. Moreover, we have exponential convergence to α of the MCNE's.*

Our strongest result of convergence depends on the survival capacity, defined next. We recall that α is associated to some extinction rate λ :

$$\forall\, t \geq 0,\ \mathbb{P}_\alpha(t < \tau_\partial) = e^{-\lambda\, t} \quad \text{so that as a QSD, } \alpha \text{ satifies } \alpha P_t = e^{-\lambda\, t}\,\alpha \tag{2.2.2}$$

- cf e.g. Theorem 2.2 in [CMS13]. This motivates the definition :

$$h_t(x) \;:= e^{\lambda\, t}\mathbb{P}_x(t < \tau_\partial) = \mathbb{P}_x(t < \tau_\partial)/\mathbb{P}_\alpha(t < \tau_\partial).$$

Theorem 2.2.2. *Under $(\mathbf{AF})$, $(h_t)_{t \geq 0}$ converges at exponential rate in the supremum norm to a positive bounded function h, vanishing on ∂. h belongs to the domain of the* *infinitesimal generator $\mathcal{L}$, associated with the semi-group $(P_t)_{t\geq 0}$ on $(B(\mathcal{X}\cup\{\partial\}); \|.\|_\infty)$, and satisfies :*

$$\mathcal{L}\,h = -\lambda\,h, \qquad so \quad \forall\,t \geq 0,\; P_t\,h = e^{-\lambda\,t}h. \tag{2.2.3}$$

Remark : As in [CV16], it is not difficult to show that there is no eigenvalue of $\mathcal{L}$ between 0 and $-\lambda$, and that h is the unique eigenvector associated to $-\lambda$.

Theorem 2.2.3. *Under* (**AF**)*, there exists $\gamma, C > 0$ such that for any $\mu \in \mathcal{M}_1(\mathcal{X})$ and $t \geq 0$:*

$$\|\,\mathbb{P}_\mu\,[\,X_t \in dx \mid t < \tau_\partial] - \alpha(dx)\,\|_{TV} \leq C\frac{\inf_{u>0}\|\mu - u\alpha\|_{TV}}{\langle\mu\,\big|\,h\rangle}\; e^{-\gamma\,t}. \tag{2.2.4}$$

Theorem 2.2.4. *Under again the same assumptions* (**AF**)*, we have :*

(i) **Existence of the Q-process :**

There exists a family $(\mathbb{Q}_x)_{x\in\mathcal{X}}$ of probability measures on Ω defined by :

$$\lim_{t\to\infty}\mathbb{P}_x(\Lambda_s\,\big|\,t < \tau_\partial) = \mathbb{Q}_x(\Lambda_s), \tag{2.2.5}$$

for all $\mathcal{F}_s$-measurable set Λ_s. The process $(\Omega; (\mathcal{F}_t)_{t\geq 0}; (X_t)_{t\geq 0}; (\mathbb{Q}_x)_{x\in\mathcal{X}})$ is an $\mathcal{X}$-valued homogeneous strong Markov process.

(ii) Transition kernel :

The transition kernel of the Markov process X under $(\mathbb{Q}_x)_{x\in\mathcal{X}}$ is given by :

$$q(x;t;dy) = e^{\lambda\,t}\,h(y)\;p(x;t;dy)/h(x), \tag{2.2.6}$$

where $p(x;t;dy)$ is the transition kernel of the Markov process X under $(P_x)_{x\in\mathcal{X}}$.

(iii) Exponential ergodicity of the Q-process :

The measure $\beta(dx) := h(x)\,\alpha(dx)$ is the unique invariant probability measure under $\mathbb{Q}$.

Moreover, for any $\mu \in \mathcal{M}_1(\mathcal{X})$ satisfying $\langle\mu\,\big|\,1/h\rangle < \infty$ and $t \geq 0$:

$$\|\mathbb{Q}_\mu\,[\,X_t \in dx] - \beta(dx)\|_{TV} \leq C\,\inf_{u>0}\|\mu - u\,\beta\|_{1/h}\,e^{-\gamma\,t}, \tag{2.2.7}$$

$$\text{where } \mathbb{Q}_\mu(dw) \;:= \textstyle\int_{\mathcal{X}}\mu(dx)\,\mathbb{Q}_x(dw), \quad \|\mu\|_{1/h} := \|\dfrac{\mu(dx)}{h(x)}\|_{TV}$$

Remark : • We refer to Chapter 1 for other equivalent statements on the transition kernel.

For simplicity, we will say that the exponential quasi-stationarity holds when the results of these three theorems hold (see [BR99] for more details on this notion of quasi-ergodic measure).

• The constants involved in the convergences are explicitly related to the ones in Assumption (**AF**). Although the precise relation is very intricate, it implies the following theorem of approximation :

Theorem 2.2.5. *Assume that* (**AF**) *holds, all the preceding results hold true with the same constants involved in the convergences when τ_∂ is replaced by $\tau_\partial^L := \tau_\partial \wedge T_{\mathcal{D}_L}$ for any $L \geq 1$ sufficiently large. The associated QSD α^L, extinction rates λ_L and survival capacities h^L converge respectively to α (in total variation), to λ and to h (punctually at least). Also, we deduce $\rho_S = \lambda$.*

Remark : Theorem 2.2.5 also holds when one assumes (**A**) of Chapter 1 instead of (**AF**).

2.2.3 Refined versions for the coupling approach including failures

Because our last assumption may appear too technical, let us give some insight on where it comes from. Recall that the last assumption needed in Chapter 1 is the relaxed version of (2.1.1) :

$(A3)$: *There exists $c, u > 0$ such that :*

$$\forall\, x \in E,\ \forall\, t \geq u, \quad \mathbb{P}_x(t < \tau_\partial) \leq c\, \mathbb{P}_\zeta(t < \tau_\partial),$$

where ζ is the measure involved in $(A1)$.

Notably, rather than the uniformity on $\mathcal{X}$ as in inequality (2.1.1), the uniformity on E is sufficient thanks to assumption $(A2)$. Our approach actually consists in proving the property $(A3)$, given (**AF**), and this is the only purpose of assumption $(A3_F)$. The results presented in previous subsection would still hold if one replaces assumption $(A3_F)$ in (**AF**) by assumption $(A3)$ (with the same restrictions on E).

One way to verify $(A3)$ is to get the process with initial condition ζ to come close enough to any test initial condition x with a given time-shift (and a given efficiency). Then, any set of long-time surviving trajectories of the process with initial condition x can in some sense be coupled to behave like some of the long-time surviving trajectories of the process with initial condition ζ. Such containment property is what we call "absorption" in the following.

Generally, it means that, for an absorption in one step to be deduced (with efficiency parameter $1/c$), we need to associate to any x two stopping times U_A and V such that U_A is uniformly bounded and :

$$\mathbb{P}_x\Big(X_{U_A} \in dx \ ;\ U_A < \tau_\partial\Big) \leq c\, \mathbb{P}_\zeta\Big(X_V \in dx \ ;\ V < \tau_\partial\Big). \tag{2.2.8}$$

Originally for continuous-time and continuous-space Markov processes, this property is deduced from the local Harnack inequality (see [CV17c] or Chapter 1). Another approach has been introduced in [DV16] in the context of a time-periodic diffusion, with a more explicit coupling.

Yet, it might be very difficult to ensure in practice such a strong property as (2.2.8). We might need for instance in some multidimensional model to wait for a jump on a specific dimension to happen, while there may be positive probability for very singular behavior to happen meanwhile on the other dimensions. This issue is in some sense avoided in the simple application of Subsection 2.3.2, thanks to our extension $(A3_F)$. In general, the purpose of such an extension is to handle singularities, that is problematic events which are rare in probability. When they happen, we say that the attempt to obtain absorption as in (2.2.8) has failed. For instance, if the process is already diffusive between jumps, it may be interesting to consider the first jump as the time of the first failure, and exploit the Harnack inequality in a time-scale so short that jumps happen at that time-scale with a small enough probability.

Although such a failure is rare in the short term, the survival itself becomes even more rare in the long term. New attempts will thus be needed while looking at large time asymptotic, until one happens to succeed in an absorption. The constants in $(A3_F)$ are chosen such that the "cost" of waiting for a new attempt is covered (in terms of probability) by the exceptionality of the previous failure.

In the second ArXiv version associated to this [Chapter 2], one can find a refinement of assumption $(A3_F)$ where U_A is more loosely bounded. The proof of how to include $(A3_F)$ is then deduced from its refined version. It gives more insight on the role of the constants involved. Yet, it also requires to specify the times at which failures are stated, since there is no more reason for each step to end before time t. The statement is thus much more technical. And in practice, by reducing ρ (while keeping $\rho' > \rho_S$), one can always truncate the value of U_A and deal with the event that U_A takes large value as exceptional.

While considering the conditions provided in [CV17c] (as an alternative to the ones in Chapter 1), the reader may wonder if our new assumption $(A3_F)$ could also replace their assumption $(F3)$. For our argument to hold, we need however to consider only confined trajectories of the process (our conditions involving $T_{\mathcal{D}_L}$), be it for our survival estimate (2.2.1) or our mixing assumption $(A1)$. While any constructive proof of $(F1)$ in [CV17c] certainly ensures in fact $(A1)$, it is clearly not as clear for the survival estimate, corresponding to the $\gamma_2 > 0$ involved in their condition $(F2)$. But still, our criterion may also be improved for cases where the QSD is not unique, in which case we need more precision on the cost of "failed attempts". And to estimate such cost, the best way would probably be to express it in term of a Lyapunov function like the ψ_1 given in their $(F2)$.

We conclude by the following remarks on the formulation of $(A3_F)$:

⋆ With the above notations : $\{U_A < \tau_\partial\} = \{U_A < t\} = \{U_A = \infty\}^c$.

$\star$ The r.v. U_A and V are expected to depend upon ϵ and x, and to be related to t and c, while these two constants must be uniform in x.

$\star$ The control on the failures is an upper-bound on their probability via ϵ that takes into account the time we waited hoping for a success. In fact, we will exploit this assumption for a given value of $\epsilon > 0$, which is explicitly but not so simply related to the other parameters (cf Subsection 2.4.1.1).

$\star$ The condition of existence of such U_A^∞ comes from the fact that we want to apply the property inductively (on the number of failed attempts) It has mainly to do with the regularity of the definition of U_A with respect to the initial condition x. Like in our applications, the definition of U_A^∞ shall come quite naturally from the conditions required by U_A and the proof that both are stopping times should be similar.

$\star$ Although the law of U_A^∞ is defined uniquely (which is what we need), it is a priori unclear how to define it generally. If U_A is defined through strong Markov processes with independent increments like Brownian Motions, Poisson Point Processes or other Levy processes, U_A^∞ can be expressed through these increments in the time-intervals $[\tau_E^i, \tau_E^{i+1}]$, for $i \in [\![0, \infty[\![$, where recursively :

$$\tau_E^{i+1} := \inf\{s \geq \tau_E^i + t : X_s \in E\} \wedge \tau_\partial, \text{ and } \tau_E^{(0)} = 0. \tag{2.2.9}$$

The Markov property on the incremental process shall then imply the condition on $U_A^\infty - \tau_E^t$.

2.3 Two models to which our results apply

2.3.1 Mutations compensating a drift leading to extinction

2.3.1.1 Description of the process

We consider a process X taking values in $\mathbb{R}^d$ and governed by a Piecewise-Deterministic Markov Process. The mixing assumption $(A1)$ shall hold true as a consequence of the jumps. By having a death rate going to infinity for large values of X, the confinement property $(A2)$ shall hold whatever $\rho > 0$. In order to prove $(A3_F)$, V shall necessarily be larger than the first jump time T_J. In this simple example, choosing $V = T_J$, provided T_J is not too large and the jump not too large, will be sufficient. Given ρ, E and ϵ, we can adjust the two associated thresholds so that most realizations are included. Note that T_J is bounded here by the fact that the flow (without jumps) drives the process to areas with a large killing rate.

For simplicity, the drift is assumed to be constant, which leads us to the following system :

$$X_t = x - v\,t\,e_1 + \int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} w\,\mathbf{1}_{\left\{u\le g(X_{s^-},w)\right\}}\,M(ds,dw,du) \tag{2.3.1}$$

as long as $t < \tau_d$. Here, M is a Poisson Point Process over $\mathbb{R}_+\times\mathbb{R}\times\mathbb{R}_+$, with intensity $\pi(ds,dw,du) = ds\;dw\;du$, while $g(x,w)$ describes the jump rate from x to $x+w$. Between jumps, the process is translated at constant speed $v>0$ along the first coordinate (i.e. along e_1). Moreover, the extinction occurs at time τ_∂ with the rate $\rho_e(x), x\in\mathbb{R}^d$, i.e. :

$$\tau_\partial := \inf\{t\ge 0\ ;\ \textstyle\int_0^t\int_0^{\rho_e(X_s)} M^E(ds,du)\ge 1\},$$

where M^E is again a PPP, with intensity $ds\;du$, independent of M.

2.3.1.2 Biological motivations

This process is to be seen as a model for the adaptation of a population to a changing environment. In this view, v corresponds here to the translation speed of the fitness landscape due to the environmental change.

We can consider a first description where X is a summary of the individual characters of the population. Then, the jumps come from the fixation of new mutations in the population, whose rate depends on the adaptation of the mutant subpopulation (with trait $X_{t-}+w$) as compared to the resident individuals (with trait X_{t-}). A much more detailed description is given in Chapter 3. There, we extend this proof to a coupled process involving continuous fluctuations of the population size.

As discussed in next Subsection, this model can also represent the specific adaptation of a uniformly chosen individual in the population. In this case, h simply describes the mutation rate of the individual from character x to $x+w$. Although the formalism is similar, the shape of h will be much different between these models. For instance, contrary to the first case, h can reasonably be assumed to be symmetrical in w in this second description.

2.3.1.3 Quasi-stationarity results

— Assumption (**PD**) (for Piecewise-Deterministic) :

(PD^1) ρ_e is locally bounded and $\lim_{\|x\|\to\infty}\rho_e(x) = +\infty$.
Also, explosion implies extinction : $\tau_\partial \le \sup_{\{\ell\ge 1\}} T_{\mathcal{D}_\ell}$.

h is measurable and for any compact $K\subset\mathbb{R}^d$, there exists :

(PD^2) an upper-bound $\rho_J^\vee$ on the jump-rate :
$\forall\, x\in K,\quad 0<\rho_J(x) := \int_{\mathbb{R}^d} g(x,w)\,dw \le \rho_J^\vee$.

(PD^3) a lower-bound $g_\wedge > 0$ on jumps able to compensate the drift, with some $0 < \delta S < S$:
$\forall\, x \in K,\ \forall\, w \in B(S\, e_1, \delta S), \quad g(x, w) \geq g_\wedge.$

(PD^4) *a tightness estimate for the jumps.* For any $\epsilon > 0$, there exists $w_\vee$ such that :
$\forall\, x \in K, \quad \int_{\mathbb{R}^d} g(x, w)/\rho_J(x)\, \mathbf{1}_{\{\|w\| \geq w_\vee\}}\, dw \leq \epsilon.$

(PD^5) an upper-bound $g_\vee$ on the density for each jump :
$\forall\, x \in K,\ \forall\, w \in \mathbb{R}^d, \quad g(x, w) \leq g_\vee\, \rho_J(x).$

Theorem 2.3.1. *Consider the process X given by equation (2.3.1) under Assumption (PD). Then, with $\mathcal{D}_\ell := \bar{B}(0, \ell), \ell \geq 1$,* (**AF**) *holds, thus also the exponential quasi-stationarity and the approximation property.*

Remark : *We do not need to assume any comparison between v and h. The extinction rate of the QSD can thus be very large, notably for large values of v.*

2.3.1.4 Connection with reaction-diffusion equations

In the following discussion, what I present is still merely conjectural. It is supported by similar results coming from analytic results in situations probably simpler to handle. Another project in preparation is to have a clear proof of these results, notably for the existence, uniqueness and regularity of the solutions to the following evolution equations.

This quasi-stationary regime is expected be related to the behavior of the solution $(u(t, x))_{t \geq 0, x \in \mathbb{R}^d}$ at low densities ($u \approx 0$) to reaction-diffusion evolution equations of the form :

$$\begin{aligned}\partial_t u(t, x) :=& v\, \partial_{x_1} u(t, x) + \int_{\mathbb{R}^d} g(y, x - y)\, u(t, y)\, dy - \left(\int_{\mathbb{R}^d} g(x, w)\, dw\right) u(t, x) \\ &+ r(x, u(t, x))\, u(t, x).\end{aligned} \tag{2.3.2}$$

At low density, the growth rate $r(x, u)$ can be approximated by $r_0(x) := r(x, 0)$, so as to linearize (2.3.2). The invasion at low density is presumably associated to the principal eigenvalue of the operator :

$$\mathcal{L}^\star f(x) := v\, \partial_{x_1} f(x) + \int_{\mathbb{R}^d} g(y, x - y)\, f(y)\, dy - [\int_{\mathbb{R}^d} g(x, w)\, dw]\, f(x) + r_0(x)\, f(x),$$

since results of this kind are obtained for instance in [BHR05], [CDM08].

If we consider $r_0(x)$ rather as some upper-bound of $r(x)$, the solution $\bar{u}$ to $\partial_t \bar{u}(t, x) = \mathcal{L}^\star \bar{u}(t, x)$. It shall provide an upper-bound of u by maximum principle approaches. If the eigenvalue $\lambda^\star$ of $\mathcal{L}^\star$ is negative, the solution u is expected to asymptotically decline at least quicker than at rate $-\lambda^\star$, provided that the initial condition is sufficiently confined (we may expect weaker conditions than compactly supported). Thus, results such as our have implications as a criteria for non-persistence, as in [BCV16], [BD+09], [BHR05], [BHR08], [BHR09], [CC98].

Finally, eigenvalue problems may arise while looking at large populations at equilibrium. If the competition exerted on the individuals is uniform, the distribution of individual characters shall be at equilibrium with the QSD. The competition pressure should then exactly compensate the growth rate of this QSD.

Assuming that r_0 is bounded by R, we can then define $\rho_e(x) := R - r_0(x)$. From Theorem 2.3.1, Assumption (PD) implies the existence of a unique QSD $\alpha(dx)$ and survival capacity $h(x)$ which can be seen as the eigenvector of respectively $\mathcal{L}^\star$ and $\mathcal{L}$, associated to the eigenvalue $R - \lambda$ (where λ is the asymptotic extinction rate with the local one being ρ_e). Moreover, we expect that, as long as $r(x, u)$ is well-approximated by r_0, the solution $u(t, x)$ should stabilize at an exponential rate quicker than γ towards the profile :

$$u(t, x)\, dx \sim \exp[(R - \lambda)\, t] \left[\int_{\mathbb{R}} u(0, y)\, h(y)\, dy\right] \alpha(dx).$$

This corresponds to the view that the individuals do not interact at the limit of low density, while they are sufficiently numerous for u to grow (or shrink) driven by the law of a typical individual (taking reproduction and death events into account). At very low density of individuals, demographic fluctuations cannot be neglected. Referring to studies on branching processes (e.g. a review on branching processes for biology in [KA06]), we still expect that, conditionally upon the survival of the population, the population size tends to grow at rate $R - \lambda$ with a profile of individuals stabilizing to α. This is stated in the so-called Kesten-Sigmund theorem for multi-type branching processes in [KL+97]. There is a scaling that depends both on the initial condition and on the randomness in the first events of reproduction and displacements.

Note that it is not crucial for the proof to assume as strongly as $\rho(x) \to \infty$, i.e. $r_0(x) \to -\infty$, as $x \to \infty$. It suffices that $-r_0$ is strictly larger than ρ_S outside of some compact set.

Several authors are looking at characterizing such eigenvalue when there is possibly no regular eigenvector (see e.g. [C10], [CLW17], [GR09], [IRS12], [CH20]). We hope to justify in a future work that, as soon as Theorem 2.3.1 applies, they all coincide to $R - \lambda$. Considering for instance each of the definitions presented in [CH20], the inequality on one side is generally easily derived from the fact that h is an eigenvector of $\mathcal{L}$. The other inequality is more challenging, since it shall require to approximate $\alpha(dx)$ by some $\varphi(x)\, dx$, where φ is C^1 and a sub-solution (or a super-solution) of $\mathcal{L}^\star$.

2.3.1.5 Proof of Theorem 2.3.1

$(A0)$ is clear. Since the extinction rate outside of $\mathcal{D}_\ell$ tends to infinity while $\ell \to \infty$, for any $\rho > 0$, we can find some $\mathcal{D}_\ell$ for which assumption $(A2)$ is clear (cf. Subsection 1.4.1.2 in Chapter 1). The proof of $(A1)$ is deduced from the following lemma, whose proof is deferred to the end of this subsection :

Lemma 2.3.1.1. *Under $(PD^{1,2,3})$, for any $\ell \geq 1$, with $L := \ell + 1$, there exists $c, t > 0$ such that :*

$$\forall\, x \in \mathcal{D}_\ell, \qquad \mathbb{P}_x\left[X_t \in dx \;;\; t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\, \mathbf{1}_{\mathcal{D}_\ell}(dx).$$

In particular, Assumption $(A1)$ holds with ζ uniform over $\mathcal{D}_1$.

Remark : 1) The main idea is that we are able to compensate the drift with the jumps and introduce a bit of fluctuations at each step, and thus we are able to ensure a dispersion to any part of the space (while keeping a bound on the norm).

2) Besides $(A1)$, we need a lower-bounded density on any $\mathcal{D}_\ell$ in order to prove assumption $(A3_F)$ (with a reaching time and efficiency of course dependent on this set).

Proof of Assumption $(A3_F)$ Consider a compact set $E = \bar{B}(0, \ell_E)$ and a test initial condition $x \in E$. We also assume that $\epsilon, \rho_S > 0$ are given. W.l.o.g. we assume that $\forall\, y \notin E, \quad , \rho_e(y) \geq \rho_E > \rho_S$. We first define $t_{\barwedge}$ by the relation :

$$\exp[\hat{\rho}_S \times (2\; \ell_E / v) - (\rho_E - \hat{\rho}_S) \times (t_{\barwedge} - 2\; \ell_E / v)] = \epsilon/2.$$

The l.h.s. is decreasing and converges to 0 when $t_{\barwedge} \to \infty$, so that $t_{\barwedge}$ is well-defined. Let T_J be the first jump time of X. On the event $\{t_{\barwedge} < T_J\}$, we set $U_A = \infty$. The choice of $t_{\barwedge}$ is done to ensure that the probability associated to the failure is indeed exceptional enough (with threshold $\epsilon/2$ and time-penalty $\hat{\rho}_S$). Any jump occurring before $t_{\barwedge}$ thus occur from a position $X(T_J-) \in \bar{B}(0, \ell_E + v\, t_{\barwedge}) := K$. By Assumption (PD), we can then define $w_\vee$ such that :

$$\forall\, x \in K, \quad \textstyle\int_{\mathbb{R}^d} g(x, w)\, \mathbf{1}_{\{\|w\| \geq w_\vee\}}\, dw \leq \epsilon/2 \times \exp[\hat{\rho}_S\, t_{\barwedge}].$$

A jump size larger than $w_\vee$ is then the other criterion of failure.

On the event $\{T_J \leq t_{\barwedge}\} \cap \{T_J < \tau_\partial\} \cap \{\|W\| \leq w_\vee\}$, where W is the size of the first jump (at time T_J), we set $U_A := T_J \leq t_{\barwedge}$. Otherwise $U_A := \infty$.

The proof that U_A extends to some U_A^∞ as stated in $(A3_F)$ is elementary and left to the reader. On the event $\{U_A^\infty < \tau_\partial\}$, U_A^∞ will be the first jump time T'_J to satisfy both that for some $i \geq 0$, $\tau_E^i \leq T'_J \leq \tau_E^i + t_{\barwedge}$, where τ_E^i is defined by (2.2.9), and that $\|\Delta X_{T'_J}\| \leq w_\vee$.

In particular, $\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\}$ is clear.

We prove next that the failures are indeed exceptional enough :

$$\mathbb{P}_x(U_A = \infty,\ t_{\bar{\wedge}} < \tau_\partial) \leq \mathbb{P}_x(t_{\bar{\wedge}} \leq T_J \wedge \tau_\partial) + \mathbb{P}_x(T_J < \tau_\partial \wedge t_{\bar{\wedge}}\ ,\ \|W\| > w_\vee).$$

By the definition of $w_\vee$, we deal with the second term :

$$\begin{aligned}\mathbb{P}_x(T_J < \tau_\partial \wedge t_{\bar{\wedge}}\ ,\ \|W\| > w_\vee) &\leq \mathbb{P}_x(\|W\| > w_\vee \,\big|\, T_J < \tau_\partial \wedge t_{\bar{\wedge}}) \\ &\leq \epsilon/2 \times \exp[\hat{\rho}_S\, t_{\bar{\wedge}}].\end{aligned}$$

On the event $\{t_{\bar{\wedge}} \leq T_J \wedge \tau_\partial\} : \forall\, x \leq t_{\bar{\wedge}}, \quad X_t = x - v\, t\, e_1$. Thus, X is outside of E in the time-interval $[2\ \ell_E/v\ ,\ t_{\bar{\wedge}}]$, with an extinction rate at least ρ_E. By the definition of $t_{\bar{\wedge}}$:

$$\mathbb{P}_x(t_{\bar{\wedge}} \leq T_J \wedge \tau_\partial) \leq \exp[-\rho_E\ (t_{\bar{\wedge}} - 2\ \ell_E/v)] \leq \epsilon/2 \times \exp[\hat{\rho}_S\, t_{\bar{\wedge}}].$$

This concludes :

$$\mathbb{P}_x(U_A = \infty,\ t_{\bar{\wedge}} < \tau_\partial) \leq \epsilon \times \exp[\hat{\rho}_S\, t_{\bar{\wedge}}]. \tag{2.3.3}$$

On the other hand, since the density of jumps and the jump rate are bounded on K :

$$\mathbb{P}_x(X(U_A) \in dx' \,\big|\, T_J < t_{\bar{\wedge}} \wedge \tau_\partial\ ,\ \|W\| \leq w_\vee) \leq g_\vee\, \mathbf{1}_{\{x' \in \bar{B}(0, \ell_E + v\, t_{\bar{\wedge}} + w_\vee)\}} dx'.$$

We know also from Lemma 2.3.1.1 that there exists $t_M, c_M > 0$ such that :

$$\mathbb{P}_\zeta\Big(X(t_M) \in dx\ ;\ t_M < \tau_\partial\Big) \geq c_M\, \mathbf{1}_{\{x' \in \bar{B}(0, \ell_E + v\, t_{\bar{\wedge}} + w_\vee)\}} dx'.$$

With $V = t_M$, $c = g_\vee / c_M$ and inequality (2.3.3), this concludes the proof of $(A3_F)$, thus of Theorem 2.3.1 (given Lemma 2.3.1.1). $\square$

Proof of Lemma 2.3.1.1 We wish to decompose the proof into elementary steps. The first idea is that we cut $\mathcal{D}_\ell$ into very small pieces, both for the initial value (to ensure the uniformity), and the final marginal (to ensure a density), around some reference values, resp. x_I and x_F. We also cut the path from the neighborhood of the initial value x_I to the neighborhood of the final value x_F into elementary steps with only one jump in each. We see here that imposing one jump of size in $B(S\, e_1,\ \delta S)$ in a time-interval of length $\Delta t = S/v$ gives to the marginal at time Δt a bit of dispersion in a neighborhood of x_F.

Let us formalize the above explanation. Changing the scale of X, we may assume w.l.o.g. that $S = 1$. We also consider a characteristic length of dispersion by $r := \delta S/4$.

Given some $\ell \geq 1$, $x_I \in \mathcal{D}_\ell$, $L := \ell + 2$ and $c > 0$, we define :

$$\mathcal{R}^{(L)}(c) := \Big\{(t, x_F) \in \mathbb{R}_+ \times \mathbb{R}^d \,\Big|\, \forall\, x_0 \in B(x_I, r),\\ \mathbb{P}_{x_0}\left[X_t \in dx \;;\; t < T_{\mathcal{D}_L}\right] \geq c\, \mathbf{1}_{B(x_F,r)}(x)\, dx\Big\},. \tag{2.3.4}$$

The proof then relies on the following three elementary lemmas :

Lemma 2.3.1.2. *Given any (ℓ, L) with $\ell \geq 1$, $L = \ell + 2$, $x_I \in \mathcal{D}_\ell$, $t, c, c' > 0$ and $x \in B(x_I, r)$:*

$$(t, x) \in \mathcal{R}^{(L)}(c) \Rightarrow \{(t+s, y) \,\big|\, (s, y) \in \mathcal{R}^{(L)}(c')\} \subset \mathcal{R}^{(L)}(c \times c').$$

Lemma 2.3.1.3. *Given any $L = \ell + 2$, $x_I \in \mathcal{D}_\ell$, there exists $c_0, t_0, \delta t > 0$, such that :*

$$[t_0, t_0 + \delta t] \times \{x_I\} \subset \mathcal{R}^{(L)}(c_0).$$

Lemma 2.3.1.4. *For any $x \in B(0, 2\,L - 2)$, there exists $\Delta t, c' > 0$ such that :*

$$(t, x) \in \mathcal{R}^{(L)}(c) \quad (\textit{for } t, c > 0) \quad \Rightarrow \quad \{t + \Delta t\} \times B(x, r) \subset \mathcal{R}^{(L)}(c \times c').$$

Lemma 2.3.1.2 is just an application of the Markov property. The proofs of Lemmas 2.3.1.3 and 2.3.1.4 are postponed to Appendix A. Next we show how to deduce Lemma 2.3.1.1 from them.

Proof of Lemma 2.3.1.1 *Step 1 :* from the vicinity of any $x_I \in \mathcal{D}_\ell$ to the vicinity of any $x_F \in \mathcal{D}_\ell$.

Let $K > \lfloor \|x_F - x_I\|/r \rfloor$. For $0 \leq k \leq K$, let $x_k := x_I + k\,(x_F - x_I)/K$. Exploiting Lemma 2.3.1.3, we choose (t_0, c_0) such that $(t_0, x_I) \in \mathcal{R}^{(L)}(c_0)$. Since $x_{k+1} \in B(x_k, r)$ for each k, by induction for any $k \leq K$ exploiting Lemma 2.3.1.4, there exists $t_k, c_k > 0$ such that : $(t_k, x_k) \in \mathcal{R}^{(L)}(c_k)$. In particular with $k = K$, we get some $t_f, c_f > 0$ such that $(t_f, x_F) \in \mathcal{R}^{(L)}(c_f)$.

Step 2 : a uniform time and efficiency.

By compactness, there exists $(x^j)_{j \leq J}$ such that : $\mathcal{D}_\ell \subset \bigcup_{j \leq J} B(x^j, r)$. Let $t_\veebar$ be the larger time t_f (as in step 1) needed to reach the vicinity of any $x_F \in \{x^j\}$ from any $x_I \in \{x^{j'}\}$. To adjust the arrival time, we make the process stay some time around x_I. Let $t_A := t_0 \times \lfloor 1 + t_0/\delta t \rfloor$. Then, for any $t \geq t_A$, we can find some $c > 0$ for which $(t, x_I) \in \mathcal{R}^{(L)}(c)$. We let the reader deduce this corollary from Lemmas 2.3.1.3 and 2.3.1.2. Combining this corollary with Lemma 2.3.1.2 ensures, with $t_M := t_\veebar + t_A$, a global lower-bound $c_M > 0$ on the efficiency :

$$\forall\, j, j' \leq J,\ \forall\, x_0 \in B(x_i^{j'}, r),$$
$$\mathbb{P}_{x_0}\left[X(t_M) \in dx\ ;\ t_M < T_{\mathcal{D}_L}\right] \geq c_M\ \mathbf{1}_{B(x_f^\ell, r)}(x)\, dx.$$

Since $\mathcal{D}_\ell \subset \bigcup_{j \leq J} B(x^j, r)$, this completes the proof of Lemma 2.3.1.1. □

2.3.2 Pure jump processes, a coordinate at a time

2.3.2.1 Description of the process :

For this process, we do not assume anymore that the extinction rate explodes for the process conditioned on not having any jump for a long time. The confinement property $(A2)$ shall thus hold only for a bounded range of ρ. We thus have to ensure a better survival rate while confined. We also add another difficulty in the fact that each jump affects only one component of the process at a time. For the comparison assumption $(A3_F)$, we thus need to wait at least until there has been a jump for each of these components. Again, our assumptions make us able to prove that the associated stopping time has an exponential moment greater than the survival rate. This condition $(A3_F)$ makes sure that the component of the marginal law that is singular to Lebesgue's measure vanishes at an exponential rate.

Let $(X_t)_{t \geq 0}$ be the pure jump process on $\mathcal{X} := \mathbb{R}^d$ defined by :

$$X_t := x + \sum_{i \leq d} \int_{[0,t]\times\mathbb{R}\times\mathbb{R}_+} w\, \mathbf{e_i}\, \mathbf{1}_{\left\{u \leq g_i(X_{s-}, w)\right\}} M_i(ds,\, dw,\, du) \tag{2.3.5}$$

where M_i are mutually independent PPP with intensities $ds\, dw\, du$. The process is still associated to an extinction rate ρ_e.

(X_t) is a Markov Process with piecewise constant trajectories. Conditionally upon $X_t = x$, the waiting-time and size of the next jump are independent, the law of the waiting-time is exponential of rate $\rho_J(x) := \sum_{i \leq d} \rho_J^i(x)$, where $\rho_J^i(x) := \int_{\mathbb{R}^d} g_i(x, w)\ dw < \infty$. The jump occurs on the i-th coordinate with probability $\rho_J^i(x)/\rho_J(x)$, then with size given by $g_i(x, w)/\rho_J^i(x)\, dw$.

— **Assumption (JC) :** (for Jumps by Coordinate) For any $i \leq d$, g_i is measurable and for any compact $K \subset \mathbb{R}^d$, we have :

(JC^1) An upper-bound $\rho_J^\vee$ on the jump-rate : $\sup_{\{x \in K\}} \rho_J(x) \leq \rho_J^\vee$.

(JC^2) A lower-bound $g_\wedge > 0$ on the jumps of size lower than r :
$\forall\, x \in K,\ \forall\, i \leq d,\ \forall\, w \in B(0, r), \quad g_i(x, w) \geq g_\wedge$.

(JC^3) A tightness estimate for the jumps :
for any $\epsilon > 0$, there exists $w_\vee$ such that :
$\inf_{\{x \in K,\, i \leq d\}} \left\{ \int_{\mathbb{R}} g_i(x, w)\, \mathbf{1}_{\{\|w\| \geq w_\vee\}}\, dw / \rho_J(x) \right\} \leq \epsilon.$

(JC^4) An upper-bound $g_\vee$ on the density for each jump :
$\forall\, x \in K,\ \forall\, i \le d,\ \forall\, w \in \mathbb{R}, \quad g_i(x, w)/\rho_J(x) \le g_\vee.$

(JC^5) There is also a global lower-bound $p_\wedge$ on the probability that each direction gets involved in the jump :
$\sup_{\{x\in\mathbb{R}^d,\, i\le d\}} \{\rho_J^i(x)/\rho_J(x)\} \ge p_\wedge.$

(JC^6) ρ_e is lower-bounded by $\rho > \rho_S$ outside some compact set. Moreover, ρ_e is locally bounded and explosion implies extinction : $\tau_\partial \le \sup_{\{\ell\ge 1\}} T_{\mathcal{D}_\ell}$.

(JC^7) No stable subset : $\rho_S < \inf_{\{x\in\mathbb{R}^d,\, i\le d\}} \{\rho_J^i(x) + \rho_e(x)\} \ := \rho_F.$

For this example, we can still consider $\mathcal{D}_\ell = \bar{B}(0, \ell)$, for $\ell \ge 1$ with rather here the supremum norm $\|.\|_\infty$.

Theorem 2.3.2. *Consider the process X given by equation (2.3.5) under Assumption (**JC**). Then, (**AF**) holds, thus also the exponential quasi-stationarity and the approximation property.*

Remark : The purpose of assumption (JC^7) is to bound the time T_c at which either extinction occurs or all of the coordinates have changed. Assumption (JC^7) indeed ensures an exponential moment with parameter $\hat{\rho}_S$. Since there is a finite number d of coordinates to be reassigned, we can decompose the interval $[0, T_c]$ into at most d intervals whose law is simpler. The transition from one sub-interval to the next occurs when there is a jump affecting a coordinate not previously reassigned, as long as the extinction has still not occurred. The lengths of these intervals can be easily upper-bounded by some exponential laws. The main constraint on $\hat{\rho}_S$ is then given by the worst condition on these steps, namely where one waits for the last coordinate to be reassigned.

We also refer to Subsections 2.3.1.4 for the connection with reaction-diffusion equations, which holds in the same way, this time for a non-local dispersion operator of the form :

$$c_M f(x) := \sum_{i\le d} \left[\int_{\mathbb{R}} g_i(x - w_i\, e_i,\, w_i)\, f(x - w_i\, e_i)\, dw_i - \left(\int_{\mathbb{R}} g(x, w_i)\, dw_i \right)\, f(x) \right]$$

2.3.2.2 Proof of Theorem 2.3.2 :

Assumption $(A0)$ is clearly satisfied. Assumption $(A2)$ is clearly implied by Assumpion (JC^6). E is chosen in such a way that the extinction rate is larger than some $\rho > \rho_S$ as long as the process has not come back into E. The proof of $(A1)$ is deduced from the following lemma :

Lemma 2.3.2.1. *Assumptions $(JC^{1,2,5})$ imply assumption $(A1)$, with ζ the uniform distribution over $\mathcal{D}_1$. More generally, for any $\ell \ge 1$, we can find $L > \ell$ and $t, c > 0$*

such that : $\forall\, x \in \mathcal{D}_\ell$,

$$\mathbb{P}_x\left[X(t) \in dy \ ; \ t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\, \mathbf{1}_{\{y \in \mathcal{D}_\ell\}}\, dy$$

For the proof of Lemma 2.3.2.1, we can rely on exactly the same lemmas as in the proof of Lemma 2.3.1.1 to prove the mixing on any coordinate. It is then not difficult to combine the estimates on the different coordinates. The proof is again left to the reader.

The proof of Theorem 2.3.2 is thus achieved with :

Lemma 2.3.2.2. *Assumption* (**JC**) *implies that for any* $E \in \mathbf{D}$, *assumption* $(A3_F)$ *holds.*

Proof of Lemma 2.3.2.2 We consider here three types of "failed attempts". Either the process has not done all of its required jumps despite a very long time of observation, or there are too many of these jumps, or at least one of these jumps is too large.

Definition of the stopping times and time of observation

For $k \leq d$, let T_J^k the first time at which (at least) k jumps have occurred in different coordinates. On the event $\left\{T_J^k < \tau_\partial\right\}$ (for $0 \leq k \leq d-1$), and conditionally on $\mathcal{F}_{T_J^k}$, we know from assumption (JC^7) that $T_J^{k+1} \wedge \tau_\partial - T_J^k$ is upper-bounded by an exponential variable with rate parameter $\rho_F > \hat{\rho}_S$. Thus, with $\hat{\rho}'_S := (\rho_F + \hat{\rho}_S)/2$:

$$\begin{aligned} &\forall\, x \in \mathbb{R}^d, \quad \mathbb{E}_x \exp[\hat{\rho}'_S\, (T_J^d \wedge \tau_\partial)] \leq \left[2\, \rho_F/(\rho_F - \hat{\rho}_S)\right]^d := e_f < \infty, \\ &\mathbb{P}_x[T_J^d \wedge \tau_\partial > t]\, \exp[\hat{\rho}_S\, t] \leq e_f\, \exp[-(\rho_F - \hat{\rho}_S)\, t/2] \underset{t\to\infty}{\longrightarrow} 0. \end{aligned} \tag{2.3.6}$$

Let ; $\epsilon > 0$. By inequality (2.3.6), we can find $t_{\bar{\wedge}} > 0$ such that :

$$\exp[\hat{\rho}_S\, t_{\bar{\wedge}}]\, \mathbb{P}_x(t_{\bar{\wedge}} < T_J^d \wedge \tau_\partial) \leq \epsilon/3. \tag{2.3.7}$$

On the event $\left\{t_{\bar{\wedge}} < T_J^d\right\}$, we set $U_A := \infty$. This clearly implies that $U_A \leq t_{\bar{\wedge}}$ on the event $\{t_{\bar{\wedge}} < U_F\}$.

Upper-bound on the number of jumps

Thanks to assumption (JC^5), at each new jump, conditionally on the past until the previous jump, there is a lower-bounded probability that a new coordinate gets modified. The number N_J of jumps before T_J^d (on $T_J^d < \tau_\partial$) is thus upper-bounded by a sum of d mutually independent geometric random variables. Therefore, we can define $n_J^\vee \geq 1$ such that :

$$\forall\, x \in \mathbb{R}^d, \quad \mathbb{P}_x(n_J^\vee \leq N_J\ ,\ T_J^d < \tau_\partial) \leq \epsilon\, \exp[-\hat{\rho}_S\, t_{\bar{\wedge}}]/3. \tag{2.3.8}$$

We thus declare a failure if the $n^{\vee}_J$-th jump occurs while T^d_J still is not reached.

Upper-bound on the size of the jumps
The crucial argument on the jump size is given by the following lemma.

Lemma 2.3.2.3. *Suppose that assumption* (JC^3) *holds. Given some initial condition* x*, with* $\|x\| \leq L$, $L > 0$*, let* $(W_i, i \geq 0)$ *the time-ordered sequence of jump effects. Let also* $N \geq 1$ *and* $\epsilon > 0$*. Then, there exists* $w_{\vee} > 0$ *such that :* $\mathbb{P}_x(\sup_{i \leq N} \|W_i\| \geq w_{\vee}) \geq 1 - \epsilon$.

The proof is based on an induction over N, where one needs to adjust at each step ϵ and $w_{\vee}$. The initialization is clearly given by assumption (JC^3). Assume that, with a lower-bound $1 - \epsilon/2$ on the probability of the following event, the first $N-th$ jumps are upper-bounded by $w^N_{\vee} > 0$. On this event, $\|X(T_N)\| \leq L + N \times w^N_{\vee}$ so that we can apply assumption (JC^3) to make the event $\|W_{N+1}\| \leq w^{N+1}_{\vee}$ occur with probability greater than $1 - \epsilon/2$. The intersection of both conditions is included in $\{\sup_{i \leq N+1} \|W_i\| \geq w^{N+1}_{\vee}\}$ and occurs with probability greater than $1 - \epsilon$. The induction concludes the proof of the lemma.

From this lemma, we can deduce a value $w_{\vee}$ such that for any $x \in E$, the probability for the process starting from x that there is a jump with size larger than $w_{\vee}$ before the $n^{\vee}_J$-th jump and $T^d_J \wedge \tau_\partial$ is upper-bounded by $\epsilon \exp[-\hat{\rho}_S\, t_{\bar{\wedge}}]/3$. We thus declare a failure if a jump larger than $w_{\vee}$ occurs. From this, we deduce that the process has stayed in some $\mathcal{D}_L := B(0, \ell_E + n^{\vee}_J \times w_{\vee})$ on the event $\{U^1_F = \infty\}$.

On the event that at time $T^d_J < \tau_\partial$, none of the three following conditions have been violated :
(i) T^d_J still has not occurred at time $t_{\bar{\wedge}}$,
or (ii) the $n^{\vee}_J$-th jump has occurred,
or (iii) a jump of size larger than n_W has occurred (before time T^d_J),
we set $U_A := T^d_J$. Otherwise $U_A := \infty$.

Given our construction (see (2.3.7), (2.3.8) and the above definition of $w_{\vee}$), it is clear that :

$$\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\} \quad \text{and} \quad \mathbb{P}_x(U_A = \infty,\, t < \tau_\partial) \leq \epsilon \exp(-\rho\, t),$$

As previously, the proof of Lemma 2.3.2.2 is then completed with Lemma 2.3.2.1 and :

Lemma 2.3.2.4. *Assume that* (JC^{3-7}) *hold, with the preceding notations. Then, there exists* $c > 0$ *such that :*

$$\mathbb{P}_x(X(U_A) \in dx' \;;\; U_A < \tau_\partial) \leq c\, \mathbf{1}_{\{x' \in \mathcal{D}_L\}} dx'.$$

The proof of Lemma 2.3.2.4, which is technical, is presented in detail in Appendix B. □

2.4 Proofs of Theorems 2.2.1-4

2.4.1 Absorption with failures

For the proofs, we directly exploit Theorems 1.2.1-3 in Chapter 1 thanks to the following theorem. The latter extends in fact Theorem 1.5.2 in Chapter 1 and their proofs share many similarities. Similarly, we exploit Corollary 1.5.2.1 in Chapter 1 to deduce that under Assumption (**AF**), there exists $t_S, c_S > 0$ and $\hat{\rho}_S \geq \rho_S$ such that :

$$\forall\, u \geq 0,\; \forall\, t \geq u + t_S, \qquad P_\zeta(t - u < \tau_\partial) \leq c_S\, e^{\hat{\rho}_S\, u}\, \mathbb{P}_\zeta\left(t < \tau_\partial\right), \tag{2.4.1}$$

As mentioned in Chapter 1, $(A3)$ is not required for the proof of Corollary 1.5.2.1, so that (2.4.1) is indeed implied by the present Assumption (**AF**).

Theorem 2.4.1. *Assume that there exist $\zeta \in\; \mathcal{M}_1\left(\mathcal{X}\right)$, $t_S, c_S > 0$, $\rho \geq \hat{\rho}_S > \rho_S > 0$, and $E \in \mathbf{D}$ such that inequality (2.4.1), assumptions $(A2)$ and $(A3_F)$ hold. Then, for some $c, u > 0$:*

$$\forall\, x \in \mathcal{X},\; \forall\, t \geq u, \quad \mathbb{P}_x(t < \tau_\partial) \leq c\, \mathbb{P}_\zeta(t < \tau_\partial).$$

Remark : The following proof simply ensures $(A3)$, and we refer to Theorem 1.5.2 in Chapter 1 to see how it implies Theorem 2.4.1. The conclusion is actually inequality (2.1.1) presented in the Introduction.

2.4.1.1 Proof of Theorem 2.4.1

In the following, every time we will apply assumption $(A3_F)$, we will exploit the parameter :

$$\epsilon \;:= \left(2\, c_S\, e_{\mathcal{T}}\right)^{-1}, \qquad \text{where } e_{\mathcal{T}} := \sup_{\{x \in \mathcal{X}\}}\; \mathbb{E}_x\left(\exp\left[\rho\left(\tau_\partial \wedge \tau_E\right)\right]\right).$$

For $t_{\bar{\wedge}}$ the associated deterministic upper-bound on U_A, we define $t_A \;:= t_{\bar{\wedge}} + t_S$.

The first step is the following lemma :

Lemma 2.4.1.1. *Assume that inequality (2.4.1) and assumption $(A3_F)$ hold with the above parameters. Then, there exists $C_0 > 0$ such that :*

$$\forall\, x \in E,\; \forall\, t \geq t_A, \quad \mathbb{P}_x(U_A < t < \tau_\partial) \leq C_0\, \mathbb{P}_\zeta(t < \tau_\partial)$$

For ϵ sufficiently small, we expect this quantity to be the leading part of $\mathbb{P}_x(t < \tau_\partial)$. We will prove indeed that the extension of survival during the failed coupling procedure, for a time-length $t_{\bar{\wedge}}$, then outside E (before τ_E^1) is not sufficient to compensate the cost of such a failure.

Our idea is to distinguish the events according to the number of failures, and treat them inductively by replacing x by $X(\tau_E^1)$ and t by $t - \tau_E^1$. Therefore, the induction is done not exactly depending on J, but more precisely on the value of the r.v. :

$$J(t) \ := \sup\left\{j \geq 0,\ \tau_E^j < (t - t_A) \wedge \tau_\partial \wedge U_A^\infty\right\}, \tag{2.4.2}$$

where we recall from (2.2.9) :

$$\tau_E^{i+1} := \inf\{s \geq \tau_E^i + t_{\bar{\wedge}} : X_s \in E\} \wedge \tau_\partial, \text{ and } \tau_E^0 = 0.$$

To end the initialization ($J(t) = 0$), we thus prove :

Lemma 2.4.1.2. *Assume that assumption* $(A2)$*, inequalities* (2.4.1) *and assumption* $(A3_F)$ *hold with* $\rho > \hat{\rho}_S$*. Then, there exists* $C_F > 0$ *such that :*

$$\forall\, x \in E,\ \forall\, t \geq t_A, \quad \mathbb{P}_x(U_A = \infty\ ,\ t \leq \tau_E^1\ ,\ t < \tau_\partial) \leq C_F\ \mathbb{P}_\zeta(t < \tau_\partial)$$

For any $t \geq 0$, $J(t) \leq t/t_{\bar{\wedge}} < \infty$, so that the induction is completed with the following lemma.

Lemma 2.4.1.3. *Assume that assumption* $(A2)$*, inequalities* (2.4.1) *and assumption* $(A3_F)$ *hold with* $\rho > \hat{\rho}_S$*. If there exists* $j \geq 0$ *and* $C_j > 0$ *such that :*

$$\forall\, x \in E,\ \forall\, t \geq t_A, \quad \mathbb{P}_x(t < \tau_\partial\ ,\ J(t) = j) \leq C_j\ \mathbb{P}_\zeta(t < \tau_\partial), \tag{2.4.3}$$
$$\textit{Then :} \quad \forall\, x \in E,\ \forall\, t \geq t_A, \quad \mathbb{P}_x(t < \tau_\partial\ ,\ J(t) = j+1) \leq \tfrac{C_j}{2}\ \mathbb{P}_\zeta(t < \tau_\partial)$$

Lemmas 2.4.1.1-3 imply (2.1.1) : With $C := 2\,(C_0 + C_F)$, by Lemmas 2.4.1.1-3, we deduce that $\forall\, x \in E,\ \forall\, t \geq t_A,$:

$$\forall\, j \geq 0, \quad \mathbb{P}_x(t < \tau_\partial\ ,\ J(t) = j) \leq 2^{-j-1}\ C\ \mathbb{P}_\zeta(t < \tau_\partial)$$
$$\mathbb{P}_x(t < \tau_\partial) = \textstyle\sum_{\{j \geq 0\}} \mathbb{P}_x(t < \tau_\partial\ ,\ J(t) = j) \leq C\ \mathbb{P}_\zeta(t < \tau_\partial),$$

which means that assumption $(A3)$ in Chapter 1 holds. We refer to the proof of Theorem 1.5.2 in Chapter 1 (inspired by [CV17b], cf. p.13) to see how it implies inequality (2.1.1) with again inequality (2.4.1) and assumption $(A2)$. $\square$

Proof of Lemma 2.4.1.1

$$\begin{aligned}
\mathbb{P}_x(U_A < t < \tau_\partial) &= \mathbb{E}_x\left[\mathbb{P}_{X[U_A]}(t - U_A < \widetilde{\tau_\partial})\ ;\ U_A < \tau_\partial\right] \\
&\leq C\ \mathbb{E}_\zeta\left[\mathbb{P}_{X[V]}(t - t_{\bar{\wedge}} < \widetilde{\tau_\partial})\ ;\ V < \tau_\partial\right] \\
&\leq C\ \mathbb{P}_\zeta\left[t - t_{\bar{\wedge}} < \tau_\partial\right],
\end{aligned}$$

where we exploited assumption $(A3_F)$ with $U_A \le t_{\bar{\wedge}}$ on the event $\{U_A < \tau_\partial\}$. By inequality (2.4.1), with $t \ge t_{\bar{\wedge}} + t_S$:

$$\mathbb{P}_x(U_A < t < \tau_\partial) \le C\, c_S\, e^{\rho\, t_{\bar{\wedge}}}\, \mathbb{P}_\zeta\left[t < \tau_\partial\right].$$

Lemma 2.4.1.1 is thus satisfied with : $C_0 \ := c_S C\, e^{\rho\, t_{\bar{\wedge}}}$. □

Proof of Lemma 2.4.1.2 First, we exploit the Markov inequality, and assumption $(A2)$ that states that :

$$\text{a.s. on } \{U_A = \infty, t_{\bar{\wedge}} < \tau_\partial\} \ : \qquad \mathbb{E}_{X_{t_{\bar{\wedge}}}}\left[\exp\left(\rho[\widetilde{\tau}_E \wedge \widetilde{\tau_\partial}]\right)\right] \le e_{\mathcal{T}}.$$

$$\begin{aligned}
&\mathbb{P}_x(U_A = \infty\,,\ t - \tau_E \le \tau_E^1\,,\ t < \tau_\partial) \\
&\quad = \mathbb{E}_x\left[\mathbb{P}_{X_{t_{\bar{\wedge}}}}(t - t_A - t_{\bar{\wedge}} \le \widetilde{\tau}_E\ ;\ t - t_{\bar{\wedge}} < \widetilde{\tau_\partial})\ ;\ t_{\bar{\wedge}} < \tau_\partial\,,\ U_A = \infty\right] \\
&\quad\quad \le e_{\mathcal{T}}\ e^{-\rho[t - t_A - t_{\bar{\wedge}}]} \mathbb{P}_x\left[t_{\bar{\wedge}} < \tau_\partial\,,\ U_A = \infty\right].
\end{aligned}$$

From (2.4.1), we can relate the decay $e^{-\rho t}$ to $\mathbb{P}_\zeta(t < \tau_\partial)$. Exploiting also $(A3_F)$, we conclude the proof of Lemma 2.4.1.2 where :

$$C_F \ := \frac{e_{\mathcal{T}}\ c_S\ \epsilon}{e^{\rho\, t_S}\ \mathbb{P}_\zeta(t_S < \tau_\partial)} = (2 e^{\rho\, t_S}\ \mathbb{P}_\zeta(t_S < \tau_\partial))^{-1} > 0.$$ □

Proof of Lemma 2.4.1.3 By the Markov property assumed on U_A^∞ (see assumption $(A3_F)$) and definition (2.4.2) :

$$\begin{aligned}
&\mathbb{P}_x(t < \tau_\partial\,,\ J(t) = j + 1) \\
&\quad = \mathbb{E}_x\left[\mathbb{P}_{X[\tau_E^1]}(t - \tau_E^1 < \widetilde{\tau_\partial}\,,\ \widetilde{J}(t - \tau_E^1) = j)\ ;\ J(t) \ge 1\right] \\
&\quad \le C_j\ \mathbb{E}_x\left[\mathbb{P}_\zeta(t - \tau_E^1 < \widetilde{\tau_\partial})\ ;\ J(t) \ge 1\right],
\end{aligned}$$

where we exploited (2.4.3) and the fact that $J(t) \ge 1 \Rightarrow t - \tau_E^1 \ge t_A$. Then, by inequality (2.4.1) :

$$\begin{aligned}
&\mathbb{P}_x(t < \tau_\partial\,,\ J(t) = j + 1) \le C_j\ c_S\ \mathbb{P}_\zeta(t < \tau_\partial) \times \mathbb{E}_x\left[\exp(\rho\, \tau_E^1);\ J(t) \ge 1\right] \\
&\quad \le C_j\, c_S\, \mathbb{P}_\zeta(t < \tau_\partial)\ \mathbb{E}_x\Big[\mathbb{E}_{X_{t_{\bar{\wedge}}}}\left(\exp(\rho\, \widetilde{\tau}_E)\ ;\ \widetilde{\tau}_E < (t - t_{\bar{\wedge}} - t_A) \wedge \widetilde{\tau_\partial}\right) \\
&\qquad\qquad \times \exp(\rho t_{\bar{\wedge}})\ ;\ U_A = \infty\,,\ t_{\bar{\wedge}} < \tau_\partial\Big],
\end{aligned}$$

where we decomposed τ_E^1 into $t_{\bar{\wedge}} + \widetilde{\tau}_E$.

Now, by assumption $(A2)$ we have $a.s.$ on $\{U_A = \infty\,,\; t_{\bar{\wedge}} < \tau_\partial\}$:

$$\mathbb{E}_{X_{t_{\bar{\wedge}}}}(\exp(\rho\,\widetilde{\tau}_E)\;;\;\widetilde{\tau}_E < (t - t_{\bar{\wedge}} - t_A) \wedge \widetilde{\tau_\partial}) \le e_{\mathcal{T}};$$

$$\mathbb{E}_x\left[\exp(\rho t_{\bar{\wedge}})\;;\;U_A = \infty\right] \le \epsilon = \frac{1}{2\,c_S\,e_{\mathcal{T}}}. \qquad \text{by assumption } (A3_F).$$

$$\text{Thus, } \forall\, x \in E,\ \forall\, t \ge t_A, \quad \mathbb{P}_x(t < \tau_\partial\,,\; J(t) = j+1) \le \tfrac{C_j}{2}\,\mathbb{P}_\zeta(t < \tau_\partial),$$

which yields the result of Lemma 2.4.1.3. □

2.4.2 A more precise convergence result

Assumption (**A**) of Chapter 1 is thus implied by our Assumption (**AF**). By Theorems 1.2.1-3 in Chapter 1, we deduce most of the results of the present Theorems 2.2.1-3, except for the convergence results towards α and β.

What is obtained in Chapter 1 is that there exists $\gamma > 0$ such that there exists $C_h > 0$ such that :

$$\forall\, t > 0,\ \|h_t - h\|_\infty \le C_h\, e^{-\gamma\, t}. \tag{2.4.4}$$

Moreover, for any $\ell \ge 1$ and $\xi \in (0,1]$, there exists $C_\alpha = C_\alpha(\ell,\, \xi) > 0$ such that :

$$\forall\, t > 0,\ \forall\, \mu \in \mathcal{M}_{\ell,\,\xi},\ \|\, \mu A_t(dx) - \alpha(dx)\,\|_{TV} \le C_\alpha(\ell, \xi)\, e^{-\gamma\, t}. \tag{2.4.5}$$

As stated in the following Proposition, these property directly imply the remaining conclusions of our theorems.

Proposition 2.4.2.1. *Assume* (2.4.4) *and* (2.4.5)*. Then, there exists* $\gamma, C > 0$ *such that :*

$$\forall\, \mu \in \mathcal{M}_1(\mathcal{X}),\ \forall\, t > 0, \quad \|\, \mathbb{P}_\mu\left[\, X_t \in dx \mid t < \tau_\partial\right] - \alpha(dx)\,\|_{TV} \le C \frac{\inf_{u>0} \|\mu - u\,\alpha\|_{TV}}{\langle \mu \,|\, h\rangle}\, e^{-\gamma\, t}.$$

Likewise for the Q-process, with $h_*\mu(dx) := h(x)\,\mu(dx)\,/\langle \mu \,|\, h\rangle$ *:*

$$\forall\, \mu \in \mathcal{M}_1(\mathcal{X}),\ \forall\, t > 0, \quad \|\, \mu B[h] Q_t - \beta\,\|_{TV} \le C \frac{\inf_{u>0} \|\mu - u\,\alpha\|_{TV}}{\langle \mu \,|\, h\rangle}\, e^{-\gamma\, t}.$$

Assume that $\nu \in \mathcal{M}_1(\mathcal{X})$ satisfies $\langle \nu \,|\, 1/h\rangle < \infty$. We may define $\mu(dx) := \dfrac{\nu(dx)}{h(x)\,\langle \nu \,|\, 1/h\rangle}$, which trivially satisfies : $\mu \in \mathcal{M}_1(\mathcal{X})$ and $\mu B[h] = \nu$. Moreover, by the definition of β and μ :

$$\frac{\inf_{u>0} \|\mu - u\,\alpha\|_{TV}}{\langle \mu \,|\, h\rangle} = \inf_{u>0} \|\frac{\nu}{\langle \nu \,|\, 1/h\rangle} - u\,\beta\|_{1/h} \times \frac{\langle \nu \,|\, 1/h\rangle}{\langle \nu \,|\, \mathbf{1}\rangle}$$
$$= \inf_{u'>0} \|\nu - u'\,\beta\|_{1/h}.$$

The proof of Theorems 2.2.1-3 is thus complete with this proposition.

As one can see in the following subsection, the proof of the proposition comes from the following lemma, which may be interesting by itself.

Lemma 2.4.2.2. *Assume* (2.4.5) *and* (2.4.4)*. Then, there exists* $\gamma, C > 0$ *such that for any signed measure* μ *with* $\|\mu\|_{TV} = 1$ *(in particular for any probability measure) :*

$$\left\| e^{-t\,\lambda}\,\mu.P_t - \langle \mu \,|\, h\rangle \alpha \right\|_{TV} \leq C\,\exp[-\gamma\,t].$$

As a corollary, for any signed measure μ *with finite total variation and such that* $\langle \mu \,|\, h\rangle = 0$ *:*

$$\left\| e^{-t\,\lambda}\,\mu P_t \right\|_{TV} \leq C\,\exp[-\gamma\,t]\,\|\mu\|_{TV}\,.$$

Likewise, for any bounded measurable function ψ *such that* $\langle \alpha \,|\, \psi\rangle = 0$*, we have :*

$$\|e^{-t\,\lambda}\,P_t\psi\|_\infty \leq C\,\exp[-t\,\gamma]\,\|\psi\|_\infty.$$

The proof of the Lemma is deferred after the one of the Proposition.

2.4.2.1 Proof of Proposition 2.4.2.1

Let any $u \in \mathbb{R}_+$, and define $\bar{\mu} = \mu - u\,\alpha$. We recall that by definition, $\langle \alpha \,|\, h_t\rangle = \langle \alpha \,|\, h\rangle = 1$.

$$\mu A_t - \alpha = \frac{\exp[t\lambda]\,\mu P_t - \langle \mu \,|\, h\rangle\,\alpha}{\langle \mu \,|\, h\rangle} + \frac{\langle \mu \,|\, h - h_t\rangle}{\langle \mu \,|\, h\rangle \times \langle \mu \,|\, h_t\rangle}\,\exp[t\lambda]\,\mu P_t$$
$$= \frac{\exp[t\lambda]\,\bar{\mu} P_t - \langle \bar{\mu} \,|\, h\rangle\,\alpha}{\langle \mu \,|\, h\rangle} + \frac{\langle \bar{\mu} \,|\, h - h_t\rangle}{\langle \mu \,|\, h\rangle}\,\mu A_t.$$

By Lemma 2.4.2.2, this immediately implies the estimate on the convergence to α by optimizing the estimate in u. Likewise for the convergence to β :

$$\langle \mu B[h] Q_t \,\big|\, f\rangle - \langle \beta \,\big|\, f\rangle = \frac{\langle \mu \,\big|\, e^{t\lambda} P_t \,\big|\, h \times f\rangle - \langle \mu \,\big|\, h\rangle \times \langle \alpha \,\big|\, h \times f\rangle}{\langle \mu \,\big|\, h\rangle}$$
$$= \frac{\langle \bar{\mu} \,\big|\, e^{t\lambda} P_t \,\big|\, h \times f\rangle - \langle \bar{\mu} \,\big|\, h\rangle \times \langle \alpha \,\big|\, h \times f\rangle}{\langle \mu \,\big|\, h\rangle}.$$

Again, by Lemma 2.4.2.2, this immediately implies Proposition 2.4.2.1 by noting that h is bounded and optimizing the estimate in u. □

2.4.2.2 Proof of Lemma 2.4.2.2 :

Assume first that $\|\mu\|_{TV} = 1$. Denote by μ_+ (resp. μ_-) the positive (resp. negative) component of μ so that : $\mu = \mu_+ - \mu_-$ and $\|\mu\|_{TV} = 1 = \mu_+(\mathcal{X}) + \mu_-(\mathcal{X})$. Let $y \in \mathcal{D}_1$ and define :

$$\hat{\mu}_+(dx) := \frac{1}{1 + \mu_+(\mathcal{X})}\,[\delta_y + \mu_+(dx)] \geq 0 \;,\; \hat{\mu}_-(dx) := \frac{1}{1 + \mu_-(\mathcal{X})}\,[\delta_y + \mu_-(dx)] \geq 0.$$

They satisfy :

$$\mu = [1 + \mu_+(\mathcal{X})]\,\hat{\mu}_+ - [1 + \mu_-(\mathcal{X})]\,\hat{\mu}_- \;,\; \hat{\mu}_+(\mathcal{X}) = \hat{\mu}_-(\mathcal{X}) = 1,$$
$$\hat{\mu}_+(\mathcal{D}_1) \wedge \hat{\mu}_-(\mathcal{D}_1) \geq \frac{1}{1 + \|\mu\|_{TV}} = 1/2.$$

Thus, first by (2.4.5), then by (2.4.4), there exists ζ, C independent from μ such that :

$$\|\, \hat{\mu}_+ A_t(dx) - \alpha(dx)\,\|_{TV} \vee |\, \hat{\mu}_- A_t(dx) - \alpha(dx)\,\|_{TV} \leq C\, e^{-\zeta\, t},$$
$$|\langle \hat{\mu}_+ \,\big|\, h_t - h\rangle| \vee |\langle \hat{\mu}_- \,\big|\, h_t - h\rangle| \leq C\, e^{-\zeta\, t}.$$

With the fact that h and the family h_t are uniformly bounded :

$$\mu.(e^{-t\,\lambda}\, P_t) = [1 + \mu_+(\mathcal{X})]\, \langle \hat{\mu}_+ \,\big|\, h_t\rangle\, \hat{\mu}_+ .A_t\Big| - [1 + \mu_-(\mathcal{X})]\, \langle \hat{\mu}_- \,\big|\, h_t\rangle\, \hat{\mu}_- .A_t\Big|$$
$$= \Big([1 + \mu_+(\mathcal{X})]\, \langle \hat{\mu}_+ \,\big|\, h\rangle - [1 + \mu_-(\mathcal{X})]\, \langle \hat{\mu}_- \,\big|\, h\rangle\Big)\; \alpha + O_{TV}(e^{-t\,\zeta})$$
$$= \langle \mu \,\big|\, h\rangle\, \alpha + O_{TV}(e^{-t\,\zeta}).$$

On any signed measure such that $\|\mu\|_{TV} = 1$ and $\langle \mu \,\big|\, h\rangle = 0$, we thus have indeed for some $C' > 0$: $\Big\|\mu.(e^{-t\,\lambda}\, P_t)\Big\|_{TV} \leq C' \exp[-t\,\zeta]$. By linearity of P_t, we conclude the case of other values for the total variation.

Concerning the last estimate, we just have to remark that for any signed measure μ :

$$\begin{aligned}\Big|\,\langle\mu\,\big|\,e^{-t\,\lambda}\,P_t\,\big|\,\psi\rangle - \langle\mu\,\big|\,h\rangle\times\langle\alpha\,\big|\,\psi\rangle\,\Big| &\leq \Big\|\mu.(e^{-t\,\lambda}\,P_t) - \langle\mu\,\big|\,h\rangle\alpha\Big\|_{TV}\times\|\psi\|_\infty\\ &\leq C\,\exp[-t\,\zeta]\,\|\psi\|_\infty\times\|\mu\|_{TV}\,.\end{aligned}$$

This concludes the proof of Lemma 2.4.2.2. □

Alternatively, one could adapt the previous proof by noticing that :

$$e^{-t\,\lambda}\,P_t\psi(x)/2 = (e^{-t\,\lambda}\,P_t\psi(x) + \langle\alpha\,\big|\,\psi\rangle)/2 = \langle\alpha_x\,\big|\,h_t\rangle\times\langle\alpha_x A_t\,\big|\,\psi\rangle,$$

where $\alpha_x := (\delta_x+\alpha)/2$ satisfies $\alpha_x(\mathcal{D}_1)\geq\alpha(\mathcal{D}_1)/2>0$.

2.4.3 Proof of Theorem 2.2.5

Recall that we wish to describe the approximations of the previous dynamics when extinction happens at $\tau_\partial^L := \tau_\partial\wedge T_{\mathcal{D}_L}$ instead of τ_∂.

As mentioned in the remark, there is an explicit relation between all the constants introduced in the proofs of Theorems 2.2.1-4 (cf. also the proof in Chapter 1). Moreover, the proof actually relies on a single value of $\rho>\rho_S$ and a specific set E. Note that for any L such that $E\subset\mathcal{D}_L$, we have :

$$\sup_{\{x\in\mathcal{X}\}}\,\mathbb{E}_x\left(\exp\left[\rho\,(\tau_\partial^L\wedge\tau_E)\right]\right)\leq\sup_{\{x\in\mathcal{X}\}}\,\mathbb{E}_x\left(\exp\left[\rho\,(\tau_\partial\wedge\tau_E)\right]\right) := e_{\mathcal{T}}$$

Likewise, Assumption $(A3_F)$ extends naturally for τ_∂^L. The largest set involved in the application of $(A1)$ is actually $\mathcal{D}_\circ$ (cf. notably the definition of the constants exploited for the convergences in Section 1.5.3 of Chapter 1). So it suffices to take L such that $\mathcal{D}_\circ\subset\mathcal{D}_L$ to ensure that all the results extend for the extinction time τ_∂^L instead of τ_∂. Under this condition, our proof ensures that the exponential quasi-stationarity also holds for the process with extinction at time τ_∂^L and that the constants involved in the convergences can be taken uniformly over these L.

To compare λ to λ_L, we can observe that for any $t>0$:

$$\frac{-1}{t}\log\mathbb{P}_\zeta(t<\tau_\partial)\leq\frac{-1}{t}\log\mathbb{P}_\zeta(t<\tau_\partial^L)\qquad\text{so that}\qquad\lambda\leq\lambda_L$$

by taking the limit $t\to\infty$ and exploiting the convergence of the survival capacities. The same argument ensures that λ_L is an increasing sequence in L.

Assume then by contradiction that there exists $\delta>0$ such that $\lim_L\lambda_L\geq\lambda+\delta$. Let also $\rho>\lambda_L$ for L sufficiently large (for instance the one in $(A2)$). By Lemma 2.4.2.2

and the analogous result with τ_∂^L, for t sufficiently large :

$$e^{\lambda t}\mathbb{P}_\zeta(t < \tau_\partial) \geq \frac{1}{2}\langle \zeta \,\big|\, h\rangle, \qquad e^{\lambda_L t}\mathbb{P}_\zeta(t < \tau_\partial^L) \leq 2\langle \zeta \,\big|\, h^L\rangle. \tag{2.4.6}$$

Exploiting also the properties of δ and ρ, and recalling that one has an explicit upper-bound $\|h_*\|_\infty$ that is also valid for the function h^L, we deduce :

$$0 < \langle \zeta \,\big|\, h\rangle \leq 4\,\|h_*\|_\infty e^{-\delta t} + 2e^{(\rho-\delta)t}\,|\mathbb{P}_\zeta(t < \tau_\partial) - \mathbb{P}_\zeta(t < \tau_\partial^L)|. \tag{2.4.7}$$

The first term in the upper-bound becomes negligible uniformly over L by taking t sufficiently large. In order to obtain a contradiction, we merely have to prove that $\mathbb{P}_\zeta(t < \tau_\partial^L)$ converges to $\mathbb{P}_\zeta(t < \tau_\partial)$ at any fixed time t. The difference is $\mathbb{P}_\zeta(T_{\mathcal{D}_L} < t < \tau_\partial) \leq \mathbb{P}_\zeta(T_{\mathcal{D}_L} < t)$, so we shall prove that a.s. $\lim_L T_{\mathcal{D}_L} = \infty$. Assume by contradiction that the limit T_∞ of this increasing sequence is at a finite value. Then, by $(A0)$ and the fact that X is cadlag, $X_{T_\infty -} \in \mathcal{X}$ is in some $int(\mathcal{D}_M)$. Thus, there exists a vicinity to the left of T_∞ on which $T_{\mathcal{D}_L}$ for $L > M$ cannot happen. Yet, this precisely contradicts the definition of T_∞. Consequently, a.s. $\lim_L T_{\mathcal{D}_L} = \infty$ and $\mathbb{P}_\zeta(t < \tau_\partial^L)$ converges to $\mathbb{P}_\zeta(t < \tau_\partial)$ as $L \to \infty$. The contradiction with (2.4.7) makes us conclude that λ_L tends to λ as $L \to \infty$.

The next step is to look at the survival capacities, by exploiting again Lemma 2.4.2.2 (with the measures evaluated on $\mathcal{X}$). For any $x \in \mathcal{X}$:

$$|h(x) - h_L(x)| \leq e^{\lambda t}|\mathbb{P}_x(t < \tau_\partial) - \mathbb{P}_x(t < \tau_\partial^L)| + |e^{\lambda t} - e^{\lambda_L t}| + C\,e^{-\gamma t}.$$

Again, we can choose t sufficiently large to make $C\,e^{-\gamma t}$ negligible. We already know that λ_L tends to λ and as previously, we prove that $\mathbb{P}_x(t < \tau_\partial^L)$ tends to $\mathbb{P}_x(t < \tau_\partial)$, as $L \to \infty$. This concludes the punctual convergence of h_L to h. The conclusion would be the same if one replaces x by any probability measure μ, for instance α.

Concerning the QSD :

$$\begin{aligned}\|\alpha - \alpha_L\|_{TV} \leq \left\|e^{\lambda_L t}\delta_\alpha P_t^L - \langle \alpha \,\big|\, h_L\rangle \alpha_L\right\|_{TV} + |e^{\lambda_L t} - e^{\lambda t}| \\ + |\langle \alpha \,\big|\, h_L - h\rangle| + e^{\lambda t}\left\|\delta_\alpha P_t^L - \delta_\alpha P_t\right\|_{TV} \\ \text{where } \left\|\delta_\alpha P_t^L - \delta_\alpha P_t\right\|_{TV} = \mathbb{P}_\alpha(T_{\mathcal{D}_L} < t < \tau_\partial) \to 0 \text{ as } L \to \infty \qquad (t \text{ fixed})\end{aligned}$$

Exploiting again Lemma 2.4.2.2 and the previous convergence results, the r.h.s. can be made negligible by taking t then L sufficiently large, concluding the convergence of α_L to α in total variation. This concludes the proof of Theorem 2.2.5. □

Appendix A : Proofs of Lemmas 2.3.1.3 and 2.3.1.4

Proof of Lemma 2.3.1.3

Let $t_0 := S/v$, $x_0 \in B(x_I, r)$. Let T_J be the first time of jump of X and let f be a positive and measurable test function. Concerning the constraint $t_0 < T_{\mathcal{D}_L}$, note that :

$$\Big\{x_0 - v\, s\, e_1 \,\Big|\, s \le t_0\Big\} \cup \Big\{x_0 - v\, s\, e_1 + w \,\Big|\, s \le t_0 \text{ , } w \in B(S\, e_1,\ 4r)\Big\} \subset \mathcal{D}_L.$$

Thus, making only such jumps will keep the process inside of $\mathcal{D}_L$. By Palm's formula (see e.g. Proposition 13.1.VII in [DV08]), conditionally on M having a Dirac mass on (s, w, u), the law of M is simply the one of $\widetilde{M} + \delta_{(s,w,u)}$, where $\widetilde{M}$ is an independent copy of M. By this way, we can deal with this constraint $t_0 < T_{\mathcal{D}_L}$, referring to the process $\hat{X}$ driven by $\widetilde{M} + \delta_{(s,w,u)}$:

$$\begin{aligned}
&\mathbb{E}_{x_0}\left(f[X(t_0)] \text{ ; } t_0 < T_{\mathcal{D}_L}\right) = f[x_0 - v\ t_0\ e_1] \\
&\qquad + \mathbb{E}_{x_0}\Big(\int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} (f[X_{s^-} + w - v(t_0 - s)e_1] - f[X_{s^-} - v(t_0-s)e_1]) \\
&\qquad\qquad\qquad\qquad \mathbf{1}_{\left\{u \le g(X_{s^-}, w)\right\}}\, \mathbf{1}_{\left\{t_0 < \hat{T}_{\mathcal{D}_L}\right\}}\, M(ds, dw, du)\Big) \\
&= f[x_0 - v\ t_0\ e_1] + \int_{[0,t]\times\mathbb{R}^d} \mathbb{E}_{x_0}\Big((f[X_{s^-} + w - v\ (t_0 - s)\ e_1] \\
&\qquad\qquad - f[X_{s^-} - v\ (t_0 - s)\ e_1]) \times g(X_{s^-}, w)\, \mathbf{1}_{\left\{t_0 < \hat{T}_{\mathcal{D}_L}\right\}}\Big)\, ds\, dw.
\end{aligned}$$

Here we specify by the hat (for $t_0 < \hat{T}_{\mathcal{D}_L}$) that we consider the above-mentioned process $\hat{X}_0$, with $u = 0$, remarking that $\hat{X}$ (with (s, w, u)) equals $\hat{X}_0$ (with $(s, w, 0)$) on the event $\{u \le g(X_{s^-}, w)\}$. Note that $\hat{X}_{s-} = X_{s-}$. We now define :

$$\begin{aligned}
H(x_0) &:= \int_{[0,t_0]\times\mathbb{R}^d} g(x - v\, s\, e_1, w)\, ds\, dw \\
\text{and } p(x_0) &:= \mathbb{P}_{x_0}(t_0 < U^j) = \exp\left[-H(x_0)\right] \ge \exp[-\rho_J^\vee\ t_0],
\end{aligned} \tag{2.5.1}$$

by assumption (JC^1). With a similar approach as above, we can describe with Palm's formula $\mathbb{E}_{x_0}\left(f[X(t_0)] \text{ ; } t_0 < T_{\mathcal{D}_L} \text{ , } X \text{ has done two jumps before } t_0\right)$ so as to obtain :

$$\begin{aligned}
&\mathbb{E}_{x_0}\left(f[X(t_0)] \text{ ; } t_0 < T_{\mathcal{D}_L}\right) \\
&\quad = p(x_0)\, f[x_0 - v\ t_0\ e_1] + p(x_0)\, H(x_0)\, f[x_0 - v\ t_0\ e_1] \\
&\quad + \int_{[0,t_0]\times\mathbb{R}^d} p(x_0)\, (f[x_0 - v\ t_0\ e_1 + w] - f[x_0 - v\ t_0\ e_1])\, g(x_0 - v\ s\ e_1, w)\, ds\, dw \\
&\quad + \mathbb{E}_{x_0}\left(f[X(t_0)] \text{ ; } t_0 < T_{\mathcal{D}_L} \text{ , } X \text{ has done two jumps before } t_0\right).
\end{aligned}$$

Thus, by assumption (JC^2) and the fact that f is positive :

$$\mathbb{E}_{x_0}\left(f[X(t_0)]\ ;\ t_0 < T_{\mathcal{D}_L}\right) \geq p(x_0)\, h_\wedge\, t_0 \int_{B(S\, e_1,\ 4r)} f[x_0 - v\ t_0\ e_1 + w]\ dw.$$

Now, with inequality (2.5.1), the facts that $x_0 \in B(x_I, r)$, and that f can be any positive and measurable function :

$$\mathbb{P}_{x_0}[X(t_0) \in dx\ ;\ t_0 < T_{\mathcal{D}_L} \wedge \tau_\partial] \geq \exp[-\rho_J^\vee\, t_0]\ h_\wedge\ t_0\ \mathbf{1}_{\{x \in B(x_I, r)\}}\ dx.$$

This concludes the proof of Lemma 2.3.1.3 with $c_0 := \exp[-\rho_J^\vee\ t_0]\ h_\wedge\ t_0$. □

Proof of Lemma 2.3.1.4

This proof follows similar lines as that of Lemma 2.3.1.3, so we just mention the adjustments. First, $\Delta t := S/v = t_0$. Of course, the result relies on the Markov property combined with a uniform estimate on the transitions starting from x_I with $x_I \in B(x, r)$. For any $x_F \in B(x,\, r)$, since $|x_I - x_F| \leq 2r$, we ensure that X_{t_0} has a lower-bounded density on $B(x_F, r)$. As previously, still with the fact that :

$$\left\{x_I - v\ s\ e_1 \,\middle|\, s \leq t_0\right\} \cup \left\{x_I - v\ s\ e_1 + w \,\middle|\, s \leq t_0\ ,\ w \in B(S\, e_1,\ 4r)\right\} \subset \mathcal{D}_L,$$

we deduce :

$$\mathbb{P}_{x_I}\left(X(t_I) \in dy\ ;\ t_0 < T_{\mathcal{D}_L}\right) \geq c_0\, \mathbf{1}_{\{y \in B(x_I,\ 4r)\}}\, dy \geq c_0\, \mathbf{1}_{\{y \in B(x_F,\ r)\}}\, dy.$$

By the Markov property, if $(t, x) \in \mathcal{R}^{(L)}(c)$, then $(t + t_0, x_F) \in \mathcal{R}^{(L)}(c \times c_0)$. It concludes the proof of Lemma 2.3.1.4. □

Appendix B : Control on the the pure jump process

Proof of Lemma 2.3.2.4

The proof is based on an induction on the coordinates affected by jumps in the time-interval $[0, t_{\bar{\wedge}}]$. We recall that, thanks to our criterion of exceptionality, we can restrict ourselves to trajectories where any coordinate is affected by at least one jump in the time-interval $[0, t_{\bar{\wedge}}]$, while at most $n_J^\vee$ jumps have occurred in this time-interval. There is clearly a finite number of possible sequences of directions that the process follow at each successive jumps. In order to deduce the upper-bound on the density of $X(U_A)$ presented in Lemma 2.3.2.4, we merely need to prove the restricted versions for any such possible sequence of directions.

So let U_J^k be the k-th jump of X. Let also $i(k) \in [\![1, d]\!]$ for $k \leq n_J \leq n_J^\vee$ be a sequence of directions such that, at $k = n_J$, all the directions have been listed. Let also $I(k) \in [\![1, d]\!]$ for $k \leq n_J^\vee$, be the sequence of random directions that the successive jumps of X follow. Since in our model, all directions are defined in a similar way, we can simplify a bit our notations without loss of generality by relabeling some of the directions. Since we will go backwards to progressively forget about the conditioning, we order the coordinates by the time they appear for the last time in $(i(k))_{k\leq n_J}$. It means that $i(n_J) = d$ and $\{i(k) \ ; \ K \leq k \leq n_J\} = [\![j(K), d]\!]$. Let then $K(j)$ be the largest integer k for which $j(k) \geq j$.

Remark : In our case, n_J is naturally chosen as the first integer for which all the directions have been listed. Yet, our induction argument is more clearly stated if we do not assume this condition on n_J.

Then, for any positive and measurable functions $(f_j)_{j\leq d}$ and any $x \in E$:

$$\begin{aligned} E^d := \mathbb{E}_x\Bigg[\prod_{j\leq d} f_j[X^j(U_J^{n_J})] \ ; \ U_J^{n_J} < \tau_\partial \wedge t_{\bar\wedge}, \\ \forall\, k \leq n_J,\ I(k) = i(k) \ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Bigg] \end{aligned}$$

With $\mathcal{F}^*_{U_J^{n_J}} := \sigma(\mathcal{F}_{U_J^{n_J-1}}, \{I(n_J) = d\} \cap \{U_J^{n_J} < \tau_\partial \wedge t_{\bar\wedge}\}$, the latter is upper-bounded by :

$$\mathbb{E}_x\Bigg[\prod_{j\leq d-1} f_j[X^j(U_J^{n_J-1})] \times \mathbb{E}_x\Big[f_d[X^d(U_J^{n_J})] \ ; \ |\Delta X^d(U_J^{n_J})| \leq w_\vee \,\Big|\, \mathcal{F}^*_{U_J^{n_J}} \Big] \ ; \tag{2.6.1}$$
$$U_J^{n_J} < \tau_\partial \wedge t_{\bar\wedge} \ , \ I(n_J) = d \ , \ \forall\, k \leq n_J - 1,\ I(k) = i(k) \ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Bigg],$$

Note that $X(U_J^{n_J}-) = X(U_J^{n_J-1})$ is $\mathcal{F}_{U_J^{n_J-1}}$-measurable, since we consider a pure jump process. By the Markov property, the law of the next jump only depends on $x' = X(U_J^{n_J-1})$ through the functions $(w \mapsto g_j(x', w))_{j\leq d}$. With the σ-algebra $\mathcal{F}^*_{U_J^{n_J}}$, we include the knowledge of the direction of the jump at time $U_J^{n_J}$, so that only the size of this jump (possibly negative) remains random. Noting that : $\|X^d(U_J^{n_J}-)\|_\infty \vee \|X^d(U_J^{n_J})\|_\infty \leq \ell_E + n_J^\vee \times w_\vee := L$ (independent of n_J or of the particular choice of $i(k)$), we deduce from assumption (JC^4) that on the event $\{I(n_J) = d\} \cap \{U_J^{n_J} < \tau_\partial \wedge t_{\bar\wedge}\}$:

$$\mathbb{E}_x\Big[f_d[X^d(U_J^{n_J})] \ ; \ |\Delta X^d(U_J^{n_J})| \leq w_\vee \,\Big|\, \mathcal{F}^*_{U_J^{n_J}} \Big] \leq g_\vee \int_{[-L,L]} f_d(x^d)\, dx^d. \tag{2.6.2}$$

In the following, we upper-bound by 1 the probability of the event $\{I(n_J) = d\} \cap \{U_J^{n_J} < \tau_\partial \wedge t_{\bar\wedge}\}$. Combining inequalities (2.6.1), (2.6.2), and our ordering with the definition of $K(j)$, we deduce :

$$\begin{aligned} E^d \leq g_\vee \int_{[-L,L]} f_d(x^d)\, dx^d \times \mathbb{E}_x\Big[\prod_{j\leq d-1} f_j[X^j(U_J^{n_J-1})] \; ; \; U_J^{n_J-1} < \tau_\partial \wedge t_A, \\ \forall\, k \leq n_J - 1,\ I(k) = i(k)\ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Big] \\ \leq g_\vee \int_{[-L,L]} f_d(x^d)\, dx^d \times \mathbb{E}_x\Big[\prod_{j\leq d-1} f_j[X^j(U_J^{K(d-1)})] \; ; \; U_J^{K(d-1)} < \tau_\partial \wedge t_A, \\ \forall\, k \leq K(d-1),\ I(k) = i(k)\ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Big], \end{aligned}$$

where in particular $i(\, K(d-1)\,) = d-1$, $K(d-1) \leq n_J^\vee$.

We see that quite an immediate recursion ensures :

$$\begin{aligned} E^{(d-1)} = \mathbb{E}_x\Big[\prod_{j\leq d-1} f_j[X^j(U_J^{K(d-1)})] \; ; \; U_J^{K(d-1)} < \tau_\partial \wedge t_A, \\ \forall\, k \leq K(d-1),\ I(k) = i(k)\ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Big] \\ \leq g_\vee \int_{[-L,L]} f_{d-1}(x^{d-1})\, dx^{d-1} \times \mathbb{E}_x\Big[\prod_{j\leq d-2} f_j[X^j(U_J^{K(d-2)})] \; ; \\ U_J^{K(d-2)} < \tau_\partial \wedge t_A\ , \ \forall\, k \leq K(d-2),\ I(k) = i(k)\ , \ \|\Delta X(U_J^k)\| \leq w_\vee \Big], \end{aligned}$$

and so on until finally :

$$E^d \leq (g_\vee)^d \times \prod_{i\leq d} \left(\int_{[-L,L]} f_i(x)\, dx \right).$$

We then sum over all possible sequences $i(k)$ for $k \leq n_J \leq n_J^\vee$ where n_J is the first integer for which all the d coordinates have been visited. There are clearly less than $d^{n_J^\vee}$ possibilities (surjection from the set of all sequences of length $n_J^\vee$). Since for any positive and measurable functions $(f_j)_{j\leq d}$, we have for any $x \in E$:

$$\mathbb{E}_x\left[\prod_{j=1}^d f_j[X(U_A)] \; ; \; U_A < \tau_\partial \right] \leq d^{n_J^\vee} \times (g_\vee)^d \int_{\bar{B}(0,L)} \prod_{j=1}^d f_j(x_j) dx_1 ... dx_d,$$

it is classical that it implies :

$$\forall\, x \in E, \quad \mathbb{P}_x\left[X(U_A) \in dx \; ; \; U_A < \tau_\partial \right] \leq d^{n_J^\vee} \times (g_\vee)^d \, \mathbf{1}_{\left\{x\in \bar{B}(0,L)\right\}}\, dx.$$

It concludes the proof of Lemma 2.3.2.4. □

Remark : It is quite conceivable that the distinction between all the possible choices of directions introduces quite artificially the combinatorial factor $d^{n_J^{\vee}}$*, which is certainly very rough. This distinction is however very efficient in order to simplify the proof.*

3 Adaptation of a population to a changing environment under the light of quasi-stationarity

This chapter is taken from the preprint with the same name whose ArXiv reference can be found at the end of the bibliography (here is the link for the pdf version : [Chapter 3]). Additional results of simulation are provided in Appendix at the end of the manuscript, just before the bibliography.

Abstract

We consider a model of diffusion with jumps intended to illustrate the adaptation of a population to the variation of its environment. A specific process for adaptation is coupled to a Feller logistic diffusion whose growth rate declines as the optimal trait is further away. Extinction is then more likely to occur. Assuming that our deterministic environment is changing regularly in a constant direction, we obtain the existence and uniqueness of the quasi-stationary distribution, the associated survival capacity and the Q-process. Our approach provides moreover several results of exponential convergence (in total variation for the measures). From these summary information, we can characterize the efficiency of internal adaptation (i.e. renewal of the population from the invasions of mutants). When the latter is lacking, there is still stability, yet due to the high level of population extinction. Different features then emerge.

3.1 Introduction

3.1.1 Eco-evolutionary motivations

We aim at studying the relative contribution of mutations with various strong effects to the adaptation of a population. So our purpose is to analyze a model as simple as possible on which these mutations are filtered depending on the advantage they provide. This advantage can be immediately significant (better growth rate of the mutant sub-population) or play a role in the future adaptation (the population is doomed without mutants). The stochastic model under consideration takes those two aspects into account. It extends the one first introduced by [KH09] and then more formally described in [NP17] and [KNP18].

Likewise, we assume that the population is described by some trait $\hat{x} \in \mathbb{R}^d$. In a view for a simple theoretical model, spatial dispersion as well as phenotypic heterogeneity (at least for the traits of interest) are neglected. So we assume that the population is monomorphic at all times and $\hat{x}$ shall then represent the phenotype of the individuals in the population. Nonetheless, we allow variations of this trait $\hat{x}$ due to stochastic events, namely when a subpopulation issued from a mutant with trait $\hat{x} + w$ manages to subsist and invade the "resident" population. In the model, such events are assumed to occur instantaneously.

The main novelty of our approach is that we couple this "adaptative" process with a Feller diffusion N with a logistic drift. This diffusion shall describe the dynamics of the population size in a limit where it is large. We mean here that individual events of birth and death have negligible impact, but that the accumulation of these events has a visible and stochastic effect. Introducing the "size" in the model enables us notably to easily translate the notion of mal-adaptation, in the form of a poor growth rate.

For the long-time dynamics, we are mainly interested in considering only surviving populations, that is conditioning the process upon the fact that the population size has not come down to 0. The implication of considering the size is then twofold. On the one hand, the extinction occurs way more rapidly when adaptation is poor. Indeed, the population size is then very rapidly declining. One may thus observe an effect of natural selection at the level of the population. On the other hand, the better the adaptation is, the larger the population size is able to reach and the more frequently the birth of new mutants occurs in the population. In our simple model, a trait that is better suited for the survival of the whole population shall also have a greater probability to invade a resident population, once a single mutant is introduced. Compared to the case of a fixed size as in [NP17] and [KNP18], this second implication means a stabilizing effect for the phenotype when the population size is sufficiently large ; but also a destabilizing effect when the population size declines. This opposes to the natural selection at the level of the individuals among the population (which is the main effect detailed in [KH09]). Indeed, when the adaptation is already quite optimal, very few among the mutants arising among the population can manage to maintain themselves and finally invade the resident trait.

Let us assume here that the mutations can allow the individuals to survive in these new environments. In this context, how resilient is the population while facing environmental changes ? Is there a clear threshold to the rate of change such population can manage ? How could we describe the interplay between the above-mentioned properties ?

To begin at answering these issues, and like [KH09], we assume for simplicity that the environmental change is given by a translation at constant speed v of the profile of fitness. In practice, it means that the growth rate of the population at time t is expressed as a function of $x := \hat{x} - v\,t$, for a monomorphic population with trait $\hat{x}$

at time t. Naturally, the phenotypic lag x becomes the main quantity of interest for varying t. Likewise, we can express as a function of x and w the probability that an individual mutant, with mutation w, manages to invade a resident population with trait $\hat{x}$ at time t. This probability should indeed be related to the difference between the growth rate at x and at $x + w$, although we will not require any precise relationship in the results of this paper. Moreover, we assume that the distribution of the additive effect for the new mutations is constant in time and independent of the trait $\hat{x}$ of the population before the mutation (thus independent of x in the moving referential).

In this context, we can exploit the notion of quasi-stationary distribution (QSD) to characterize what would be an equilibrium for this dynamics previous to extinction. The main contribution of the current paper is to ensure that this notion is here defined without ambiguity. To our knowledge, this is the first time that existence and uniqueness of the QSD is proved for a piecewise deterministic process coupled with a diffusion.

By our proof, we also provide a justification of the notion of typical relaxation time and extinction time. The quasi-stationary description is well suited provided the latter is way longer than the former. As can be checked from simulations, typical convergence to the QSD is exponential in such cases. Still, the marginal starting from some initial conditions might take long before it approximates the QSD, yet mainly in cases where extinction is initially large. Several implications from the theory of QSD to this application shall be discussed in Section 3.2.3 in this evolutionary perspective, and some limitations as well.

In the next subsections of the introduction, we present the stochastic process under consideration then some elementary notations. The main results are described in Section 3.2, starting with our hypothesis in Subsection 3.2.1 and the statement of the Theorem in Section 3.2.2. In Subsection 3.2.3, we discuss its interpretation in terms of ecology and evolution. Its connection to related models of adaptation is given in Subsection 3.2.4, and to the classical techniques of quasi-stationarity in Subsection 3.2.5. The rest of the paper deals with the proofs. We prove existence and uniqueness of the process in Section 3.3, and introduce in next Section 3.4 the main Theorems on which our main Theorem 3.2.1 rely. Two sets of hypothesis are considered, with some variations in the proofs. We choose to gather the Theorems in the three following Sections depending on the features of the process they imply for both sets of hypothesis. Some of the remaining proofs are left to the Appendix.

3.1.2 The stochastic model

Following [KH09] as explained in the introduction for the definition of the adaptive component, the system that describes the combined evolution of the population size and of its phenotypic lag is then given by :

$$(S_0)\begin{cases} X_t = x - v\,t\,\mathbf{e_1} + \displaystyle\int_{[0,t]\times\mathbb{R}^d\times(\mathbb{R}_+)^2} w\,\varphi_0\left(X_{s^-},\,N_s,\,w,\,u_f,\,u_g\right) M(ds, dw, du_f, du_g) \\ N_t = n + \displaystyle\int_0^t \Big(r(X_s)\;N_s - \gamma_0 \times (N_s)^2\Big)\,ds + \sigma \int_0^t \sqrt{N_s}\;dB_s, \end{cases}$$

where N_t describes the size of the population and X_t the phenotypic lag of this population.

Here, $v > 0$, B_t is a standard $\mathcal{F}_t$ Brownian motion and M is a Poisson Point Process (PPP) on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+$, also adapted to $\mathcal{F}_t$, with intensity :

$$\pi(ds,\,dw,\,du_f,\,du_g) = ds\;\nu(dw)\;du_f\;du_g,$$

where $\nu(dw)$ is a measure describing the distribution of new mutations, and :

$$\varphi_0(x,\,n,\,w,\,u_f,\,u_g) = \mathbf{1}_{\left\{u_f \le f_0(n)\right\}} \times \mathbf{1}_{\{u_g \le g(x,w)\}}.$$

In the model of the moving optimum originally considered in [KH09], $X = 0$ corresponds to the optimal state in terms of reproductive value, while this function is symmetric and decreasing with $\|X\|$. X is described as the phenotypic lag because $X_t + v\,t$ is the character of the individuals at time t in the population while in this original model, the mobile optimum is located at trait $v\,t$. These assumptions on the fitness landscape are natural, and we abide by them in our simulations. Nonetheless, they are mainly assumed for simplicity and we have chosen here to be as general as possible in the definition of r. X_t is thus a lag as compared to the trait $v\,t$ that is merely a reference value.

$g(X_t, w)$ is the mutation kernel that describes the rate of fixation at which a mutant sub-population of trait $X_t + v\,t + w$ invades a resident population of trait $X_t + v\,t$. Although the rate at which the mutations occur in one individual can reasonably assumed to be symmetrical in w, it is clearly not the case for g. In a large population, the filtering of considering only fixing mutations highly restricts the occurrence of strongly deleterious mutations, strongly favors strongly advantageous mutations. For mutations with little effects, there is only a slight bias. To cover both of these situations, we consider in our analysis both the case where any mutation effect is permitted and the case where only advantageous ones are. Although the latter case will raise more difficulty in terms of accessibility of the domain, the core of the argument is quite the same and the simulations seem to provide similar results in both cases.

The term $f(N_t)$ is introduced to model the fact that for a constant mutation rate by individual, the mutation rate for the population is all the larger than the population size is large. $f(N_t) := N_t$ is the first reasonable choice, but we may also be interested in introducing an effect of the population size in the fixation rate.

N follows the equation of a Feller logistic diffusion where the growth rate r at time t only depends on X_t, while the strength of competition c and the coefficient of diffusion σ are kept constant. Such a process is the most classical ones for the dynamics of a large population size in a continuous space setting and such that explosion is prevented. It is described in [La05] (with fixed growth rate), notably as a limit of some individual-based model. σ is related to the proximity between to uniformly sampled individuals in terms of their filiation links : $1/\sigma^2$ scales as the population size and is sometimes describes as the "effective population size".

From a biological perspective, X has no reason to explode. Under our assumption $[H11]$ below, such explosion is clearly prevented. Yet, we won't focus on conditions ensuring non-explosion for X. Indeed, it would mean (by assumption $[H8]$ below) that the growth rate becomes extremely negative. It appears very natural to consider that it would lead to the extinction of the population. So, we define the extinction time as :

$$\tau_\partial := \inf\{t \geq 0,\ N_t = 0\} \wedge \sup_{\{k\geq 1\}} T_X^k, \qquad \text{where } T_X^k := \inf\{t \geq 0,\ \|X_t\| \geq k\}. \tag{3.1.1}$$

Because it simplifies many of our calculations, in the following, we will consider $Y_t := \frac{2}{\sigma}\sqrt{N_t}$ rather than N_t.

Lemma 3.1.2.1. *With the previous notations, (X, Y) satisfies the following SDE :*

$$(S) \begin{cases} X_t = x - v\, t\, \mathbf{e_1} + \displaystyle\int_{[0,t]\times\mathbb{R}^d\times(\mathbb{R}_+)^2} w\, \varphi\left(X_{s^-},\, Y_s,\, w,\, u_f,\, u_g\right)\ M(ds, dw, du_f, du_g) \\ Y_t = y + \displaystyle\int_0^t \psi\left(X_s,\, Y_s\right) ds + B_t \end{cases}$$

where we define : $\psi(x,y) = -\dfrac{1}{2\,y} + \dfrac{r(x)\,y}{2} - \gamma\, y^3,$ *with* $\gamma := \dfrac{\gamma_0\,\sigma^2}{8}$

$$\varphi(x,\, y,\, w,\, u_f,\, u_g) := \varphi_0\left(x,\, \sigma^2 y^2/4,\, w,\, u_f,\, u_g\right),$$

Thus with $f(y) := f_0[\sigma^2 y^2/4]$ *,* $\qquad \varphi(x,\, y,\, w,\, u_f,\, u_g) = \mathbf{1}_{\left\{u_f \leq f(y)\right\}} \times \mathbf{1}_{\{u_g \leq g(x,w)\}}.$

We let the reader prove this Lemma by a simple application of the Ito formula.

The aim of the following theorems is to describe the law of the marginal of the process (X, Y) at large time t conditionally upon the fact that the extinction has not occurred, in short the MCNE at time t. Considering the conditioning at the current time leads to considering properties of quasi-stationarity ; while a conditioning at a much more future time leads to a Markov process usually referred to as the Q-process, in some sense the process conditioned on never going extinct. The two aspects are clearly complementary and our approach will treat both in the same framework, in the spirit initiated by [CV16].

3.1.3 Elementary notations

In the following, the notation $k \geq 1$ is to be understood as $k \in \mathbb{N}$ while $t \geq 0$ –resp. $c > 0$– should be understood as $t \in \mathbb{R}_+ := [0, \infty)$ –resp. $c \in \mathbb{R}_+^* := (0, \infty)$. In this context (with $m \leq n$), we denote classical sets of integers by : $\mathbb{Z}_+ := \{0, 1, 2...\}$, $\mathbb{N} := \{1, 2, 3...\}$, $[\![m, n]\!] := \{m, m+1, ..., n-1, n\}$, where the notation $:=$ makes explicit that we define some notation by this equality. For maxima and minima, we usually denote : $s \vee t := \max\{s, t\}$, $s \wedge t := \min\{s, t\}$. Accordingly, for a function φ, $\varphi^\wedge$ –resp. $\varphi^\vee$– will be used for a lower-bound –resp. for an upper-bound– of φ. Numerical indices are rather indicated in superscript, while specifying notations are often in subscript.

3.2 Exponential convergence to the QSD

3.2.1 Hypothesis

We will consider two different sets of assumptions, including or rejecting the possibility for deleterious mutations to invade the population. We gather all of them in this subsection :

[$H1$] $f \in \mathcal{C}^0\Big(\mathbb{R}_+^*, \mathbb{R}_+\Big)$ (the condition is equivalent for f_0).

[$H2$] r is locally Lipschitz-continuous on $\mathbb{R}^d$.

[$H3$] $g \in \mathcal{C}^0\Big(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+\Big)$ and is bounded on any $K \times \mathbb{R}^d$, where K is a compact set of $\mathbb{R}^d$.
(1 is the natural bound with the biological interpretation, yet an extension may be needed when g is not exactly the fixation probability, cf. Corollary 3.2.2.2)

[$H4$] $\nu(\mathbb{R}^d) < \infty$.
Under [$H4$], on any finite time-interval, only a finite number of mutations can occur.

[$H5$] $\forall\, y > 0, \quad f(y) > 0$
It is quite natural to assume that $f(0) = 0$ and $f(y) \to \infty$ as $y \to \infty$, but we will not need those assumptions.

[$H6$] There exist $S, \nu_\wedge > 0$ and $0 < \delta S < S$ such that $\nu(dw) \geq \nu_\wedge\, \mathbf{1}_{B(S+\delta S) \setminus B(S-\delta S)}\, dw$, where $B(R)$, for $R > 0$, denotes the ball of radius R centered at the origin.

[$H7$] $r(x) \to -\infty$, as $\|x\| \to \infty$.

[$H8$] g is positive.
This means that even deleterious mutations can invade the population.

[$H9$] This assumption states that deleterious mutations cannot invade the population, but that advantageous ones always have some chance to do so :

$$\forall\, x, w \in \mathbb{R}^d, \quad \begin{array}{lcl} \|x+w\| < \|x\| & \Rightarrow & g(x,w) > 0 \\ \|x+w\| \geq \|x\| & \Rightarrow & g(x,w) = 0. \end{array}$$

[$H10$] $\nu(dw) << dw$ and the density $g(x,w)\,\nu(w)$ (for a jump from x to $x+w$), of the jump size law w.r.t. Lebesgue's measure, satisfies :

$$\forall\, x_\vee > 0, \quad \sup\left\{ \frac{g(x,w)\,\nu(w)}{\int_{\mathbb{R}^d} g(x,w')\,\nu(w')\,dw'} \;;\; \|x\| \leq x_\vee,\, w \in \mathbb{R}^d \right\} < \infty.$$

This last hypothesis will only be needed in the case $d \geq 2$.

3.2.2 Statement of the main Theorem

Proposition 3.2.2.1. *Assume that* [H1-4] *hold. Then, for any initial condition* $(x,y) \in \mathbb{R}^d \times \mathbb{R}_+^*$, *there is a unique strong solution* $(X_t, Y_t)_{t \geq 0}$ *satisfying* (S) *for any* $t < \tau_\partial$, *and* $X_t = Y_t = 0$ *for* $t \geq \tau_\partial$, *where* $\tau_\partial := \sup_{\{n\geq 1\}} T_Y^n \wedge \sup_{\{n \geq 1\}} T_X^n$, $T_Y^n := \inf\{t \geq 0,\ Y_t \leq 1/n\}$, $T_X^n := \inf\{t \geq 0,\ \|X_t\| \geq n\}$.

Theorem 3.2.1. *Assume that* [H1-7] *hold. Suppose that either* [H8] *or* [H9] *holds. If* $d \geq 2$, *assume finally that* [H10] *holds. Then, there exists a unique QSD* α. *Moreover, we have exponential quasi-ergodicity, i.e. :*

(i) Exponential convergence to the survival capacity

With λ *the extinction rate associated to the QSD, let :* $h_t(x,y) := e^{\lambda\, t}\mathbb{P}_{x,y}(t < \tau_\partial)$. *This sequence of functions converges exponentially in the uniform norm to a function* h, *that we call the survival capacity.* h *is positive and bounded on* $\mathbb{R}^d \times \mathbb{R}_+^*$ *and vanishes on* ∂. *It also belongs to the domain of the infinitesimal generator* $\mathcal{L}$, *associated with the semi-group* (P_t) *on* $(B(\mathbb{R}^d \times \mathbb{R}_+^* \cup \{\partial\}); \|.\|_\infty)$

$$\textit{with } \mathcal{L}\, h = -\lambda\, h, \qquad \textit{so} \quad \forall\, t \geq 0, \quad P_t\, h = e^{-\lambda\, t} h$$

(ii) Exponential convergence to the QSD

For some $\gamma, C > 0$ *and with the survival capacity* h *:*

$$\forall\, \mu \in \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}_+^*),\ \forall\, t > 0, \tag{3.2.1}$$
$$\|\, \mathbb{P}_\mu\left[\, (X_t, Y_t) \in (dx, dy) \mid t < \tau_\partial\right] - \alpha(dx,dy)\, \|_{TV} \leq C \frac{\inf_{u>0} \|\mu - u\alpha\|_{TV}}{\langle \mu \,\big|\, h \rangle}\, e^{-\gamma\, t}.$$

The relation between α *and its extinction rate is the following :*

$$\forall\, t \geq 0, \quad \alpha P_t = e^{-\lambda\, t}\, \alpha, \qquad \text{and in particular } \mathbb{P}_\alpha(t < \tau_\partial) = e^{-\lambda\, t} \tag{3.2.2}$$

(iii) Existence of the Q-process and associated transition kernel
There exists a family $(\mathbb{Q}_{x,y})_{(x,y)\in\mathbb{R}^d\times\mathbb{R}_+^*}$ of probability measures on Ω defined by

$$\lim_{t\to\infty}\mathbb{P}_{x,y}(\Lambda_s \,\big|\, t<\tau_\partial) = \mathbb{Q}_{x,y}(\Lambda_s) \tag{3.2.3}$$

for all $\mathcal{F}_s$-measurable set Λ_s. The process $(\Omega; (\mathcal{F}_t)_{t\geq 0}; (X_t, Y_t)_{t\geq 0}; (\mathbb{Q}_{x,y})_{(x,y)\in\mathbb{R}^d\times\mathbb{R}_+^*})$ is a homogeneous strong Markov process. Its transition kernel is given by :

$$q(x,y;t;dx',dy') = e^{\lambda\, t}\,\frac{h(x',y')}{h(x,y)}\,p(x,y;t;dx',dy'), \tag{3.2.4}$$

where $p(x,y;t;dx',dy')$ is the transition kernel of the Markov process (X,Y) under $(\mathbb{P}_{x,y})$.

(iv) Exponential ergodicity of the Q-process :
There is a unique invariant distribution of X under $\mathbb{Q}$, given by : $\beta(dx,dy) := h(x,y)\,\alpha(dx,dy)$.
Moreover, with the same constants $\zeta>0$ and C as in (3.2.1) :

$$\|\mathbb{Q}_\mu\,[\,(X_t,Y_t)\in(dx,dy)] - \beta(dx,dy)\|_{TV} \leq C\,\inf_{u>0}\|\mu - u\,\beta\|_{1/h}\,e^{-\gamma\,t}, \tag{3.2.5}$$

$$\text{where } \mathbb{Q}_\mu(dw) := \textstyle\int_{\mathcal{X}}\mu(dx,dy)\,\mathbb{Q}_{(x,y)}(dw), \quad \|\mu\|_{1/h} := \|\dfrac{\mu(dx,dy)}{h(x,y)}\|_{TV}$$

What is more, all the preceding results hold true with the same constants involved in the convergences when τ_∂ is replaced by $\tau_\partial^L := \tau_\partial\wedge T_{\mathcal{D}_L}$ for any $L\geq 1$ sufficiently large. The associated QSD α^L, extinction rates λ_L and survival capacities h^L converge respectively to α (in total variation), λ and h (punctually at least) as $L\to\infty$.

Remarks 3.2.2.1. • *For the total variation norm, considering (X,Y) or (X,N) is equivalent.*

• *Since r goes to $-\infty$ as $\|x\|$ goes to infinity, it is natural to assume that mutations leading X to be large have a very small probability of fixation. Notably, it means that we highly expect the upper-bound of g in $[H3]$, uniform over w.*

• *Under hypothezis $[H9]$, one may expect the real probability of fixation $g(x,w)$ to be at most of order $O(\|w\|)$ for small values of w (and locally in x). In such a case, we can allow ν to satisfy a smaller integrability condition than $[H4]$ while forbidding observable accumulation of mutations.*

Corollary 3.2.2.2. *Assume that $[H1\text{-}3]$, $[H5\text{-}7]$ and $[H9]$ hold. Suppose that $\int_{\mathbb{R}}(|w|\wedge 1)\,\nu(dw)<\infty$ while $\widetilde{g}:(x,w)\mapsto g(x,w)/(|w|\wedge 1)$ is bounded on any $K\times\mathbb{R}^d$ for K a compact set of $\mathbb{R}^d$. If $d\geq 2$, assume additionally that $[H10]$ holds. Then, all the results of Theorem 3.2.1 hold.*

Proof (X, Y) is solution of (S) iff it is solution of :

$$(S) \begin{cases} X_t = x - v\, t\, \mathbf{e_1} + \int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} w\, \widetilde{\varphi}\,(X_{s^-},\, Y_s,\, w,\, u_f,\, \widetilde{u_g})\;\; \widetilde{M}(ds, dw, du_f, du_g), \\ Y_t = y + \int_0^t \psi\,(X_s,\, Y_s)\, ds + B_t, \end{cases}$$

where $\widetilde{M}$ is a PPP of intensity $ds\; \widetilde{\nu}(dw)\; du_f\; d\widetilde{u_g}$,

$$\widetilde{\nu}(dw) := \nu(dw)/(\|w\| \wedge 1)\,, \qquad \widetilde{\varphi}(x,\, y,\, w,\, u_f,\, \widetilde{u_g}) = \varphi(x,\, y,\, w,\, u_f,\, \widetilde{u_g}\times(\|w\| \wedge 1)),$$

where $\widetilde{\varphi}$ is defined as φ with g replaced by $\widetilde{g}$.

By the condition on ν, $[H4]$ holds with $\widetilde{\nu}$ instead of ν. By the condition on g, $[H3]$ still holds with $\widetilde{g}$ instead of g. Conditions $[H9]$ and $[H10]$ are equivalent for the systems (g, ν) and $(\widetilde{g}, \widetilde{\nu})$. Consequently, if we prove Theorem 3.2.1 with $[H3]$ and $[H4]$, the results follow under the assumptions of Corollary 3.2.2.2. $\square$

3.2.3 Eco-evolutionary implications of these results

One of the major motivation for the present analysis is to make a distinction, as rigorous as possible, between an environmental change that the population can spontaneously adapt to and one that imposes too much a pressure. We recall that in [NP17], the authors obtain a clear and explicit threshold on the speed of this environmental change. Namely, above this speed, the Markov process that they consider is transient, whereas it is recurrent below this critical speed. Thus, it might seem a bit frustrating that such a distinction cannot be observed in the previous theorems (which seem not to care about this speed). At least, these results prove that the distinction is not based on the existence nor the uniqueness of the QSD, and even not on the exponential convergence per se. In fact, this threshold is so distinct in [NP17] because their model relies on the following underlying assumption : The poorer the current adaptation is, the more effectively mutations are able to fix, provided that they are then beneficial. In our case, a population too poorly adapted is almost doomed to a quick extinction, because the population size cannot be maintained at large values. The survival is rather triggered by dynamics that maintain the population adapted.

Looking back at the history of surviving populations, we shall observe that the process has mostly stayed confined outside of deadly areas. This can be seen by comparing the QSD α to the stationary distribution of the Q-process. The latter can also be seen as the stationary distribution of some related backward in time process. This comparison sheds light on which components of the QSD are to disappear very quickly due to extinction. ... First results of simulations indicate a strong effect of confinement when the extinction of the populations starts to play a significant role. It even affects areas where, assuming a constant level of mal-adaptation, the population could survive for quite long at this smaller growth rate. These regions are visibly more unstable than the core areas where the backward process stays confined in. This is probably due to this decline in population size when the level of mal-adaptation gets larger. This confinement by the conditioning upon survival only weakens in the recent past, because the very probable extinction may come only in the future.

In order to establish this distinction between environmental changes that are sustainable and those which endanger the population, we need a criterion that quantifies the stability of such core regions. Our results provide two exponential rates whose comparison is enlightening : if the extinction rate is of the same order as the convergence rate or larger, it means that the dynamics is strongly dependent upon the initial condition. If the convergence happens much quicker, the dynamics shall rapidly become similar whatever the initial condition. This is at least the case for not too risky initial conditions (i.e. where h is not too small). This criterion takes into account the intrinsic sustainability of the mechanisms involved in the adaptation to the current environmental change, but does not involve the specific initial state of adaptation.

Looking at simulation results, the convergence in total variation indeed appears to happen at some exponential rate, provided that extinction does not erase brutally a large part of the distribution at a given time. To obtain a generic estimation of this exponential rate, we plan to exploit the decay in time of the correlations of X and N starting from the QSD profile. Our interest in these correlations in time comes from the fact that the exponential decay starts immediately and appears quite stable in our simulations. On the other hand, the extinction rate is much easier to estimate. At least in the case where $\mathcal{X}$ is of dimension one, we can directly estimate the dynamics of the density and then of the extinction rate.

We have to mention that the relevance of such a criterion is not necessarily guaranteed for any type of growth rate function. We shall exclude notably the case where there are several basins of attraction separated by deadly areas. Indeed, in such a case, the adaptation might be well-observed for each of the occupied basin, while the convergence rate ζ is very small because of the transitions between the basins. The transition rates are possibly even smaller than λ itself. Nonetheless, when there is a unique basin, our simulations tend to confirm the relevance of this criteria. Besides, it is quite reassuring to see that including or not deleterious mutations (for which the invasion probability is expected be positive but very small) is not crucial in the present proof. We do not see much difference while looking at simulations.

The project of detailing these simulation results is in progress, but let us mention a few other hints from the aspect of the QSD and the survival capacity. For simplicity, suppose that the environment is sufficiently beneficial for a well-adapted population to sustain quite a large carrying capacity, so that the risk of extinction is almost negligible as long as the adaptation does not become too poor. Notably, confronted to a moderate speed, the population size given by the QSD is still quite large with a high probability, meaning a very small extinction rate (by (3.2.2) λ is also the "instantaneous" death rate of the QSD). The phenotypic lag appears in some sense pushed all the more than v increases. Actually, neglecting the extinction, the emergence of successful mutations shall compensate the effect of the drift v on average over the QSD, i.e. (from $\langle \alpha \,\big|\, \mathcal{L}h \rangle \approx 0$ with $h(x,y) = x$) :

$$v \approx \int_{\mathbb{R}^d \times \mathbb{R}} \int_{\mathbb{R}^d} w\, f(y)\, g(x,w)\, \nu(dw)\, \alpha(dx,dy). \tag{3.2.6}$$

This can only be done by increasing the mean value of $g(x,w)$, i.e. the probability of fixation, especially for large jumps w. Yet, larger probability of fixation can only be obtained when the growth rate of the resident is notably smaller than the one of the mutant, and a fortiori than the optimal growth rate. Values of x for which $g(x,w)$ is large are therefore associated to a decline of the population size, implying an overall reduction of the arrival of mutations (the term $f(y)$). At some point, the increased probability that large mutations invade is in fact completely compensated by the associated reduction of $f(y)$ (on average given x). Then, the role of extinction becomes quite suddenly essential. For large v, we thus observe that the population size is with some large probability very close to zero, which illustrates this large extinction rate.

This effect can be interpreted from the survival capacity, in its relation with the Q-process. When we condition on the survival in the long run, it shall have nearly no effect on the randomness observed in an already well-adapted population able to sustain the future change. It means that $h(x+w,y)/h(x,y)$ shall be close to 1 for (x,y) typical to the QSD, and w around the range of $\|x\|$ or smaller. Likewise, h shall not vary much in the y direction (as long as y is not too small). Indeed, we observe for small v a domain supporting most of the QSD where h is very flat. On the other hand, since the fixation of mutations is needed to compensate the environmental change, while the population size is expected to be quite small, these events of fixation must be largely amplified in the Q-process as compared to the original process. It means that for typical values of (x,y) and at least some effect sizes, we have $h(x+w,y)/h(x,y)$ much larger than 1. And it can be seen in the simulations that the derivative of $\log[h]$ along the x axis becomes very large for large values of v. This increase of $\log[h]$ along the x axis is observable in a large range of values : at least from the left border of the QSD to beyond the optimal trait to the right. It appears only limited to the right by the immediate risk of extinction in areas far from the optimum. Since the trajectories of X get drifted towards the left, the further is the initial condition to the right (provided that it does not vanish right away), the longer the population can maintain reasonably adapted.

3.2.4 Quasi-ergodicity of related models

The current paper completes the illustrations given in Sections 1.4 of Chapter 1 and 2.3 of Chapter 2. If the model of the current paper was in fact the original motivation for the techniques presented in these two papers, we can focus more closely on each of the difficulties thanks to these various illustrations. In any of them, the adaptation of the population to its environment is described by some process X solution to some SDE of the form :

$$X_t = x - \int_0^t V_s\, ds + \int_0^t \Sigma_s \cdot dB_s + \int_{[0,t]\times\mathbb{R}^d\times\mathbb{R}_+} w\, \mathbf{1}_{\{u\leq U_s(w)\}}\, M(ds, dw, du),$$

where B is a $\mathcal{F}_t$-adapted Brownian Motion and M a $\mathcal{F}_t$-adapted PPP. V_s and Σ_s a priori depend on X_s, U_s on X_{s-} and possibly on a coupled process N_t describing the population size. Like the product $f(Y_t)\, g(X_{t-}, w)$ in equation (S), one specifies in $U_t(w)$ the rate at which a mutations of effect w invades the population at time s. V_s both relates to the speed of the environmental change and to the mean effects of the mutations invading the population at time s in a limit of very frequent mutations of very small effects. Σ_s then relates to the undirected fluctuations both of the environment and in the effects of this large number of small fixating mutations.

Although X lives in an unbounded domain, the mal-adaptation of the process when it is far from the optimal position constraints X to stay confined conditionally upon survival. This effect of the mal-adaptation has been modeled either directly on the growth rate of the coupled process N_t or with some averaged description in term of some extinction rate. Such confinement property for the coupled process is in fact the main novelty of Chapter 1 and notably illustrated in Subsection 1.4. For simplicity, we considered there a locally elliptic process, for which the Harnack inequality is known to greatly simplify the proof, as observed previously for instance in [CV17b]. The proof of this confinement is of course simpler when there is a death rate going to infinity outside of compact sets. For the adaptation to this case, we refer to Subsection 2.3.1 of Chapter 2.

As discussed at the end of Subsection 2.3.1.1 of Chapter 2, we can relate the coupling of X and N to an approximation given by the autonomous dynamics of a process Y similar to X. For the approximation to be as valid as possible, the law of Y should be biased by some extinction rate (depending at time t on the value Y_t) and its jump rate should be adjusted. By these means, we would take an implicit account of what would be the fluctuations of N if X would be around the value of Y_t. This approximation is particularly reasonable when the characteristic fluctuations of N around its quasi-equilibrium are much quicker than the effect of the growth rate changing over time with the adaptation. Its validity is less clear when the extinction has a strong effect on the establishment of the quasi-equilibrium.

Yet, it shall have a quite limited effect for our concern, which is to compare the extinction rate to the rate of stabilization to equilibrium, see Subsection 3.2.3. Indeed, as long as the extinction rate is not way larger than the rate of stabilization to equilibrium, such domains of mal-adaptation are strongly avoided when looking in the past of surviving populations. We conjecture that this rate of stabilization shall therefore be almost independent of the dynamics in these domains, which is to be confirmed in future simulations. On the other hand, the population is almost doomed when it enters these domains. Thus, we also conjecture that the extinction rate shall not depend so much either on the dynamics in these deadly regions.

Assume for now that the fluctuations of N are much quicker than the change of the growth rate in the domain where the population is well-adapted. Then we conjecture that considering the autonomous process Y (including the bias by the extinction rate) instead of the coupled process (X, N) would produce very similar results : the extinction rates and the rates of stabilization to equilibrium should be close between these models, while the QSD profile of X should be similar to the one of Y.

3.2.5 The mathematical perspective on quasi-stationarity

The subject of quasi-stationarity is now quite vast and an extensive literature is dedicated to it, as suggested by the bibliography collected by Pollett [Po15]. Some overview on the subject can be found in general surveys like [CMS13], [DP13] or more specifically for population dynamics [MV12]. Yet, it seems that, even recently, very little is known for strong Markov processes both on a continuous space and in a continuous time, without some property of reversibility. This is all the more true when the process is discontinuous (because of the jumps in X) and multidimensional, since the property of reversibility becomes all the more stringent and new difficulties arise (cf e.g. Appendix A of [CCM17]).

Thus, ensuring the existence and uniqueness of the quasi-stationary distribution –QSD in the following– is already some breakthrough, and we are even able to ensure an exponential rate of convergence in total variation to the QSD and similar results on the Q-process. This model is in fact a very interesting illustration of the new technique which we exploit.

Our approach relies on the general result presented in Chapter 2, which, as a continuation of Chapter 1, has been originally motivated by this problem. In Chapter 1, the generalization of Harris recurrence property at the core of the results of [CV16] is extended to deal with exponential convergence which are not uniform with respect to the initial condition. The fine control over the MCNE has opened the way for the approach developed in Chapter 2 to deal with continuous-time and continuous-space strong Markov processes with discontinuous trajectories.

After their seminal article [CV16], these same authors have obtained quite a large range of extensions, for instance with diffusions in several dimensions [CCV18], processes inhomogeneous in time [CV18b], and different examples of processes in denumerable space notably with the use of Lyapunov functions, cf. [CV17b] or [CV17c].

Exploiting the result of [CV17c], it may be possible to ensure the properties of exponential quasi-ergodicity for such a discontinuous process as the one of this article, keeping some dependency on the initial condition. Yet, in the approach of [CV17c], the study of continuous-time and continuous space Markov process relies strongly on the Harnack inequality, and no alternative is proposed. For discontinuous processes, this inequality generally does not hold true, while the alternative given in Chapter 2 is here very efficient.

This dependency on the initial condition is biologically expected, although its crucial importance shall arise when the population is already very likely to disappear. For a broader comparison of this approach with the general literature, we refer to the introductions of [CV17c], Chapter 1 and, more specifically for discontinuous processes, of Chapter 2.

3.3 Proof of Proposition 3.2.2.1

Uniqueness :

step 1 : a priori upper-bound on the number of jumps

Assume that we have a solution $(X_t, Y_t)_{t\leq T}$ to (S) until some (stopping) time T (i.e. for any $t < T$) satisfying $T \leq t_\vee \wedge T_Y^m \wedge T_X^n$ for some $t_\vee > 0$, $m, n \geq 1$ (see Definition (3.1.1)). We know from $[H2]$ that the growth rate of the population remains necessarily upper-bounded by some $r^\vee > 0$ until T. Thus, we deduce a stochastic upper-bound $(Y_t^\vee)_{t\geq 0}$ on Y :

$$Y_t^\vee = y + \int_0^t \psi^\vee(Y_s)\,ds + B_t \quad \text{where} \quad \psi^\vee(y) = -\frac{1}{2\,y} + \frac{r^\vee\,y}{2} - \gamma\,y^3, \tag{3.3.1}$$

which is thus independent of M. Since $\psi^\vee(y) \leq r^\vee\,y/2$, it is classical that $Y^\vee$ –and a fortiori Y– cannot explode before T, see for instance Lemma 3.3 in [BM15] or [La05] where such a process is described in detail.

Under $[H3]$, the jump rate of X is uniformly bounded until T by :

$$\nu(\mathbb{R}^d)\times\sup\left\{g(x', w)\ ;\ x' \in \overline{B}(0,n), w \in \mathbb{R}^d\right\}\times\sup\{f(y')\ ;\ y' \leq \sup_{s\leq t_\vee} Y_s^\vee\} < \infty\ a.s.$$

Under $[H9]$, the jump rate is uniformly bounded in fact until T by :

$$\nu(\mathbb{R}^d)\times\sup\left\{\int_{w\in B(2\,\|x\|+2\,v\,t_\vee)} g(x', w)dw\ ;\ x' \in B(\|x\| + v\,t_\vee)\right\}\times\sup\{f(y)\,;\,y \leq \sup_{s\leq t_\vee} Y_s^\vee\} < \infty\ a.s.$$

where this bound, clear a priori on the first jump, can be extended inductively to the next ones since $\|X\|$ is necessarily decreasing at each jump.

step 2 : identification until T In any case, this means that the behavior of X until T is determined by the value of M on a (random) domain associated to an $a.s.$

finite intensity. Thus, we need a priori to consider only a finite number of "potential" jump, that we can write as $(T_J^1, W^1, U_f^1, U_g^1), ..., (T_J^K, W^K, U_f^K, U_g^K)$ in the increasing order of the times T_J^i.

From the a priori estimates, we know that :
for any $t < T_J^1$ or $t < T$ (for $K = 0$) : $\qquad X_t = x - v\, t.$
Consider the solution $\hat{Y}$ of : $\qquad \hat{Y}_t = y + \int_0^t \psi \left(x - v\, s \,\textbf{,}\, \hat{Y}_s\right) ds + B_t.$
It is not difficult to adjust the proof of [YW71] to this time-inhomogeneous setting, with $[H2]$, so as to prove the existence and uniqueness of such a solution until any stopping time $T \leq \hat{\tau}_\partial$, where $\hat{\tau}_\partial := \inf\{t \geq 0, \hat{Y}_t = 0\}$. Besides, $\hat{Y}$ is independent of M and must coincide with Y until $T_f^1 > 0$ defined just above. Since $T \leq T_Y^m$, the event $\{\hat{\tau_\partial} < T_f^1\}$ is necessarily empty. If there is no potential jump before T, i.e. $K = 0$, we have identified (X_t, Y_t) for $t \leq T$ as $X_t = x - v\, t$, $Y_t = \hat{Y}_t$. Else, at time T_J^1, we check whether $U_f^1 \leq f(\hat{Y}(T_J^1))$ and $U_g^1 \leq g(x - v\, T_J^1 \,\textbf{,}\, W^1)$. If it holds, necessarily $X(T_J^1) = x - v\, T_J^1 + W^1$, else $X(T_J^1) = x - v\, T_J^1$. Doing the same inductively for the following time-intervals $[T_J^k, T_J^{k+1}]$, we identify the solution (X, Y) until T.

step 3 : uniqueness of the global solution

Now, consider two solutions (X, Y) and (X', Y') of (S) defined up to respectively τ_∂ and τ'_∂ as in Proposition 3.2.2.1 with in addition $X_t = Y_t = 0$ for $t \geq \tau_\partial$, and $X'_t = Y'_t = 0$ for $t \geq \tau'_\partial$.

On the event $\{\sup_m T_y^m = \tau_\partial \wedge \tau'_\partial\}$, we deduce by continuity of Y' that $T_y^m = T_y^{\prime m}$ so that $\tau_\partial = \tau'_\partial$. On the event $\{\sup_n T_X^n = \tau_\partial \leq \tau'_\partial < \infty\}$, for any n and $t_\vee > 0$ there exists $m \geq 1$ and $n' \geq n$ such that $T_X^n \wedge t_\vee < T_Y^m \wedge T_Y^{\prime m}$ and $\|X(T_X^n \wedge t_\vee)\| \vee \|X'(T_X^n \wedge t_\vee)\| < n' < \infty$. By step 2, (X, Y) and (X', Y') must coincide until $T = (t_\vee + 1) \wedge T_Y^m \wedge T_Y^{\prime m} \wedge T_X^{n'} \wedge T_X^{\prime n'}$, where the previous definitions ensure $T_X^n \wedge t_\vee < T$ (with the fact that X and X' are right-continuous). This proves that $T_X^n \wedge t_\vee = T_X^{\prime n} \wedge t_\vee$, and with $t_\vee, n \to \infty$ that $\tau'_\partial = \tau_\partial$.

By symmetry between the two solutions, we have a.s. $\tau_\partial = \tau'_\partial, \quad \forall\, t < \tau_\partial,\ X_t = X'_t$ and $\forall\, t \geq \tau_\partial, \quad X_t = X'_t = 0$. It concludes the proof of the uniqueness.

Existence : We see that the identification obtained for the uniqueness clearly defines the solution (X, Y) until some $T = T(t_\vee, n)$ such that either $T = t_\vee$ or $Y_T = 0$ or $\|X_T\| \geq n$. By the uniqueness property and the a priori estimates, this solution coincide with the ones for larger values of $t_\vee$ and n. Thus, it indeed produces a solution up to time τ_∂. $\square$

3.4 Main properties leading to the proof of Theorem 3.2.1

The proof of Theorem 3.2.1 relies on the criteria presented in Chapter 2, which are the same as the one in Chapter 1 except for the last one (Absorption with failures). Let :

$$\mathcal{D}_\ell := B(0, \ell) \times [1/\ell, \ell]. \tag{3.4.1}$$

These assumptions are proved to hold true under the assumptions of Theorem 3.2.1 in the following Theorems 3.4.1-6. We see in Subsection 3.4.1.1 how these Theorems together with Theorems 2.2.1-4 in Chapter 2 imply Theorem 3.2.1. In the next subsections, we then prove Theorems 3.4.1-6. By mentioning first the mixing estimate, we wish to highlight the constraint on the reachable domain under hypothesis $[H9]$. The order of the proof is different and done for clarity, noting that the mixing estimates are directly exploited in the proofs of the absorption estimates, while the escape estimates are very close to the ones of previously considered models.

Remark : With the notations of Chapter 2, the multiplicative constant $C/\langle\mu \,\big|\, h\rangle$ in Theorem 3.2.1 will be uniformly upper-bounded over the sets :

$$\mathcal{M}_{\ell,\xi} := \left\{\mu \in \mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}_+^*) \ ; \ \mu\,(\mathcal{D}_\ell) \geq \xi\right\}, \qquad \text{with } \ell \geq 2, \xi \in (0, 1].$$

Forbidding deleterious mutations in the case of dimension one $\mathcal{X}$ will make our proof a bit more complicated. This case is thus treated later on. The expression "with deleterious mutation" will be used a bit abusively to discuss the model under $[H8]$. On the other hand, the expression "with only advantageous mutation" will refer to the case where $[H9]$ holds.

3.4.1 The whole space is accessible : with deleterious mutations or $d \geq 2$

3.4.1.1 Mixing property and accessibility

With deleterious mutations, the whole space becomes accessible. It is in fact also the case with only advantageous mutations, provided $d \geq 2$:

Theorem 3.4.1. *Assume* $[H1\text{-}6]$. *For* $d \geq 2$, *suppose either* $[H8]$ *or* $[H9]$; *for* $d = 1$, *suppose* $[H8]$. *Then, for any* $\ell_I, \ell_M \geq 1$, *there exists* $L > \ell_I \vee \ell_M$ *and* $c, t > 0$ *such that, recalling* $T_{\mathcal{D}_L} := \inf\{t \geq 0 \ ; \ (X, Y)_t \notin \mathcal{D}_L\}$ *:*

$$\forall\, (x_I, y_I) \in \mathcal{D}_{\ell_I}, \quad \mathbb{P}_{(x_I, y_I)}\left[(X, Y)_t \in (dx, dy) \ ; \ t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\, \mathbf{1}_{\mathcal{D}_{\ell_M}}(x, y)\, dx\, dy \tag{3.4.2}$$

Remark : Theorem 3.4.1 implies in particular that the density w.r.t. Lebesgue's measure of any QSD is uniformly lower-bounded on any $\mathcal{D}_\ell$.

3.4.1.2 Escape from the Transitory domain

Theorem 3.4.2. *Assume [H1-4], [H7]. Then, for any $\rho > 0$, there exists $\ell_E \geq 1$ such that, with $E := \mathcal{D}_{\ell_E}$ and recalling $\tau_E := \inf\{t \geq 0\ ;\ (X,Y)_t \in E\}$:*

$$\sup_{(x,y)\in\mathbb{R}^d\times\mathbb{R}_+^*} \mathbb{E}_{(x,\,y)}\left(\exp\left[\rho\left(\tau_E \wedge \tau_\partial\right)\right]\right) < \infty.$$

3.4.1.3 Absorption with failures

We need some reference set on which our reference measure has positive density. Let :

$$\varDelta := B(-S\,\mathbf{e_1}\ ,\ \delta S) \times [1/2, 2]. \tag{3.4.3}$$

This choice (rather arbitrary), is made in such a way that the uniform distribution on $\varDelta$ can be taken as the lower-bound in the conclusions of Theorems 3.4.4 and 3.4.1.

Including deleterious mutations or with $d \geq 2$, we will consider any $E \in \mathbf{D}$. The corresponding value for ℓ_E shall be determined according to $(A2)$.

Theorem 3.4.3. *Assume [H1-6] . For $d = 1$, assume [H8]. For $d \geq 2$, assume [H10] and either [H8] or [H9]. Then, given $\rho > 0$ and $E \in \mathbf{D}$, for any $\epsilon \in (0, 1)$, we can find $t, c > 0$ such that for any $(x, y) \in E$ and $(x_\zeta, y_\zeta) \in \varDelta$, there exists a stopping time U_A such that :*

$$\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\} \qquad \text{and} \qquad \mathbb{P}_{(x,y)}(U_A = \infty,\, t < \tau_\partial) \leq \epsilon\, \exp(-\rho\, t),$$

while for some stopping time V :

$$\mathbb{P}_{(x,y)}\Big[(X(U_A), Y(U_A)) \in (dx', dy')\ ;\ U_A < \tau_\partial\Big] \leq c\,\mathbb{P}_{(x_\zeta,y_\zeta)}\Big[(X(V), Y(V)) \in (dx', dy')\ ;\ V < \tau_\partial\Big].$$

Moreover, there exists a stopping time U_A^∞ satisfying the following properties :

- *$U_A^\infty := U_A$ on the event $\{\tau_\partial \wedge U_A < \tau_E^t\}$, where $\tau_E^t := \inf\{s \geq t : (X_s, Y_s) \in E\}$.*
- *On the event $\{\tau_E^t < \tau_\partial\} \cap \{U_A = \infty\}$, and conditionally on $\mathcal{F}_{\tau_E^t}$, the law of $U_A^\infty - \tau_E^t$ coincides with the one of $\widetilde{U}_A^\infty$ for the solution $(\widetilde{X}, \widetilde{Y})$ of :*

$$\begin{cases} \widetilde{X}_r = X(\tau_E^t) - v\, r\, \mathbf{e_1} + \displaystyle\int_{[0,r]\times\mathbb{R}^d\times(\mathbb{R}_+)^2} w\, \varphi\left(\widetilde{X}_{s^-},\, \widetilde{X}_s,\, w,\, u_f,\, u_g\right)\ \widetilde{M}(ds, dw, du_f, du_g) \\ \widetilde{Y}_r = Y(\tau_E^t) + \displaystyle\int_0^r \psi\left(\widetilde{X}_s,\, \widetilde{Y}_s\right) ds + \widetilde{B}_r, \end{cases} \tag{3.4.4}$$

where $r \geq 0$, $\widetilde{M}$ and $\widetilde{B}$ are independent copies of respectively M and B.

3.4.1.4 Proof of Theorem 3.2.1 as a consequence of Theorems 3.4.1-3

— First, it is clear that the sets $\mathcal{D}_\ell$ are closed and satisfy assumption $(A0)$ in Chapter 2 :
$\forall\, \ell \geq 1,\ \mathcal{D}_\ell \subset \mathcal{D}^\circ_{\ell+1}$ and $\bigcup_{\ell \geq 1} \mathcal{D}_\ell = \mathbb{R}^d \times \mathbb{R}^*_+$.

— From Theorem 3.4.1, assumption $(A1)$ holds true, where ζ is the uniform distribution over Δ –cf (3.4.3).

— Theorem 3.4.2 implies $(A2)$ for any ρ, and we also require that ρ is chosen such that :

$$\rho > \rho_S := \sup\left\{\gamma \geq 0 \,\middle|\, \sup_{L \geq 1} \liminf_{t>0}\ e^{\gamma t}\, \mathbb{P}_\zeta(t < \tau_\partial \wedge T_{\mathcal{D}_L}) = 0\right\} \vee 0.$$

From Lemma 1.3.0.2 in Chapter 1 and $(A1)$, we know that ρ_S is upper-bounded by some value $\widetilde{\rho}_S$. In order to satisfy $\rho > \rho_S$, we set $\rho := 2\widetilde{\rho}_S$. From Theorem 3.4.2, we deduce $E = \mathcal{D}_{\ell_E}$ such that assumption $(A2)$ holds for this value of ρ.

— Finally, Theorem 3.4.3 states that assumption $(A3_F)$ holds true, for E and ρ.

Referring to Theorems 2.2.1-4 in Chapter 2, this concludes the proof of the assumption (AF), thus of Theorem 3.2.1.

3.4.2 No deleterious mutations in the uni-dimensional case

3.4.2.1 Mixing property and accessibility

When only advantageous mutations are allowed and $d = 1$, as soon as the size of jumps get bounded, the process can't access some portion of space (there is a limit in the X direction). We could prove that the limit is related to the quantity : $L_A := \sup\{M\ ;\ \nu[2\,M, +\infty) > 0\} \in (S/2,\, \infty]$.
The accessible domains with maximal extension would then be rather of the form : $[-\ell\,,\ L_A - 1/\ell] \times [1/\ell, \ell]$, with $\ell \geq 1$.

But for simplicity, we will rather consider domains of the form :

$$E := [-L, 0] \times [1/\ell, \ell], \quad \text{with } L, \ell \geq 1. \tag{3.4.5}$$

Sets of this form will be gathered under the notation Δ_E. In fact, the limit L_A will not appear in the next statements. We just wanted to point out this potential constraint on the visited domain.

Theorem 3.4.4. *Assume $d = 1$, $[H1\text{-}6]$ and $[H9]$. Then, for any $\ell_I \geq 1$ and $E \in \Delta_E$, there exists $L > \ell_I$ and $c, t > 0$ such that :*

$$\forall\,(x_I,\, y_I) \in \mathcal{D}_{\ell_I}, \quad \mathbb{P}_{(x_I, y_I)}\left[(X_t,\, Y_t) \in (dx,\, dy)\ ;\ t < \tau_\partial \wedge T_{\mathcal{D}_L}\right] \geq c\,\mathbf{1}_E(x,\, y)\,dx\,dy \tag{3.4.6}$$

Remark : Theorem 3.4.4 implies that the density w.r.t. Lebesgue's measure of any QSD is uniformly lower-bounded on any E.

3.4.2.2 Escape from the Transitory domain

Theorem 3.4.5. *Assume $d = 1$, $[H1\text{-}4]$, $[H7]$ and $[H9]$. Then, for any $\rho > 0$, there exists $E \in \Delta_E$ such that :*

$$\sup_{(x,y)\in\mathbb{R}\times\mathbb{R}_+} \mathbb{E}_{(x,\,y)}\left(\exp\left[\rho\left(\tau_E \wedge \tau_\partial\right)\right]\right) < \infty.$$

3.4.2.3 Absorption with failures

Theorem 3.4.6. *Assume $[H1\text{-}6]$ and $[H9]$. Then, given any $\rho > 0$, $E \in \Delta_E$ and $\epsilon \in (0,\,1)$, we deduce the same conclusions as in Theorem 3.4.3.*

3.4.2.4 Proof of Theorem 3.2.1 as a consequence of Theorems 3.4.4-6

- We deduce $(A1)$ from Theorem 3.4.4 (with the same ζ).
- As with Theorem 3.4.2, we deduce from Theorem 3.4.5 that we can define $E \in \Delta_E$ such that $(A2)$ holds with some value $\rho > \rho_S$.
- Finally, from Theorem 3.4.3, we deduce $(A3_F)$ for these choices of ρ and E.

Again, applying Theorems 2.2.1-4 in Chapter 2 concludes the proof of Theorem 3.2.1 in this case.

3.4.3 Structure of the proof

To allow for fruitful comparison, the proofs are gathered according to the assumptions of Chapter 2 they ensure. We first prove Theorems 3.4.2 and 3.4.5 in Section 3.5 since they are the simplest and the closest to the proofs in Chapter 2 and the remaining theorems are more closely related. We then prove Theorems 3.4.1 and 3.4.4 in Section 3.6, and finally Theorems 3.4.3 and 3.4.6 in Section 3.7.

3.5 Escape from the transitory domain

3.5.1 With deleterious mutations or $d \geq 2$

The proof of Theorem 3.4.2 for any given ρ can be easily adapted from Subsection 1.4.2.4 of Chapter 1. Indeed, the proofs rely on uniform couplings which ensure –for any of the considered cases– that the population size decreases sufficiently quickly, and makes large increase very exceptionally. Specifically, this proof does not depend at all on the dynamics of X. The proof developed in the next subsection is an extension of the latter and illustrates the technique.

3.5.2 Without deleterious mutations, $d = 1$

In this section, we prove Theorem 3.4.5, i.e. :

Assume $d = 1$, [H1-4] , [$H7$] and [$H9$]. Then :

$$\forall\, \rho > 0, \quad \exists\, E \in \Delta_E, \qquad \sup_{(x,y) \in \mathbb{R} \times \mathbb{R}_+} \mathbb{E}_{(x,\,y)} \left(\exp\left[\rho\left(\tau_E \wedge \tau_\partial\right)\right]\right) < \infty$$

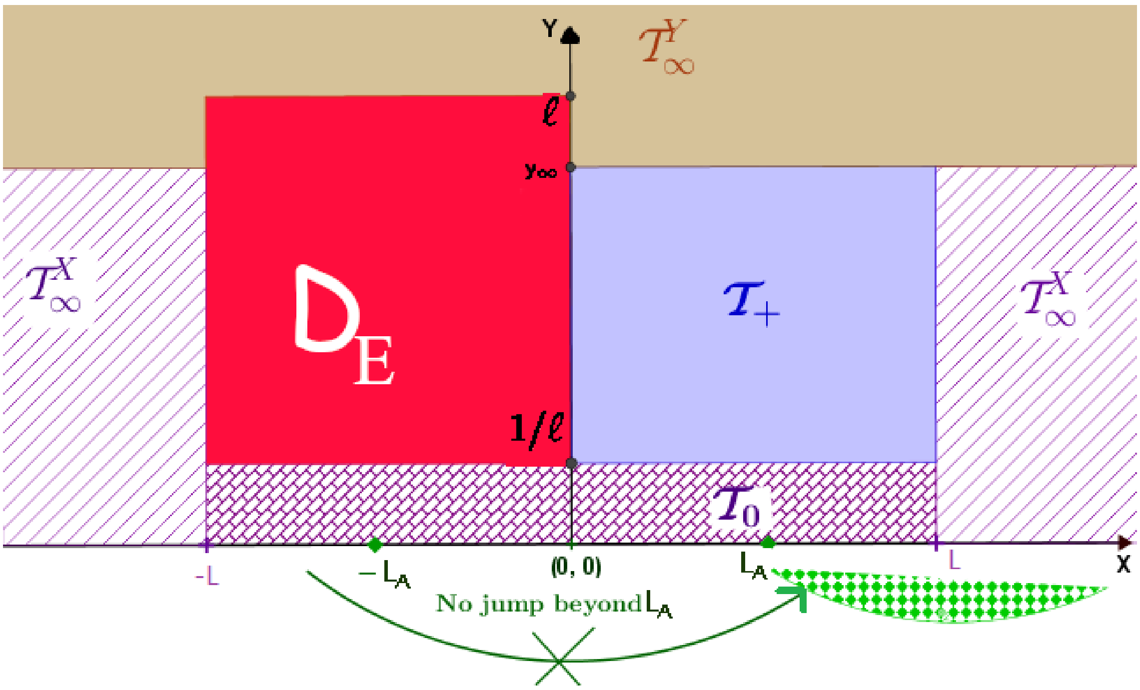


FIGURE 3.1 – subdomains for A2
As long as $X_t > 0$, $\|X_t\|$ must decrease (see Lemma (A) in Appendix A.2). Once the process has escaped $\{x \geq L_A\}$, there is no way (by allowed jumps and v) that it reaches it afterwards.

3.5.2.1 Decomposition of the transitory domain

The proof is very similar to the one of Subsection 1.4.2.4 of Chapter 2 except that, due to Theorem 3.4.6, the domain E cannot be chosen as large. We thus need to consider another subdomain of $\mathcal{T}$, that will be treated specifically thanks to $[H9]$.

The complementary $\mathcal{T}$ of E is then made up of 4 subdomains : "$y = \infty$", "$y = 0$", "$x > 0$", and "$\|x\| = \infty$", according to the figure 3.1. Thus, we define :

— $\mathcal{T}_\infty^Y := \{(-\infty,\, -L) \cup (0, \infty)\} \times (y_\infty, \infty) \bigcup [-\ell, 0] \times [\ell, \infty)$ ("$y = \infty$")

— $\mathcal{T}_0 := (-L,\, L) \times [0, 1/\ell]$ ("$y = 0$")

— $\mathcal{T}_+ := (0, L) \times (1/\ell, y_\infty]$ ("$x > 0$")

— $\mathcal{T}_\infty^X := \{\mathbb{R} \setminus (-L,\, L)\} \times (1/\ell, y_\infty]$ ("$|x| = \infty$")

With some threshold $t_\vee$ (to ensure the finiteness but that shall go to ∞), let us first introduce the exponential moments of each area (remember that τ_E is the hitting time of E) :

— $V_E := \tau_E \wedge \tau_\partial \wedge t_\vee$

— $\mathcal{E}_\infty^Y := \sup_{(x,\,y)\in\mathcal{T}_\infty^Y} \mathbb{E}_{(x,y)}[\exp(\rho\, V_E)]$, — $\mathcal{E}_\infty^X := \sup_{(x,\,y)\in\mathcal{T}_\infty^X} \mathbb{E}_{(x,y)}[\exp(\rho\, V_E)]$,

— $\mathcal{E}_0 := \sup_{(x,\,y)\in\mathcal{T}_0} \mathbb{E}_{(x,y)}[\exp(\rho\, V_E)]$, — $\mathcal{E}_X := \sup_{(x,\,y)\in\mathcal{T}_+} \mathbb{E}_{(x,y)}[\exp(\rho\, V_E)]$.

Implicitly, $\mathcal{E}_\infty^Y$, $\mathcal{E}_\infty^X$, $\mathcal{E}_X$ and $\mathcal{E}_0$ are functions of ρ, L, ℓ, y_∞ that need to be specified.

3.5.2.2 A set of inequalities

As in Subsection 1.4.2.4 in Chapter 1, we first state some inequalities between these quantities, summarized in Proposition 3.5.2.1, 3.5.2.2, 3.5.2.4 and 3.5.2.3 that follow. Thanks to these inequalities, we prove in Appendix C that those quantities are bounded. This will end the proof of Theorem 3.4.5.

Proposition 3.5.2.1. *Suppose $[H1\text{-}4]$. Then, given any $\rho > 0$, we can find $y_\infty > 0$ and $C_\infty^Y \geq 1$ such that any choice $\ell > y_\infty$ and $L > 0$ ensures :*

$$\mathcal{E}_\infty^Y \leq C_\infty^Y \left(1 + \mathcal{E}_\infty^X + \mathcal{E}_X\right) \tag{3.5.1}$$

Proposition 3.5.2.2. *Suppose $[H1\text{-}4]$ and $[H7]$. Then, given any $\rho > 0$, there exists $C_\infty^X \geq 1$ such that whatever ϵ^X, $y_\infty > 0$, we can find $L > 0$ and $\ell^X > y_\infty$ such that choosing $\ell \geq \ell^X$ ensures :*

$$\mathcal{E}_\infty^X \leq C_\infty^X\, (1 + \mathcal{E}_0 + \mathcal{E}_X) + \epsilon^X\, \mathcal{E}_\infty^Y \tag{3.5.2}$$

Proposition 3.5.2.3. *Suppose $[H1\text{-}4]$ and $[H9]$. Then, given any ρ, $L > 0$, there exists $C_X \geq 1$ such that for any ϵ^+, $y_\infty > 0$, choosing ℓ sufficiently large ($\ell \geq \ell^+ > y_\infty$) ensures :*

$$\mathcal{E}_X \leq C_X\, (1 + \mathcal{E}_0) + \epsilon^+\, \mathcal{E}_\infty^Y \tag{3.5.3}$$

Proposition 3.5.2.4. *Suppose* $[H1\text{-}4]$. *Then, given any* ρ, ϵ^0, $y_\infty > 0$, *there exists* $C_0 \geq 1$ *such that any choice of* L *and of* ℓ *sufficiently large* ($\ell \geq \ell^0 > y_\infty$) *ensures :*

$$\mathcal{E}_0 \leq C_0 + \epsilon^0 \left(\mathcal{E}^Y_\infty + \mathcal{E}^X_\infty + \mathcal{E}_X \right) \tag{3.5.4}$$

Again, the proofs of Proposition 3.5.2.1, 3.5.2.2 and 3.5.2.4 are nearly identical to the ones of Propositions respectively 1.4.2.1, 1.4.2.2 and 1.4.2.3 of Chapter 1, and are thus left to the reader. In Appendix C, we will prove first how to deduce Theorem 3.4.5 and it generalizes naturally what the similar proof of $(A2)$ in Chapter 1. Then, we will prove Proposition 3.5.2.3. The core idea is that jumps are not allowed here to increase the mal-adaptation. Thus, the worst-case scenario for the exit time of $\mathcal{T}_+$ is that the process gets simply drifted by the environmental change at speed v until X gets negative.

3.6 Mixing properties and accessibility

3.6.1 Mixing for Y independently of the jumps through a Girsanov thansform

The idea of this subsection is to prove that we can think of Y as a Brownian Motion up to some stopping time which will bound U_A. If we get a lower bound for the probability of events in this simpler setup, this will prove that we also get a lower bound in the general setup.

3.6.1.1 Construction of the change of probability under $[H4]$:

The limits of our control :

$$\text{Let} \quad : \qquad t_G,\, x_\vee > 0, \quad 0 < y_\wedge < y_\vee, \quad N_J \geq 1.$$

Our aim is to simplify the law of $(Y_t)_{t\in[0,t_G]}$ as long as Y stays in $[y_\wedge, y_\vee]$, $\|X\|$ stays in $\overline{B}(0, x_\vee)$, and at most N_J jumps have occurred. Thus, let :

$$T_X := \inf \{t \geq 0,\, \|X_t\| \geq x_\vee\}\ , \qquad T_Y := \inf \{t \geq 0,\, Y_t \notin [y_\wedge, y_\vee]\}\,. \tag{3.6.1}$$

$$g_\vee := \sup \left\{ g(x, w)\ ;\ \|x\| \leq x_\vee, w \in \mathbb{R}^d \right\}, \qquad f_\vee := \sup \{ f(y)\ ;\ y \in [y_\wedge,\, y_\vee]\} \tag{3.6.2}$$

$$\mathcal{J} := \left\{ (w, u_g, u_f) \in \mathbb{R}^d \times [0, f_\vee] \times [0, g_\vee] \right\}, \quad \text{so that } \nu \otimes du_g \otimes du_f(\mathcal{J}) = \nu(\mathbb{R}^d)\, g_\vee\, f_\vee < \infty.$$

Our Girsanov's transform alter the law of Y until the stopping time :

$$T_G := t_G \wedge T_X \wedge T_Y \wedge U_{N_J}, \tag{3.6.3}$$

$$\text{where } U_{N_J} := \inf \{ t\ ;\ M([0,t] \times \mathcal{J}) \geq N_J + 1 \}\,. \tag{3.6.4}$$

Note that the $(N_J + 1)$-th jump of X will then necessarily occur after T_G.

The change of probability : We define :

$$L_t := -\int_0^{t\wedge T_G} \psi(X_s, Y_s) dB_s, \quad \text{and } D_t := \exp\left[L_t - < L >_t /2\right]. \tag{3.6.5}$$

the exponential local martingale associated with (L_t).

Theorem 3.6.1. *Suppose* $[H1\text{-}4]$ *. Then, for any* t_G*,* $x_\vee > 0$*, and* $y_\vee > y_\wedge > 0$*, there exists* $C_G > c_G > 0$ *such that a.s.* $c_G \leq D_\infty \leq C_G$*. In particular,* D_t *is a uniformly integrable martingale and* $\beta_t = B_t - < B\, ,\, L >_t$ *is a Brownian Motion under :* $\mathbb{P}^G_{(x,y)} := D_\infty \cdot \mathbb{P}_{(x,y)}$*. Moreover :*

$$\forall\, (x,\, y) \in \mathbb{R}^d \times \mathbb{R}_+, \qquad c_G\, \mathbb{P}^G_{(x,\, y)} \leq \mathbb{P}_{(x,\, y)} \leq C_G\, \mathbb{P}^G_{(x,\, y)},$$

On the event $\{t \leq T_G\}$, $Y_t = y + \beta_t$, i.e. Y has the law of a Brownian Motion under $\mathbb{P}^G_{(x,y)}$ up to time T_G. This means that we can have bounds of the probability for events involving Y as in our model by considering Y as a simple Brownian Motion. Meanwhile, the independence between its variations as a Brownian and the Poisson Process still hold due to the following proposition.

Proposition 3.6.1.1. *A Brownian Motion and a PPP that are adapted to the same filtration and such that their increase after time t is independent from* $\mathcal{F}_t$ *are necessarily independent.*

The proof of Theorem 3.6.1 is postponed in Appendix D so we turn now to the proof of Proposition 3.6.1.1.

Proof : By Theorem 2.1.8 of [DiT13], if X_1, X_2 are additive functionals and semi-martingales with respect to a common filtration, both starting from zero, and such that their quadratic covariations $[X_1, X_2]$ is a.s. zero, then the random vector $(X_1(t) - X_1(s), X_2(t) - X_2(s))$ is independent of $\mathcal{F}_s$, for every $0 \leq s \leq t$. Moreover, the vector (X_1, X_2) of additive processes is independent.

Note B the Brownian Motion and M the PPP on $\mathbb{R}_+ \times \mathcal{X}$ For any test function $F : \mathcal{X} \mapsto \mathbb{R}$, define $Z(t) := \int_{[0,t]\times\mathcal{X}} F(x)\; M(ds, dx)$. Both Z and B are additive functionals and semi-martingales with respect to the filtration $\mathcal{F}_t$, both starting from zero. Z being a jump process and B continuous, their quadratic covariation equals a.s. 0. Since it applies to any F, exploiting Theorem 2.18 of [DiT13] implies that B and M are independent. $\square$

3.6.1.2 Mixing for Y

The proof will rely on Theorem 3.6.1 and on the following property of Brownian Motion :

Lemma 3.6.1.2. *Let $(B_t)_{t\geq 0}$ be a Brownian Motion with initial condition $b_I \in [0, b_\vee]$. Then, given any $\epsilon > 0$, $b_\vee > 0$ and $0 < t_0 \leq t_1$, there exists $c_B > 0$ such that for any $b_I \in [0, b_\vee]$ and $t \in [t_0, t_1]$:*

$$\mathbb{P}_{b_I}\left(B_t \in db\,,\ \min_{s\leq t_1} B_s \geq -\epsilon,\ \max_{s\leq t_1} B_s \leq b_\vee + \epsilon\right) \geq c_B \times \mathbf{1}_{[0,\, b_\vee]}(b)\, db.$$

Thanks to this lemma and Theorem 3.6.1, we will be able to control Y to prove that it indeed diffuses and that it stays in some closed interval I_Y away from 0. We can then control the behavior of X independently of the trajectory of Y by appropriate conditioning of M –the PPP– so as to ensure the jumps we need (conditionally that it remains in I_Y).

Proof : Consider the collection of marginal laws of B_t, with initial condition $b \in (-\epsilon, b_\vee + \epsilon)$, killed when it reaches $-\epsilon$ or $b_\vee + \epsilon$. It is classical that these laws have a density $u(t; b, b')$, $t > 0$, $b' \in [-\epsilon, b_\vee + \epsilon]$, w.r.t. the Lebesgue measure (cf e.g. Section 2.4 in Bass [Ba95] for more details). u is a solution to the Cauchy problem with Dirichlet boundary conditions :

$$\begin{aligned}
&\partial_t u(t; b_I, b) = \Delta_b u(t; b_I, b), && \text{for } t > 0, b_I, b \in (-\epsilon, b_\vee + \epsilon),\\
&u(t; b_I, -\epsilon) = u(t; b_I, b_\vee + \epsilon) = 0, && \text{for } t > 0.
\end{aligned}$$

By the maximum principle (cf e.g. Theorem 4, Subs 2.3.3. in Evans [Ev98]), $u > 0$ on $\mathbb{R}_+^* \times [0, b_\vee] \times (-\epsilon, b_\vee + \epsilon)$ and since u is continuous in its three variables, it is lower-bounded by some c_B on the compact subset $[t_0, t_1] \times [0, b_\vee] \times [0, b_\vee]$. □

3.6.2 Mixing for X with deleterious mutations

For clarity, we decompose the "migration" along X into different kinds of elementary steps, as already done in Chapter 2. Let :

$$\mathcal{A} := B(-S\,\mathbf{e_1}\,,\ \delta S/2)\,, \qquad \tau_{\mathcal{A}} := \inf\{t \geq 0\ ;\ X_t \in \mathcal{A}\,,\ Y_t \in [2,3]\}\,, \tag{3.6.6}$$

where we assume w.l.o.g. that $\delta S \leq S/8$ ($[2,3]$ is chosen arbitrarily). We also recall that for any $\ell \geq 1$, $T_{\mathcal{D}_\ell} := \inf\{t \geq 0\ ;\ (X, Y)_t \notin \mathcal{D}_\ell\}$.

The first step is to prove that, with a lower-bounded probability for any initial condition in $\mathcal{D}_\ell$, $\tau_{\mathcal{A}}$ is upper-bounded by some constant $t_{\mathcal{A}}$. The second step is to prove that the process is sufficiently diffuse and that time-shifts are not a problem. In the third step, we specify which sets we can reach from $\mathcal{A}$.

More precisely, each step corresponds to each of the following :

Lemma 3.6.2.1. *Assume that [H1-6] and [H8] hold. Then, for any $\ell_I > 0$, there* *exists* $c_{\mathcal{A}}, t_{\mathcal{A}} > 0$ *and* $L \geq \ell_I$ *such that :*

$$\forall\,(x_I, y_I) \in \mathcal{D}_{\ell_I}, \quad \mathbb{P}_{(x_I,y_I)}(\tau_{\mathcal{A}} \leq t_{\mathcal{A}} \wedge T_{\mathcal{D}_L}) \geq c_{\mathcal{A}}, \tag{3.6.7}$$

Lemma 3.6.2.2. *Assume that* $[H1\text{-}6]$ *and* $[H8]$ *hold. Then, there exists* $\ell_{\mathcal{A}} \geq 1$ *and* $t_R > 0$ *such that we can associate to any* $t_\vee > t_R$ *some* $c_R > 0$ *such that for any* $t \in [t_R, t_\vee]$ *and* $(x_0, y_0) \in \mathcal{A} \times [2, 3]$ *:*

$$\mathbb{P}_{(x_0,y_0)}\left[(X,Y)_t \in (dx, dy)\ ;\ t < T_{\ell_{\mathcal{A}}}\right] \geq c_R\, \mathbf{1}_{\mathcal{A}}(x)\, \mathbf{1}_{[2,3]}(y)\, dx\, dy.$$

For the third step, for $L \geq 3$ and $t, c > 0$, let :

$$\begin{aligned}\mathcal{R}^{(L)}(t,c) := \Big\{ x_F \in \mathbb{R}^d \,\Big|\, &\forall\,(x_0, y_0) \in \mathcal{A} \times [1/L,\ L], \\ &\mathbb{P}_{(x_0,y_0)}\left[(X,Y)_t \in (dx, dy)\ ;\ t < T_{2\,L}\right] \geq c\, \mathbf{1}_{B(x_F, \delta S/2)}(x)\, \mathbf{1}_{[1/L,\, L]}(y)\, dx\, dy \Big\}.\end{aligned} \tag{3.6.8}$$

Lemma 3.6.2.3. *Assume that* $[H1\text{-}6]$ *and* $[H8]$ *hold. Then, for any* $m \geq 3$*, we can find* $L \geq m$ *such that :* $\quad B(0, m) \subset \cup_{\{t,c>0\}} \mathcal{R}^{(L)}(t, c)$.

In the following Subsection, we prove Theorem 3.4.1 as a consequence of these. In the next one, these three lemmas are related to three others that are much more elementary. The proof of those and how they imply the first three ones is deferred to Appendix E.

Remark : • In view of Lemma 3.6.1.2, choosing a different value for L will only change the lower-bound involved.
• The first two steps would be sufficient to prove Theorem 3.4.1 and 3.4.4 with the uniform distribution over $\mathcal{A} \times [2, 3]$ instead of respectively $\mathcal{D}_{\ell_m}$ and E.

3.6.2.1 Theorem 3.4.1 as a consequence of Lemmas 3.6.2.1-3

The proof is quite naturally adapted from the one of Lemma 2.3.2.1 in Chapter 2. Let $\ell_I, \ell_M \geq 0$. Let also $\ell_{\mathcal{A}}$ and t_R be defined according to Lemma 3.6.2.2. According to Lemma 3.6.2.1, we can find $c_{\mathcal{A}}, t_{\mathcal{A}}$ and $L \geq 2\,(\ell_{\mathcal{A}} \vee \ell_I \vee \ell_M)$ such that (3.6.7) holds true.

Since $\mathcal{D}_{\ell_M} \subset \mathcal{D}_L$ is compact, we can find a finite sequence $(x_k)_{k \leq K} \in (\mathcal{D}_{\ell_M})^K$ such that :

$$\mathcal{D}_{\ell_M} \subset \textstyle\bigcup_{k \leq K} B(x_k\ ,\ \delta S/2).$$

From Lemma 3.6.2.3, we thus deduce that there exists c_F and $t_F > 0$ such that :

$$\begin{aligned}&\forall\, 1 \leq k \leq K,\ \exists\, t_k \leq t_F, \quad \forall\,(x_0, y_0) \in \mathcal{A} \times [y_r,\ 2\,y_r], \\ &\mathbb{P}_{(x_0,y_0)}\left[(X,Y)_{t_k} \in (dx, dy)\ ;\ t_k < T_{\mathcal{D}_L}\right] \geq c_F\, \mathbf{1}_{B(x_k, \delta S/2)}(x)\, \mathbf{1}_{[1/L,\, L]}(y)\, dx\, dy.\end{aligned} \tag{3.6.9}$$

Let $t_M := t_{\mathcal{A}} + t_R + t_F$ and $c_M := c_{\mathcal{A}} \times c_R \times Leb(\mathcal{A}) \times c_F$ where c_R is the value deduced from Lemma 3.6.2.2 with $t_\vee := t_M$. In particular, note that :

$$\mathbb{E}_{(x_I,y_I)}\Big(\mathbb{P}_{(X,Y)[\tau_{\mathcal{A}}]}\Big[(\widetilde{X},\widetilde{Y})[t_M - t_k - \tau_{\mathcal{A}}] \in \mathcal{A} \times [2,\,3]\ ;\ t_M - t_k - \tau_{\mathcal{A}} < \widetilde{T}_{\ell_{\mathcal{A}}}\Big] \Big| \mathcal{F}_{\tau_{\mathcal{A}}}\ ;\ \tau_{\mathcal{A}} \le t_{\mathcal{A}} \wedge T_{\mathcal{D}_L}\Big) \ge c_R \times Leb(\mathcal{A}), \tag{3.6.10}$$

Note also that the time needed for this second step is random –cf (3.6.10). It depends on the time needed to reach $\mathcal{A}$ from (x_I, y_I) and on the time needed from $\mathcal{A}$ to the vicinity of x_k, but can always be chosen in $[t_R,\, t_M]$.

Thanks to the Markov property, with (3.6.7), (3.6.9) and (3.6.10), we conclude that for any $(x_I, y_I) \in \mathcal{D}_{\ell_I}$:

$$\mathbb{P}_{(x_I,y_I)}\left[(X,Y)_{t_M} \in (dx,dy)\ ;\ t_M < T_{\mathcal{D}_L}\right] \ge c_M\, \mathbf{1}_{B(x_k,\delta S/2)}(x)\, \mathbf{1}_{[1/L,\,L]}(y)\, dx\, dy.$$

Since t_M does not depend on the initial condition nor on k, this immediately implies :

$$\forall\, (x_I,y_I) \in \mathcal{D}_{\ell_I}, \quad \delta_{(x_I,y_I)} P_{t_M}(dx,dy) \ge c_M\, \mathbf{1}_{\mathcal{D}_{\ell_M}}(x)\, \mathbf{1}_{[1/L,\,L]}(y)\, dx\, dy.$$

This proves immediately Theorem 3.4.1 as a consequence of Lemmas 3.6.2.1-3. □

3.6.2.2 Elementary steps for the proof of Lemmas 3.6.2.1-3

Lemmas 3.6.2.2-3 are deduced by induction from the following elementary properties.

Lemma 3.6.2.4. *Assume that [H1-4] hold. Then, for any $t, c > 0$, $L \ge 3$ and $x \in B(0, 2\,L)$:*

$$x \in \mathcal{R}^{(L)}(t,c) \Rightarrow \exists\, c' > 0, \quad \forall\, 0 \le u < u_\vee, \quad x - v\,u\,\mathbf{e_1} \in \mathcal{R}^{(L)}(t+u, c'),$$
$$\text{where } u_\vee := \sup\{u \ge 0 \,\big|\, (x - v\,u\,\mathbf{e_1}) \in B(0, 2\,L)\}$$

Lemma 3.6.2.5. *Assume that [H1-6] and [H8] hold. Then, there exists $c_0, t_0 > 0$ such that for any $L \ge 3$: $\mathcal{A} \subset \mathcal{R}^{(L)}(t_0, c_0)$.*

Lemma 3.6.2.6. *Assume that [H1-6] and [H8] hold. Then, there exists $t_P, c_P > 0$ such that for any $t, c > 0$, $L \ge 3$:*

$$\{x\ :\ d(x\ ,\ B(0, 2\,L - 2\,S) \cap \mathcal{R}^{(L)}(t,c)\ ) \le \delta S/4\} \subset \mathcal{R}^{(L)}(t + t_P, c \times c_P).$$

Remarks 3.6.2.1. *A direct application of the Markov property also implies that for any $t, t', c, c' > 0$:*

$$\mathcal{A} \cap \mathcal{R}^{(L)}(t,c) \ne \emptyset \Rightarrow \mathcal{R}^{(L)}(t',c') \subset \mathcal{R}^{(L)}(t+t', c \times c').$$

To prove Lemma 3.6.2.4, one just needs to get a uniform upper-bound on the jump rate and combine Lemma 3.6.1.2 with Theorem 3.6.1 (and Lemma 3.6.1.1) to deal with Y. We leave these proofs to the reader, and will rather detail in Appendix E the ones of Lemma 3.6.2.5 and 3.6.2.6, and how they imply Lemmas 3.6.2.3, 3.6.2.2 and 3.6.2.1.

3.6.3 Mixing for X with only advantageous mutations for a unidimensional phenotype ($d = 1$)

In this case, in place of Lemma 3.6.2.3, we prove :

Lemma 3.6.3.1. *Assume that $d = 2$, [H1-6] and [H9] hold. Then, for any $L \geq 3$:* $[-L, 0] \subset \cup_{\{t,c>0\}} \mathcal{R}^{(L)}(t, c)$.

The only difference here is that Lemma 3.6.2.6 does not hold anymore for any x, but still under the restriction that x is not too far from 0 in $\mathbb{R}_+$:

Lemma 3.6.3.2. *For any $L \geq 3$, there exists $t_P, c_P > 0$ such that for any $t, c > 0$:*

$$\{x \ : d(x, [-L, S/4] \cap \mathcal{R}^{(L)}(t, c)\,) \leq \delta S/4\} \subset \mathcal{R}^{(L)}(t + t_P, c \times c_P).$$

The proof of Lemma 3.6.2.2 now follows from this new Lemma, while the one of Lemma 3.6.3.1 is easily adapted. In the same way, to prove Lemma 3.6.2.1, either x_I is bigger than $x_0 + \delta S/2$ and we just let X comes down from x_I to $(x_0 - \delta S/2, x_0 + \delta S/2)$ as in Lemma 3.6.2.4 ; or with the technique of Lemma 3.6.2.6, we use jumps to go up from x_I to $(x_0 - \delta S/2, x_0 + \delta S/2)$. Again, the proof of 3.6.3.2 is deferred to Subsection E.5 of the Appendix.

3.6.4 Mixing for X with only advantageous mutations for a multi-dimensional phenotype ($d \geq 2$)

To prove Lemma 3.6.2.1 is not much more difficult when we restrict ourselves to advantageous jumps. We only need to get close to 0 and then let v bring the process back in $B(x_0, \delta S/2)$. In the same way, the proof of Lemma 3.6.2.2 is quite exactly the same with the adapted version of Lemma 3.6.2.6.

When $d \geq 2$, we are still able to prove Lemma 3.6.2.3, but the argument is more tricky. Recall that we shall prescribe jumps close to $S\,\mathbf{e_1}$ that necessarily reduce the norm if we want these jump to happen with a lower-bounded probability. Yet, the larger is the component orthogonal to $\mathbf{e_1}$, the less margin we have to do so. Nonetheless, it is not very difficult to reach any x such that $x^1 \leq 0$. But then, starting around $n\,\mathbf{e_2}$ with n large, we can make X_1 increase by jumps while reducing the second component sufficiently fast, so as to reach any value on the first coordinate.

Of course, it is the same if we replace $\mathbf{e_2}$ by any other direction orthogonal to $\mathbf{e_1}$.

To simplify notations, given any direction $\mathbf{u}$ on the sphere S^d of radius 1, we denote its orthogonal component by :

$$x^{(\perp \mathbf{u})} := x - \langle x, \mathbf{u} \rangle \mathbf{u}\ , \qquad \text{and specifically for } \mathbf{e_1} : \quad x^{(\perp 1)} := x - \langle x, \mathbf{e_1} \rangle \mathbf{e_1}. \tag{3.6.11}$$

Instead of Lemma 3.6.2.6, what we can get is the following lemma :

Lemma 3.6.4.1. *Assume that $d \geq 2$, [H1-6] and [H9] hold. Then, for any $m \geq 3 \vee (2\,S)$, there exists $\epsilon \leq \delta S/2$ such that for any $x \in B(0, m)$ with $\langle x, \mathbf{e_1} \rangle \leq 0$, there exists $t_P, c_P > 0$ such that :*

$$\forall\, t, c > 0, \quad x \in \mathcal{R}^{(L)}(t, c) \Rightarrow \overline{B}(x, \epsilon) \subset \mathcal{R}^{(L)}(t + t_P, c \times c_P).$$

This Lemma is in fact a consequence of Lemma 3.6.2.4 and (with $u = \mathbf{e_1}$) :

Lemma 3.6.4.2. *Assume that $d \geq 2$, [H1-6] and [H9] hold. Then, for any $x_\vee > 0$, there exists $\epsilon \leq \delta S/2$ such that for any $L \geq 3 \vee (2\,S)$, $x \in B(0, L)$ and $\mathbf{u} \in S^d$ such that both $\langle x, \mathbf{u} \rangle \geq S$ and $\|x^{(\perp \mathbf{u})}\| \leq x_\vee$, there exists $t_P, c_P > 0$ such that for any $t, c > 0$:*

$$x \in \mathcal{R}^{(L)}(t, c) \Rightarrow \overline{B}(x - S\,\mathbf{u}, \epsilon) \subset \mathcal{R}^{(L)}(t + t_P, c \times c_P).$$

Since they are so close, we leave to the reader the proof of Lemma 3.6.4.1 and defer the proof of Lemma 3.6.4.2 to the Appendix E.6.

3.6.4.1 Lemma 3.6.2.3 as a consequence of Lemmas 3.6.4.1 and 3.6.4.2

$$\text{First, we define :} \qquad \mathbf{u} := x^{(\perp 1)} / \|x^{(\perp 1)}\|. \tag{3.6.12}$$

Step 1 is to prove that for any $L \geq 3 \vee (2S)$ and $x \in B(0, L)$ satisfying $\langle x, \mathbf{e_1} \rangle := -S$, there exists $t, c > 0$ such that $x \in \mathcal{R}^{(L)}(t, c)$.

We consider the value of ϵ given by Lemma 3.6.4.1 and define :

$$K := \left\lfloor \|x^{(\perp 1)}\| / \epsilon \right\rfloor, \quad \text{for } 0 \leq k \leq K, \quad x_k := -S\,\mathbf{e_1} + k\,\epsilon\,\mathbf{u}, \quad x_{K+1} := x.$$

By Lemma 3.6.2.5, $x_0 \in \mathcal{R}^{(L)}(t_0, c_0)$. Thus, by induction with Lemma 3.6.4.1, since $x_k \in B(0, L)$ and $\|x_{k+1} - x_k\| \leq \epsilon$, $x_k \in \mathcal{R}^{(L)}(t_0 + k\,t_P,\ c_0\,[c_P]^k)$. In particular, there exists $t, c > 0$ such that $x \in \mathcal{R}^{(L)}(t, c)$.

Note that in order to reach any $x \in B(0, m)$ such that $\langle x, \mathbf{e_1} \rangle \leq -S$, it suffices to reach first $x' := -S\,\mathbf{e_1} + x^{(\perp 1)}$ –also in $B(0, m)$– thanks to Step 1, and then apply

Lemma 3.6.2.4.

We consider now the general case. Assume solely that $x \in \mathcal{B}(0, m)$. We consider the value of ϵ given by Lemma 3.6.4.2 for $x_\vee := |\langle x, \mathbf{e_1}\rangle| \vee S$. Let :

$$K := \left\lfloor \frac{\langle x, \mathbf{e_1}\rangle + S}{\epsilon} \right\rfloor, \qquad K \leq \frac{m+S}{\epsilon} \tag{3.6.13}$$

$$\text{for } 0 \leq k \leq K, \quad x_k := (-S + k\,\epsilon)\,\mathbf{e_1} + \Big[(K+1-k)\,S + \|x^{(\perp 1)}\|\Big]\;u\,, \qquad x_{K+1} := x.$$

Note that we can indeed find an upper-bound $L \geq 3\vee(2S)$ on the sequence $(\|x_k\|)_{k\leq K}$ uniform over any such $x_F \in \mathcal{B}(0, m)$ (with the bound on L and $\|x^{(\perp 1)}\| \leq m$).

Since $\langle x_0, \mathbf{e_1}\rangle = -S$, we can use Step 1 to prove that there exists $t_0, c_0 > 0$ such that $x_0 \in \mathcal{R}^{(L)}(t_0, c_0)$. By Lemma 3.6.4.2 and induction on k, we deduce : $x_k \in \mathcal{R}^{(L)}(t_0 + k\,t_P,\, c_0\,[c_P]^k)$. In particular, there exists $t, c > 0$ such that $x \in \mathcal{R}^{(L)}(t, c)$. □

3.7 Absorption with failures

3.7.1 Proof of Theorem 3.4.3 in the case $d = 1$

3.7.1.1 Propositions 3.7.1.1-4

We consider a first process (X, Y) with some initial condition $(x_E,\, y_E) \in E$. We will prove that considering $U_A = t$ is sufficient, except for exceptional behavior of the process. Given ϵ, t shall be chosen sufficiently small to ensure that, with probability close to 1 (according to ϵ) no jump has occurred before time t, and that the population size has not changed too much.

We define :

$$\delta y := \Big(3\,\ell_E(\ell_E+1)\Big)^{-1}, \quad y_\wedge := 1/(\ell_E+1) = 1/\ell_E - 3\,\delta y, \quad y_\vee := \ell_E + 1 > \ell_E + 3\,\delta y,$$
$$T_{\delta y} := \inf\{t\ ;\ |Y_t - y_E| \geq 2\,\delta y\} < \tau_\partial \tag{3.7.1}$$

We recall that we can upper-bound the first jump time of X by :

$$T_J := \inf\{t\ ;\ M([0,t]\times\mathcal{J}) \geq 1\}\,, \tag{3.7.2}$$

where $\mathcal{J}$ is defined as in Subsection 3.6.1.1.

- On the event $\{t < T_{\delta y} \wedge T_J\}$, we set $U_A := t$.
- On the event $\{T_{\delta y} \wedge T_J \wedge \tau_\partial \leq t\}$, we set $U_A := \infty$.

On the event $\{t < T_{\delta y} \wedge T_J\}$, the law of $(X,\, Y)_{U_A}$ is quite easy to upper-bound. Indeed, $X(t) = x_E - v\,t$; while $Y(t)$ has a bounded density with support in $[y_E - \delta y, y_E + \delta y]$.

On the other hand, we need to ensure that the event $\{T_{\delta y} \wedge T_J < t\}$ is sufficiently exceptional.

Proposition 3.7.1.1. *Assume [H1-4] . Then, for any $\ell_E, t > 0$, there exists $D^X > 0$ such that for any $x_E \in [-\ell_E, \ell_E]$ and $y_E \in [1/\ell_E, \ell_E]$:*

$$\mathbb{P}_{(x_E, y_E)} [(X, Y)(U_A) \in (dx, dy), U_A < \tau_\partial] \leq D^X \mathbf{1}_{[y_E - 2\delta y, y_E + 2\delta y]}(y)\, \delta_{x_E - v\, t}(dx)\, dy.$$

Proposition 3.7.1.2. *Assume [H1-4] . Then, given any $\ell_E, \rho > 0$, $\epsilon \in (0, 1)$, and $t > 0$ sufficiently small ($t \leq t^\vee \leq 1/v$), we have for any $(x_E, y_E) \in E$:*

$$\mathbb{P}_{(x_E, y_E)} (T_{\delta y} < t) \vee \mathbb{P}_{(x_E, y_E)} (T_J < t) \leq \epsilon\, e^{-\rho}/2 \leq \epsilon\, \exp(-\rho\, t)/2.$$

Note that our choice ensures :

$$\forall\, t < t \wedge T_{\delta y} \wedge T_J, \quad (X_t, Y_t) \in [-\ell_E - 1,\, \ell_E] \times [y_\wedge,\, y_\vee].$$

In the following, we look for the definition of V in the upper-bound of Theorem 3.4.3 In view of exploiting Theorem 3.4.1 and the Markov property, we first assume that the initial condition (x', y') satisfies :

$$x' \in [x_E,\, x_E + v\, t]\,, \qquad |y' - y_E| \leq \delta y. \tag{3.7.3}$$

Then, we define : $$V' = V'(x', x_E) := t + (x' - x_E)/v \quad \in [t, 2t], \tag{3.7.4}$$

and adapt the proof of Proposition 3.7.1.1 to ensure the reversed inequality :

Proposition 3.7.1.3. *Assume [H1-4] . Then, given any $\ell_E > 0$ and $t > 0$, there exists $d^X > 0$ such that :*

$$\begin{aligned} &\forall\, (x_E,\, y_E) \in E, \quad \forall\, x' \in [x_E,\, x_E + v\, t], \quad \forall\, y' \in [y_E - \delta y/2\,,\, y_E + \delta y/2], \\ &\qquad \mathbb{P}_{(x', y')} [(X, Y)(V') \in (dx, dy)] \geq d^X \mathbf{1}_{[y_E - 2\delta y, y_E + 2\delta y]}(y)\, \delta_{x_E - v\, t}(dx)\, dy. \end{aligned}$$

Lemma 3.7.1.4. *We can define a stopping time U_A^∞ extending the above definition of U_A as described in Theorem 3.4.3.*

The next Subsection concludes the proof of Theorem 3.4.3 and the Propositions 3.7.1.1-3 are proved afterwards by order of occurrence. The technical proof of Lemma 3.7.1.4 is given in Appendix F.

test

3.7.1.2 Proof of Theorem 3.4.3 as a consequence of Propositions 3.7.1.1-4

Let $\ell_E \geq 1$. With ζ the uniform distribution over $\mathcal{D}_1$, we deduce from Theorem 3.4.1 that there exists $c_M, t_M > 0$ such that :

$$\mathbb{P}_\zeta\left[(X,Y)_{t_M} \in (dx',\, dy')\right] \geq c_M\, \mathbf{1}_{\left\{(x',y')\in\mathcal{D}_{\ell_E+1}\right\}}\, dx'\, dy'.$$

With the Markov property and Proposition 3.7.1.3, we conclude that for any $t > 0$, there exists d^X such that for any $x_E \in [-\ell_E, \ell_E]$ and $y_E \in [1/\ell_E,\, \ell_E]$, there exists a stopping time V (of the form $t_M + V'(X_{t_M}, x_E)$) such that :

$$\mathbb{P}_\zeta\left[(X,\, Y)(V) \in (dx,\, dy)\right] \geq d^X\, c_M\, \mathbf{1}_{[y_E-2\,\delta y,\, y_E+2\,\delta y]}(y)\; \delta_{x_E - v\,t}(dx)\; dy.$$

This concludes the proof of Theorem 3.4.3 thanks to Proposition 3.7.1.1 and 3.7.1.2, with : $c := D^X/(d^X\, c_M)$. □

3.7.1.3 Proof of Proposition 3.7.1.1,

$\{\mathbf{t} < \mathbf{T}_{\delta \mathbf{y}} \wedge \mathbf{T_J}\}$: On this event, we have : $X_{U_A} = x_E - v\,t$ and $Y_{U_A} \in [y_E - \delta y, y_E + \delta y]$. Indeed, as in the proof of Lemma 3.6.2.5, we have chosen our stopping times to ensure that no jump for X can occur before time $T_J \wedge t \wedge T_{\delta y}$. We also rely on the Girsanov transform and Theorem 3.6.1 to prove that, during the time-interval $[0, t]$, Y is indeed sufficiently diffused (since we care now for an upper-bound, we can neglect the effect of assuming $t < T_{\delta y}$). □

3.7.1.4 Proof of Proposition 3.7.1.2 :

By Theorem 3.6.1,

$$\mathbb{P}_{(x_E,\, y_E)}\left(T_{\delta y} < t\right) \leq C_G\, \mathbb{P}^G_{(x_E,\, y_E)}\left(T_{\delta y} < t\right) \leq C_G\, \mathbb{P}^G_0\left(T_{\delta y} < t\right) \to 0 \text{ as } t \to 0.$$

Moreover :

$$\mathbb{P}_{(x_E,\, y_E)}\left(T_J < t\right) \leq \mathbb{P}\left(M([0,t] \times \mathcal{J}) \geq 1\right) \leq \nu(\mathbb{R})\; f_\vee\, t \to 0 \text{ as } t \to 0. \qquad \square$$

3.7.1.5 Proof of Proposition 3.7.1.3

The idea is to let X decrease until it reaches $x_E - v\,t$ by ensuring that no jump occurs. We then know the time V needed for this to happen. Then, thanks to Theorem 3.6.1 and Lemma 3.6.1.2, we deduce a lower-bound on the density of Y on $[y_E - 2\,\delta y, y_E + 2\,\delta y]$. We have already proved a stronger result for Lemma 3.6.2.4, that we let the reader adapt. □

3.7.2 Proof of Theorem 3.4.6

Except that we use Theorem 3.4.4 instead of 3.4.1, which constrains the shape of E, the proof is quite immediately adapted from the previous Subsection 3.7.1. $\square$

3.7.3 Proof of Theorem 3.4.3 in the case $d \geq 2$

The difficulty in this case is that, as long as no jump has occurred, X_t stays confined in the line $x + \mathbb{R}_+.\mathbf{e_1}$. The "absorption" thus cannot occur before a jump. Thus, we first wait for a jump to diffuse on $\mathbb{R}^d$ and then let Y diffuse independently in the same way as in Subsection 3.7.1. These two steps are summarized in the following :

Proposition 3.7.3.1. *Given any $\rho > 0$, $E \in \mathbf{D}$ and $\epsilon_X \in (0, 1)$, there exists t^X, c^X, $x^X_\vee > 0$ and $0 < y^X_\wedge < y^X_\vee$ such that for any $(x_E, y_E) \in E$, there exists a stopping time U^X such that :*

$$\left\{\tau_\partial \wedge t^X \leq U^X\right\} = \left\{U^X = \infty\right\}, \quad \mathbb{P}_{(x_E,y_E)}(U^X = \infty,\, t^X < \tau_\partial) \leq \epsilon_X \exp(-\rho\, t^X),$$
$$\text{and } \mathbb{P}_{(x_E,y_E)}\Big(X(U^X) \in dx \;,\; Y(U^X) \in [y^X_\wedge \;,\; y^X_\vee] \;,\; U^X < \tau_\partial\Big) \leq c^X\, \mathbf{1}_{B(0,x^X_\vee)}(x)\, dx.$$

We defer the proof in Subsection 3.7.3.2.

Proposition 3.7.3.2. *Given any ρ, $x^X_\vee > 0$, $0 < y^X_\wedge < y^X_\vee$ and $\epsilon_Y \in (0, 1)$, there exists t^Y, $c^Y > 0$ and $0 < y^Y_\wedge < y^Y_\vee$ such that for any $(x, y) \in B(0, x^X_\vee) \times [y^X_\wedge \;,\; y^X_\vee]$ there exists a stopping time T^Y such that :*

$$\mathbb{P}_{(x,y)}(T^Y \leq t^Y \wedge \tau_\partial) \leq \epsilon_Y \exp(-\rho\, t^Y),$$
$$\text{and } \mathbb{P}_{(x,y)}\Big((X, Y)\,(t^Y) \in (dx, dy) \;;\; t^Y < T^Y \wedge \tau_\partial\Big) \leq c^Y\, \delta_{\{x - v\, t^Y\, \mathbf{e_1}\}}(dx)\; \mathbf{1}_{[y^Y_\wedge \,,\, y^Y_\vee]}(y)\, dy.$$

The proof is almost exactly the same as in Subsection 3.7.1, thus left to the reader. It leads to define U_A as below.

• $U_A := U^X + t^Y$ on the event $\left\{U^X < t^X \wedge \tau_\partial\right\} \cap \{t^Y < \widetilde{\tau_\partial} \wedge \widetilde{T}^Y\}$, where $\widetilde{\tau_\partial}$ and $\widetilde{T}^Y$ are defined as respectively τ_∂ and T^Y for the solution $(\widetilde{X}_t, \widetilde{Y}_t)$, defined on the event $\left\{U^X < t^X \wedge \tau_\partial\right\}$, of :

$$\begin{cases} \widetilde{X}_t = X(U^X) - v\, t\, \mathbf{e_1} + \displaystyle\int_{[U^X, U^X+t] \times \mathbb{R}^d \times (\mathbb{R}_+)^2} w\, \varphi\left(\widetilde{X}_{s^-},\, \widetilde{Y}_s,\, w,\, u_f,\, u_g\right)\; M(ds, dw, du_f, du_g) \\ \widetilde{Y}_t = Y(U^X) + \displaystyle\int_0^t \psi\left(\widetilde{X}_s,\, \widetilde{Y}_s\right) ds + \int_{U^X}^{U^X+t} dB_r. \end{cases}$$

• Else $U_A := \infty$.

Lemma 3.7.3.3. *There exists a stopping time U^∞_A extending the above definition of U_A as described in Theorem 3.4.3 (with $t = t^X + t^Y$ here).*

This Lemma 3.7.3.3 is a bit technical but totally elementary from the way we define U^X and T^Y and easily adapted from the proof given in Appendix F. The proof is left to the reader.

3.7.3.1 Proof of Theorem 3.4.3 as a consequence of Propositions 3.7.3.1-3

Given E, ρ and some $\epsilon \in (0,1)$, we define $\epsilon_X := \epsilon/4$ and deduce from Proposition 3.7.3.1 the values t^X, c^X, $x^X_\vee$, $y^X_\wedge$, $y^X_\vee$ and the definition for the stopping times U^X with the associated properties.

With $\epsilon_Y := \epsilon \, \exp(-\rho\, t^X)/2$, we then deduce from Proposition 3.7.3.2 the values t^Y, c^Y, $y^Y_\wedge$, $y^Y_\vee$ and the stopping time T^Y with the associated properties. Defining, for some $(x,y) \in E$, U_A as in Proposition 3.7.3.3 and combining these results :

$$\left\{\tau_\partial \wedge (t^X + t^Y) \le U_A\right\} = \{U_A = \infty\}\,, \tag{3.7.5}$$

$$\mathbb{P}_{(x,y)}\left[(X,Y)\,(U_A) \in (dx,dy)\ ;\ U_A < \tau_\partial\right] \le c^X\, c^Y\, \mathbf{1}_{B(0, x^X_\vee + v\, t^Y)}(x)\, \mathbf{1}_{[y^Y_\wedge\,,\, y^Y_\vee]}(y)\, dx\, dy, \tag{3.7.6}$$

$$\mathbb{P}_{(x,y)}\Big(U_A = \infty, t^X + t^Y < \tau_\partial\Big) \le \epsilon_X\, \exp(-\rho\, t^X) + \epsilon_Y\, \exp(-\rho\, t^Y) \le \epsilon\, \exp(-\rho\,[t^X + t^Y]), \tag{3.7.7}$$

where we used the definitions of ϵ_X, ϵ_Y and that $t^Y \le \ln(2)/\rho$ (i.e. $1/2 \le \exp(-\rho\, t^Y)$) in the last inequality.

For the opposite upper-bound, we recall first that ζ is chosen to be uniform over the compact space Δ, that is included in some $\mathcal{D}_\ell$. Exploiting Theorem 3.4.4 on this set $\mathcal{D}_\ell$, we deduce that there exists $t, c > 0$ such that :

$$\mathbb{P}_\zeta\Big[(X,Y)\,(t) \in (dx,dy)\ ;\ t < \tau_\partial\Big] \ge c\, \mathbf{1}_{B(0, x^X_\vee + v\, t^Y)}(x)\, \mathbf{1}_{[y^Y_\wedge\,,\, y^Y_\vee]}(y)\, dx\, dy. \tag{3.7.8}$$

Combining (3.7.5)–(3.7.8) ends the proof of Theorem 3.4.3 in the case $d \ge 2$. $\square$

3.7.3.2 Proof of Proposition 3.7.3.1

For readability, note that most of the subscripts "X" (except for t^X) from Proposition 3.7.3.1 are removed in this proof.

First, remark that without any jump, $\|X\|$ goes to infinity, which makes the population almost doomed to extinction. We can thus find some time-limit $t_\vee$ such that, even with an amplification of order $\exp(\rho\, t_\vee)$, the event that the population survived without any mutation occurring in the time-interval $[0, t_\vee]$ is exceptional enough. With this time-scale, we can find an upper-bound $y_\vee$ on Y : that the population reaches such size before $t_\vee$ is an exceptional enough event. For the lower-bound, we use the fact that extinction is very strong when the population size is too small. Thus, that the population has survived –at least for a bit– after declining below this lower-bound $y_\wedge$ is also an exceptional enough event.

The last part is to ensure that this first jump is indeed diffuse in X (which is why we need $\nu(dw)$ to have a density w.r.t. Lebesgue with the bound of [H10]).

For $y_\vee > \ell_E > 1/\ell_E > y_\wedge > 0$, $t_\vee, w_\vee > 0$ and initial condition $(x, y) \in E$, let :

$$T_J := \inf\{t \geq 0 \ ; \ \Delta X_t \neq 0\} \tag{3.7.9}$$

$$T_Y^\vee := \inf\{t \geq 0 \ ; \ Y_t = y_\vee\}\,, \qquad T_Y^\wedge := \inf\{t \geq 0 \ ; \ Y_t = y_\wedge\} < \tau_\partial, \tag{3.7.10}$$

On the event $\{T_J < t_\vee \wedge T_Y^\vee \wedge T_Y^\wedge\} \cap \{\|\Delta X_{T_J}\| < w_\vee\}$, we define $U := T_J$. Else $U := \infty$.

To choose $y_\wedge, y_\vee$, $t_\vee$ and $w_\vee$, we refer to the following lemmas :

Lemma 3.7.3.4. *Assume* [H1-4] *and* [H7]*. Then, for any $\rho, \epsilon_1 > 0$, there exists $t_\vee > 0$ such that :*

$$\forall\, (x, y) \in E, \quad \mathbb{P}_{(x,y)}(t_\vee < T_J \wedge \tau_\partial) \leq \epsilon_1 \exp(-\rho\, t_\vee).$$

Lemma 3.7.3.5. *Assume* [H1-4] *and* [H7]*. Then, for any $t_\vee, \epsilon_2 > 0$, there exists $y_\vee > 0$ such that :*

$$\forall\, (x, y) \in E, \quad \mathbb{P}_{(x,y)}(T_Y^\vee < t_\vee \wedge \tau_\partial) \leq \epsilon_2.$$

Lemma 3.7.3.6. *Assume* [H1-4] *. Then, for any $t_S, \epsilon_3 > 0$, there exists $y_\wedge > 0$ such that :*

$$\forall\, x \in \mathbb{R}^d, \quad \mathbb{P}_{(x,y_\wedge)}(t_S < \tau_\partial) \leq \epsilon_3.$$

Lemma 3.7.3.7. *Assume* [H1-4] *. Then, for any $t_\vee, \epsilon_4 > 0$, there exists $w_\vee > 0$ such that :*

$$\forall\, (x, y) \in E, \quad \mathbb{P}_{(x,y)}(\|\Delta X_{T_J}\| \geq w_\vee \ ; \ T_J < t_\vee) \leq \epsilon_4.$$

Lemma 3.7.3.8. *Assume* [H1-4] *and* [H10]*. Assume either* [H11] *or* [H9] *Then, for any $t_\vee > 0$, and any $y_\vee > \ell_E > 1/\ell_E > y_\wedge > 0$, there exists $c, x_\vee > 0$ such that :*

$$\forall\, (x, y) \in E, \quad \mathbb{P}_{(x,y)}\Big(X(U) \in dx \ ; \ U < \tau_\partial\Big) \leq c\, \mathbf{1}_{B(0,x_\vee)}(x)\, dx.$$

Proof of Proposition 3.7.3.1 as a consequence of Lemmas 3.7.3.4-7 Let $\ell_E, \rho, \epsilon > 0$. We first deduce $t_\vee$ from Lemma 3.7.3.4 such that :

$$\forall\, (x, y) \in E, \quad \mathbb{P}_{(x,y)}(t_\vee < T_J \wedge \tau_\partial) \leq \epsilon \exp(-\rho\, t_\vee)/6. \tag{3.7.11}$$

We could take any value for t_S (so possibly 1), yet $\quad t_S = \log(2)/\rho \quad$ seem somewhat more natural. By Lemma 3.7.3.5, we deduce some $y_\vee > 0$ such that :

$$\forall\,(x,y) \in E, \quad \mathbb{P}_{(x,y)}(T_Y^\vee < t_\vee \wedge \tau_\partial) \leq \epsilon\, \exp(-\rho\, t_\vee)/6. \tag{3.7.12}$$

We then deduce $y_\wedge$ from Lemma 3.7.3.6 such that :

$$\sup_{\{x\in\mathbb{R}^d\}} \mathbb{P}_{(x,y_\wedge)}(t_S < \tau_\partial) \leq \epsilon\, \exp(-\rho\, t_\vee)/6.$$

This implies that for any $(x,y) \in E$:

$$\begin{aligned} &\mathbb{P}_{(x,y)}\left(t_\vee + t_S < \tau_\partial\,,\ T_Y^\wedge < t_\vee \wedge \tau_\partial \wedge T_Y^\vee \wedge T_J\right) \\ &\quad \leq \mathbb{E}_{(x,y)}\left(\mathbb{P}_{(X_{T_Y^\wedge}, y_\wedge)}(t_S < \tau_\partial)\ ;\ T_Y^\wedge < t_\vee \wedge \tau_\partial \wedge T_Y^\vee \wedge T_J\right) \\ &\quad \leq \epsilon\, \exp(-\rho\, t_\vee)/6. \end{aligned} \tag{3.7.13}$$

By Lemma 3.7.3.8, we can now choose c and $x_\vee$ such that :

$$\forall\,(x,y) \in E, \quad \mathbb{P}_{(x,y)}\Big(X(U) \in dx\ ;\ U < \tau_\partial\Big) \leq c\,\mathbf{1}_{B(0,x_\vee)}(x)\,dx.$$

By construction of U, and noting $t^X := t_\vee + t_S$, it is clear that $U \geq \tau_\partial \wedge t^X$ is equivalent to $U = \infty$. Combining (3.7.11), (3.7.12) and (3.7.13) :

$$\begin{aligned} \mathbb{P}_{(x,y)}(U = \infty\,,\ t_\vee + t_S < \tau_\partial) &\leq \mathbb{P}_{(x,y)}(t_\vee < T_J \wedge \tau_\partial) + \mathbb{P}_{(x,y)}(T_Y^\vee < t_\vee \wedge \tau_\partial) \\ &+ \mathbb{P}_{(x,y)}\left(\|\Delta X_{T_J}\| \geq w_\vee\right) + \mathbb{P}_{(x,y)}\left(t_\vee + t_S < \tau_\partial\,,\ T_Y^\wedge < t_\vee \wedge \tau_\partial \wedge T_Y^\vee \wedge T_J\right) \\ &\leq \epsilon\, \exp(-\rho\, t_\vee)/2 = \epsilon\, \exp(-\rho\, t^X). \end{aligned}$$

□

Proof of Lemma 3.7.3.4 Exploiting assumption $[H7]$, as long as $\|X\|$ is sufficiently large, we can ensure that the growth rate of Y is largely negative, leading to a quick extinction. The proof is similar to the one of Lemma 2.3.2.2 in Chapter 1, where more details can be found. We consider the autonomous process Y^D as an upper-bound of Y where the growth rate is replaced by r_D. For any t_D and ρ, there exists r_D (a priori negative) such that whatever y_D the initial condition of Y^D, survival of Y^D until t_D (i.e. $t_D < \tau_\partial^D$) happens with a probability smaller than $\exp(-2\,\rho\, t_D)$. By assumption $[H7]$, we define $x_\vee$ such that for any x, $\|x\| \geq x_\vee$ implies $r(x) \leq r_D$. We then deduce :

$$\forall\,(x,y), \quad \mathbb{P}_{(x,y)}(\forall\, t \leq t_D,\ \|X_t\| \geq x_\vee\ ;\ t_D < \tau_\partial) \leq \sup_{y_D > 0} \mathbb{P}_{y_D}\left(t_D < \tau_\partial^D\right) \leq \exp(-2\,\rho\, t_D).$$

Let $t_E := (x_\vee + \ell_E)/v$ and assume $t_\vee \geq t_E$. A.s. on $\{t_\vee < T_J \wedge \tau_\partial\}$ for any $(x,y) \in E$:

$$\forall\, t_E \leq t \leq t_\vee, \quad \|X(t)\| = \|x - v\,t\,\mathbf{e_1}\| \geq x_\vee.$$

Using inductively the Markov property at times $t_\vee := t_E + k\, t_D$ for $k \geq 1$, we obtain :

$$\forall\, (x, y), \qquad \exp[\rho\, t_\vee]\, \mathbb{P}_{(x,y)}(t_\vee < T_J \wedge \tau_\partial) \leq \exp(\rho\, [t_E - k\, t_D]) \underset{k\to\infty}{\longrightarrow} 0. \qquad \square$$

Proof of Lemma 3.7.3.5 : This is an immediate consequence of the fact that Y is upper-bounded by the process $Y^\vee$ given in (3.3.1) with initial condition ℓ_M. This bound is uniform in the dynamics of X_t and M and uniform for any $(x, y) \in E$. It is classical that a.s. $\sup_{t \leq t_\vee} Y^\vee_t < \infty$, which proves the Lemma, see e.g. Lemma 3.3 in [BM15]. $\square$

Proof of Lemma 3.7.3.6 : Like in the proof of Proposition 1.4.2.3 in Chapter 1, cf Appendix D, we use $r_\vee$ as the upper-bound of the growth rate of the individuals to relate to the formulas for Continuous State Branching Processes. Referring for instance to [Pa16] Subsection 4.2, notably Lemma 5, it is classical that 0 is an absorbing boundary for these processes (we even have explicit formulas for the probability of extinction). This directly entails the result of the present Lemma that the probability of extinction goes uniformly to zero as the initial population size goes to zero. $\square$

Proof of Lemma 3.7.3.8 : For $x_\vee := \ell_E + v\, t_\vee$, let :

$$c := \sup \left\{ \frac{g(x, w)\, \nu(w)}{\int_{\mathbb{R}^d} g(x, w')\, \nu(w')\, dw'} \;;\; \|x\| \leq x_\vee \;,\; w \in \mathbb{R}^d \right\} < \infty. \tag{3.7.14}$$

Conditionally on $\mathcal{F}^*_{T_J}$ on the event $\{U < \tau_\partial\} \in \mathcal{F}^*_{T_J}$(cf Appendix G), the law of $X(T_J)$ is given by :

$$\frac{g(X[T_J-]\;,\; x - X[T_J-])\; \nu(x - X[T_J-])}{\int_{\mathbb{R}^d} g(X[T_J-], w')\; \nu(w')\, dw'}\, dx.$$

Note also that a.s. $\|X[T_J-]\| \leq \ell_E + v\, t_\vee = x_\vee$ (since no jump has occurred yet).

Since $\|\Delta X_{T_J}\| \leq w_\vee$ on the event $\{U < \tau_\partial\}$, with $\bar{x}_\vee := x_\vee + w_\vee$, we get the following upper-bound of the law of $X(T_J)$:

$$\begin{aligned} \mathbb{P}_{(x,y)}\left(X(U) \in dx \;;\; U < \tau_\partial\right) &= \mathbb{P}_{(x,y)}\left(\mathbb{E}\left[X(U) \in dx \,\middle|\, \mathcal{F}^*_{T_J}\right] \;;\; U < \tau_\partial\right) \\ &\leq c\, \mathbf{1}_{B(0, \bar{x}_\vee)}(x)\, dx. \end{aligned} \qquad \square$$

What is left in the proofs of Theorem 3.1 is by now either left to the reader or in the following Appendices. We recall that Appendix C is dedicated to the results needed to prove $(A2)$; Appendix D to Theorem 3.6.1 related to the Girsanov transform ; Appendix E to the proofs for $(A1)$; Appendix F to the proof that U^∞_A is well-defined for $(A3_F)$; and finally Appendix G to the filtration up to the jumping time.

Appendix C : Proof of Theorem 3.4.5

C.1 Combine all the inequalities, proof of Theorem 3.4.5

We first prove that the inequalities (3.5.2), (3.5.3) and (3.5.4) given by Propositions 3.5.2.1-4 imply an upper-bound on $\mathcal{E}_\infty^Y \wedge \mathcal{E}_\infty^X \wedge \mathcal{E}_X \wedge \mathcal{E}_0$ for sufficiently small ϵ^X, ϵ^+ and ϵ^0.
Assuming first that $\epsilon^X \leq (2\, C_\infty^Y)^{-1}$, we have :

$$\mathcal{E}_\infty^X \leq C_\infty^X\, (3 + 3\, \mathcal{E}_X + 2\, \mathcal{E}_0)\,, \qquad \mathcal{E}_\infty^Y \leq C_\infty^Y\, C_\infty^X\, (4 + 4\, \mathcal{E}_X + 2\, \mathcal{E}_0)\,.$$

Assuming further that $\epsilon^+ \leq (8\, C_\infty^Y\, C_\infty^X)^{-1}$:

$$\mathcal{E}_X \leq C_X\, (2 + 3\, \mathcal{E}_0)\,, \qquad \mathcal{E}_\infty^X \leq C_\infty^X\, C_X\, (9 + 11\, \mathcal{E}_0)\,, \qquad \mathcal{E}_\infty^Y \leq C_\infty^Y\, C_\infty^X\, (12 + 14\, \mathcal{E}_0)\,.$$

Assuming further that $\epsilon^0 \leq (60\, C_\infty^Y\, C_\infty^X\, C_X)^{-1}$: $\qquad (2 \times [14 + 11 + 3] \leq 60)$

$$\mathcal{E}_0 \leq 50\, C_0, \quad \mathcal{E}_X \leq 152\, C_X\, C_0, \quad \mathcal{E}_\infty^X \leq 559\, C_\infty^X\, C_X\, C_0, \quad \mathcal{E}_\infty^Y \leq 712\, C_\infty^Y\, C_\infty^X\, C_0,$$

In particular $\sup_{(x,y)\in\mathbb{R}\times\mathbb{R}_+} \mathbb{E}_{(x,\,y)}\,(\exp{[\rho\,(\tau_E \wedge \tau_\partial)]}) = \mathcal{E}_\infty^Y \wedge \mathcal{E}_\infty^X \wedge \mathcal{E}_X \wedge \mathcal{E}_0 < \infty.$

Let us now specify the choice of the various parameters involved. For any given ρ, we obtain from Proposition 3.5.2.1 the constant y_∞, and C_∞^Y which gives us a value $\epsilon^X := (2\, C_\infty^Y)^{-1}$. We then deduce, thanks to Proposition 3.5.2.2, some value for C_∞^X, ℓ^X and L. We can then be fix $\epsilon^+ := (8\, C_\infty^Y\, C_\infty^X)^{-1}$, and deduce, according to Proposition 3.5.2.3, some value C_X and $\ell^+ > 0$. Now we fix $\epsilon^0 := (60\, C_\infty^Y\, C_\infty^X\, C_X)^{-1}$ and choose, according to Proposition 3.5.2.4, some value C_0 and $\ell^0 > 0$. To make the inequalities (3.5.2), (3.5.3) and (3.5.4) hold, we can just take $\ell := \ell^X \vee \ell^+ \vee \ell^0$. With the calculations above, we then conclude Theorem 3.4.5. □

C.2 Proof of Proposition 3.5.2.3 : phenotypic lag pushed towards the negatives,

Since the norm of X decreases at rate at least v as long as the process stays in $\widetilde{\mathcal{T}}_+ := [0, L] \times \mathbb{R}_+^*$, we know that the process cannot stay in this area during a time-interval larger than $t_\vee := \frac{L}{v}$. This effect will give us the bound $C_X := \exp{(\rho\, L/v)}$.
Moreover, we need to ensure that the transitions from $\mathcal{E}_X$ to $\mathcal{E}_\infty^Y$ are exceptional enough. This is done exactly as for Proposition 1.4.2.2 in Chapter 1, by taking ℓ^+ sufficiently larger than y_∞. The event of having the process reaching ℓ^+ in the time-interval $[0, t_\vee]$ is then exceptional enough.
More precisely, given L and $\ell > y_\infty \geq 1$ and initial condition $(x, y) \in \mathcal{T}_+$, let :

$$C_X := \exp\left(\frac{\rho\, L}{v}\right), \qquad T := \inf\{t \geq 0\ ;\ X_t \leq 0\} \wedge V_E \tag{3.8.1}$$

Lemma (C) *Assume that* $[H1\text{-}4]$ *and* $[H9]$ *hold.*
Then, for any initial condition $(x, y) \in \mathcal{T}_+$, $(X, Y)_T \notin \mathcal{T}_\infty^X$ *a.s. and :*

$$\forall\, t < T, \qquad X_t \leq x - v\, t \leq L - v\, t \qquad \text{so that } T \leq t_\vee := L/v.$$

Because of assumption $[H4]$, it can be proved by induction on the number of jumps previous to $T \wedge t$. The proof is elementary and left to the reader. Thanks to this Lemma,

$$\begin{aligned}
\mathbb{E}_{(x,y)}[\exp(\rho V_E)] &= \mathbb{E}_{(x,y)}\Big[\exp(T)\ ;\ T = V_E\Big] + \mathcal{E}_0\, \mathbb{E}_{(x,y)}\Big[\exp(T)\ ;\ (X,Y)_T \in \mathcal{T}_0\Big] \\
&\quad + \mathcal{E}_\infty^Y\, \mathbb{E}_{(x,y)}\Big[\exp(T)\ ;\ (X,Y)_T \in \mathcal{T}_\infty^Y\Big] \\
&\leq C_X\ (1 + \mathcal{E}_0) + C_X\, \mathcal{E}_\infty^Y\, \mathbb{P}_{y_\infty}(T_\uparrow \leq t_\vee)
\end{aligned}$$

where $T_\uparrow := \inf\left\{t \geq 0\ ;\ Y_t^\uparrow \geq \ell\right\}$, and $Y^\uparrow$is the solution of :

$$Y_t^\uparrow := y_\infty + \int_0^t \psi_\vee\left(Y_s^\uparrow\right)\, ds + B_t \qquad (\text{again } \psi_\vee(y) := -\frac{1}{2\, y} + \frac{r_\vee\, y}{2} - \gamma\, y^3). \quad (3.8.2)$$

We conclude the proof of Proposition 3.5.2.3 by noticing that : $\mathbb{P}_{y_\infty}(T_\uparrow \leq t_\vee) \underset{\ell \to \infty}{\longrightarrow} 0.$
□

Appendix D : Proof of Theorem 3.6.1

D.1 Proof in the case where r is C^1

$$\text{Let} \qquad \|r\|_\infty^G := \sup\left\{|r(x)|\ ;\ x \in \overline{B}(0, x_\vee)\right\}, \quad (3.9.1)$$

$$\|r'\|_\infty^G := \sup\left\{|r'(x)|\ ;\ x \in \overline{B}(0, x_\vee)\right\}. \quad (3.9.2)$$

With $\psi_G^\vee$ an upper-bound of ψ on $\overline{B}(0, x_\vee) \times [y_\wedge, y_\vee]$ (deduced from $[H2]$) and recalling that (X, Y) belongs to this subset until T_G (see (3.6.3)) :

$$< L >_\infty = \int_0^{T_G} \psi(X_s, Y_s)^2 ds \leq t_G \times (\psi_G^\vee)^2. \quad (3.9.3)$$

In the following, we look for bounds on $\int_0^{T_G} \psi(X_s, Y_s) dY_s$, noting that :

$$L_{T_G} + \int_0^{T_G} \psi(X_s, Y_s) dY_s = \int_0^{T_G} \psi(X_s, Y_s)^2 ds \in [0, t_G \times (\psi_G^\vee)^2].$$

$$\int_0^{T_G} \psi(X_s, Y_s) dY_s = \int_0^{T_G} \left(-\frac{1}{2Y_s} + \frac{r(X_s)\, Y_s}{2} - \gamma\, (Y_s)^3 \right) dY_s \tag{3.9.4}$$

Now, thanks to Itô's formula :

$$\begin{aligned} \ln(Y_{T_G}) &= \ln(y) + \int_0^{T_G} \frac{1}{Y_s} dY_s - \frac{1}{2} \int_0^{T_G} \frac{1}{(Y_s)^2} ds \\ \text{thus } \left| \int_0^{T_G} \frac{1}{Y_s} dY_s \right| &\le 2\, (|\ln(y_\wedge)| \vee |\ln(y_\vee)|) + \frac{t_G}{2\, (y_\wedge)^2} < \infty. \end{aligned} \tag{3.9.5}$$

$$\begin{aligned} (Y_{T_G})^4 &= y^4 + 4 \int_0^{T_G} (Y_s)^3 dY_s + 6 \int_0^{T_G} (Y_s)^2 ds \\ \text{thus } \left| \int_0^{T_G} (Y_s)^3 dY_s \right| &\le (y_\vee)^4/4 + 3\, t_G\, (y_\vee)^2/2 < \infty. \end{aligned} \tag{3.9.6}$$

$$\begin{aligned} r(X_{T_G-})\, (Y_{T_G})^2 &= r(x)\, y^2 + 2 \int_0^{T_G} r(X_s)\, Y_s\, dY_s + \int_0^{T_G} r(X_s)\, ds - v \int_0^{T_G} r'(X_s)\, (Y_s)^2\, ds \\ &+ \int_{[0,T_G)\times\mathbb{R}^d\times\mathbb{R}_+} \Big(r(X_{s-} + w) - r(X_{s-}) \Big) \times (Y_s)^2 \\ &\qquad \times \mathbf{1}_{\left\{u_f \le f(Y_s)\right\}}\, \mathbf{1}_{\left\{u_g \le g(X_{s-}, w)\right\}} M(ds, dw, du_f, du_g). \end{aligned} \tag{3.9.7}$$

Since $\forall s \le T_G$, $Y_s \in [y_\wedge, y_\vee]$, we get from $[H3]$ and (3.6.2) :

$$\forall\, s \le T_G,\ \forall\, w \in \mathbb{R}^d, \qquad g(X_{s-}, w) \le g_\vee\ ,\ f(Y_s) \le f_\vee, \qquad \text{and } T_G \le U_{N_J}.$$

Since moreover $T_G \le T_X$:

$$\begin{aligned} &\left| \int_{[0,T_G)\times\mathbb{R}^d\times\mathbb{R}_+} \Big(r(X_{s-} + w) - r(X_{s-}) \Big)\, (Y_s)^2\, \mathbf{1}_{\left\{u_f \le f(Y_s)\right\}}\, \mathbf{1}_{\left\{u_g \le g(X_{s-}, w)\right\}} M(ds, dw, du_f, du_g) \right| \\ &\qquad \le 2\, N_J\, \|r\|_\infty^G\, (y_\vee)^2, \end{aligned}$$

so that (3.9.7) leads to :

$$2 \left| \int_0^{T_G} r(X_s)\, Y_s\, dY_s \right| \le \Big(2\, (N_J + 1)\, \|r\|_\infty^G + \|r'\|_\infty^G\, v\, t_G \Big) \times (y_\vee)^2 + \|r\|_\infty^G\, t_G < \infty. \tag{3.9.8}$$

Inequalities (3.9.5), (3.9.6), (3.9.8) combined with (3.9.3) conclude that L_∞ and $< L >_\infty$ are uniformly bounded. This proves the existence of $0 < c_G < C_G$ such that a.s. $c_G \leq D_\infty \leq C_G$. The rest of the proof is only classical application of Girsanov's tranform theory.

D.2 Extension to the case where r is only Lipschitz-continuous

(3.9.5) and (3.9.6) are still true, so we show that we can find the same bound on $\left|\int_0^{T_G} r(X_s)\, Y_s\, dY_s\right|$ where we replace $\|r'\|_\infty^G$ by the Lipschitz-constant $\|r\|_{Lip}^G$ of r on $\bar{B}(0, x_G^v ee)$, by approximating r by C^1 functions that are $\|r\|_{Lip}^G$-Lipschitz continuous.

Lemma (D). *Suppose r is Lipschitz continuous on $\overline{B}(0,\, x_\vee)$.*
Then there exists $r_n \in C^1\Big(\overline{B}(0,\, x_\vee), \mathbb{R}\Big), n \geq 1$ such that :

$$\|r_n - r\|_\infty^G \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and } \forall\, n \geq 1,\ \|r_n'\|_\infty^G \leq \|r\|_{Lip}^G.$$

D.2.1 Proof that Lemma (D) and the case $r \in C^1$ proves Theorem 3.6.1 :

We just have to prove (3.9.8) with $\|r\|_{Lip}^G$ instead of $\|r'\|_\infty^G$. If we apply this formula for r_n and use Lemma (D), we see that there will be some $C = C(t_G, y_\vee, N_J) > 0$ such that :

$$2\,\left|\int_0^{T_G} r_n(X_s)\, Y_s\, dY_s\right| \leq \Big(2\,(N_J+1)\,\|r\|_\infty^G + \|r\|_{Lip}^G\, v\, t_G\Big)\,(y_\vee)^2 + r_\infty\, t_G + C\,\|r - r_n\|_\infty^G.$$

Thus, it remains to bound :

$$\left|\int_0^{T_G} (r_n(X_s) - r(X_s))\, Y_s\, dY_s\right| \leq t_G\, y_\vee\, \psi_G^\vee\, \|r - r_n\|_\infty^G + |M_n|\,,$$

where $M_n := \displaystyle\int_0^{T_G} (r_n(X_s) - r(X_s))\, Y_s\, dB_s$ has mean 0 and variance :

$$\begin{aligned}\mathbb{E}\left((M_n)^2\right) &= \mathbb{E}\left(\int_0^{T_G} (r_n(X_s) - r(X_s))^2\, {Y_s}^2\, ds\right)\\ &\leq t_G\,(y_\vee)^2\,(\|r - r_n\|_\infty^G)^2 \underset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Thus, we can extract some subsequence $M_{\phi(n)}$ which converges a.s. towards 0. So that a.s. :

$$\begin{aligned}\left|\int_0^{T_G} r(X_s)\, Y_s\, dY_s\right| &\leq \liminf_{n\to\infty}\left\{\left|\int_0^{T_G} r_{\phi(n)}(X_s)\, Y_s\, dY_s\right| + t_G\, y_\vee\, \psi_G^\vee\, \|r - r_{\phi(n)}\|_\infty^G + \left|M_{\phi(n)}\right|\right\}\\ &\leq \frac{1}{2}\Big(2\,(N_J+1)\,\|r\|_\infty^G + \|r\|_{Lip}^G\, v\, t_G\Big)\,(y_\vee)^2 + \frac{1}{2}\,\|r\|_\infty^G\, t_G < \infty \qquad \square\end{aligned}$$

D.2.2 Proof of Lemma (D) :

We begin by extending r on $\mathbb{R}^d$ with $r_G(x) := r \circ \Pi_G(x)$, where Π_G is the projection on $B(0, x_\vee)$ (it is well-known that r can be extended on $\bar{B}(0, x_\vee)$ with the same Lischitz constant). Note that this extension r_G is still $\|r\|^G_{Lip}$-Lipschitz continuous. If we define now :
$r_n := r_G * \phi_n \in C^1$, where (ϕ_n) is an approximation of the identity of class C^1, then :

$$\begin{aligned}\forall\, x, y,\ |r_n(x) - r_n(y)| &= \left|\int_{\mathbb{R}^d} (r_G(x-z) - r_G(y-z))\phi_n(z)dz\right| \\ &\le \|r\|^G_{Lip}\, \|x-y\| \int_{\mathbb{R}^d} \phi_n(z)dz = \|r\|^G_{Lip}\, \|x-y\|.\end{aligned}$$

Thus $\qquad \forall\, n \ge 1, \quad \|r'_n\|^G_\infty \le \|r\|^G_{Lip}, \quad \|r_n - r_G\|^G_\infty \xrightarrow[n\to\infty]{} 0.$ $\square$

Appendix E : Elementary mixing properties

E.1 Proofs of Lemmas 3.6.2.1-3 as a consequence of Lemmas 3.6.2.4 to 3.6.2.6

Let $x_F \in B(0, L)$ and $K > \lfloor 4\, \|x_F + S\, \mathbf{e_1}\| / \delta S \rfloor$.

For $0 \le k \le K$, let $x_k := -S\,\mathbf{e_1} + k/K\,(x_F + S\,\mathbf{e_1})$.

$x_0 = -S\,\mathbf{e_1} \in B(0, 2\,L - 2\,S) \cap \mathcal{R}^{(L)}(t_0, c_0)$ by Lemma 3.6.2.5. By Lemma 3.6.2.6, and since $x_{k+1} \in B(x_k, \delta S/4)$ we deduce by immediate induction over $k \le K$ that there exist $t_k, c_k > 0$ such that : $x_k \in B(0, 2\,L - 2\,S) \cap \mathcal{R}^{(L)}(t_k, c_k)$. In particular with $k = K$, Lemma 3.6.2.3 is proved.

In particular, with still $x_0 := -S\,\mathbf{e_1}$ and $x_1 := (-S + \delta S/2)\,\mathbf{e_1}$, there exists $t_1, c_1 > 0$ such that :

$$\{x_0 + u\,\mathbf{e_1}\ ;\ u \in [\delta S/6\ ,\ 5\,\delta S/6]\} \subset \mathcal{R}^{(L)}(t_1, c_1).$$

From Lemma 3.6.2.4, there exists $t_2, c_2 > 0$ such that : $\forall\, t \in [t_2\ ,\ t_2 + 2\,\delta S/(3\,v)]$, $x_0 \subset \mathcal{R}^{(L)}(t, c_2)$.

By Remark 3.6.2.1, for any $k \ge 1$: $\forall\, t \in [k\,t_2\ ,\ k\,t_2 + 2\,k\,\delta S/(3\,v)]$, $x_0 \subset \mathcal{R}^{(L)}(t, [c_2]^k)$.

In particular for k sufficiently large, applying once more Lemma 3.6.2.6 there exists $t_3, c_3 > 0$ such that : $\forall\, t \in [t_3\ ,\ t_3 + t_0]$, $\quad \mathcal{A} \subset \mathcal{R}^{(L)}(t, c_3)$.

Finally, by induction with Remark 3.6.2.1, and since $x_0 \in \mathcal{R}^{(L)}(t_0, c_0)$:

$$\forall\, k \ge 1,\ \forall\, t \in [t_3\ ,\ t_3 + k\,t_0], \quad \mathcal{A} \subset \mathcal{R}^{(L)}(t, c_3\,[c_0]^k).$$

Lemma 3.6.2.2 is proved.

The proof of Lemma 3.6.2.1 is very similar, except that one goes backward by defining for $\ell_I \geq 2$, $t \geq 0$, $c > 0$:

$$\mathcal{R}'(t, c) := \left\{(x_I, y_I) \in \mathcal{D}_{\ell_I} \ : \mathbb{P}_{(x_I, y_I)}\left[\tau_{\mathcal{A}} \leq t \wedge T_{2\,\ell_I}\right] \geq c\right\}.$$

It is not difficult to adapt the proof of Lemma 3.6.2.6 to ensure the following lemma.

Lemma (E.1) *There exists $t_P, c_P > 0$ such that for any $t \geq 0$, $c > 0$:*

$$\{(x, y) \in \mathcal{D}_{\ell_I} \ : d(x \ , \ \mathcal{R}'(t, c)\ ) \leq \delta S/4\} \subset \mathcal{R}'(t + t_P, c \times c_P).$$

The proof of this lemma is left to the reader. Noting that $\mathcal{A} \subset \mathcal{R}'(0, 1)$ and that for any $(x, y) \in \mathcal{D}_{\ell_I}, d(x, \mathcal{A}) \leq \ell_I + S$, we deduce by immediate induction that $\mathcal{D}_{\ell_I} \subset \mathcal{R}'(k\,t_P, (c_P)^k)$ for any $k \geq 4(\ell_I + S)/\delta S$. This concludes the proof of Lemma 3.6.2.1 with $L := 2\,\ell_I$. □

E.2 Proof of Lemma 3.6.2.5

Since the result holds also under the assumption $[H9]$, we give a proof valid under the two sets of assumptions.

Let $t_0 := S/v$, $y_\wedge := 1/(2\,m)$, $y_\vee := 2\,m$. Let then :

$$\begin{aligned} g_\wedge := \inf\Big\{g(x, w) \ ; \ x \in \Big\{x^1\,\mathbf{e_1} + \delta x \ ; \ x^1 \in [-2\,S, -S] \ , \ \delta x \in \overline{B}(0, \delta S/2)\Big\}\ , \\ w \in \overline{B}(S\,\mathbf{e_1}, \delta S)\Big\}, \end{aligned} \tag{3.10.1}$$

$$T^Y := \inf\left\{t \geq 0 \ ; \ Y_t \notin [y_\wedge, \, y_\vee]\right\}, \qquad f_\wedge := \inf\left\{f(y) \ ; \ y \in [y_\wedge, \, y_\vee]\right\}, \tag{3.10.2}$$

$$g_\vee := \sup\left\{g(x, w) \ ; \ \|x\| \leq 3\,S\right\}, \qquad f_\vee := \sup\left\{f(y) \ ; \ y \in [y_\wedge, \, y_\vee]\right\} \tag{3.10.3}$$

$f_\vee$ is finite due to $[H1]$. It also implies with $[H5]$ that $f_\wedge$ is positive. $g_\wedge$ is positive due to $[H1]$ and $[H8]$. It is also true with $[H9]$ instead of $[H8]$, since $g_\wedge$ is the infimum of a continuous function on a compact set where it does not vanish. Indeed, with the notations of (3.10.1) :

$$\begin{aligned} \|x^1\,\mathbf{e_1} + \delta x\|^2 - \|x^1\,\mathbf{e_1} + \delta x + w\|^2 &= 2\langle(-x^1)\,\mathbf{e_1} - \delta x \ , \ w\rangle - \|w\|^2 \\ &\geq 2\left[S \times (S - \delta S) - \delta S/2 \times (S + \delta S)\right] - (S + \delta S)^2 = S^2 - 5\,S \times \delta S - 2\,(\delta S)^2 > 0, \end{aligned}$$

since $\delta S \leq S/8$, as assumed above, just after (3.6.6).

On the event $\{t_0 < T^Y\}$ the values of X on $[0, t_0]$ are given as functions of M restricted to the subset :

$$\mathcal{X}^M := [0 \ , \ t_0] \times \mathbb{R}^d \times [0, f_\vee] \times [0, g_\vee]. \tag{3.10.4}$$

Let $x_0 := -S\,\mathbf{e_1} + \delta x_0$, with $\delta x_0 \in B(0, \delta S/2)$ and $y_0 \in [1/m,\, m]$.
To ensure one jump of size around S, removing the effect of δx_0, let :

$$\mathcal{J} := [0, t_0] \times B(S\,\mathbf{e_1} - \delta x_0,\, \delta S/2) \times [0, f_\wedge] \times [0, g_\wedge] \tag{3.10.5}$$

We partition $\mathcal{X}^M = \mathcal{J} \cup \mathcal{N}$, where : $\mathcal{N} := \mathcal{X}^M \setminus \mathcal{J}$.
We consider the event :

$$\mathcal{W} = \mathcal{W}^{(x_0,y_0)} := \left\{t_0 < T^Y\right\} \cap \{M(\mathcal{J}) = 1\} \cap \{M(\mathcal{N}) = 0\}\,. \tag{3.10.6}$$

By Theorem 3.6.1, with $L = 3\,S$, $t_G = t_0$, and the values for $y_\wedge$ and $y_\vee$, there exists $c_G > 0$ such that :

$$\begin{aligned}\mathbb{P}_{(x_0,y_0)}\left((X,Y)_{t_0} \in (dx, dy)\ ;\ \mathcal{W}\right)\\ \geq c_G\, \mathbb{P}^G_{(x_0,y_0)}\left((X,Y)_{t_0} \in (dx, dy)\ ;\ \mathcal{W}\right)\end{aligned} \tag{3.10.7}$$

Under the law $\mathbb{P}^G_{(x_0,y_0)}$, the condition $\{M(\mathcal{J}) = 1\}$ is independent of $\{M(\mathcal{N}) = 0\}$, of $\left\{t_0 < T^Y\right\}$ and of Y_{t_0}, cf Lemma 3.6.1.1. Thus, on the event $\mathcal{W}$, the only "jump" coded in the restriction of M on $\mathcal{J}$ is given as $(T_J\ ,\ S\,\mathbf{e_1} - \delta x_0 + W\ ,\ U_f\ ,\ U_g)$, where T_J, U_f and U_g are chosen uniformly and independently on respectively $[0, t_0]$, $[0, f_\wedge]$ and $[0, g_\wedge]$, and $S\,\mathbf{e_1} - \delta x_0 + W$ independently according to the restriction of ν on $B(S\,\mathbf{e_1} - \delta x_0,\, \delta S/2)$ (see notably chapter 2.4 in [DV08]). By $[H6]$, W has a lower-bounded density d_W on $B(0,\, \delta S/2)$.

The following lemma, whose proof is postponed in the next Subsection E.3, motivates this description :

Lemma (E.2) *Under $\mathbb{P}^G_{(x_0,y_0)}$ on the event $\mathcal{W}$, $X_{t_0} = -S\,\mathbf{e_1} + W$.*

Note finally that under $\mathbb{P}^G$, $\{M(\mathcal{N}) = 0\}$ is also independent of $\left\{t_0 < T^Y\right\}$ and of Y_{t_0}, so that :

$$\begin{aligned}\mathbb{P}^G_{(x_0,y_0)}\left[(X,Y)_{t_0} \in (dx, dy)\ ;\ \mathcal{W}\right] = \mathbb{P}[M(\mathcal{N}) = 0]\, \mathbb{P}^G_{y_0}\left(Y_{t_0} \in dy\ ;\ t_0 < T^Y\right)\\ \times \mathbb{P}(M(\mathcal{J}) = 1)\, d_W\, \mathbf{1}_{B(-S\,\mathbf{e_1},\, \delta S/2)}(x)\, dx.\end{aligned} \tag{3.10.8}$$

From (3.10.4) and (3.10.5) :

$$\begin{aligned}\mathbb{P}(M(\mathcal{N}) = 0)\ \mathbb{P}(M(\mathcal{J}) = 1) = t_0\, f_\wedge\, g_\wedge\, \nu\{B(S\,\mathbf{e_1} - \delta x_0,\, \delta S/2)\} \times \exp[-t_0\, f_\vee\, g_\vee\, \nu(\mathbb{R}^d)]\\ \geq t_0\, f_\wedge\, g_\wedge\, d_S\, Leb\{B(0, \delta S/2)\} \times \exp[-t_0\, f_\vee\, \nu(\mathbb{R}^d)] := c_X\end{aligned} \tag{3.10.9}$$

where the lower-bound c_X is independent of x_0 and y_0.

By Lemma 3.6.1.2 (recall the definitions of $y_\wedge$ and $y_\vee$ at the beginning of this subsection),

$$\mathbb{P}^G_{y_0}\left(Y_{t_0} \in dy \ ; \ t_0 < T^Y\right) \geq c_B \, \mathbf{1}_{[1/L,\, L]}(y)\, dy. \tag{3.10.10}$$

Again, c_B is independent of x_0 and y_0.

With (3.10.7), (3.10.8), (3.10.9), (3.10.10) and the value : $c_0 := c_G\, c_B\, c_X\, d_W > 0$, we conclude the proof of Lemma 3.6.2.5 by noting that :

$$\begin{aligned}&\forall\, x_0 \in B(-S\,\mathbf{e_1}, \delta S/2),\ \forall\, y_0 \in [1/L,\, L],\\ &\quad \mathbb{P}_{(x_0,y_0)}\left[(X,Y)_{t_0} \in (dx, dy)\right] \geq c_0\, \mathbf{1}_{B(-S\,\mathbf{e_1},\, \delta S/2)}(x)\, \mathbf{1}_{[1/L,\, L]}(y)\; dx\, dy.\end{aligned}$$

□

E.3 Proof of Lemma (E.2)

Step 1 is to prove that on the event $\mathcal{W}$ defined by (3.10.6) :

$$\forall\, t < T_J, \quad X_t := x_0 - v\, t\, \mathbf{e_1}. \tag{3.10.11}$$

Indeed, $t_0 < T^Y$ implies that for any $t \leq T_J$, $Y_t \in [y_\wedge,\, y_\vee]$. By (3.10.3), any "potential jump" (T'_J, W', U'_f, U'_g) such that $T'_J \leq T_J$ and either $U'_f > f_\vee$ or $U'_g > g_\vee$ will be rejected. By the definition of T_J, with (3.10.4), (3.10.5) and (3.10.6), no other jump can occur, thus (3.10.11) holds.

Note that, in order to prove this rejection very rigorously, we would like to consider the first one of such jumps. This cannot be done however for (X, Y) directly, but is easy to prove for any approximation of M where u_f and u_g are bounded. Since the result does not depend on these bounds and the approximations converge to (X, Y) (and even equal to it before T_J for bounds larger than $(f_\vee, g_\vee)$), (3.10.11) indeed holds.

Step 2 is to prove that the jump at time T_J is surely accepted.

By (3.10.1), (3.10.2), and the definition of (T_J, W, U_f, U_g) :

$$\begin{aligned}U_f \leq f_\wedge \leq f(Y_{T_J})\ , \quad U_g &\leq g_\wedge \leq g(x_0 - v\, T_J\, \mathbf{e_1}\ ,\ S\,\mathbf{e_1} - \delta x_0 + W)\\ &= g(X_{T_J-}\ ,\ S\,\mathbf{e_1} - \delta x_0 + W)\end{aligned}$$

Thus $X_{T_J} = -S\,\mathbf{e_1} + \delta x_0 - v\, T_J\,\mathbf{e_1} + S\,\mathbf{e_1} - \delta x_0 + W = -v\, T_J\,\mathbf{e_1} + W.$

Step 3 is to prove that no jump can be accepted after T_J, as in step 1.

This means : $\forall\, T_J \leq t \leq t_0, \quad X_t = X_{T_J} - v\,(t - T_J)\,\mathbf{e_1} = -v\, t\,\mathbf{e_1} + W.$

This proves in particular Lemma (E.2) with $t = t_0 = S/v$. □

E.4 Proof of Lemma 3.6.2.6

This proof follows the same principles as the one of Lemma 3.6.2.5, so we just mention the adjustments. Of course, the result relies on the Markov property combined with a uniform estimate on the transitions starting from $(x_I,\, y_I)$ with $x_I \in B(x, \delta S/4)$. For any $x_F \in B(x,\, \delta S/4)$, we will ensure that X_{t_0} has a lower-bounded density on $B(x_F, \delta S/2)$.

Let $\delta x := x_F - x_I \in B(0, \delta S/2)$. A lower-bound on the rate of acceptation along the new trajectories is obtained from :

$$g_\wedge := \inf \Big\{ g(x', w) \;;\; x' \in \overline{B}(x, 2\,S) \;,\; w \in \overline{B}(S\,\mathbf{e_1}, \delta S) \Big\} \quad > 0 \text{ by } [H8],$$

so that the accepted jump shall be given in :

$$\mathcal{J} := [0, t_0] \times B(S\,\mathbf{e_1} - \delta x,\, \delta S/2) \times [0, f_\wedge] \times [0, g_\wedge].$$

In the same way as for Lemma (E.2), under $\mathbb{P}^G_{(x_I, y_I)}$ and conditionally on the event $\mathcal{W}$, $X_{t_0} = x_F + W$, where W has a density on $B(0, \delta S/2)$ lower-bounded by some $d_W > 0$. The rest of the proof is almost the same, except for Theorem 3.6.1 where we need to choose $L = \|x\| + S$. $\square$

E.5 Proof of Lemma 3.6.3.2

To adapt the proof of Lemma 3.6.2.6, we only need to remark that g is strictly lower-bounded on $\overline{B}(x - S, \delta S/2) \times \overline{B}(S, \delta S)$ (while the upper-bound $g_\vee$ follows from $[H3]$). Since $g \in C^0$, cf $[H1]$, and by $[H9]$, we only need to ensure that :

$$\begin{aligned}
&\forall\, \delta x, \delta w \in [-\delta S, \delta S], \quad |x - S + \delta x| > |x + \delta x + \delta w|.\\
&\text{For } x \le 0, \quad |x - S + \delta x| \ge S - x - \delta S > -x + 2\,\delta S \ge |x + \delta x + \delta w|.\\
&\text{For } 0 \le x \le S/4, \quad |x - S + \delta x| \ge S - x - \delta S > x + 2\,\delta S \ge |x + \delta x + \delta w|.
\end{aligned}$$

We recall that we have assumed $\delta S \le S/8$, cf (3.6.8). With $g_\wedge$ the associated lower-bound, we need of course to adapt $\mathcal{J}$ to ensure that jumps occur while $X_t \in \overline{B}(x - S, \delta S/2)$.

Let $\delta x_I \in \overline{B}(0, \delta S/8)$, $\delta x_F \in \overline{B}(0, \delta S/4)$ and $t_P := t_0 + \delta S/(8v)$. Note that this ensures : $B(S' \,,\, \delta S/2) \subset B(S, \delta S)$, where $\quad S' := S + \delta x_F - \delta x_I + \delta S/8$.

$$x + \delta x_I - v\,t \in \overline{B}(x - S, \delta S/2) \Leftrightarrow t \in \overline{B}(t_0 + \delta x_I/v \;,\; \delta S/(2v)).$$

Thus, it is natural to choose :

$$\mathcal{X}^M := [0\,,\ t_P] \times \mathbb{R} \times [0, f_\vee] \times [0, g_\vee].$$
$$\mathcal{J} := [t_0 + \delta x_I/v - \delta S/(2v)\,,\ t_0 + \delta x_I/v] \times B(S'\,,\ \delta S/2) \times [0, f_\wedge] \times [0, g_\wedge].$$

The rest is done almost as in Lemma 3.6.2.5. In particular :

$$X_{t_P} = x + \delta x_I - v\, t_P + S + \delta x_F - \delta x_I + \delta S/8 + W = x + \delta x_F + W.$$

Of course, in the lower-bound associated to the probability that $M(\mathcal{J}) = 1$, the factor t_0 is now replaced by the new length of the time-interval, i.e. $\delta S/(2v)$. ☐

E.6 Proof of Lemma 3.6.4.2

Compared to Lemma 3.6.2.5, the first main difference is that the jump is now almost instantaneous. The second is that, in order that $g_\wedge > 0$, we have way less choice in the value of w when $\|x^{(u)}\|$ is large. In particular, the variability of any particular jump will not be sufficient to wipe out the initial diffusion around x deduced from $x \in \mathcal{R}^{(L)}(t, c)$, but will rather make it even more diffuse. To see this, let us compute, for $\delta x \in B(0, \delta S)$, $\delta w \in B(0, \epsilon)$, with $\delta S \wedge \epsilon \leq S/8$:

$$\begin{aligned}\|x + \delta x\|^2 - \|x + \delta x - S\,\mathbf{u} + \delta w\|^2 &= 2\,\langle x + \delta x\,,\ S\,\mathbf{u} - \delta w\rangle - \|S\,\mathbf{u} - \delta w\|^2\\ &\geq (\,{}^7\!/_4 - {}^9\!/_8 \times ({}^1\!/_4 + {}^9\!/_8))\, S^2 - 2\,\epsilon\, x_\vee,\end{aligned}$$

where we used that $\langle \mathbf{u}\,,\ S\,\mathbf{u} - \delta w\rangle \geq 7\,S/8$, and note that $c := {}^7\!/_4 - {}^9\!/_8 \times ({}^1\!/_4 + {}^9\!/_8) > 0$. By taking $\epsilon := c\,S^2/(4\,x_\vee) \wedge S/8$, we thus ensure that $\|x + \delta x\|^2 > \|x + \delta x - S\,\mathbf{u} + \delta w\|^2$.

Let $x_I \in B(x, \delta S/2)$,

$$\begin{aligned}&g_\wedge := \inf\left\{g(x', w)\ ;\ x' \in \overline{B}(x, \delta S)\,,\ w \in \overline{B}(S\,\mathbf{u}, \epsilon)\right\} > 0\\ &\mathcal{X}^M := [0, t_P] \times \mathbb{R}^d \times [0, f_\vee] \times [0, g_\vee], \qquad \text{with } t_P := \epsilon/(2\,v)\\ &\mathcal{J} := [0, t_P] \times B(S\,\mathbf{u} + \epsilon/2\,\mathbf{e_1}\,,\ \epsilon/2) \times [0, f_\wedge] \times [0, g_\wedge].\end{aligned}$$

With the same reasoning as for Subsection E.2, we obtain :

$$X_{t_P} = x_I - \epsilon/3\,\mathbf{e_1} + S\,\mathbf{u} + \epsilon/3\,\mathbf{e_1} + W \quad = x_I + S\,\mathbf{u} + W$$

where the density of W is lower-bounded by d_W on $B(0, \epsilon/2)$, uniformly over x_I (given x), and $y_I \in [1/L,\ L]$ (first under $\mathbb{P}^G_{(x_I, y_I)}$ but we have already seen how to deduce it for $\mathbb{P}$).

Let $x_0, y_0 \in \mathcal{A} \times [1/L,\, L]$. By the Markov property and since $x \in \mathcal{R}^{(L)}(t, c)$:

$$\begin{aligned}
\delta_{(x_0,y_0)} P_{t+t_P}\,(dx_F, dy_f) &\geq c \int_{B(x,\delta S/2)} dx_I \int_{[1/L,\, L]} dy_I\; [\delta_{(x_I,y_I)} P_{t_P}]\,(dx_F, dy_f)\\
&\geq c \times c_I \int_{B(x,\delta S/2)} dx_I \int_{[1/L,\, L]} dy_I\; \mathbf{1}_{B(x_I+S\,\mathbf{u},\epsilon/2)}(x_F)\,\mathbf{1}_{[1/L,\, L]}(y_f)\, dx_F\, dy_f,
\end{aligned}$$

where we used the same reasoning as in Subsection E.2 to deduce $c_I > 0$.

$$\begin{aligned}
\delta_{(x_0,y_0)} P_{t+t_P}\,(dx_F, dy_f) &\geq c \times c_I \times (L - 1/L) \times Leb\{B(x + S\,\mathbf{u}\;,\; \delta S/2) \cap B(x_F\;,\; \epsilon/2)\}\\
&\qquad \times \mathbf{1}_{[1/L,\, L]}(y_f)\, dx_F\, dy_f\\
&\geq c \times c_P\, \mathbf{1}_{B(x+S\,\mathbf{u},\;\delta S/2+\epsilon/3)} \mathbf{1}_{[1/L,\, L]}(y_f)\, dx_F\, dy_f,\\
\text{where } c_P &:= c_I \times (L - 1/L) \times Leb\{B(0\;,\; \delta S/2) \cap B([\delta S/2 + \epsilon/3]\,\mathbf{e_1}\;,\; \epsilon/2)\} > 0.
\end{aligned}$$

In particular $B(x + S\,\mathbf{u}\;,\; \epsilon/3) \in \mathcal{R}^{(L)}(t + t_P, c \times c_P)$.
We just need to replace $\epsilon/3$ by ϵ' to deduce Lemma 3.6.4.2. □

Appendix F : Extension of the stopping time

Recall (with simplified notations) that considering the process (X, Y) with initial condition (x, y), we define for some $t > 0$: $U_A := t$ on the event $\{t < T_{\delta y} \wedge T_J\}$, $U_A := \infty$ otherwise,

$$\begin{aligned}
\text{where } T_{\delta y} &:= \inf\{s\;;\; |Y_s - y| \geq 2\,\delta y\} < \tau_\partial &&\text{for some } \delta y > 0,\\
T_J &:= \inf\{s\;;\; M([0, s] \times \mathcal{J}) \geq 1\}\,,\\
\mathcal{J} &:= \mathbb{R}^d \times [0, f_\vee] \times [0, g_\vee] &&\text{for some } f_\vee, g_\vee > 0.
\end{aligned}$$

Recursively, we also define :

$$\tau_E^{i+1} := \inf\{s \geq \tau_E^i + t : X_s \in E\} \wedge \tau_\partial, \text{ and } \tau_E^0 = 0,$$

and on the event $\{\tau_E^i < \tau_\partial\}$, for any i, we set :

$$\begin{aligned}
T_{\delta y}^i &:= \inf\left\{s \geq \tau_E^i\;;\; |Y_s - Y(\tau_E^i)| \geq 2\,\delta y\right\}\\
U_j^i &:= \inf\left\{s\;;\; M([\tau_E^i, \tau_E^i + s] \times \mathcal{J}) \geq 1\right\},\\
U_A^\infty &:= \inf\{\tau_E^i + t\; : \tau_E^i < \infty\;,\; \tau_E^i + t < T_{\delta y}^i \wedge U_j^i\},
\end{aligned}$$

where in this notation, the infimum equals ∞ if the set is empty, $T_{\delta y}^i := \infty$ and $U_j^i = \infty$ on the event $\{\tau_\partial \leq \tau_E^i\}$.
The proof that all these random times define stopping times is very classical and details are left to the reader. The main point is that there is a.s. a positive gap between

any of these iterated stopping times. We can thus ensure recursively in I that for some sequence of stopping times with discrete values

$$(\tau_E^{i,(n)}, T_{\delta y}^{i,(n)}, U_j^{i,(n)})_{i\leq I, n\geq 1},$$

such that a.s. for n sufficiently large, and $1 \leq i \leq I$:

$$\tau_E^i \leq \tau_E^{i,(n)} \leq \tau_E^i + 1/n < \tau_E^i + t,$$
$$T_{\delta y}^i \leq T_{\delta y}^{i,(n)} \leq T_{\delta y}^i + 1/n\,, \qquad U_j^i \leq U_j^{i,(n)} U_j^i + 1/n$$

It is obvious that U_A^∞ coincide with U_A on the event $\{U_A \wedge \tau_\partial \leq \tau_E^1\}$, while the Markov property at time τ_E^1 and the way U_A^∞ is defined entails that on the event $\{\tau_E^1 < U_A \wedge \tau_\partial\}$, $U_A^\infty - \tau_E^1$ has indeed the same law as $\widetilde{U}_A^\infty$ associated the process $(\widetilde{X}, \widetilde{Y})$ solution of the system (3.4.4) with initial condition $(X(\tau_E^1),\, Y(\tau_E^1)\,)$. □

Appendix G : A specific filtration for jumps

This Appendix extends to our case the result already presented in Chapter 2 : $\mathcal{F}_t$ contains the information carried by M and B until the jump time except the realization of the jump itself.

Let T_J be the first jump time of X, with additive effect W. We then define :

$$\mathcal{F}_{T_J}^* := \sigma\left(A_s \cap \{s < T_J\}\ ;\ s > 0,\ A_s \in \mathcal{F}_s\right).$$

Properties of $\mathcal{F}_{T_J}^*$: If Z_s is $\mathcal{F}_s$-measurable and $s < t \in (0, \infty]$, $Z_s\,\mathbf{1}_{\{s<T_J\leq t\}}$ is $\mathcal{F}_{T_J}^*$-measurable.

Lemma (G1). *For any left-continuous and adapted process Z, Z_{T_J} is $\mathcal{F}_{T_J}^*$-measurable. Reciprocally, $\mathcal{F}_{T_J}^*$ is in fact the smallest σ-algebra generated by these random variables.*
In particular, for any stopping time T, $\{T_J \leq T\} \in \mathcal{F}_{T_J}^*$.

Lemma (G2). *For any $h : \mathbb{R} \to \mathbb{R}_+$ measurable, $(x, y) \in (-L, L) \times \mathbb{R}_+$:*

$$\mathbb{E}_{(x,y)}\left[h(W)\,\middle|\,\mathcal{F}_{T_J}^*\right] = \frac{\int_{\mathbb{R}} h(w)\, f(Y_{T_J}) g(X_{T_J-}, w)\ \nu(dw)}{\int_{\mathbb{R}} f(Y_{T_J}) g(X_{T_J-}, w')\ \nu(dw')}$$

Proof of Lemma (G1) :

For any left-continuous and adapted process Z, $\quad Z_{T_J} = \lim_{n\to\infty} \sum_{k\leq n^2} Z_{\frac{k-1}{n}}\ \mathbf{1}_{\left\{\frac{k-1}{n}<T_J\leq\frac{k}{n}\right\}}$, where by previous property and the fact that Z is adapted : $Z_{\frac{k-1}{n}}\ \mathbf{1}_{\left\{\frac{k-1}{n}<T_J\leq\frac{k}{n}\right\}}$ is

$\mathcal{F}^*_{T_J}$-measurable for any k, n. Reciprocally, for any $s > 0$ and $A_s \in \mathcal{F}_s$:

$$\mathbf{1}_{A_s \cup \{s<T_J\}} = \lim_{n\geq 1} Z^n_{T_J} \qquad \text{where } Z^n_t := \{1 \wedge [n\,(t-s)_+]\} \times \mathbf{1}_{A_s}$$

Now, for any stopping time T, and any $t \geq 0$, $\{t \leq T\} \in \mathcal{F}_t$ and $\{t \leq T\} = \bigcap_{s<t} \{s \leq T\}$, thus $\{T_J \leq T\} \cap \{T_J < \infty\} \in \mathcal{F}^*_{T_J}$. Similarly :

$$\{T_J = T = \infty\} = \bigcap_{s>0} \{s < T\} \cap \{s < T_J \leq \infty\} \in \mathcal{F}^*_{T_J}.$$

Proof of Lemma (G2) :

Let :

$$Z_t := \frac{\int_{\mathbb{R}} h(w')\, f(Y_t) g(X_{t-}, w')\; \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-}, w'')\; \nu(dw'')},$$

which is a left-continuous and adapted process. By Lemma (G1), Z_{T_J} is $\mathcal{F}^*_{T_J}$-measurable.

We note the two following identities :

$$h(W) = \int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} h(w)\; \mathbf{1}_{\{t=T_J\}}\; M(dt,\, dw,\, du_f,\, du_g)$$

$$\frac{\int_{\mathbb{R}} h(w)\, f(Y_{T_J}) g(X_{T_J-}, w)\; \nu(dw)}{\int_{\mathbb{R}} f(Y_{T_J}) g(X_{T_J-}, w')\; \nu(dw')} = \int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} \frac{\int_{\mathbb{R}} h(w')\, f(Y_t) g(X_{t-}, w')\; \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-}, w'')\; \nu(dw'')}\; \mathbf{1}_{\{t=T_J\}}\; M(dt,\, dw,\, du_f,\, du_g),$$

Then, we use Palm's formula to prove that their product with any $Z_s\; \mathbf{1}_{\{s<T_J\leq r\}}$, $s < r$, Z_s $\mathcal{F}_s$-measurable, has the same average :

$$\begin{aligned}
&\mathbb{E}_{(x,y)} \left[h(W)\, Z_s\;;\; s < T_J \leq r\right] \\
&\quad = \mathbb{E}_{(x,y)} \left[Z_s \int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} h(w)\; \mathbf{1}_{\{t=T_J\}}\; M(dt,\, dw,\, du_f,\, du_g)\;;\; s < T_J \leq r\right] \\
&\quad = \mathbb{E}_{(x,y)} \left[\int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} Z_s\, h(w)\; \mathbf{1}_{(s,\,r]}(t)\; \mathbf{1}_{\{t=T_J\}}\; M(dt,\, dw,\, du_f,\, du_g)\right] \\
&\quad = \mathbb{E}_{(x,y)} \left[\int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} \mathbf{1}_{(s,\,r]}(t)\; Z_s\, h(w)\; \mathbf{1}_{\left\{t=\widehat{T_J}\right\}}\; dt\, \nu(dw)\, du_f\, du_g\right],
\end{aligned}$$

where, according to Palm's formula, $\widehat{T_J}$ is the first jump of the process $(\widehat{X},\, \widehat{Y})$ encoded by $M + \delta_{(t,w,u)}$ and B (cf e.g. [DV08] Proposition 13.1.VII). Since $(\widehat{X},\, \widehat{Y})$ coincide with $(X,\, Y)$ at least up to time $t > s$, Z_s was not affected by this change. Moreover :

$$\left\{t = \widehat{T_J}\right\} = \{t \leq T_J\} \cap \{u \leq f(Y_t)\, g(X_{t-}, w)\}\,.$$

Thus :

$$
\begin{aligned}
&\mathbb{E}_{(x,y)}\left[h(W)\, Z_s \;;\; s < T_J \le r\right] \\
&= \mathbb{E}_{(x,y)}\left[\int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} \mathbf{1}_{(s,r]}(t)\, Z_s\, h(w)\ \mathbf{1}_{\left\{u_f \le f(Y_t)\right\}}\, \mathbf{1}_{\{u_g \le g(X_{t-},w)\}}\, \mathbf{1}_{\{t \le T_J\}}\, dt\, \nu(dw)\, du_f\, du_g\right], \\
&\quad = \mathbb{E}_{(x,y)}\left[Z_s \int_s^r \int_{\mathbb{R}} \mathbf{1}_{\{t \le T_J\}}\ h(w)\ f(Y_t)\, g(X_{t-},w)\ \nu(dw)\, dt\right].
\end{aligned}
$$

On the other hand, and with the same spirit :

$$
\begin{aligned}
&\mathbb{E}_{(x,y)}\left[\frac{\int_{\mathbb{R}} h(w')\, f(Y_{T_J}) g(X_{T_J-},w')\ \nu(dw')}{\int_{\mathbb{R}} f(Y_{T_J}) g(X_{T_J-},w'')\ \nu(dw'')}\, Z_s \;;\; s < T_J \le r\right] \\
&\quad = \mathbb{E}_{(x,y)}\Bigg[Z_s \int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} \frac{\int_{\mathbb{R}} h(w')\, f(Y_t) g(X_{t-},w')\ \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-},w'')\ \nu(dw'')} \\
&\qquad\qquad\qquad \times \mathbf{1}_{\{t=T_J\}}\ M(dt,\, dw,\, du_f,\, du_g) \;;\; s < T_J \le r\Bigg] \\
&\quad = \mathbb{E}_{(x,y)}\Bigg[\int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} Z_s\, \mathbf{1}_{(s,r]}(t)\, \frac{\int_{\mathbb{R}} h(w')\, f(Y_t) g(X_{t-},w')\ \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-},w'')\ \nu(dw'')} \\
&\qquad\qquad\qquad \times \mathbf{1}_{\{t=T_J\}}\ M(dt,\, dw,\, du_f,\, du_g)\Bigg] \\
&\quad = \mathbb{E}_{(x,y)}\Bigg[\int_{[0,t]\times(\mathcal{X}\setminus\mathcal{D}_{\ell j})\times\mathbb{R}_+} Z_s\, \mathbf{1}_{(s,r]}(t)\, \frac{\int_{\mathbb{R}} h(w')\, f(Y_t) g(X_{t-},w')\ \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-},w'')\ \nu(dw'')} \\
&\qquad\qquad\qquad \times \mathbf{1}_{\{t \le T_J\}}\ \mathbf{1}_{\left\{u_f \le f(Y_t)\right\}}\, \mathbf{1}_{\{u_g \le g(X_{t-},w)\}}\, dt\, \nu(dw)\, du_f\, du_g\Bigg] \\
&\quad = \mathbb{E}_{(x,y)}\left[Z_s \int_s^r \int_{\mathbb{R}} \mathbf{1}_{\{t \le T_J\}}\ h(w')\ f(Y_t)\, g(X_{t-},w')\ \nu(dw')\, dt\right],
\end{aligned}
$$

which is indeed the same integral as for $h(W)$. $\square$

4 Two level natural selection with a quasi-stationarity approach

This chapter is taken from the preprint with the same name whose ArXiv reference can be found at the end of the bibliography (here is the link for the pdf version : [Chapter 4]). Additional results of simulation are provided in Appendix at the end of the manuscript, just before the bibliography.

Abstract

In a view for a simple model where natural selection at the individual level is confronted to selection effects at the group level, we consider some individual-based models of some large population subdivided into a large number of groups. We then obtain the convergence to the law of a stochastic process with some Feynman-Kac penalization. To analyze the limiting behavior of this law, we use a recent approach, designed for the convergence to quasi-stationary distributions, that generalizes the principles of Harris recurrence. We are able to deal with the fixation of the stochastic process and relate the convergence to equilibrium to the one where fixation implies extinction. We notably establish different regimes of convergence. Besides the case of an exponential rate (the rate being uniform over the initial condition), critical regimes with convergence in $1/t$ are also to notice. We finally address the relevance of such limiting behaviors to predict the long-time behavior of the individual-based model and describe more specifically the cases of weak selection. Consequences in term of evolutive dynamics are also derived, where such a competition is assumed to occur repeatedly with each de novo mutation.

Introduction

We consider a model of two alleles competing in groups of individuals without inter-group migration. This model is derived as a new limit of large population (both within and between groups) from the more realistic individual-based model presented in [Luo13], so as to shed light on the dynamics of the latter. Different scenarii can be observed depending on the effect of alleles on the replication within each group of individuals carrying it and on the replication of groups as a whole (where the groups duplicate identically). The focus is especially on two conflicting behaviors : either the allele favoring replication at group level is less favorable to the individuals carrying it for the competition within groups (case of an altruistic trait) ; or selective effects at group level favors polymorphism while those at individual level favors a specific allele.

We study a specific limiting behavior when the populations sizes are large (both for the number of groups and of individuals within each group). Contrary to [LM15], we allow for stochastic neutral fluctuations of type frequencies within the groups and show that it leads in fact to non-trivial effects of selection. Notably, the strength of selection between groups depends very much upon sufficiently high levels of stochasticity for the dynamics within groups.

This mainly explains the discrepancy between our results and the ones of [LM15], where only the selective effects are kept to the limit. Considering the specifications of their model, we look at the competition between an altruistic allele and an egoistic allele. Type D individuals (the egoistic "Defectors") perform always better inside their group while type C individuals (the altruistic "Cooperators") enhance the global survival of their group. In our case, the equilibrium with a domination of altruistic groups is shown to possibly have a much larger basin of attraction : contrary to the purely selective case (described in [LM15]), it may reasonably be the endpoint of the dynamics even though the altruistic allele constitutes initially a majority in none of the groups. By reasonably, we mean that the convergence is likely to be seen in the individual-based models, as we discuss in Section 4.3. The crucial issue for the relevance of this approximation is that we shall not observe the emergence of pure groups of a specific type if the invasion of one group by individuals of this type is too exceptional to be reasonably seen in the population of m groups, cf Section 4.3.2.

Of course, it also highly depends on the range of parameters involved. Namely, when the level of within group stochasticity is small, we shall retrieve a sharp transition similar to the one obtained in [LM15]. The basin of attraction of altruistic domination is in practice limited to initial conditions where a non-negligible proportion of groups have a large majority of C individuals. The rest of the initial conditions belongs to the basin of attraction of defection, where D individuals have fixed in every group.

In Section 4.3, we will provide a much more detailed presentation of the implication of our results to describe the dynamics of the individual-based model. This discussion is a significant part of this work and clarifies the implications of our rigorous mathematical analysis of the asymptotic model. In this section 4.3, we will also discuss implications of our study in terms of evolutionary dynamics with well-separated mutations : the central question is then to quantify the probability of invasion when one type (the mutant) starts with only one individual in an otherwise homogeneous population (of residents).

But first of all, we will introduce in much more details both the individual-based models (IBM) under consideration and several asymptotic descriptions in the limit of large population sizes. These descriptions takes the form of measure-valued processes that approximate the empirical distribution in the groups of the proportion of one type. This will be done in Section 4.1 for the description of the individual-based model and in Section 4.1.1 for the definitions of these processes and their relation to the IBM. In Section 4.1.2, we focus on a different characterization of the process that will be the main subject of study for the propositions presented in Section 4.2. It is expressed in terms of the law of a stochastic process with a specific conditioning. The propositions which follow provide an exhaustive description of the long-time behavior of this asymptotic solution, with a variety of potential behaviors. Some subsequent results for parameters going either to 0 or infinity are also given, so as to introduce the already-mentioned discussion of Section 4.3.

The proofs of the convergence result to our asymptotic solution and of the propositions of Section 4.2 are deferred respectively to Section 4.5 and Section 4.4. For clarity, the latter are kept in the order of appearance of the Theorems, while the more classical techniques at use in Section 4.5 deserve less attention.

4.1 The derivation of the limiting model from the individual-based model

The individual-based model

Each group contains $n \in \mathbb{N}$ individuals. There are two types of individuals : type D individuals (for Defectors) are selectively advantageous at individual level while type C individuals (for Cooperators) are favorable to the group to which they belong. Replication and selection occur concurrently at the individual and group level according to the Moran process, as presented in Figure 1 of [LM15]. Type C individuals replicate at rate $\bar{\gamma}_I \geq 0$ and type D individuals at rate $\bar{\gamma}_I\,(1+\bar{s}), \bar{s}$ with $\bar{s} \geq 0$. When an individual gives birth, another individual in the same group is selected uniformly at random to die. To reflect the antagonism at the higher level of selection, groups replicate at a rate which depends on the proportion of type C individuals they contain. As a simple case, we take this rate to be of the form $\bar{\gamma}_G \times [1+\bar{r}(k/n)]$, where k/n is the fraction of individuals in the group that are of type C and $\bar{r}(x), x \in [0,1]$ is a non-negative bounded function measuring selective advantage at group level. Similarly as at individual level, the number of groups is maintained at the value m by selecting a group uniformly at random to die whenever a group replicates. The offspring of groups are assumed to be identical to their parent. We refer to [Luo13] for a general presentation of the biological motivations for such models.

Let X^i_t be the number of type C individuals in group i at time t. Then

$$\mu^{m;n}_t := \frac{1}{m} \sum_{i \leq m} \delta_{X^i_t/n}$$

is the empirical measure at time t –of the proportion of type C by group– for a given number of groups m and individuals per group n. $\delta_x(y) = 1$ if $x = y$ and zero otherwise. The X_t^i are divided by n so that $\mu_t^{m;n}$ is a probability measure on $E_n := [0; 1/n; ...; 1]$. For fixed $T > 0$, $\mu_t^{m;n} \in D([0;T]; \mathcal{M}_1(E_n))$, the set of càdlàg processes on $[0;T]$ taking values in $\mathcal{M}_1(E_n)$, where $\mathcal{M}_1(S)$ is the set of probability measures on a set S. The particle process being as described above, the generator of the Markov process $\mu_t^{m;n}$ takes the form :

$$(\mathcal{L}^{m;n}\psi)(v) = \sum_{i,j}(\bar{\gamma}_I\, R_I^{i,j} + \bar{\gamma}_G\, R_G^{i,j})(v) \times \Big(\psi\Big[v + 1/m\,(\delta_{j/n} - \delta_{i/n})\Big] - \psi[v]\Big)$$

where $\psi \in C_b(\mathcal{M}_1([0;1]))$ is a bounded continuous functions, and $v \in \mathcal{M}_1(E_n) \subset \mathcal{M}_1([0;1])$. The transition rates $(\bar{\gamma}_I\, R_I^{i,j} + \bar{\gamma}_G\, R_G^{i,j})$ are given by

$$R_I^{i;j}(v) := \begin{cases} m\, v(i/n)\, i\,(1 - i/n)\,(1 + \bar{s}) & \text{if } j = i - 1; i < n, \\ m\, v(i/n)\, i\,(1 - i/n) & \text{if } j = i + 1; i > 0 \\ 0 \text{ otherwise} \end{cases} \tag{4.1.1}$$

$$\text{and} \qquad R_G^{i,j}(v) := m\, v(i/n)\, v(j/n)\,(1 + \bar{r}[j/n]). \tag{4.1.2}$$

$R_I^{i,j}$ represents individual-level events while $R_G^{i,j}$ represents group-level events.

4.1.1 Different limiting behaviors

In [LM15], Luo and Mattingly consider extensively the limit as $n, m \to \infty$ of the measure valued process $(\mu_t^{m;n})$ with fixed parameters $\bar{\gamma}_I$, $\bar{\gamma}_G$, $\bar{s}$ and $\bar{r}$. The limiting measure π_t satisfies :

$$\partial_t \left\langle \pi_t \middle| f \right\rangle = -\bar{\gamma}_I\, s \left\langle \pi_t \middle| x(1-x)f' \right\rangle + \bar{\gamma}_G \left[\left\langle \pi_t \middle| r\, f \right\rangle - \left\langle \pi_t \middle| f \right\rangle \left\langle \pi_t \middle| r \right\rangle\right]. \tag{4.1.3}$$

They also proved that, with the scaling : $\bar{\gamma}_I = n\,\gamma_I$, $\bar{\gamma}_G = m\,\gamma_G$, $\bar{\gamma}_I\,\bar{s} = s$, $\bar{\gamma}_G\,\bar{r} = r$, as $n, m \to \infty$, the process $(\mu_t^{m;n})$ converges weakly to ν_t, where ν_t satisfies the following martingale problem :

$$\text{for any } f \in \mathcal{C}_b^2, \quad \text{with } \mathcal{L}_{WF} f(x) := x\,(1-x)\left[\gamma_I\, \partial_{xx}^2 f(x) - s\, \partial_x f(x)\right], \tag{4.1.4}$$

$$N_t^f = \left\langle \nu_t \middle| f \right\rangle - \left\langle \nu_0 \middle| f \right\rangle - \textstyle\int_0^t \left\langle \nu_u \middle| \mathcal{L}_{WF} f \right\rangle du + \gamma_G \textstyle\int_0^t \left\langle \nu_u \middle| (\rho - \left\langle \nu_u \middle| \rho \right\rangle) \times f \right\rangle du$$

is a martingale with conditional quadratic variation :

$$< N^f >_t = (\gamma_G)^2 \textstyle\int_0^t \left\{ \left\langle \nu_u \middle| f^2 \right\rangle - \left\langle \nu_u \middle| f \right\rangle^2 \right\} du.$$

There is actually an intermediate limit between these two, which will be the main focus of the current paper. In this limit, the fluctuations inside groups still play a role while the fluctuations between groups are neglected (rather in order to simplify the following analysis than for biological relevance) :

Theorem 4.1.1. *Suppose that as $n, m \to \infty$, we have the convergence of the rates $\bar{\gamma}_I/n \to \gamma_I$, $\bar{\gamma}_I\, \bar{s} \to s$, while* $\limsup \bar{\gamma}_G < \infty$ *and $\{\bar{\gamma}_G\, \bar{r}(x)\}_{x\in[0,1]} \equiv \{r(x)\}_{x\in[0,1]}$ is the same bounded measurable function for any n, m. Suppose that $\mu_0^{m;n}$ is defined by assigning independently the proportion of type C in each group according to the measure $\bar{\mu}_0^{m;n}$, where $\bar{\mu}_0^{m;n} \to \mu_0$ as $m, n \to \infty$. Then, $\mu_t^{m;n}$ converges weakly in $D([0;T]; \mathcal{M}_1([0;1]))$ to μ_t, where μ_t is the unique solution to satisfy for any $f \in C_b^2$ the following equation :*

$$\partial_t \left\langle \mu_t \,\middle|\, f \right\rangle = \left\langle \mu_t \,\middle|\, \mathcal{L}_{WF} f \right\rangle + \left\langle \mu_t \,\middle|\, r\, f \right\rangle - \left\langle \mu_t \,\middle|\, f \right\rangle \times \left\langle \mu_t \,\middle|\, r \right\rangle , \qquad \mu_0 = \mu_0. \tag{4.1.5}$$

Since γ_I is the only diffusion term left in this limit, we shall drop the subscript I from now on. The uniqueness of the solution is proved in the next Section 2.1 as part of Proposition 4.1.2.1, after we identify a more convenient way to describe it. The tightness of the above sequences and the fact that any limiting measure is indeed a solution of (4.1.5) is deferred to Section 4.5.

4.1.2 Definition of the solution of (4.1.5) as a conditional law

Consider X_t the $[0, 1]$-valued solution of the following SDE, with initial condition $X_0 \sim \mu_0$:

$$dX_t := -s\, X_t\, (1 - X_t)\, dt + \sqrt{2\gamma\, X_t\, (1 - X_t)}\, dB_t. \tag{4.1.6}$$

The existence and uniqueness of such a process can be found e.g. in in chapter 5.3.1 of [Daw10].

We will describe the solution of equation (4.1.5) at time t as the marginal distribution of X_t with a Feynman-Kac penalization. To relate the process to our results on convergence towards quasi-stationary distributions, we will also represent this penalization as a conditioning upon survival of the stochastic process.

Since subtracting a constant to r does not change the value of $\left\langle \mu_t \,\middle|\, r\, f \right\rangle - \left\langle \mu_t \,\middle|\, f \right\rangle \left\langle \mu_t \,\middle|\, r \right\rangle$, and recalling that r is bounded, we can easily rewrite (4.1.5) in terms of $\rho(x) = \|r\|_\infty - r(x)$, which is non-negative and bounded :

$$\partial_t \left\langle \mu_t \,\middle|\, f \right\rangle = \left\langle \mu_t \,\middle|\, \mathcal{L}_{WF} f \right\rangle - \left\langle \mu_t \,\middle|\, (\rho - \left\langle \mu_t \,\middle|\, \rho \right\rangle)\, f \right\rangle , \qquad \mu_0 = \mu_0 \tag{4.1.7}$$

We then consider the following Feynman-Kac penalization :

$$Z_t := \exp[-\textstyle\int_0^t \rho(X_s)\, ds] \tag{4.1.8}$$

Proposition 4.1.2.1. *Define for each $t \geq 0$ the probability measure μ_t by :*

$$\langle \mu_t \,\big|\, f \rangle := \mathbb{E}\left[f(X_t)\, Z_t\right] / \,\mathbb{E}\left[Z_t\right], \forall\, f \in \mathcal{C}([0,1]), \; . \tag{4.1.9}$$

$(\mu_t)_{t\geq 0}$ is the unique solution of equation (4.1.5) (and equation (4.1.7)).

This penalization can then be interpreted as the probability that the process has survived while confronted to a death rate of ρ, conditionally on $(X_t)_{t\geq 0}$. More precisely, with T_∂ an exponential r.v. with rate 1 that is independent from X, we define the extinction time as :

$$\tau_\partial := \inf\left\{t \geq 0 \;;\; -\ln(Z_t) \geq T_\partial\right\}, \tag{4.1.10}$$

Clearly, 0 and 1 are absorbing for the dynamics of X. We will also treat these fixation events as another kind of extinction. The hitting times of 0 and 1 are denoted τ_0 and τ_1, and we consider any combination :

$$\tau_{0,\partial} := \tau_\partial \wedge \tau_0, \quad \tau_{1,\partial} := \tau_\partial \wedge \tau_1, \quad \tau_{0,1} := \tau_0 \wedge \tau_1, \quad \tau_{0,1,\partial} := \tau_\partial \wedge \tau_0 \wedge \tau_1. \tag{4.1.11}$$

The extinction rates of δ_0, i.e. $\rho_0 = \rho(0)$, and δ_1, i.e. $\rho_1 = \rho(1)$ will play a crucial role in the long-time behavior of μ_t.

Proposition 4.1.2.2. *With the above notations, we then define for any $t \geq 0$ the probability measure μ_t by :*

$$\mu_t = x_t^0\, \delta_0 + x_t^1\, \delta_1 + x_t^\xi\, \xi_t \tag{4.1.12}$$

where $x_t^0 := \dfrac{\mathbb{E}\left[Z_{\tau_0}\, \exp[-\rho_0(t-\tau_0)]\,;\, \tau_0 < t\right]}{\mathbb{E}\left[Z_t\right]}, \qquad x_t^1 := \dfrac{\mathbb{E}\left[Z_{\tau_1}\, \exp[-\rho_1(t-\tau_1)]\,;\, \tau_1 < t\right]}{\mathbb{E}\left[Z_t\right]},$

$$x_t^\xi := \frac{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\right]},$$

$$\langle \xi_t \,\big|\, f \rangle := \frac{\mathbb{E}\left[f(X_t)\, Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]} = \mathbb{E}\left[f(X_t)\,\middle|\, t < \tau_{0,1,\partial}\right], \forall\, f \in \mathcal{C}([0,1]), \; .$$

$(\mu_t)_{t\geq 0}$ is the unique solution of equation (4.1.5), which is equivalently given by : $\mu_t(dx) := \mathbb{P}_{\mu_0}(X_t \in dx \,\big|\, t < \tau_\partial)$.

Remarks 4.1.2.1. *The solution of (4.1.5) will thus generally be denoted $\mu_0 A_t$ in the following statements. Expressing the dynamics in terms of an extinction rate is done mainly to simplify notations with conditional laws. We just adjusted the reference growth rate, here $\|r\|_\infty$, to ensure that the associated semi-group is sub-conservative.*

Remarks 4.1.2.2. *In practice, it means that one weights specifically any potential trajectory for the proportion of type C individuals inside a group. In order to obtain* *the dynamics that is typical while looking in the past of a uniformly sampled group, the weight of such trajectories is related to the mean number of lineages that are expected to follow this dynamics. For instance, spending time where the reproduction rate is high gives more opportunities for at least one group to follow the trajectory until the end. In this view, note that the solution to equation (4.1.6) is well-known to describe the evolving proportion of an allele under selection in a population without any group (hence no selective effects between groups), cf for instance [Ew04].*

Proof of Proposition 4.1.2.1 : By the Ito formula, for any $f \in \mathbb{C}_b^2$:

$$\begin{aligned}\mathbb{E}\left[f(X_t)\, Z_t\right] &= \langle \mu_0 \,\big|\, f\rangle + \textstyle\int_0^t \mathbb{E}\left[\mathcal{L}_{WF} f(X_s)\, Z_s\right] ds - \int_0^t \mathbb{E}\left[f(X_s)\, \rho(X_s)\, Z_s\right] ds,\\ \mathbb{E}\left[Z_t\right] &= 1 - \textstyle\int_0^t \mathbb{E}\left[\rho(X_s)\, Z_s\right] ds,\end{aligned}$$

Thus :

$$\begin{aligned}\partial_t \langle \mu_t \,\big|\, f\rangle &= \frac{\mathbb{E}\left[\mathcal{L}_{WF} f(X_t)\, Z_t\right]}{\mathbb{E}\left[Z_t\right]} - \frac{\mathbb{E}\left[f(X_t)\, \rho(X_t)\, Z_t\right]}{\mathbb{E}\left[Z_t\right]} + \frac{\mathbb{E}\left[f(X_t)\, Z_t\right]}{\mathbb{E}\left[Z_t\right]} \times \frac{\mathbb{E}\left[\rho(X_t)\, Z_t\right]}{\mathbb{E}\left[Z_t\right]}\\ &= \langle \mu_t \,\big|\, \mathcal{L}_{WF} f\rangle + \langle \mu_t \,\big|\, r\, f\rangle - \langle \mu_t \,\big|\, f\rangle \times \langle \mu_t \,\big|\, r\rangle.\end{aligned}$$

(μ_t) is indeed solution to equation (4.1.5).

Now, we turn to uniqueness. Let $\bar{\mu}$ be a solution to equation (4.1.5), P_t the semi-group associated to X_t, the Wright-Fisher diffusion defined by (4.1.6), $f_0 \in C_b^2([0,1])$, and for $0 \le s \le t$:

$$\bar{n}_t := \exp\left[\textstyle\int_0^t \langle \bar{\mu}_s \,\big|\, r\rangle\, ds\right], \qquad f_s^t(x) = \bar{n}_s \times P_{t-s} f_0(x),$$

$$\text{so that :} \qquad \partial_s f_s^t(x) := \langle \bar{\mu}_s \,\big|\, r\rangle\, f_s^t\,(x) - \mathcal{L}_{WF} f_s^t\,(x),$$

$$\begin{aligned}\langle \bar{\mu}_t \,\big|\, \bar{n}_t\, f_0\rangle = \langle \bar{\mu}_t \,\big|\, f_t^t\rangle := \langle \bar{\mu}_0 \,\big|\, P_t\, f_0\rangle + \textstyle\int_0^t \Big[&\langle \bar{\mu}_s \,\big|\, \mathcal{L}_{WF} f_s^t\rangle + \langle \bar{\mu}_s \,\big|\, r\, f_s^t\rangle - \langle \bar{\mu}_s \,\big|\, f_s^t\rangle \times \langle \bar{\mu}_s \,\big|\, r\rangle\\ &+ \langle \bar{\mu}_s \,\big|\, \langle \bar{\mu}_s \,\big|\, r\rangle \times f_s^t\rangle - \langle \bar{\mu}_s \,\big|\, \mathcal{L}_{WF} f_s^t\rangle\Big]\, ds,\end{aligned}$$

$$\text{so that } \bar{\nu}_t(dx) := \bar{n}_t\, \bar{\mu}_t(dx) \quad \text{solves} \quad \langle \bar{\nu}_t \,\big|\, f_0\rangle = \langle \bar{\nu}_0 \,\big|\, P_t\, f_0\rangle + \textstyle\int_0^t \langle \bar{\nu}_s \,\big|\, r \times P_{t-s}\, f_0\rangle\, ds. \tag{4.1.13}$$

Recalling that we already have a solution μ_t defined through equation (4.1.9), we define similarly :

$$n_t := \exp\left[\textstyle\int_0^t \langle \mu_s \,\big|\, r\rangle\, ds\right], \quad \nu_t(dx) := n_t\, \mu_t(dx). \tag{4.1.14}$$

As previously, ν is also solution to (4.1.13), and we deduce :

$$\begin{aligned}|\langle \nu_t - \bar{\nu}_t \,\big|\, f_0\rangle| &\leq \textstyle\int_0^t |\langle \nu_s - \bar{\nu}_s \,\big|\, r \times P_{t-s}\, f_0\rangle|\, ds\\ &\leq 2\,\|f_0\|_\infty \times \|r\|_\infty \textstyle\int_0^t \|\nu_s - \bar{\nu}_s\|_{TV}\, ds,\end{aligned}$$

with the convention :

$$\|\nu\|_{TV} = \sup_{f_0 \in C_b^2([0,1])} \frac{|\langle \nu \,\big|\, f_0\rangle|}{2\|f_0\|_\infty}$$

Since this is true for any $f_0 \in C_b^2([0,1])$:

$$\|\nu_t - \bar{\nu}_t\|_{TV} = \frac{|\langle \nu_t - \bar{\nu}_t \,\big|\, f_0\rangle|}{2\|f_0\|_\infty} \leq \|r\|_\infty \textstyle\int_0^t \|\nu_s - \bar{\nu}_s\|_{TV}\, ds.$$

By Gronwall's Lemma (with the total variation uniformly bounded), this proves that $\nu_t = \bar{\nu}_t$ for any $t > 0$. Since μ_t (resp. $\bar{\mu}_t$) is deduced from ν_t (resp. $\bar{\nu}_t$) by renormalization, $\bar{\mu}_t = \mu_t$ for any $t > 0$. □

Proof of Proposition 4.1.2.2 : First of all, we note that $\{t < \tau_\partial\} = \{\int_0^t \rho(X_s) ds < T_\partial\}$. By the independence between X and T_∂, $\mathbb{P}(t < \tau_\partial \,\big|\, X) = Z_t$. Thus :

$$\mathbb{E}\left[f(X_t)\ ;\ t < \tau_\partial\right] = \mathbb{E}\left[f(X_t)\, Z_t\right] \text{ and in particular } \mathbb{P}(t < \tau_\partial) = \mathbb{E}(Z_t).$$

This proves that for any $t \geq 0$ $\mathbb{P}_{\mu_0}(X_t \in dx \,\big|\, t < \tau_\partial)$ defines a solution to equation (4.1.9). By Proposition 4.1.2.1, it coincides with the solution with equation (4.1.5).

Since 0 and 1 are absorbing for the process X, we have on the event $\{\tau_0 < t\}$ (resp. $\{\tau_1 < t\}$) the fact that $Z_t := Z_{\tau_0} \exp[-\rho_0(t - \tau_0)]$ and $X_t = 0$ (resp. $Z_t := Z_{\tau_0} \exp[-\rho_1(t - \tau_1)]$ and $X_t = 1$). From this, Proposition 4.1.2.2 is elementary. □

4.2 QSDs and exponential convergence

We know from Proposition 4.1.2.2 that the quasi-stationary distributions for X with extinction at time τ_∂ correspond exactly to the initial conditions for which the solution of equation (4.1.7) is constant in time. Such a QSD will be called stable if it is the quasi-limiting distribution for any initial condition that is close enough in total variation distance.

Since 0 and 1 are absorbing states for X, δ_0 and δ_1 necessarily belong to these QSDs, with extinction rate respectively ρ_0 and ρ_1.

We define the semi-groups associated to our different extinctions :

$$\mu P_t(dx) := \mathbb{P}_\mu(X_t \in dx \;;\; t < \tau_\partial)\,, \qquad \mu A_t(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_\partial)$$
$$\mu P_t^{01}(dx) := \mathbb{P}_\mu(X_t \in dx \;;\; t < \tau_{0,1,\partial})\,, \qquad \mu A_t^{01}(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_{0,1,\partial})$$
$$\mu P_t^{1}(dx) := \mathbb{P}_\mu(X_t \in dx \;;\; t < \tau_{1,\partial})\,, \qquad \mu A_t^{1}(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_{1,\partial})$$

Proposition 4.2.0.1. *There exists a unique QSD $\alpha \in \mathcal{M}_1[(0,1)]$ and a unique capacity of survival h associated to the extinction $\tau_{0,1,\partial}$. With the associated extinction rate ρ_α, it means first that :*

$$\forall\, t > 0, \quad \alpha P_t^{01}(dx) = \exp[-\rho_\alpha\, t]\, \alpha(dx)\,,\; P_t^{01} h = \exp[-\rho_\alpha\, t]\, h$$

Moreover, for any $\mu \in \mathcal{M}_1[(0,1)]$ and $x \in (0,1)$:

$$\alpha(dx) = \lim_{t\to\infty} \mu A^{01}(dx)\,, \qquad h(x) = \lim_{t\to\infty} h_t(x)$$
$$\text{where } h_t(x) := \exp[\rho_\alpha\, t]\, \mathbb{P}_x(t < \tau_{0,1,\partial}). \tag{4.2.1}$$

The convergence are uniformly exponential, in the sense that there exists $\chi, C > 0$ such that :

$$\forall\, \mu \in \mathcal{M}_1[(0,1)], \quad \left\|\mu A_t^{01} - \alpha\right\|_{TV} \vee \|h_t - h\|_\infty \le C\, \exp[-\chi\, t]. \tag{4.2.2}$$
$$\text{in particular} \quad \|\bar{h}\|_\infty := \sup_{\{x\in(0,1),\, t>0\}} \exp[\rho_\alpha\, t]\, \mathbb{P}_x(t < \tau_{0,1,\partial}) < \infty \tag{4.2.3}$$

Moreover, for any $n \ge 2$, h is lower-bounded by a positive constant on $[1/n, 1 - 1/n]$.

We show in the following Subsections that the long-time behavior of the process with only the local extinction rate depends mainly on ρ_α, ρ_0 and ρ_1. In the convergences that follow, we will often have uniform bounds for probability measures belonging for some $n \ge 1$ and $\xi \in (0,1)$ to :

$$\mathcal{M}_{n,\xi} := \left\{\mu \in \mathcal{M}_1([0,1]) \,\middle|\, \mu[1/n, 1] \ge \xi\right\}\,,\; \cup_{n,\xi} \mathcal{M}_{n,\xi} = \mathcal{M}_1([0,1]) \setminus \{\delta_0\}. \tag{4.2.4}$$

$$\text{or in} \quad \mathcal{M}_{n,\xi}^{0,1} := \left\{\mu \in \mathcal{M}_1([0,1]) \,\middle|\, \mu[1/n,\, 1 - 1/n] \ge \xi\right\}\,, \qquad (n \ge 3, \xi > 0) \tag{4.2.5}$$
$$\bigcup_{n,\xi} \mathcal{M}_{n,\xi}^{0,1} = \mathcal{M}_1([0,1]) \setminus \{x\, \delta_0 + (1-x)\, \delta_1 \,\big|\, x \in [0,1]\}.$$

But first, the following Lemma provides some elementary properties of the extinction rate in terms of the function ρ.

Lemma 4.2.0.2. *Assume that μ_∞ is a QSD of the Markov process X for the extinction time τ_∂ defined by (4.1.10). Then, its extinction rate is given by $\langle \mu_\infty \,\big|\, \rho\rangle$.*

The QSD α given in Proposition 4.2.0.1 satisfies :

$$\langle \alpha \,\big|\, \rho \rangle = \rho_\alpha \, \mathbb{P}_\alpha(\tau_{0,1,\partial} = \tau_\partial) < \rho_\alpha.$$

Proof of Lemma 4.2.0.2 Let λ be the extinction rate of the QSD μ_∞. We thus know that for any $t > 0$:

$$\begin{aligned}\lambda &= \frac{-1}{t} \log \mathbb{P}_{\mu_\infty}(t < \tau_\partial) = \frac{-1}{t} \log \mathbb{E}_{\mu_\infty}(Z_t) \\ &= \frac{-1}{t} \log \mathbb{E}_{\mu_\infty}(\exp[-\int_0^t \rho(X_s) ds])\end{aligned}$$

We can simply look at the limit of this expression as t tends to 0 to deduce Lemma 4.2.0.2. With the Fubini Theorem, the expression can also be identified with fixed t, because :

$$\begin{aligned}\mathbb{E}_{\mu_\infty}(\exp[-\int_0^t \rho(X_s) ds]) &= 1 - \int_0^t ds\, \mathbb{E}_{\mu_\infty}(\rho(X_s)\, \exp[-\int_0^s \rho(X_u) du]) \\ &= 1 - \int_0^t ds\, \langle \mu_\infty P_t \,\big|\, \rho \rangle = 1 - \langle \mu_\infty \,\big|\, \rho \rangle \times \frac{1 - \exp -\lambda t}{\lambda}.\end{aligned}$$

Concerning α, we can exploit Theorem 2.6 in [CMS13], that proves that the exit state is independent from the exit time when the initial condition is a QSD, with an exponential law for the exit time. This implies notably that for any $t \geq 0$:

$$\mathbb{E}_\alpha(\exp[\rho_\alpha \tau_{0,1,\partial}] \;;\; \tau_{0,1,\partial} = \tau_\partial \leq t) = \rho_\alpha^t \, \mathbb{P}_\alpha(\tau_{0,1,\partial} = \tau_\partial). \tag{4.2.6}$$

Then, we can prove :

$$\begin{aligned}\langle \alpha \,\big|\, \rho \rangle &= \lim_{t \to 0} \tfrac{1}{t} \mathbb{P}_\alpha(T_\partial < \textstyle\int_0^t \rho(X_s) ds) = \lim_{t \to 0} \tfrac{1}{t} \mathbb{P}_\alpha(\tau_\partial \leq t < \tau_{0,1}) \\ &= \lim_{t \to 0} \tfrac{1}{t} \mathbb{E}_\alpha(\exp[\rho_\alpha \tau_{0,1,\partial}] \;;\; \tau_{0,1,\partial} = \tau_\partial \leq t).\end{aligned}$$

By (4.2.6), this proves $\langle \alpha \,\big|\, \rho \rangle = \rho_\alpha \, \mathbb{P}_\alpha(\tau_{0,1,\partial} = \tau_\partial) < \rho_\alpha$.

4.2.1 $\rho_1 < \rho_0 < \rho_\alpha$: Group selection favoring one allele with a quick fixation

Proposition 4.2.1.1. *Assume that $\rho_1 < \rho_0 < \rho_\alpha$. δ_1 is then the only stable QSD, with convergence rate $\rho_0 - \rho_1$, i.e. :*

$$\forall\, n \geq 1,\ \forall\, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}_{n,\xi}, \quad \|\mu A_t - \delta_1\|_{TV} \leq C_{n,\xi}\, \exp[-(\rho_0 - \rho_1)\, t].$$

In order to obtain this convergence, we considered the dynamics of μA_t restricted on $[0, 1)$, which happens to be given by μA_t^1 because the process is fixed at 1 until its extinction, once 1 is reached. Looking at the dynamics of μA^1_t, we deduce an additional level of convergence for the disappearance of polymorphic groups :

Proposition 4.2.1.2. *Assume that $\rho_0 < \rho_\alpha$. Then, there exists $C > 0$ such that :*

$$\forall \mu \in \mathcal{M}_1([0,1]) \setminus \{\delta_1\}, \quad \left\| \mu A^1_t - \delta_0 \right\|_{TV} \leq C \exp[-(\rho_\alpha - \rho_0)\, t].$$

With the notations of Proposition 4.1.2.2, the two previous propositions imply that :

$$x^0_t \leq 1 - x^1_t \leq C_{n,\xi} \exp^{-(\rho_0 - \rho_1)\, t} \;; \quad x^\xi_t \leq C\, x^0_t \exp^{-(\rho_\alpha - \rho_0)\, t} \leq C \times C_{n,\xi} \exp^{-(\rho_\alpha - \rho_1)\, t}.$$

Remarks 4.2.1.1. *In this limiting model, whatever the selection effects inside the groups, the selective effects at group level favoring any of the pure groups always dominate in the long run. The convergence to the pure C groups population happens in total variation, with at the end an exponential rate of convergence. This rate is given by the competition between pure groups.*

Remarks 4.2.1.2. *We shall precise in Section 4.3 and more specifically in Section 4.3.2 the limits of this description, in particular when one wishes to relate it in terms of the individual-based model. Even if pure C groups happen to dominate in the long run, expecting an exponential convergence rate might be misleading : the initial proportion of pure C groups may be so small that pure D groups would be dominant for a very long time. Some illustrations obtained by simulations of such a case are given in Figure 6.8-6.11 of the Appendix of the manuscript. The main quantities of interest are then the time needed for the competition between groups to compensate the initial domination by pure D groups, assuming that it happens, and the probability that this transition actually occurs for the IBM.*

4.2.2 $\rho_1 < \rho_\alpha < \rho_0$: Group selection favoring one allele with a slow fixation

Proposition 4.2.2.1. *Assume that $\rho_1 < \rho_\alpha < \rho_0$. Then, δ_1 is again the only stable QSD, with convergence rate $\rho_\alpha - \rho_1$, i.e. :*

$$\forall n \geq 1,\ \forall \xi > 0,\ \exists C_{n,\xi} > 0, \quad \forall \mu \in \mathcal{M}_{n,\xi}, \quad \| \mu A_t - \delta_1 \|_{TV} \leq C_{n,\xi} \exp[-(\rho_\alpha - \rho_1)\, t].$$

Again, we have an additional level of convergence, and the quasi-equilibrium is precisely described in terms of the polymorphic quasi-stationary distribution :

Proposition 4.2.2.2. *Assume that $\rho_\alpha < \rho_0$. Then :*

$$\begin{aligned} &\exists \chi^1 > 0, \quad \forall n \geq 1,\ \forall \xi > 0,\ \exists C_{n,\xi} > 0, \quad \forall \mu \in \mathcal{M}_{n,\xi} \setminus \{\delta_1\}, \\ &\qquad \left\| \mu A^1_t - \alpha_1 \right\|_{TV} \leq C_{n,\xi} \exp[-\chi^1\, t], \end{aligned} \tag{4.2.7}$$

where the QSD α_1 has extinction rate ρ_α and is given as $\alpha_1 = y_0\,\delta_0 + y_\alpha\,\alpha$ with the relations :

$$\frac{y_0}{y_\alpha} = \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})}{(\rho_0 - \rho_\alpha)}\,, \qquad y_0 + y_\alpha = 1 \tag{4.2.8}$$

$$\text{and thus } y_\alpha := \frac{(\rho_0 - \rho_\alpha)}{\rho_0 - \rho_\alpha \times \mathbb{P}_\alpha(\tau_{1,\partial} = \tau_{0,1,\partial})}\,, \quad y_0 := \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})}{\rho_0 - \rho_\alpha \times \mathbb{P}_\alpha(\tau_{1,\partial} = \tau_{0,1,\partial})}.$$

Moreover, we know the associated capacity of survival $h^1 := h/y_\alpha$ ($h^1(0) = 0$) and :

$$\forall\, n, \xi,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}_{n,\xi},$$
$$|\exp[\rho_\alpha\, t]\, \mathbb{P}_\mu(t < \tau_{1,\partial}) - \langle \mu \,\big|\, h^1 \rangle| \leq C_{n,\xi}\, \exp[-\chi^1\, t] \tag{4.2.9}$$
$$\text{and} \quad \|\bar{h}^1\|_\infty := \sup_{\{x \in [0,1),\, t > 0\}} \exp[\rho_\alpha\, t]\, \mathbb{P}_x(t < \tau_{1,\partial}) < \infty. \tag{4.2.10}$$

With the notations of Proposition 4.1.2.2, the two previous propositions notably imply that :

$$x^0_t + x^\xi_t = 1 - x^1_t \leq C_{n,\xi}\, \exp^{-(\rho_0 - \rho_1)\, t}\ ; \quad \frac{x^\xi_t}{x^0_t} \to \frac{y_\alpha}{y_0}.$$

Remarks 4.2.2.1. *Some illustrations obtained by simulations of this situation are given in Figure 6.17-6.19 of the Appendix. The maintenance of pure D groups in the population is here mainly due to the fixation of polymorphic groups, so that their proportion relative to polymorphic groups tends to stabilize, while both vanish. It does not affect the asymptotic profile of these polymorphic groups in this limit. Only the proportion of these groups is adjusted : extinction of pure D groups exactly compensates their generation by fixation of polymorphic ones.*

Remarks 4.2.2.2. *If we were to include a small effect of neutral replacements of groups, or consider the individual-based models, this size reduction would imply a higher rate of these fluctuations for the polymorphic profile, as compared to having fixation implying extinction of the groups.*

Remarks 4.2.2.3. *It may happen that the polymorphic QSD actually emerges after a long domination of the other pure groups D. First results of simulations indicate that this quasi-equilibrium may not even be noticeable when looking at the population as a whole since it can emerge at almost the same time as the pure groups, cf Section 4.3.2. Some illustrations obtained by simulations of such a situation are also given in Figure 6.12-6.14 of the Appendix of the manuscript.*

4.2.3 $\rho_1 < \rho_0 = \rho_\alpha$: Group selection favoring one allele with a critical fixation rate

Proposition 4.2.3.1. *Assume that $\rho_1 < \rho_0 = \rho_\alpha$. Then, δ_1 is again the only stable QSD, with convergence rate $\rho_0 - \rho_1$, i.e. :*

$$\forall\, n \geq 1,\ \forall\, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}_{n,\xi}, \quad \|\mu A_t - \delta_1\|_{TV} \leq C_{n,\xi} \times (1+t)\, \exp[-(\rho_0 - \rho_1)\, t].$$

For the next level of convergence, 0 is still dominant, yet the proportion of polymorphic states is vanishing in comparison only at rate $1/t$:

Proposition 4.2.3.2. *Assume that $\rho_0 = \rho_\alpha$. Then, δ_0 is the only QSD with extinction $\tau_{1,\partial}$. Moreover :*

$$\begin{aligned} &\exists\, t_\vee, C > 0, \quad \forall\, t \geq t_\vee,\ \forall\, \mu \in \mathcal{M}_1([0,1]), &&\left\|\mu A^1_t - \delta_0\right\|_{TV} \leq C/(1+t),\\ &\forall\, n, \xi,\ \exists\, t_{n,\xi}, C_{n,\xi} > 0, \quad \forall\, t \geq t_{n,\xi},\ \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi}, &&\left\|\mu A^1_t - \delta_0\right\|_{TV} \geq C_{n,\xi}/t. \end{aligned}$$

With the notations of Proposition 4.1.2.2, the two previous propositions notably imply that provided $x^\xi_0 > 0$:

$$x^0_t \leq 1 - x^1_t \leq C_{n,\xi}(1+t)\, \exp^{-(\rho_0-\rho_1)\, t}\ ; \qquad \frac{x^\xi_t}{x^0_t} = O_{t\to\infty}(1/t).$$

Remarks 4.2.3.1. *In this critical case, pure D groups happen to dominate the transient dynamics, but only because of an asymptotically linear increase of their proportion. This increase is actually due to the fixation of polymorphic groups : these groups, still asymptotically distributed as α, act as a generator of pure D groups that stay conserved. Asymptotically, the selective pressure is indeed uniform between pure D groups and the polymorphic groups as a whole.*

Remarks 4.2.3.2. *The same issue of relevance as in Remark 4.2.1.2 may be noted : the polymorphic QSD might only be reached once pure C groups have already appeared non-negligible.*

4.2.4 $\rho_0 = \rho_1 < \rho_\alpha$: selective effects at group level favoring fixation

Again, the convergence rate of the distribution is given by the competition between the pure groups and the polymorpic QSD. There is in this case a strong dependency on the initial condition regarding the final equilibrium.

Proposition 4.2.4.1. *Assume that $\rho_0 = \rho_1 < \rho_\alpha$. Then, any convex combination of δ_0 and δ_1 is a QSD, with extinction rate ρ_1. The convergence still happens with* *convergence rate* $\rho_\alpha - \rho_1$*, i.e. :*

$$\exists\, C > 0, \quad \forall\, \mu \in \mathcal{M}_1([0,1]),\ \exists\, x \in [0,1], \quad \|\mu A_t - (x\,\delta_0 + (1-x)\,\delta_1)\|_{TV} \le C\,\exp[-(\rho_\alpha - \rho_1)\,t].$$

Moreover, the proportion x *for the limiting QSD is :*

$$x(\mu) := \mathbb{E}_\mu\left[\exp(\rho_1\,\tau_{0,1,\partial})\ ;\ \tau_{0,1,\partial} = \tau_0\right] / \mathbb{E}_\mu\left[\exp(\rho_1\,\tau_{0,1,\partial})\ ;\ \tau_{0,1,\partial} = \tau_{0,1}\right]. \tag{4.2.11}$$

With the notations of Proposition 4.1.2.2, this implies that for x depending on the initial condition :

$$|x^0_t - x| \le C\,\exp^{-(\rho_\alpha - \rho_1)\,t}\ ; \quad |x^1_t - 1 + x| \le C\,\exp^{-(\rho_\alpha - \rho_1)\,t}\ ; \quad |x^\xi_t| \le C\,\exp^{-(\rho_\alpha - \rho_1)\,t}$$

The next level of convergence (with extinction $\tau_{0,1,\partial}$) is the already known convergence to α at exponential rate.

Remarks 4.2.4.1. *Here, the selective effects at group level are allowed to dictate the dynamics of polymorphic groups. For a polymorphic (intermediate) initial condition, it may well happen (depending on* r *and* γ*) that* $x(\mu)$ *is close to 0.*

4.2.5 $\rho_\alpha < \rho_0 \wedge \rho_1$: selective effects at group level strongly favoring polymorphism

Proposition 4.2.5.1. *Assume that* $\rho_\alpha < \rho_0 \wedge \rho_1 := \rho$*. Then, there is only one stable QSD* $\alpha^{0,1}$*, with convergence rate* $\rho - \rho_\alpha$*, i.e. :*

$$\forall\, n \ge 1,\ \forall\, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi}, \quad \left\|\mu A_t - \alpha^{0,1}\right\|_{TV} \le C_{n,\xi}\,\exp[-(\rho_\alpha - \rho)\,t],$$

where $\alpha^{0,1}$ *has extinction rate* ρ_α *and is given as* $\alpha^{0,1} = y_0\,\delta_0 + y_1\,\delta_1 + y_\alpha\,\alpha$ *with :*

$$\frac{y_0}{y_\alpha} = \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})}{(\rho_0 - \rho_\alpha)}\,, \qquad \frac{y_1}{y_\alpha} = \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_1 = \tau_{0,1,\partial})}{(\rho_1 - \rho_\alpha)}\,, \tag{4.2.12}$$

and of course $y_0 + y_1 + y_\alpha = 1$*. Regarding the capacity of survival* $h^{0,1}$ *:*

$$\forall\, n, \xi,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}_{n,\xi},$$
$$|\exp[\rho_\alpha\,t]\,\mathbb{P}_\mu(t < \tau_{1,\partial}) - \langle \mu \,|\, h^{0,1}\rangle| \le C_{n,\xi}\,\exp[-\chi^1\,t] \tag{4.2.13}$$
$$\text{and} \quad \|\bar{h}^{0,1}\|_\infty := \sup_{\{x \in [0,1],\, t > 0\}} \exp[\rho_\alpha\,t]\,\mathbb{P}_x(t < \tau_\partial) < \infty. \tag{4.2.14}$$

where for $x \in (0,1)$*,* $h^{0,1}(x) = h(x)/y_\alpha$ *while* $h^{0,1}(0) = h^{0,1}(1) = 0$*.*

If $\rho_1 < \rho_0$, for any initial condition $\mu = x\,\delta_0 + (1-x)\,\delta_1$ with $x \in (0,1)$, μA_t converges at rate $\rho_0 - \rho_1$ to δ_1 as $t \to \infty$.

If $\rho_1 = \rho_0$, then any such distribution is a QSD with the extinction rate ρ_0.

Remarks 4.2.5.1. *Like in Proposition 4.2.2.2, the QSD $\alpha^{0,1}$ is actually obtained by the stabilization of the profile of polymorphic groups (towards α), then by the compensation between the fixation of pure groups by α and their extinction. Since $h^{0,1}$ is null at 0 and 1 (it vanishes also at their vicinity), we see that the contribution of fixed groups to the survival of the population becomes negligible. These pure groups are in fact driven by the polymorphic groups, as one could see from their lineages : it would not take long to come back to polymorphic ancestors.*

Remarks 4.2.5.2. *Considering an individual selection depending on the frequency in the group, we could easily extend our model to describe the case of a balancing selection. In such an extension, as soon as fixation is not too exceptional, selective effects at group level are however needed to maintain polymorphism without transmission between groups.*

4.2.6 $\rho_1 = \rho_\alpha < \rho_0$: critical vanishing of the polymorphic QSD

Proposition 4.2.6.1. *Assume that $\rho_1 = \rho_\alpha < \rho_0$. Then, δ_1 is again the only stable QSD, yet the convergence is not exponential, and more precisely :*

$$\begin{aligned}
&\exists\, C > 0, \quad \forall\, \mu \in \mathcal{M}_1([0,1]), \quad \|\mu A_t - \delta_1\|_{TV} \le C/(1+t),\\
&\forall\, n \ge 2,\ \forall\, \xi > 0, \quad \exists\, c_{n,\xi} > 0,\ \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi},\\
&\qquad \|\mu A_t - \delta_1\|_{TV} \ge c_{n,\xi}/(1+t).
\end{aligned}$$

For the next level of convergence, we refer to Proposition 4.2.2.2.

Remarks 4.2.6.1. *This case corresponds to a very specific compensation of the parameters, where selective effects at group level exactly compensate for the fixation events. The rate of convergence is slow, because it is driven by the polymorphic groups becoming negligible as compared to the fixed lineages that they generate.*

4.2.7 $\rho_0 = \rho_1 = \rho_\alpha$:

This case is the most counter-intuitive, since any polymorphic component in the initial distribution imposes a predictable final equilibrium without polymorphism.

Proposition 4.2.7.1. *Assume that $\rho_0 = \rho_1 = \rho_\alpha$. Then, any convex combination of δ_0 and δ_1 is a QSD, with extinction rate ρ_1. They are the only ones, and among them, only one is stable :*

$$\forall\, n, \xi, \quad \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi}, \quad \|\mu A_t - (x\,\delta_0 + (1-x)\,\delta_1)\|_{TV} \le C_{n,\xi}/(1+t).$$

where the proportion x for the limiting QSD is :

$$x := \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})/\mathbb{P}_\alpha(\tau_{0,1} = \tau_{0,1,\partial}).$$

Remarks 4.2.7.1. *The distribution inside the interval vanishes so slowly that its flux to* 0 *and* 1 *governs the final equilibrium (with a much quicker stabilization of $\mu A_t^{0,1}$ to α).*

4.2.8 Limits of the parameters

Proposition 4.2.8.1. *Given any $s > 0$ and any bounded function r, $\lim_{\gamma\to\infty} \rho_\alpha(\gamma) = +\infty$.*

Proposition 4.2.8.2. *Given any $\gamma > 0$, $s \geq 0$, and a continuous (and negative) function r^0 with its maximum only in the interior of $(0,1)$, there exists a critical value $R_\vee > 0$ such that for any $R > R_\vee$ and considering the system with $r = R\,r^0$, it holds $\rho_\alpha < \rho_0 \wedge \rho_1$.*

Polymorphism is maintained by any sufficiently large selective effects at group level favoring it.

Proposition 4.2.8.3. *Conversely, given any $\gamma > 0$, $s \geq 0$, and a bounded function r^0, there exists a critical value $R_\wedge > 0$ such that for any $R < R_\wedge$ and considering the system with $r = R\,r^0$, it holds $\rho_0 \wedge \rho_1 < \rho_\alpha$.*

When the selective effects at group level is too small, polymorphism cannot maintain itself.

Remarks 4.2.8.1. *One could expect $\rho_\alpha(\gamma)$ to be first a decreasing function of γ and then increasing. Yet, it seems not to hold true for any general r. Think for instance of two types of equilibria that compete inside $(0,1)$, i.e. r with two localized modes, with a specific optimal value $\gamma_1 < \gamma_2$ for each. In the range $\gamma = \gamma_1$ to $\gamma = \gamma_2$, the QSD shifts from the first mode, where the extinction is becoming much larger as γ increases, to the second mode, where such increase is much less significant. It can happen if there is a very strong mode of r close to a border, that is responsible for the first equilibrium. We may thus observe $\rho_\alpha(\gamma) > \rho_\alpha(\gamma_1) \vee \rho_\alpha(\gamma_2)$ for $\gamma \in (\gamma_1, \gamma_2)$, which contradicts the predicted profile of ρ_α.*

We conjecture that $\lim_{\gamma\to 0} \rho_\alpha(\gamma) = \infty$ also holds for any $s > 0$ and any bounded function r. To ensure this, one should study the behavior of μ_t around the boundary $x = 1$ for very small γ. Yet, when the process X stays close to 1, there is a non-trivial competition between the amplification through the Feynman-Kac penalization and the fixation rate at 1, This analysis is beyond the reach of this work. Even if our conjecture were false, the survival of the QSD would mainly rely on a vicinity of 1, since :

Proposition 4.2.8.4. *For any* $s, \epsilon > 0$ *and measurable bounded* r*, for any* $t \geq 1$ *sufficiently large :* $\mathbb{P}_{1-\epsilon}(t < \tau_\epsilon \,\big|\, t < \tau_\partial) \to 1 \, as \, \gamma \to 0$*, where* $\tau_\epsilon := \inf\{t \geq 0 : X_t \leq \epsilon\}$.

Note that in the deterministic limit ($\gamma = 0$), there is no more extinction but a convergence to 0 at exponential rate. Thus, $\mathbb{P}_{1-\epsilon}(t < \tau_{0,1}) \to 1$ as $\gamma \to 0$ is to be expected.

The previous conjecture would imply the following result :
"Given any bounded function r, there is a critical value $s_\vee$ such that for any $s \geq s_\vee$ and $\gamma \in \mathbb{R}_+$, $\rho_0 \wedge \rho_1 < \rho_\alpha$."
Such result would imply that polymorphism cannot subsist when the selection at the individual level is too large.

4.3 Discussion on these results with regards to the individual-based model

The following discussion aims at answering to the next questions :

- Why is it helpful to consider the limiting model μ_t when we are primarily interested in understanding the dynamics of real populations of finite sizes ?
- Do the asymptotic results of convergence to QSDs and their stability properties provide an accurate idea of the basins of attraction ?
- Is the convergence at exponential rate obtained in the previous convergences representative of the observed dynamics ?
- Are the intra-group fluctuations effectively able to make the inter group selective effects overcome the intra-group selective effects ?

We first present in Subsection 4.3.1 our motivation for the analysis in terms of the conditional distribution μA_t. Notably, the qualitative asymptotic results of the model in finite population size do not depend on the parameter values. It provides a poor understanding of the confronting powers.

Subsection 4.3.2 is focused on the two next questions in the case where the scarcity of intra-group fluctuations might raise some issues. We deal with the last one in Subsections 4.3.3. The numerical results presented in Subsection 4.3.3.1 are replaced in Subsection 4.3.3.2 in the perspective of the discussions on the relevance of group selection models and compared to the more classical results of weak selection in Subsection 4.3.3.3.

An evolutionary perspective is also considered in Subsection 4.3.4, in a situation where mutations are rare. We thus imagine that a population with a resident type dominates until it gets replaced by a mutant subpopulation. Of course, such a replacement is generally rare among the numerous mutations that are continuously generated, but we look on a long time-scale where numerous of these events have happened. For simplicity, we assume that the mutations that emerge in individuals are as likely to promote cooperation than egoism. We then try to answer the following question : Which types of mutations effectively drive the evolutive dynamics of the system ?

Our prediction is based on the expected stability properties of the process with a finite yet large population size around the stable quasi-stationary distributions of the limiting process. Based on these expected stability properties, we also conjecture in Subsection 4.3.5 that in a regime where a polymorphic QSD is stable for the dynamics of μA_t, it is also a very stable a attractor for the dynamics in finite yet large population size. Finally, we conclude in Subsection 4.3.6 by the main conclusions of Section 4.3.

4.3.1 General motivations.

Our purpose in this analysis of the long-time behavior of μ_t is to highlight common features with the individual-based models $\mu_t^{m,n}$, for m and n reasonably large. Note that contrary to μ_t, $\mu_t^{m,n}$ is a priori a random process evolving on the discrete grid :

$$\mathcal{M}^{m,n}([0,1]) := \{\mu \in \mathcal{M}_1([0,1]) \,\big|\, \mu(\textstyle\bigcup_{0\le k\le n}\{k/n\}) = 1\,;\, \forall\, 0\le k\le n,\ m{\times}\mu(\{k/n\}) \in [\![0,m]\!]\}. \tag{4.3.1}$$

So we generally expect a greater diversity of possible scenarii for the individual-based models. The mathematical description of its limiting behavior is however not very informative unless one gets precise quantitative estimates, as we shall see in the next lemma.

Lemma 4.3.1.1. *For any $m, n \ge 1$, δ_0 and δ_1 are the only absorbing points of $\mu_t^{m,n}$. Denoting $\tau_{0,1}^{m,n}$ this absorption time, there exists a unique associated QSD $A^{m,n} \in \mathcal{M}_1(\mathcal{M}^{m,n}([0,1]))$ and a unique capacity of survival $H^{m,n} \in L_\infty(\mathcal{M}^{m,n}([0,1]))$ with extinction rate ρ_A. Moreover, there exists C, γ such that the following convergence results hold :*

$$\begin{aligned}
&\forall\, \mu_0, \mu_1 \in \mathcal{M}^{m,n}([0,1]) \setminus \{\delta_0, \delta_1\},\\
&0 \le \mathbb{P}_{\mu_0}(\exists\, s\ge 0,\ \mu_s^{m,n} = \delta_0) - \mathbb{P}_{\mu_0}(\mu_t^{m,n} = \delta_0) \le C\exp(-\rho_A\, t),\\
&0 \le \mathbb{P}_{\mu_0}(\exists\, s\ge 0,\ \mu_s^{m,n} = \delta_1) - \mathbb{P}_{\mu_0}(\mu_t^{m,n} = \delta_1) \le C\exp(-\rho_A\, t),\\
&|\exp(\rho_A\, t)\mathbb{P}_{\mu_0}(\mu_t^{m,n} = \mu_1) - H^{m,n}(\mu_0) \times A^{m,n}(\{\mu_1\})| \le C\exp(-\gamma\, t).
\end{aligned}$$

The proof of this lemma is deferred to the Appendix. In this proof, the parameters C, ρ_A and γ strongly depend on m and n. So there is no clear dependency on the parameters r and s in this lemma, unless one gets by other means precise estimates of C, ρ_A and γ as well as $\mathbb{P}_{\mu_0}(\exists\, s\ge 0,\ \mu_s^{m,n} = \delta_0)$ and $\mathbb{P}_{\mu_0}(\exists\, s\ge 0,\ \mu_s^{m,n} = \delta_1)$. The study of $A^{m,n}$ and $H^{m,n}$ is also much more complicated than the one of α and h of Proposition 4.2.0.1 because of its detailed state space $\mathcal{M}^{m,n}([0,1])$.

This was our motivation for the study of the μ_t in order to describe the dynamics of $(\mu_t^{m,n})$, as a complement to the study provided in [LM15] of the solution (π_t) of equation (4.1.3). The solution (ν_t) of the martingale problem given by equation (4.1.14) is likely to behave more closely to $(\mu_t^{m,n})$. Yet, we expect a poorly informative limiting behavior similar to the one of Lemma 4.3.1.1 with a proof that seems too technical for now. Furthermore, regarding the limitations that we plan to describe in the next subsection in inferring the probable dynamics of $(\mu_t^{m,n})$ from the one of μ_t, we shall have similar issues in the connection of $(\mu_t^{m,n})$ to ν_t.

To gain some perspective on our analysis on the long time behavior of the solution μ_t of equation (4.1.5), we have used a numerical approximation. The subject would require a much more complete study to be more quantitative, but these simulations already provide illustrations that our convergence results can be informative.

4.3.2 Close to the purely selective case.

We first begin the comparison by connecting to the results of [LM15]. We thus focus on the vicinity of their limits, namely when γ is quite small as compared to r and s. The first prediction that we can get from [LM15] is that μ_t goes to δ_0 provided that $\mu_0([1-\epsilon_\gamma, 1]) = 0$ for a small value of ϵ_γ that we could let tend to 0 as $\gamma \to 0$. Some illustrations obtained by simulations for this case are given in Figure 6.8-6.21 of the Appendix of the manuscript.

4.3.2.1 First conjecture and limitations.

Looking at Proposition 4.2.8.4 and our conjecture that $\lim_{\gamma\to 0} \rho_\alpha(\gamma) = \infty$ also holds for any $s > 0$ and any bounded function r, we would be in the case $\rho_1 < \rho_0 < \rho_\alpha$. From Proposition 4.2.1.1 and 4.2.1.2, one would a priori predict an eventual convergence to δ_1, with the vast majority of the other groups being fixed as pure D groups. We can however expect that the proportion of groups that get fixed as pure C groups might be very tiny for times that are not so large. This would possibly be too tiny for the approximation of $(\mu_t^{m,n})$ by (μ_t) to be really valid in this vicinity of 1 for a large range of values of n and m. Notably, if m is not so large, we can not observe in $(\mu_t^{m,n})$ a proportion less than $1/m$ since it is only one group and the risk of extinction is very large for small sub-populations of groups.

4.3.2.2 Simulations with various initial conditions.

For this purpose, we tried to derive the dynamics starting from Dirac initial conditions with various positions. Our simulations seem indeed to generally indicate that for small values of γ, and as t goes on, the distribution μ_t seems first to be attracted by the vicinity of 0, where type D individuals prevail. $\mu_t(0)$ is then very close to 1 after a time that depends on the initial condition. While looking at the delay in this fixation time (say at a proportion $1-\epsilon$, with a tiny ϵ) between different Dirac initial conditions, we see that it is close to the time needed for the deterministic flow to bring the condition furthest to 0 to the closest one. We also considered the law of $\mu_0 A_t^{0,1}$, i.e. the one of μ_t conditioned on polymorphic groups. It usually stabilizes for long in some kind of attractive state, quite concentrated around 0, cf figure 6.11 in the Appendix of the manuscript, before the actual QSD emerges, with a concentration around 1. When γ is sufficiently small, the transition from one attractor to the other seems very brutal and unexpected while looking at densities. As previously noted, this transition usually occur after the emergence and domination of pure C groups, and possibly long after !

4.3.2.3 The convergence result of $\mu A_t^{0,1}$ to its QSD might be of little significance.

As stated in Proposition 4.2.0.1, there is strictly speaking only one QSD for the extinction time $\tau_{0,1,\partial}$. In the context of the previous paragraph, this exact QSD seems not to play a significant role unless there is initially a non-negligible proportion of polymorphic groups with a vast majority of C types. Asymptotically in the diffusion model, the domination by pure C groups seems not to be following the convergence to the QSD but rather to happen concurrently. With the notations of Proposition 4.1.2.2, we mean that at any time t where the QSD is representative of μA_t^1, x_t^1 is already close to one. The rate of convergence to 1 of x_t^1 might thus be only approach $\rho_\alpha - \rho_1$ at a very late stage.

4.3.2.4 An often observed alternative QSD.

In case m would be sufficiently large for the transitions of certain polymorphic groups to pure C groups to be considered as likely, an alternative distribution, located around 0, might play a role similar to the one of a QSD. Yet, it would be meaningful mainly if one is interested in exceptional events : namely, this "observed QSD" is associated to a large fixation rate towards pure D groups.

4.3.2.5 Scarcity of the transitions towards more cooperation, including towards the real QSD.

The real QSD appears in fact at very low density at the beginning. Since it survives at a better rate than the observed distribution, it simply ends up dominating the whole. Transitions from the first effective attractor to the real QSD can however be shown to be very rare. To see it, we use some small value(s) η that will serve as a threshold of exceptionality. For such η, we define an analogous version $\bar{\mu}_t^\eta$ of the discretized solution of μ_t, with the main difference that we set iteratively to 0 the densities below η. The transition may then be delayed or even unobserved, possibly even at very low values of truncation (10^{-20} for instance). The previously mentioned observed QSD then arises, concentrated near the pure D group type.

This method is questionable, notably because it depends on the time- and space-discretization. Yet, with space-grid of order 100 and stabilization in less than 10,000 steps, we could indeed exhibit and truncate exceptional transitions. We also expect that the delay between the untruncated dynamics and the truncated one is due to the fact that the front towards groups enriched in C types is pulled by these exceptional transitions. The bulk of more polymorphic groups is not so much involved in pushing the proportion of types towards more cooperation. A more advanced numerical scheme would certainly be helpful to quantify the exceptionality of the trajectories leading the front. In this view, this distinction between pulled versus pushed wave could be observed by introducing neutral markers whose density would be followed as in [RG+12].

4.3.2.6 As a conclusion, a non-forgettable dependency on the initial conditions.

When γ is small and there is no group for which C types constitute a large majority, then, from an ecological point of view, D individuals have fixed in almost every group. And even conditionally upon the fact that this fixation has not occurred (or with mutations generating new type C individuals), there is a very stable equilibrium with groups dominated by D individuals.

4.3.3 The contribution of the intra-group fluctuations for intermediate γ.

4.3.3.1 A surprising numerical observation

There is another interesting behavior for not so small values of γ, a linear growth rate r (increasing) and a Dirac mass in the middle of the interval as initial condition. At the beginning of its dynamics, μ_t is close to a Gaussian distribution with an expanding variance and a drift. Except that the variance is larger, it seems first to behave as in the case of small γ and for initial conditions rather close to 0, not much difference can be observed. Yet, although the drift is always first directed towards 0, we may see a u-turn after a while, with a drift now seemingly directed towards 1. It creates the impression that the drift at the individual level is changing, while it is in fact the selective effects at group level that starts to play a significant role. The more diverse the distribution is, the more these effects differentiate between these different realizations and the larger is this additional drift.

In such a case, the role of the real QSD can be much more significant as can be seen in Figures 6.12-6.14 in the Appendix of the manuscript. The convergence to 1 is much more robust against the truncations that we can implement as mentioned in the previous paragraph. We thus expect the estimation of the typical dynamics of $\mu^{m,n}$ by the one of μ_t to be relatively well-justified in this context. In a future work, we plan to validate notably this point by simulations and be more specific regarding the associated range of parameters (including the initial condition). This would demonstrate the crucial importance of having sufficiently large intra-group fluctuations for selective effects at group level to be significant.

4.3.3.2 The contested notion of "group selection".

For such a model where the selective effects are clearly associated with two hierarchical levels without transmission between groups, the notion of selective effects at group level does not appear so ambiguous. It is closely related to the more common expression of "group selection", except that the interpretation of the latter has been quite diverse depending on the authors. Notably, we have to mention the recurrent discussions on the confusions brought about by this notion (see [WGG07], [WGG07b], [WMG11], [LK+07]). Notably, it has been argued that one shall rather relate to the kin selection formalism, which means that one shall weight the selective advantage that Cooperators give to other individuals by the relatedness they share with the recipient. On the other hand, our formalism focus on the group level by specifying the dynamics of a randomly chosen group. We are convinced that a formalism in terms of relatedness could produce an equivalent description but we have not found it particularly helpful. We would happily welcome any suggestion in this regard, as we find it very intricate in our model. Notably, the notion of relatedness is here much more difficult to describe, even in the simplification of weak selection. This is strongly linked to the fact that individuals cannot transfer from one group to another while the internal selective advantage is relative to the mean selective advantage inside the group. In addition, fixation events inside groups are a specificity of this model. The difficulty is also to be expected given that we do not assume, contrary to the usual considerations on the direction of selection, that the selective effects are on different time-scales (see e.g. [O'F08]) nor weak (see e.g. [ML14], [TN06]). Already without such population structure, different selection intensities may lead to qualitatively different relations of dominance, see for instance [ML14], [WG+13], [WD19].

4.3.3.3 Weak selective effects : r and s small as compared to γ.

For simplicity, consider the case where r is linear : $r(x) := r_1\, x$, for $r_1 > 0$. We then may think of the selective effect as type C individuals distributing a reproduction benefit of r_1/n to all the n individuals in their group. The specificity is however that the whole group is to be duplicated at once at the reproduction event at group level. Since we forbid any transmission between groups, one could expect a relatedness of 1 between any two individuals of the same group, and 0 between individuals of different groups. Under weak selection, this may lead to the prediction that the mutation is positively selected provided that $r_1 > s$ (in agreement with eq. 1 in [TN06], with a much larger number of groups than the size of each group, since only inter-group fluctuations are kept). Since in our model, we assumed that the random fluctuations in group reproduction are negligible, we cannot really approach the case of weak selection.

Such weak mutation assumption corresponds nonetheless also to $\gamma \to \infty$. So what happens in practice in our model in this limit ? Note first that this implies a separation in the time scales of fixation inside one group and among groups. Such time-scale separation is classically assumed in the context of weak selection, recall e.g. [O'F08]. In the time scale of interactions between groups, we can assume with almost no restriction that all the groups have fixed, so that only remains the competition between pure groups. Indeed, all the individual in the group are then strongly related (with a very close common ancestor). But the crucial parameter of interest for predicting the outcome is the initial proportion of pure C groups that can be established early on. If this is non-negligible, pure C groups will very probably prevail whatever the respective values of r and s are.

4.3.4 Evolutionary context.

One may however be interested in looking at an evolutionary scenario where mutations that are well-separated in time have each some probability to invade a population with a given type and replace the resident type. In this context, one has first to choose if C or D plays the role of the resident type. Let us first say it is D type. We thus consider the case where only one individual is of type C and compare the probability that an homogeneous population of type C emerges to the neutral case where that probability would be $1/(m \times n)$. This value is due to the fact that at some large time, every individual in the population will be issued from a unique common ancestor at generation 0, which is taken uniformly at random under neutrality. Note that in this limit, the invasion probability is respectively of order $1/n$ and $1/m$ for the fixation within the group and of the group among the population of groups.

4.3.4.1 Estimation of the probability of invasions at individual level.

Because there is this separation of time-scales, one can exploit classical results for estimating the probability of invasion in each of these cases. At least for the selection within group, we may exploit the explicit formula for probability of invasion starting from a proportion $x \in (0, 1)$ in the solution (X_t) of the limiting equation (4.1.6). The formula, first obtained by Malécot and often presented in the context of Kimura's diffusive approximation, takes the form :

$$\mathbb{P}_x(\tau_1 < \tau_0) = \frac{\exp[sx/\gamma] - 1}{\exp[s/\gamma] - 1}. \tag{4.3.2}$$

This can be obtained by identifying it as the only solution of $\mathcal{L}_{WF} u = 0$ with boundary conditions $u(x) = 0$ and $u(1) = 1$. This expression is then equivalent as x tends to

0 to $(s/\gamma)\times(e^{s/\gamma}-1)^{-1}\times x$. Although a precise justification would require a careful analysis of the process when the C sub-population is still negligible, we can expect that selective effects do not play a consequent role in this first step. We may thus expect a fixation probability starting from only one type C individuals to be well-approximated by

$$\pi^I_{D\mapsto C} = \frac{s/\gamma}{n\,(e^{s/\gamma}-1)} \approx (1/n)\times(1-\frac{s}{2\gamma}) \quad \text{with } s/\gamma \ll 1. \tag{4.3.3}$$

Note that except in this last step of approximation, we only exploited the approximation by X for n large and the separation of time-scale without restriction on s or γ. By changing s into $-s$, we would obtain the probability of invasion of C type residents by D type mutants.

4.3.4.2 Estimations of the probability of invasions under a weak selection assumption.

With a weak selection at both group and individual level, we need to keep a quantification of the fluctuations at group level with the parameter γ_G. The same reasoning can then be applied for the subsequent fixation of the pure C group in a resident population of pure D groups, leading to an overall invasion probability well-approximated by :

$$\begin{aligned}
\pi_{D\mapsto C} &= \pi^I_{D\mapsto C}\times\pi^G_{D\mapsto C} = \frac{s/\gamma}{n\,(e^{s/\gamma}-1)}\times\frac{(\rho_0-\rho_1)/\gamma_G}{m\,(e^{(\rho_0-\rho_1)/\gamma_G}-1)} \\
&= \frac{1}{n\times m}\times\frac{s/\gamma}{e^{s/\gamma}-1}\times\frac{r_1/\gamma_G}{1-e^{-r_1/\gamma_G}} \approx \frac{1}{n\times m}\times(1+(1/2)\times(r_1/\gamma_G-s/\gamma)), \\
\pi_{C\mapsto D} &= \frac{1}{n\times m}\times\frac{s/\gamma}{1-e^{-s/\gamma}}\times\frac{r_1/\gamma_G}{e^{r_1/\gamma_G}-1} \approx \frac{1}{n\times m}\times(1-(1/2)\times(r_1/\gamma_G-s/\gamma)),
\end{aligned} \tag{4.3.4}$$

where the last approximations assume $(r_1/\gamma_G)\vee(s/\gamma)\ll 1$. Assuming that mutations in both directions happen with the same law and intensity, well-separated and with small selective effects, we would then expect an evolutionary drift in the direction of cooperation provided $r_1/\gamma_G > s/\gamma$.

4.3.4.3 A specific criterion for the direction of selection.

We see that the above-mentioned prediction is not the one we find here, with an additional implication of the levels of fluctuations γ and γ_G. Note that the inverse of these quantities is usually referred to as the effective population size. This can be derived from Theorem 4.1.1 by recalling that $\gamma\approx\bar{\gamma}/n$ with $\bar{\gamma}$ the actual reproduction rate of the individuals. Assuming $\bar{\gamma}$ of order 1 then implies that γ scales as the inverse of the population size. Depending on the respective values of n and m, we thus do not see any reason for γ to be close to γ_G. So our condition $r_1/\gamma_G > s/\gamma$ seems to invalidate the naive one $r_1 > s$. Besides, assuming this separation of time-scales in a framework of weak-selection means essentially assuming $\gamma \gg \gamma_G$ because fixation is all the quicker if γ is large. So we see that even small selective effects at group level can effectively outcompete much larger selective effects at individual level.

4.3.4.4 Stronger selective effects : main differences.

This is something we observe also in our limiting results relying on QSD, and the separation of time-scale extends to the case where we only assume $\gamma \gg 1$ and not necessarily $\gamma_G \gg 1$.

Amplification of the invasion by C-type mutants. Close to our setting with γ_G very small, the selective effects at group level are not expected to play any significant role in the probability of fixation of the first mutant in his group. Recalling equation (4.3.3), this value only slightly deviates from neutrality when s/γ is small. On the other hand, since its selective advantage is strong against the other groups :

$$\pi^G_{D\mapsto C} \approx (1/m) \times (r_1/\gamma_G), \tag{4.3.5}$$

meaning that its probability of fixation is largely amplified.

Robustness to invasion by D-type mutants. The situation is clearly not symmetric, since on the other hand :

$$\pi^G_{C\mapsto D} \approx (1/m) \times (r_1/\gamma_G) \times e^{-r_1/\gamma_G} \ll (1/m). \tag{4.3.6}$$

Now we have the situation of a ratchet, where invasions by Defectors is completely negligible as compared to invasions by Cooperators. These latter events would then drive the dynamics of selection regarding strong selective effects.

This robustness shall extend even when γ is not so large. The situation is much more complicated to analyze in this case where γ is not so large, so that both selective effects are competing simultaneously. Nonetheless, in the limiting process we described, as soon as there is a non-negligible proportion of pure C groups in the population, D individuals simply cannot completely replace C type individuals. The complete invasion by D type individuals is impossible, even in the case where $0 < r_1 \ll s$ (nor $r_1 \ll (s/\gamma)$). With a large yet finite number m of groups, we expect that it would be possible to interpret such invasion as a large deviation result of the process ν (see section 1). Referring to classical literature on the subject of large deviations, notably Section 5 of [DZ98], the associated probability is thus likely to be exponentially small with increasing m. At least, this rate of decay is what we have obtained in (4.3.6), recalling $m = O(1/\gamma_G)$. So we conjecture that this strong resistance to invasion by Defectors is very general as long as γ_G is sufficiently small, that is m sufficiently large.

The invasion by C type mutants is not as strongly selected against when γ is not so large. When γ is sufficiently large to keep non-negligible the probability of invasion of its group by a mutant C type individual, we shall retrieve the ratchet effect : neglecting the effects of weak mutations, invasions by Cooperative individuals should drive the dynamics of selection towards more cooperation, while invasions by Defectors scarcely occur. We would have the same effect if mutations towards more cooperation were strong and simply more advantageous in an homogeneous population (without group structure). Globally, the dynamics is driven also by these weak mutations and the contribution of both weak and strong selection a priori depends on the specific situation of study.

Robustness to invasion by C-type mutants when γ is small. On the contrary, we also have a similar resistance when γ is sufficiently small, as noted in our Section 4.3.2. Even if $r_1 \gg s > 0$, the invasion of some group with mainly type D individuals by some type C individuals relies on so exceptional events that it seems biologically almost impossible. Now, we expect that it would be possible to interpret such invasion as a large deviation result of the process X, cf (4.1.6). In practice, we thus predict that the probability of such invasion shall decrease exponentially with $\gamma = O(1/n)$.

The case where r_1 is of order s/γ is possibly more intricate and would require further consideration. It might lead to a specific optimization problem as described in [CH19], where the cost of deviating X has to be balanced with the amplification through r. The shape of the function r would then play a much more significant role.

A general robustness to invasion when γ is small. Considering both directions of invasion, it seems that strong selective effects are strongly selected against in very large populations, whatever the level of selection they favor, as long as they are detrimental for one level of selection. Indeed, both invasion probabilities scale in the exponentially with the population sizes (in the detrimental level of selection), which is much more stringent than the order $O(1/(n \times m))$ of nearly neutral mutations. We expect it to extend with possibly more levels of "selection". Thus, beside the effect of mutations favoring its carriers at both levels (but not necessarily equivalently), the trade-off between selective effects at different levels shall be driven mostly by weak selective effects.

Note that these conclusions seem quite robust to more general forms of functions r, provided $\rho_1 < \rho_0 < \rho_\alpha$ with a QSD α very concentrated around 1. Based on our first simulation results, this effect of concentration seems to be robust as long as r is increasing. This suggests also that it is not so crucial that 1 is an absorbing state.

Thus, the above conclusions shall be maintained even with some transmission between groups, as long as their rate is sufficiently small.

4.3.5 A similar robustness for polymorphic QSD.

Similarly, the fixation of a polymorphic population with profile α such that $\rho_\alpha < \rho_0 \wedge \rho_1$ is likely to be an exceptional event as compared to the time-scale at which the transitory profile evolves. We mean that the profile of $\mu^{m,n}$ shall remain very close to α, with a much quicker regulation of the random perturbations when m is large. Our confidence originates again from the comparison with Large Deviations results. Even the events of fixation, that are possibly much less negligible for finite m, are not expected to be significant. The pure groups have a lower progeny and do not contribute much to the dynamics on the long term.

Simulations of the dynamics of μ_t are provided in the Appendix of the manuscript in Figures 6.28-6.30.

By construction, this case corresponds to r being a function with at least a strict maximum inside $(0, 1)$. Selective effects must favor polymorphism directly and not only conflict with selective effects at individual level. Otherwise, $\rho_\alpha < \rho_1$ is excluded by Lemma 4.2.0.2. We could also imagine more general selective effects at individual level with a frequency-dependency. This would possibly also entail a stable polymorphic QSD.

Note that we also assume here that γ is not negligible. Again, referring to [CH19], it is possible that the description gets much more tricky in the case where γ is small but s also so that the function r scales as s/γ.

4.3.5.1 Difficulties in relating to kin selection.

Considering more general r, notably with a maximal value in the interior $(0, 1)$, the approach of kin selection becomes even less clear. It seems required to deal with another definition of relatedness, like the one given in [Gr06], with much more complexity.

4.3.6 Main conclusion of Subsection 4.3

— In any case where δ_0 is stable, for an initial condition with enough highly cooperative groups so that the diffusion in them generates a non-negligible proportion of pure C groups, these groups do eventually invade the population with a high probability.

— If intra-group fluctuations of population size are small, a very large proportion of the groups become visibly increasingly dominated by D individuals, even though the pure D groups are the worst at reproducing. Introducing in the model rare mutations from individuals D to individuals C probably wouldn't make much of a difference.

- The case $\rho_1 < \rho_\alpha < \rho_0$ characterizes the fact that a subpopulation of polymorphic groups is capable of maintaining itself better than pure D groups. This means that domination by D individuals would still be prevented, even if we disrupt the model by introducing rare events of migration between groups or mutations from C individuals to D individuals. This should hold true as long as the polymorphic QSD α is rapidly approached.
- Likewise, in the case of $\rho_\alpha < \rho_1 < \rho_0$, the polymorphic state is a priori very stable and μ_t tends to him provided that the proportion of polymorphic groups was not violently reduced at first.
- However, it is unclear that the area of attraction of this quasi-stationary distribution is far beyond the vicinity of the purely cooperative state. Convergence towards it and its influence has potentially no significant effect if the proportion of the groups with a majority of C individuals is too small. A transient attractor is then likely to appear for μ_t restricted to $(0, 1)$, widely supported on a neighborhood of the purely D state. It can be interpreted in practice as an alternative quasi-stationary distribution, with an extinction rate generally higher than ρ_0. This happens when the transition from a group dominated by D individuals to one with mostly C type individuals is too costly as a deviation from the process X for its probability to be quickly compensated for by the difference in extinction rates. Transitions in the vicinity of the less stable "pseudo"-QSD to the "real" QSD (more stable and around 1) take a non-negligible time so that the attraction to the pseudo-QSD is clearly visible.
- In an evolutionary perspective, for very strong selection effects as compared to genetic fluctuations, the status quo situation is more likely to prevail : the probability of fixing a mutation which puts a burden on its holder at some selection level is only slightly compensated by the advantage that this mutation could bring at another level.
- Still from an evolutionary point of view, cooperation is favored when increasing the level of intra-group genetic fluctuations (i.e. kinship between individuals in the same group). Mutations that put the group of those who carry them at a disadvantage have a much lower probability of fixation compared to the mutations that put those who carry them at a disadvantage inside their group : the relation of comparison between the strength of these effects in the weak selection context seems to lose its relevance here when the effects combine.

4.4 Proof of the results of Section 4.2

4.4.1 Proof of Proposition 4.2.0.1 : characterization of α on $(0,1)$

We rely on the method used in Chapter 1 and more precisely on the proof of the second illustration presented in Subsection 1.4.2 of Chapter 1 to ensure that :

$$\exists\,\chi > 0, \quad \forall\, n \geq 1, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi}, \quad \left\| \mu A^{01}_t - \alpha \right\|_{TV} \leq C_{n,\xi} \exp[-\chi\, t]. \tag{4.4.1}$$

The diffusion is indeed regular on any $\mathcal{D}_n := [1/n, 1-1/n]$ (for $n \geq 3$) so that applying the Harnack inequality, we prove similarly as in Chapter 1 that for any choice of $0 < t_M < t_c$, there exists $c_M > 0$ such that for any $x \in \mathcal{D}_n$:

$$\mathbb{P}_x \left(X_{t_M} \in dx \ ;\ t_M < \tau_\partial^{n+1} \right) \geq c_M \, \mathbb{P}_{1/2} \left(X_{t_c} \in dx \,\middle|\, t_c < \tau_\partial^3 \right) := c_M \, \zeta(dx), \tag{4.4.2}$$
$$\text{with } \tau_\partial^n := \inf \left\{ t > 0 \,\middle|\, X_t \notin \mathcal{D}_n \right\}.$$

We refer to the step 4 of the proof given in Sect. 4 of [CV17b] to ensure that for any $n \geq 3$ and $t > 0$, there exists $c_n > 0$ such that :

$$\forall\, x, y \in \mathcal{D}_n, \quad \mathbb{P}_x \left(X_t \in dx \ ;\ t < \tau_{0,1,\partial} \right) \leq c_n \, \mathbb{P}_y \left(X_t \in dx \ ;\ t < \tau_{0,1,\partial} \right). \tag{4.4.3}$$

Next, we prove that the process X cannot maintain itself close to the boundary :

Lemma 4.4.1.1. *For any $\rho > 0$, there exists $E = \mathcal{D}_{n_E}$ such that :*

$$\sup_{x \in (0,1)} \mathbb{E}_x \exp[\rho V_E] < \infty \qquad \text{where } V_E := \tau_{0,1,\partial} \wedge \inf \left\{ t > 0 \ ;\ X_t \in E \right\}. \tag{4.4.4}$$

Applying Theorem 1.2.1 in Chapter 1 with (4.4.2), (4.4.3) and (4.4.4), noting also that condition (A0) on $\{\mathcal{D}_n\}$ is clearly satisfied, concludes the proof of (4.4.1). The results on the capacity of survival comes from Theorem 1.2.2.

To end the proof of Proposition 4.2.0.1, it is sufficient to ensure the following lemma

Lemma 4.4.1.2. *There exists $n_B \geq 3$, $\xi_B, t_B > 0$ such that :*

$$\forall\, \mu \in \mathcal{M}_1[(0,1)], \quad \mu A^{01}_{t_B} \in \mathcal{M}_{n_B, \xi_B}.$$

This can be done exactly as in step 1, Section 5.1 of [CV18a], by handling precisely with the vicinities of 0 and of 1.

Proof of Lemma 4.4.1.1 The core of the proof is the well-known fact that for any $t > 0$,

$$\mathbb{P}_x(t < \tau_{0,1}) \to 0 \text{ as } x \to 0\,, \quad \mathbb{P}_x(t < \tau_{0,1}) \to 0 \text{ as } x \to 1 \tag{4.4.5}$$

(see notably Theorem 3.4 and 3.7 in [CV18a] for a much more precise estimate of the extinction on the boundaries).

Let $\rho > 0$. We fix then arbitrarily $t = 1$ and deduce from (4.4.5) that for n_E sufficiently large :

$$\forall\, x \in (0,1), \quad \mathbb{P}_x(t < V_E) \leq e^{-\rho t}/2.$$

By induction on $k \geq 1$ with the Markov property, we deduce that for any k :

$$\forall\, x \in (0,1), \quad \mathbb{P}_x(k\, t < V_E) \leq e^{-k\rho t}/2^k.$$

We know conclude the proof of Lemma 4.4.1.1 by noting :

$$\begin{aligned}\mathbb{E}_x(\exp[\rho V_E]) &\leq \textstyle\sum_{\{k\geq 0\}} e^{\rho t\,[k+1]}\mathbb{P}_x(V_E \in [k\,t, (k+1)\,t))\\ &\leq e^{\rho t}\textstyle\sum_{\{k\geq 0\}} 2^{-k} = 2e^{\rho t} < \infty.\end{aligned}$$

□

4.4.2 Proof of Proposition 4.2.1.1 : convergence to δ_1 for $\rho_1 < \rho_0 < \rho_\alpha$

For $\mu \in \mathcal{M}_{n,\xi}$, we have the following lower-bound of the mass absorbed at 1 before time 1 :

$$\mu P_1\{1\} = \mathbb{P}_\mu(\tau_1 \leq 1 \leq \tau_{0,\partial}) \geq \xi\, \mathbb{P}_{1/n}(\tau_1 \leq 1)\, \exp[-\|r\|_\infty]$$

By the Markov property, this implies with $C_{n,\xi} := \xi\, \mathbb{P}_{1/n}(\tau_1 \leq 1)\, \exp[-(\|r\|_\infty - \rho_1)]$:

$$\mu P_t\{1\} = \mu P_1\{1\}\exp[-\rho_1(t-1)] \geq C_{n,\xi}\, \exp[-\rho_1\, t]. \tag{4.4.6}$$

Since $\tau_0 \leq \tau_{0,1,\partial}$ and the extinction rate is ρ_0 once X has reached 0, then exploiting (4.2.3) :

$$\begin{aligned}\mathbb{P}_\mu(\tau_0 \leq t < \tau_\partial) &\leq \mathbb{E}_\mu\left[\exp[-\rho_0\,(t - \tau_{0,1,\partial})]\ ;\ \tau_{0,1,\partial} \leq t\right]\\ &\leq \exp[-\rho_0\, t]\, \left[1 + \rho_0 \textstyle\int_0^t ds\, \exp[\rho_0\, s]\mathbb{P}_\mu(s < \tau_{0,1,\partial})\right]\\ &\leq \exp[-\rho_0\, t]\, \left[1 + \|\bar{h}\|_\infty \times \dfrac{\rho_0}{\rho_\alpha - \rho_0}\right].\end{aligned} \tag{4.4.7}$$

With again (4.2.3), and (4.4.6) and the fact that both μA_t and δ_1 are probability measure :

$$\begin{aligned}\|\mu A_t - \delta_1\|_{TV} &= \sup_D |\mu A_t(D) - \delta_1(D)| = \mu A_t[0,1) = \mu P_t[0,1)/(\mu P_t[0,1) + \mu P_t\{1\}) \\ &\leq \frac{\mathbb{P}_\mu(\tau_0 \leq t < \tau_\partial) + \mathbb{P}_\mu(t < \tau_{0,1,\partial})}{\mu P_t\{1\}} \leq C'_{n,\xi} \exp[-(\rho_0 - \rho_1)\, t] \\ &\text{where } C'_{n,\xi} := \left[1 + \|\bar{h}\|_\infty \times \rho_\alpha/(\rho_\alpha - \rho_0)\right]/C_{n,\xi}\end{aligned}$$

□

4.4.3 Proof of Proposition 4.2.1.2 : conditional convergence to δ_0 for $\rho_0 < \rho_\alpha$

Let $t \geq 1$ and assume first that $\mu([0,x]) \geq \xi$ for $x \in (0,1)$ and $\xi > 0$. From (4.2.3),

$$\mathbb{P}_\mu(t < \tau_{0,1,\partial}) \leq \|\bar{h}\|_\infty \exp[-\rho_\alpha\, t]$$

With the rough lower-bound $\mu\, P_1^1\{0\} \geq \exp[-\|r\|_\infty]\, \mathbb{P}_\mu(\tau_0 \leq 1)$:

$$\begin{aligned}\mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial}) &\geq \exp[-\|r\|_\infty]\, \mathbb{P}_\mu(\tau_0 \leq 1) \times \exp[-\rho_0\,(t-1)] \\ &\text{with } C := \frac{\|\bar{h}\|_\infty \exp[\|r\|_\infty - \rho_0]}{\xi\, \mathbb{P}_x(\tau_0 \leq 1)} > 0, \\ \left\|\mu A_t^1 - \delta_0\right\|_{TV} = \mu A_t^1(0,1) &= \frac{\mathbb{P}_\mu(t < \tau_{0,1,\partial})}{\mathbb{P}_\mu(t < \tau_{0,1,\partial}) + \mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial})} \\ &\leq C \exp[-(\rho_\alpha - \rho_0)\, t]. \qquad (4.4.8)\end{aligned}$$

The case where μ has support on $\{0,1\}$ is trivial, since then $\mu A_t^1 = \delta_0$.

Finally, for the general case of $\mu \in \mathcal{M}_1([0,1]) \setminus \{\delta_1\}$, where $\mu(0,1) > 0$, remark that, for any $s > 0$, there exists $m_s \in (0,1)$ such that :

$$\begin{aligned}&\mu A_s^1 = m_s\, \mu A_s^{01} + (1 - m_s)\, \delta_0 \\ &\text{where for any } x > 0\,, \qquad \mu A_s^{01}([0,x]) \underset{s\to\infty}{\longrightarrow} \alpha([0,x]) > 0.\end{aligned}$$

by Proposition 4.2.0.1, with the rate of convergence uniform over μ. Thus, we deduce some $t_\vee > 0$ such that, with $x = 1/2$:

$$\forall\, \mu \in \mathcal{M}_1([0,1]) \setminus \{\delta_1\}, \quad \mu\, A_{t_\vee}^1([0,x]) \geq \mu A_{t_\vee}^{01}([0,x]) \geq \alpha([0,x])/2 := \xi.$$

Thus, for any $t \geq t_\vee$, by the Markov property, then (4.4.8) with the initial condition $\mu A_{t_\vee}^{01}$:

$$\left\|\mu A_t^1 - \delta_0\right\|_{TV} = \left\|\mu A_{t_\vee}^1\, A_{t-t_\vee}^1 - \delta_0\right\|_{TV} \leq C \exp[(\rho_\alpha - \rho_0)\, t_\vee] \exp[-(\rho_\alpha - \rho_0)\, t].$$

□

4.4.4 Proof of Proposition 4.2.2.1 and 4.2.2.2 : the case $\rho_1 < \rho_\alpha < \rho_0$

For this Proposition, we need to adapt the proof given in Subsection 1.5.3 of Chapter 1. The main step is to prove that the mass on the interval $(0,1)$ does not vanish :

Lemma 4.4.4.1. *Assume that $\rho_\alpha < \rho_0$. Then, there exists $n_\circ \geq 2, \xi_\circ > 0$ such that :*

$$\begin{aligned}&\forall\, n \in \mathbb{N},\ \forall\, \xi > 0,\ \exists\, t_\circ > 0,\\ &\forall\, \mu \in \mathcal{M}_{n,\xi},\ \forall\, t \geq t_\circ, \qquad \mu A^1_t(1/n_\circ,\, 1 - 1/n_\circ) \geq \xi_\circ.\end{aligned}$$

We also need to ensure the persistence of the component issued from the coupling, which is done with the following lemma.

Lemma 4.4.4.2. *Assume that $\rho_\alpha < \rho_0$ and $\zeta \in \mathcal{M}_1[(0,1)]$. Then, there exists $t_P, c_P > 0$ such that :*

$$\forall\, x \in [0,1),\ \forall\, t \geq t_P, \quad \mathbb{P}_x(t < \tau_{1,\partial}) \leq c_P\, \mathbb{P}_\zeta(t < \tau_{1,\partial}). \tag{4.4.9}$$

The measure ζ comes from a mixing estimate that we recall –cf (4.4.2) :

Lemma 4.4.4.3. *Let $n \geq 2$ and $\xi > 0$. Then, there exists $\zeta \in \mathcal{M}_1[(0,1)]$, $t_M, c_M > 0$ such that for any $\mu \in \mathcal{M}_1([0,1])$ satisfying $\mu(1/n,\, 1 - 1/n) \geq \xi$:*

$$\mu A^1_{t_M}(dx) \geq c_M\, \zeta(dx).$$

4.4.4.1 Proof of Proposition 4.2.2.1 :

The proof of Proposition 4.2.2.1 follows as the one of Proposition 4.2.1.2 from (4.2.10), which is a consequence of Proposition 4.2.2.2. □

4.4.4.2 Proof of Proposition 4.2.2.2 with Lemmas 4.4.4.1-3 :

Combining these tree lemmas and applying exactly the same reasoning as in Subsection 1.5.3 of Chapter 1 proves that there exists a unique QSD α_1 associated to $\tau_{1,\partial}$, with the convergence stated in Proposition 4.2.2.2.

Moreover, as stated in Proposition 4.2.2.2, we can identify α_1 and h^1.

Let $\alpha_1^y := y\,\alpha + (1-y)\,\delta_0$. For any $t \geq 0$:

$$\alpha_1^y P^1_t(dx) = y\, \exp[-\rho_\alpha\, t]\, \alpha(dx) + [(1-y)\, \exp[-\rho_0\, t] + y\, \mathbb{P}_\alpha(\tau_0 \leq t < \tau_{1,\partial})]\, \delta_0(dx). \tag{4.4.10}$$

It follows from Theorem 2.6 in [CMS13] that the exit state is independent from the exit time when the initial condition is a QSD, with an exponential law for the exit time. Thus :

$$\begin{aligned}\mathbb{P}_\alpha(\tau_0 \leq t < \tau_{1,\partial}) &= \mathbb{E}_\alpha\left[\exp[-\rho_0\,(t-\tau_0)]\ ;\ \tau_0 = \tau_{0,1,\partial} \leq t\right]\\ &= \mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right)\,\mathbb{E}_\alpha\left[\exp[-\rho_0\,(t-\tau_{0,1,\partial})]\ ;\ \tau_{0,1,\partial} \leq t\right]\\ &= \mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right)\,\textstyle\int_0^t \exp[-\rho_0\,(t-s)] \times \rho_\alpha\,\exp[-\rho_\alpha\,s]\,ds\\ &= \left(\exp[-\rho_\alpha\,t] - \exp[-\rho_0\,t]\right)\frac{\rho_\alpha\,\mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right)}{\rho_0 - \rho_\alpha}.\end{aligned} \tag{4.4.11}$$

With our choice (4.2.8), i.e. $\frac{1-y_\alpha}{y_\alpha} = \frac{\rho_\alpha\,\mathbb{P}_\alpha\left(\tau_0=\tau_{0,1,\partial}\right)}{\rho_0-\rho_\alpha}$, we see that we obtain indeed :

$$\alpha_1^{y_\alpha}\,P_t^1 = \exp[-\rho_\alpha\,t]\,\alpha_1^{y_\alpha}.$$

The proof that h^1 is uniquely defined, the convergences in (4.2.7) and (4.2.9) and the upper-bound in (4.2.10) are exactly the same as in Chapter 1. It remains to identify h^1. Clearly :

$$h^1(0) := \lim_{t\to\infty} \exp[\rho_\alpha\,t]\,\mathbb{P}_0(t < \tau_{1,\partial}) = \lim_{t\to\infty} \exp[-(\rho_0 - \rho_\alpha)\,t] = 0. \tag{4.4.12}$$

We define h_t^1 similarly as h_t (cf (4.2.1)) : for $x \in [0,1)$

$$h_t^1(x) := \exp[\rho_\alpha\,t]\,\mathbb{P}_x(t < \tau_{1,\partial}). \tag{4.4.13}$$

Decomposing according to the state of X at time 1, and recalling that since X stays at 0 once it is hit, we have :

$$h^1_{2\,t}(x) = h_t(x)\,\langle \delta_x\,A_t^{01}\,\big|\,h_t^1\rangle + h_t^1(x)\,\delta_x\,A_t^1\{0\}\,h_t^1(0) \tag{4.4.14}$$

From (4.4.12) and (4.2.10), the second term in the right-hand side is clearly negligible. Because of (4.2.2), we are interested in the asymptotic as $t \to \infty$ of $\langle \alpha\,\big|\,h_t^1\rangle$. Since α is the QSD associated to the extinction rate ρ_α with extinction at time $\tau_{0,1,\partial}$, we already know that $\mathbb{P}_\alpha(t < \tau_{0,1,\partial}) = \exp[-\rho_\alpha t]$. We will deal with the component that has fixed at 0 before time t thanks to (4.4.11). Concluding with (4.2.8), we mean that :

$$\begin{aligned}\langle \alpha\,\big|\,h_t^1\rangle &= \exp[\rho_\alpha\,t]\,\mathbb{P}_\alpha(t < \tau_{0,1,\partial}) + \exp[\rho_\alpha\,t]\,\mathbb{P}_\alpha(\tau_0 \leq t < \tau_{1,\partial})\\ &= 1 + \left(1 - \exp[-(\rho_0 - \rho_\alpha)\,t]\right)\frac{\rho_\alpha\,\mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right)}{\rho_0 - \rho_\alpha}\\ &\underset{t\to\infty}{\longrightarrow} 1 + (1-y_\alpha)/y_\alpha = 1/y_\alpha.\end{aligned} \tag{4.4.15}$$

From (4.4.14), (4.4.15), (4.2.2), and (4.2.10), we conclude :

$$h^1_{2\,t}(x) \underset{t\to\infty}{\longrightarrow} h(x)/y_\alpha = h^1(x). \qquad \square$$

4.4.4.3 Proof of Lemma 4.4.4.2

: From (4.2.9), with the notation (4.4.13) :

$$\mathbb{P}_\zeta(t < \tau_{1,\partial}) = \langle \zeta \,\big|\, h^1_t \rangle \; \exp[-\rho_\alpha\, t] \quad \text{where } \langle \zeta \,\big|\, h^1_t \rangle \xrightarrow[t\to\infty]{} \langle \zeta \,\big|\, h \rangle / y_\alpha.$$

We know from Proposition 4.2.0.1 that h is lower-bounded by a positive constant on any $\mathcal{D}_n$ for $n \geq 2$. It implies in particular $\langle \zeta \,\big|\, h \rangle > 0$. From (4.2.2), let thus $t_P > 0$ be such that :

$$\forall\, t \geq t_P, \quad \langle \zeta \,\big|\, h^1_t \rangle \geq \langle \zeta \,\big|\, h \rangle / 2 > 0. \tag{4.4.16}$$

(4.4.12) is clearly true and implies with (4.2.3) that (4.4.9) holds for $x = 0$.

For $x \in (0,1)$ and any $t > 0$:

$$\begin{aligned}
\mathbb{P}_x(\tau_0 \leq t < \tau_{1,\partial}) &= \mathbb{E}_x\left[\exp[-\rho_0\,(t - \tau_0)]\;;\; \tau_0 = \tau_{0,1,\partial} \leq t\right] \\
&\leq \mathbb{E}_x\left[\exp[-\rho_0\,(t - \tau_{0,1,\partial})]\;;\; \tau_{0,1,\partial} \leq t\right] \\
&= \exp[-\rho_0\, t]\, \left(1 + \rho_0 \textstyle\int_0^t \exp[\rho_0\, s] \times \mathbb{P}_x(s \leq \tau_{0,1,\partial} \leq t)\, ds\right) \\
&\leq \exp[-\rho_0\, t]\, \left(1 + \rho_0\, \|\bar{h}\|_\infty \textstyle\int_0^t \exp[(\rho_0 - \rho_\alpha)\, s] ds\right) \\
&\leq \exp[-\rho_0\, t] + \frac{\rho_0\, \|\bar{h}\|_\infty}{\rho_0 - \rho_\alpha} \exp[-\rho_\alpha\, t].
\end{aligned} \tag{4.4.17}$$

Combining (4.4.16), (4.4.17) and (4.2.3) ends the proof of Lemma 4.4.4.2. $\square$

4.4.4.4 Proof of Lemma 4.4.4.1

: Let $n_\circ \geq 3$ such that :

$$\alpha(1/n_\circ\,,\, 1 - 1/n_\circ) \geq 1/2 \tag{4.4.18}$$

From (4.2.2), we can find $t_S > 0$ such that for any μ with $\mu(0,1) > 0$:

$$\begin{gathered}
\forall\, t \geq t_S, \quad \mathbb{P}_\mu(t < \tau_{0,1,\partial}) \geq \langle \mu \,\big|\, h \rangle / 2 \times \exp[-\rho_\alpha\, t], \\
\left\| \mu A^{01}_t - \alpha \right\|_{TV} \leq 1/4 \quad \text{thus} \quad \mu A^{01}_t(1/n_\circ\,,\, 1 - 1/n_\circ) \geq 1/4.
\end{gathered} \tag{4.4.19}$$

Since 0 is absorbing and by (4.4.19) :

$$\begin{aligned} &\forall\, t \geq 0, \quad \mu A^1_t(dx) = \mu A^1_t(0,1) \times \mu A^{01}_t(dx) + \Big[1 - \mu A^1_t(0,1)\Big]\, \delta_0(dx), \\ &\qquad \forall\, t \geq t_S, \quad \mu A^1_t(1/n_\circ\,,\, 1 - 1/n_\circ) \geq \mu A^1_t(0,1)/4 \quad \text{where} \;: \end{aligned} \tag{4.4.20}$$

$$\mu A^1_t(0,1) = \frac{\mathbb{P}_\mu(t < \tau_{0,1,\partial})}{\mathbb{P}_\mu(t < \tau_{0,1,\partial}) + \mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial})} = \left(1 + \frac{\mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial})}{\mathbb{P}_\mu(t < \tau_{0,1,\partial})}\right)^{-1}. \tag{4.4.21}$$

Assume first that $\mu[1/n, 1 - 1/n] \geq \xi$ for some $n \geq 3$ and $\xi > 0$. Since h is positive on $(0,1)$, this implies, with (4.2.2), (4.4.17), (4.4.20) and (4.4.21), a lower-bound $\xi_\circ$ that only depends on n and ξ such that :

$$\forall\, t \geq t_S, \quad \mu A^1_t(1/n_\circ\,,\, 1 - 1/n_\circ) \geq \xi_\circ. \tag{4.4.22}$$

Lemma 4.4.1.2 completes the proof. Indeed, consider $\mu \in \mathcal{M}_{n,\xi}$ (w.l.o.g. $\mu\{1\} = 0$ since it vanishes immediately).
Either $\mu(1/n, x_\vee) \geq \xi/2$ and we deduce the result from (4.4.22),
or $\mu[x_\vee, 1) \geq \xi/2$ and we deduce from Lemma 4.4.1.2 and (4.4.22) :

$$\forall\, t \geq t_S + t_B, \quad \mu A^1_t(1/n_\circ\,,\, 1 - 1/n_\circ) = [\mu A^1_{t_B}] A^1_{t - t_B}(1/n_\circ\,,\, 1 - 1/n_\circ) \geq \xi'_\circ. \quad \square$$

4.4.5 Proof of Proposition 4.2.3.1 : the case $\rho_1 < \rho_0 = \rho_\alpha$

The calculations leading to (4.4.17) gives for the case $\rho_0 = \rho_\alpha$:

$$\forall\, \mu \in \mathcal{M}_1([0,1]), \quad \mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial}) \leq \exp[-\rho_0\, t]\, \Big(1 + \rho_0\, \|\bar{h}\|_\infty\, t\Big)\,. \tag{4.4.23}$$

With (4.4.23) instead of (4.4.7), like in the proof of Proposition 4.2.1.1 (i.e. with (4.2.3) and (4.4.6)), we deduce Proposition 4.2.3.1. $\square$

4.4.6 Proof of Proposition 4.2.3.2 : conditional convergence to δ_0 when $\rho_0 = \rho_\alpha$

Thanks to Proposition 4.2.0.1, there is for any $n \geq 2$ a positive lower-bound of h in $\mathcal{D}_n$, $\langle \mu \,\big|\, h \rangle$ is uniformly lower-bounded for $\mu \in \mathcal{M}^{0,1}_{n,\xi}$ (for any $n \geq 3, \xi > 0$). By (4.2.2), for t sufficiently large and any $\mu \in \mathcal{M}^{01}_{n,\xi}$:

$$\mathbb{P}_\mu(t < \tau_{0,1,\partial}) \geq c_{n,\xi}\, \exp[-\rho_0\, t].$$

Combining this with (4.4.21) and (4.4.23) concludes the proof that for t sufficiently large :

$$\left\|\mu A_t^1 - \delta_0\right\|_{TV} \geq C_{n,\xi}/t. \qquad \square$$

Remark : Adapting the proof of step 1, Section 5.1 of [CV18a], one can prove that there exists $t_B, c_B, n_B > 0$ such that for any $x \in [1/2, 1)$, $\delta_x A^1_{t_B} \in [1/4, 1 - 1/n_B)$. Exploiting the above proof of Proposition 4.2.3.2, this implies that the convergence is uniform for any μ such that $\mu[1/n, 1) \geq \xi$.

For the reverse inequality, assume first that $\mu \in \mathcal{M}^{01}_{n,\xi}$. From the Markov property

$$\begin{aligned}
\mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial}) &= \mathbb{E}_\mu\left[\exp[-\rho_0\,(t-\tau_0)]\ ;\ \tau_0 = \tau_{0,1,\partial} \leq t\right] \\
&= \mathbb{E}_\mu\left[\exp[-\rho_0\, t]\,\left(1 + \rho_0 \textstyle\int_0^{\tau_{0,1,\partial}} \exp[\rho_0\, s]\, ds\right)\ ;\ \tau_0 = \tau_{0,1,\partial} \leq t\right] \\
&\geq \rho_0\ \exp[-\rho_0\, t]\int_0^t \exp[\rho_0\, s]\,\mathbb{P}_\mu(\tau_0 = \tau_{0,1,\partial} \in [s,t])\, ds, \\
&\geq c_{n,\xi}\exp[-\rho_0\, t]\int_0^t \mathbb{P}_{\mu A^{01}_s}(\tau_0 = \tau_{0,1,\partial} \leq t-s)\, ds,
\end{aligned} \tag{4.4.24}$$

where we exploited once more Proposition 4.2.0.1 in the last inequality, to obtain a uniform lower-bound $c_{n,\xi}$ on $\langle \mu \,\big|\, h_s\rangle$.

Since $\mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial}) > 0$ and by monotone convergence, there exists $t_\vee > 0$ such that :

$$\forall\, t \geq t_\vee, \quad \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial} \leq t) \geq \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial} \leq t_\vee) := m_0 > 0. \tag{4.4.25}$$

Now, according to (4.2.2), we choose $t_S > 0$ such that :

$$\begin{aligned}
&\forall\,\mu \in \mathcal{M}_1[(0,1)],\ \forall\, s \geq t_S, \quad \left\|\mu A^{01}_s - \alpha\right\|_{TV} \leq m_0/2 \\
\text{which implies}\quad &\forall\, s \geq t_S,\ \forall\, t-s \geq t_\vee, \quad \mathbb{P}_{\mu A^{01}_s}(\tau_0 = \tau_{0,1,\partial} \leq t-s) \geq m_0/2.
\end{aligned} \tag{4.4.26}$$

Thus, (4.4.26) and (4.4.24) imply that for any $t \geq t_S + t_\vee$:

$$\mathbb{P}_\mu(\tau_0 \leq t < \tau_{1,\partial}) \geq c'_{n,\xi}\exp(-\rho_0\, t) \times (t - t_S - t_\vee).$$

With (4.2.3) and (4.4.21), this concludes the proof that :

$$\mu[1/n, 1-1/n] \geq \xi \quad \Rightarrow \quad \forall\, t \geq t_S + t_\vee, \quad \left\|\mu A_t^1 - \delta_0\right\|_{TV} \leq C_{n,\xi}/t. \tag{4.4.27}$$

Now, we prove that such upper-bound is in fact uniform with respect to $\mathcal{M}_1[(0,1)]$

thanks to Lemma 4.4.1.2. Indeed

$$\mu A^1_{t_B}(dx) = \mu A^1_{t_B}(0,1) \times \mu A^{01}_{t_B}(dx) + \left[1 - \mu A^1_{t_B}(0,1)\right] \delta_0(dx),$$
$$\text{where } \exists \xi_B > 0,\ \exists n_B \geq 2, \quad \forall \mu \in \mathcal{M}_1[(0,1)], \quad \mu A^{01}_{t_B}(1/n_B, 1-1/n_B) \geq \xi_B$$
$$\text{Thus by (4.4.27) :} \qquad \forall t \geq t_S + t_\vee, \quad \left\| [\mu A^{01}_{t_B}] A^1_t - \delta_0 \right\|_{TV} \leq c_{n_B, \xi_B}/t. \tag{4.4.28}$$

We also note that there exists $y_t \in (0,1)$ such that :

$$\mu A^1_{t_B + t}(dx) = y_t\, [\mu A^{01}_{t_B}] A^1_t + (1 - y_t)\delta_0.$$

In fact, our comparison of the survival from 0 and from μ gives us a uniform upper-bound $C > 0$ such that :

$$y_t = \mu A^1_t(0,1) \times \frac{\left\langle \mu A^{01}_{t_B} \,\middle|\, \mathbb{P}_.(t < \tau_{1,\partial}) \right\rangle}{\left\langle \mu A^1_{t_B} \,\middle|\, \mathbb{P}_.(t < \tau_{1,\partial}) \right\rangle} \quad \leq C\, \mu A^1_t(0,1)$$

Hence, we have more precision on the convergence :

$$\left\| \mu A^1_{t+t_B} - \delta_0 \right\|_{TV} \leq \mu A^1_{t_B}(0,1) \times C/t.$$

And at least, (4.4.28) concludes the proof of Proposition 4.2.3.2 (where $t_S + t_\vee$ replaces $t_\vee$). □

4.4.7 Proof of Proposition 4.2.4.1 : the case $\rho_0 = \rho_1 < \rho_\alpha$

Since $\rho_0 = \rho_1$, it is straightforward that any convex combination of δ_0 and δ_1 is a QSD, with extinction rate ρ_1.

It is then not difficult to adapt the proof of Proposition 4.2.1.2, and since $\mathbb{P}_\mu(\tau_{0,1} \leq 1)$ is lower-bounded uniformly over any $\mu \in \mathcal{M}_1([0,1])$, we obtain

$$\forall \mu \in \mathcal{M}_1([0,1]), \quad \mu A_t(0,1) \leq C\, \exp[-(\rho_\alpha - \rho_0)\, t].$$

$$\begin{aligned}
\mu A_t\{0\} &= \frac{\mathbb{E}_\mu\left[\exp[-\rho_1\,(t - \tau_{0,1,\partial})]\ ;\ \tau_{0,1,\partial} = \tau_0 \leq t\right]}{\mathbb{P}_\mu(t < \tau_{0,1,\partial}) + \mathbb{E}_\mu\left[\exp(-\rho_1\,(t - \tau_{0,1,\partial}))\ ;\ \tau_{0,1,\partial} = \tau_{0,1} \leq t\right]} \\
&= \frac{\mathbb{E}_\mu\left[\exp[\rho_1\, \tau_{0,1,\partial}]\ ;\ \tau_{0,1,\partial} = \tau_0 \leq t\right]}{\mathbb{E}_\mu\left[\exp[\rho_1\, \tau_{0,1,\partial}]\ ;\ \tau_{0,1,\partial} = \tau_{0,1} \leq t\right]} \times \left(1 + \frac{\exp[\rho_1\, t]\, \mathbb{P}_\mu(t < \tau_{0,1,\partial})}{\mathbb{E}_\mu\left[\exp(\rho_1\, \tau_{0,1,\partial})\ ;\ \tau_{0,1,\partial} = \tau_{0,1} \leq t\right]}\right)^{-1}.
\end{aligned} \tag{4.4.29}$$

The limit as $t \to \infty$ is well-defined and the convergence occurs at exponential rate since :

$$\begin{aligned}
0 &\leq \mathbb{E}_\mu\left[\exp[\rho_1\, \tau_{0,1,\partial}]\ ;\ \tau_{0,1,\partial} = \tau_1\right] - \mathbb{E}_\mu\left[\exp[\rho_1\, \tau_{0,1,\partial}]\ ;\ \tau_{0,1,\partial} = \tau_1 \leq t\right] \\
&\leq \mathbb{E}_\mu\left[\exp[\rho_1\, \tau_{0,1,\partial}]\ ;\ t < \tau_{0,1,\partial}\right] \\
&\leq \|\bar{h}\|_\infty\, \exp[-\rho_\alpha\, t] \left[1 + \rho_1 \int_{\mathbb{R}_+} \exp[\rho_1\, s]\, \mathbb{P}_{\mu A_t^{01}}(s < \tau_{0,1,\partial})\, ds\right] \\
&\leq \|\bar{h}\|_\infty \left[1 + \frac{\rho_1\, \|\bar{h}\|_\infty}{\rho_\alpha - \rho_1}\right] \exp[-\rho_\alpha\, t] := C\, \exp[-\rho_\alpha\, t].
\end{aligned}$$

The same holds of course for the case $\{\tau_{0,1,\partial} = \tau_{0,1}\}$ and $\mathbb{E}_\mu\left[\exp(\rho_1\, \tau_{0,1,\partial})\ ;\ \tau_{0,1,\partial} = \tau_{0,1} \leq t\right]$ converges with exponential rate. Therefore with (4.4.29) –and the well-defined notation (4.2.11)– we can define some $C > 0$ such that $\forall \mu \in \mathcal{M}_1([0,1])$:

$$|\mu A_t\{1\} - (1 - x(\mu))| \vee |\mu A_t\{0\} - x(\mu)| \vee |\mu A_t(0,1)| \leq C\, \exp[-(\rho_\alpha - \rho_0)\, t],$$

which concludes the proof of Proposition 4.2.4.1. □

4.4.8 Proof of Proposition 4.2.5.1 : the case $\rho_\alpha < \rho_0 \wedge \rho_1$

This proof is very similar to the one of Proposition 4.2.2.2, so we won't go into much detail. Lemmas 4.4.4.3 and 4.4.4.2 are of course replaced by :

Lemma 4.4.8.1. *Assume that $\rho_\alpha < \rho := \rho_0 \wedge \rho_1$. Then, there exists $n_\circ \geq 3, \xi_\circ > 0$ such that :*

$$\begin{aligned}
&\forall\, n \in \mathbb{N},\ \forall\, \xi > 0,\ \exists\, t_\circ > 0, \\
&\forall\, \mu \in \mathcal{M}^{01}_{n,\xi},\ \forall\, t \geq t_\circ, \qquad \mu A_t(1/n_\circ,\, 1 - 1/n_\circ) \geq \xi_\circ.
\end{aligned}$$

Lemma 4.4.8.2. *Assume that $\rho_\alpha < \rho := \rho_0 \wedge \rho_1$ and $\zeta \in \mathcal{M}_1[(0,1)]$. Then, there exists $t_P, c_P > 0$ such that :*

$$\forall\, x \in [0,1],\ \forall\, t \geq t_P, \quad \mathbb{P}_x(t < \tau_\partial) \leq c_P\, \mathbb{P}_\zeta(t < \tau_\partial).$$

We leave the proofs to the reader, and just mention that we can take as an upper-bound for $\mathbb{P}_x(\tau_1 \leq t < \tau_\partial)$ the same formula as for $\mathbb{P}_x(\tau_0 \leq t < \tau_\partial) = \mathbb{P}_x(\tau_0 \leq t < \tau_{1,\partial})$, with ρ_1 instead of ρ_0 (cf (4.4.17)).

For the rest of the proof, we remark that, for $\alpha^y := y_\alpha\, \alpha + y_0\, \delta_0 + y_1\, \delta_1$ with $y_\alpha + y_0 + y_1 = 1$, (4.4.10) has to be replaced by :

$$\begin{aligned}
\alpha^y P_t(dx) = y_\alpha\, \exp[-\rho_\alpha\, t]\, \alpha(dx) &+ [y_0\, \exp[-\rho_0\, t] + y_\alpha\, \mathbb{P}_\alpha(\tau_0 \leq t < \tau_{1,\partial})]\, \delta_0(dx) \\
&+ [y_1\, \exp[-\rho_1\, t] + y_\alpha\, \mathbb{P}_\alpha(\tau_1 \leq t < \tau_{1,\partial})]\, \delta_1(dx). \qquad (4.4.30)
\end{aligned}$$

Again : $\alpha^y P_t(dx) = \exp[-\rho_\alpha\, t]\, \alpha^y(dx)$ iff the conditions in (4.2.12) are satisfied. □

4.4.9 Proof of Proposition 4.2.6.1 : the case $\rho_\alpha = \rho_1 < \rho_0$

Let us first prove that we only need to control $\|\mu A_t^0 - \delta_1\|_{TV}$ like it is done in Proposition 4.2.3.2. From Proposition 4.2.2.2, we know that for some $\alpha_1 := y_\alpha\, \alpha + y_0\, \delta_0$, with $y_\alpha, y_0 \in (0,1)$, there exists $C^1, \chi^1 > 0$ such that :

$$\left\|\mu A_t^1 - \alpha_1\right\|_{TV} \leq C^1 \exp[-\chi^1\, t]. \tag{4.4.31}$$

Consequently, for t sufficiently large :

$$\frac{y_0}{2\, y_\alpha} \leq \frac{\mu A_t\{0\}}{\mu A_t(0,1)} \leq \frac{2\, y_0}{y_\alpha}. \tag{4.4.32}$$

On the other hand, with the notation $\quad \mu A_t^0(dx) := \mathbb{P}_\mu\left(X_t \in dx \,\middle|\, t < \tau_{0,\partial}\right)$:

$$\|\mu A_t - \delta_1\|_{TV} = \left[1 + \frac{\mu A_t\{1\}}{\mu A_t(0,1) + \mu A_t\{0\}}\right]^{-1}, \qquad \left\|\mu A_t^0 - \delta_1\right\|_{TV} = \left[1 + \frac{\mu A_t\{1\}}{\mu A_t(0,1)}\right]^{-1}.$$

Consequently, (4.4.32) implies that $\|\mu A_t - \delta_1\|_{TV}$ has the same rate of convergence as $\|\mu A_t^0 - \delta_1\|_{TV}$ (as long as it indeed converges to 0).
Now, from the proof of Proposition 4.2.3.2, we deduce quite immediately :

$$\begin{gathered}\exists\, t_\vee, C > 0, \quad \forall\, t \geq t_\vee,\ \forall\, \mu \in \mathcal{M}_1([0,1]), \quad \left\|\mu A_t^0 - \delta_1\right\|_{TV} \leq C/t,\\ \forall\, n \geq 3,\ \forall\, \xi > 0, \quad \exists\, t_{n,\xi}, c_{n,\xi} > 0,\ \forall\, t \geq t_{n,\xi},\ \forall\, \mu \in \mathcal{M}_{n,\,\xi}^{0,1},\\ \left\|\mu A_t^0 - \delta_1\right\|_{TV} \geq c_{n,\xi}/t. \end{gathered}$$

□

4.4.10 Proof of Proposition 4.2.7.1 : the most critical case $\rho_\alpha = \rho_0 = \rho_1$

Any convex combination of δ_0 and δ_1 is clearly a QSD with extinction rate $\rho := \rho_0 = \rho_1 = \rho_\alpha$.
For $t \geq 0$ and $x \in [0,1]$, let :

$$h_t(x) := \exp[\rho\, t]\, \mathbb{P}_x(t < \tau_{0,1,\partial}) \;,\; E_0^t(x) := \mathbb{E}_x\left[\exp[\rho\, \tau_{0,1,\partial}] \;;\; \tau_{0,1,\partial} = \tau_0 \leq t\right] \tag{4.4.33}$$

$$E_1^t(x) := \mathbb{E}_x\left[\exp[\rho\, \tau_{0,1,\partial}] \;;\; \tau_{0,1,\partial} = \tau_1 \leq t\right], \tag{4.4.34}$$

Let then $k \geq 1$ and $\mu \in \mathcal{M}_1([0,1])$ with $\mu(0,1) > 0$, so that $\langle \mu \,|\, h\rangle > 0$. Then :

$$\langle \mu \,|\, E_0^k\rangle = \sum_{j=0}^{k-1} \langle \mu \,|\, h_j\rangle\; \langle \mu A_j^{01} \,|\, E_0^1\rangle,$$

where by (4.2.2), (with the upper-bound e^ρ of E_0^1), there exists $C > 0$ such that :

$$|\langle\mu \,\big|\, h_j - h\rangle\,| \leq C\,\exp[-j\,\chi]\,, \qquad |\langle\mu A_j^{01} - \alpha \,\big|\, E_0^1\rangle| \leq C\,\exp[-j\,\chi].$$

Consequently :

$$|\langle\mu \,\big|\, E_0^k\rangle - k\,\langle\mu \,\big|\, h\rangle\,\langle\alpha \,\big|\, E_0^1\rangle| \leq 2\,C/(1+\exp[-\chi]) < \infty. \tag{4.4.35}$$

Likewise $|\langle\mu \,\big|\, E_0^k + E_1^k\rangle - k\,\langle\mu \,\big|\, h\rangle\,\langle\alpha \,\big|\, E_0^1 + E_1^1\rangle| \leq 4\,C/(1+\exp[-\chi]) < \infty.$

From (4.4.29) and (4.2.3), we deduce that there exists $C' > 0$ such that :

$$\left|\mu A_k\{0\} - \frac{\langle\alpha \,\big|\, E_0^1\rangle}{\langle\alpha \,\big|\, E_0^1 + E_1^1\rangle}\right| \leq \frac{C'}{k\,\langle\mu \,\big|\, h\rangle} \tag{4.4.36}$$

The symmetrical result for $\mu A_k\{1\}$ holds of course true, and since the sum of the limits equals 1, we deduce also

$$|\mu A_k(0,1)| \leq \frac{C'}{k\,\langle\mu \,\big|\, h\rangle}. \tag{4.4.37}$$

Again, from Theorem 2.6 in [CMS13], the exit state is independent from the exit time when the initial condition is a QSD, with an exponential law for the exit time. Thus :

$$\langle\alpha \,\big|\, E_0^1\rangle = \mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right)\;\textstyle\int_0^1 \exp[\rho\,s]\,\rho\,\exp[-\rho\,s]\,ds = \mathbb{P}_\alpha\left(\tau_0 = \tau_{0,1,\partial}\right).$$

To end the proof, just remark that $\langle\mu \,\big|\, h\rangle$ is lower-bounded for any $\mu \in \mathcal{M}_{n,\xi}^{01}$. $\square$

4.4.11 Proof of Proposition 4.2.8.1 : $\rho_\alpha(\gamma) \to \infty$ as $\gamma \to \infty$

We assume first that $r \equiv 0$ and choose arbitrary some t, for instance $t := 1$.

Consider $T_\delta := \inf\{u \geq 0\ ;\ X_u\,(1 - X_u) \leq \delta\}$, which can possibly be 0. Given any $\epsilon > 0$, we want to prove that choosing δ sufficiently small ensures, uniformly for $\gamma \geq 1$: $\mathbb{P}_x(T_\delta \leq t_0\,,\; 2\,t_0 < \tau_{0,1}) \leq \epsilon$.

We can notice that :

$$X(t_0) = x_0 - s\,.\,T(t_0) + \gamma\,\widetilde{B}[T(t_0)] \quad \text{ with } T(t_0) := \int_0^{t_0\wedge\tau_{0,1}} X_u\,(1 - X_u)\,du, \tag{4.4.38}$$

and $\widetilde{B}$ has the law of a Brownian Motion. Indeed, define $\widetilde{B}_v := B(T^{-1}(v) \wedge \tau_{0,1}) + \mathbf{1}_{\{\tau_{0,1} < T^{-1}(v)\}}\,(\hat{B}(T^{-1}(v)) - \hat{B}(\tau_{0,1}))$, with $\hat{B}$ another Brownian Motion independent of

B. Since for any $v > 0$, $T^{-1}(v) := \inf\{t \geq 0 : \int_0^t X_u\,(1 - X_u)\,du > v\}$ is a stopping times, $\hat{B}$ is indeed a continuous martingale with respect to the filtration $\mathcal{F}_{T^{-1}(v)}$. Finally, the change of variable $w = T(u)$ ensures that $\mathbb{E}[(\widetilde{B}_v - \widetilde{B}_{v'})^2] = v - v'$ for any $v > v'$.

On the other hand, by Itô's formula :

$$\mathbb{E}_x\left(X_t - x - \int_0^{t_0 \wedge T_\delta} s\,X_u\,(1 - X_u)\,du\right)^2 = \mathbb{E}_x\left(\int_0^{t_0 \wedge T_\delta} \gamma\,\sqrt{X_u\,(1 - X_u)}\,dB_u\right)^2$$
$$= \mathbb{E}_x\left[\int_0^{t_0 \wedge T_\delta} \gamma^2\,X_u\,(1 - X_u)\,du\right]$$
$$\text{thus } \mathbb{P}_x(t_0 < T_\delta) \times t_0\,\gamma^2\,\delta \leq (2 + s\,t_0/2)^2, \quad \text{independent from } x.$$

For γ sufficiently large, it implies that $\mathbb{P}_x(t_0 < T_\delta)$ is indeed lower than ϵ.

Thus $\mathbb{P}_x(t_0 < \tau_{0,1}) \leq \mathbb{P}_x(T_\delta \leq t_0\,,\,2\,t_0 < \tau_{0,1}) + \mathbb{P}_x(t_0 < T_\delta) \leq 2\epsilon$.

In the general case of bounded r, we deduce for the QSD α that, for γ large enough :

$$\mathbb{P}_\alpha(2\,t_0 < \tau_{0,1}) = \exp[-2\,\rho_\alpha\,t_0] \leq 2\,\exp[2\,\|r\|_\infty\,t_0]\,\epsilon.$$

It indeed proves that $\rho_\alpha \to \infty$ as $\gamma \to \infty$. □

4.4.12 Proof of Proposition 4.2.8.2 : $\rho_\alpha < \rho_0 \wedge \rho_1$ for r sufficiently strong

Define r_2, r_3 such that $\max r(x) < r_3 < r_2 < r(1) \wedge r(0)$ and the open sets $A := r^{-1}([0, r_3)) \subset B := r^{-1}([0, r_2)) \subset (0, 1)$ (recall that r is assumed to be continuous). We choose arbitrary t_0. A classical result on diffusion ensures that there exists $\rho > 0$ such that :

$$\inf_{x \in A} \mathbb{P}_x(X_{t_0} \in A\,,\ \forall\, s \leq t,\ X_s \in B) \geq \exp[-\rho\,t_0].$$

Then, it implies by the Markov property :

$$\inf_{x \in A} \mathbb{P}_x(\forall\, s \leq t,\ X_s \in B\ ;\ t_0 < \tau_{0,1,\partial}) \geq C\,\exp[-(\rho + R\,r_2)\,t].$$

From the Harnack inequality, we know that $\alpha^{(R)}$ has a lower-bounded density on any open set of $(0, 1)$ so that $\alpha^{(R)}(A) > 0$ and

$$\alpha^{(R)}(B) \geq \exp[\rho_\alpha^{(R)}\,t]\mathbb{P}_{\alpha^{(R)}}(\forall\, s \leq t,\ X_s \in B\ ;\ t < \tau_{0,1,\partial}) \geq C\,\alpha^{(R)}(A)\,\exp[-(\rho + R\,r_2 - \rho_\alpha^{(R)})\,t].$$

This proves $\rho_\alpha^{(R)} \leq \rho + R\,r_2 < R\,(r(0) \wedge r(1)) = \rho_0^{(R)} \wedge \rho_1^{(R)}$ for R sufficiently large. □

4.4.13 Proof of Proposition 4.2.8.3 : $\rho_0 \wedge \rho_1 < \rho_\alpha$ for r sufficiently weak

Let $\rho_\alpha^{(0)}$ be the death rate of the QSD for the Wright-Fisher diffusion conditioned not to touch the boundary with $r = 0$. Since $\rho_0^{(R)} \wedge \rho_1^{(R)} = R \times [r^0(0) \vee r^0(1)] \to 0$ as $R \to 0$, it is sufficient to prove that $\liminf_R \rho_\alpha^{(R)} \geq \rho_\alpha^{(0)}$ to deduce Proposition 4.2.8.3.

By Proposition 4.2.0.1, for any $x \in (0,1)$ and $t \geq 0$:

$$\mathbb{P}_x(t < \tau_{0,1}) \leq \|\bar{h}\|_\infty \exp[-\rho_\alpha^{(0)}\, t]\mathbb{P}_x(t < \tau_{\partial,0,1}) \leq \|\bar{h}\|_\infty \exp[-(\rho_\alpha^{(0)} - R\,\|r\|_\infty)\, t]$$

and in particular, with the QSD $\alpha(R)$ as initial condition, we deduce $\rho_\alpha^{(R)} \geq \rho_\alpha^{(0)} - R\,\|r\|_\infty$.

□

Remark : In fact, $\rho_\alpha^{(R)} \to \rho_\alpha^{(0)} > 0$ as $R \to 0$, because :

$$\exp[-(\rho_\alpha^{(0)} + R\,\|r\|_\infty - \rho_\alpha^{(R)})\, t] \leq \exp[+(\rho_\alpha^{(R)})\, t], \mathbb{P}_{\alpha^{(0)}}(t < \tau_{\partial,0,1}) \underset{t\to\infty}{\longrightarrow} \langle \alpha^{(0)} \,\big|\, h^{(R)} \rangle,$$

which implies $\rho_\alpha^{(R)} \leq \rho_\alpha^{(0)} + R\,\|r\|_\infty$.

4.4.14 Proof of Proposition 4.2.8.4 : concentration towards 0 as $\gamma \to 0$

Since r is bounded, the probability of the event $\{t < \tau_\epsilon\}$ with r is at most $\exp(\|r\|_\infty\, t)$ times the probability with $r \equiv 0$. If we prove that the latter converges to 0 (as a limit of this parameter γ), it will be the same for the former. We can thus assume without loss of generality that $r \equiv 0$.

We recall (see (4.4.38)) that for any $t \geq 0$ and initial condition $1 - \epsilon$:

$$X(t) = 1 - \epsilon - s\,.\,T(t) + \gamma\,\widetilde{B}[T(t)] \quad \text{with } T(t_0) := \int_0^{t_0 \wedge \tau_{0,1}} X_u\,(1 - X_u)\,du,$$

and $\widetilde{B}$ has the law of a Brownian Motion.

Fix some $M > 0$ ans assume that $\gamma \leq \epsilon/(2\,M)$ and that we are conditionally on the event $\{\sup_{u \leq 1/s} |\widetilde{B}_u| \leq M\}$. Then, $T(\infty) < 1/s$, since $T^{-1}(1/s)$ would be well-defined otherwise and would satisfy :

$$X(T^{-1}(1/s)) \leq 1 - \epsilon - 1 + \epsilon/2 \leq -\epsilon/2 < 0,$$

which contradicts the classical property that X takes its value on $[0,1]$. By the definition of $T(t)$, we see that there exists $c = c(\epsilon) > 0$ such that :

$$1/s > T(\tau_{\epsilon/2} \wedge \tau_{1-\epsilon/2}) \geq c\,\tau_{\epsilon/2} \wedge \tau_{1-\epsilon/2}. \tag{4.4.39}$$

We also deduce that for any $u \leq T(\infty)$, since $u < 1/s$:

$$X(T^{-1}(u)) \leq 1 - \epsilon + \epsilon/2 \leq 1 - \epsilon/2.$$

This implies that $\tau_{1-\epsilon/2} = \infty$ on the event $\{\sup_{u \leq 1/s} |\widetilde{B}_u| \leq M\}$ and by (4.4.39) that $\tau_{\epsilon/2} \leq t := 1/(cs)$. This directly implies :

$$\forall \gamma \leq \epsilon/(2M), \quad \mathbb{P}_{1-\epsilon}(t < \tau_\epsilon) \leq \mathbb{P}(\sup_{u \leq 1/s} |B_u| \geq M).$$

Letting M tend to ∞ concludes the proof of Proposition 4.2.8.4. □

4.5 Proof of Theorem 4.1.1

Like in [LM15], the proof follows a standard procedure [FM04], [JM86], [CFM06] in which we prove : (i) the tightness of the sequence of stochastic processes – which implies a subsequential limit, and (ii) the uniqueness of this limit. For the tightness of $\{\mu_t^{m,n}\}_{m,n}$ on $D([0,\ T],\ \mathcal{P}([0,1]))$, it is sufficient, by Theorem 14.26 in Kallenberg [Ka97] to show that $\{\langle \mu_t^{m,n} \,|\, f \rangle\}$ is tight on $D([0,\ T],\ \mathbb{R})$ for any test function f from a countably dense subset of continuous, positive functions on $[0,\ 1]$.

4.5.1 Semimartingale property of multilevel selection process

It will be useful for what follows to treat $\langle \mu_t^{m,n} \,|\, f \rangle$ as a semimartingale. We exploit the following discrete derivatives of f, with span $1/n$:

$$\begin{aligned} &D_x^+ f(x) := n\,(f(x+1/n) - f(x))\,, \quad D_x^- f(x) := n\,(f(x) - f(x-1/n)), \\ &D_{xx} f(x) := n^2 \times (f(x+1/n) + f(x-1/n) - 2f(x)) = n \times (D_x^+ f(x) - D_x^- f(x)). \end{aligned}$$

We recall that in our limit, $n, m \to \infty$, $\bar{\gamma}_I/n \to \gamma_I$, $\bar{\gamma}_I\,\bar{s} \to s$, $\{\bar{\gamma}_G \bar{r}(x)\}$ is fixed and $\bar{\gamma}_G$ bounded. It is easy to adapt the proof of [LM15] in order to state :

Lemma 4.5.1.1. *For $f \in C^2([0,1])$ and $\mu_t^{m,n}$ with generator $L^{m,n}$ defined in (1),*

$$\langle \mu_t^{m,n} \,|\, f \rangle - \langle \mu_0^{m,n} \,|\, f \rangle \ = A_t^{m,n}(f) + M_t^{m,n}(f)$$

where $A_t^{m,n}(f)$ is a process of finite variation, $A_t^{m,n}(f) := \int_0^t a_s^{m,n}(f) ds$, with :

$$\begin{aligned} a_t^{m,n}(f) = &\sum_i \mu_t^{m,n}\left({}^{i}\!/_{n}\right) \times {}^{i}\!/_{n}\left(1 - {}^{i}\!/_{n}\right)\ \left[\left({}^{\bar{\gamma}_I}\!/_{n}\right) D_{xx} f\left({}^{i}\!/_{n}\right) - \bar{\gamma}_I\,\bar{s}\ D_x^- f\left({}^{i}\!/_{n}\right)\right] \\ &+ \bar{\gamma}_G \left\{\sum_j \mu_t^{m,n}\left({}^{j}\!/_{n}\right) \bar{r}\left({}^{j}\!/_{n}\right)\ f\left({}^{j}\!/_{n}\right) - \sum_i \mu_t^{m,n}\left({}^{i}\!/_{n}\right) f\left({}^{i}\!/_{n}\right) \sum_j \mu_t^{m,n}\left({}^{j}\!/_{n}\right) \bar{r}\left({}^{j}\!/_{n}\right)\right\} \end{aligned}$$

and $M_t^{m,n}(f)$ *is a càdlàg martingale with (conditional) quadratic variation :*

$$\langle M^{m,n}(f)\rangle_t = \frac{1}{m}\int_0^t \Bigg\{ \frac{\bar\gamma_I}{n}\sum_i \mu_s^{m,n}\left({}^i/_n\right)\ {}^i/_n\ \left(1-{}^i/_n\right)\ \left[\left(D_x^+ f\left({}^i/_n\right)\right)^2 + (1+\bar s)\left(D_x^- f\left({}^i/_n\right)\right)^2\right]$$
$$+\bar\gamma_G \sum_{i,j}\mu_s^{m,n}\left({}^i/_n\right)\mu_s^{m,n}\left({}^j/_n\right)\ \left(1+\bar r({}^j/_n)\right)\left[f\left({}^i/_n\right) - f\left({}^j/_n\right)\right]^2 \Bigg\} ds$$

4.5.2 Proof of the convergence to our limit

We prove here that the drift term is tight while the martingale converges to zero. For the finite variation term $A_t^{m,n}(f)$, assuming w.l.o.g. $\bar\gamma_I/n \le 2\,\gamma_I$, $\bar\gamma_I\,\bar s \le 2\,s$:

$$|a_t^{m,n}(f)| \ \le\ 2\gamma_I\|f''\|_\infty + 2\,s\ \|f'\|_\infty + 2\ \|r\|_\infty\|f\|_\infty := G_f$$
$$\text{therefore : } \sup_{t\in[0,T]} |A_t^{m,n}(f)| \ \le G_f T$$

where G_f is a constant that depends on f. Moreover, for any prescribed ϵ, we can always choose $\delta_\vee$ to be sufficiently small so that, for any $0 \le t \le t+\delta$ with $\delta \le \delta_\vee$, for any n, m : $|A_{t+\delta}^{m,n} - A_t^{m,n}| \le \delta G_f \le \epsilon$. By Proposition 3.26, Chapter 3 in [JS03], this proves immediately that the sequence $(A_t^{m,n})_{t\le T}$ is tight, and any limit is continuous. For the martingale part, assuming w.l.o.g. $\bar s \le 1$:

$$\langle M_t^{m,n}(f)\rangle_t \le \ \frac{T}{m}\left\{6\,\gamma_I\,\|f'\|_\infty + (\bar\gamma_G + \|r\|_\infty)\,\|f\|_\infty^2\right\} := J_f/m \underset{m\to\infty}{\longrightarrow} 0,$$

where J_f is a constant only depending on $T>0$ and $f \in C^2([0,1])$. From Burkholder-Davis-Gundy's inequality, since the jumps of $M_t^{m,n}(f)$ are bounded by $\|f\|_\infty/m$:

$$\mathbb{E}\left[\sup_{t\le T}(M_t^{m,n}(f))^2\right] \le C\,J_f/m + \|f\|_\infty^2/m^2 \underset{n,m\to\infty}{\longrightarrow} 0.$$

This proves that $M_t^{m,n}(f)$ converges to 0 in such a way that $(\langle \mu_t^{m,n} \,\big|\, f\rangle - \langle \mu_0^{m,n} \,\big|\, f\rangle)_{t\le T}$ is tight and any associated limit is continuous.

By construction and the Law of Large Numbers, $\langle \mu_0^{m,n} \,\big|\, f\rangle$ converges to $\langle \mu_0 \,\big|\, f\rangle$. Thus, the sequence $(\mu_t^{m,n})_{t\le T}$ for $n, m \ge 1$ is tight in $D([0;T]; \mathcal{M}_1([0;1]))$.

So, we consider a subsequence $(\mu_t^{(k)})_{t\le T} = (\mu_t^{m_k,n_k})_{t\le T}$ such that $m_k, n_k \to \infty$ as $k\to\infty$ and such that $(\mu_t^{(k)})$ converges to $(\mu_t)_{t\le T}$ in $D([0;T]; \mathcal{M}_1([0;1]))$. Necessarily, μ_0 coincide with the law of the initial condition provided in the assumptions of Theorem 4.1.1. For any $f \in C_2([0,1])$ and $t \le T$, it is not difficult to see that as

$k \to \infty$:

$$\begin{aligned}
&\langle \mu_t^{(k)} \big| f\rangle \to \langle \mu_t \big| f\rangle\,, \qquad \langle \mu_0^{(k)} \big| f\rangle \to \langle \mu_0 \big| f\rangle,\\
&a_t^{(k)}(f) \to \langle \mu_t \big| \mathcal{L}_{WF} f\rangle + \langle \mu_t \big| r \times f\rangle - \langle \mu_t \big| f\rangle \times \langle \mu_t \big| r\rangle\,, \qquad M_t^{(k)}(f) \to 0.
\end{aligned}$$

Thus (μ_t) is a solution to equation (4.1.5). From the uniqueness property that we proved in Proposition 4.1.2.1, and the tightness of the sequence, we conclude that $(\mu_t^{m,n})_{t\leq T}$ converges globally to this solution. This concludes the proof of Theorem 4.1.1 and more globally the proofs presented in this paper. □

Appendix : Proof of Lemma 4.3.1.1

$\mu_t^{m,n}$ is a Markov process evolving on a finite state, so that the analysis of the QSD is much more elementary. The exit from any site $\mu \in \mathcal{M}^{m,n}([0,1]) \setminus \{\delta_0, \delta_1\}$ is given by an exponential law whose rate is lower-bounded by some positive constant. So it is natural to look at the induced Markov Chain $(M_k^{m,n})_{k\in\mathbb{Z}_+}$ which specify the state of $\mu^{m,n}$ just after its k-th jump (and $M_0^{m,n} = \mu_0$). Provided that $M_k^{m,n}$ satisfies that for any $\mu_1, \mu_2 \in \mathcal{M}^{m,n}([0,1]) \setminus \{\delta_0, \delta_1\}$, there exists $k \geq 1$ such that

$$\mathbb{P}_{\mu_1}(M_k^{m,n} = \mu_2) > 0, \tag{4.6.1}$$

it is classical, see for instance Section 3.1 in [CMS13], that the transition matrix of $\mu^{m,n}$ satisfies the conditions of Perron-Frobenius theorem, namely that it is irreducible (with non-negative coefficients).

The proof of (4.6.1) is elementary. We only give a few comments for the interested reader. The main point is that the Moran transitions allow to change any polymorphic group to any other type of group : if you require more C types, simply impose several events of replacement of D types by reproducing C types. Likewise, it is not difficult to get to a situation where any group is polymorphic by replacing pure groups. Handling independently each group (without group reproduction) as described two sentences before, we can make the process reach any state from there.

Then, the Perron-Frobenius theorem implies that there exists a unique associated QSD $A^{m,n} \in \mathcal{M}_1(\mathcal{M}^{m,n}([0,1]))$ and a unique capacity of survival $H^{m,n} \in L_\infty(\mathcal{M}^{m,n}([0,1]))$ with extinction rate ρ_A. Moreover, there exists C, γ such that the following convergence results hold :

$$\begin{aligned}
&\forall\, \mu_0, \mu_1 \in \mathcal{M}^{m,n}([0,1]) \setminus \{\delta_0, \delta_1\},\\
&|\exp(\rho_A\, t)\mathbb{P}_{\mu_0}(\mu_t^{m,n} = \mu_1) - H^{m,n}(\mu_0) \times A^{m,n}(\{\mu_1\})| \leq C \exp(-\gamma\, t).
\end{aligned}$$

Fixing the initial condition μ_0, let us denote $f_0(t) := \mathbb{P}_{\mu_0}(\mu_t^{m,n} = \delta_0)$ and resp.

$f_1(t) := \mathbb{P}_{\mu_0}(\mu_t^{m,n} = \delta_1)$. Since δ_0 (resp. δ_1) is absorbing, f_0 (resp. f_1) is an increasing function. Moreover, since $(\mu^{m,n})$ is a conservative process on $\mathcal{M}_1^{m,n}([0,1])$ and similarly as we had in (4.2.3) :

$$1 - f_0(t) - f_1(t) = \mathbb{P}_{\mu_0}(\mu_t^{m,n} \notin \{\delta_0, \delta_1\}) = \mathbb{P}_{\mu_0}(t < \tau_{0,1}^{m,n}) \leq C \exp[-\rho_A t].$$

It proves that $f_0(\infty) + f_1(\infty) = 1$ and that :

$$0 \leq f_0(\infty) - f_0(t) \leq 1 - f_0(t) - f_1(t) \leq C \exp[-\rho_A t].$$

This ends the proof of Lemma 4.3.1.1. $\square$

5 Metastability between the clicks of the Muller ratchet

This chapter has been completed very recently and it shall be submitted soon.

Abstract

We prove the existence and uniqueness of a quasi-stationary distribution for three stochastic processes derived from the model of the Muller ratchet. This model has been originally introduced to quantify the limitations of a purely asexual mode of reproduction in preventing, only through natural selection, the fixation and accumulation of deleterious mutations. As we can see by comparing the proofs, not relying on the discreteness of the system or an imposed upper-bound on the number of mutations makes it much more necessary to specify the behavior of the process with more realistic features. The third process under consideration is clearly non-classical, as it is a stochastic diffusion evolving on an irregular set of infinite dimension with hard killing at an hyperplane. We are nonetheless able to prove results of exponential convergence in total variation to the quasi-stationary distribution even in this case. The parameters in this last result of convergence are directly related to the core parameters of the Muller ratchet effect (although the relation is very intricate). The speed of convergence to the quasi-stationary distribution deduced from the infinite dimensional model extends to the approximations with a large yet finite number of potential mutations. In this dynamics, a crucial role is played by the moments in the empirical distribution for the number of carried mutations. From our estimates that these moments quickly decay, we deduce an upper-bound on their expectation under the QSD. As previously, these upper-bounds extend to the models where the number of mutations is large yet bounded.

5.1 Introduction

5.1.1 General presentation

Since deleterious mutations occur much more frequently than beneficial ones, it is crucial to understand how the fixation of these deleterious mutations is regulated. Notably, it is very exceptional that another mutation replaces a deleterious one, so that only natural selection can maintain some purity in the population. In this respect, there is a major distinction to be made between sexual and asexual reproduction : in a purely asexually reproducing population, a deleterious mutation can only be purged when the lineages carrying it go extinct. In a sexually reproducing population, such a deleterious mutation can be avoided through recombination, without getting rid of the whole set of other mutations carried by the lineages. There is actually no strong evidence that deleterious mutations are specifically targeted during this process of recombination. It appears sufficient that at random some lineages do not carry the mutation any longer and that natural selection comes into play. Such an advantage for sexual reproduction is however to be confronted with the cost (in terms of reproduction efficacy) of requiring two parents. This aspect of purging deleterious mutations is often cited to explain the success of sexual reproduction (see [Ma78] for more details).

Although we base the following models on the above scheme of purging deleterious mutations, the situation is more intricate in reality. In many asexual populations, there is evidence of the role of horizontal gene transfers, for instance with plasmids ([KP08], [MJ10], [OLG00], [TR+08]) It can be seen as a weak form of recombination. Moreover, the fact that mutations are deleterious is due to a change in the physiology that may be compensated by other means. It might even happen that after subsequent mutations, the carriers of an initially deleterious mutation become more adapted that the wild types [SB+14].

In the current paper, we wish to justify the existence and uniqueness of a metastable state in which selective effects are able to maintain the population from having deleterious mutations fixed. The rigorous definition of such metastable state where no click occurs can be obtained in a broad generality by a conditioning of the stochastic process.

There is a classical issue of specifying the conditions for metastability to be the most common observable. A generally accepted answer is to compare the so-called relaxation time t_R, which quantifies the rate at which the dependency in the past conditions vanish, and average clicking time t_C of the system. Metastability between clicks would be the most common observation provided $t_R \ll t_C$, so that a sequence of i.i.d. exponential law provides a an accurate description of the sequence of intervals between clicks. If t_R is of the same order as t_C or larger, we a priori can not exclude that trains of short interdependent intervals could alter this observed distribution of interval length. However, if t_R is of the same order as t_C, there shall still be long realizations of inter-click intervals after which we can say that the dependency in the past is forgotten.

Our theorems provide a proper definition of these two main quantities, by relying on recent techniques for ensuring the convergence to a unique quasi-stationary distribution (QSD). Notably, we prove that this QSD is approached at an exponential rate, from which we derive t_R, by the marginal law of the process conditioned upon the fact that the click has not occurred. We establish these results for several versions of the model. In the context of large population that we are interested in for such metastability to be reasonably likely, our primary interest is in the proof of quasi-stationary results for the diffusive limiting model without any artificial constraint in the number of deleterious mutations. Notably, the parameters involved in the convergence results only depend on the mutation rate λ and the strength of selection α (although the relation is too intricate to be reasonably given). We shall also mention that, by choosing the right time-scale for simplicity, we arbitrarily fix demographic fluctuations to be of order 1.

As compared to the other models that we also treat, the proof is more difficult because it additionally exploits the effect of selection to obtain a bound on the maximal number of accumulated mutations that we need to handle. Such a bound is already needed for the proof of Theorem 5.2.1. But the fact that the process concerns there a discrete population greatly simplifies the argument at the expense of an artificial dependency on the population size in the convergence parameters. In the diffusive limit, an additional difficulty arises in that the diffusion is degenerate in a non-smooth boundary that is partly absorbing and partly repulsive. We thus introduce such a bound on mutations in the context of Section 5.2.3 and the proofs in Section 5.4 in order to simplify the analysis and relate to simulations for which it may be done. It means of course that the obtained constants depend a priori strongly on the introduced threshold. It is however proved finally that we can forget about this dependency, at least for large thresholds.

In the next Section 5.1.2, we specify the stochastic processes under consideration, first with the individual-based model in Section 5.1.2.1 and then with the diffusive limits in Section 5.1.2.2. We conclude the introduction by a justification for the study of quasi-stationarity in Section 5.1.3 and for not introducing a bound on the number of mutations in Section 5.1.4. Our results of quasi-stationarity are presented in Section 5.2. Starting in Section 5.2.1 with the general notion of exponential quasi-stationarity that we aim to prove, we treat resp. in Sections 5.2.2, 5.2.3 and 5.2.4 each of the three stochastic processes mentioned above, in increasing order of difficulty (and generality). Then, we discuss more precisely the interpretation of these results in Section 5.2.5. The rest of the paper is dedicated to the proofs. Sections 5.3, 5.4 and 5.6 are devoted to the proofs of quasi-stationarity for each of the three processes, while Section 5.5 is devoted to upper-bounds on the moments of the QSDs. Such controls of the moments are exploited in Section 5.6, so that this ordering appears natural. We defer to the Appendix some elements of these proofs that we judge elementary and naturally expected.

Please note that the notations for the parameters involved in our statements may differ between Sections 5.4 and 5.6. Although their definitions have been adapted, the notations for the subsets $\mathcal{D}_\ell^{(d)}$, $E^{(d)}$ for the irreductibility measure $\zeta^{(d)}$ have been kept in order to more easily relate to our assumptions while keeping the notations light.

5.1.2 The mathematical model of the Müller ratchet

5.1.2.1 The individual-based model as a guideline

A simplified mathematical model has been proposed by Haigh in [Ha78] to quantify the regulation of deleterious mutations in an asexual population. Since in finite populations, the ultimate fixation of deleterious mutations cannot be avoided (unless by the extinction of the population), it has been called Muller's ratchet : assuming a constant deleterious effect of mutations, at each time the fittest individuals disappear, the ratchet clicks in the sense that an additional deleterious effect has been shared by the whole population (present and future).

This first model with discrete generations and fixed population size N evolves as follows. Mutations occur at the birth of individual and are only deleterious. The number of mutations carried by the offspring is the number carried by its parent plus the realization of a Poisson random variable with average $\lambda > 0$. The strength of this drawback is quantified by $\alpha \in (0,1)$. Assume that the current population is distributed with N_i individuals carrying i mutations and consider an individual from the next generation. Each shall choose its parent independently where the probability that he chooses a specific parent carrying i mutations is :

$$\frac{(1-\alpha)^i}{\sum_{k\geq 0} N_k (1-\alpha)^k}.$$

In addition to the mutations of its parent, each newborn gains ξ deleterious mutations, where ξ is a Poisson random variable with mean λ.

5.1.2.2 The stochastic diffusion under consideration

In the following, we also consider a description of the model that corresponds to a limit of large population size, small effects and mutation rate and an accelerated time-scale (for which time is continuous). In the following statements, $d \in \mathbb{N} \cup \{\infty\}$ defines an upper-bound on the number of deleterious mutations that can be carried by an individual. If $d := \infty$ in the following expression, $i \in [\![0, d]\!]$ has to be understood as $i \in \mathbb{N}$.

We are interested in the following Fleming-Viot system of SDEs for the $X_i^{(d)}(t)$'s, $i \in [\![0, d]\!]$, where $X_i^{(d)}(t)$ denotes the proportion of individuals in the population who carry exactly i deleterious mutations at time t (with $X_{-1}^{(d)} \equiv 0$) :

$$\begin{aligned}\forall\, i \leq d, \quad dX_i^{(d)}(t) &= \alpha(M_1^{(d)}(t) - i)\, X_i^{(d)}(t)\, dt + \lambda(X_{i-1}^{(d)}(t) - \mathbf{1}_{\{i<d\}} X_i^{(d)}(t))\, dt \\ &\quad + \sqrt{X_i^{(d)}(t)}\, dW_t^i - X_i^{(d)}(t)\, dW_t^{(d)} \qquad (5.1.1 : S^{(d)})\end{aligned}$$

where $W_t^{(d)} := \sum_{j=0}^d \int_0^t \sqrt{X_j^{(d)}(s)} dW_s^j, \quad M_1^{(d)}(t) := \sum_{i=0}^d i\, X_i^{(d)}(t),$

with $(W^i)_{i\geq 0}$ a family of mutually independent Brownian Motions.

Remarks 5.1.2.1. *Generally, the subscripts* (d) *will refer to the upper-bound* d *in the number of mutations. We usually keep it, except when the notation gets really too heavy (or by mistake).*

This process has been introduced in [EPW09] and it has been shown in [AP13] that clicks occur a.s. in finite time. In [PSW12], a closely related process with compensatory mutations is considered. We refer to this article for a detailed presentation of the connection to related individual-based models and only sketch next the interpretation of the parameters.

The selective effect of the deleterious mutations is the term proportional to α in the drift term. Since the population size is fixed, the growth rate of the individuals is shifted to be 0 on average over the population. As we assume that deleterious mutations carry the same burden, the growth rate of individuals carrying i mutations is proportional to the difference between i and the average number of mutations, i.e. $M_1^{(d)}(t)$. The occurrence of new mutations is modeled by the term proportional to λ in the drift term. λ corresponds to the rate at which individuals carrying i mutations obtain an additional one and become individuals carrying $i+1$ mutations (without dependency on i). Finally, the neutral choice of the individuals replaced at each birth events give rise to the last martingale part. For simplicity, we consider the time-scale at which this term has no coefficient. It corresponds to the rescaling of time $t \mapsto t'/N_e$, where N_e is the "effective population size".

Remarks 5.1.2.2. *This notion of "effective population size" has been largely considered to extend the properties of unstructured homogeneous individual-based models to individual-based models with a population structure that differentiates the individuals. So it is meant to be applied to real populations under ecological study. Notably, it provides the scaling of the genealogies that makes it approximate the standard Kingman's coalescent [Ki82]. Thus, it gives an estimate on the time at which lives the most recent common ancestor of a given sample in the population. It is of course natural that this quantity plays a role in such modeling of heritable factors.*

Remarks 5.1.2.3. *For practical reasons, the current formulation of the martingales is different from the one that is considered in [AP13] in the aftermath of [EPW09]. One can easily check however that the brackets of these martingale parts coincide, so that the models are actually the same.*

5.1.3 Motivation in the study of quasi-stationarity between clicks

The fact that such quasi-stationary regime can be defined does not automatically imply that this regime is likely to be observed. The next clicks could happen too quickly after the previous ones so that the distribution of mutations might remain in such transitory regime for a long period of time. In such a case, the time-interval between two following clicks could depend much on this distribution just after the former click. Provided that the process in the metastable regime keeps an optimal sub-population concentrated at a large size, the time between clicks can be much longer than in the above-described transitory mode. Indeed, in this metastable regime, the click would be the consequence of an exceptional deviation of the process away from the metastable attractor.

In this view, it shall be noted that, on large time-scales and broadly speaking, effects of natural selection on the mutation rate can be observed. In the context of a rapid succession of clicks, the population would be likely to get extinct quite early on as compared to populations able to reach a metastable regime between each click. This distinction between a dynamics with a rapid succession of clicks and a one with a succession of long metastable time-intervals is both highly dependent on the mutation rate and significant in terms of survival of the population. After generations under selection, The mutation rate is thus likely to be constrained. It should be sufficiently small in order for a population under normal conditions of life not to be subject to such a rapid succession of clicks. Also, advantageous mutations do not occur so frequently as compared to deleterious ones to compensate for this interest in reducing the mutation rate. The second scenario with metastability is thus expected to be the more likely for stable asexual populations, although the other one cannot be excluded for destabilized populations or too small ones. The interest in this metastable regime is thus biologically motivated by its benefice in term of survival.

The study of this quasi-stationary regime arises naturally when one wishes to estimate the rate at which the ratchet clicks. In the following, the clicking time at which the last individual with no deleterious mutations gets one will be considered as another type of extinction and denoted τ_∂. To obtain quantitative estimates, several authors have justified their approach by assuming that the typical clicking time t_C is much larger than the typical relaxation time t_R of the system, usually with an empirical reference for the latter ([EPW09], [ME13]). It is argued that there is a simpler approximation of this process in term of an uni-dimensional autonomous diffusion that is very accurate in the context of large populations. This autonomous diffusion simply corresponds to the proportion of fittest individuals. The rest of the distribution shall maintain itself close to a deterministic profile that is simply a function of this proportion of fittest individuals. Actually, the dependency in this number of fittest individuals only occurs in the normalizing factor of this distribution. Rigorously, the latter argument of concentration should rely on Large Deviation results that are not treated in the current paper. We rather focus on this objective of rigorously defining the relaxation time and the clicking time. This average clicking time t_C is derived as the inverse of the extinction rate of a specific quasi-stationary distribution (QSD), while the relaxation time t_R is derived as the inverse of the convergence rate to this

QSD.

In [ME13], an estimation of t_C in the context where $t_R \ll t_C$ is obtained through the characteristic equation of some QSD ν, of the form $\mathcal{L}\nu = -\lambda\nu$, with $\mathcal{L}$ the infinitesimal generator and some constant λ. This QSD that they study is not the general QSD, for which the description is reasonably argued to be too intricate, but the one of a one-dimensional approximation of the process under metastability. Note that the justification for this approximation relies notably upon the fact that $t_R \ll t_C$, where t_R has to be related to the complete QSD. The approximation is also based on a estimate of concentration as in Large Deviation theory : it shall ensure that, properly rescaled, the profile of the number of mutations carried by non-optimal individuals stay close under metastability to a deterministic one.

This profile is considered in the limit of large population. While still keeping a discrete description of the QSD, a continuous function of the proportion is introduced to specify its asymptotics. The approximation by the diffusion system is expected to be well-suited for population sizes way smaller than what is required to obtain such concentration effects (except for the classes of individuals carrying too many mutations, yet they shall hardly contribute).

For the first discrete model and including an upper-bound on the number of potential mutations, the existence and uniqueness of a quasi-stationary distribution would not be difficult to prove, as well as the exponential rate of convergence in total variation. By truncation, at rank d, we mean that we replace at each new generation the individuals carrying more than d mutations by individuals carrying exactly d as we have done for the definition of (5.1.1 : $S^{(d)}$). We may alternatively replace the individual by a new independent sample until the condition of having less than d mutations is satisfied. This case would correspond to the immediate killing of individuals with too many mutations rather than a saturation as we consider for simplicity. The details of the truncation are not crucial, especially since it shall hardly concern any individual for a sufficiently large rank. With the same argument as in Proposition 5.3.0.1, the transition matrix of the system has positive entries while restricted to states for which there is at least one individual without mutation. The result is then classically deduced from the Perron-Frobenius Theorem. We shall see in Section 5.2.2 that these results extend to the case of an unbounded number of mutations (thus an unbounded domain). To our knowledge, even the existence and uniqueness of a QSD has not been rigorously proved until now, although the authors of [ME13] rely on this notion for their approximations.

We also prove that the effect of the previously mentioned truncation (that may be useful in simulations) is vanishing in terms of the clicking rate and our convergence rate. We expect that it holds more generally for the relaxation rate, yet a precise estimate is a priori out of reach. Besides, it would not be much more difficult to extend the argument with overlapping generations.

Since the results of [ME13] or [EPW09] largely exploit the fact that the population is very large for their estimation, we wish to extend the justification of the relaxation time and the clicking rate for a limiting model where individuals are not distinguished. In this limit, the constants involved in the convergence are not contingent of the discreteness of the system.

It has been shown in [AP13] that clicks happen in a.s. finite time even for the model with a diffusion for an unbounded number of types. With τ_∂ the first hitting time of 0 by X_0, it means that τ_∂ is a.s. finite. Assuming that the population size is sufficiently large (which corresponds to the limit where α and λ tend to infinity as discussed above Remark 5.1.2.2), this event can however be made exceptional. We thus may expect that the initial condition of the process has been long forgotten at this time. On the event $\{t < \tau_\partial\}$ for quite large t, we may expect to see the profile of $(X_k(t))_{k \in Z_+}$ close to the deterministic Poisson distribution. This distribution, with mean λ/α, is shown in [EPW09] to be the equilibrium of the above infinite system of equations without martingale parts. A next step of the analysis consists in quantifying, thanks to the Large Deviation theory, the rate at which such deviations away from the deterministic limit happen.

In conclusion, our objective in the current paper is to prove that such a notion of quasi-stationarity between clicks can indeed be defined, to the expense of searching for a distribution on the space of population empirical distributions. Provided $t_C \gg t_R$, the successive intertime T^C_k, $k \in \mathbb{Z}_+$, between clicks are expected to be nearly distributed as a sequence of independent realizations. The tail of this distribution shall decay almost exponentially, at a rate that is the clicking rate of the QSD. Of course, this loss of dependency is invalidated for the realizations of T^C_k that are of order t_R. Because the clicking is already an exceptional dynamics of the class of optimal individuals, the new shifted profile at the clicking time is not likely to be particularly prone to a new click. Notably, it means that it is unlikely that T^C_k is negligible as compared to t_C. Yet, provided it is much larger than t_R, it means that the metastable regime is reached, so that the next clicking time can be assumed to happen at rate $1/t_C$. The average of T^C_k shall thus be nearly t_C and we do not expect accumulations of these exceptional events where T^C_k is smaller than a few t_R. Thus, these events should not matter much while considering the ultimate rate of fixation of deleterious mutations.

5.1.4 Motivation for looking at an unbounded number of deleterious mutations

In order to prove quasi-stationarity results, the case where $d < \infty$ can be treated more easily and provides an introduction to the case $d = \infty$. Nonetheless, the arguments for having a convergence at a given rate becomes more and more artificial as d tends to infinity : the constant involved in the Harnack inequalities that are crucial for our proof go to zero as the dimension increase. By considering the case $d = \infty$, we actually handle as a whole the case where d is sufficiently large. By these means, we are able to prove that the rate of convergence can be upper-bounded by a quantity that does not depend on the specific value of d. This is to be expected since, even when a large number of deleterious mutations is permitted, we expect individuals carrying a large number of mutations to remain negligible.

Referring for instance to [EPW09], it is not difficult to prove that in the deterministic limit of large population, the empirical measure of the number of mutations in the population tends to a Poisson distribution. The tail of the distribution is quickly disappearing. This deterministic limit corresponds to a limiting time-change of equation (5.1.1 : $S^{(d)}$) of the form $t' = t/\epsilon$ with $\alpha = \alpha'/\epsilon$, $\lambda = \lambda'/\epsilon$ as ϵ tends to 0. The Poisson distribution has a mean of $\lambda'/\alpha' = \lambda/\alpha$ so that it may be possible to quantify much more precisely as we do the threshold in the number of deleterious mutations after which differentiating individuals is not so crucial. This could make it possible to obtain quantitative estimates from our results on the extinction rate and the convergence rate in a context of very large population (in the vicinity of the deterministic limit). In the spirit of [CCM16], one could also specify asymptotic results on the concentration of the QSD.

5.2 Exponential quasi-stationarity results

5.2.1 Exponential quasi-stationarity

The conclusions of the following Theorems are expressed in terms of the notion of exponential quasi-stationarity that we describe now. As in [Chapter 2], for a process X on $\mathcal{X}$ with extinction τ_∂ to be exponentially quasi-stationary with a unique QSD ν means the following :

(i) Exponential convergence to the survival capacity :
With ρ_0 the extinction rate associated to the QSD, let, for any $x \in \mathcal{X}$:

$$h_t(x) \ := e^{\rho_0\, t}\mathbb{P}_x(t < \tau_\partial).$$

This sequence of functions converges exponentially in the uniform norm to a function h, that we call the survival capacity. h is positive and bounded. It also belongs to the domain of the infinitesimal generator $\mathcal{L}$, associated with the semi-group (P_t) (acting on the set $B(\mathcal{X})$ of bounded measurable function from $\mathcal{X}$ to $\mathbb{R}$ with the supremum norm $\|.\|_\infty$) and is actually the eigenvector of $\mathcal{L}$ with eigenvalue $-\rho$

$$\text{with } \mathcal{L}\, h = -\rho_0\, h, \qquad \textit{so} \quad \forall\, t \geq 0, \quad P_t\, h = e^{-\rho_0\, t} h.$$

(ii) Exponential convergence to the QSD :
For some $\gamma, C > 0$ and with the survival capacity h :

$$\forall \mu \in \mathcal{M}_1(\mathcal{X}),\ \forall t > 0, \qquad \| \mathbb{P}_\mu [X_t \in dx \mid t < \tau_\partial] - \nu(dx) \|_{TV} \leq C \frac{\inf_{u>0} \|\mu - u\nu\|_{TV}}{\langle \mu \,\big|\, h \rangle} e^{-\gamma t}. \tag{5.2.1}$$

The relation between ν and its extinction rate ρ_0 is the following :

$$\forall t \geq 0, \quad \nu P_t = e^{-\rho_0 t} \nu, \qquad \text{and in particular } \mathbb{P}_\nu(t < \tau_\partial) = e^{-\rho_0 t} \tag{5.2.2}$$

This relation is what characterize ν as a QSD since it implies that for any $t \geq 0$, $\mathbb{P}_\nu [X_t \in dx \mid t < \tau_\partial] = \nu(dx)$. There is an additional related notion when we wish to describe the behavior of the process with the requirement of a long inter-click interval.

(iii) Existence of the Q-process and its associated transition kernel :
There exists a family $(\mathbb{Q}_x)_{x \in \mathcal{X}}$ of probability measures on Ω defined by :

$$\lim_{t\to\infty} \mathbb{P}_x(\Lambda_s \,\big|\, t < \tau_\partial) = \mathbb{Q}_x(\Lambda_s) \tag{5.2.3}$$

for any $\mathcal{F}_s$-measurable set Λ_s. The process $(\Omega; (\mathcal{F}_t)_{t\geq 0}; (X_t)_{t\geq 0}; (\mathbb{Q}_x)_{x\in\mathcal{X}})$ is a homogeneous strong Markov process. Its transition kernel is given by :

$$q(x; t; dy) = e^{\rho_0 t} \frac{h(y)}{h(x)} p(x; t; dy), \tag{5.2.4}$$

where $p(x; t; dy)$ is the transition kernel of the Markov process (X) under $(\mathbb{P}_x)$.

(iv) Exponential ergodicity of the Q-process : :
The probability measure $\beta(dx) := h(x)\,\nu(dx)$ is the unique invariant distribution of X under $\mathbb{Q}$. Moreover, with the same constants $\gamma > 0$ and C as in (5.2.1) :

$$\|\mathbb{Q}_\mu [X_t \in dx] - \beta(dx)\|_{TV} \leq C \inf_{u>0} \|\mu - u\,\beta\|_{1/h}\, e^{-\gamma t}, \tag{5.2.5}$$

$$\text{where } \mathbb{Q}_\mu(dw) := \textstyle\int_{\mathcal{X}} \mu(dx)\, \mathbb{Q}_x(dw), \quad \|\mu\|_{1/h} := \| \dfrac{\mu(dx)}{h(x)} \|_{TV}$$

5.2.2 The discrete population case

Let $N \geq 1$ be the number of individuals in the population, and $D_n(t)$ for $n \leq N$ and $t \in \mathbb{Z}_+$ be the number of mutations carried by the n-th individual. We do not care about the precise numbering of the individuals and consider the empirical measure process defined as follow :

$$\mathcal{Z}_t^N := (1/N) \textstyle\sum_{n \leq N} \delta_{D_n(t)}, \tag{5.2.6}$$

so that $\mathcal{Z}_t^N(i) \in [\![0, N]\!]$ specifies the number of individuals with exactly i mutations (since everything is discrete, we identify $\mathcal{Z}_t^N$ as a function on $\mathbb{Z}_+$). From the rules describing the next generation from the previous one, $\mathcal{Z}^N$ is clearly a Markov process evolving on $\mathcal{M}_1^N(\mathbb{Z}_+)$, where :

$$\mathcal{M}_1^N(\mathbb{Z}_+) := \{(1/N) \textstyle\sum_{i \leq N} \delta_{d_i} \,;\, d_i \in \mathbb{Z}_+\} \equiv \{z : \mathbb{Z}_+ \mapsto (1/N) \times [\![0, N]\!] \,;\, \sum_{i \in \mathbb{Z}_+} z(i) = 1\}. \tag{5.2.7}$$

The extinction time under consideration comes from the extinction of the fittest individuals, i.e. :

$$\tau_\partial := \inf\{t \geq 0 \ ;\ \mathcal{Z}_t^N(0) = 0\} = \inf\{t \geq 0 \ ;\ \mathcal{Z}_t^N \notin \mathcal{M}_1^{(0),N}(\mathbb{Z}_+)\} \tag{5.2.8}$$

$$\text{where } \mathcal{M}_1^{(0),N}(\mathbb{Z}_+) = \{z \in \mathcal{M}_1^N(\mathbb{Z}_+), z(0) \geq 1/N\}. \tag{5.2.9}$$

The process $\bar{\mathcal{Z}}_t^N := \mathcal{Z}_t^N \mathbf{1}_{\{t < \tau_\partial\}} + \partial \mathbf{1}_{\{\tau_\partial \leq t\}}$ can easily be seen as a Markov process. For the process $\bar{\mathcal{Z}}^N$ living on $\mathcal{M}_1^{(0),N}(\mathbb{Z}_+) \cup \{\partial\}$, τ_∂ is the hitting time of the absorbing state ∂ (the cemetery), so that the results of quasi-stationarity apply to this context. In the following, we will not care about this distinction and simply consider Z^N and consider τ_∂ as its extinction time.

Theorem 5.2.1. *Consider for any N the Markov process Z^N whose transitions are prescribed as in Section 5.1.2.1, with extinction time τ_∂. Then, we have exponential quasi-stationarity as described in Section 5.2.1, with a positive lower-bound of the survival capacity h, thus a uniform convergence to ν and β in resp.* (5.2.1) *and* (5.2.5). *In particular, ν is the unique QSD.*

Remarks 5.2.2.1. *The proof of this Theorem strongly exploits the arguments given in the proofs in [CV16] and generalized in [Chapter 1]. It provides an elementary understanding of how we exploit in [Chapter 1] the criteria of persistence* $(A3)$ *that we will introduce in the proofs of Theorem 5.2.2 and 5.2.3.*

5.2.3 The finite dimensional case

In this section, we denote by $\tau_\partial^{(d)}$ the clicking time of the process $X^{(d)}$ solution of the system ($5.1.1 : S^{(d)}$), that is :

$$\tau_\partial^{(d)} := \inf\{t \geq 0 \ ;\ X_0^{(d)}(t) = 0\}.$$

The system of SDE then evolves for finite d on :

$$\mathcal{X}_d := \{(x_k)_{k \in [\![0,d]\!]} \in [0,1]^{d+1} \ ;\ \textstyle\sum_{k=0}^d x_k = 1\}$$

Theorem 5.2.2. *Consider the system of SDEs* ($5.1.1 : S^{(d)}$) *for any* $d \in \mathbb{N}$*, with extinction time* $\tau_\partial^{(d)}$*. We have exponential quasi-stationarity as described in Section 5.2.1 with a uniform convergence towards the unique QSD* $\nu^{(d)}$ *for initial conditions at a lower-bounded distance from the boundary.*

The proof of this theorem applies quite directly the ideas that we have previously exploited in [Chapter 1]. As in [CV17b], we rely mainly on the Harnack inequality to ensure that the dependency on the initial condition gets progressively forgotten and to compare the survival from different initial conditions. In addition, we shall be cautious in the way we handle together the absorbing and repulsive boundary conditions.

Moreover, it is then not too difficult to obtain the following controls on the moments of the QSDs $\nu^{(d)}$, for $d \in \mathbb{N}$. Notably, it proves the tightness of the sequence. Given the following theorem, the sequence $(\nu^{(d)})$ is expected to converge as $d \to \infty$ to the unique QSD $\nu^{(\infty)}$ for the infinite system (for which the control extends). This control on the moments is actually crucial for the proof of uniqueness.

Proposition 5.2.3.1. *For any* $k \geq 1$*, we have uniform tightness in* d *over the moments of order* k *of the unique QSDs* $\nu^{(d)}$ *associated to the solution of* ($5.1.1 : S^{(d)}$)*, i.e. :*

$$\sup_{d\in\mathbb{N}} \int_{\mathcal{X}_d} \nu^{(d)}(dx) \mathbf{1}_{\{M_k(x) \geq m_k\}} \to 0 \text{ as } m_k \to \infty \qquad \text{where } M_k(x) := \sum_{i \in [\![0,d]\!]} i^k\, x_i.$$

In particular, the sequence $\hat{\nu}_d$*, where the remaining components are put to 0, is tight in* $\mathcal{M}_1(\mathbb{R}_+^{\mathbb{Z}})$*.*

5.2.4 The infinite dimensional case

We consider now the infinite dimensional case, for which we require existence of moments. Let

$$\mathcal{X}^\eta := \left\{ (x_k)_{k\in\mathbb{Z}_+} \in [0,1]^\infty \ ; \ \sum_{k=0}^{\infty} x_k = 1 \ , \ \sum_{k=0}^{\infty} k^\eta x_k < \infty \right\}$$

Thanks to Theorem 3 in [AP13], we know that for any initial condition x that belongs to $\mathcal{X}^\eta$, for some $\eta > 2$, $(S^{(\infty)})$ has a unique weak solution which is a. s. continuous with values in $\mathcal{X}^\eta$. In practice, we will need to control moments of order η' strictly larger than 2 while exploiting the finiteness of moments of order $2\eta'$. For simplicity, we will thus restrict ourselves to $\mathcal{X}^6$ although our proof could certainly be adapted provided $\eta > 4$. Anyways, the core of our proof is based on the intuition that the slower the decay of the tail the more rapidly it gets erased and renewed. So we do not expect large tails to play a significant role.

Theorem 5.2.3. *Consider the system of SDEs* $(S^{(\infty)})$*, i.e. with* $d = \infty$*, defined on* $\mathcal{X}^6$ *with extinction time* $\tau_\partial^{(\infty)}$*. We have exponential quasi-stationarity as described* *in Section 5.2.1. Besides, there exist $C, \gamma, d_\wedge > 0$ such that for any $d \geq d_\wedge$, the convergences stated in (5.2.1), (5.2.5) and their counterpart for $h^{(\infty)}$ hold true with these constants for $(5.1.1 : S^{(d)})$.*

5.2.5 Discussion on these results

For the last process, no other parameter than α and λ are introduced. We deduce from Theorem 5.2.3 that the QSD and the survival capacity depend only on α and λ, as well as the values $C, \gamma > 0$ in (5.2.1) and (5.2.5). For the interpretation of our results, let us consider a new time-scale such that the mutation rate is set at 1. Then, the coefficient before the martingale term is $1/\lambda$. We recall that it then corresponds to $\sqrt{1/N_e}$, where N_e is the effective population size mentioned in Remark 5.1.2.2 and quantifies the relatedness of uniformly sampled individuals in the population. A large population size thus corresponds to letting λ go to infinity. This parameter $1/\lambda$ indicates in some way the level of fluctuations around the deterministic equilibrium that is a function of α/λ. As already noted by Haigh in [Ha78], α/λ is the average number of deleterious mutations that is established in the deterministic limit (neglecting neutral fluctuations).

From this notion of exponential quasi-stationarity, it is quite natural to define the expected time between clicks to be $t_C := \rho_0^{-1}$. On the other hand, our result shows that the following definition for the relaxation time would be meaningful :

$$t_R := \inf\{t_r > 0 \ ; \ \exists C > 0, \ \forall \mu, \quad \|\mu A_t - \nu\|_{TV} \leq (C/\langle \mu \,\big|\, h\rangle) \times e^{-t/t_r}\} \leq 1/\gamma. \tag{5.2.10}$$

For any $t_r > t_R$, we can deduce as in [Chapter 1] that the convergence to h and β would also occur at rate quicker than $1/t_r$. This kind of dependency in the initial condition is expected from the linearity of the semi-group (P_t) without renormalization. More general dependencies could nonetheless be imagined, relying for instance on Lyapunov functions as in [CV17c] or in [BCGM19]. We simply do not think it would change the value of t_R because the confinement is mainly due to extinction and immediate repulsion from the boundaries.

By relying on the arguments of Theorem 5.2.3 and Proposition 5.2.3.1, we expect that truncating the number of accumulated mutations is not likely to alter much this value of t_R provided the threshold is sufficiently large. Since we cannot evaluate t_R precisely and only provide an upper-bound, we do not claim it for sure. But substantial increase of these last components are proved to be rare from Proposition 5.2.3.1 and not so significant when we look at Section 5.6.7.

Provided $t_R \ll t_C$, we clearly expect to be in the quasi-stationary regime between clicks. It is classical that with the QSD as an initial condition, the extinction time and extinction state are independent, the former being exponentially distributed, as it has been established in Theorem 2.6 of [CMS13]. Assuming that we start the analysis at a new click after a long time-interval without click, it implies that the profile of mutations just after the click is distributed as the restriction of the QSD to the hyperplane $\{X_0 = 0\}$. Since having large values of M_1 makes it actually harder for the process to reach the hyperplane, we expect that, under the QSD restricted to $\{X_0 = 0\}$, M_1 tend to be rather smaller than the prediction $1 + \lambda/\alpha$ derived from the deterministic limit. Besides, the fittest individuals are altered by first changing into types with only one mutation. So we expect also that under this law, there is an over-representation in the proportion of individuals carrying a single mutation (the new optimal trait). Thus, we expect the distribution just after the click to be less prone to a future click than would be the QSD itself. Since $t_R \ll t_C$, the quasi-stationary regime is then rapidly reached. Expecting an exponential law for the inter-click intervals and the independence between them is thus fairly accurate.

Let us also imagine a dramatic situation where some clicks would rapidly succeed each others. Then, it would imply that these fittest classes of individuals are rapidly wiped off, while not letting much time for the others to change much. Since we have seen that we have very strong controls of moments under the QSD, such succession of clicks cannot hold for long : a class that is not prone to a quick extinction should be reached quite early and generate a new quasi-stationary regime. Such dramatic situations are thus expected to be very isolated and of limited impact, while of course very rare.

As we discuss in [Chapter 3], one can also conclude whether or not the QSD profile is likely to be observed without conditioning by comparing ν to the survival capacity h. If quasi-stationarity is stable, we do not expect that the conditioning on having a click in the far future shall substantially alter the dynamics. In most trajectories, the Q-process shall thus behave as the original process. So h should be mostly constant on the support of $\beta(dx) = h(x)\,\nu(dx)$, implying $h \approx 1$ where the density of ν is large. Yet, the QSD and the survival capacity are certainly quite difficult to specify with simulations because they live on a large dimensional space. Likewise, the convergence in total variation exploited in (5.2.10) is probably not very practical for numerical estimation. To be very accurate, the family of Wasserstein distances would be easier to deal with, while comparing different initial conditions by pairs. Namely, it corresponds to run independent simulations and compare the difference in the set of outcomes, depending on the initial condition. These distances are a priori dominated by the total variation distance, because it would be surprising to choose an unbounded distance for the comparison.

But the best indicator here is probably the decay in time of the correlations between X_0 and the other components. This means evaluating for instance

$$\frac{\mathbb{E}_\nu[X_k(t)X_0(0)] - \mathbb{E}_\nu[X_k(0)] \times \mathbb{E}_\nu[X_0(0)]}{\sqrt{\mathbb{E}_\nu[X_k(0)^2] - \mathbb{E}_\nu[X_k(0)]^2} \times \sqrt{\mathbb{E}_\nu[X_0(0)^2] - \mathbb{E}_\nu[X_0(0)]^2}},$$

for t small enough for extinction not to be significant, and k being either small $(1, 2, 3..)$ or close to $n_* := \lambda/\alpha$ where the QSD shall be concentrated.

We conjecture that, as long as extinction stays of negligible effect, the exponential decay that we expect to observe coincide with the one of the convergence in total variation of μA_t to the QSD. For a system where the coordinates are linked in such a specific way, this i however quite uncertain.

5.3 Proof of Theorem 5.2.1

The proof of Theorem 5.2.1 relies on the criteria presented in [Chapter 1]. We mainly require the two following propositions, whose proofs are deferred after we deduce Theorem 5.2.1. In this specific case the convergence is uniform. By exploiting [Chapter 1], we implicitly deduce the two criteria presented in [CV16].

Proposition 5.3.0.1. *For any $N \geq 1$, $\alpha \geq 0$, $\lambda > 0$ and $z \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+)$:*

$$\inf\left\{\mathbb{P}_{z_0}(Z^N(1) = z) \,\middle|\, z_0 \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+)\right\} > 0.$$

Proposition 5.3.0.2. *For any $N \geq 1$, $\alpha, \lambda > 0$ and $\epsilon > 0$, there exists $K \geq 1$ such that :*

$$\forall\, z \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+), \quad \mathbb{P}_z(Z^N(1) \notin E) \leq \epsilon,$$
$$\textit{where } E := \{z \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+) \;;\; z([\![K, \infty[\![) = 0\}$$

5.3.1 Proof of Theorem 5.2.1 with Propositions 5.3.0.1 and 5.3.0.2

We see that Proposition 5.3.0.1 clearly implies Assumption $(A1)$ of [Chapter 1], where the sequence $\mathcal{D}_\ell$ is here constant at $\mathcal{M}_1^{(0),N}(\mathbb{Z}_+)$, so that $(A0)$ is also satisfied. Next we prove $(A2)$, namely that for any $\rho > 0$, there exists E such that, with τ_E its first hitting time :

$$\sup_z \mathbb{E}_z(\exp[\rho(\tau_\partial \wedge \tau_E)]) < \infty.$$

This is elementary deduced from Proposition 5.3.0.2 though the Markov property and an induction over $k \geq 1$ to have upper-bound on $\mathbb{P}_z(k < \tau_\partial \wedge \tau_E)$. Then, for the last criterion $(A3)$, we remark that E as defined in Proposition 5.3.0.2 is finite. By Proposition 5.3.0.1 and the Markov property, we deduce that there exists $c > 0$ such that that for any $t \geq 1$:

$$\mathbb{P}_{\delta_0}(t < \tau_\partial) \geq c \sup_{z \in E} \mathbb{P}_z(t - 1 < \tau_\partial) \geq c \sup_{z \in E} \mathbb{P}_z(t < \tau_\partial).$$

This concludes $(A3)$ and that $(\mathcal{A})$ is satisfied. Theorem 5.2.1 is then deduced as an application of Theorems 2.1-3 of [Chapter 1]. □

Remarks 5.3.1.1. *If we were to impose mutations to occur one by one, Proposition 5.3.0.1 would still hold with the restriction of $z = \delta_0$, which is the only case we need. It would extend to any z provided we change the time 1 by the maximal number of mutations in z. The proof would not be much more difficult with overlapping generations, except that individuals should then be removed one by one. The proof of the equivalent of Proposition 5.3.0.2 would be slightly more difficult. The details are left to the reader.*

5.3.2 Proof of Proposition 5.3.0.1

We simply impose that all the individuals of the next generation are the offspring of an individual without any mutation, and prescribe the number of mutations that they get from the profile of z. We obtain a positive lower-bound uniform over any $z_0 \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+)$ by noticing that the probability of choosing a fittest individual as a parent is : $z_0(0)/(\sum_{i\geq 0} z_0(i)\times(1-\alpha)^i)$, which is necessarily larger than $1/N$. The number of mutations is then chosen independently of z_0, and there is indeed a positive probability that the sequence of independent Poisson distributed random variable has an empirical law distributed as z. This concludes the proof of Proposition 5.3.0.1. □

5.3.3 Proof of Proposition 5.3.0.2

We first prove that with a high probability, the sub-population of individuals carrying a large number of mutations leave no progeny. Let $K \geq 1$ for the threshold in the number of mutations. The probability that an individual chooses a parent with more than K mutations is upper-bounded by $N\times(1-\alpha)^K$, since $z(0)\geq 1$. For any $\epsilon > 0$, there exists indeed $K \geq 1$ such that, with a probability greater than $1-\epsilon/2$, no individual in the next generation descends from an individual with more than K mutations. Likewise, there exists $K' \geq 1$ such that, with a probability greater than $1-\epsilon/2$, the number of additional mutations is less than K' (for any individual, independently of the initial condition z). We deduce that :

$$\begin{aligned}&\forall\, z \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+), \quad \mathbb{P}_z(Z^N(1) \notin E) \leq \epsilon,\\ &\qquad \text{where } E := \{z \in \mathcal{M}_1^{(0),N}(\mathbb{Z}_+)\ ;\ z([\![K+K',\infty[\![) = 0\}\end{aligned}$$

which concludes the proof of Proposition 5.3.0.2. □

The proof of Theorem 5.2.1 is now complete.

5.4 Proof of Theorem 5.2.2

In the whole Section, d is fixed and this dependency is not closely considered in this section. On the contrary, in Sections 5.5 and 5.6, the dependency in d should not be forgotten. As a compromise between readability and the recall of this dependency, we have chosen to rather keep a subscript (d) for instance for the subsets $\mathcal{D}^{(d)}$, $E^{(d)}$, the irreducibility measure $\zeta^{(d)}$ and the probability laws $\mathbb{P}^{(d)}$, but to usually omit it for the process X and the associated stopping times.

5.4.1 Main properties leading to the proof

The proof of Theorem 5.2.2 relies also on the criteria presented in [Chapter 1], with here a non-uniform convergence. Let :

$$\mathcal{D}_\ell^{(d)} := \left\{ x = (x_i)_{0\le i\le d} \in [(2\ell d)^{-1}, 1]^d \ : 1 - \textstyle\sum_{0\le i\le d-1} x_i \ge (2\ell d)^{-1} \right\}, \tag{5.4.1}$$

which are of non-empty interior. The three following propositions state that X with extinction time τ_∂ satisfies the conditions given in [Chapter 1] to ensure exponential quasi-stationarity, as we show in Section 5.4.2 before the proof of each of them, in the order of appearance.

Proposition 5.4.1.1. *For any $d \in \mathbb{N}$ and $t > 0$, there exists $\zeta^{(d)} \in \mathcal{M}_1(\mathcal{X}_d)$ with support in $\mathcal{D}_2^{(d)}$ such that for any $\ell \ge 1$, there exists $c = c(d, \ell) > 0$ such that :*

$$\forall\, x \in \mathcal{D}_\ell^{(d)}, \quad \mathbb{E}_x^{(d)} \left(X_t \in dy \ ; \ t < \tau_\partial \wedge T_{\mathcal{D}_{\ell+1}^{(d)}} \right) \ge c\, \zeta^{(d)}(dy).$$

Proposition 5.4.1.2. *For any $d \in \mathbb{N}$ and $\ell \ge 1$:*

$$\limsup_{t\to\infty} \sup_{x,x' \in \mathcal{D}_\ell^{(d)}} \frac{\mathbb{P}_x^{(d)}\left(t < \tau_\partial\right)}{\mathbb{P}_{x'}^{(d)}\left(t < \tau_\partial\right)} < \infty.$$

Proposition 5.4.1.3. *For any $d \in \mathbb{N}$ and $\rho > 0$, there exists $\ell \ge 1$ such that :*

$$\sup_{x\in\mathcal{X}_d} \mathbb{E}_x^{(d)} \exp[\rho\, (\tau_{\mathcal{D}_\ell^{(d)}} \wedge \tau_\partial)] < \infty.$$

Remarks 5.4.1.1. *By Theorem 2.1 in [Chapter 1], we know also that there is for any $\ell \ge 1$ a positive lower-bound of $h^{(d)}$ on $\mathcal{D}_\ell^{(d)}$. With the notations of [Chapter 1], it implies that the multiplicative constant $C/\langle \mu \,\big|\, h^{(d)} \rangle$ in Theorem 5.2.2 is uniformly lower-bounded over the sets :*

$$\mathcal{M}_{\ell,\xi}^{(d)} := \left\{ \mu \in \mathcal{M}_1(\mathcal{X}_d) \ ; \ \mu\left(\mathcal{D}_\ell^{(d)}\right) \ge \xi \right\}, \qquad \text{with } \ell \ge 2, \xi \in (0, 1].$$

Remarks 5.4.1.2. *In our result, we do not distinguish much between the two types of boundaries, as can be seen in our definitions of $\mathcal{D}_\ell^{(d)}$ and $\mathcal{M}_{\ell,\xi}^{(d)}$. This is certainly not optimal for initial conditions close to the repulsive boundaries but far from the absorbing one. If one wishes to extend these results, it should not be too difficult to obtain an estimate of the form $\delta_x A_t^{(d)} \in \mathcal{M}_{\ell,\xi}^{(d)}$ where t, ℓ and ξ are uniform for x in $(\mathcal{X}_d \setminus \mathcal{D}_m^{(d)}) \cap \pi_0^{-1}(\epsilon, 1]$ with ϵ fixed and m possibly large enough. Indeed, we know from Proposition 5.4.1.3 that for m large, $\mathcal{X}_d \setminus \mathcal{D}_m^{(d)}$ is quickly escaped, while the extinction can be proven to be negligible in this time-scale provided x_0 is sufficiently large. This would imply that $h^{(d)}$ is actually lower-bounded on any $\pi_0^{-1}(\epsilon, 1]$ and give more insight on the convergence.*

5.4.2 Proof of Theorem 5.2.2 with these Propositions

It is easily seen from their definition in (5.4.1) that the sets $\mathcal{D}_\ell^{(d)}$ satisfy :

$$\forall \ell \geq 1,\ \mathcal{D}_\ell^{(d)} \subset int(\mathcal{D}_{\ell+1}^{(d)}) \quad \text{and} \quad \bigcup_{\ell \geq 1} \mathcal{D}_\ell^{(d)} = \mathcal{X}_d.$$

In addition, for any $\ell \geq 1$, $\mathcal{D}_\ell^{(d)}$ is a closed subset of $int(\mathcal{D}_{\ell+1}^{(d)})$ in $\mathcal{X}_d$. Assumption $(A0)$ in [Chapter 1] clearly holds true.

Propositions 5.4.1.1, 5.4.1.2 and 5.4.1.3 ensure respectively $(A1)$, $(A3)$ and $(A2)$ of the set of assumption (A) in [Chapter 1]. Notably, for $(A3)$, since $\zeta^{(d)}$ has support in $\mathcal{D}_2^{(d)}$, for any $\ell \geq 2$ (chosen given $\rho^{(d)} > 0$ by Proposition 5.4.1.3), by Proposition 5.4.1.2 :

$$\limsup_{t \to \infty} \sup_{x \in \mathcal{D}_\ell^{(d)}} \frac{\mathbb{P}_\zeta^{(d)}(t < \tau_\partial)}{\mathbb{P}_x^{(d)}(t < \tau_\partial)} < \infty.$$

Theorem 5.2.2 is then a consequence of Theorems 2.1-3 in [Chapter 1].

5.4.3 Harnack inequalities for Propositions 5.4.1.1 and 5.4.1.2

The proofs of Propositions 5.4.1.1 and 5.4.1.2 are actually similar to those in Subsection 4.2.2 of [Chapter 1]. They exploit the Harnack inequality –the following assumption (H)– classically deduced for such an elliptic diffusion.

In the following, we say that a process (Y_t) on $\mathcal{Y}$ with generator $\mathcal{L}$ (including possibly an extinction rate ρ_c) satisfies Assumption (H) if :

For any compact sets $K, K' \subset \mathcal{Y}$ with C^2 boundaries s.t. $K \subset int(K')$, $0 < t_1 < t_2$ and positive C^2 constraints : $u_{\partial K'} : (\{0\} \times K') \cup ([0, t_2] \times \partial K') \to \mathbb{R}_+$, there exists a unique strong non-negative solution $u(t, x) \in C^{1,2}((0, t_2) \times K', \mathbb{R}_+) \cap C^0([0, t_2] \times \overline{K'}, \mathbb{R}_+)$

to the Cauchy problem :

$$\partial_t u(t,x) = \mathcal{L}u(t,x) \text{ on } [0,t_2]\times K'\,;$$
$$u(t,x) = u_{\partial K'}(x) \text{ on } (\{0\}\times K')\cup([0,t_2]\times\partial K'),$$

and it satisfies, for some $C = C(t_1,t_2,K,K') > 0$ independent of $u_{\partial K'}$:

$$\inf_{x\in K} u(t_2,x) \geq C\, \sup_{x\in K} u(t_1,x).$$

On any $\mathcal{D}_n^{(d)}$, $\sigma^{(d)}$ is uniformly elliptic while $\sigma^{(d)}$ and $b^{(d)}$ are uniformly Lipschitz. Assumption (H) holds for the generator $\mathcal{L}^{(d)}$ of any finite dimensional process $X^{(d)}$, while restricted on some $\mathcal{D}_n^{(d)}$. This can be proved as in Section 4.2.2 of [Chapter 1] by referring first to Corollary 2 in Section 4, Chapter 3 of [Fr08] for the existence and uniqueness, then Theorem 1 in Chapter 2 of [Fr08] for the positivity, and finally Theorem 1.1 in [KS80] for the Harnack inequality itself.

5.4.4 Proof of Proposition 5.4.1.1

We apply assumption (H) to $u(t,y) := \mathbb{E}_y^{(d)}\left(f(Y_t)\ ;\ t<\tau_\partial^{n+1}\right)$, where f is any C^∞ function with support in $\mathcal{D}_n = K$, and $\tau_\partial^{n+1} := \inf\{t\geq 0 : X_t \notin \mathcal{D}_{n+1}^{(d)}\}$. It implies that for any $y\in\mathcal{D}_n^{(d)}$, $y_0\in\mathcal{D}_1^{(d)}$ and $0<t_0<t$:

$$\mathbb{E}_y^{(d)}\left(f(Y_t)\ ;\ t<\tau_\partial^{n+1}\right) \geq C_n\, \mathbb{E}_{y_0}^{(d)}\left(f(Y_{t_0})\ ;\ t_0<\tau_\partial^{2}\right).$$

Since $\inf_{y_0\in\mathcal{D}_1^{(d)}} \mathbb{P}_{y_0}^{(d)}\left(Y_t\in\mathcal{D}_1^{(d)}\ ;\ t<\tau_\partial^2\right) > 0$ (by choosing arbitrary y_0 and $t_0 = t/2$), we can obtain a probability measure $\zeta^{(d)}$ with support on $\mathcal{D}_2^{(d)}$, independent of n, s.t. (since C_n does not depend on f) :

$$\forall\, y\in\mathcal{D}_n^{(d)},\quad \mathbb{E}_y^{(d)}\left(Y_t\in dy\ ;\ t<\tau_\partial^{n+1}\right) \geq c_n\,\zeta^{(d)}(dy). \qquad \square$$

5.4.5 Proof of Proposition 5.4.1.2

The proof is a bit similar but more technical because the reference measure is now in the upper-bound, so that we can no longer neglect trajectories exiting $\mathcal{D}_{n+1}^{(d)}$. We can choose $t_1 := 1$ and find two compact sets $K, K'\subset\mathcal{Y}$ with C^2 boundaries s.t. $\mathcal{D}_n^{(d)}\subset K\subset int(K')\subset int(\mathcal{D}_{n+1}^{(d)})$. We want to approximate the function :

$$u(t,y) := \mathbb{E}_y^{(d)}\left(f(Y_t)\ ;\ t<\tau_\partial\right),\quad \text{with }\ t\geq t_1,\ y\in K'$$

defined for some $f\in\mathbb{C}^\infty(\mathcal{Y})$. Although we can prove (refering to Thm 5.1.15 in [Lu95]) that u is continuous, it is a priori not regular enough to apply the Harnack inequality directly. Thus, we approximate it on the parabolic boundary $[t_1,\infty)\times\partial K'$

$\bigcup \{t_1\} \times K'$ by the family $(U_k)_{k\geq 1}$ of smooth $-\mathcal{C}_+^\infty$ w.l.o.g.– functions. We then deduce approximations of u in $[t_1, \infty) \times K'$ by (smooth) solutions of :

$$\begin{aligned} \partial_t u_k(t,y) - \mathcal{L} u_k(t,y) = 0, \qquad & t \geq t_1,\ y \in int(K') \\ u_k(t,y) = U_k(t,y), \quad t \geq t_1,\ y \in \partial K', \quad & \text{or } t = t_1,\ y \in K'. \end{aligned}$$

By Assumption (H), the constant involved in the Harnack inequality does not depend on the values on the boundary. Thus, it applies with the same constant for the whole family of approximations u_k. We refer to the proof in [CV17b], Section 4, step 4, to state that the Harnack inequality then extends to the approximated function u, where the regularity of $u \in C^{1,2}$ is required to apply the Itô formula on the process $u(t-s, X_s)$. With $t_2 := 2$ and $t_3 := 3$, we deduce that there exists $C_n > 0$ such that for any $f \in \mathbb{C}^\infty(\mathcal{Y})$:

$$\forall\, y, y' \in \mathcal{D}_n^{(d)},\ \mathbb{E}_y^{(d)} \left(f(Y_{t_2}) \ ;\ t_2 < \tau_\partial \right) \leq C_n\, \mathbb{E}_{y'}^{(d)} \left(f[Y(t_3)] \ ;\ t_3 < \tau_\partial \right)$$

It thus extends to any measurable and bounded f. Applying this result to $f(t,y) := \mathbb{P}_y^{(d)}(t - t_2 < \tau_\partial)$, and applying the Markov property :

$$\forall\, y, y' \in \mathcal{D}_n^{(d)},\ \forall\, t \geq t_2, \quad \mathbb{P}_y^{(d)} \left(t < \tau_\partial \right) \leq C_n\, \mathbb{P}_{y'}^{(d)} \left(t + t_3 - t_2 < \tau_\partial \right) \leq C_n\, \mathbb{P}_{y'}^{(d)} \left(t < \tau_\partial \right)$$

concluding the proof of Proposition 5.4.1.2. $\square$

5.4.6 Proof of Proposition 5.4.1.3

Proposition 5.4.1.3 is proved by recursively ensuring that the k-th first coordinates have escaped from the value 0. For any $y > 0$, and $0 \leq k \leq d$, let :

$$T_y^k := \inf\{s \geq 0 \ ;\ \forall j \leq k,\ X_j(s) \geq y\}.$$

The proof of Proposition 5.4.1.3 is achieved below with the help of the two following lemmas.

Lemma 5.4.6.1. *For any $d \in \mathbb{N}$ and $t > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < \tau_\partial \right) \middle| x \in \mathcal{X}_d, x_0 \leq y_0 \right\} \to 0 \ \text{as}\ y_0 \to 0.$$

This Lemma is a trivial consequence of the fact that X_0 is upper-bounded by the solution Y of :

$$dY_t = \alpha\, d\, Y_t dt + \sqrt{Y_t(1-Y_t)} dB_0(t)\ ,\ Y_0 = y_0,$$

for which it is known that 0 is an absorbing value.

Lemma 5.4.6.2. *For any $y \in (0,1), \epsilon > 0$ and $1 \leq k \leq d$, there exists $t > 0$, $y' \in (0,1)$ such that :*

$$\inf \left\{ \mathbb{P}_x^{(d)} \left(T_{y'}^k < t \wedge \tau_\partial \right) \middle| x \in \mathcal{X}_d, \forall\, j \leq k-1,\ x_j \geq y \right\} \geq 1 - \epsilon.$$

The proof of this Lemma is deferred to Section 5.7.0.1 of the Appendix.

Given $\rho > 0$, let $t_0 := \log(2)/\rho$. We can find $y_0 \in (0,1)$ by Lemma 5.4.6.1 such that for any $x = (x_i)$ satisfying $x_0 \leq y_0$, it holds :

$$\mathbb{P}_x^{(d)} \left(t_0 < \tau_\partial \right) \leq \exp[-\rho\, t_0]/2 = 1/4.$$

By the Markov property and an induction, for any $k \geq 1$:

$$\begin{aligned}
\mathbb{P}_x^{(d)}(k\, t_0 \leq \tau_\partial \wedge T_{y_0}^0) &\leq \mathbb{P}_x^{(d)}(k\, t_0 \leq \tau_\partial\,,\ (k-1)\, t_0 \leq T_{y_0}^0) \leq 1/4^k, \\
\sup_x \mathbb{E}_x^{(d)} \exp[\rho\, (T_{y_0}^0 \wedge \tau_\partial)] &\leq \sum_{k \geq 0} e^{\rho t_0 [k+1]}\, \mathbb{P}_x^{(d)}(k\, t_0 \leq \tau_\partial \wedge T_{y_0}^0) \\
&\leq \sum_{k \geq 0} 2^{k+1}/4^k = 4 < \infty.
\end{aligned}$$

By the Markov property, Lemma 5.4.6.2 and by induction on $0 \leq k \leq d$, there exists y_k such that, on the event $\{T_{y_0}^0 \leq \tau_\partial\}$ and with $\epsilon = 1/8$:

$$\mathbb{P}_x^{(d)} - \text{a.s.} \quad \mathbb{P}_x^{(d)} \left(T_{y_k}^k \leq (T_{y_0}^0 + k\, t_0/d) \wedge \tau_\partial \,|\mathcal{F}_{T_{y_0}^0} \right) \geq 1 - k/(8d). \tag{5.4.2}$$

To deduce this induction, note that without loss of generality, we can choose t_k as small as needed, in particular $t_k \leq t_0/d$, when we apply Lemma 5.4.6.2 for $T_{y_k}^k$. The probability is indeed decreasing with t.

Then, for some large value of $t > 0$, let $V_t := \tau_\partial \wedge T_{y_d}^d \wedge t$, and consider

$$E_t^{(d)} := \sup_x \mathbb{E}_x^{(d)} \exp[\rho V_t] < \infty$$

For any x such that $x_0 \geq y_0$, we deduce from the Markov property :

$$\begin{aligned}
\mathbb{E}_x^{(d)} \exp[\rho V_t] &\leq e^{\rho t_0} \left(1 + \mathbb{E}_x^{(d)} [\mathbb{E}_{X(t_0)}^{(d)} \exp[\rho \widetilde{V}_t]\ ;\ t_0 < V_t]\right) \\
&\leq 2 \left(1 + E_t^{(d)} \times [1 - \mathbb{P}_x^{(d)}(T_{y_d}^d \leq t_0 \wedge \tau_\partial)]\right) \\
&\leq 2 + E_t^{(d)}/4,
\end{aligned}$$

where we used inequality (5.4.2). On the other hand, for any general x :

$$\begin{aligned}\mathbb{E}_x^{(d)} \exp[\rho V_t] &\leq \mathbb{E}_x^{(d)} \left(\exp[\rho(T_{y_0}^0 \wedge \tau_\partial)]; V_t \leq T_{y_0}^0\right) \\ &\qquad + \mathbb{E}_x^{(d)} \left(\exp[\rho T_{y_0}^0]\, \mathbb{E}_{X(T_{y_0}^0)}^{(d)} \exp[\rho \widetilde{V}_t] \ ; \ T_{y_0}^0 < V_t\right) \\ &\leq (2 + E_t^{(d)}/4) \times \mathbb{E}_x^{(d)} \left(\exp[\rho(T_{y_0}^0 \wedge \tau_\partial)]\right) \\ &\leq 4 + E_t^{(d)}/2,\end{aligned}$$

where we used the previous estimate with the fact that $X_0(T_{y_0}^0) \geq y_0$. Taking the supremum over x, and since $E_t^{(d)} < \infty$, we deduce : $E_t^{(d)} \leq 8$.

The limit where $t \to \infty$ ensures $\sup_x \mathbb{E}_x^{(d)} \exp[\rho(T_{y_d}^d \wedge \tau_\partial)] \leq 8$. This concludes the proof of Proposition 5.4.1.3. □

5.5 Proof of Proposition 5.2.3.1

The proof of Proposition 5.2.3.1 relies on uniform estimates that descent from large values of the moment happen quickly while a sudden and too large increase of the moment is unlikely even for large d. It is proved in Section 5.5.1 thanks to the two following propositions.

Proposition 5.5.0.1. *For any $t > 0$ and $k \geq 1$, with $T_k(m) := \inf\{t \geq 0\ ;\ M_k(X_t) \leq m\}$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < \tau_\partial \wedge T_k(m)\right) \middle| d \in \mathbb{N}, x \in \mathcal{X}_d \right\} \underset{m \to \infty}{\longrightarrow} 0.$$

Proposition 5.5.0.2. *For any $k, t, \epsilon, m > 0$, there exists $m' > 0$ such that for any $d \in \mathbb{N}$ and initial condition $x \in \mathcal{X}_d$ such that $M_k(x) \leq m$:*

$$\mathbb{P}_x^{(d)} \left(\sup_{s \leq t} M_k(X_s) \geq m'\right) \leq \epsilon.$$

We will first prove Proposition 5.2.3.1 thanks to these two propositions, then Proposition 5.5.0.2 and finally prove the one of Proposition 5.5.0.1. This last proof relies on 3 steps of descent, the last one being iterated for each moment between 2 and k. The main result for each of these steps is given by the 3 following lemmas, whose proofs are deferred to Appendix 5.7 :

Lemma 5.5.0.3. *For any $t > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < \tau_\partial\right) \middle| d \in \mathbb{N}, x \in \mathcal{X}_d, M_1(x) \geq 1, x_0\, M_1(x) \leq \delta \right\} \to 0 \ \textit{as}\ \delta \to 0.$$

Lemma 5.5.0.4. *Given any $t, y_0 > 0$:*

$$\sup\left\{\mathbb{P}^{(d)}_x\left(t \le T_1(m_1)\wedge\tau_\partial\right)\middle| d\in\mathbb{N}, x\in\mathcal{X}_d, x_0\ge y_0\right\}\underset{m_1\to\infty}{\longrightarrow}0.$$

Lemma 5.5.0.5. *Given any $k \ge 1$ and t, $m_k > 0$:*

$$\sup\left\{\mathbb{P}^{(d)}_x\left(t \le T_{k+1}(m_{k+1})\wedge\tau_\partial\right)\middle| d\in\mathbb{N}, x\in\mathcal{X}_d, M_k(x)\le m_k\right\}\underset{m_{k+1}\to\infty}{\longrightarrow}0.$$

5.5.1 Proof of Proposition 5.2.3.1

First of all, we show that we have a uniform upper-bound on the extinction rate : $\sup_{d\in\mathbb{N}}\rho_d < \infty$. Indeed, whatever $d\in\mathbb{N}$, we can find $x^{(d)}\in\mathcal{X}_d$ such that $x^{(d)}_0 \ge 1/2$ so that X_0 under $\mathbb{P}^{(d)}_x$ is lower-bounded by the solution Y to :

$$dY_s = -\lambda\, ds + \sqrt{Y_s\,(1-Y_s)}dB_0(s)\,,\qquad Y(0)=1/2.$$

Note that the boundary $y = 1$ is entrance for this process so that it exits $(0,1)$ only through 0, cf e.g. Subsection 3.3.3 in [JT17]. The semi-group governing Y, with extinction at τ^Y_0, corresponds exactly to the system ($5.1.1 : S^{(d)}$) with $d = 1$, $\alpha = 0$, $X'_0 = Y$ and $X'_1 = 1 - Y$. We know from Theorem 5.2.2 that it is exponentially quasi-ergodic with extinction rate $\rho_\vee$. Denoting $\mathbb{P}^{(Y)}_{1/2}$ the law of Y, we deduce from the convergences of the survival capacities :

$$\rho_d = \lim_{t\to\infty}\tfrac{-1}{t}\log\mathbb{P}^{(d)}_{x^{(d)}}(t<\tau_\partial) \le \lim_{t\to\infty}\tfrac{-1}{t}\log\mathbb{P}^{(Y)}_{1/2}(t<\tau^Y_0) := \rho_\vee.$$

Thanks to Proposition 5.5.0.1, we can find m such that $T_k(m)$ satisfies :

$$\sup_{d\ge1}\sup_{x\in\mathcal{X}_d}\mathbb{E}^{(d)}_x\exp[(\rho_\vee+1)\,(T_k(m)\wedge\tau_\partial)] \le C < \infty.$$

In particular, it implies that for any $t > 0$ and $d\in\mathbb{N}$:

$$\mathbb{P}^{(d)}_{\nu^{(d)}}(t<T_k(m)\wedge\tau_\partial)\le C\,\exp[-(\rho_\vee+1)\,t] \tag{5.5.1}$$

Then, for any $\epsilon > 0$, consider $t := -\log(\epsilon/(2\,C))$. Thanks to Proposition 5.5.0.2, we can find some $m' > 0$ such that for any initial condition x such that $M_k(x)\le M$:

$$\mathbb{P}^{(d)}_x\left(\sup_{s\le t}M_k(X_s)\ge m'\right)\le \epsilon/2\,\exp[-\rho_\vee\,t]. \tag{5.5.2}$$

Then, for any $d \in \mathbb{N}$: (with the QSD $\nu^{(d)}$)

$$\begin{aligned}\nu^{(d)}(\{M_k \geq m'\}) &= \exp[\rho^{(d)}\, t]\, \mathbb{P}^{(d)}_{\nu^{(d)}}(M_k(X_t) \geq m'\, ;\, t \leq \tau_\partial)\\ &\leq \exp[\rho_\vee\, t] \left(\mathbb{P}^{(d)}_{\nu^{(d)}}(T_k(m) > t\, ;\, t < \tau_\partial) \right.\\ &\qquad \left. + \mathbb{E}^{(d)}_{\nu^{(d)}} \left[\mathbb{P}^{(d)}_{X(T_k(m))} \left(\sup_{s \leq t} M_k(X_s) \geq m' \right) ;\, T_k(m) < t \wedge \tau_\partial \right] \right)\\ &\leq C \times \epsilon/(2\, C) + \epsilon/2 \leq \epsilon,\end{aligned}$$

by inequalities (5.5.1), (5.5.2) and the definition of t. □

Remark : We see quite naturally with Lemma 5.5.0.5 that we are in fact able to bound any moment with large probability under the QSD $\nu^{(d)}$ uniformly on $d \in \mathbb{N}$. These of course will extend to the limiting QSD on $[0, 1]^{\mathbb{Z}}$.

5.5.2 Proof of Proposition 5.5.0.2

Let $k, t > 0$,$m' \geq m$, $d \in \mathbb{N}$ and $x \in \mathcal{X}_d$ such that $M_k(x) \leq m$ be fixed.
We consider the semi-martingale decomposition of M_k :

$$dM_k(t) = V_k(t)\, dt + d\mathcal{M}_k(t), \tag{5.5.3}$$

where $\mathcal{M}_k$ is a continuous martingale starting from 0, whose quadratic variation is

$$\langle \mathcal{M}_k \rangle_t = \int_0^t (M_{2\,k}(s) - M_k(s)^2)\, ds,$$

and V_k is a bounded variation process defined as :

$$V_k := \alpha(M_1 \times M_k - M_{k+1}) + \lambda \sum_{\ell=0}^{d-1} (\ell + 1)^k X_\ell - \lambda(M_k - d^k X_d). \tag{5.5.4}$$

Thanks to the Hölder inequality, considering a random variable Y such that $Y = j$ with probability X_j :

$M_1 = \mathbb{E}(Y) \leq \mathbb{E}(Y^{k+1})^{1/(k+1)}$; $M_k = \mathbb{E}(Y^k) \leq \mathbb{E}(Y^{k+1})^{k/(1+k)}$ thus $M_1 \times M_k \leq M_{k+1}$.

Exploiting also that $(\ell+1)^k \leq 2^k \times \ell^k$ for $\ell \geq 1$, we deduce, with $C = C(k) = \lambda(2^k - 1)$:

$$V_k \leq C\, M_k + \lambda. \tag{5.5.5}$$

To obtain an upper-bound on the probability that $\sup_{s \leq t} M_k(s)$ is large, we want to exploit the Doob inequality on a non-negative sub-martingale that is an upper-bound of M_k. It leads us to consider the solution of the following equation :

$$\widehat{M}_k(t) := m + \lambda t + C \int_0^t \widehat{M}_k(s)\, ds + \mathcal{M}_k(t), \tag{5.5.6}$$

because classical results of comparison then implies that for any $t \geq 0$, $\widehat{M}_k(t) \geq M_k(t)$, see for instance Proposition 3.12 in [PR14]. The fact that $\widehat{M}_k$ is non-negative comes from the fact that M_k is non-negative. As a solution to equation (5.5.6), $\widehat{M}_k$ is clearly a sub-martingale. Since it is upper-bounded by d^k, we can also apply the Gromwall Lemma to deduce that for any initial condition x such that $M_k(x) \leq m$:

$$\mathbb{E}_x^{(d)}\left[\sup_{s\leq t} \widehat{M}_k(s)\right] \leq (m + \lambda t)\, e^{Ct}. \tag{5.5.7}$$

By exploiting Doob's inequality on $\widehat{M}_k$, then inequality (5.5.7) with $C_M := (1 + \lambda t)\, e^{Ct}$, we obtain :

$$\begin{aligned}\mathbb{P}_x^{(d)}(\sup_{s\leq t} M_k(s) > m') &\leq \mathbb{P}_x^{(d)}(\sup_{s\leq t} \widehat{M}_k(s) > m')\\ &\leq \frac{\mathbb{E}_x^{(d)}[\widehat{M}_k(t)]}{m'} \leq \frac{C_M\, m}{m'}.\end{aligned}$$

This concludes the proof of Proposition 5.5.0.2. □

5.5.3 Proof of Proposition 5.5.0.1

Let $t, \epsilon > 0$. From Lemma 5.5.0.3, we can find $\delta > 0$ and $m_1^\vee \geq (2\,\lambda)/\alpha$ such that :

$$\sup\left\{\mathbb{P}_x^{(d)}\left(t < \tau_\partial\right) \middle| d \in \mathbb{N}, x \in \mathcal{X}_d, M_1(x) \geq m_1^\vee, x_0\, M_1(x) \leq \delta\right\} \leq \epsilon. \tag{5.5.8}$$

Recall $T_1(m_1) := \inf\{t \geq 0\ ;\ M_1(X_t) \leq m_1\} \leq T_1(m_1^\vee)$ for any $m_1 \geq m_1^\vee$. The value of m_1 will be fixed in (5.5.12), but we first need to prove that with a probability close to 1, X_0 has escaped from the boundary $x_0 = 0$ provided that M_1 has not been small. Let :

$$T_{01}(\delta) := \inf\{t \geq 0\ ;\ X_0(t)\, M_1(X_t) \leq \delta\}, \tag{5.5.9}$$

By the Markov property, we deduce from (5.5.8) that for any $x \in \mathcal{X}_d$ and $d \geq 1$:

$$\mathbb{P}_x^{(d)}\left(T_{01}(\delta) \leq t \leq T_1(m_1^\vee)\ ,\ 2\,t < \tau_\partial\right) \leq \epsilon. \tag{5.5.10}$$

Recalling that $m_1^\vee \geq (2\,\lambda)/\alpha$, on the event $\{t \leq T_{01}(\delta) \wedge T_1(m_1^\vee) \wedge \tau_\partial\}$, we deduce that for any $s \leq t$: $(\alpha M_1(s) - \lambda) X_0(s) \geq \alpha\, \delta/2$. Thus, a.s. on this event, X_0 is lower-bounded by the solution Y to :

$$dY_s = \alpha\,\delta/2\,ds + \sqrt{Y_s\ (1-Y_s)}dB_0(s)\ , \qquad Y(0) = 0. \tag{5.5.11}$$

Note that Y is independent of d and x and 0 is an entrance boundary for Y, cf e.g. Subsection 3.3.3 in [JT17]. So we can find $y_0 > 0$ such that : $\quad \mathbb{P}(Y(t) \leq y_0) \leq \epsilon$. This implies that for any d and x :

$$\begin{aligned}\mathbb{P}^{(d)}_x(X_0(t) \leq y_0\ ,\ t \leq T_1(m_1^\vee)\ ,\ 2\,t < \tau_\partial) &\leq \mathbb{P}^{(d)}_x(T_{01}(\delta) \leq t \leq T_1(m_1^\vee)\ ,\ 2\,t < \tau_\partial)\\ &\quad + \mathbb{P}^{(d)}_x(Y(t) \leq y_0\ ,\ t \leq T_{01}(\delta) \wedge T_1(m_1^\vee) \wedge \tau_\partial)\\ &\leq 2\,\epsilon.\end{aligned}$$

From Lemma 5.5.0.4, we can find some $m_1 \geq m_1^\vee$ associated to y_0. We thus deduce, with the Markov property at time t :

$$\begin{aligned}\mathbb{P}^{(d)}_x(2\,t < T_1(m_1) \wedge \tau_\partial) &\leq \mathbb{P}^{(d)}_x(X_0(t) \leq y_0\ ,\ t \leq T_1(m_1^\vee)\ ,\ 2\,t < \tau_\partial)\\ &\quad + \mathbb{E}^{(d)}_x[\mathbb{P}^{(d)}_{X(t)}(t \leq \widetilde{T}_1(m_1) \wedge \widetilde{\tau_\partial})\ ;\ X_0(t) \geq y_0]\\ &\leq 3\,\epsilon.\end{aligned} \tag{5.5.12}$$

From Lemma 5.5.0.5, we can find some $m_2 > 0$ associated to m_1 (with $T_2(m_2) := \inf\{t\ \geq\ 0\ ;\ M_2(X_t) \leq m_2\}$). and some $m_3 > 0$ associated to m_2 such that :

$$\mathbb{P}^{(d)}_x(3\,t < T_2(m_2) \wedge \tau_\partial) \leq 4\,\epsilon\ , \qquad \mathbb{P}^{(d)}_x(4\,t < T_3(m_3) \wedge \tau_\partial) \leq 5\,\epsilon.$$

More generally, we can prove inductively that there exists m_k such that :

$$\mathbb{P}^{(d)}_x((k+1)\,t < T_k(m_k) \wedge \tau_\partial) \leq (k+2)\,\epsilon,$$

so as to treat any moment. Since t is arbitrary, this concludes the proof of Proposition 5.5.0.1. ☐

5.6 Proof of Theorem 5.2.3 : the infinite dimensional case

Remarks 5.6.0.1. *As one can imagine, this proof is much more technical than the previous ones. For instance, there is no explicit reference measure that seems to be exploitable as $\zeta^{(\infty)}$: the Lebesgue measure cannot be extended on an infinite dimensional space ! The core idea behind the proof is that the individuals carrying many mutations are actually wiped out very rapidly, implying rapid shuffle of the last coordinates. Quite unexpectedly, the criteria we developed to deal with jump events has proved to be very effective in this context : notably, we could deal with moments increasing too largely as* *exceptional events.*

5.6.1 Main properties leading to the proof

The proof of Theorem 5.2.3 relies on the criteria presented in [Chapter 2], which are the same as the one in [Chapter 1] except for the last one (Absorption with failures). We will treat both the case of large yet finite values of d and $d = \infty$, for which we recall that any $x \in \mathcal{X}_\infty$ has a finite sixth moment.

For the purpose of Theorem 5.2.3, we replace the notation given in (5.4.1) by the following one :

$$\mathcal{D}_\ell^{(d)} := \{x \in \mathcal{X}_d \,\big|\, M_3(x) \leq \ell \,\text{,}\, x_0 \in ((3\ell)^{-1}, 1 - (3\ell)^{-1})\}$$

We prove Theorem 5.2.3 thanks to the following Theorems 5.6.1-3, ordered by difficulty. We see in Section 5.6.1 how these Theorems together with Theorems 2.1-4 in [Chapter 2] imply Theorem 5.2.3. In the next subsections, we then prove Theorems 5.6.1-3 by order of appearance.

Escape from the Transitory domain

Theorem 5.6.1. *For any $t, \epsilon, \eta, m_3^\wedge > 0$, there exists $m_3 \geq m_3^\wedge$, $y > 0$ such that, with*

$$E^{(d)} := \{x \in \mathcal{X}_d \;;\; M_3(x) \leq m_3 \,\text{,}\, \forall j \leq \lfloor m_3/\eta \rfloor + 1,\; x_j \geq y\} \tag{5.6.1}$$

and $\tau_{E^{(d)}} := \inf\{t \geq 0 \;;\; X(t) \in E^{(d)}\}$:

$$\sup_{d \in \mathbb{N} \cup \{\infty\}} \sup_{x \in \mathcal{X}_d} \mathbb{P}_x^{(d)} \left(t < \tau_\partial \wedge \tau_{E^{(d)}}\right) \leq \epsilon.$$

In particular, for any $\rho, \eta, m_3^\wedge > 0$ we can find such m_3 and y such that :

$$\sup_{d \in \mathbb{N} \cup \{\infty\}} \sup_{x \in \mathcal{X}_d} \mathbb{E}_x^{(d)} \left(\exp[\rho\left(\tau_\partial \wedge \tau_{E^{(d)}}\right)]\right) \leq 2.$$

Mixing property and accessibility

Theorem 5.6.2. *There exists for any $d \in \mathbb{N} \cup \{\infty\}$ a probability measure $\zeta^{(d)}$ satisfying the following uniform mixing condition. For any $\ell \geq 1$ and $t > 0$, there exists $L > \ell$ and $c > 0$ such that for any $d \in \mathbb{N} \cup \{\infty\}$ and $x \in \mathcal{D}_\ell$, with $T_{\mathcal{D}_L^{(d)}}$ the exit time out of $\mathcal{D}_L^{(d)}$:*

$$\mathbb{P}_x^{(d)}(X(t) \in dx \;;\; t < T_{\mathcal{D}_L^{(d)}}) \geq c\,\zeta^{(d)}(dx).$$

Absorption with failures We will consider any $E^{(d)}$ of the form prescribed by Theorem 5.6.1.

Theorem 5.6.3. *Given any $\rho > 0$, $m_3, \eta, y > 0$, and any $\epsilon \in (0, 1)$, there exists $t, t', c > 0$ such that for any $d \in \mathbb{N} \cup \{\infty\}$ and $x \in E^{(d)}$, there exists a stopping time U_A such that :*

$$\{\tau_\partial \wedge t \leq U_A\} = \{U_A = \infty\} \qquad \text{and} \qquad \mathbb{P}_x^{(d)}(U_A = \infty,\, t < \tau_\partial) \leq \epsilon \exp(-\rho\, t),$$

while, with $\zeta^{(d)}$ as defined for Theorem 5.6.2 in (5.6.15) :

$$\mathbb{P}_x^{(d)}\Big[X(U_A) \in dy \ ;\ U_A < \tau_\partial\Big] \leq c\, \mathbb{P}_{\zeta^{(d)}}^{(d)}\Big[X(t') \in dy \ ;\ t' < \tau_\partial\Big].$$

Moreover, there exists a stopping time U_A^∞ satisfying the following properties :

- *$U_A^\infty := U_A$ on the event $\Big\{\tau_\partial \wedge U_A < \tau_{E^{(d)}}^t\Big\}$, where $\tau_{E^{(d)}}^t := \inf\{s \geq t : X_s \in E^{(d)}\}$.*
- *On the event $\Big\{\tau_{E^{(d)}}^t < \tau_\partial\Big\} \cap \{U_A = \infty\}$, and conditionally on $\mathcal{F}_{\tau_{E^{(d)}}^t}$, the law of $U_A^\infty - \tau_{E^{(d)}}^t$ coincides with the one of $\widetilde{U}_A^\infty$ for the shifted process $(\widetilde{X}_s)_{s\geq 0} := (X_{\tau_{E^{(d)}}^t + s})_{s \geq 0}$.*

For the proof of this Theorem, it is helpful to replace the initial condition $\zeta^{(d)}$ by $x_\zeta \in E^{(d)}$. Without this issue of having an estimate uniform in d sufficiently large, we could simply impose additionally in our choice of $m_3 \geq 1$ and $y > 0$ that $\zeta^{(d)}(E^{(d)}) \geq 1/2$. A priori, there is no reason to expect that it could not hold globally in d, yet we do not see how to justify it clearly. We thus require the following Lemma, whose proof is deferred in Section 5.8.5 of the Appendix.

Lemma 5.6.1.1. *There exists $m_3, y, c, t > 0$ such that for any d and $\zeta^{(d)}$ as defined in (5.6.15), we have $\mathbb{P}_{\zeta^{(d)}}^{(d)}(X(t) \in E^{(d)}) \geq c$.*

5.6.2 Proof of Theorem 5.2.3 as a consequence of Theorems 5.6.1-3

— First, it is clear that the sets $\mathcal{D}_\ell^{(d)}$ are closed and satisfy assumption $(A0)$ in [Chapter 2] :
$\forall\, \ell \geq 1,\ \mathcal{D}_\ell^{(d)} \subset int(\mathcal{D}_{\ell+1}^{(d)})$ and $\bigcup_{\ell \geq 1} \mathcal{D}_\ell^{(d)} = \mathcal{X}_d$.

— From Theorem 5.6.2, there is a reference measure $\zeta^{(d)}$ for which assumption $(A1)$ holds true.

— Theorem 5.6.1 implies $(A2)$ for any ρ, and we also require that ρ is chosen such that :

$$\rho > \rho_S := \sup\Big\{\gamma \geq 0 \,\Big|\, \sup_{L \geq 1} \liminf_{t > 0}\ e^{\gamma t}\, \mathbb{P}_{\zeta^{(d)}}^{(d)}(t < \tau_\partial \wedge T_{\mathcal{D}_L^{(d)}}) = 0\Big\} \vee 0.$$

From Lemma 3.0.2 in [Chapter 1] and $(A1)$, we know that ρ_S is upper-bounded by some value $\widetilde{\rho}_S$ independent of d. In order to satisfy $\rho > \rho_S$, we set $\rho := 2\widetilde{\rho}_S$.

From Theorem 5.6.1, we deduce $E^{(d)}$ such that assumption $(A2)$ holds for this value of ρ.

— Finally, Theorem 5.6.3 states that assumption $(A3_F)$ holds true, for $E^{(d)}$ and ρ.

Referring to Theorems 2.1-4 in [Chapter 2], this concludes the proof of the set of assumptions (AF), thus of Theorem 5.2.3.

Since all the parameters can be chosen independently of d sufficiently large, this is also the case of any parameter involved in the convergences. One can indeed check that there are intricate yet explicit relations between all the parameters introduced in [Chapter 2].

5.6.3 Proof of Theorem 5.6.1

The proof of Theorem 5.6.1 relies on the following Lemmas, easily adapted from the uniform escape and the uniform descent of the moments in the finite dimensional systems.

Lemma 5.6.3.1. *For any $d \in \mathbb{N} \cup \{\infty\}$, $t > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < \tau_\partial\right) \middle| x \in \mathcal{X}_d \, , \; M_1(x) \in (1, \infty) \, , \; x_0 \, M_1(x) \leq \delta \right\} \to 0 \text{ as } \delta \to 0.$$

Lemma 5.6.3.2. *Given any $d \in \mathbb{N} \cup \{\infty\}$, $t, y_0 > 0$, and recalling $T_1(m_1) := \inf\{t \geq 0 \; ; \; M_1(X_t) \leq m_1\}$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t \leq T_1(m_1) \wedge \tau_\partial\right) \middle| x \in \mathcal{X}_d \, , \; x_0 \geq y_0 \right\} \underset{m_1 \to \infty}{\longrightarrow} 0.$$

Lemma 5.6.3.3. *For any $d \in \mathbb{N} \cup \{\infty\}$, $t, m_1 > 0$, :*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < \tau_\partial\right) \middle| x \in \mathcal{X}_d \, , \; M_1(x) \leq m_1 \, , \; x_0 \leq \delta \right\} \to 0 \text{ as } \delta \to 0.$$

From the finite dimensional case, we adapt the definition, for any $y > 0$ and $k \geq 0$, of :

$$T_y^k := \inf\{s \geq 0 \; ; \; \forall j \leq k, \; X_j(s) \geq y\}.$$

Recall also that $k \in [\![1, d]\!]$ has to be understood as $k \in \mathbb{N}$ if $d = \infty$.

Lemma 5.6.3.4. *Given any $d \in \mathbb{N} \cup \{\infty\}$, $k \in [\![1, d]\!]$, and t, $m_1, y_0 > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(t < T_y^k \wedge \tau_\partial\right) \middle| x \in \mathcal{X}_d \, , \; x_0 \geq y_0 \, , \; M_1(x) \leq m_1 \right\} \to 0 \text{ as } y \to 0.$$

Lemma 5.6.3.5. *Given any $d \in \mathbb{N} \cup \{\infty\}$, $k \in [\![1, d]\!]$ and t, $m_1, y > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(\exists\, j \leq k,\ \exists\, s \leq 3\,t,\ X_j(s) \leq y' \;,\; 4\,t < \tau_\partial \right) \middle| x \in \mathcal{X}_d, \right.$$
$$\left. M_1(x) \leq m_1 \;;\; \forall\, j \leq k,\ x_j \geq y \right\} \to 0 \text{ as } y' \to 0.$$

Lemma 5.6.3.6. *Given any $d \in \mathbb{N} \cup \{\infty\}$, $k \in [\![1, d]\!]$ and t, $m_k > 0$:*

$$\sup \left\{ \mathbb{P}_x^{(d)} \left(\inf_{s \leq t} M_{k+1}(X_s) \geq m_{k+1} \right) \middle| x \in \mathcal{X}_d, M_{2\,k}(x) < \infty \;,\; M_k(x) \leq m_k \right\} \to 0 \text{ as } m_{k+1} \to \infty.$$

One can very easily adapt the proofs of Lemmas 5.5.0.3, 5.5.0.4, 5.4.6.2 and 5.5.0.5 respectively for Lemmas 5.6.3.1, 5.6.3.2, 5.6.3.4 and 5.6.3.6. The details are left to the reader.

As a generalization of Lemma 5.4.6.1, Lemma 5.6.3.3 is a consequence of the fact that X_0 is upper-bounded on the event $\{\sup_{s \leq t} M_1(s) \leq m_1'\}$ by the solution Y of :

$$dY_s = \alpha\, m_1'\, dt + \sqrt{Y_t(1 - Y_t)} dB_0(t) \;,\; Y_0 = \delta,$$

for which it is known that 0 is an absorbing value. Thanks to Lemma 3.2 in [AP13], we know an upper-bound going to 0 as m_1' goes to ∞, uniform in $x \in \mathcal{X}_d$ such that $M_1(x) \leq m_1$, of $\mathbb{P}_x(\sup_{s \leq t} M_1(s) \geq m_1')$. The details are left to the reader.

Finally, we defer the proof of Lemma 5.6.3.5 to the end in Section 5.6.3.2.

5.6.3.1 How to deduce Theorem 5.6.1

With exactly the same reasoning as for Proposition 5.5.0.1, with Lemmas 5.6.3.1 and 5.6.3.2 instead of 5.5.0.3 and 5.5.0.4, we deduce that for any $x \in \mathcal{X}_d$, $t > 0$ and ϵ, we can find m_1 such that :

$$\mathbb{P}_x^{(d)}(\mathcal{E}_1) \leq 3\,\epsilon, \quad \text{with } \mathcal{E}_1 := \{2\,t < T_1(m_1) \wedge \tau_\partial\}$$

From Lemma 5.6.3.3, we can find y_0 such that :

$$\mathbb{P}_x(\mathcal{E}_2) \leq \epsilon, \quad \text{with } \mathcal{E}_2 := \{X_0(T_1(m_1)) \leq y_0\} \cap \{T_1(m_1) \leq 2\,t\} \cap \{3\,t < \tau_\partial\}.$$

Again with Lemma 3.2 in [AP13] and the Markov property, we find $m_1' > 0$ such that :

$$\mathbb{P}_x^{(d)}(\mathcal{E}_4) \leq \epsilon, \quad \text{with } \mathcal{E}_4 := \{T_1(m_1) + t < \tau_\partial\}$$
$$\cap \{\exists\, s \in [T_1(m_1), T_1(m_1) + t],\ M_1(s) \geq m_1'\}$$

Before we ensure that many components escape 0 during this time-interval $[T_1(m_1), T_1(m_1)+$

$t]$, we need to know the number of components needed, which is given by the next step. Thus, iterating twice Lemma 5.6.3.6 with the Markov property (first at $T_1(m_1)+t$ and then at some time $T_2(m_2)$ for $m_2 > 0$ sufficiently large), we obtain some $m_3 > 0$ such that :

$$\begin{aligned}\mathbb{P}_x^{(d)}(\mathcal{E}_6) \leq \epsilon, \quad & \text{with } \mathcal{E}_6 := \{T_1(m_1)+t < \tau_\partial\} \cap \{M_1(T_1(m_1)+t) \leq m_1'\} \\ & \cap \{T_1(m_1) + 3\,t \leq \widetilde{T}_3(m_3)\} \\ \text{and } \widetilde{T}_3(m_3) &:= \inf\{s \geq T_1(m_1)+t \;;\; M_3(s) \leq m_3\}\end{aligned}$$

Now, we can define $k := \lfloor m_3/\eta \rfloor + 1$ and find by Lemma 5.6.3.4 some y such that :

$$\begin{aligned}\mathbb{P}_x^{(d)}(\mathcal{E}_3) \leq \epsilon, \quad & \text{with } \mathcal{E}_3 := \{T_1(m_1) < \tau_\partial\} \cap \{X_0(T_1(m_1)) \geq y_0\} \\ & \cap \{T_1(m_1)+t \leq \widetilde{T}_y^k\} \\ \text{where } \widetilde{T}_y^k &:= \inf\{s \geq T_1(m_1) \;;\; \forall\, j \leq k,\ X_j(s) \geq y\}.\end{aligned}$$

Finally, we choose y' by Lemma 5.6.3.5 such that :

$$\begin{aligned}\mathbb{P}_x^{(d)}(\mathcal{E}_5) \leq \epsilon, \quad & \text{with } \mathcal{E}_5 := \{\widetilde{T}_y^k + 4\,t < \tau_\partial\} \cap \{M_1(\widetilde{T}_y^k) \leq m_1'\} \\ & \cap \{\exists\, s \in [\widetilde{T}_y^k, \widetilde{T}_y^k + 3\,t],\ \exists\, j \leq k,\ X_j(s) \leq y'\}\end{aligned}$$

Provided that we prove that the event $\mathcal{E} := \{6\,t < \tau_\partial \wedge \tau_{E^{(d)}}\}$ (with y' instead of y in the definition of $\tau_{E^{(d)}}$) is necessarily included in the union of the exceptional events we have just defined, this ensures : $\forall\, x \in \mathcal{X}_d, \quad \mathbb{P}_x^{(d)}(6\,t < \tau_\partial \wedge \tau_{E^{(d)}}) \leq 8\,\epsilon$ and concludes the proof since t and ϵ have been arbitrary chosen.

On $\mathcal{E} \setminus \mathcal{E}_1$, we know $T_1(m_1) \leq 2\,t$.

On $\mathcal{E} \setminus \cup_{i=1}^2 \mathcal{E}_i$, we deduce also $X_0(T_1(m_1)) \geq y_0$.

On $\mathcal{E} \setminus \cup_{i=1}^3 \mathcal{E}_i$: $\quad \widetilde{T}_y^k \leq T_1(m_1)+t \leq 3\,t$.

On $\mathcal{E} \setminus \cup_{i=1}^4 \mathcal{E}_i$: $\quad M_1(\widetilde{T}_y^k) \vee M_1(T_1(m_1)+t) \leq m_1'$.

On $\mathcal{E} \setminus \cup_{i=1}^5 \mathcal{E}_i$: $\quad \forall\, s \in [\widetilde{T}_y^k, \widetilde{T}_y^k + 3\,t],\ \forall\, j \leq k,\ X_j(s) \geq y'$.

On $\mathcal{E} \setminus \cup_{i=1}^6 \mathcal{E}_i$: $\quad \widetilde{T}_3(m_3) \leq T_1(m_1) + 3\,t \leq 5\,t$.

Since moreover $\widetilde{T}_y^k \leq T_1(m_1)+t$, while, by definition of $\widetilde{T}_3(m_3)$, $\widetilde{T}_3(m_3) \geq T_1(m_1)+t$, we deduce : $\widetilde{T}_3(m_3) \in [\widetilde{T}_y^k, \widetilde{T}_y^k + 3\,t]$. As a consequence : $\forall\, j \leq k,\ X_j(\widetilde{T}_3(m_3)) \geq y'$. Then, it would imply $\tau_{E^{(d)}} \leq \widetilde{T}_3(m_3) \leq 5\,t$, which contradicts the definition of $\mathcal{E}$. Thus : $\mathcal{E} \subset \cup_{i=1}^6 \mathcal{E}_i$, and the conclusion of Theorem 5.6.1 is proved. □

5.6.3.2 Proof of Lemma 5.6.3.5

Let $k \geq 1,\ t, m_1, y > 0$. For $\delta > 0$, that we will choose sufficiently small, let : $\tau_\delta := \inf\{t \geq 0 \;;\; X_0(t) \leq \delta\}$. By Lemma 3.2 in [AP13], $M_1(\tau_\delta) \leq m_1'$ with large probability for m_1' sufficiently large and independently of δ and of d. Thus, by Lemma 5.6.3.3, choosing δ sufficiently small ensures that on the event $\{\tau_\delta \leq 3t\}$ extinction before $4\,t$ happens with a probability close to one.

We restrict ourselves in the following to the event $\{3\,t < \tau_\delta\}$. Now, for $1 \leq i \leq k$, and $y_{i-1} > 0$, let :

$$\tau^{i-1}(y_{i-1}) := \inf\{t \geq 0 \ ; \ \exists\, j \leq i,\ X_j(t) \leq y_{i-1}\}.$$

The proof relies on an induction over the coordinates $1 \leq i \leq k$ that there exits $0 < y_i \leq y_{i-1}$ such that $3\,t < \tau^i(y_i)$ with a probability close to 0 conditionally on the event $\{3\,t < \tau^{i-1}(y_{i-1})\}$.

On the event $\{3t < \tau^{i-1}(y_{i-1})\}$, we can observe :

$$dX_i(t) \geq \lambda\, y_{i-1}\, dt - (i\,\alpha + \lambda)\, X_i(t)\, dt + \sqrt{X_i(t)\,(1 - X_i(t))}\, dB_i(t),$$

for some standard Brownian Motion B_i (these are clearly not independent for different values of i and their correlations depend on d through $X^{(d)}$). By some comparison principles, for instance Proposition 3.12 in [PR14], X_i is lower-bounded (uniformly in x) by the solution to the SDE :

$$dY_i(t) = \lambda\, y_{i-1}\, dt - (i\,\alpha + \lambda)\, Y_i(t)\, dt + \sqrt{Y_i(t)\,(1 - Y_i(t))}\, dB_i(t),$$

with $Y_i(0) = y$ and absorption at 1. Note that Y_i cannot be absorbed at 1 before $\tau^{i-1}(y_{i-1})$ by the definition of the latter and that the law Y_i does not depend on d.

Now, for any i, 0 is an entrance boundary for Y_i, cf e.g. Subsection 3.3.3 in [JT17], so that there exists $0 < y_i \leq y_{i-1}$ such that with a probability close to 1 conditionally on the event $\{3t \leq \tau^{i-1}(y_{i-1})\}$:

$$\inf_{s \leq t} Y_i(s) \geq y_i \qquad \text{thus } t \leq \tau^i(y_i).$$

More precisely, for any ϵ, the above arguments shows by induction that there exists a decreasing sequence $(y_i)_{1\leq i\leq k} \in (\mathbb{R}_+^*)^k$, with $y_0 = \delta$, such that :

$$\sup\left\{\mathbb{P}_x^{(d)}\left(4t < \tau_\partial \,\middle|\, 3t \leq \tau^0_{y_0}\right) \middle| x \in \mathcal{X}_d, M_1(x) \leq m_1 \ ; \ \forall\, j \leq k,\ x_j \geq y\right\} \leq \epsilon/2,$$
$$\sup\left\{\mathbb{P}_x^{(d)}\left(\tau^i_{y_i} < 3t \,\middle|\, 3t \leq \tau^{i-1}_{y_{i-1}}\right) \middle| x \in \mathcal{X}_d, M_1(x) \leq m_1 \ ; \ \forall\, j \leq k,\ x_j \geq y\right\} \leq \epsilon/2^{i+1}.$$

Now, since an immediate induction ensures :

$$\mathbb{P}_x^{(d)}(\tau^k(y_k) \leq 3\,t\,,\ 4\,t < \tau_\partial) \leq \mathbb{P}_x(4\,t < \tau_\partial \,|\, \tau_\delta \leq 3t) + \sum_{i=1}^{k} \mathbb{P}_x^{(d)}\left(\tau^i_{y_i} \leq 3t \,\middle|\, 3t < \tau^{i-1}_{y_{i-1}}\right),$$

we can indeed conclude that the probability of $\{\tau^k(y_k) \leq 3\,t\} \cap \{4\,t < \tau_\partial\}$ is uniformly upper-bounded by ϵ. This ends the proof of Lemma 5.6.3.5. □

Transformation of the system of SDEs

These changes in the description of the system will be crucial for both the proofs of Theorems 5.6.3 and 5.6.2. Up to a multiplicative constant in the probabilities, they makes it possible to gather the last coordinates in one specific block while keeping a Markovian description. Our aim is then to prove that the dependency in the initial values of these last coordinates vanishes very quickly. We split the system of SDEs to distinguish the "descendants" of these doomed lineages from the unlucky newcomers that have acquired additional mutations (whose traits are predictable).

5.6.4 Aggregation of the last coordinates

In the following, for any $d \in \mathbb{N} \cup \{\infty\}$ and $k \leq d$, we denote by $d\mathbb{P}^{(k,d)}$ the law of the solution to :

$$\begin{aligned}\forall\, i \leq d, \quad dX_i(t) &= \alpha(M_1^{(k)}(t) - i \wedge k)\, X_i(t)\, dt + \lambda(X_{i-1}(t) - \mathbf{1}_{\{i<d\}}\, X_i(t))\, dt \\ &\quad + \sqrt{X_i(t)}\, dW_t^i - X_i(t)\, dW_t \qquad\qquad (5.6.2 : S^{(k,d)})\end{aligned}$$

$$\text{where } W_t := \textstyle\sum_j \int_0^t \sqrt{X_j(s)} dW_s^j,$$

$$M_1^{(k)}(t) := \textstyle\sum_i (i \wedge k)\, X_i(t) \quad = \sum_{i \leq k-1} i\, X_i(t) + k\, \sum_{i \geq k} X_i(t),$$

with $(W^i)_{i \geq 0}$ a family of mutually independent Brownian Motions.

For the following proposition, we will exploit a control on moments of order δ relying on the stopping time :

$$\tau_m^\delta := \inf\{s \geq 0\ ;\ M_\delta(s) \geq m\}.$$

Proposition 5.6.4.1. *Given any $t, \epsilon > 0$, $\delta > 2$, there exists $C_M, C_G > 0$ for which the following holds. For any $m \geq 1$, with $m' := C_M \times m$, for any $d \in \mathbb{N} \cup \{\infty\}$, $k \leq d$, and any $x \in \mathcal{X}_d \cap \mathcal{X}^{2\delta}$ such that $M_\delta(x) \leq m$, there exists a coupling between $\mathbb{P}^{(k,d)}$ and $\mathbb{P}^{(d)}$ such that :*

$$\text{on the event } \{t < \tau_{m'}^\delta\} : \quad \left| \log\left(\frac{d\mathbb{P}_x^{(k,d)}}{d\mathbb{P}_x^{(d)}}_{|[0,t]} \right) \right| \leq C_G \frac{m}{k^{\delta-2}},$$

$$\text{where } m' \text{ is such that } \mathbb{P}_x^{(d)}\left(\tau_{m'}^\delta \leq t\right) \leq \epsilon.$$

An analogous result holds for $\delta := 2$, except that the upper-bound on the log-ratio of densities is then $C_G \times m$.

Remarks 5.6.4.1. *In practice, we will exploit Proposition 5.6.4.1 only for $\delta = 3$. Yet, the proof is almost the same for any moment provided $\delta > 2$, while we mentioned earlier that the requirement that $M_6(x) < \infty$ could be replaced by the requirement $M_{2\delta}(x) < \infty$. So we treat the Proposition for this generality and let the interested reader extend the argument.*

This transform is naturally associated to the following projection π_k from $\mathcal{X}_d$ to $\mathcal{X}_k$, given by :

$$\pi_k(x)_i \begin{cases} = x_i \text{ if } i \leq k-1, \\ = 1 - \sum_{j=k}^{d} x_j \text{ if } i = k, \end{cases} \tag{5.6.3}$$

For the following proposition, we also define $(X_i^{[F]})_{i\in[\![k,d]\!]}$ as the solution to :

$$\begin{aligned} dX_i^{[F]}(t) := \lambda \frac{X_{k-1}(t)}{X_{(k)}(t)}(\mathbf{1}_{\{i=k\}} - X_i^{[F]}(t))\, dt + \lambda(X_{i-1}^{[F]}(t)\mathbf{1}_{\{i\geq k+1\}} - X_i^{[F]}(t))\, dt \\ + \sqrt{\frac{X_i^{[F]}(t)}{X_{(k)}(t)}} dW_t^{[F],i} - \frac{X_i^{[F]}(t)}{\sqrt{X_{(k)}(t)}} dW_t^{[F]} \end{aligned} \tag{5.6.4}$$

$$\text{where } X_{(k)}(t) := 1 - \sum_{i\leq k-1} X_i(t)\, ,\; W_t^{[F]} := \sum_{i\geq k} \int_0^t \sqrt{\frac{X_i^{[F]}(t)}{X_{(k)}(t)}} dW_t^{[F],i},$$

where $(W^{[F],i})_{i\in[\![k,d]\!]}$ is a sequence of independent Brownian Motion that are mutually independent from the $(W^i)_{i\in[\![0,d]\!]}$. $X^{[F]}$ shall play the role of the renormalized sequence of the last coordinates, and F stands for "Final".

Proposition 5.6.4.2. *For any $k \geq 1$, $\pi_k(X)$ has the same autonomous law under any $\mathbb{P}_x^{(k,d)}$, with $d \in [\![k, \infty[\![\cup \{\infty\}$. The vector X under the law $\mathbb{P}_x^{(k,d)}$ has the same law as the vector $(X_i : 0 \leq i \leq k-1\ ;\ X_{(k)} \times X_i^{[F]} : i \geq k)$, where the $X_i^{[F]}$ are defined in (5.6.4).*

The control of the increase in the moments in Proposition 5.6.4.1 relies on the following lemma, with another similar Lemma exploiting Proposition 5.6.4.2. Their proofs are commonly deferred to Sections 5.8.1-3 of the Appendix, given how similar they are to the one of Proposition 5.5.0.2.

Proposition 5.6.4.3. *For any $t > 0$, there exists $C \geq 1$ such that for any m, m', $d \in \mathbb{N} \cup \{\infty\}$ and $x \in \mathcal{X}_d$ such that $M_3(x) \leq m$,*

$$\mathbb{P}_x^{(d)}\left(\tau_{m'}^3 \leq t\right) \leq \frac{Cm}{m'}.$$

Similarly, exploiting the decomposition in Proposition 5.6.4.2, we define :

$$M_3^{[F]}(t) := \sum_{i\geq k} i^3\, X_i^{[F]}(t) \in [k^3, \infty) \tag{5.6.5}$$

$$\tau_m^{[F],3} := \inf\{s \geq 0 \ ; \ M_3^{[F]}(s) \geq m\}, \quad m > 0. \tag{5.6.6}$$

For clarity, we define $\mathcal{F}^{(k)} = \sigma(W^i : i \leq k-1 \ ; \ W)$. Recall that the process $X_i^{[F]}$ is driven by Brownian Motions $(W^{[F],i} : i \geq k)$ that are independent of $\mathcal{F}^{(k)}$. The inclusion $\sigma(\pi_k(X)) \subset \mathcal{F}^{(k)}$ is directly obtained through the autonomous set of equation verified by $\pi_k(X)$. The following control on $M_3^{[F]}$ exploits the filtration $\mathcal{F}_t^{(k)} := \mathcal{F}^{(k)} \vee \mathcal{F}_t$.

Proposition 5.6.4.4. *For any $t > 0$, there exists $C \geq 1$ such that for any m, m', $d \in \mathbb{N} \cup \{\infty\}$ and $x \in \mathcal{X}_d$ such that $M_3^{[F]}(x) \leq m$,*

$$\mathbb{P}_x^{(k,d)}\left(\tau_{m'}^{[F],3} \leq t \,\middle|\, \mathcal{F}^{(k)}\right) \leq \frac{Cm}{m'}.$$

5.6.4.1 Proof of Proposition 5.6.4.1

In view of applying the Girsanov formula to a well-suited exponential martingale, we denote :

$$R_1^{(k)}(t) := \sum_{i\geq k+1} (i-k)\, X_i(t)\, , \qquad R_2^{(k)}(t) := \sum_{i\geq k+1} (i-k)^2\, X_i(t)$$

One can notice that they correspond to the expectation and variance of the vector $(1 - \sum_{i=0}^k X_i\, ; \, X_{k+1}; X_{k+2}; ...)$. By the Cauchy-Schwarz inequality, this implies that $R_2^{(k)} \geq (R_1^{(k)})^2$.

The following lemma gives the coupling we shall consider. Its proof is deferred to Section 5.8.4 of the Appendix.

Lemma 5.6.4.5. *There exists a coupling between $\mathbb{P}^{(k,d)}$ and $\mathbb{P}^{(d)}$ such that :*

$$\begin{aligned}\log \frac{d\mathbb{P}_x^{(k,d)}}{d\mathbb{P}_x^{(d)}}_{|[0,t]} &= \alpha\, R_1^{(k)}(0) - \alpha\, R_1^{(k)}(t) + \alpha^2 \int_0^t (M_1(s) - k)\, R_1^{(k)}(s)\, ds \\ &- \alpha^2 \int_0^t R_2^{(k)}(s)\, ds + \alpha\, \lambda \int_0^t X_{(k)}(s)\, ds - \frac{\alpha^2}{2} \int_0^t \left[R_2^{(k)}(s) - R_1^{(k)}(s)^2\right] ds.\end{aligned}$$

In order to bound these terms, we note that :

$$0 \leq R_1^{(k)} \leq k^{-(\delta-1)} \sum_{i\geq k+1} i^\delta\, X_i(t) \leq k^{-(\delta-1)} M_\delta, \tag{5.6.7}$$

$$\text{Likewise} \quad 0 \leq M_1 \times R_1^{(k)} \leq k^{-(\delta-2)} M_1 \times M_{\delta-1} \leq k^{-(\delta-2)} M_\delta, \tag{5.6.8}$$

where we used Hölder's inequality to deduce : $M_1 \leq M_\delta^{1/\delta}$ and $M_{\delta-1} \leq M_\delta^{(\delta-1)/\delta}$. Similarly, $0 \leq X_{(k)} \leq k^{-\delta} M_\delta$, $0 \leq R_2^{(k)} \leq k^{-(\delta-2)} M_\delta$ while $(R_1^{(k)})^2 \leq R_2^{(k)}$ by the Cauchy-Schwarz inequality. Thanks to Lemma 5.6.4.5, this proves that on the event $\{t < \tau_{m'}^\delta\}$, for some constant C_1 only depending on δ and t (also on α and λ) :

$$\Big| \log \frac{d\mathbb{P}_x^{(k,d)}}{d\mathbb{P}_x^{(d)}}_{|[0,t]} \Big| \leq C_1 \, \frac{m'}{k^{\delta-2}}. \tag{5.6.9}$$

For ϵ, with the constant C_2 coming from Proposition 5.6.4.3, to prove $\mathbb{P}_x^{(d)}\left(\tau_{m'}^\delta \leq t\right) \leq \epsilon$, it suffices to choose $C_M := C_2/\epsilon$, where we recall $m' = C_M \times m$. With $C_G = C_1 \times C_2/\epsilon$, we deduce then from inequality (5.6.9) that on the event $\{t < \tau_{m'}^\delta\}$:

$$\Big| \log \frac{d\mathbb{P}_x^{(k,d)}}{d\mathbb{P}_x^{(d)}}_{|[0,t]} \Big| \leq C_G \, \frac{m}{k^{\delta-2}}.$$

The proof is quite the same for the case $\delta = 2$, except that we only require $(k \vee M_1) R_1^{(k)} \vee R_2^{(k)} \leq M_2$. This concludes the proof of Proposition 5.6.4.1. □

5.6.4.2 Proof of Proposition 5.6.4.2

By defining $X_{(k)}(t) := \sum_{i \geq k} X_i(t)$, we can remark that : $M_1^{(k)}(t) := \sum_{i \leq k-1} i\, X_i(t) + k\, X_{(k)}(t)$. Moreover, under $\mathbb{P}^{(k,d)}$:

$$\begin{aligned}
dX_{(k)}(t) &= \alpha(M_1^{(k)}(t) - k)\, X_{(k)}(t)\, dt + \lambda X_{k-1}(t)\, dt + \sqrt{X_{(k)}(t)}\, dW_t^{(k)} - X_{(k)}(t)\, dW_t, \\
&\text{where } W_t^{(k)} := \sum_{j \geq k} \int_0^t \sqrt{\frac{X_j(s)}{X_{(k)}(s)}}\, dW_s^j, \\
W_t &:= \sum_i \int_0^t \sqrt{X_i(s)} dW_s^i = \sum_{i \leq k-1} \int_0^t \sqrt{X_i(s)} dW_s^i + \int_0^t \sqrt{X_{(k)}(s)} dW_s^{(k)}.
\end{aligned}$$

Here, $W^{(k)}$ indeed defines a Brownian Motion at least until τ_∂, since the mutation term ensures that $X_{(k)}$ stays positive, as it has been stated in Lemmas 5.6.3.4 and 5.6.3.5. (the way we may extend it afterwards plays no role). The correlation between $W^{(k)}$ and the $(W^i)_{i \leq k-1}$ remain zero, while they constitute a system of Brownian Motions under the same filtration. $W^{(k)}$ is thus independent from $\sigma(W^i; i \leq k-1)$, so that the system of equations satisfied by $\pi_k(X)$ is equivalent for any $\mathbb{P}_x^{(k,d)}$.

Concerning $X^{[F]}$, we define :

$$d\bar{X}_i^{[F]}(t) := \lambda \frac{X_{k-1}(t)}{X_{(k)}(t)} (\mathbf{1}_{\{i=k\}} - \bar{X}_i^{[F]}(t))\, dt + \lambda(\bar{X}_{i-1}^{[F]}(t) \mathbf{1}_{\{i \geq k+1\}} - \bar{X}_i^{[F]}(t))\, dt \\ + \sqrt{\frac{\bar{X}_i^{[F]}(t)}{X_{(k)}(t)}} dW_t^i - \frac{\bar{X}_i^{[F]}(t)}{\sqrt{X_{(k)}(t)}} dW_t^{(k)}.$$

We will first check by the Itô formula that, for $i \geq k$, X_i coincide with $\bar{X}_i := X_{(k)} \times \bar{X}_i^{[F]}$, by looking at their differentials. For $i \leq k-1$, we denote also $\bar{X}_i \equiv X_i$.

$$\begin{aligned} d\bar{X}_i(t) &= \lambda(\bar{X}_{i-1}(t) - \bar{X}_i(t))\, dt - \lambda\, X_{k-1}(t)\, \bar{X}_i(t)\, dt + \alpha(M_1^{(k)} - k)\, \bar{X}_i(t)\, dt \\ &\quad + \lambda\, \bar{X}_{k-1}(t)\, \bar{X}_i(t)\, dt + \sqrt{\bar{X}_i(t)} dW_t^i - \sqrt{X_{(k)}(t)}\, \sqrt{\bar{X}_i^{[F]}(t)} dW_t^{(k)} \\ &\quad + \sqrt{\bar{X}_i^{[F]}(t)}\, \sqrt{X_{(k)}(t)} dW_t^{(k)} - \bar{X}_i(t) dW_t^i + \tfrac{1}{2}\, d\langle \mathcal{M}^{(k)}, \mathcal{M}^{[F],i} \rangle_t, \end{aligned} \tag{5.6.10}$$

where $\mathcal{M}_t^{(k)}$ and $\mathcal{M}_t^{[F],i}$ denote the martingale parts of resp. $X_{(k)}(t)$ and $\bar{X}_i^{[F]}(t)$. In fact, this covariation vanishes because, from the definitions of $(W^j, j \in \mathbb{Z}_+)$ and $W^{(k)}$, we first deduce :

$$\sqrt{\bar{X}_i^{[F]}(t)} d\langle W^i, W^{(k)} \rangle_t - \bar{X}_i^{[F]}(t) d\langle W^{(k)}, W^{(k)} \rangle_t \equiv 0. \tag{5.6.11}$$

Concerning then the covariation with W and recalling $dW_t = \sum_{j \leq k-1} \sqrt{X_j(t)} dW^j(t) + \sqrt{X_{(k)}(t)} dW^{(k)}(t)$, since for any $j \in [\![0, k-1]\!]$, $\langle W^i, W^j \rangle \equiv 0$ and $\langle W^{(k)}, W^j \rangle \equiv 0$, we can conclude that $d\langle \mathcal{M}^{(k)}, \mathcal{M}^{[F],i} \rangle_t \equiv 0$. After simplification and replacing $dW_t^{(k)}$ by its expression involving $(\bar{X}_i)_{i \geq k}$, the system of equations (5.6.10) satisfied by $(\bar{X}_i)_{i \in [\![0,d]\!]}$ coincide with the system satisfied by $(X_i)_{i \in [\![0,d]\!]}$. By uniqueness of the whole system, X_i coincide with $\bar{X}_i$.

Now, we have exploited in the previous calculation that the martingale component of $\bar{X}_i^{[F]}$, for $i \geq k$, has a quadratic co-variation with $W^{(k)}$ that stays null. Since these semi-martingales are also adapted to a common filtration $(\mathcal{F}_t)$ and their increments after time t is independent of $\mathcal{F}_t$, we deduce from Theorem 2.1.8 in [DiT13] that $W^{(k)}$ is actually independent of the martingale components driving $\bar{X}^{[F]}$. Moreover, $\sigma(W^i : i \leq k-1)$ is by construction independent of $\sigma(W^i : i \geq k)$. Thus, considering $\sigma(W^i : i \leq k-1; W^{(k)})$, for which W is measurable, we deduce that it is independent of the family of martingale driving $\bar{X}^{[F]}$. We can thus replace the latter by the expression with a system of independent copies $W^{[F],i}$ as in Proposition 5.6.4.2 without changing the law of the vector. This ends the proof of Proposition 5.6.4.2.

□

5.6.5 Splitting of the solution

For any $d \in \mathbb{N} \cup \{\infty\}$ and $k \in [\![1, d]\!]$, consider the solution to :

$$\begin{aligned}\forall\, i \leq d, \quad dX_i^{[G]}(t) = \alpha(M_1^{(k)}(t) - i \wedge k)\, X_i^{[G]}(t)\, dt + \lambda(X_{i-1}^{[G]}(t) - \mathbf{1}_{\{i \neq d\}} X_i^{[G]}(t))\, dt \\ + \sqrt{X_i^{[G]}(t)}\, dW_t^{(n),i} - X_i^{[G]}(t)\, dW_t, \qquad (S^{[G]})\end{aligned}$$

$$\begin{aligned}\text{where } X_i^{[G]}(0) := x_i\, \mathbf{1}_{\{i \leq k-1\}}\ ; \quad X_{[R]}(t) := 1 - \sum_{i=0}^{d} X_i^{[G]}(t)\,; \\ W_t := \sum_i \int_0^t \sqrt{X_i^{[G]}(s)} dW_s^{(n),i} + \int_0^t \sqrt{X_{[R]}(s)} dW_s^{[R]}, \\ M_1^{(k)}(t) := \textstyle\sum_{i \leq d} (i \wedge k)\, X_i^{[G]}(t) + k\, X_{[R]}(t).\end{aligned}$$

Here, the $(W^{[G],i}, i \geq 0\ ;\ W^{[R]})$ defines a mutually independent family of standard Brownian Motions. $[G]$ stands for "Generative" while $[R]$ stands for "Rest", with the idea that the $[R]$ component shall quickly get extinct. Notably, 0 is an absorbing state for $X_{[R]}$, whose absorption time is denoted $\tau_\partial^{[R]}$.

The following solutions are well-defined in the time interval $[0, \tau_\partial^{[R]})$.

$$\forall\, i \leq d, \quad dX_i^{[R]}(t) = \lambda\, (X_{i-1}^{[R]}(t) - \mathbf{1}_{\{i \neq d\}}\, X_i^{[R]}(t))\, dt + \sqrt{\frac{X_i^{[R]}(t)}{X_{[R]}(t)}}\, dW_t^{[R],i} - \frac{X_i^{[R]}(t)}{\sqrt{X_{[R]}(t)}}\, dW_t^{[R]}$$

$$\text{where } W_t^{[R]} := \sum_{i \geq 0} \int_0^t \sqrt{X_i^{[R]}(s)} dW_s^{[R],i}, \quad X_i^{[R]}(0) = \frac{x_i\, \mathbf{1}_{\{i \geq k\}}}{\sum_{j \geq k} x_j}. \qquad (S^{[R]})$$

Again, the $(W^{[R],i})$ define a mutually independent family of Brownian Motions, also mutually independent from the family $(W^{[G],i})$ (and not with $W^{[R]}$ of course).

Looking at the equations for $X^{[G]}$, we see that it describes an autonomous system. We thus deduce the two following lemmas, the first one being immediately verified.

Lemma 5.6.5.1. *Considering two initial conditions x and x' such that $\forall\, i \leq k-1,\ x_i = x'_i$, $X^{[G]}$ under $\mathbb{P}_x$ has the same law as $\bar{X}^{[G]}$ under $\mathbb{P}_{x'}$.*

Lemma 5.6.5.2. *For any t, $\sum_j X_j^{[R]}(t) = 1$. Thus, $W^{[R]}$ is indeed a Brownian Motion, clearly independent from the family $(W^{[G],i})$. We can thus indeed choose this same Brownian Motion that couples the dynamics of $X^{[R]}$ and $X^{[G]}$.*

Lemma 5.6.5.3. *$X_{[R]}$ is solution to :*

$$dX_{[R]}(t) = \alpha\, (M_1^{(k)}(t) - k)\, X_{[R]}(t)\, dt + \sqrt{X_{[R]}(t)\, (1 - X_{[R]}(t))}\, d\widehat{W}_t^{[R]},$$

for some Brownian Motion $\widehat{W}^{[R]}$. Looking more precisely at the interactions with $X^{[G]}$, *it is actually solution to :*

$$dX_{[R]}(t) = \alpha\,(M_1^{(k)}(t) - k)\,X_{[R]}(t)\,dt + \sqrt{X_{[R]}(t)}\,dW_t^{[R]} - X_{[R]}(t)\,dW_t. \tag{5.6.12}$$

We see in the next lemma how these solutions are related to our initial problem.

Lemma 5.6.5.4. *The* $(X_i^{[G]} + X_{[R]} \times X_i^{[R]})_{i\leq d}$ *defines a solution to* (5.6.2 : $S^{(k,d)}$).

Moreover, an analogue of Proposition 5.6.4.3 can also be obtained in this setting. Let :

$$M_3^{[R]}(t) := \textstyle\sum_{\{i\geq k\}} i^3\, X_i^{[R]}(t), \tag{5.6.13}$$

$$\tau_m^{[R],3} := \inf\{s \geq 0\ ;\ M_3^{[R]}(s) \geq m\}, \quad m > 0. \tag{5.6.14}$$

Lemma 5.6.5.5. *For any* $t > 0$*, there exists* $C \geq 1$ *such that for any* m, m', $d \in \mathbb{N}\cup\{\infty\}$ *and* $x \in \mathcal{X}_d$ *such that* $x_{(k)} > 0$ *and* $M_3^{[F]}(x) \leq m$,

$$\mathbb{P}_x^{(k,d)}\left(\tau_{m'}^{[R],3} \leq t \,\middle|\, \mathcal{F}^{[G]}\right) \leq \frac{Cm}{m'}.$$

The proof of this lemma relies on the same ingredient as the ones of Propositions 5.6.4.3 and 5.6.4.4. The main difference is that the sequence $(T_k^{[R]})$, i.e. the analogue to the sequence (T_k) of Lemma 5.8.1.1, now satisfies $\lim_{k\to\infty} T_k^{[R]} \geq \tau_\partial^{[R]} := \inf\{t \geq 0 : X_{[R]}(t) = 0\}$. The details of the proof are left to the reader.

Remark : • To define $(X_i^{[R]})_{i\geq 0}$, one can extend the well-defined solutions on $[0, \tau_n^{[R]}]$ where $\tau_n^{[R]} := \inf\{t \geq 0 \,\big|\, X_{[R]}(t) \leq 2^{-n}\}$.

• By construction, $X_i^{[R]}(t) = 0$ for any $t < \tau_\partial^{[R]}$ and $i \leq k-1$.

• $X^{[R]}$ gathers all the information related to dependency in the initially large components. It is doomed to disappear quickly when k is large, because it concerns only a very small fraction of the population, and the source term has been driven towards $X_k^{[G]}$. Note that $M_1^{(k)}(t) \leq k$ so that even the drift term will not help.

5.6.5.1 Proof of Lemma 5.6.5.3

We deduce equation 5.6.12 from $(S^{[G]})$ since $\sum_{i\leq d}(x_{i-1} - \mathbf{1}_{\{i\neq d\}} x_i(t)) = 0$ for any $x \in \mathcal{X}_d$ and :

$$\begin{aligned}\sum_i [M_1^{(k)}(t) - (i\wedge k)]\, X_i^{[G]}(t) &= M_1^{(k)}(t)\,[1 - X_{[R]}(t)] - \sum_i (i\wedge k)\, X_i^{[G]}(t)\\ &= k\, X_{[R]}(t) - M_1^{(k)}(t)\, X_{[R]}(t)\end{aligned}$$

From this, we deduce :

$$dX_{[R]}(t) = \alpha(M_1^{(k)}(t) - k)\, X_{[R]}(t) + \sqrt{X_{[R]}(t)\,(1 - X_{[R]}(t))} d\widehat{W}_t^{[R]},$$
$$\text{where } d\widehat{W}_t^{[R]} := \mathbf{1}_{\{t<\tau_\partial\}} \left(\sqrt{1 - X_{[R]}(t)}\, dW_t^{[R]} - \sqrt{X_{[R]}(t)} \sum_{i \le d} \sqrt{\frac{X_i^{[G]}(t)}{1 - X_{[R]}(t)}}\, dW_t^{[G],i} \right)$$
$$+ \mathbf{1}_{\{t \ge \tau_\partial\}}\, dW_t^{(e)},$$

is indeed a Brownian Motion for $W^{(e)}$ a Brownian Motion, noting that $1 - X_{[R]} \ge X_0^{[G]} > 0$ as soon as $t < \tau_\partial := \inf\{t \ge 0 \ ;\ X_0^{[G]}(t) = 0\}$. □

5.6.5.2 Proof of Lemma 5.6.5.4

For $i \le d$, denote $\widetilde{X}_i := X_i^{[G]} + X_{[R]} \times X_i^{[R]}$. We deduce the system of equations it satisfies from the Itô formula. Note that the martingale parts of $X_{[R]}$ and $X^{[R]}$ have a null covariation. The component of bounded variation in the equation of $\widetilde{X}_i$ is thus :

$$\begin{aligned}
&\alpha(M_1^{(k)}(t) - i \wedge k)\, X_i^{[G]}(t) + \lambda\, (X_{i-1}^{[G]}(t) - \mathbf{1}_{\{i \ne d\}}\, X_i^{[G]}(t)) \\
&\quad + \alpha\, (M_1^{(k)}(t) - k)\, X_{[R]}(t)\, X_i^{[R]} + X_{[R]}(t)\, \lambda\, (X_{i-1}^{[R]}(t) - \mathbf{1}_{\{i \ne d\}}\, X_i^{[R]}(t)) \\
&= \alpha(M_1^{(k)}(t) - i \wedge k)\, \widetilde{X}_i(t) + +\lambda\, (\widetilde{X}_{i-1}(t) - \mathbf{1}_{\{i \ne d\}}\, \widetilde{X}_i(t)).
\end{aligned}$$

For the martingale part, we find :

$$\begin{aligned}
&d\mathcal{M}_i(t) := \sqrt{X_i^{[G]}(t)}\, dW_t^{[G],i} - X_i^{[G]}(t)\, dW_t \\
&+ \sqrt{X_{[R]}(t)} \left[\sqrt{X_i^{[R]}(t)}\, dW_t^{[R],i} - X_i^{[R]}(t)\, dW_t^{[R]} \right] + X_i^{[R]}(t) \left[\sqrt{X_{[R]}(t)}\, dW_t^{[R]} - X_{[R]}(t)\, dW_t \right] \\
&= \sqrt{X_i^{[G]}(t)}\, dW_t^{[G],i} + \sqrt{X_{[R]}(t)\, X_i^{[R]}(t)}\, dW_t^{[R],i} - (X_i^{[G]}(t) + X_{[R]}(t)\, X_i^{[R]}(t))\, dW_t
\end{aligned}$$

Here, we thus define the family $\widetilde{W}_i$ by

$$W_t := \sum_i \int_0^t \sqrt{X_i^{[G]}(s)} dW_s^{[G],i} + \int_0^t \sqrt{X[R](s)} dW_s^{[R]}$$

Then, (X_i) indeed defines solutions to (5.6.2 : $S^{(k,d)}$) since for any $i \le d$ and decomposition $y = y^r + y^n$, we have : $x_{i-1}^r - \mathbf{1}_{\{i \ne d\}} x_i^r(t) + x_{i-1}^n - \mathbf{1}_{\{i \ne d\}} x_i^n(t) = x_{i-1} - \mathbf{1}_{\{i \ne d\}} x_i(t)$. In particular, we have a.s. $\sum_j X_j^{[G]} + X_j^{[R]} = 1$.

□

5.6.6 Proof of Theorem 5.6.2

We recall the definition of $\mathcal{D}_\ell^{(d)}$:

$$\mathcal{D}_\ell^{(d)} := \{x \in \mathcal{X}_d \,\big|\, M_3(x) \leq \ell \,,\; x_0 \in ((3\ell)^{-1}, 1-(3\ell)^{-1})\}$$

The mixing will be achieved in two steps, with each step being completed after a time-interval of length t_M. t_M is arbitrarily taken as $t_M = 1$. We will exploit upper-bounds of the third moments given in Propositions 5.6.5.5 and 5.6.4.4 for the specific case where $k = 1$. In this proof, considering second moments instead of third ones would have been sufficient. Yet, estimates of the third moment shall be required for Theorem 5.6.3 and we wish to emphasize the similarity between them.

We will now first state two Lemmas, then show that Theorem 5.6.2 follows from them, and finally prove the two Lemmas.

For some $m > \ell$ and $y < 1/(2\ell)$ to be fixed later, and noting $M_3^{[F]} = M_3/(1 - X_0)$, let :

$$\tau_m^{[F],3} := \inf\{t \geq 0 \,\big|\, M_3(t) \geq m\,(1 - X_0(t))\} \,,\; T_y^0 := \inf\{t \geq 0 \,\big|\, X_0(t) \notin (y, 1-y)\} < \tau_\partial.$$

Lemma 5.6.6.1. *For any $\ell \geq 1$ and $t_M > 0$, there exists $m > \ell$ such that for any $y \in (0, 1/2\,\ell)$ there exists $c > 0$ such that for any $x \in \mathcal{D}_\ell^{(d)}$:*

$$\mathbb{P}_x^{(d)}(X_0(t_M) \in dx_0 \;;\; t_M < \tau_m^{[F],3} \wedge T_y^0) \geq c\,\mathbf{1}_{\{x_0 \in (2\,y, 1-2\,y)\}}\, dx_0.$$

For the following lemma, we base ourselves on the splitting presented in subsection 5.6.5 with $k = 1$. With these definitions of $X^{[G]}$ and $X^{[R]}$, we also denote :

$$\begin{aligned}
&\tau_\partial^{[R]} := \inf\{t \geq 0 \,\big|\, X_{[R]}(t) = 0\}\\
&\tau_{m_G}^{[G],3} := \inf\{t \geq 0 \,\big|\, M_3^{[G]}(t) \geq m_G\}\,, \qquad \tau_{m_R}^{[R],3} := \inf\{t \geq 0 \,\big|\, M_3^{[R]}(t) \geq m_R\},\\
&\qquad \text{where } M_3^{[G]} := \textstyle\sum_{i\geq 0} i^3\, X_i^{[G]}\,, \qquad M_3^{[R]} := \textstyle\sum_{i\geq 0} i^3\, X_i^{[R]}.
\end{aligned}$$

Moreover, let $\mathcal{F}^{[G]}$ be the filtration generated by the family $(W^{[G],i}, i \geq 0 \;;\; W^{[R]})$. In particular, the event $\{\tau_\partial^{[R]} \leq t_M < \tau_{m_G}^{[G],3} \wedge \tau^0\}$ is $\mathcal{F}^{[G]}$-measurable.

Lemma 5.6.6.2. *For any $t_M > 0$,there exists $m_G, c > 0$, $y \in (0, 1/10)$, $y' \in (0, y)$, such that for any $x \in \mathcal{X}_d$ such that $x_0 \in (1-3\,y, 1-2\,y)$, with $\tau^0 := \inf\{t \geq 0 \,\big|\, X_0(t) \notin (1/10, 1-y')\}$:*

$$\mathbb{P}_x^{(1,d)}(\tau_\partial^{[R]} \leq t_M < \tau_{m_G}^{[G],3} \wedge \tau^0) \geq c.$$

Thanks to these two Lemmas and Lemma 5.6.5.5, we will be able to prove Theorem 5.6.2 with the following (intricate) definition of $\zeta^{(d)}$. In this formula, the values of y

and m_G are deduced from Lemma 5.6.6.2 with the (arbitrary) choice $t_M := 1$.

$$\zeta^{(d)}(dx) := \tfrac{1}{y}\int_{1-3\,y}^{1-2y} dx_0\,\mathbb{P}^{(1,d)}_{\bar{x_0}}(X(t_M)\in dx \,\big|\, \tau^{[R]}_\partial \le t_M < \tau^{[G],3}_{m_G}\wedge\tau^0)\,; \qquad (5.6.15)$$
$$\bar{x_0} := x_0\delta_0 + (1-x_0)\,\delta_1.$$

Remark : For simplicity, we will apply the cutting and the splitting for $k=1$. The proof of Theorem 5.6.3 will exploit a generalization of this result for k large, with the first step ensuring a lower-bounded density of $(X_i(t_M)\ ;\ i\le k-1)$ on any $\mathcal{Y}_k(y)$, where $\mathcal{Y}_k(y)$ is defined for $y\in(0,1/k)$ by :

$$\mathcal{Y}_k(y) := \{x\in\mathcal{X}_d \,\big|\, (\wedge_{i\le k-1} x_i)\wedge(1-\textstyle\sum_{i\le k-1} x_i) > y\}.$$

5.6.6.1 Lemmas 5.6.6.1 and 5.6.6.2 imply Theorem 5.6.2

We first define $m_G, c_G>0$, $y\in(0,1/10)$, $y'\in(0,y)$ thanks to Lemma 5.6.6.2 such that for any $x\in\mathcal{X}_d$ such that $x_0\in(1-3\,y,1-2\,y)$:

$$\mathbb{P}^{(1,d)}_x(\tau^{[R]}_\partial \le t_M < \tau^{[G],3}_{m_G}\wedge\tau^0) \ge c_G. \qquad (5.6.16)$$

Given $\ell\ge 1$, we then define $m_F, c_\ell>0$ such that for any $x\in\mathcal{D}^{(d)}_\ell$:

$$\mathbb{P}^{(d)}_x(X_0(t_M)\in dx_0\ ;\ t_M<\tau^{[F],3}_{m_F}\wedge T^0_y) \ge c_\ell\,\mathbf{1}_{\{x_0\in(1-3\,y,1-2\,y)\}}\,dx_0. \qquad (5.6.17)$$

We define also $m_R>0$ thanks to Lemma 5.6.5.5, so that :

$$\mathbb{P}^{(1),d}_x(t_M<\tau^{[R],3}_{m_R}\,\big|\,\mathcal{F}^{[G]}) \ge 1/2, \qquad (5.6.18)$$

provided x satisfies $M^{[R]}(0)\le 2m_F$, in particular when $x_0\ge 1/2$ and $M_3(x)\le m_F$.

By choosing L sufficiently large, we ensure $L\ge m_F\vee(m_G+m_R)\vee(1/y')$. Recalling that $T_{\mathcal{D}^{(d)}_L}$ denotes the exit time out of $\mathcal{D}^{(d)}_L$, and that $M_3\le M_3^{[F]}$ (here $k=1$), it proves that $\{t_M<\tau^{[F],3}_{m_F}\wedge T^0_y\}\subset\{t_M<T_{\mathcal{D}^{(d)}_L}\}$. Likewise, since $M_3\le M_3^{[G]}+M_3^{[R]}$, $\{t_M<\tau^{[G],3}_{m_G}\wedge\tau^{[R],3}_{m_R}\wedge\tau_0\}\subset\{t_M<T_{\mathcal{D}^{(d)}_L}\}$.

By the Markov property :

$$\mathbb{P}^{(d)}_x(X(2\,t_M)\in dx'\ ;\ 2\,t_M<T_{\mathcal{D}^{(d)}_L}) \ge \int_{\mathcal{X}^{(d)}}\nu_x(dz)\mathbb{P}^{(d)}_z(X(t_M)\in dx'\ ;\ t_M<T_{\mathcal{D}^{(d)}_L}),$$
$$\text{where } \nu^{(d)}_x(dz) := \mathbb{P}^{(d)}_x(X(t_M)\in dz\ ;\ t_M<\tau^{[F],3}_{m_F}\wedge T^0_y).$$

The previous r.h.s. is itself lower-bounded by

$$\int_{\mathcal{X}^{(d)}}\nu^{(d)}_x(dz)\mathbf{1}_{\{z_0\in(1-3y,1-2y)\}}\mathbb{P}^{(d)}_z(X(t_M)\in dx'\ ;\ \tau^{[R]}_\partial\le t_M<\tau^{[G],3}_{m_G}\wedge\tau^0\wedge\tau^{[R],3}_{m_R}).$$

Note that on the event $\{\tau_\partial^{[R]} \leq t_M\}$, we know from Proposition 5.6.4.2 that $X(t_M) = X^{[G]}(t_M)$. Both $X^{[G]}(t_M)$ and $\{\tau_\partial^{[R]} \leq t_M\}$ are $\mathcal{F}^{[G]}$-measurable. Moreover, on the event $\{t_M < \tau^{[G],3}_{m_G} \wedge \tau^0 \wedge \tau^{[R],3}_{m_R}\}$:

$$\sup_{s\leq t_M} M_3(s) \leq \sup_{s\leq t_M} M_3^{[G]}(s) + \sup_{s\leq t_M\wedge \tau_\partial^{[R]}} M_3^{[R]}(s) \leq m_G + m_R.$$

Using Lemma 5.6.4.5 with some uniform upper-bound on the exponential martingale (with $k=1$) on the event $\{\tau_\partial^{[R]} \leq t_M < \tau^{[G],3}_{m_G} \wedge \tau^0 \wedge \tau^{[R],3}_{m_R}\}$, we deduce that there exists $c_E > 0$ such that :

$$\begin{aligned}
&\mathbb{P}_x^{(d)}(X(2\,t_M) \in dx' \;;\; 2\,t_M < T_{\mathcal{D}_L^{(d)}}) \\
&\quad\geq c_E \int_{\mathcal{X}^{(d)}} \nu_x^{(d)}(dz) \mathbf{1}_{\{z_0\in(1-3y,1-2y)\}} \mathbb{E}_z^{(1,d)}\Big[\mathbb{P}_z^{(1,d)}(t_M < \tau^{[R],3}_{m_R} \,\Big|\, \mathcal{F}^{[G]}) \;;\; X^{[G]}(t_M) \in dx', \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \tau_\partial^{[R]} \leq t_M < \tau^{[G],3}_{m_G} \wedge \tau^0\Big] \\
&\quad\geq (c_E/2) \int_{\mathcal{X}^{(d)}} \nu_x^{(d)}(dz) \mathbf{1}_{\{z_0\in(1-3y,1-2y)\}} \mathbb{E}_{\bar{z}_0}^{(1,d)}\Big[X^{[G]}(t_M) \in dx' \,,\; \tau_\partial^{[R]} \leq t_M < \tau^{[G],3}_{m_G} \wedge \tau^0\Big]
\end{aligned}$$

where we used that both $X^{[G]}$ and the event $\{\tau_\partial^{[R]} \leq t_M < \tau^{[G],3}_{m_G} \wedge \tau^0\}$ are $\mathcal{F}^{[G]}$-measurable, that they depend on z only through z_0, and (5.6.18). We thus have the same laws for initial condition z and $\bar{z}_0$. Thanks to (5.6.16) and (5.6.17), this concludes the proof of Theorem 5.6.2.

□

5.6.6.2 Proof of Lemma 5.6.6.1

Thanks to Lemma 5.6.5.5, we can fix some $m > 0$ such that for any $x \in \mathcal{D}_\ell^{(d)}$:

$$\mathbb{P}_x^{(1,d)}(t_M < \tau^{[R],3}_m \,\Big|\, \sigma(\widehat{W}^0)) \geq 1/2. \tag{5.6.19}$$

On the event $\{t_M < \tau^{[R],3}_m \wedge T_y^0\}$, one has for any $s < t_M$:

$$\sup\{X_{(1)}(s) \;;\; M_1(s) \;;\; R_1^{(1)}(s) \;;\; R_2^{(1)}(s)\} \leq M_3(s) \leq m.$$

Applying Lemma 5.6.4.5, we thus deduce that there exists $C > 0$ only depending on m and t_M such that :

$$\mathbb{P}_x^{(d)}(X_0(t_M) \in dx_0 \;;\; t_M < \tau^{[R],3}_m \wedge T_y^0) \geq C\, \mathbb{P}_x^{(1,d)}(X_0(t_M) \in dx_0 \;;\; t_M < \tau^{[R],3}_m \wedge T_y^0).$$

Under $\mathbb{P}^{(1,d)}$, X_0 is solution to the following autonomous equation :

$$dX_0(t) := \alpha X_0(t)\,(1 - X_0(t))\,dt - \lambda X_0(t)\,dt + \sqrt{X_0(t)\,(1 - X_0(t))}\,d\widehat{W}_t^0.$$

This property can be deduced as in Lemma 5.6.5.3. It is then classical for such an elliptic diffusion that $X_0(t_M)$ has a lower-bounded density on $(2\,y, 1-2\,y)$ on the event $\{t_M < T^0_y\}$, uniformly over any initial condition such that $x_0 \in (1/\ell, 1-1/\ell)$. Because of (5.6.19), this concludes the proof of Lemma 5.6.6.1. □

5.6.6.3 Proof of Lemma 5.6.6.2

From the Harnack inequalities, recalling that the equation for $X_0 = X_0^{[G]}$ is autonomous, one can prove that there exists $c_0 > 0$ such that for any $x \in \mathcal{X}_d$ satisfying $x_0 \in [1/2, 1]$:

$$\mathbb{P}^{(1,d)}_x(t_M < \tau_{1/10}) \geq c_0, \quad \text{where } \tau_{1/10} := \inf\{t \geq 0 \,\big|\, X_0(t) \leq 1/10\}. \tag{5.6.20}$$

Likewise, there is an autonomous equation for $X_{(1)}$ from which one can deduce that there exists y sufficiently small such that for any $x \in \mathcal{X}_d$ satisfying $x_0 \in [1-2\,y, 1-y]$ (i.e. $X_{(1)}(0) \in [y, 2\,y]$) :

$$\mathbb{P}^{(1,d)}_x(t_M < \tau^{[R]}_\partial) \leq c_0/4. \tag{5.6.21}$$

Since the border 1 is an entrance boundary for X_0, cf e.g. Subsection 3.3.3 in [JT17], there exists $y' \in (0, y)$ (again independent of d because X_0 is autonomous under $\mathbb{P}^{(1,d)}$), such that for any $x \in \mathcal{X}_d$ satisfying $x_0 \in [1/2, 1-y]$:

$$\mathbb{P}^{(1,d)}_x(t_M < \tau^0) \geq 3c_0/4. \tag{5.6.22}$$

By Proposition 5.6.4.3, there exists m_G sufficiently large such that for any $x \in \mathcal{X}_d$:

$$\mathbb{P}^{(1,d)}_x(\tau^{[G],3}_{m_G} \leq t_M) = \mathbb{P}^{(1,d)}_{\bar{x}_0}(\tau^{[G],3}_{m_G} \leq t_M) \leq c_0/4, \tag{5.6.23}$$

where we exploited that $M_3(\bar{x}_0) \leq 1$ because $\bar{x}_0$ has support on $\{0, 1\}$.

Combining the inequalities (5.6.22), (5.6.21), (5.6.23), we obtain the following inequality :

$$\mathbb{P}^{(1,d)}_x(\tau^{[R]}_\partial \leq t_M < \tau^{[G],3}_{m_G} \wedge \tau^0) \geq c_0/2.$$

This concludes the proof of Lemma 5.6.6.2. □

5.6.7 Proof of Theorem 5.6.3

5.6.7.1 Choice of the parameters

Before we do this splitting, we need first to associate the law of the first k coordinates between the processes with different initial conditions.

We fix arbitrarily $t_0 = 1$. In the first time-interval of length $t_H \leq t_0$ to be fixed below, we couple the first k coordinates between two processes with different initial conditions. Then, we aggregate the last coordinates in $X_{[R]}$ and impose that $X_{[R]}$ get extinct in the next time-interval of length t_0 while $X^{[N]}$ evolves independently.

By choosing η sufficiently small in the definition of $E^{(d)}$ (with m_3 sufficiently large), we shall get that $k := \lfloor m_3/\eta \rfloor + 1$ is sufficiently large for $X_{[R]}$ to be initialized at a small value. In view of Lemma 5.6.5.3, we wish to control extinction of an upper-bound of the form :

$$\text{for } t \geq t_0, \quad dZ_t := \sqrt{Z_t\,(1 - Z_t)}\,dW_t\,, \qquad Z_{t_0} = z, \tag{5.6.24}$$

where W is a Brownian Motion. Namely, for any $\epsilon > 0$, we choose z in such a way that :

$$\mathbb{P}_z(t_0/2 < \tau_\partial^Z) \leq \epsilon, \qquad \text{where } \tau_\partial^Z := \inf\{t \geq 0\ ;\ Z_t = 0\}. \tag{5.6.25}$$

Now, with the constants C_G, C_M appearing in Proposition 5.6.4.1 for $\delta = 3$ and $t = t_0$, we choose η such that :

$$\eta < (z/C_M) \wedge (\epsilon/(C_G)^2). \tag{5.6.26}$$

Given some $\rho > 0$, we deduce from Theorem 5.6.1 that we can find $m_3 \geq 1$ and $y > 0$ such that, recalling :

$$E^{(d)} := \{x \in \mathcal{X}_d\ ;\ M_3(x) \leq m_3\,,\ \forall\, j \leq \lfloor m_3/\eta \rfloor + 1,\ x_j \geq y\}, \tag{5.6.27}$$

$$\forall\, x \in \mathcal{X}_d, \quad \mathbb{E}_x^{(d)}\left(\exp[\rho\,(\tau_\partial \wedge \tau_{E^{(d)}})]\right) \leq 2. \tag{5.6.28}$$

Exploiting Proposition 5.6.4.1 with $m' = C_M m_3$,, we deduce that for any $x \in E^{(d)}$:

$$\mathbb{P}_x^{(d)}(\tau_{m'}^3 \leq t_0) \leq \frac{C_G m_3}{k} \leq \epsilon. \tag{5.6.29}$$

Recalling that $k := \lfloor m_3/\eta \rfloor + 1$ and $\eta \leq z/C_M$, we deduce that on the event $\{t_0 < \tau_{m'}^3\}$, for any $t_H < t_0$:

$$X_{(k)}(t_H) \leq \frac{C_M \times m_3}{k^3} \leq z. \tag{5.6.30}$$

This provides the initialisation of $X_{[R]}$, that we couple to the original process from time t_H onward thanks to the Markov property. Notably, $X_{[R]}$ is upper-bounded by Z (see definition (5.6.24)).

Note that for any $x \in E^{(d)}$, recalling (5.6.3) :

$$\pi_k(x) \in \mathcal{Y}(y) \qquad \text{where } \mathcal{Y}(y) := \{x \in \mathcal{X}_k \,\Big|\, (\bigwedge_{i[\![0,k]\!]} x_i) > y\}.$$

on which the diffusion term for $(S^{(k)})$, i.e. ($5.1.1 : S^{(d)}$) with d replaced by k, is uniformly elliptic. In practice, we need a bit more space for the Harnack inequality to hold, so that we consider the exit time :

$$T_H := \inf\{t \geq 0 \ ; \ \pi_k(X(t)) \notin \mathcal{Y}(y/2)\} < \tau_\partial. \tag{5.6.31}$$

The probability of such an escape is required to be very small, uniformly in $x \in E^{(d)}$, according to the following lemma.

Lemma 5.6.7.1. *With the above definitions, and $\mathbb{P}^{(k)}$ the law of the system given by $(S^{(k)})$:*

$$\sup_{x\in E^{(d)}} \mathbb{P}^{(k)}_{\pi_k(x)}(T_H \leq t_H) \to 0 \text{ as } t_H \to 0.$$

Since the system $(S^{(k)})$ is uniformly elliptic on $\mathcal{Y}(y/3)$, and recalling Proposition 5.6.4.2, this result is easily deduced from classical results as for instance Proposition V.2.3 in [Ba98].

Thanks to Proposition 5.6.4.2, we can thus find $t_H \leq t_0/2$ sufficiently small such that :

$$\sup_{x\in E^{(d)}} \mathbb{P}^{(k,d)}_x(T_H \leq t_H) \leq \epsilon. \tag{5.6.32}$$

5.6.7.2 Definition of U_A with a control of exceptional events

In this context, with the splitting starting at time t_H, $\tau_\partial^{[R]} := \inf\{t \geq t_H : X_{[R]}(t) = 0\}$.

In view of Theorem 5.6.3, we define $U_A := t_0$ on the event $\{t_H < T_H\} \cap \{t_0 < \tau_3^{m'} \wedge \tau_\partial\} \cap \{\tau_\partial^{[R]} \leq t_0\}$, and otherwise $U_A := \infty$.

Exploiting Proposition 5.6.4.1 and recalling the definition of $E^{(d)}$, m' and t_H, we deduce :

$$\mathbb{P}_x(U_A = \infty, t_0 < \tau_\partial) \leq 2\,\mathbb{P}^{(k,d)}_x(t_H \leq T_H) + 2\,\mathbb{P}^{(k,d)}_x(t_0 < \tau_\partial^{[R]}) + \mathbb{P}^{(d)}_x\left(\tau_3^{m'} \leq t_0\right) \leq 5\,\epsilon.$$

For Theorem 5.6.3, it means that the threshold is obtained with ϵ' such that $\epsilon = \epsilon' \times \exp[-\rho\, t_0]/5$. Since ϵ is freely chosen, so is ϵ'.

5.6.7.3 Comparison of densities

Since the problem is reduced to a finite dimensional one, the Harnack inequality states, as in Section 5.4.3, that there exists $C_H > 0$ such that :

$$\begin{aligned}
&\inf_{x\in\mathcal{X}_d:\pi_k(x)\in\mathcal{Y}(y)} \mathbb{P}_x^{(k,d)}(\pi_k(X_{t_H}) \in dx' \,,\, t_H < T_H \wedge \tau_{m'}^{(k),3}) \\
&\qquad \geq C_H \sup_{x'\in\mathcal{X}_d:\pi_k(x')\in\mathcal{Y}(y)} \mathbb{P}_{x'}^{(k,d)}(\pi_k(X'_{t_H}) \in dx' \,,\, t_H < \tau_\partial \wedge \tau_{m'}^{(k),3}), \qquad (5.6.33)\\
&\quad \text{where } \tau_{m'}^{(k),3} := \inf\{t \geq 0 \ ;\ \sum_{i\geq 0}(i\wedge k)^3 X_i(t) \geq m'\}.
\end{aligned}$$

Let $x, x_\varsigma \in E^{(d)}$. Because of Proposition 5.6.4.1 :

$$\begin{aligned}
&\mathbb{P}_x^{(d)}(X_{U_A} \in dx' \,,\, U_A < \infty) \leq 2\,\mathbb{P}_x^{(k,d)}(X_{t_0} \in dx' \,,\, U_A < \infty)\\
&\quad = 2\mathbb{E}_x^{(k,d)}[\mathbb{P}_{X_{t_H}}^{(k,d)}(X^{[G]}(t_0-t_H) \in dx' \ ;\ \tau_\partial^{[R]} \leq t_0 - t_H < \tau_\partial \wedge \tau_{m'}^3) \ ;\ t_H < T_H \wedge \tau_{m'}^3]\\
&\quad \leq 2\mathbb{E}_x^{(k,d)}[\mathbb{P}_{X(t_H)}^{(k,d)}(X^{[G]}(t_0-t_H) \in dx' \ ;\ \tau_\partial^{[R]} \leq t_0 - t_H < \tau_\partial \wedge \tau_{m'}^{[G],3})\\
&\qquad\qquad\qquad ;\ t_H < T_H \wedge \tau_{m'}^{(k),3}]\\
&\quad \leq 2C_H\,\mathbb{E}_{x_\varsigma}^{(k,d)}[\mathbb{P}_{X(t_H)}^{(k,d)}(X^{[G]}(t_0-t_H) \in dx' \ ;\ \tau_\partial^{[R]} \leq t_0 - t_H < \tau_\partial \wedge \tau_{m'}^{[G],3})\\
&\qquad\qquad\qquad ;\ t_H < T_H \wedge \tau_{m'}^{(k),3}],
\end{aligned}$$

because of (5.6.33) and Lemma 5.6.5.1, noting that τ_∂ and $\tau_\partial^{[R]}$ are measurable with respect to $\sigma(X^{[G]})$.

To go back to $\mathbb{P}^{(d)}$, we will use again Proposition 5.6.4.1. So we need to again ensure upper-bounds on the third moments for the last components for which we lost the information.

For the time-interval $[0, t_H]$, in order to exploit independence as much as possible, we will exploit the representation given in Proposition 5.6.4.2. Since $x_\varsigma \in E^{(d)}$, we have $M_3^{[F]}(0) \leq m_3/y$. From Proposition 5.6.4.4, we thus define m_H such that for any $x_\varsigma \in E^{(d)}$:

$$\mathbb{P}_{x_\varsigma}^{(k,d)}\left(\tau_{m_H}^{[F],3} \leq t_H \,\middle|\, \mathcal{F}^{(k)}\right) \leq 1/2. \qquad (5.6.34)$$

Note that $M_3^{[R]}$, as in Lemma 5.6.5.5, is initialized (at time t_H in this context) by $M_3^{[R]}(t_H) = M_3^{[F]}(X^{[R]}(t_H)) \leq m_H$. Depending on the context for the start of the splitting with $X^{[F]}$, the definition of $\tau_{m_R}^{[F],3}$ may be adapted accordingly.

From Lemma 5.6.5.5, with $t = t_0 - t_H$, we then define m_R such that for any $x \in \mathcal{X}_d$ such that $M_3^{[R]}(x) \leq m_H$:

$$\mathbb{P}_x^{(k,d)}\left(\tau_{m_R}^{[R],3} \leq t_0 - t_H \,\middle|\, \mathcal{F}^{[G]}\right) \leq 1/2. \qquad (5.6.35)$$

From these results, we can come back to (5.6.38) and deduce first from (5.6.35) that, on the event $\{t_H < T_H \wedge \tau^{(k),3}_{m'}\}$:

$$\begin{aligned}&\mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \wedge \tau^{[F],3}_{m_R} \leq t_0 - t_H < \tau_\partial)\\ &\qquad \geq (1/2) \times \mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \leq t_0 - t_H < \tau_\partial)\end{aligned}$$

Then more generally, since $X^{[G]}$ is independent of what happens to $X^{[F]}$ in the time-interval $[0, t_H]$:

$$\begin{aligned}&\mathbb{E}^{(k,d)}_{x_\varsigma}[\mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \leq t_0 - t_H < \tau_\partial \wedge \tau^{[G],3}_{m'} \wedge \tau^{[F],3}_{m_R})\\ &\qquad ; \ t_H < T_H \wedge \tau^{(k),3}_{m'} \wedge \tau^{[F],3}_{m_H}]\\ &\quad \geq (1/2)\, \mathbb{E}^{(k,d)}_{x_\varsigma}[\mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \leq t_0 - t_H < \tau_\partial \wedge \tau^{[G],3}_{m'})\\ &\qquad ; \ t_H < T_H \wedge \tau^{(k),3}_{m'} \wedge \tau^{[F],3}_{m_H}]\\ &\quad \geq (1/4)\, \mathbb{E}^{(k,d)}_{x_\varsigma}[\mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \leq t_0 - t_H < \tau_\partial \wedge \tau^{[G],3}_{m'})\\ &\qquad ; \ t_H < T_H \wedge \tau^{(k),3}_{m'}] \end{aligned} \tag{5.6.36}$$

The upper-bound is then simplified, with the Markov property and the fact that $M_3 \leq M^{[G]}_3 + M^{[R]}_3$. Then, we exploit again Proposition 5.6.4.1 to state that there exists $C_G > 0$, independent of x, such that :

$$\begin{aligned}&\mathbb{E}^{(k,d)}_{x_\varsigma}[\mathbb{P}^{(k,d)}_{X(t_H)}(X^{[G]}(t_0 - t_H) \in dx' \ ; \ \tau^{[R]}_\partial \leq t_0 - t_H < \tau_\partial \wedge \tau^{[G],3}_{m'} \wedge \tau^{[F],3}_{m_R})\\ &\qquad ; \ t_H < T_H \wedge \tau^{(k),3}_{m'} \wedge \tau^{[F],3}_{m_H}]\\ &\quad \leq \mathbb{P}^{(k,d)}_{x_\varsigma}(X(t_0) \in dx' \ ; \ t_0 < \tau_\partial \wedge \tau^{3}_{m'+m_R})\\ &\quad \leq C_G\, \mathbb{P}^{(d)}_{x_\varsigma}(X(t_0) \in dx' \ ; \ t_0 < \tau_\partial). \end{aligned} \tag{5.6.37}$$

Combining (5.6.38), (5.6.36) and (5.6.37) yields that, with $C := 8\, C_H\, C_G > 0$, for any $x, x_\varsigma \in E^{(d)}$:

$$\mathbb{P}^{(d)}_x(X_{U_A} \in dx' \ , \ U_A < \infty) \leq C_G\, \mathbb{P}^{(d)}_{x_\varsigma}(X(t_0) \in dx' \ ; \ t_0 < \tau_\partial).$$

We finally deduce from Lemma 5.6.1.1 and the Markov property that there exists $t_E > 0$ and $C > 0$ such that for any d and $x \in E^{(d)}$:

$$\begin{aligned}&\mathbb{P}^{(d)}_x(X_{U_A} \in dx' \ , \ U_A < \infty)\\ &\qquad \leq C\, \mathbb{P}^{(d)}_{\zeta^{(d)}}(X(t_0 + t_E) \in dx' \ ; \ t_0 + t_E < \tau_\partial).\end{aligned} \tag{5.6.38}$$

This concludes the proof of Theorem 5.6.3. □

5.7 Appendix for the finite dimensional diffusion

We gather here the proofs of respectively Lemmas 5.4.6.2, 5.5.0.3, 5.5.0.4 and 5.5.0.5.

5.7.0.1 Proof of Lemma 5.4.6.2

Given k, ϵ, y, we choose t sufficiently small such that :

$$\inf\left\{\mathbb{P}^{(d)}_x\left(t < U^{k-1}_{y/2}\right)\middle| x \in \mathcal{X}_d, \forall\, j \leq k-1,\ x_j \geq y\right\} \geq 1 - \epsilon/2,$$
$$\text{where } U^{k-1}_{y/2} := \inf\{s \geq 0\ ;\ \exists\, j \leq k-1,\ X_j(s) \leq y/2\}.$$

To ensure roughly the uniformity in such x, we can simply lower-bound X_j for $j \leq k-1$ by solutions to the equation :

$$dY^j_s = -(\lambda + \alpha j)\, dt + \sqrt{Y_t(1-Y_t)} dB(t)\ ,\ Y^j_0 = y,$$

and choose t such that Y^j stays above $y/2$ on the time-interval $[0, t]$ with probability greater than $1 - \epsilon/(2k)$.

Let $y_1 := \lambda\, y/(4\lambda + 4\alpha\, k)$. Then, for any $s \leq U^{k-1}_{y/2} \wedge T^k_{y_1}$:

$$(\alpha M_1(s) - \alpha k - \lambda)\, X_k(s) + \lambda\, X_{k-1}(s) \geq \lambda\, y/2 - (\alpha k + \lambda)\, y_1 \geq \lambda\, y/4,$$

so that X_k is lower-bounded by the solution Y_k of :

$$dY_k(s) = \lambda\, y/4\, dt + \sqrt{Y_t(1-Y_t)} dB(t)\ ,\ Y_k(0) = 0.$$

Since 0 is an entrance boundary for this process, cf e.g. Subsection 3.3.3 in [JT17], there exists $0 < y' \leq y_1 \wedge (y/2)$ such that :

$$\mathbb{P}(\sup_{\{s \leq t\}} Y_k(s) < y') \leq \epsilon/2.$$

On the event $\left\{\sup_{s\leq t} Y_k(s) \geq y'\right\} \cap \left\{t < U^{k-1}_{y/2}\right\}$, which occurs with probability greater than $1-\epsilon$, the condition $T^k_{y'} < t \wedge \tau_\partial$ is indeed satisfied. This ends the proof of Lemma 5.4.6.2.

□

5.7.1 Proof of Lemma 5.5.0.3

This proof is an extension of the one of Proposition 3.8 in [AP13]. W.l.o.g., we assume $\delta \leq \delta_\vee := 1/(16\, \alpha)$. Consider an initial condition x such that $m_1 := M_1(x)$ and $x_0\, m_1 \leq \delta$, where δ is to be fixed later. Thus, on the event $\{\sup_{s\leq t} X_0(s)\, M_1(s) \leq 2\, \delta_\vee\} \cap \{\sup_{s\leq t} X_0(s) \leq 1/2\}$, we have for any $s \leq t$, $\quad (\alpha\, M_1(s) - \lambda)\, X_0(s) \leq 2\, \alpha\, \delta_\vee \leq$

$1/4\,(1 - X_0(s))$. X_0 is thus upper-bounded on this event by the solution Y to :

$$dY_s = (1 - Y_s)/4\, ds + \sqrt{Y_s\ (1 - Y_s)} dB_0(s)\,, \qquad Y(0) = y_0 := \delta/m_1.$$

The main interest of this upper-bound is that it is explicitly given as :

$$Y_t := y_0\ \exp\left[-\int_0^t \frac{1 - Y_s}{4\, Y_s}\, ds + \int_0^t \sqrt{\frac{1 - Y_s}{Y_s}}\, dB_0(s) \right],$$

which is an immediate consequence of Itô's formula. We then define the time-change :

$$\rho_t := \int_0^t \frac{1 - Y_s}{Y_s}\, ds\,, \qquad W_t := M(\rho^{-1}(t)), \text{where } M_t := \int_0^t \sqrt{\frac{1 - Y_s}{Y_s}}\, dB_0(s).$$
$$Y(\rho^{-1}(t)) := y_0\ \exp\left[-t/4 + W_t\right]$$

We can easily check from the quadratic variations that the martingale W is in fact a Brownian Motion. Through conditions on the law of $\exp\left[-t/4 + W_t\right]$ (independent of the parameters), we will thus constrain Y, then X_0.

$$(\rho^{-1})'(t) = (\rho' \circ \rho^{-1}(t))^{-1} = \frac{y_0\ \exp\left[-t/4 + W_t\right]}{1 - y_0\ \exp\left[-t/4 + W_t\right]}.$$

For any $y > 0$, let : $\quad \tau_y^Y := \inf\{t \geq 0\ ;\ Y_t = y\}$ and remark that for any $\mu > 0$:

$$\{t < \tau_0^Y\} = \{t < \rho^{-1}(\infty)\} = \left\{ t < \int_0^\infty \frac{y_0\ \exp\left[-r/4 + W_r\right]}{1 - y_0\ \exp\left[-r/4 + W_r\right]} dr \right\}$$

On the event $\{\tau_0^Y < \tau_{y_0+\mu}^Y\}$, for any $t \geq 0$: $y_0 e^{-t/4 + W_t} < y_0 + \mu$, so that one can have an explicit upper-bound of $(1 - y_0 e^{-t/4+W_t})^{-1}$. On the event $\{\tau_{y_0+\mu}^Y < \tau_0^Y\}$, there must exist $t \geq 0$ such that $y_0 e^{-t/4+W_t} = y_0 + \mu$. From these, we deduce :

$$\mathbb{P}_{y_0}(t < \tau_0^Y < \tau_{y_0+\mu}^Y) \leq \mathbb{P}\left(\frac{t\,(1 - y_0 - \mu)}{y_0} < \int_0^\infty \exp\left[-r/4 + W_r\right] dr \right), \tag{5.7.1}$$

$$\mathbb{P}_{y_0}(\tau_{y_0+\mu}^Y < \tau_0^Y) = \mathbb{P}\left((y_0 + \mu)/y_0 < \sup_{r \geq 0} \exp\left[-r/4 + W_r\right] \right). \tag{5.7.2}$$

Let $\epsilon > 0$. Since $W_t/t \xrightarrow[t\to\infty]{} 0$, we can define $c_1, c_2 \geq 1$ such that

$$\mathbb{P}\left(c_1 < \int_0^\infty \exp\left[-r/4 + W_r\right]\, dr \right) \leq \epsilon\,,\ \mathbb{P}\left(c_2 < \sup_{r \geq 0} \exp\left[-r/4 + W_r\right] \right) \leq \epsilon.$$

Likewise, from Lemma 3.2 in [AP13], we can find $c_3 > 0$ such that for any d and

$x \in \mathcal{X}_d$:

$$\mathbb{P}_x(\sup_{\{s\le t\}} M_1(s) - M_1(0) \le \lambda\, t + c_3) \le \epsilon.$$

This motivates : $\quad m_1' := m_1 + \lambda\, t + c_3$, $\mu := \delta_\vee / m_1'$.

We choose also $\delta \le \delta_\vee$ sufficiently small to ensure, with $m_1 \ge 1$:

$$\frac{t\,(1-y_0-\mu)}{y_0} \ge \frac{m_1\, t}{\delta} \times (1 - 2\frac{\delta_\vee}{m_1'}) \ge c_1\,, \qquad \frac{y_0+\mu}{y_0} \ge \frac{\mu}{y_0} \ge \frac{\delta_\vee}{\delta} \times (1+\lambda\, t + c_3)^{-1} \ge c_2.$$

Thus, from equations (5.7.1) and (5.7.2) and the above definitions :

$$\mathbb{P}_x(\mathcal{A}) \ge 1 - 3\,\epsilon, \qquad \text{where } \mathcal{A} := \left\{\sup_{s\le t} M_1(s) \le m_1'\right\} \cap \left\{\tau_0^Y < t \wedge \tau_{y_0+\mu}^Y\right\}.$$

To check the upper-bound by Y, let $T_{01} := \inf\{s \ge 0\ ;\ X_0(s)\, M_1(s) \ge 2\,\delta_\vee\}$. Then, on the event $\mathcal{A}$, for any $s \le t \wedge T_{01}$ (with $m_1 \ge 1$) :

$$X_0(s) \le Y_s \le y_0 + \mu\ ,\ M_1(s) \le m_1'$$
$$\text{so } X_0(s)\, M_1(s) \le \delta\,(1 + \lambda\, t + c_3) + \delta_\vee < 2\,\delta_\vee,$$

where we can use the inequality on $c_2 > 1$ to deduce the last upper-bound. By continuity of $X_0\, M_1$, $T_{01} < t$ is incompatible with $\mathcal{A}$, so that we indeed have $\forall s \le t,\ X_0(s) \le Y_s$, thus $\tau_\partial \le t$. In conclusion, for any x such that $m_1 \ge 1$ and $x_0\, m_1 \le \delta$:

$$\mathbb{P}_x^{(d)}(\tau_\partial \le t) \ge \mathbb{P}_x^{(d)}(\mathcal{A}) \ge 1 - 3\,\epsilon. \qquad \square$$

5.7.2 Proof of Lemma 5.5.0.4

On the event $\{\inf_{s\le t} M_1(X_s) \ge m_1\}$, X_0 is lower-bounded on $[0,t]$ by the solution Y to :

$$dY_s = r(m_1)\, Y_s\, ds + \sqrt{Y_s\ (1-Y_s)}dB_0(s)\ , \qquad Y(0) = y_0,$$
$$\text{where} \quad r(m_1) := \alpha\, m_1 - \lambda \underset{m_1\to\infty}{\longrightarrow} \infty.$$

Since $M_1(s) = 0$ as soon as $X_0 = 1$, this lower-bound cannot hold until $T_1^Y := \inf\{t \ge 0\ ;\ Y_t \ge 1\}$. We thus only have to prove that $\mathbb{P}(t < T_1^Y) \to 0$ as $m_1 \to \infty$.

Let $\epsilon, t_1 > 0$. The quadratic variation of the martingale part $\mathcal{M}_s$ until time $t_1 \le t$ is upper-bounded by t_1, so that the Doob inequality implies :

$$\mathbb{P}_{y_0}(\sup_{s\le t_1} |\mathcal{M}_s| > y_0/2) \le 8t_1/y_0^2. \tag{5.7.3}$$

By choosing t_1 sufficiently small, we can assume $8t_1/y_0^2 \le \epsilon$.

On the event $\left\{\sup_{s\leq t_1} |\mathcal{M}_s| \leq y_0/2\right\}$, it is clear that Y stays above $y_0/2$ on the time-interval $[0, t_1]$. The drift term can thus be lower-bounded by $r(m_1)\, s\, y_0/2$ for any $s \leq t_1 \wedge T_1^Y$. Since it cannot exceed $1 - y_0/2$ before T_1^Y, it necessarily implies that for $r(m_1)$ sufficiently large (that is m_1 sufficiently large), we must have $T_1^Y < t_1$ on the event $\left\{\sup_{s\leq t_1} |\mathcal{M}_s| \leq y_0/2\right\}$. With (5.7.3) and $t_1 \leq t$, this clearly implies $\mathbb{P}(t < T_1^Y) \to 0$ as $m_1 \to \infty$ and concludes the proof of Lemma 5.5.0.4. □

5.7.3 Proof of Lemma 5.5.0.5

For any $k \geq 1$:

$$dM_k(t) = \alpha\, (M_1(t)M_k(t) - M_{k+1}(t))\, dt + \lambda \sum_{j\leq k-1} c_j^k\, M_j(t)dt + d\mathcal{M}_k(t),$$

where $\mathcal{M}_k(t)$ is a continuous martingale, and $\langle \mathcal{M}_k \rangle_t = \int_0^t (M_{2\,k}(s) - M_k(s))^2\, ds$.

For some $m_1^{[k]}$ to be defined later, depending on ϵ, let $\tau_1^{[k]} := \inf\{t \geq 0\,;\, M_1(t) \geq m_1^{[k]}\}$ so that, since M_k is increasing with $k \geq 1$:

$$0 \leq \mathbb{E}_x^{(d)}(M_k(t\wedge \tau_1^{[k]}) \leq M_k(x) - \alpha\, \mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_{k+1}(s)\, ds\right) + C_k\, \mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_k(s)\, ds\right)$$

$$\mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_{k+1}(s)\, ds\right) \leq m_k/\alpha + C_k'\, \mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_k(s)\, ds\right).$$

By some immediate induction :

$$\begin{aligned}\mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_{k+1}(s)\, ds\right) &\leq (k-1)\, m_k/\alpha + C_k''\, \mathbb{E}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_1(s)\, ds\right)\\ &\leq (k-1)\, m_k/\alpha + C_k''\, t\, m_1^{[k]}.\end{aligned}$$

For any $\epsilon > 0$, we then use Lemma 3.2 in [AP13] together with $M_1(x) \leq m_k$ to find $m_1^{[k]}$ such that : $\quad \mathbb{P}_x^{(d)}(\tau_1^{[k]} < t) \leq \epsilon \quad$ for any $x \in \mathcal{X}_d$ such that $M_k(x) \leq m_k$. Now that $m_1^{[k]}, \tau_1^{[k]}$ is clearly defined, we can find, by the Markov inequality, m_{k+1} such that :

$$\mathbb{P}_x^{(d)}\left(\int_0^{t\wedge\tau_1^{[k]}} M_{k+1}(s)\, ds \geq t\, m_{k+1}\right) \leq \epsilon.$$

This concludes the proof since :

$$\left\{\inf_{s\leq t} M_{k+1}(s) \geq m_{k+1}\right\} \cap \left\{t \leq \tau_1^{[k]}\right\} \subset \left\{\int_0^{t\wedge\tau_1^{[k]}} M_{k+1}(s)\, ds \geq t\, m_{k+1}\right\}. \qquad \square$$

5.8 Appendix for the general unbounded case

We gather here the proofs of respectively Propositions 5.6.4.3, 5.6.4.4, then Lemmas 5.6.4.5 and 5.6.1.1, while Section 5.8.1.1 is involved in the proof of Proposition 5.6.4.3.

5.8.1 Proof of Proposition 5.6.4.3

Let $t > 0$, $\delta > 1$, $m' \geq m$, $d \in \mathbb{N} \cup \{\infty\}$ and $x \in \mathcal{X}_d$ such that $M_\delta(x) \leq m$ be fixed. The proof generalizes the one of Lemma 5.5.0.2 in the case where the martingale part is a priori only local.

We again consider the semi-martingale decomposition of M_δ :

$$dM_\delta(t) = V_\delta(t)\, dt + d\mathcal{M}_\delta(t), \tag{5.8.1}$$

where $\mathcal{M}_\delta$ is a continuous local martingale starting from 0, whose quadratic variation is

$$\langle \mathcal{M}_\delta \rangle_t = \int_0^t (M_{2\,\delta}(s) - M_\delta(s)^2)\, ds,$$

and V_δ is a bounded variation process. Thanks to Theorem 3 in [AP13], since $M_{2\delta}(x) < \infty$, we know that $(M_{2\delta}(t))_{t\geq 0}$ is a.s. finite. This expression for the quadratic variation is thus well-defined. Similarly as in Proposition 5.5.0.2 and thanks to the Hölder inequality, there exists $C = C(\delta) = \lambda(2^\delta - 1)$ such that :

$$V_\delta \leq C\, M_\delta + \lambda. \tag{5.8.2}$$

To obtain an upper-bound on the probability that $\sup_{s\leq t} M_\delta(s)$ is large, we wish to exploit the Doob inequality on a positive sub-martingale. The lemma below, whose proof is deferred to the next Subsection, makes it possible.

Lemma 5.8.1.1. *Let $\delta, t > 0$ be given. There exists $C_M = C > 0$ such that for any $m \geq 1$, $x \in \mathcal{X}^{2\delta}$ such that $M_\delta(x) \leq m$, there exists a sequence of positive sub-martingale $\widehat{M}_\delta^{(k)}$ stopped at times T_k, where $T_k \to \infty$, such that for any $s \leq t \wedge T_k$, $M_\delta(s) \leq \widehat{M}_\delta^{(k)}(s)$, and such that :*

$$\mathbb{E}_x[\widehat{M}_\delta^{(k)}(t)] \leq C_M\, m.$$

By exploiting Doob's inequality on $\widehat{M}_\delta^{(k)}$, we obtain :

$$\mathbb{P}_x(\sup_{s\leq t\wedge T_k} M_\delta(s) > m') \leq \mathbb{P}_x(\sup_{s\leq t} \widehat{M}_\delta^{(k)}(s) > m')$$
$$\leq \frac{\mathbb{E}_x[\widehat{M}_\delta^{(k)}(t)]}{m'} \leq \frac{C_M\, m}{m'}.$$

We know let $T_k \to \infty$ and conclude the proof of Proposition 5.6.4.3 by showing that :

$$\mathbb{P}_x(\tau_{m'}^\delta \leq t) \leq \frac{C_M\, m}{m'}.$$

□

5.8.2 Proof of Lemma 5.8.1.1

The sequence T_k, for $k \geq 1$, is introduced to localize $\mathcal{M}_\delta$ and have an upper-bound on $M_\delta^{(k)}$.

$$T_k := \inf\{s \geq 0 \,\big|\, \langle\mathcal{M}_\delta\rangle_s \geq k \;,\; M_\delta^{(k)}(s) \geq k\}.$$

Recalling Theorem 3 in [AP13], for the fact that $(M_{2\delta}(t))_{t\geq 0}$ is a.s. finite, we easily deduce that $T_k \to \infty$ as $k \to \infty$.

We wish to characterize $\widehat{M}_\delta^{(k)}$ as the solution to the following equation, valid for $s \leq t \wedge T_k$:

$$\widehat{M}_\delta^{(k)}(s) = m + \int_0^s (C\widehat{M}_\delta^{(k)}(r) + \lambda)dr + \mathcal{M}_\delta^{(k)}(s). \tag{5.8.3}$$

We use Duhamel's formula to first define what will be $\widehat{M}_\delta^{(k)}(s) - M_\delta^{(k)}(s)$ on the event $\{s \leq T_k\}$:

$$E(s) := (m - M_\delta(x))\, e^{Cs} + e^{Cs} \int_0^s e^{-Cr}(CM_\delta(r) + \lambda - V_\delta(r))dr.$$

Note that for any $s < T_k$, because of inequality (5.8.2) and the definition of T_k, this expression is both positive and upper-bounded by a deterministic constant. Let us check that $\widehat{M}_\delta^{(k)}(s) := M_\delta^{(k)}(s \wedge T_k) + E(s \wedge T_k)$ is indeed solution to equation (5.8.3).

Let $s \leq T_k$ and compute :

$$\begin{aligned}
E(s) = e^{Cs} \times (e^{-Cs}E(s)) &= (m - M_\delta(x)) + \int_0^s e^{Cr} \times e^{-Cr}(CM_\delta(r) + \lambda - V_\delta(r))dr \\
&\quad + \int_0^s Ce^{Cr} \times (e^{-Cr}E(r))dr \\
&= (m - M_\delta(x)) + \int_0^s [C(E(r) + M_\delta(r)) + \lambda - V_\delta(r)]dr \\
&= \widehat{M}_\delta^{(k)}(0) - M_\delta(0) - \int_0^s V_\delta(r)dr + \int_0^s [C\widehat{M}_\delta^{(k)}(r) + \lambda]dr,
\end{aligned}$$

from which it is clear that $\widehat{M}_\delta^{(k)}(s)$ is indeed solution.

Recalling that E is positive, we immediately deduce that $\widehat{M}_\delta^{(k)}(s) > M_\delta(s)$ for any $s \leq T_k$. Since M_δ is non-negative by definition, it is also the case for $\widehat{M}_\delta^{(k)}$. Because of (5.8.3), with the fact that $\widehat{M}_\delta^{(k)}$ stays fixed after T_k, this proves that $\widehat{M}_\delta^{(k)}$ is a positive sub-martingale and that for any $s \leq t$:

$$\mathbb{E}_x[\widehat{M}_\delta^{(k)}(s)] \leq (m + \lambda t) + C \int_0^s \mathbb{E}_x[\widehat{M}_\delta^{(k)}(r)]dr.$$

Recall that both $M_\delta(s)$ and $E(s)$, are upper-bounded for any $s \leq t \wedge T_k$, by a uniform constant depending on t and k. This implies a similar upper-bound on $\mathbb{E}_x[\widehat{M}_\delta^{(k)}(r)]$, that guaranties that we are in conditions to apply Gromwall's Lemma, see for instance Proposition 6.59 in [PR14]. From this we deduce :

$$\mathbb{E}_x[\widehat{M}_\delta^{(k)}(t)] \leq (m + \lambda t)\, e^{Ct}.$$

This concludes the proof of Proposition 5.6.4.3 with $C_M := (1 + \lambda t)\, e^{Ct}$ (recalling $m \geq 1$).

□

5.8.3 Proof of Proposition 5.6.4.4

Under $\mathbb{P}^{(k,d)}$, we exploit the Itô formula and distinguish the part involving the Brownian Motions. $M_3^{[F]}$ is solution of :

$$dM_3^{[F]}(t) := V_3^{[F]}(t)dt + d\mathcal{M}_3^{[F]}(t), \tag{5.8.4}$$

where $V_3^{[F]}$ is a bounded variation process defined as :

$$V_3^{[F]} := \lambda \frac{X_{k-1}}{X_{(k)}}(k^3 - M_3^{[F]}(t)) + \lambda \sum_{\ell \geq k} (\ell + 1)^3 X_\ell^{[F]} - \lambda M_3^{[F]}. \tag{5.8.5}$$

Note that whatever the values of $\frac{X_{k-1}}{X_{(k)}}$, with the rough estimate $(\ell+1)^3 \leq 8\ell^3$ for $\ell \geq 1$, we always have $V_3^{[F]} \leq 7\lambda\, M_3^{[F]}$. On the other hand, $\mathcal{M}_3^{[F]}$ is defined as :

$$\mathcal{M}_3^{[F]} := \sum_{i\geq k} i^3 \left[\sqrt{\frac{X_i^{[F]}(t)}{X_{(k)}(t)}} dW_t^{[F],i} - \frac{X_i^{[F]}(t)}{\sqrt{X_{(k)}(t)}} dW_t^{[F]} \right]. \tag{5.8.6}$$

Relying on the same calculations as for M_3, $\mathcal{M}_3^{[F]}$ is a continuous local martingale starting from 0 for the filtration $\mathcal{F}_t^{(k)}$ whose quadratic variation is

$$\langle \mathcal{M}_3^{[F]} \rangle_t = \int_0^t \frac{\mathcal{M}_6^{[F]}(s) - (\mathcal{M}_3^{[F]}(s))^2}{X_{(k)}(s)}\, ds.$$

The rest of the proof of Proposition 5.6.4.4 can be easily adapted from the one of Proposition 5.6.4.3 (with $C = \exp[7\lambda t_H] \vee 1$). □

5.8.4 Proof of Lemma 5.6.4.5

While applying the Girsanov formula, we shall consider the exponential martingale of :

$$L_t^{(k)} := -\alpha \sum_{i\geq k+1} (i-k) \int_0^t \sqrt{X_i(s)}\, dW_s^i + \alpha \int_0^t R_1^{(k)}(s)\, dW_s$$

By this choice, we obtain the following equalities :

$$\begin{aligned} d\langle L^{(k)}, W^i\rangle_s &= \alpha \left[M_1(s) - M_1^{(k)}(s) - (i-k)_+ \right] \sqrt{X_i(s)}\, ds, \\ d\langle L^{(k)}, W\rangle_s &= -\alpha \sum_{i\geq k+1} (i-k)\, X_i(s)\, ds + \alpha\, R_1^{(k)}(s)\, ds = 0, \\ d\langle L^{(k)}\rangle_s &= -\alpha \sum_{i\geq k+1} (i-k) \sqrt{X_i(s)} d\langle L^{(k)}, W^i\rangle_s = \alpha^2 \left[R_2^{(k)}(t) - R_1^{(k)}(t)^2 \right] \end{aligned}$$

The coupling can thus indeed be given by the exponential martingale associated to $L^{(k)}$, i.e. :

$$\begin{aligned} \log \frac{d\mathbb{P}_x^{(k,d)}}{d\mathbb{P}_x^{(d)}}_{|[0,t]} &= -\alpha \sum_{i\geq k+1} (i-k) \left[\int_0^t \sqrt{X_i(s)}\, dW_s^i - \int_0^t X_i(s)\, dW_s \right] \\ &\quad - \frac{\alpha^2}{2} \int_0^t \left[R_2^{(k)}(s) - R_1^{(k)}(s)^2 \right] ds. \end{aligned}$$

This expression can be stated as in Lemma 5.6.4.5 in term of the solution to $\mathbb{P}^{(d)}_x$ by noting the following equality˜ :

$$dR_1^{(k)}(s) = \alpha(M_1(s) - k)\, R_1^{(k)}(s)\, ds + \alpha\, R_2^{(k)}(s) + \lambda\, X_{(k)}(s)\, ds \\ + \sum_{i \geq k+1} (i-k) \left[\int_0^t \sqrt{X_i(s)}\, dW_s^i - \int_0^t X_i(s)\, dW_s \right].$$

□

5.8.5 Proof of Lemma 5.6.1.1

We have thanks to Theorem 5.6.1 a uniform control on the time of coming back to $E^{(d)}$, provided we can handle the survival starting from $\zeta^{(d)}$. From (5.6.15) and the definition of τ^0, we know that $X_0 \geq 1/10\ \zeta^{(d)}$-a.s. Whatever d, we deduce that under $\mathbb{P}^{(d)}_{\zeta^{(d)}}$, $(X_0(s))$ is lower-bounded by the solution Y to the equation :

$$dY_s = -\lambda\, ds + \sqrt{Y_s\ (1 - Y_s)} dB_0(s)\, , \qquad Y(0) = 1/10.$$

Thus, denoting $c := \mathbb{P}_{1/10}(Y_{t/2} > 0)/2 > 0$ independent of d, we have uniformly :

$$\mathbb{P}^{(d)}_{\zeta^{(d)}}(t/2 < \tau_\partial) \geq 2c. \tag{5.8.7}$$

Thanks to Theorem 5.6.1, we then deduce $m_3, y > 0$ such that for any d and $x \in \mathcal{X}_d$:

$$\mathbb{P}^{(d)}_{\zeta^{(d)}}(t/2 < \tau_\partial \wedge \tau_{E^{(d)}}) \leq c.$$

Combined with (5.8.7), this concludes the proof of Lemma 5.6.1.1. □

6 Individual-based models under various time-scales

This chapter is taken from the preprint with the same name whose ArXiv reference can be found at the end of the bibliography (here is the link for the pdf version : [Chapter 6]). It has recently been accepted for ESAIM Proceedings et Surveys and shall be released with the 68th volume.

Abstract

This article is a presentation of specific recent results describing scaling limits of individual-based models. Thanks to them, we wish to relate the time-scales typical of demographic dynamics and natural selection to the parameters of the individual-based models. Although these results are by no means exhaustive, both on the mathematical and the biological level, they complement each other. Indeed, they provide a viewpoint for many classical time-scales. Namely, they encompass the time-scale typical of the life-expectancy of a single individual, the longer one wherein a population can be characterized through its demographic dynamics, and at least four interconnected ones wherein selection occurs. The limiting behavior is generally deterministic. Yet, since there are selective effects on randomness in the history of lineages, probability theory is shown to be a key factor in understanding the results. Besides, randomness can be maintained in the limiting dynamics, for instance to model rare mutations fixing in the population.

Les résultats récents présentés dans cet article décrivent des limites d'échelles de modèles individus-centrés. Grâce à eux, nous allons mettre en lumière les différentes échelles de temps caractéristiques des variations démographiques et de la sélection naturelle. L'objectif n'est pas d'être exhaustif, tant au niveau mathématique que biologique. Néanmoins, ces résultats sont très complémentaires les uns des autres. Ils fournissent un aperçu des principales échelles de temps d'intérêt. Ils englobent notamment l'échelle de temps de la vie d'un individu, celle plus longue où la population peut être décrite comme une entité avec ses caractéristiques propres qui dirigent les dynamiques démographiques, et au moins quatre autres échelles imbriquées sur lesquelles la sélection joue un rôle. La dynamique limite est généralement déterministe. Pour autant, puisque l'action de la sélection se base sur l'aléa présent dans l'histoire

des individus, les probabilités apparaissent comme un élément clé pour comprendre ces résultats. Par ailleurs, la stochasticité peut aussi être conservée à la limite, par exemple pour modéliser de rares événements de fixation de mutations.

Introduction

The origin of this paper is the session "Stochastic Processes and Biology" during the conference Journées MAS in Dijon at the end of August 2018. Four talks were presented, with a variety of methods and probabilistic approaches. Their common feature, that motivates this presentation, is the fact that they start from similar Individual-Based Models (IBM) and justify an asymptotic behavior as both the population size and the observation time are large. The time-scales wherein these observations can be obtained are nonetheless different, yet interconnected. The structure of the paper aims both at presenting the principal objects of the four talks of the session and at reflecting this increasing sequence of time-scales involved in adaptation of populations.

For more than fifteen years, there has been a significant activity around a new probabilistic framework for mathematical modeling of ecology, population genetics and trait evolution. Traits are any features of the individuals whose transmission the modeler is interested in, and we refer to the beginning of Chapter 1 for some classical examples. They may vary during the life-time of the individuals, at births or only at rare mutation events. The population is described by the distribution of its traits, from which one may for instance deduce its growth rate. Any effect of natural selection should depend on these traits and affect their distribution.

The IBM are a priori the best formulation for validation of a macroscopic model of population dynamics. These processes detail at individual level births, deaths and interactions in the population. The description may also specify migration patterns, aging or competition for resources between individuals. This setting is clearly the closest to actual simulations designed by computational biologists to validate theoretical models, or simply some predictions. In this view, such setting is also the one with the least number of simplifying assumptions. It is also certainly the one where calculations of probabilities of events are the most complicated. One of the first aims of such probabilistic modeling has actually been to connect these individual-based models and simulations to much simpler systems of Ordinary Differential Equations (ODE) or Partial Differential Equation (PDE) models.

To describe natural selection on heritable traits, systems of ODE usually provide the most classical models to simply express their effects. They have been refined by introducing PDE to deal with a continuum of traits in populations. Thanks to these models, one can cover most of contemporary mathematical models of ecology, population genetics and character evolution. Yet, this deterministic approach does not take into account the variability observed when reproducing similar experiments. A part of this variability is due to external perturbations. Stochastic Differential Equations (SDE) have been introduced to model these effects. They usually assume that the parameters appearing in the systems of ODE / PDE are themselves subject to stochasticity. The other main component of this noise is internal and inherent to the discreteness of the system : selection between traits emerges through competition between a finite number of individuals. Taking this component into account is the purpose of these IBM.

To justify simpler models from the IBM, the usual assumption is that populations are large enough so that the law of large numbers makes random fluctuations negligible. Given the finite population sizes ecologists are interested in, an elementary description may well be accurately designed while a finer structuring of the population is much more demanding.

It is rather natural for mathematicians to try to describe simplified behavior through an asymptotic on large time-scales (i.e. for large number of generations). It is at the core of Darwin's theory of natural selection to assume a separation of time-scale between apparently incidental randomness among contemporary individuals and the evolution of traits, and even of species. The goal is surely to obtain exact results of convergence to a non-trivial process for the new dynamics in a new time-scale. It is however very challenging, and perhaps an approximation cannot be justified in this way, while still being helpful. There is also no reason to restrict ourselves purely to deterministic limiting processes : a first order of fluctuations can be inferred from central limit theorems. For instance, it may take the form of a solution to a specific SDE. Such a next order notably helps to specify the conditions of validity of the deterministic approximation.

An archetypal example is the stochastic approximation of the population size in a continuous setting (i.e. with Brownian-type fluctuations). Beyond such a result is the idea that individual events (of births or deaths) have quite a negligible impact on the population as a whole, yet randomness still occurs through varying accumulation of such effects. Thus, to obtain the convergence towards a stochastic continuous process, one needs (besides the population size going to infinity) to amplify the frequency of births and deaths also. Notably, this means that selective effects appear in a much longer time-scale than the life-expectancy of individuals (cf. in particular chapter 3 in [BM15]). The scale of population size provides the essential relation between the parameters at individual time-scale (on this time-scale, the birth rate is around 1) and at population level (cf. section 6.1). In case there is more than one parameter to adjust, analysis may be even less intuitive and may even lead to different behaviors associated with different scalings.

Each of the four talks given during the above-mentioned session constitutes the foundation of each of the four sections to follow. The results that we will present are taken from the referenced articles, and the reader is spared some details on the underlying assumptions. These choices and the discussions that complement these results express my personal view on the subject, and do not engage the authors of the articles. My aim while reporting these results is to highlight the variety of time-scales involved in population models. These population models are simpler to analyze and quicker to simulate than the IBM for large populations. We will thus specify how large the population sizes shall be in order to have valid estimations of individual-based models through population models.

In Section 6.1, we shall present the work of Hoffmann and Marguet [HM19]. IBM are introduced in terms of traits that structure the population and evolve in this section during the life-time of individuals. The estimation of this trait dynamics is the purpose of their paper, in which specific estimators are proposed and their accuracy evaluated. From this evaluation, we will obtain some insight into the time-scale required for different populations (with different trait dynamics) to be distinguished. The following sections are devoted to evolution and the associated models rely on traits that are much more accurately transmitted to descendants. Although this previously described variability is usually neglected in evolutionary studies (only some average effect is considered), we shall see in Section 6.2 that it may play a significant role when two components of selection induce competing effects. Both the strength and the time-scale of selection may in fact depend on these fluctuation effects that are generally considered neutral for evolution. This is one of the main conclusions that I wished to discuss based on my work in Chapter 4. Assuming however that this variability has only a minor effect on the 'local' reproductive value of individuals and rescaling time properly, Champagnat and Henry describe in [CH19] the process of evolution as an almost deterministic behavior. This work and the connections to the two previous sections will be our concern in Section 6.3. Finally, natural selection may rely on the emergence of very rare mutations, as in Tran's talk. The associated time-scale depends on the occurrence rate and the mean effect of successful mutations (that invade the population). Among several possibilities, I have chosen to present the models from [BF+15] and [BBC17] in Section 6.4.

6.1 Recognition of population traits from individual dynamics

This section is devoted to the results of Hoffmann and Marguet. For simplicity, the model neglects any interaction between individuals other than transmission of specified traits from parents to offspring, what we call a model of branching population. Assuming that changing traits induce a structure on this branching population, they propose a statistical estimation of this dynamics of traits [HM19]. This is the right place to discuss this notion of individual-based models and how to relate individual histories to the law that governs these behaviors. It gives some insight into the number of generations involved before one can actually distinguish distinct population models. On a shorter time-scale, natural selection may favor some trajectories by mere chance.

Yet those lucky behaviors should be observed with close frequencies between distinct populations. This selection shall thus hardly lead to any heritable effect. For instance, as presented in [HM19], any statistical estimation of traits requires a sufficiently large population history. Likewise, a long-term history is required for natural selection to favor one subpopulation over another (with different inheritance). Moreover, an initially very favorable combination of trait and environment might not be so beneficial in the long term : most descendants might be exposed later on to a much less favorable environment given their traits. This "sample size" shall thus be larger when there are strong correlations between successive generations. In cases where such correlations are sufficiently weak, Hoffmann and Marguet (cf. [HM19]) study how to take them into account in statistical analysis. They indicate the accuracy of some estimators in terms of the size of the genealogical tree on which observations are indexed. The focus in [HM19] is on the random process governing birth events, for which they propose the first statistical analysis in a structured population.

At an individual-based level, each individual is characterized by some value $x \in \mathcal{X}$, that we call a trait. Typically $\mathcal{X} \subset \mathbb{R}^k$ or $\mathcal{X} \subset \mathbb{R}^k \times F$, where we may consider k characteristics such as size, spatial position, age, the amount of certain proteins, of certain ressources, of parasites and so on, while F denotes a finite (or numerable) set of classes, such as sex, eye color etc. Before the death of the individual, this trait x may be constant, but it can also evolve according to some stochastic differential equation (SDE). Interaction with other individuals can then appear in parameters of this equation, possibly depending on the trait x' of the other individuals around. Death of individuals happens at a rate d, that may depend first on the trait x of the individual, but also on the whole population and eventually its effect on the environment (as is the case in Section 3). We recall that we could include the age of the individual in the information carried by x, so that there is almost no restriction. Death at a fixed age is also not difficult to include in the model. Same kind of dependencies may be considered for births, and apply to the number of offspring and their states at birth. Independent exponential variables or Poisson Point Processes are thus usually a very convenient way to encode all of these events. In a nutshell, one can a priori represent with stochastic IBM models any computational model that one could design for validation of biological predictions, as soon as one can follow the individuals one by one. I refer notably to Subsection 6.4.2 for a more concrete description of such measure-valued process. In the following, we will focus the analysis on continuous trait-spaces, for which one can exploit the regularity of the estimated functions.

6.1.1 Discussion on the assumptions in [HM19]

6.1.1.1 A process on an incomplete tree

The main specificity of this statistical approach is to deal with the dependency given by the genealogy of individuals. For instance, looking at the size as a trait, the larger the mother-cell is, the larger its daughter-cells shall be. With the experiments where one follows each cell individually, one can also easily obtain the entire associated lineages. Yet, many of these cells are no longer observable because of the design of experimental processes. Although the authors do not mention it, death of some cells is likely to be included as well. The statistical analysis of the traits thus relies on data that is indexed by some tree that is "incomplete" as compared to the complete binary tree. A more detailed definition on the assumption on the tree is given below. In order to relate the accuracy of the estimators to the number of birth events, the authors consider both the cases of a bounded population size at each generation and the case of an population size expanding at a given growth rate :

Definition 2. *Consider the complete binary tree, on which our tree will be indexed. Define the generation n of this complete tree as G_n. Note that $Card(G_n) = 2^n$.*

A regular incomplete tree is a family of subsets U_n of the complete tree up to generation n such that :

(i) the parent of any individual of U_n is also in U_n

and (ii) for some $0 \leq \rho \leq 1$:

$0 < \liminf_{n\to\infty} 2^{-\rho n} Card(U_n \cap G_n) \leq \limsup_{n\to\infty} 2^{-\rho n} Card(U_n \cap G_n) < \infty$.

In the case $\rho \in (0, 1]$, the data characterize a growing population of cells. For instance, it corresponds to an experiment where one lets these cells duplicate freely in a rich medium. In the case $\rho = 0$, one rather expects some population size at equilibrium, although it is not required here. It also corresponds to the experiments where only one cell is followed after each division (which is the case in recent experiments in a microfluidic environment : the mother-machine).

In the following results, this indexation tree is considered as given. It means that we consider the trait dynamics conditionally on the realization of the tree. Yet, the interest of the experimenter is a priori rather on intrinsic parameters without conditioning. Thus, there is an implicit assumption that the shape of the indexation tree is independent of the trait dynamics.

With a single lineage experiment, this is not really an issue. But if one has to consider natural death events while estimating variable birth rates, assuming the independence seems much more questionable. It appears a bit surprising to have a birth rate depending on the trait while the probability of giving birth in one life-span is independent of it. A constant death rate would be a more natural assumption than a prescribed tree. Yet, it would be much more difficult to analyze and would possibly not reflect the design of the experiment very well.

6.1.1.2 Ergodicity of the trait

The other main assumption in [HM19] is the very rapid convergence to the ergodic measure ν in the transmission of the trait at birth from the parent to its offspring. Assuming that $u \in U_n \cap G_n$ describes the parental index of an individual with index $v \in U_{n+1} \cap G_{n+1}$, the trait at birth X_v of the latter is given by $Q(X_u, dx_v)$ where X_u is the trait at birth of the parent. The convergence of the estimators is stated uniformly as long as there exist $C > 0$ and a positive weight-function V such that for any $m \geq 1$, and φ satisfying $\|\varphi\|_V := \sup_x |\varphi(x)|/[1 + V(x)] < \infty$:

$$\|Q^m(\varphi) - \nu(\varphi)\|_V \leq C\, 2^{-m}\, \|\varphi - \nu(\varphi)\|_V. \tag{6.1.1}$$

The set of such Q is denoted $\mathbb{Q}_{1/2}$. Note in this view that the authors choose exploit exclusively traits at birth for their statistical estimations.

6.1.1.3 Definition of the stochastic process, regularity and confinement

The model in [HM19] is concerned with a trait supposedly governed during the lifetime of each individual by a stochastic flow of the form : $dX_t = b(X_t)\, dt + \sigma(X_t) dW_t$, where W refers to a standard Brownian Motion, and (X_t) evolves on $\mathcal{X} \subset \mathbb{R}^k$. Birth occurs at time s at rate $B(X_s)$. At reproduction event, the authors assume that the value of the trait is distributed between two offspring with the following mechanism : given an independent r.v. θ, the trait at birth of the two offspring is given respectively by θy and $(1 - \theta) y$. θ is drawn according to $\kappa(y) dy$, for some probability density function $\kappa(y)$ on $[0, 1]$.

This model can for instance represent the size of the cell or the propagation of a parasite (that we take here as an archetype). The main issue is here to estimate the effect of the amount of parasites in a cell on its reproduction. Note that the individual level is the one of a given cell, and not of a given parasite, as it would be in a completely discretized model. Yet, the number of parasites is assumed to be so large that a continuous description by a random process is a well-justified simplification. The estimation is focused on the effect of these parasites on the birth rates $(B(x) : x \in \mathcal{X})$ of the hosts and on the law of distribution $(\kappa(y) : y \in [0, 1])$ between offspring.

Assumptions are specifically designed in [HM19] for such a model. In a broader perspective, we rather focus on the core principles for which they are introduced and refer to [HM19] for the precise statements. By Assumption 2 on the drift and diffusion coefficient, the authors ensure that the trait $\mathcal{X}$ stays somewhat confined around 0, with a uniformly elliptic and non-singular diffusion. By Assumption 3 on the birth rate, they ensure some regularity in x, a boundedness condition on the potential explosion of births and they prevent the vanishing of these events. Finally, by Assumption 4 on the splitting of x at birth between the two newborns, they ensure a lower-bounded density on the fragmentation parameter, and prevent too asymmetric partitioning.

The authors mention that this assumption could probably be relaxed.

6.1.2 Main results for the estimations of the Generation kernel and the birth rate

The first step of the analysis consists in the estimation of the kernel Q and its stationary distribution ν. This step should be easily generalized in a broader perspective of kernels Q.

For any function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and $y \in \mathcal{X}$, consider $\psi_\star(y) := \sup_{x\in\mathcal{X}} |\psi(x, y)|$, $\psi^\star(x) := \sup_{y\in\mathcal{X}} |\psi(x, y)|$. Denoting by $\wedge$ the minimum, we also define for any positive measure ρ on $\mathcal{X}$:

$$|\psi|_\rho := \int_{\mathcal{X}^2} |\psi(x, y)|\, \rho(dx)\, dy + \left(\int_{\mathcal{X}^2} |\psi(x, y)|\, dx\, dy \wedge \int_{\mathcal{X}} |\psi_\star(y)|\, dy\right)$$
$$\mathcal{M}_{\mathrm{U}_n}(\psi) := \frac{1}{Card(\mathrm{U}_n^\star)} \sum_{u\in\mathrm{U}_n^\star} \psi(X_{u-}, X_u),$$

where $\mathrm{U}_n^\star$ is U_n deprived from the root, and X_{u-}, X_u denotes respectively the trait at birth of the parent of u and the one of individual u itself. Recall that the transition from X_{u-} to X_u is given by $Q(X_{u-}, dx)$.

Proposition 6.1.2.1. *Let Assumptions 2, 3 and 4 be satisfied. 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. If U_n is a regular incomplete tree, the following estimate holds true :*

$$\mathbb{E}_\mu[(\mathcal{M}_{\mathrm{U}_n}(\psi) - \nu(Q\psi))^2] \lesssim Card(\mathrm{U}_n)^{-1}(|\psi^2|_{\mu+\nu} + |\psi^\star\psi|_\mu + (1 + \mu(V^2))|\psi_\star|_1|\psi|_\nu), \tag{6.1.2}$$

where the symbol $\lesssim$ means up to an explicitly computable constant that depends only on supp $(\psi_\star)$ as long as $\mathrm{Q} \in \mathbb{Q}_{1/2}$. More generally, it would also depend on Q.

For the following theorem, we assume that the operator Q has a density $(q(x, y))_{(x,y)\in\mathcal{X}^2}$ w.r.t. Lebesgue measure on $\mathcal{X}$, with some Hölder regularity. I refer to [HM19] for the exact definition of the sets $\mathbb{Q}_{1/2}^{\alpha,\beta}(R)$, where the value $R > 0$ defines a bound on the Hölder regularity of respectively order α in x and β in y. The "1/2" refers to the property (6.1.1). Given such a regularity, the authors propose to adjust the order of the estimation kernel and explain how to choose the associated window sizes given $|\mathrm{U}_n|$ (for some large $n \geq 1$). The estimation of q also depends on a threshold ϖ_n that is to adjust. I refer to [HM19] for the exact definitions of the estimators $\hat{\nu}_n(y)$ and $\hat{q}_n(x, y)$ of respectively $\nu(x)$ and $q(x, y)$ (with the knowledge of U_n).

Theorem 6.1.1. *Let Assumptions 2, 3 and 4 be satisfied. Assume that the initial distribution μ is absolutely continuous w.r.t. the Lebesgue measure with a locally* *bounded density and satisfies* $\mu(V^2) < \infty$. *Let* $\alpha, \beta > 0$. *Then, for any* ϱ*-regular incomplete tree* U_n *and any* $R > 0$,

$$\sup_{Q \in \mathbb{Q}^{\alpha,\beta}_{1/2}(R)} (\mathbb{E}_\mu[(\hat{\nu}_n(y) - \nu(y))^2])^{1/2} \lesssim Card(\mathrm{U}_n)^{-\beta/(2\beta+1)}.$$

and

$$\sup_{Q \in \mathbb{Q}^{\alpha,\beta}_{1/2}(R)} (\mathbb{E}_\mu[(\hat{q}_n(x,y) - q(x,y))^2])^{1/2} \lesssim \varpi_n^{-1} Card(\mathrm{U}_n)^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$$

hold true, where $s(\alpha,\beta)^{-1} := (\alpha \wedge \beta)^{-1} + \beta^{-1}$ *is the effective anisotropic smoothness associated with* (α, β).

Note that the rate of convergence depends on the regularity of q. The higher it is, the more useful information we can gather from observed transitions from vicinities of x to vicinities of y. The accuracy of the estimators is thus better supported.

The estimation of the division rate is much more complicated. Additional assumptions are required. The study is restricted to diffusion of the trait on a compact space, so that (6.1.2) can be strengthened with a right-hand side only depending on the L^∞-norm of ψ. Again, we restrict the study to $\mathrm{Q} \in \mathbb{Q}_{1/2}$. A parametric approach is considered, where B is encoded by parameters $\vartheta \in \Theta$. The associated Fisher matrix $\Psi(\vartheta)$ is assumed to be non-singular to avoid issues of identifiabilty. They also assume some upper-bounds on the derivative of B along ϑ up to the third order. Moreover, the authors prevent any degeneracy and singularities of the birth rate. Finally, this birth rate B is assumed to be a globally monotone function of ϑ (uniformly in $\mathcal{X}$).

Theorem 6.1.2. *Let the above-mentioned assumptions be satisfied. For every* ϑ *in the interior of* Θ, *if* U_n *is a* ϱ*- regular incomplete tree :*

$$\sqrt{Card(\mathrm{U}_n)} \times (\hat{\vartheta}_n - \vartheta) \to \mathrm{N}(0, \Psi(\vartheta)^{-1})$$

in distribution as $n \to \infty$, *where* $\mathrm{N}(0, \Psi(\vartheta)^{-1})$ *denote the* d*- dimensional Gaussian distribution with mean* 0 *and covariance the inverse of* $\Psi(\vartheta)$.

Note that we find in this case the classical rate of convergence with order the square root of the number of observations, as for classical Markov Chains. It hints at the fact that every observed transition from parents to offspring contributes to the estimation of ϑ.

The monotonicity condition of B along ϑ seems very restrictive, especially in the case of multidimensional parameter. Relaxing this condition would be a natural direction to look at for improvement.

Concluding remarks :

As mentioned after the assumption on the genealogical tree, considering this tree as prescribed is much more convenient. The fact that the convergence results of Theorems

6.1.1 and 6.1.2 depend on U_n only through its cardinal indicates that these results shall be quite robust concerning the specific realization of the tree.

However, it does not exclude that the estimators are slightly biased by the fact that certain transmission patterns might lead to a greater probability of survival (of the lineages). For instance, it may be beneficial to divide early, in order not to let the parasites divide for long before the next division. Traits favoring early division are thus more likely to be transmitted. This bias depends on how strongly these effects may have shaped the genealogical tree. Given that the design of the experimental process has a strong effect on the shape of the tree, including this dependency would though be very difficult and likely to introduce even more bias. Besides, the estimation of the biased birth rate may be the main interest. It only means that one infers the dynamics of a "successful" lineage, with luck playing a role in this success, as well as environmental conditions.

This issue of estimating individual parameters by looking at the population as a whole leads to similar issues of independence. Notably, the population around the focal individual can be considered as a component of its environment. One may wish that the parameters of its dynamics depend on the interaction it has with this extended environment. The models become much simpler if one can neglect the detailed interactions between the individuals, and replace them by average effects. In the limit of a large population, for a system at equilibrium, such assumption may be justified by the Law of Large Number. Indeed, the effects of the interactions are averaged, as long as they are not too local, and thus almost globally constant. In results of propagation of chaos, such property is generalized to cases where the population is not at equilibrium. I refer to Sznitman's lecture at Saint-Flour [Sz91] for an overview on this topic. In such a limiting case, known as McKean-Vlasov equations, the law of the process itself acts on the individual dynamics, together with fluctuations specific to this individual. At equilibrium, this law is all the more stable because the population size is large. Thus, we can deal with the effects of these interactions through some hidden parameters (like an effective death or birth rate). And as for the analysis of this section, estimating them by simply looking at death and reproduction times would produce a bias if they depend on heritable factors. But this bias is presumably small if heritability is weak.

To study evolution of population traits, that is for instance ϑ if mutations could alter this value, it is very classical to assume that the population size is large. Since one usually assumes a separation of time-scales between demographic dynamics and evolutionary processes, there is clearly time for individual fluctuations to be averaged. Heritable traits can leave a sufficiently clear mark for natural selection to be effective. When the environment itself is affected by the traits of individuals, as in the next model, it might however not be so clear what an average effect would be.

6.2 Selection with two levels

The foundation of this section is my work on the ability of selective effects acting at a group level to compensate for those acting inside each group at an individual level (cf Chapter 4). For simplicity, we focus here on competition between two types in a population of fixed and large size. In real populations and IBM models, there are fluctuations in the proportion due to the inexact compensation between births and deaths events occurring with the same large rate. Biologists usually refer to it as genetic drift, and often neglect its effect in the case of large populations. One shall see however that these fluctuations may not be neutral at all in a model where two selective effects are considered : the first one favors some individuals inside their groups (individual level selection) while the other favors some groups depending on the individuals they gather (group level selection). Reducing random fluctuations inside each group strongly hinders response to selection at group level. For clarity, we shall assume that these selective effects are conflicting. For instance, one may ask if and how the inefficient or cheater individuals can be regulated through natural selection at this group level.

The individual-based model is taken from [LM15], where a formalism for group selection is introduced. All groups have the same size $n \in \mathbb{N}$. There are two types of individuals : C and D. Type D individuals have a better reproduction at the individual level (D for defectors) while type C individuals are positively selected at the group level (C for cooperators). Replication and selection occur concurrently at individual and group levels according to a "nested Moran process", as introduced in [Du08] and recalled next. Type C individuals replicate at rate w_I and type D individuals at rate $w_I\,(1+s), s \geq 0$. When an individual gives birth, another individual in the same group is selected uniformly at random to die and be replaced, so that the population size remains constant. To reflect antagonism at the higher level of selection, groups replicate at a rate that depends on the number of type C individuals they contain. We take this rate to be $w_G \times [1+r(k/n)]$, where k/n is the fraction of type C individuals in the group and $r(x), x \in [0,1]$ is the selection coefficient at group level. The number of groups is maintained at m by selecting a group uniformly at random to die whenever a group replicates. The two offspring of groups are assumed to be identical to their parent.

Two limits of large population are justified from the individual-based model in [LM15]. At least four different effects may contribute to the limiting dynamics. The selective effect at individual level is due to difference in growth rate between D and C individuals ; the selective effect at group level to difference in the growth rate of groups (depending on the proportion of C individuals). Consider the number of replacement events in a Moran model of a population made up of a constant number of identical individuals. In a limit of large populations as stated in the central limit theorem, this number shall increase with a linear rate with random fluctuations that may be well approximated by a Brownian Motion. Similar fluctuation effects generate variations in the proportion of trait carriers in a population. One can include such effects, that we call "random fluctuations", in the limiting large population process, both at individual and group levels.

Both limits in [LM15] include a term for each selective force, but only the second includes random fluctuations, both within the groups and between groups. In Chapter 4, another limit is considered, where random fluctuations are kept only inside the groups, together with selective forces. Although such limit can be mathematically justified, it is undoubtedly more realistic to keep also random fluctuations at group level. Yet, the analysis is notably much more technical with much less clear interpretation. In this more complete model, one type go almost surely extinct in finite time. Such event of "ultimate fixation" has a positive probability for each type. For large enough group populations, the limit obtained by neglecting these fluctuations provides an interesting view on the main features of the dynamics.

Let X^i_t be the number of type C individuals in group i at time t. Then :

$$\mu^{m;n}_t := \tfrac{1}{m} \textstyle\sum_{i\leq m} \delta_{X^i_t/n}$$

is the empirical measure at time t of the proportion of type C by group, with m the number of groups and n the number of individuals per group. Here, δ_x is the Dirac at x. The X^i_t are divided by n so that $\mu^{m;n}_t$ is a probability measure on $E_n := [0; 1/n; ...; 1]$. For fixed $T > 0$, $(\mu^{m;n}_t)_{t\leq T} \in D([0;T]; \mathcal{M}_1(E_n))$, the set of càdlàg processes on $[0;T]$ taking values in $\mathcal{M}_1(E_n)$ (the set of probability measures on E_n). With the particle process described above, $\mu^{m;n}_t$ has generator

$$(\mathcal{L}^{m;n}\psi)(v) = \textstyle\sum_{i,j}(w_I\, R^{i,j}_I + w_G\, R^{i,j}_G)(v) \times \Big(\psi\big[v + 1/m\,(\delta_{j/n} - \delta_{i/n})\big] - \psi[v]\Big)$$

where $\psi \in C_b(\mathcal{M}_1([0;1]))$ is a bounded continuous function, and $v \in \mathcal{M}_1(E_n) \subset \mathcal{M}_1([0;1])$.

The transition rates $(w_I\, R^{i,j}_I + w_G\, R^{i,j}_G)$ are given by

$$R^{i,j}_I(v) := \begin{cases} m\, v(i/n)\, i\,(1 - i/n)\,(1+s) & \text{if } j = i-1; i < n, \\ m\, v(i/n)\, i\,(1-i/n) & \text{if } j = i+1; i > 0, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{and} \qquad R^{i,j}_G(v) := m\, v(i/n)\, v(j/n)\,(1 + r[j/n]). \tag{6.2.3}$$

$R^{i,j}_I$ and $R^{i,j}_G$ are the rates of respectively individual- and group-level reproductive events.

Theorem 6.2.1. *Suppose that $w_I/n \to \omega_I$, $n\, s \to \sigma$ as $n, m \to \infty$, while w_G and $\{r(x)\}_{x\in[0,1]}$ are kept constant. Suppose the particles in the process $\mu^{m;n}_t$ are initially* *independently and identically distributed according to the measure $\mu_0^{m;n}$, where $\mu_0^{m;n} \to \mu_0$ as $m, n \to \infty$. Then, $\mu_t^{m;n}$ converges weakly to $\mu_t \in D(\mathbb{R}_+; \mathcal{M}_1([0;1]))$, where μ_t is the unique solution with initial condition μ_0 of the differential equation :*

$$\partial_t \langle \mu_t \,\big|\, f\rangle = \langle \mu_t \,\big|\, \mathcal{L}_{WF} f\rangle + \langle \mu_t \,\big|\, r\, f\rangle - \langle \mu_t \,\big|\, f\rangle \times \langle \mu_t \,\big|\, r\rangle, \tag{6.2.4}$$

$$\text{where } \mathcal{L}_{WF} f(x) = -s\, x\,(1-x)\,\partial_x f(x) + (\sigma^2/2)\,.\,x\,(1-x)\,\partial^2_{xx} f(x). \tag{6.2.5}$$

6.2.1 Definition as a conditional law

We describe the solution of such equation through a Feynman-Kac penalization of the stochastic process with generator $\mathcal{L}_{WF}$. Let us define X as the solution of the SDE :

$$dX_t := -s\, X_t\,(1-X_t)\, dt + \sigma \sqrt{X_t\,(1-X_t)}\, dB_t, \qquad X_0 \sim \mu_0 \tag{6.2.6}$$

The existence and uniqueness of such process can be found e.g. in chapter 5.3.1 of [Daw10]. It is linked there with an individual-based model, namely the neutral 2-allele Wright-Fischer Markov Chains. We also consider the following Feynman-Kac penalization :

$$Z_t := \exp \textstyle\int_0^t r(X_s)\, ds,$$

Note that r is bounded so that for any $t > 0$, $\mathbb{E}(Z_t) \in (0, \infty)$.

Proposition 6.2.1.1. *With the above definitions, we characterize μ_t by the fact that for any $f \in \mathcal{C}_b^2$:*

$$\langle \mu_t \,\big|\, f\rangle := \mathbb{E}\left[f(X_t)\, Z_t\right] / \, \mathbb{E}\left[Z_t\right],$$

I refer to Chapter 4 for the proof of the proposition, which is easily derived from the Ito formula.

We next give another view on this law μ_t, that will simplify our notations and clarify our point. Since subtracting a constant to r does not change the value of $\langle \mu_t \,\big|\, r\, f\rangle - \langle \mu_t \,\big|\, f\rangle\ \langle \mu_t \,\big|\, r\rangle$, we assume in the following that $r \le 0$, so that we can consider it as a death rate. Z_t can then be interpreted as the probability that the process has survived until time t, while confronted to a death rate of r, conditionally on $(X_u)_{u \ge 0}$. More formally, with T_∂ an independent exponential r.v. with mean 1, we can define the extinction time as :

$$\begin{aligned} &\tau_\partial := \inf\left\{t \ge 0\,;\, -\ln(Z_t) \ge T_\partial\right\}, \\ &\quad \text{so that } \mathbb{P}(t < \tau_\partial \,\big|\, X_u, u > 0) = \mathbb{P}\left(-\ln(Z_t) < T_\partial \,\big|\, X_u, u \le t\right) = Z_t \\ &\qquad\qquad \langle \mu_t \,\big|\, f\rangle = \mathbb{E}\left[f(X_t)\,;\, t < \tau_\partial\right] / \, \mathbb{P}\left[t < \tau_\partial\right] = \mathbb{E}(f(X_t) \,\big|\, t < \tau_\partial). \end{aligned}$$

Any equilibrium of μ_t is thus a quasi-stationary distribution (QSD) of X under the death rate r. Applying recent results on QSD makes it possible to obtain the limiting behavior and the speed of convergence of μ_t.

Similarly, the hitting times of 0 and 1 are denoted τ_0 and τ_1. Since they are absorbing, it is natural to be interested in the law of the marginal with these extended extinction times :

$$\tau_{0,\partial} := \tau_\partial \wedge \tau_0, \quad \tau_{1,\partial} := \tau_\partial \wedge \tau_1, \quad \tau_{0,1} := \tau_0 \wedge \tau_1, \quad \tau_{0,1,\partial} := \tau_\partial \wedge \tau_0 \wedge \tau_1.$$

Proposition 6.2.1.2. *With the above definitions, we can also characterize μ_t by the fact that for any $f \in \mathcal{C}_b^2$:*

$$\mu_t = x_t^0\, \delta_0 + x_t^1\, \delta_1 + x_t^\xi\, \xi_t$$

$$\textit{where } x_t^\xi := \frac{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\right]}, \qquad \left\langle \xi_t \,\middle|\, f \right\rangle := \frac{\mathbb{E}\left[f(X_t)\, Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]} = \mathbb{E}\left[f(X_t)\,|\, t < \tau_{0,1,\partial}\right],$$
$$x_t^0 := \frac{\mathbb{E}\left[Z_{\tau_0}\, \exp[-r_0(t-\tau_0)]\,;\, \tau_0 < t\right]}{\mathbb{E}\left[Z_t\right]}, \qquad x_t^1 := \frac{\mathbb{E}\left[Z_{\tau_1}\, \exp[-r_1(t-\tau_1)]\,;\, \tau_1 < t\right]}{\mathbb{E}\left[Z_t\right]}.$$

The proof is elementary and left to the reader.

Remarks : • It is not difficult to generalize the model to include frequency dependent effects of selection at individual level. We would then replace $s > 0$ by some smooth function $(s(x))_{x\in[0,1]}$. Generalizations of σ as a function are also not a mathematical issue. Yet a priori, it does not seem biologically justified.

• This Feynman-Kac penalization is analogous to the many-to-one formula described by [BD+11] in the case of Markov processes associated to branching Galton-Watson Trees. Behind this description is the idea that there is a bias towards larger offspring in the genealogy of a typical individual in the population, favoring the individuals with a larger birth rate. Contrary to [BD+11], where no interaction occurs between lineages, one cannot expect to represent the law of μ_t directly through another Markov Process, i.e. without penalization (note in (6.2.4) the quadratic term in μ_t).

6.2.2 QSDs and exponential convergence

Note first that in any case, δ_0 and δ_1 are QSDs for the extinction time τ_∂, i.e. stable distributions for the dynamics given by (6.2.4). If the initial condition μ supported on $\{0, 1\}$, the dynamics is immediately deduced from the death rates in 0 and 1.

For other initial conditions μ (not supported on $\{0, 1\}$), we define the following semi-groups associated to our different extinctions :

$$\mu A_t(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_\partial), \quad \mu A_t^{01}(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_{0,1,\partial}),$$
$$\mu A_t^{1}(dx) := \mathbb{P}_\mu(X_t \in dx \,\big|\, t < \tau_{1,\partial})$$

Proposition 6.2.2.1. *There exists a unique QSD $\alpha \in \mathcal{M}_1[(0,1)]$ and a survival capacity of η associated to the extinction time $\tau_{0,1,\partial}$. With the associated extinction rate ρ_α, it means :*

$$\forall\, t > 0, \quad \mathbb{P}_\alpha(X_t \in dx\,;\, t < \tau_{0,1,\partial}) = \exp[-\rho_\alpha\, t]\, \alpha(dx),$$
$$\forall\, x,\, \eta(x) = \exp[\rho_\alpha\, t]\, \mathbb{E}_x(\eta(X_t)\,;\, t < \tau_\partial)$$

Moreover, we have the following exponential convergences at rate $\zeta > 0$:

$$\exists\, C > 0,\ \forall\, \mu \in \mathcal{M}_1[(0,1)], \quad \left\| \mu A_t^{01} - \alpha \right\|_{TV} \le C\, \exp[-\zeta\, t]. \tag{6.2.7}$$

$$\exists\, C' > 0,\ \forall\, \mu \in \mathcal{M}_1[(0,1)], \quad |\exp[\rho_\alpha\, t]\, \mathbb{P}_\mu(t < \tau_{0,1,\partial}) - \langle \mu \,\big|\, \eta\rangle| \le C'\, \exp[-\zeta\, t] \tag{6.2.8}$$

A fortiori

$$\eta(x) := \lim_{t\to\infty} \exp[\rho_\alpha\, t]\, \mathbb{P}_x(t < \tau_{0,1,\partial})$$
$$\textit{and } \|\eta_\bullet\| := \sup_{\{x \in (0,1),\, t > 0\}} \exp[\rho_\alpha\, t]\, \mathbb{P}_x(t < \tau_{0,1,\partial}) < \infty \tag{6.2.9}$$

Let $\rho_0 = -r_0$ ($\rho_1 = -r_1$) the extinction rate of δ_0 (resp. δ_1). We show in the following that the long-time behavior of the process with only the local extinction rate depends mainly on ρ_α, ρ_0 and ρ_1.

In the following convergences, we will often have uniform bounds for probability measures belonging for some $n \ge 1$ and $\xi > 0$ to :

$$\mathcal{M}^0_{n,\xi} := \left\{ \mu \in \mathcal{M}_1([0,1]) \,\middle|\, \mu[1/n, 1] \ge \xi \right\},\ \textstyle\bigcup_{n,\xi} \mathcal{M}^0_{n,\xi} = \mathcal{M}_1([0,1]) \setminus \{\delta_0\}.$$
$$\text{or in} \quad \mathcal{M}^{0,1}_{n,\xi} := \left\{ \mu \in \mathcal{M}_1([0,1]) \,\middle|\, \mu[1/n,\, 1 - 1/n] \ge \xi \right\}, \quad (n \ge 3, \xi > 0)$$
$$\textstyle\bigcup_{n,\xi} \mathcal{M}^{0,1}_{n,\xi} = \mathcal{M}_1([0,1]) \setminus \{u\, \delta_0 + (1-u)\, \delta_1 \,\big|\, u \in [0,1]\}.$$

6.2.2.1 Despite the fixation events, polymorphic groups maintain themselves in the population

Proposition 6.2.2.2. *Assume that $\rho_\alpha < \rho_0 \wedge \rho_1 := \rho$. Then, there is only one stable QSD $\alpha_{0,1}$, with convergence rate $\rho - \rho_\alpha$, i.e. :*

$$\forall\, n \ge 1,\ \forall\, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}^{0,1}_{n,\xi}, \quad \|\mu A_t - \alpha_{0,1}\|_{TV} \le C_{n,\xi}\, \exp[-(\rho_\alpha - \rho)\, t],$$

where $\alpha_{0,1}$ has extinction rate ρ_α and is given as $\alpha_{0,1} = y_0\,\delta_0 + y_1\,\delta_1 + y_\alpha\,\alpha$ with :

$$\frac{y_0}{y_\alpha} = \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})}{(\rho_0 - \rho_\alpha)}\,, \qquad \frac{y_1}{y_\alpha} = \frac{\rho_\alpha \times \mathbb{P}_\alpha(\tau_1 = \tau_{0,1,\partial})}{(\rho_1 - \rho_\alpha)}\,,$$

and of course $y_0 + y_1 + y_\alpha = 1$.

If $\rho_1 < \rho_0$, for any initial condition $\mu_0 = u\,\delta_0 + (1-u)\,\delta_1$ with $u \in (0,1)$, μ_t converges at rate $\rho_0 - \rho_1$ to δ_1.
If $\rho_1 = \rho_0$, then any such distribution is a QSD with the extinction rate ρ_0.

Pure groups are continuously generated from polymorphic groups without any reversed transition. Yet, in this case, these polymorphic groups are sufficiently selected upon through their better survival to persist in the population. Pure groups are like remnants of these polymorphic groups : their proportion reaches a steady state where their faster decay compensate for the fixation rate from polymorphic groups. The stabilization of polymorphic profile induces the convergence of both this fixation rate and of the maintenance rate of polymorphic groups to ρ_α.

In any case, polymorphism is maintained by any sufficiently large group selection favoring it, since :

Proposition 6.2.2.3. *Given any $\sigma > 0$, $s \geq 0$, and a bounded continuous function r^0 with its maximum only in the interior of $(0,1)$, there exists a critical value $R_\vee > 0$ such that for any $R > R_\vee$ and considering the system with $r = R\,r^0$, we indeed have $\rho_\alpha < \rho_0 \wedge \rho_1$.*

Conversely, when group selection is too small, polymorphic groups cannot be maintained :

Proposition 6.2.2.4. *Conversely, given any $\sigma > 0$, $s \geq 0$, and a bounded measurable function r^0, there exists a critical value $R_\wedge > 0$ such that for any $R < R_\wedge$ and considering the system with $r = R\,r^0$, it holds $\rho_0 \wedge \rho_1 < \rho_\alpha$.*

Too strong neutral fluctuations also make the fixation of the groups hardly avoidable, so that :

Proposition 6.2.2.5. *Given any $s > 0$ and any bounded function r, $\lim_{\sigma\to\infty} \rho_\alpha(\sigma) = +\infty$.*

Conditions for polymorphism to be maintained : For this maintenance rate ρ_α to be higher than the decay of pure groups, it is clearly necessary that r is maximal in $(0,1)$. Of course, it is not sufficient, because first of random fluctuations (σ) and second because of selection effects inside each group. As stated in Proposition 2.6, too large σ would induce a too large rate of fixation (to 0 or 1). Even strong effects of selection through r would then be unable to make their reproduction large enough to compensate for this loss.

But on the other hand, when there are internal selective effects pushing towards type D individuals, having only small random fluctuations limits the effectiveness of selection at group level. Indeed, all groups with similar initial condition evolve too closely for such effects of selection to really distinguish between them : they are essentially driven by the flow of the ODE : $\partial_t x_t := -s\, x_t\,(1 - x_t)$. Even if perturbations are amplified by this selection towards the opposite direction, strong deviations would be much too costly.

6.2.2.2 Fixation on either side is the most stable case and the type C is favored by group selection

Proposition 6.2.2.6. *Assume that $\rho_1 < \rho_0 < \rho_\alpha$. Then, δ_1 is the only stable QSD, with convergence rate $\rho_0 - \rho_1$:*

$$\forall\, n \geq 1,\ \forall\, \xi > 0,\ \exists\, C_{n,\xi} > 0, \quad \forall\, \mu \in \mathcal{M}^0_{n,\xi}, \quad \|\mu A_t - \delta_1\|_{TV} \leq C_{n,\xi}\, \exp[-(\rho_0 - \rho_1)\, t].$$

We also have an additional level of convergence :

Proposition 6.2.2.7. *Assume that $\rho_0 < \rho_\alpha$. Then, there exists $C > 0$ s.t. :*

$$\forall\, \mu \in \mathcal{M}_1([0,1]) \setminus \{\delta_1\}, \quad \left\|\mu A^1_t - \delta_0\right\|_{TV} \leq C\, \exp[-(\rho_\alpha - \rho_0)\, t].$$

Both results are asymptotic and might not reflect exactly the dynamics on a short time-scale. Yet, their justification gives us some insight into what can happen.

When polymorphism gets quickly negligible : If σ is large, except for initial conditions very close to 0, a non-negligible proportion of pure type C groups quickly emerges and dominates the distribution. For initial conditions very close to 0, the emergence time of these pure C groups mainly depends on the proportion of groups able to quickly escape such vicinity of 0. Soon, the growth in the proportion of pure C groups is essentially due to the difference in growth rate between these groups and the rest of the population, meaning that, at this time, the fixation of polymorphic groups plays a negligible role. Finally, one observes the competition between the two types of pure groups, with the initially rare C groups outnumbering the first dominant D.

When trajectories are drifted with little fluctuations : As explained in the previous subsection, the flow of the equation $\partial_t x_t := -s\, x_t\,(1 - x_t)$ dominates the dynamics as long as μ_t stays localized. If σ is rather small, the profile μ_t is essentially given by the integration of the growth rate along the trajectories of the flow. Notably, consider for simplicity the case where the initial condition is supported on $[0, 1-2\epsilon]$. The pure flow brings the group from a proportion $1-\epsilon$ to ϵ in a deterministic time t_ϵ. If σ is small enough, then with a probability close to 1 the process is upper-bounded by this deterministic flow starting from $1-2\epsilon$. It implies that μ_{t_ϵ} is mainly concentrated in $[0, 2\epsilon]$. If the proportion of pure C groups is still negligible at this time, we shall observe a decay in the proportion of polymorphic groups larger than an exponential rate of $\rho_\alpha - \rho_0$. It is a priori unclear that the rate $\rho_\alpha - \rho_0$ is actually observed, since the QSD α might be localized around 1 and very difficult to reach for initial conditions with many less cooperative groups. One may expect some stabilization to occur, where μ_t restricted to $(0,1)$ gets close to some $\tilde{\alpha}$. This distribution $\tilde{\alpha}$ is presumably supported mainly close to 0, with a much larger extinction rate $\rho_{\tilde{\alpha}}$ than ρ_α (and ρ_0). One shall have the right intuition by considering $\tilde{\alpha}$ (resp. $\rho_{\tilde{\alpha}}$) instead of α (resp. ρ_α) in the reasoning of Propositions 6.2.2.6 and 6.2.2.7. This view is supported by first hints of simulations (not detailed in the article).

As stated in Proposition 6.2.2.6, pure C groups shall prevail even in that case. A rate $\rho_1 - \rho_0$ of decay of remaining groups is very likely to be seen, but possibly after a domination by type I groups for a large period of time. The duration of this domination depends actually much on the initial distribution, and particularly on its tail near 1. Indeed, transitions to 1 is especially costly in term of its probability of occurrence.

Compensate the flow of invasion by D individuals : Section 3 provides an evaluation of the strength of this group selection needed to compensate the flow in the limit of vanishing σ. In this large deviation regime, the process seems to evolve for most of the time according to a modification of the initial flow. Yet, it is not as simple : for instance, one might observe the abrupt emergence of a type which was so far negligible, whose growth rate is much higher than the previously dominant type.

This is exactly what shall presumably happen in this model in the case $\rho_1 < \rho_0$. Undoubtedly, the initial proportion of C groups or the neutral fixation for initially almost pure C groups concerns a very tiny proportion of ancestors. Yet, the ancestors initially drifted towards larger proportion of D individuals soon loose any C individual (with very few exceptions). So there is a point in time at which we can no longer neglect the more prolific descendants the former exceptional ancestor groups will have. This shall certainly correspond to the time at which C individuals eventually dominate. Are the ones that finally dominate necessarily pure groups ? Due to the potential long-term persistence of the process in the vicinity of 1 when the random fluctuations are very small, the behavior of ρ_α as σ tends to 0 is quite unclear. So it might happen that some almost pure C groups actually dominate.

Selection upon the initial condition : Still, the selection between different initial conditions may be effective if those are sufficiently apart. It may postpone for some significant time the trend towards 0. Yet, for the polymorphism to persist for long, a very specific form of the law of the initial condition is required. This has been specified in [LM15] in the limit where the random fluctuations are neglected. Such long-term persistence is effectively possible because the flow is vanishing in the vicinity of 1. The authors consider only functions r that are linear in the proportion of type C individuals. But it should generalize to a much more general cases, provided r is larger near 1 than near 0.

Note however that in such regime, the groups dominating at a given time are essentially not the ancestors of those dominating at a much larger time. The descendants of the former have been transported by the flow towards 0, where this sub-population has decayed faster. On the other hand, the ancestors of the latter must have stayed for long in a very tiny and specific region very close to 1, and are thus very few (while the former dominate). Indeed, going backward through the ancestry lines means essentially following the flow backwards. This backward flow goes quickly towards 1, which it approaches at exponential speed. With initial conditions irregular in the vicinity of 1, we thus might observe a surprising sequence of vanishing and reemerging polymorphism between periods of domination by D individuals.

Is this observable in the individual-based model ? Any transition involving a reasonable amount of groups is expected to be indeed observed. This can be estimated through the number of ancestors from the initial population upon which such transitions shall rely. The fluctuations shall vanish with the number of them. Provided the neutral fluctuations in the births and deaths of groups are small, the approximation should be qualitatively valid with approximately 20 ancestries.

It means also that too exceptional transitions are very unlikely to be observed. For instance, the escape from a too close vicinity of 0 happens with a too small probability. The most likely is to observe the complete fixation. For the fixation of pure C groups to be observed, the most probable is then to have one group escaping the vicinity of 0, reaching the other boundary and generating a sufficiently large family there for the extinction to become negligible. Only after such exceptional realization becomes the fixation of pure C groups likely to occur. Given Theorem 6.2.1, larger population sizes makes the event more likely to occur. Yet, in order for one group to behave in a way so different from the typical one, it might be required that the population size is at a largely unrealistic level. Similarly, when σ is so small that transitions towards 1 become negligible, and for an initial condition with a light tail (towards 1), the fixation of pure C groups happens after a very exceptional behavior.

This mathematical complexity is presumably not so relevant in terms of the biology. As soon as the initial condition has a sufficiently light tail in this vicinity of 1, one mainly observe a massive proportion of the groups fixing as pure D types and becoming dominant for a very long time.

6.2.2.3 Polymorphic groups are more stable than type I groups but less than pure type C groups

This case is also treated in Chapter 4. The result is a combination of the ones in the two previous subsections. The dynamics for the domination by pure C groups relies on similar principles as in Proposition 6.2.2.6. The main difference is that the intermediate convergence is stated towards a polymorphic QSD rather than pure D groups. This polymorphic QSD α_1 is described as in Proposition 6.2.2.2 when one subtracts pure C groups before the renormalization. The asymptotic rate of convergence towards the Dirac at pure C groups is deduced from this intermediate convergence result : $\rho_1 - \rho_\alpha$ (smaller than $\rho_1 - \rho_0$).

Again, for the approximated IBM, one may be faced to the same limitations regarding the origin of the first pure C groups as in Proposition 6.2.2.6. In practice, the convergence towards the polymorphic QSD might also not actually reflect the main dynamics of convergence. It might happen that the "emergence" of pure type 1 groups can actually be almost concomitant to the emergence of α^1. Looking at simulations for small values of σ, an alternative metastable regime around 0 may dominate for a significant time the marginal law restricted to $(0, 1)$. Comparing with the case of initial condition close to 1, this distribution $\tilde{\alpha}$ is very different from the actual QSD α, with very separate supports. Even if $\rho_\alpha < \rho_0$, it would not be surprising that for the alternative distribution $\rho_{\tilde{\alpha}} > \rho_0$. For a large range of initial conditions, even if $\rho_\alpha < \rho_0$, the initially observed dynamics is rather the one described in Propositions 6.2.2.6 and 6.2.2.7 by the interplay between 1, 0 and $\tilde{\alpha}$ where $\rho_1 < \rho_0 < \rho_{\tilde{\alpha}}$. The first step of the dynamics is thus obviously a convergence to 0 with a rate expected to stabilize towards $\rho_{\tilde{\alpha}} - \rho_0$.

6.2.2.4 Validation by some simulations ?

In order to evaluate the exceptionality of transitions towards 1, one can propose the following simulation experiment, which is a work in progress. Let us consider some parameters for which most of the marginals tend towards 0, with σ sufficiently small. To observe the upheaval of the group population by type C groups, the marginals should be encoded rather not by the masses at the different grid points, but by the logarithms of these quantities. The case of the absorbing states is treated separately from the marginal restricted to $(0, 1)$. For a better accuracy , it may be useful to refine the grid in the vicinities of 0 and 1.

We do observe the mass towards 1 increasing up to the point of exceeding the mass towards 0. Yet, it is then unclear at which concentration the regeneration of these cooperative groups exceeds the effect of having more and more exceptional transitions leading there. Since the quickest transitions are expected to bring less mass towards 0, our idea is to truncate densities to prevent the most exceptional transitions. So at each simulation step, we suppress from the marginal the mass on states than contain less than the threshold. By varying the threshold, we should have a better view on the number of groups required to observe the transition from 0 to 1 in individual-based models.

If the dynamics is almost unchanged after the truncation, we could conclude that the cost of the transitions that mainly contribute to this upheaval is smaller than the threshold. If the upheaval arises later on, it would mean that less costly transitions could have been sufficient to make type C emerge. Yet, their contribution becomes negligible when compared to quicker transitions. Finally, if the marginal becomes supported on some interval that does not approach 1, it means that any transition towards 1 would be at least as costly as the threshold.

6.2.3 Conclusion of the section

Generally, the interplay between different traits happens in the time-scale at which their carriers can be differentiated. Yet, we have seen in this example that the trade-off between different kinds of advantages can be particularly tricky. The a priori neutral genetic drift might happen to be strongly coupled to the efficiency of some components of selection. In a broader view, we can see this model as an illustration that selective effects might be strongly dependent upon details of the local ecological dynamics, and not only upon the average behavior. If the local subdivision constitutes a sufficiently stable entity with the ability to reproduce itself, natural selection may act. Its strength depends on the level of variability between those entities, as if they were individuals.

Crucial requirements for the presented confrontation are the independence of the groups bewteen successive splittings and a strong heritability for the groups when they split (the two descendants are assumed to be very similar to the "parent"). Even small interactions between these communities (notably migration between groups) is known to greatly disrupt the stability of cooperative strategies (cf. for instance [WG+06]). So we clearly do not claim that such simplified effect is prevalent, because such lack of interaction between the local dynamics is not so common.

Also, the selection at the group level might rely on exceptional transitions of the process X. Although this exceptionality can be compensated in the long run by a larger asymptotic growth rate, two aspects should be remembered : first, it might be much too unlikely for actual populations that some groups experience those rare transitions, so that the mathematical model could be misleading ; second, it takes a very long time for their descendants to invade. It thus raises the question concerning real life whether no other event happens to disturb any of the sub-populations before the emergence of cooperative groups.

Finally, this competition model provides a fruitful insight in an evolutionary perspective (that is the subject of the following Section 6.4). A main quantity of interest is notably the probability that the descendants of a single mutant individual with trait y invade and replace the whole population of "residents" with trait x. This probability is usually compared to the one for the invasion by mutants identical to the residents (the neutral case). In this model, for small values of σ, as long as the mutant trait has a strong deleterious effect, either at the individual- or at the group-level, the invasion is much more difficult than neutral : for the invasion of cooperatives, the random fluctuations inside the groups have to lead the process away from the very stable cheater quasi-equilibrium ; while, for the invasion of cheaters, the genetic drift between groups has to disrupt the also very stable cooperative quasi-equilibrium. Such large deviations are known to be generally especially costly. It means that natural selection of such traits should be much more constraint.

6.3 Large deviation estimate for the adaptation by mutations of weak effect

In this section, I present the work of Champagnat and Henry on some effects of Dirac concentration in non-local models of adaptation with several resources [CH19]. The focus is on the trajectories of evolution, with the selection of favorable mutations.

Interestingly, the authors of [CH19] show a mathematical similarity between two limiting behaviors : a first one where the trait variations are very small ; and the second where the mutation rate is very small (as compared to the selective effects). Notably, the first case can be associated to a limit where mutations of very small effects accumulate. The main requirement for the proofs is that natural selection is very strong in the timescale where one observes the trait dispersion under neutrality.

Again, the population size is assumed to be sufficiently large to include the whole range of the stochastic variations of the trait. The resulting purely deterministic model shall provide a valid approximation to the dynamics of trait proportions in this population. This relation to the trait proportions can be retrieved from the convergence result deduced from the time-scale separation.

6.3.1 The continuous-space limiting behavior

One considers the family $(u^\epsilon)_\epsilon$ of deterministic solutions to the parabolic SDE :

$$\begin{aligned}&\partial_t u^\epsilon(t,x) = \frac{\epsilon}{2}\,\Delta u^\epsilon(t,x) + \frac{R(x,\psi_t^\epsilon)}{\epsilon}\,u^\epsilon(t,x), \quad x\in\mathbb{R}^d\,,\, t\geq 0, \qquad (6.3.10)\\ &-\epsilon\log(u^\epsilon(0,x)) = h_\epsilon(x)\end{aligned}$$

where the competition effect $\psi^\epsilon_t = (\psi^{i,\epsilon}_t)_{i\leq r}$ is defined, for r resources, with a competition kernel $\Psi_i : \mathbb{R}^d \to \mathbb{R}_+$ for resource i by :

$$\psi^{i,\epsilon}_t := \textstyle\int_{\mathbb{R}^d} \Psi^i(y)\, u^\epsilon(t,y)\, dy. \tag{6.3.11}$$

In this model, the traits $x \in \mathbb{R}^d$ characterize the ability to exploit the different resources. Considering mutation effects through some heat kernel corresponds to the case where lots of mutations with very small effects occur and are dispersed throughout the population. In asexual populations, we rather expect mutations with a strong selective effect to invade and fix one after the other. Such a model is rather well-suited for sexual populations, where many recombination events occur and many different alleles with small selective effects may coexist. This is the principles of the so-called infinitesimal model for which we refer notably to [BEV17]. For instance, it is known that many Human traits are affected by many different loci in the genome. Two individuals with almost the same phenotype have possibly very different alleles in these loci and the recombination of alleles along the lineages creates variability. The heat kernel seems then quite relevant to characterize the variability of response for such a trait.

Relying on the arguments presented in [FM04], [CFM06] and [CFM08], the dynamics of u^ϵ as a solution to (6.3.10), can be justified as a limiting description of individual-based models. Theorem 6.2.1 has also been inspired by these results.

The following assumptions are required in [CH19] in order to justify a limit to $\varphi_\epsilon := -\epsilon \log u^\epsilon$ through a variational representation.

1. *Assumptions on* Ψ_i *:* For any $1 \leq i \leq r$, $\Psi_i \in W^{2,\infty}(\mathbb{R}^d)$.
Moreover, there exist $0 < \Psi_{\min} < \Psi_{\max}$ such that

$$\forall\, 1 \leq i \leq r,\ ,\forall\, x \in \mathbb{R}^d, \quad \Psi_{\min} \leq \Psi_i(x) \leq \Psi_{\max}.$$

2. *Assumptions on* R *:* (a) R is continuous on $\mathbb{R}^d \times \mathbb{R}^r$.
(b) There exists $A > 0$ such that

$$\forall\, 1 \leq i \leq r,\ \forall\, x \in \mathbb{R}^d,\ \forall\, (v_\ell) \in \mathbb{R}^r, \quad -A \leq \partial_{v_i} R(x, (v_\ell)_{\ell\leq r}) \leq -A^{-1}$$

(c) There exist two positive constants $0 < v_{\min} < v_{\max}$ such that $\min_{x\in\mathbb{R}^d} R(x,v) > 0$ as soon as $\|v\|_1 < v_{\min}$,
and $\max_{x\in\mathbb{R}^d} R(x,v) < 0$ as soon as $\|v\|_1 > v_{\max}$, where $\|v\|_1 = \sum_{\ell\leq r} |v_\ell|$.
(d) Let H denotes the annulus $B(x, 2v_{\max}) \setminus B(x, v_{\min}/2)$ (for the $\|.\|_1$ norm). Then $\sup_{v\in H} \|R(\mathring{u}, v)\|_{W^{2,\infty}} < \infty$.

3. *Assumptions on* h_ϵ *:* (a) h_ϵ is Lipschitz-continuous on $\mathbb{R}^d$, uniformly with respect to $\epsilon > 0$.

(b) h_ϵ converges uniformly as ϵ tends to 0 to a function h.
(c) For all $\epsilon > 0$ and all $1 \leq i \leq d$, $v_{\min} \leq \int_{\mathbb{R}^d} \Psi_i(x) \exp(-h_\epsilon(x)/\epsilon)dx \leq v_{\max}$.
In particular, $u_\epsilon(0, x)$ is bounded in $L^1(\mathbb{R}^d)$.

Remark : Such competition may at first sight appear as a global competition where every trait is competing against each other for the same resources. The first assumption done in [CH19] requires indeed for the Ψ_i to be lower-bounded by a strictly positive constant. However, in a limit where r is large and Ψ_i is very concentrated, the model may become a valid approximation of a local competition (in the trait space). Thus, one may hope to extend the model naturally to such a framework.

6.3.2 Results for the continuous case

I gather here the main results presented in [CH19], where I have changed some notations for clarity, notably ψ_t and φ :

Theorem 6.3.1. ***Feynman-Kac representation of the solution of** (6.3.10)*
Let u^ε be the unique weak solution of (6.3.10), then, under the assumptions given in [CH19] :

$$\forall (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad u^\varepsilon(t, x) = \mathbb{E}_x \left[\exp\left(\frac{-h_\varepsilon(X_t^\varepsilon)}{\varepsilon} + \frac{1}{\varepsilon} \int_0^t R(X_t^\varepsilon, \psi_{t-s}^\varepsilon) ds \right) \right], \tag{6.3.12}$$

where for all $x \in \mathbb{R}^d, \mathbb{E}_x$ is the expectation associated to the probability measure $\mathbb{P}_x$, under which $X_0^\varepsilon = x$ almost surely and the process $B_t = (X_t^\varepsilon - x)/\sqrt{\varepsilon}$ is a standard Brownian motion in $\mathbb{R}^d$.

Remark : This representation of the solution through a penalization of a stochastic process is in practice of the same kind as the one presented for the empirical law μ_t in Section 6.2. There is only a little difference in the fact that X_t describes now the process when we look at the lineage backward in time. A similar convergence result should thus justify the interpretation of $u^\epsilon(t, x)\, dx$ as the limit of some individual-based measure-valued processes $(\nu_t^{\epsilon,K})$. In particular, $(1/\epsilon) \,.\, R(x, \psi_t^{\epsilon,K})$ should be the additional growth rate of individuals with trait x, where for $i \leq r$, $\quad \psi_t^{i,\epsilon,K} := \langle \nu_t^{\epsilon,K} \,|\Psi^i\rangle$. With fixed ϵ, this is a priori a specific case of the McKean-Vlasov equations mentioned in the conclusion of Section 6.1, where the law of the process itself acts on the individual dynamics (again, I refer to [Sz91]).

Lemma 6.3.2.1. *The function $I_t^{R,\varepsilon} : C([0, t]) \to \mathbb{R}$ defined by $\quad I_t^{R,\varepsilon}(y) = \int_0^t R(y_s, \psi_s^\varepsilon)ds$ is Lipschitz continuous on $C([0, t])$ endowed with the L^∞-norm. The Lipschitz constant is uniform with respect to ε for ε small enough. Moreover, there exists a kernel $\mathcal{M}$ on $\mathbb{R}_+ \times \mathcal{B}(\mathbb{R}^k)$ such that, along a subsequence $(\varepsilon_k)_{k\geq 1}$ converging to 0 :*

$$\forall y \in C([0,t]) \quad I_t^R(y) := \lim_{k\to\infty} I_t^{R,\varepsilon_k}(y) = \int_0^t \int_{\mathbb{R}^k} R(y_s,\psi)\mathcal{M}_s(d\psi)ds.$$

By a kernel $\mathcal{M}$, we specify here a function from $\mathbb{R}_+ \times \mathcal{B}(\mathbb{R}^k)$ into $\mathbb{R}_+$ such that, for all $s \in \mathbb{R}_+, \mathcal{M}_s$ is a measure on $\mathcal{B}(\mathbb{R}^k)$ and, for all $A \in \mathcal{B}(\mathbb{R}^k)$, the function $s \to \mathcal{M}_s(A)$ is measurable.

Theorem 6.3.2. *For all (t,x) in $\mathbb{R}_+ \times \mathbb{R}^d$,*

$$\varphi(t,x) \ := \lim_{k\to\infty} \ \varepsilon_k \log u^{\varepsilon_k}(t,x) = \sup_{y\in\mathcal{G}_{t,x}} \{-h(y_0) + I_t^R(y) - I_t^L(y)\}$$

where the convergence holds uniformly on compact sets and the limit $\varphi(t,x)$ is Lipschitz w.r.t. $(t,x) \in \mathbb{R}_+\times\mathbb{R}^d$ while $I_t^L(y) = \begin{cases} \frac{1}{2}\int_0^t \|y'(s)\|^2 ds & \text{if } y \text{ is absolutely continuous,} \\ +\infty & \text{otherwise.}\end{cases}$

$\mathcal{G}_{t,x}$ denotes the set of continuous functions from $[0,t]$ to $\mathbb{R}^d$ such that $y_t = x$, and I_t^R and $(\varepsilon_k)_{k\geq 1}$ are associated as in Lemma 6.3.2.1.

Remarks : • In this optimization, y defines the main ancestral line of the individuals dominant with trait x at time t. Namely, the Large Deviation theory ensures that as $\epsilon \to 0$, their ancestral lines are very concentrated on such specific and deterministic histories.

• Given a lineage described by y, $I^R(y)$ encodes an average effect of selection due to the perceived growth rate. As one can infer from (6.3.12) when ϵ is very small, the density around such a path is approximately amplified by the exponential of this quantity divided by ϵ. Such large amplification is the core of the Large Deviation theory, where the stochastic behavior is concentrated close to specific paths.

• The dependency of the competition kernel $\mathcal{M}_s$ on φ is implicit and unclear in the limiting model. The level of competition depends on the global composition in traits of the population, while φ only indicates which traits are non-negligible (the x for which $\varphi(x) = 0$). Detailed information on the density is lost in the limit $\epsilon \to 0$, and that is why subsequences ϵ_k are considered. For the same reason, the uniqueness of solutions to such equations is not easy to establish (when it holds).

Again in [CH19], the authors finally relate this variational limit to the solution of the Hamilton-Jacobi equation :

Theorem 6.3.3. *Under some assumptions on the behavior of R and h_ϵ, ψ^{ε_k} converges in $L^1_{loc}(\mathbb{R}_+)$ along the subsequence ε_k of Lemma 6.3.2.1 to a nondecreasing limit $\overline{\psi}$, the kernel $\mathcal{M}$ satisfies*

$$\forall s \geq 0, \qquad \mathcal{M}_s(d\psi) = \delta_{\overline{\psi}_s}(d\psi),$$

and the limit φ of Theorem 6.3.2 solves in the viscosity sense :

$$\begin{cases} \partial_t\varphi(t,x) = R(x,\overline{\psi}_t) + \frac{1}{2}|\nabla\varphi(t,x)| \ , & \forall t \geq 0, x \in \mathbb{R}^d, \\ \max_{x\in\mathbb{R}^d} \varphi(t,x) = 0, & \forall t \geq 0. \end{cases}$$

Behavior of the solution to Hamilton-Jacobi equation

From Theorem 6.3.2, one deduces the following approximation for small $\epsilon : u^\epsilon(t,x) \approx C(t,x)\ \exp[\varphi(t,x)/\epsilon]$. We thus expect to observe a concentration of the individual traits on typically a unique value or a few well-separated values. In the limiting model, these values may be driven by a continuous displacement in the direction of natural selection. They may also split, which would in a sense describe a model of speciation. It is in fact the original purpose of such Hamilton-Jacobi analysis in [DJ+05] to extend the model of Adaptive Dynamics (cf. Subsection 6.4.5). The reverse can happen too, with two subpopulations merging into a single one : the traits are concentrated around a moving value specific to each subpopulation until these two values coalesce. Finally, there could also be some jumps, that is a brutal change of the trait composition. A wide variety of events are thus observable with this simplified model.

Discussion on the involved time-scales

At a given $\epsilon > 0$, looking at equation (6.3.10), we can infer from the growth rate that the stabilization of the densities occurs in a time-scale of order ϵ, as compared to the time-scale at which the traits evolve. The variance of the trait evolves on the contrary at a much larger time scale of order ϵ^{-1}, as in the case of neutral traits. Of course, the detailed description of the demography as presented in Section 1 above and in [HM19] should occur in the short time-scale of order ϵ. In such a large population, the life-expectancy of any particular individual is on an even shorter time-scale, depending on the ratio between the birth and the death rates. The closer it is to one, the shorter this time-scale. A separation of time-scale with the evolutionary trajectory can still be justified for the competition between different effects of selection. But as we have seen in Section 6.2, this competition might last for very long times.

In practice, the description given by equation (6.3.10) is not so easily related to the macroscopic observation of the response to selection, because the neutral variability is difficult to scale (the factor ϵ before the Laplacian). We can reasonably assume that we know the density $u^e(t,x) = u^\epsilon(t,x)$ (as long as it is not too negligible) and the fitness effect $R^e_t(x) = (1/\epsilon)\,.\,R(x,\psi_t)$. For at least positive values, it can be inferred in some laboratory experiment by artificially introducing individuals with trait x in a much larger sample of the population at time t and see how they grow. One could then estimate the value ϵ for which $\varphi^\epsilon(t,x) = \epsilon \log u^e(t,x)$ seems reasonably close to the solution φ of the Hamilton-Jacobi equation :

$$\partial_t\varphi(t,x) = \epsilon\, R^e_t(x) + \tfrac{1}{2}|\nabla\varphi(t,x)|, \tag{6.3.13}$$

or the associated control problem.

Note that a Gaussian distribution for u^e corresponds to $\varphi(t,x) \sim -(x-x_t)^2/\sigma_t^2$, that is the Second order Taylor expansion around x_t, with x_t the optimal trait and $\epsilon\,\sigma_t^2$

the variance of the distribution. Moreover, if φ is a solution to equation (6.3.13), then for $\lambda > 0$, $\hat{\varphi}(x,t) := \varphi(\lambda\, t, \lambda\, x)$ is a solution to (6.3.13) with ϵ replaced by $\lambda\, \epsilon$. Since we expect the dependency in σ_0 on σ_t to vanish very quickly, the identification of ϵ from σ_t should not be a too difficult option. Yet, as mentioned in the next paragraph, the speed of response to selection might be driven by an exceptional proportion of the population at time t. An estimate such as σ_t that summarizes rather the core of the distribution $u^e(t,x)\, dx$ might thus not be so relevant.

Connection with individual-based models

To justify the connection with individual-based models, concentration effects may appear favorable. Yet, as we can see from Theorem 6.3.2, the adaptation of the population is driven by exceptional profiles of stochastic variations. The stochastic techniques of Large Deviations is then crucial to obtain the limiting behavior. Notably, it means that selection is mainly driven by individuals at the front. This optimization provides a very interesting insight on the history of the genealogies that led to the traits observed generally in the population at time t. Notably, we can observe cases where the traits at time $s \leq t$ for the ancestors of the dominants at time t may not be typical at all as compared to the dominant traits at time s. This of course raises the issue of the biological relevance of such transitions. We face similar limitations as in previous Section 6.2 for the relevance of events involving very small population sizes. Still, it is much easier to relate the above description to some individual-based model than from the more classical description only relying on Hamilton-Jacobi equations (cf. the conclusion of the section).

6.3.3 Results for the discrete case

This analysis of Feynman-Kac operators and Large Deviation principles can be adapted to a discrete space. Simplifying assumptions are then more easily obtained and makes it possible to have a clearer view on the implications of this model.

In [CH19], the authors consider the following system of ordinary differential equations, for $k \in E$ and $t \in [0,T]$:

$$\begin{cases} \partial_t u^\varepsilon(t,k) = \sum_{j \in E\setminus\{k\}} \exp\Big(\dfrac{-\mathfrak{T}(k,j)}{\epsilon}\Big)(u^\varepsilon(t,j) - u^\varepsilon(t,k)) + \dfrac{1}{\varepsilon} u^\varepsilon(t,k) R(k, \psi^\varepsilon_t) \,, \\ u^\varepsilon(0,k) = \exp(-\frac{h(k)}{\varepsilon}) \,, \qquad \psi^{i,\varepsilon}_t = \sum_{k\in E} \Psi^i(k)\, u^\varepsilon(t,k), \quad \forall i \leq r \end{cases}$$

In such a model, mutations from state j to state i happen at rate $\exp(-\epsilon^{-1}\mathfrak{T}(k,j))$. As mentioned in the introduction of [CH19], this is not the classical form of the solution for the system of ODE describing population densities. The former can however be obtained from the latter by a slight adjustment in the definition of R (vanishing as $\epsilon \to 0$) and is more practical for the following analysis. The growth rate of the individuals in state k is (nearly) $\epsilon^{-1}R(k,\psi)$, where $\psi = (\psi^i)$ specifies the amount of available resources.

First, the solution u^ε of the system is described in [CH19] by using an integral representation similar to (6.3.12). Let $(X^\varepsilon_s,\ s \in [0,T])$ be the Markov process in E with infinitesimal generator :

$$L^\varepsilon f(k) = \textstyle\sum_{j\in E} \exp[\frac{-\mathfrak{T}(k,j)}{\varepsilon}](f(j) - f(k))$$

i.e. the continuous-time Markov process whose jump rate from state $i \in E$ to $j \neq i$ is $\exp(-\mathfrak{T}(i,j)/\varepsilon)$.

Proposition 6.3.3.1. ***Integral representation***
For any positive real number t and any element i of E, we have

$$u^\varepsilon(t,i) = \mathbb{E}_i\left[\exp\left(\frac{-h(X^\varepsilon_t)}{\varepsilon} + \frac{1}{\varepsilon}\int_0^t R\ (X^\varepsilon_s, \psi^\varepsilon_{t-s})ds\right)\right]$$

The interpretation is very similar to the one of Theorem 6.3.1.

Proposition 6.3.3.2. ***Weak LDP***
$(X^\varepsilon)_{\varepsilon\geq 0}$ satisfies a weak LDP with rate function

$$I_T : \left\{ \begin{array}{lll} \mathrm{D}([0,T],E) & \to & \mathbb{R} \\ y & \mapsto & \sum_{\ell=1}^{N_y} \mathfrak{T}(y_{t^y_\ell -}, y_{t^y_\ell}) \end{array} \right. ,$$

where $\mathrm{D}([0,T],E)$ is the space of càdlàg functions from $[0,T]$ to E, N_y is the number of jumps of y and $(t^y_\ell)_{1\leq\ell\leq N_y}$ the increasing sequence of jump times of y.

We shall use the notation $\quad I_T(y) := \sum_{0<s\leq T} \mathfrak{T}(y_{s-}, y_s) \quad$ with the implicit convention that $\mathfrak{T}(i,i) = 0$ for all $i \in E$.

Theorem 6.3.4. *Lemma 6.3.2.1 is also satisfied in the discrete case and for all (t,i) in $(0,+\infty)\times E$, with the associated subsequence (ε_k) :*

$$\begin{aligned} \varphi(t,i) \ &:= \lim_{k\to\infty} \varepsilon_k \log u^{\varepsilon_k}(t,i) \\ &= \sup_{y\in\mathrm{D}([0,t],E) s.t.\ y_0=i} \{-h(y_t) + \int_0^t \int_{\mathbb{R}^r} R(y_s,\psi)\mathcal{M}_{t-s}(d\psi)ds - \sum_{0<s\leq t} \mathfrak{T}(y_{s-}, y_s)\} \ . \end{aligned}$$

As in the expression in Theorem 6.3.2, one shall remark that the optimal y in this expression describes the transition of states for the lineage of a typical individual at time t. This evolution is backward in time, as can be also seen in the term $R(y_s,\psi)\mathcal{M}_{t-s}(d\psi)$, where $\mathcal{M}_{t-s}(d\psi)$ provides the law of ψ_{t-s}.

Theorem 6.3.5. *For all subsequence $(\varepsilon_k)_{k\geq 1}$ as in Theorem 6.3.4, the limit $\varphi(t,i)$ of $\varepsilon \log u^\varepsilon(t,i)$ is Lipschitz with respect to the time variable t on $(0,+\infty)$. In addition, if $h(i) \leq h(j) + \mathfrak{T}(i,j)$ for all $i \neq j$, the function φ is Lipschitz on $\mathbb{R}_+$.*

We see that all the previous results in continuous space have an equivalent for discrete space.

In the following, we focus on a specific case where the uniqueness can be obtained and the whole dynamics described with much more accuracy. The result presented in [CH19] relies on some assumption on the stability of the dynamics restricted to any subset of E, which is called Assumption (H). I present next the main intuition behind this assumption and refer to [CH19] for the exact formulation.

By this Assumption (H), the authors notably ensure that, for any subset A containing k different types, there exists a unique strongly attractive equilibrium. Since the focus is on the dynamical system, the estimation of the dynamics is valid as soon as the k types are initially in non-negligible proportion. Yet, some of the components of the steady-state may be 0, meaning an exponential decay of the mass of those. Assumption (H) also ensures that any other (unstable) equilibrium is quickly escaped. The dominant traits at time t (for which $\varphi(t,.) = 0$) are then proved to stay piecewise constant, while the emergence of favorable competitors is easily computed. Once such a competitor has emerged, it disrupts the equilibrium of traits, and assumption (H) enables us to predict the issue of the equilibrium that follows.

Given any set A of present types, the equilibrium is given by $(u^*_{A,j})_{j\in A}$. Thus, the competition exerted on the i-th resource by this eco-system is $\Psi^i(A) := \sum_{j=1}^r \Psi^i(j)\,.\,u^*_{A,j}$.

Remark : To satisfy assumption (H), these steady states need to satisfy a condition of compatibility regarding their vanishing components. By restricting the dynamics on some subset B of A that contains all the non-vanishing components of the steady state associated with A, the steady state for B is necessarily the restriction on B of the steady state for A.

Proposition 6.3.3.3. *Assume that (H) is satisfied. Let $(\varepsilon_k)_{k\geq 1}$ be as in Theorem 6.3.4. For any $t \geq 0$, there exists $\rho_t > 0$ such that, for all $s \in (t, t+\rho_t]$, ψ^ε_s converges to $\psi_s = \psi_t := \Psi(\{\varphi(t,\cdot) = 0\})$, where the convergence is uniform in all compact subsets of $(t, t+\rho_t]$ and where we define for any $A \subset E$: $F(A) = (\sum_{j=1}^r \eta_i(j) u^*_{A,j})_{1\leq i\leq r}$.*

In particular, the weak limit $\mathcal{M}_s$ of $\delta_{\psi^{\epsilon_k}_s}$ obtained in Lemma 6.3.2.1 satisfies $\mathcal{M}_s = \delta_{\psi_t}$, for almost all $s \in (t, t+\rho_t)$ and the function $t \mapsto \psi_t$ is right-continuous.

As in the previous case, we shall observe a concentration effect. It means here that there are brutal transitions between one demographic equilibrium and the next one generated by the first invasion by a mutant sub-population.

Proposition 6.3.3.4. *Assume that (H) is satisfied. Any limit φ of $\varepsilon_k \log u^{\varepsilon_k}$ along a subsequence as in Theorem 6.3.4 satisfies for all $i \in E$ and for all $t \geq 0$: $\varphi(0,i) = -h(i)$*

$$\varphi(t,i) = \sup_{y(0)=i} \{-h(y(t)) + \int_0^t R(y(u), \psi_{t-u})\, du - I_t(y)\},$$

and its dynamic programming version

$$\varphi(t,\ i) = \sup_{y(s)=i} \{\varphi(s, y(t)) + \int_0^t R(y(u), \psi_{t+s-u})\ du - I_{s,t}(y)\}. \tag{6.3.14}$$

In addition, the problem (6.3.14) admits a unique solution s.t. $t \mapsto \psi_t := F(\{\varphi(t, \cdot) = 0\})$ *is right-continuous. In particular, the full sequence* $(\varepsilon \log u^\varepsilon)_{\varepsilon>0}$ *converges to this unique solution when* $\varepsilon \to 0$.

Theorem 6.3.6. *Under Hypothesis* (H) *and assuming that* $h(i) \le h(j) + \mathfrak{T}(i,j)$ *for all* $i \neq j$*, the problem*

$$\begin{cases} \partial_t \varphi(t,i) = \sup\{R(j, \psi_t) \,|\, j \in E \, s.t.\, \varphi(t,j) - \mathfrak{T}(j,i) = \varphi(t,i)\}, \\ \forall i \in E\,, \quad \varphi(0,i) = h(i) \qquad\qquad (\text{ with the convention } \mathfrak{T}(i,i) = 0) \end{cases}$$

admits a unique solution such that $t \mapsto \psi_t = F(\{\varphi(t, \cdot) = 0\})$ *is right-continuous and it is the unique solution to the variational problem (6.3.14).*

The system follows a succession of equilibria (with possibly different types that equilibrate), with the population headcount of mal-adapted types vanishing linearly in their logarithm, while this headcount logarithm increases for the rare adapted ones. It goes on until one adapted type reaches the threshold for a non-negligible frequency. Then occurs a kind of "catastrophe", where the whole equilibrium is immediately renewed (in the time-scale of evolution). Given the similarity between the description in this discrete case and the continuous case presented in Subsection 6.3.2, one can infer that similar "catastrophes" might happen even in the continuous case. This is less expected, since we see meanwhile a dynamics for the dominant trait, but it cannot be excluded a priori that a brutal change of the traits happen.

Without this assumption of a unique stable equilibrium, it would not be clear what happens to the population during such a "catastrophe" event. Namely, the only information provided by $\varphi(., t)$ is the knowledge of the non-negligible types, for which $\varphi(x, t) = 0$, with a priori no means to infer otherwise the headcounts at equilibrium. Notably, different issues or an unstable behavior could lead to different competition effects, thus different dynamics after the "catastrophe".

The main restriction for this model is that the tails of the distribution are not too much involved. Indeed, when there is some barrier, where the growth rate is very low, the trajectory that gets selected upon through the optimization procedure in Theorem 6.3.2 might not be realistic. It might well be inferred from this optimization that the ancestors of the population at time t shall represent at time $s \le t$ a proportion less than 10^{-9} of the entire population at time s. Of course, it would only be reasonable for populations with a size quite larger than 10^9! And all the more since families with small population size have a high risk of going extinct. Considering the actual population size considered, such optimization procedure could be an efficient way to truncate too rare such transitions as we have proposed for the analysis in Section 6.2. Namely, for a threshold $-\varphi_\vee$, we would like to define some $\widetilde{\varphi}(t,x)$ as the supremum over the same quantity as in Theorem 6.3.2 under the additional condition on $y \in \mathcal{G}_{t,x}$ that :

$$\forall\, s \leq t, \quad -h(y_0) + I^R_s(y) - I_s(y) > -\varphi_\vee, \qquad \text{where } I^R_s(y) := \int_0^s \int_{\mathbb{R}^k} R(y_u, \psi)\widetilde{\mathcal{M}}_u(d\psi)du.$$

Naturally, we want to impose $\widetilde{\varphi}(t,x) = -\infty$ if no function $y \in \mathcal{G}_{t,x}$ is able to satisfy this condition. By this way, we forbid transitions that rely on too exceptional ancestors to be treated in such a deterministic way. While $\widetilde{\mathcal{M}}_u$ shall represent the availability of resources at time u, it shall actually depend on $\widetilde{\varphi}$, as in the case without truncation. This makes the analysis a priori much trickier. Yet, the transitions we forbid are rare so there is a lag between the time at which they occur and the time at which they have an effect on $\widetilde{\mathcal{M}}$. The estimation of $\widetilde{\mathcal{M}}$ shall thus not be more difficult than the one of $\mathcal{M}$ in the case without truncation.

Alternative strategies have been proposed in [PG10] and [MB+12] with more regularities than the proposed truncation, yet not much more biological justification. To prevent such exceptional densities, the authors use a singular term for the growth rate, which gets very negative when the density of the state is too small.

To be even more realistic, one should treat such events of crossing barriers as punctual events occurring at a very low rate. In fact, large deviation approaches may be well suited to estimate these rates. One should only remark that the growth rate on the other side of the barrier is not responsible for any increase on the rate at which such transitions occur. It only increases the probability that such exceptional crossing event leads actually to an invasion. And that is why such events are much more exceptional than with the asymptotic model given by φ, where the growth rate immediately and regularly increases the log-density on this other side. The punctual transitions that we are to describe in the next section is thus in fact well suited to deal also with such rare crossing events.

6.4 Mutations as the limiting factor

Although this section is clearly referring to the talk given by Tran, the presentation he made was not dedicated to a specific paper. It was rather a general introduction to the framework of IBM at use to prove classical scaling limits of adaptive dynamics. The focus was mostly on models where the traits are highly conserved along the lineages until exceptional events of mutation in a new-born. Relying on two articles of my choice that are in the spirit of Tran's talk, I intend to offer a broad perspective on the limiting descriptions of IBM. Given the progression of the whole paper, I favored in my choice (besides the commitment of the speaker) the coupling of several time-scales over the intrinsic complexity in the asymptotic dynamics.

This last Section 4 addresses the longest time-scales, where the limiting factor for evolution is the emergence of mutations. By its emergence, we mean not only that some individuals carry such mutations at time t but rather that it has been stabilized at a non-negligible frequency in the population. For simplicity, many authors assume that selective traits cannot coexist in the same population, and that the population size is very stable at a value depending on this dominant trait. Namely, demographic fluctuations that has been crucial in Section 6.2 are neglected, and a fortiori the individual variability detailed in Section 6.1 is only seen through an average effect. In such a deterministic approximation, the fixation or the decay of an invasive trait is generally described, as soon as the proportion of invaders is not negligible anymore, by Lotka-Volterra equations. For elementary models of interactions (notably not frequency-dependent), it is then easy to impose conditions ensuring that the fixation of one of the alleles is the only stable equilibrium of the dynamical system.

Given the assumption that such ecological transitions happen in a much shorter time-scale than the time between occurrences of new mutations (at first in a unique individual), such conditions, called "invasion implies fixation", ensures that the population stays almost always monomorphic in the evolutionary time-scale. This means in particular that the invasion of new mutations can be treated as punctual events that refreshes the equilibrium. In this view, this behavior is similar to the evolution of traits given for the discrete space in Section 6.3 (cf. the concluding remarks for more details on this link). Yet, these punctual fixation events happen at unpredictable times, corresponding to the occurrence of a successful mutation.

Note that this shall hold true even if coexistence of traits are allowed in the model. Yet, the rate at which an invasion occurs and its issue might depend on any variations of the frequencies (like oscillation of frequencies) and of the population size, which makes the analysis tricky. Of course, one could also generalize these models to include competition for resources as in Section 6.3. Yet, the specificity of these more complicated models can only be seen when several traits can coexist. If one wishes to include events of speciation, this is however a very reasonable way to justify it (specialization of two sub-populations).

For strong selection effects, one can prove negligible the time-interval between the arrival of the first mutant and the invasion of the population by its descendants (at least reaching a non-negligible frequency). Namely, a favorable mutant has a larger expectancy of offspring than individuals of resident type and shall thus generate a family that either dies out quite early or thrive at an exponential rate. Contrary to the model of discrete trait evolution described in Subsection 6.3.3, the time-scale for the occurrence of mutations is chosen here in such a way that this growth period is not a limiting factor. By "successful mutation", we mean the occurrence of the mutation in a first carrier followed by a successful invasion by its descendants of the resident population. By assuming that mutation events are sufficiently rare, we infer that successful mutations happen nearly independently of the time waiting for it (where the monomorphic population stays the same) and of unsuccessful invasions. This explains why the time-interval between two successive events of invasion is given by an exponential law without memory.

Such a process with piecewise constant population equilibria, including a dominant trait, has been originally introduced as the Trait Substitution Sequence in [MG+96] and more formally related to individual-based models in [Ch06]. A large family of extensions has emerged, notably to include coexistence of traits (cf. [CM11], [BB18]), or more complicated interactions, for instance with horizontal transfer [BC+16] or aging [MT09]. Since they deal with two additional time-scales, we rather focus on [BF+15], where events of invasion affect the dynamics of some marker, and [BBC17], which demonstrates that adaptive dynamics can indeed be an accurate approximation of some individual-based models.

6.4.1 A model combining a marker dynamics with the evolution of traits

The next four subsections present the results of Billiard, Ferrière, Méléard and Tran on the stochastic dynamics of neutral markers coupled to the one of adaptive traits [BF+15]. The authors consider an asexual population driven by births and deaths where each individual is characterized by hereditary types : a phenotypic trait under selection and a neutral marker. This analysis is motivated by research on the prevalence of selection between various species or clades that could be based on the observed variability of neutral markers.

The trait and marker spaces $\mathcal{X}$ and $\mathcal{U}$ are assumed to be compact subsets of $\mathbb{R}$. The type of individual i is thus a pair (x_i, u_i), $x_i \in \mathcal{X}$ being the trait value and $u_i \in \mathcal{U}$ its neutral marker. The individual-based microscopic model from which we start is a stochastic birth and death process with density-dependence whose demographic parameters are functions of the trait under selection and are independent of the marker. We assume that the population size scales with an integer parameter K tending to infinity so that individuals are weighted with $\frac{1}{K}$ to observe a non-trivial limit of the empirical measure. The state of the population at time $t \geq 0$, rescaled by K, is described by the point measure :

$$\nu_t^K = \frac{1}{K}\sum_{i=1}^{N_t^K} \delta_{(x_i,u_i)} = X_t^K(dx)\,\pi_t^K(x,du),$$

$$\text{where } X_t^K = \frac{1}{K}\sum_{i=1}^{N_t^K} \delta_{x_i} \text{ and } \pi_t^K(x,du) = \frac{\sum_{i=1}^{N_t^K} \mathbf{1}_{x_i=x}\delta_{u_i}}{\sum_{i=1}^{N_t^K} \mathbf{1}_{x_i=x}}$$

are respectively the trait marginal and the marker distribution for a given trait value x. Here, $\delta_{(x,u)}$, δ_x are respectively the Dirac measure at (x, u) and x.

With the following definitions, the authors ensure that the mutations happen at different time scales for the trait and for the marker, both longer than the individuals lifetime scale. Thus, the limiting behavior results from the interplay of these three time scales : births and deaths, trait mutations and marker mutations. To justify such a separation of time scales for the mutations, the proof relies strongly on the fact that the population size remains around the equilibrium of some dominant trait(s) when a mutation on the trait occurs. Although the authors are able to include fluctuations of the parameter u, it is only possible because u has no effect on the probability that the mutation succeeds to invade the population.

6.4.2 The individual-based model

Definition 3. • *An individual with trait x and marker u reproduces with birth rate given by $0 \leq b(x) \leq \overline{b}$, the function b being continuous and $\overline{b}$ a positive real.*

• *Reproduction produces a single offspring which usually inherits the trait and marker of its ancestor except when a mutation occurs. Mutations on trait and marker occur independently with probabilities p_K and q_K respectively. Mutations are rare and the marker mutates much more often than the trait. We assume that*

$$q_K = p_K r_K \ , \text{ with } p_K = \frac{1}{K^2}, q_K \underset{K \longrightarrow \infty}{\rightarrow} 0,\ r_K \underset{K \longrightarrow \infty}{\rightarrow} +\infty.$$

• *When a trait mutation occurs, the new trait of the descendant is $x + k \in \mathcal{X}$ with k chosen according to the probability measure $m(x, k)dk$.*

• *When a marker mutation occurs, the new marker of the descendant is $u + h \in \mathcal{U}$ with h chosen according to the probability measure $G_K(u, dh)$. For any $u \in \mathcal{U}, G_K(u, .)$ is approximated as follows when K tends to infinity :*

$$\lim_{K \to +\infty} \sup_{u \in \mathcal{U}} \left| \frac{r_K}{K} \int_{\mathcal{U}} (\phi(u + h) - \phi(u)) G_K(u, dh) - A\phi \right| = 0,$$

where $(A, \mathcal{D}(A))$ is the generator of a Feller semigroup and $\phi \in \mathcal{D}(A) \subseteq C_b(\mathcal{U}, \mathbb{R})$, the set of continuous bounded real functions on $\mathcal{U}$.

• *An individual with trait x and marker u dies with intrinsic death rate $0 \leq d(x) \leq \overline{d}$, the function d being continuous and $\overline{d}$ a positive real. Moreover, the individual experiences competition the effect of which is an additional death rate*

$$\eta(x) \, . \, C * \nu_t^K(x) = \frac{\eta(x)}{K} \sum_{i=1}^{N_t^K} C(x - x_i).$$

The quantity $C(x - x_i)$ describes the competition pressure exerted by an individual with trait x_i on an individual with trait x. We assume that the functions C and η are continuous and that there exists $\underline{\eta} > 0$ such that

$$\forall x, y \in \mathcal{X},\ \eta(x)\ C(x-y)\ \geq \underline{\eta} > 0. \ \textit{(2.6)}$$

A classical choice of competition function C is $C \equiv 1$ which is called "mean field case" or "logistic case". In that case the competition death rate is $\eta(x)N_t^K/K$.

As mentioned in the beginning of the Section 4, the authors work under the simplifying assumption that follows, ensuring that the population remains monomorphic between two events of invasion by a new trait.

Assumption "Invasion implies fixation" For all $x \in \mathcal{X}$ and for almost every $y \in \mathcal{X}$:
either $\dfrac{b(y)-d(y)}{\eta(y)C(y-x)} < \dfrac{b(x)-d(x)}{\eta(x)C(0)}$,
or $\dfrac{b(y)-d(y)}{\eta(y)C(y-x)} > \dfrac{b(x)-d(x)}{\eta(x)C(0)}$ and $\dfrac{b(x)-d(x)}{\eta(x)C(x-y)} < \dfrac{b(y)-d(y)}{\eta(y)C(0)}$.

Remark : $\hat{\mathrm{n}}_x := \frac{b(x)-d(x)}{\eta(x)C(0)}$ is the equilibrium of the dynamical system that a large population size of individuals with trait x approximate.

Moreover, in the case of logistic populations with a constant C, this assumption is satisfied as soon as $x \to \hat{\mathrm{n}}_x$ is strictly monotonous.

6.4.3 Flemming-Viot process and marker evolution

The dynamics of the marker is first defined for a constant trait x via a Flemming-Viot process as defined below. This process generalizes to a potential infinity of markers the measure-valued process describing two markers : $F_t(dv) = X_t\,\delta_0(dv) + (1-X_t)\,\delta_1(dv)$ with X the neutral Wright-Fisher diffusion defined in Section 6.2, cf. (6.2.6) with $s = 0$ and remarks below.

In the sequel, we denote by $\mathcal{P}(\mathcal{U})$ and $\mathcal{P}(\mathcal{X} \times \mathcal{U})$ the probability measure spaces respectively on $\mathcal{U}$ and on $\mathcal{X} \times \mathcal{U}$, while $\langle \nu |\ \phi \rangle$ denotes the integral of the measurable function ϕ against the measure ν. Also, $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$ is the set of finite measures on $\mathcal{X} \times \mathcal{U}$.

Given $x \in \mathcal{X}$ and $u \in \mathcal{U}$, the Fleming-Viot process $(F_t^u(x, t\ \geq\ 0)$ indexed by x, started at time 0 with initial condition δ_u and associated with the mutation operator A is the $\mathcal{P}(\mathcal{U})$- valued process whose law is characterized as the unique solution of the following martingale problem. For any $\phi \in \mathcal{D}(A)$,

$$M_t^x(\phi) = \langle F_t^u(x,\,.)|\,\phi\rangle - \phi(u) - b(x)\ \textstyle\int_0^t \langle F_s^u(x,\,.)|\,A\phi\rangle\,ds \tag{6.4.15}$$

is a continuous square integrable martingale with quadratic variation process

$$\langle M^x(\phi)\rangle_t = \frac{2b(x)}{\hat{n}_x}\int_0^t (\langle F_s^u(x,\,.)|\,\phi^2\rangle - \langle F_s^u(x,\,.)|\,\phi\rangle^2)ds.$$

Remark : The model presented in Section 6.2 provides simple illustrations for such kind of processes. For the solution X of (6.2.6), define :

$$F_t(dv) := X_t\, \delta_0(dv) + (1 - X_t)\, \delta_1(dv).$$

Then, with $A \equiv 0$ (the only transitions for the traits are between 0 and 1), S a continuous function such that $S(0) = s$ (the parameter of selection at individual level) and $S(1) = 0$, for any φ measurable :

$$\begin{aligned} &M_t(\phi) := \langle F_t | \, \phi \rangle - \langle F_0 | \, \phi \rangle - (\langle F_t | \, S \times \phi \rangle - \langle F_t | \, S \rangle \times \langle F_t | \, \phi \rangle) \\ &= [\phi(0) - \phi(1)]\, (X_t - x - \textstyle\int_0^t s\, X_r (1 - X_r)\, dr) = [\phi(0) - \phi(1)]\, \sigma\, \textstyle\int_0^t \sqrt{X_r (1 - X_r)}\, dB_r, \end{aligned}$$

where we used Ito's formula, is clearly a square-integrable martingale with quadratic variation :

$$\langle M(\phi) \rangle_t = [\phi(0) - \phi(1)]^2\, \sigma^2 \int_0^t X_r (1 - X_r) dr = \sigma^2 \int_0^t (\langle F_r | \, \phi^2 \rangle - \langle F_r | \, \phi \rangle^2)\, dr.$$

Note that s is here a selective effect on the distribution that is not present in (6.4.15) because the marker is neutral.

Moreover, recalling the equation that described the state μ_t of the population of groups in Section 6.2, one can relate it to equation (6.4.15) with $b(x)\, A = \mathcal{L}_{WF}$ and the martingale $M(\phi)$ being identically zero (i.e. with zero quadratic variations). In [LM15], the authors in fact derive another description of the population of groups in the limit of large population sizes (intra-groups and inter-groups). This limit is also described as such Flemming-Viot process, with a non-zero martingale because one does no longer neglect the non-selective birth and death events of groups. Again, its quadratic variations satisfies : $d \langle M(\phi) \rangle_t \propto \int_0^t (\langle F_u | \, \phi^2 \rangle - \langle F_u | \, \phi \rangle^2)\, du$. Of course, there is still the additional term involving r in the finite variation part (6.4.15). I refer for instance to [Et00] and [Daw96] for a detailed presentation of Flemming-Viot processes.

6.4.4 Convergence to the Substitution Flemming-Viot Process

We can now state the main theorem that describes the slow-fast dynamics of adaptive traits and neutral markers at the (trait) evolutionary time scale.

Theorem 6.4.1. *We work under definition 3 and the assumption that "invasion implies fixation". Consider the initial condition* $\nu_0^K(dy, dv) = n_0^K \delta_{(x_0, u_0)}(dy, dv)$ *with* $\lim_{K \to \infty} n_0^K = \hat{n}_{x_0}$ *and* $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^3) < +\infty$*. Then, the population process* $(\nu_{Kt}^K,\ t \geq 0)$ *converges in law to the* $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$*- valued process* $(V_t(dy, dv),\ t \geq 0)$*.*

To define this Markov process, with initial condition $\hat{n}_{x_0}\, \delta_{x_0}\, \delta_{u_0}$*, we only need to describe it until the first jump of the trait, which is given by an exponential law. Namely,* *the trait jumps from x_0 to $x_0 + k$ with rate :*

$$b(x_0)\, \hat{\mathrm{n}}_{x_0} \frac{[f(x_0+k;x_0)]_+}{b(x_0+k)} m(x_0,\, k)dk.$$

Then, given that this first jump occurs at time T, the law of the new marker is given by :

$$U \sim F_T^{u_0}(x_0, du),\ \textit{ so that } V_T(dy,\, dv) = \hat{n}_{x_0+k}\, \delta_{x_0+k}\, \delta_U.$$

Then, for any $t < T$, we have :

$$V_t(dy, dv) = \hat{\mathrm{n}}_{x_0}\delta_{x_0}(dy) F_t^{u_0}(x_0, dv),$$

and the process is defined recursively like this before the next jumps.

The convergence holds in the sense of finite dimensional distributions on $\mathcal{M}_F(\mathcal{X}\times\mathcal{U})$. In addition, the convergence also holds in the sense of occupation measures, i.e. the measure $\nu^K_{KT}(dy;\, dv)\, dt$ on $\mathcal{X}\times\mathcal{U}\times[0;T]$ converges weakly to the measure $V_t(dy;\, dv)\, dt$ for any $T > 0$.

This process is called by the authors the Substitution Fleming-Viot Process (SFVP). It generalizes the Trait Substitution Process (TSP) introduced in [MG+96] and also obtained as a limit of individual-based model in [Ch06]. The TSP is in fact the marginal of the SFVP on the trait space. The jump rate of the TSP can be easily interpreted. The mutation rate of any resident is given by $b(x_0)\, m(x_0,\, k)dk$. The number of such residents at equilibrium is nearly $\hat{n}_{x_0}\, K$. While in competition with the resident population, the survival of the lineage of the mutant depends mainly on the period where the associated sub-population is too small to disrupt the resident population. From classical results of Branching process, it survives with probability $[f(x_0+k;x_0)]_+/b(x_0+k)$. Note that only beneficial mutations pass through, so that they shall invade quickly after reaching a non-negligible proportion in the population. The product provides the rate of occurrence of such successful mutation in the whole population, that gets divided by K in the new time-scale.

Given the recent and impressive progress in sequencing and comparing genetic data between species, one has partly access to the marker dynamics. The selective dynamics is however much more difficult to follow, since one would have to evaluate mutation effects and the advantage they bring in the past eco-systems. It would thus be of high interest to be able to infer strong selective effects from the dynamics of the marker.

At each sweep, a very specific marker is selected. This effect is referred to as a genetic hitchhiking. It shall be chosen according to the law $F_T^{u_0}(x_0, du)$. We may expect that numerous selective sweeps should increase the variability of the markers. Yet, without assuming any effect on the marker dynamics from the selective component, it is not very clear that doing these steps more frequently shall increase the variance in the selected marker (after a comparable time span).

Nonetheless, such hitchhiking events usually leave a mark, at least for sexual reproduction species. Because of frequent recombination events, hitchhiking effect is mostly effective for markers closely linked to the selected allele. The diversity of variants become small that one gets closer in the genome to the selected allele. A well-known example is given by the selection for genes that favor the digestion of milk. As one can imagine, the analysis is then much more demanding than the convergence to the SFVP.

6.4.5 The last time-scale of adaptive-dynamics

As the last time-scale presented in the current paper, Adaptive Dynamics is probably the one that gives to natural selection the most predictable effect. Let us follow Baar, Bovier and Champagnat [BBC17] : the connection of Adaptive dynamics with the individual-based model is demonstrated through a single step of convergence (as long as no singularity is reached). In this context, the canonical equation of adaptive dynamics (CEAD) states that the population of interest can be considered monomorphic, and its trait x_t follows an ordinary differential equation of first order. Namely, the speed of the trait involves the mutation rate, the population size at equilibrium, some derivative of the fitness of invasion, and the squared effect of mutations in the direction of invasion. Remark that this last term is not the mean effect of mutations in this direction because the more effective is a mutation, the more likely it is to fix. In order to obtain such deterministic behavior, we again need the assumptions for the TSS, i.e. rare mutations as compared to the ecological time-scale, with negligible fluctuations around the size equilibrium and invasion implying fixation (cf. previous Subsection). Moreover, the CEAD relies on the assumption that mutations have infinitesimal effects, so that the trait evolves continuously by the accumulation of large number of such mutations. The connection of the TSS to the CEAD involves a coupled rescaling of time and of fitness effects, which is rather natural. Yet, for the actual connection with the individual-based model, it introduces the major difficulty that any favorable mutation step shall be quite insignificant and yet shall replace effectively the dominant trait. In the same idea, the dominant population shall prevent deleterious yet almost neutral mutations to invade and filter favorable mutation with an invasion success still proportional to the mutation effect.

Nonetheless, the authors of [BBC17] actually manage to demonstrate a regime of convergence to the CEAD, where these issues are rigorously controlled. Their individual-based model is quite close to the previous one, except that there is no marker anymore and that the possible mutation steps are assumed to be on some discrete and finite grid (whose mesh size goes to 0), preventing large mutations. I thus use the same notations as previously (rather than the one of [BBC17]) and refer to Subsection 6.4.2. The main difference is also a scaling parameter for the mutation effect :

When a trait mutation occurs (in a population with trait x), the new trait of the descendant is $x + \sigma_K\, k \in \mathcal{X}$ with k chosen according to the probability measure $\{m(x,k)\}_{k\in[\![-A,A]\!]}$. The mutation rate is also allowed to depend on x, and is thus given by $q_K\, M(x)$.

Besides the other assumptions we have made in Subsection 6.4.2, there are additional issues of regularity for the birth rate b, the death rate d, the mutation rate M, the sensibility to competition η and the competition kernel C, for which assumptions are required. It is also assumed that $b(x) > d(x)$ and that $C(x,x)$ is uniformly upper-bounded for any $x \in \mathcal{X}$.

There is a last assumption to ensure the absence of singularity, based on the invasion fitness of a mutant y in a resident population x :

$$f(y,x) := b(y) - d(y) - \eta(y)\, C(y,x)\, \hat{n}_x.$$

It indicates the mean growth rate of a mutant population with trait y as long as it is still negligible as compared to the resident population with trait x. As stated in the convergence to the TSS, the invasion probability of a mutant population initiated by a single individual with trait y tends to $f(y,x)_+/b(y)$ (for large population size). Here, f_+ is the positive part of f, meaning that deleterious mutations cannot invade.

Assumption 3 : For all $x \in \mathcal{X}$, $\qquad \partial_1 f(x,x) \neq 0.$

Assumption 3 implies that either $\forall x \in \mathcal{X} : \partial_1 f(x,x) > 0$ or $\forall x \in \mathcal{X} : \partial_1 f(x,x) < 0$. Therefore, coexistence of two traits is not possible. Without loss of generality, we can assume that, $\forall x \in \mathcal{X}, \partial_1 f(x,x) > 0$.

Theorem 6.4.2. *Assume that Assumptions 1 and 3 hold and that there exists a small $\alpha > 0$ such that*

$$\begin{aligned} &K^{-1/2+\alpha} \ll \sigma_K \ll 1 \\ \textit{and} \qquad &\exp(-K^\alpha) \ll q_K \ll \frac{\sigma_K^{1+\alpha}}{K\,\ln K}, \qquad \textit{as } K \to \infty. \end{aligned}$$

Fix $x_0 \in \mathcal{X}$ and let $(N_0^K)_{K\geq 0}$ be a sequence of $\mathbb{N}$- valued random variables such that N_0^K/K converges in law, as $K \to \infty$, to the positive constant $\hat{\mathrm{n}}(x_0)$ and is bounded in L^p, for some $p > 1$.

For each $K \geq 0$, let ν_t^K be the process generated by $\mathcal{L}^K$ with monomorphic initial state $(N_0^K/K).\,\delta_{\{x_0\}}$. Then, for all $T > 0$, the sequence of rescaled processes, $(\nu^K_{t/(K\, q_K\, \sigma_K^2)})_{0\leq t\leq T}$, converges in probability, as $K \to \infty$, with respect to the Skorokhod topology on $\mathbb{D}([0,T],\mathrm{M}(\mathcal{X}))$ to the measure-valued process $\hat{\mathrm{n}}(x_t)\,\delta_{x_t}$, where $(x_t)_{0\leq t\leq T}$ is given as a solution of the CEAD,

$$\frac{dx_t}{dt} = \sum_{k=-A}^{A} k\,[k\,M(x_t)\,\hat{\mathrm{n}}(x_t)\,\partial_1 f(x_t,x_t)\,]_+ m(x_t,k), \qquad \textit{with initial condition } x_0.$$

Remarks :

(i) The main result of the paper actually holds under weaker assumptions. More precisely, Assumption 3 can be replaced by the following :

Assumption $3'$. The initial state v_0^K has a.s. support $\{x_0\}$ with $x_0 \in \mathcal{X}$ satisfying $\partial_1 f(x_0, x_0) \neq 0$.

The reason is that, since $x \mapsto \partial_1 f(x,x)$ is continuous, Assumption 3 is satisfied locally. Since moreover $x \mapsto \partial_1 f(x,x)$ is Lipschitz-continuous, the CEAD never reaches in finite time an evolutionary singularity (i.e. a value $y \in \mathcal{X}$ such that $\partial_1 f(y,y) = 0$). In particular, for a fixed $T > 0$, the CEAD only visits traits in some interval D of $\mathcal{X}$ where $\partial_1 f(x,x) \neq 0$. By modifying the parameters of the model out of D in such a way that $\partial_1 f(x,x) \neq 0$ everywhere in $\mathcal{X}$, we can apply Theorem 6.4.2 to this modified process $\tilde{\nu}$. Then, we deduce that $\tilde{\nu}_{t/(Ku_K\sigma_K^2)}$ has support included in D for $t \in [0,T]$ with high probability, and hence coincides with $\nu_{t/(Ku_K\sigma_K^2)}$ on this time interval.

(ii) The condition $q_K \ll \dfrac{\sigma_K^{1+\alpha}}{K \ln K}$ allows mutation events during an invasion phase of a mutant trait, but ensures that there is no "successful mutation" event during this phase.

(iii) The fluctuations of the resident population are of order $K^{-1/2}$, thus $K^{-1/2+\alpha} << \sigma_K$ ensures that the sign of the initial growth rate is not influenced by the fluctuations of the population size. If a mutant trait y appears in a monomorphic population with trait x, then its initial growth rate is $b - d(y) - \eta(y)\, C(y,x)\langle \nu_t^K \mid 1\rangle = f(y,x) + o(\sigma_K) = (y-x)\,\partial_1 f(x,x) + o(\sigma_K)$ since $y - x = O(\sigma_K)$.

(iv) $\exp(K^\alpha)$ is the time during which the resident population stays with high probability in an $O(\varepsilon\sigma_K)$-neighborhood of an attractive domain. This is a moderate deviation result. Thus, the condition $\exp(-K^\alpha) \ll u_K$ ensures that the resident population is still in this neighborhood when a mutant occurs.

(vi) The time scale is $(K\,q_K\sigma_K^2)^{-1}$ since the expected time for a mutation event is $(K\,q_K)^{-1}$, the probability that a mutant invades is of order σ_K and one needs $O(\sigma_K^{-1})$ mutant invasions to see an $O(1)$ change of the resident trait value.

Still, such a strong filtering of mutations is probably the most questionable issue of realism concerning the modeling of evolution. The fluctuations around the deterministic system shall be extremely small and slightly deleterious mutations shall be well-separated for such conclusions to be satisfied. It seems unlikely that selective effects are so dominant even for rather large populations (a million of individuals or so). Especially since there is usually a structuring of the population in term of non-heritable or loosely heritable characteristics, that are not neutral for survival. Think for instance of the individual positions or their level of infection by parasites. As we have seen in

Section 1, it may induce much more variability compared to the case where all the individuals are identical. In this time-scale of infinitesimal mutations, we may expect to see, in addition to these selective effect mainly driven by favorable mutations, also some noise due to the fixation of almost neutral mutations. The trait of the population is still quite likely to follow the direction given by invasion fitness, yet its displacement might be quite different from the one given by the CEAD and not as regular.

6.5 Conclusion

As we have seen, there is a large class of processes that can be rigorously obtained as limits of individual-based models appropriately rescaled. This is to be expected since this representation fits the closest to simulations of populations, with the minimal set of assumptions to include any interaction of interest. Nonetheless, the proof of convergence results are quite challenging and impose to be very specific on the way time-scales are well-separated. By the coupled observation both of the proofs and the simulations, the main weaknesses of the models usually appear much more salient.

Notably, we have evaluated the difficulty in estimating the birth rate in Section 6.1 from the sole knowledge of the trait at birth (because there is a lot of fluctuations until the birth event). In Sections 6.2 and 6.3, the main issue appears to be that the predicted selection effects might be driven by too exceptional realizations of the stochastic process describing the dynamics of a typical individual. In simulations and actual life, such transitions from a very stable equilibrium to another one quite separated shall not happen exactly the way these models predict, notably for the time at which they occur. Although corrections can be made by some truncation, or by adding another term to the equation governing the density, the most realistic first step would possibly be to consider the Flemming-Viot representation introduced in Section 6.4. Yet, such measured-valued stochastic process is quite more challenging to describe. In Section 6.4, the assumption that mutations have only an infinitesimal effect appears difficult to combine with the fact that the population stays monomorphic and that the mutations are filtered depending on their effects.

This is the core of science to start with the most elementary models, like the ODE defining the growth rate of a population, and then to progressively incorporate more realistic features. With the current probabilistic tools at our disposal, it is clearly time to relate most of these models describing the dynamics of densities to the individual-based measure-valued processes. The main requirement is clearly that one averages over a large number of individuals, but this can be obtained in very various ways depending on the interactions of interest. In the case where a Central Limit Theorem holds, it can be exploited to confirm the stability of the less noisy dynamics. It may also provide another dynamics, a priori closer to IBM, with different qualitative properties as in Section 2. Moreover, the convergence results can be stated for very diverse time-scales, from the rapid adaptation of cells to the propagation of parasites and the evolution of species over millions of years. They provide an elementary way to unify the models of micro-biology, ecology and evolution. Thus, they enable to justify more rigorously the separation of the related time-scales or on the contrary motivate interesting couplings. Notably, in Section 6.2 the selective effects are closely linked to random demographic fluctuations ; while in Section 6.4, the evolution of a marker as a measure-valued process is coupled in a very specific way to the punctual events of fixation for the traits under selection.

Appendice : Illustrations par simulations

Illustrations du modèle couplé d'optimum mobile

Rappel du modèle

La dynamique du système dont on souhaite ici regarder les simulations est la suivante :

$$(S_0) \begin{cases} X_t = x - v\,t + \displaystyle\int_{[0,t]\times\mathbb{R}\times(\mathbb{R}_+)^2} w\,\mathbf{1}_{\left\{u_f \le f(N_s)\right\}} \times \mathbf{1}_{\{u_g \le g(X_{s-},w)\}} M(ds, dw, du_f, du_g) \\ N_t = n + \displaystyle\int_0^t \Big(r(X_s)\; N_s - \gamma_0 \times (N_s)^2\Big)\, ds + \sigma \int_0^t \sqrt{N_s}\; dB_s, \end{cases}$$

où X est décrit comme l'écart phénotypique (ici uni-dimensionnel pour simplifier) et N la taille de population. B_t est un $\mathcal{F}_t$ Mouvement Brownien standard et M un Processus Ponctuel de Poisson (PPP) sur $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+$, adapté à $\mathcal{F}_t$, d'intensité :

$$\pi(ds,\, dw,\, du_f,\, du_g) = ds\; \nu(dw)\; du_f\; du_g.$$

Les paramètres $v, \gamma_0, \sigma > 0$, la distribution ν et les fonctions f et g sont à ajuster et je reviens sur les valeurs prises après la présentation de la méthode numérique.

Détails sur la méthode numérique

Ces simulations ont été obtenues en calculant l'évolution des densités elles-mêmes. Cette méthode se rattache à celles de volumes finis, avec un schéma numérique explicite et une renormalisations des estimations de densité à chaque pas de temps pour prendre en compte la non-linéarité. Les transitions sur X et sur N sont opérées successivement pour réduire le temps de calculs.

J'ai aussi construit une méthode alternative avec une description de type Flemming-Viot. On suit alors une population de particules qui évoluent chacune indépendamment selon la dynamique du processus de Markov (X, N) jusqu'à l'extinction de l'une d'elle. Lorsqu'une extinction advient, on remplace indépendamment l'état de chaque particule qui vient de s'éteindre par celui d'une autre particule choisie uniformément parmi les survivantes. Cela permet de gérer la non-linéarité comme il a été montré dans [Vi11].

La dynamique de chaque particule est codée de manière explicite en réexploitant les probabilités de transition du modèle à densité. Je ne conserve donc en mémoire que les coordonnées des particules sur la grille. J'imagine qu'un schéma numérique de type Euler-Maruyama explicite donnerait des résultats similaires en temps comparable, mais je n'ai pas vérifié.

Les deux méthodes produisent des résultats similaires, même si la première (volumes finis) est intrinsèquement plus lissée et donne de meilleures estimations des faibles densités. La deuxième (Flemming-Viot) est plus rapide. Dans le cadre d'un X unidimensionnel, j'obtiens des résultats similaires en un peu plus de 20 minutes pour les volumes finis contre environ 5 minutes pour Flemming-Viot. Il va sans dire que je n'ai pas cherché à obtenir une très grande précision mais plus une analyse qualitative. Si on cherche à obtenir une description pour un espace d'état de X de plus grande dimension, l'écart en terme de rapidité et de contrainte de mémoire serait plus clairement en faveur de cette deuxième méthode. J'ai privilégié la méthode de volumes finis pour les illustrations pour leur intérêt en termes de visualisation bidimensionnelle.

Entre ces simulations, seul le taux d'arrivée des mutations a été modifié. Cela devrait être assez équivalent en changeant les vitesses, mais les grille spatiales et temporelles sont ajustées à celles-ci.

Choix des paramètres

Les détails des paramètres utilisés sont les suivants. Je me suis arrangé pour que X évolue essentiellement entre -1 et un peu au-delà de 0, en gardant l'intervalle symétrique autour de l'optimum mobile pour la visualisation. Pour la dynamique de taille de population, le taux de croissance en fonction de x est ainsi donné par le profil : $r(x) = 4 - 30 \times x$. Un profil parabolique donne des résultats très similaires (vu sur des simulations antérieures, mais une erreur a été détectée et je n'ai pas refait). Le taux de compétition est de $c = 0,1$, ce qui conduit à des tailles de population à l'équilibre (la capacité soutenable ou "carrying capacity") proche de 40 (en unité arbitraire). Le coefficient de diffusion σ vaut 2, tandis que la vitesse de l'environnement vaut 0.6. On observe donc des fluctuations rapides de la taille de population par rapport aux variations d'adaptation.

Pour représenter le changement environnemental, la grille en espace est translatée d'une unité tous les 7 pas de temps, ou chaque pas de temps dure $Dt = 2.10^{-3}$. Cela correspond à une grille en x espacée de $Dx = 7 \times v \times Dt = 8,4.10^{-3}$, avec environ 300 points. Pour mieux représenter le terme diffusif, la discrétisation est régulière pour la racine carrée de la taille de population. J'ai choisi d'utiliser une grille de 200 points. J'ai pris 100 comme borne sur la taille de populations, sachant qu'une visualisation préalable me laissait voir que la QSD avec un trait optimal fixé a très peu de masse au-delà de 60.

Le profil des effets additifs de mutations est donné par $\nu(dw) = \frac{1}{2w_0} \exp(-|w|/w_0)$.

Il est donc exponentiel symétrique, avec $w_0 = 0.03$, donc avec beaucoup de petites mutations. L'effet de la taille de population sur le taux de fixation est simplement proportionnel $f_N(n) = m \times n$. Le taux de mutation m est le seul paramètre modifié ici entre les 3 jeux simulations : il prend les valeurs $m = 0.85$, $m = 0.55$ et $m = 0.25$.

Pour définir la probabilité d'invasion $g(x, w)$, j'ai utilisé la dérivée en petites proportions de sous-population mutante de la probabilité de fixation pour le modèle de diffusion à taille de population fixée. Pour tenir compte des fluctuations de cette taille de population, j'ai utilisé la moyenne harmonique $N_H(x)$ de cette taille de la QSD à x fixé, dont on peut voir l'estimation en figure 6.1. Dans le cas où le taux d'extinction de celle-ci est grand, ces estimées sont plus discutables, mais cela ne devrait pas beaucoup changer les résultats. Dans l'expression de la probabilité d'invasion intervient l'écart de taux de croissance $\Delta r = r(x + w) - r(x)$:

$$g(x, w) := \frac{N_H(x) \times \Delta r/\sigma}{1 - \exp[-N_H(x) \times \Delta r/\sigma]}.$$

Les mutations délétères sont permises mais leur probabilité de fixation très réduite si elles sont fortes par rapport aux fluctuations de populations.

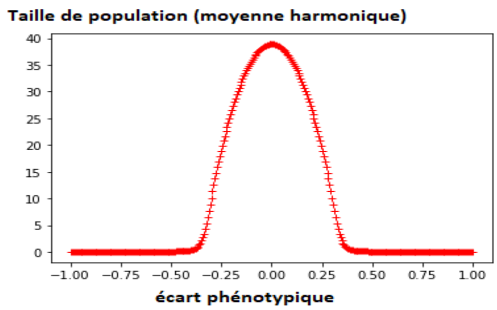


FIGURE 6.1 – $N_H(x)$: moyennes harmoniques de la taille de population pour les QSD à x fixé

Visualisation de la convergence

Je me suis aussi intéressé à évaluer le taux exponentiel de convergence de μA_t vers α via des corrélations conditionnelles. Cela fournit une mesure de référence de la perte de dépendance dans le régime métastable. Celles-ci définies comme suit pour deux

La convergence s'observe d'ici 2000 itérations, ce qui prend environ 20 minutes à calculer. En arrêtant l'algorithme après 4000 itérations, soit au temps $t = 8$, et en partant d'une condition mal-adaptée mais avec une population stable $x = -1$ et $n = 40$ pour la condition initiale, on observe les profils suivants sur l'écart en variation totale entre les distributions aux temps t courant et au temps final :

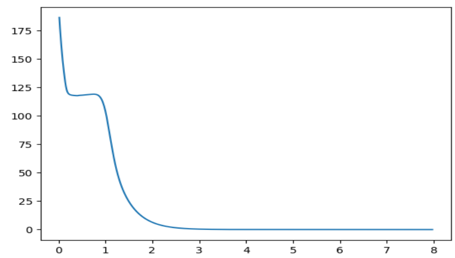


Figure 6.2 – Ecart en variation totale comparée à la dernière itération

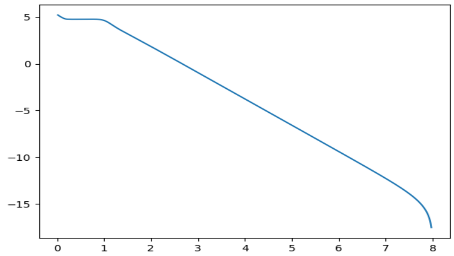


Figure 6.3 – logarithme de la distance en variation totale comparée à la dernière itération

Après une très rapide extinction qui amène à un pallier, il faut un certain temps pour qu'une part non-négligeable de la distribution atteigne la zone de survie favorable, cf profils suivants de la mesure quasi-ergodique. A partir de ce moment, on observe bien une convergence à taux manifestement exponentiel (sachant qu'il est aussi attendu comme exponentiel pour une version discrétisée du modèle). La rupture de pente qui termine la figure 6.3 est un artifice du fait que l'on regarde la distance comparée à la dernière itération plutôt qu'à la QSD. Elle témoigne de la corrélation temporelle entre les positions. Vu ces courbes, il est clair qu'on approche précisément la QSD.

fonctions f et g :

$$Cor_t(g, f) := \frac{\mathbb{E}_\alpha[g(Y_0)\, f(Y_t) \,\big|\, t < \tau_\partial] - \mathbb{E}_\alpha[g(Y_0) \,\big|\, t < \tau_\partial] \times \mathbb{E}_\alpha[f(Y_t) \,\big|\, t < \tau_\partial]}{Var_\alpha[g(Y_t^{\leftarrow})] \times Var_\alpha[f]} \tag{6.5.16}$$

où $(Y_t^{\leftarrow})_{t\geq 0}$ désigne le processus en temps arrière formellement défini pour toute fonction f et g mesurables bornées par :

$$\begin{aligned}\mathbb{E}_\alpha[g(Y_0)\, f(Y_t) \,\big|\, t < \tau_\partial] &= \mathbb{E}_\alpha[g(Y_t^{\leftarrow})\, f(Y_0^{\leftarrow})] \\ &= \lim_{T\to\infty} \mathbb{E}_\mu[g(Y_{T-t})\, f(Y_T) \,\big|\, T < \tau_\partial].\end{aligned} \tag{6.5.17}$$

Introduire ce processus me permet de gérer le problème qui se pose en temps direct avec le conditionnement à survie au temps t. J'ai donc aussi construit un schéma numérique qui décrit cette dynamique en temps arrière en se donnant une estimation de la QSD.

Pourvu que α ait une densité (aussi notée α) par rapport à la mesure de Lebesgue, la densité $(p^{\leftarrow}_t(x,y))_{t\geq 0, x,y\in\mathbb{R}_+\times\mathbb{R}}$ du semi-groupe de $Y^{\leftarrow}$ peut être reliée comme suit à celle du semi-groupe de départ $(p_t(x,y))_{t\geq 0, x,y\in\mathbb{R}_+\times\mathbb{R}}$, avec λ le taux d'extinction :

$$\begin{aligned} p^{\leftarrow}_t(x,y) &:= \frac{p_t(y,x)}{\int \alpha(z) p_t(z,x)\,dz}\,\alpha(y) \\ &= \exp[\lambda\, t]\frac{p_t(y,x)}{\alpha(x)}\,\alpha(y) \end{aligned}$$

Les transitions sur les estimations de densité sont ainsi ajustées selon cette formule. C'est de cette manière que j'ai pu obtenir une estimation des corrélations entre taille de population et écart phénotypique à deux temps distincts selon les formules (6.5.16) et (6.5.17). Notamment, la figure "Taille de population après décalage phénotypique" considère les espérances données via $f(X_t, N_t) := N_t$ et $g(X_0, N_0) = X_0$ pour évaluer $Cor_t(g,f)$ à chaque temps t.

Evolution des corrélations avec adaptation spontanée

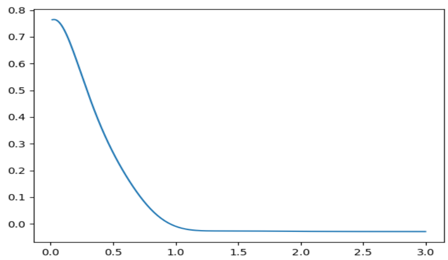


FIGURE 6.4 – $Cor_t(X,X)$: auto-corrélation du décalage phénotypique

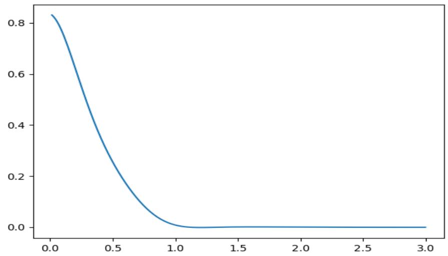


FIGURE 6.5 – $Cor_t(X,N)$: Corrélation de la taille de population selon le décalage phénotypique précédent

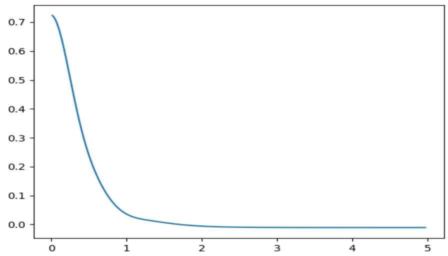


FIGURE 6.6 – $Cor_t(N,X)$: Corrélation du décalage phénotypique selon la taille de population précédente

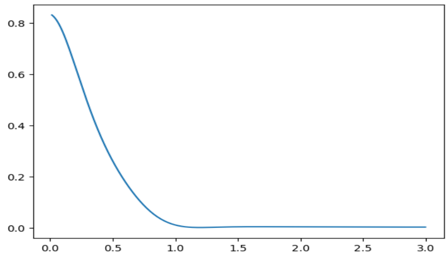
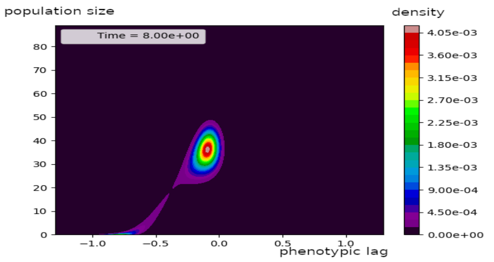


FIGURE 6.7 – $Cor_t(N,N)$: auto-corrélation de la taille de population

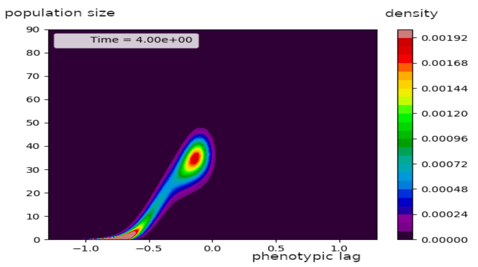
Profils caractéristiques du régime quasi-stationnaire

Profils de la "QSD"

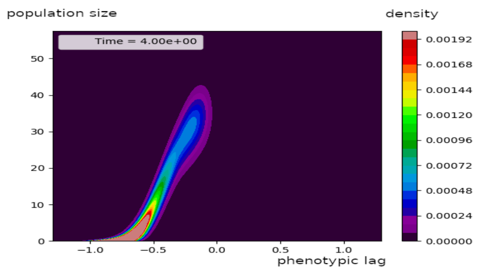


$$\mathbb{P}_{(x,n)}\left[(X,N)_t \in (dx,dn) \,\middle|\, t < \tau_\partial\right] \xrightarrow[t\to\infty]{} \alpha(dx,dn)$$

Adaptation spontanée

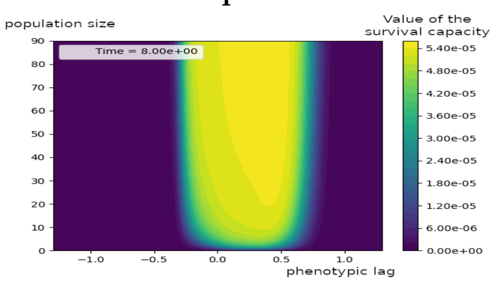


Régime intermédiaire



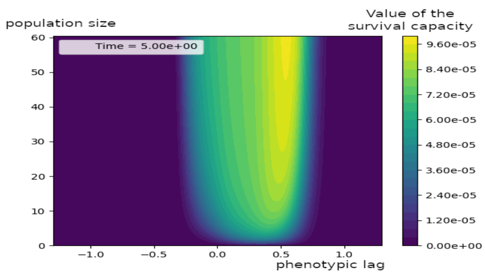
pseudo-adaptation : sélection par l'extinction

Profils de la capacité de survie

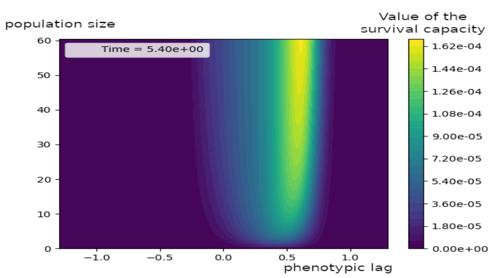


$$h(x,n) := \lim_{t\to\infty} \frac{\mathbb{P}_{(x,n)}(t < \tau_\partial)}{\mathbb{P}_{\alpha}(t < \tau_\partial)}$$

Adaptation spontanée

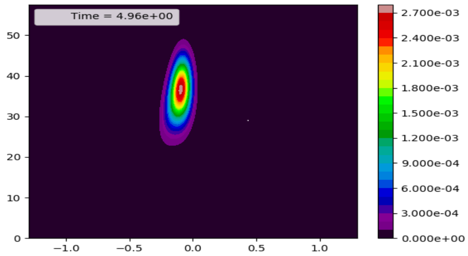


Régime intermédiaire



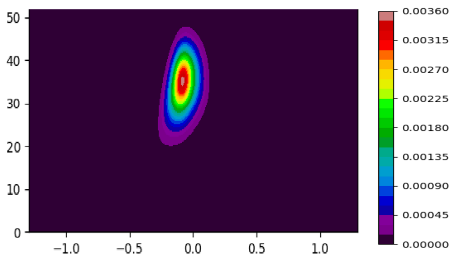
pseudo-adaptation : sélection par l'extinction

Profils de la mesure quasi-ergodique, mesure invariante du Q-processus

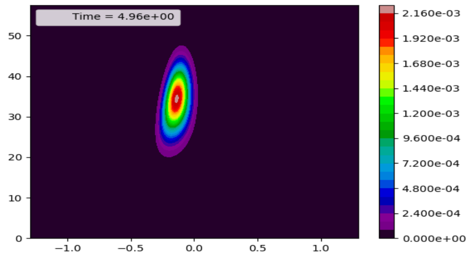


$$\lim_{t\to\infty} \lim_{T\to\infty} \mathbb{P}_{(x,n)}\left[(X,N)_t \in (dx,dn) \middle| T < \tau_\partial\right] = \beta(dx,dn)$$

Adaptation spontanée



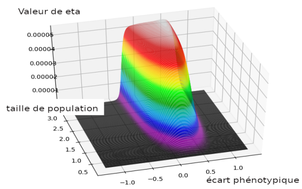
Régime intermédiaire



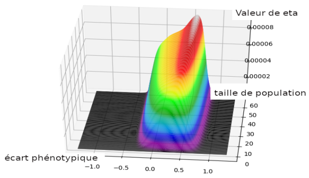
pseudo-adaptation : sélection par l'extinction

Etant donné les profils très différents que l'on avait obtenu pour les QSD, il est tout à fait remarquable que les mesures quasi-ergodiques soient aussi semblables. On voit en particulier que l'histoires des populations survivantes est encore marquée par un maintien de ces populations à de grandes tailles avec des traits proches de ceux optimaux, même lorsque de tels profils ont quasiment disparu de la quasi-stationnaire.

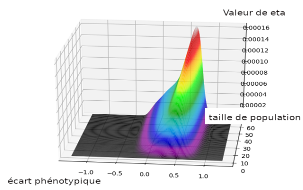
Profils 3D de la capacité de survie



L'extinction améliore peu l'adaptation ici



Régime intermédiaire



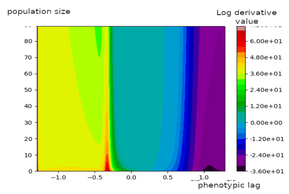
pseudo-adaptation : sélection par l'extinction

Aperçu de l'effet de conditionnement

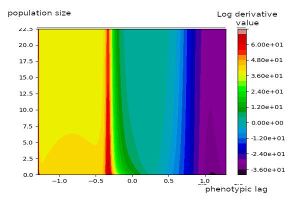
Pour mieux observer la force du biais dans la dynamique du Q-processus par rapport à ce qui résulterait du seul semi-groupe, j'ai trouvé intéressant les graphes présentant la dérivée du logarithme de la capacité de survie. Cela permet de se faire une idée du rôle de la normalisation selon la direction donnée par X via le ratio

$$h(y,n)/h(x,n) = \exp[(y-x)\int_0^1 \partial_x \log h((1-t)x+ty,n)dt].$$

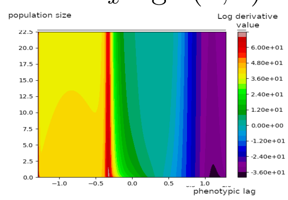
Log-dérivée selon x de la capacité de survie : $\partial_x \log h(x,z)$



Adaptation spontanée

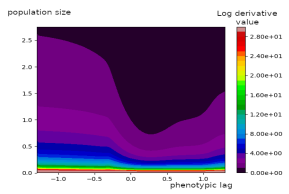


Régime intermédiaire

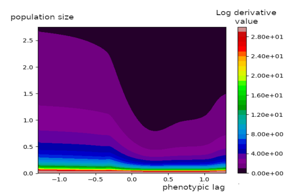


pseudo-adaptation : sélection par l'extinction

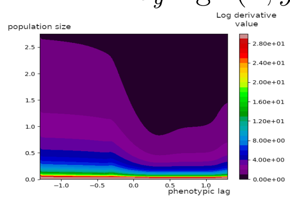
Log-dérivée selon $y \propto \sqrt{n}$ de la capacité de survie : $\partial_y \log h(x,y)$



Adaptation spontanée



Régime intermédiaire



pseudo-adaptation : sélection par l'extinction

Peu de différences selon le taux de mutations (avec des fluctuations démographiques rapides).

Le pic d'intensité situé vers la droite $x = -0.4$ est tout à fait remarquable. Cette position correspond sur le profil de la QSD à une zone "tendue" où elle semble se scinder en deux parties de plus grande densité. Il semble que cette zone est critique pour la survie : avec un niveau de mal-adaptation au-delà de ce seuil, il devient très difficile pour la population de maintenir son adaptation. Le déclin de la taille de population doit alors être trop violent, notamment vu la figure 6.1. Il n'est plus compensé par l'augmentation de la proportion de mutations favorables.
Ce pic est déjà présent même lorsque la QSD correspond au cas stable. Quand on baisse le taux de mutations en affaiblissant cette stabilité, l'intensité de ce pic semble croître relativement peu mais l'étendue de son effet devient de plus en plus grande.

Un pic antagoniste plus faible est aussi présent du côté des écarts phénotyiques positifs. Il est situé proche de la droite $x = 1.1$, ce qui le rend très surprenant. A un tel niveau de mal-adaptation (fixé), la population est d'après la figure 6.1 en danger immédiat d'extinction, sans que cela corresponde à ce seuil.

De même, la dérivée du logarithme de la densité de la QSD devrait nous informer sur la dynamique du processus en temps arrière.

Illustrations du modèle de sélection à deux niveaux

Rappel du modèle

Le modèle considéré est celui du processus μ_t obtenu au 4.1.2.1 comme limite d'un modèle individu-centré à deux niveau hiérarchique et comme l'unique solution à satisfaire pour tout $f \in C_b^2$:

$$\partial_t \langle \mu_t \,\big|\, f\rangle = \langle \mu_t \,\big|\, \mathcal{L}_{WF} f\rangle + \langle \mu_t \,\big|\, r\, f\rangle - \langle \mu_t \,\big|\, f\rangle \times \langle \mu_t \,\big|\, r\rangle \,, \qquad \mu_0 = \mu_0.$$
$$\text{où } \mathcal{L}_{WF} f(x) = \gamma_I \, x\,(1-x) \partial_{xx}^2 f(x) - s\, x\,(1-x)\, \partial_x f(x).$$

Pour l'estimer, on va exploiter la représentation donnée via la Proposition 4.1.2.2 en fonction du processus X solution de l'équation stochastique :

$$dX_t := -s\, X_t\,(1-X_t)\, dt + \sqrt{2\gamma\, X_t\,(1-X_t)}\, dB_t, \quad X_0 \sim \mu_0,$$

avec B un Mouvement Brownien standard. On exploite aussi la fonctionnelle mutiplicative : $Z_t := \exp[-\int_0^t \rho(X_s)\, ds]$, τ_0, τ_1 les temps d'atteinte par X de resp. 0 et 1, et $\tau_{0,1} := \tau_0 \wedge \tau_1$. On rappelle alors l'expression :

$$\mu_t = x_t^0\, \delta_0 + x_t^1\, \delta_1 + x_t^\xi\, \xi_t$$

$$\text{où } x_t^0 := \frac{\mathbb{E}\left[Z_{\tau_0}\, \exp[-\rho_0(t-\tau_0)]\,;\, \tau_0 < t\right]}{\mathbb{E}\left[Z_t\right]}, \qquad x_t^1 := \frac{\mathbb{E}\left[Z_{\tau_1}\, \exp[-\rho_1(t-\tau_1)]\,;\, \tau_1 < t\right]}{\mathbb{E}\left[Z_t\right]},$$
$$x_t^\xi := \frac{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\right]}, \quad \langle \xi_t \,\big|\, f\rangle := \frac{\mathbb{E}\left[f(X_t)\, Z_t\,;\, t < \tau_{0,1}\right]}{\mathbb{E}\left[Z_t\,;\, t < \tau_{0,1}\right]}, \forall\, f \in \mathcal{C}([0,1]), \; .$$

Enjeux de la méthode numérique

L'objectif de ces simulations est d'obtenir de bonnes représentations qualitatives du système jusqu'à la convergence vers l'équilibre quasi-stationnaire tout en optimisant la rapidité de calculs. L'algorithme utilisé se rattache à la famille des volumes finis, avec un schéma explicite, et la non-linéarité simplement traitée par renormalisation. Etant donné le besoin de prendre en compte des densités très faibles pour observer certaines transitions, les expressions des densités et des transitions sont traduites en échelle logarithmique. Par ce choix de volumes finis, il ne serait pas très difficile de transformer l'algorithme pour adopter une approche de type Flemming-Viot.

5 régimes de paramètres sont présentés. Pour les 3 premiers, les effets de sélection entre groupes favorisent simplement une plus grande proportion de type C avec une fonction r linéaire. On a alors nécessairement $\rho(1) < \rho(0) \wedge \rho_\alpha$. Dans les deux derniers,

on tient compte d'un effet supplémentaire en faveur du polymorphisme. Dans les 2 premiers, le terme de diffusion est faible, ce que l'on caractérise qualitativement par l'exceptionalité des transitions depuis un état médian de type C vers un état dominé par les types C. On distingue d'abord le cas $\rho(0) < \rho_\alpha$ où la QSD a un taux de survie moins bon que l'état des groupes purement composés de D. On regarde ensuite un cas où $\rho(0) < \rho_\alpha$, sans que les résultats qualitatifs ne soient très différent. En amplifiant la dérive, on illustre par la suite un autre cas où $\rho_\alpha < \rho(0)$ pour lequel le rôle de la QSD polymorphique α est bien plus conséquent. On poursuit la logique dans les deux derniers cas d'un rôle toujours plus important joué par la dynamique polymorphique, en passant du cas $\rho(1) < \rho_\alpha < \rho(0)$ à $\rho_\alpha < \rho(1) < \rho(0)$. La question de l'allure de cette QSD α (concentrée ou non sur un bord) et de voir si on peut observer des QSD transitoires fait partie de l'analyse.

Au voisinage du cas purement sélectif : $\rho_1 < \rho_0 < \rho_\alpha$

Dans ce cas, on constate en effet que la QSD polymorphe ne joue en fait aucun rôle dans la dynamique observée car le logarithme de $\mu A_t(0,1)$ reste toujours bien inférieur à 0. La valeur de $r_1 = \rho_0 - \rho_1$ peut être déduite de la pente d'augmentation des purs groupes C avant l'invasion et leur domination finale, ainsi que de la pente de diminution des groupes de type D. Concernant la décroissance des groupes polymorphiques, la première pente correspond à $\rho_{\alpha'} - \rho_0$, où α' est la QSD observée dans la troisième figure, tandis que la seconde pente est $\rho_\alpha - \rho_0$ et la troisième $\rho_\alpha - \rho_1$. Bien que les QSD réelle et transitoire semblent concentrées autour de 0, On voit bien que la première apparaît bien après la seconde, presque lorsque les groupes de purs C émergent. Cela indique que la survie des premiers dépend principalement de sa queue de distribution du côté des purs groupes C.

Parce que la proportion croissante de purs groupes C part de très bas, on ne s'attend pas à ce que leur domination finale soit visible même pour une population de groupes raisonnablement importante.

Les paramètres utilisées pour ces simulations sont les suivants : $\sqrt{2\gamma} = 5.10^{-3}$, $s = 0.1$ et $r(x) = 5.10^{-3} \times x$. La discrétisation dans l'espace est ajustée au niveau de diffusion à chaque itération avec une grille d'environ 100 points. Par des estimations numériques, l'idée est de pouvoir se ramener à une grille telle que la diffusion induit une répartition sur un nombre similaire de voisins quelle que soit la position sur $(0,1)$. Ainsi, à proximité de 0, le terme de martingale de $\sqrt{X_t}$ est presque constant. On veut donc produire plus de points de références aux voisinages de 0 et 1. Néanmoins, pour 100 points, il n'est pas si clair que ce travail est vraiment essentiel. La discrétisation en temps est ajustée pour avoir à chaque discrétisation une répartition sur essentiellement 5 points de la grille, ce que l'on obtient avec $Dt = 2$.

Nous prenons δ_0 avec une croissance nulle, δ_1 a un taux de croissance de 5.10^{-3}

alors que le taux d'extinction de α est estimé à $2,39.10^{-3}$. La condition initiale est une masse de Dirac à environ 0,4. L'estimation numérique $\mu_t^{[N]}$ est rapidement déportée vers 0, jusqu'à ce qu'elle atteigne la QSD transitoire. L'émergence de la QSD réelle est assez brutale avec une augmentation uniforme dans l'espace.

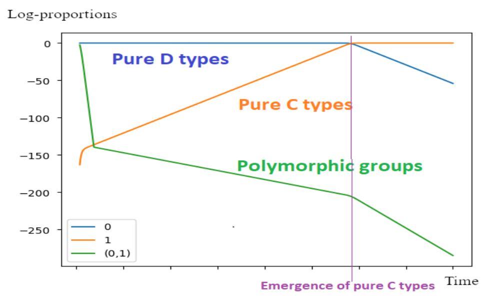


FIGURE 6.8 – Dynamique en temps des log-proportions, pures et polymorphiques Avec les notations de la Proposition 4.1.2.2, les courbes en temps présentées sont resp. $\{\log(x_t^0)\}_{t\geq 0}$ en bleu, $\{\log(x_t^1)\}_{t\geq 0}$ en orange et $\{\log(x_t^\xi)\}_{t\geq 0}$ en vert.

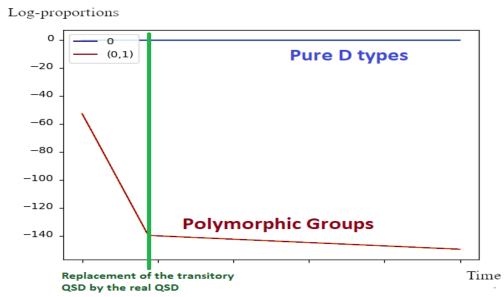


FIGURE 6.9 – Dynamique en temps des log-proportions, pures D et polymorphiques
Les courbes en temps présentées sont resp. $\{\log(x_t^0) - \log(x_t^0 + x_t^\xi)\}_{t\geq 0}$ en bleu et $\{\log(x_t^\xi) - \log(x_t^0 + x_t^\xi)\}_{t\geq 0}$ en rouge.

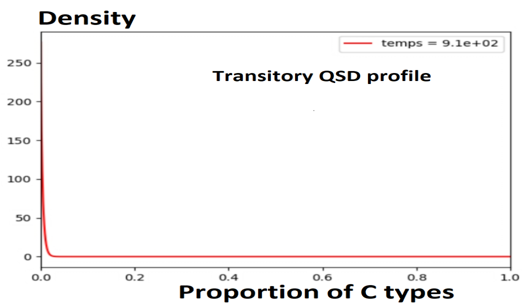


FIGURE 6.10 – Profil de la QSD transitoire de groupes polymorphiques

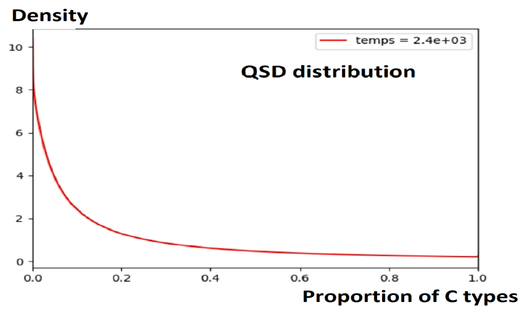


FIGURE 6.11 – Profil de la QSD réelle et finale de groupes polymorphiques

Au voisinage du cas purement sélectif mais avec $\rho_1 < \rho_\alpha < \rho_0$

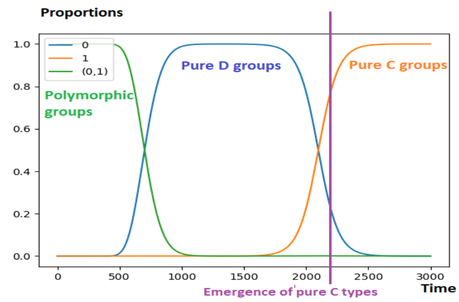


FIGURE 6.12 – Dynamique des proportions
Les courbes en temps présentées sont resp. $\{x_t^0\}_{t\geq 0}$ en bleu, $\{x_t^1\}_{t\geq 0}$ en orange et $\{x_t^\xi\}_{t\geq 0}$ en vert.

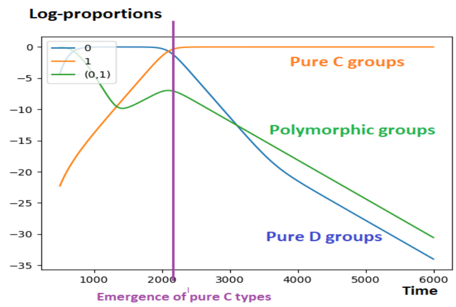


FIGURE 6.13 – Dynamique des log-proportions
Les courbes en temps présentées sont resp. $\{\log(x_t^0)\}_{t\geq 0}$ en bleu, $\{\log(x_t^1)\}_{t\geq 0}$ en orange et $\{\log(x_t^\xi)\}_{t\geq 0}$ en vert.

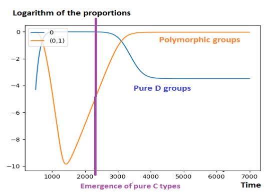


FIGURE 6.14 – Sans les purs groupes C

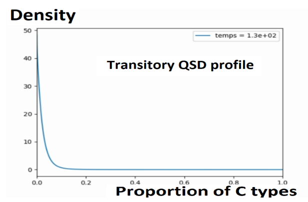


FIGURE 6.15 – fausse QSD transitoire

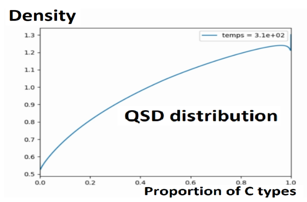


FIGURE 6.16 – QSD réelle

On peut observer ici que même si $\rho_\alpha < \rho_0$, la QSD joue encore un rôle négligeable parce qu'elle émerge trop lentement par rapport aux purs groupes C. Là encore, la domination finale par les purs groupes C ne serait pas aussi probable dans une population finie.

Les paramètres utilisés pour ces simulations sont les suivants : $\sqrt{2\gamma} = 0,02$, $s = 0,3$ et $r(x) = 0,1 \times x$. La discrétisation en espace est ajustée avec une grille d'environ 100 points. La discrétisation en temps est ajustée pour avoir des sauts vers un endroit proche avec $Dt = 0,125$.

Nous prenons δ_0 avec une croissance nulle, δ_1 a un taux de croissance de 0,1 tandis que le taux de croissance de α est estimé à $5,02.10^{-2}$.

La condition initiale est une masse de Dirac à environ 0,6. $\mu_t^{[N]}$ se laisse déporter vers 0 avec un peu plus de diffusion que dans le cas précédent tant qu'elle reste polymorphe, jusqu'à atteindre la QSD transitoire. L'émergence de la QSD réelle est à nouveau assez brutale avec une augmentation uniforme dans l'espace. Maintenant que $\rho_\alpha > \rho(0)$, la QSD semble concentrée autour de 1. Il serait intéressant de vérifier si cela pourrait se généraliser, mais cela semble naturel.

Pour observer un retournement apparent de la dérive, on peut changer les paramètres en $\sqrt{2\gamma} = 5.10^{-3}$, $s = 0.03$ et $r(x) = 0.1 \times x$, ce qui me conduit à $Dt = 2$

Étape intermédiaire où la QSD est conséquente $\rho_1 < \rho_\alpha < \rho_0$

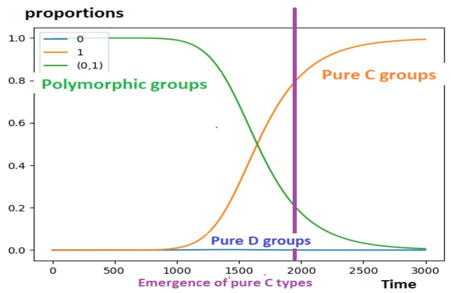


FIGURE 6.17 – Dynamique des proportions

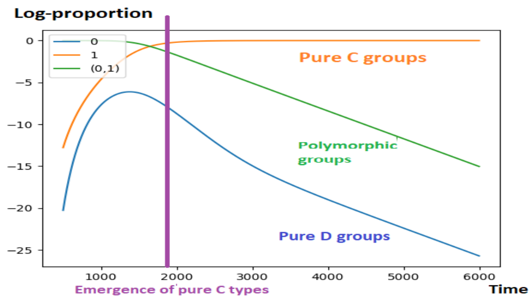


FIGURE 6.18 – Dynamique des log-proportions

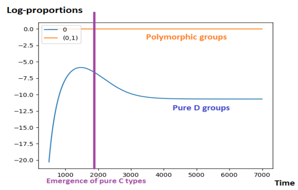


FIGURE 6.19 – Sans les purs groupes C

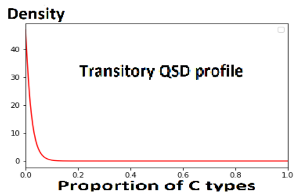


FIGURE 6.20 – "fausse" QSD transitoire

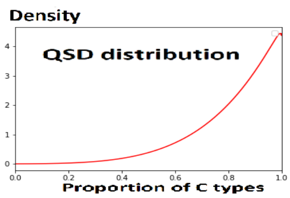


FIGURE 6.21 – vraie QSD

On peut observer ici que même si $\rho_\alpha < \rho_0$, la QSD joue encore un rôle négligeable parce qu'elle émerge trop tard par rapport aux purs groupes C. Là encore, la domination finale par les purs groupes C ne serait pas aussi probable dans une population finie. Comme indiqué dans la proposition 4.2.2.2, $\log(\mu A_t(0,1)/\mu A_t\{0\})$ tend vers :

$$\log(y_\alpha/y_0) = \log((\rho_0 - \rho_\alpha)/\rho_\alpha) - \log(\mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})). \tag{6.1}$$

La taille de l'écart entre les deux asymptotiques dans la figure 6.19 fournit donc une estimation de $\mathbb{P}_\alpha(\tau_0 = \tau_{0,1,\partial})$, qui est ici particulièrement faible.

Les paramètres utilisés pour ces simulations sont les suivants : $\sqrt{2\gamma} = 0,02$, $s = 0,1$ et $r(x) = 0,1 \times x$. La discrétisation dans l'espace est ajustée avec une grille d'environ 100 points. La discrétisation en temps est ajustée avec $Dt = 0,125$.

Nous prenons δ_0 avec une croissance nulle, δ_1 a un taux de croissance de 0,1 tandis que le taux de croissance de α est estimé à $7,35.10^{-2}$.

Avec une masse de Dirac à environ 0,6 comme condition initiale, la QSD se démarque depuis $\mu_t^{[N]}$. avant qu'elle ne s'approche de 0, alors que le profil est encore très polymorphe. Pour obtenir la QSD transitoire, nous partons d'une masse de Dirac très proche de 0.

Effets sélectifs en faveur du polymorphisme

Dans le cadre de ces simulations, il y a des effets sélectifs pour les groupes qui sont à un degré intermédiaire de polymorphisme. Une QSD transitoire peut alors être observée des deux côtés, tandis que la QSD réelle se situe plutôt au milieu.

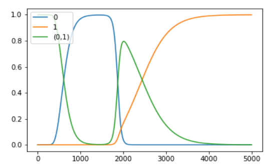


FIGURE 6.22 – Dynamique des proportions

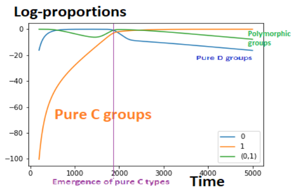


FIGURE 6.23 – Dynamique des log-proportions

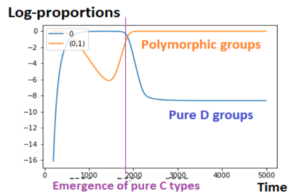


FIGURE 6.24 – Sans les purs groupes C

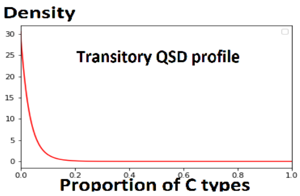


FIGURE 6.25 – QSD transitoire

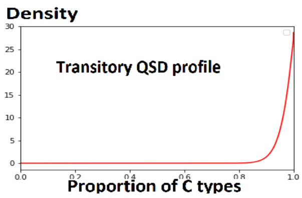


FIGURE 6.26 – Autre QSD transitoire

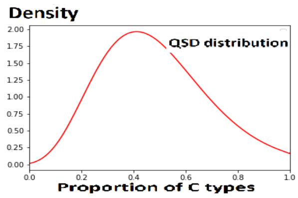


FIGURE 6.27 – QSD réelle

Les paramètres utilisés pour ces simulations sont les suivants : $\sqrt{2\gamma} = 0,03$, $s = 0,5$ et $r(x) = 0,5 \times x + x \times (1 - x)$. La discrétisation en espace est ajustée avec une grille d'environ 100 points. La discrétisation en temps est ajustée avec $Dt = 5.56.10^{-2}$.
En prenant δ_0 avec une croissance nulle, δ_1 a un taux de croissance de 0,5 tandis que le taux de croissance de α est estimé à $0,446$, ce qui est assez proche de celui de 1.

Le cas $\rho_\alpha < \rho(1) < \rho(0)$

L'extension au cas $\rho_\alpha < \rho(1) < \rho(0)$ n'apporte pas particulièrement de nouvelles idées, si ce n'est que le QSD interne finit par dominer. Voici ce qui se passe si nous fixons $r(x) = 0,5{\times}x + 2\,x{\times}(1-x)$, de sorte que α ait un taux de croissance autour de $0,52$:

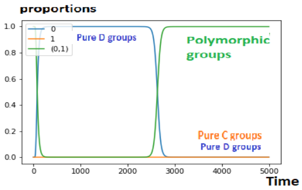


FIGURE 6.28 – Dynamique des proportions

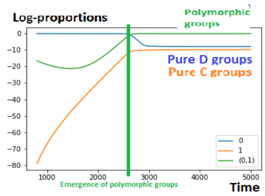


FIGURE 6.29 – Dynamique des log-proportions

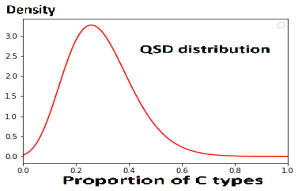


FIGURE 6.30 – QSD réelle

Bibliographie

[AFP88] Aldous, D., Flannery, B., Palacios, J.L. ; Two applications of urn processesthe fringe analysis of search trees and the simulation of quasi-stationarydistributions of markov chains. Probab. in the Engin. and Inform. Sciences, V.2, N.3, pp.293–307 (1988)

[ANT80] Arjas, E., Nummelin, E., Tweedie, R. L. ; Semi-Markov processes on a general state space : α-theory and quasi-stationarity. J. Austral. Math. Soc., Ser. A, V.30(2), pp.187–200 (1980)

[AP13] Audiffren, J., Pardoux, E. ; Muller's ratchet clicks in finite time, Stoch. Proc. and their Appl., V.123, pp.2370–2397 (2013)

[AMZ20] Aurzada, F., Mukherjee, S., Zeitouniv, O. ; Persistence exponents in Markov chains, ArXiv :1703.06447

[BP12] Ba, M., Pardoux, E. ; The effect of competition on the height and length of the forest of genealogical trees of a large population, Malliavin Calculus and Related Topics, Springer Proceedings in Mathematics and Statistics, V.34, pp.445-467 (2012).

[BB18] Baar, M., Bovier, A. ; The polymorphic evolution sequence for populations with phenotypic plasticity, Electron. J. Probab., V.23, N.72, pp.1-27 (2018)

[BBC17] Baar, M., Bovier, A., Champagnat, N. ; From stochastic, individual-based models to the canonical equation of adaptive dynamics in one step. Ann. Appl. Probab., V.27, N. 2, pp. 1093-1170. (2017)

[BCG19] Bansaye, V., Cloez, B., Gabriel, P. ; Ergodic behavior of non-conservative semigroups via generalized Doeblin's conditions, P. Acta Appl Math, V.166, N.1, pp.29-72. (2019)

[BCGM19] Bansaye, V., Cloez, B., Gabriel, P., Marguet, A. ; A non-conservative Harris' ergodic theorem, preprint on ArXiv : 1903.03946 (2019)

[BCMM19] Bansaye, V., Collet,P., Martinez, S., Méléard, S., San Martin, J. ; Diffusions from Infinity, Trans. Amer. Math. Soc., V.372, pp. 5781-5823 (2019)

[BD$^+$11] Bansaye, V., Delmas,J-F., Marsalle, L., Tran, V. ; Limit theorems for Markov processes indexed by continuous time Galton–Watson trees Ann. Appl. Probab., V.21, N.6, pp. 2263-2314 (2011)

[BM15] Bansaye, V., Méléard, S. ; Stochastic Models for Structured Populations, Scaling Limits and Long Time Behavior, Stochastics in Biological Systems, V.1, N.4, Springer Switzerland (2015)

[BP10] Barbour, A.D., Pollett, P.K. ; Total variation approximation for quasi-stationary distributions, J. of Appl. Probab., V. 47, pp. 934–946 (2010)

[BEV17] Barton N.H., Etheridge A.M., Véber A. ; The infinitesimal model : Definition, derivation, and implications. Theor Popul Biol. ; V.118, pp.50-73 (2017)

[Ba95] Bass, R. ; Probabilistic Techniques in Analysis, Probab. and its Appl., Springer (1995)

[Ba98] Bass, R. F. ; Diffusions and Elliptic Operators, Probab. and Its Applications, Springer, New York (1998)

[Ba10] Bass, R. F. ; The measurability of hitting times, Elec. Comm. in Probab., V.15, pp.99-105 (2010)

[BCV20] Benaïm, M., Champagnat, N., Villemonais, D. ; Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain ArXiv :1904.08620

[BCP18] Benaim, M., Cloez, B., Panloup, F. ; Stochastic approximation of quasi-stationary distributions on compact spaces and applications, Ann. Appl. Probab., V.28, N.4, pp.2370-2416 (2018)

[Be19] Bertoin, J ; On a Feynman-Kac approach to growth-fragmentation semigroups and their asymptotic behaviors, J. Funct. Anal., V.277, N.11, pp. (2019)

[BCL17] Bonnefon, O., Coville, J., Legendre, G : Concentration phenomenon in some non-local equation, Discrete and Continuous Dynamical Systems, V.22, N.3, p.763-781 (2017)

[BCV16] Berestycki, H., Coville, J., Vo, H-H. ; Persistence criteria for populations with non-local dispersion, J. Math. Biol., V.72, pp.1693-1745 (2016)

[BD$^+$09] Berestycki, H., Diekmann, O., Nagelkerke, C. J., Zegeling, P. A. ; Can a species keep pace with a shifting climate ?, Bull. Math. Biol., V.71, N.2, pp.399–429 (2009)

[BHR05] Berestycki, H., Hamel, F., Roques, L. ; Analysis of the periodically fragmented environment model. I. Species persistence, J. Math. Biol., V.51, N.1, pp.75–113. (2005)

[BHR08] Berestycki, H., Hamel, F., Roques, L. ; Reaction-diffusion equations for population dynamics with forced speed. I. The case of the whole space, Discrete Contin. Dyn. Syst., V.21, N.1, pp.41–67 (2008),

[BHR09] Berestycki, H., Hamel, F., Roques, L. ; Reaction-diffusion equations for population dynamics with forced speed. II : Cylindrical-type domains, Discrete Contin. Dyn. Syst., V.25, N.1, pp.19–61 (2009)

[BRZ19] Bierkens, J., Roberts, G.O., Zitt, P-A. ; Ergodicity of the zigzag process, Ann. Appl. Probab., V.29, N.4, pp.2266-2301 (2019)

[BC+16] Billiard, S., Collet,P., Ferrière, R., Méléard,S., Tran, V. ; The effect of competition and horizontal trait inheritance on invasion, fixation and polymorphism. Journal of Theoretical Biology, Elsevier, V.411, pp.48-58 (2016)

[BF+15] Billiard, S., Ferrière, R., Méléard, S., Tran, V. ; Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks J. Math. Biol. 71 : 1211-1242 (2015)

[BCGL17] Bourgeron, T., Calvez, V., Garnier, J., Lepoutre, T. ; Existence of recombination-selection equilibria for sexual populations. preprint : hal-01489695v2 (2017)

[BL95] Bürger, R., Lynch, M. ; Evolution and Extinction in a Changing Environment : A Quantitative-Genetic Analysis ; Evolution, V.49, N.1, pp.151-163 (1995)

[BR99] Breyer L.A., Roberts, G.O ; A quasi-ergodic theorem for evanescent processes ; Stoch. Proc. and their Appl., V.84, N.2, pp.177-186 (1999)

[CC98] Cantrell, R. S., Cosner, C. ; On the effects of spatial heterogeneity on the persistence of interacting species, J. Math. Biol., V.37, N.2, pp.103–145. (1998)

[CC+09] Cattiaux, P., and all ; Quasi-Stationary Distributions and Diffusion Models in Population Dynamics, The Annals of Probab., V.37, N.5, pp.1926–1969 (2009)

[CG16] Cattiaux, P., Guilin, A. ; Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity, J. of Funct. Analysis, V.272, N.6, pp.2361-2391 (2016)

[CGZ10] Cattiaux, P., Guilin, A., Zitt, P.A. ; Poincaré inequalities and hitting times, Annales de l'institut Henri Poincare (B) Probab. and Stat., V. 49, pp. 95-118 (2010)

[Ch06] Champagnat, N. ; A microscopic interpretation for adaptative dynamics trait substitution sequence model. Stoch Process Appl 116 :1127–1160 (2006)

[CCV18] Champagnat, N., Coulibaly-Pasquier, K., Villemonais, D. ; Convergence to quasi-stationary distribution for multi-dimensional diffusion processes, Séminaire de Probabilités XLIX, V.2215, pp.165-182, Springer (2018)

[CFM06] Champagnat, N., Ferrière, R., Méléard, S. ; Unifying evolutionary dynamics : from individual stochastic processes to macroscopic models. Theoretical Population Biology, Elsevier, V.69, N.3, pp.297-321 (2006)

[CFM08] Champagnat, N., Ferrière, R., Méléard, S. From individual stochastic processes to macroscopic models in adaptive evolution. Stoch. Models, V.24, Suppl.1, pp.2-44 (2008)

[CH19] Champagnat, N., Henry, B. ; A probabilistic approach to Dirac concentration in non-local models of adaptation with several resources, Ann. Appl. Probab., V.29, N.4, pp.2175-2216 (2019)

[CM11] Champagnat, N., Méléard, S. ; Polymorphic evolution sequence and evolutionary branching. Probab. Theory Relat. Fields, V.151, pp.45–94 (2011)

[CV16] Champagnat, N., Villemonais, D. ; Exponential convergence to quasi-stationary distribution and Q-process, Probab. Theory Relat. Fields, V. 164, pp. 243-283 (2016)

[CV17] Champagnat, N., Villemonais, D. ; Exponential convergence to quasi-stationary distribution for absorbed one-dimensional diffusions with killing. ALEA - Lat. Am. J. of Probab. and Math. Stat., V. XIV, pp. 177-199 (2017).

[CV18a] Champagnat, N., Villemonais, D. ; Uniform convergence of conditional distributions for absorbed one-dimensional diffusions, Adv. in Appl. Probab., V.50, N.01, pp.178-203 (2018)

[CV18b] Champagnat, N., Villemonais, D. ; Uniform convergence of time-inhomogeneous penalized Markov processes, ESAIM : Probab. and Stat., V.22, pp.129-162 (2018)

[CV17b] Champagnat, N., Villemonais, D. ; Lyapunov criteria for uniform convergence of conditional distributions of absorbed Markov processes, preprint on ArXiv : 1704.01928 (2017)

[CV17c] Champagnat, N., Villemonais, D. ; General criteria for the study of quasi-stationarity, preprint on ArXiv : 1712.08092 (2017)

[CCM16] Chazottes, R., Collet, P., Méléard, S. ; Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes, Probab. Theory Relat. Fields, V.164, N.1, pp. 285–332 (2016)

[CCM17] Chazottes, R., Collet, P., Méléard, S. ; On time scales and quasi-stationary distributions for multitype birth-and-death processes, Ann. Inst. H. Poincaré Probab. Statist., V.55, N.4, pp.2249-2294 (2019)

[CLJ12] Chen, J., Li, H., Jian, S. ; Some limit theorems for absorbing Markov processes ; J. Phys. A : Math. Theor., V.45, N.34 (2012)

[Cl17] Cloez, B. ; Limit theorems for some branching measure-valued processes ; Adv. Appl. Probab., V.49, N.2, pp. 549-580 (2017)

[CM+11] Collet, P., Martínez, S., Méléard, S., San Martín, J. ; Quasi-stationary distributions for structured birth and death processes with mutations, Probab. Theory Relat. Fields, V. 151, pp. 191–231 (2011)

[CMS13] Collet, P., Martínez, S., San Martín, J. ; Quasi-Stationary Distributions, Probab. and Its Appl., Springer, Berlin Heidelberg (2013)

[C10] Coville, J., On a simple criterion for the existence of a principal eigenfunction of some nonlocal operators J. Differential Equations, V.249, pp.2921-2953 (2010)

[CDM08] Coville, J., Dávila, J., Martínez, S. ; Existence and uniqueness of solutions to a nonlocal equation with monostable nonlinearity. SIAM J. Math. Anal., V.39, pp.1693–1709 (2008)

[CH20] Coville, J, Hamel, F. ; On generalized principal eigenvalues of non-local operators with a drift, Nonlinear Analysis, V.193 (2020)

[CLW17] Coville, J., Li, F., Wang, X. ; On eigenvalue problems arising from nonlocal diffusion models, Discrete Contin.Dyn. Syst., V.37, pp.879-903 (2017)

[DV08] Daley, D.J., Vere-Jones, D. ; An Introduction to the Theory of Point Processes : Volume II : General Theory and Structure ; Probab. and Its Appl., Springer, 2nd ed. (2008)

[Daw96] Dawson, D. A. ; Mesure-valued markov processes. v. 1541 of Lectures Notes in Math., Ecole d'Eté de probabilités de Saint-Flour XXI, Springer, pp.1–260, New York (1993)

[Daw10] Dawson, D.A. ; Introductory Lectures on Stochastic Population Systems, Technical Report Series No. 451, Laboratory for Research in Stat. and Probab. (2010)

[DM11] Dellacherie, C., Meyer, P.A. ; Probabilities and potential, North Holland (2011)

[DV16] Del Moral, P., Villemonais, D. ; Exponential mixing properties for time inhomogeneous diffusion processes with killing, Bernouli, V.24, N.2, pp.1010-1032 (2018)

[DZ98] Dembo, A., Zeitouni, O. ; Large Deviation Techniques and Applications. Second Edition. Springer-Verlag, New York (1998).

[DiT13] Di Tella, P., On the Predictable Representation Property of Martingales Associated with Levy Processes, PhD thesis (2013)

[DM15] Diaconis, P., Miclo, L. ; On quantitative convergence to quasi-stationarity ; Ann. Fac. Sci. Toulouse Math., Sér. 6, V. 24, N. 4, pp. 973-1016 (2015)

[DJ+05] Diekmann, O., Jabin, P.E., Mischler, S., Perthame, B. ; The dynamics of adaptation : An illuminating example and a Hamilton-Jacobi approach. Theor. Pop. Biol., 67 :257–271 (2005)

[DV75] Donsker, M.D., Varadhan, S.R.S. ; Asymptotic evaluation of certain Markov process expectations for large time. I. Comm. Pure Appl. Math., V.28, pp. 1-47 (1975)

[DV83] Donsker, M.D., Varadhan, S.R.S. ; Asymptotic evaluation of certain Markov process expectations for large time. IV. Comm. Pure Appl. Math., V.36, N., pp.183–212 (1983)

[DP13] van Doorn, E.A., Pollett, P.K. ; Quasi-stationary distributions for discrete-state models, Eur. J. of Operational Research, V. 230, pp. 1-14 (2013)

[DMT95] Down, D., Meyn, S.P., Tweedie, R. L. ; Exponential and Uniform Ergodicity of Markov Processes, The Annals of Probab., V. 23, pp. 1671-1691 (1995)

[Du08] Durrett, R. ; Probability Models for DNA Sequence Evolution, 2nd ed. Springer (2008)

[E84] Ellis, R.S. ; Large deviations for a general class of random vectors. Ann. Probab., V.12, N.1, pp.1–12 (1984)

[Et00] Etheridge, A.. An introduction to superprocesses, University Lecture Series, v. 20. American Mathematical Society, Providence (2000)

[EPW09] Etheridge, A., Pfaffelhuber, P., Wakolbinger, A. ; How often does the ratchet click ? Facts, heuristics, asymptotics, in : Trends in Stochastic Analysis, in : London Math. Soc. Lecture Notes Series, Cambridge Univ. Press (2009)

[Ev98] C. Evans, L. ; Partial Differential Equations ; Graduate Studies in Mathematics, V. 19, Am. Math. Society (1998)

[Ew04] Ewens, W. ; Mathematical Population Genetics, I, 2nd ed., Interdiscip. Appl. Math. 27, Springer-Verlag, Berlin (2004)

[FK+95] Ferrari, P. A., Kesten, H., Martinez S., Picco, P. ; Existence of quasi-stationary distributions. A renewal dynamical approach ; Ann. of Probab., V.23, N.2, pp.501–521 (1995)

[FK+96] Ferrari, P. A., Kesten, H., Martinez S. ; R-positivity, quasi-stationary distributions and ratio limit theorems for a class of probabilistic automata ; Ann. Appl. Probab. V.6, Nbr. 2, 577-616 (1996)

[Fe20] Ferré, G. ; Large Deviations Theory in Statistical Physics : Some Theoretical and NumericalAspects. PhD Mathematics [math]. Université Marne La Vallée, (2019)

[FRS19] Ferré, G., Rousset, M., Stoltz, G. ; More on the long time stability of Feynman-Kac semigroups ; preprint on ArXiv : 1807.00390 (2019)

[FM04] Fournier, N., Méléard, S. ; A microscopic probabilistic description of a locally regulated population and macroscopic approximations. Ann. Appl. Probab., V.14, N.4, pp.1880-1919 (2004)

[Fr08] Friedman, A. ; Partial differential equations of parabolic type, Dover publications, New York (2008)

[Gä77] Gärtner, J. ; On large deviations from the invariant measure. Theor. Probab. Appl., V.22, N.1, pp.24–39, (1977)

[Gr06] Grafen, A. ; Optimisation of inclusive fitness ; J. Theor. Biol., V.238, pp.541–563 (2006)

[GR09] Garcia-Melian, J., Rossi, J.D. ; On the principal eigenvalue of some nonlocal diffusion problems, J. Diff. Equations, V.246 , pp.21-38 (2009)

[Ha78] Haigh, J. ; The accumulation of deleterious genes in a population — Muller's ratchet, Theoret. Population Biol., V.14, pp.251–267 (1978)

[HH09] Hardy, R., Harris, S. C. ; A spine approach to branching diffusions with applications to Lp-convergence of martingales. In Séminaire de Probabilités XLII. Lecture Notes in Math. 1979 pp.281–330. Springer, Berlin (2009)

[HZZ19] He G., Zhang H., Zhu Y. ; On the quasi-ergodic distribution of absorbing Markov processes Stat. and Probab. Letters, V.149 , pp. 116-123 (2019)

[Ho09] Hoepfner, R. : An extension of the Yamada-Watanabe condition for pathwise uniqueness to stochastic differential equations with jumps. Electron. Commun. Probab., V14, 43, 447–456 (2009)

[Ho99] Marc Hoffmann. Adaptive estimation in diffusion processes. Stochastic Process. Appl., V.79, N.1, pp.135-163 (1999)

[HM19] Hoffmann, M., Marguet, A. ; Statistical estimation in a randomly structured branching population, Stochastic Process. Appl., V.129, N.12, pp.5236-5277 (2019)

[IRS12] Ignat, L.I., Rossi, J.D., San Antolin, A. ; Lower and upper bounds for the first eigenvalue of nonlocal diffusion problems in the whole space, J. Diff. Equations, V.252, pp.6429-6447 (2012)

[IW81] Ikeda, N., Watanabe, S. : Stochastic Differential Equations and Diffusion Processes, V. 24, North-Holland Mathematical Library (1981)

[JS03] Jacod, J., N.Shiryaev, A. ; Limit Theorems for Stochastic Processes. Second Edition. Grundlehren der mathematischen Wissenschaft, V.288, Springer (2003)

[JS17] Jenkins, P. A., Spano, D : Exact simulation of the Wright-Fisher diffusion. Ann. Appl. Prob., V.27, N.3, pp.1478-1509 (2017)

[JT17] Jilani Ben Naouara, N.J.B., Trabelsi, F. ; Boundary classification and simulation of one-dimensional diffusion processes, Int. J. Mathematics in Operational Research, V.11, N.1 (2017)

[JM86] Joffe, A., Metivier, M. ; Weak convergence of sequences of semimartingales with applications to multitype branching processes. Advances in Appl. Probab., V.18, pp.20-65 (1986)

[Ka97] Kallenberg, O. ; Foundations of Modern Probability. Springer, 1st edition (1997)

[KP08] Keeling, P., Palmer, J. ; Horizontal gene transfer in eukaryotic evolution, Nat. Rev. Genet. V.9, pp. 605–618 (2008)

[KKT16] Kim, D., Kuwae, K., Tawara, Y. ; Large deviation principles for generalized Feynman-Kac functionals and its applications, Tohoku Math. J. (2) V. 68, N. 2, pp. 161-197 (2016)

[KA06] Kimmel, M., Axelrod, D. E. ; Branching Processes in Biology, Springer, New York, NY (2006)

[Ki82] Kingman, J.F.C. ; The coalescent, Stochastic Process. Appl. V.13, pp.235-248 (1982)

[KM03] Kontoyiannis, I., Meyn, S. P. ; Spectral theory and limit theorems for geometrically ergodic Markov processes. Ann. Appl. Probab., V.13, pp.304–362 (2003)

[KM05] Kontoyiannis, I., Meyn, S. P. ; Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. Electron. J. Probab., V.10, N.3, pp.61–123 (2005)

[Ku13] Kutoyants, Y. A. ; Statistical inference for ergodic diffusion processes. Springer Science and Business Media (2013)

[KH09] Kopp, M., Hermisson, J. ; The genetic basis of phenotypic adaptation II : The distribution of adaptative substitutions of the moving optimum model ; Genetics, V.183, pp. 1453-1476 (2009)

[KNP18] Kopp, M., Nassar, E., Pardoux, E. ; Phenotypic lag and population extinction in the moving-optimum model : insights from a small-jumps limit, J. Math. Biol., V.7, N.5, pp.1431-1458. (2018).

[KS80] Krylov, N. V., Safonov,M. V. ; A property of the solutions of parabolic equations with measurable coefficients. Izv. Akad. Nauk SSSR Ser. Mat., V.44, N.1, pp.161–175 (1980)

[KL+97] Kurtz, T., Lyons, R., Pemantle, R., Peres, Y. ; A Conceptual Proof of the Kesten-Stigum Theorem for Multi-Type Branching Processes. In : Athreya K.B., Jagers P. (eds) Classical and Modern Branching Processes. The IMA Volumes in Mathematics and its Applications, V.84. Springer, New York, NY (1997)

[La05] Lambert, A. ; The branching process with logistic growth ; Ann. Appl. Probab., V.15, N.2, pp.1506-1535 (2005)

[LK+07] Lehmann, L., Keller, L., West, S. A., Roze, D. ; Group selection and kin selection. Two concepts but one process. Proceed. of the National Acad. of Sciences of the U.S.A., V.104, pp.6736–6739 (2007)

[Lu95] Lunardi., A. ; Analytic semigroups and optimal regularity in parabolic problems, v. 16 of Progress in Nonlinear Diff. Eq. and their Appl. ; Birkhauser Verlag, Basel (1995)

[Luo13] Luo, S. ; A unifying framework reveals key properties of multilevel selection, J. Theor. Biol. V.341, pp.41-52 (2013)

[LM15] Luo, S., Mattingly, J. ; Scaling limits of a model for selection at two scales. Nonlinearity. V.30, N.4, pp.1682-1707 (2015)

[MV20] Mailler, C., Villemonais, D. ; Stochastic approximation on non-compact measure spaces and application to measure-valued Pólya processes. ArXiv :1809.01461

[Ma78] Maynard Smith, J. ; The Evolution of Sex, Cambridge University Press (1978)

[Al19] Marguet, A. ; Uniform sampling in a structured branching population, Bernoulli V.25, N.4A, pp.2649-2695 (2019)

[MSV14] Martínez, S., San Martin, J., Villemonais, D. ; Existence and uniqueness of a quasi-stationary distribution for Markov processes with fast return from infinity, J. of Appl. Probab., V.51, N. 3, pp. 756-768 (2014)

[Mé16] Méléard, S. ; Modèles aléatoires en Ecologie et Evolution, Math. et appl., V. 77, Springer, Berlin Heidelberg (2016)

[MT09] Méléard, S., Tran, V. ; Trait Substitution Sequence process and Canonical Equation for age-structured populations. J. Math. Biology, V.58, N.6, pp.881-921 (2009)

[MV12] Méléard, S., Villemonais, D. ; Quasi-stationary distributions and population processes, Probab. Surveys, V. 9, pp. 340-410 (2012)

[MG+96] Metz, J.A.J., Geritz, SAH, and al. ; Adaptative dynamics, a geometrical study of the consequences of nearly faithful reproduction. S.J. Van Strien and S.M. Verduyn Lunel (ed.), Stochastic and Spatial Structures of Dynamical Systems, V.45, pp.183–231 (1996)

[ME13] Metzger, J.J., Eule, S. ; Distribution of the Fittest Individuals and the Rate of Muller's Ratchet in a Model with Overlapping Generations. PLoS ComputBiol, V.9, N.11 (2013)

[MT93] Meyn, S.P., Tweedie, R. L. ; Markov Chains and Stochastic Stability, Commun. and Control Eng. Series, Springer, London (1993)

[MMP05] Michel, P., Mischler, S., Perthame, B. ; General relative entropy inequality : an illustration on growth models. J. Math. Pures Appl., V.84, N.9, pp.1235–1260 (2005)

[MB+12] Mirrahimi S., Barles G., Perthame B., Souganidis P. E. : Singular Hamilton-Jacobi equation for the tail problem. SIAM J. Math. Anal., V.44, N.6, pp.4297–4319 (2012)

[MJ10] Moran, M., Jarvik, T. ; Lateral transfer of genes from fungi underlies carotenoid production in aphids, Science V.328, pp.624–627 (2010)

[ML14] Mullon, C., Lehmann, L. ; The robustness of the weak selection approximation for the evolution of altruism against strong selection. J. Evol. Biol., V.27, pp.2272-2282 (2014)

[NP17] Nassar, E., Pardoux, E. ; On the long time behavior of the solution of an SDE driven by a Poisson Point Process ; Adv. in Appl. Probab., V.49, pp.344-367 (2017)

[NP19] Nassar E., Pardoux, E. ; Small jumps asymptotic of the moving optimum Poisonian SDE, Stoch. Proc. and Appl., V.129, N.7, pp.2320-2340. (2019)

[Nu84] Nummelin, E. ; General Irreducible Markov Chains and Nonnegative Operators. Cambridge University Press, Cambridge, (1984)

[O'F08] O'Fallon, B. ; Population Structure, Levels of Selection, and the Evolution of Intracellular Symbionts ; Evolution, V.62, N.2, pp.361-373 (2008)

[Oc18] Ocafrain, W. ; Quasi-stationarity and quasi-ergodicity for discrete-time Markov chains with absorbing boundaries moving periodically ALEA-Latin Amer. J. Probab. Math. Stat., V.15, N.1, pp.3445-3476 (2018)

[Oc19] Ocafrain, W. ; Q-processes and asymptotic properties of Markov processes conditioned not to hit moving boundaries, Stochastic Processes and their Applications, V.130, N.6, pp.3445-3476 (2019)

[Oc20] Ocafrain, W. ; Quasi-stationarity for one-dimensional renormalized Brownian motion accepted in Esaim : Probab. Stat. (2020)

[OLG00] Ochman, H., Lawrence, J., Groisman, E. ; Lateral gene transfer and the nature of bacterial innovation, Nature V.405, pp.299–304 (2000)

[OV17] Ocafrain, W., Villemonais, D. ; Convergence of a non-failable mean-field particle system, Stoch. Ana. Appl., V.35, N.4, pp.587-603 (2017)

[PG10] Perthame, B., Gauduchon, M. ; Survival thresholds and mortality rates in adaptive dynamics : conciliating deterministic and stochastic simulations. Math. Med. Biol., V.27, pp.195–210 (2010)

[Pa16] Pardoux, E ; Probabilistic Models of Population Evolution : Scaling Limits, Genealogies and Interactions ; Springer (2016)

[PR14] Pardoux, E., Răşcanu, A. ; Stochastic Differential Equations, Backward SDEs, Partial Differential Equations. Stochastic Modelling and Applied Probability, 69, Springer (2014)

[PP04] Pascucci, A., and Polidoro, S. ; On the Harnack inequality for a class of hypoelliptic evolution equations, Transactions of the American Mathematical Society ,V. 356, N.11, pp.4383-4394 (2004)

[PSW12] Pfaffelhuber, P., Staab, P.R., Wakolbinger, A. ; Muller's ratchet with compensatory mutations, Ann. Appl. Probab., V.22, N.5, pp.2108–2132 (2012)

[Po15] Pollett, P. K. ; Quasi-stationary distributions : A bibliography., available at people.smp.uq.edu.au/PhilipPollett/papers/qsds/qsds.html (2015)

[PF+16] Pollock, M., Fearnhead, P., Johansen, A.M., Roberts, G.O. ; The Scalable Langevin Exact Algorithm : Bayesian Inference for Big Data, preprint on ArXiv : 1609.03436 (2016)

[RW00] Rogers, L. C. G., Williams, D. ; Diffusions, Markov processes, and martingales ; V. 1 ; Cambridge Math. Library (2000)

[RG+12] Roques, L., Garnier, J., Hamel, F., Klein, E.K. ; Allee effect promotes diversity in traveling waves of colonization, PNAS June 5, 109 (23), 8828-8833 (2012)

[RVY06] Roynette, B., Vallois, P., Yor, M. ; Some penalisations of the Wiener measure ; Jpn. J. Math., V. 1, N. 1, pp. 263–290 (2006)

[SGE13] Schraiber, J., Griffiths, R. C., Evans, S. N. ; Analysis and rejection sampling of Wright-Fisher diffusion bridges.Theoretical Population Biology, V.89, pp.64–74 (2013)

[SV96] Seneta, E., Vere-Jones, D. ; On quasi stationary distributions in discrete-time Markov chains with a denumerable infinity of states ; J. Appl. Prob., V.3, pp.403-434 (1996)

[SB+14] Szamecz, B., Boross, G. and al. ; The Genomic Landscape of Compensatory Evolution PLoS Biology, V.12, N.8 :e1001935 (2014)

[Sz91] Sznitman, A.S. ; Topics in propagation of chaos. In Ecole d'Eté de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math. V.1464, pp 165–251 (1991).

[TH13] Touchette, H., Harris, R.J. ; Large deviation approach to nonequilibrium systems, in R. Klages, W. Just, C. Jarzynski (eds), Noneq. Stat. Physics of Small Systems : Fluct. Relat. and Beyond, Wiley-VCH, Weinheim (2013)

[TR+08] Tettelin, H., Riley, D., Cattuto, C., Medini, D. ; Comparative genomics : the bacterial pan-genome, Curr. Opin. Microbiol. V.11, pp.472–477 (2008)

[TN06] Traulsen, A., Nowak, M.A. : Evolution of cooperation by multilevel selection. Proceedings of the National Academy of Sciences, 103, 29 (2006)

[TCV11] Turitsyn, K.S., Chertkov, M., Vucelja, M. ; Irreversible Monte Carlo algorithms for efficient sampling. Physica D : Nonlinear Phenomena, V. 240, pp. 410–414 (2011)

[Tw74a] Tweedie, R.L. ; R-Theory for Markov Chains on a General State Space I : Solidarity Properties and R-Recurrent Chains, Ann. Probab., V. 2, N. 5,pp. 840-864 (1974)

[Tw74b] Tweedie, R.L. ; Quasi-Stationary Distributions for Markov Chains on a General State Space, J. of Appl. Probab., V. 11, N. 4, pp. 726-741 (1974)

[Vi11] Villemonais, D. ; Interacting particle systems and Yaglom limit approximation of diffusions with unbounded drift, Elec. J. Probab., V.16, pp.1663-1692 (2011)

[Vi13] Villemonais, D. ; Uniform tightness for time-inhomogeneous particle systems and for conditional distributions of time-inhomogeneous diffusion processes, Markov Proc. and Relat. Fields, V.19, N.3, pp.543-562 (2013)

[Vi14] Villemonais, D. ; General approximation method for the distribution of Markov processes conditioned not to be killed, Esaim : Probab. Stat., V.18, pp.441-467 (2014)

[WD19] Wang, Z., Durrett, R. J. ; Extrapolating Weak Selection in Evolutionary Games, Math. Biol. V.78, N.1-2, pp.135-154 (2019)

[WGG07] West, S.A., Griffin, A.S., Gardner, A. ; Social semantics : altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. V.20, pp.415–432 (2007)

[WGG07b] West, S.A., Griffin, A.S., Gardner, A. ; Social semantics : how useful has group selection been ? J. Evol. Biol. V.21, pp.374–385 (2007)

[WG+06] West, SA., Griffin, AS., Gardner A., Diggle SP. Social evolution theory for microorganisms. Nat. Rev. Microbiol. V.4, pp.597–607 (2006)

[WMG11] West, S.A., El Mouden, C., Gardner, A. ; Sixteen common misconceptions about the evolution of cooperation in humans, Evol. and Human Behavior V.32, N.4, pp.231-262 (2011)

[WG$^+$13] Wu, B., Garcia, J., Hauert, C., Traulsen, A. ; Extrapolating Weak Selection in Evolutionary Games, PLoS Comput. Biol., V.9, N.12, pp.1-7 (2013)

[YW71] Yamada, T., Watanabe, S. ; On the uniqueness of solutions of stochastic differential equations. J. Math. Kyoto Univ. V.11, pp.155–167 (1971)

Personal preprints

[Chapter 1] Velleret, A. ; Unique quasi-stationary distribution, with a possibly stabilizing extinction ; ArXiv :1802.02409 (2018)

[Chapter 2] Velleret, A. ; Exponential quasi-ergodicity for processes with discontinuous trajectories, preprint on ArXiv : 1902.01441 (2019)

[Chapter 3] Velleret, A. ; Adaptation of a population to a changing environment under the light of quasi-stationarity ; preprint on ArXiv : 1903.10165 (2019)

[Chapter 4] Velleret, A. ; Two level natural selection with a quasi-stationarity approach ; preprint on ArXiv : 1903.10161 (2019)

[Chapter 6] Velleret, A. ; Individual-based models under various time-scales ; ESAIM :Proceedings and Surveys, V.68, 123-152 (2020)

[Ve20] Velleret, A. ; Large deviation of surviving trajectories of general Markov processes, now under corrections, ArXiv :2003.01026 (2020)