



**HAL**  
open science

## Une ballade en forêts aléatoires

Viet Chi Tran

► **To cite this version:**

Viet Chi Tran. Une ballade en forêts aléatoires. Probabilités [math.PR]. Université Lille 1, 2014.  
tel-01087229

**HAL Id: tel-01087229**

**<https://theses.hal.science/tel-01087229v1>**

Submitted on 25 Nov 2014

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

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

Document de synthèse

présenté par

**TRAN Viet Chi**

en vue de l'obtention de

**l'Habilitation à diriger des recherches**

**Spécialité : Mathématiques appliquées**

**Université des Sciences et Technologies Lille 1**

Ecole doctorale Sciences pour l'Ingénieur (SPI)

**Une ballade en forêts aléatoires**

*Théorèmes limites pour des populations structurées et leurs généalogies, étude probabiliste et statistique de modèles SIR en épidémiologie, contributions à la géométrie aléatoire*

Date de soutenance : 24 novembre 2014

Rapporteurs :

Thomas G KURTZ    Professeur (University of Wisconsin-Madison)  
Etienne PARDOUX    Professeur (Université de Provence)  
Anita WINTER    Professeur (Universität Duisburg-Essen)

Jury de la soutenance :

Antoine AYACHE    Professeur (Université Lille 1)  
Jean-Stéphane DHERSIN    Professeur (Université Paris 13)  
Jean-François MARCKERT    Directeur de recherches (Université Bordeaux 1)  
Sylvie MÉLÉARD    Professeur (Ecole Polytechnique)  
Etienne PARDOUX    Professeur (Université de Provence)  
Vincent RIVOIRARD    Professeur (Université Paris Dauphine)



# Table des matières

<b>Introduction</b>	<b>9</b>
<b>1 Populations structurées par âge et généalogies</b>	<b>13</b>
1.1 Populations structurées en âge . . . . .	14
1.1.1 Construction du processus individu-centré . . . . .	15
1.1.2 Critères de tension . . . . .	17
1.1.3 Approximation par l'EDP de McKendrick-Von Foerster . . . . .	20
1.1.4 Approximation par une diffusion de Feller avec coefficients moyennés . . . . .	24
1.2 Evolution de généalogies . . . . .	30
1.2.1 Processus historiques . . . . .	30
1.2.2 Trajectoire ancestrale d'un individu tiré au hasard . . . . .	41
1.3 Autres modèles de populations structurées . . . . .	46
1.3.1 Evolution de la diversité génétique d'un marqueur lié à un trait sous sélection . . . . .	46
1.3.2 Plantes à fleurs avec reproduction auto-incompatible . . . . .	48
<b>2 Modèles SIR stochastiques en épidémiologie</b>	<b>57</b>
2.1 Estimation pour des modèles SIR compartimentaux . . . . .	58
2.1.1 Modèle individu-centré et EDO de Kermack-McKendrick . . . . .	58
2.1.2 Maximum de vraisemblance avec données complètes . . . . .	60
2.1.3 Estimation par ABC en présence de données manquantes . . . . .	61
2.2 Modèles SIR évoluant sur un graphe . . . . .	66
2.2.1 Epidémies se propageant sur un modèle de configuration et équations de Volz . . . . .	66
2.2.2 Prise en compte de la structure de graphe dans l'épidémie cubaine . . . . .	73
2.3 Conclusion . . . . .	75
<b>3 Deux problèmes de géométrie aléatoire</b>	<b>77</b>
3.1 Espérances d'ensembles aléatoires et moyennes de Vorob'ev associées . . . . .	78
3.1.1 Espérance de Vorob'ev . . . . .	78
3.1.2 Estimateur des ensembles de niveau . . . . .	79
3.1.3 Estimation de l'espérance de Vorob'ev . . . . .	81
3.2 Arbre couvrant radial sur un processus ponctuel de Poisson . . . . .	82
3.2.1 Forêt dirigée . . . . .	84
3.2.2 Arbre radial . . . . .	88
3.3 Conclusion . . . . .	94
<b>Bibliographie</b>	<b>97</b>

## Remerciements

La balade à laquelle je vous invite nous emmène dans des contrées que j'ai explorées au cours des dernières années grâce à de merveilleux compagnons et guides, que je souhaite remercier ici.

Tout d'abord, je souhaite remercier du fond du cœur Sylvie Méléard qui a guidé mes tous premiers pas en recherche et avec qui nous avons tant fait depuis ! Sylvie, je suis très heureux que nous ayons pu travailler ensemble sur autant de sujets passionnants. Ton plaisir de faire des mathématiques est communicatif. Ton énergie, ta curiosité, ton engagement sont pour moi une source de motivation et d'inspiration inégalée.

Tom Kurtz, Etienne Pardoux et Anita Winter m'ont fait l'honneur de rapporter mon travail, et je les en remercie infiniment. En plus d'être des figures charismatiques des probabilités et des personnalités enthousiasmantes, vos travaux et les discussions que nous avons eues ont eu des impacts directs sur mes recherches. Merci pour votre expertise. Etienne (et Sylvie), merci de m'avoir ouvert les portes de l'Afrique !

Merci à Antoine Ayache, Jean-Stéphane Dhersin, Jean-François Marckert et Vincent Rivoirard pour avoir accepté de participer à mon jury d'HDR. Antoine, Jess et Vincent, travailler avec vous a été une grande chance, et vous avez toujours été des aînés attentifs. Jean-François, merci pour l'intérêt que tu portes à nos travaux encore tous récents en géométrie aléatoire !

Je souhaite également remercier l'ensemble de mes co-auteurs : Diala Abu Awad, François Baccelli, Vincent Bansaye, Sylvain Billiard, Michael G. Blum, Stéphane Cléménçon, Marc Constant, David Coupier, Anthony Cousien, Miraine Davilà, Hector De Arazoza, Laurent Decreusefond, Jean-François Delmas, Sylvie Deuffic-Burban, Jean-Stéphane Dhersin, Régis Ferrière, Ankit Gupta, Philippe Heinrich, Marie Jauffret-Roustide, Pauline Lafitte-Godillon, Laurence Marsalle, Sylvie Méléard, Hans Metz, Pascal Moyal, Kilian Raschel, Fabrice Rossi, Radu S. Stoica, Francis Vasseur, Yazdan Yazdanpanah. Pour moi, l'aventure mathématique s'est toujours doublée d'une aventure humaine et sans vous, ce travail n'aurait pas du tout la même saveur ! On retrouve de vous un peu partout dans ce mémoire.

Bakari Traoré, Diala Abu Awad, Anthony Cousien, Hoang Van Hà, vous êtes les premiers doctorants que j'ai co-encadrés. Merci de m'avoir fait confiance. Je suis vraiment très fier de vous et vous souhaite des soutenance de thèse prochaines et brillantes ! Merci aussi à mes co-encadrants, Jean-Stéphane Dhersin, Sylvain Billiard, Yazdan Yazdanpanah, Vincent Rivoirard et Thanh Mai Pham Ngoc.

Depuis 2007, j'ai la chance de travailler au laboratoire Paul Painlevé à Lille 1, où j'ai pu bénéficier d'un cadre de travail convivial et exceptionnel. Je souhaiterais remercier l'ensemble des collègues du LPP et de l'UFR de mathématiques, en particulier : les membres de l'équipe proba-stat, l'équipe pédagogique du master ingénierie mathématique et l'ensemble de notre personnel BIATSS.

A mon arrivée à Lille s'est créé un groupe de travail autour de la géométrie aléatoire, sous l'impulsion de Youri Davydov. Ce groupe a été un moteur incroyable dont sont nées des collaborations, des thèses et de nombreux projets, comme le GdR GeoSto (GdR 3477). Merci à mes collègues du GT et merci également à Pierre Calka pour porter le GdR GeoSto, que nous avons monté, lui, David Coupier et moi.

Trois autres projets ont également été décisifs pour mon travail : la Chaire MMB (portée par Sylvie) et les ANR MAEV et MANEGE (ANR-09-BLAN-0215) portées par Etienne et Sylvie successivement. Les rencontres organisées ont été déterminantes dans la maturation de mon travail à l'interface entre mathématiques et biologie. Les collaborations avec plusieurs de mes co-auteurs ont commencé par des discussions en marge des exposés de ces journées, toujours extrêmement enrichissantes. Je mesure la chance d'avoir été, en tant que jeune chercheur, dans un cadre aussi stimulant ! Grâce à la Chaire MMB, j'ai également pu bénéficier d'un poste de chercheur associé à l'Ecole Polytechnique de 2009 à 2012. J'en profite pour remercier mes collègues du CMAP et Nasséra pour leur accueil.

Mes recherches ont aussi bénéficié d'autres financements tels que ceux du Labex CEMPI (ANR-11-LABX-0007-01, porté par Stephan De Bièvre) ou de l'ANR Viroscopy (ANR-08-SYSC-016-03, portée par Stéphane Cléménçon). Les données cubaines sur lesquelles nous avons travaillées ont été mises disposition par le Dr. J. Perez du National Institute of Tropical Diseases à Cuba.

J'ai également une pensée pour les bénévoles d'Opération Postes, en particulier les anciens que j'ai côtoyés entre 2008 et 2012, et pour Maria Esteban et Alain Prignet, de la SMAI, qui nous ont défendus et protégés en 2012.

Mes remerciements finaux, vont à ma famille et mes amis, chers à mon cœur ! J'envoie un clin d'oeil aux mélomanes qui auront remarqué les deux 'ℓ' du titre... Petite Mai Phuong, cette thèse t'est dédiée ♡

## Liste des travaux

### Journaux internationaux avec comités de lecture

- (1) V.C. Tran, A wavelet particle method for McKean-Vlasov and Navier-Stokes statistical solutions, **Stochastic Processes and their Applications**, vol. 118/2, pp. 284-318 (2008).
- (2) V.C. Tran Large population limit and time behaviour of a stochastic particle model describing an age-structured population, **ESAIM : P&S** vol. 12, pp. 345-386 (2008).
- (3) S. Cléménçon, V.C. Tran, H. De Arazoza, A stochastic SIR model with contact-tracing : large population limits and statistical inference, **Journal of Biological Dynamics**, vol. 2/4, pp. 392-414 (2008).
- (4) S. Méléard, V.C. Tran, Trait Substitution Sequence process and Canonical Equation for age-structured populations, **Journal of Mathematical Biology**, vol. 58/6, pp. 881-921 (2009).
- (5) M.G.B. Blum, V.C. Tran, HIV with contact-tracing : a case study in Approximate Bayesian Computation, **Biostatistics**, vol. 11(4), pp. 644-660 (2010).
- (6) V. Bansaye, V.C. Tran, Branching Feller diffusion for cell division with parasite infection, **ALEA**, Vol. 8, 95-127 (2011).
- (7) V. Bansaye, J.-F. Delmas, L. Marsalle, V.C. Tran, Limit theorems for Markov processes indexed by continuous time Galton-Watson trees, **Annals of Applied Probability**, Vol. 21, No. 6, 2263-2314 (2011).
- (8) S. Méléard, V.C. Tran, Slow and fast scales for superprocess limits of age-structured populations, **Stochastic Processes and their Applications**, Vol. 122, No. 1, 250-276 (2012).
- (9) S. Billiard, V.C. Tran, A general stochastic model for sporophytic self-incompatibility, **Journal of Mathematical Biology**, Vol. 64, No. 1-2, 163-210 (2012).
- (10) L. Decreusefond, J.-S. Dhersin, P. Moyal, V.C. Tran, Large graph limit for a SIR process in random network with heterogeneous connectivity, **Annals of Applied Probability**, Vol. 22, No. 2, 541-575 (2012).
- (11) S. Méléard, V.C. Tran, Nonlinear historical superprocess approximations for population models with past dependence, **Electronic Journal of Probability**, Vol. 17, No. 47, 1-32 (2012).
- (12) Ph. Heinrich, R.S. Stoica, V.C. Tran, Level sets estimation and Vorob'ev expectation of random compact sets, **Spatial Statistics**, Vol. 2, 47-61 (2012).
- (13) M. Constant, V.C. Tran, B. Benoit, F. Vasseur, New First-trimester crown-rump length's equations from a French general population, **Fetal Diagnosis and Therapy**, Vol. 32, No. 4, 277-287 (2012).
- (14) D. Coupier, V.C. Tran, The 2D-Directed Spanning Forest is a tree, **Random Structures and Algorithms**, Vol. 42, No. 1, 59-72 (2013).
- (15) J.A.J. Metz, V.C. Tran, Daphnias : from the stochastic individual based model to the large population equation, **Journal of Mathematical Biology**, special Issue in honor of Odo Diekmann, Vol. 66, No. 4-5, 915-933 (2013).
- (16) P. Lafitte-Godillon, K. Raschel, V.C. Tran, Extinction probabilities for a distylous plant population modeled by an inhomogeneous random walk on the positive quadrant, **SIAM Journal on Applied Mathematics** (SIAP), Vol. 73, No. 2, 700-722 (2013).
- (17) F. Baccelli, D. Coupier, V.C. Tran, Semi-infinite paths of the 2d-Radial Spanning Tree, **Advances in Applied Probability**, Vol. 45, No. 4, 895-1201 (2013).
- (18) A. Gupta, J.A.J. Metz, V.C. Tran, A new proof for the convergence of an individual based model to the trait substitution sequence, **Acta Applicandae Mathematicae**, Vol. 131, No. 1, 1-27 (2014).
- (19) A. Cousien, M. Jauffret-Roustide, J.-S. Dhersin, S. Deuffic-Burban, Y. Yazdanpanah, Dynamic modelling of HCV transmission among people who inject drugs : a methodological review, **Journal of Viral Hepatitis**, (2014).
- (20) S. Billiard, R. Ferrière, S. Méléard, V.C. Tran, Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks, **Journal of Mathematical Biology**.
- (21) S. Cléménçon, A. Cousien, M. Davilà, V.C. Tran, On computer-intensive simulation and estimation methods for rare event analysis in epidemic models, soumis.

- (22) S. Cléménçon, H. De Arazoza, F. Rossi, V.C. Tran, A network analysis of the HIV/AIDS epidemics in Cuba, soumis.
- (23) D. Abu Awad, S. Billiard, V.C. Tran, The effect of the timing of selection on the mutation load, inbreeding depression and population size, soumis.

#### Actes de conférences

- (24) R. Ferrière, V.C. Tran, Stochastic and deterministic models for age-structured populations with genetically variable traits, **ESAIM : Proceedings**, actes du CANUM 2008, vol. 27, pp 289-310 (2009).
- (25) S. Cléménçon, H. De Arazoza, F. Rossi, V.C. Tran, Hierarchical clustering for graph visualization, **Proceedings of XVIIIth European Symposium on Artificial Neural Networks (ESANN 2011)**, 227-232, Bruges, Belgium (April 2011).
- (26) S. Cléménçon, H. De Arazoza, F. Rossi, V.C. Tran, Visual Mining of Epidemic Networks, in Joan Cabestany, Ignacio Rojas, and Gonzalo Joya, editors, **Advances in Computational Intelligence**, Lecture notes in Computer Science, J. Cabestany I. Rojas and G. Joya eds., Springer Berlin/Heidelberg, vol. 6692, pp 276-283 (2011).
- (27) S. Méléard, J.A.J. Metz, V.C. Tran, Limiting Feller diffusions for logistic populations with age-structure, **Proceedings of the 58th World Statistics Congress of the International Statistical Institute (ISI 2011)**, Dublin (August 2011).

#### Autres travaux

- (28) V.C. Tran, Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques, Thèse de doctorat, Université Paris 10, 302 pages (2006). Directrice : S. Méléard, Rapporteurs : D.A. Dawson, D. Talay. Jury : S. Cléménçon, R. Ferrière, A. Lambert, C. Léonard, S. Méléard, D. Talay.
- (29) F. Alabau, G. Allaire, A. de Bouard, S. Cordier, E. Godlewski, L. Halpern, Ph. Helluy, J. Jacques, P. Lafitte, F. Lagoutière, A. Lejay, F. Murat, M. Postel, E. de Rocquigny, V.C. Tran, Analyse des résultats du sondage organisé par la Société de Mathématiques Appliquées et Industrielles (SMAI) auprès de ses adhérents du 21 au 31 octobre 2012.

**Notations :**

Dans l'ensemble de ce document, on considérera un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .

On note  $\mathcal{C}(E, \mathbb{R})$  (resp.  $\mathcal{C}_b(E, \mathbb{R})$ ,  $\mathcal{C}_0(E, \mathbb{R})$ ,  $\mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$ ,  $\mathcal{C}_K(E, \mathbb{R})$ ) l'espace des fonctions réelles continues sur  $E$  (resp. continues bornées, continues tendant vers 0 à l'infini, bornées et de classe  $\mathcal{C}^1$  sur  $\mathbb{R}^d$  à dérivées bornées, continues à support compact dans  $E$ ). Pour une fonction  $f \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$ , on notera

$$\|f\|_{\mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})} = \|f\|_\infty + \sum_{j=1}^d \|\partial_j f\|_\infty.$$

On note  $\mathbb{D}(I, E)$ , pour  $I \subset \mathbb{R}_+$  l'ensemble des fonctions càdlàg de  $\mathbb{R}_+$  dans  $E$ . Cet espace est muni de la topologie de Skorokhod. Cette topologie est métrisable et on notera par  $\mathbf{d}_{\text{Sk}}$  la distance de Skorokhod (cf. [Bil68, JM86] pour une présentation). Lorsque  $I = \mathbb{R}_+$ , on notera parfois  $\mathbb{D}_E$  pour alléger les notations.

Pour  $E$  un espace polonais, on notera  $\mathcal{M}_F(E)$  l'espace des mesures finies sur  $E$ . On pourra munir cet espace de la topologie de la convergence étroite ou de la topologie de la convergence vague (cf. [Kal83]). Lorsque la précision sera nécessaire, on notera  $(\mathcal{M}_F(E), \dot{\epsilon})$  et  $(\mathcal{M}_F(E), v)$  pour signifier que  $\mathcal{M}_F(E)$  est muni de la topologie de la convergence étroite ou vague. On notera  $\mathcal{P}(E)$  l'ensemble des mesures de probabilité sur  $E$ . Pour une mesure  $\mu$  sur  $E$  et  $f$  une fonction réelle sur  $E$  dont l'intégrale par  $\mu$  est bien définie, on note  $\langle \mu, f \rangle = \int_E f \mu$ . Pour une fonction  $f(t, x) \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , on notera parfois  $f_t(x) = f(t, x)$  et  $\langle \mu, f_t \rangle = \int_E f(t, x) \mu(dx)$ .





# Introduction

Au cours de mes travaux, je me suis principalement intéressé à des modèles aléatoires individu-centrés pour la dynamique des populations et l'épidémiologie, ainsi qu'à des problèmes issus de la géométrie aléatoire. Ce document se structure en trois parties relatives à chacun de ces domaines. Je retrace ici un bref historique et présente les idées principales qui ont guidé ma recherche. Une exposition plus détaillée et plus rigoureuse sera donnée dans le texte.

## **Modèles probabilistes et leurs limites en dynamique des populations et en évolution**

C'est grâce à Sylvie Méléard que j'ai découvert en 2003 des modèles aléatoires de systèmes vivants individu-centrés et en interaction, sujet qui m'a tout de suite plu. Dès le début de ma thèse, Sylvie Méléard m'a fait rencontrer Régis Ferrière qui m'a proposé de travailler sur les structures d'âges. Le but est d'enrichir les dynamiques individuelles en soulignant la dimension individuelle et en permettant de prendre en compte des histoires de vie - le fait que les individus peuvent avoir différents comportements au cours de leur vie - et ainsi modéliser des phénomènes tels que la maturation, la sénescence, la sélection de parentèle ou leurs évolutions (e.g. [Cha94]).

Nous considérons des processus à valeurs mesures aléatoires et non linéaires pour décrire des populations d'individus pouvant se reproduire, entrer en compétition ou mourir, à des taux dépendant de leurs âges et des traits qui les caractérisent. De tels processus, sans compétition ou avec des interactions non logistiques, avaient déjà fait l'objet d'une abondante littérature (e.g. [AN70, Jag75, Oel90]). En grandes populations, les processus à valeurs mesures correspondants peuvent être approchés par des solutions d'EDP du type McKendrick-Von Foerster.

Dans une série de plusieurs articles de ma thèse [MT09, FT09], en considérant des accélérations du temps combinées avec une structure de traits héréditaires et avec mutations rares, on obtient à la limite des processus de sauts généralisant, pour des populations structurées par âges, le *Trait Substitution Sequence* introduit par Metz et al. [MGM<sup>+</sup>96] et dont la justification mathématique a ensuite été établie par Champagnat et al. [Cha03, CFM06b] (voir aussi [GMT14]). Ces résultats pour les populations avec structures d'âges sont à mettre en regard des articles de Durinx *et al.* [DMM08] et Metz [Met11] pour des populations "structurées générales".

Après ma thèse nous avons cherché des limites diffusives de ces objets. En accélérant le temps, mais sans hypothèse sur les mutations des traits, la composante rapide, l'âge, se stabilise dans une mesure d'équilibre dépendant des traits et liée aux EDP mentionnées plus haut. Nous utilisons pour cela des méthodes de martingales introduites par Kurtz [Kur92] et répondons ainsi à des questions étudiées par Dynkin [Dyn91], Dawson et al. [DGL02], Kaj et ses co-auteurs [BK95, KS98], Athreya et al. [AAI11] entre autres, mais non résolues dans les cas qui nous intéressent : avec de la compétition de type logistique et des taux dépendant des âges et traits.

Ensuite, nous avons construit des processus de naissances et morts historiques permettant de généraliser la prise en compte du passé, et en avons étudié les limites diffusives. Ces processus décrivent l'évolution des généalogies des individus vivants. Nous adaptons des méthodes introduites par Dawson et Perkins [DP91, Per95] pour prendre en compte les non-linéarités et la dépendance des paramètres par rapport au passé. La loi de la généalogie d'un individu choisi au hasard dans la population est, lorsqu'il y a des interactions, difficile à obtenir. L'étude dans le cas sans in-

teraction correspond au travail réalisé avec Vincent Bansaye, Jean-François Delmas et Laurence Marsalle. Nous avons établi des représentations *many-to-one* pour des processus de Markov indexés par des Galton-Watson en temps continu. Le processus auxiliaire exhibé dans ces représentations correspond à la généalogie d'un individu choisi uniformément dans la population à un instant donné, et fait apparaître des phénomènes de biais. Nous donnons des applications à des modèles multi-niveaux.

Dans des modèles plus généraux, avec Sylvain Billiard, Régis Ferrière et Sylvie Méléard, nous avons considéré les effets de la démographie, dirigée par les traits, sur la génération et le maintien de la diversité neutre, faisant apparaître des Fleming-Viot ou des Wright-Fisher des avec goulots d'étranglement. Ceci nous apporte un éclairage nouveau sur les généalogies et les phylogénies que l'on peut reconstruire à partir du polymorphisme génétique au temps présent.

J'ai également travaillé sur l'introduction de démographie et d'aléa dans des modèles de génétique des populations. Avec Sylvain Billiard, nous avons considéré des systèmes de reproduction auto-incompatibles chez les plantes à fleurs. Dans le cas d'espèces distyles, nous avons cherché des critères pour classer les processus comme sous ou sur-critiques. Pour des petites populations, nous avons été amenés, avec Pauline Lafitte et Kilian Raschel, à étudier des problèmes difficiles de marches aléatoires non homogènes sur le quart de plan positif, absorbées sur le bord.

**Modèles mathématiques et statistiques en épidémiologie** Avec Stéphan Cléménçon et Hector De Arazoza, nous avons travaillé sur les données collectées depuis 1986 à Cuba par le système de *contact-tracing* pour le SIDA. Nous avons proposé et estimé un premier modèle compartimental, en utilisant les observations des 6 premières années, supposées complètes. Avec Michael Blum, grâce à une méthode Bayésienne, l'ABC, mieux adaptée au format et à la taille de nos données, nous avons pu prendre en compte le caractère incomplet des observations entre 1986 et 2006 afin de faire des prévisions et estimer le nombre de séropositifs non dépistés. Avec Stéphan Cléménçon, Hector De Arazoza et Fabrice Rossi, nous avons exploité par des méthodes de *clustering* fondées sur la maximisation de modularité, les informations sur les graphes d'infection et de détection présentes dans la base cubaine (5389 noeuds reliés par 4073 arêtes, avec une composante géante de 2386 noeuds et 3168 arêtes, ce qui est unique dans la littérature sur le SIDA).

D'un point de vue plus probabiliste, nous avons étudié avec Laurent Decreusefond, Jean-Stéphane Dhersin et Pascal Moyal des problèmes limites liés aux modèles d'épidémies se propageant sur des graphes aléatoires de distributions de degrés données (cf. [BN08] pour l'un des premiers travaux dans cette direction). C'est à un workshop du DIMACS organisé par Todd Parsons que Lindi Wahl m'a parlé de l'article de Volz [Vol08] qui a pu résumer l'évolution de telles épidémies par un système de 5 EDO. Il m'apparaissait clairement que les objets sous-jacents dans le raisonnement de Volz étaient des distributions de degrés. A nouveau, les processus à valeurs mesures n'étaient pas loin! Nous avons ainsi donné une justification rigoureuse des EDO de Volz, qui ont eu par la suite un écho important dans le domaine. Notre théorème limite est valable lorsque la population est de grande taille avec une proportion d'infectieux strictement positive, et lorsque les degrés des individus admettent des moments d'ordre 5. Nous complétons ce résultat par la considération du début de l'épidémie. On peut alors montrer que, partant de distributions de degrés n'admettant que des moments d'ordre 2, on obtient des moments exponentiels lorsque l'on arrive dans le cadre de validité du théorème limite précédent. Des couplages habituels avec des processus de branchement linéaires nous permettent également de calculer le nombre de reproduction  $R_0$ . Le lien entre ces modèles théoriques et les données cubaines est finalement abordé.

**Contribution à deux problèmes issus de la Géométrie aléatoire** Depuis 2007, je participe au groupe de travail "Géométrie aléatoire" organisé à Lille 1. C'est dans ce cadre que nous avons été amenés, avec Radu Stoica et Philippe Heinrich à étudier les espérances d'ensembles aléatoires. Pour les applications qui intéressent Radu Stoica, nous nous sommes penchés sur l'espérance de Vorob'ev (e.g. [Mol05] pour une présentation). Nous améliorons l'estimateur proposé par Kovyazin [Kov86] en ajoutant une discrétisation spatiale pour le rendre implémentable numériquement. Les

propriétés statistiques de ces estimateurs nous amènent à étudier la littérature sur l'estimation des ensembles de niveaux et nous nous inspirons des articles de Cuevas [CF97, CGMRC06] en particulier.

Avec François Baccelli, David Coupier, Kumarjit Saha et Anish Sarkar, nous nous intéressons à un arbre radial, enraciné en l'origine  $O$ , introduit par Baccelli et Bordenave [BB07] : le RST (*Radial Spanning Tree*). Ce graphe géométrique et planaire, construit sur un processus de Poisson homogène dans  $\mathbb{R}^2$ , exhibe des structures de dépendances spatiales complexes qui empêchent l'utilisation des techniques classiques fondées sur les martingales, par exemple. L'étude de ses propriétés nécessite alors l'utilisation d'outils très variés.

Pour étudier les géodésiques de cet arbre, nous utilisons sa propriété d'être *straight*, au sens de Howard et Newman [HN01]. Pour les arbres ayant cette propriété, ces auteurs obtiennent des résultats fins sur les directions asymptotiques des géodésiques : dans chaque direction  $\theta \in [0, 2\pi)$ , il existe au moins une géodésique, et l'ensemble des  $\theta$  qui sont directions asymptotiques pour au moins 2 géodésiques est dense dans  $[0, 2\pi)$ . Pour le RST, nous avons affiné ce résultat en montrant qu'il n'existe dans une direction  $\theta$  fixée qu'une unique géodésique presque sûrement et que l'ensemble des directions admettant plusieurs géodésiques est aléatoire. En utilisant le résultat de Howard et Newman et des couplages de géodésiques, on peut étudier les interfaces du RST, avec une jolie application à un problème de coloriage des sous-arbres du RST.

L'étude du nombre de géodésiques que l'on rencontre à un rayon  $r$  donné croît sous-linéairement en  $r$ , en espérance. Pour montrer cela, nous utilisons un couplage local, loin de  $O$ , du RST avec une forêt dirigée, la DSF (*Directed Spanning Forest*), également introduite dans [BB07]. Nous montrons que la DSF est en réalité un arbre, en utilisant un argument de percolation de Burton et Keane [BK89], ce qui nous permet de conclure. Enfin, mentionnons que la DSF a des liens avec la toile Brownienne [Arr79, TW98, FINR04] que nous sommes actuellement en train d'explorer.



# Chapitre 1

## Populations structurées par âge et généalogies

### Sommaire

---

<b>1.1 Populations structurées en âge</b> . . . . .	<b>14</b>
1.1.1 Construction du processus individu-centré . . . . .	15
1.1.2 Critères de tension . . . . .	17
1.1.3 Approximation par l'EDP de McKendrick-Von Foerster . . . . .	20
1.1.4 Approximation par une diffusion de Feller avec coefficients moyennés . . . . .	24
<b>1.2 Evolution de généalogies</b> . . . . .	<b>30</b>
1.2.1 Processus historiques . . . . .	30
1.2.2 Trajectoire ancestrale d'un individu tiré au hasard . . . . .	41
<b>1.3 Autres modèles de populations structurées</b> . . . . .	<b>46</b>
1.3.1 Evolution de la diversité génétique d'un marqueur lié à un trait sous sélection . . . . .	46
1.3.2 Plantes à fleurs avec reproduction auto-incompatible . . . . .	48

---

La modélisation de populations structurées, dont les individus sont caractérisés par différentes covariables qui influent sur leurs comportements ou sur leurs capacités reproductives ou de survie, a toujours été importante en écologie, en théorie de l'évolution ou en épidémiologie. Dans ce chapitre, nous allons considérer des modèles en temps continu, aléatoires et individu-centrés - fondés sur les dynamiques individuelles "microscopiques" - pour décrire l'évolution de telles populations, sous l'effet des morts, des naissances (la reproduction considérée ici sera asexuée) et du vieillissement. Nous étudierons les différentes dynamiques macroscopiques émergeant des hypothèses d'échelles qui seront faites.

Les variables décrivant les individus sont les suivantes. Un premier type de variables, les traits, correspond à des caractères morphologiques, physiologiques, génétiques ou comportementaux qui sont supposés être transmis héréditairement en l'absence de mutation (on peut penser à des caractères tels que la taille, la capacité à tirer des ressources du milieu, la position spatiale pour des plantes...). D'autres caractères peuvent varier dans le temps, par exemple de façon croissante et avec remise à une certaine valeur initiale à chaque naissance (taille, âge par exemple), ou de façon plus stochastique avec partage entre le parent et le descendant (parasites contenus dans une cellule et qui se répartissent entre mère et fille, au moment de la division cellulaire par exemple). D'autres aspects, tels que la spatialité, la reproduction sexuée, etc. sont explorés dans la littérature, mais nous ne les considérerons pas ici.

Dans une première section, nous portons notre attention sur des populations structurées en traits et en âges. Les structures d'âges permettent d'enrichir les dynamiques individuelles en soulignant la dimension individuelle, et rendent possible la prise en compte des histoires de vie et ainsi la modélisation de phénomènes tels que la maturation, la sénescence, la sélection de parentèle... (e.g. [Cha94]). Une particularité des variables d'âge est qu'elles sont fonctions croissantes du temps et sont "remises à zéro" à chaque naissance. De plus, outre le caractère non-markovien des processus introduits, une autre difficulté est que l'histoire particulière des individus s'inscrit dans une échelle de temps *a priori* différente de celle décrivant l'évolution de la population. En grandes populations, il est possible d'approcher le processus aléatoire par la solution d'une équation aux dérivées partielles (EDP) avec bord, du type McKendrick-Von Foerster [McK26, Foe59] - Gurtin MacCamy [GM74]. Lorsque l'on regarde des dynamiques diffusives, des phénomènes d'homogénéisation apparaissent, que nous traiterons avec des techniques dues à Kurtz [Kur92].

Dans une seconde section, nous considérons un modèle plus général où la dynamique d'un individu dépend de sa lignée ancestrale. Nous introduisons le processus historique lié au processus de la Section 1.1 et en considérons sa limite diffusive. Dans le cas particulier où il n'y a pas d'interaction entre les individus, la loi de la lignée ancestrale d'un individu tiré au hasard est caractérisée. Il apparaît un phénomène de biais par la taille, que nous décrirons et qui nous permet d'obtenir des théorèmes limites nous permettant, par exemple, de répondre à certaines questions concernant des populations hôtes-parasites.

Dans une dernière section, nous présentons un modèle stochastique de populations de plantes à fleurs distyles avec reproduction auto-incompatible, et un modèle décrivant l'évolution de la diversité génétique d'un marqueur neutre lié à un trait sous sélection.

## 1.1 Populations structurées en âge

Dans cette section, nous présentons, dans un cadre simplifié, des approximations en grandes populations pour des populations structurées en traits et âges obtenues dans [Tra06, Tra08, MT09, FT09, MT12b, MT13]. Pour simplifier l'exposition, nous "oublions" la structure de traits et ne considérons que l'âge physique  $a$ , à valeurs dans  $\mathbb{R}_+$  et qui est associé à une vitesse de vieillissement de 1. A chaque naissance, les individus naissent avec l'âge 0 et un individu né à la date  $c$  a l'âge  $t - c$  au temps  $t$ . Les résultats présentés peuvent se généraliser en incluant des structures de traits et en considérant un vecteur d'âges, pour prendre en compte différents phénomènes dépendant du passé, avec des vitesses de vieillissement éventuellement non linéaires en temps (e.g. [Tra06, MT13]).

Nous considérerons différentes hypothèses sur les échelles de temps et la taille et la fréquence des mutations. Les ordres de grandeur de ces quantités seront exprimés en fonction de l'ordre de grandeur  $K$  de la taille de la population initiale, qui sera un paramètre que nous ferons tendre vers l'infini. Il sera alors naturel de renormaliser le poids de chaque individu en  $1/K$ . Intuitivement, on peut comprendre  $K$  comme étant la surface ou le volume du milieu dans lequel la population est confinée, et que nous faisons tendre vers l'infini en conservant la densité constante. Ce volume est également l'ordre de grandeur de la capacité de charge, c'est-à-dire la taille de la population maximale que l'environnement peut soutenir. La population est représentée par une mesure ponctuelle

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{a_i(t)}(da) \quad (1.1.1)$$

où  $N_t^K$  est le nombre d'individus vivants au temps  $t$  et où  $a_i(t)$  est l'âge du  $i$ ème individu dans la population au temps  $t$ , les individus étant classés par âges croissants. Durant leurs vies, les individus vieillissent, interagissent, donnent naissance à de nouveaux individus et meurent. On définit ainsi un processus  $(X_t^K)_{t \in \mathbb{R}_+}$  càdlàg, à valeurs mesures, sautant aux instants de naissances

et de morts. La dynamique que nous allons décrire ci-dessous prend en compte de la compétition dans le terme de mort. Ce terme d'interaction introduit des non-linéarités dans les équations qui cassent les propriétés de branchement souvent supposées dans la littérature (c'est-à-dire le fait que les descendances de deux individus de la population vivant au même instant sont indépendantes, e.g. [AN70, Har63, Jag75]).

Un individu d'âge  $a \in \mathbb{R}_+$  donne naissance à un individu d'âge 0 au taux  $b(a)$ , que l'on suppose continu par morceaux et borné par  $\bar{b}$ . Un individu d'âge  $a$ , dans la population représentée par la mesure  $X \in \mathcal{M}_F(\mathbb{R}_+)$ , meurt au taux

$$d(a, X) = d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X(d\alpha). \quad (1.1.2)$$

La fonction  $d_0$  est le taux de mort naturelle. On la suppose continue, bornée par  $\bar{d}_0$  et telle que  $\int_0^{+\infty} d_0(a) da = +\infty$ , ce qui implique que la durée de vie d'un individu non soumis à la compétition est finie presque sûrement. Le noyau  $U(a, \alpha)$  traduit la compétition exercée par l'individu  $\alpha$  sur l'individu  $a$  et  $\eta > 0$  permet de moduler la force de cette compétition. On suppose que  $U(a, \alpha)$  définit une fonction continue bornée par  $\bar{U}$  et minorée par  $\underline{U} > 0$ . Notons que si  $X = X_t^K$ , alors

$$\int_{\mathbb{R}_+} U(a, \alpha) X(d\alpha) = \frac{1}{K} \sum_{i=1}^{N_t^K} U(a, a_i(t)).$$

Il est à remarquer que, les individus étant pondérés par  $1/K$ , la pression de compétition qu'ils exercent sur d'autres individus est elle aussi divisée par  $K$ .

**Exemple 1.1.1.** Dans [MT09], nous considérons deux exemples particuliers :

- Si  $U(a, \alpha) = \bar{U}$  est constante, alors  $\int_{\mathbb{R}_+} U(a, \alpha) X(d\alpha) = \bar{U} \langle X, 1 \rangle$  et on retrouve une population *logistique* où le terme de compétition est proportionnel à la taille  $\langle X, 1 \rangle$  de la population.
- Si  $U(a, \alpha) = V(a)$  ne dépend que de l'âge  $a$ , c'est-à-dire que la pression de la compétition ne dépend pas de l'âge du compétiteur, alors  $\int_{\mathbb{R}_+} U(a, \alpha) X(d\alpha) = V(a) \langle X, 1 \rangle$ , modèle que l'on a qualifié d'*âge-logistique* avec Sylvie Méléard dans [MT09].

### 1.1.1 Construction du processus individu-centré

Soit une condition initiale  $X_0^K$  de la forme (1.1.1) telle que  $\mathbb{E}(N_0^K) < +\infty$ . Il est possible de construire trajectoriellement un processus à valeurs mesures qui décrit l'évolution d'une population régie par la dynamique décrite en début de section. Cette construction trajectorielle correspond à une généralisation de l'algorithme de Gillespie [Gil76, Gil77] utilisé par les biologistes pour leurs simulations (cf. aussi les travaux de Kendall [Ken50] et de Bartlett [Bar53] qui avaient déjà utilisé auparavant ces algorithmes). En comparaison des constructions utilisées pour des populations structurées uniquement en traits (cf. [FM04, CFM06a]) une difficulté réside dans le fait que les masses de Dirac représentant les individus se "déplacent" à cause du vieillissement. Nous exploitons le fait que le vieillissement est déterministe et avec vitesse 1, impliquant qu'on a une évolution déterministe par morceaux entre les événements de naissances et de morts. Ainsi, un individu d'âge  $a_i(0)$  au temps 0 a l'âge  $a_i(0) + t$  au temps  $t$ , un individu d'âge  $a$  au temps  $s$  devrait avoir l'âge  $a + (t - s)$  au temps  $t$  s'il est toujours vivant. Dans [Tra06] et [MT13], des cas plus compliqués impliquant des vitesses de vieillissement non linéaires sont traités, avec structure de traits.

Soit une mesure de Poisson  $Q(ds, di, d\theta)$  sur  $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}_+$  de mesure intensité  $ds \otimes n(di) \otimes d\theta$  où  $ds$  et  $d\theta$  sont les mesures de Lebesgue sur  $\mathbb{R}_+$  et où  $n(di)$  est la mesure de comptage sur  $\mathbb{N}^*$ . La population  $X_t$  au temps  $t$  peut-être obtenue en :

- considérant les individus initiaux avec leurs âges au temps  $t$ . Si un individu a l'âge  $a_i(0)$  au temps 0, on ajoute une masse de Dirac  $a_i(0) + t$ .



- en ajoutant des masses de Dirac correspondant aux âges à  $t$  des individus nés entre 0 et  $t$ . Si un individu naît au temps  $s$ , on ajoute une masse de Dirac  $t - s$ .
- en effaçant les masses de Dirac relatives aux âges à  $t$  des individus qui sont finalement morts avant le temps  $t$ . Si un individu d'âge  $a$  meurt, on retire une masse de Dirac  $a + (t - s)$ .

On a alors :

$$X_t^K(da) = \frac{1}{K} \sum_{i=1}^{N_0^K} \delta_{a_i(0)+t}(da) + \int_{[0,t] \times \mathbb{N}^* \times \mathbb{R}_+} \mathbb{1}_{i \leq N_{s-}} \left[ \mathbb{1}_{\{0 \leq \theta \leq b(a_i(s-))\}} \frac{1}{K} \delta_{t-s} \right. \\ \left. - \mathbb{1}_{\{b(a_i(s-)) < \theta \leq b(a_i(s-)) + d(a_i(s-)) + \eta \int U(a_i(s-), \alpha) X_{s-}^K(d\alpha)\}} \frac{1}{K} \delta_{a_i(s-)+t-s} \right] Q(ds, di, d\theta) \quad (1.1.3)$$

De façon similaire à [FM04, CFM06a, Tra06] on peut prouver l'existence et l'unicité d'une solution de (1.1.3). Pour cela, et pour la suite, on peut montrer en s'inspirant de [FM04, CFM06a] :

**Lemme 1.1.2.** *S'il existe  $p \in \mathbb{N}^*$  tel que  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^p) < +\infty$ , alors pour tout  $T > 0$ ,*

$$\sup_{K \in \mathbb{N}^*} \mathbb{E} \left( \sup_{t \leq T} \langle X_t^K, 1 \rangle^p \right) < +\infty \quad \text{et} \quad \mathbb{E} \left( \sup_{K \in \mathbb{N}^*, t \leq T} \langle X_t^K, 1 \rangle^p \right) < +\infty. \quad (1.1.4)$$

**Proposition 1.1.3.** *Soit  $K$  fixé. Soit une condition initiale aléatoire  $X_0^K$  telle que  $\mathbb{E}(\langle X_0^K, 1 \rangle) < +\infty$ , et soit la mesure de Poisson  $Q(ds, di, d\theta)$  introduite précédemment. Il existe une unique solution trajectorielle à l'équation différentielle stochastique (EDS) (1.1.3).  $\square$*

Il existe d'autres façons d'écrire ces processus, à l'aide de processus de Poisson changés de temps [EK86, Kur80]. L'écriture (1.1.3) nous permet d'utiliser plus facilement le calcul stochastique et reflète la procédure d'acceptation-rejet réalisée lors de l'implémentation numérique. Remarquons que les taux dépendent de l'âge et varient au cours du temps, d'où la nécessité ici de passer par cette procédure d'acceptation-rejet pour la simulation, qui avait déjà été proposée dans [FM04, CFM06a].

**Proposition 1.1.4.** *(i) Considérons des fonctions test  $f_t(a) = f(t, a)$  dans  $\mathcal{C}_b^1(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ . Alors :*

$$\langle X_t^K, f_t \rangle = \langle X_0^K, f_0 \rangle + \int_0^t \langle X_s^K, \partial_s f_s + \partial_a f_s \rangle ds + \frac{1}{K} \int_{[0,t] \times \mathbb{N}^* \times \mathbb{R}_+} \mathbb{1}_{i \leq N_{s-}} \left[ \mathbb{1}_{\{0 \leq \theta \leq b(a_i(s-))\}} f(s, 0) \right. \\ \left. - \mathbb{1}_{\{b(a_i(s-)) < \theta \leq b(a_i(s-)) + d(a_i(s-)) + \eta \int U(a_i(s-), \alpha) X_{s-}^K(d\alpha)\}} f(s, a_i(s)) \right] Q(ds, di, d\theta) \quad (1.1.5)$$

*(ii) Si de plus la condition initiale  $X_0^K$  satisfait  $\mathbb{E}(\langle X_0^K, 1 \rangle^2) < +\infty$ , alors*

$$M_t^{K,f} := \langle X_t^K, f_t \rangle - \langle X_0^K, f_0 \rangle - \int_0^t \langle X_s^K, \partial_s f_s + \partial_a f_s \rangle ds \\ - \int_0^t \int_{\mathbb{R}_+} \left[ b(a) f(s, 0) - (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X_s^K(d\alpha)) f(s, a) \right] X_s^K(da) ds \quad (1.1.6)$$

*est une martingale de carré intégrable issue de zéro et de variation quadratique prévisible :*

$$\langle M^{K,f} \rangle_t := \frac{1}{K} \int_0^t \int_{\mathbb{R}_+} \left[ b(a) f^2(s, 0) + (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X_s^K(d\alpha)) f^2(s, a) \right] X_s^K(da) ds. \quad (1.1.7)$$

Le processus microscopique est facile à simuler, mais son comportement sur des grandes échelles de temps est complexe. Pour le comprendre, on peut essayer d'approcher le processus (1.1.1) par un objet - solution d'EDP, superprocessus ou processus à valeurs dans un espace d'état plus simple - plus facile à étudier. Nous commençons par présenter un résultat de tension général, que nous avons appliqué pour établir ces différentes limites.

### 1.1.2 Critères de tension

Considérons un cadre plus général que celui de cette section. Soit  $E$  un espace vectoriel normé localement compact à base dénombrable (par exemple  $\mathbb{R}_+$  ou  $\mathbb{R}^d$ ). L'espace  $\mathcal{M}_F(E)$  est muni de la topologie de la convergence étroite mais on aura besoin de considérer également la topologie de la convergence vague sur cet espace. Pour bien faire la distinction, nous noterons  $(\mathcal{M}_F(E), \acute{e})$  et  $(\mathcal{M}_F(E), v)$  quand cela sera nécessaire. Soit  $(X^K)_{K \in \mathbb{N}^*}$  une suite de processus càdlàg à valeurs mesures sur  $E$ , i.e. dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(E))$  muni de sa topologie de Skorokhod.

Le schéma classique pour obtenir la convergence en loi de  $(X^K)_{K \in \mathbb{N}^*}$  est le suivant :

- On commence par montrer que la suite  $(X^K)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathbb{D}([0, T], (\mathcal{M}_F(E), \acute{e}))$ , pour tout  $T > 0$ . On peut alors utiliser le théorème de Prohorov qui nous dit que la famille des lois de  $(X^K)_{K \in \mathbb{N}^*}$  est relativement séquentiellement étroitement compacte.
- On caractérise les valeurs d'adhérence de la suite  $(X^K)_{K \in \mathbb{N}^*}$  de façon à montrer l'unicité de la valeur d'adhérence.

Dans un contexte très général, des critères pour des processus càdlàg sont énoncés dans Jakubowski [Jak86, Th. 3.1 et 4.6] et Ethier et Kurtz [EK86, Th. 9.1] :

**Théorème 1.1.5.** *Soit  $\mathfrak{E}$  un espace Polonais. Soit  $\mathfrak{F}$  une famille de fonctions continues réelles sur  $\mathfrak{E}$  telle que :*

- $\mathfrak{F}$  sépare les points de  $\mathfrak{E}$ ,
- $\mathfrak{F}$  est fermée pour l'addition, i.e. si  $f, g \in \mathfrak{F}$ ,  $f + g \in \mathfrak{F}$ .

*La suite des lois de  $(X^K)_{K \in \mathbb{N}^*}$  sur  $\mathbb{D}(\mathbb{R}_+, \mathfrak{E})$  est tendue si et seulement si, pour tout  $\varepsilon > 0$  et  $T > 0$  :*

- (i) *Il existe un compact  $\Gamma_{\varepsilon, T} \subset \mathfrak{E}$  tel que*

$$\exists K_0 \in \mathbb{N}^*, \quad \inf_{K \geq K_0} \mathbb{P}(\forall t \in [0, T], X_t^K \in \Gamma_{\varepsilon, T}) \geq 1 - \varepsilon \quad (1.1.8)$$

- (ii) *Pour tout  $f \in \mathfrak{F}$ , la suite  $(f(X^K))_{K \in \mathbb{N}^*}$  est tendue dans  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ .*

Pour le point (ii), des critères de tension pour les processus càdlàg réels sont connus (voir par exemple [EK86, JM86] pour une exposition). Par exemple, la condition suffisante d'Aldous est rappelée en Th. 1.1.6. La difficulté est de montrer le point (i). Pour les processus à valeurs mesures, on dispose de critères pour obtenir la condition (1.1.8) pour des compacts de  $(\mathcal{M}_F(E), v)$  et ensuite pour vérifier que la masse des mesures  $X_t^K$  ne peut pas s'échapper à l'infini lorsque  $K \rightarrow +\infty$ .

**Théorème 1.1.6** (Critère d'Aldous). *La suite des lois de  $(Y^K)_{K \in \mathbb{N}^*}$  de  $\mathbb{D}(\mathbb{R}_+, \Xi)$ , où  $(\Xi, d)$  est un espace métrique Polonais, est tendue si :*

- (i) *Pour tout  $t$  appartenant à un sous-ensemble dense de  $\mathbb{R}_+$ , les lois de  $(Y_t^K)_{K \in \mathbb{N}^*}$  sont uniformément tendues, dans  $\mathcal{P}(\Xi)$ .*
- (ii) *Pour tout  $T > 0$ ,  $\varepsilon > 0$  et  $\eta > 0$ , il existe  $\delta > 0$  et  $K_0 \in \mathbb{N}^*$  tels que pour toute suite  $(\sigma^K, \tau^K)_{K \in \mathbb{N}^*}$  de temps d'arrêt satisfaisant  $\sigma^K \leq \tau^K \leq T$  et  $\tau^K - \sigma^K \leq \delta$ ,*

$$\sup_{K \geq K_0} \mathbb{P}(d(Y_{\sigma^K}^K, Y_{\tau^K}^K) > \eta) < \varepsilon. \quad (1.1.9)$$

Si  $(Y^K)_{K \in \mathbb{N}^*}$  est une suite de semi-martingales de carré intégrable, de décomposition de Doob  $Y^K = A^K + M^K$ , où  $A^K$  est un processus à variations finies et  $M^K$  une martingale de carré intégrable, alors le critère d'Aldous-Rebolledo (cf e.g. [JM86] ou [Eth00, Th. 1.17]) nous assure que la suite des lois de  $(Y^K)_{K \in \mathbb{N}^*}$  est tendue dès que (1.1.9) est satisfaite par  $(A^K)_{K \in \mathbb{N}^*}$  et  $(M^K)_{K \in \mathbb{N}^*}$ .

Bien sûr, puisque  $\mathcal{M}_F(E)$  est Polonais, on pourrait essayer d'appliquer directement ce critère, en choisissant une bonne distance  $d$  et de bons compacts, mais on présente un critère fondé sur le Théorème 1.1.5 et utilisant la spécificité des espaces de mesures. En particulier, Kurtz [Kur81] a caractérisé les compacts de  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$  :

**Théorème 1.1.7.** *Soit  $T > 0$ . Un ensemble  $\mathfrak{K}$  de  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$  est relativement compact si et seulement si*

- (i) *L'ensemble  $\{X_t, X \in \mathfrak{K}, t \in [0, T]\}$  est relativement compact dans  $(\mathcal{M}_F(E), v)$ , c'est-à-dire qu'il est composé de mesures uniformément bornées.*
- (ii) *Il existe une suite dense  $(f_m)_{m \in \mathbb{N}}$  de  $\mathcal{C}_0(E, \mathbb{R})$  telle que*

$$\lim_{\delta \rightarrow 0} \sup_{X \in \mathfrak{K}} w'_{\langle X, f_m \rangle}(\delta) = 0$$

où pour  $y \in \mathbb{D}([0, T], \mathbb{R})$ ,

$$w'_y(\delta) = \sup_{t \in [0, T-\delta]} \left[ \inf_{\{t_i\}} \max_{0 \leq i \leq r} \left( \sup_{t \leq u < v \leq t+\delta} |y_u - y_v| \right) \right]$$

est le module de continuité pour la topologie de Skorokhod sur  $\mathbb{D}([0, T], \mathbb{R})$ , l'infimum étant pris sur les subdivisions finies de  $[0, T]$  de pas supérieur à  $\delta$  et  $r$  étant le nombre d'intervalles de cette subdivision (cf. [Bil99] ou [JM86]).

Avant d'énoncer le critère de compacité qui nous intéresse, nous introduisons une famille de fonctions  $(\varphi_k)_{k \in \mathbb{N}^*}$ , de classe  $\mathcal{C}^2$  et qui approchent les indicatrices  $\mathbb{1}_{\{x \geq k\}}$ , en suivant les idées de [JMW12] (voir aussi [MT09]). Cette famille nous permettra de transformer la tension dans  $(\mathcal{M}_F(E), v)$  en tension dans  $(\mathcal{M}_F(E), \acute{e})$ .

$$\forall x \in \mathbb{R}, \quad \varphi_0(x) = 1, \quad \varphi_k(x) = \psi(0 \vee (x - (k-1)) \wedge 1), \quad (1.1.10)$$

avec  $\psi(x) = 6x^5 - 15x^4 + 10x^3$ . Ces fonctions satisfont  $\psi(0) = \psi'(0) = \psi''(0) = 1 - \psi(1) = \psi'(1) = \psi''(1) = 0$  et pour tout  $x \in \mathbb{R}$  :

$$\begin{aligned} \mathbb{1}_{\{x \geq k\}} &\leq \varphi_k(x) \leq \mathbb{1}_{\{x \geq k-1\}} && \text{et} \\ \varphi'_k(x) &\leq \left( \sup_{x \in [k-1, k]} |\varphi'_k(x)| \right) \mathbb{1}_{\{x \geq k-1\}} \leq \left( \sup_{x \in [k-1, k]} |\varphi'_k(x)| \right) \varphi_{k-1}(x). \end{aligned}$$

**Théorème 1.1.8.** *Soit  $(X^K)_{K \in \mathbb{N}^*}$  une suite de  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(E))$ . Soit  $F$  une famille de fonctions continues stable pour l'addition, dense dans  $\mathcal{C}_0(E, \mathbb{R})$ , et à laquelle on ajoute la fonction constante égale à 1. Une condition suffisante de tension des lois de  $(X^K)_{K \in \mathbb{N}^*}$  dans  $\mathcal{P}(\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), \acute{e})))$  est que pour tout  $T > 0$  :*

- (a) *Pour toute fonction  $f \in F$ , la suite  $(\langle X^K, f \rangle)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathbb{D}([0, T], \mathbb{R})$ .*
- (b) *Pour les fonctions  $(\varphi_k)$  définies en (1.1.10), on a :*

$$\lim_{k \rightarrow +\infty} \limsup_{K \rightarrow +\infty} \sup_{t \leq T} \int_E \varphi_k(|x|) X_t^K(dx) = 0 \quad \text{en proba.} \quad (1.1.11)$$

- (c) *Les valeurs d'adhérences de  $(X^K)_{K \in \mathbb{N}^*}$ , si elles existent, appartiennent à  $\mathcal{C}([0, T], (\mathcal{M}_F(E), \acute{e}))$  p.s.*

*Idée de la preuve du Théorème 1.1.8.* Remarquons que pour montrer le point (a), on peut utiliser le critère d'Aldous-Rebolledo. Le point (ii) du critère énoncé au Théorème 1.1.5 est satisfait avec la famille de fonction  $\mathfrak{F} = \{\nu \in \mathcal{M}_F(E) \mapsto \langle \nu, f \rangle, f \in F\}$ . Il reste à montrer la condition (1.1.8). Dans la suite, nous introduisons une numérotation  $(f_m)_{m \in \mathbb{N}}$  de  $F$ , avec par convention  $f_0 \equiv 1$ .

Roelly [Roe86] a montré que le point (a) impliquait la tension des lois de  $(X^K)_{k \in \mathbb{N}^*}$  dans  $\mathcal{P}(\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), v)))$ . En effet, la tension des lois de  $(\langle X^K, 1 \rangle)_{K \in \mathbb{N}^*}$ , implique qu'il existe un compact  $\mathcal{K}_0$  de  $\mathbb{D}([0, T], \mathbb{R})$  (composé en particulier de processus uniformément bornés) tel que

$$\forall K \in \mathbb{N}^*, \quad \mathbb{P}(\langle X^K, 1 \rangle \in \mathcal{K}_0) \geq 1 - \frac{\varepsilon}{2}.$$

De même, la tension des lois de  $(\langle X^K, f_m \rangle)_{K \in \mathbb{N}^*}$ , pour chaque  $m \in \mathbb{N}^*$ , implique l'existence d'un compact  $\mathcal{K}_m$  de  $\mathbb{D}([0, T], \mathbb{R})$  tel que

$$\forall K \in \mathbb{N}^*, \mathbb{P}(\langle X^K, f_m \rangle \in \mathcal{K}_m) \geq 1 - \frac{\varepsilon}{2^{m+1}}, \quad \text{et} \quad \lim_{\delta \rightarrow 0} \sup_{y \in \mathcal{K}_m} w'_y(\delta) = 0.$$

En notant  $\pi_f : X \in \mathcal{M}_F(E) \mapsto \langle X, f \rangle \in \mathbb{R}$  et  $\pi_f^{-1}$  sa réciproque, on définit  $\mathcal{K} = \pi_{f_0}^{-1} \mathcal{K}_0 \cap (\bigcap_{m \geq 1} \pi_{f_m}^{-1} \mathcal{K}_m)$ .  $\mathcal{K}$  est un ensemble compact de  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$  par le Théorème 1.1.7, et contient  $X^K$  avec probabilité supérieure à  $1 - \varepsilon$  pour tout  $K \in \mathbb{N}^*$ .

Considérons alors une valeur d'adhérence  $\bar{X}$  et une sous-suite  $(X^{u_K})_{K \in \mathbb{N}^*}$  qui converge en loi dans  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), v))$  vers  $\bar{X}$ . Par le point (a), et puisque  $F$  contient la fonction constante égale à 1, la suite  $(\langle X^{u_K}, 1 \rangle)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ . Quitte à extraire de  $(X^{u_K})_{K \in \mathbb{N}^*}$  une autre sous-suite que l'on appelle encore  $(X^{u_K})_{K \in \mathbb{N}^*}$ , on peut supposer que  $(\langle X^{u_K}, 1 \rangle)_{K \in \mathbb{N}^*}$  converge aussi en loi dans  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ . Pour une suite satisfaisant (b) et (c), ceci est équivalent à la convergence en loi de  $(X^{u_K})_{K \in \mathbb{N}^*}$  vers  $\bar{X}$  dans  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), \acute{e}))$ . Cette équivalence fait l'objet du lemme suivant.  $\square$

**Lemme 1.1.9.** *Soit  $T > 0$ . Soit  $(X^K)_{K \in \mathbb{N}}$  une suite de processus de  $\mathbb{D}([0, T], (\mathcal{M}_F(E), \acute{e}))$ , et soit  $\bar{X}$  un processus de  $\mathcal{C}([0, T], (\mathcal{M}_F(E), \acute{e}))$  satisfaisant (1.1.11). Alors  $(X^K)_{K \in \mathbb{N}}$  converge en loi vers  $\bar{X}$  dans  $\mathbb{D}([0, T], (\mathcal{M}_F(E), \acute{e}))$  si et seulement si*

1.  $(X^K)_{K \in \mathbb{N}}$  converge en loi vers  $\bar{X}$  dans  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$
2. la suite de processus réels  $(\langle X^K, 1 \rangle)_{K \in \mathbb{N}}$  converge en loi dans  $\mathbb{D}([0, T], \mathbb{R})$ .

*Démonstration.* Seul le sens  $\Leftarrow$  doit être prouvé. Par le théorème de représentation de Skorokhod (e.g. [Bil95, Th. 25.6 p.333]), il est possible de trouver, sur un espace de probabilité  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ , une suite de processus  $(\tilde{X}^K)_{K \in \mathbb{N}}$  de  $\mathbb{D}([0, T], (\mathcal{M}_F(E), \acute{e}))$ , un processus  $\tilde{X}$  de  $\mathcal{C}([0, T], (\mathcal{M}_F(E), \acute{e}))$  et un processus réel  $\tilde{x}$  de  $\mathcal{C}([0, T], \mathbb{R})$  tels que

- $(\tilde{X}^K, \langle \tilde{X}^K, 1 \rangle)$  converge presque-sûrement vers  $(\tilde{X}, \tilde{x})$  dans  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v)) \times \mathbb{D}([0, T], \mathbb{R})$
- pour tout  $K \in \mathbb{N}$ ,  $(\tilde{X}^K, \langle \tilde{X}^K, 1 \rangle)$  a même loi que  $(X^K, \langle X^K, 1 \rangle)$
- $\tilde{X}$  a même loi que  $\bar{X}$ .

Remarquons que puisque les trajectoires de  $(\tilde{X}, \tilde{x})$  sont continues, la convergence donnée par le point 1 est en fait une convergence uniforme sur  $[0, T]$  (e.g. [Bil68, p.112]).

Si on montre que  $\tilde{x} = \langle \tilde{X}, 1 \rangle$ , alors, en utilisant un critère de Méléard et Roelly [MR93], on obtiendra le résultat.

On a :

$$\begin{aligned} |\langle \tilde{X}_t^K, 1 \rangle - \langle \tilde{X}_t, 1 \rangle| &\leq |\langle \tilde{X}_t^K, 1 \rangle - \int_E (1 - \varphi_k(|x|)) \tilde{X}_t^K(dx)| \\ &\quad + |\int_E (1 - \varphi_k(|x|)) \tilde{X}_t^K(dx) - \int_E (1 - \varphi_k(|x|)) \tilde{X}_t(dx)| \\ &\quad + |\int_E (1 - \varphi_k(|x|)) \tilde{X}_t(dx) - \langle \tilde{X}_t, 1 \rangle| := A_t^{K,k} + B_t^{K,k} + C_t^k. \end{aligned} \quad (1.1.12)$$

On a  $A_t^{K,k} = \int_E \varphi_k(|x|) \tilde{X}_t^K(dx)$  et par l'hypothèse (1.1.11),  $\lim_k \limsup_K \sup_{t \leq T} A_t^{K,k} = 0$ .

La convergence p.s. et uniforme en temps de  $(\tilde{X}^K)_{K \in \mathbb{N}^*}$  vers  $\tilde{X}$  dans  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$  et le fait que  $1 - \varphi_k(|x|)$  soit à support compact assurent que  $\lim_{K \rightarrow +\infty} \sup_{t \leq T} B_t^{K,k} = 0$  pour tout  $k \in \mathbb{N}$ .

Enfin, pour  $C_t^k = \int_E \varphi_k(|x|) \tilde{X}_t(dx)$ , on peut écrire :

$$\begin{aligned} \sup_{t \leq T} C_t^k &= \sup_{t \leq T} \int_E \lim_{\ell \rightarrow +\infty} \varphi_k(|x|)(1 - \varphi_\ell(|x|)) \tilde{X}_t(dx) \\ &\leq \sup_{t \leq T} \liminf_{\ell \rightarrow +\infty} \int_E \varphi_k(|x|)(1 - \varphi_\ell(|x|)) \tilde{X}_t(dx) \quad \text{par le lemme de Fatou} \\ &= \sup_{t \leq T} \liminf_{\ell \rightarrow +\infty} \lim_{K \rightarrow +\infty} \int_E \varphi_k(|x|)(1 - \varphi_\ell(|x|)) \tilde{X}_t^K(dx) \\ &\leq \sup_{t \leq T} \limsup_{K \rightarrow +\infty} \int_E \varphi_k(|x|) \tilde{X}_t^K(dx) \leq \limsup_{K \rightarrow +\infty} \sup_{t \leq T} \int_E \varphi_k(|x|) \tilde{X}_t^K(dx) \end{aligned}$$

d'où par (1.1.11),  $\lim_{k \rightarrow +\infty} \sup_{t \leq T} C_t^k = 0$ . Donc  $\langle \tilde{X}_t^K, 1 \rangle$  converge en proba (et donc aussi p.s.) uniformément vers  $\langle \tilde{X}_t, 1 \rangle$ , qui est donc égal à  $\tilde{x}$ . Ceci conclut la preuve.  $\square$

**Remarque 1.1.10.** On remarque que si l'on dispose de conditions de moments telles que

$$\sup_{K \in \mathbb{N}^*} \mathbb{E}(\sup_{t \leq T} \langle \tilde{X}_t^K, 1 \rangle) < +\infty,$$

par exemple, alors on peut affaiblir (1.1.11) en :

$$\lim_{k \rightarrow +\infty} \limsup_{K \rightarrow +\infty} \mathbb{E}(\sup_{t \leq T} \int_E \varphi_k(|x|) X_t^K(dx)) = 0. \quad (1.1.13)$$

En effet, on peut alors montrer avec le lemme de Fatou que  $\mathbb{E}(\sup_{t \leq T} \tilde{x}_t) < +\infty$ , puis par convergence dominée et en utilisant (1.1.13) que

$$\lim_{K \rightarrow +\infty} \mathbb{E}(\sup_{t \leq T} |\langle \tilde{X}_t^K, 1 \rangle - \langle \tilde{X}_t, 1 \rangle|) = 0.$$

### 1.1.3 Approximation par l'EDP de McKendrick-Von Foerster

Revenons à la suite  $(X^K)_{K \in \mathbb{N}^*}$  introduite en (1.1.1) dont l'évolution est décrite en (1.1.3). Dans [Tra08], on étudie la suite de processus  $(X^K)_{K \in \mathbb{N}^*}$  lorsque  $K$  tend vers l'infini, sans hypothèse particulière sur le taux de mutation ou sur les échelles de temps. A la Proposition 1.1.4, on remarque dans (1.1.6)-(1.1.7) que  $\langle X_t^K, f_t \rangle$  est une semi-martingale dont le crochet est d'ordre  $1/K$ . Lorsque  $K \rightarrow +\infty$ , si l'on reste dans l'échelle de temps des naissances et morts, on s'attend à voir un phénomène de moyennage dans lequel la partie martingale disparaît à la limite. On montre la convergence vers la solution d'une EDP généralisant les équations de McKendrick-Von Foerster [McK26, Foe59] et Gurtin-McCamy [GM74] qui décrivent l'écologie à l'échelle de la population. Dans ces solutions, les trajectoires individuelles sont perdues. Plusieurs théorèmes limites similaires avaient été obtenus auparavant. Wang [Wan77], Solomon [Sol87] considèrent des taux de mort ne dépendant pas de l'âge. Oelschläger [Oel90] regarde des taux dépendant de l'âge, mais bornés, ce qui exclut les compétitions de type logistique. Jagers et Klebaner [JK00] considèrent une non-linéarité qui disparaît lorsque la population est de grande taille. En 2011, ces auteurs [JK11] proposent un modèle de population structurée par âge similaire à celui considéré dans [Tra08], avec dépendance par rapport à la taille  $N_t^K$  de la population, et en étudient le comportement en temps grand.

**Théorème 1.1.11.** *S'il existe  $\xi_0 \in \mathcal{M}_f(\mathbb{R}_+)$  telle que la suite  $(X_0^K)_{K \in \mathbb{N}^*}$  converge étroitement, en probabilité, vers  $\xi_0$  et si  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ , alors, pour tout  $T > 0$ , la suite  $(X^K)_{K \in \mathbb{N}^*}$  converge dans  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{R}_+))$  vers la solution  $\xi \in \mathcal{C}([0, T], \mathcal{M}_F(\mathbb{R}_+))$  de l'équation suivante,*

où  $f_t(a) \in \mathcal{C}_b^1(\mathbb{R}_+, \mathbb{R})$  est une fonction test

$$\begin{aligned} \langle \xi_t, f_t \rangle &= \langle \xi_0, f_0 \rangle + \int_0^t \langle \xi_s, \partial_s f_s + \partial_a f_s \rangle ds \\ &\quad + \int_0^t \int_{\mathbb{R}_+} [b(a)f_s(0) - (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) \xi_s(d\alpha)) f_s(a)] \xi_s(da) ds. \end{aligned} \quad (1.1.14)$$

Si de plus  $\xi_0(da) = n_0(a)da$  est une mesure déterministe admettant la densité  $n_0(a)$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}_+$ , alors pour tout  $t \geq 0$ ,  $\xi_t(da) = n(t, a)da$  admet une densité  $n(t, a)$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}_+$  qui satisfait l'EDP :

$$\begin{aligned} \partial_t n(t, a) + \partial_a n(t, a) &= -(d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) n(t, \alpha) d\alpha) n(t, a) \\ n(t, 0) &= \int_{\mathbb{R}_+} b(a) n(t, a) da, \quad n(0, a) = n_0(a). \end{aligned} \quad (1.1.15)$$

*Idée de la preuve.* On commence par montrer la tension de la suite des lois de  $(X^K)_{K \in \mathbb{N}^*}$  dans  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{R}_+))$  en utilisant le Théorème 1.1.8 et la Remarque 1.1.10. On caractérise les valeurs d'adhérence de la suite  $(X^K)_{K \in \mathbb{N}^*}$  comme les solutions de (1.1.14). L'existence d'une unique solution à cette équation nous permettra d'obtenir l'unicité de la valeur d'adhérence de  $(X^K)_{K \in \mathbb{N}^*}$  et ainsi la convergence de cette suite.

**Etape 1 :** Nous commençons donc par montrer que pour toute fonction  $f \in \mathcal{C}_b^1(\mathbb{R}_+, \mathbb{R})$ , la suite  $(\langle X^K, f \rangle)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathbb{D}([0, T], \mathbb{R})$ . Par la Proposition 1.1.4, pour tout  $K \in \mathbb{N}^*$ ,  $\langle X^K, f \rangle$  est une semi-martingale réelle  $\langle X_t^K, f \rangle = \langle X_0^K, f \rangle + A_t^{K,f} + M_t^{K,f}$  dont la partie martingale de carré intégrable est donnée en (1.1.6) avec comme crochet :

$$\langle M^{K,f} \rangle_t = \frac{1}{K} \int_0^t \int_{\mathbb{R}_+} [b(a)f^2(0) + (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X_s^K(d\alpha)) f^2(a)] X_s^K(da) ds \quad (1.1.16)$$

et dont la partie à variation finie prévisible est :

$$A_t^{K,f} = \int_0^t \int_{\mathbb{R}_+} [f'(a) + b(a)f(0) - (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X_s^K(d\alpha)) f(a)] X_s^K(da) ds. \quad (1.1.17)$$

En utilisant (1.1.4) pour  $p = 2$ , on a pour tous  $\varepsilon > 0$ ,  $\eta > 0$ ,  $n$  suffisamment grand,  $\delta$  suffisamment petit et pour tout couple de temps d'arrêt  $(T_K, S_K)$  tels que  $S_K < T_K < (S_K + \delta) \wedge T$ , que  $\mathbb{P}(|A_{T_n}^{n,f} - A_{S_n}^{n,f}| > \eta) \leq \varepsilon$  et  $\mathbb{P}(|\langle M^{n,f} \rangle_{T_n} - \langle M^{n,f} \rangle_{S_n}| > \eta) \leq \varepsilon$ . Le critère d'Aldous-Rebolledo donne alors la tension uniforme des lois de  $(\langle \bar{X}^K, f \rangle)_{K \in \mathbb{N}^*}$ . En particulier, ceci vaut pour  $f \equiv 1$ .

**Etape 2 :** Soit  $\bar{X}$  une valeur d'adhérence. Comme  $\sup_{t \in [0, T]} \sup_{f, \|f\|_\infty \leq 1} |\langle X_t^K, f \rangle - \langle X_{t-}^K, f \rangle| \leq \frac{1}{K}$  alors  $\bar{X} \in \mathcal{C}([0, T], (\mathcal{M}_F(\mathbb{R}_+), v))$  presque-sûrement. Notons que si  $\bar{X}$  est continu de  $[0, T]$  dans  $(\mathcal{M}_F(\mathbb{R}_+), v)$ , il est également continu de  $[0, T]$  dans  $(\mathcal{M}_F(\mathbb{R}_+), \hat{e})$ .

**Etape 3 :** Montrons maintenant (1.1.11). Pour cela, on commence par montrer que

$$\lim_{k \rightarrow +\infty} \limsup_{K \rightarrow +\infty} \mathbb{E}(\langle X_t^{u_K}, \varphi_k \rangle) = 0, \quad (1.1.18)$$

où  $(\varphi_k)_{k \in \mathbb{N}}$  est la suite de fonctions introduite en (1.1.10). On note que  $\mathbb{E}(\langle X_t^{u_K}, \varphi_k \rangle) \leq \mathbb{E}(\langle X_t^{u_K}, 1 \rangle) \leq \mathbb{E}(\langle X_0^{u_K}, 1 \rangle) \exp(\bar{b}T)$ . En utilisant (1.1.6) où l'on néglige les termes de mort négatifs et en notant

que dès que  $k \leq 1$ ,  $\varphi_k(0) = 0$ , on obtient

$$\begin{aligned}
\mathbb{E}(\langle X_t^{u_K}, \varphi_k \rangle) &\leq \mathbb{E}(\langle X_0^{u_K}, \varphi_k \rangle) + \mathbb{E}\left(\int_0^t \langle X_s^{u_K}, \varphi'_k \rangle ds\right) \\
&\leq \mathbb{E}(\langle X_0^{u_K}, \varphi_k \rangle) + C \int_0^t \mathbb{E}(\langle X_s^{u_K}, \varphi_{k-1} \rangle) ds \\
&\leq \sum_{\ell=0}^{k-1} \mathbb{E}(\langle X_0^{u_K}, \varphi_{k-\ell} \rangle) \frac{(Ct)^\ell}{\ell!} + \frac{(C \int_0^t \mathbb{E}(\langle X_s^{u_K}, 1 \rangle) ds)^k}{k!} \\
&\leq \mathbb{E}(\langle X_0^{u_K}, \varphi_{[k/2]} \rangle) e^{Ct} + \mathbb{E}(\langle X_0^{u_K}, 1 \rangle) \sum_{\ell=[k/2]+1}^{+\infty} \frac{(Ct)^\ell}{\ell!} + \frac{(C't)^k}{k!}
\end{aligned}$$

en séparant la somme sur  $\ell$  en deux parties pour  $\ell \leq [k/2]$  et  $\ell > [k/2]$ . Chacun des termes de la majoration converge vers 0 lorsque  $k \rightarrow +\infty$ , ce qui prouve (1.1.18). Comme

$$\mathbb{E}\left(\sup_{t \leq T} \langle X_t^{u_K}, \varphi_k \rangle\right) \leq \mathbb{E}(\langle X_0^{u_K}, \varphi_k \rangle) + \int_0^t C \mathbb{E}(\langle X_s^{u_K}, \varphi_{k-1} \rangle) ds + \mathbb{E}\left(\sup_{t \leq T} M_t^{K, \varphi_k}\right),$$

on obtient alors (1.1.11), en utilisant (1.1.18) et l'inégalité de Doob, combinée avec le fait que  $\mathbb{E}(\langle M^{K, \varphi_k} \rangle_T)$  converge vers 0 en  $1/K$  par (1.1.7). Ceci conclut la preuve de la tension de  $(X^K)_{K \in \mathbb{N}^*}$  dans  $\mathbb{D}([0, T], (\mathcal{M}_F(\mathbb{R}_+), \acute{e}))$ .

**Etape 4 :** Soit  $\bar{X}$  une valeur d'adhérence de la suite  $(X^K)_{K \in \mathbb{N}^*}$  et soit une sous-suite, encore notée  $(X^K)_{K \in \mathbb{N}^*}$  par abus de notation, qui converge vers  $\bar{X}$ . Pour  $n \in \mathbb{N}^*$ ,  $0 \leq t_1 < \dots < t_n < t \leq T$ ,  $f(t, a) \in \mathcal{C}_b^1(\mathbb{R}_+^2, \mathbb{R})$ , introduisons la fonctionnelle  $\Psi$  définie pour  $X \in \mathbb{D}([0, t], \mathcal{M}_F(\mathbb{R}_+))$  par :

$$\begin{aligned}
\Psi_{t,f}(X) &= \langle X_t, f_t \rangle - \langle X_0, f_0 \rangle \\
&\quad - \int_0^t \int_{\mathbb{R}_+} \left[ \partial_s f_s(a) + \partial_a f_s(a) + b(a) f(0) - (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) X_s(d\alpha)) f(a) \right] X_s(da) ds.
\end{aligned}$$

Comme  $X^K$  converge vers  $\bar{X}$  qui est continue,  $\lim_{K \rightarrow +\infty} \Psi_{t,f}(X^K) = \Psi_{t,f}(\bar{X})$ . Comme

$$|\Psi_{t,f}(X^K)| \leq \sup_{t \in [0, T]} \langle X_t^K, 1 \rangle \|f\|_{\mathcal{C}_b^1(\mathbb{R}_+^2, \mathbb{R})} (2 + T + \bar{b}T + \bar{d}_0 T) + \eta \bar{U} T \|f\|_{\mathcal{C}_b^1(\mathbb{R}_+^2, \mathbb{R})} \sup_{t \in [0, T]} \langle X_t^K, 1 \rangle^2$$

alors par l'hypothèse de moments d'ordre 3 pour la condition initiale,  $|\Psi_{t,f}(X^K)|$  est uniformément intégrable et  $\lim_{K \rightarrow +\infty} \mathbb{E}(|\Psi_{t,f}(X^K)|) = \mathbb{E}(|\Psi_{t,f}(\bar{X})|)$ . Par ailleurs on sait que  $\Psi_{t,f}(X^K) = M_t^{K,f}$ , d'où l'on conclut par Doob, (1.1.7) et (1.1.4) que  $\mathbb{E}(|\Psi_{t,f}(\bar{X})|) = 0$ . On en déduit que  $M_t^{K,f} = 0$  presque-sûrement et ainsi, les valeurs d'adhérence de  $(X^K)_{K \in \mathbb{N}^*}$  sont solutions de (1.1.14).

Cela prouve au passage l'existence d'une solution à (1.1.14).

**Etape 5 :** Soient  $t \in [0, T]$ ,  $\varphi \in \mathcal{C}_b^1(\mathbb{R}, \mathbb{R})$  et soit  $f(s, a) = \varphi(t + a - s)$ . On a  $f(t, a) = \varphi(a)$ ,  $f(0, a) = \varphi(a + t)$  et  $\partial_s f(s, a) + \partial_a f(s, a) = 0$ . En appliquant (1.1.14) avec cette fonction test,

$$\begin{aligned}
\langle \xi_t, \varphi \rangle &= \int_{\mathbb{R}_+} \varphi(a + t) \xi_0(da) \\
&\quad + \int_0^t \int_{\mathbb{R}_+} [b(a) \varphi(t - s) - (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) \xi_s(da)) \varphi(a + t - s)] \xi_s(da) ds. \quad (1.1.19)
\end{aligned}$$

Nous pouvons utiliser cette reformulation pour prouver l'unicité de la solution de (1.1.14). On a alors unicité de la valeur d'adhérence, ce qui prouve la convergence en loi de  $(X^K)_{K \in \mathbb{N}^*}$  vers la

solution de (1.1.14).

**Etape 6 :** Si la condition initiale  $\xi_0$  admet une densité  $n_0(a)$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}_+$ , alors (1.1.19) utilisée avec une fonction test  $\varphi$  positive nous fournit que :

$$0 \leq \langle \xi_t, \varphi \rangle \leq \int_0^{+\infty} \varphi(\alpha) [\mathbb{1}_{\alpha \geq t} n_0(\alpha - t) + \mathbb{1}_{\alpha < t} \langle \xi_{t-\alpha}, b \rangle] d\alpha.$$

On peut remarquer que dans le majorant, le premier terme correspond aux individus présents à la date initiale, tandis que le second terme correspond aux individus nés après le temps  $t$ . Donc  $\xi_t$  est dominée par une mesure absolument continue par rapport à la mesure de Lebesgue et admet une densité sur  $\mathbb{R}_+$ . En reprenant (1.1.19) et en réexprimant  $\xi_s(da) = n(s, a)da$ , on obtient la formulation exacte de la densité  $n(t, \alpha)$  par identification, qui est le crochet de (1.1.20) :

$$\begin{aligned} \langle \xi_t, \varphi \rangle = \int_0^{+\infty} \varphi(\alpha) & \left[ \mathbb{1}_{\alpha \geq t} n_0(\alpha - t) + \mathbb{1}_{\alpha < t} \int_{\mathbb{R}_+} (b(a) - d_0(a)) n(t - \alpha, a) da \right. \\ & \left. - \int_{t-\alpha}^t (d_0(\alpha - t + s) + \eta \int_{\mathbb{R}_+} U(\alpha + s - t, u) \xi_s(du)) ds \right] d\alpha. \end{aligned} \quad (1.1.20)$$

**Etape 7 :** Considérons (1.1.14) avec l'information que  $\xi_t(da) = n(t, a)da$ , et avec une fonction test  $f(a) \in \mathcal{C}_b^1(\mathbb{R}_+, \mathbb{R})$ . Supposons que  $(s, a) \mapsto n(s, a)$  soit de classe  $\mathcal{C}^1$ . Puisque  $\xi_s(da)$  est une mesure finie et positive, on a nécessairement que  $\lim_{a \rightarrow +\infty} n(s, a) = 0$  pour tout  $s \in [0, T]$ . Alors, par intégration par partie, on obtient :

$$\begin{aligned} \int_{\mathbb{R}_+} f(a) n(t, a) da &= \int_{\mathbb{R}_+} f(a) n_0(a) da + \int_0^t \left[ -f(0) n(s, 0) - \int_{\mathbb{R}_+} f(a) \partial_a n(s, a) \right. \\ & \left. + f(0) b(a) n(s, a) - f(a) (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) n(s, \alpha) d\alpha) n(s, a) \right] da ds. \end{aligned} \quad (1.1.21)$$

Si l'on choisit une fonction test  $f(a)$  dont le support est inclus dans  $(0, +\infty)$ , alors  $f(0) = 0$  et par identification des intégrands :

$$n(t, a) = n_0(a) - \int_s^t ds \left[ \partial_a n(s, a) + (d_0(a) + \eta \int_{\mathbb{R}_+} U(a, \alpha) n(s, \alpha) d\alpha) n(s, a) \right] \quad (1.1.22)$$

ce qui fournit en dérivant par rapport à  $t$  la première équation de (1.1.15) avec  $n(a, 0) = n_0(a)$ . Maintenant, si 0 est inclus dans le support de  $f(a)$ , (1.1.21) et (1.1.22) fournissent la condition au bord  $n(t, 0) = \int_{\mathbb{R}_+} b(a) n(t, a) da$ .

Or l'EDP de McKendrick-Von Foerster (1.1.15) admet une unique solution régulière [McK26, Foe59, Web85]. On en déduit, par unicité de la solution de (1.1.14), que la densité de  $\xi_t(da)$  est la solution de (1.1.15) et qu'elle est bien  $\mathcal{C}^1$ .  $\square$

L'équation (1.1.15) est une équation de transport qui a été beaucoup étudiée dans la littérature (e.g. [McK26, Foe59, Web85]). Pour la suite, nous calculons les solutions stationnaires de l'EDP limite dans le cas où  $U(a, \alpha) = U(a)$  ne dépend que de l'âge  $a$  de l'individu qui subit la pression de compétition. Une solution stationnaire  $\hat{n}(a)$  est solution de :

$$\frac{d\hat{n}(a)}{da} = -(d_0(a) + \eta U(a) \hat{N}) \hat{n}(a) \quad (1.1.23)$$

$$\hat{N} = \int_{\mathbb{R}_+} \hat{n}(\alpha) d\alpha, \quad \hat{n}(0) = \int_0^{+\infty} b(a) \hat{n}(a) da. \quad (1.1.24)$$

La fonction nulle  $\hat{n}(a) = 0$  est solution. Cherchons une solution qui ne soit pas nulle pour tout  $a$ .



**Proposition 1.1.12.** *L'EDO (1.1.23)-(1.1.24) admet une unique solution non triviale si et seulement si*

$$R_0 := \int_0^{+\infty} b(a) e^{-\int_0^a d_0(\alpha) d\alpha} da > 1. \quad (1.1.25)$$

Cette solution est alors :

$$\hat{n}(a) = \frac{\hat{N} e^{-\int_0^a (d_0(\alpha) + \eta U(\alpha) \hat{N}) d\alpha}}{\int_0^{+\infty} e^{-\int_0^a (d_0(\alpha) + \eta U(\alpha) \hat{N}) d\alpha} da}, \quad (1.1.26)$$

où  $\hat{N}$  est l'unique solution de

$$1 = \int_0^{+\infty} b(a) e^{-\int_0^a (d_0(\alpha) + \eta U(\alpha) \hat{N}) d\alpha} da. \quad (1.1.27)$$

Si  $R_0 < 1$ , il n'existe pas de solution stationnaire non triviale. Si  $R_0 = 1$ , il ne peut exister une solution stationnaire non triviale que si  $U \equiv 0$  et  $b(a) = d(a) := r(a)$  pour tout  $a \in \mathbb{R}_+$ . Les solutions stationnaires sont alors paramétrées par la condition initiale  $\hat{n}(0) : \hat{n}(a) = \hat{n}(0) e^{-\int_0^a r(\alpha) d\alpha}$ .

*Démonstration.* L'équation (1.1.23) s'intègre en

$$\hat{n}(a) = \hat{n}(0) \exp \left( - \int_0^a (d_0(\alpha) + \eta U(\alpha) \hat{N}) d\alpha \right). \quad (1.1.28)$$

En injectant (1.1.28) dans la condition au bord de (1.1.24), on obtient (1.1.27), dont le membre de droite représente le nombre moyen de descendants dans une population à l'équilibre, quand le taux de naissance est  $b$  et que le taux de mort est  $d_0(\alpha) + \eta U(\alpha) \hat{N}$ . On voit que le membre de droite de (1.1.27) définit une fonction strictement décroissante de  $\hat{N}$  et qu'il ne peut exister de solution à (1.1.27) que si  $R_0 > 1$ . Dans ce cas, la solution  $\hat{N}$  est unique. En injectant (1.1.28) dans la définition de  $\hat{N}$ , on trouve  $\hat{n}(0)$  en fonction de  $\hat{N}$  et ceci fournit (1.1.26).  $\square$

#### 1.1.4 Approximation par une diffusion de Feller avec coefficients moyennés

Nous considérons maintenant limite diffusive, qui correspond à une limite en grande population avec une accélération du temps de (1.1.3) paramétrée par la taille  $K$  de la population initiale, et lorsque les taux de naissance et de mort ne diffèrent qu'à l'ordre  $1/K$ . Cette question, naturelle au vu de ce qui avait été étudié pour des populations structurées sans âge [FM04, CFM06a] m'avait été posée par Sylvie Méléard durant ma thèse, mais nous ne l'avons résolue que plus tard [MT12b, MMT11]. Dans [MT09], nous avons considéré une population structurée en traits et âges, avec des taux de naissance et de mort qui dépendent des deux types de variables. On montre alors que la suite  $(X_K^K)_{K \in \mathbb{N}^*}$  converge vers un superprocessus. Pour simplifier l'exposition des résultats, nous ne considérons, comme dans la section précédente, que la structure d'âge avec un âge réel. Dans ce cadre simplifié, on montre que la distribution d'âge se stabilise autour d'une structure d'âge stationnaire tandis que la taille de la population est décrite à la limite par une diffusion de Feller dont les coefficients sont des moyennes par rapport à cette structure d'âge.

La principale difficulté est que la structure d'âge, qui va déterminer les taux de naissance et de mort globalement au niveau de la population, évolue dans une échelle de temps relative à l'individu et que cette échelle diffère de celle dans laquelle on se place lorsque l'on accélère le temps en  $Kt$ . Dans l'échelle  $Kt$ , la vitesse de vieillissement devient  $K$  et les âges des individus évoluent de plus en plus vite. Comment faire pour obtenir une limite macroscopique diffusive qui "ne voit" plus les individus lorsque les mécanismes sur lesquels repose la dynamique sont complètement déterminés par cette échelle individuelle? Intuitivement, un phénomène d'homogénéisation doit se produire,

mais cette question n'avait, à notre connaissance, jamais été résolue dans la littérature, bien que plusieurs auteurs aient tourné autour. Une approche avec les grandes déviations a été utilisée par Laredo et Rouault [LR83] dans un cas sans interaction mais la distribution d'âge limite n'est pas caractérisée. Athreya et al. [AAI11], Bose et Kaj [BK95] ont considéré ce problème dans un cas où les taux de naissance et de mort ne dépendent pas de l'âge. Ces auteurs montrent alors que la distribution d'âge converge vers une distribution exponentielle. La présence d'interactions et de densité dépendance, en plus de la dépendance en âge des taux, dans nos modèles, rendent difficile le contrôle de la structure d'âge. Dans le cas sans interaction [BK00, DGL02, Dyn91, FVW02, KS98, Wan77], où la durée de vie des individus suit une distribution donnée indépendante des traits, ou est gouvernée par une fonctionnelle continue additive, on peut exploiter les équations satisfaites par les transformées de Laplace. Cela n'est plus possible en présence d'interactions.

Une autre difficulté est le renouvellement lié au fait que les nouveaux individus ont un âge "remis à zéro". Les interactions empêchent de définir pour chaque individu sa durée de vie *a priori*, et de considérer le temps qui lui reste à vivre. Dans [ES07], la quantité de substances néfastes contenues dans des cellules en division (et à nouveau sans interaction) est vue comme un âge. Cependant, ces âges ne sont pas réinitialisés, mais plutôt répartis aléatoirement entre les deux cellules filles.

On considère la population décrite par la mesure (1.1.1). La taille de la population initiale est toujours d'ordre  $K$  et les individus ont une biomasse en  $1/K$ . L'accélération du temps peut correspondre à la considération de démographies allométriques, avec des taux de naissance et de mort d'ordre  $K$ , de sorte que les temps de gestation et les durées de vie soient du même ordre que la biomasse de l'individu, c'est-à-dire  $1/K$  :

$$b^K(a) = K r(a) + \beta(a), \quad d^K(a, X_t^K) = K r(a) + \eta \langle X_t^K, 1 \rangle.$$

Le terme d'ordre  $K$  est le même dans le taux de naissance et dans le taux de mort, de façon à maintenir la balance démographique sans avoir une explosion des naissances ou des morts lorsque  $K \rightarrow +\infty$ . Nous supposons que la fonction  $r(a)$  est positive, continue, bornée et telle que

$$\int_0^{+\infty} r(a) da = +\infty \tag{1.1.29}$$

afin que les durées de vie soient finies presque-sûrement. Pour simplifier, nous considérons ici une compétition logistique  $\eta \langle X_t^K, 1 \rangle$  où  $\eta \geq 0$ . Le cas général est traité dans [MT12b]. Le terme  $\beta(a)$ , supposé positif continu et borné, peut être vu comme le taux de croissance de la population en l'absence d'interaction. Comme dit précédemment, la vitesse de vieillissement est  $K$  de sorte que l'âge à  $t$  d'un individu né au temps  $c$  est  $a = K(t - c)$ . Il est donc difficile de conserver la trace d'un individu et de passer à la limite.

Comme pour la section précédente, nous aurons besoin d'estimées sur les moments de la masse pour obtenir la tension et la convergence de  $(X_t^K)_{K \in \mathbb{N}^*}$ . Partant de (1.1.3) et en suivant la preuve dans [CFM08], nous pouvons montrer que :

**Proposition 1.1.13.** *Si  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ , alors pour tout  $T > 0$ ,*

$$\sup_{K \in \mathbb{N}^*} \sup_{t \leq T} \mathbb{E}(\langle X_t^K, 1 \rangle^3) < +\infty \quad \text{and} \quad \sup_{K \in \mathbb{N}^*} \mathbb{E}(\sup_{t \leq T} \langle X_t^K, 1 \rangle^2) < +\infty. \tag{1.1.30}$$

et pour toute fonction test  $f \in \mathcal{C}_b^1(\mathbb{R}_+, \mathbb{R})$  une fonction test,

$$M_t^{f,K} = \langle X_t^K, f \rangle - \langle X_0^K, f \rangle - \int_0^t \langle X_s, K f'(\cdot) + (K r(\cdot) + \beta(\cdot)) f(0) - (K r(\cdot) + \eta \langle X_s^K, 1 \rangle) f(\cdot) \rangle ds, \tag{1.1.31}$$

est une martingale de carré intégrable, issue de 0, et dont le crochet est :

$$\langle M^{f,K} \rangle_t = \int_0^t \left( \langle X_s^K, r(\cdot) f^2(0) + r(\cdot) f^2(\cdot) + \frac{\beta(\cdot) f^2(0) + \eta \langle X_s^K, 1 \rangle f^2(\cdot)}{K} \rangle \right) ds. \quad (1.1.32)$$

Dans (1.1.31), on voit apparaître les différentes échelles de temps : la vitesse de vieillissement et les termes dominants des naissances et des morts sont en  $K$ . Si l'on choisit  $f \equiv 1$ , ces termes d'ordre  $K$  disparaissent : on ne voit pas le vieillissement si l'on ne fait que compter les individus, et on a compensation des termes de naissances et de morts d'ordre  $K$ . Cependant, le terme de croissance  $\beta(a)$  est fonction de l'âge et on a besoin de connaître les distributions d'âge limites pour espérer obtenir une convergence.

**Théorème 1.1.14.** *Supposons qu'en probabilité*

$$\lim_{K \rightarrow +\infty} X_0^K(da) = n_0(a)da, \quad \text{avec} \quad \sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty. \quad (1.1.33)$$

Alors :

(i) la suite de processus  $(\langle X_t^K, 1 \rangle)_{K \geq 1}$  converge en loi vers la diffusion de Feller :

$$\bar{X}_t = \int_{\mathbb{R}_+} n_0(a)da + \int_0^t (\hat{\beta} \bar{X}_s - \eta \bar{X}_s^2) ds + \int_0^t \sqrt{2\hat{r} \bar{X}_s} dB_s \quad (1.1.34)$$

où  $B$  est un mouvement Brownien réel, avec  $\hat{\beta} = \int_{\mathbb{R}_+} \beta(a) \hat{n}(a) da$  et  $\hat{r} = \int_{\mathbb{R}_+} r(a) \hat{n}(a) da$ , où :

$$\hat{n}(a) = \frac{\exp(-\int_0^a r(\alpha) d\alpha)}{\int_0^{+\infty} \exp(-\int_0^a r(\alpha) d\alpha) da}. \quad (1.1.35)$$

(ii) pour tout  $t > 0$ , la suite de mesures  $(X_t^K(da))_{K \geq 1}$  converge en loi vers  $\bar{X}_t \hat{n}(a) da$ .

Le Théorème 1.1.14 met en évidence un phénomène d'homogénéisation pour la distribution en âge, qui est la composante rapide. Pour obtenir cette homogénéisation, nous utilisons une idée de Kurtz [Kur92] qui se fonde sur le problème de martingale (1.1.31)-(1.1.32) (voir aussi Ball et al. [BKPR06] ou Gupta et al. [GMT14]). Comme les distributions d'âge fluctuent de plus en plus vite et qu'on ne peut plus les suivre dans le temps, l'idée de Kurtz est d'oublier la dimension "processus" et de considérer  $(X_t^K(da))_{t \in [0, T]}$  plutôt comme une mesure  $\Gamma_t^K(da, dt) := X_t^K(da) dt$  sur  $\mathcal{M}_F(\mathbb{R}_+ \times \mathbb{R}_+)$ . On va chercher à montrer la convergence de cette mesure pour identifier la distribution d'âge limite. Cette distribution se stabilise au voisinage de son équilibre non-trivial  $\hat{n}(a)$ , qui est solution stationnaire de l'équation de McKendrick-Von Foerster, exhibée à la Proposition 1.1.12. La composante lente, la taille de la population, est décrite par une diffusion de Feller dont les coefficients ont été moyennés par rapport à  $\hat{n}(a)$ .

**Proposition 1.1.15.** *Pour la diffusion de Feller (1.1.34) :*

(i) si  $\eta > 0$ , on a extinction presque-sûre en temps fini.

(ii) si  $\eta = 0$ ,  $\mathbb{P}(\exists t \in \mathbb{R}_+, \bar{X}_t = 0) = \exp(-\hat{\beta} \int_0^{+\infty} n_0(a) da / \hat{r})$ .

*Démonstration.* Remarquons que 0 est un point absorbant. Lorsque  $\eta > 0$ , nous pouvons prouver l'extinction presque-sûre en exhibant une fonction de Lyapounov (cf. [IW89, p.235] ou [MT12b]). Pour  $\eta = 0$ , nous pouvons adapter le calcul de [IW89, p.236] :

$$\mathbb{P}(\bar{X}_t = 0) = \exp\left(-\frac{e^{\hat{\beta}t} \int_0^{+\infty} n_0(a) da}{\hat{r}(e^{\hat{\beta}t} - 1)/\hat{\beta}}\right).$$

Lorsque  $t \rightarrow +\infty$ , nous obtenons le résultat annoncé.  $\square$

*Preuve du Théorème 1.1.14.* La preuve du Théorème 1.1.14 se divise en plusieurs étapes.

**Étape 1 :** Comme expliqué précédemment, en choisissant  $f \equiv 1$ , les termes d'ordre  $K$  dans (1.1.31) disparaissent ou se compensent. Ainsi,

$$M_t^{1,K} = \langle X_t^K, 1 \rangle - \langle X_0^K, 1 \rangle - \int_0^t \left( \langle X_s^K, \beta \rangle - \eta \langle X_s^K, 1 \rangle^2 \right) ds \quad (1.1.36)$$

définit une martingale locale issue de 0 et de crochet

$$\langle M^{1,K} \rangle_t = \int_0^t \left( 2 \langle X_s^K, r \rangle + \frac{\langle X_s^K, \beta \rangle + \eta \langle X_s^K, 1 \rangle^2}{K} \right) ds.$$

Grâce à (1.1.30), la famille de variables aléatoires  $(\langle X_t^K, 1 \rangle)_{K \geq 1, t \in [0, T]}$  est tendue. De plus, (1.1.30) et le critère d'Aldous-Rebolledo nous fournissent la tension de la suite des lois de  $(\langle X_t^K, 1 \rangle)_{K \geq 1}$  dans  $\mathbb{D}([0, T], \mathbb{R}_+)$ .

**Remarque 1.1.16.** Dans le cas où l'on a également une structure en trait [MT09], on utilise, pour l'étape 1, des fonctions tests qui ne dépendent que du trait et on montre, en utilisant le Théorème 1.1.8 et la Remarque 1.1.10, la tension de la famille des marginales en trait.

**Étape 2 :** Montrons maintenant la tension uniforme des lois de  $(X_t^K)_{K \geq 1, t \in [0, T]}$  dans  $\mathcal{M}_F(\mathbb{R}_+)$ . En modifiant les indicatrices dans (1.1.3), il est possible de coupler, pour chaque  $K$ , le processus  $X^K$  avec un processus  $Y^K$  sans compétition, où les individus se reproduisent au taux  $K r(a) + \beta(a)$  et meurent au taux  $K r(a)$ . Les individus de  $X^K$  sont présents dans  $Y^K$  avec des durées de vie indépendantes et plus longues. La durée de vie de l'individu  $i$  vivant au temps  $t$  et né après les temps 0 est ainsi dominée par une variable aléatoire  $D_i^K(t)$  dont la loi est caractérisée par sa fonction de survie  $S^K(\ell) = \exp\left(-\int_0^\ell K r(Ku) du\right)$ . Comme la vitesse de vieillissement est  $K$ , la probabilité qu'un individu vive jusqu'à l'âge  $\alpha$  est  $S^K(\alpha/K) = \exp\left(-\int_0^\alpha r(a) da\right)$ . Pour l'individu  $i$  au temps 0, le temps restant à vivre peut être dominé par une variable aléatoire  $\Delta_i^K$  telle que conditionnellement au temps 0 :

$$\mathbb{P}\left(\Delta_i^K > \ell\right) = \exp\left(-\int_0^\ell K r(A_i(0) + Ku) du\right).$$

Alors, pour  $\alpha, \varepsilon$  et  $n > 0$  :

$$\begin{aligned} \mathbb{P}(X_t^K([0, \alpha]^c) > \varepsilon) &\leq \mathbb{P}\left(\sum_{i=1}^{N_t^K} \mathbb{1}_{A_i(t) \geq \alpha} > K\varepsilon, N_t^K \leq nK\right) + \mathbb{P}(N_t^K > nK) \\ &\leq \mathbb{P}\left(\sum_{i=1}^{nK} \mathbb{1}_{D_i(t) \geq \alpha/K} > \frac{K\varepsilon}{2}\right) + \frac{\mathbb{E}(\langle X_t^K, 1 \rangle)}{n} + \mathbb{P}\left(\sum_{i=1}^{N_0^K} \mathbb{1}_{A_i(0) + K\Delta_i^K \geq \alpha} > \frac{K\varepsilon}{2}\right). \end{aligned} \quad (1.1.37)$$

Le second terme est majoré par  $\varepsilon/3$  pour  $n$  suffisamment grand, grâce à (1.1.30). Pour le premier terme, on utilise l'inégalité de Bernstein qui nous fournit que (cf *e.g.* [HKPT87, p.241]) :

$$\mathbb{P}\left(\sum_{i=1}^{nK} \mathbb{1}_{D_i(t) \geq \alpha/K} > \frac{K\varepsilon}{2}\right) \leq \exp\left(-\frac{K^2(\varepsilon/2 - nS^K(\alpha/K))^2}{2\left(\frac{nK}{4} + \frac{2K(\varepsilon/2 - nS^K(\alpha/K))}{3}\right)}\right)$$

On choisit  $\alpha = \alpha(n)$  suffisamment grand, de sorte que  $(\varepsilon/2 - nS^K(\alpha/K)) > \varepsilon/4$ . Il est alors possible de trouver  $K_0 = K_0(\alpha, n)$  suffisamment grand tel que le membre de droite soit borné par

$\varepsilon/3$  pour tout  $K \geq K_0$ . Pour le 3ème terme de (1.1.37)

$$\begin{aligned} & \mathbb{P}\left(\sum_{i=1}^{N_0^K} \mathbb{1}_{A_i(0)+K\Delta_i^K \geq \alpha} > \frac{K\varepsilon}{2} \mid X_0^K\right) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{N_0^K} \left\{ \mathbb{1}_{A_i(0)+K\Delta_i^K \geq \alpha} - 1 \wedge e^{-\int_{A_i(0)}^{\alpha} r(a)da} \right\} > K\left\{ \frac{\varepsilon}{2} - \int_{\mathbb{R}_+} 1 \wedge e^{-\int_a^{\alpha} r(u)du} X_0^K(da) \right\} \mid X_0^K\right) \end{aligned} \quad (1.1.38)$$

Soit  $\phi(\alpha) \leq \alpha$  telle que  $\phi(\alpha)$  et  $\int_{\phi(\alpha)}^{\alpha} r(a)da$  tendent vers  $+\infty$  lorsque  $\alpha \rightarrow +\infty$ . Comme on a la limite suivante en probabilité

$$\begin{aligned} \lim_{K \rightarrow +\infty} \int_{\mathbb{R}_+} 1 \wedge e^{-\int_a^{\alpha} r(u)du} X_0^K(da) &= \int_{\mathbb{R}_+} 1 \wedge e^{-\int_a^{\alpha} r(u)du} m_0(a)da \\ &\leq \int_{\mathbb{R}_+} \left( \mathbb{1}_{[\phi(\alpha), +\infty)}(a) + e^{-\int_{\phi(\alpha)}^{\alpha} r(u)du} \right) m_0(a)da \end{aligned}$$

où le membre de droite tend vers 0 lorsque  $\alpha \rightarrow +\infty$ , alors pour  $\alpha$  et  $K \geq K_1 = K_1(\alpha)$  suffisamment grands, nous pouvons majorer le membre de droite de (1.1.38) par  $\varepsilon/3$  en utilisant l'inégalité de Bernstein et en procédant comme pour le terme précédent. Ceci conclut la preuve. Remarquons que le majorant obtenu ne dépend pas de  $t \in [0, T]$ . Ainsi, les choix de  $\alpha$  et  $K_0$  sont indépendants de  $t \in [0, T]$ .

**Etape 3 :** Nous considérons maintenant la composante rapide dans  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{R}_+))$ . En suivant les idées de Kurtz [Kur92], nous oublions l'aspect processus et considérons plutôt les mesures  $(\Gamma^K)_{K \geq 1}$  in  $\mathcal{M}_F([0, T] \times \mathbb{R}_+)$ . Alors, pour  $\alpha > 0$  assez grand,

$$\begin{aligned} \Gamma^K(([0, T] \times \mathbb{R}_+) \setminus ([0, T] \times [0, \alpha])) &= \mathbb{E}\left(\int_0^T X_t^K([0, \alpha]^c)dt\right) \\ &\leq \int_0^T \left( \varepsilon \mathbb{P}(X_t^K([0, \alpha]^c) \leq \varepsilon) + \mathbb{E}(\langle X_t^K, 1 \rangle \mathbb{1}_{X_t^K([0, \alpha]^c) > \varepsilon}) \right) dt \\ &\leq T\varepsilon + \int_0^T \mathbb{E}(\langle X_t^K, 1 \rangle^2)^{1/2} \mathbb{P}(X_t^K([0, \alpha]^c) > \varepsilon)^{1/2} dt \leq T\varepsilon + C(T)\sqrt{\varepsilon}. \end{aligned}$$

Ceci prouve que la suite  $(\Gamma^K)_{K \geq 1}$  est tendue dans  $\mathcal{M}_F([0, T] \times \mathbb{R}_+)$ .

**Etape 4 :** Par les résultats précédents, la suite des lois de  $(\Gamma^K, \langle X^K, 1 \rangle)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathcal{M}_F([0, T] \times \mathbb{R}_+) \times \mathbb{D}([0, T], \mathbb{R}_+)$  puisque chacune de ses composantes est tendue. On peut en extraire une sous-suite qui converge en loi vers une valeur d'adhérence dans  $\mathcal{M}_F([0, T] \times \mathbb{R}_+) \times \mathcal{C}([0, T], \mathbb{R})$ , que l'on note  $(\Gamma, \bar{X})$ .

Soit  $\varphi(s)$  une fonction continue bornée. Par définition de  $\Gamma^K$  nous avons pour tout  $t \in [0, T]$  :

$$\int_0^t \int_{\mathbb{R}_+} \varphi(s) \Gamma^K(da, ds) = \int_0^t \varphi(s) \langle X_s^K, 1 \rangle ds.$$

En faisant  $K \rightarrow +\infty$ , nous obtenons que  $\bar{X}_s ds$  est la mesure marginale de  $\Gamma(da, ds)$ . Il existe alors une probabilité de transition  $\gamma_s(da)$  telle que  $\Gamma(da, ds) = \gamma_s(da) \bar{X}_s ds$ . En divisant (1.1.31) par  $K$ ,

$$\begin{aligned} \frac{M_t^{f,K}}{K} &= A_t^K + B_t^K \\ A_t^K &= \frac{\langle X_t^K, f \rangle - \langle X_0^K, f \rangle}{K} + \int_0^t \int_{\mathbb{R}_+} \left( \frac{\beta(a)}{K} f(0) - \frac{\eta \langle X_s^K, 1 \rangle}{K} f(a) \right) X_s^K(da) ds \\ B_t^K &= \int_0^t \int_{\mathbb{R}_+} (f'(a) + r(a)(f(0) - f(a))) X_s^K(da) ds. \end{aligned}$$

Par (1.1.30),  $\mathbb{E}(\sup_{t \leq T} A_t^K) \rightarrow 0$  quand  $K \rightarrow +\infty$ . Comme  $\Xi \mapsto \int_0^t \int_{\mathbb{R}_+} (f'(a) + r(a)(f(0) - f(a))) \Xi(da, ds)$  est continue en  $\Xi = \Gamma$ , le dernier terme converge en loi pour tout  $t \in [0, T]$  vers

$$\bar{M}_t^f = \int_0^t \int_{\mathbb{R}_+} (f'(a) + r(a)(f(0) - f(a))) \bar{X}_s \gamma_s(da) ds. \quad (1.1.39)$$

Soit  $H_s \in \mathcal{F}_s$  une variable aléatoire bornée. Comme  $M^{f,K}$  est une martingale de carré intégrable, la famille  $(M_t^{f,K}/K)_{K \geq 1}$  est uniformément intégrable pour tout  $t \in [0, T]$ . Alors :

$$\mathbb{E}(\bar{M}_t^f H_s) = \mathbb{E}\left(\lim_{K \rightarrow +\infty} \frac{M_t^{f,K}}{K} H_s\right) = \lim_{K \rightarrow +\infty} \mathbb{E}\left(\frac{M_t^{f,K}}{K} H_s\right) = \lim_{K \rightarrow +\infty} \mathbb{E}\left(\frac{M_s^{f,K}}{K} H_s\right) = \mathbb{E}(\bar{M}_s^f H_s).$$

Ceci montre que  $\bar{M}^f$  est une martingale, mais par (1.1.39) c'est également un processus continu à variations finies. Ainsi  $\bar{M}^f$  est indistinguable du processus nul. Alors, presque-sûrement et  $dt$ -presque partout :

$$\int_{\mathbb{R}_+} (f'(a) + r(a)(f(0) - f(a))) \bar{X}_t \gamma_t(da) = 0. \quad (1.1.40)$$

En choisissant une fonction test de la forme  $f(a) = \varphi(0) + \int_0^a \varphi(\alpha) d\alpha$ , pour une fonction  $\varphi$  continue bornée, on obtient :

$$\bar{X}_t \langle \gamma_t, \varphi \rangle = \bar{X}_t \int_{\mathbb{R}_+} \varphi(\alpha) \left( \int_{\alpha}^{+\infty} r(\alpha) \gamma_t(da) \right) d\alpha.$$

Si  $\bar{X}_t = 0$ , alors  $\Gamma(da, ds) = 0$ . Sur l'événement  $\{\bar{X}_t > 0\}$ , l'égalité précédente nous indique que  $\gamma_t(da)$  est absolument continue par rapport à la mesure de Lebesgue avec comme densité  $\hat{n}_t(a) = \int_a^{+\infty} r(\alpha) \hat{n}_t(\alpha) d\alpha$ . La fonction  $a \mapsto \hat{n}_t(a)$  est alors infiniment différentiable et résout l'équation  $\partial_a \hat{n}_t(a) = -r(a) \hat{n}_t(a)$  qui caractérise la solution stationnaire de l'EDP de McKendrick-von Foerster, que nous avons vue en Proposition 1.1.12. Nous obtenons donc (1.1.35) et toute valeur d'adhérence de  $(\Gamma^K)_{K \geq 1}$  est donc nécessairement de la forme  $\bar{X}_t \hat{n}(a) da$ . Il nous reste donc à identifier  $\bar{X}$  comme l'unique solution forte de (1.1.31) pour montrer que la suite  $(\Gamma^K)_{K \geq 1}$  n'a qu'une seule valeur d'adhérence vers laquelle elle converge en loi.

**Etape 5 :** Soit  $n \in \mathbb{N}^*$ , et soient  $t_1 < \dots < t_n < s < t$  et  $\psi_1, \dots, \psi_n$  des fonctions continues bornées. Considérons la fonctionnelle :

$$\Psi_{s,t}(Y) = \left[ Y_t - Y_s - \int_s^t (\hat{\beta} Y_s - \eta Y_s^2) ds \right] \psi_1(Y_{t_1}) \dots \psi_n(Y_{t_n}).$$

Comme les valeurs d'adhérence de  $(\langle X^K, 1 \rangle)_{K \geq 1}$  sont nécessairement continues,  $\Psi$  est continue en  $\bar{X}$ . Nous avons donc :

$$\begin{aligned} \Psi_{s,t}(\langle X^K, 1 \rangle) &= \psi_1(\langle X_{t_1}^K, 1 \rangle) \dots \psi_n(\langle X_{t_n}^K, 1 \rangle) \left[ A_{s,t}^K + B_{s,t}^K \right] \\ A_{s,t}^K &= \int_s^t (\langle X_s^K, \beta \rangle - \hat{\beta} \bar{X}_s) ds, \quad B_{s,t}^K = \int_s^t \eta (\langle X_s^K, 1 \rangle^2 - \bar{X}_s^2) ds. \end{aligned}$$

Comme  $\beta(a)$  est continue et comme  $\Gamma^K$  converge vers  $\hat{n}(a) da \bar{X}_s ds$ ,  $A_{s,t}^K$  converge vers zéro. Grâce à (1.1.30),  $\mathbb{E}(|A_{s,t}^K|) \rightarrow 0$  lorsque  $K \rightarrow +\infty$ . Un résultat similaire est satisfait pour  $E(|B_{s,t}^K|)$  car  $\bar{X}$  est valeur d'adhérence de  $\langle X^K, 1 \rangle$ . Nous en déduisons que pour tous  $s$  et  $t$ ,  $\mathbb{E}(|\Psi_{s,t}(\langle X^K, 1 \rangle)|) \rightarrow 0$  et pour tout  $t$ ,  $M_t = \bar{X}_t - \int_{\mathbb{R}_+} n_0(a) da - \int_0^t (\hat{\beta} \bar{X}_s - \eta \bar{X}_s^2) ds$  est une martingale. Le calcul de son crochet est classique, en appliquant la formule d'Itô à  $M$  et à (1.1.3), ce qui nous fournit par identification  $\langle M \rangle_t = 2\hat{r} \int_0^t \bar{X}_s ds$ . Il est alors classique de montrer que sur un espace de probabilité élargi, il existe un mouvement Brownien  $(B_t)_{t \in [0, T]}$  tel que  $M_t = \int_0^t \sqrt{2\hat{r} \bar{X}_s} dB_s$ . Ainsi,  $\bar{X}$  résout l'équation de Feller (1.1.34), pour laquelle on a l'unicité trajectorielle des solutions, pour une condition initiale et un mouvement Brownien  $B$  donnés.  $\square$

## 1.2 Evolution de généalogies

### 1.2.1 Processus historiques

#### 1.2.1.1 Définitions

Les structures d'âges - âge physique (temps depuis la naissance), mais aussi taille, âge biologique (stade de maturation intrinsèque), âge d'une maladie etc. - sont une façon de prendre en compte la dépendance de la dynamique vis-à-vis du passé des individus. Avec Sylvie Méléard, nous avons souhaité généraliser cette dépendance [MT12a]. Nous étudions des populations structurées par un trait  $x \in \mathbb{R}$  (pour simplifier la présentation) transmis héréditairement sauf lorsqu'il y a mutation, et dont la dynamique dépend de la trajectoire passée du processus au lieu de passer par des quantités intermédiaires, les âges, qui refléteraient les événements du passé. Comme pour les sections précédentes, l'évolution est en temps continu et résulte de l'occurrence des morts et reproductions (asexuées).

On associe ainsi à chaque individu  $i$  au temps  $t$  sa lignée ancestrale  $(y_i(s))_{s \in [0,t]}$  qui est un processus càdlàg associant à chaque temps  $s \in [0,t)$  la valeur du trait de son plus proche ancêtre (en terme de générations) vivant au temps  $s$ . On peut par convention prolonger la trajectoire ancestrale de l'individu  $i$  à  $t$ , de trait  $x_i(t)$ , par  $y_i(s) = x_i(t)$ , pour  $s \geq t$ . La trajectoire  $y_i$  ainsi obtenue est un processus de sauts pur, constant après  $t$ . Plus précisément, si l'individu  $i$  vivant à  $t$  et né au temps  $S_m$  a  $m-1$  ancêtres nés après le temps 0, aux instants  $S_1 = 0 < S_2 < \dots < S_{m-1}$  et de traits  $(x_1, x_2, \dots, x_{m-1})$ , alors

$$y_t = \sum_{j=1}^{m-1} x_j \mathbf{1}_{S_j \leq t < S_{j+1}} + x \mathbf{1}_{S_m \leq t}. \quad (1.2.1)$$

On définit par  $\mathcal{L}$  l'ensemble de toutes les lignées ancestrales possibles, de la forme (1.2.1).

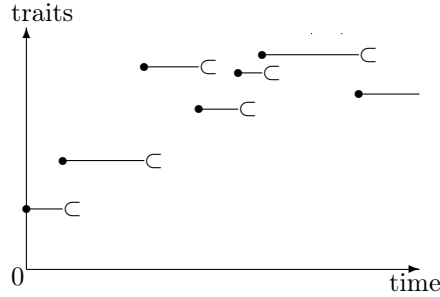


FIGURE 1.2.1 – Exemple de trajectoire ancestrale.

**Notation :** pour une fonction  $y \in \mathbb{D}(\mathbb{R}_+, \mathbb{R})$  et pour  $t > 0$ , on notera  $y^t$  la trajectoire stoppée au temps  $t$ , définie pour tout  $s \in \mathbb{R}_+$  par  $y^t(s) = y(s \wedge t)$  et par  $y^{t-}$  la fonction définie par  $y^{t-}(s) = \lim_{r \uparrow t} y^r(s)$ . Nous écrirons parfois  $y_t = y(t)$  pour la valeur de la fonction au temps  $t$ . Pour deux trajectoires  $y, w \in \mathbb{D}(\mathbb{R}_+, \mathbb{R})$  et  $t \in \mathbb{R}_+$ , on définit  $(y|t|w) \in \mathbb{D}(\mathbb{R}_+, \mathbb{R})$  la trajectoire obtenue en concaténant  $y$  et  $w$  au temps  $t$  :

$$(y|t|w) = \begin{cases} y_u & \text{if } u < t \\ w_{u-t} & \text{if } u \geq t. \end{cases} \quad (1.2.2)$$

Lorsque  $w$  est une fonction constante, avec  $\forall u \in \mathbb{R}_+, w_u = x$ , on notera avec un abus de notation  $(y|t|x)$ .

L'ensemble des trajectoires ancestrales des individus vivants au temps  $t$  est décrit par :

$$X_t^K := \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{y_i^{\wedge t}} \in \mathcal{M}_F(\mathcal{L}) \subset \mathcal{M}_F(\mathbb{D}(\mathbb{R}_+, \mathbb{R})), \quad (1.2.3)$$

où, comme dans les sections précédentes,  $K$  est un paramètre représentant la capacité de charge, et que l'on va faire tendre vers l'infini. On étudie la convergence de la suite de processus  $(X^K)_{K \in \mathbb{N}^*}$ , dans une échelle de temps diffusive comme à la Section 1.1.4, en suivant le schéma de tension-unicité proposé par Dawson and Perkins [DP91] pour le super-Brownien historique (sans interaction). En présence de densité-dépendance, la propriété de branchement n'est plus valable et on ne peut pas utiliser d'argument fondé sur les transformées de Laplace. On considère un problème de martingale infini-dimensionnel. Les processus que nous étudions sont discontinus et à valeurs dans l'espace des mesures sur les trajectoires càdlàg, ce qui nécessite de définir de nouvelles classes de fonctions test. Nous obtenons à la limite un superprocessus historique non-linéaire qui généralise les travaux de Perkins [Per95]. La dynamique des lignées peut prendre en compte une dépendance par rapport au passé et de la compétition.

L'évolution des généalogies a fait l'objet de beaucoup de travaux. En l'absence d'interaction, on peut se référer à la théorie des processus coalescents, décrivant les généalogies dans le sens indirect du temps (voir Berestycki [Ber09] pour une présentation). Donnelly et Kurtz [DK99a, DK99b] étudient des modèles généalogiques et des équations de diffusion (en particulier, le processus de Fleming-Viot). Les processus MMMS (*Marked Measure Metric Space*) introduits par Greven et al. [GPW11] et Depperschmidt et al. [DGP11] fournissent un cadre général pour décrire l'évolution de généalogies, en particulier celles issues de processus de coalescence (par exemple les  $\Lambda$ -coalescents). Plus récemment, Barton, Etheridge, Véber et leurs co-auteurs [BEV10, EV12, BEV13, EVY14] étudient l'évolution des généalogies dans une version spatiale du processus  $\Lambda$ -Fleming-Viot et en considèrent plusieurs limites. Tous ces modèles prennent en compte de la sélection et des mutations (voir aussi e.g. [KN97, BE04, EPW06]) mais pas de compétition entre les individus. Récemment, suite à notre travail avec Sylvie Méléard, Kliem et Winter [KW14] ont introduit un processus MMMS avec densité dépendance et taille de population variable. La formulation par des processus historiques fournit une description de l'évolution des généalogies dans le sens direct du temps, et permet de prendre en compte de la compétition et des tailles de population variables, donc également de possibles extinctions.

Un individu de trajectoire ancestrale  $y$  au temps  $t$  se reproduit au taux

$$b^K(t, y) = K r(t, y) + b(t, y),$$

où les fonctions  $b$  et  $r$  sont continues positives sur  $\mathbb{R}_+ \times \mathbb{D}_{\mathbb{R}}$ . Un exemple de fonction  $b$  est :

$$b(t, y) = B \left( \int_0^t y_{t-s} \nu_b(ds) \right) \quad (1.2.4)$$

avec  $B$  une fonction continue bornée par  $\bar{B} > 0$  et  $\nu_b$  une mesure de Radon sur  $\mathbb{R}_+$ . De même, la fonction allométrique  $r$  peut-être prise de la forme (1.2.4) en remplaçant  $B$  par  $R$ , une fonction Lipschitzienne, bornée inférieurement et supérieurement par  $\underline{R} > 0$  et  $\bar{R} > 0$ , et  $\nu_b$  par  $\nu_r$ , une autre mesure de Radon sur  $\mathbb{R}_+$ .

Cette formulation avec des mesures de Radon englobe plusieurs modèles. Si  $\nu_b(ds) = \delta_0(ds)$ , alors  $\int_0^t y_{t-s} \nu_b(ds) = y_t$  est le trait de l'individu de la lignée  $y$  qui vivait au temps  $t$ . Si  $\nu_b(ds) = e^{-\alpha s} ds$ , avec  $\alpha > 0$ , alors  $\int_0^t y_{t-s} \nu_b(ds) = \int_0^t e^{-\alpha(t-s)} y_s ds$ . Ceci permet de modéliser l'influence des ancêtres passés, d'autant plus forte que ces ancêtres sont récents. De tels termes sont utiles pour représenter des interactions sociales telles que les liens de parentèles, où les ancêtres contribuent à protéger



et élever leurs descendants indirects. Enfin, en étendant l'espace des traits de façon à introduire des couleurs, on peut retrouver les dépendances par rapport à l'âge physique étudiées aux sections précédentes (voir [MT09]).

Lors de la reproduction, le nouvel individu est soit un clone avec probabilité  $1 - p \in [0, 1]$ , de même trait  $y_{t-}$  et de même trajectoire ancestrale  $y$ , ou un mutant avec probabilité  $p \in [0, 1]$ . Dans ce cas, le nouveau trait est  $y_{t-} + h$ , où  $h$  est tiré dans la distribution  $k^K(h) dh$ , que l'on supposera dans la suite être une gaussienne centrée de variance  $\sigma^2/K$  par souci de simplicité. La lignée ancestrale du nouvel individu est alors  $(y|t|y_{t-} + h)$ . On posera dans la suite  $M^K(dh) = pk^K(h)dh + (1-p)\delta_0(dh)$ .

Un individu de trajectoire ancestrale  $y$  au temps  $t$ , dans la population représentée par le processus  $X \in \mathbb{D}_{\mathcal{M}_F(\mathbb{D}_{\mathbb{R}})}$ , meurt au taux

$$d^K(t, y, X) = K r(t, y) + d(t, y, X),$$

où

$$d(t, y, X) = D(t, y) + \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} U(t, y, y') X_{t-s}(dy') \nu_d(ds), \quad (1.2.5)$$

avec le noyau d'interaction  $U \in \mathcal{C}_b(\mathbb{R}_+ \times \mathbb{D}_{\mathbb{R}}^2, \mathbb{R}_+)$  (borné par  $\bar{U}$ ), le taux de mort naturelle  $D \in \mathcal{C}_b(\mathbb{R}_+ \times \mathbb{D}_{\mathbb{R}}, \mathbb{R}_+)$  (borné par  $\bar{D}$ ) et  $\nu_d$  une mesure de Radon.

On remarque que la partie d'ordre  $K$ , dans les taux de naissance et de mort, fait intervenir la même fonction allométrique  $r$ , pour préserver la balance démographique. Dans la suite, on notera  $\gamma(t, y, \bar{X}^t)$  le taux de croissance des individus  $y$  à  $t$  dans l'environnement décrit par  $\bar{X}$  :

$$\gamma(t, y, \bar{X}^t) = b(t, y) - d(t, y, \bar{X}^t). \quad (1.2.6)$$

De même que dans la Section 1.1.1, on peut écrire le processus  $X^K$ , pour  $K$  fixé, comme solution d'une EDS dirigée par une mesure ponctuelle de Poisson. On peut montrer l'existence et l'unicité forte d'une solution, pour une condition initiale de masse finie en espérance et une mesure de Poisson données.

**Proposition 1.2.1.** *Supposons  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ . Alors,*

(i) *Pour tout  $T > 0$ ,*

$$\sup_{K \in \mathbb{N}^*} \mathbb{E} \left( \sup_{t \in [0, T]} \langle X_t^K, 1 \rangle^3 \right) < +\infty. \quad (1.2.7)$$

(ii) *Pour toute fonction mesurable bornée  $\varphi$ ,*

$$\begin{aligned} \langle X_t^K, \varphi \rangle &= \langle X_0^K, \varphi \rangle + M_t^{K, \varphi} + \int_0^t ds \int_{\mathbb{D}_{\mathbb{R}}} X_s^K(dy) \left[ \right. \\ &\quad \left. K r(s, y) \left( \int_{\mathbb{R}^d} \varphi(y|s|y_s + h) M^K(y_s, dh) - \varphi(y) \right) \right. \\ &\quad \left. + b(s, y) \int_{\mathbb{R}^d} \varphi(y|s|y_s + h) M^K(y_s, dh) - d(s, y, (X^K)^s) \varphi(y) \right] \end{aligned} \quad (1.2.8)$$

où  $M^{K, \varphi}$  est une martingale de carré intégrable issue de 0 et de variation quadratique prévisible

$$\begin{aligned} \langle M^{K, \varphi} \rangle_t &= \frac{1}{K} \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} \left[ (K r(s, y) + b(s, y)) \int_{\mathbb{R}^d} \varphi^2(y|s|y_s + h) M^K(y_s, dh) \right. \\ &\quad \left. + (K r(s, y) + d(s, y, (X^K)^s)) \varphi^2(y) \right] X_s^K(dy) ds. \end{aligned} \quad (1.2.9)$$

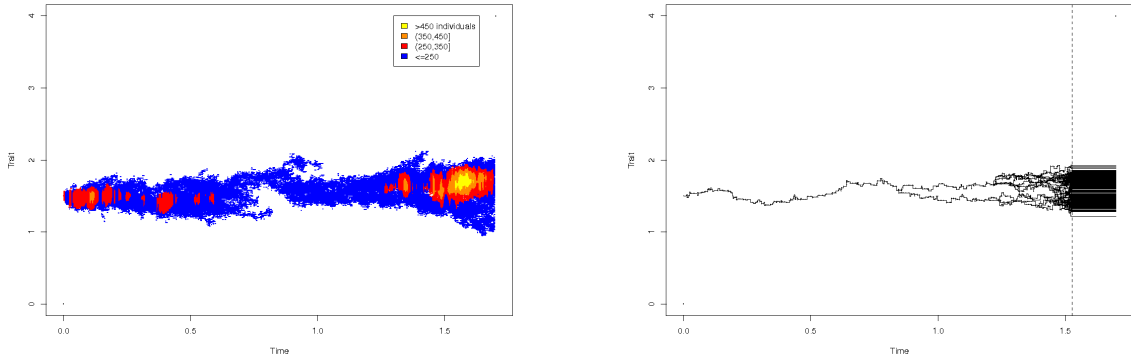


FIGURE 1.2.2 – *Modèle de Dieckmann-Doebeli (cf. [CM11]). Le trait est  $x \in [0, 4]$  et les taux sont les suivants :  $r(t, y) = 1$ ,  $b(t, y) = \exp\left(-\frac{(y_t - 2)^2}{2\sigma_b^2}\right)$ ,  $d(t, y, X) = \int_{\mathbb{D}} \exp\left(-\frac{(y_t - y'_t)^2}{2\sigma_V^2}\right) X(dy')$  avec  $\sigma = 0.4$ ,  $\sigma_b = 0.4$ ,  $\sigma_V = 0.3$ ,  $x_0 = 4$ ,  $p = 0.5$ ,  $K = 300$ . On démarre la simulation avec 300 particules de trait 1.5. On montre à droite le processus historique au temps  $t$  (trait vertical en pointillé) et à gauche le support du superprocessus correspondant à la même simulation.*

### 1.2.1.2 Simulations

Les simulations suivantes illustrent l'intérêt des processus historiques, sur un exemple de Doebeli-Dieckmann (Fig. 1.2.2) où il n'y a pas de dépendance au passé dans la dynamique. Les limites des processus non-historiques associés ont été étudiées par Champagnat et al. [CFM08]. Le processus historique ne montre que les trajectoires ancestrales des individus vivant au temps  $t$ , tandis que le support du superprocessus conserve les branches mortes. Ces deux représentations se complètent : le superprocessus perd la dimension individuelle et on ne peut pas retrouver les lignées ancestrales en considérant son support seul. Il nous montre, par contre, la biodiversité de la population à tous les temps passés. Le superprocessus historique révèle les structures généalogiques et souligne qu'au temps  $t$  représenté en Fig. 1.2.2, le continuum de traits observé correspond en fait à deux sous-populations dont le plus récent ancêtre commun est relativement éloigné (vers le temps 0.8). Pour un exemple où la dynamique dépend du passé, nous renvoyons à l'exemple des chèvres de [AT98] (voir aussi [Per95, MT12a]).

### 1.2.1.3 Limite diffusive

Nous considérons ici ce qui se passe sur un intervalle de temps  $[0, T]$  avec  $T > 0$ . Nous commençons par introduire une classe de fonctions test qui serviront à écrire le problème de martingale que vérifie la limite en loi de  $(X^K)_{K \in \mathbb{N}^*}$ .

**Définition 1.2.2.** Pour deux fonctions  $\mathcal{C}_b^2$ ,  $g$  sur  $\mathbb{R}_+ \times \mathbb{R}$  et  $G$  sur  $\mathbb{R}$ , on définit la fonction continue  $G_g$  sur  $\mathbb{D}_{\mathbb{R}}$  par

$$G_g(y) = G\left(\int_0^T g(s, y_s) ds\right). \quad (1.2.10)$$

On peut montrer que la classe engendrée par les combinaisons linéaires finies de telles fonctions est stable par addition et sépare les points. Remarquons également que si  $y$  est une trajectoire

càdlàg arrêtée en  $t \in [0, T]$  alors

$$G_g(y) = G\left(\int_0^t g(s, y_s) ds + \int_t^T g(s, y_t) ds\right).$$

Pour une telle fonction  $G_g$ , nous définissons, pour  $t \in [0, T]$  et  $y \in \mathbb{D}_{\mathbb{R}}$  :

$$\begin{aligned} \mathcal{D}^2 G_g(t, y) &= G'\left(\int_0^T g(s, y_s) ds\right) \int_t^T \Delta_x g(s, y_t) ds \\ &\quad + G''\left(\int_0^T g(s, y_s) ds\right) \sum_{i=1}^d \left(\int_t^T \partial_{x_i} g(s, y_t) ds\right)^2, \end{aligned} \quad (1.2.11)$$

qui généralise le Laplacien dans notre cadre infini-dimensionnel. Pour  $G(x) = x$  et une fonction  $g(s, y) = g(y)$  qui ne dépend pas du temps,  $\mathcal{D}^2 G_g(t, y) = (T - t)\Delta g(y_t)$ .

Remarquons que Dawson [Daw93, p.203] et Etheridge [Eth00, p.24] introduisent une autre classe de fonctions test de la forme

$$\varphi(y) = \prod_{j=1}^m g_j(y_{t_j}), \quad (1.2.12)$$

pour  $m \in \mathbb{N}^*$ ,  $0 \leq t_1 < \dots < t_m$  et  $\forall j \in \{1, \dots, m\}$ ,  $g_j \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$ . Comme nous allons travailler avec des processus càdlàg, ces fonctions ne sont pas continues pour la topologie de Skorokhod et donc peu pratiques. Cependant, nous montrerons qu'à la limite, les trajectoires deviennent continues et le problème de martingale limite peut également s'écrire à l'aide de ces fonctions test.

Si  $y$  est une trajectoire continue arrêtée en  $t$ , alors  $\varphi(y) = \prod_{j=1}^m g_j(y_{t_j \wedge t})$ . Pour  $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ , et  $t > 0$ , on peut alors définir

$$\tilde{\Delta}\varphi(t, y) = \sum_{k=0}^{m-1} \mathbb{1}_{[t_k, t_{k+1}[}(t) \left( \prod_{j=1}^k g_j(y_{t_j}) \Delta \left( \prod_{j=k+1}^m g_j \right)(y_t) \right), \quad (1.2.13)$$

où  $t_0 = 0$ .

Soit  $\varphi$  une fonction test de la forme (1.2.12). Il existe une suite de fonctions test  $(\varphi_q)$  de la forme (1.2.10) telle que pour tous  $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R})$  et  $t \in [0, T]$ , les suites  $(\varphi_q(y))_{q \in \mathbb{N}^*}$  et  $(\mathcal{D}^2 \varphi_q(t, y))_{q \in \mathbb{N}^*}$  sont bornées uniformément en  $q$ ,  $t$  et  $y$ , et convergent vers  $\varphi(y)$  et  $\tilde{\Delta}\varphi(t, y)$  respectivement.

**Théorème 1.2.3.** *Supposons qu'il existe  $\bar{X}_0 \in \mathcal{M}_F(\mathbb{D}_{\mathbb{R}})$ , tel que  $\lim_{K \rightarrow +\infty} X_0^K = \bar{X}_0$  en loi et supposons que  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X_0^K, 1 \rangle^3) < +\infty$ . Alors, la suite  $(X^K)_{K \in \mathbb{N}^*}$  converge en loi dans  $\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  vers le superprocessus  $\bar{X} \in \mathcal{C}([0, T], \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  caractérisé par le problème de martingale suivant. Pour des fonctions test  $G_g$  de la forme (1.2.10) et  $t \in [0, T]$ ,*

$$\begin{aligned} M_t^{G_g} &= \langle \bar{X}_t, G_g \rangle - \langle \bar{X}_0, G_g \rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} \left( pr(s, y) \frac{\sigma^2}{2} \mathcal{D}^2 G_g(s, y) \right. \\ &\quad \left. + \gamma(s, y, \bar{X}^s) G_g(y) \right) \bar{X}_s(dy) ds \end{aligned} \quad (1.2.14)$$

est une martingale de carré intégrable issue de 0 et de variation quadratique prévisible :

$$\langle M^{G_g} \rangle_t = \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} 2r(s, y) G_g^2(y) \bar{X}_s(dy) ds, \quad (1.2.15)$$

*Démonstration.* La preuve proposée dans [MT09] contenait deux erreurs, relevées et corrigées par Kliem [Kli14]. Nous présentons ici une esquisse de la preuve révisée, toujours basée sur des arguments de compacité-unicité.

**Etape 1 :** Nous commençons par montrer que  $(X^K)_{K \in \mathbb{N}^*}$  est tendue. Notons que  $\mathbb{D}_{\mathbb{R}}$  est Polonais et que pour  $\mathcal{K} \subset \mathbb{D}_{\mathbb{R}}$  compact,  $\mathcal{K}_T = \{y^t, y^{t-} \mid y \in \mathcal{K}, t \in [0, T]\}$  est également un compact de  $\mathbb{D}_{\mathbb{R}}$  (cf. Dawson and Perkins [DP91, Lemme 7.6]). Comme la famille de fonctions test (1.2.10) sépare les points et contient l'adhérence des fonctions test (1.2.12), nous pouvons établir :

**Lemme 1.2.4.** *Si :*

(i)  $\forall \varepsilon > 0, \exists \mathcal{K} \subset \mathbb{D}_{\mathbb{R}}$  compact,

$$\sup_{K \in \mathbb{N}^*} \mathbb{P}(\exists t \in [0, T], X_t^K(\mathcal{K}_T^c) > \varepsilon) \leq \varepsilon, \quad (1.2.16)$$

(ii)  $\forall G_g$  de la forme (1.2.10), la famille des lois de  $((X^K, G_g))_{K \in \mathbb{N}^*}$  est tendue dans  $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}))$ , alors, la famille des lois de  $(X^K)_{K \in \mathbb{N}^*}$  est tendue dans  $\mathcal{P}(\mathbb{D}([0, T], \mathcal{M}_F(\mathbb{D}_{\mathbb{R}})))$ .

Le point (ii) se prouve de façon similaire à ce qui a été présenté à l'Etape 1 de la preuve du Théorème 1.1.11. On vérifie le critère d'Aldous-Rebolledo en s'appuyant sur des développements de Taylor et les contrôles de moments (1.2.7).

**Etape 2 :** Pour prouver le point (i), nous commençons par coupler, pour chaque  $K \in \mathbb{N}^*$ , notre processus  $X^K$  avec interaction avec un processus sans interaction, qui le domine. Le processus dominant correspond à une population dont les taux de naissance et de mort sont

$$\bar{b}^K(t, y) = K r(t, y) + b(t, y), \quad \bar{d}^K(t, y, X) = K r(t, y).$$

Il est possible de construire le couplage de façon trajectorielle en donnant des durées de vie plus longues aux individus du nouveau processus qui existent aussi dans  $X^K$ . Remarquons que (1.2.16) est satisfaite dès que le processus sans interaction satisfait la même inégalité, ce que nous allons montrer. Dans les Etapes 2 à 5, nous ne considérerons plus que ce processus sans interaction et par abus de notation, nous l'appellerons également  $X^K$ .

Nous nous sommes débarrassés de la densité-dépendance, mais le fait que la fonction allométrique  $r$  dépende de  $y$  et  $t$  nous empêche d'utiliser directement les résultats de Dawson et Perkins [DP91] qui montrent la convergence d'un système historique de particules branchantes Browniennes vers le super-mouvement Brownien historique.

Soient  $\varepsilon > 0$  et  $\mathcal{K}$  un compact de  $\mathbb{D}_{\mathbb{R}}$ . Nous définissons  $\mathcal{K}^t = \{y^t \mid y \in \mathcal{K}\} \subset \mathbb{D}_{\mathbb{R}}$  l'ensemble des trajectoires de  $\mathcal{K}$  arrêtées en  $t$ . Introduisons le temps d'arrêt suivant :

$$S_\varepsilon^K = \inf\{t \in \mathbb{R}_+ \mid X_t^K(\mathcal{K}_T^c) > \varepsilon\}. \quad (1.2.17)$$

On a alors,

$$\begin{aligned} \mathbb{P}(\exists t \in [0, T], X_t^K(\mathcal{K}_T^c) > \varepsilon) &= \mathbb{P}(S_\varepsilon^K < T) \\ &= \mathbb{P}\left(S_\varepsilon^K < T, X_T^K((\mathcal{K}^T)^c) > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(S_\varepsilon^K < T, X_T^K((\mathcal{K}^T)^c) \leq \frac{\varepsilon}{2}\right) \\ &\leq \frac{2}{\varepsilon} \mathbb{E}(X_T^K((\mathcal{K}^T)^c)) + \mathbb{P}\left(S_\varepsilon^K < T, X_T^K((\mathcal{K}^T)^c) \leq \frac{\varepsilon}{2}\right) \end{aligned} \quad (1.2.18)$$

en introduisant la quantité plus facilement contrôlable  $X_T^K((\mathcal{K}^T)^c)$ . Si nous montrons qu'il existe  $\mathcal{K}$  compact de  $\mathbb{D}_{\mathbb{R}}$  et  $\eta \in (0, 1)$  tels que pour  $K$  suffisamment grand,

$$\mathbb{P}\left(S_\varepsilon^K < T, X_T^K((\mathcal{K}^T)^c) \leq \frac{\varepsilon}{2}\right) \leq \mathbb{P}(S_\varepsilon^K < T)(1 - \eta). \quad (1.2.19)$$

et

$$\mathbb{E}(X_T^K((\mathcal{K}^T)^c)) < \frac{\varepsilon^2 \eta}{2}, \quad (1.2.20)$$

alors (1.2.18) implique (1.2.16).

**Etape 3 :** Montrons (1.2.19). Sur  $\{S_\varepsilon^K < T\}$ ,  $y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}$  implique que  $y^T \notin \mathcal{K}^T$ . D'où

$$X_T^K((\mathcal{K}^T)^c) \geq X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) \quad (1.2.21)$$

où le membre de droite est la masse au temps  $S_\varepsilon^K + t$  de la population issue des particules hors de  $\mathcal{K}$  au temps  $S_\varepsilon^K$ .

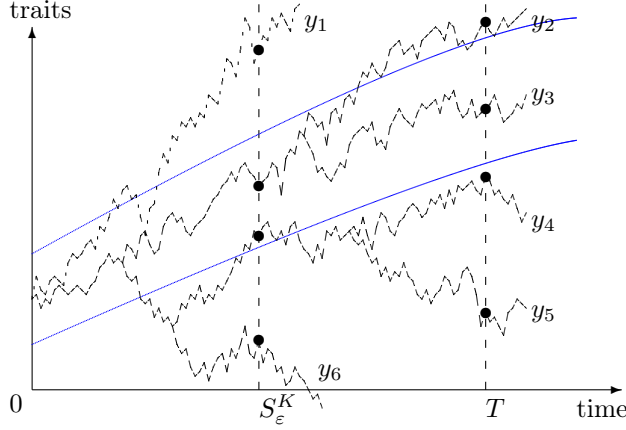


FIGURE 1.2.3 – Le compact  $\mathcal{K}$  est la région entre les deux lignes pleines. Les trajectoires dessinées correspondent au support de  $X_T^K$ .  $X_T^K((\mathcal{K}^T)^c)$  compte les trajectoires qui n'appartiennent pas à  $\mathcal{K}$  entre 0 et  $T$  : ici,  $y_1$ ,  $y_2$ ,  $y_4$ ,  $y_5$  et  $y_6$ .  $X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\})$  compte les trajectoires, au temps  $T$ , qui n'appartiennent pas à  $\mathcal{K}$  entre 0 et  $S_\varepsilon^K$  : ici,  $y_1$ ,  $y_4$ ,  $y_5$  et  $y_6$  ; en effet, la trajectoire  $y_2$  considérée entre les temps 0 et  $S_\varepsilon^K$  appartient à  $\mathcal{K}$ . Pour compter les trajectoires qui vont donner  $X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\})$ , on compte donc les descendants à  $T$  des 3 points au temps  $S_\varepsilon^K$  correspondant aux trajectoires  $y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}$ . On peut également vérifier sur la figure que la relation (1.2.21) est satisfaite.

Alors :

$$\begin{aligned} \mathbb{P}\left(S_\varepsilon^K < T, X_T^K((\mathcal{K}^T)^c) \leq \frac{\varepsilon}{2}\right) &\leq \mathbb{P}\left(S_\varepsilon^K < T, X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) \leq \frac{\varepsilon}{2}\right) \\ &= \mathbb{E}\left(\mathbb{1}_{S_\varepsilon^K < T} \mathbb{P}\left(X_{S_\varepsilon^K + (T - S_\varepsilon^K)}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) \leq \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right)\right). \end{aligned} \quad (1.2.22)$$

La probabilité sous l'espérance est la probabilité que la population issue des particules hors de  $\mathcal{K}$  au temps  $S_\varepsilon^K$ , soit de masse plus petite que  $\varepsilon/2$  au temps  $T$ . Du fait de l'indicatrice sous l'espérance, on travaillera dans toute la suite de cette étape sous l'hypothèse que  $S_\varepsilon^K < T$ . Notre but est de majorer cette probabilité par  $1 - \eta$  avec  $\eta \in (0, 1)$  ou de façon équivalente, comme

$$\mathbb{P}\left(X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) \leq \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right) \leq 1 - \mathbb{P}\left(\inf_{s \in [S_\varepsilon^K, T]} X_s^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) > \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right) \quad (1.2.23)$$

de minorer la probabilité apparaissant dans le second membre par  $\eta \in (0, 1)$ .

Sur  $\{S_\varepsilon^K < T\}$ , et conditionnellement à  $\mathcal{F}_{S_\varepsilon^K}$ ,  $X_{S_\varepsilon^K + \cdot}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\})$  définit un processus de naissances et morts en temps continu, et dont la masse au temps  $S_\varepsilon^K$  est

$$Z_{S_\varepsilon^K}^K = X_{S_\varepsilon^K}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) \geq X_{S_\varepsilon^K}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}_{S_\varepsilon^K}\}) \geq X_{S_\varepsilon^K}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}_T\}) = X_{S_\varepsilon^K}^n(\mathcal{K}_T^c) > \varepsilon.$$

Comme on cherche une minoration, on va minorer stochastiquement  $X_{S_\varepsilon^K+}^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\})$  par un processus  $Z_{S_\varepsilon^K+}^K$  de naissances et morts de taux  $K r(t, y)$  et  $K r(t, y) + D_0$  avec  $D_0 > 0$ , dont on verra plus tard le rôle, et on pourra supposer sans restriction que  $Z_{S_\varepsilon^K}^K = ([K\varepsilon] + 1)/K$ . Alors,

$$\mathbb{P}\left(\inf_{s \in [S_\varepsilon^K, T]} X_T^K(\{y^{S_\varepsilon^K} \notin \mathcal{K}^{S_\varepsilon^K}\}) > \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right) > \mathbb{P}\left(\inf_{s \in [S_\varepsilon^K, T]} Z_s^K > \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right).$$

Lorsque  $K$  est grand, on peut approcher  $Z_{S_\varepsilon^K+}^K$  par une diffusion. Pour  $t \in \mathbb{R}_+$ ,

$$Z_{S_\varepsilon^K+t}^K = Z_{S_\varepsilon^K}^K - D_0 \int_0^t Z_{S_\varepsilon^K+s}^K ds + M_t^{K,Z} \quad (1.2.24)$$

où  $M^{K,Z}$  est une martingale de carré intégrable telle que pour tout  $s \leq t$ ,

$$\begin{aligned} 2\underline{R} \int_s^t Z_{S_\varepsilon^K+u}^K du &\leq \langle M^{K,Z} \rangle_t - \langle M^{K,Z} \rangle_s = \int_s^t \left\langle X_{S_\varepsilon^K+u}^K, 2r(S_\varepsilon^K + u, \cdot) + \frac{D_0}{K} \right\rangle du \\ &\leq (2\bar{R} + \frac{D_0}{K}) \int_s^t Z_{S_\varepsilon^K+u}^K du. \end{aligned} \quad (1.2.25)$$

En utilisant à nouveau les contrôles des moments et le critère d'Aldous-Rebolledo, la famille des lois de  $(Z_{S_\varepsilon^K+}^K, \langle M^{K,Z} \rangle)$  est uniformément tendue dans  $\mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^2))$ . On montre que les valeurs d'adhérence  $(\mathcal{Z}, \mathcal{A})$  sont nécessairement continues et telles que :

$$\mathcal{N}_t = \mathcal{Z}_t - \varepsilon + D_0 \int_0^t \mathcal{Z}_s ds, \quad (1.2.26)$$

est une martingale de carré intégrable et de crochet  $\mathcal{A}$ , par [JS87, Th. 6.1 p.341].

En utilisant le théorème de représentation de Skorokhod, il existe une suite aléatoire  $(\tilde{Z}^K, \tilde{A}^K)_{K \in \mathbb{N}^*}$  et un couple aléatoire  $(\tilde{\mathcal{Z}}, \tilde{\mathcal{A}})$  définis sur le même espace de probabilité et de mêmes lois que  $(Z_{S_\varepsilon^K+}^K, \langle M^{K,Z} \rangle)_{K \in \mathbb{N}^*}$  et  $(\mathcal{Z}, \mathcal{A})$  respectivement, tels que

$$\lim_{K \rightarrow +\infty} (\tilde{Z}^K, \tilde{A}^K) = (\tilde{\mathcal{Z}}, \tilde{\mathcal{A}}) \quad \text{p.s.} \quad (1.2.27)$$

On déduit alors de (1.2.25) que p.s., pour  $0 \leq s \leq t$ ,

$$2\underline{R} \int_s^t \tilde{\mathcal{Z}}_u du \leq \tilde{\mathcal{A}}_t - \tilde{\mathcal{A}}_s \leq 2\bar{R} \int_s^t \tilde{\mathcal{Z}}_u du. \quad (1.2.28)$$

Ceci implique (e.g. Rudin [Rud87, Chapter 8]) que  $\tilde{\mathcal{A}}$  est une fonction absolument continue presque sûrement et qu'il existe une fonction aléatoire  $(\mathcal{F}_t)$ -adaptée,  $\rho(\cdot)$ , telle que

$$\forall u \in \mathbb{R}_+, 2\underline{R} \leq \rho(u) \leq 2\bar{R}, \quad \text{et} \quad \tilde{\mathcal{A}}_t = \int_0^t \rho(u) \tilde{\mathcal{Z}}_u du \quad \text{p.s.} \quad (1.2.29)$$

Il existe alors un mouvement Brownien standard  $(B_t)_{t \in \mathbb{R}_+}$  tel que l'on ait p.s.

$$\tilde{\mathcal{Z}}_t = \varepsilon - D_0 \int_0^t \tilde{\mathcal{Z}}_s ds + \int_0^t \sqrt{\rho(u)} \tilde{\mathcal{Z}}_u dB_u. \quad (1.2.30)$$

**Étape 4 :** Maintenant que nous disposons d'une limite diffusive pour  $Z_{S_\varepsilon^K+}^K$ , revenons à la minoration de la probabilité dans le second membre de (1.2.23) :

$$\mathbb{P}\left(\inf_{s \in [S_\varepsilon^K, T]} Z_s^K > \frac{\varepsilon}{2} \mid \mathcal{F}_{S_\varepsilon^K}\right) \mathbb{1}_{S_\varepsilon^K < T} \geq \mathbb{P}\left(\inf_{u \in [0, T]} (y|s| \tilde{Z}^K)_{s+u} > \frac{\varepsilon}{2} \Big|_{y=Z_{\cdot \wedge S_\varepsilon^K}^K, s=S_\varepsilon^K}\right) \mathbb{1}_{S_\varepsilon^K < T}. \quad (1.2.31)$$

avec pour tout  $y \in \mathbb{D}_{\mathbb{R}}$  et  $s \in [0, T]$ ,

$$\lim_{K \rightarrow +\infty} \mathbb{P} \left( \inf_{u \in [0, T]} (y|s|\tilde{Z}^K)_{s+u} > \frac{\varepsilon}{2} \right) = \mathbb{P} \left( \inf_{u \in [0, T]} (y|s|\tilde{Z})_{s+u} \geq \frac{\varepsilon}{2} \right) = \mathbb{P}_{y,s} \left( \inf_{u \in [0, T]} \tilde{Z}_u \geq \frac{\varepsilon}{2} \right), \quad (1.2.32)$$

où la notation  $\mathbb{P}_{y,s}$  est utilisée pour nous rappeler que la loi de  $\tilde{Z}$  dépend de  $\rho$ , qui lui-même peut dépendre de  $(y, s)$ . Comme nous avons besoin d'une minoration uniforme en  $(y, s)$  pour la limite en (1.2.32), nous allons montrer une "continuité" par rapport à  $(y, s)$ . On note  $B((z, r), \zeta)$  la boule ouverte de centre  $(z, r) \in \mathbb{D}_{\mathbb{R}} \times [0, T]$  et de rayon  $\zeta > 0$ . Il existe  $\zeta > 0$  suffisamment petit pour que pour tous  $(y, s) \in \mathcal{B}((z, r), \zeta)$  et  $Z \in \mathbb{D}$ ,  $\mathbf{d}_{\text{Sk}}((y|s|Z), (z|r|Z)) < \varepsilon/4$ . Pour ce choix de  $\zeta$ , on a pour tout  $K \in \mathbb{N}^*$ ,

$$\mathbb{P} \left( \inf_{u \in [0, T]} (z|r|\tilde{Z}^K)_{r+u} > \frac{3\varepsilon}{4} \right) \leq \mathbb{P} \left( \inf_{u \in [0, T]} (y|s|\tilde{Z}^K)_{s+u} > \frac{\varepsilon}{2} \right). \quad (1.2.33)$$

Soit  $\xi > 0$ . Les lois de  $(Z^K)_{K \in \mathbb{N}^*}$  étant uniformément tendues, il existe un compact  $\mathcal{K}_{\xi}$  de  $\mathbb{D}_{\mathbb{R}}$  tel que pour  $K$  suffisamment grand,  $\mathbb{P}(Z^K \notin \mathcal{K}_{\xi}) < \xi$ . Comme  $\mathcal{K}_{\xi} \times [0, T]$  est compact, il existe une suite finie  $(z_i, r_i)_{1 \leq i \leq M}$  avec  $M = M(\xi) \in \mathbb{N}^*$  telle que

$$\mathcal{K}_{\xi} \times [0, T] \subset \bigcup_{i=1}^{M(\xi)} B((z_i, r_i), \zeta).$$

Un argument similaire à (1.2.32), nous donne l'existence, pour chaque  $i \in \{1, \dots, M\}$ , d'un entier  $K_i$  tel que pour tout  $K \geq K_i$ ,

$$\mathbb{P} \left( \inf_{u \in [0, T]} (z_i|r_i|\tilde{Z}^K)_{r_i+u} > \frac{3\varepsilon}{4} \right) > \frac{1}{2} \mathbb{P}_{z_i, r_i} \left( \inf_{u \in [0, T]} \tilde{Z}_u \geq \frac{3\varepsilon}{4} \right). \quad (1.2.34)$$

Ainsi, pour tout  $(y, s) \in \mathcal{K}_{\xi} \times [0, T]$  et  $K \geq \max_{1 \leq i \leq M(\xi)} K_i$ ,

$$\begin{aligned} \mathbb{P} \left( \inf_{u \in [0, T]} (y|s|\tilde{Z}^K)_{s+u} \geq \frac{\varepsilon}{2} \right) &\geq \min_{1 \leq i \leq M(\xi)} \mathbb{P} \left( \inf_{u \in [0, T]} (z_i|r_i|\tilde{Z}^K)_{r_i+u} > \frac{3\varepsilon}{4} \right) \\ &> \min_{1 \leq i \leq M(\xi)} \frac{1}{2} \mathbb{P}_{z_i, r_i} \left( \inf_{u \in [0, T]} \tilde{Z}_u \geq \frac{3\varepsilon}{4} \right) =: 2\eta(\xi). \end{aligned} \quad (1.2.35)$$

$\eta$  défini en (1.2.35) est positif. En effet, comme il s'agit du minimum sur un nombre fini de terme, il suffit de minorer chacun d'eux. Pour  $(z, r) \in \mathbb{D}_{\mathbb{R}}$ , le fait que  $\mathbb{P}_{z,r} \left( \inf_{u \in [0, T]} \tilde{Z}_u \geq \frac{3\varepsilon}{4} \right) > 0$  est obtenu par l'utilisation du théorème de Girsanov, pour  $D_0 < \frac{4R}{T}$  (cf. [MT09, Kli14]).

On déduit de (1.2.31) et (1.2.35) que :

$$\mathbb{E} \left( \mathbb{1}_{S_{\varepsilon}^K < T} \mathbb{P} \left( \inf_{s \in [S_{\varepsilon}^K, T]} Z_s^K > \frac{\varepsilon}{2} \mid \mathcal{F}_{S_{\varepsilon}^K} \right) \right) \geq 2\eta(\xi) \mathbb{P}(S_{\varepsilon}^K < T, Z^K \in \mathcal{K}_{\xi}). \quad (1.2.36)$$

Le terme  $\mathbb{P}(S_{\varepsilon}^K < T, Z^K \in \mathcal{K}_{\xi})$  converge vers  $\mathbb{P}(S_{\varepsilon}^K < T)$  lorsque  $\xi$  tend vers zéro, et il existe  $\xi_0 > 0$  suffisamment petit tel que la probabilité du membre de droite de (1.2.36) puisse être minorée par  $\mathbb{P}(S_{\varepsilon}^K < T)/2$  pour tout  $\xi < \xi_0$ . On a donc minoré (1.2.36) par  $\eta(\xi)\mathbb{P}(S_{\varepsilon}^K < T)$ , pour  $0 < \xi < \xi_0$ .

**Étape 5 :** Il reste maintenant à prouver (1.2.20). En suivant la correction que Kliem [Kli14] a apporté à notre preuve dans [MT09], nous considérons un couplage de  $X^K$  avec un processus  $\tilde{X}^K$  qui le domine (toujours sans interaction), et dont les taux de naissance et de mort sont respectivement, pour  $K \in \mathbb{N}^*$  fixé :

$$\tilde{b}^K(t, y) = \left( K + \frac{\bar{B}}{R} \right) r(t, y), \quad \tilde{d}^K(t, y) = K r(t, y).$$

Nous avons pour tout  $T > 0$  et pour tout compact  $\mathcal{K} \subset \mathbb{D}$ ,  $\mathbb{E}(X_T^K((\mathcal{K}^T)^c)) \leq \mathbb{E}(\tilde{X}_T^K((\mathcal{K}^T)^c))$ .

L'arbre sous-jacent à  $\tilde{X}^K$  peut-être obtenu par élagage d'un arbre de Yule avec traits. Les particules se divisent en deux au taux  $\tilde{b}^K(t, y) + \tilde{d}^K(t, y)$ , l'une étant le parent et l'autre le descendant dans le processus initial. Les deux nouvelles particules sont étiquetées suivant les notation d'Harris-Ulam-Neveu et on appelle  $Y^{K,u}$ , pour  $u \in \mathcal{I} = \bigcup_{m=0}^{+\infty} \{0, 1\}^{m+1}$ , la lignée de la particule  $u$ . On notera  $u \succ i$  pour signifier que  $u$  est un descendant de  $i$ , et  $|u|$  la longueur de l'étiquette, i.e. la génération de l'individu correspondant. Pour  $k$  fixé, les lignées  $Y^{K,u}$  où  $|u| = k$  sont échangeables et leur loi commune est celle d'un processus de sauts pur sur  $\mathbb{R}$  stoppé au  $k$ ième saut. On notera  $\mathbb{P}_x^K$  la loi du processus de sauts issu de  $x \in \mathbb{R}$ , dont le taux de saut est  $\tilde{b}^K(t, y) + \tilde{d}^K(t, y)$  et dont la loi des sauts est  $\frac{1}{2}\delta_0(dh) + \frac{1}{2}M^K(dh)$ .

L'élagage est alors réalisé de la façon suivante pour obtenir  $\tilde{X}^K$  à partir de l'arbre de Yule : à chaque noeud  $(t, y)$  (c'est-à-dire à chaque individu  $y$  essayant de se diviser au temps  $t$  dans l'arbre de Yule), on conserve la descendance avec probabilité

$$\tilde{p}^K = \tilde{b}^K(t, y) / (\tilde{b}^K(t, y) + \tilde{d}^K(t, y)) = \frac{1}{2} \frac{1 + \frac{\bar{B}}{K\bar{R}}}{1 + \frac{\bar{B}}{2K\bar{R}}}, \quad (1.2.37)$$

qui est indépendante de  $(t, y)$ , et on supprime le sous arbre enraciné en  $(t, y)$  avec probabilité  $1 - \tilde{p}^K$ . Remarquons que lorsque  $K$  est grand, il existe  $C$  tel que

$$\tilde{p}^K \leq \frac{1}{2} \left(1 + \frac{C}{K}\right). \quad (1.2.38)$$

En notant  $V_t^K \subset \mathcal{I}$  l'ensemble (aléatoire) des étiquettes des individus vivants à  $t$  (pour  $X^K$ ),

$$\begin{aligned} \mathbb{E}(\tilde{X}_T^K((\mathcal{K}^T)^c)) &= \mathbb{E} \left( \frac{1}{K} \sum_{i=1}^{N_0^K} \sum_{u \succ i} \mathbb{P}(Y^{n,u} \notin \mathcal{K}^T, \quad u \in V_T^K) \right) \\ &= \mathbb{E} \left( \frac{1}{K} \sum_{i=1}^{N_0^K} \mathbb{E} \left( \sum_{u \succ i} \mathbb{P}(u \in V_T^K \mid Y^{K,u}) \mathbb{1}_{Y^{K,u} \notin \mathcal{K}^T} \right) \right). \end{aligned}$$

On appelle  $N_T^{K,u}$  le nombre de divisions le long de  $Y^{K,u}$  (dans l'arbre de Yule) entre les temps 0 et  $T$ .  $N_T^{K,u}$  est stochastiquement dominé par une variable de Poisson  $Z^K$  de paramètre  $2K\bar{R} + \bar{B}\bar{R}/\bar{R}$ , tronquée par  $|u|$  puisqu'il ne peut pas y avoir plus de  $|u|$  divisions le long de la trajectoire ancestrale de  $u$ . On a :

$$\mathbb{P}(u \in V_T^K \mid Y^{K,u}) = \mathbb{1}_{N_T^u = |u|} (\tilde{p}^K)^{|u|} \leq \mathbb{1}_{N_T^u = |u|} \left(\frac{1}{2}\right)^{|u|} \left(1 + \frac{C}{K}\right)^{|u|} \leq \frac{e^{C|u|/K}}{2^{|u|}}$$

pour  $K$  assez grand, et avec la constante  $C$  de (1.2.38).

Comme  $\{u \succ i\} = \bigcup_{k=1}^{+\infty} \{u \succ i, |u| = k\}$  et comme dans un arbre binaire, la génération  $k$  compte  $2^k$  individus, alors on a, pour  $A \in \mathbb{N}^*$  avec  $A > 2\bar{R}$  :

$$\begin{aligned} \mathbb{E} \left( \sum_{u \succ i} \mathbb{P}(u \in V_T^K \mid Y^{K,u}) \mathbb{1}_{Y^{K,u} \notin \mathcal{K}^T} \right) &\leq \mathbb{E} \left( e^{CA} \mathbb{P}_{X_0^K}^K((\mathcal{K}^T)^c) \right) + \sum_{k > KA} (2\tilde{p}^K)^k \mathbb{P}(Z^K = k) \\ &\leq e^{CA} \mathbb{E} \left( \mathbb{P}_{X_0^K}^K((\mathcal{K}^T)^c) \right) + \mathbb{E} \left( (2\tilde{p}^K)^{Z^K} \mathbb{1}_{\frac{Z^K}{K} > A} \right) \end{aligned}$$

Par le choix du noyau de mutation  $M^K(dh)$ , la famille  $(\mathbb{P}_x^K)_{K \in \mathbb{N}^*, x \in C}$  est tendue pour tout compact  $C$  de  $\mathbb{R}$ . Il est donc possible de trouver un compact  $\mathcal{K}^T$  de sorte à ce que le premier terme du membre de droite soit aussi petit que souhaité. Pour le second terme, remarquons que par la loi des grands nombres,  $Z^K/K \rightarrow 2\bar{R}$  lorsque  $K \rightarrow +\infty$  et comme  $Z^K$  est une variable de Poisson,

$$\mathbb{E} \left( (2\tilde{p}^K)^{Z^K} \right) = \exp \left( (2K\bar{R} + \frac{\bar{B}\bar{R}}{R}) (2\tilde{p}^K - 1) \right) \leq e^{3\bar{R}C} < +\infty.$$



Alors, par convergence dominée le second membre tend vers 0 lorsque  $K \rightarrow +\infty$ . Ceci conclut la preuve de la tension de  $(X^K)_{K \in \mathbb{N}^*}$ .

**Etape 6 :** Soit  $\bar{X} \in \mathcal{C}([0, T], \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  une valeur d'adhérence, nécessairement continue. En procédant de façon similaire à l'identification des valeurs d'adhérence dans les preuves des Théorèmes 1.1.11 et 1.1.14, on peut montrer que  $\bar{X}$  satisfait le problème de martingale (1.2.14)-(1.2.15). Comme  $\bar{X}$  est continu, on peut utiliser les fonctions test (1.2.12) et reformuler le problème de martingale. On montre alors que pour  $\bar{X}$ ,

$$M_t^\varphi = \langle \bar{X}_t, \varphi \rangle - \langle \bar{X}_0, \varphi \rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} \left( pr(s, y) \frac{\sigma^2}{2} \tilde{\Delta} \varphi(s, y) + \gamma(s, y, \bar{X}^s) \varphi(y) \right) \bar{X}_s(dy) ds \quad (1.2.39)$$

est une martingale de carré intégrable, issue de 0 et de crochet

$$\langle M^\varphi \rangle_t = \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} 2r(s, y) \varphi^2(y) \bar{X}_s(dy) ds. \quad (1.2.40)$$

On montre l'unicité de la solution du problème de martingale (1.2.39)-(1.2.40), dont on déduit l'unicité de la solution de (1.2.14)-(1.2.15). Soit  $\mathbb{P}$  une solution de (1.2.39)-(1.2.40) et soit  $\bar{X}$  le processus canonique sur  $\mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$ . Par le théorème de Dawson-Girsanov (cf. [Daw78, Section 5], [EP94, Theorem 2.3]), il existe une probabilité  $\mathbb{Q}$  sur  $\mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$  équivalente à  $\mathbb{P}$  sous laquelle, pour toute fonction test  $\varphi$  de la forme (1.2.12),

$$\tilde{M}_t^\varphi = \langle \bar{X}_t, \varphi \rangle - \langle \bar{X}_0, \varphi \rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}}} \frac{pr(s, y) \sigma^2}{2} \tilde{\Delta} \varphi(s, y) \bar{X}_s(dy) ds \quad (1.2.41)$$

est une martingale de crochet (1.2.40). Comme on s'est débarrassé des non-linéarités, on peut considérer la transformée de Laplace de  $\bar{X}$  sous  $\mathbb{Q}$ ,  $L(s, t, y, \varphi) = \mathbb{E}^{\mathbb{Q}}(\exp(-\langle \bar{X}_t, \varphi \rangle) | \bar{X}_s = \delta_{y^s})$ . L'étude de  $L(s, t, y, \varphi)$  est liée à celle de la diffusion  $Y_t = Y_0 + \int_0^t \sqrt{\sigma^2 pr(s, Y^s)} dB_s$  et du processus historique associé  $W \in \mathcal{C}(\mathbb{R}_+, \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d))$ , défini par  $W_t = Y^t = (Y_{t \wedge s})_{s \in \mathbb{R}_+}$ .  $W$  n'est pas homogène, mais c'est un processus de Markov fort (à valeurs trajectoires), de semi-groupe défini pour tout  $s \leq t$  et tout  $\varphi \in \mathcal{C}_b(\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d), \mathbb{R})$  par  $S_{s,t} \varphi(y) = \mathbb{E}^{\mathbb{Q}}(\varphi(W_t) | W_s = y^s)$ . On montre alors, en adaptant [Daw93, Th 12.3.1.1 p.207], que  $L(s, t, y, \varphi)$  est l'unique solution de

$$L(s, t, y, \varphi) = e^{-V_{s,t} \varphi(y)}, \quad \text{où } V_{s,t} \varphi(y) = S_{s,t} \varphi(y) - \int_s^t \frac{p\sigma^2}{2} S_{s,u} \left( r(u, \cdot) (V_{u,t} \varphi(\cdot))^2 \right) (y) du.$$

□

Dans la preuve de l'Etape 6 du Théorème 1.2.3, on voit que la solution du problème de martingale linéaire (1.2.41) peut se décrire comme un bouquet de diffusions obtenues en considérant une EDS dirigée par un Brownien  $y$  que l'on va choisir comme étant l'une des trajectoires du super-mouvement Brownien historique  $(H_t(dy))_{t \in \mathbb{R}_+}$  (cf. [DP91]) :

$$Y_t(y) = Y_0(y) + \int_0^t \sqrt{\sigma^2 pr(s, Y^s(y))} dy_s \quad (1.2.42)$$

$$\langle \bar{X}_t, \varphi \rangle = \int_{\mathbb{D}_{\mathbb{R}}} \varphi(Y(y)^t) H_t(dy) \quad (1.2.43)$$

pour une fonction test  $\varphi$ . Cette représentation permet de dire que les trajectoires des individus tirés au hasard au temps  $t$  ont la loi de la trajectoire  $(Y_s)_{s \in [0, t]}$ . Ce résultat a été généralisé pour des problèmes de martingales historiques non-linéaires par Perkins [Per95], mais qui n'incluent pas les interactions de type logistique. Pour l'instant, nous n'avons dans [MT09] que des réponses

partielles à ce problème extrêmement intéressant. Ceci pourrait avoir des applications en théorie de l'évolution pour décrire des processus canoniques sans les hypothèses de mutations rares, comme le suggère la simulation de la Fig. 1.2.2(a) : lorsque l'on considère les temps précédant l'instant présent  $t$ , le nombre de trajectoires ancestrales devient rapidement très petit, et caractériser la loi des trajectoires observées, c'est-à-dire des trajectoires ancestrales des individus ayant survécu jusqu'au temps  $t$ , serait une grande avancée.

Remarquons que lorsque  $r(s, y)$  et  $\gamma(s, y, \bar{X}^s)$  sont des fonctions constantes égales à  $\bar{R}$  et  $\bar{\gamma}$ , les lignées ancestrales des individus tirés au hasard, sous  $\mathbb{E}(\bar{X}_t(\cdot)/\langle \bar{X}_t, 1 \rangle)$  sont des mouvements Browniens. Considérons des fonctions test  $\phi(\mu) = \langle \mu \otimes \mu, \chi \rangle / \langle \mu, 1 \rangle^2$  où  $\chi(y, z)$  est une fonction de deux variables telle que  $y \mapsto \chi(y, z)$  et  $z \mapsto \chi(y, z)$  soient de la forme (1.2.12), et où  $\langle \mu_t \otimes \mu_t, \chi \rangle = \int_{\mathbb{D}_R} \int_{\mathbb{D}_R} \chi(y, z) \mu_t(dy) \mu_t(dz)$ . Le générateur de  $\bar{X}$  s'écrit :

$$L^{FV} \phi(X) = \frac{p\bar{R}\sigma^2}{2} \left\langle \frac{X}{\langle X, 1 \rangle} \otimes \frac{X}{\langle X, 1 \rangle}, \tilde{\Delta}^{(2)} \chi \right\rangle + \frac{2\bar{R}}{\langle X, 1 \rangle} \left( \int_{\mathbb{D}_R} \chi(y, y) \frac{X(dy)}{\langle X, 1 \rangle} - \left\langle \frac{X}{\langle X, 1 \rangle} \otimes \frac{X}{\langle X, 1 \rangle}, \chi \right\rangle \right),$$

où  $\tilde{\Delta}^{(2)} \chi(y, z) = \tilde{\Delta}(y \mapsto \chi(y, z)) + \tilde{\Delta}(z \mapsto \chi(y, z))$ . Il s'agit d'un processus de Fleming-Viot (avec un terme de ré-échantillonnage non-linéaire) similaire à celui introduit par [GPW11].

Dans un cas sans interaction un peu plus général, où  $\gamma(s, y, \bar{X}^s) = \gamma(s, y)$  n'est pas nécessairement constant, il est plus facile de considérer  $\mathbb{E}(\bar{X}_t(\cdot)/\mathbb{E}(\langle \bar{X}_t, 1 \rangle))$ , au lieu de la mesure intensité de  $\bar{X}_t(\cdot)/\langle \bar{X}_t, 1 \rangle$ . Nous n'avons pas pu décrire cette probabilité sur  $\mathcal{C}([0, t], \mathbb{R})$  comme la loi d'un processus, mais néanmoins, il est possible de donner une représentation du type Feynman-Kac et ainsi d'exprimer cette probabilité à partir de la loi d'un processus à valeurs réelles. Pour toute fonction test  $\varphi$  de la forme (1.2.12),

$$\langle \mathbb{E}\nu_t, \varphi \rangle = \frac{\mathbb{E}(\langle \bar{X}_t, \varphi \rangle)}{\mathbb{E}(\langle \bar{X}_t, 1 \rangle)} = \frac{\mathbb{E}\left(\varphi(W_t) e^{\int_0^t \gamma(s, W_s) ds}\right)}{\mathbb{E}\left(e^{\int_0^t \gamma(s, W_s) ds}\right)}, \quad (1.2.44)$$

où  $(W_t)_{t \in \mathbb{R}_+}$  est le processus à valeurs trajectoires qui est le processus historique de la diffusion  $Y$  introduite en (1.2.43), comme dans l'Etape 6 de la preuve du Théorème 1.2.3.

Dans le cas logistique avec  $r(s, y) = \bar{R}$  tel que  $p\sigma^2\bar{R} = 1$  (pour simplifier) et  $\gamma(s, y, X) = \alpha - \eta\langle X, 1 \rangle$ , où  $\alpha, \eta > 0$ , on a la représentation de type Feynman-Kac suivante :

$$\langle \mathbb{E}\nu_t, \varphi \rangle = \frac{\mathbb{E}\left(\varphi(W_t) e^{\int_0^t (\alpha N_s - \eta N_s^2) ds}\right)}{\mathbb{E}\left(e^{\int_0^t (\alpha N_s - \eta N_s^2) ds}\right)}, \quad (1.2.45)$$

où  $(W_t)_{t \in \mathbb{R}_+}$  est le processus Brownien historique et où  $N$  est la solution de

$$N_t = \langle \bar{X}_0, 1 \rangle + \int_0^t (2\bar{R} + \alpha N_s - \eta N_s^2) ds + \int_0^t \sqrt{2\bar{R}N_s} dB_t, \quad (1.2.46)$$

avec  $B$  un mouvement Brownien standard indépendant de  $W$ .

## 1.2.2 Trajectoire ancestrale d'un individu tiré au hasard

### 1.2.2.1 Diffusions indexées par un processus de Galton-Watson en temps continu

L'idée de caractériser la loi de la trajectoire ancestrale d'un individu tiré au hasard parmi les individus vivants au temps  $t$ , développée à la fin de la Section 1.2.1.3, a été largement étudiée pour

des populations sans interaction. Pour des superprocessus, on peut citer par exemple Engländer et Turaev [ET02], Engländer et Winter [EW06] ou Evans et Steinsaltz [ES07]. Avec Vincent Bansaye, Jean-François Delmas et Laurence Marsalle [BDMT11], nous avons considéré des populations discrètes en temps continu.

Chaque individu a une durée de vie exponentielle de paramètre  $r$  et à sa mort, est remplacé par un nombre aléatoire  $\nu$  de descendants. La loi de  $\nu$  est une probabilité sur  $\mathbb{N}$ , et on note  $p_k = \mathbb{P}(\nu = k)$ . Ceci décrit un processus de Galton-Watson en temps continu. Dans la suite, on suppose que la loi de  $\nu$  admet une espérance  $m$  et une variance  $\sigma^2$  finies. On ne considérera ici que le cas sur-critique  $m > 1$ , pour simplifier. La population survit alors en temps long avec une probabilité strictement positive (e.g. [AN70]).

Au cours de sa vie, l'individu est caractérisé par une covariable ou état, à valeurs dans  $\mathbb{R}$  et qui évolue suivant un processus de Markov  $Y$  de générateur défini, pour  $f \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$ , par  $Lf(x) = b(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x)$ , avec  $b$  et  $\sigma$  des fonctions Lipschitziennes bornées. Dans [BDMT11], nous avons considéré des d'états espaces et des générateurs plus généraux. Lorsqu'un individu meurt à l'état  $x$  et laisse  $k$  descendants, les états de ces derniers sont donnés par le vecteur  $(F_1^{(k)}(x, \Theta), \dots, F_k^{(k)}(x, \Theta))$  où  $\Theta$  est une variable aléatoire qui paramètre l'aléa dans les états de départ des descendants. On peut supposer dans cette section, sans restriction, que  $\Theta$  est uniforme sur  $[0, 1]$ . Remarquons que variables  $F_i^{(k)}$  précédentes ne sont pas nécessairement indépendantes. Les descendants évoluent ensuite de façon complètement indépendante.

Comme dans les sections précédentes, on peut associer à chaque individu une étiquette en suivant les notations d'Ulam-Harris-Neveu. Soit  $\mathcal{U} = \bigcup_{m=0}^{+\infty} (\mathbb{N}^*)^m$  avec la convention  $(\mathbb{N}^*)^0 = \{\emptyset\}$ . La racine est étiquetée par  $\emptyset$  et lorsqu'un individu d'étiquette  $u$  est remplacé par  $k$  descendants, ceux-ci portent les étiquettes  $(u, 1), \dots, (u, k)$  obtenues en concaténant à  $u$  un entier supplémentaire (1 à  $k$ ). En notant  $V_t$  l'ensemble des étiquettes des individus vivants au temps  $t$ , la population à  $t$  peut-être décrite par la mesure :  $X_t(dx) = \sum_{u \in V_t} \delta_{Y_t^u}(dx) \in \mathcal{M}_F(\mathbb{R})$  où  $Y_t^u$  est l'état de l'individu  $u$  au temps  $t$ . Le processus historique associé est

$$H_t(dy) = \sum_{u \in V_t} \delta_{W_t^u}(dy) \in \mathcal{M}_F(\mathbb{D})$$

avec  $W_t^u$  la trajectoire ancestrale de  $u$ , c'est-à-dire que  $W^u$  est le processus à valeurs trajectoires défini pour  $s \leq t$  par  $W_t^u(s) = Y_s^{u(s)}$  où  $u(s)$  est l'ancêtre de  $u$  vivant à  $s$ , et prolongé pour  $s > t$  par  $W_t^u(s) = Y_t^u$ . Enfin, on introduit également  $\mathcal{T}$  le sous-ensemble des étiquettes de  $\mathcal{U}$  appartenant aux individus présents dans l'arbre de Galton-Watson.

Le processus  $(X_t)_{t \in \mathbb{R}_+}$ , dont la dynamique est décrite ci-dessus, est un processus fortement Markovien de générateur  $\mathcal{L}$  défini pour les fonctions test  $\phi_f$  de la forme  $\phi_f(X) = \phi(\langle X, f \rangle)$ , avec  $f \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$  et  $\phi \in \mathcal{C}_b^2(\mathbb{R})$  (cf. [Daw93, Th. 3.2.6]), par :

$$\mathcal{L}\phi_f(X) = \mathcal{L}_1\phi_f(X) + \mathcal{L}_2\phi_f(X), \quad (1.2.47)$$

où les générateurs  $\mathcal{L}_1$  et  $\mathcal{L}_2$  correspondent aux parties de branchement et d'évolution suivant  $Y$  :

$$\begin{aligned} \mathcal{L}_1\phi_f(X) &= r \int_E \int_0^1 \sum_{k \in \mathbb{N}} \left( \phi \left( \langle X, f \rangle + \sum_{j=1}^k f(F_j^{(k)}(x, \theta)) - f(x) \right) - \phi_f(X) \right) p_k d\theta Z(dx), \\ \mathcal{L}_2\phi_f(X) &= \langle Z, Lf \rangle \phi'(\langle X, f \rangle) + \langle X, \sigma(x)f''(x) \rangle \phi''(\langle X, f \rangle). \end{aligned}$$

De tels processus de branchement-diffusion ont été considérés par Dawson [Daw93, Section 2.10] ou Roelly et Rouault [RR90] par exemple, mais pour des branchements locaux, c'est-à-dire que

les descendants d'une particule apparaissent à l'emplacement de cette dernière. Des processus de branchement-diffusion avec branchement non-local, similaires à celui présenté ici ont été considérés par Asmussen et Hering [AH83], Dawson et ses co-auteurs [Daw93, DGL02]. Suite à notre travail [BDMT11], Cloez [Clo11] a également étudié des systèmes de particules en mouvement, avec branchements non locaux et dépendances spatiales possibles pour le taux de mort et la loi de reproduction. Alors que ces articles considèrent des théorèmes limites, pour des grandes densités d'individus, de petites masses et des branchements parfois rapides, nous ne renormalisons cette fois pas la mesure  $X_t$ .

Le fait de conserver un arbre discret nous permet de faire le lien entre ces modèles et certaines données récupérées par les biologistes, afin de faire des statistiques (voir par exemple Stewart et al. [SMPT05]). L'arbre binaire peut par exemple correspondre à des cellules en division, et la diffusion le long de chaque branche à l'évolution d'un continuum de parasites à l'intérieur de chaque cellule (cf. [BT11] ou Evans et Steinsaltz [ES07]). En temps discret, le pendant des processus que nous introduisons sont les chaînes de Markov bifurcantes, voir par exemple Athreya et Kang [AK98], Benjamini et Peres [BP94], Bansaye [Ban13]. Des applications statistiques pour ces chaînes de Markov bifurcantes ont été développées entre autres par Guyon [Guy07], Bansaye [Ban08], Delmas et Marsalle [DM08], Bercu et al. [BSGP09], De Saporta Gégout-Petit et Marsalle [SGPM14, SGPM13, SGPM12, SGPM11].

Notre objectif est de chercher à caractériser la loi de la trajectoire ancestrale d'un individu tiré au hasard au temps  $t$ . Plus précisément, on s'intéresse aux mesures empiriques

$$\mathbb{1}_{N_t > 0} \frac{X_t}{N_t} \quad \text{et} \quad \mathbb{1}_{N_t > 0} \frac{H_t}{N_t},$$

où  $N_t$  est la taille de la population au temps  $t$ . Si  $U(t)$  est une variable aléatoire uniforme sur  $V_t$ , conditionnellement à  $V_t$ , alors pour de bonnes fonctions test  $\varphi$  :

$$\mathbb{E} \left[ \frac{\mathbb{1}_{\{N_t > 0\}}}{N_t} \sum_{u \in V_t} \varphi(W_t^u) \right] = \mathbb{E} [\mathbb{1}_{\{N_t > 0\}} \varphi(W_t^{U(t)})],$$

et il en est de même en remplaçant  $W^u$  par  $Y^u$ . Nous montrons dans [BDMT11] que la loi de cette trajectoire tirée au hasard peut-être approchée par celle d'un processus  $(\mathcal{Y}_t)_{t \in \mathbb{R}_+}$  que nous appellerons processus auxiliaire :

**Proposition 1.2.5.** *Soit  $(\mathcal{Y}_t)_{t \in \mathbb{R}_+}$  le processus de saut-diffusion réel dont le générateur est donné, pour une fonction test  $f \in \mathcal{C}_b^2(\mathbb{R}, \mathbb{R})$ , par :*

$$Af(x) = Lf(x) + rm \sum_{k=1}^{+\infty} \frac{p_k}{m} \int_0^1 \sum_{j=1}^k \left( f(F_j^{(k)}(x, \theta)) - f(x) \right) d\theta. \quad (1.2.48)$$

*Soit  $(\mathcal{W}_t)_{t \in \mathbb{R}_+}$  le processus historique associé, à valeurs trajectoires : pour tout  $t > 0$ ,  $\mathcal{W}_t = \mathcal{Y}^t \in \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ . Alors, pour tout  $T > 0$  et pour toute fonction positive  $\varphi \in \mathcal{B}(\mathbb{D}([0, T], \mathbb{R}), \mathbb{R}_+)$ ,*

$$\frac{\mathbb{E}(\sum_{u \in V_T} \varphi(W_T^u))}{\mathbb{E}(N_T)} = \mathbb{E}(\varphi(\mathcal{W}_T)). \quad (1.2.49)$$

*Cette identité reste également vraie en remplaçant  $W^u$  par  $Y^u$  et  $\mathcal{W}$  par  $\mathcal{Y}$ .*

La preuve de la Proposition 1.2.5 s'appuie sur un théorème de Girsanov pour les processus de Poisson composés. En fait, le théorème de Girsanov que l'on applique résulte d'une identité

simple plus profonde, liée aux propriétés des lois exponentielles : si  $\tau$  est une variable aléatoire exponentielle de paramètre  $r$ , alors pour toute fonction  $g$  positive,

$$\mathbb{E}\left(r \int_0^\tau g(t) dt\right) = \mathbb{E}(g(\tau)).$$

Le processus auxiliaire  $\mathcal{Y}$  a le même générateur  $L$  que le processus  $Y$  décrivant l'évolution le long des branches du Galton-Watson, auquel s'ajoutent des sauts. Il apparaît un phénomène de biais dans la loi des sauts et dans le taux de saut. Ceci est intuitif, car lorsque l'on choisit un individu uniformément parmi  $V_T$ , on a tendance à le tirer dans les lignées ayant davantage de générations ou possédant des ancêtres plus prolifiques. Tout se passe comme si la lignée qu'on avait tirée sautait au taux  $rm$  (au lieu de  $r$ ), et comme si les événements de branchement que l'on rencontrait le long de cette lignée donnaient  $k$  descendants avec probabilité  $kp_k/m$ . Ce phénomène de biais avait déjà été constaté pour des processus de branchement diffusion par Chauvin et al. [CRW91], Hardy et Harris [HH09], Harris et Roberts [HR12, HH06]. La proposition 1.2.5 permet de remplacer l'étude des processus à valeurs mesures  $X$  ou  $H$  par celle de processus réels  $\mathcal{Y}$  ou  $\mathcal{W}$ . Harris et ses co-auteurs ont pour cela introduit l'appellation de *many-to-one* pour ce genre d'identité.

Dans [BDMT11], nous avons ajouté au modèle des naissances non-locales et pouvant être dépendantes. De plus, notre résultat s'énonce dans un cadre trajectorien (avec le processus historique), et nous l'avons généralisé dans plusieurs autres directions :

- Il est possible de faire dépendre les fonctions  $\varphi$  de  $T$  dans (1.2.49).
- Nous pouvons aussi étendre (1.2.49) à des sommations, dans le membre de gauche, portant sur tous les noeuds de l'arbre. Plus précisément, si  $\beta(u)$  dénote le temps de mort de l'individu  $u$  (où il est remplacé par un nombre aléatoire de descendants), alors :

$$\frac{\mathbb{E}\left(\sum_{u \in \mathcal{T}} \varphi(\beta(u), W_{\beta(u)}^u)\right)}{\mathbb{E}(N_T)} = r \int_0^{+\infty} e^{r(m-1)s} \mathbb{E}(\varphi(s, \mathcal{W}_s)) ds. \quad (1.2.50)$$

- Enfin, nous avons également étudié la loi jointe des trajectoires ancestrales d'un nombre fini d'individus tirés uniformément sans remise dans  $V_T$ . Nous renvoyons à [BDMT11] pour l'énoncé de ce résultat qui demande d'introduire beaucoup de nouvelles notations.

Appliquons ces formules *many-to-one* à l'étude d'une loi des grands nombres pour  $X_t/N_t$  lorsque  $t \rightarrow +\infty$ .

**Théorème 1.2.6.** *Si le processus  $\mathcal{Y}$  est ergodique, de mesure invariante  $\pi$ , alors pour toute fonction réelle bornée  $f$  :*

$$\lim_{t \rightarrow +\infty} \frac{\mathbb{1}_{N_t > 0}}{N_t} \langle X_t, f \rangle = \mathbb{1}_{\chi > 0} \int_{\mathbb{R}} f(x) \pi(dx), \quad (1.2.51)$$

en probabilité, en notant  $\chi = \lim_{t \rightarrow +\infty} N_t / E(N_t)$  p.s.

En particulier on déduit de ce résultat que

$$\lim_{t \rightarrow +\infty} \mathbb{E}(f(X_t^{U(t)}) \mid N_t > 0) = \int_{\mathbb{R}} f(x) \pi(dx)$$

où  $U(t)$  conditionnellement à  $V_t$  est une variable uniforme sur  $V_t$ . De ce point de vue, on donne ainsi une interprétation probabiliste à la distribution quasi-stationnaire d'une lignée ancestrale tirée au hasard. Dans les travaux d'Engländer Turaev et Winter [ET02, EW06] ou d'Evans et Steinsaltz [ES07], des résultats similaires sont énoncés, pour des superprocessus, et la limite est obtenue par des résultats spectraux. Ici, on caractérise la limite comme la mesure stationnaire d'un processus auxiliaire  $\mathcal{Y}$  reflétant un biais lié au temps continu et au caractère aléatoire du nombre de descendants de chaque individu.

Le Théorème 1.2.6 est généralisable pour des conditions plus faibles que l'ergodicité de  $\mathcal{Y}$ , mais permettant de considérer des classes de fonctions  $f$  plus larges et en particulier à croissance polynomiale plutôt que bornées. On peut également en donner une version trajectorielle en remplaçant  $\langle X_t, f \rangle = \sum_{u \in V_t} f(X_t^u)$  par  $\sum_{u \in V_t} \varphi(X_s^u, t-h \leq s < t)$  pour  $h > 0$  fixé, ce qui revient à considérer la loi d'une portion de lignée ancestrale dans une fenêtre de temps de largeur  $h$ .

*Idée de la preuve du Théorème 1.2.6.* Quitte à ajouter une constante à  $f$ , on peut supposer que  $\langle \pi, f \rangle = 0$ , et il suffit donc de montrer que  $\langle X_t, f \rangle / N_t$  converge vers 0. Comme  $N_t / \mathbb{E}(N_t)$  converge presque sûrement vers  $\chi$  (cf. Athreya-Ney [AN70, Th.1 et 2 p.111]), alors il suffit de montrer que  $\langle X_t, f \rangle / \mathbb{E}(N_t)$  converge vers 0, ce que nous faisons dans  $L^2$ . Remplacer  $N_t$  par  $\mathbb{E}(N_t) = e^{r(m-1)t}$  correspond à ce qui est fait pour les super-processus ([ES07], par exemple), et c'est la renormalisation qui nous fournit le processus auxiliaire dans (1.2.5).

$$\mathbb{E} \left[ \frac{\langle X_t, f \rangle^2}{\mathbb{E}(N_t)^2} \right] = \frac{1}{\mathbb{E}(N_t)^2} \mathbb{E} \left( \sum_{u \in V_t} f^2(Y_t^u) \right) + \frac{1}{\mathbb{E}(N_t)^2} \mathbb{E} \left( \sum_{\substack{(u,v) \in V_t^2 \\ u \neq v}} f(Y_t^u) f(Y_t^v) \right). \quad (1.2.52)$$

Ainsi, le premier terme du membre de gauche est égal à  $e^{-r(m-1)t} \mathbb{E}(f^2(\mathcal{Y}_t))$  qui converge vers 0 lorsque  $t \rightarrow +\infty$ . Pour le second terme, l'idée heuristique est que lorsque l'on tire  $u$  et  $v$  au hasard, le plus récent ancêtre commun est éloigné en terme de générations et les états de  $u$  et  $v$  sont presque indépendants. Alors, comme on a supposé que  $\langle \pi, f \rangle = 0$ , le second terme est petit.  $\square$

Il est enfin également possible d'établir un théorème central limite. Soit  $h > 0$  fixé. On peut considérer le processus défini pour  $t > 0$  et  $f \in \mathcal{B}_b(\mathbb{R}, \mathbb{R})$  :

$$\langle \eta_t^T, f \rangle = \sqrt{\mathbb{E}(N_{t+h})} \left( \frac{\langle X_{t+h}, f \rangle}{\mathbb{E}(N_{t+h})} - \frac{\langle X_h, \mathbb{E}(f(\mathcal{Y}_t)) \rangle}{\mathbb{E}(N_t)} \right).$$

La mesure  $\eta_t^T$  n'est plus une mesure positive, et l'espace des mesures signées muni de la topologie de la convergence étroite n'est plus métrisable. En suivant les approches de Métivier [M84] et Méléard [M98], on plonge l'ensemble des mesures signées dans des espaces de Sobolev, dans lesquels le processus des fluctuations converge vers un processus Gaussien.

### 1.2.2.2 Application aux modèles hôtes-parasites

Avec Vincent Bansaye [BT11], nous avons utilisé les formules *many-to-one* présentées précédemment pour étudier des systèmes hôtes-parasites. Nous considérons un processus de Yule en temps continu, représentant une population de cellules qui se divisent, au bout de temps aléatoires exponentiels de paramètre  $r > 0$ , en exactement deux cellules filles. La diffusion le long des branches décrit l'évolution de la quantité de parasites contenue dans chaque cellule. Nous choisissons ici une diffusion de Feller driftée  $Y$  de générateur  $Lf(x) = g x f'(x) + \sqrt{2\sigma^2 x} f''(x)$ . On s'intéresse ici au cas surcritique où  $g > 0$ . Alors avec la probabilité positive  $1 - e^{-g x_0 / \sigma^2}$ , le processus  $Y$  issu de  $x_0$  ne touche pas 0 et  $Y_t$  tend vers  $+\infty$  lorsque  $t \rightarrow +\infty$ . Au moment de la division, la répartition des parasites est  $F_1^{(2)}(x, \Theta) = \Theta x$ , et  $F_1^{(2)}(x, \Theta) = (1 - \Theta)x$ , où  $\Theta$  est une variable de loi symétrique sur  $[0, 1]$ . Ce modèle comporte des similarités avec les modèles de fragmentation d'intervalles (e.g. [BD86, BD87]), mais la taille de la population totale des parasites est ici variable.

Le processus auxiliaire  $\mathcal{Y}$  est ici une diffusion de Feller de générateur  $L$ , fragmentée, au taux  $2r$ , par multiplication par des variables aléatoires ayant la même loi que  $\Theta$ . L'étude fine de ce processus de saut-diffusion, en particulier de la probabilité avec laquelle il touche 0, point absorbant, nous fournit, grace aux formules *many-to-one* des critères de guérison :

**Proposition 1.2.7.** *Si  $g \leq 2r \mathbb{E}(\log(1/\Theta))$ , alors on a guérison au sens où presque sûrement,*

$$\lim_{t \rightarrow +\infty} \frac{\text{Card}(u \in V_t, X_t^u = 0)}{N_t} = 1.$$

*Si  $g > 2r \mathbb{E}(\log(1/\Theta))$ , alors presque sûrement, pour tout  $0 < \kappa < g - 2r\mathbb{E}(\log(1/\Theta))$ ,*

$$\left\{ \limsup_{t \rightarrow +\infty} \frac{\text{Card}(u \in V_t, X_t^u \geq e^{\kappa t})}{N_t} > 0 \right\} = \{\forall t \geq 0, \mathcal{Y}_t > 0\}.$$

*Ceci signifie que l'on va trouver une quantité de parasites croissant exponentiellement vite dans une fraction positive de cellules.*

Remarquons que la loi de répartition des parasites au moment de la division, c'est-à-dire la loi de  $\Theta$ , va permettre la guérison même pour des croissances  $g$  fortes du parasites si  $\mathbb{E}(\log(1/\Theta))$  est grand (répartition très inégale). Le facteur 2 qui apparaît dans le critère  $g \leq 2r \mathbb{E}(\log(1/\Theta))$  vient du fait que le processus auxiliaire  $\mathcal{Y}$  saute au taux  $2r$  du fait des phénomènes de biais qui apparaissent. Grâce à ce facteur 2, on constate qu'il peut y avoir guérison même dans des cas où presque sûrement, en partant de la cellule initiale et en choisissant à chaque division une fille au hasard avec probabilité  $1/2$  on a une quantité de parasites tendant vers  $+\infty$ . En effet, procédant ainsi, on obtient une diffusion de Feller fractionnée au taux  $r$  et non pas  $2r$ ...

Pour les applications, le cas d'un taux de division  $r(x)$  non constant serait très intéressant, mais les critères de guérison dans ce cas restent ouverts. Enfin, la thèse de Hoang V.H., commencé en 2013, est issue de questions statistiques liées à ces modèles.

## 1.3 Autres modèles de populations structurées

### 1.3.1 Evolution de la diversité génétique d'un marqueur lié à un trait sous sélection

Les discussions avec Sylvain Billiard, Régis Ferrière et Sylvie Méléard ont fait ressortir que la modélisation jointe des interactions entre écologie, évolution et génétique des populations, était extrêmement importante pour reconstruire les phylogénies des espèces. En génétique des populations, les phylogénies sont reconstruites à partir de marqueurs neutres, c'est-à-dire de séquences du génome sans valeur sélective. Pour comprendre les phylogénies à l'échelle des populations et dans un cadre de dynamique adaptative, nous avons considéré, dans [BFMT14], une population caractérisée par un trait  $x$  et un marqueur  $u$ . Pour simplifier, nous supposons dans cette présentation que  $x \in \mathcal{X}$  un compact de  $\mathbb{R}$  et  $u \in \{A, a\}$ . Soit

$$X_t^K(dx, du) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{(x_i, u_i)}(dx, du).$$

Nous nous intéressons aux limites de la suite de processus  $(X_{t/K}^K)_{K \in \mathbb{N}^*}$  en grandes populations, avec des mutations rares des traits et des marqueurs, mais telles que les marqueurs mutent plus rapidement. Le processus est accéléré pour obtenir une limite décrivant l'évolution sur de grandes échelles de temps. Les résultats que nous obtenons nous apportent un éclairage nouveau sur les phylogénies que l'on cherche à inférer à partir du polymorphisme génétique au temps présent, par exemple.

Le marqueur  $u$  est neutre, c'est-à-dire que la dynamique ne dépend que de la structure de traits. Les individus se reproduisent au taux  $b(x)$  et meurent au taux  $d(x) + \eta(x)C * X_t^K(x) = d(x) + \frac{\eta(x)}{K} \sum_{i=1}^{N_t^K} C(x - x_i)$ . La reproduction est asexuée. Le marqueur est lié au trait. Au moment de la reproduction, le couple  $(x, u)$  est transmis du parent au descendant, sauf s'il se produit

une mutation du trait ou du marqueur. On suppose que les mutations des traits et marqueurs sont indépendantes et qu'elles se produisent respectivement avec les probabilités  $p_K = 1/K^2$  et  $q_K = p_K r_K$ , avec  $r_K/K \rightarrow \bar{r} > 0$  et  $q_K(\log K)^2 \rightarrow 0$ . L'hypothèse que les mutations du marqueur sont plus fréquentes que celles du trait, tout en restant à une échelle de temps plus lente que celle des naissances et morts, peut se justifier par le fait que la plupart des mutations sont neutres ou quasiment neutres, comme celles affectant les microsattellites, par exemple. Enfin, le noyau de mutation pour le trait est donné par la mesure de probabilité  $m(x, k)dk$ , tandis que le noyau de mutation du marqueur est donné par :

$$G_K(u, dv) = \mathbb{1}_{u=a} q_a \delta_A(dv) + \mathbb{1}_{u=A} q_A \delta_a(dv),$$

avec  $q_a + q_A = 1$ .

On va se placer sous les hypothèses suivantes :

**Hypothèse 1.3.1.** (i) "Invasion implique fixation". Pour tout  $x, y \in \mathbb{R}$ ,

$$\begin{aligned} \text{soit } \frac{b(y) - d(y)}{\eta(y)C(y-x)} &< \frac{b(x) - d(x)}{\eta(x)C(0)}, \\ \text{soit } \frac{b(y) - d(y)}{\eta(y)C(y-x)} &> \frac{b(x) - d(x)}{\eta(x)C(0)} \quad \text{et} \quad \frac{b(x) - d(x)}{\eta(x)C(x-y)} < \frac{b(y) - d(y)}{\eta(y)C(0)}. \end{aligned}$$

(ii) Conditions initiales. On suppose que l'on part d'une population monomorphique  $(x_0, u_0)$ ,  $X_0^K(dx, du) = n_0^K \delta_{(x_0, u_0)}(dx, du)$ , où  $\lim_{K \rightarrow \infty} n_0^K = \hat{n}_{x_0} = \frac{b(x_0) - d(x_0)}{\eta(x_0)C(0)}$  avec  $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^3) < +\infty$ .

L'hypothèse 1.3.1 (i) garantit que l'EDO approximante, en grande population et sans accélération du temps, n'admet pas d'équilibre stable avec deux traits. Autrement dit, lorsqu'un trait mutant  $y$  apparaît, soit sa descendance envahit la population et remplace le trait résident  $x$ , soit elle disparaît. La probabilité d'invasion est alors

$$\frac{[f(y; x)]_+}{b(y)} \quad \text{où } f(y; x) = b(y) - d(y) - \eta(y)C(y-x)\hat{n}_x$$

est le taux de croissance initial de la population de trait  $y$  dans la population résidente de trait  $x$  à l'équilibre  $\hat{n}_x$ , et où  $[f(y; x)]_+$  est la partie positive de  $f(y; x)$ . Sous cette hypothèse, l'évolution de la population sera une succession d'états monomorphiques en trait. Le processus de saut pur décrivant la suite des traits qui se succèdent dans la population est le *Trait Substitution Sequence* (TSS) introduit par Metz et al. [MGM<sup>+</sup>96] puis rigoureusement obtenu par Champagnat [Cha06]. Sans cette hypothèse, Champagnat et Méléard [CM11] ont montré que l'on pouvait remplacer le TSS par le *Polymorphic Evolutionary Sequence* (PES). La généralisation du résultat présenté ici au PES est décrite dans [BFMT14].

**Théorème 1.3.2.** *Sous les hypothèses 1.3.1, la suite  $(X_K^K)_{K \in \mathbb{N}^*}$  converge en loi vers le processus à valeurs dans  $\mathcal{M}_F(\mathcal{X} \times \{a, A\})$*

$$(\hat{n}_{Y_t} (W_t^a \delta_{(Y_t, a)}(dx, du) + (1 - W_t^a) \delta_{(Y_t, A)}(dx, du)), t \geq 0),$$

où  $(Y_t, t \geq 0)$  est le TSS qui saute de  $x$  à  $x+k$  avec la mesure de saut  $b(x) \hat{n}_x \frac{[f(x+k; x)]_+}{b(x+k)} m(x, k)dk$  et où  $(W_t^a, t \geq 0)$  est une diffusion avec sauts, qui décrit l'évolution de la proportion d'allèles  $a$ .  $W^a$  évolue entre deux sauts de  $Y$  comme la diffusion habituelle de Wright-Fisher avec mutations :

$$dW_t^a = \bar{r} b(Y_t) (q_A(1 - W_t^a) - q_a W_t^a) dt + \sqrt{\frac{2b(Y_t)}{\hat{n}_{Y_t}}} W_t^a (1 - W_t^a) dB_t \quad (1.3.1)$$

où  $(B_t, t \geq 0)$  est un mouvement Brownien standard. Le processus  $W^a$  saute avec le TSS  $Y$ , et, au temps  $t$  du saut,  $(W_t^a, 1 - W_t^a) = (1, 0)$  avec probabilité  $W_{t-}^a$  et  $(0, 1)$  avec probabilité  $1 - W_{t-}^a$ .



Au moment où le trait résident est remplacé par le trait mutant,  $W^a$  saute en  $(0, 1)$  ou  $(1, 0)$ . Ce phénomène, qualifié de prise en “auto-stop” par les biologistes (*hitchhiking* en anglais), a été mis en évidence par Kojima and Schaffer [KS67], Maynard Smith et Haigh [SH74], Ohta Kimura [OK75], Kaplan et al. [KHL89]. Au moment du remplacement de trait, la diversité neutre est “purgée” aux loci liés au trait sous sélection. Les marqueurs liés au trait envahisseur se fixent donc également, malgré leur neutralité. Il existe une littérature abondante sur l’impact de la sélection sur le polymorphisme neutre (Barton [Bar00], Etheridge et al. [EPW06], Durrett-Schweinsberg [DS05], Chevin et al. [CH08]...), mais qui suppose souvent la taille de la population et les coefficients de sélection constants, ou que la taille de la population évolue de façon indépendante de la valeur sélective des individus. Le résultat précédent montre une co-variation des traits, de la taille de la population et des termes exprimant la sélection.

La diffusion de Wright-Fisher (1.3.1) décrit l’évolution et la reconstitution de la diversité neutre une fois le remplacement effectué. On voit dans le facteur multiplicatif  $2b(Y_t)/\hat{n}_{Y_t}$  du coefficient de diffusion que les fluctuations dépendent du trait  $Y_t$ , malgré le fait qu’il s’agisse du marqueur neutre. Notons que la forme  $2b(Y_t)/\hat{n}_{Y_t}$  correspond à la formulation habituelle “variance/taille efficace” des biologistes, puisqu’ici  $2b(Y_t) = b(Y_t) + d(Y_t) + \eta(Y_t)C(0)\hat{n}_{Y_t}$  correspond à la variance induite par les naissances et morts. On voit ainsi l’importance de la prise en compte de la démographie et des dépendances en traits des paramètres.

Par ailleurs, lorsque l’on relaxe l’hypothèse d’“invasion implique fixation”, il apparaît que les fluctuations des fréquences du marqueur neutre dépendent de l’ensemble des traits présents dans la population. L’extinction ou l’apparition d’un trait a pour conséquence une variation de la dynamique de la diversité neutre, pour l’ensemble des sous-populations de traits.

L’un des projets que nous avons en cours est de proposer, à partir de ces théorèmes limites, de nouvelles façons d’inférer les phylogénies, en généralisant les méthodes fondées sur des coalescents ou des processus de branchement dont les taux seraient homogènes sur les branches correspondant aux individus de même trait, différents entre branches, et pouvant changer sur différents intervalles de temps.

*Idée de la preuve du Théorème 1.3.2.* Le point le plus délicat est de montrer que, pendant la période transitoire où la population mutante envahit et se fixe, la distribution du marqueur reste proche d’une masse de Dirac chargeant le marqueur  $u_0$  du mutant initial. Sous les hypothèses fixées, le nombre de mutations de marqueurs, chez les porteurs du trait mutant, n’est pas nécessairement négligeable pendant l’invasion. Cependant, comme la masse d’un arbre se concentre autour des feuilles, ces mutations se produisent surtout lorsque l’invasion est presque terminée. Les sous-populations de marqueurs associées sont alors petites par rapport à la sous-population de marqueur  $u_0$ .  $\square$

Le résultat du Théorème 1.3.2 peut être généralisé à des espaces discrets plus riches (par exemple, l’ensemble des séquences de  $n$  nucléotides choisis parmi  $\{A, T, G, C\}$ ), où même à des espaces continus. Dans ce cas, la diffusion de Wright-Fisher doit être remplacée par un processus de Fleming-Viot (cf. [BFMT14]).

### 1.3.2 Plantes à fleurs avec reproduction auto-incompatible

Avec Sylvain Billiard, Pauline Lafitte et Kilian Raschel, nous avons étudié des modèles de populations de plantes à fleurs avec reproduction auto-incompatible. La reproduction des plantes à fleurs est une reproduction sexuée. Les plantes produisent du pollen qui peut féconder les stigmates d’autres plantes. A priori, le pollen d’une plante pourrait féconder ses propres stigmates, et c’est ce qui se produit parfois. Pour empêcher cela, certaines espèces ont développé un mécanisme appelé l’auto-incompatibilité, qui permet à un individu de ne pas se reproduire qu’avec des partenaires

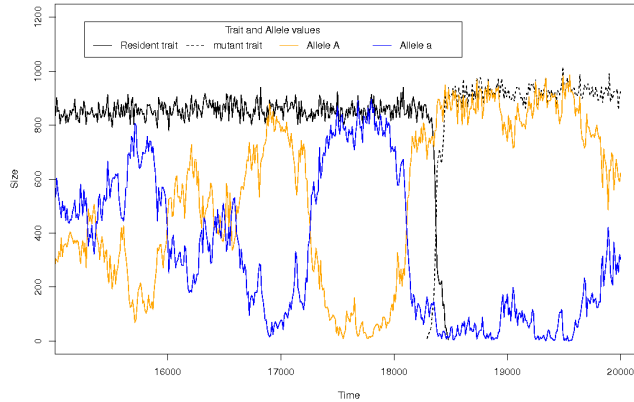


FIGURE 1.3.1 – Evolution des tailles des populations d'allèles  $a$  et  $A$  au cours du temps. Les fréquences des allèles  $a$  et  $A$  suivent une diffusion de Wright-Fisher tandis que la taille totale de la population fluctue autour d'un équilibre déterminé par le trait présent dans la population. Un mutant apparaît autour du temps 18290, envahit et se fixe. Au moment où le mutant apparaît, la fréquence d'allèles  $A$  est 85%, et il n'est pas surprenant que le mutant porte le marqueur  $A$ . Celui-ci bénéficie alors du phénomène d'auto-stop : vers le temps 18490, la fixation a eu lieu et il n'y a plus d'allèle  $a$  dans la population. Celui est alors régénéré par les mutations du marqueur.

compatibles.

Dans Billiard et Tran [BT12], un modèle général est introduit, mais pour la présentation, nous considérons un cas particulier, celui des plantes distyles et sans limitation de pollen. Chaque plante est diploïde et on les caractérise par leurs deux allèles au locus  $S$  intervenant dans le mécanisme d'auto-incompatibilité. Dans le cas distyle, seuls deux allèles existent, disons  $A$  et  $a$ . Les génotypes possibles pour les plantes sont donc  $AA$ ,  $Aa$  et  $aa$ . Le phénotype de la plante correspond aux protéines de reconnaissance portées par le pollen et les stigmates. Le seul cas intéressant est celui où  $A$  est dominant sur  $a$  (cf. [BT12]). Les phénotypes correspondant aux trois génotypes précédents sont alors respectivement  $A$ ,  $A$  et  $a$ .

Au moment de la rencontre, seules les protéines de phénotypes différents peuvent donner des graines viables. Ainsi, seules les rencontres de  $AA$  et  $aa$ , ou de  $Aa$  et  $aa$ , sont viables. Le pollen ou les stigmates sont des gamètes qui ne portent qu'un seul allèle. La rencontre de  $AA$  et  $aa$  donne un individu  $Aa$  avec probabilité 1 tandis que la rencontre de  $Aa$  avec  $aa$  donne un individu  $aa$  avec probabilité  $1/2$  et un individu  $Aa$  avec probabilité  $1/2$ . On voit donc que les graines  $AA$  ne sont jamais créées et les individus  $AA$  sont amenés à disparaître. On peut donc considérer sans restriction une population constituée uniquement des génotypes  $Aa$  (de phénotype  $A$ ), dont on notera  $X_t$  la taille au temps  $t$ , et  $aa$  (de phénotype  $a$ ), de taille  $Y_t$ .

L'évolution de la population est décrite par le processus de sauts pur  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  à valeurs dans  $\mathbb{N}^2$  (voir Fig. 1.3.4). On supposera que les ovules sont produits en temps continu, au taux  $r > 0$  et ont une durée de vie exponentielle de paramètre  $d > 0$  (pas de compétition). Les taux de transition de  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  sont décrits à la Fig. 1.3.3 (a), et :

$$\begin{aligned} X_t &= X_0 + \int_0^t \left( \frac{r}{2} (X_{s-} + Y_{s-}) \mathbb{1}_{X_{s-} > 0, Y_{s-} > 0} - d X_{s-} \right) ds + M_t^X, \\ Y_t &= Y_0 + \int_0^t \left( \frac{r}{2} (X_{s-} + Y_{s-}) \mathbb{1}_{X_{s-} > 0, Y_{s-} > 0} - d Y_{s-} \right) ds + M_t^Y \end{aligned} \quad (1.3.2)$$

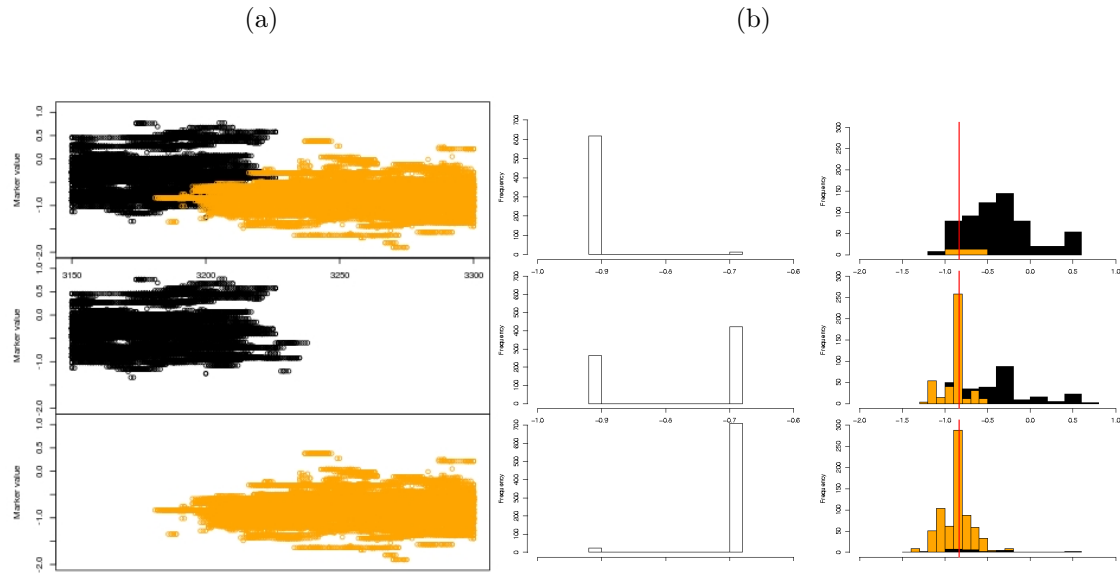


FIGURE 1.3.2 – Cas de marqueurs appartenant à un espace continu. On considère une population résidente monomorphe en trait (en noir) dans laquelle un trait mutant (en jaune) apparaît et se fixe. (a) Evolution du support de la distribution du marqueur en fonction du temps. Les populations mutantes sont représentées simultanément (graphique du haut) et séparément (milieu et bas). (b) Distributions des traits (gauche) et des marqueurs (droite) à trois instants de l'invasion. Lorsque le mutant apparaît (haut), pendant l'invasion, lorsque les deux populations résidentes et mutantes sont de tailles comparables (milieu) et lorsque la population résidente s'éteint (bas). On constate que pendant la durée de l'invasion, la distribution du marqueur, dans la population mutante (en jaune) reste très piquée autour de la valeur du marqueur du premier mutant, ce qui illustre l'idée expliquée dans la preuve du Théorème 1.3.2.

où  $(M_t^X, M_t^Y)_{t \in \mathbb{R}_+}$  est une martingale de carré intégrable que l'on peut écrire à partir de mesures de Poisson, et dont le crochet est calculé [BT12]. On peut associer au processus en temps continu une chaîne de Markov, dont les probabilités de transition sont représentées à la Fig. 1.3.3 (b). Par abus de notation, on appellera encore  $(X_t, Y_t)_{t \in \mathbb{N}}$  cette chaîne de Markov.

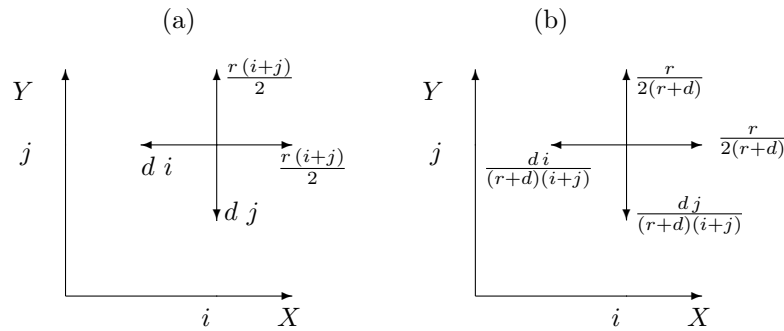


FIGURE 1.3.3 – (a) Taux de transition de  $(X_t, Y_t)_{t \in \mathbb{R}_+}$ . (b) Probabilités de transition de la chaîne de Markov associée, dans le quadrant positif.

On s'intéresse ici à déterminer quand ce processus est sous ou sur-critique, et à estimer les pro-

babilités d'extinction de la population dans ce dernier cas. Ces problématiques sont naturelles pour les biologistes, qui se demandent comment les phénomènes d'auto-incompatibilité, de dépression de consanguinité, de compétition, agissent sur l'évolution. En particulier, l'hypothèse que ce système de reproduction a été développé pour empêcher la dépression de consanguinité a été avancée, et des comparaisons ont été réalisées dans [BT12]. D'autres problèmes, faisant également intervenir des mutations, considèrent la fixation d'allèles délétères dans ces populations (e.g. Coron et al. [CMPR13, Cor14] ou Abu Awad et al. [ABT14], qui, dans sa thèse, étudie l'impact des mutations délétères récurrentes via les fardeaux génétique et numérique).

Mathématiquement, la difficulté de l'étude de  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  ou  $\mathbb{N}$  vient du fait que le processus est non-homogène dans l'espace, et absorbé aux bords du quadrant positif. Les techniques développées dans la littérature sont essentiellement valables pour des marches aléatoires homogènes (voir par exemple Fayolle et al. [FIM99], Klein Haneveld and Pittenger [HP90], Kurkova and Raschel [KR11], Walraevens et al. [WvLB10]).

Soit  $\tau = \inf\{t \in \mathbb{N}, X_t = 0 \text{ ou } Y_t = 0\}$  le nombre d'événements avant que l'un des deux génotypes  $Aa$  ou  $aa$  ne disparaisse. La perte de l'un des deux génotypes implique que plus aucune reproduction n'est possible et alors la population s'éteint. On remarque que l'extinction en temps discret est équivalente à l'extinction en temps continu :  $\{\tau < +\infty\} = \{\exists t \in \mathbb{R}_+, X_t = 0 \text{ ou } Y_t = 0\}$ . On note

$$p_{i,j} = \mathbb{P}(\tau < +\infty \mid X_0 = i, Y_0 = j) \quad (1.3.3)$$

Par symétrie,  $p_{i,j} = p_{j,i}$ . D'autre part,  $\forall i \in \mathbb{N}, p_{i,0} = p_{0,i} = 1$ .

**Proposition 1.3.3.** *Partant de la condition initiale  $(i, j) \in \mathbb{N}^* \times \mathbb{N}^*$ ,*

(i) *Cas sous-critique ou critique :  $\mathbb{P}(\tau_0 < +\infty) = 1$  si et seulement si  $r \leq d$ .*

(ii) *Cas sur-critique : Si  $r > d$ ,  $0 < p_{i,j} < 1$  avec :*

$$\left(\frac{d}{r}\right)^{i+j} \leq p_{i,j} \leq \left(\frac{d}{r}\right)^i + \left(\frac{d}{r}\right)^j - \left(\frac{d}{r}\right)^{i+j} \quad (1.3.4)$$

*Idée de preuve.* La preuve de la Proposition 1.3.3 s'appuie sur des couplages avec d'autres processus sur le quart de plan positif. Pour la partie (i) et la minoration de (ii), on peut dominer la population par un Galton-Watson bi-type  $(\tilde{X}_t, \tilde{Y}_t)_{t \in \mathbb{R}_+}$ , sans les indicatrices dans (1.3.2). Ce Galton-Watson ne s'éteint pas lorsqu'il atteint un bord, seulement lorsqu'il touche l'origine  $(0, 0)$ .

Pour la majoration de (ii), il faut minorer stochastiquement  $(X_t, Y_t)_{t \in \mathbb{R}_+}$ . Du fait de la forme des transitions (voir Fig. 1.3.3), il est naturel d'introduire un processus  $(\hat{X}_t, \hat{Y}_t)_{t \in \mathbb{R}_+}$  qui a les mêmes taux de reproduction que  $(X_t, Y_t)_{t \in \mathbb{R}_+}$ , mais où les individus  $Aa$  ou  $aa$  meurent avec le même taux  $\frac{d}{2}(X_{t-} + Y_{t-})$ . Heuristiquement, le processus  $(\hat{X}_t, \hat{Y}_t)_{t \in \mathbb{R}_+}$  a tendance à se déplacer davantage vers les bords du quadrant que  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  : dans l'octant inférieur  $\{i > j\}$ , le taux de mort des individus  $aa$  dans  $\hat{Y}_t$ ,  $d(i+j)/2$ , est supérieur à celui,  $dj$ , dans  $Y_t$  et de façon similaire, le taux de mort des  $Aa$  dans  $\hat{X}_t$  est supérieur à celui dans  $X_t$  lorsque l'on est dans l'octant supérieur  $\{i < j\}$ . Cependant, un couplage naïf ne marche pas, comme le montre la Fig. 1.3.5 (a). Il nous faut encadrer le processus  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  par deux copies du processus  $(\hat{X}_t, \hat{Y}_t)_{t \in \mathbb{R}_+}$ , notées  $(\hat{X}_t^-, \hat{Y}_t^-)_{t \in \mathbb{R}_+}$  et  $(\hat{X}_t^+, \hat{Y}_t^+)_{t \in \mathbb{R}_+}$ , symétriques par rapport à la première bissectrice. Le couplage a lieu avec la copie qui appartient au même octant que le processus.  $\square$

Dans le cas  $r > d$ , on peut chercher à être plus précis sur l'expression de  $p_{i,j}$ . Considérons la chaîne de Markov en temps discret. En conditionnant par rapport à  $t = 1$ , on montre que les  $p_{i,j}$  sont solutions du problème de Dirchlet suivant. Pour tous  $i, j \in \mathbb{N}^*$ ,

$$p_{i,j} = \frac{di}{(r+d)(i+j)} p_{i-1,j} + \frac{dj}{(r+d)(i+j)} p_{i,j-1} + \frac{r}{2(r+d)} p_{i,j+1} + \frac{r}{2(r+d)} p_{i+1,j} \quad (1.3.5)$$

avec les conditions au bord  $p_{i,0} = p_{0,i} = 1$ . (1.3.5) n'admet pas de solution unique (la fonction constante égale à 1 est solution, mais il y en a d'autres si  $r > d$ ).

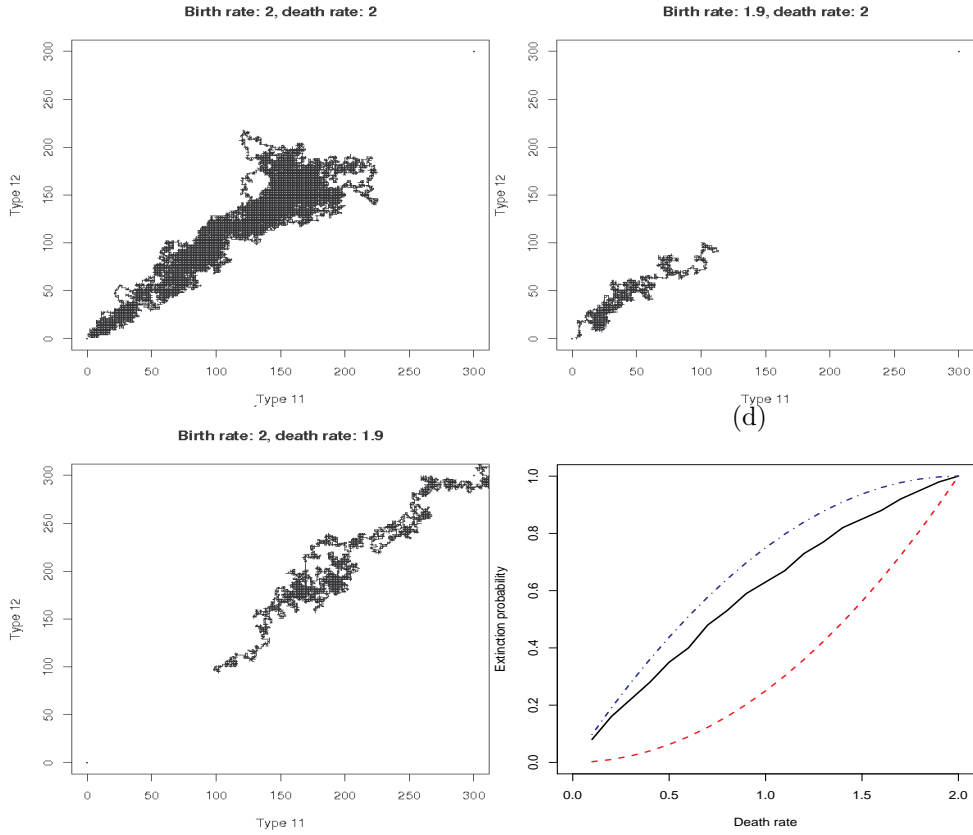


FIGURE 1.3.4 – (a)-(c) : Simulation des trajectoires de  $(X_t, Y_t)_{t \geq 0}$  dans les cas critique ( $r = d$ ), sous-critique ( $r < d$ ) et sur-critique ( $r > d$ ). Pour ces simulations, la condition initiale est  $(100, 100)$ . (d) : Pour la condition initiale  $(1, 1)$ , estimation par Monte-Carlo de la probabilité d'extinction de  $(X_t, Y_t)_{t \geq 0}$  (trait continu), de la probabilité d'extinction de  $(\tilde{X}_t, \tilde{Y}_t)_{t \geq 0}$  (pointillés rouges) et de  $(\hat{X}_t, \hat{Y}_t)_{t \geq 0}$  (traits-pointillés bleus).

**Proposition 1.3.4.** 1. Les probabilités d'extinction  $(p_{i,j})_{i,j \geq 1}$  constituent la plus petite solution positive de (1.3.5) avec les conditions au bord  $p_{i,0} = p_{0,i} = 1$ .

2. Si on se donne les probabilités d'extinction  $(p_{i,1})_{i \geq 1}$ , alors les probabilités  $(p_{i,j})_{i,j \geq 1}$  sont entièrement déterminées.

Considérons la fonction génératrice  $P(x, y) = \sum_{i,j \geq 1} p_{i,j} x^i y^j$ , définie à partir de la solution  $(p_{i,j})_{i,j \in \mathbb{N}}$ . Nous avons montré avec Pauline Lafitte-Godillon et Kilian Raschel [LGRT13] que  $P(x, y)$  était solution d'une EDP que nous pouvons résoudre formellement, en fonction des  $(p_{i,1})_{i \geq 1}$ . Malheureusement, l'expression que nous obtenons n'est pas exploitable numériquement.

**Proposition 1.3.5.** (i) La fonction  $P(x, y)$  est solution de :

$$AP(x, y) = h(x, y, P), \quad (1.3.6)$$

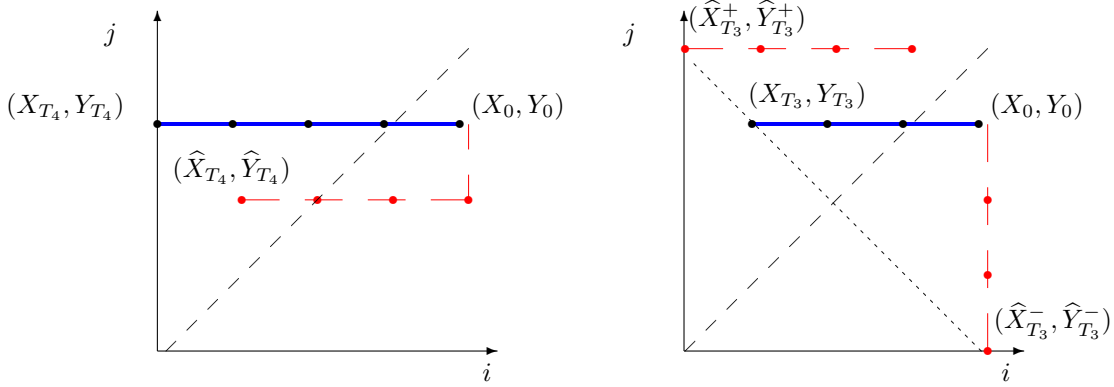


FIGURE 1.3.5 – (a) un exemple de couplage naïf entre le processus original  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  (trait continu bleu) et le processus auxiliaire  $(\hat{X}_t, \hat{Y}_t)_{t \in \mathbb{R}_+}$  (en pointillés rouge) où le processus original atteint le bord du quadrant positif avant le processus censé le minorer stochastiquement. Les deux processus sont issus de l’octant inférieur  $\{i > j\}$  et au premier pas, la marche aléatoire originale fait un pas vers la gauche tandis que le processus auxiliaire est attiré par l’axe des abscisses. Ensuite, le processus original se dirige vers la gauche, comme son couplage, et atteint donc l’axe des ordonnées en premier. (b) Couplage avec deux processus auxiliaires symétriques  $(\hat{X}_t^-, \hat{Y}_t^-)_{t \in \mathbb{R}_+}$  et  $(\hat{X}_t^+, \hat{Y}_t^+)_{t \in \mathbb{R}_+}$ , de même loi que  $(\hat{X}_t, \hat{Y}_t)_{t \in \mathbb{R}_+}$ . Le processus original est couplé avec la marche auxiliaire qui se trouve dans le même octant que lui. Comme  $(X_t, Y_t)_{t \in \mathbb{R}_+}$  est issu d’un point de l’octant inférieur, le couplage se fait initialement avec  $(\hat{X}_t^-, \hat{Y}_t^-)_{t \in \mathbb{R}_+}$ . Dès qu’il passe dans l’octant supérieur, le couplage se fait avec  $(\hat{X}_t^+, \hat{Y}_t^+)_{t \in \mathbb{R}_+}$ . Cette fois, on voit que les processus auxiliaires atteignent bien les bords du quadrant positif en premier.

avec

$$AP(x, y) = Q(x, y) \frac{\partial P}{\partial x}(x, y) + Q(y, x) \frac{\partial P}{\partial y}(x, y) + R(x, y)P(x, y), \quad (1.3.7)$$

$$Q(x, y) = (r + d)x - \frac{r}{2} - \frac{r}{2} \frac{x}{y} - dx^2, \quad R(x, y) = \frac{r}{2x} + \frac{r}{2y} - dx - dy,$$

$$h(x, y, P) = -\frac{r}{2} \left( x \frac{\partial^2 P}{\partial x \partial y}(x, 0) + y \frac{\partial^2 P}{\partial y \partial x}(0, y) \right) + dxy \left( \frac{1}{1-x} + \frac{1}{1-y} \right). \quad (1.3.8)$$

(ii) Etant donnée une suite  $(p_{i,1})_{i \geq 1}$ , il existe une unique solution de (1.3.6)-(1.3.8) sur  $]0, 1[ \times ]0, 1[$ .

Dans le cas de marches aléatoires homogènes, l’opérateur  $AP(x, y)$  sur lequel on tombe habituellement a la forme  $R(x, y)P(x, y)$  (cf. [KR11]). L’inhomogénéité nous fait donc passer d’équations algébriques à des EDP. Remarquons que la fonction  $h$  dans (1.3.8) ne dépend que des bords ( $x = 0$  ou  $y = 0$ ), ce qui fait de cette EDP une EDP non classique.

L’opérateur  $A$  est du premier ordre et donc lié à des équations de transport. Il est donc naturel de chercher à identifier les courbes caractéristiques liées à cet opérateur,

$$\begin{cases} \dot{x}_s &= \frac{dx}{ds}(s) = Q(x_s, y_s), \\ \dot{y}_s &= \frac{dy}{ds}(s) = Q(y_s, x_s), \end{cases} \quad (1.3.9)$$

**Proposition 1.3.6.** Pour toute condition initiale  $(x_0, y_0) \in \mathbb{R}^2$  telle que  $x_0, y_0 \neq 0$ , il existe une

unique solution à (1.3.9), définie pour  $s \in \mathbb{R}$  par :

$$x_s = \frac{\lambda r e^{rs} + \mu d e^{ds}}{d(\lambda e^{rs} + \mu e^{ds} + 1)}, \quad y_s = \frac{\lambda r e^{rs} + \mu d e^{ds}}{d(\lambda e^{rs} + \mu e^{ds} - 1)}, \quad (1.3.10)$$

où :

$$\lambda = \frac{2d x_0 y_0 - d(x_0 + y_0)}{(x_0 - y_0)(r - d)}, \quad \mu = \frac{-2d x_0 y_0 + r(x_0 + y_0)}{(x_0 - y_0)(r - d)}. \quad (1.3.11)$$

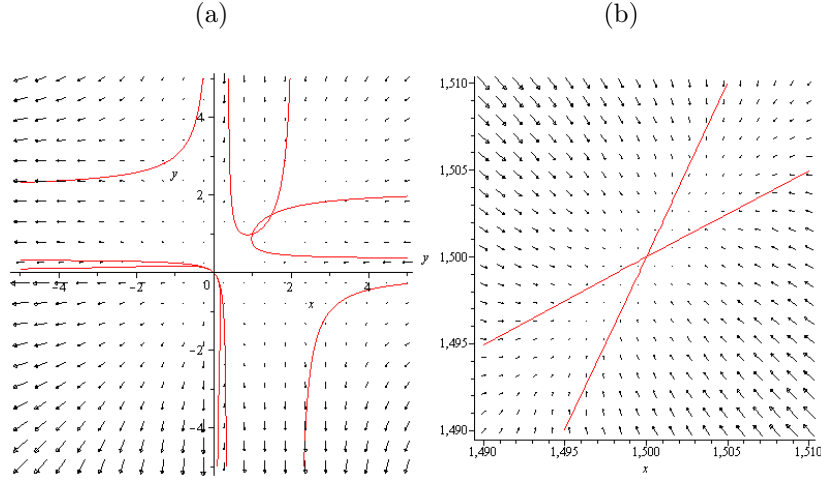


FIGURE 1.3.6 – (a) Champ de vecteur  $(Q(x, y), Q(y, x))$  associé au système (1.3.9). Les lignes pleines correspondent aux ensembles  $\{Q(x, y) = 0\}$  et  $\{Q(y, x) = 0\}$ . (b) Voisinage de l'équilibre attractif  $(r/d, r/d)$ .

$P(x, y)$  est solution de (1.3.6) si  $g_s = P(x_s, y_s)$  est solution de  $\dot{g}_s + R(x_s, y_s)g_s = h(x_s, y_s, P)$ . Cette équation ressemble à une équation différentielle ordinaire (EDO), excepté que  $h$  dépend de la condition au bord de  $P$ . On obtient grâce à cela une expression pour la solution :

**Proposition 1.3.7.** Soit  $(x_0, y_0) \in (0, 1)^2$ , et soit  $s_0 = \log(-\frac{\mu d}{\lambda r})/(r - d) > 0$ . Alors :

$$\begin{aligned} P(x_0, y_0) &= \frac{r}{2} \sum_{i \geq 1} p_{i,1} \int_0^{s_0} i(x_u)^i e^{\int_0^u R(x_\alpha, y_\alpha) d\alpha} du \\ &+ \frac{r}{2} \sum_{j \geq 1} p_{1,j} \int_0^{s_0} j(y_u)^j e^{\int_0^u R(x_\alpha, y_\alpha) d\alpha} du \\ &- d \int_0^{s_0} x_u y_u \left( \frac{1}{1 - x_u} + \frac{1}{1 - y_u} \right) e^{\int_0^u R(x_\alpha, y_\alpha) d\alpha} du. \end{aligned} \quad (1.3.12)$$

On peut montrer que le membre de gauche de (1.3.12) est bien défini et ne dépend que des  $(p_{i,1})$ . Le temps  $s_0 \in (0, +\infty)$  correspond au temps tel que  $\lim_{s \rightarrow s_0} (x_s, y_s) = (0, 0)$ . Comme  $(x_0, y_0) \in (0, 1)^2$  implique que pour tout  $u \in (0, s_0)$ ,  $(x_u, y_u) \in (0, 1)^2$ , il est possible d'injecter dans (1.3.12) des développements limités des  $p_{i,1}$ , par exemple  $p_{1,i} = 2d/(ri) + o(1/i)$  lorsque

$i \rightarrow +\infty$ . Alors, pour  $I_0$  suffisamment grand,

$$\begin{aligned}
 P(x_0, y_0) &= \frac{r}{2} \sum_{i=1}^{I_0} p_{i,1} \int_0^{s_0} i((x_u)^i + (y_u)^i) e^{\int_0^u R(x_\alpha, y_\alpha) d\alpha} du \\
 &\quad + d \int_0^{s_0} \left( \frac{x_u(x_u^{I_0} - y_u)}{1 - x_u} + \frac{y_u(y_u^{I_0} - x_u)}{1 - y_u} \right) e^{\int_0^u R(x_\alpha, y_\alpha) d\alpha} du + o(x_0^{I_0+1} + y_0^{I_0+1}). \quad (1.3.13)
 \end{aligned}$$

Cette dernière expression justifie que numériquement, on peut se restreindre à la prise en compte des  $p_{i,1}$  pour un nombre fini d'entiers  $i$ . Des simulations sont réalisées dans [LGR13].





## Chapitre 2

# Modèles SIR stochastiques en épidémiologie

### Sommaire

---

<b>2.1</b>	<b>Estimation pour des modèles SIR compartimentaux . . . . .</b>	<b>58</b>
2.1.1	Modèle individu-centré et EDO de Kermack-McKendrick . . . . .	58
2.1.2	Maximum de vraisemblance avec données complètes . . . . .	60
2.1.3	Estimation par ABC en présence de données manquantes . . . . .	61
<b>2.2</b>	<b>Modèles SIR évoluant sur un graphe . . . . .</b>	<b>66</b>
2.2.1	Epidémies se propageant sur un modèle de configuration et équations de Volz . . . . .	66
2.2.2	Prise en compte de la structure de graphe dans l'épidémie cubaine . . .	73
<b>2.3</b>	<b>Conclusion . . . . .</b>	<b>75</b>

---

L'épidémie du SIDA à Cuba présente plusieurs spécificités. Le SIDA a été introduit à Cuba en 1986 et se propage essentiellement par voie sexuelle. Depuis le début de l'épidémie et jusqu'en 2006, une base de donnée recense les individus dépistés séropositifs ainsi que leurs partenaires sexuels des deux dernières années. Cette information sur les partenaires est recueillie dans le cadre du dépistage par *contact-tracing*, qui consiste à rechercher les individus à risque dans l'entourage des nouveaux séropositifs pour leur proposer un test de dépistage. En conséquence, en plus des caractéristiques démographiques, sociologiques et médicales habituelles (âge, sexe, lieu de résidence, orientation sexuelle...) nous disposons d'indications nous permettant de relier les individus et de connaître partiellement le réseau social le long duquel la maladie s'est propagée. Parallèlement, les méthodes de dépistage usuelles (méthodes dites "aléatoires") sont également utilisées. Nous disposons ainsi d'une information souvent absente dans le cas des pays d'Afrique ou d'Amérique du Sud. Bien sûr, les observations manquantes sont inévitables puisque par définition seuls les individus dépistés et séropositifs figurent dans la base de donnée ; les partenaires sains ne sont pas enregistrés.

Dans un premier temps, nous avons considéré un modèle compartimental dans [CTA08, BT10]. Le but était d'étudier des pendants aléatoires aux modèles compartimentaux déterministes étudié par De Arazoza et ses co-auteurs sur le sujet [HALJ05, HWA<sup>+</sup>05, ALHI00, AL96, LA99] en introduisant de l'aléa pour enrichir la modélisation. Avoir des modèles aléatoires individu-centrés permettait également de prendre en compte des covariables propres à chaque individu. Par exemple, dans [CTA08] nous associons à chaque individu dépisté le temps depuis lequel il est dépisté, ce qui permet de prendre en compte la contribution de chacun de ces individus au système de *contact-tracing* : un individu dépisté séropositif depuis 2 semaines n'a par exemple pas le même rôle qu'un

individu dépisté depuis 5 ans. Enfin, les modèles individu-centrés fournissent un formalisme très adapté aux statistiques. La calibration des modèles, la construction d’intervalles de confiance et les tests jouent bien sûr un rôle important en épidémiologie et en santé publique pour la prise de décision. Dans [CTA08], nous utilisons les 6 premières années disponibles dans les données et qui sont “à peu près complètes”, c’est-à-dire que tous les séropositifs non dépistés de l’époque (1986-1992) peuvent être considérés comme trouvés en 2006. Nous utilisons alors des méthodes de maximum de vraisemblance pour estimer les paramètres du modèle considéré. Dans [BT10], nous utilisons cette fois le jeu de données entier, et prenons en compte le fait que la population des infectieux n’est que partiellement observée. Nous utilisons une méthode Bayésienne, l’*Approximate Bayesian Computation* (ABC) proposée initialement pour pratiquer l’inférence en génétique des populations [TBGD97, BZB02]. Cette méthode n’est pas fondée sur la vraisemblance, mais sur des simulations numériques et sur la comparaison entre des statistiques simulées et observées.

Pour des maladies sexuellement transmissibles, l’hypothèse de mélange, où n’importe quel infectieux peut contaminer n’importe quel susceptible, est une hypothèse forte, que l’on va pouvoir raffiner avec les données dont on dispose. La prise en compte de la structure de graphe sous-jacente nous permet de mieux comprendre les mécanismes de propagation de l’épidémie, ainsi que de reconstruire son histoire. Ceci est important pour modéliser l’évolution future et faire des prédictions, dans des contextes de santé publique. Dans la Section 2.2, nous commençons par étudier un modèle théorique inspiré de Volz [Vol08]. Sous l’hypothèse que le graphe est généré par un modèle de configuration, et en considérant des quantités arête-centrées plutôt que noeud-centrées pour décrire l’évolution d’un modèle SIR, Volz a réussi à résumer l’évolution de l’épidémie à 5 EDO. Certains des arguments de Volz ne sont pas rigoureux mathématiquement. Il apparaît que les objets naturellement sous-jacents dans le raisonnement de Volz sont des distributions de degré. Avec Laurent Decreusefond, Jean-Stéphane Dhersin et Pascal Moyal nous avons reformulé le problème en terme de processus à valeurs mesures et avons prouvé rigoureusement les théorèmes limites qui liaient le modèle SIR individu-centré décrivant la propagation de l’épidémie sur un graphe de configuration et les EDO de Volz.

Bien sûr, il est naturel de se demander dans quelle mesure l’hypothèse de modèle de configuration est valable pour des graphes réels. Avec Stéphan Cléménçon, Hector De Arazoza et Fabrice Rossi, nous avons exploré le graphe obtenu à partir des données de *contact-tracing* dans la population Cubaine [CART10]. Nous présentons ces résultats en fin de chapitre.

## 2.1 Estimation pour des modèles SIR compartimentaux

Un problème important, pour les prévisions et pour l’explication des données collectées, est l’estimation des modèles que le mathématicien va utiliser. Les données Cubaines sont très difficiles à traiter car de grande dimension, avec une importante part d’information inobservée. Dans un but de simplification, nous allons présenter les problèmes et méthodes statistiques employées sur un modèle simple, le modèle SIR, plutôt que le modèle introduit dans [CTA08]. Nous renvoyons à [Bai50, AM91, AB00] pour une revue et une introduction.

### 2.1.1 Modèle individu-centré et EDO de Kermack-McKendrick

Nous considérons, en temps continu, une population de taille fixe  $n$ . La population est divisée en 3 compartiments, qui regroupent respectivement les individus sains et susceptibles d’être infectés, les individus infectieux, les individus rémis ou retirés (morts ou guéris) qui ne peuvent plus être infectés. On note respectivement  $S_t^n$ ,  $I_t^n$  et  $R_t^n$  les tailles de ces populations au temps  $t \geq 0$ . Les proportions de susceptibles, infectieux et rémis seront notées  $s_t^n = S_t^n/n$ ,  $i_t^n = I_t^n/n$  et  $r_t^n = R_t^n/n$ .

Au temps  $t$ , les individus susceptibles peuvent être infectés au taux (global)  $\lambda^n S_t^n I_t^n$ , avec  $\lambda^n = \lambda/n$ . On remarque que le taux d’infection est proportionnel au nombre  $S_t^n I_t^n$  de couples que l’on peut faire avec une personne susceptible et une personne infectieuse. De façon sous-jacente

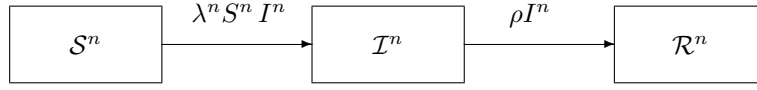


FIGURE 2.1.1 – Modèle SIR

se trouve donc l’hypothèse qu’une personne infectieuse peut infecter n’importe quelle personne susceptible au taux  $\lambda^n$  et indépendamment de toutes autres possibilités de propager l’infection. La division de  $\lambda$  par  $n$  signifie que le risque est plus faible dans des grandes populations. Un taux global d’infection de la forme  $\lambda^n S_t^n I_t^n$  est qualifié de *mass action* dans la littérature. Il existe aussi des modèles où le taux global d’infection est fréquence-dépendant,  $\lambda I_t^n$  ou  $\lambda S_t^n I_t^n / (S_t^n + I_t^n)$ , signifiant que le taux d’infection individuel n’est pas proportionnel au nombre de personnes infectieuses, mais à leur proportion.

**Remarque 2.1.1.** Le nombre d’infections dues à une personne joue un rôle particulier en épidémiologie. Pour un taux d’infection  $\lambda S_t^n I_t^n / (S_t^n + I_t^n)$ , chaque individu infectieux rencontre une autre personne au taux  $\lambda$  et cette rencontre a lieu avec un susceptible avec probabilité  $S_t^n / (S_t^n + I_t^n)$ . Pour un taux  $\lambda S_t^n I_t^n = \lambda (S_t^n + I_t^n) \frac{S_t^n}{S_t^n + I_t^n} I_t^n$ , les rencontres ont lieu au taux  $\lambda (S_t^n + I_t^n)$ , et la personne rencontrée est infectieuse avec probabilité  $I_t^n / (S_t^n + I_t^n)$ . Dans la première forme, il y a une “limitation” des rencontres, alors que pour un taux *mass action*, il y a d’autant plus de rencontres que la population est grande. Le remplacement de  $\lambda$  par  $\lambda/n$  permet de retrouver un ordre de grandeur similaire au premier modèle, avec un taux de rencontre  $\lambda (s_t^n + i_t^n)$ .  $\square$

Dans cette section, nous considérerons un modèle “mass action” pour le taux d’infection, mais les raisonnements sont reproductibles pour une forme fréquence-dépendante.

Chaque personne infectieuse est retirée au taux individuel  $\rho$ , indépendamment des autres personnes infectieuses, si bien qu’au niveau de la population, les événements de rémission ont lieu au taux  $\rho I_t^n$  (voir Fig. 2.1.1).

Comme pour le chapitre précédent, nous pouvons décrire l’évolution de cette population au moyen d’EDS dirigées par des mesures ponctuelles de Poisson. On suppose que la condition initiale est donnée par  $(S_0^n, I_0^n)$ , avec  $S_0^n + I_0^n = n$  et  $R_0^n = 0$ . La population étant de taille fixe  $n$ , avec pour tout  $t$ ,  $S_t + I_t + R_t = n$ , les taux globaux d’infection et de rémission sont majorés respectivement par  $\lambda n^2$  et  $\mu n$  ce qui implique un contrôle direct de tous les moments de  $(S_t^n, I_t^n, R_t^n)_{t \geq 0}$ .

**Proposition 2.1.2.** (i) Le processus  $(M_t^{S,n}, M_t^{I,n}, M_t^{R,n})_{t \geq 0}$  défini par

$$\begin{aligned}
 M_t^{S,n} &= s_t^n - s_0^n + \int_0^t \lambda s_u^n i_u^n du \\
 M_t^{I,n} &= i_t^n - i_0^n - \int_0^t (\lambda s_u^n i_u^n - \rho i_u^n) du \\
 M_t^{R,n} &= r_t^n - \int_0^t \rho i_u^n du
 \end{aligned}$$

est une martingale de carré intégrable, issue de  $(0, 0, 0)$  et dont les processus de covariance sont :

$$\begin{aligned}\langle M^{S,n} \rangle_t &= \frac{1}{n} \int_0^t \lambda s_u^n i_u^n du, & \langle M^{I,n} \rangle_t &= \frac{1}{n} \int_0^t (\lambda s_u^n i_u^n + \rho i_u^n) du, \\ \langle M^{R,n} \rangle_t &= \frac{1}{n} \int_0^t \rho i_u^n du, & \langle M^{S,n}, M^{I,n} \rangle_t &= -\frac{1}{n} \int_0^t \lambda s_u^n i_u^n ds, \\ \langle M^{I,n}, M^{R,n} \rangle_t &= -\frac{1}{n} \int_0^t \rho i_u^n du, & \langle M^{S,n}, M^{R,n} \rangle_t &= 0.\end{aligned}$$

(ii) Soit  $T > 0$ . Si  $\lim_{n \rightarrow +\infty} s_0^n = s_0 > 0$  et  $\lim_{n \rightarrow +\infty} i_0^n = i_0 > 0$  en probabilité, alors, lorsque  $n \rightarrow +\infty$ , la suite de processus  $(s^n, i^n, r^n)_{n \in \mathbb{N}^*}$  converge en probabilité dans  $\mathbb{D}([0, T], \mathbb{R}_+^3)$  vers l'unique solution du système d'équations différentielles ordinaires :

$$\frac{ds}{dt}(t) = -\lambda s(t)i(t), \quad s(0) = s_0, \quad (2.1.1)$$

$$\frac{di}{dt}(t) = \lambda s(t)i(t) - \rho i(t), \quad i(0) = i_0 \quad (2.1.2)$$

$$\frac{dr}{dt}(t) = \rho i(t), \quad r(0) = 0. \quad (2.1.3)$$

La démonstration de (i) et (ii) suit les étapes vues dans le Chapitre 1, preuve du Théorème 1.1.14. Le système (2.1.1)-(2.1.3) a été introduit par Kermack et McKendrick [KM27]. Il n'est pas résoluble explicitement, mais on peut par exemple exprimer en  $s(t)$  et  $i(t)$  en fonction de  $r(t)$  :

$$s(t) = s_0 e^{-\frac{\lambda}{\rho} r(t)}, \quad i(t) = s_0 (1 - e^{-\frac{\lambda}{\rho} r(t)}) + i_0 - r(t).$$

De plus,  $s(t)$  et  $r(t)$  sont des fonctions respectivement décroissantes et croissantes du temps, à valeurs dans  $[0, 1]$ . On en déduit donc que ces fonctions, et donc également  $i(t)$  puisque  $s(t) + i(t) + r(t) = 1$ , admettent une limite lorsque  $t \rightarrow +\infty$ . L'épidémie s'arrête lorsque  $i(t) = 0$ . Donc  $(s(t), i(t), r(t))$  converge vers  $(s_\infty, 0, r_\infty) \in [0, 1]^3$  avec  $s_\infty + r_\infty = 1$  et  $0 = s_0(1 - e^{\lambda r_\infty / \rho}) - r_\infty$ , ce qui caractérise entièrement la limite.

Il est également possible d'établir le théorème central limite suivant associé à la convergence :

**Proposition 2.1.3.** *Soit  $T > 0$ . Définissons*

$$\eta_t^{S,n} = \sqrt{n}(s_t^n - s_t), \quad \text{et} \quad \eta_t^{I,n} = \sqrt{n}(i_t^n - i_t). \quad (2.1.4)$$

Supposons que  $(\eta_0^{S,n}, \eta_0^{I,n})$  converge en loi, lorsque  $n \rightarrow +\infty$  vers  $(\eta_0^s, \eta_0^i)$ . Alors la suite de processus  $(\eta_t^{S,n}, \eta_t^{I,n})_{n \in \mathbb{N}^*}$  converge en loi, dans  $\mathbb{D}([0, T], \mathbb{R}^2)$  vers le processus gaussien caractérisé comme l'unique solution de l'EDS suivante, issue de  $(\eta_0^s, \eta_0^i)$  :

$$\begin{aligned}d\eta_t^s &= -\lambda(i_t \eta_t^s + s_t \eta_t^i)dt - \sqrt{\lambda i_t s_t} dB_t^1 \\ d\eta_t^i &= \lambda(i_t \eta_t^s + s_t \eta_t^i)dt + \rho \eta_t^i dt + \sqrt{\lambda i_t s_t} dB_t^1 - \sqrt{\gamma i_t} dB_t^2.\end{aligned} \quad (2.1.5)$$

## 2.1.2 Maximum de vraisemblance avec données complètes

On peut utiliser les résultats précédents pour établir la convergence et la normalité asymptotique d'estimateurs du maximum de vraisemblance (cf. [CTA08] dans un cas plus complexe) dans le cas où l'on dispose de données complètes. Supposons que l'on connaisse à  $t$  et pour chaque individu, les temps éventuels d'infection et de rémission. La vraisemblance des observations peut alors se définir comme étant la densité sur  $\mathbb{D}([0, T], \mathbb{R}_+^3)$  de la loi du processus  $(S^n, I^n, R^n)$  par rapport à la loi du processus d'infections-rémissions pour lequel les infections et les rémissions se produiraient aux taux

(globaux) de  $n/2$  (cf. [CTA08]). Ceci nous fournit les estimateurs du maximum de vraisemblance suivants pour  $(\lambda, \rho)$  :

$$\widehat{\lambda}_n = \frac{1}{n} \frac{\text{Card}\{\text{infections}\}}{\int_0^T s_s^n i_s^n ds}, \quad \widehat{\rho}_n = \frac{1}{n} \frac{\text{Card}\{\text{rémissions}\}}{\int_0^T i_s^n ds}. \quad (2.1.6)$$

A l'aide des théorèmes limites précédents, on peut montrer que lorsque  $n \rightarrow +\infty$ , ces estimateurs sont convergents et normalement asymptotiques :

$$\sqrt{n} \begin{pmatrix} \widehat{\lambda}_n - \lambda \\ \widehat{\rho}_n - \rho \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0_{\mathbb{R}^2}, I^{-1}(\lambda, \rho)), \quad (2.1.7)$$

où l'information de Fisher  $I(\lambda, \rho)$  s'écrit en fonction des solutions  $(s_t, i_t, r_t)_{t \in [0, T]}$  de (2.1.1)-(2.1.3) :

$$I(\lambda, \rho) = \begin{pmatrix} \frac{\int_0^T s_s i_s ds}{\lambda} & 0 \\ 0 & \frac{\int_0^T i_s ds}{\rho} \end{pmatrix}.$$

Les estimateurs  $\widehat{\lambda}_n$  et  $\widehat{\rho}_n$  sont calculables, à condition que les processus  $S^n$  et  $I^n$  soient complètement observés sur  $[0, T]$ , impliquant en particulier la connaissance des nombres d'infections et de rémissions sur l'intervalle de temps  $[0, T]$ .

### 2.1.3 Estimation par ABC en présence de données manquantes

Le problème des données manquantes est un problème récurrent en épidémiologie. Les méthodes *Markov Chain Monte-Carlo* (MCMC) sont devenues extrêmement populaires pour traiter ces problèmes [OR99, O'N02, CCV<sup>+</sup>04] mais peuvent se révéler très coûteuses numériquement lorsque le jeu de données est de grande taille, voir par exemple [CF04, SSF09]. Dans [SSF09], l'estimation MCMC requiert plusieurs jours sur un système parallélisé. C'est le cas pour la base cubaine (8662 individus observés entre 1986 et 2006),

Avec Michaël Blum [BT10], nous avons étudié l'application de la méthode ABC, alternative aux MCMC et mieux adaptée au format et à la taille de nos données. Des essais numériques sur nos données avec des méthodes MCMC n'ont pas abouti.

#### 2.1.3.1 Principe de l'ABC

Rappelons le principe de l'ABC (cf. [MPRR12] pour une revue). Soient  $\mathbf{x}$  les données dont on dispose (par exemple le processus  $(R_t)_{t \in [0, T]}$  et éventuellement une partie des infectieux) et soit  $\theta = (\lambda, \rho)$  les paramètres du modèle SIR que l'on cherche à estimer. L'ABC est une méthode Bayésienne. Soit  $\pi(d\theta) = \pi(\theta)d\theta$  la loi *a priori* de  $\theta$  sur  $\mathbb{R}_+^2$ . Dans la suite, par souci de simplicité pour la présentation, on supposera que les lois sont absolument continues par rapport à la mesure de Lebesgue. Le principe de l'ABC repose sur 2 approximations.

**Remplacement des observations par des descripteurs** (*summary statistics* en anglais). Au lieu de chercher à estimer la densité *a posteriori*  $\pi(\theta | \mathbf{x})$ , le but de l'ABC est d'approcher la densité *a posteriori* partielle (ou cible)  $p(\theta | S(\mathbf{x}) = s_{obs}) \propto \Pr(s_{obs} | \theta) \pi(\theta)$  où  $S$  est une statistique descriptive à valeurs dans un espace normé et où  $s_{obs}$  est la valeur observée, sur les données, de cette statistique  $S$ . Dans les cas que nous considérerons, la statistique descriptive  $S$  pourra être un vecteur de dimension  $d$  ou une variable infini-dimensionnelle comme une fonction  $L^1$  par exemple. Dans le cas où  $S$  est une statistique exhaustive, les densités *a posteriori* et cible coïncident.

**Approximations basées sur les simulations** Une fois que la statistique descriptive choisie, la seconde approximation consiste à approcher la densité *a posteriori* partielle  $p(\theta | S(\mathbf{x}) = s_{obs})$  par

des estimateurs à noyaux, en se fondant sur des simulations.

L'algorithme initialement proposé par [BZB02] est le suivant :

1. On génère  $N$  couples indépendants  $(\theta_i, s_i)$ ,  $i = 1, \dots, N$ . Pour cela, on tire le paramètre  $\theta_i$  dans la loi *a priori*  $\pi$ . Puis, on simule des données  $\mathbf{x}^i$  en utilisant ce paramètre  $\theta_i$  et on calcule la statistique descriptive  $S$  avec ces données, ce qui nous fournit  $s_i$ .
2. On associe à la  $i$ ème simulation le poids  $W_i = K_\delta(s_i - s_{obs})$ , où  $\delta$  est un seuil de tolérance et  $K_\delta$  un noyau.
3. La distribution  $(\sum_{i=1}^N W_i \delta_{\theta_i}) / (\sum_{i=1}^N W_i)$  est alors une approximation de la loi *a posteriori* partielle.

Le paramètre  $\delta$  sera choisi au vu d'études préliminaires des données *in silico* dont les paramètres sont connus. Pour le noyau  $K$ , plusieurs choix sont possibles (Gaussien, Epanechnikov etc.). Remarquons que si l'on choisit pour  $K$  une indicatrice,  $K_\delta(u) = \frac{\mathbb{1}_{|u| \leq \delta}}{2\delta}$ , alors, la pondération agit comme une procédure de rejet : on attribue un poids  $W_i = 1/(2\delta)$  si la simulation  $i$  nous donne  $s_i$  à distance  $\delta$  de  $s_{obs}$  et  $W_i = 0$  sinon.

Une fois que l'on a approché la distribution *a posteriori* partielle, on peut obtenir les estimateurs ponctuels de  $\lambda$  et  $\rho$ . Par exemple, si l'on note  $\theta_i = (\lambda_i, \rho_i)$ , on a pour  $\lambda$  :

$$\tilde{\lambda}_N = \frac{\sum_{i=1}^N \lambda_i W_i}{\sum_{i=1}^N W_i} = \frac{\sum_{i=1}^N \lambda_i K_\delta(s_i - s_{obs})}{\sum_{i=1}^N K_\delta(s_i - s_{obs})}, \quad (2.1.8)$$

qui est l'estimateur de Nadaraya-Watson de l'espérance conditionnelle  $\mathbb{E}(\lambda | s_{obs})$ . Notons qu'au lieu de considérer l'espérance de la loi *a posteriori* partielle, on peut considérer sa médiane ou son mode. Il est aussi possible d'obtenir des intervalles de confiance à 95% dans la distribution *a posteriori* partielle estimée.

Aux deux approximations mentionnées précédemment, peut s'ajouter une troisième astuce numérique lorsque  $S$  est un vecteur. L'estimateur de l'espérance conditionnelle (2.1.8) est un estimateur convergent si le seuil de tolérance satisfait  $\lim_{N \rightarrow +\infty} \delta_N = 0$  (pour que le biais tende vers 0) et si  $\lim_{N \rightarrow +\infty} N\delta_N^d = +\infty$  (afin que la variance converge vers 0) [Fan92]. Lorsque  $d$  croît, le seuil de tolérance doit être choisi plus grand pour que la variance soit contrôlée, ce qui a pour conséquence de faire tendre le biais plus lentement vers 0. Ce phénomène s'appelle la "malédiction de la dimension". La correction proposée par Beaumont et al. [BZB02] a pour but de remédier à ce problème. L'objectif des corrections est de produire à partir du couple  $(\theta_i, s_i)$  une nouvelle variable aléatoire dont la distribution serait  $p(\theta | s_{obs})$ . L'idée est de construire un couplage entre les distributions  $p(\theta | s_i)d\theta$  et  $p(\theta | s_{obs})d\theta$ . Les corrections se réalisent composante de  $\theta$  par composante. Pour un paramètre  $\theta$  réel, la correction s'obtient en supposant qu'il existe entre  $\theta$  et le descripteur  $s$  une relation de la forme

$$\theta = G(s, \varepsilon) =: G_s(\varepsilon), \quad (2.1.9)$$

pour une fonction  $G$  et une variable aléatoire  $\varepsilon$  dont la loi ne dépend pas de  $s$ . Par exemple, l'inverse (généralisée)  $G_s = F_s^{-1}$  de la fonction de répartition de  $p(\theta | s)d\theta$  convient. Dans ce cas,  $\varepsilon = F_s(\theta)$  est une variable aléatoire uniforme sur  $[0, 1]$ . Etant donnée cette relation (2.1.9), on peut remplacer  $\theta_i$  par

$$\theta_i^* = G_{s_{obs}}^{-1}(G_{s_i}(\theta_i)) \quad i = 1, \dots, N, \quad (2.1.10)$$

qui sont bien i.i.d. de loi  $p(\theta | s_{obs})d\theta$ . En pratique, on ne peut pas choisir  $G_s = F_s^{-1}$  car il faut choisir en pratique une fonction  $G$  facile à estimer. Beaumont et al. [BZB02] proposent une correction par régressions linéaires locales (LOCL), en supposant que l'on peut choisir pour  $G$  un modèle linéaire, au moins au voisinage de  $s_{obs}$  :  $G(s, \varepsilon) = \alpha + s^t \beta + \varepsilon$ . On peut estimer les paramètres  $\alpha$  et  $\beta$  par

moindres carrés pondérés en minimisant  $\sum_{i=1}^N K_\delta(s_i - s_{obs})(\theta_i - (\alpha + (s_i - s_{obs})^T \beta))^2$ . Alors

$$\theta_i^* = \theta_i - (s_i - s_{obs})^T \hat{\beta}, \quad i = 1, \dots, N. \quad (2.1.11)$$

La consistance asymptotique des estimations de la loi *a posteriori* partielle avec cette correction a été obtenue par [Blu10].

Pour relacher les hypothèses d'homoscédasticité et de linéarité sous-jacentes aux régressions linéaires locales de Beaumont et al., Blum et François [BF10] ont proposé d'approcher  $G$  au voisinage de  $s_{obs}$  par  $G(s, \varepsilon) = m(s) + \sigma(s) \times \varepsilon$  où  $m(s)$  et  $\sigma^2(s)$  sont l'espérance et la variance de  $p(\theta|s)d\theta$  dont on peut obtenir des estimateurs en ajustant deux réseaux de neurones.

$$\theta_i^* = \hat{m}(s_{obs}) + (\theta_i - \hat{m}(s_i)) \times \frac{\hat{\sigma}(s_{obs})}{\hat{\sigma}(s_i)}, \quad i = 1, \dots, N. \quad (2.1.12)$$

En résumé :

1. En utilisant les  $(\theta_i, s_i)$ 's avec les pondérations  $W_i$ , on peut approcher, au moins au voisinage de  $s_{obs}$ , la fonction  $G$ .
2. On remplace alors les  $\theta_i$  par les valeurs ajustées  $\theta_i^*$ . L'échantillon pondéré  $(\theta_i^*, W_i)$ ,  $i = 1, \dots, N$ , est alors un échantillon que l'on peut considérer tiré dans la loi  $p(\theta|s_{obs})d\theta$ .

Notons que des raffinements des algorithmes exposés existent. En particulier, des algorithmes adaptatifs pour l'ABC ont été proposés, et se révèlent très efficaces pour des modèles avec beaucoup de paramètres [BCMR09, SFT07, BMCR09, TWS<sup>+</sup>09].

### 2.1.3.2 Comparaisons entre l'ABC et les méthodes MCMC pour les modèles SIR

Dans [BT10], nous comparons les estimations de  $(\lambda, \rho)$  par MCMC et par ABC. Les données sont les temps de rémission et nous supposons qu'aucun temps d'infection n'est observé. Pour l'ABC, nous choisissons comme statistique descriptive la courbe du nombre cumulé de rémissions en fonction du temps, qui est une statistique exhaustive. Ainsi, les lois *a posteriori* fournies par les deux méthodes doivent coïncider. 100 000 simulations sont réalisées. Les lois *a priori* choisies sont des lois peu informatives (uniformes sur une échelle log), afin de refléter l'incertitude sur l'étendue de la distribution des paramètres. Pour le MCMC, nous avons implémenté la méthode d'O'Neill et Roberts [OR99]. 10 000 pas de temps sont considérés après une période transitoire (*burn-in*) de 5 000 pas.

Deux exemples sont considérés. Le premier exemple a été étudié par [OR99]. Pour le second exemple, nous avons simulé une trajectoire du modèle SIR avec des paramètres choisis, ce qui nous a donné un jeu de 29 temps de rémission. Les résultats montrent que dans les deux cas, les distributions *a posteriori* estimées par ABC sont très proches de celles obtenues par MCMC dès lors que le seuil  $\delta$  est choisi suffisamment petit.

Nous avons ensuite considéré le cas où les dates de rémissions ne sont pas observées individuellement. On ne dispose que du nombre de rémissions par année, sur la période d'observation. On remplace alors la statistique descriptive infini-dimensionnelle précédente par un vecteur de statistiques descriptives, qui ne sont alors plus exhaustives. La loi *a posteriori* partielle diffère alors de la loi *a posteriori*.

Pour chacun des  $M = 200$  jeux de données réalisés, la distribution *a posteriori* partielle est estimée grâce à  $N = 5000$  simulations. Pour comparer les différentes méthodes, nous avons calculé l'erreur quadratique moyenne renormalisée (RMSE) ou les intervalles de confiance obtenus (RMCI), en moyennant sur les  $M$  estimations réalisées. Les valeurs les plus faibles du RMSE sont toujours obtenues pour les choix des seuils de tolérance  $\delta$  les plus faibles et lorsqu'on applique la correction (2.1.12). L'étude des RMCI indique que la largeur de l'intervalle de confiance augmente de façon importante avec le seuil  $\delta$  sauf si une correction ((2.1.11) ou (2.1.12)) est appliquée.



### 2.1.3.3 Application à l'épidémie cubaine avec contact-tracing

Nous avons ensuite appliqué les méthodes ABC testées à la section précédente à un modèle SIR avec *contact-tracing* inspiré de [CTA08] (voir par exemple Müller, Kretzschmar et Dietz [MKD00] pour d'autres modèles de *contact-tracing*). Dans notre modèle, la classe des rémis est séparée en deux sous-groupes, suivant que les individus ont été dépistés par *contact-tracing* ou non. Nous avons  $T = 21.5$  années d'observation, avec beaucoup d'information inobservée. Les statistiques descriptives sont les nombres de rémis détectés par méthodes aléatoires d'une part, ou par *contact-tracing* d'autre part, respectivement  $(R_t^1)_{t \in [0, T]}$  et  $(R_t^2)_{t \in [0, T]}$ , que nous considérons comme des fonctions de  $L^1$ . Nous avons estimé les paramètres du modèle en faisant tourner un algorithme ABC avec 100 000 simulations et des lois *a priori* non informatives, uniformes sur une échelle log. Pour fixer les seuils de tolérance  $\delta$ , nous considérons les 15 premières années de l'épidémie comme échantillon d'apprentissage. Nous calibrons le modèle avec les données de cette période et prédisons l'état de la population au temps  $T$ . Nous choisissons alors les seuils de tolérance de façon à minimiser l'erreur de prédiction sur les nombres de rémis au temps  $T$  :

$$\text{Pred Error} = \mathbb{E} \left[ \frac{|R_{pred, \delta, 21.5}^1 - R_{obs, 21.5}^1|}{R_{obs, 21.5}^1} + \frac{|R_{pred, \delta, 21.5}^2 - R_{obs, 21.5}^2|}{R_{obs, 21.5}^2} \right]. \quad (2.1.13)$$

Pour les seuils choisis, nous refaisons alors une estimation ABC en utilisant l'ensemble des données.

Pour valider l'estimation, nous avons simulé des trajectoires pour le modèle SIR avec des paramètres tirés dans la loi *a posteriori* partielle obtenue et avons comparé les distributions *a posteriori* des descripteurs avec leurs valeurs observées. Les résultats sont représentés à la Figure 2.1.2.

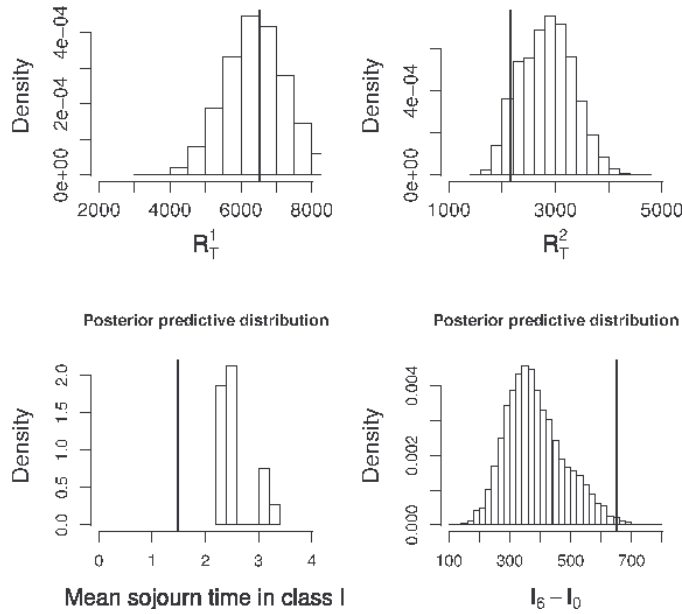


FIGURE 2.1.2 – Distributions *a posteriori* prédites de  $R_{21.5}^1$ ,  $R_{21.5}^2$ ,  $I_6$ , et du temps de séjour moyen dans l'état  $I$ . Les distributions *a posteriori* sont obtenues par un algorithme ABC avec des seuils de tolérance correspondant au rejet de 99% des simulations pour le calcul de chaque poids  $W_i$ .

Les nombres cumulés d'individus dépistés par *contact-tracing* ou méthode aléatoire sont bien dans l'intervalle de confiance construit à partir de la distribution *a posteriori*. On trouve que le temps moyen pour détecter un individu séropositif est environ 3,8 ans. Par contre, le temps moyen

durant lequel un individu reste infectieux est décentré par rapport à sa loi *a posteriori* estimée. De même, le nombre observé d'individus infectés sur les 6 premières années se situe dans la queue de distribution à droite de sa loi *a posteriori*. Les paramètres nous disent qu'un individu infectieux non dépisté met environ 3,1 ans pour infecter une nouvelle personne (95%CI [0, 72; 4, 34]) ; d'après la comparaison avec les données, ce temps pourrait être plus court. Ces écarts pourraient être liés au fait qu'il est important de prendre en compte une structure d'âge dans le taux d'infection afin de capturer des effets non-markoviens.

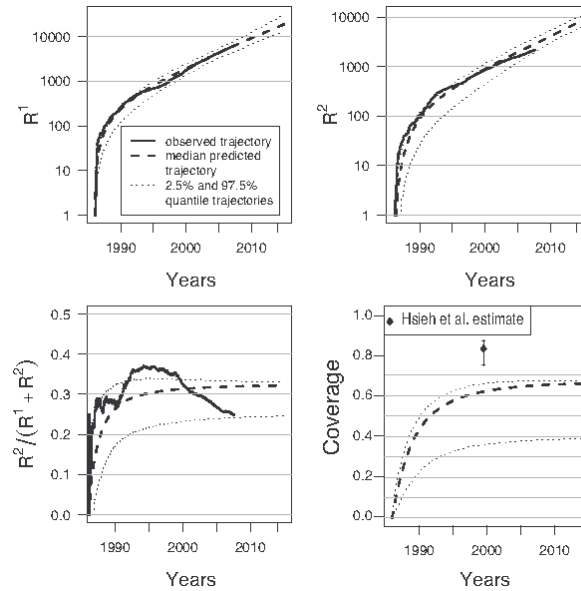


FIGURE 2.1.3 – Médiane et intervalle de confiance à 95% pour la distribution *a posteriori* de  $R_t^1$ ,  $R_t^2$ ,  $R_t^1/(R_t^1 + R_t^2)$  et de la couverture  $R_t/(I_t + R_t)$  entre 1986 et 2015.

La Figure 2.1.3 montre les prédictions, par le modèle SIR, des nombres d'individus détectés par dépistage aléatoire, par *contact-tracing*, ainsi que la courbe reconstruite du nombre de personnes infectieuses. Il y a une très bonne adéquation entre les nombres de dépistages aléatoires observés et prédits sur la période sauf entre 1992 et 1995. Cette période correspond à la période de crise qui a suivi l'effondrement de l'Union Soviétique où l'effort sur la détection du HIV a reçu moins de soutien. On observe aussi, depuis 2000, une diminution des détections par *contact-tracing*, par rapport à ce qui est prédit. Ceci est l'indicateur d'un effort moins soutenu pour le *contact-tracing* sur les années récentes. Une explication possible est qu'un nouveau mode de détection, la "captation" a été développé (H. De Arazoza, communication personnelle).

**Performance du contact-tracing** Pour tester la performance du *contact-tracing*, [HdALC02] ont calculé la couverture de l'épidémie, définie comme étant le pourcentage d'individus ayant été détectés  $(R^1 + R^2)/(I + R^1 + R^2)$ . Nous prédisons une couverture de 62% (IC à 95% [36%; 66%]) en 2000 (Figure 2.1.3) ce qui est bien plus bas que le taux de couverture 83% (IC95% [75%; 87%]) estimé par [HdALC02]. A cela, on ajoute que la Figure 2.1.2 montre que le modèle SIR prédit moins d'infections que dans nos données, ce qui laisse penser que le taux de couverture pourrait encore être surestimé.

En utilisant l'estimation des couvertures, on peut comparer les taux de détection par *contact-tracing* et par dépistage aléatoire. On retrouve la constatation de [CART10] qu'en fin de période

d'observation, 1 détection sur 2 environ est due au *contact-tracing*. L'examen plus en détail des données nous indique que le *contact-tracing* est plus efficace chez les femmes, tandis que les hommes sont plutôt dépistés par tests aléatoires.

**Prédictions** Les simulations nous fournissent également des prédictions, valables sous les hypothèses du modèle bien sûr. Dans [BT10], nous avons prédit qu'en 2015, 42 000 individus (IC à 95% [29 000; 107 000]) auront été infectés depuis le début de l'épidémie. Parmi ces individus, 45% (95%IC [29%; 46%]) auront été détectés par dépistage aléatoire et 21% (95%CI [10%; 22%]) par *contact-tracing*. La proportion d'individus détectés par *contact-tracing* tendrait vers 32% (95%CI [25%; 33%]). Le nombre d'infectés en 2015 correspond à 27 000 (95%CI [19 000; 80 000]) nouveaux cas entre juillet 2007 et janvier 2015.

## 2.2 Modèles SIR évoluant sur un graphe

Pour les maladies sexuellement transmissibles, supposer qu'on a un modèle mélangeant avec une force d'infection de la forme  $\lambda^n S_t^n I_t^n$  proportionnelle au nombre de couples susceptible-infectieux que l'on peut créer n'est pas très réaliste. L'introduction d'hétérogénéités dans le nombre de contacts complexifie grandement les modèles et cela a motivé une littérature abondante (cf. Durrett [Dur07], Newman [New03], House [Hou12] pour des revues sur le sujet).

Dans la littérature, concernant la théorie, des approximations en champ moyen (e.g. [PSV02, BBPSV05, Dur07]) ou en grande population (e.g. [And99, BN08] ou [And98, Eq. (3)]) ont été étudiées et fournissent une description fondée sur un système infini dénombrable d'équations. D'autres travaux supposent des relations approximantes entre les moments pour fermer les équations (e.g. Eames et Keeling [EK02]).

Tous ces articles supposent en général des structures simples pour les réseaux : graphes réguliers où chacun des sommets a le même degré [DJM98, KLLM08], petits mondes [KG99, MN00], modèles de configuration [ML01, Vol08, VM07, K GK13], graphes d'intersection [BDLL08, BST14], etc. Nous considérons, à la suite de Volz, un modèle SIR sur les modèles de configuration. Enfin, nous conclurons cette section par la considération des données Cubaines.

### 2.2.1 Epidémies se propageant sur un modèle de configuration et équations de Volz

En 2008, Volz [Vol08] a proposé un système de seulement cinq équations pour décrire la propagation d'une épidémie de type SIR sur un graphe de configuration, en partant d'un modèle individu-centré et en considérant de grandes tailles de population (voir également Miller [Mil11] pour une autre dérivation heuristique). Nous commençons par présenter une preuve du résultat de Volz que nous avons élaborée avec Laurent Decreusefond, Jean-Stéphane Dhersin et Pascal Moyal. Les équations de Volz décrivent s'appliquent lorsqu'initialement la population infectée est déjà de grande taille. Pour montrer la convergence du modèle individu-centré vers les EDO proposées par Volz, des moments d'ordre 5 pour les conditions initiales sont requis. Ceci peut sembler contraignant, mais en considérant le début de l'épidémie, on peut montrer que dans le cas particulier d'épidémies se propageant sur des modèles de configuration, des moments exponentiels apparaissent dès que la fraction des individus infectés devient strictement positive. C'est ce que nous regarderons dans un deuxième temps.

Les modèles de configuration [Bol01, MR95] (*configuration model* en anglais, abrégé par CM) sont des graphes caractérisés par la distribution de degrés  $(p_k)_{k \in \mathbb{N}}$  de leurs sommets et une connectivité aléatoire. On peut construire un CM de taille  $n$  en se donnant une suite indépendante et identiquement distribuée de variables  $d_1, \dots, d_n$  aléatoires de distribution  $(p_k)_{k \in \mathbb{N}}$ . La variable  $d_i$

peut être comprise comme étant le nombre de demi-arêtes attachées à l'individu  $i$ . On forme alors les arêtes en choisissant uniformément deux demi-arêtes parmi les demi-arêtes non encore appariées (cf. [dH14] pour une revue).

Dans [DDMT12], nous montrons qu'au coeur du problème, derrière les cinq équations de Volz, se cachent en fait trois équations à valeurs mesures. En effet, comme ce sont les distributions de degrés qui caractérisent les graphes de configuration, la dynamique se déduit naturellement de l'évolution de trois distributions de degrés particulières. Soient  $\mathcal{S}_t^n$ ,  $\mathcal{I}_t^n$  et  $\mathcal{R}_t^n$  les ensembles de sommets susceptibles, infectés et rémis au temps  $t$ . On notera les tailles de ces classes  $S_t^n$ ,  $I_t^n$  et  $R_t^n$  respectivement. Pour un individu  $i$ , son degré est noté  $d_i$  et le nombre d'arêtes qu'il a vers  $\mathcal{S}_t^n$  est  $d_i(\mathcal{S}_t^n)$ . On introduit :

- la distribution de degrés des susceptibles,  $\mu_t^{n,S}(dk) = \sum_{i \in \mathcal{S}_t^n} \delta_{d_i}(dk) \in \mathcal{M}_F(\mathbb{N})$ ,
- la distribution du nombre d'arêtes que les individus infectés ont vers les susceptibles,  $\mu_t^{n,IS}(dk) = \sum_{i \in \mathcal{I}_t^n} \delta_{d_i(\mathcal{S}_t^n)}(dk)$ ,
- la distribution du nombre d'arêtes que les individus rémis ont vers les susceptibles,  $\mu_t^{n,RS}(dk) = \sum_{i \in \mathcal{R}_t^n} \delta_{d_i(\mathcal{S}_t^n)}(dk)$ .

A partir de ces mesures, on peut décrire les tailles  $S_t^n = \langle \mu_t^{n,S}, 1 \rangle$ ,  $I_t^n$  et  $R_t^n$  ou encore le nombre d'arêtes reliant un infecté à un susceptible  $N_t^{n,IS} = \langle \mu_t^{n,IS}, k \rangle = \sum_{i \in \mathcal{I}_t^n} d_i(\mathcal{S}_t^n)$ . Un exemple est donné à la figure 2.2.1. Mais bien sûr, l'information contenue par ces trois mesures ne permet pas de reconstruire le graphe sur lequel se propage l'épidémie, ni de savoir quels sont les individus qui sont connectés ensemble. Elle est cependant suffisante pour décrire l'évolution de l'épidémie en grande population, quand le graphe est un CM.

Nous avons proposé un système dynamique qui permette de retrouver les équations de Volz en construisant simultanément l'épidémie et le graphe sur lequel elle se propage (voir Fig. 2.2.1).

- Seules les arêtes entre  $\mathcal{I}^n$  et  $\mathcal{R}^n$  sont observées. Les degrés de tous les individus sont connus.
- A chaque individu infectieux, on associe une horloge exponentielle de taux  $\rho$ , qui détermine l'instant de sa rémission.
- A chaque demi-arête libre (arête dirigée vers  $\mathcal{S}^n$ ), on associe une horloge exponentielle de taux  $\lambda$ . Le taux d'infection global est  $\lambda N_{t-}^{n,IS}$ .
- Lorsqu'une horloge liée à une arête sonne, une demi-arête d'un susceptible est choisie uniformément parmi celles disponibles et détermine quel nouvel individu susceptible est infecté. Chaque demi-arête ayant la probabilité  $1/N_{t-}^{n,S}$  d'être choisie, la probabilité de choisir un susceptible de degré  $k$  est  $k/N_{t-}^{n,S}$ .

Le taux d'infection d'un susceptible de degré  $k$  est alors :  $\lambda k p_{t-}^{n,I}$ , où  $p_{t-}^{n,I} = N_{t-}^{n,IS} / N_{t-}^{n,S}$  est la proportion des demi-arêtes liées aux susceptibles qui sont susceptibles de les infecter.

- Il reste ensuite à déterminer parmi les demi-arêtes restantes du nouvel infecté lesquelles sont liées à  $\mathcal{I}^n$  ou  $\mathcal{R}^n$ . Si le nouvel infecté est de degré  $k$ , la probabilité que les  $k - 1$  arêtes restantes soient liées à  $\mathcal{I}^n$  ou  $\mathcal{R}^n$  est :

$$p(j, \ell, m | k - 1, t) = \frac{\binom{j}{N_{t-}^{n,IS} - 1} \binom{\ell}{N_{t-}^{n,RS}} \binom{m}{N_{t-}^{n,SS}}}{\binom{k-1}{N_{t-}^{n,S} - 1}} \mathbb{1}_{j+\ell+m=k-1} \mathbb{1}_{j < N_{t-}^{n,SI}} \mathbb{1}_{\ell \leq N_{t-}^{n,SR}}.$$

Nous avons proposé 3 EDS à valeurs mesures pour décrire l'évolution de  $(\mu^{n,S}, \mu^{n,IS}, \mu^{n,RS})_{n \in \mathbb{N}^*}$ , dont on peut montrer l'existence et l'unicité forte des solutions. On considère  $\mu_t^{(n),S} = \frac{1}{n} \mu_t^{n,S}$ ,  $\mu_t^{(n),IS} = \frac{1}{n} \mu_t^{n,IS}$  et  $\mu_t^{(n),RS} = \frac{1}{n} \mu_t^{n,RS}$ . Notre but est d'étudier la convergence de la suite  $\mu^{(n)} := (\mu_t^{(n),S}, \mu_t^{(n),IS}, \mu_t^{(n),RS})_{n \in \mathbb{N}^*}$  lorsque le nombre de sommets  $n$  tend vers l'infini sans changer les distributions de degré.

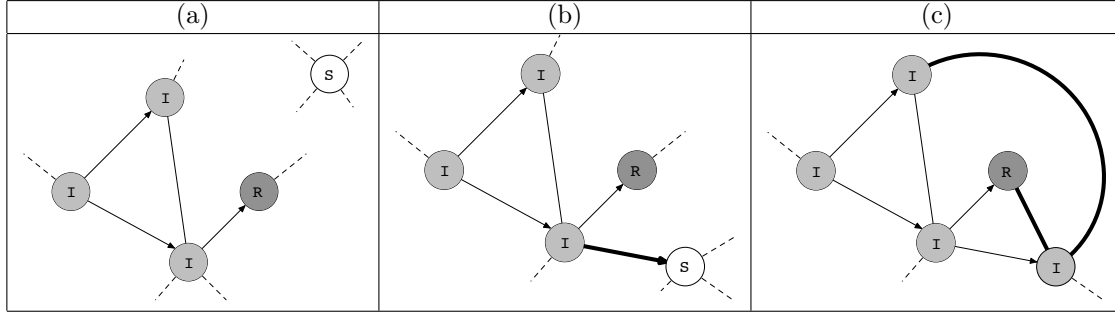


FIGURE 2.2.1 – *Processus d'infection.* Les individus susceptibles, infectieux et rémis sont respectivement en blanc, gris et noir. (a) Le graphe est découvert au fur et à mesure de l'infection : les individus susceptibles ne sont pas rattachés au cluster. Dans cette vignette :  $\mu_t^{n,IS} = 2\delta_1 + \delta_2$  et  $\mu_t^{n,RS} = \delta_1$ . On voit que  $N_t^{n,IS} = \langle \mu_t^{n,IS}, k \rangle = 2 \times 1 + 1 \times 2 = 4$ . (b) Une infection se produit lorsque l'une des demi-arêtes issues d'un infectieux propage l'épidémie. On choisit alors une demi-arête appartenant à un susceptible uniformément parmi celles disponibles. Ce susceptible devient alors infectieux. (c) Les demi-arêtes du susceptible nouvellement contaminé sont appariées à celles des rémis ou infectieux encore libres, ou restent libres, pouvant alors continuer de propager l'épidémie.

Nous introduisons, pour  $A > 0$  et  $\varepsilon > 0$  l'espace

$$\mathcal{M}_{\varepsilon, A} = \left\{ \nu \in \mathcal{M}_F(\mathbb{N}) ; \sum_{k \in \mathbb{N}} k^5 \nu(k) \leq A \text{ et } \sum_{k \in \mathbb{N}} k \nu(k) \geq \varepsilon \right\} \quad (2.2.1)$$

et posons  $\mathcal{M}_{0+, A} := \cup_{\varepsilon > 0} \mathcal{M}_{\varepsilon, A}$ .

**Théorème 2.2.1.** *Si les conditions initiales  $(\mu_0^{(n),S}, \mu_0^{(n),IS}, \mu_0^{(n),RS})$  sont dans  $(\mathcal{M}_{0, A})^3$  pour tout  $n$ , et si elles convergent vers  $(\bar{\mu}_0^S, \bar{\mu}_0^{IS}, \bar{\mu}_0^{RS}) \in (\mathcal{M}_F(\mathbb{N}))^3$ , avec  $\sum_{k \in \mathbb{N}} k \bar{\mu}_0^{IS}(k) > 0$ , alors la suite  $(\mu^{(n)})_{n \in \mathbb{N}^*}$  converge en probabilité dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_{0+, A}^3)$ , lorsque  $n \rightarrow +\infty$ , vers l'unique solution déterministe  $(\bar{\mu}_t^S, \bar{\mu}_t^{IS}, \bar{\mu}_t^{RS})_{t \in \mathbb{R}_+}$  des équations suivantes dans  $\mathcal{C}(\mathbb{R}_+, \mathcal{M}_{0, A} \times \mathcal{M}_{0+, A} \times \mathcal{M}_{0, A})$ , pour toute fonction  $f \in \mathcal{B}_b(\mathbb{N}, \mathbb{R})$  :*

$$\langle \bar{\mu}_t^S, f \rangle = \sum_{k \in \mathbb{N}} \bar{\mu}_0^S(k) \theta_t^k f(k), \quad (2.2.2)$$

$$\langle \bar{\mu}_t^{IS}, f \rangle = \langle \bar{\mu}_0^{IS}, f \rangle - \int_0^t \rho \langle \bar{\mu}_s^{IS}, f \rangle ds \quad (2.2.3)$$

$$\begin{aligned} & + \int_0^t \sum_{k \in \mathbb{N}} \lambda k \bar{p}_s^I \sum_{\substack{j, \ell, m \in \mathbb{N} \\ j + \ell + m = k - 1}} \binom{k-1}{j, \ell, m} (\bar{p}_s^I)^j (\bar{p}_s^R)^\ell (\bar{p}_s^S)^m f(m) \bar{\mu}_s^S(k) ds \\ & + \int_0^t \sum_{k \in \mathbb{N}} \lambda k \bar{p}_s^I (1 + (k-1) \bar{p}_s^I) \sum_{k' \in \mathbb{N}^*} (f(k') - 1) \frac{k' \bar{\mu}_s^{IS}(k')}{\bar{N}_s^{IS}} \bar{\mu}_s^S(k) ds, \end{aligned}$$

$$\langle \bar{\mu}_t^{RS}, f \rangle = \langle \bar{\mu}_0^{RS}, f \rangle + \int_0^t \rho \langle \bar{\mu}_s^{IS}, f \rangle ds \quad (2.2.4)$$

$$+ \int_0^t \sum_{k \in \mathbb{N}} \lambda k \bar{p}_s^I (k-1) \bar{p}_s^R \sum_{k' \in \mathbb{N}^*} (f(k') - 1) \frac{k' \bar{\mu}_s^{RS}(k')}{\bar{N}_s^{RS}} \bar{\mu}_s^S(k) ds,$$

où  $\theta_t = \exp(-\lambda \int_0^t \bar{p}_s^I ds)$  peut se comprendre comme étant la probabilité qu'un sommet de degré 1 soit encore susceptible au temps  $t$ , où  $\bar{N}_t^{IS} = \langle \bar{\mu}_t^{IS}, k \rangle$ , et où  $\bar{p}_t^I = \bar{N}_t^{IS} / \bar{N}_t^S$ .

Nous déduisons des équations (2.2.2)-(2.2.4) les équations de Volz.

**Proposition 2.2.2.** *Soit  $g(z) = \sum_{k \in \mathbb{N}} \bar{\mu}_0^S(k) z^k$  la fonction génératrice de  $\bar{\mu}_0^S$ , alors*

$$\begin{aligned} \theta_t &= \exp\left(-\lambda \int_0^t \bar{p}_s^I ds\right) \\ \bar{S}_t &= g(\theta_t), \quad \bar{I}_t = \bar{I}_0 + \int_0^t \left(\lambda \bar{p}_s^I \theta_s g'(\theta_s) - \lambda \times \bar{N}_s^{IS} - \rho \bar{I}_s\right) ds \\ \bar{p}_t^I &= \frac{\bar{N}_t^{IS}}{\bar{N}_t^S} = \bar{p}_0^I + \int_0^t \left(\lambda \bar{p}_s^I \bar{p}_s^S \theta_s \frac{g''(\theta_s)}{g'(\theta_s)} - \lambda \bar{p}_s^I (1 - \bar{p}_s^I) - \rho \bar{p}_s^I\right) ds. \\ \bar{p}_t^S &= \frac{\bar{N}_t^{SS}}{\bar{N}_t^S} = \bar{p}_0^S + \int_0^t \lambda \bar{p}_s^I \bar{p}_s^S \left(1 - \theta_s \frac{g''(\theta_s)}{g'(\theta_s)}\right) ds \end{aligned}$$

constituent les équations de Volz. □

On peut remarquer que le terme  $\lambda \bar{p}_s^I \theta_s g'(\theta_s)$  qui apparaît dans l'expression de  $\bar{I}_t$  est égal à  $\lambda \bar{N}_t^{IS}$ . Pour les modèles compartimentaux, on peut rappeler que l'on avait  $\lambda \bar{s}_t \bar{i}_t$  à la place de ce terme.

*Idée de la preuve du théorème 2.2.1.* Il s'agit encore d'un résultat de tension-unicité. Dans la preuve, on voit que l'épidémie reste large pourvu que le nombre d'arêtes de  $\mathcal{I}_t^n$  vers  $\mathcal{S}_t^n$  reste de l'ordre de  $n$ . Pour  $\varepsilon > 0$  et  $n \in \mathbb{N}^*$ , on est alors amené à définir  $\tau_\varepsilon^n = \inf\{t \geq 0, \sum_{k \in \mathbb{N}} k \mu_t^{(n), IS}(k) < \varepsilon\}$ . Nous montrons d'abord la tension de  $(\mu^{(n)})_{n \in \mathbb{N}^*}$  dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_{0,A}^3)$  de façon classique, et en déduisons la tension de  $(\mu_{\wedge \tau_\varepsilon^n}^{(n)})_{n \in \mathbb{N}^*}$ .

Notons  $\Psi_t^{S,f}$ ,  $\Psi_t^{IS,f}$  et  $\Psi_t^{RS,f}$  les applications de  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_{0,A}^3)$  dans  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$  telles que (2.2.2)-(2.2.4) s'écrivent :

$$\begin{aligned} &(\langle \bar{\mu}_t^S, f \rangle, \langle \bar{\mu}_t^{IS}, f \rangle, \langle \bar{\mu}_t^{RS}, f \rangle) \\ &= \left( \Psi_t^{S,f}(\bar{\mu}^S, \bar{\mu}^{IS}, \bar{\mu}^{RS}), \Psi_t^{IS,f}(\bar{\mu}^S, \bar{\mu}^{IS}, \bar{\mu}^{RS}), \Psi_t^{RS,f}(\bar{\mu}^S, \bar{\mu}^{IS}, \bar{\mu}^{RS}) \right). \end{aligned} \quad (2.2.5)$$

On montre l'unicité de la solution de ce système dans  $\mathcal{C}([0, T], \mathcal{M}_{0,A} \times \mathcal{M}_{\varepsilon,A} \times \mathcal{M}_{0,A})$ . Il reste donc à montrer que les valeurs d'adhérences de  $(\mu^{(n)})_{n \in \mathbb{N}^*}$  satisfont (2.2.5).

C'est dans l'identification de l'équation satisfaite par ces valeurs d'adhérence qu'on a besoin de minorer  $\sum_{k \in \mathbb{N}} k \mu_t^{(n), IS}(k)$  par  $\varepsilon$  (et donc de considérer le processus arrêté) et qu'on a besoin de montrer que les moments  $\sum_{k \in \mathbb{N}} k^4 \mu_0^{(n), S}(k)$  convergent vers 0 (et donc de contrôler les moments d'ordre 5 avec (2.2.1)). On montre ainsi, pour  $\mu^{(n), IS}$ , que

$$\langle \mu_{t \wedge \tau_\varepsilon^n}^{(n), IS}, f \rangle = \Psi_{t \wedge \tau_\varepsilon^n}^{IS,f}(\mu^{(n), S}, \mu^{(n), IS}, \mu^{(n), RS}) + \Delta_{t \wedge \tau_\varepsilon^n}^{n,f} + M_{t \wedge \tau_\varepsilon^n}^{(n), IS,f}, \quad (2.2.6)$$

où  $\Delta_{\wedge \tau_\varepsilon^n}^{n,f}$  converge vers 0 en probabilité et uniformément sur tout intervalle de temps compact, et où  $M^{(n), IS,f}$  est une martingale de carré intégrable qui converge dans  $L^2$  et en probabilité vers 0. Nous avons des décompositions similaires pour  $\mu^{(n), S}$  ou  $\mu^{(n), RS}$ .

Pour conclure, nous devons nous débarrasser des temps d'arrêt  $\tau_\varepsilon^n$  dans (2.2.6). Soit  $\tilde{\mu} := (\tilde{\mu}^S, \tilde{\mu}^{IS}, \tilde{\mu}^{RS})$  une valeur d'adhérence et soit une sous-suite, de nouveau notée  $\mu^{(n)}$  par abus de notation, telle que  $\mu_{\wedge \tau_\varepsilon^n}^{(n)}$  converge en loi vers cette valeur d'adhérence. Par le théorème de Skorokhod, on peut, quitte à changer d'espace de probabilité, supposer que cette convergence est presque sûre. Comme le processus  $(N_{\wedge \tau_\varepsilon^n}^{(n), IS})$  converge alors presque sûrement dans  $\mathbb{D}([0, T], \mathbb{R}_+)$

vers  $\tilde{N}^{IS} := \sum_{k \in \mathbb{N}} k \tilde{\mu}^{IS}(k)$ , qui est continu, la convergence est uniforme sur tout intervalle de temps compact et on en déduit que presque sûrement,

$$\inf_{t \in [0, T]} \tilde{N}_t^{IS} = \lim_{n \rightarrow +\infty} \inf_{t \in [0, T]} N_{t \wedge \tau_\varepsilon^n}^{(n), IS}. \quad (2.2.7)$$

Pour  $\varepsilon' > \varepsilon$ , soit  $t_{\varepsilon'} := \inf\{t \geq 0, \tilde{N}_t^{IS} < \varepsilon'\}$ , dont on ne sait pas encore s'il s'agit d'un temps déterministe. On a  $\varepsilon' \leq \inf_{t \in [0, T]} \tilde{N}_{t \wedge t_{\varepsilon'}}^{IS} = \lim_{n \rightarrow +\infty} \inf_{t \in [0, T]} N_{t \wedge \tau_\varepsilon^n \wedge t_{\varepsilon'}}^{(n), IS}$ , et par le lemme de Fatou,

$$\begin{aligned} 1 &= \mathbb{P}\left(\inf_{t \in [0, t_{\varepsilon'}]} \tilde{N}_t^{IS} \geq \varepsilon'\right) = \mathbb{E}\left(\mathbb{1}_{\lim_{n \rightarrow +\infty} \inf_{t \in [0, T \wedge t_{\varepsilon'}]} N_{t \wedge \tau_\varepsilon^n}^{(n), IS} \geq \varepsilon'}\right) \\ &\leq \mathbb{E}\left(\lim_{n \rightarrow +\infty} \mathbb{1}_{\inf_{t \in [0, T \wedge t_{\varepsilon'}]} N_{t \wedge \tau_\varepsilon^n}^{(n), IS} > \varepsilon}\right) \leq \lim_{n \rightarrow +\infty} \mathbb{P}\left(\inf_{t \in [0, T \wedge t_{\varepsilon'}]} N_{t \wedge \tau_\varepsilon^n}^{(n), IS} > \varepsilon\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(\tau_\varepsilon^n > T \wedge t_{\varepsilon'}). \end{aligned}$$

Ceci nous permet de montrer en écrivant que

$$\Psi_{\cdot \wedge \tau_\varepsilon^n \wedge t_{\varepsilon'} \wedge T}^{IS, f}(\mu^{(n)}) = \Psi_{\cdot \wedge \tau_\varepsilon^n \wedge T}^{IS, f}(\mu^{(n)}) \mathbb{1}_{\tau_\varepsilon^n \leq t_{\varepsilon'} \wedge T} + \Psi_{\cdot \wedge t_{\varepsilon'} \wedge T}^{IS, f}(\mu^{(n)}) \mathbb{1}_{\tau_\varepsilon^n > t_{\varepsilon'} \wedge T},$$

que  $\tilde{\mu}^{IS}$  résout (2.2.3) sur  $[0, t_{\varepsilon'} \wedge T]$  et coïncide donc sur cet intervalle avec  $\bar{\mu}^{IS}$ . En conséquence,  $t_{\varepsilon'}$  est déterministe et on peut montrer la convergence uniforme de  $(\mu^{(n)})_{n \in \mathbb{N}^*}$  vers  $(\bar{\mu}^S, \bar{\mu}^{IS}, \bar{\mu}^{RS})$ .  $\square$

### 2.2.1.1 Début de l'épidémie : évolution de $\mu^{(n), S}$

L'hypothèse de moments d'ordre 5 dans l'énoncé du Théorème 2.2.1 peut sembler restrictive. Janson Luczak et Windridge [JLW14] ont travaillé pour lever cette hypothèse en considérant le processus  $(S_t^n, I_t^n, R_t^n, N_t^{n, S}, N_t^{n, I}, N_t^{n, R})_{t \in \mathbb{R}_+}$  où  $N_t^{n, S} = \langle \mu_t^{n, S}, k \rangle$ ,  $N_t^{n, I}$  et  $N_t^{n, R}$  sont respectivement les nombres de demi-arêtes des susceptibles, infectieux et remis encore libres. Ils montrent qu'avec ce processus, qui contient encore moins d'information que notre processus à valeurs mesures  $(\mu^{(n), S}, \mu^{(n), IS}, \mu^{(n), RS})$ , il suffit de demander à ce que les moments d'ordre 2 soient uniformément bornés en  $n$ . En fait, si l'on accepte que le graphe généré par le CM puisse avoir des auto-boucles et des arêtes multiples, alors il suffit de demander l'intégrabilité uniforme de la distribution des degrés d'un susceptible choisi au hasard, ce qui semble être une condition minimale...

Après discussion avec Pieter Trapman, il est apparu, en considérant le début de l'épidémie, que l'hypothèse de moments d'ordre 5 n'est pas tant restrictive. Considérons la distribution des degrés des susceptibles, par exemple. Si cette distribution n'a pas de moment jusqu'à l'ordre 5 c'est à cause d'individus ayant un nombre de voisins trop élevé. Or lorsqu'on apparie les demi-arêtes au hasard, on tire les individus proportionnellement à leurs nombres de demi-arêtes, et ces individus sont ceux qui sont infectés en premier. Après les  $[\varepsilon n]$  premières infections, avec  $\varepsilon > 0$ , nous allons montrer que les susceptibles de plus haut degré ont été infectés et que la distribution des degrés des susceptibles restants admet des moments exponentiels.

Pour une population de taille  $n$  dont les degrés des individus sont  $d_1, \dots, d_n$ , posons, pour tout  $k \in \mathbb{N}$ ,  $N_k^n = \text{Card}\{i \in \{1, \dots, n\}, d_i = k\}$ .  $N_k^n$  est le nombre de sommets de degré  $k$  parmi les  $n$  sommets du graphe. On associe à chacune des  $d_i$  demi-arêtes de l'individu  $i$  des variables aléatoires uniformes sur  $[0, 1]$ . Le sommet  $i$  est infecté avant le sommet  $j$  si la valeur minimale  $Z_i$  des demi-arêtes de  $i$  est plus petite que la valeur minimale  $Z_j$  sur les demi-arêtes de  $j$ . Cette construction a été utilisée par Riordan [Rio12] et est liée aux rangements biaisés par la taille.

**Proposition 2.2.3.** (i) La distribution de degrés  $(\hat{p}_k^{\varepsilon, n})_{k \geq 1}$  des susceptibles restants après les  $[\varepsilon n]$  premières infections est :

$$\hat{p}_k^{\varepsilon, n} = \frac{1}{n - [\varepsilon n]} \sum_{i=1}^n \mathbb{1}_{d_i = k} \mathbb{1}_{Z_i > Z_{([\varepsilon n])}} \quad (2.2.8)$$

où  $(Z_{(1)}, \dots, Z_{(n)})$  sont les statistiques d'ordre de  $(Z_1, \dots, Z_n)$ , et où  $\mathbb{P}(Z_i \leq z \mid d_i) = 1 - (1 - z)^{d_i}$ .  
(ii) Pour  $z \in (0, 1)$ , soit  $M(z)$  la fonction de survie de la loi des  $Z_i$  et soit  $M_n(z)$  la fonction de survie empirique de  $(Z_1, \dots, Z_n)$  :

$$M_n(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Z_i > z}, \quad \text{et} \quad M(z) = \sum_{k \geq 0} p_k (1 - z)^k = g(1 - z),$$

où  $g(z) = \sum_{k \geq 0} p_k z^k$  est la fonction génératrice de la distribution des degrés du graphe. Soit  $\varepsilon$  tel que  $z^\varepsilon = \inf\{z \in (0, 1), M(z) \geq \varepsilon\}$  soit le quantile d'ordre  $\varepsilon$ . Alors dès que  $\varepsilon$  est tel que  $z^\varepsilon$  est un point où  $M$  est continue et strictement croissante,

$$\lim_{n \rightarrow +\infty} Z_{[\varepsilon n]} = z^\varepsilon \quad \text{presque sûrement.}$$

(iii) Pour un tel  $\varepsilon$ , la distribution des degrés des susceptibles restant après les  $[\varepsilon n]$  premières infections converge étroitement vers :

$$\lim_{n \rightarrow +\infty} \sum_{k \geq 0} \widehat{p}_k^{\varepsilon, n} \delta_k = \frac{1}{1 - \varepsilon} \sum_{k \geq 0} p_k (1 - z^\varepsilon)^k \delta_k, \quad (2.2.9)$$

où  $z^\varepsilon$  est solution de  $1 - \varepsilon = g(1 - z^\varepsilon)$ . On a de plus la convergence des moments d'ordre 5 :

$$\lim_{n \rightarrow +\infty} \sum_{k \geq 0} k^5 \widehat{p}_k^{\varepsilon, n} = \frac{1}{1 - \varepsilon} \sum_{k \geq 0} k^5 p_k (1 - z^\varepsilon)^k < +\infty. \quad (2.2.10)$$

En particulier, la distribution limite (2.2.9) admet des moments de tous ordres.

*Démonstration.* Montrons (ii). Soit  $z \in [0, 1]$ . La proportion de sommets de degrés  $k$  dont la valeur minimale des  $Z_i$  est inférieure à  $z$  est  $M_k^n(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{d_i=k} \mathbb{1}_{Z_i > z}$ . Par la loi des grands nombres,  $\lim_{n \rightarrow +\infty} M_k^n(z) = p_k (1 - z)^k$  p.s., ce qui implique que

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Z_i > z} \delta_{d_i} = \sum_{k \geq 0} p_k (1 - z)^k \delta_k$$

pour la convergence de la topologie vague et pour la topologie de la convergence étroite, en vérifiant la convergence des masses par convergence dominée.

Supposons que  $\varepsilon > 0$  soit tel que  $z^\varepsilon$  soit un point où  $M$  est continu et strictement croissante. Alors,  $M(z^\varepsilon) = \varepsilon$ . Soit  $\delta > 0$  et  $\eta = \min(|M(z^\varepsilon - \delta) - M(z^\varepsilon)|, |M(z^\varepsilon + \delta) - M(z^\varepsilon)|)$ . Par le théorème de Kolmogorov-Smirnov :  $\lim_{n \rightarrow +\infty} \|M_n - M\|_\infty = 0$ . Alors, il existe  $\mathbb{P}(d\omega)$ -p.s. un entier  $n_0(\omega)$  suffisamment grand tel que pour tout  $n \geq n_0$ ,  $\|M_n - M\|_\infty < \eta/2$ . Comme  $M$  est croissante et comme  $Z_{[\varepsilon n]}$  est tel que  $M_n(Z_{[\varepsilon n]}) = \frac{[\varepsilon n]}{n}$ , alors, pour  $n \geq \max(n_0, 2/\eta)$ ,  $|M(Z_{[\varepsilon n]}) - \varepsilon| < \eta$  et donc  $Z_{[\varepsilon n]} \in (z^\varepsilon - \delta, z^\varepsilon + \delta)$  p.s. Donc,  $(Z_{[\varepsilon n]})_{n \geq 1}$  converge p.s. vers  $z^\varepsilon$ .  
Comme

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n |\mathbb{1}_{Z_i > Z_{[\varepsilon n]}} - \mathbb{1}_{Z_i > z^\varepsilon}| \mathbb{1}_{d_i=k} \right) \\ = \lim_{n \rightarrow +\infty} \mathbb{P} \left( Z_1 \in [Z_{[\varepsilon n]} \wedge z^\varepsilon, Z_{[\varepsilon n]} \vee z^\varepsilon], d_1 = k \right) = 0, \end{aligned} \quad (2.2.11)$$

et comme  $n/(n - [\varepsilon n])$  converge vers  $1/(1 - \varepsilon)$ , on obtient (2.2.9).



Pour la convergence des moments d'ordre 5, on remarque que pour  $K \in \mathbb{N}^*$  grand,

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{k \geq 0} k^5 M_k^n(Z_{[\varepsilon n]}) - \sum_{k \geq 0} k^5 p_k (1 - z^\varepsilon)^k \right| \right) \\ & \leq \mathbb{E} \left( \left| \sum_{k \leq K} k^5 (M_k^n(Z_{[\varepsilon n]}) - p_k (1 - z^\varepsilon)^k) \right| \right) + \mathbb{E} \left( \sum_{k > K} k^5 M_k^n(Z_{[\varepsilon n]}) \right) + \sum_{k > K} k^5 p_k (1 - z^\varepsilon)^k. \end{aligned}$$

Le premier terme du membre de droite converge vers 0 par les arguments précédents. Le troisième terme est contrôlé pour  $K$  suffisamment grand. Pour le second terme, on utilise que pour tout  $z \in (0, 1)$ ,  $\mathbb{E}(\sum_{k > K} k^5 M_k^n(z)) = \sum_{k > K} k^5 p_k (1 - z)^k$  et que  $Z_{[\varepsilon n]}$  converge p.s. vers  $z^\varepsilon$ .  $\square$

### 2.2.1.2 Début de l'épidémie : nombre de reproduction

Les EDO proposées par Volz et le système (2.2.2)-(2.2.4) sont valables lorsque la proportion d'infectieux dans la population est strictement positive. D'autres approximations sont valables pour le début de l'épidémie, en particulier des approximations par des processus de branchement linéaires ([AB00] pour une revue). En calculant le nombre espéré de personnes contaminées par un unique infectieux dans une population saine, on peut classer ces processus de branchement en sous ou sur-critiques. Ce nombre est appelé nombre de reproduction ou  $R_0$  (voir [DHM90]). Une abondante littérature s'est focalisée sur les définitions et calculs de ce nombre de reproduction ou de quantités équivalentes. Mentionnons enfin les travaux de Barbour [Bar75b, Bar75a] ou de Barbour et Reinert [BR13], qui ont "recollé" les différentes approximations du processus décrivant l'épidémie pour obtenir une vue globale de sa propagation.

Au début de l'épidémie, les infectés sont négligeables dans la population et heuristiquement, le processus  $(I_t^n)_{t \in \mathbb{R}_+}$  peut-être approché par un processus de naissances et morts linéaire  $(X_t)_{t \in \mathbb{R}_+}$ , où la durée de vie des individus est une exponentielle de paramètre  $\rho$ , où chaque individu infectieux est muni d'un nombre d'arêtes tiré dans la distribution de degré  $(p_k)_{k \in \mathbb{N}}$  (dont on note  $g$  est la fonction génératrice) et où chacune de ces arêtes peut transmettre la maladie au taux  $\lambda$ .

**Proposition 2.2.4.** *Supposons que la distribution de degrés  $(p_k)_{k \in \mathbb{N}}$  admet des moments d'ordre 2. Soit*

$$R_0 = \frac{\lambda}{\lambda + \rho} \frac{g''(1)}{g'(1)}. \quad (2.2.12)$$

(i) *Le taux de croissance Malthusien de  $(X_t)_{t \geq 0}$  est  $r = \lambda \mathbb{E}(\tilde{D} - 2) - \rho$ , où  $\tilde{D}$  a pour distribution  $(k p_k / g'(1))_{k \in \mathbb{N}}$ .*

(ii) *Si  $R_0 \leq 1$ , alors le processus  $(X_t)_{t \geq 0}$  s'éteint presque sûrement.*

(iii)  *$R_0 = 1$  si et seulement si  $\mathbb{E}(\tilde{D} - 1) = (\lambda + \rho) / \lambda$ .*

(iv) *Si  $R_0 > 1$ , alors le processus  $(X_t)_{t \geq 0}$  s'éteint une probabilité  $z \in (0, 1)$ , solution de*

$$z = \frac{\rho}{g'(1)} \int_{\mathbb{R}_+} g'(z + e^{-\lambda y}(1 - z)) e^{-\rho y} dy. \quad (2.2.13)$$

(v) *Soient  $S_0 = \inf\{t \geq 0 \mid X_t = 0\}$  et  $S_{\varepsilon n} = \inf\{t \geq 0 \mid X_t \geq \varepsilon n\}$ . Si  $R_0 > 1$ , alors pour toute suite  $(t_n)_{n \in \mathbb{N}}$  telle que  $\lim_{n \rightarrow +\infty} t_n / \log(n) = +\infty$ ,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(S_0 \leq t_n \wedge S_{\varepsilon n}) = z \quad (2.2.14)$$

$$\lim_{n \rightarrow +\infty} \mathbb{P}(S_{\varepsilon n} \leq t_n \wedge S_0) = 1 - z. \quad (2.2.15)$$

*Démonstration.* Le taux de croissance Malthusien de  $(X_t)_{t \in \mathbb{R}_+}$  est défini comme la valeur  $r > 0$  telle que  $\int_0^{+\infty} e^{-rt} \mu(dt) = 1$ , où  $\mu(dt) = \sum_{k=1}^{+\infty} (k-1) \frac{k p_k}{g'(1)} e^{-\rho t} \lambda e^{-\lambda t} dt$  est la mesure intensité du processus ponctuel décrivant l'arrivée des nouvelles infections causées par un infectieux fixé (e.g.

[Jag75]). Conditionnellement au fait d’avoir un degré  $k$  et une durée de vie  $y$ , un individu génère un nombre de nouveaux infectieux suivant une loi binomiale de paramètre  $k - 1$  et  $1 - e^{-\lambda y}$ . L’intégration de ces quantités fournit le  $R_0$  et l’équation pour  $z$  annoncés.

La preuve du point (v) est une adaptation du Lemme of Méléard and Tran [MT09]. Le début de l’épidémie est donc heuristiquement une période de durée d’ordre  $\log(n)$ .  $\square$

**Exemple 2.2.5.** (i) **Poisson :** Pour une loi de Poisson de paramètre  $a$ , on trouve  $R_0 = \lambda a / (\lambda + \rho)$ . On a  $R_0 > 1$  si et seulement si  $a > 1 + \rho / \lambda$ .

(ii) **Géométrique :** Dans ce cas,  $f(s) = (1 - e^{-1/a}) / (1 - se^{-1/a})$ , et  $R_0 = \frac{2\lambda}{\lambda + \rho} \left( \frac{e^{-1/a}}{1 - e^{-1/a}} \right)$ . Alors  $R_0 > 1$  si et seulement si  $a > -\ln \left( \frac{\lambda + \rho}{3\lambda + \rho} \right)^{-1}$ .  $\square$

Le lien plus précis entre les processus  $X$  et  $I^n$  demande plus de travail, en particulier pour savoir quand le couplage n’est plus possible, et nous renvoyons à Barbour et Reinert [BR13] ou Janson, Luczak et Windridge [JLW14]. Pour l’épidémie, on a, comme pour le processus de branchement, une dichotomie. Lorsque  $R_0 > 1$ , on a avec grande probabilité que soit l’épidémie est petite avec peu d’infectés, soit on atteint une proportion positive de personnes infectées et les approximations déterministes deviennent valables. Lorsque  $R_0 \leq 1$ , on a avec probabilité 1 une “petite” épidémie.

## 2.2.2 Prise en compte de la structure de graphe dans l’épidémie cubaine

Supposer que la structure sociale est régie par un modèle de configuration ou tout autre modèle de graphe mathématiquement tractable (Erdős-Rényi, graphes d’intersection, petit-monde etc.) semble peu réaliste. Considérons les données cubaines.

**Présentation des données :** En 2006, nous disposons de 8 662 individus. Pour chacun d’eux, nous disposons de covariables telles que le sexe, l’orientation sexuelle, la province de résidence, le type de détection (“contact-tracing”, “captation” si l’individu s’est fait dépister suite à un suivi de son médecin de famille, ou “spontanée” si l’individu est allé se faire dépister de lui-même - on parle alors aussi de dépistage “aléatoire”), l’âge à la détection, la date de détection, la date possible d’infection lorsqu’elle est disponible, la date à laquelle le patient a développé le SIDA et la date de mort si elles existent. Nous disposons également pour ces individus du nombre de partenaires sexuels durant les deux années précédant le dépistage, et du nombre de partenaires dépistés par *contact-tracing*. Si ces partenaires sont séropositifs, alors ils figurent également dans la base de données et un lien entre les deux personnes est indiqué. Nous nous référons à [AJL<sup>+</sup>07, CART10] pour une description de la base et de sa construction.

Si l’on nettoie la base, on obtient un réseau de 5389 noeuds (les individus diagnostiqués séropositifs entre 1986 et 2006) : 1109 femmes (20.58%) et 4280 hommes (79.42%) dont 566 (10.50%) sont hétérosexuels et 3714 (68.92%) sont des hommes ayant des relations sexuelles avec des hommes (HSH). Ces sommets sont reliés par 4073 arêtes. Le graphe ainsi obtenu comporte une “composante géante” de 2386 noeuds (44% des individus recensés) et 3168 arêtes (78% du nombre global d’arêtes), ce qui est très intéressant statistiquement : dans la littérature des maladies sexuellement transmissibles, les composantes géantes sont souvent de tailles plus petites, ou ne contiennent qu’un faible nombre de personnes infectées. Par exemple, dans l’étude de Rothenberg et al. [RWP<sup>+</sup>95], la population compte 2200 individus, mais la composante géante seulement 965 personnes avec un très faible nombre de séropositifs. Dans l’étude de Wylie et Jolly [WJ01], 4544 individus forment la base de données, mais la composante géante ne compte que 82 personnes.

**Statistiques descriptives sur le graphe :** Pour une arête orientée, on appelle *ego* le départ de la flèche et *alter* l’arrivée. L’analyse statistique plus fine des *alter* et *ego* de chaque arête nous fournit des informations. Par exemple, lorsqu’un individu donne ses contacts, 77% en moyenne

sont testés. Ces proportions sont plus fortes chez les individus dépistés eux-mêmes par *contact-tracing* : 81%. Le temps de dépistage lorsque l'*ego* a été dépisté par *contact-tracing* ou *captation* est également plus court (699 et 428 jours) que lorsque l'*ego* a été dépisté de façon aléatoire (726 jours). Globalement, 22% des contacts se révèlent être positifs. Une analyse plus poussée de toutes les variables dont on dispose et considérant également les arbres d'infection et de détection est proposée dans [CART10].

**Clustering de la composante géante :** Du fait de la taille du graphe cubain, les représentations habituelles échouent à nous donner une représentation lisible des données. Pour analyser la connectivité et l'hétérogénéité du graphe, nous réalisons un *clustering* des données, qui consiste à rechercher une partition de la population maximisant une mesure de modularité, avec des groupes fortement intra-connectés et faiblement inter-connectés. Le résultat de cette maximisation (qui est un problème NP complet) est représenté en Fig. 2.2.2, obtenue dans [CART10].

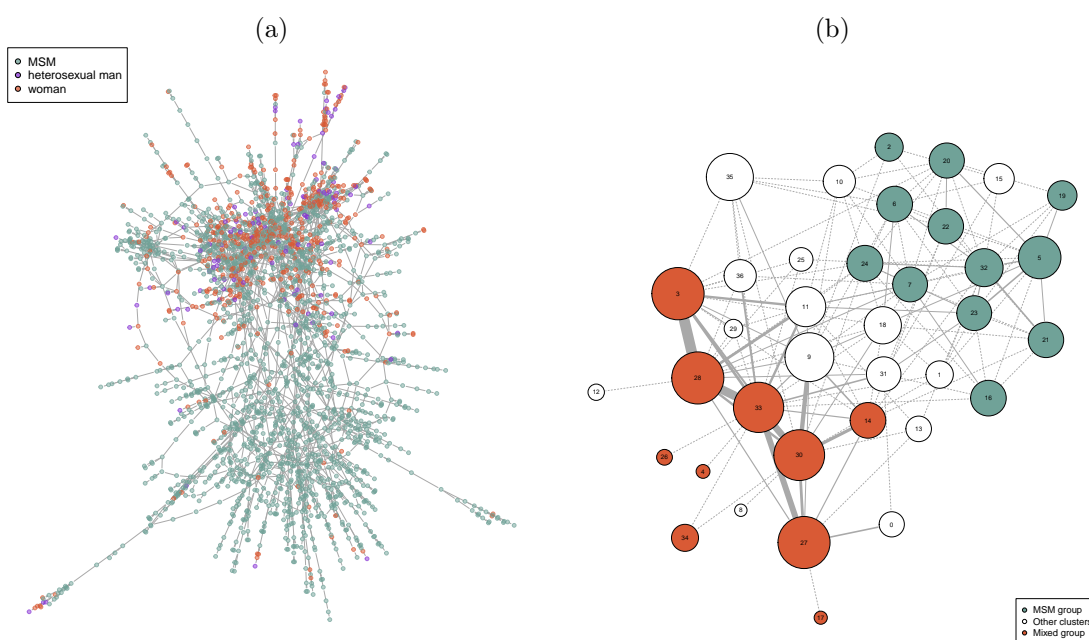


FIGURE 2.2.2 – (a) Représentation de la composante géante du graphe cubain. Les femmes sont en rose, les hommes hétérosexuels en bleu et les HSH en vert. Du fait du grand nombre d'individus, les structures ressortent très peu. (b) Représentation des 37 clusters, pour la composante géante du graphe cubain. Les clusters dans lesquels on trouve une majorité de HSH sont en rose, ceux où l'on trouve une majorité d'hétérosexuels en vert.

On peut calculer différents indicateurs statistiques pour décrire la structure mise en évidence, comme par exemple les coefficients d'association, de clustering, les distances géodésiques, la résilience etc. (cf. [CART10] pour les détails). On observe 37 *clusters* très intra-connectés mais très peu connectés entre eux. Le graphe se révèle être très hétérogène, avec une densité globalement faible, mais certains *clusters* sont beaucoup plus denses que la moyenne. Fortunato et Barthélemy [FB07] ont montré que la résolution limite pour la modularité était  $\sqrt{2m}$  où  $m$  est le nombre d'arêtes de la composante analysée. Ici, cette résolution limite est d'environ 80. Parmi les 37 *clusters* obtenus, 28 ont plus de 80 arêtes et 6 parmi eux ont une sous-structure significative. Pour les autres, l'approximation du modèle de configuration est valable (voir par exemple le *cluster 3* de la Fig. 2.2.3). On peut alors envisager d'utiliser les équations de Volz pour décrire l'évolution de la maladie dans

ces *clusters*.

En plus de permettre une meilleure représentation des données, le *clustering* met également en évidence des phénomènes ayant participé à la formation du graphe. Nous illustrons cela en commentant la répartition des orientations sexuelles dans les *clusters* obtenus. Sur les 37 *clusters*, 22 présentent une composition, en terme de répartition des orientations sexuelles, qui est significativement différente de la répartition globale sur l'ensemble de la population. Sur la Fig. 2.2.2 (b), on voit que le bas du graphe est composé de *clusters* à dominance hétérosexuelle, tandis que le haut du graphe est composé de *clusters* à dominance HSH. Ceci reflète en fait la propagation historique de l'épidémie : importée sur l'île par des militaires hétérosexuels, la maladie s'est propagée aux femmes, puis aux bisexuels et aux HSH.

## 2.3 Conclusion

En 2012, parce qu'il avait eu connaissance de ces travaux, l'un de mes anciens étudiants de master, Anthony Cousien, alors ingénieur d'étude à l'INSERM, m'a fait rencontrer Yazdan Yazdanpanah (INSERM) avec qui nous nous sommes mis à travailler sur la modélisation de l'épidémie du VHC chez les usagers de drogues. De nos discussions, et avec Jean-Stéphane Dhersin, Sylvie Burban (statisticienne de l'INSERM) et Marie Jauffret-Roustide (sociologue de l'INVS) est né un projet financé par l'ANRS (Agence Nationale de Recherche sur le SIDA et les hépatites virales) et organisé autour de la thèse qu'a entreprise Anthony que je co-dirige avec Yazdan Yazdanpanah. Le but est de modéliser la propagation de l'hépatite C chez les usagers de drogues, en prenant en compte les réseaux d'injection, diverses covariables, ainsi qu'un niveau de résolution suffisamment détaillé sur la situation de l'individu dans le réseau de santé et sur son stade dans la maladie afin de tester différents scénarii de santé publique. L'étude est motivée par la sortie prochaine de nouveaux traitements contre le VHC et la volonté de réaliser une étude coût-efficacité précise pour comparer différentes stratégies possibles (réduction des risques, augmentation du nombre de traitements, augmentation de l'efficacité des traitements et association de ces stratégies). Le test de scénarii se fait par des simulations qui ont été implémentées par Anthony. La calibration des paramètres du modèle développé s'appuie sur la littérature et sur des estimations ABC. Une difficulté réside dans le choix du réseau sur lequel on propage la maladie dans les simulations. Très peu d'information statistique est disponible sur les réseaux d'usagers de drogues et ainsi une collaboration a été initiée avec la sociologue Marie Jauffret-Roustide pour lancer une enquête *respondent driven sampling* (RDS) qui permettra de reconstituer partiellement une partie de ce réseau et d'en considérer les propriétés. Au cours de ce projet, différentes questions théoriques liées à l'analyse de sensibilité ou à la modélisation du réseau RDS qui sera découvert par l'enquête ont émergées et font l'objet de travaux en cours.

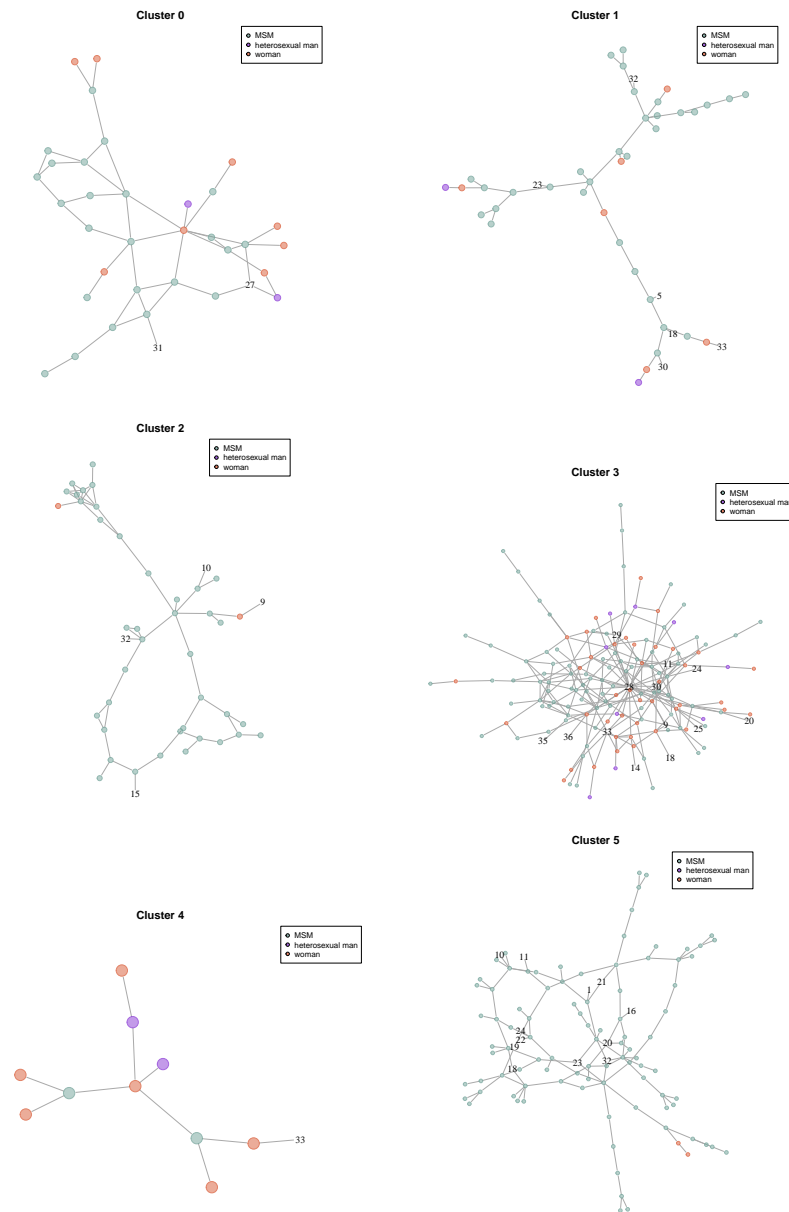


FIGURE 2.2.3 – 6 premiers clusters de la composante géante.

# Chapitre 3

## Deux problèmes de géométrie aléatoire

### Sommaire

---

<b>3.1</b>	<b>Espérances d'ensembles aléatoires et moyennes de Vorob'ev associées</b>	<b>78</b>
3.1.1	Espérance de Vorob'ev . . . . .	78
3.1.2	Estimateur des ensembles de niveau . . . . .	79
3.1.3	Estimation de l'espérance de Vorob'ev . . . . .	81
<b>3.2</b>	<b>Arbre couvrant radial sur un processus ponctuel de Poisson</b>	<b>82</b>
3.2.1	Forêt dirigée . . . . .	84
3.2.2	Arbre radial . . . . .	88
<b>3.3</b>	<b>Conclusion</b>	<b>94</b>

---

Dans ce chapitre, deux problèmes différents issus de la géométrie aléatoire seront abordés. Dans la Section 3.1, nous considérons une suite de  $n$  ensembles aléatoires  $(X_i)_{i \in \{1, \dots, n\}}$  et étudions l'espérance de ces ensembles, au sens de Vorob'ev. L'espérance, au sens de Vorob'ev, est étroitement liée à l'étude des ensembles de niveaux qui ont fait l'objet d'une abondante littérature statistique (Cuevas et al. [CF97, CGMRC06], Molchanov [Mol87, Mol98], Rigollet et Vert [RV09], Tsybakov [Tsy97], Singh et al. [SSN09]). Puis, nous proposons un estimateur numériquement implémentable de cette espérance, en nous inspirant des travaux de Kovyazin [Kov86]. Cet estimateur s'appuie sur la distribution empirique des  $X_i$  et sur une discrétisation spatiale de l'espace.

Dans une seconde section, nous présentons des résultats sur l'arbre radial introduit par Baccelli et Bordenave [BB07] : le RST (*Radial Spanning Tree*). Il s'agit d'un graphe géométrique, au sens où la position des sommets dans  $\mathbb{R}^2$  détermine la présence ou non d'arêtes entre eux. Le RST est construit sur un processus de Poisson homogène dans le plan  $\mathbb{R}^2$  et est conditionné à contenir l'origine  $O$ . Cet arbre présente, de part sa construction, des dépendances géométriques qui rendent son étude complexe et nécessitent, pour comprendre sa structure, l'utilisation d'outils très variés. Le RST possède des propriétés d'invariance en loi par rotation autour de l'origine, et c'est un arbre planaire, au sens où ses arêtes ne se coupent pas. Nous commencerons par étudier les chemins semi-infinis ou géodésiques du RST. Pour cela, nous utiliserons une propriété d'étirement qui a été introduite par Howard et Newman [HN01] et qui nous fournit des résultats puissants sur les directions asymptotiques de ces chemins. Ceci et la construction de chemins semi-infinis auxiliaires nous permettront de compléter le résultat de Baccelli et Bordenave [BB07] inspiré de [HN01], en montrant qu'il n'existe presque sûrement, pour  $\theta \in [0, 2\pi)$  fixé, qu'une unique géodésique de direction asymptotique  $\theta$  dans le RST. En modifiant localement le processus de Poisson des

sommets, nous pourrions également décrire plus précisément les interfaces du RST et appliquerons ces résultats à un problème de coloriage de sous-arbres. Enfin, le couplage local, loin de  $O$ , entre le RST et une forêt dirigée nous permettra de montrer que le nombre de chemins semi-infinis que l'on rencontre à une distance  $r$ , grande, de  $O$  croît de façon sous-linéaire en espérance.

Les prolongements possibles seront expliqués en conclusion.

### 3.1 Espérances d'ensembles aléatoires et moyennes de Vorob'ev associées

La définition de l'espérance d'un ensemble aléatoire  $X$  ou de la moyenne d'un échantillon de  $n$  ensembles aléatoires  $(X_i)_{i \in \{1, \dots, n\}}$  est une question naturelle. Elle peut émerger dans des problèmes d'écologie, d'évolution ou d'épidémiologie tels que ceux présentés aux chapitres précédents. Pour Radu Stoica [SGK07, SMS07, SMS10], l'objectif était de détecter des motifs dans des données spatiales de filaments galactiques. A très grande échelle, on peut assimiler les galaxies à des points qui forment visuellement, sous le jeu des forces gravitationnelles, des objets géométriques complexes : clusters, filaments...

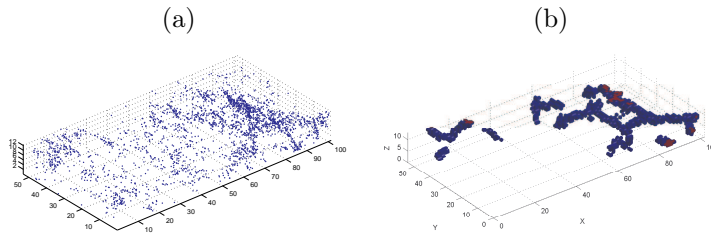


FIGURE 3.1.1 – Echantillon NGP150 de l'étude 2dFGRS (a) et l'estimateur de Vorob'ev de son espérance (b).

Soit  $X$  un compact aléatoire de  $[0, 1]^d$  (pour simplifier). Ce compact aléatoire est une fonction de l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  dans l'ensemble  $\mathfrak{C}$  des compacts de  $[0, 1]^d$  mesurable dans le sens où  $\forall C \in \mathfrak{C}, \{\omega \in \Omega, X(\omega) \cap C \neq \emptyset\} \in \mathcal{F}$ .

Différentes espérances de  $X$  peuvent être définies (cf. par exemple [SM97, BM98, Mol99] ou [Mol05] pour une revue). Dans beaucoup de cas, même si  $X$  n'est pas convexe ou même connexe, on attend comme espérance un ensemble convexe et l'espérance d'Aumann convient. Si le volume de  $X$  n'est pas d'espérance nulle et que l'on cherche une espérance reflétant davantage une notion d'occupation moyenne et d'ensemble de niveaux, alors il est naturel de considérer l'espérance de Vorob'ev [Vor77].

L'espérance de Vorob'ev en elle-même a été relativement peu étudiée. Kovyazin [Kov86] propose un estimateur de l'espérance de Vorob'ev lorsque l'on dispose d'un échantillon i.i.d.  $(X_i)_{i \in \{1, \dots, n\}}$  de  $n$  ensembles aléatoires de même loi que  $X$ , mais qui fut peu cité. Par contre, l'étude des ensembles de niveaux, notion à laquelle est liée l'espérance de Vorob'ev, ont fait l'objet d'une abondante littérature.

#### 3.1.1 Espérance de Vorob'ev

En notant  $\lambda(dx)$  la mesure de Lebesgue, remarquons que  $\mathbb{E}\lambda(X) = \int_{[0,1]^d} \mathbb{P}(x \in X) \lambda(dx)$ . Il est naturel de considérer la fonction de couverture  $p(x) = \mathbb{P}(x \in X)$  et son inverse généralisée. On peut alors définir pour  $\alpha \in [0, 1]$ , l'ensemble de niveau  $\alpha$  associé à  $p(x)$  comme l'ensemble (déterministe) :

$$Q_\alpha = \{x \in [0, 1]^d : p(x) > \alpha\}$$

En choisissant  $\alpha$  de sorte à ce que le volume de  $Q_\alpha$  corresponde au volume moyen de  $X$  (l'égalité n'est pas toujours possible), on obtient l'espérance de Vorob'ev  $\mathbb{E}_V(X)$  (e.g. [Kov86, Mol99, Mol05, SS94]).

**Définition 3.1.1.** Soit  $F(\alpha) = \lambda(Q_\alpha) = \lambda\{p > \alpha\}$  et  $\alpha^* = \inf\{\alpha \in [0, 1] : F(\alpha) \leq \mathbb{E}\lambda(X)\}$ . L'espérance de Vorob'ev de  $X$  est l'ensemble mesurable  $\mathbb{E}_V(X)$  tel que  $\lambda(\mathbb{E}_V(X)) = \mathbb{E}\lambda(X)$  et

$$\{p > \alpha^*\} \subset \mathbb{E}_V(X) \subset \{p \geq \alpha^*\}. \quad (3.1.1)$$

L'unicité de  $\mathbb{E}_V(X)$  n'est assurée que si  $F$  est continue en  $\alpha^*$ , auquel cas  $\mathbb{E}_V(X) = \{p \geq \alpha^*\}$ , qui est compact par semi-continuité supérieure de  $p(x)$ . En général,  $F$  est càdlàg.

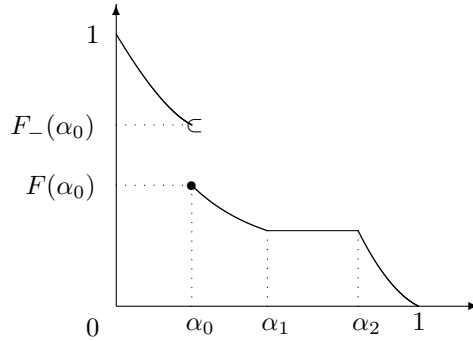


FIGURE 3.1.2 – Exemple de fonction  $F(\alpha) = \lambda(Q_\alpha)$ . Les plateaux de  $p(x)$  forment les discontinuités de  $F(\alpha)$  tandis que les discontinuités de  $p(x)$  forment les plateaux de  $F(\alpha)$ . Un saut de  $F$  en  $\alpha_0$  signifie que la mesure de Lebesgue des points qui sont recouverts par  $X$  avec la probabilité  $\alpha_0$  est non nulle, plus précisément :  $\lambda\{p = \alpha_0\} = F_-(\alpha_0) - F(\alpha_0)$ . Un plateau  $[\alpha_1, \alpha_2]$  de  $F$  signifie que les points recouverts avec une probabilité comprise entre  $\alpha_1$  et  $\alpha_2$  sont de mesure de Lebesgue nulle :  $\lambda\{p > \alpha_1\} = \lambda\{p > \alpha\}$  pour tout  $\alpha \in [\alpha_1, \alpha_2]$ .

Il est possible de montrer facilement ([Mol05, Theorem 2.3. p.177]) que  $\mathbb{E}_V(X)$  minimise  $B \mapsto \mathbb{E}\lambda(B \triangle X)$  sous la contrainte  $\lambda(B) = \mathbb{E}\lambda(X)$ , où  $B \triangle X = (B \setminus X) \cup (X \setminus B)$  est la différence symétrique de  $B$  et  $X$ .

Ni  $Q_\alpha$  ni  $\mathbb{E}_V(X)$  ne sont directement implémentables numériquement. Premièrement, la fonction de couverture  $p(x)$  n'est pas toujours disponible sous forme analytique. Pour pallier à cela, des estimateurs *plug-in*, où  $p(x)$  est remplacée par une fonction de couverture empirique, ont été étudiés par Molchanov [Mol87, Mol90, Mol98], Cuevas et al [CF97, CGMRC06] qui démontrent la consistance en norme  $L^1$  de ces estimateurs sous des hypothèses faibles, sans par exemple supposer la continuité de  $p(x)$ . Deuxièmement, le calcul de ces ensembles nécessite une évaluation en tout les points  $x \in [0, 1]^d$ . Nous allons considérer un estimateur introduisant une discrétisation spatiale. Enfin, pour estimer l'espérance de Vorob'ev, il est également nécessaire d'estimer  $\alpha^*$  (qui dépend de  $X$ ). Ce problème a été considéré par Kovyazin, mais à nouveau sans discrétisation spatiale.

### 3.1.2 Estimateur des ensembles de niveau

La contrepartie empirique de  $p(x)$  est :

$$p_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \in X_i\}}.$$

Nous pouvons alors définir l'estimateur de  $Q_\alpha$ , pour  $\alpha \in [0, 1]$  fixé par  $Q_{n,\alpha} = \{p_n > \alpha\}$ .



Dans la littérature, différentes distances peuvent être utilisées pour les ensembles fermés. Nous nous intéressons à la pseudo-distance définie pour  $A, B \in \mathcal{B}([0, 1]^d)$  par  $d(A, B) = \lambda(A \Delta B)$ . Des résultats similaires à ceux présentés ici ont été obtenus pour la distance de Hausdorff par exemple, sous des hypothèses de régularité supplémentaire pour  $p(x)$  (e.g. [Mol90] et [CGMRC06, Th. 1 et 2]). Dans le cas où  $p(x)$  est une densité de probabilité, on peut choisir des estimateurs à noyaux au lieu de  $p_n(x)$ . Les propriétés des estimateurs (taux de convergence, normalité asymptotique) ont été considérés dans [Mol98, RV09, Tsy97, MP09]. L'estimation adaptative a été étudiée par [SSN09].

Pour les applications, nous ne souhaitons pas supposer la continuité de  $p(x)$ . Nous suivons ainsi l'approche de [CGMRC06], et généralisons leurs résultats en ajoutant une discrétisation spatiale - nécessaire à toute implémentation numérique.

**Définition 3.1.2.** Soit  $B$  un ensemble mesurable de  $[0, 1]^d$  et soit  $r \in 2^{-\mathbb{N}} = \{2^{-k}, k \in \mathbb{N}\}$ . La discrétisation de pas  $r$  de  $B$  est définie comme étant

$$B^r = \bigsqcup_{x \in B \cap r\mathbb{Z}^d} x + [0, r]^d.$$

Un nouvel estimateur naturel de  $Q_\alpha$  est donc  $Q_{n,\alpha}^r$ .

Il est naturel que la régularité du bord  $\partial Q_\alpha$  intervienne dans l'étude des propriétés statistiques de l'estimateur  $Q_{n,\alpha}^r$  :

**Définition 3.1.3.** (cf. [Fal97, p.38-39]) Soit  $B$  un ensemble mesurable de  $[0, 1]^d$  et soit

$$N_r(\partial B) = \text{Card}\{x \in r\mathbb{Z}^d : (x + [0, r]^d) \cap \partial B \neq \emptyset\}.$$

La dimension de Kolmogorov (*upper box counting dimension*) de  $\partial B$  est

$$\overline{\dim}_{\text{box}}(\partial B) = \limsup_{r \rightarrow 0} \frac{\log N_r(\partial B)}{-\log r}.$$

Nous étendons alors le résultat de Cuevas et al. [CGMRC06, Th. 3] comme suit :

**Proposition 3.1.4.** *Supposons qu'il existe  $\kappa > 0$  tel que  $\overline{\dim}_{\text{box}}(\partial X) \leq d - \kappa$  p.s. Alors pour tout  $\alpha$  tel que  $\lambda\{p = \alpha\} = 0$ ,*

1. *avec probabilité 1,*

$$\lim_{\substack{r \rightarrow 0 \\ n \rightarrow \infty}} d(Q_{n,\alpha}^r, Q_\alpha) = 0.$$

2. *pour tout  $\varepsilon > 0$ ,  $\mathbb{E} d(Q_{n,\alpha}^r, Q_\alpha) \leq r^\kappa + 2\varepsilon^{-2n\varepsilon^2} + F(\alpha - \varepsilon) - F(\alpha + \varepsilon)$ .*

*Idée de preuve.* La preuve de ce résultat s'appuie sur l'inégalité triangulaire

$$\mathbb{E} d(Q_{n,\alpha}^r, Q_\alpha) \leq \mathbb{E} d(Q_{n,\alpha}^r, Q_{n,\alpha}) + \mathbb{E} d(Q_{n,\alpha}, Q_\alpha). \quad (3.1.2)$$

Le premier terme du membre de droite de (3.1.2) se majore en remarquant que l'hypothèse  $\overline{\dim}_{\text{box}}(\partial X) \leq d - \kappa$  implique que pour tout  $\varepsilon$ , il existe  $r_\varepsilon > 0$  tel que pour tout  $r \in (0, r_\varepsilon)$ ,  $d(Q_{n,\alpha}^r, Q_{n,\alpha}) \leq r^{d - \overline{\dim}_{\text{box}}(\partial Q_{n,\alpha}) - \varepsilon} < r^{\kappa - \varepsilon}$ . En faisant tendre  $\varepsilon$  vers 0, comme le membre de gauche ne dépend pas de  $\varepsilon$ , on peut majorer le premier terme par  $r^\kappa$ . Pour le second terme de (3.1.2), en utilisant l'inégalité de Bernstein, on a pour tout  $\varepsilon > 0$

$$\begin{aligned} \mathbb{E} d(Q_{n,\alpha}, Q_\alpha) &= \mathbb{E}(\lambda\{p_n > \alpha, p \leq \alpha\} + \lambda\{p_n \leq \alpha, p > \alpha\}) \\ &\leq \int_{[0,1]^d} \mathbb{P}\{|p_n(x) - p(x)| \geq |p(x) - \alpha|\} \lambda(dx) \\ &\leq \int_{\{|\alpha - p| \geq \varepsilon\}} 2\varepsilon^{-2n|\alpha - p(x)|^2} \lambda(dx) + \int_{\{|\alpha - p| < \varepsilon\}} \lambda(dx) \\ &\leq 2\varepsilon^{-2n\varepsilon^2} + F(\alpha - \varepsilon) - F(\alpha + \varepsilon). \end{aligned}$$

□

### 3.1.3 Estimation de l'espérance de Vorob'ev

L'espérance de Vorob'ev est l'ensemble de niveau dont le volume est  $\mathbb{E}\lambda(X)$ . Soit  $F_n(\alpha) = \lambda\{p_n > \alpha\}$  la fonction de couverture empirique et  $\Lambda_n = \frac{1}{n} \sum_{i=1}^n \lambda(X_i)$  le volume moyen. L'idée naturelle de Kovyazin est de définir

$$\alpha_n^* = \inf\{\alpha \in [0, 1] : F_n(\alpha) \leq \Lambda_n\}.$$

**Définition 3.1.5.** La “moyenne de Kovyazin” est un ensemble aléatoire mesurable  $K_n$  défini par :  $\lambda(K_n) = \Lambda_n$  et  $\{p_n > \alpha_n^*\} \subset K_n \subset \{p_n \geq \alpha_n^*\}$ .

Cet ensemble moyen n'est pas nécessairement unique. Kovyazin [Kov86] a montré que si  $\lambda\{p = \alpha^*\} = 0$ , alors  $d(K_n, \mathbb{E}_v(X)) \xrightarrow[n \rightarrow \infty]{} 0$  p.s.

Introduisons maintenant une discrétisation spatiale. Une première idée pourrait être de discrétiser  $X$  ou  $K_n$ , mais on obtiendrait alors un ensemble dont le volume ne serait pas nécessairement  $\Lambda_n$ .

**Définition 3.1.6.** Soit

$$\alpha_{n,r}^* = \inf\{\alpha \in [0, 1] : (\{p_n > \alpha\}^r) \leq \Lambda_n\}. \quad (3.1.3)$$

Nous définissons l'estimateur discrétisé  $K_{n,r}$  de l'espérance de Vorob'ev comme l'ensemble mesurable aléatoire de volume  $\Lambda_n$  tel que

$$\{p_n > \alpha_{n,r}^*\}^r \subset K_{n,r} \subset \{p_n \geq \alpha_{n,r}^*\}^r. \quad (3.1.4)$$

Pour calculer  $K_{n,r}$  en pratique, le cube  $[0, 1]^d$  est discrétisé suivant une grille de pas  $r$ . A chaque cellule  $x + [0, r)^d$ , on associe la couverture estimée  $p_n(x)$  au point  $x$ . Nous calculons le volume moyen  $\Lambda_n$ , qui nous fournit le nombre de cellules à choisir. Ces dernières sont choisies suivant leur taux de couverture estimé.

Pour énoncer la consistance de l'estimateur, nous définissons également  $\beta^* = \sup\{\alpha \in [0, 1] : F(\alpha) \geq \mathbb{E}\lambda(X)\}$ .

**Théorème 3.1.7.** *Supposons que*

1. *il existe  $\kappa > 0$  tel que*

$$\mathbb{P}(\overline{\dim_{\text{box}}(\partial X)} \leq d - \kappa) = 1. \quad (3.1.5)$$

2.  $\lambda\{p = \alpha^*\} = \lambda\{p = \beta^*\} = 0$ .

*Alors, presque sûrement,  $d(K_{n,r}, \mathbb{E}_V(X)) \xrightarrow[r \rightarrow 0]{n \rightarrow \infty} 0$ .*

L'idée de la preuve est la suivante. Par inégalité triangulaire,  $d(K_{n,r}, \mathbb{E}_V(X)) \leq d(K_{n,r}, K_n) + d(K_n, \mathbb{E}_V(X))$ . Le second terme est majoré en utilisant le résultat de Kovyazin. La preuve du premier terme s'appuie sur le lemme suivant :

**Lemme 3.1.8.** *Supposons que (3.1.5) soit satisfaite. Alors, presque sûrement,*

- (i)  $\alpha^* \leq \liminf_{r \rightarrow 0} \alpha_{n,r}^* \leq \limsup_{r \rightarrow 0} \alpha_{n,r}^* \leq \beta^*$ , *d'où l'on déduit également que*

$$\alpha^* \leq \liminf_{n \rightarrow \infty} \alpha_n^* \leq \limsup_{n \rightarrow \infty} \alpha_n^* \leq \beta^*.$$

- (ii)  $\limsup_{r \rightarrow 0} d(K_{n,r}, K_n) \leq 2 \left[ \lim_{\substack{\alpha \rightarrow \alpha^* \\ \alpha < \alpha^*}} F(\alpha) - F(\beta^*) \right]$ .

### 3.2 Arbre couvrant radial sur un processus ponctuel de Poisson

Baccelli et Bordenave [BB07] ont introduit un arbre planaire radial couvrant un processus ponctuel de Poisson homogène, auquel nous nous sommes intéressés avec David Coupier, ce qui a motivé plusieurs travaux avec également François Baccelli, Anish Sarkar et Kumarjit Saha.

L'arbre planaire couvrant radial introduit par Baccelli et Bordenave [BB07], appelé RST (*Radial Spanning Tree*) est construit de la façon suivante. Soit  $N$  un processus ponctuel de Poisson homogène dans le plan  $\mathbb{R}^2$  muni de sa métrique euclidienne et considéré dans sa version de Palm, c'est-à-dire conditionné à contenir l'origine  $O$ . On pourra supposer ce processus d'intensité 1. Dans la suite,  $\mathbb{R}^2$  sera muni de sa base canonique orthonormale  $(O, e_x, e_y)$  et on pourra caractériser de façon classique tout point de  $\mathbb{R}^2$  par son abscisse et son ordonnée  $(x, y)$ . On sera également amené à identifier  $\mathbb{R}^2$  au plan complexe, et un point sera aussi décrit par son rayon et son angle  $re^{i\theta}$ .

**Définition 3.2.1.** Le RST est le graphe aléatoire  $\mathcal{T}$  dont les sommets sont les atomes de  $N$  et dont les arêtes sont les paires  $\{X, Y\}$  avec  $X, Y \in N$  et  $X \neq O$  qui vérifient

$$|Y| < |X| \text{ and } N \cap B(O, |X|) \cap B(X, |X - Y|) = \emptyset \quad (3.2.1)$$

où  $B(c, r)$  est la boule de centre  $c$  et de rayon  $r$ .  $\square$

Pour chaque sommet  $X \in N \setminus \{O\}$ , il existe presque-sûrement un unique  $Y \in N$  vérifiant (3.2.1), que l'on appelle l'ancêtre de  $X$  et qui sera aussi noté  $A(X)$ . C'est, parmi les sommets de  $N \cap B(O, |X|)$ , celui qui est le plus proche en distance euclidienne de  $X$ . Nous obtenons ainsi presque-sûrement un arbre enraciné en  $O$ . Par convention, on pose  $A(O) = O$ .

**Définition 3.2.2.** Une suite  $(X_n)_{n \geq 0}$  de sommets de  $N$  est un chemin semi-infini du RST  $\mathcal{T}$  si pour tout  $n$ ,  $X_n = A(X_{n+1})$ . On dit que ce chemin semi-infini a une direction asymptotique  $\theta \in [0, 2\pi)$  si

$$\lim_{n \rightarrow \infty} \frac{X_n}{|X_n|} = e^{i\theta}.$$

Il existe plusieurs autres modèles d'arbres radiaux : Pimentel [Pim07], Bonichon et Marckert [BM11], Norris Turner et leurs co-auteurs [VST12, VST13, NT12, NT08, NT], Coletti et Valencia [CV13]. La difficulté dans notre cas réside dans la définition locale des arêtes du RST (3.2.1), générant des dépendances complexes entre voisins, ce qui empêche l'utilisation de processus Markoviens et de théorèmes de convergence de martingales pour étudier les propriétés de cet arbre.

Cet arbre présente cependant certaines bonnes propriétés. Un exemple est le fait d'être planaire et de ne pas avoir de croisement entre différentes arêtes. Un autre est le fait d'être étiré (*straight*), dont nous verrons les conséquences plus tard. Nous utiliserons cette propriété d'être *straight* pour l'étude des propriétés fines des chemins semi-infinis du RST et de leurs directions asymptotiques, si elles existent. Ces chemins sont liés aux notions d'interfaces et de géodésiques de l'arbre (cf. par exemple [Cou11, CH12, CH11] dans le cas d'arbres obtenus par percolation de dernier passage) et sont très importants. Dans le cas du RST, une jolie motivation de l'étude de ces chemins est la suivante. Les contraintes géométriques associées à la construction du RST impliquent que, presque sûrement,  $O$  a au plus 5 descendants directs. De chacun de ces descendants de  $O$  est issu un arbre, constitué des descendants de ce point, cet arbre étant fini ou infini. Il est possible de colorier chacun de ces arbres d'une couleur différente et d'étudier leurs formes. Les séparations entre ces arbres sont des interfaces qui sont étroitement liées aux chemins semi-infinis du RST... On montrera que l'on peut avoir  $m$  arbres colorés infinis pour  $m \in \{1, \dots, 5\}$ . Ceci est illustré à la Figure 3.2.1. Il est bien sûr aussi possible que certains arbres colorés soient bornés (Figure 3.2.2).

Par ailleurs, dans une région relativement petite du plan et très éloignée de l'origine  $O$ , une approximation intuitive est de dire que les branches de l'arbre ne sont pas dirigées vers un point  $O$ ,

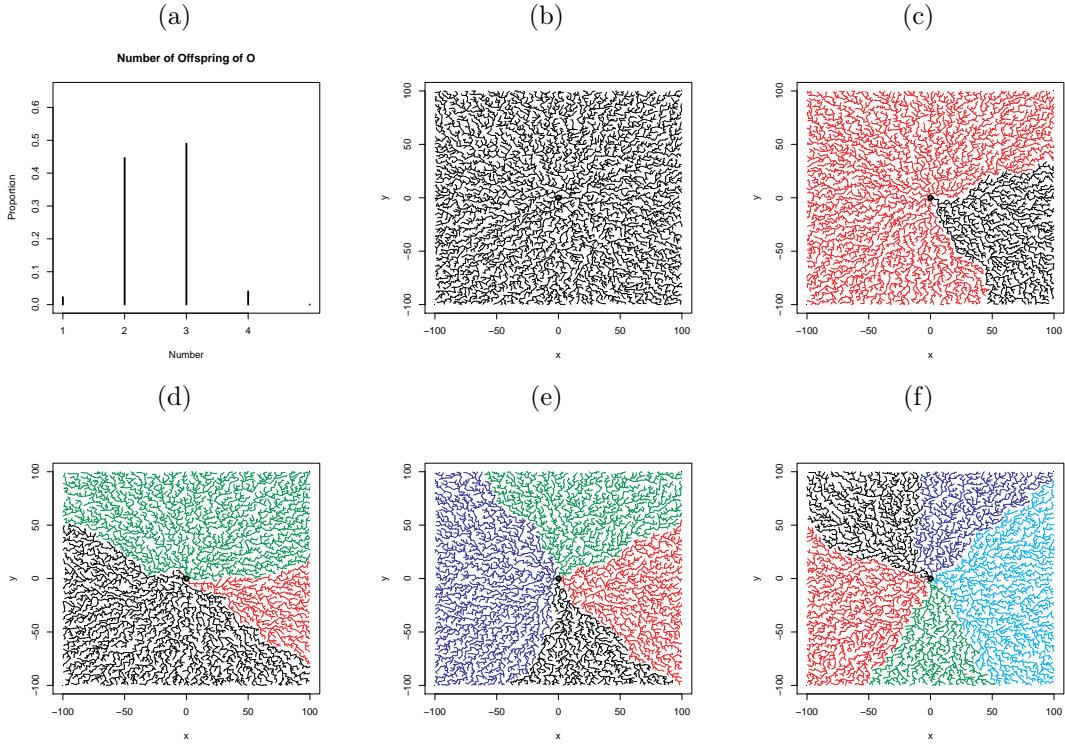


FIGURE 3.2.1 – (a) Distribution empirique du nombre de descendants directs de  $O$ , obtenue à partir de 5000 simulations. (b)-(f) : Réalisations du RST à  $m = 1, \dots, 5$  sous-arbres colorés enracinés sur les descendants directs de  $O$ .

mais dans une direction. Ceci a amené Baccelli et Bordenave [BB07] à introduire une forêt dirigée, la DSF (*Directed Spanning Forest*) dont les propriétés nous fournissent des outils pour étudier le RST.

**Définition 3.2.3.** Soit  $N$  un processus ponctuel de Poisson sur  $\mathbb{R}^2$  d'intensité 1. La DSF de direction  $e_x$ , notée  $\mathcal{T}_{e_x}$  est le graphe aléatoire dont les sommets sont les atomes de  $N$  et dont les arêtes sont définies comme suit : l'ancêtre de  $X \in N$  est le plus proche point de  $N$  (au sens de la distance euclidienne) dont l'abscisse est strictement supérieure à celle de  $X$ .  $\square$

Chaque sommet  $X$  n'ayant presque sûrement qu'un seul ancêtre, nous définissons ainsi une forêt aléatoire. Cette forêt est une approximation du RST "loin de  $O$ " au sens suivant :

**Lemme 3.2.4.** *Considérons le RST  $\mathcal{T}$  et la DSF  $\mathcal{T}_{-e_x}$  construits sur le même processus ponctuel de Poisson  $N$ . Pour tout  $0 < \alpha < 1/3$  :*

$$\lim_{r \rightarrow +\infty} \mathbb{P}(\mathcal{T} \cap B((r, 0), r^\alpha) = \mathcal{T}_{-e_x} \cap B((r, 0), r^\alpha)) = 1. \quad (3.2.2)$$

En particulier, la limite précédente reste valable pour un voisinage de rayon constant  $R$  de  $(r, 0)$ .

Nous utiliserons cette approximation pour quantifier le nombre de chemins semi-infinis du RST qui intersectent le cercle de rayon  $r$ . Nous montrerons que ce nombre croît de façon sous-linéaire. Dans la littérature, d'autres forêts dirigées ont été étudiées (e.g. [ARS08, FP05b, GRS04, LN96] sur  $\mathbb{Z}^2$  ou [FLT04, HN01] sur des processus de Poisson homogènes). Dans ces articles, la règle de construction des arêtes implique que le choix de l'ancêtre d'un point  $(x, y)$  ne dépend pas de ce qui se passe avant l'abscisse  $x$ . Pour la DSF (cf. Fig. 3.2.3), si  $Y$  est un descendant de  $(x, y)$  et

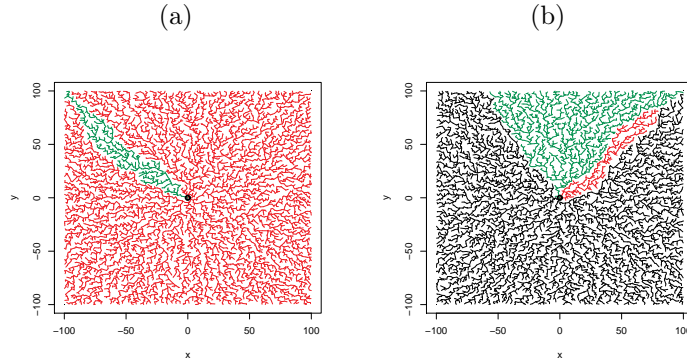


FIGURE 3.2.2 – (a) Deux arbres, dont l'un est borné. (b) Trois arbres, dont l'un est borné.

$Y' = A(Y)$  est son ancêtre, la boule  $B(Y, |Y - Y'|)$  qui est vide de points de  $N$  par construction, peut intersecter le demi-plan  $\{(x', y') \in \mathbb{R}^2, x' > x\}$  générant de la non-Markoviannité puisque l'ancêtre de  $(x, y)$  ne peut alors pas appartenir à l'intersection de ce demi-plan et de l'ensemble des boules associées au passé de la forêt.

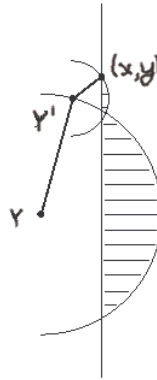


FIGURE 3.2.3 – La détermination de l'ancêtre du point  $Y$  nous apporte l'information que la boule  $B(Y, |Y - A(Y)|)$  est vide. La trajectoire peut ainsi dépendre de façon forte de son passé si elle avait de très grandes arêtes par exemple.

Dans Roy Saha et Sarkar [RSS14], un modèle discret analogue à la DSF est introduit et les auteurs utilisent la distance de graphe, ce qui simplifie la forme des intersections cassant la markoviannité.

### 3.2.1 Forêt dirigée

Considérons la DSF  $\mathcal{T}_{e_x}$  définie précédemment. Comme elle est construite sur le processus de Poisson homogène  $N$ , isotropique, on peut généraliser les résultats suivants à une direction  $u \in \mathbb{R}^2$  quelconque.

Pour un point  $X \in N$ , définissons sa trajectoire ancestrale comme étant la suite  $(A^n(X))_{n \in \mathbb{N}}$  avec  $A^0(X) = X$  et  $A^n(X) = A(A^{n-1}(X))$  pour  $n \geq 1$ . Chaque point ayant un et un seul ancêtre, on a des trajectoires infinies quand on les parcourt dans le sens des ancêtres, dans la direction de  $e_x$ . Une trajectoire  $\gamma$  est dite bi-infinie si elle est aussi infinie quand on la parcourt dans la direction

$-e_x$ , impliquant que tout point  $X \in N \cap \gamma$  est l'ancêtre d'un autre point de  $N$ , qui appartient aussi à  $\gamma$ .

Avec David Coupier, nous avons établi [CT13] :

**Théorème 3.2.5.** *Presque sûrement :*

(i) *La DSF est un arbre, c'est-à-dire que les trajectoires ancestrales de  $X$  et  $Y$  coalescent presque-sûrement.*

(ii) *Il n'y a pas de trajectoire bi-infinie, c'est-à-dire que les branches de l'arbre se terminent toutes presque-sûrement avec un point qui n'a pas de descendant.*

Le point (i) du Théorème 3.2.5 est illustré à la Figure 3.2.4. Ce résultat n'est pas surprenant en regard de ce qui est connu pour les arbres obtenus par percolation de premier et dernier passage sur  $\mathbb{Z}^2$  (cf. Licea et Newman [LN96], Ferrari et Pimentel [FP05a]) ou sur un processus de Poisson homogène (Howard et Newman [HN97]). Dans le même esprit, Alexander [Ale95] étudie le nombre de bouts topologiques pour des forêts couvrantes minimales. Des résultats similaires sont obtenus par Gangopadhyay et al. [GRS04], Athreya *et al.* [ARS08] pour des forêts dirigées sur un sous-ensemble de  $\mathbb{Z}^2$ , ou par Ferrari et al. [FFW05] sur un processus de Poisson et avec des voisinages rectangulaires particuliers. Dans chacun de ces travaux, la difficulté présentée en à la Figure 3.2.3 n'existe pas. Si l'on considère deux trajectoires ancestrales, le résultat est obtenu en considérant le processus défini par l'écart des ordonnées des points de même abscisse de ces trajectoires, et en utilisant des techniques de martingales.

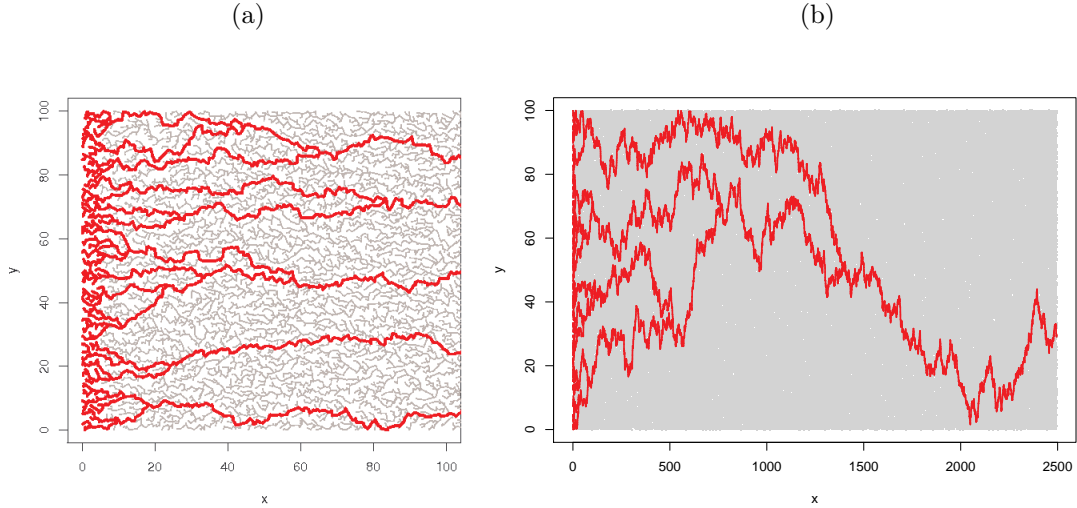


FIGURE 3.2.4 – *Simulations de la DSF. Les trajectoires ancestrales de direction  $e_x$  issues des sommets d'abscisses  $0 \leq x \leq 5$  et d'ordonnées  $0 \leq y \leq 100$  sont représentées en rouge.*

Pour montrer le Théorème 3.2.5 (i), nous utilisons une idée de percolation due à Burton et Keane [BK89], et également utilisée par [FP05a, HN97, LN96]. Nous nous appuyons pour cela sur 2 propriétés essentielles du processus de Poisson : son invariance par translations et l'indépendance des restrictions à des ensembles disjoints. Cependant, cette indépendance doit être utilisée prudemment car du fait des interactions liées à la construction de la DSF, des modifications locales peuvent avoir des conséquences au niveau macroscopique.

Dans [CT13], nous étendons le Théorème 3.2.5 à un environnement aléatoire où des points du processus de Poisson  $N$  sont effacés s'ils appartiennent à un processus Booléen.

*Idée de la preuve du Théorème 3.2.5.* Pour montrer le point (i), nous allons montrer que pour deux points  $X$  et  $Y \in N \cap \mathcal{T}$ , les trajectoires  $\gamma_X = (A^n(X))_{n \geq 0}$  et  $\gamma_Y = (A^n(Y))_{n \geq 0}$  issues

de ces points coalescent presque-sûrement. Pour cela, on raisonne par l'absurde. On introduit  $\mathcal{N} \in \{1, 2, \dots, +\infty\}$  le nombre de trajectoires infinies disjointes dans la DSF. On suppose que  $\mathbb{P}(\mathcal{N} \geq 2) > 0$ . Par ergodicité de  $N$ , la variable aléatoire  $\mathcal{N}$  est presque-sûrement constante et on peut en fait même supposer

$$\mathbb{P}(\mathcal{N} \geq 2) = 1. \quad (3.2.3)$$

Il nous faut montrer que ceci amène à une contradiction.

Pour  $m, M \in \mathbb{N}^*$  et  $z = (z_1, z_2) \in \mathbb{R}^2$ , introduisons la cellule  $C_{m,M}^z = z + [-m, m] \times [-M, M]$  ainsi que l'événement  $F_{m,M}^z$  suivant : il existe une trajectoire  $\gamma_X$  de la DSF issue de  $X \in C_{m,M}^z$  et qui ne rencontre aucune des trajectoires  $\gamma_Y$  pour  $Y \in \{x < z_1 + m\} \setminus C_{m,M}^z$  :

$$F_{m,M}^z = \{\exists X \in N \cap C_{m,M}^z, \forall Y \in N \cap \{x < z_1 + m\} \setminus C_{m,M}^z, \gamma_X \cap \gamma_Y = \emptyset\}. \quad (3.2.4)$$

Par invariance par translation, nous obtenons la minoration

$$\mathbb{E}(\eta_{L,m,M}) \geq (2L + 1)^2 \mathbb{P}(F_{m,M}^O), \quad (3.2.5)$$

où  $\eta_{L,m,M}$  est le nombre d'arêtes de la DSF sortant du rectangle  $\mathcal{R}_{L,m,M} = [-2mL, 2mL] \times [-2ML, 2ML]$ . Ceci peut se comprendre car chaque trajectoire issue d'un point  $X$  réalisant  $F_{m,M}^z$ , pour  $z = (2mk, 2M\ell)$  avec  $-L \leq k, \ell \leq L$ , est également une trajectoire qui sort du rectangle  $\mathcal{R}_{L,m,M}$ .

On peut montrer le

**Lemme 3.2.6.** *L'hypothèse (3.2.3) implique l'existence d'entiers  $m, M$  tels que  $\mathbb{P}(F_{m,M}^O) > 0$ .*

Donc, si (3.2.3) est vraie, (3.2.5) implique que  $\mathbb{E}(\eta_{L,m,M})$  augmente en  $L^2$  au moins. Or le nombre d'arêtes sortant du rectangle  $\mathcal{R}_{L,m,M}$  est heuristiquement proportionnel à la surface de ce rectangle, et donc devrait être en  $L$ . Nous n'arrivons pas à obtenir exactement cet ordre  $L$ , mais montrons que

$$\mathbb{E}(\eta_{L,m,M}) \leq CL^{3/2}, \quad (3.2.6)$$

ce qui suffit à contredire (3.2.5). Le contrôle (3.2.6) s'établit de la façon suivante. Les arêtes de longueur inférieure à  $\sqrt{L}$  qui sortent de  $\mathcal{R}_{L,m,M}$  ont l'une de leurs extrémités dans une bande de largeur  $\sqrt{L}$  qui borde  $\mathcal{R}_{L,m,M}$ , et le nombre de points de  $N$  dans cette bande est en  $L^{3/2}$ . Les arêtes de longueur supérieure à  $\sqrt{L}$  qui sortent de  $\mathcal{R}_{L,m,M}$  engendrent un demi-disque vide, de rayon  $\sqrt{L}$ . Cette contrainte implique que le nombre de ces arêtes sortant de  $\mathcal{R}_{L,m,M}$  est au plus de l'ordre de  $L$ . Ceci explique (3.2.6).

Il reste donc à expliquer d'où vient le Lemme 3.2.6. La preuve de ce résultat se divise en deux étapes.

Tout d'abord, sous l'hypothèse (3.2.3), on peut montrer qu'il existe deux entiers  $m, M$  tels qu'avec une probabilité positive, trois trajectoires infinies sont issues de  $C_{m,M}^O$ . Pour ce faire, on montre que sous (3.2.3), il existe  $M'$  tel que la probabilité que deux trajectoires infinies sortent de  $C_{m,M'}^O$  par son côté est soit positive. Alors, en utilisant l'invariance par translation verticale et quitte à prendre  $M$  suffisamment grand, on a une probabilité positive que  $C_{m,M}^O$  contienne deux boîtes  $C_{m,M'}^{(0,2M'k)}$  et  $C_{m,M'}^{(0,2M'\ell)}$ , pour  $k, \ell \in \mathbb{Z}$ , dont sortent de chacune 2 trajectoires infinies par le côté est (voir Figure 3.2.5). Comme les trajectoires ne peuvent pas se croiser sans coalescer, on peut alors montrer qu'il existe au moins trois trajectoires infinies disjointes parmi ces quatre trajectoires.

La seconde étape consiste à considérer les trois trajectoires infinies issues de  $C_{m,M}^O$ , qu'on appelle  $\gamma_X$ ,  $\gamma_Y$  et  $\gamma_Z$ . Dans le demi-plan  $\{x \geq m\}$ , la trajectoire du milieu, disons  $\gamma_X$ , est piégée entre les deux autres trajectoires. On peut alors dans un second temps modifier localement le processus de Poisson  $N$  en construisant un "bouclier" pour empêcher toute trajectoire  $\gamma_Y$  issue de

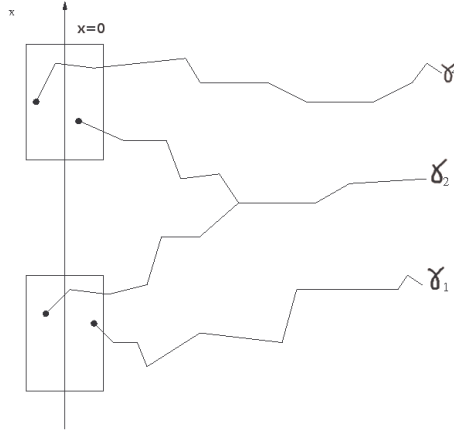


FIGURE 3.2.5 – Représentation des quatre trajectoires  $\gamma_1, \gamma_2, \gamma_3, \gamma_4$  issues des rectangles  $C_{m, M'}^{(0, 2M'k)}$  et  $C_{m, M'}^{(0, 2M'\ell)}$ . Les trajectoires  $\gamma_2$  et  $\gamma_3$  coalescent, mais il reste 3 trajectoires disjointes puisque par construction,  $\gamma_1$  et  $\gamma_2$  d'une part,  $\gamma_3$  et  $\gamma_4$  d'autre part sont disjointes. La sortie des trajectoires par le côté est des rectangles se comprend : ceci empêche par exemple  $\gamma_3$  de se faufiler par le haut et d'aller coalescer avec  $\gamma_1$  tandis que  $\gamma_2$  coalescerait avec  $\gamma_4$ .

$Y \in \{x < m\} \setminus C_{m, M}^O$  de toucher  $\gamma_X$  (voir Figure 3.2.6), ce qui prouve le Lemme 3.2.6 et le point (i) du Théorème 3.2.5. Les détails techniques de cette construction sont présentés dans [CT13]. Heuristiquement, dans une boule suffisamment grande autour de  $C_{m, M}^O$ , on remplace le processus  $N$  par un processus de Poisson  $N'$  indépendant et de même intensité. On considère l'événement où les points de  $N'$  forment une configuration qui empêche  $C_{m, M}^O$  d'être traversé par des trajectoires issues de points situés hors de  $C_{m, M}^O$  et qui ne modifie pas les trois trajectoires  $\gamma_X, \gamma_Y$  et  $\gamma_Z$ .

Considérons le point (ii). Introduisons  $\mathcal{T}_X$  le sous-arbre de la DSF issu d'un point  $X \in \mathcal{T}$  et composé des trajectoires bi-infinies passant par  $X$ . Soit  $I$  et  $J$  deux intervalles verticaux de  $\mathbb{R}^2$ , tels que l'abscisse des points de  $I$  soit inférieure à celle des points de  $J$ . Considérons le nombre de points d'intersection  $R_I$  des trajectoires bi-infinies avec un intervalle vertical  $I$  de  $\mathbb{R}^2$  :

$$R_I = \text{Card}\{(x, y) \in I; (x, y) \text{ appartient à une chemin bi-infini}\}, \quad (3.2.7)$$

et  $\tilde{R}_{I, J}$  le nombre de points d'intersection  $X$  de trajectoires bi-infinies avec  $J$  et dont le sous-arbre  $\mathcal{T}_X$  associé intersecte  $I$ .

$$\tilde{R}_{I, J} = \text{Card}\{X \in J; X \text{ appartient à la DSF et } \mathcal{T}_X \cap I \neq \emptyset\}. \quad (3.2.8)$$

$R_I$  et  $\tilde{R}_{I, J}$  sont intégrables dès que  $I$  est de mesure de Lebesgue finie.

Par stationnarité, il est possible de montrer que pour tout  $L > 0$  :

$$\mathbb{E}(R_{\{0\} \times [0, L]}) = \mathbb{E}(\tilde{R}_{\{0\} \times [0, L], \{x\} \times \mathbb{R}}). \quad (3.2.9)$$

Comme  $(\tilde{R}_{\{0\} \times [0, L], \{x\} \times \mathbb{R}})_{x > 0}$  est une suite décroissante d'entiers, par le point (i), sa limite lorsque  $x \rightarrow +\infty$  existe presque-sûrement et plus petite ou égale à 1, d'où :

$$1 \geq \mathbb{E}(R_{\{0\} \times [0, L]}) = L\mathbb{E}(R_{\{0\} \times [0, 1]}). \quad (3.2.10)$$

Comme cette inégalité est valable pour tout  $L$ , on a nécessairement  $\mathbb{E}(R_{\{0\} \times [0, 1]}) = 0$  impliquant  $\mathbb{P}(R_{\{0\} \times \mathbb{R}} = 0) = 1$ , ce qui est le résultat recherché.

Remarquons qu'un argument alternatif fondé sur la notion de point bifurquant a été proposé par Gangopadhyaya et al. [GRS04, Th. 2.2].  $\square$



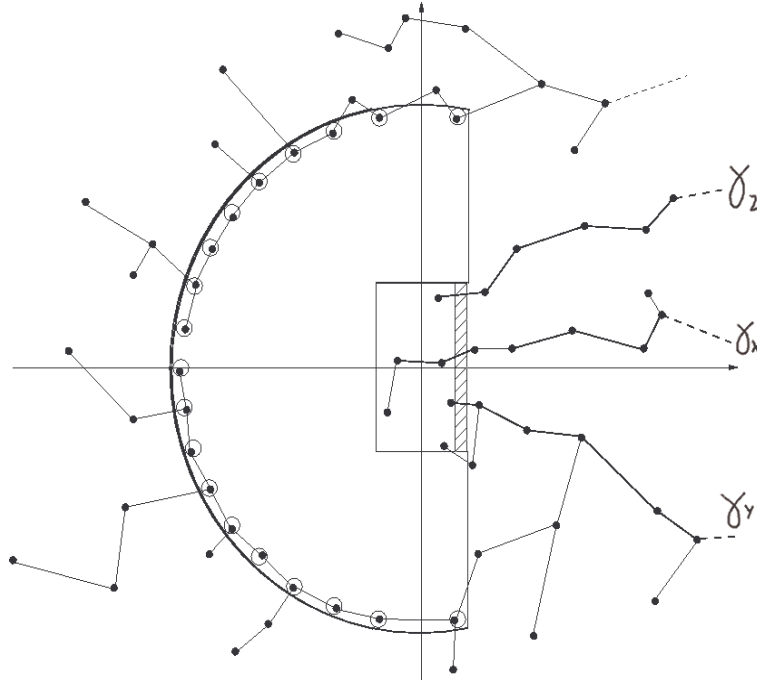


FIGURE 3.2.6 – Modification du processus ponctuel de Poisson autour du rectangle  $C_{m,M}^O$ . Sur cette configuration,  $\gamma_x$ , protégé par  $\gamma_y$  et  $\gamma_z$  ne touche aucun chemin issu d'un point de  $\{x < m\} \setminus C_{m,M}^O$ .

## 3.2.2 Arbre radial

### 3.2.2.1 Directions asymptotiques

Revenons à l'étude de l'arbre radial introduit par Baccelli et Bordenave [BB07]. Dans le problème des couleurs, on peut montrer que les interfaces entre chacun des arbres issus des descendants directs de  $O$  sont encadrées par deux branches semi-infinies du RST. Il s'agit de deux suites infinies  $(X_n)_{n \in \mathbb{N}}$  de  $N$  telles que  $X_0 = O$  et pour tout  $n > 0$ ,  $X_n = A(X_{n+1})$ , qui sont la branche semi-infinie juste "à gauche" et "à droite" de l'interface. Les propriétés des chemins semi-infinis du RST sont donc importantes. Elles se révèlent difficiles à étudier, du fait des interactions découlant de la construction de l'arbre.

Un résultat important concernant les chemins semi-infinis du RST découle du fait que cet arbre est *straight*.

**Définition 3.2.7.** Soit  $f$  une fonction positive sur  $\mathbb{R}_+$ . Pour un point  $x$ , soit  $C(x, f(x)) = \{y \in \mathbb{R}^2, |x\widehat{O}y| \leq f(x)\}$  le cône de sommet  $x$  et d'ouverture  $f(x)$ . Un arbre  $\mathcal{T}$  est *f-straight* en  $O$  si pour tous ses sommets  $X \neq O$  (sauf éventuellement un nombre fini d'entre eux)  $\mathcal{T}_X \subset C(X, f(X))$ .

Baccelli Bordenave [BB07, Theorem 5.4] ont montré que le RST était *f-straight* avec  $f(x) = |x|^{-1/2+\varepsilon}$  pour tout  $\varepsilon > 0$ . Il découle alors du résultat de Howard Newman [HN01, section 2.3] pour les arbres *straight*, le résultat suivant [BB07, Theorem 2.1] :

**Théorème 3.2.8.** *Presque sûrement, pour le RST :*

- (i) toute trajectoire semi-infinie de  $\mathcal{T}$  admet une direction asymptotique.
- (ii) pour tout  $\theta \in [0; 2\pi)$ , il existe au moins une trajectoire semi-infinie admettant  $\theta$  comme direction asymptotique.
- (iii) l'ensemble des  $\theta \in [0, 2\pi)$  qui sont les directions asymptotiques d'au moins deux trajectoires semi-infinies est dense dans  $[0, 2\pi)$ .

Avec François Baccelli et David Coupier, nous complétons ce résultat pour le RST :

**Proposition 3.2.9.** *Pour tout  $\theta \in [0, 2\pi)$ , il existe presque sûrement une unique trajectoire semi-infinie de direction asymptotique  $\theta$ .*

Pour le RST, l'ensemble des directions asymptotiques dans lesquelles il existe au moins 2 trajectoires semi-infinies est donc dense dans  $[0, 2\pi)$  (on peut même démontrer qu'il est dénombrable), mais aléatoire et la loi de ces directions asymptotiques particulières qui admettent plusieurs trajectoires semi-infinies est sans atome. Dans une direction  $\theta \in [0, 2\pi)$  fixée, comme au point (ii) du Théorème 3.2.8, nous précisons qu'il n'existe, pour le RST, qu'une unique trajectoire semi-infinie. De plus, en utilisant le Théorème 3.2.8, il est facile de montrer que les trajectoires bordant les interfaces dans le problème des couleurs ont la même direction asymptotique, qui est également celle de l'interface. Les directions partagées par plusieurs trajectoires semi-infinies sont donc les directions dans lesquelles on peut trouver les interfaces. Nous montrons qu'il peut exister avec probabilité positive 2, 3, 4 ou 5 arbres infinis enracinés en  $O$  (ou de façon équivalente 2, 3, 4 ou 5 interfaces issues de  $O$ ). La loi des angles entre les interfaces, autrement dit la loi de la largeur des arbres, n'est pas connue, mais nous présentons quelques simulations.

*Esquisse de la preuve de la Proposition 3.2.9.* Considérons l'arbre  $\mathcal{T}_X$  issu de  $X$  et constitué de tous ses descendants. Lorsque  $\mathcal{T}_X$  n'est pas borné et a une infinité de noeuds, il est possible de définir ses trajectoires semi-infinies les plus à droite et à gauche,  $\underline{\gamma}_X$  and  $\bar{\gamma}_X$ . Posons  $\bar{X}_0 = \underline{X}_0 = X$ . Soit

$$\bar{K}_0 = \text{Card}\{Y \in N, \mathcal{A}(Y) = \bar{X}_0 \text{ et } \mathcal{T}_Y \text{ est non borné}\}$$

et  $\underline{K}_0$  défini de la même façon en remplaçant  $\bar{X}_0$  par  $\underline{X}_0$  (pour l'indice 0,  $\bar{K}_0 = \underline{K}_0$ ).  $\bar{K}_0$  est le nombre de descendants directs de  $\bar{X}_0$  ayant eux-même une infinité de descendants. Comme le nombre de descendants directs d'un sommet est fini p.s. ([BB07, Section 3.3.2.]) alors  $\bar{K}_0 \geq 1$  et  $\underline{K}_0 \geq 1$  p.s. Il est possible de ranger les descendants  $\bar{X}_0^1, \dots, \bar{X}_0^{\bar{K}_0}$  de  $\bar{X}_0$ , par ordre croissant des angles orientés  $\mathcal{A}(\bar{X}_0)\bar{X}_0^k$  pour  $k \in \{1, \dots, \bar{K}_0\}$ . Soit  $\bar{X}_1$  le descendant direct de  $\bar{X}_0$  correspondant au plus grand angle. De même, on peut ranger les descendants  $\underline{X}_0^1, \dots, \underline{X}_0^{\underline{K}_0}$  de  $\underline{X}_0$  suivant les angles orientés  $\mathcal{A}(\underline{X}_0)\underline{X}_0^k$  pour  $k \in \{1, \dots, \underline{K}_0\}$  et choisir pour  $\underline{X}_1$  celui correspondant au plus petit angle. Par induction, on peut construire deux chemins semi-infinis du RST  $\bar{\gamma}_X = (\bar{X}_n)_{n \in \mathbb{N}}$  et  $\underline{\gamma}_X = (\underline{X}_n)_{n \in \mathbb{N}}$  issus de  $X$  et tels que tous les chemins semi-infinis de  $\mathcal{T}_X$  sont compris entre  $\underline{\gamma}_X$  et  $\bar{\gamma}_X$  (dans le sens trigonométrique).

La preuve de la Proposition 3.2.9 utilise alors un argument classique (cf. [HN97] pour la percolation de premier passage sur un processus de Poisson homogène sur  $\mathbb{R}^2$ , [FP05a] dans le cas de la percolation de dernier passage sur  $\mathbb{Z}^2$ ). Soit  $\theta \in [0, 2\pi)$  et soit  $U(\theta)$  l'événement :

$$U(\theta) = \{ \text{Il existe au moins deux trajectoires semi-infinies} \\ \text{différentes dans le RST de direction asymptotique } \theta \}.$$

Sur  $U(\theta)$ , soient  $\gamma_1$  et  $\gamma_2$  deux trajectoires semi-infinies de direction asymptotique  $\theta$ . Soit  $X$  un point de  $N$  appartenant à  $\gamma_1$  mais pas à  $\gamma_2$ . Ainsi,  $\gamma_1$  est une trajectoire de  $\mathcal{T}_X$  mais pas  $\gamma_2$ . Alors, l'un des chemins semi-infinis  $\underline{\gamma}_X$  ou  $\bar{\gamma}_X$  est piégé entre  $\gamma_1$  et  $\gamma_2$ , par planarité, et admet donc également  $\theta$  comme direction asymptotique. En notant  $\lambda$  la mesure de Lebesgue de  $[0, 2\pi)$ ,

$$\begin{aligned} \mathbb{E}\lambda\{\theta; U(\theta)\} &= \int_{\Omega} \int_0^{2\pi} \mathbb{1}_{U(\theta)}(\omega) d\theta d\mathbb{P}(\omega) \\ &\leq \int_{\Omega} \sum_{X \in N} \mathbb{1}_{\mathcal{T}_X \text{ non borné}} \int_0^{2\pi} \mathbb{1}_{\substack{\underline{\gamma}_X \text{ ou } \bar{\gamma}_X \text{ admet } \theta \text{ comme} \\ \text{direction asymptotique}}}(\omega) d\theta d\mathbb{P}(\omega) \end{aligned}$$

Pour  $X$  et  $\omega$  fixés, l'indicatrice portant sur  $\underline{\gamma}_X$  et  $\bar{\gamma}_X$  ne vaut 1 que pour au plus deux angles de  $[0, 2\pi)$ . Son intégrale vaut donc zéro et par le théorème de Fubini,

$$\int_0^{2\pi} \mathbb{P}(U(\theta)) d\theta = \mathbb{E}\lambda\{\theta; U(\theta)\} = 0 .$$

Ceci implique que  $\mathbb{P}(U(\theta)) = 0$  pour  $\lambda$ -presque tout  $\theta \in [0, 2\pi[$ . Par isotropie, on peut montrer que cela est vrai pour tout  $\theta \in [0, 2\pi[$ . Ceci combiné avec le point (ii) du Théorème 3.2.8 conclut la preuve.  $\square$

### 3.2.2.2 Arbres colorés du RST

Nous considérons maintenant les interfaces entre les arbres colorés issus des descendants directs de  $O$ , dont on rappelle qu'ils sont au plus 5 p.s. [BB07, Lemme 3.2]. Les couleurs sont un label  $i, j, k \dots \in \{1, \dots, 5\}$  que l'on attribue à chaque descendant direct de  $O$  et qui est hérité par tous ses descendants. On notera  $\mathcal{T}_i$  l'arbre de couleur  $i$ , vu comme le sous-ensemble de  $\mathbb{R}^2$  constitué des descendants du point initial de couleur  $i$  et des arêtes les reliant. Par planarité, et parce que les branches du RST ne s'intersectent pas, on peut montrer que les arbres  $(\mathcal{T}_i)_{i \in \{1, \dots, 5\}}$  ne s'interpénètrent pas au sens où, pour un couple ordonné de deux couleurs  $(i, j)$  telles que  $i \neq j$  et pour  $r > 0$ , il existe au plus un couple  $(\theta, \theta') \in [0; 2\pi)^2$  tel que

$$re^{i\theta} \in \mathcal{T}_i, re^{i\theta'} \in \mathcal{T}_j \quad \text{et} \quad r a(\theta, \theta') \cap \mathcal{T} = \emptyset ,$$

où  $a(\theta, \theta')$  est l'arc du cercle unité allant, dans le sens trigonométrique, de  $e^{i\theta}$  à  $e^{i\theta'}$ . Lorsque ce couple existe, les deux couleurs  $i$  et  $j$  sont adjacentes à la distance  $r$  de  $O$  et on peut définir l'angle "bissecteur" comme étant l'angle (direct)  $\theta_r(i, j) \in [0; 2\pi)$  correspondant à la ligne  $\Delta$  passant par  $O$  et coupant  $a(\theta, \theta')$  en deux parts égales. L'interface entre  $\mathcal{T}_i$  et  $\mathcal{T}_j$  est définie par la courbe

$$\varphi(i, j) = \{re^{i\theta_r(i, j)} \in \mathbb{C}, \beta(i, j) < r < \partial(i, j)\} ,$$

où  $\beta(i, j)$  et  $\partial(i, j)$  sont respectivement les infimum et supremum de l'ensemble  $\{r > 0, \theta_r(i, j) \text{ existe}\}$ .  $\beta(i, j)$  est à la distance de  $O$  à partir de laquelle  $\mathcal{T}_i$  et  $\mathcal{T}_j$  deviennent adjacents (0 ou strictement positif si  $\mathcal{T}_i$  et  $\mathcal{T}_j$  ne sont initialement pas adjacents mais qu'un arbre borné intermédiaire est mort).  $\partial(i, j)$  est le rayon où  $i$  et  $j$  cessent d'être adjacents ( $+\infty$  ou fini lorsque l'un des arbres  $\mathcal{T}_i$  et  $\mathcal{T}_j$  est borné).

Remarquons que l'application  $r \mapsto \theta_r(i, j)$  peut-être discontinue et que  $\theta_r(i, j) \neq \theta_r(j, i)$  (l'un peut exister et l'autre non, ou il peut s'agir des deux frontières entre les arbres  $\mathcal{T}_1$  et  $\mathcal{T}_2$  dans le cas où  $O$  n'a que 2 descendants directs). Nous distinguerons donc les interfaces  $\varphi(i, j)$  et  $\varphi(j, i)$ .

**Théorème 3.2.10.** *Pour chaque  $m \in \{1, 2, 3, 4, 5\}$ , la probabilité qu'il existe exactement  $m$  arbres colorés non bornés dans  $\mathcal{T}$  est strictement positive.*

*Idée de la preuve.* Nous montrerons uniquement le cas  $m = 5$ . Pour les autres cas, des techniques similaires sont utilisées et nous référons à [BCT13]. L'idée est de trouver un ensemble de configurations de  $N$  dont la probabilité de réalisation est strictement positive, et sur laquelle  $O$  a 5 descendants directs qui donnent chacun un arbre non borné.

Pour chaque  $k \in \{1, \dots, 5\}$ , le point  $(i)$  du Théorème 3.2.8 nous assure l'existence p.s. d'un chemin semi-infini  $\gamma_k$  de direction asymptotique  $2k\pi/5$ . Alors, pour  $\varepsilon > 0$  et pour chaque  $k \in \{1, \dots, 5\}$ , il existe p.s. un rayon  $r_k$  (qui dépend de  $\omega$ ) tel que  $\gamma_k$  soit inclus dans le cône

$$C_{2k\pi/5, \varepsilon, r_k} = \{\rho e^{i\theta} ; \rho > r_k \text{ et } |\theta - 2k\pi/5| < \varepsilon\} .$$

Sans perte de généralité, on peut demander à ce que  $r_k$  soit un entier et à ce que  $\gamma_k$  soit issu d'un sommet  $X_k \in N$  dont le module satisfait  $r_k < |X_k| \leq r_k + 1$ . Ceci nous permet de montrer

l'existence de rayons déterministes  $r_1, \dots, r_5 \in \mathbb{N}$  tels qu'avec probabilité positive,  $\gamma_k$  appartienne au cône  $C_{2k\pi/5, \varepsilon, r_k}$  et soit issu d'un sommet  $X_k$  tel que  $r_k < |X_k| \leq r_k + 1$ . Nous appelons  $A_\varepsilon(r_1, \dots, r_5)$  cet événement.

Définissons  $R = \max\{r_k + 1; 1 \leq k \leq 5\}$  et

$$V_\varepsilon(r_1, \dots, r_5) = B(O, R) \setminus \left[ \left( \bigcup_{k=1}^5 C_{2k\pi/5, \varepsilon, r_k} \right) \cup \{0\} \right],$$

le complémentaire des 5 cônes dans la boule  $B(O, R)$ . Il est alors possible de changer la configuration  $N$  dans  $V_\varepsilon(r_1, \dots, r_5)$  afin d'assurer que les  $X_k$  soient de couleurs différentes.

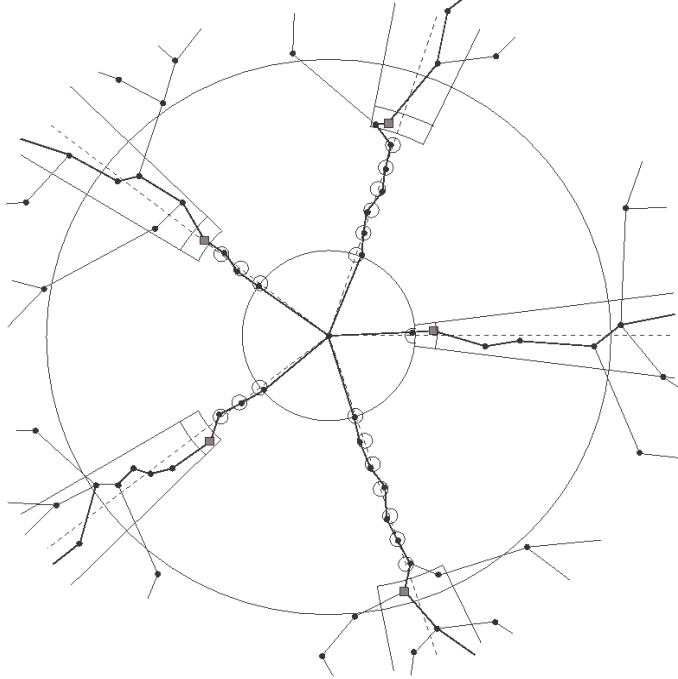


FIGURE 3.2.7 – RST construit sur une réalisation du processus de Poisson  $N$  qui satisfait  $A_\varepsilon(r_1, \dots, r_5)$  et  $B_\varepsilon(r_1, \dots, r_5)$ . Les  $X_k$  sont représentés par des losanges gris tandis que les autres points de  $N$  sont des cercles noirs.

Remarquons tout d'abord que l'occurrence de  $A_\varepsilon(r_1, \dots, r_5)$  ne dépend pas de la configuration dans la région  $V_\varepsilon(r_1, \dots, r_5)$ . On définit ensuite un événement  $B_\varepsilon(r_1, \dots, r_5)$  qui ne concerne que les points de la région  $V_\varepsilon(r_1, \dots, r_5)$  : dans cette région, les seuls points sont dans des petites boules qui relient les  $X_k$  à  $O$  de façon à s'assurer que chacun des chemins  $\gamma_1, \dots, \gamma_5$  soit d'une couleur différente (voir Figure 3.2.7). Cet événement a une probabilité positive de se réaliser et on peut montrer que  $A_\varepsilon(r_1, \dots, r_5) \cap B_\varepsilon(r_1, \dots, r_5)$  a également une probabilité positive de se réaliser, ce qui conclut la preuve.  $\square$

Soit  $\Omega(i, j)$  l'événement correspondant à l'existence d'une interface  $\varphi(i, j)$  non bornée, i.e. avec  $\partial_{i, j} = +\infty$ . D'après le Théorème 3.2.10, cet événement a une probabilité positive.

**Proposition 3.2.11.** *Sur  $\Omega(i, j)$ , la suite  $(\theta_r(i, j))_{r > \beta(i, j)}$  converge p.s. vers un angle aléatoire  $\theta(i, j) \in [0; 2\pi[$  dont la loi n'a pas d'atome par rapport à la mesure de Lebesgue.*

*Démonstration.* Soient  $X(i)$  et  $X(j)$  les descendants directs de  $O$  de couleurs  $i$  et  $j$ . Sur  $\Omega(i, j)$ , les arbres  $\mathcal{T}_{X(i)}$  et  $\mathcal{T}_{X(j)}$  sont non bornés. Par définition de  $\bar{\gamma}_{X(i)}$  et  $\underline{\gamma}_{X(j)}$ , la région délimitée par

ces trajectoires et qui contient l'interface  $\varphi(i, j)$  ne contient que des trajectoires bornées de  $\mathcal{T}$ . Alors, par les points (i) et (ii) du Théorème 3.2.8,  $\bar{\gamma}_{X(i)}$  et  $\underline{\gamma}_{X(j)}$  doivent avoir la même direction asymptotique, que l'on notera  $\theta(i, j)$ . C'est également nécessairement la direction asymptotique de  $\varphi(i, j)$ . L'absolue continuité de la loi de  $\theta(i, j)$  par rapport à la mesure de Lebesgue découle de la Proposition 3.2.9.  $\square$

Pour l'instant, nous n'avons pas identifié la loi du vecteur  $(\theta(1, 2), \dots, \theta(m-1, m), \theta(m, 1))$  conditionnellement à l'existence de  $m$  arbres colorés non bornés. Pour conserver l'échangeabilité des arbres colorés, nous associons à chacun des descendants directs de  $O$  une variable aléatoire uniforme et labellisons par 1 l'arbre issu du descendant possédant la plus petite variable. Les autres labels sont donnés en suivant le sens trigonométrique. Du fait de l'invariance par rotation autour de  $O$ , les lois marginales sont faciles à identifier.

On prendra la convention  $\theta(m, m+1) = \theta(m, 1)$  et  $\theta(m+1, m+2) = \theta(1, 2)$ .

**Proposition 3.2.12.** *Conditionnellement au fait d'avoir  $m$  arbres non bornés, pour  $i \in \{1, \dots, m\}$ ,  $\theta(i, i+1)$  suit la loi uniforme sur  $[0, 2\pi)$ .*

L'étude des lois jointes du vecteur  $(\theta(1, 2), \dots, \theta(m-1, m), \theta(m, 1))$  conditionnellement à  $m$  est équivalente à l'étude de la loi des secteurs  $(\phi(i+1) := \theta(i+1, i+2) - \theta(i, i+1), i \in \{1, \dots, m\})$  qui correspondent aux "largeurs" limites des arbres colorés.

**Lemme 3.2.13.** *Conditionnellement au fait d'avoir  $m$  arbres non bornés, les angles des secteurs entre interfaces convergent vers des variables identiquement distribuées d'espérance  $2\pi/m$ .*

**Conjecture 3.2.14.** *Conditionnellement à  $m \in \{2, 3, 4, 5\}$ , le vecteur  $(\phi(1), \dots, \phi(m))$  a une distribution proche de la loi de Dirichlet symétrique d'ordre  $m$  sur  $[0, 2\pi) : \text{Dir}(m, [0, 2\pi), \alpha)$ .*

Les distributions de Dirichlet symétriques d'ordre  $m$  et de paramètre  $\alpha > 0$  sur  $[0, 2\pi)$  sont des distributions de probabilité sur  $\mathbb{R}^m$  dont le support est inclus dans  $\Lambda = \{\eta = (\eta_1, \dots, \eta_m) \in \mathbb{R}^m, \sum_{i=1}^m \eta_i = 2\pi\}$  et dont la densité par rapport à la mesure de Lebesgue sur  $\Lambda$  est :

$$f(\eta_1, \dots, \eta_m; \alpha) = \frac{1}{\mathbf{B}(\alpha)} \prod_{i=1}^m \left(\frac{\eta_i}{2\pi}\right)^{\alpha-1}, \quad \text{où } \mathbf{B}(\alpha) = \frac{\left(\int_0^{+\infty} t^{\alpha-1} e^{-t} dt\right)^m}{\int_0^{+\infty} t^{m\alpha-1} e^{-t} dt}$$

est la fonction Beta.

Si la loi de  $(\theta(1, 2), \dots, \theta(m-1, m), \theta(m, 1))$  suivait une telle Dirichlet conditionnellement à  $m$ , les lois marginales des secteurs seraient des lois  $\mathbf{B}(\alpha, (m-1)\alpha)$  sur  $[0, 2\pi)$  d'espérance  $2\pi/m$ . Cela montrerait également que les directions asymptotiques  $(\theta(1, 2), \dots, \theta(m-1, m), \theta(m, 1))$  ne dépendraient que de  $m$  et pas du nombre de descendants directs de  $O$  qui est un événement local et qui serait "oublié" loin de l'origine.

Cependant, des simulations ont été réalisées (Fig. 3.2.8) et laissent penser que ce qui se passe localement au voisinage de l'origine impacte les lois asymptotiques des secteurs. D'où le "proche" de la conjecture. Dans les simulations, nous calculons les angles entre interfaces et calibrons les distributions Beta qui correspondraient si la conjecture était vérifiée. Les estimateurs du maximum de vraisemblance n'ont pas de forme analytique, mais on peut utiliser les estimateurs de moments suivants :

$$\hat{\alpha} = \frac{\bar{x}}{2\pi} \left( \frac{\bar{x}(2\pi - \bar{x})}{\text{Var}(x)} - 1 \right), \quad \hat{\beta} = \frac{2\pi - \bar{x}}{2\pi} \left( \frac{\bar{x}(2\pi - \bar{x})}{\text{Var}(x)} - 1 \right). \quad (3.2.11)$$

Pour les valeurs  $m = 2, 3$  et  $4$ , des tests d'adéquation de Kolmogorov-Smirnov ont été effectués pour tester  $H_0$  : la loi des secteurs est une  $B(\hat{\alpha}, (m-1)\hat{\beta})$ . L'hypothèse  $H_0$  est systématiquement rejetée avec des p-valeurs plus petites que  $2.2e - 16$  malgré les ressemblances graphiques. D'autres tests confirment le rejet de  $H_0$  (voir [BCT13]).

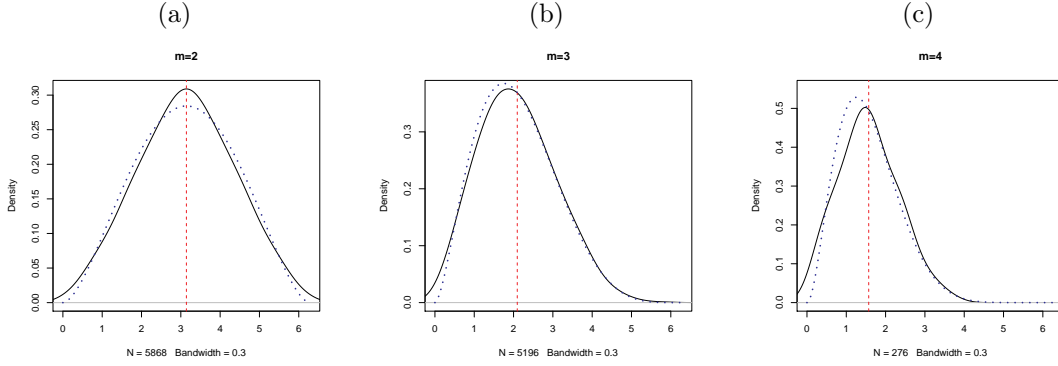


FIGURE 3.2.8 – Densité estimée pour les angles entre 2 interfaces, conditionnellement au nombre  $m$  d'arbres non bornés. Les distributions se concentrent autour de  $2\pi/m$  (ligne en rouge pointillée) et ont une variance plus petite pour les grandes valeurs de  $m$ . Densités des lois de Beta estimées (ligne bleue pointillée et épaisse). Des tests d'adéquation de Kolmogorov-Smirnov avec les lois Beta supposées par la conjecture sont effectués (statistique de test  $D$ ). (a)  $\hat{\alpha} = \hat{\beta} = 2.74$ ,  $D = 96\%$ ; (b)  $\hat{\alpha} = 2.69$ ,  $\hat{\beta} = 5.38$ ,  $D = 91.77\%$ ; (c)  $\hat{\alpha} = 2.99$ ,  $\hat{\beta} = 8.99$ ,  $D = 88.79\%$ .

### 3.2.2.3 Nombre de chemins infinis que l'on peut rencontrer au rayon $r$

Le RST admet une infinité de chemins semi-infinis, mais le nombre  $\chi_r$  de chemins-infinis qui traversent le cercle de centre  $O$  et de rayon  $r$  est fini. En utilisant les liens entre DSF et RST lorsque  $r \rightarrow +\infty$ , nous montrons que :

**Théorème 3.2.15.** *Le nombre de chemins infinis rencontrés sur la sphère  $S(O, r)$  de rayon  $r$  croît de façon sous-linéaire en espérance :*

$$\lim_{r \rightarrow \infty} \mathbb{E} \left( \frac{\chi_r}{r} \right) = 0 .$$

*Idee de la preuve.* Pour  $r > 0$ , on peut introduire les points  $A_r = re^{i/r}$  et  $B_r = re^{-i/r}$  de la sphère  $S(O, r)$ . On notera  $[A_r B_r]$  le segment d'extrémités  $A_r$  et  $B_r$  et par  $a(A_r, B_r) = \{re^{i\theta}, \theta \in [-1/r, 1/r]\}$  l'arc d'extrémités  $A_r$  et  $B_r$  qui contient le point  $(r, 0)$ . Cet arc est de longueur 2 et soit  $\tilde{\chi}_r$  le nombre de points d'intersection entre cet arc et les chemins semi-infinis du RST. L'invariance par rotation de  $N$  nous garantit que  $\mathbb{E}(\chi_r) = \pi r \mathbb{E}(\tilde{\chi}_r)$ . Il est possible de prouver que  $\mathbb{E}(\tilde{\chi}_r^2) < +\infty$ , d'où  $\mathbb{E}(\tilde{\chi}_r) \leq \mathbb{E}(\tilde{\chi}_r^2)^{1/2} \mathbb{P}(\tilde{\chi}_r \geq 1)^{1/2}$ . Pour prouver le théorème, il nous reste donc à montrer que

$$\lim_{r \rightarrow \infty} \mathbb{P}(\tilde{\chi}_r \geq 1) = 0 . \quad (3.2.12)$$

Lorsque  $r$  est grand, le RST peut-être localement approché, au voisinage de  $(r, 0)$ , par la DSF  $\mathcal{T}_{-e_x}$  de direction  $-e_x = -(1, 0)$ . La probabilité  $\mathbb{P}(\tilde{\chi}_r \geq 1)$  qu'il existe un chemin semi-infini intersectant  $a(A_r, B_r)$  est alors proche de la probabilité qu'une trajectoire de la DSF traverse  $\{0\} \times [-1, 1]$  et soit très longue dans la direction  $e_x$ .

Plus précisément, soit  $R > 0$  grand. Posons  $\tilde{r} = r \cos(1/r)$ . Considérons une trajectoire du RST intersectant le segment  $\{\tilde{r}\} \times [-1, 1]$  et  $S((\tilde{r}, 0), R)$ . On montre que cette trajectoire est incluse dans la boule  $B((\tilde{r}, 0), 2R)$  avec grande probabilité (cf. Figure 3.2.9). Pour  $r$  grand, par le Lemme 3.2.4, le RST peut-être approché par la DSF. Alors, pour  $\varepsilon > 0$  et  $r$  et  $R$  suffisamment grands,

$$\begin{aligned} \mathbb{P}(\tilde{\chi}_r \geq 1) &\leq \mathbb{P} \left( \begin{array}{l} \text{il existe une trajectoire du RST incluse dans} \\ B((\tilde{r}, 0), 2R) \text{ et intersectant } \{\tilde{r}\} \times [-1, 1] \text{ et } S((\tilde{r}, 0), R), \end{array} \right) + \varepsilon \\ &\leq \mathbb{P} \left( \begin{array}{l} \text{il existe une trajectoire de la DSF} \\ \text{intersectant } \{0\} \times [-1, 1] \text{ et } S(O, R) \end{array} \right) + 2\varepsilon . \end{aligned} \quad (3.2.13)$$

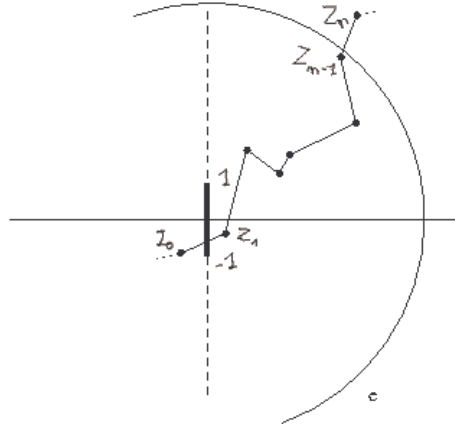


FIGURE 3.2.9 – Exemple de trajectoire  $Z_0, \dots, Z_n$  du RST intersectant le segment  $\{\tilde{r}\} \times [-1, 1]$  (en gras) et la sphère  $S((\tilde{r}, 0), R)$ .

Le fait que la DSF n'ait pas de chemin bi-infini [CT13] implique que la probabilité du membre de droite de (3.2.13) tend vers 0.  $\square$

### 3.3 Conclusion

Les théorèmes limites liés aux renormalisations du RST ou de la DSF et la convergence vers des objets géométriques continus est une suite naturelle de l'étude de ces arbres, et qui pourrait faire le lien avec les problématiques du Chapitre 1. Pour des forêts radiales dirigées plus simples que la DSF, il a été montré des convergences vers la toile Brownienne (*Brownian web*), cf. Fontes et al. [FINR04], Newman et al. [NRS05], Ferrari et al. [FFW05], Coletti et al. [CDF09, CV11].

**Toile Brownienne :** Heuristiquement, la toile Brownienne  $\mathcal{W}$  est la collection de mouvements Browniens standards réels issus de tous les points  $(x, t) \in \mathbb{R}^2$  et qui coalescent lorsqu'ils se rencontrent (voir Figure 3.2.4). Cet objet a été introduit par Arratia [Arr79], Tóth et Werner [TW98] puis Fontes et al. [FINR04] dont nous donnons la définition de la toile Brownienne. On considère l'espace  $\Pi = \bigcup_{t_0 \in \mathbb{R}} \mathcal{C}[t_0] \times \{t_0\}$  où  $\mathcal{C}[t_0]$  est l'ensemble des trajectoires continues de  $[t_0, +\infty)$  dans  $\mathbb{R}$ . Pour  $(y_1, t_1), (y_2, t_2) \in \Pi$ , on définit la distance

$$d_{\Pi}((y_1, t_1), (y_2, t_2)) = |\tanh(t_1) - \tanh(t_2)| \vee \left( \sup_{t \in \mathbb{R}} \left| \frac{\tanh(y_1(t \vee t_1))}{1 + |t \vee t_1|} - \frac{\tanh(y_2(t \vee t_2))}{1 + |t \vee t_2|} \right| \right)$$

qui en fait un espace Polonais.

L'espace  $\mathcal{H}$  des compacts de  $(\Pi, d_{\Pi})$  est muni de la distance de Hausdorff, qui en fait un espace Polonais. Pour tous  $K_1, K_2 \in \mathcal{H}$ ,

$$d_{\mathcal{H}}(K_1, K_2) = \sup_{(y_1, t_1) \in K_1} \inf_{(y_2, t_2) \in K_2} d_{\Pi}((y_1, t_1), (y_2, t_2)) \vee \sup_{(y_2, t_2) \in K_2} \inf_{(y_1, t_1) \in K_1} d_{\Pi}((y_1, t_1), (y_2, t_2)).$$

Alors la toile Brownienne  $\mathcal{W}$  est une variable aléatoire à valeurs dans  $(\mathcal{H}, d_{\mathcal{H}})$  caractérisée par ses distributions marginales :

**Théorème-Définition 3.3.1.** *Il existe une variable aléatoire  $\mathcal{W}$  de  $(\mathcal{H}, d_{\mathcal{H}})$  dont la loi  $\mathbb{W}$  est caractérisée de façon unique par les trois propriétés suivantes :*

- (i) *Pour tout  $(x, t) \in \mathbb{R}^2$ , il existe p.s. une unique trajectoire  $W_{x,t}$  de  $\mathcal{W}$  issue de  $(x, t)$ .*

- (ii) Pour tout  $n$ -uplet  $(x_1, t_1), \dots, (x_n, t_n)$ , avec  $n \in \mathbb{N}^*$ , la loi de  $(W_{x_1, t_1}, \dots, W_{x_n, t_n})$  est celle de  $n$  mouvements Browniens standards coalescents, issus de  $(x_1(t_1), t_1), \dots, (x_n(t_n), t_n)$ .
- (iii)  $\mathcal{W}$  est p.s. la fermeture dans  $(\mathcal{H}, d_{\mathcal{H}})$  de  $\{W_{x,t}, (x,t) \in \mathcal{D}\}$  pour tout ensemble  $\mathcal{D}$  dénombrable dense déterministe de  $\mathbb{R}^2$ .

**Espace de Schramm-Smirnov et autre définition de la toile Brownienne :** Récemment, Berestycki Garban et Sen [BGS14] ont donné une autre définition de la toile Brownienne, vue comme une variable aléatoire à valeurs dans l'espace de Schramm-Smirnov. Soit un homéomorphisme  $\varphi$  de  $[0, 1]^2$  dans  $\mathbb{R}^2$  tel qu'il existe  $t_0 < t_1$  tel que  $\varphi([0, 1] \times \{0\}) \subset \mathbb{R} \times \{t_0\}$  et  $\varphi([0, 1] \times \{1\}) \subset \mathbb{R} \times \{t_1\}$ . On peut alors définir le tube  $T$  de  $\mathbb{R}^2$  associé à  $\varphi$  comme le triplet  $([T], \partial_0 T, \partial_1 T) := (\varphi([0, 1]^2), \varphi([0, 1] \times \{0\}), \varphi([0, 1] \times \{1\}))$ . L'espace des tubes  $\mathfrak{T}$  peut être muni de la distance

$$d_{\mathfrak{T}}(T_1, T_2) = d_{\text{Haus}}([T_1], [T_2]) + d_{\text{Haus}}(\partial_0 T_1, \partial_0 T_2) + d_{\text{Haus}}(\partial_1 T_1, \partial_1 T_2)$$

où  $d_{\text{Haus}}$  est la distance de Hausdorff dans  $\mathbb{R}^2$ .

Un tube  $T$  est traversé par une trajectoire  $(y, t) \in \Pi$  si  $t \leq t_0$ ,  $(y(t_0), t_0) \in \partial_0 T$ ,  $(y(t_1), t_1) \in \partial_1 T$  et  $\forall s \in (t_0, t_1)$ ,  $(y(s), s) \in [T]$ . On peut alors définir une relation d'ordre :  $T_1 < T_2$  s'il existe des voisinages ouverts  $U_1$  et  $U_2$  de  $T_1$  et  $T_2$  dans  $(\mathfrak{T}, d_{\mathfrak{T}})$  tels que pour tous  $T'_1 \in U_1$  et  $T'_2 \in U_2$ ,  $T'_1$  est traversé par n'importe quelle trajectoire qui traverse  $T'_2$ . Un sous-ensemble  $\mathfrak{S}$  de  $\mathfrak{T}$  est héréditaire si pour tout  $T \in \mathfrak{S}$  et  $T' \in \mathfrak{T}$ ,  $T' < T \Rightarrow T' \in \mathfrak{S}$ .

L'espace  $\mathfrak{H}$  des sous-ensembles héréditaires fermés de  $\mathfrak{T}$  est l'espace de Schramm-Smirnov. Muni de la topologie engendrée par la base constituée des  $\{H \in \mathfrak{H}, H \cap U \neq \emptyset\}$  pour  $U$  ouvert de  $(\mathfrak{T}, d_{\mathfrak{T}})$  et des  $\{H \in \mathfrak{H}, T \notin H\}$  pour  $T \in \mathfrak{T}$ ,  $\mathfrak{H}$  est complet, séparable et même compact.

On peut associer à toute famille de trajectoires le sous-ensemble héréditaire des tubes qui sont traversés par au moins l'une d'entre elles. En particulier, si on considère  $n$  trajectoires Browniennes coalescentes et que l'on fait tendre  $n \rightarrow +\infty$ , on peut montrer que la famille des tubes associée converge en loi vers ce que Berestycki, Garban et Sen appellent la toile Brownienne.

**Convergence du DSF vers la toile Brownienne :** Les deux caractérisations précédentes de la toile Brownienne, par [FINR04] et [BGS14], s'accompagnent de critères permettant de montrer des théorèmes limites. Les forêts considérées dans les articles ci-dessus peuvent être associées à des martingales qui permettent de vérifier ces critères, ce qui n'est pas notre cas à cause de la non-markoviannité expliquée en Fig. 3.2.3. En particulier, les deux critères nécessitent de montrer que pour deux trajectoires de la DSF issues de  $X$  et  $Y \in N$ ,  $\mathbb{P}(\tau_{X,Y} > t) \leq C/\sqrt{t}$ , où  $\tau_{X,Y}$  est le temps de coalescence de ces trajectoires. Ceci est un travail en cours avec Coupier, Saha et Sarkar.

Remarquons que pour une version discrète sur  $\mathbb{Z}^2$  de la DSF, Saha Sarkar et Roy [RSS14] ont réussi à prouver une convergence vers la toile Brownienne. Les auteurs jouent sur le fait que la distance de graphe sur  $\mathbb{Z}^2$  produit des boules qui sont triangulaires, et permettent un bon contrôle des régions vides.

Enfin, les limites continues du RST sont également un problème ouvert. Il y a peut-être des relations à creuser avec les limites établies pour les modèles DLA (*Diffusion Limited Aggregation*) étudiés par Norris Turner et leurs co-auteurs [VST12, VST13, NT12, NT08, NT] ou les limites des arbres discrets de Coletti et Valencia [CV13].





# Bibliographie

- [AAI11] K. Athreya, S. Athreya, and S. Iyer. Super-critical age dependent branching Markov processes and their scaling limits. *Bernoulli*, 17(1) :138–154, 2011.
- [AB00] H. Anderson and T. Britton. *Stochastic Epidemic models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer, New York, 2000.
- [ABT14] D. Abu Awad, S. Billiard, and V.C. Tran. The effect of the timing of selection on the mutation load, inbreeding depression and population size. page submitted, 2014.
- [AH83] S. Asmussen and H. Hering. *Branching processes*, volume 3 of *Progress in Probability and Statistics*. Birkhäuser, Boston, 1983.
- [AJL<sup>+</sup>07] H. De Arazoza, J. Joanes, R. Lounes, C. Legeai, S. Cléménçon, J. Perez, and B. Auvert. The HIV/AIDS epidemic in Cuba : description and tentative explanation of its low prevalence. *BMC Infectious Disease*, 7 :130, 2007.
- [AK98] K.B. Athreya and H.J. Kang. Some limit theorems for positive recurrent branching Markov chains I, II. *Advances in Applied Probability*, 30(3) :693–710, 1998.
- [AL96] H. De Arazoza and R. Lounes. A two-sex model for the AIDS-epidemic. application to the Cuban national program for HIV-AIDS. *Investigacion Operacional*, 17(3-4) :95–100, 1996.
- [Ale95] K.S. Alexander. Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.*, 23(1) :87–104, 1995.
- [ALHI00] H. De Arazoza, R. Lounes, T. Hoang, and Y. Interian. Modeling HIV epidemic under contact-tracing - the Cuban case. *Journal of Theoretical Medicine*, 2(4) :267–274, 2000.
- [AM91] R.M. Anderson and R.M. May. *Infectious Diseases of Humans : Dynamics and Control*. Oxford University Press, 1991.
- [AN70] K.B. Athreya and P.E. Ney. *Branching Processes*. Springer, 1970.
- [And98] H. Andersson. Limit theorems for a random graph epidemic model. *Annals of Applied Probability*, 8(4) :1331–1349, 1998.
- [And99] H. Andersson. Epidemic models and social networks. *Mathematical Scientist*, 24(2) :128–147, 1999.
- [Arr79] R. Arratia. *Coalescing Brownian motions on the line*. PhD thesis, Univ. Wisconsin, Madison, 1979.
- [ARS08] S. Athreya, R. Roy, and A. Sarkar. Random directed trees and forest - drainage networks with dependence. *Electronic journal of Probability*, 13 :2160–2189, 2008. Paper no.71.
- [AT98] R. Adler and R. Tribe. Uniqueness for a historical SDE with a singular interaction. *Journal of Theoretical Probability*, 11(2) :515–533, 1998.
- [Bai50] N.T.J. Bailey. A simple stochastic epidemic. *Biometrika*, 37 :193–202, 1950.
- [Ban08] V. Bansaye. Proliferating parasites in dividing cells : Kimmel’s branching model revisited. *Annals of Applied Probability*, 2008.
- [Ban13] V. Bansaye. Ancestral lineages and limit theorems for branching Markov chains. 2013. submitted.
- [Bar53] M.S. Bartlett. Stochastic processes or the statistics of change. *J. Roy. Statist. Sec., Ser. C*, 2(1) :44–64, 1953.
- [Bar75a] A.D. Barbour. The duration of the closed stochastic epidemic. *Biometrika*, 62 :477–482, 1975.
- [Bar75b] A.D. Barbour. A note on the maximum size of a closed epidemic. *J. Roy. Stat. Soc. B*, 37 :459–460, 1975.
- [Bar00] N.H. Barton. Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B*, 355 :1553–1562, 2000.
- [BB07] F. Baccelli and C. Bordenave. The radial spanning tree of a Poisson point process. *Annals of Applied Probability*, 17(1) :305–359, 2007.
- [BBPSV05] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*, 235 :275–288, 2005.

- [BCMR09] M.A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C.P. Robert. Adaptive approximate Bayesian computation. *Biometrika*, 96(4) :983–990, 2009.
- [BCT13] F. Baccelli, D. Coupier, and V.C. Tran. Semi-infinite paths of the 2d-radial spanning tree. *Advances in Applied Probability*, 45(4) :895–1201, 2013.
- [BD86] M.D. Brennan and R. Durrett. Splitting intervals. *Annals of Probability*, 14 :1024–1036, 1986.
- [BD87] M.D. Brennan and R. Durrett. Splitting intervals ii : Limit laws for lengths. *Probability Theory and Related Fields*, 75 :109–127, 1987.
- [BDLL08] T. Britton, M. Deijfen, A.N. Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45 :743–756, 2008.
- [BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle, and V.C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *Annals of Applied Probability*, 21(6) :2263–2314, 2011.
- [BE04] N.H. Barton and A.M. Etheridge. The effect of selection on genealogies. *Genetics*, 166 :1115–1131, 2004.
- [Ber09] N. Berestycki. *Recent progress in coalescent theory*, volume 16. Ensaos Matemáticos, Rio de Janeiro, sociedade brasileira de matemática edition, 2009. 193pp.
- [BEV10] N.H. Barton, A.M. Etheridge, and A. Véber. A new model for evolution in a spatial continuum. *Electronic Journal of Probability*, 15(7) :162–216, 2010.
- [BEV13] N. Berestycki, A. Etheridge, and A. Véber. Large scale behaviour of the spatial Lambda-Fleming-Viot process. *Annales de l’Institut Henri Poincaré, Probability and Statistics*, 49 :374–401, 2013.
- [BF10] M.G.B. Blum and O. François. Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, 20 :63–73, 2010.
- [BFMT14] S. Billiard, R. Ferrière, S. Méléard, and V.C. Tran. Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks. *Journal of Mathematical Biology*, 2014. in revision.
- [BGS14] N. Berestycki, C. Garan, and A. Sen. Coalescing Brownian flows : a new approach. *Annals of Probability*, page to appear, 2014.
- [Bil68] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.
- [Bil95] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3 edition, 1995.
- [Bil99] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, second edition edition, 1999.
- [BK89] R.M. Burton and M.S. Keane. Density and uniqueness in percolation. *Communication in Mathematical Physics*, 121 :501–505, 1989.
- [BK95] A. Bose and I. Kaj. Diffusion approximation for an age-structured population. *Annals of Applied Probabilities*, 5(1) :140–157, 1995.
- [BK00] A. Bose and I. Kaj. A scaling limit process for the age-reproduction structure in a Markov population. *Markov Processes and Related Fields*, 6(3) :397–428, 2000.
- [BKPR06] K. Ball, T.G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Annals of Applied Probability*, 16(4) :1925–1961, 2006.
- [Blu10] M.G. Blum. Approximate Bayesian Computation : a non-parametric perspective. *Journal of the American Statistical Association*, 105 :1178–1187, 2010.
- [BM98] A. J. Baddeley and I. S. Molchanov. Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision*, 8 :79–92, 1998.
- [BM11] N. Bonichon and J.-F. Marckert. Asymptotic of geometrical navigation on a random set of points of the plane. *Advances in Applied Probability*, 43(4) :889–942, 2011.
- [BMCR09] M.A. Beaumont, J.-M. Marin, J.-M. Cornuet, and C.P. Robert. Adaptive approximate bayesian computation. *Biometrika*, 96(4) :983–990, 2009.
- [BN08] F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Mathematical Biosciences*, 212 :69–87, 2008.
- [Bol01] B. Bollobás. *Random graphs*. Cambridge University Press, 2 edition, 2001.
- [BP94] I. Benjamini and Y. Peres. Markov chains indexed by trees. *Annals of Probability*, 22(1) :219–243, 1994.
- [BR13] A.D. Barbour and G. Reinert. Approximating the epidemic curve. *Electronic Journal of Probability*, 18(54) :2557, 2013.
- [BSGP09] B. Bercu, B. De Saporta, and A. Gégout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability*, 14(87) :2492–2526, 2009.
- [BST14] F. Ball, D. Sirl, and P. Trapman. Epidemics on random intersection graphs. *Annals of Applied Probability*, 24(3) :1081–1128, 2014.
- [BT10] M. Blum and V.C. Tran. HIV with contact-tracing : a case study in Approximate Bayesian Computation. *Biostatistics*, 11(4) :644–660, 2010.

- [BT11] V. Bansaye and V.C. Tran. Branching Feller diffusion for cell division with parasite infection. *ALEA*, 8 :95–127, 2011.
- [BT12] S. Billiard and V.C. Tran. A general stochastic model for sporophytic self-incompatibility. *Journal of Mathematical Biology*, 64(1-2) :163–210, 2012.
- [BZB02] M.A. Beaumont, W. Zhang, and D.J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162 :2025–2035, 2002.
- [CART10] S. Cléménçon, H. De Arazoza, F. Rossi, and V.C. Tran. Exploratory analysis of the cuban aids contact-tracing database. 2010. in progress.
- [CCV<sup>+</sup>04] S. Cauchemez, F. Carrat, C. Viboud, A. J. Valleron, and P. Y. Boelle. A Bayesian MCMC approach to study transmission of influenza : application to household longitudinal data. *Statistics in Medicine*, 23 :3469–3487, 2004.
- [CDF09] C.F. Coletti, E.S. Dias, and L.R.G. Fontes. Scaling limit for a drainage network model. *Journal of Applied Probability*, 46(4) :1184–1197, 2009.
- [CF97] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25 :2300–2312, 1997.
- [CF04] S. Cauchemez and N.M. Ferguson. Likelihood-based estimation of continuous-time epidemic models from time-series data : application to measles transmission in london. *Journal of the Royal Society Interface*, 5 :885–897, 2004.
- [CFM06a] N. Champagnat, R. Ferrière, and S. Méléard. Individual-based probabilistic models of adaptive evolution and various scaling approximations. In *Proceedings of the 5th seminar on Stochastic Analysis, Random Fields and Applications*, Probability in Progress Series, Ascona, Suisse, 2006. Birkhauser.
- [CFM06b] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics : from individual stochastic processes to macroscopic models via timescale separation. *Theoretical Population Biology*, 69 :297–321, 2006.
- [CFM08] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive dynamics. *Stochastic Models*, 24 :2–44, 2008.
- [CGMRC06] A. Cuevas, W. Gonzalez-Manteiga, and A. Rodriguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48 :7–19, 2006.
- [CH08] L.-M. Chevin and F. Hospital. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180 :1645–1660, 2008.
- [CH11] D. Coupier and P. Heinrich. Stochastic domination for the last passage percolation tree. *Markov Processes and Related Fields*, 17(1) :37–48, 2011.
- [CH12] D. Coupier and P. Heinrich. Coexistence probability in the last passage percolation model is  $6-8\log 2$ . *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4) :973–988, 2012.
- [Cha94] B. Charlesworth. *Evolution in Age structured Population*. Cambridge University Press, 2 edition, 1994.
- [Cha03] N. Champagnat. Convergence and existence for polymorphic adaptive dynamics jump and degenerate diffusion models. Preprint Laboratoire MODAL'X 03/7, 03 2003.
- [Cha06] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic Processes and their Applications*, 116 :1127–1160, 2006.
- [Clo11] B. Cloez. Limit theorems for some branching measure-valued processes. 2011. arXiv :1106.0660.
- [CM11] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, 151(1-2) :45–94, 2011.
- [CMPR13] C. Coron, S. Méléard, E. Porcher, and A. Robert. Quantifying the mutational meltdown in diploid populations. *The American Naturalist*, 181(5) :623–636, 2013.
- [Cor14] C. Coron. Stochastic modeling of density-dependent diploid populations and extinction vortex. *Advances in Applied Probability*, 2014. to appear.
- [Cou11] D. Coupier. Multiple geodesics with the same direction. *Electronic Communications in Probability*, 16(46) :517–527, 2011.
- [CRW91] B. Chauvin, A. Rouault, and A. Wakolbinger. Growing conditioned trees. *Stochastic Processes and their Applications*, 39 :117–130, 1991.
- [CT13] D. Coupier and V.C. Tran. The 2d-directed spanning forest is almost surely a tree. *Random Structures and Algorithms*, 42(1) :59–72, 2013.
- [CTA08] S. Cléménçon, V.C. Tran, and H. De Arazoza. A stochastic SIR model with contact-tracing : large population limits and statistical inference. *Journal of Biological Dynamics*, 2(4) :391–414, 2008.
- [CV11] C.F. Coletti and G. Valle. Convergence to the Brownian web for a generalization of the drainage network model. 1, 2011. arXiv:1109.3517.
- [CV13] C.F. Coletti and L.A. Valencia. The radial Brownian web. 1, 2013. arXiv:1310.6929.

- [Daw78] D.A. Dawson. Geostochastic calculus. *Canadian Journal of Statistics*, 6(2) :143–168, 1978.
- [Daw93] D. A. Dawson. Measure-valued Markov processes. In Springer, editor, *Ecole d'Eté de probabilités de Saint-Flour XXI*, volume 1541 of *Lectures Notes in Math.*, pages 1–260, New York, 1993.
- [DDMT12] L. Decreusefond, J.-S. Dhersin, P. Moyal, and V.C. Tran. Large graph limit for a sir process in random network with heterogeneous connectivity. *Annals of Applied Probability*, 22(2) :541–575, 2012.
- [DGL02] D.A. Dawson, L.G. Gorostiza, and Z. Li. Nonlocal branching superprocesses and some related models. *Acta Applicandae Mathematicae*, 74 :93–112, 2002.
- [DGP11] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Tree-valued Fleming-Viot dynamics with mutation and selection. *Annals of Applied Probability*, 2011. in press.
- [dH14] R. Van der Hofstad. *Random Graphs and Complex Networks*. 2014. Lecture notes in prep.
- [DHM90] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio  $r_0$  in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28 :365–382, 1990.
- [DJM98] O. Diekmann, M.C.M. De Jong, and J.A.J. Metz. A deterministic epidemic model taking account of repeated contacts between the same individuals. *Journal of Applied Probability*, 35(2) :448–462, 1998.
- [DK99a] P. Donnelly and T.G. Kurtz. Genealogical processes for fleming-viot models with selection and recombination. *Annals of Applied Probability*, 9(4) :1091–1148, 1999.
- [DK99b] P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Annals of Probability*, 27(1) :166–205, 1999.
- [DM08] J.-F. Delmas and L. Marsalle. Detection of cellular aging in a Galton-Watson process. *Preprint*, 2008. <http://arxiv.org/abs/0807.0749>.
- [DMM08] M. Durinx, J.A.J. Metz, and G. Meszéna. Adaptive dynamics for physiologically structured models. *Journal of Mathematical Biology*, 56(5) :673–742, 2008.
- [DP91] D.A. Dawson and E.A. Perkins. *Historical Processes*, volume 93. American Mathematical Society, Memoirs of the American Mathematical Society edition, 1991.
- [DS05] R. Durrett and J. Schweinsberg. Random partitions approximating the coalescence of lineages during a selective sweep. *Annals of Applied Probability*, 15(3) :1591–1651, 2005.
- [Dur07] R. Durrett. *Random graph dynamics*. Cambridge University Press, New York, 2007.
- [Dyn91] E.B. Dynkin. Branching particle systems and superprocesses. *Annals of Probability*, 19 :1157–1194, 1991.
- [EK86] S.N. Ethier and T.G. Kurtz. *Markov Processus, Characterization and Convergence*. John Wiley & Sons, New York, 1986.
- [EK02] K.T.D. Eames and M.J. Keeling. Modelling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20) :13330–13335, 2002.
- [EP94] S.N. Evans and E.A. Perkins. Measure-valued branching diffusions with singular interactions. *Canadian Journal of Mathematics*, 46 :120–168, 1994.
- [EPW06] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. An approximate sampling formula under genetic hitchhiking. *Annals of Applied Probability*, 16(2) :685–729, 2006.
- [ES07] S.N. Evans and D. Steinsaltz. Damage segregation at fissioning may increase growth rates : A super-process model. *Theoretical Population Biology*, 71 :473–490, 2007.
- [ET02] J. Engländer and D. Turaev. A scaling limit theorem for a class of superdiffusions. *Annals of Probability*, 30(2) :683–722, 2002.
- [Eth00] A. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, 2000.
- [EV12] A. Etheridge and A. Véber. The spatial Lambda-Fleming-Viot process on a large torus : genealogies in the presence of recombination. *Annals of Applied Probability*, 22 :2165–2209, 2012.
- [EVY14] A. Etheridge, A. Véber, and F. Yu. Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. 2014. [arXiv:1406.5884](https://arxiv.org/abs/1406.5884).
- [EW06] J. Engländer and A. Winter. Law of large numbers for a class of superdiffusions. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(2) :171–185, 2006.
- [Fal97] K. Falconer. *Fractal Geometry : Mathematical foundations and applications*. Wiley & Sons, 1997.
- [Fan92] J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420) :998–1004, 1992.
- [FB07] S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1) :36–41, 2007.
- [FFW05] P.A. Ferrari, L.R.G. Fontes, and X.-Y. Wu. Two-dimensional Poisson trees converge to the Brownian web. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 41(5) :851–858, 2005.

- [FIM99] G. Fayolle, R. Iasnogorodski, and V. Malyshev. *Random walks in the quarter plane*. Springer-Verlag, Berlin, 1999.
- [FINR04] L.R.G. Fontes, M. Isopi, C. Newman, and K. Ravishankar. The Brownian web : characterization and convergence. *Annals of Probability*, 32(4) :2857–2883, 2004.
- [FLT04] P.A. Ferrari, C. Landim, and H. Thorisson. Poisson trees, succession lines and coalescing random walks. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 40(2) :141–152, 2004.
- [FM04] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4) :1880–1919, 2004.
- [Foe59] H. Von Foerster. Some remarks on changing populations. In Grune & Stratton, editor, *The Kinetics of Cellular Proliferation*, pages 382–407, New York 1959.
- [FP05a] P. A. Ferrari and L. P. R. Pimentel. Competition interfaces and second class particles. *Ann. Probab.*, 33(4) :1235–1254, 2005.
- [FP05b] P.A. Ferrari and L.P.R. Pimentel. Competition interfaces and second class particles. *Annals of Probability*, 33(4) :1235–1254, 2005.
- [FT09] R. Ferrière and V.C. Tran. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM : Proceedings*, 27 :289–310, 2009.
- [FVW02] K. Fleischmann, V.A. Vatutin, and A. Wakolbinger. Branching systems with long-living particles at the critical dimension. *Theoretical Probability and its Applications*, 47(3) :429–454, 2002.
- [Gil76] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4) :403–434, 1976.
- [Gil77] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25) :2340–2361, 1977.
- [GM74] M.E. Gurtin and R.C. MacCamy. Nonlinear age-dependent population dynamics. *Arch. Rat. Mech. Anal.*, 54 :281–300, 1974.
- [GMT14] A. Gupta, J.A.J. Metz, and V.C. Tran. A new proof for the convergence of an individual based model to the trait substitution sequence. *Acta Applicandae Mathematicae*, 131(1) :1–27, 2014.
- [GPW11] A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics. martingale problems and applications. *Probability Theory and Related Fields*, 2011. in press.
- [GRS04] S. Gangopadhyay, R. Roy, and A. Sarkar. Random oriented trees : a model of drainage networks. *Ann. App. Probab.*, 14(3) :1242–1266, 2004.
- [Guy07] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Annals of Applied Probability*, 17 :1538–1569, 2007.
- [HALJ05] Y.-H. Hsieh, H. De Arazoza, R. Lounes, and J. Joanes. A class of methods for HIV contact tracing in Cuba : implications for intervention and treatment. In *Deterministic and Stochastic Models for AIDS and HIV with Interventions*, Singapore : World Scientific, 2005.
- [Har63] T.E. Harris. *The Theory of Branching Processes*. Springer, Berlin, 1963.
- [HdALC02] Y.H. Hsieh, H. de Arazoza, S.M. Lee, and C.W. Chen. Estimating the number of Cubans infected sexually by human immunodeficiency virus using contact tracing data. *Int. J. Epidemiol.*, 31(3) :679–83, 2002.
- [HH06] R. Hardy and S.C. Harris. A new formulation of the spine approach to branching diffusions. 2006. preprint <http://arxiv.org/abs/math.PR/0611054>.
- [HH09] R. Hardy and S.C. Harris. A spine approach to branching diffusions with applications to  $L^p$ -convergence of martingales. In Springer, editor, *Séminaire de Probabilités*, volume XLII of *Lectures Notes in Math.*, pages 281–330, Berlin, 2009.
- [HKPT87] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation, and Statistical Applications*, volume 129 of *Lecture Notes in Statistics*. Springer, New York, 1987.
- [HN97] C.D. Howard and C.M. Newman. Euclidean models of first-passage percolation. *Probab. Theory Related Fields*, 108(2) :153–170, 1997.
- [HN01] C.D. Howard and C.M. Newman. Geodesics and spanning trees for Euclidean first-passage percolation. *Annals of Probability*, 29 :577–623, 2001.
- [Hou12] T. House. Modelling epidemics on networks. *Contemporary Physics*, 53(3) :213–225, 2012.
- [HP90] L.A. Klein Haneveld and A.O. Pittenger. Escape time for a random walk from an orthant. *Stochastic Processes and their Applications*, 35 :1–9, 1990.
- [HR12] S. Harris and M. Roberts. Branching Brownian motion : almost sure growth along scaled paths. *Séminaire de Probabilités XLIV*, pages 375–399, 2012.
- [HWA<sup>+</sup>05] Y.-H. Hsieh, H.-C. Wang, H. De Arazoza, R. Lounes, S.-J. Twu, , and H.-M. Hsu. Ascertaining HIV underreporting in low prevalence countries using the approximate ratio of underreporting. *Journal of Biological Systems*, 13(4) :441–454, 2005.

- [IW89] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*, volume 24. North-Holland Publishing Company, 1989. Second Edition.
- [Jag75] P. Jagers. *Branching Processes with Biological Applications*. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, London-New York-Sydney, 1975.
- [Jak86] A. Jakubowski. On the Skorokhod topology. *Annales de l'Institut Henri Poincaré*, 22(3) :263–285, 1986.
- [JK00] P. Jagers and F. Klebaner. Population-size-dependent and age-dependent branching processes. *Stochastic Processes and their Applications*, 87 :235–254, 2000.
- [JK11] P. Jagers and F. Klebaner. Population size dependent, age structured branching processes linger around their carrying capacity. *Journal of Applied Probability*, 48 :249–260, 2011.
- [JLW14] S. Janson, M. Luczak, and P. Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Annals of Applied Probability*, 2014. accepted.
- [JM86] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, 18 :20–65, 1986.
- [JMW12] B. Jourdain, S. Méléard, and W.A. Woyczynski. Lévy flights in evolutionary ecology. *Journal of Mathematical Biology*, 65(4) :677–707, 2012.
- [JS87] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.
- [Kal83] O. Kallenberg. *Random Measures*. Academic Press, 1983.
- [Ken50] D.G. Kendall. An artificial realization of a simple "birth-and-death" process. *J. Roy. Statist. Sec., Ser. B*, 12(1) :116–119, 1950.
- [KG99] A. Kleczkowski and B.T. Grenfell. Mean-field-type equations for spread of epidemics : The small world model. *Physica A*, 274 :355–360, 1999.
- [KKG13] I.Z. Kiss, D.M. Green, and R.R. Kao. Infectious disease control using contact tracing in random and scale-free networks. *J. R. Soc. Interface*, 3(6) :55–62, 2013.
- [KHL89] N.L. Kaplan, R.R. Hudson, and C.H. Langley. The hitchhiking effect revisited. *Genetics*, 123(4) :887–899, 1989.
- [Kli14] S. Kliem. A compact containment result for nonlinear historical superprocess approximations for population models with trait-dependence. *Electronic Communication in Probability*, 2014. [arXiv:1405.0815](https://arxiv.org/abs/1405.0815).
- [KLLM08] T.G. Kurtz, E. Lebensztayn, A.R. Leichsenring, and F.P. Machado. Limit theorems for an epidemic model on the complete graph. *ALEA*, 4 :45–55, 2008.
- [KM27] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. A*, 115 :700–721, 1927.
- [KN97] S. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51 :210–237, 1997.
- [Kov86] S.A. Kovyazin. On the limit behavior of a class of empirical means of a random set. *Theory of Probability and its Applications*, 30(4) :814–820, 1986. Translated from Russian by J. Malek.
- [KR11] I. Kurkova and K. Raschel. Random walks in  $(\mathbb{Z}_+)^2$  with non-zero drift absorbed at the axes. *Bulletin de la Société Mathématique de France*, 139 :341–387, 2011.
- [KS67] K.I. Kojima and H.E. Schaffer. Survival process of linked mutant genes. *Evolution*, 21 :518, 1967.
- [KS98] I. Kaj and S. Sagitov. Limit processes for age-dependent branching particle systems. *Journal of Theoretical Probability*, 11(1) :225–257, 1998.
- [Kur80] T.G. Kurtz. Representations of Markov processes as multiparameter time changes. *Annals of Probability*, 8(4) :682–715, 1980.
- [Kur81] T.G. Kurtz. *Approximation of Population Processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Soc. for Ind. and Appl. Math. (SIAM), Philadelphia, 1981.
- [Kur92] T.G. Kurtz. Averaging for martingale problems and stochastic approximation. In Springer, editor, *Applied stochastic analysis (New Brunswick, NJ, 1991)*, volume 177 of *Lectures Notes in Control and Inform. Sci.*, pages 186–209, Berlin, 1992.
- [KW14] S. Kliem and A. Winter. 2014. personal communication, work in progress.
- [LA99] R. Lounes and H. De Arazoza. A two-type model for the Cuban national program on HIV-AIDS. *IMA. J. Math. Appl. Med. Biol.*, 16 :143–154, 1999.
- [LGRT13] P. Lafitte-Godillon, K. Raschel, and V.C. Tran. Extinction probabilities for a distylous plant population modeled by an inhomogeneous random walk on the positive quadrant. *SIAM Journal on Applied Mathematics (SIAP)*, 73(2) :700–722, 2013.
- [LN96] C. Licea and C. M. Newman. Geodesics in two-dimensional first-passage percolation. *Annals of Probability*, 24(1) :399–410, 1996.

- [LR83] C. Laredo and A. Rouault. Grandes déviations, dynamique de populations et phénomènes malthusiens. *Annales de l'Institut Henri Poincaré*, XIX(4) :323–350, 1983.
- [M84] M. Métivier. Convergence faible et principe d'invariance pour des martingales à valeurs dans des espaces de sobolev. *Annales de l'IHP*, 20(4) :329–348, 1984.
- [M98] S. Méléard. Convergence of the fluctuations for interacting diffusions with jumps associated with boltzmann equations. *Stochastics and Stochastics Reports*, 63 :195–225, 1998.
- [McK26] A.G. McKendrick. Applications of mathematics to medical problems. *Proc. Edin. Math.Soc.*, 54 :98–130, 1926.
- [Met11] J.A.J. Metz. Thoughts on the geometry of meso-evolution : collecting mathematical elements for a post-modern synthesis. In F.A.C.C. Chalub and J.F. Rodrigues, editors, *The Mathematics of Darwins Legacy*, Basel, 2011. Birkhauser.
- [MGM<sup>+</sup>96] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs, and J.S. Van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *S.J. Van Strien & S.M. Verdun Lunel (ed.), Stochastic and Spatial Structures of Dynamical Systems*, 45 :183–231, 1996.
- [Mil11] J.C. Miller. A note on a paper by Erik Volz : SIR dynamics in random networks. *Journal of Mathematical Biology*, 62(3) :349–358, 2011. <http://arxiv.org/abs/0909.4485>.
- [MKD00] J. Müller, M. Kretzschmar, and K. Dietz. Contact tracing in stochastic and deterministic epidemic models. *Mathematical Biosciences*, 164 :39–64, 2000.
- [ML01] R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64 :066112, 2001.
- [MMT11] S. Méléard, J.A.J. Metz, and V.C. Tran. Limiting feller diffusions for logistic populations with age-structure. In *58th World Statistics Congress of the International Statistical Institute (ISI 2011)*, Dublin Ireland, 2011.
- [MN00] C. Moore and M.E.J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61 :5678–5682, 2000.
- [Mol87] I. Molchanov. Uniform laws of large numbers for empirical associated functionals of random closed sets. *Theory of Probability and its Applications*, 32 :556–560, 1987.
- [Mol90] I. Molchanov. Empirical estimation of distribution quantiles of random closed sets. *Theory of Probability and its Applications*, 35 :594–600, 1990.
- [Mol98] I. Molchanov. A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*, 25 :235–242, 1998.
- [Mol99] I. Molchanov. Random closed sets : results and problems. In O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic geometry, likelihood and computation*. CRC Press/Chapman and Hall, Boca Raton, 1999.
- [Mol05] I. Molchanov. *Theory of random sets*. Springer, 2005.
- [MP09] D. Mason and W. Polonik. Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability*, 19(3) :1108–1142, 2009.
- [MPRR12] J.-M. Marin, P. Pudlo, C.P. Robert, and R. Ryder. Approximate Bayesian computation methods. *Statistics and Computing*, 22(6) :1167–1180, 2012.
- [MR93] S. Méléard and S. Roelly. Sur les convergences étroite ou vague de processus à valeurs mesures. *C.R.Acad.Sci.Paris, Serie I*, 317 :785–788, 1993.
- [MR95] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures and algorithms*, 6 :161–180, 1995.
- [MT09] S. Méléard and V.C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Mathematical Biology*, 58(6) :881–921, 2009.
- [MT12a] S. Méléard and V.C. Tran. Nonlinear historical superprocess approximations for population models with past dependence. *Electronic Journal of Probability*, 17(47), 2012.
- [MT12b] S. Méléard and V.C. Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications*, 122(1) :250–276, 2012.
- [MT13] J.A.J. Metz and V.C. Tran. Daphnias : from the individual based model to the large population equation. *Journal of Mathematical Biology*, 66(4-5) :915–933, 2013. Special issue in honor of Odo Diekmann.
- [New03] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45 :167–256, 2003.
- [NRS05] C. Newman, K. Ravishankar, and R. Sun. Convergence of coalescing nonsimple random walks to the Brownian web. *Electronic Journal of Probability*, 10(2) :21–60, 2005.
- [NT] J. Norris and A.G. Turner. Weak convergence of the localized disturbance flow to the coalescing Brownian flow. *Annals of Probability*. to appear, [arXiv:1106.3252](https://arxiv.org/abs/1106.3252).
- [NT08] J. Norris and A.G. Turner. Planar aggregation and the coalescing Brownian flow. [arXiv:0810.0211](https://arxiv.org/abs/0810.0211), 2008.



- [NT12] J. Norris and A.G. Turner. Hastings-levitov aggregation in the small-particle limit. *Communication in Mathematical Physics*, 316(3) :809–841, 2012.
- [Oel90] K. Oelschläger. Limit theorem for age-structured populations. *The Annals of Probability*, 18(1) :290–318, 1990.
- [OK75] T. Ohta and M. Kimura. Effects of selected linked locus on heterozygosity of neutral alleles (hitchhiking effect). *Genetical Research*, 25(3) :313–326, 1975.
- [O’N02] P.D. O’Neill. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Mathematical Biosciences*, 180 :103–114, 2002.
- [OR99] P.D. O’Neill and G.O. Roberts. Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society : Series A*, 162 :121–129, 1999.
- [Per95] E. Perkins. *On the Martingale Problem for Interactive Measure-Valued Branching Diffusions*, volume 115(549). American Mathematical Society, Memoirs of the American Mathematical Society edition, May 1995.
- [Pim07] L.P.R. Pimentel. Multitype shape theorems for first passage percolation models. *Advances in Applied Probability*, 39(1) :53–76, 2007.
- [PSV02] R. Pastor-Satorras and A. Vespignani. Epidemics and immunization in scale-free networks. In *Handbook of Graphs and Networks : From the Genome to the Internet*, pages 113–132, Berlin, 2002. Wiley-VCH.
- [Rio12] O. Riordan. The phase transition in the configuration model. *Combinatorics, Probability and Computing*, 21(1-2) :265–299, 2012.
- [Roe86] S. Roelly. A criterion of convergence of measure-valued processes : Application to measure branching processes. *Stochastics*, 17 :43–65, 1986.
- [RR90] S. Roelly and A. Rouault. Construction et propriétés de martingales des branchements spatiaux interactifs. *International Statistical Review*, 58(2) :173–189, 1990.
- [RSS14] R. Roy, K. Saha, and A. Sarkar. Random directed forest and the Brownian web. 2014.
- [Rud87] W. Rudin. *Real and Complex Analysis*. McGraw-Hill International Editions, third edition edition, 1987.
- [RV09] P. Rigollet and R. Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15 :1154–1178, 2009.
- [RWP+95] R.B. Rothenberg, D.E. Woodhouse, J.J. Potterat, S.Q. Muth, W.W. Darrow, and A.S. Klovdahl. Social networks in disease transmission : The Colorado Springs study. In R.H. Needle, S.L. Coyle, S.G. Genser, and R.T. Trotter II, editors, *Social networks, drug abuse and HIV transmission*, volume 151 of *Research Monographs*, pages 3–18. National Instit, 1995.
- [SFT07] S.A. Sisson, Y. Fan, and M. Tanaka. Sequential Monte Carlo without likelihoods. *Proc. Nat. Acad. Sci. USA*, 104 :1760–1765, 2007.
- [SGK07] R. S. Stoica, E. Gay, and A. Kretzschmar. Cluster detection in spatial data based on monte carlo inference. *Biometrical Journal*, 49(2) :1–15, 2007.
- [SGPM11] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Parameter estimation for asymmetri bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics*, 5 :1313–1353, 2011.
- [SGPM12] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *Statistics and Probability Letters*, 82(7) :1439–1444, 2012.
- [SGPM13] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Random coefficients bifurcating autoregressive processes. *ESAIM P&S*, 2013. in press.
- [SGPM14] B. De Saporta, A. Gégout-Petit, and L. Marsalle. Statistical study of asymmetry in cell lineage data. *Computational Statistics & Data Analysis*, 69 :15–39, 2014.
- [SH74] J. Maynard Smith and J. Haigh. The hitchhiking effect of a favourable gene. *Genetical Research*, 23 :23–35, 1974.
- [SM97] D. Stoyan and I.S. Molchanov. Set-valued means of random particles. *Journal of Mathematical Imaging and Vision*, 7 :111–121, 1997.
- [SMPT05] E.J. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and death in a organism that reproduces by morphologically symmetric division. *PLoS Biol.*, 3(2), 2005.
- [SMS07] R. S. Stoica, V. J. Martinez, and E. Saar. A three dimensional object point process for detection of cosmic filaments. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 55 :189–205, 2007.
- [SMS10] R. S. Stoica, V. J. Martinez, and E. Saar. Filaments in observed and mock catalogs. *Astronomy and Astrophysics*, 510,A38 :1–12, 2010.
- [Sol87] W. Solomon. Representation and approximation of large population age distributions using poisson random measures. *Stochastic Processes and their Applications*, 26 :237–255, 1987.

- [SS94] D. Stoyan and H. Stoyan. *Fractals, random shapes and point fields*. Wiley, Chichester, 1994.
- [SSF09] I. Chis Ster, B.K. Singh, and N.M. Ferguson. Epidemiological inference for partially observed epidemics : the example of the 2001 foot and mouth epidemic in Great Britain. *Epidemics*, 1 :21–34, 2009.
- [SSN09] A. Singh, C. Scott, and R. Nowak. Adaptive Hausdorff estimation of density level sets. *Annals of Statistics*, 37 :2760–2782, 2009.
- [TBGD97] S. Tavaré, D.J. Balding, R.C. Griffiths, and P. Donnelly. Inferring coalescence times from DNA sequence data. *Genetics*, 145(2) :505–518, 1997.
- [Tra06] V.C. Tran. *Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques*. Phd thesis, Université Paris X - Nanterre, 12 2006. <http://tel.archives-ouvertes.fr/tel-00125100>.
- [Tra08] V.C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM : P&S*, 12 :345–386, 2008.
- [Tsy97] A. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25 :948–969, 1997.
- [TW98] B. Tóth and W. Werner. The true self-repelling motion. *Probability Theory and Related Fields*, 111 :375–452, 1998.
- [TWS<sup>+</sup>09] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M.P. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, 6 :187–202, 2009.
- [VM07] E. Volz and L. Ancel Meyers. Susceptible-infected-recovered epidemics in dynamic contact networks. *Proceeding of the Royal Society B*, 274 :2925–2933, 2007.
- [Vol08] E. Volz. SIR dynamics in random networks with heterogeneous connectivity. *Mathematical Biology*, 56 :293–310, 2008.
- [Vor77] O.Y. Vorob'ev. On set-theoretical characteristics of random processes of spread. *Izv. Sibirsk. Otdel. Akad. Nauk SSSR Ser. Tehn. Nauk*, 153(3) :3–7, 1977. (Russian).
- [VST12] F. Johansson Viklund, A. Sola, and A.G. Turner. Scaling limits of anisotropic Hasting-Levitov clusters. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(1) :235–257, 2012.
- [VST13] F. Johansson Viklund, A. Sola, and A.G. Turner. Small particle limits in a regularized Laplacian random growth model. 2013. arXiv :1309.2194.
- [Wan77] F.J.S. Wang. A central limit theorem for age- and density-dependent population processes. *Stochastic Processes and their Applications*, 5 :173–193, 1977.
- [Web85] G.F. Webb. *Theory of Nonlinear Age-Dependent Population Dynamics*, volume 89 of *Monographs and Textbooks in Pure and Applied mathematics*. Marcel Dekker, inc., New York - Basel, 1985.
- [WJ01] J.L. Wylie and A. Jolly. Patterns of Chlamydia and Gonorrhoea infection in sexual networks in Manitoba, Canada. *Sexually transmitted diseases*, 28(1) :14–24, January 2001.
- [WvLB10] J. Walraevens, J.S.H. van Leeuwen, and O.J. Boxma. Power series approximations for two-class generalized processor sharing systems. *Queueing Systems. Theory and Applications*, 66 :107–130, 2010.

**Résumé** Cette habilitation contient trois parties. Le premier chapitre s'intéresse à modèles individuels centrés de populations structurées, où les dynamiques individuelles (reproductions, compétitions, morts, vieillissement) sont supposées dépendre du passé, au travers des âges des individus ou de leurs lignées ancestrales, et de covariables héréditaires (sauf lorsqu'il y a mutation), appelées traits. En grandes populations, les processus aléatoires à valeurs mesures décrivant ces populations peuvent être approchés par des solutions d'EDP ou par des processus diffusifs, suivant les échelles de temps en jeu. Dans ce deuxième cas de figure, des phénomènes d'homogénéisation se produisent pour le vieillissement. Des méthodes de martingales introduites par Kurtz sont utilisées pour séparer échelles de temps lentes et rapides, répondant ainsi à des questions étudiées entre autres par Dynkin, Kaj et ses co-auteurs... mais non résolues dans les cas considérés ici. Pour décrire l'évolution des généalogies, des processus de naissances et morts historiques, au sens de Dawson et Perkins, sont proposés et nous en étudions les limites super-processus. La loi de la trajectoire ancestrale d'un individu tiré au hasard à un instant donné est difficile à décrire. Dans le cas sans interaction, nous pouvons répondre à cette question, en utilisant des représentations 'many-tone', mettant en évidence des phénomènes de biais. Une application pour les modèles multi-niveaux est considérée.

La seconde partie de ce chapitre s'intéresse à l'introduction de démographie et d'aléa dans deux modèles de génétique des populations. Un premier modèle, pour expliquer la génération et le maintien de la diversité neutre en présence de traits sous sélection, nous apporte un éclairage nouveau sur les phylogénies qu'il est possible de reconstruire à partir du polymorphisme génétique observé à un temps donné. Ensuite, une étude de populations possédant des systèmes de reproduction auto-incompatibles (par exemple les plantes à fleurs) est exposée, amenant à des problèmes de marches aléatoires non homogènes sur le quart de plan positif et absorbées sur la frontière.

Dans le second chapitre, nous présentons des modèles probabilistes et statistiques pour l'épidémiologie, motivés par les données collectées entre 1986 et 2006 à Cuba par le système de contact-tracing pour le SIDA. L'estimation statistique, par exemple pour évaluer les efficacités respectives de différents modes de dépistage, est rendue difficile par le fait que les personnes infectieuses non dépistées sont inobservées. L'ABC, une méthode bayésienne, apparaît comme adaptée au problème et au format et à la taille des données. Nous avons également une observation partielle des graphes d'infections et de détections pour Cuba (5389 noeuds reliés par 4073 arêtes, avec une composante géante de 2386 noeuds et 3168 arêtes, ce qui est unique dans la littérature sur le SIDA). L'étude statistique de ce graphe et la mise en évidence de communautés par des méthodes de 'clustering' ont été réalisées. D'un point de vue plus probabiliste, des théorèmes limites liés aux épidémies se propageant sur des grands graphes aléatoires de distributions de degrés données sont établis, fournissant trois équations à valeurs mesures pour décrire les distributions de degrés sous-jacentes au problème, et justifiant ainsi rigoureusement les EDO proposées par Volz qui ont eu un écho important dans le domaine.

Le troisième chapitre décrit deux contributions en géométrie aléatoire. La première concerne l'approximation de l'espérance de Vorob'ev d'ensembles aléatoires. La seconde est motivée par l'étude du 'Radial Spanning Tree' (RST) et de la 'Directed Spanning Forest' (DSF) introduits par Baccelli et Bordenave. Ces objets exhibent des structures de dépendance complexes qui empêchent l'utilisation des techniques classiques fondées sur les martingales, par exemple. Pour montrer que la DSF est un arbre sans chemin bi-infini, ou pour étudier les géodésiques et les interfaces du RST et montrer qu'il admet avec probabilités strictement positives 1, 2, 3, 4 ou 5 sous-arbres enracinés en l'origine, des arguments géométriques ou venant de la percolation sont invoqués.

---

**Summary** This thesis contains three parts. The first chapter considers individual-centered models of structured populations, where the individual dynamics (reproductions, competitions, deaths and aging) depend on the past, either through individuals' ages or through their ancestral lineages, and on hereditary covariates, called traits, that are transmitted from parent to offspring unless mutations occur. In large populations, the random measure-valued processes describing these populations can be approximated by solutions of PDEs or by diffusive processes, given the time-scales involved. In the second case, averaging phenomena are observed for the aging. Martingale methods introduced by Kurtz are used to separate the slow and fast time-scales, thus answering questions tackled among others by Dynkin, Kaj and coauthors... but unanswered in the cases studied here. To describe the evolution of the genealogies, historical births and deaths processes, à la Dawson et Perkins, are introduced. Their superprocess limits are studied. The distribution of the ancestral lineage of an individual chosen randomly at a given time is difficult to characterise. In the case without interaction, we can however answer this question by using many-to-one formulas that exhibit biases. An application to multi-level models is considered.

The second part of this chapter focuses on two models of population genetics, with demography and randomness. In the first model, we study how the neutral diversity is affected by selection and adaptation in an eco-evolutionary framework. This sheds a new light on the phylogenies that can be reconstructed from the genetical polymorphism observed at a given time. The second model considers populations with self-incompatible reproduction systems (such as some flower populations), leading to the study of inhomogeneous random walks on the quarter plane and absorbed at the boundary.

In chapter two, probabilistic and statistical models for epidemiology are presented. The questions are motivated by the HIV-AIDS data collected between 1986 and 2006 in Cuba by contact-tracing. Statistical estimation, to compare for instance the relative efficiencies of different detection methods, is made difficult as infectious but non-detected individuals are unobserved. A Bayesian technique called ABC is well suited for our problem and for size of our database. The infection and detection graphs are also partially observed (5,389 vertices linked by 4,073 edges, with a giant component of 2,386 vertices and 3,168 edges, which is unique in the literature on AIDS). The statistical study of this graph and its community structure obtained by clustering methods are shown. From a more probabilistic point of view, limit theorems linked with disease propagating on large random graphs with fixed degree distributions are established. This provides three measure-valued equations to describe important degree distributions underlying the propagation, and gives a rigorous proof of the ODEs proposed by Volz.

The third chapter describes two contributions in stochastic geometry. The first one deals with the approximation of the Vorob'ev expectation of random sets. The second one is motivated by the study of the Radial Spanning Tree (RST) and of the Directed Spanning Forest (DSF) introduced by Baccelli and Bordenave. These objects exhibit complex dependencies that prevent the use of classical techniques based on martingales, for instance. To show that the DSF is a tree without bi-infinite path or to study the geodesics and interfaces of the RST and show that it admits 1, 2, 3, 4 or 5 subtrees rooted at the origin with positive probability, arguments from geometry or percolation are used.

---

**Keywords** Measure-valued birth and death processes, limit theorems, age distributions, separation of slow and fast time-scales, historical particle systems, many-to-one formulas, population dynamics and eco-evolutionary models, random SIR models, epidemics on configuration networks, Vorob'ev expectation, random spanning tree, directed spanning forest.

---