HAL
open science

Structures aléatoires de branchement et applications en génétique des populations

Julien Berestycki

To cite this version:

Julien Berestycki. Structures aléatoires de branchement et applications en génétique des populations. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2010. tel-00552025

HAL Id: tel-00552025

https://theses.hal.science/tel-00552025v1

Submitted on 5 Jan 2011

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.

Habilitation à diriger des recherches

Structures aléatoires de branchement et applications en génétique des populations

Julien Berestycki

Rapporteurs :

Jean-François Delmas	Ecole Nationale des Ponts et Chaussées
Alison Etheridge	University of Oxford
Anton Wakolbinger	J.W. Goethe-Universität

Soutenue le 3 décembre devant un jury composé de :

Erwin Bolthausen	Universität Zürich
Jean Bertoin	Université Pierre et Marie Curie
Brigitte Chauvin	Université de Versailles St Quentin en Yvelines
Bernard Derrida	Ecole Normale Supérieure
Andreas Kyprianou	University of Bath
Jean-François Le Gall	Université Paris-Sud 11
Anton Wakolbinger	J.W. Goethe-Universität

À Livia

Remerciements

C'est un plaisir d'avoir l'occasion ici de remercier les nombreux collègues dont l'example et souvent l'amitié sont l'un des moteurs du métier qui est le nôtre.

I would like to start by warmly thanking Alison Etheridge, Anton Wakolbinger and Jean-François Delmas who refereed this thesis. I am deeply honoured and grateful to Erwin Bolthausen, Jean Bertoin, Brigitte Chauvin, Bernard Derrida, Andreas Kyprianou, Jean-François Le Gall and Anton Wakolbinger who are present today in the Jury.

J'ai eu un très grand plaisir à travailler avec tous mes collaborateurs, aux qualités et à la patience desquels je rend ici hommage. Merci à Jason Schweinsberg, Vlada Limic, Andreas Kyprianou, Simon Harris, Eric Brunet, John Harris, Matt Roberts, Vincent Bansaye, Zhan Shi, Elie Aidekon, Lorenzo Zambotti, Leonid Mytnik et Leif Doering. Une large part des travaux que je présente dans cette thèse est le produit de collaborations avec mon frère Nathanaël. Je tiens à lui exprimer ici ma gratitude pour tout ce que notre travail m'a apporté.

Je remercie tout particulièrement Eric Brunet. Depuis plusieurs années il nous propose avec Bernard Derrida des intuitions et des conjectures fascinantes. Nos discussions, ses questions et sa ténacité m'ont beaucoup influencé et continuent d'orienter ma recherche.

Sylvie Méléard ne peut pas être présente aujourd'hui, mais je tiens à lui dire toute mon amitié et ma reconnaissance pour ses enseignements stimulants, son acceuil chalereux au sein du CMAP à l'École Polytechnique et son soutien constant.

Merci a Jean Bertoin d'avoir guidé mes premiers pas dans la recherche en probabilité et de continuer à me faire l'amitié de ses conseils.

In the past years I have become a regular visitor at the Probability Laboratory of Bath University, where, thanks to Andreas Kyprianou, Simon Harris and all its past and present members, I have always found a vibrant scientifical life and a warm and welcoming atmosphere. I am equally indebted to Alison Etheridge for her constant support and encouragements. Thanks to her and the people in her group, I have had the opportunity to make several visits to the Statistic Department of Oxford University, each time a fruitful and friendly experience.

Finalement, j'aimerais remercier tous mes amis et collègues de Paris 6, du CMI à Marseille et de la comunauté mathématique en général.

Merci à mes amis, à ma famille et à Valentine pour leur soutien constant.

Table des matières

Table des matières i

Introduction iii

1 Comportement asymptotique des coalescents 1
1.1 Définition et premières propriétés 1
1.1.1 Définition 1
1.1.2 Le théorème de Pitman 2
1.1.3 La construction Poissonienne 3
1.1.4 La descente de l'infini 3
1.1.5 Fréquences asymptotiques et poussière 4
1.1.6 Les Beta-coalescents 5
1.2 Résultats asymptotiques 6
1.2.1 Nombre de blocs 6
1.2.2 Longueur totale de l'arbre de coalescence 7
1.2.3 Le chemin de descente de l'infini 9
1.2.4 Taille des blocs 10
1.2.5 Spectre multifractal de coalescence 12

2 Coalescents, processus de branchement et arbres continus 15
2.1 Définitions et premières propriétés 15
2.1.1 Processus de branchement à espace d'état continu 15
2.1.2 Le changement de temps de Lamperti 17
2.1.3 Les conditions de Grey 17
2.2 Généalogie des CSBP 18
2.2.1 Processus de hauteurs et arbres continus 18
2.2.2 Le processus *lookdown* 19
2.2.3 Le *lookdown* dans le processus de hauteur 23
2.2.4 Le Beta-coalescent dans l'arbre stable 25
2.2.5 Beta-coalescents : Heuristique pour le nombre de blocs et le spectre multifractal 27

2.3 Le cas des Λ-coalescents généraux et le lien avec les CSBP 27
2.3.1 Processus de Fleming-Viot généralisés et leur généalogie . . . 28
2.3.2 Couplage 29
2.4 Le cas du coalescent de Kingman 30

3 Formules d'échantillonage 35
3.1 Rappels 36
3.1.1 Le coalescent de Kingman comme limite d'échelle des généalogie dans les modèles de Cannings 36
3.1.2 Le coalescent de Kingman et la formule d'échantillonnage de Ewens 37
3.2 Le concept de taille effective 40
3.3 Au-delà du coalescent de Kingman 41
3.4 Formules asymptotiques pour le spectre allélique et le spectre de mutations 43

4 Mouvement Brownien avec absorption : généalogie et survie 47
4.1 Les conjectures de Brunet-Derrida 47
4.1.1 Les marches aléatoires branchantes avec sélection 47
4.1.2 Universalité 48
4.1.3 Les résultats rigoureux 49
4.1.4 Polymères dirigés 50
4.1.5 Le lien avec l'équation FKPP bruitée 51
4.2 Résultats principaux 53
4.2.1 Mouvement Brownien branchant avec absorption 53
4.2.2 Survie 55
4.2.3 Généalogie et CSBP de Neveu 56
4.2.4 Quelques idées de la preuve des Théorèmes 35, 36 et 37 58
4.2.5 Un bref aperçut de la preuve du Théorème 34 60

5 Trois aspects des processus de branchements. 63
5.1 Analogue de l'équation FKPP pour les fragmentations 63
5.2 Grandes déviations pour les processus de branchement en environnement aléatoire 68
5.3 Le squelette des superprocessus 72

Introduction

L'objet de ce mémoire est de présenter de façon succincte les travaux que j'ai menés et auxquels j'ai collaboré depuis la fin de ma thèse. Ces travaux sont reliés par le thème central de la *structure arborescente aléatoire* ou du *processus de branchement*.

Les arbres aléatoires apparaissent dans de nombreux contextes mathématiques, biologiques, informatiques ou physiques pour ne citer que quelques uns des champs où ces modèles trouvent à s'appliquer. En biologie évolutionnaire on les utilise pour modéliser les généalogies de populations et analyser le polymorphisme génétique observé (voir les livres [134, 68] ou [80]) ; en physique, les arbres aléatoires apparaissent dans de nombreux systèmes comme l'analyse de la structure ultra-métrique des modèles de verres de spin en champs moyen ([54, 108]), les polymères dirigés en environnement aléatoire ([94, 92]), les chocs de turbulences en dimension 1 ([85, 123]) ou les phénomènes de fragmentation ([36]). En informatique, les arbres représentent des structures de données ou d'algorithmes (voir [62]).

D'un point de vue mathématique, l'étude des processus de branchement remonte aux racines mêmes de la théorie des probabilités, avec les études de Galton et Watson et présente aujourd'hui un panorama d'une grande richesse.

Le premier chapitre concerne le comportement asymptotique en temps petit des processus de coalescence à collisions multiples, également appelés Λ-coalescents. Les processus de coalescence sont des modèles stochastiques qui décrivent des phénomènes dans lesquels des particules coalescent en blocs (ou clusters) de plus en plus grands à mesure que le temps s'écoule. Leur loi est entièrement caractérisée par une mesure finie Λ sur $[0, 1]$. De nombreux travaux leur ont été consacrés ces dernières années et il s'agit d'un champs de recherche extrêmement vivant aujourd'hui. Les coalescents se sont révélés particulièrement utiles, entre autre en génétique des populations, puisqu'ils permettent de décrire comment les lignées ancestrales d'un échantillon pris dans une population coalescent lorsque l'on remonte le temps. Ces applications sont une motivation majeure pour l'étude de l'évolution du nombre de blocs d'un coalescent. L'un des résultats principaux de cette première partie décrit ainsi la *vitesse de descente de l'infini,* c'est-à-dire le comportement asymptotique du nombre de blocs quand celui-ci diverge lorsque $t \searrow 0$.

Plus précisément, les coalescents commencent leur vie à l'instant 0 dans un état

où une infinité de particules sont présentes (chaque particule est isolée et constitue un bloc). Sous certaines conditions sur la mesure Λ la coalescence est assez rapide pour que pour tout temps $t > 0$ le nombre de blocs devienne fini. On dit alors que le coalescent *descend de l'infini*. Considérons $(\Pi(t), t \geq 0)$ un Λ-coalescent et notons $N(t)$ le nombre de blocs à l'instant t. On a alors $N(0) = \infty$ alors que $\forall t > 0 N(t) < \infty$ et lorsque $t \to 0$ on a $N(t) \nearrow \infty$. Dans [21, 22, 19, 20] nous étudions (pour une certaine classe de mesure Λ puis pour toutes les mesures Λ) la vitesse de descente de l'infini, c'est à dire que nous exhibons une fonction $t \mapsto v(t)$ telle que presque sûrement

$$\frac{N(t)}{v(t)} \to C_\Lambda$$

où C_Λ est une constante explicite qui ne dépend que de Λ. Outre ce résultat qui donne donc le comportement asymptotique du nombre de blocs, nous obtenons également des résultats concernant la taille asymptotique des blocs, le nombre de blocs perdus à chaque collision, le nombre de blocs de taille atypique, etc....

Un outil essentiel pour comprendre et prouver les résultats donnés au premier chapitre est le lien qui existe entre coalescents et processus de branchement à espace d'états continu. Les processus de branchement à espace d'états continu (ou CSBP pour utiliser l'acronyme anglais d'usage) sont la limite d'échelle des processus de Galton-Watson critiques. Ce sont des processus de Markov à valeur dans $\mathbb{R}^+$ qui modélisent la taille d'une grande population. Si l'on sait bien qu'il est aisé de retrouver la généalogie d'un processus de Galton-Watson (il suffit de prendre l'arbre correspondant), le passage à la limite qui permet d'obtenir le CSBP "oublie" la généalogie. Il existe cependant plusieurs approches qui permettent de "retrouver" la généalogie de ces processus. Poursuivant un travail de Le Gall et Le Jan [104] Duquesne et Le Gall [63] ont donné une extension du théorème de Ray-Knight qui permet de voir les CSBP comme le temps local d'un certain processus d'excursion. La forêt d'arbres codée par ces excursions contient la généalogie du processus. Une autre approche due à Donnelly et Kurtz [58, 59] consiste à construire une système de particules dont la mesure empirique donne la composition de la population. La notion de généalogie pour ce système de particules est alors claire. C'est en exploitant cette dernière approche que Birkner et al. [50] ont prouvé que la généalogie des CSBP alpha-stables était donnée par les Beta-coalescents. Les résultats que je présente dans ce second chapitre montrent que les deux notions de généalogies pour les CSBP mentionnés ci-dessus coïncident dans le sens où le système de particules de Donnelly et Kurtz peut se construire à l'aide du processus d'excursions de Duquesne et Le Gall. Ce résultat permet de "voir" les Beta coalescent dans le processus d'excursion et d'utiliser les propriétés de ce dernier pour étudier le comportement des premiers.

Les autres Λ-coalescents (i.e. ceux qui ne sont pas de Beta coalescents) ne sont jamais la généalogie d'un CSBP (cf. [50]). Cependant Bertoin et Le Gall [29, 31] ont montré qu'il existe un processus à valeur mesure dual (les processus de Fleming-Viot généralisés) qui présente une certaine similarité avec les CSBP. Je montrerai comment on peut coupler les généalogies des CSBP et de ces processus de Flemming-

Viot généralisés afin d'en déduire des résultats sur le comportement de coalescents généraux. Ces résultats présentent également un intérêt en eux même et apportent un nouvel éclairage sur une séries de travaux récents ([50, 29, 31, 22] etc...)

Le troisième chapitre concerne les applications à la génétique des populations. Le coalescent de Kingman ([98, 99]), introduit en 1982 par Kingman, est le modèle standard de coalescence utilisé dans ce domaine. Il peut être informellement décrit de la façon suivante : pour un échantillon de taille n fixée, chaque paire de lignées ancestrale coalesce à taux 1. Le coalescent de Kingman est utilisé pour quantifier le polymorphisme génétique attendu dans une population neutre et non-structurée. Ainsi, pour une taille d'échantillon donnée, la loi de la partition allélique (nombre d'allèles distincts, fréquences des différents allèles présents,..) est décrite par la célèbre formule d'échantillonnage de Ewens [80].

Il y a cependant de bonnes raisons de penser que dans certaines situations -certaines espèces marines (voir, e.g., [106]) ou les populations de virus à l'intérieur d'un patient par exemple- d'autres coalescents permettant des collisions multiples reflètent mieux la réalité.

Il y a donc une forte motivation biologique pour essayer d'obtenir des analogues de la formules d'échantillonnage de Ewens pour le cas des Λ-coalescents, i.e. de décrire la loi de partition allélique d'un échantillon de taille n (voir, e.g. [60, 75, 110, 137, 10]). Il semble que seul le coalescent de Kingman permette d'obtenir des solutions exactes (cf. par exemple [80] ou [68]), et qu'en général il faille plutôt chercher de bonnes approximations. Un ingrédient essentiel pour l'obtention de ces approximations est l'étude fine du comportement en temps petit des Λ-coalescents (chapitre 1).

Le chapitre suivant concerne une direction de recherche plus récente pour moi et qui trouve son origine dans une série de travaux de Brunet, Derrida et leurs coauteurs [39, 41, 42, 43, 44, 46, 47, 48, 49]. L'un des objets considérés dans ces travaux -initialement motivés par l'étude de la version bruitée de l'équation FKPP- est le modèle de population en évolution et sous sélection suivant. La population est de taille constante N et évolue par générations discrètes. Les individus sont identifiés avec des points sur la ligne réelle. À chaque génération, chaque individu produit k enfants dont le déplacement par rapport à sa propre position est donné par des variables iid. On sélectionne les N individus les plus à droite pour former la génération suivante. Brunet, Derrida, Mueller et Munier [43, 46] formulent alors la conjecture remarquable que sur une échelle de temps d'ordre $(\log N)^3$ la généalogie de cette population est le coalescent de Bolthausen-Sznitman. Du fait de la non-indépendance entre les particules, il parait encore difficile de prouver exactement ce résultat. On s'intéresse donc à un modèle proche, le mouvement Brownien branchant avec absorption (ou marche branchante avec absorption si on souhaite rester en temps discret). Dans ce nouveau modèle, la taille de la population n'est plus fixée, les particules sont tuées dés qu'elles passent sous une barrière linéaire, qui impose ainsi une “vitesse" déterministe. Plus précisément, le modèle que l'on considère est un système de particules sur $\mathbb{R}^+$ qui se déplacent selon des mouvements Browniens

indépendants avec une dérive d'intensité $\mu \geq 0$ vers 0, branchent à taux 1 en deux particules et qui sont immédiatement tuées quand elles atteignent 0. Il existe une valeur critique de μ ($\mu_c = \sqrt{2}$ dans notre cas) à partir de laquelle la dérive est trop forte et le processus finit par mourir alors que si $\mu < \mu_c$ on a une probabilité positive de survivre. Nous avons pu montrer pour ce modèle que, quand $\mu \searrow \mu_c$, dans une échelle de temps qui correspond bien à $(\log N)^3$ la généalogie est en effet le coalescent de Bolthausen-Sznitman. Ce résultat suggère ainsi qu'un bon modèle pour la généalogie d'une population sous sélection serait le coalescent de Bolthausen-Sznitman.

Le mouvement Brownien branchant avec absorption (ou les marches aléatoire branchantes avec absorption) sont par ailleurs des objets qui font actuellement l'objets de nombreuses recherches, citons [83, 11, 12, 90]. L'une des questions essentielles concernant ces modèles (et qui occupe tous les articles cités ci-dessus) est de décrire le comportement de la probabilité de survie lorsque $\mu = \mu_c - \epsilon$ et que l'on part d'une particule à x. Nous apportons à cette question une réponse très précise qui valide des prédictions de [57].

Afin de ne pas trop alourdir ce mémoire, les trois articles qui composent le dernier chapitre sont présentés plus brièvement. Le lien entre certaines marches aléatoires branchantes et l'équation aux dérivés partielles de FKPP (dans sa version bruitée ou non) est un sujet de recherche bien développé et qui a connu un fort regain d'intérêt avec les travaux de Derrida et Brunet évoqués ci-dessus. Le premier travail que je présente dans ce chapitre montre comment les résultats classiques (dans le cadre du mouvement Bownien branchant et de marches branchantes simples) qui lient solution de l'équation et martingales multiplicatives peuvent se généraliser au cas des fragmentations homogènes. Nous abordons ensuite un résultat de grande déviation pour les processus en environnement aléatoire et enfin nous donnons une décomposition trajectorielle pour les superprocessus surcritiques.

Chapitre 1

Comportement asymptotique des coalescents

Les résultats présentés dans cette section sont issus des articles [21, 22, 19, 20] et concernent le comportement en temps petit des Λ-coalescents. Motivés par des questions issues de la génétique des populations (et qui sont explorées plus en détail au chapitre 3) nous avons en particulier analysé le comportement du nombre de blocs en temps petit et la longueur totale de l'arbre du n-coalescent.

1.1 Définition et premières propriétés

1.1.1 Définition

Soit $\mathcal{P} := \{\text{partitions de } \mathbb{N}\}$. Un élément $\pi \in \mathcal{P}$ peut être vu comme une relation d'équivalence $\sim_\pi$ sur $\mathbb{N}$ ou bien comme une collection de blocs $\pi = (B_1, B_2, \ldots)$ où les B_i sont des sous-ensembles de $\mathbb{N}$ tels que $\cup_i B_i = \mathbb{N}$ et $B_i \cap B_j = \emptyset$. Étant donné $\pi \in \mathcal{P}$ on aura besoin de la restriction de π à $\{1, \ldots, n\}$ (qui est la partition de cet ensemble défini par la même relation d'équivalence que π. Une partition (aléatoire) est dite échangeable si sa loi est laissée inchangée par l'action d'une permutation.[1]

Définition 1 *Un coalescent à collisions multiples (ou Λ-coalescent) est une processus de Markov $(\Pi_t, t \geq 0)$ sur l'espace d'état $\mathcal{P}$ issu de la partition triviale en singletons*

$$\Pi_0 = \{1\}, \{2\}, \ldots$$

satisfaisant :

(i) Le processus est échangeable.

[1] Pour $\pi \in \mathcal{P}$ et σ une injection de $\mathbb{N} \mapsto \mathbb{N}$ on définit $\sigma(\pi)$ par la relation d'équivalence $i \sim_{\sigma(\pi)} j$ si et seulement si $\sigma(i) \sim_\pi \sigma(j)$.

(ii) Si $n \in \mathbb{N}$, alors la restriction $(\Pi_t^{(n)}, t \geq 0)$ de $(\Pi_t, t \geq 0)$ à $[n] := \{1, \ldots, n\}$ est une chaîne de Markov en temps continu qui prends ses valeurs dans $\mathcal{P}_n$ l'ensemble de partitions de $[n]$;

(iii) Les seules transitions possibles des chaînes $\Pi^{(n)}$ correspondent à des événements de coagulation par lesquels plusieurs blocs coalescent en un seul.

On peut montrer (cf. [15]) que les taux de transition des chaînes $\Pi^{(n)}$ ont la propriété suivante :

Proposition 2 *Il existe des réels $(\lambda_{b,k}, b \geq 2, k \leq b)$ qui déterminent les taux de transitions des chaînes $\Pi^{(n)}$. Plus précisément si $\pi, \pi' \in \mathcal{P}_n$ sont telles que π a b blocs et π' s'obtient en fusionnant k blocs de π alors $q_n(\pi, \pi') = \lambda_{b,k}$ où $q_n(\pi, \pi')$ est le taux de transition de $\Pi^{(n)}$ de π vers π'.*

La définition ci-dessus est en fait équivalente (voir [15]) à la suivante :

Définition 3 *Un coalescent à collisions multiples (ou Λ-coalescent) est une processus de Markov $(\Pi_t, t \geq 0)$ sur l'espace d'état $\mathcal{P}$ issu de la partition triviale en singletons*

$$\Pi_0 = \{1\}, \{2\}, \ldots$$

satisfaisant :

(i) Le processus est échangeable.

(ii) Pour tout $s, t \geq 0$ la loi de $\Pi(s+t)$ est celle de $coag(\Pi(t), \Pi'(s))$ où $\Pi'(s)$ est une copie indépendante de $\Pi(s)$ et où $coag(\pi, \pi')$ est la partition que l'on obtient en fusionnant les blocs de π dont les indices sont dans le même bloc de π'.

(iii) On ne peut pas avoir deux événement de coalescence simultanés.

Sans le point (iii) la définition ci-dessus est celle des coalescents à collisions multiples simultanées, aussi appelés Ξ-coalescents ([125, 111]).

1.1.2 Le théorème de Pitman

Il est aisé de voir que les nombres $\lambda_{b,k}$ doivent nécessairement vérifier une relation de récurrence. Examinons le cas $b = 2, 3$. Le taux $\Lambda_{2,2}$ est le taux auquel on passe de l'état $\{1\}, \{2\}$ à $\{1, 2\}$. Si on considère maintenant $\Pi^{(3)}$, partant de l'état $\{1\}, \{2\}, \{3\}$ les transition vers $\{1, 2\}, \{3\}$ et $\{1, 2, 3\}$ correspondent toutes deux à $\{1\}, \{2\} \mapsto \{1, 2\}$ quand on regarde la restriction à [2]. On en conclut que l'on a nécessairement $\lambda_{2,2} = \lambda_{3,2} + \lambda_{3,3}$. On peut montrer que plus généralement

$$\lambda_{b,b} = \lambda_{b+1,b+1} + \lambda_{b+1,b}, \ b \geq 2. \tag{1.1}$$

Pitman [117] montre qu'un tableau de nombre réels positifs vérifiant ces relations est caractérisé par une unique mesure finie sur $[0,1]$ souvent notée Λ telle que

$$\lambda_{b,k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx). \tag{1.2}$$

Au vu de l'équivalence des deux définitions 1 et 3 cette mesure Λ caractérise également complètement la loi d'un coalescent à collisions multiples. Dans toute la suite nous supposerons sans perte de généralité que Λ est une mesure de probabilité

$$\Lambda[0,1] = 1. \tag{1.3}$$

En effet, au vu de l'équation (1.2), il est claire que multiplier la masse totale de Λ par un facteur constant revient simplement à accélérer (ou à ralentir) toute l'évolution du processus par le même facteur.

1.1.3 La construction Poissonienne

Dans son article fondateur [117] Pitman étudie plusieurs propriétés importantes des Λ-coalescents dont la constructions Poissonienne que nous rappelons à présent et qui explique le rôle et la signification de la mesure Λ.

Supposons que $\Lambda(\{0\}) = 0$. Soit

$$\pi(\cdot) = \sum_{i\in\mathbb{N}} \delta_{t_i,x_i}(\cdot) \tag{1.4}$$

un processus de points de Poisson sur $\mathbb{R}^+ \times (0,1)$ d'intensité $dt \otimes \nu(dx)$ où $\nu(dx) = x^{-2}\Lambda(dx)$. On construit alors $(\Pi_t, t \geq 0)$ avec la dynamique suivante : le processus Π n'évolue qu'aux instants t des atomes (t,x) de π. Pour chaque bloc de $\Pi(t-)$ on tire à pile ou face avec une pièce qui donne face avec probabilité x. On coalesce alors immédiatement tous les blocs pour lesquels la pièce est tombée sur face en un seul bloc. Les autres blocs sont laissés inchangés. Pour rendre cette construction rigoureuse il faut commencer par construire les restrictions $\Pi^{(n)}$, on ne donnera pas ici plus de détails, le lecteur pourra consulter [117] ou [17]. On voit donc que $x^{-2}\Lambda(dx)$ s'interprète comme le taux auquel une proportion x de tous les blocs fusionne.

1.1.4 La descente de l'infini

Nous rappelons à présent une propriété remarquable des Λ-coalescents. Soit E l'événement "pour tout $t > 0$ le nombre de blocs est infini" et soit F l'événement "pour tout $t > 0$ le nombre de blocs est fini"

$$E := \{\forall t \geq 0 : N(t) = \infty\},$$
$$F := \{\forall t \geq 0 : N(t) < \infty\}.$$

Pitman [117] montre que si $\Lambda(\{1\}) = 0$, seuls deux comportements sont possibles : soit $\mathbb{P}(E) = 1$ ou bien $\mathbb{P}(F) = 1$. Lorsque $\mathbb{P}(F) = 1$, on dit que le processus Π *descend de l'infini*. Le coalescent de Kingman est un exemple de coalescent qui descend de l'infini. En revanche si l'on prend $\Lambda(dx) = dx$ la mesure uniforme sur $[0, 1]$, alors le Λ-coalescent correspondant ne descend pas de l'infini. Ce dernier exemple de Λ-coalescent a été étudié par Bolthausen et Sznitman [54] en relation avec les modèles de verres de spin.

Une condition nécessaire et suffisante pour déterminer si un Λ-coalescent descend ou non de l'infini nous est fournie par Schweinsberg [124] : on définit

$$\gamma_b = \sum_{k=2}^{b} (k-1) \binom{b}{k} \lambda_{b,k},$$

alors le Λ-coalescent descend de l'infini si et seulement si

$$\sum_{b=2}^{\infty} \gamma_b^{-1} < \infty. \tag{1.5}$$

Lorsqu'un coalescent descend de l'infini, il est naturel de se demander à quelle vitesse il le fait, c'est à dire de chercher à connaître le comportement asymptotique quand $t \to 0$ du processus $t \mapsto N(t)$ où $N(t)$ est le nombre de blocs de $\Pi(t)$. Le seul cas connu était celui du coalescent de Kingman pour lequel il est aisé de montrer que

$$tN(t) \to 2 \text{ quand } t \searrow 0$$

presque sûrement.

1.1.5 Fréquences asymptotiques et poussière

Soit $(\Pi(t), t \geq 0)$ un Λ-coalescent, alors par définition $\Pi(t)$ est un processus échangeable (invariant par l'action d'une permutation des entiers naturels). Les partitions échangeables ont été étudiées par Kingman [97] (on peut voir ces travaux comme une application aux partitions échangeables de la théorie des suites de variables échangeables de De Finetti, cf. [2]). Kingman montre en particulier que si $B \subset \mathbb{N}$ est un bloc de la partition $\Pi(t)$, alors la limite

$$\lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\{i \in B\}}$$

existe presque sûrement et est appelée la fréquence asymptotique de B. Comme le nombre de blocs est au plus dénombrable, les fréquences asymptotiques des blocs existent presque sûrement pour tous les blocs et tous les t rationnels simultanément. Un simple argument de monotonie permet de renforcer ce dernier énoncé à tous les $t \geq 0$. Si pour chaque $t > 0$ la somme des fréquences asymptotiques des blocs de $\Pi(t)$

est égale à 1, on dit que le coalescent n'a pas de poussière, ou que ses fréquences sont *propres*. Pitman (voir Théorème 8 de [117]) a montré que c'est le cas si et seulement si $\int_0^1 x^{-1}\,\Lambda(dx) = \infty$. Lorsque les fréquences ne sont pas propres, presque sûrement une proportion positive des entiers naturels sont isolés dans des blocs singletons (la *poussière*). Cette proportion est alors la masse manquante pour arriver à 1 lorsque l'on somme les fréquences asymptotiques des autres blocs. Les coalescents dont la somme des fréquences asymptotiques est 1 n'ont aucun singleton.

1.1.6 Les Beta-coalescents

Les Beta-coalescents sont une famille particulière de Λ-coalescents qui joue un rôle important dans les résultats que je présente plus loin du fait des propriétés fortes d'invariance et de symétrie dont bénéficient les arbres stables associés. En outre, Schweinsberg [126] a montré que les Λ-coalescents sont, avec le coalescent de Kingman, les limites d'échelles naturelles pour les généalogies d'une large classe de modèles de population, donnant ainsi une motivation supplémentaires de considérer cette famille avec un soin particulier.

Les Beta-coalescents sont les Λ-coalescents que l'on obtient en choisissant comme mesure Λ la densité d'une variable Beta$(2-\alpha, \alpha)$ avec $0 < \alpha \leq 2$

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\,dx.$$

Les taux de transitions sont donnés par

$$\lambda_{b,k} = \frac{B(k-a, b-k+a)}{B(2-a, a)}$$

où $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. Lorsque le paramètre α varie entre 0 et 2, on a différents comportements possibles pour le processus :

- Quand $\alpha \to 2$ la mesure Λ converge faiblement vers δ_0 la masse de Dirac en 0 et le Λ-coalescent se rapproche donc (au moins formellement) du coalescent de Kingman (qui descend de l'inifini).

- Quand $1 < \alpha < 2$ le coalescent descend de l'infini.

- Quand $\alpha = 1$ on a $\Lambda(dx) = dx$ et le coalescent correspondant est le coalescent de Bolthausen-Sznitman, un processus bien connu qui apparaît dans des contextes de mécanique statistique liés aux modèles de verres de spin. Ce coalescent ne descend pas de l'infini mais il a encore des fréquences propres (i.e. les fréquences asymptotiques de ces blocs somment à 1).

- Si $\alpha < 1$ les fréquences ne sont plus propres, à tout instant $t \geq 0$ une proportion positive des entiers naturels sont encore des singletons dans la partition $\Pi(t)$.

Nous supposerons toujours dans la suite que le paramètre α vérifie $1 < \alpha < 2$.

1.2 Résultats asymptotiques

Je présente maintenant les résultats asymptotiques sur le comportement des Λ-coalescents que nous avons obtenus dans [21, 22, 19, 20]. Les articles [21, 22] se concentrent sur le cas où Λ est la distribution d'une variable Beta$(2-\alpha,\alpha)$ (ou telle que $\Lambda(dx) = f(x)\,dx$, où $f(x) \sim Ax^{1-\alpha}$) avec $\alpha \in (1,2)$. Ces hypothèses sont relâchées dans [19, 20] dont les résultats s'appliquent au cas de mesures Λ tout à fait générales.

1.2.1 Nombre de blocs

La forme la plus général du résultat concernant le nombre de blocs est donné dans [19]. À chaque mesure Λ on associe une fonction ψ définie par

$$\psi_\Lambda(q) \equiv \psi(q) := \int_{[0,1]} (e^{-qx} - 1 + qx)\nu(dx), \tag{1.6}$$

où $\nu(dx) = x^{-2}\Lambda(dx)$.

On définit

$$u_\psi(t) \equiv u(t) := \int_t^\infty \frac{dq}{\psi(q)} \in (0,\infty),\ t > 0, \tag{1.7}$$

et son inverse càdlàg (avec la convention $\inf \emptyset = \infty$)

$$v_\psi(t) \equiv v(t) := \inf\left\{ s > 0 : \int_s^\infty \frac{1}{\psi(q)} dq < t \right\},\ t > 0. \tag{1.8}$$

Quand Λ est la distribution d'une variable Beta$(2-\alpha,\alpha)$ avec $1 < \alpha < 2$,

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\,dx \tag{1.9}$$

on a $\psi(q) \sim c_1 q^\alpha$ as $q \to \infty$, et ainsi

$$v(t) \sim c_2 t^{-1/(\alpha-1)}, \quad \text{as } t \to 0,$$

où $c_1 = (\Gamma(\alpha)\alpha(\alpha-1))^{-1}$ et $c_2 = (\alpha\Gamma(\alpha))^{-1/(\alpha-1)}$. Ces calculs se généralisent aisément au cas où Λ est à variations régulières près de 0 d'indice $1 < \alpha < 2$.

Théorème 4 (B., Berestycki, Limic [19])

$$\lim_{t\to 0} \frac{N^\Lambda(t)}{v_\psi(t)} = 1,\ \textit{presque sûrement.} \tag{1.10}$$

Ce théorème est surtout intéressant pour les coalescent qui descendent de l'infini (autrement $N^\Lambda(t) = \infty$ pour tout t) mais le théorème reste formellement vrai même sans supposer que l'on descend de l'infini si l'on adopte la convention $\infty/\infty = 1$.

Ce résultat généralise celui que nous avions obtenu dans [22] et qui concerne le cas des Beta-coalescents (ou des coalescents dits "à variations régulières") et que je rappelle ci dessous.

Théorème 5 (B., Berestycki, Schweinsberg [21]) *Soit Λ une mesure finie sur $[0,1]$ telle que $\Lambda(dx) = f(x)\,dx$, où $f(x) \sim Ax^{1-\alpha}$ pour un certain $\alpha \in (1,2)$ et $\sim$ signifie que le rapport des deux quantités tends vers 1 lorsque $x \downarrow 0$. Soit $(\Pi(t), t \geq 0)$ le Λ-coalescent, et $N(t)$ le nombre de blocs de la partition $\Pi(t)$. Alors*

$$\lim_{t\downarrow 0} t^{1/(\alpha-1)} N(t) = \left(\frac{\alpha}{A\Gamma(2-\alpha)}\right)^{1/(\alpha-1)} \quad p.s. \tag{1.11}$$

En particulier, si Λ est la distribution d'une variable Beta$(2-\alpha, \alpha)$, alors

$$\lim_{t\downarrow 0} t^{1/(\alpha-1)} N(t) = (\alpha\Gamma(\alpha))^{1/(\alpha-1)} \quad p.s. \tag{1.12}$$

Comme on le verra dans le chapitre suivant, le Théorème 5 se prouve en utilisant les liens qui unissent Beta-coalescents et processus de branchement à espace d'état continus. Dans [22] , nous montrons que l'objet sous-jacent fondamental pour le coalescent et les processus de branchement est en fait une arbre continu stable. Une preuve utilisant le théorème de Kesten-Stigum peut être déduite de cette remarque. Le Théorème 4 en revanche s'obtient avec des méthodes assez différentes qui s'appuient essentiellement sur des techniques de martingales. Avec un peu de travail supplémentaire celles-ci permettent d'obtenir également un analogue du Théorème 4 en terme de convergence des moments

Théorème 6 (B.,Berestycki, Limic [19]) *Pour chaque $d \in [1,\infty)$,*

$$\lim_{s\to 0} E\left(\sup_{t\in[0,s]} \left|\frac{N^{\Lambda}(t)}{v_{\psi}(t)} - 1\right|^{d}\right) = 0. \tag{1.13}$$

1.2.2 Longueur totale de l'arbre de coalescence

Lorsqu'un coalescent descend de l'infini, la hauteur de l'arbre de coalescence, c'est à dire l'age du plus récent ancêtre commun de toute la population est une variable aléatoire finie. Par exemple dans le cas du coalescent de Kingman, il n'est pas difficile de voir que cette variable a même loi que $\sum_{i=2}^{\infty} \mathbf{e}_i / \binom{i}{2}$ où les $\mathbf{e}_i$ sont des varaibles iid de loi exponentielle et de moyenne 1, et a donc une moyenne finie égale à 2.

La longueur totale de l'arbre L_∞ est la somme des longueur de ses branches et on a

$$L_\infty = \int_0^{\zeta} N(u)du$$

où $N(u)$ est le nombre de blocs (de branches) au temps u et $\zeta = \inf\{t : N(t) = 1.\}$ Il n'est pas difficile de voir que presque sûrement $L_\infty = \infty$. Plus précisément on a le résultat suivant. Soit Λ une mesure de probabilité sur $[0,1]$ et Π le coalescent

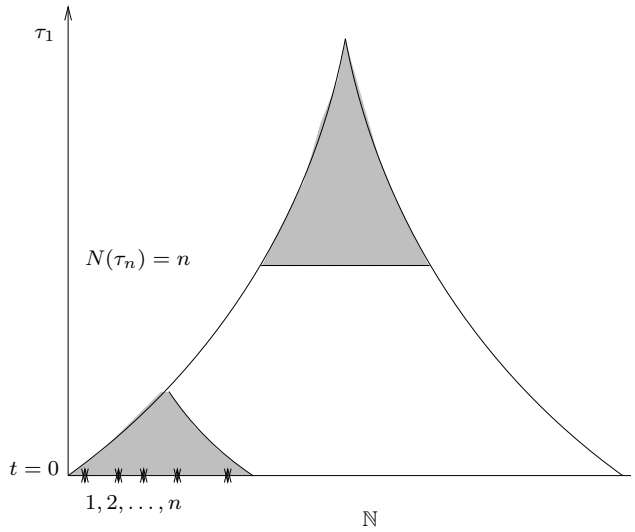

FIGURE 1.1: Conditionnellement à $N(\tau_n) = n$ (ou $\tau_n = \inf\{t : N(t) \leq n\}$) on peut voir l'arbre $\mathbf{T}_n$ du n-coalescent en deux endroits de l'arbre continu $\mathbf{T}$. 1) L'arbre $\mathbf{T}_n$ généré par les n premiers entiers (partie grise en bas à gauche de l'arbre) et 2) La partie de l'arbre au-dessus de τ_n générée par $(\Pi(t),\, t \geq \tau_n)$ (partie grise supérieure).

correspondant. On note L_n la variable aléatoire qui est la longueur totale du n-coalescent associé. Notons que les variables $L_n, n = 1, \ldots, \infty$ sont définies sur le même espace de probabilité par Π (il suffit de considérer les restrictions $\Pi^{(n)}$ pour avoir L_n) et que l'on a bien sûr $L_n \nearrow L_\infty$.

Connaître plus précisément le comportement asymptotique est important du point de vu des applications en génétique des populations, car si l'on ajoute des mutations poissoniennes sur l'arbre des lignées ancestrales, la longueur L_n nous donnera l'espérance du nombre de mutations visibles dans un échantillon de taille n. Notons $N^{\Lambda,n}$ le processus du nombre de blocs de $\Pi^{(n)}$ avec état initial $\Pi_0^{(n)} = \{\{1\}, \ldots, \{n\}\}$. Clairement

$$L_n = \int_0^{\zeta_n} N^{\Lambda,n}(u)du$$

où ζ_n est le premier temps où $\Pi^{(n)}$ n'a qu'un bloc.

Soit $\tau_n := \inf\{s > 0 : N^\Lambda(s) \leq n\}$. Posons $H_n := \{N^\Lambda(\tau_n) = n\}$ l'événement que le Λ-coalescent atteint jamais une configuration avec exactement n blocs. Alors, par la propriété de Markov forte la loi conditionnelle de $(N^\Lambda(s + \tau_n),\, s \geq 0)$ sachant $\mathcal{F}_{\tau_n}$ sur l'événement H_n, est identique à la loi de $N^{\Lambda,n}$. Soit $t_n = u_\psi(n)$ de sorte que $v_\psi(t_n) = n$.

La longueur de l'arbre au dessus de τ_n vaut $\int_{\tau_n}^{\zeta} N(u)du$ et le comportement asymptotique de cette quantité dépend de ce qui se passe près de τ_n. En admettant que l'on peut remplacer $N(u)$ par $v(u)$ grâce au Théorème 4 on voit que L_n doit se comporter

comme $\int_{t_n}^1 v(u)du$. Cette intuition est essentiellement correcte comme le montre le Théorème suivant.

Théorème 7 (B.,Berestycki, Limic [19]) *Pour chaque $s > 0$ on a*

$$\lim_{n\to\infty} \frac{\int_0^s N^{\Lambda,n}(t)\,dt}{\int_0^s v_\psi(t_n+t)\,dt} = \lim_{n\to\infty} \frac{\int_0^s N^{\Lambda,n}(t)\,dt}{\int_0^s E(N^{\Lambda,n}(t))\,dt} = 1, \text{ en probabilité.}$$

Pour les coalescents de Kingman ou les Beta$(\alpha, 2-\alpha)$-coalescents avec $1 < \alpha < 2$, la convergence ci-dessus est presque sûre.

1.2.3 Le chemin de descente de l'infini

Considérons à présent la dynamique du processus du nombre de blocs $(N(t), t \geq 0)$, plus en détail. Nous faisons dans ce paragraphe les mêmes hypothèses sur Λ que pour le Théorème 5, i.e. $\Lambda(dx) = f(x)dx$ où $f(x) \sim Ax^{1-\alpha}$.

Soit $\zeta_{n,k}$ la probabilité que si le Λ-coalescent a n blocs, il en perd exactement k au prochain événement de coalescence. Plus précisément si $T = \inf\{t : \Pi^{(n)}(t) \neq \Pi^{(n)}(0)\}$, alors $\zeta_{n,k} = P(N_n(T) = n-k)$. Si on appelle $\lambda_n = \sum_{k=2}^n \binom{n}{k}\lambda_{n,k}$ le taux total de coalescence quand le coalescent a n blocs, alors

$$\zeta_{n,k} = \binom{n}{k+1}\frac{\lambda_{n,k+1}}{\lambda_n}$$

puisque $k+1$ blocs doivent fusionner pour perdre k blocs. On voit sans peine que (cf. Lemme 4 de [33]) que

$$\lim_{n\to\infty} \zeta_{n,k} = \frac{\alpha\Gamma(k+1-\alpha)}{(k+1)!\,\Gamma(2-\alpha)}. \tag{1.14}$$

En posant $\zeta_k = \alpha\Gamma(k+1-\alpha)/[(k+1)!\,\Gamma(2-\alpha)]$, alors $\sum_{k=1}^\infty \zeta_k = 1$ et $\sum_{k=1}^\infty k\zeta_k = 1/(\alpha-1)$, (équations (39) et (40) de [33]). Il existe donc une distribution de probabilité ζ, sur $\mathbb{N}$ qui correspond à $(\zeta_k)_{k=1}^\infty$, de moyenne $1/(\alpha-1)$. Ainsi, pour des temps petits, quand le nombre de blocs est grand, les sauts successifs de $(N(t), t \geq 0)$ sont approximativement i.i.d. de loi ζ. Un argument de renouvellement nous permet de montrer le résultat suivant :

Théorème 8 (B., Berestycki, Schweinsberg [21]) *Soit Λ une mesure finie sur $[0,1]$ satisfaisant les conditions du Théorème 5. Soit $(\Pi(t), t \geq 0)$ le Λ-coalescent et $N(t)$ le nombre de blocs de $\Pi(t)$. On appelle H_n l'événement $N(t) = n$ pour un certain t. Alors*

$$\lim_{n\to\infty} P(H_n) = \alpha - 1.$$

On retrouve à nouveau le cas Kingman quand $\alpha \to 2$ puisque dans ce cas $\mathbb{P}(H_n) = 1$ pour tout n. Quand α diminue, le nombre de blocs moyen impliqués dans les coalescences augmente.

1.2.4 Taille des blocs

Considérons à présent la taille des blocs d'un Beta-coalescent.

Si $(\Pi(t), t \geq 0)$ est un Λ-coalescent, la suite ordonnée par taille décroissante des masses (i.e. des fréquences) de ses blocs, $(\Theta(t), t \geq 0)$ est ce que l'on appel un coalescents de masses ordonnées. En général, pour une mesure Λ donnée il n'y a pas de description simple de la distribution de $\Theta(t)$ pour t fixé. Le coalescent de Bolthausen-Sznitman est une exception, dans ce cas $\Theta(t)$ a la distribution de Poisson-Dirichlet avec paramètres $(e^{-t}, 0)$; voir [54, 117], ou [87] pour une preuve courte reposant sur les arbres récursifs. Le cas du coalescent de Kingman est un autre exemple. Si l'on note $T_k = \inf\{t : N(t) \leq k\}$ le premier temps auquel le coalescent a k blocs, alors la distribution de $\Theta(T_k)$ est uniforme sur le simplexe $\Delta_k = \{(x_1, \ldots, x_k) : x_1 \geq \cdots \geq x_k, x_1 + \cdots + x_k = 1\}$, (cf. [98]).

Le Théorème 9 ci-dessous donne un résultat du même type pour les Beta$(2-\alpha, \alpha)$-coalescents avec $1 < \alpha < 2$. Il est bien sûr possible -à l'inverse de ce qui se passe pour un coalescent de Kingman- que ce coalescent ne passe jamais par un état dans lequel il a k blocs. Nous calculons donc la distribution de $\Theta(T_k)$ conditionnellement à l'événement $N(T_k) = k$, i.e. il existe un temps auquel le coalescent a exactement k blocs.

Théorème 9 (B., Berestycki, Schweinsberg [21]) *Soit $(\Pi(t), t \geq 0)$ le Beta$(2-\alpha, \alpha)$-coalescent, où $1 < \alpha < 2$, et appelons $(\Theta(t), t \geq 0)$ le coalescent de masses ordonnées associé. Soit $N(t)$ le nombre de blocs de $\Pi(t)$ au temps t. Soit k un entier fixé, et soit $X'_1, \ldots, X'_k$ des variables aléatoires i.i.d. de distribution μ, où la transformée de Laplace de μest donnée par*

$$\int_0^\infty e^{-\lambda x}\, \mu(dx) = 1 - (1 + \lambda^{1-\alpha})^{-1/(\alpha-1)}. \tag{1.15}$$

Soit $X_1, \ldots, X_k$ le réarrangement par ordre décroissant des $X'_1, \ldots, X'_k$. Soit $S_k = X_1 + \cdots + X_k$. Si $g : \Delta_k \to [0, \infty)$ est une fonction non-négative mesurable, alors

$$E[g(\Theta(T_k)) | N(T_k) = k] = \frac{1}{E[S_k^{1-\alpha}]} E\left[S_k^{1-\alpha} g\left(\frac{X_1}{S_k}, \ldots, \frac{X_k}{S_k}\right)\right]. \tag{1.16}$$

Afin de mieux cerner le lien de ce résultat avec le cas du coalescent de Kingman, on peut remarquer que si $\alpha = 2$ le membre de droite de (1.15) devient $1/(1 + \lambda)$, et ainsi μ est simplement la distribution d'une variable exponentielle de moyenne 1. Si $X_1, \ldots, X_k$ sont obtenus en ordonant k variables i.i.d. exponentielles de moyenne 1 et $S_k = X_1 + \cdots + X_k$, alors S_k est indépendante de $(X_1/S_k, \ldots, X_k/S_k)$. En conséquence, le membre de droite de (1.16) devient $E[g(X_1/S_k, \ldots, X_k/S_k)]$. De plus la distribution de $(X_1/S_k, \ldots, X_k/S_k)$ est alors uniforme sur Δ_k. Enfin, la partition $\Theta(T_k)$ est indépendante de la date T_k pour le coalescent de Kingman puisque exactement deux blocs coalescent à chaque événement. Cette propriété n'est bien sur pas vraie pour d'autres coalescents.

Remarque 10 *La distribution μ apparaît dans d'autres contextes. Slack [131] est semble-t-il le premier à l'avoir étudié pour décrire les tailles des processus de Galton-Watson critiques, conditionnés à survivre, quand la loi de reproduction est à queue épaisse. Plus précisément, Yaglom [136, 100] montre que pour un processus de Galton-Watson critique de variance finie, Z_n la taille de la population à la génération n conditionnée à être positive et normalisée pour avoir une moyenne égale à 1 converge en loi vers une variable exponentielle de moyenne 1 lorsque $n \to \infty$. Quand la loi de reproduction est dans le bassin d'attraction d'une loi stable d'indice $\alpha \in (1,2)$ (donc de moyenne finie mais sans second moment), Slack a montré que Z_n conditionné à $Z_n > 0$ converge en loi vers μ lorsque $n \to \infty$, donnant ainsi un analogue au théorème de Yaglom pour le cas des variances infinies.*

Il n'est pas surprenant de voir la loi de Slack intervenir ici. En effet, les processus de branchement continus α-stables, $\alpha \in (1,2)$ s'obtiennent comme la limite d'échelle des processus de Galton Watson dont la loi de reproduction est dans la bassin d'attraction d'une loi stable (voir [103, 65]). Or comme nous l'avons déjà mentionné, les Beta coalescent sont justement la généalogie des processus de branchement α-stables. Nous ne nous servons pas directement des résultats de Slack dans notre preuve, mais ceux-ci permettent de comprendre de façon intuitive une large part des résultats obtenus dans [21].

Considérons à présent les tailles des blocs aux temps petits. En prenant la dérivé en 0 du membre de droite de (1.15), on voit que $E[X_i'] = 1$ pour chaque i. On en déduit que S_k vaut à peu près k pour k grand. Notre théorème 9 suggère donc que quand il y a k blocs, la distribution de la taille de ces blocs devrait être celle de k variables i.i.d. de loi μ, chacune étant ensuite divisée par k. Le théorème 11 ci-dessous correspond à cette intuition. Notons que ce résultat est à mettre en parallèle d'un résultat similaire obtenu par Bertoin et Le Gall (cf. Théorème 4 de [33]). Le résultat de Bertoin et Le Gall s'applique à une famille de Λ-coalescents plus grande (tous les Λ-coalescents pour lesquels la mesure Λ varie régulièrement près de 0) mais ne démontre qu'une convergence en probabilité. Nous établissons ici une convergence presque sûre dans le cas des Beta-coalescents uniquement.

Théorème 11 (B., Berestycki, Schweinsberg [21]) *Soit $(\Pi(t), t \geq 0)$ le Beta$(2-\alpha, \alpha)$-coalescent, avec $1 < \alpha < 2$. Soit $N(t,x)$ le nombre de blocs de $\Pi(t)$ dont la fréquence asymptotique est au plus x. On pose $F(x) = \mu((0,x])$ pour tout x, où μ est la distribution de probabilité définie par (1.15). Alors*

$$\lim_{t \downarrow 0} \sup_{x \geq 0} \left| t^{1/(\alpha-1)} N(t, t^{1/(\alpha-1)} x) - (\alpha\Gamma(\alpha))^{1/(\alpha-1)} F((\alpha\Gamma(\alpha))^{1/(\alpha-1)} x) \right| = 0 \quad p.s.$$

On peut remarquer que la limite en $x \to 0$ permet de retrouver le résultat du Théorème 5 pour le cas des beta coalescents. Remarquons également que si $\alpha = 2$ et μ est la distribution exponentielle de moyenne 1, alors l'expression $(\alpha\Gamma(\alpha))^{1/(\alpha-1)} F((\alpha\Gamma(\alpha))^{1/(\alpha-1)} x)$

devient $2(1-e^{-2x})$, et, comme observé dans [33],on retrouve un résultat connu pour le coalescent de Kingman (cf. section 4.2 de [6]).

Nous concluons cette section par un résultat qui décrit de façon précise le comportement asymptotique de la taille d'un bloc typique et du plus grands bloc respectivement. Par échangeabilité, la taille du bloc qui contient 1 est un tirage biaisé par la taille parmi les fréquences asymptotiques de tous les blocs présent, c'est ce que nous entendons par "bloc typique".

Proposition 12 (B., Berestycki, Schweinsberg [21]) *Soit* $(\Pi(t), t \geq 0)$ *le Beta*$(2-\alpha, \alpha)$*-coalescent, où* $1 < \alpha < 2$. *Soit* $K(t)$ *la fréquence asymptotique du bloc de* $\Pi(t)$ *qui contient* 1. *Alors*

$$(\alpha\Gamma(\alpha))^{1/(\alpha-1)} t^{-1/(\alpha-1)} K(t) \to_d X \qquad \textit{quand } t \downarrow 0,$$

où $E[e^{-\lambda X}] = (1+\lambda^{\alpha-1})^{-\alpha/(\alpha-1)}$.

Soit $W(t)$ *la taille du plus gros bloc de* $\Pi(t)$. *On a*

$$(\alpha\Gamma(\alpha)\Gamma(2-\alpha))^{1/\alpha} t^{-1/\alpha} W(t) \to_d X \qquad \textit{quand } t \downarrow 0,$$

où X *à la distribution de Fréchet d'indice* α *(i.e* $P(X \leq x) = e^{-x^{-\alpha}}$ *pour tout* $x > 0$*).*

La loi de la variable X est une version biaisée par la taille de μ, i.e. $P(X \in dx) = x\mu(dx)$ et on peut voir aisément que $E(X) = \infty$. Le résultat sur le bloc typique est une application du Théorème 11 tandis que le comportement du plus grand bloc s'étudie grâce à un théorème Taubérien qui donne des informations sur la queue de μ et des résultats de type "valeurs extrêmes". La proposition 12 montre donc que presque tous les fragments naissent et émergent de la poussière originelle en croissant comme $t^{1/(\alpha-1)}$ mais qu'il existe des blocs qui émergent avec une taille bien supérieure d'ordre $t^{1/\alpha}$.

1.2.5 Spectre multifractal de coalescence

La proposition 12 suggère donc d'étudier les fragments qui émergent de la poussière avec un taux atypique $\gamma \neq 1/(\alpha-1)$. À cette fin il est commode de considérer un espace métrique aléatoire (S, d) qui encode complètement le coalescent Π. Cet espace a été introduit par Evans [78] pour le cas du coalescent de Kingman. Il s'agit d'une nouvelle instance d'une idée plus générale qui consiste à voir les arbres continus comme des espaces métriques aléatoires. L'espace (S, d) se construit comme la complétion de $(\mathbb{N}, d)$ où $d(i, j)$ est le premier instant auquel i et j appartiennent au même bloc (leur temps de coalescence). Lorsque l'on complète $\{1, 2, \ldots\}$ de cette façon, on ajoute des points qui justement sont ceux qui se comportent de façon atypique. Dans ce contexte, nous sommes en mesure d'associer à chaque point $x \in S$ et pour tout $t \geq 0$ un nombre réel $\eta(x, t)$ qui est égal à la fréquence asymptotique de l'unique

bloc qui au temps t correspond au point x (ceci définit formellement une mesure η sur S que l'on nomme parfois *mesure de masse* telle que $\eta(x,t)$ est la masse de la boule ouverte de centre x et de rayon t). On peut maintenant formuler précisément le problème : existe-t-il des points $x \in S$ tels que la taille du bloc $B_x(t)$ qui contient x au temps t se comporte comme t^γ quand $t \to 0$, ou plus formellement tels que $\eta(x,t) \asymp t^\gamma$? (Ici $f(t) \asymp g(t)$ signifie que $\log f(t)/\log g(t) \to 1$). S'il en existe, peut on mesurer la quantité de tels points ?

Pour $\gamma \leq 1/(\alpha-1)$ on définit

$$S(\gamma) = \{x \in S : \liminf_{t\to 0} \frac{\log(\eta(x,t))}{\log t} \leq \gamma\}.$$

De même, quand $\gamma > 1/(\alpha-1)$ on pose

$$S(\gamma) = \{x \in S : \limsup_{t\to 0} \frac{\log(\eta(x,t))}{\log t} \geq \gamma\}.$$

Lorsque $\gamma \leq 1/(\alpha-1)$, $S(\gamma)$ est l'ensemble des points qui correspondent à de grands fragments. Au contraire quand $\gamma \geq 1/(\alpha-1)$, $S(\gamma)$ est l'ensemble des points qui appartiennent à des fragments asymptotiquement petits. Le Théorème suivant répond aux questions posées ci-dessus en calculant la *dimension de Hausdorff* (par rapport à la métrique de S) de l'ensemble $S(\gamma)$.

Théorème 13 (B., Berestycki, Schweinsberg [22])

1. *Si* $\frac{1}{\alpha} \leq \gamma < \frac{1}{\alpha-1}$ *alors*
$$\dim_{\mathcal{H}} S(\gamma) = \gamma\alpha - 1.$$
Si $\gamma < 1/\alpha$ *alors* $S(\gamma) = \emptyset$ *p.s. mais* $S(1/\alpha) \neq \emptyset$ *presque sûrement.*

2. *Si* $\frac{1}{\alpha-1} < \gamma \leq \frac{\alpha}{(\alpha-1)^2}$ *alors*
$$\dim_{\mathcal{H}} S(\gamma) = \frac{\alpha}{\gamma(\alpha-1)^2} - 1.$$
Si $\gamma > \frac{\alpha}{(\alpha-1)^2}$ *alors* $S(\gamma) = \emptyset$ *p.s. mais* $S(\frac{\alpha}{(\alpha-1)^2}) \neq \emptyset$ *presque sûrement.*

La valeur maximale de $\dim_{\mathcal{H}} S(\gamma)$ est obtenue quand $\gamma = 1/(\alpha-1)$ auquel cas la dimension de $S(\gamma)$ est aussi égale $1/(\alpha-1)$. On pouvait s'attendre à un tel résultat puisque il s'agit de l'exposant typique pour la taille des blocs. La valeur de la dimension correspond alors à la dimension totale de l'espace S.

On remarque également que l'on retrouve le résultat sur la taille du plus grand bloc puisque $1/\alpha$ est la plus petite valeur pour laquelle $S(\gamma) \neq \emptyset$. Il est intéressant de constater que l'on peut aussi avoir un *plus petit bloc* dont la taille est d'ordre t^γ avec $\gamma = \alpha/(\alpha-1)^2$.

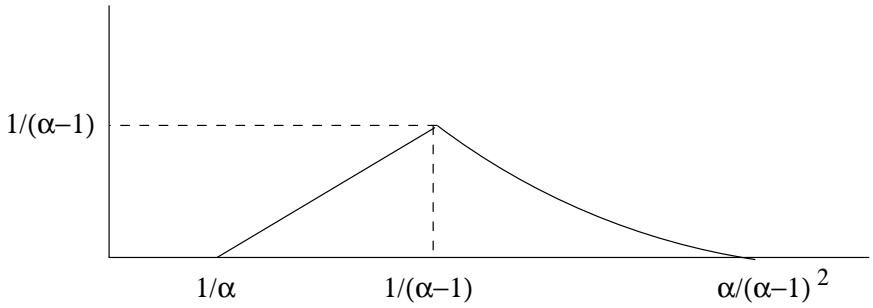

FIGURE 1.2: Spectre Multifractal $\gamma \mapsto \dim_{\mathcal{H}} S(\gamma)$. La dérivée à gauche du point critique est α tandis que la dérivée à droite est $-\alpha$.

Remarque 14 *Il est intéressant de noter la proximité de ces résultats avec l'un de mes travaux de thèse [14], dans lequel je m'étais intéressé au comportement en temps long des fragmentations homogènes. Plus précisément, je montrais dans cet article que si $F(t)$ est une fragmentation homogène de l'intervalle $(0,1)$ et que l'on note $I_x(t)$ le fragment qui contient x au temps t, alors il existe un nombre v_0 -la vitesse typique de fragmentation- tel que si U est une variable uniforme sur $(0,1)$ alors presque sûrement $|I_U(t)| \sim e^{-v_0 t}$. Pour $v \neq v_0$ dans un certain intervalle, l'ensemble aléatoire de points exceptionnels $S(v) := \{x \in (0,1) : |I_x(t)| \sim e^{-vt}\}$ est non-vide et de mesure de Lebesgue nulle. Le résultat principal de [14] donne une formule explicite pour le spectre multifractal $v \mapsto \dim_{\mathcal{H}}(S(v))$ où $\dim_{\mathcal{H}}(S)$ est la dimension de Hausdorff de S. Une différence importante est que dans [14] la dimension de Hausdorff des ensembles de points exceptionnels est calculée par rapport à la métrique induite par la mesure de Lebesgue sur $(0,1)$. Dans ce cas, le fait que le diamètre d'un bloc coïncide avec sa masse joue un rôle important. Ici, en revanche, les dimensions sont calculées par rapport à la métrique d qui s'interprète comme une distance généalogique.*

Chapitre 2

Coalescents, processus de branchement et arbres continus

L'essentiel des résultats sur le comportement asymptotique des coalescents présentés au chapitre précédente -nombre de blocs, taille des blocs, spectre multifractal- s'obtiennent grâce aux liens profonds qui unissent coalescents et processus de branchement à espace d'état continu. Nous présentons dans cette section les contributions à l'étude de ces liens que mes collaborateurs et moi-même avons apporté dans les articles [22, 20, 16]. Je commencerai par rappeler les définitions des objets que nous utiliserons ainsi qu'une brève descriptions des résultats contenus dans les articles de Birkner et al. [50] et de Bertoin et Le Gall [27].

2.1 Définitions et premières propriétés

2.1.1 Processus de branchement à espace d'état continu

Les processus de branchement à espace d'état continu peuvent être vu comme les analogues en temps et en espace continus des processus de Galton-Watson. Ils peuvent d'ailleurs être obtenus comme la limite d'échelle de processus de Galton-Watson (cf. [103]) et modélisent l'évolution de la taille d'une grande population dans laquelle on peut ignorer les interactions entre individus. Leur étude remonte à Lamperti [103]. Parmi les travaux classique les concernant citons également Grey [88] qui étudie leur comportement asymptotique et plus récemment les travaux de Le Gall et Duquesne [63] que nous détaillerons plus loin.

Formellement,

Définition 15 *On appelle processus de branchement à espace d'état continu (ou CSBP pour faire court) tout processus Markovien* $(Z_t, t \geq 0)$ *à valeurs dans* $[0, \infty]$, *dont le noyau de transition* $Q_t(x, \cdot) = P(Z_t \in \cdot | Z_0 = x)$ *vérifie*

$$Q_t(x + y, \cdot) = Q_t(x, \cdot) * Q_t(y, \cdot), \ \forall x, y \in \mathbb{R}_+. \tag{2.1}$$

Il faut penser à $Z_t \geq 0$ comme à la taille d'une population au temps t. L'équation (2.1) est simplement la propriété de branchement qui assure que Z issu de $x+y$ a la même loi que la somme de deux copies indépendantes de Z, l'une issue de x et l'autre de y.

Cette propriété s'apparente bien évidemment à l'infini divisibilité et en effet, Lamperti [103] a montré que la loi d'un CSBP est entièrement caractérisée par une fonction $\psi : [0,\infty) \to \mathbb{R}$, appelée *mécanisme de branchement*. Plus précisément, la transformée de Laplace de Z est donnée par :

$$\mathbb{E}[e^{-\lambda Z_t}|Z_0 = a] = e^{-au_t(\lambda)}, \ \forall t \geq 0, \tag{2.2}$$

où la fonction $u_t(\lambda)$ est l'unique solution de l'équation différentielle

$$\frac{\partial u_t(\lambda)}{\partial t} = -\psi(u_t(\lambda)), \qquad u_0(\lambda) = \lambda, \tag{2.3}$$

et où ψ est donné par :

$$\psi(q) = aq + \frac{b}{2}q^2 + \int_0^\infty (e^{-qx} - 1 + qx1_{x<1})\sigma(dx), \ \forall q \geq 0. \tag{2.4}$$

Ici, σ est une mesure σ-finie dont le support est inclus dans $[0,\infty)$, et telle que $\sigma(\{0\}) = 0$. En outre, σ satisfait la condition d'intégrabilité $\int (r \wedge r^2)\sigma(dr) < \infty$, et les paramètres $a \in \mathbb{R}$ et $b \geq 0$ sont appelés respectivement dérive et coefficient Gaussien de ψ. Le branchement est dit *critique* si $\psi'(0) = 0$. Quand $\mathrm{supp}(\sigma) \subset [0,1]$, alors $\psi'(0) = a$, et la criticalité est équivalente à la condition $a = 0$.

Les CSBP que nous allons considérer sont ceux qui correspondent aux Λ coalescents à travers la relation (1.6)

$$\psi(q) = \int_0^1 (e^{-qx} - 1 + qx)x^{-2}\Lambda(dx).$$

Il est aisé de voir que ce type de mécanisme de branchement est toujours critique. En outre, si $\tilde{\Lambda}$ est une mesure de probabilité telle que $\tilde{\Lambda} = (1-c)\Lambda + c\delta_0$, où Λ n'a pas d'atome en 0, on peut réécrire

$$\psi_{\tilde{\Lambda}}(q) = \frac{c}{2}q^2 + (1-c)\int_{[0,1]} (e^{-qx} - 1 + qx)x^{-2}\Lambda(dx). \tag{2.5}$$

On voit donc que la masse atomique de Λ en 0 correspond au coefficient Gaussien de ψ. Si $c = 1$ on retrouve le coalescent de Kingman, on supposera donc $c < 1$ dorénavant.

Rappelons la définition de la fonction v (cf. (1.8)) : lorsque Λ est telle que le coalescent descend de l'infini on peut définir

$$u_\psi(t) := \int_t^\infty \frac{dq}{\psi(q)} \in (0,\infty), \ t > 0,$$

et son inverse càdlàg

$$v_\psi(t) := \inf\{s > 0 : \int_s^\infty \frac{1}{\psi(q)} dq < t\},\ t > 0.$$

On écrira $u(\cdot) = u_\psi(\cdot)$ et $v(\cdot) = v_\psi(\cdot)$ quand il ne sera pas nécessaire de distinguer entre plusieurs mécanismes ψ possibles.

2.1.2 Le changement de temps de Lamperti

Les fonction ψ du type (2.4) sont aussi les exposant de Laplace de certains processus de Lévy spectralement positifs (i.e. sans sauts négatifs). Lamperti [103] et Silverstein [129] ont démontré qu'il existait une relation trajectorielle entre le CSBP $(Z_t, t \geq 0)$ et le processus de Lévy $(X_t, t \geq 0)$ associés à la même fonction ψ. On peut énoncer leur résultat de la manière suivante : soit X un processus de Lévy d'exposant ψ et tel que $X_0 = x > 0$, et soit $\eta := \inf\{t \geq 0 : X_t = 0\}$ le temps d'atteinte de 0 pour X, avec la convention habituelle que $\eta = \infty$ si $X_t > 0$ pour tout $t > 0$. Quand le branchement est critique il est aisé de montrer que $\eta < \infty$ presque sûrement Pour tout $t \geq 0$ définissons

$$U_t := \int_0^{\eta \wedge t} \frac{ds}{X_s},$$

et écrivons $U_t^{-1} = \inf\{s \geq 0 : U_s > t\}$ pour désigner l'inverse càdlàg de U_t. Alors le processus changé de temps $(X_{U_t^{-1}}, t \geq 0)$ a même loi que $(Z_t, t \geq 0)$, où Z est le CSBP de mécanisme de branchement ψ.

2.1.3 Les conditions de Grey

Soit $(Z_t, t \geq 0)$ un CSBP de mécanisme de branchement ψ (2.4). Appelons $\tau_0 = \inf\{t > 0 : Z_t = 0\}$ le temps d'extinction, et $\tau_\infty = \inf\{t > 0 : Z_t = \infty\}$ l'instant d'explosion de Z. Grey [88] a montré que $\tau_\infty = \infty$ presque sûrement si et seulement si il existe $x > 0$ tel que $\int_0^x dq/\psi(q) = \infty$. Il y a un critère analogue pour la probabilité d'extinction de Z :

$$p = P(\tau_0 < \infty) > 0 \iff \int_x^\infty \frac{dq}{\psi(q)} < \infty, \text{ pour un certain } x > 0. \tag{2.6}$$

De plus dans les cas critiques ou sous-critiques (i.e. quand $\psi'(0) \leq 0$), on a $p \in \{0, 1\}$. Dans le cas surcritique on a toujours $p < 1$, ce qui correspond au comportement de la probabilité d'extinction pour les processus de Galton-Watson.

Lorsque $\psi = \psi_\Lambda$ comme dans (1.6), il est aisé de montrer que $P(\tau_\infty = \infty) = 1$. En outre, la criticalité de ψ_Λ implique que si l'intégrale dans (2.6) est finie pour un certain (et dans ce cas en fait pour tout) $x > 0$, alors $p = 1$ et dans le cas contraire $p = 0$.

2.2 Généalogie des CSBP

Comme je l'ai indiqué plus haut, les CSBP peuvent être obtenus comme des limites d'échelle de processus de Galton Watson. Plus précisément, si $(Z_t, t \geq 0)$ est un ψ-CSBP où l'on suppose que ψ est sous-critique. Lamperti [103] a montré qu'il existe une suite de lois de reproduction μ_n telle que si l'on considère $(Z_k^n, k = 1, 2, \ldots)$ un processus de Galton-Watson discret de loi de reproduction μ_n et issu de n ancêtres, alors $(n^{-1} Z^n_{\gamma_n t}, t \geq 0)$ converge au sens des marginales fini-dimensionnelles vers $(Z_t, t \geq 0)$, où les γ_n sont des constantes d'échelle de temps bien choisies. Cependant, en passant à la limite on perd la généalogie sous-jacente au processus de Galton-Watson. Dans cette partie je vais présenter très succinctement trois approches qui ont été proposées pour retrouver cette généalogie sous-jacente au CSBP. La première est due à Le Jan et Le Gall puis à Duquesne et Le Gall [104, 63] et s'appuie sur la construction d'un *processus de hauteur* qui joue le même rôle que le processus de contour pour les arbres discrets. La seconde consiste à construire un système de particules dont la mesure empirique décrit la composition de la population du CSBP et repose essentiellement sur des travaux de Donnelly et Kurtz [58, 59]. Enfin, on peut également utiliser une approche due à Bertoin et Le Gall [27] et qui est fondée sur la subordination de Bochner. Nous conclurons en présentant l'un des principaux résultats obtenus dans [22] qui réconcilie l'approche *processus de hauteur* et l'approche *système de particules* de Donnelly et Kurtz.

2.2.1 Processus de hauteurs et arbres continus

Le Gall et Le Jan [104] ont introduit une nouvelle façon de concevoir la structure généalogique des CSBP qui a ensuite été explorée et développée par Duquesne et Le Gall [63]. Cette approche s'inspire d'un résultat bien connu de Ray et Knight concernant les temps locaux du mouvement Brownien et est reliée à la construction de l'arbre continu Brownien d'Aldous [3, 4].

Pour décrire brièvement le résultat de Ray et Knight, on introduit le processus $(L_t^x, t \geq 0, x \geq 0)$ où L_t^x est le temps local au niveau x accumulé par le mouvement Brownien réfléchi B sur l'intervalle de temps $[0, t]$. Il est connu que l'on peut choisir une version bicontinue du processus $(L_t^x, t \geq 0, x \geq 0)$ ce que je suppose fait dans la suite. On introduit également $r \mapsto T_r$ l'inverse càdlàg de $t \mapsto L_t^0$ (i.e. T_r est le premier instant où B a accumulé r unités de temps local en 0). Ray et Knight prouvent alors que pour $r > 0$ fixé, le processus $(L_{T_r}^x, x \geq 0)$ est une diffusion de Feller, c'est-à-dire un CSBP de mécanisme de branchement $\psi(q) = q^2$, issue d'une population initiale r.

Les travaux de Le Gall et Le Jan permettent justement de généraliser ce type de construction (i.e. où le CSBP s'obtient comme le temps local d'un certain processus à valeur dans $\mathbb{R}_+$) à des mécanismes de branchement critiques ou sous-critiques. Le processus qui joue le rôle du mouvement Brownien réfléchi dans un théorème de Ray-Knight sera noté $(H_t, t \geq 0)$.

Plus précisément, soit (μ_n) la suite de loi de reproduction telle que $(n^{-1}Z^n_{\gamma_n t}, t \geq 0)$ converge vers le CSBP Z, Duquesne et Le Gall ont montré que le processus de hauteur discrets $(H^n_k, k = 0, 1, \ldots)$, où H^n_k est la génération du $k$ième individu dans le processus de Galton Watson de loi de reproduction μ_n, converge une fois bien renormalisé vers un processus $(H_t, t \geq 0)$ appelé processus de hauteur. On peut également construire le processus $(H_t, t \geq 0)$ à partir du processus de Lévy $(X_t, t \geq 0)$ d'exposant ψ. Informellement, le processus de hauteur joue le même rôle pour un arbre continu que le processus de recherche en profondeur (*depth-first search*) pour un arbre discret. Bien que le processus H ne soit en général ni une semi-martingale ni un processus de Markov, l'un résultats clés obtenu par Duquesne et Le Gall [63] assure l'existence d'une version bicontinue (dés que le mécanisme de branchement ψ est critique ou sous-critique) du temps local $(L^x_s, s \geq 0, x \geq 0)$ telle que pour tout $t \geq 0$,

$$\lim_{\epsilon \to 0} E\left[\sup_{0 \leq s \leq t} \left| \frac{1}{\epsilon} \int_0^s 1_{\{x < H_r \leq x+\epsilon\}}\, dr - L^x_s \right|\right] = 0.$$

La généralisation du théorème de Ray-Knight s'énonce alors comme suit ([63, 104]). Soit $T_r = \inf\{t > 0, L^0_t > r\}$ l'inverse du temps local en 0. Pour tout $t \geq 0$, on définit

$$Z_t = L^t_{T_r}. \tag{2.7}$$

Le processus $(Z_t, t \geq 0)$ est un ψ-CSBP issu de $Z_0 = r$. Si $\psi(q) = q^2/2$, alors $(H_t, t \geq 0)$ est le mouvement Brownien réfléchi et $(Z_t, t \geq 0)$ est la diffusion de Feller, comme dans le théorème de Ray-Knight classique.

L'avantage de la construction avec le processus de hauteur H est que l'on a ainsi non-seulement le CSBP Z en prenant le temps local, mais en outre H code tout l'arbre d'évolution de la population.

On rappelle la définition d'un $\mathbb{R}$-arbre (ou arbre continu) associé avec une fonction positive H sur un intervalle $[0, T_r]$. Pour $u, v \in [0, T_r]$ posons $d_H(u, v) = H(u) + H(v) - 2\inf_{u \leq t \leq v} H(t)$, il est aisé de voir que d_H est une pseudo-distance sur $[0, T_r]$. On peut alors équiper l'espace $[0, T_r]$ quotienté par la relation $d_H(u, v) = 0$ de la distance d_H. L'espace métrique qui en résulte est un $\mathbb{R}$-arbre. Si la fonction H utilisée est le processus de hauteur de Duquesne et Le Gall $(H_s, s \leq T_r)$, ce que l'on obtient est plus précisément une collection Poissonienne d'arbres continus attachés par la racine.

2.2.2 Le processus *lookdown*

Le processus *lookdown* est un puissant outil introduit (puis développé) par Donnelly et Kurtz [58, 59] qui permet de coder la généalogie de superprocessus par des systèmes dénombrables de particules. Cette méthode offre une grande versatilité et s'adapte à de nombreux contextes. Je ne présenterai pas ici les objets dans leur plus grande généralité, en particulier je n'inclurai pas le cas où les particules peuvent avoir un mouvement dans l'espace des types.

Nous travaillons dans cette sous-partie dans le cadre suivant : soit ψ un mécanisme de branchement sans partie Gaussienne : il existe $a \in \mathbb{R}$ et une mesure de Lévy ν tels que

$$\psi(q) = aq + \int_0^\infty (e^{-qx} - 1 + qx\mathbf{1}_{x\leq 1})\nu(dx). \tag{2.8}$$

Au lieu d'utiliser ψ pour construire un CSBP avec ce mécanisme de branchement, on va commencer par construire un superprocessus M_t qui prend ses valeurs dans $\mathcal{M}_{(0,1)}$ l'espace des mesures finies sur $(0,1)$ et qui est défini à travers son générateur L. Celui-ci agit sur les fonctions $F : \mathcal{M}_{(0,1)} \mapsto \mathbb{R}$.

Définition 16 *On appelle super-CSBP de mécanisme de branchement ψ le processus $(M_t, t \geq 0)$ à valeurs dans $\mathcal{M}_{(0,1)}$ et dont le générateur L est donné par*

$$LF(\mu) = a\int_0^1 \mu(dx)F'(\mu,x) + \int_0^1 \mu(dx)\int_0^\infty \nu(dh)\big(F(\mu+h\delta_x) - F(\mu) - \mathbf{1}_{h\leq 1}hF'(\mu,x)\big). \tag{2.9}$$

On appelle processus de composition du ψ-CSBP le processus normalisé

$$R_t(\cdot) = M_t(\cdot)/M_t([0,1]).$$

La notation $F'(\mu,x)$ signifie $\lim_{\varepsilon\to 0}\varepsilon^{-1}(F(\mu+\varepsilon\delta_x) - F(\mu))$ et correspond à une modification infinitésimale de F dans la direction δ_x. Si ψ avait eu un terme quadratique il y aurait un terme supplémentaire dans le générateur, cf. equation (1.15) de [50]. On remarque que pour chaque $0 < r < 1$,

$$Z_t = M_t([0,r])$$

est un ψ-CSBP issu de $M_0([0,r])$. En effet, si l'on applique le générateur à une fonction $F(\mu) = \varphi(z)$ où $z = \mu([0,r])$ on voit immédiatement que le générateur L_1du processus Z_t est

$$\begin{aligned} L_1\varphi(z) &= a\int_0^r \mu(dx)\varphi'(x) + \int_0^r \mu(dx)\int_0^\infty \nu(dh)\big(\varphi(z+h) - \varphi(z) - h\mathbf{1}_{h\leq 1}\varphi'(z)\big) \\ &= zL_2\varphi(z) \end{aligned}$$

or la seconde intégrale ne dépend pas de x et est égale à $L_2\varphi(z)$, où L_2 est le générateur du processus de Lévy d'exposant $\psi(q)$. On conclut par le changement de temps de Lamperti.

L'interprétation de M_t est la suivante : on imagine que la population mesurée par Z_t est un continuum où chaque individu est pourvu d'un type - un nombre réel distinct entre 0 et 1 (et les individus transmettent leur type à leurs descendants). La quantité $M_t([0,a])$ est alors le nombre total d'individus au temps t qui descendent d'ancêtres dont le type au temps 0 était entre 0 et a.

Pour chaque t, R_t est une distribution de probabilité sur $(0,1)$ qui décrit la composition de la population à l'instant t. Typiquement, au moins dans le cas sous-critique,

dés que $t > 0, R_t$ est une somme pondérée de masses de Dirac $\sum_i \rho_i \delta_{x_i}$ telle que $\sum_i \rho_i = 1$. Chaque atome correspond à un groupe d'individus dans la population au temps t (représentant une proportion ρ_i de la population) qui descendent du même ancêtre au temps 0 (et dont le type était x_i).

Cette construction est à rapprocher des flots de subordinateurs de Bertoin et Le Gall [27]. Dans cet article, les auteurs montrent en effet que l'on peut définir les CSBP comme des processus de deux variables $(Z(t,a), t \geq 0, a \in [0,1])$ où t est le temps et a est la population initiale avec les propriétés suivantes : $Z(\cdot, 0) \equiv 0$ et $\forall a\, b \geq 0$ $(Z(\cdot, a+b) - Z(\cdot, a)$ est indépendant de la famille de processus $(Z(\cdot, c), 0 \leq c \leq a)$ et est un ψ-CSBP issu d'une population initiale de taille b. Si on appel M_t la mesure sur $[0,1]$ de fonction de répartition $a \mapsto Z(t,a)$ alors le processus M_t a pour générateur (2.9).

L'objet de la construction de Donnelly et Kurtz est de donner une représentation du processus de composition de la population R_t comme la limite de distributions empiriques associées avec un système dénombrable de particules. On aura ainsi accès à une notion claire de généalogie pour Z_t. La description qui suit s'inspire largement de [50] et [81, chapitre 5].

Pour définir le processus de *lookdown* (modifié) on se donne une suite d'individus $i = 1, 2, \ldots$ (ou de particules) qui sont identifiés par leur *type* $(\xi_i, i \in \mathbb{N})$. Initialement chaque individu i est de type $\xi_i(0)$ et les $\xi_i(0)$ sont de variables i.i.d. uniformes sur $(0,1)$. Pour chaque $t > 0, \xi_i(t)$ sera le type de i au temps t. Les variables $\xi_i(t)$ peuvent changer au cours du temps aux instants des *événements de naissance*. Supposons que l'on ait une configuration de points dans $[0,1] \times \mathbb{R}_+$:

$$n = \sum_i \delta_{(t_i, y_i)}$$

où $t_i \geq 0$ et $0 \leq y_i \leq 1$, et supposons que $\sum_{t_i \leq t} y_i^2 < \infty$ pour chaque $t \geq 0$. Plus loin un processus de points bien précis associé au CSBP sera spécifié.

Chaque atome (t_i, y_i) correspond à un événement de naissance et à cet instant t_i une proportion y_i des individus *participent* à l'événement. Plus précisément, chaque niveau tire à pile ou face indépendamment de tout le reste avec probabilité de succès y_i. Tout ceux pour lequel on a un succès participent à l'événement et les types des individus sont alors modifiés : supposons que les individus qui participent sont $1 \leq i_1 < i_2 < \ldots < i_k \leq n$. Au temps $t = t_i$, pour chaque $1 \leq j \leq k$, $\xi_{i_j}(t) = \xi_{i_1}(t-)$. En d'autres termes, les individus participants adoptent tous le type du plus petit d'entre eux. On ne détruit pas les types qui sont ainsi remplacés chez les individus $i_2, \ldots, i_k$ mais on attribut $\xi_{i_2}(t-)$ au premier individu (en montant) qui ne participe pas, et on continue ainsi à déplacer vers le haut les types vers le premier niveau de libre rencontré. Ce mécanisme est illustré dans la figure suivante.

Pour rendre cette construction rigoureuse il faut remarquer que comme par hypothèse $\sum_{t_i \leq t} y_i^2 < \infty$, seuls un nombre fini d'événements de naissance affectent les n premiers individus sur tout intervalle compact de temps. Les processus restreints

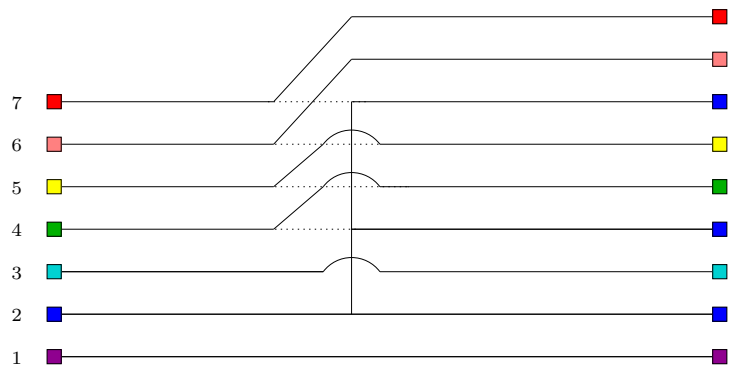

FIGURE 2.1: Représentation du processus lookdown. Les niveaux 2,4 et 5 participent à un événement de naissance. Les autres types génétiques sont poussés vers le haut. Les couleurs à gauche et à droite de la figure indiquent les types génétiques avant et après l'événement.

aux n premiers individus définis ainsi sont consistants et l'on a donc l'existence d'un processus $(\xi_i(t), t \geq 0, i = 1, 2, \ldots)$ bien défini grâce au théorème d'extension de Kolmogorov.

On a décrit la construction pour une configuration de point générale, on se restreint à présent au cas où (t_i, y_i) est donné par la construction suivante. Soit $Z_t(r)$ un ψ-CSBP où ψ a la forme (2.8), et où on a écrit la taille de la population initiale $r > 0$ comme argument de Z. Soit τ le temps d'extinction (qui n'est fini p.s. que dans le cas sous-critique, le cas important pour nous). Nous ne définissons le processus *lookdown* que jusqu'au temps $\tau-$. À chaque instant t_i où $\Delta Z_{t_i} > 0$, on associe $y_i = \Delta Z_{t_i}/Z_{t_i}$ (et l'on a donc $0 \leq y_i \leq 1$). Il est alors aisé de vérifier que si $t < \tau$,

$$\sum_{t_i \leq t} y_i^2 < \infty.$$

On peut en effet borner inférieurement Z_{t_i} par $I_t = \inf_{0 \leq s \leq t} Z_s > 0$, et ainsi cette somme est plus petite que

$$(I_t)^{-2} \sum_{t_i \leq t} (\Delta Z_{t_i})^2$$

Pour montrer que cette somme est infinie, on utilise le fait que Z est une changement de temps d'un processus de Lévy dont les sauts sont de carrés sommables (puisque $\int_0^\infty (1 \wedge x^2)\nu(dx) < \infty$). Ainsi, quand $t < \tau$ les sauts de Z sont les sauts du processus de Lévy durant un intervalle de temps aléatoire mais fini, ce qui prouve bien la convergence annoncée . On a donc l'existence d'un processus $(\xi_i(t), t \geq 0, i = 1, 2, \ldots)$ défini comme suit :

Définition 17 *Soit Z_t un ψ-CSBP. On appelle processus lookdown associé le processus $(\xi_i(t), t \geq 0, i = 1, 2, \ldots)$ construit selon la procédure lookdown modifiée décrite ci-dessus à l'aide de la configuration de points $(t_i, y_i = \Delta Z_{t_i}/Z_{t_i})$.*

On remarque que pour chaque $t \geq 0$, $(\xi_i(t), i = 1, 2, \ldots)$ est une suite échangeable et la limite

$$\rho_t = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{\infty} \delta_{\xi_i(t)}$$

est bien défini par le théorème de De Finetti.

L'intérêt de cette construction est expliqué par le résultat suivant que l'on peut trouver par exemple dans [50] (voir l'argument qui suit l'équation (2.15)).

Théorème 18 (Birkner et al. [50]) *Le processus* $(\rho_t, t \geq 0)$ *a même loi que le processus de composition* $(R_t, t \geq 0)$ *du* ψ*-CSBP avec comme distribution initiale la mesure uniforme sur* $(0, 1)$.

On peut comprendre heuristiquement ce résultat par l'argument suivant. Les sauts du CSBP $\Delta Z_t = x > 0$, sont causés par un individu dans la population qui se reproduit en créant soudain une grande quantité de descendance. La proportion d'individus qui porte le type de cet individu saute alors de $x/(Z_{t^-} + x) = \Delta Z_t / Z_t$ qui est exactement ce qui se produit dans le processus de *lookdown.*

On se restreint à présent au cas sous-critique. On suppose donc que ψ est une mécanisme de branchement comme dans (2.8) et que

$$\int_1^{\infty} \frac{dq}{\psi(q)} < \infty.$$

L'un des aspect intéressants de cette construction est que puisque Z_t s'éteint alors en temps fini, presque sûrement seul un nombre fini d'individus au temps 0 ont des descendants en vie au temps $t > 0$, ce qui signifie que la population au temps t est constituée d'un nombre fini de types, et qu'on atteint en temps fini l'instant ou il ne reste qu'un seul type en vie. Ceci est possible dans le processus *lookdown* modifié bien que l'on ne tue jamais de type. En raison de l'accumulation d'événements de naissance, il est possible de *pousser à l'infini* des types très rapidement. Ceux-ci disparaissent alors de la population visible.

2.2.3 Le *lookdown* dans le processus de hauteur

Je présente maintenant un résultat qui est au coeur de [22] et qui montre comment la construction *lookdown* de Donnelly et Kurtz peut être obtenue à partir du processus de hauteur de Duquesne et Le Gall. L'importance de ce résultat est en particulier due au fait qu'il a été montré dans [50] que, au sens de la construction lookdown, la généalogie du CSBP α-stable (i.e avec $\psi(q) = q^\alpha$) est un changement de temps du Beta$(2-\alpha, \alpha)$- coalescent. En montrant que la généalogie peut être prise au sens du processus des hauteur on s'autorise à travailler avec les outils puissants (temps locaux, théorie des excursions, etc...) pour obtenir de nouveaux résultats sur le coalescent.

Soit ψ un mécanisme de branchement critique ou sous-critique

$$\int_1^\infty \frac{dq}{\psi(q)} < \infty,$$

et H le processus de hauteur correspondant que nous avons introduit plus haut. Le processus H n'est en général pas Markovien mais, comme le mouvement Brownien réfléchi, c'est un processus d'excursions. Plus précisément, on peut associer à H une mesure infinie N sur un espace fonctionnel qui joue un rôle analogue à celui de la mesure d'Itô pour le cas du mouvement Brownien (cf. [120]). La propriété d'excursions de $(H_t, t \leq T_r)$ peut être formulée de la façon suivante. Soit $(g_i, d_i), i \in \mathcal{I}$ les intervalles d'excursion de H au dessus de zéro, de sorte que

$$\bigcup_{i\in\mathcal{I}} (g_i, d_i) = \{s \geq 0 : H_s > 0\}.$$

Pour chaque $i \in \mathcal{I}$, on définit la fonction e_i par $e_i(s) = H_{g_i+s}$ pour $0 \leq s \leq d_i - g_i$ et $e_i(s) = 0$ sinon. Une propriété essentielle du processus H, pour dire les choses de façon très schématiques, est que ses excursions forment, le long du temps local, un processus ponctuel de Poisson d'intensité donnée par une certaine mesure σ-fini sur l'espace fonctionnel $C_+([0, \infty))$.

Plus généralement, le processus H possède une telle propriété d'excursions au dessus de n'importe quel niveau $a > 0$ donné. Pour chaque $a > 0$, on appelle $(g_i^a, d_i^a), i \in \mathcal{I}^a$ les composantes connexes de l'ensemble ouvert $\{s : H_s > a\}$. Pour chaque $i \in \mathcal{I}^a$, on définit l'excursion $e_i^{(a)}$ par $e_i^{(a)}(s) = H_{g_i^a+s} - a$ quand $0 \leq s \leq d_i^a - g_i^a$ et $e_i^{(a)}(s) = 0$ sinon. Ces excursions $e_i^{(a)}$ sont également jetées le long du temps local au niveau a comme les atomes d'un processus de Poisson d'intensité la même mesure d'excursion. Ce résultat implique en particulier que le processus obtenu en rassemblant les excursiosn au dessus du niveau a a même loi que le procesus H lui même et est indépendant de ce qui se passe sous a.

Je peux à présent rappeler la construction du processus *lookdown* dans l'arbre continu que nous avions donné dans [22]. Soit $(Z_t, t \geq 0)$ un ψ-CSBP (ψ critique ou sous-critique) issu de $Z_0 = r > 0$, et supposons que Z_t est construit comme le temps local du processus de hauteur $(H_t, t \leq T_r)$ comme dans (2.7). Soit $\tilde{\xi} := ((\tilde{\xi}_j(t)), t \geq 0, j = 1, 2, \ldots)$ un processus *lookdown* construit à partir des sauts de $(Z_t, t \geq 0)$ comme dans la section précédente (i.e. à partir de la configuration de points $(t_i, \Delta Z_{t_i}/Z_{t_i})$).

Nous construisons à présent ξ une version du processus $\tilde{\xi}$ entièrement en termes du processus de hauteur $(H_t, t \leq T_r)$. Le point clé de cette construction consiste à choisir une numérotation bien particulière pour les excursions de H. Les excursions sont classées (et numérotées) selon l'ordre de leurs supremums respectifs. On notera $e_j^{(t)}$ la j-ième plus haute excursion au dessus du niveau t. On se donne une suite de variables i.i.d. $(U_i)_{i\in\mathbb{N}}$ uniformes sur $(0, 1)$ pour constituer la suites de types initiaux

dans le *lookdown*, $\xi_j(0) = U_j$ pour tout $j \geq 1$. Ainsi, pour tout $t > 0$, $\xi_j(t)$ est égal à l'un des U_i. Pour chaque $j \geq 1$ on note $k(j,t)$ l'unique entier tel que $e_j^{(t)}$ la j-ième plus haute excursion au dessus de t fait partie de $e_{k(j,t)}^{(0)}$, la $k(j,t)$-ième plus haute excursion au dessus de 0. On pose alors

$$\xi_j(t) = U_{k(j,t)}$$

et l'on dit que l'excursion $e_j^{(t)}$ possède le type $U_{k(j,t)}$.

Théorème 19 (B., Berestycki, Schweinsberg [22]) *Les processus ξ et $\tilde{\xi}$ ont même loi. Autrement dit, $((\xi_j(t)), t \geq 0, j = 1, 2, \ldots)$ a même loi que la construction du processus lookdown modifié associé au CSBP $(Z_t, t \geq 0)$.*

Voici une description de la dynamique du processus $(\xi_j(t), t \geq 0)$. Lorsque t augmente, les rangs relatifs des excursions au-dessus de t évoluent. Si $\Delta Z_t > 0$ alors avec probabilité 1 H possède une infinité de minima locaux en t, ce qui crée une infinité de nouvelles excursions au-dessus de t. En effet, par le Théorème 4.7 de [64], les sauts de Z_t correspondent toujours à une unique excursion au dessus de t^- qui se fragmente en une infinité de nouvelles excursions. En outre, tous les minimas locaux de $(H_t, t \geq 0)$ sont en fait associés à des sauts de Z_t (ce qui ne serait pas le cas si ψ avait un terme quadratique, cf. Théorème 4.7 de [64]). On dit alors qu'un événement de naissance se produit. On réordonne les excursions selon les nouveaux rangs relatifs (toujours donnés par les supremums). Les excursions qui ne fragmentent pas gardent leur type (mais peuvent changer de rang) et les excursions nouvellement créées adoptent le type de leur parent. Si l'excursion $e_j^{(t)}$ se fragmente alors cela signifie que plusieurs individus k avec $k \geq j$ adoptent le type $\xi_j(t)$. Ceux qui n'adoptent pas ce type sont poussés vers le haut en conséquence. On reconnaît la dynamique du *lookdown* décrite plus haut.

2.2.4 Le Beta-coalescent dans l'arbre stable

J'en viens à présent au résultat essentiel de [22] qui permet de prouver une grande part des résultats asymptotiques présentés plus hauts pour le cas du Beta$(2-\alpha, \alpha)$-coalescent uniquement.

Soit Z un CSBP α-stable (i.e. avec $\psi(q) = q^\alpha$) avec $1 < \alpha < 2$ obtenu selon la méthode de Duquesne et Le Gall comme le temps local du processus de hauteur $(H_t, 0 \leq t \leq T_r)$, c'est-à-dire $Z_t = L_{T_r}^t$. Pour chaque t on considère le niveau aléatoire

$$R_t = \alpha(\alpha-1)\Gamma(\alpha) \int_0^t Z_s^{1-\alpha}\, ds, \tag{2.10}$$

et on pose $R^{-1}(t) = \inf\{s : R_s > t\}$. Il est aisé de voir (cf. [50] par exemple) que $R^{-1}(t) < \infty$ p.s. pour tout t, et $\lim_{t\to\infty} R^{-1}(t) = \zeta$ où ζ est le temps d'extinction du CSBP.

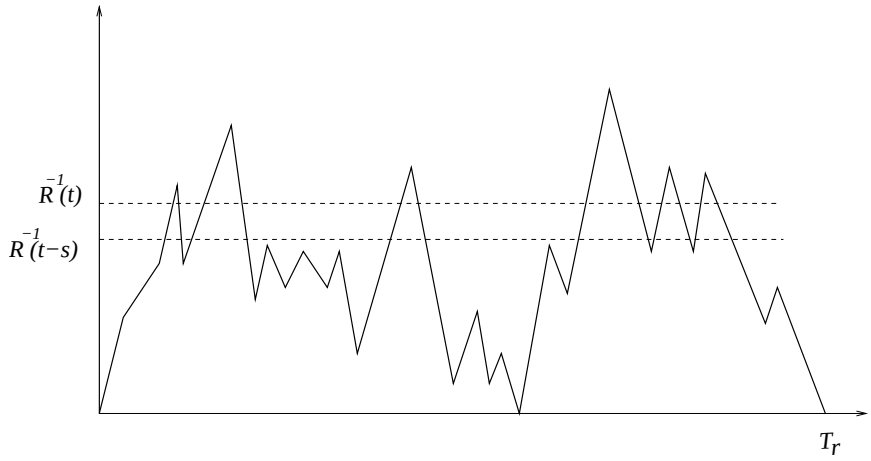

FIGURE 2.2: Un Beta-coalescent est obtenu en coalescent les excursions de $(H_t, t \leq T_r)$ au dessus de $R^{-1}(t-s)$ qui atteignent $R^{-1}(t)$. Ainsi, chaque excursion correspond à un bloc du coalescent et sa masse est donnée par le temps local de l'excursion correspondante au niveau $R^{-1}(t)$.

Soit à présent $(V_i, i = 1, 2, \ldots)$ une suite de variables dans $(0, T_r)$ telle que pour chaque $i \in \mathbb{N}$, V_i est le point le plus à gauche de la $i$ième plus haute excursion du processus H au niveau $R^{-1}(t)$. On peut ainsi définir un processus $(\Pi_s, 0 \leq s \leq t)$ à valeur dans $\mathcal{P}$ l'ensemble des partitions de $\mathbb{N}$ par :

$$i \overset{\Pi_s}{\sim} j \iff \left(\inf_{r \in [V_i, V_j]} H_r \right) > R^{-1}(t-s).$$

C'est-à-dire que i et j sont dans le même bloc de Π_s si et seulement si V_i et V_j sont dans la même excursion de H au dessus du niveau $R^{-1}(t-s)$.

Théorème 20 (B., Berestycki, Schweinsberg [22]) *Le processus* $(\Pi_s, 0 \leq s \leq t)$ *est un Beta*$(2-\alpha, \alpha)$*-coalescent qui vit jusqu'au temps* t.

Cette construction peut être décrite de la façon suivante. Quand s varie de 0 à t, le niveau $R^{-1}(t-s)$ décroît de $R^{-1}(t)$ jusqu'à 0. Les excursions de H au dessus du niveau $R^{-1}(t-s)$ coalescent car, si $s_1 < s_2$, alors plusieurs excursions de H au dessus de $R^{-1}(t-s_1)$ peuvent appartenir à la même excursion de H au dessus de $R^{-1}(t-s_2)$. Ceci se produit par exemple quand une excursion de H au dessus de $R^{-1}(t-s_1)$ a un minimum local au niveau $R^{-1}(t-s_2)$. Dans ce cas, dans le processus de coalescence correspondant on voit des blocs fusionner au temps s_2 dont les masses correspondent à la fraction de temps local au niveau $R^{-1}(t)$contenue dans chacune des excursions impliquées.

2.2.5 Beta-coalescents : Heuristique pour le nombre de blocs et le spectre multifractal

Les Théorèmes 5 et 13 du chapitre précédent concernent le cas des Beta-coalscents. Un élément clé dans la preuve de ces résultats est la notion *d'arbre réduit* associé avec le processus de hauteur H. Pour un niveau fixé a, l'arbre réduit au niveau a est l'arbre dont le nombre de branches à la hauteur $0 \leq t \leq 1$ est le nombre d'excursions de H au-dessus de t qui atteignent a. La structure généalogique est naturellement défini en disant que v est un ancêtre de w si l'excursion associée à w est contenue dans v. Duquesne et Le Gall [63] montrent que dans le cas α-stable, cet arbre est un simple changement de temps d'un arbre de Galton-Watson dont la loi de reproduction est explicite.

Lorsque l'on construit le Beta-coalescent à partir de l'arbre continu aléatoire ainsi que décrit dans le Théorème 20, le nombre de blocs $N(s)$ au temps s correspond au nombre d'excursions au dessus du niveau s' qui atteignent $R^{-1}(t)$, pour un certain s' et un certain t. On peut alors déduire le comportement asymptotique de $N(s)$ quand $s \to 0$ du comportement limite de l'arbre réduit quand $s' \to R^{-1}(t)$. Comme l'arbre réduit est un arbre de Galton-Watson, on sait que son comportement asymptotique est décrit par le Théorème de Kesten-Stigum, ce qui nous permet d'obtenir une preuve du Théorème 5. De même, le Théorème 13 s'obtient en établissant une correspondance entre le spectre multifractal de coalescence et le spectre multifractal de la mesure de branchement des arbres réduits (que l'on connaît en appliquant des résultats de Mörters et Shieh [113]).

Une difficulté technique importante à surmonter est qu'il faut montrer que les événements presque sûrs qui se produisent pour l'arbre réduit à un niveau fixé, sont également presque sûrs pour l'arbre réduit à un niveau aléatoire $R^{-1}(t)$ (ces arbres n'ont pas la même loi).

Notons encore une fois que l'approche martingale développée dans [19] permet d'obtenir des résultats plus précis et plus généraux.

2.3 Le cas des Λ-coalescents généraux et le lien avec les CSBP

Le Théorème 4 donne l'asymptotique du nombre de blocs pour tous les Λ-coalescents. Les résultats décrits ci-dessus qui s'appliquent uniquement aux Beta coalescents ou aux mesures Λ à variations régulières près de 0 ne suffisent donc pas.

Dans cette section, nous décrivons des résultats qui appartiennent à un article qui est en cours d'achèvement [20] et qui donne une explication probabiliste aux liens qui unissent le comportement en temps court des coalescents et des CSBP. Nous avons en effet vu que pour un Λ donné, $N(t)$ le nombre de blocs se comporte asymptotiquement comme $v = v_\psi(t)$ où v_ψ est donnée par (1.8). Le ψ-CSBP a un comportement très proche puisque si l'on note $N^Z(t)$ le nombre de familles en vie au

temps t (i.e. le nombre de types génétiques dans la construction *lookdown* au temps t ou le nombre d'excursions qui atteignent le niveau t dans la construction processus de hauteur) on a

$$(N^Z(t), t \geq 0) \stackrel{d}{=} (M(v(t)), t \geq 0)$$

où $t \mapsto M(t)$ est un processus de Poisson de taux 1. Bertoin et Le Gall avaient déjà remarqué dans [29] que le critère (1.5) de descente de l'infini pour un Λ-coalescent de Schweinsberg est analytiquement équivalent à (2.6).

L'explication probabiliste de cette proximité tient dans la remarque suivante : les Λ-coalescent ne décrivent pas toujours la généalogie des CSBP (en fait ceci n'est possible que pour les CSBP stables avec des Beta coalescents changés de temps), mais ils sont toujours la généalogie des *processus de Fleming-Viot généralisés*. Ces processus peuvent se construire d'une façon très semblable aux CSBP ce qui nous permet ainsi de coupler le CSBP et les Λ-coalescents.

2.3.1 Processus de Fleming-Viot généralisés et leur généalogie

Un processus $(\rho_t, t \geq 0)$ de Fleming-Viot généralisé - processus FVG- (non-spatialisé, i.e. en champs moyen) est un processus de Markov qui prend ses valeurs dans $\mathcal{M}_1([0,1])$ l'espace des mesures de probabilité sur $[0,1]$ et dont le générateur L est défini par

$$LF(\mu) = \int_{(0,1]} y^{-2}\Lambda(dy) \int_{[0,1]} \mu(dx) \left(F((1-y)\mu + y\delta_x) - F(\mu)\right), \tag{2.11}$$

où $F : \mathcal{M} \to \mathbb{R}$ est une fonction continue bornée. Autrement dit, on choisit un nombre y entre 0 et 1 au taux $y^{-2}\Lambda(dy)$. On tire un type x selon ρ_{t-} et l'on obtient ρ_t en transformant ρ_{t-} en pré-multipliant ρ_{t-} par $(1-y)$ puis en ajoutant une masse y d'individus de type x. On remarque que ρ_t reste toujours une mesure de probabilité.

En outre, Bertoin et Le Gall [29] montrent que les processus FVG sont caractérisés par la propriété suivante de leur générateur. Soit $p \in \mathbb{N}$ et soit f une fonction continue sur $[0,1]^p$. Considérons à présent la fonction G_f sur $\mathcal{M}$ définie par

$$G_f(\mu) = \int_{[0,1]^p} \mu(dx_1) \dots \mu(dx_p) f(x_1, \dots, x_p).$$

Le générateur L agit sur G_f comme suit :

$$LG_f(\mu) = \sum_{I \subset \{1,\dots,p\}, |I| \geq 2} \lambda_{p,|I|} \int \mu(dx_1) \dots \mu(dx_p)(f(R_I(x_1, \dots, x_p)) - f(x_1, \dots, x_p))$$

où $R_I(x_1, \dots, x_p) = (y_1, \dots, y_p)$ est défini par

$$y_j = x_{\min I} \text{ quand } j \in I \text{ et } y_j = x_j \text{ sinon,}$$

et les coefficients $\lambda_{p,|I|}$ sont donnés par (1.2).

Un processus de Fleming-Viot généralisé peut se construire pare la procédure *lookdown*. Le résultat suivant est le Lemme 3.6 de [50].

Lemme 21 *[50] Soit $n = \sum \delta_{(t_i,y_i)}$ un processus de Poisson d'intensité $dt \times y^{-2}\Lambda(dy)$ (par exemple le processus de points des sauts $(t, \Delta X_t)$ d'un processus de Lévy X d'exposant de Laplace ψ_λ convient). Le processus lookdown $(\Xi_t(\cdot), t \geq 0)$ construit avec n est un processus de Fleming-Viot Généralisé de générateur L donné par (2.11). En outre pour $T > 0$ fixé le processus de partition ancestrale $(\mathcal{R}^T(t), 0 \leq t \leq T)$ est un Λ-coalescent stoppé au temps T.*

Le processus de partition ancestrale $(\mathcal{R}^T(t), 0 \leq t \leq T)$ mentionné ci-dessus correspond à la relation d'équivalence $i \sim_{\mathcal{R}^T(t)} j$ si et seulement si les individus i et j au temps T dans le processus lookdown descendent du même individu au temps t.

2.3.2 Couplage

Au vu de ce dernier Lemme, il devient clair que l'on peut coupler Λ-coalescents et CSBP en les construisant à l'aide de deux processus *lookdown* couplés. Plus précisément. Soit X un processus de Lévy d'exposant de Laplace ψ. Comme on l'a vu, en utilisant les résultats de Lamperti [103] on peut changer de temps X par U_t^{-1} pour obtenir Z_t un ψ-CSBP

$$\forall t \geq 0,\ Z_t = X_{U^{-1}(t)},\ \text{où } U^{-1}(t) := \inf\left\{s > 0 : \int_0^s X_u^{-1} du > t\right\}.$$

Soit maintenant $n = \sum_{t,\Delta X_t}$ et $n' = \sum \delta_{t,\Delta Z_t/Z_t}$ deux processus de points couplés (i.e. Z et X sont comme ci-dessus). La construction *lookdown* avec n donne le processus de Fleming-Viot généralisé et si l'on utilise n' on obtient le ψ-super CSBP. On peut en outre utiliser une collection de variables unique pour décider des individus qui participent aux événements de naissance dans les deux cas. On appelle $(\xi_i(t), t \geq 0; i = 1, 2, \ldots)$ et $(\xi_i'(t), t \geq 0; i = 1, 2, \ldots)$ les deux processus *lookdown* ainsi construits (correspondant respectivement au processus de Fleming-Viot généralisé et au processus de la composition de la population d'un CSBP). Le nombre $N^X(t)$ de types distincts présents dans la collection $(\xi_i(t), t \geq 0; i = 1, 2, \ldots)$ a, par construction à même loi que $N^\Lambda(t)$ le nombre de blocs dans un Λ-coalescent au temps t. Le nombre $N^Z(t)$ de famille en vie pour le CSBP au temps t, i.e. le nombre de types distincts présents dans la collection $(\xi_i'(t), t \geq 0; i = 1, 2, \ldots)$, est une variable de Poisson de moyenne $v(t)$.

Le couplage permet de comprendre pourquoi $N^X(t)$ est proche de $N^Z(t)$ et donc de $v(t)$. Les atomes de $n' : (r, \Delta Z_r/Z_r)$ sont en bijection avec les atomes de $n : (s, \Delta X_s)$ par la transformation $(s, \Delta X_s) \mapsto (U^{-1}(s), \Delta X_s/X_s)$. Cette transformation préserve l'ordre des atomes, et comme $(X_t, t \geq 0)$ est p.s. continue en $t = 0$ on a $X_t \sim 1$, quand $t \to 0$. On en déduit que le changement de temps dans la première coordonnée est asymptotiquement proche de l'identité : $U_t^{-1} \sim t$ presque sûrement. Il s'ensuit que $\Delta Z_t/Z_t \approx \Delta X_t$ et que $\pi^X|_{(0,t)} \approx \pi^Z|_{(0,t)}$ si $t \to 0$ (j'utilise $\approx$ sans définir précisément ce symbole pour signifier "se comporte comme").

Ceci explique, au moins intuitivement, pourquoi $N^X(t) \approx N^Z(t)$. Comme $N^Z(t) \sim v(t)$ quand $t \approx 0$, la discussion ci-dessus permet de comprendre pourquoi la fonction v apparaît dans le Théorème 4. Comme nous l'avons déjà signalé, on ne peut cependant pas obtenir le résultat à partir des arguments de couplage présentés ici. On obtient en fait uniquement le résultat plus faible suivant :

Théorème 22 *[20] Pour chaque $\varepsilon \in (0,1)$,*

$$\lim_{t\to 0} \mathbb{P}\left[\frac{1}{(1+\varepsilon)^2}\cdot v\left(\frac{1+\varepsilon}{1-\varepsilon}t\right) \le N^\Lambda(t) \le \frac{1}{(1-\varepsilon)^2}\cdot v\left(\frac{1-\varepsilon}{1+\varepsilon}t\right)\right] = 1. \tag{2.12}$$

Notons que l'on peut également utiliser ce couplage pour proposer une nouvelle preuve simplifiée des résultats contenus dans [50].

La preuve du Théorème 2.3.2 produit également un certain nombre de sous-produits. On a par exemple

Corollaire 23 *Le Λ-coalescent descend de l'infini si et seulement si Z s'éteint en temps fini.*

Comme je l'ai déjà écrit plus haut, ceci avait déjà été remarqué par Bertoin et Le Gall mais uniquement de façon analytique. Le couplage donne une preuve directe et une explication probabiliste de ce fait (quand un CSBP ne s'éteint pas en temps fini $v \equiv \infty$ et le nombre de types distincts présents dans la collection $(\xi_i'(t), t \ge 0; i = 1, 2, \ldots)$ reste toujours infini).

On peut, au prix de certaines hypothèses sur la régularité de ψ (ou plus exactement sur les indices de régularité en 0 du processus de Lévy associé, cf. [119]) renforcer le Théorème et se rapprocher du Théorème 4. Je ne détaille pas ces résultats ici.

2.4 Le cas du coalescent de Kingman

Les résultats sur la taille des blocs et sur le spectre de coalescence que j'ai énoncé plus haut concernent uniquement le cas des Beta $(2-\alpha, \alpha)$-coalescents avec $\alpha \in (1,2)$. Le cas $\alpha = 2$ correspond au cas du coalescent de Kingman et on peut donc se demander quels sont les résultats qui "passent à la limite" quand $\alpha \nearrow 2$. Ces questions sont le point de départ de l'article [16] dans lequel nous examinons comment le coalescent de Kingman peut se construire dans une excursion Brownienne. Cette construction est très proche de l'esprit de travaux antérieurs de Aldous et Warren [5], [135], (et en un certain sens est déjà implicite dans ces articles), mais notre approche présente cependant des différences importantes qui nous permettent par exemple de déduire nos résultats sur le spectre de coalescence.

Nous décrivons à présent notre construction du coalescent de Kingman. Soit $(B_t, t \ge 0)$ une excursion Brownienne conditionnée à atteindre le niveau 1. Plus précisément, notons ν la mesure d'excursion de Itô sur l'espace des excursions continues $\Omega^* := \bigcup_{\zeta>0} \Omega_\zeta$, où

$$\Omega_\zeta := \{f : [0,\zeta] \to \mathbb{R} \text{ continue}, \ f(x) = 0 \iff x \in \{0,\zeta\}\}.$$

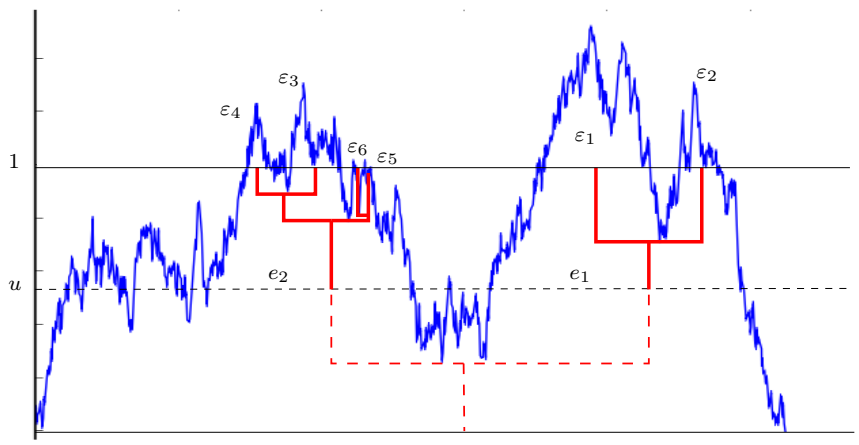

FIGURE 2.3: Le coalescent de Kingman dans l'excursion Brownienne. Ici nous avons : $\Pi_u = (\{1,2\},\{3,4,5,6\})$.

Soit alors $(B_t, 0 \leq t \leq \zeta)$ une réalisation de $\nu(\cdot|\sup_{s>0} B_s \geq 1)$. Avec un léger abus de notation on s'autorise à considérer B comme une fonction sur $[0,\infty)$ en déclarant qu'elle vaut $B(s) = 0$ pour tout $s \geq \zeta$. Soit $\{L(t,x)\}_{t\geq 0, x\geq 0}$ une version bicontinue du processus des temps locaux de B, et définissons

$$Z_x := L(\zeta, x), \quad x \geq 0.$$

Ainsi, Z_x est le temps local accumulé au niveau $x$par l'excursion $(B_s, s \geq 0)$. On définit à présent un processus $(\Pi_u, 0 \leq u \leq 1)$ à valeur dans $\mathcal{P}$ de la manière suivante. Soit $\{\varepsilon_i\}_{i=1}^{\infty}$ l'ensemble des excursions de B au dessus du niveau 1, ordonnées selon leurs hauteurs relatives. C'est-à-dire que

$$\sup_{s>0} \varepsilon_1(s) > \sup_{s>0} \varepsilon_2(s) > \dots$$

Fixons à présent $0 < u < 1$, et considérons l'ensemble des excursions $\{e_k\}_{1\leq k\leq N}$ de B au dessus de u qui atteignent le niveau 1, où $N = N(u)$ est le nombre de telles excursions. Pour chaque $i \geq 1$, ε_i appartient à exactement une excursion e_k pour un certain $1 \leq k \leq N$, et on appelle $\phi_u(i) = k \in \{1,\dots,N\}$ cet indice. On définit alors Π_u en déclarant que pour chaque $i,j \geq 1$, i et j sont dans le même bloc de Π_u si et seulement si $\phi_u(i) = \phi_u(j)$, c'est-à-dire si et seulement si ε_i et ε_j tombent dans la même excursion e_k pour un certain $k \leq N$. Le résultat principal de [16] est que Π_u est en fait un coalescent de Kingman changé de temps.

Théorème 24 (B., Berestycki [16]) *Le processus* $(\Pi_{U(t)}, t \geq 0)$ *a la même loi que le coalescent de Kingman, où pour chaque* $t > 0$,

$$U(t) = \sup\left\{ s > 0 : \int_s^1 \frac{4}{Z_u} du > t \right\}. \tag{2.13}$$

Le changement de temps $\{U(t), t \geq 0\}$ satisfait les propriétés suivantes : $U(0) = 1$, $\lim_{t\to\infty} U(t) = 0$, et U est continue et décroissante.

Il ne s'agit pas d'un résultat surprenant. Le lien entre coalescent de Kingman et excursions Browniennes avaient déjà été étudiés par plusieurs auteurs, et en particulier par Aldous [5] et Warren [135]. Ces derniers montrent comment la loi de l'arbre continu codé par une excursion Brownienne, conditionnellement au processus de temps local de celle-ci peut se décrire à l'aide d'un coalescent inhomogène construit à partir du coalescent de Kingman. Notre construction est cependant différente en ce que notre arbre est construit sous les noeuds à distance 1 de la racine (au lieu de noeuds choisis uniformément). En outre, il semble difficile d'utiliser les constructions de [6] et [135] pour en déduire directement des résultats sur le coalescent de Kingman.

Une différence cruciale entre ce résultat et le Théorème 19 est que pour les Beta coalescent dans [22] ce sont des excursions au dessus d'un niveau aléatoire $\mathcal{R}^{-1}(t)$ qui coalescent alors qu'ici ce sont des excursions au-dessus du niveau déterministe 1. Ceci explique pourquoi le Théorème ci-dessus ne peut pas s'étendre au cas des Beta-coalescent. Si tel était le cas les arbres de coalescence seraient des changement de temps d'arbres réduit pour le CSBP α-stables, or ces dernier sont des arbres de Galton-Watson (à un autre changement de temps près). Or il est possible de montrer (ce n'est pas trivial !) que la loi du nombre de blocs impliqués dans des coalescences successives durant la vie d'un Beta coalescent ne reste pas constante. L'arbre réduit ne peut donc pas être un Beta-coalescent.

Il est en fait possible d'énoncer un résultat analogue au théorème 19 pour le cas du coalescent de Kingman.

Notons à nouveau $(B_s, s \geq 0)$ un mouvement Brownien réfléchi et définissons $\tau_1 = \inf\{t > 0 : L(t, 0) > 1\}$. Soit $v > 0$ tel que $v < \sup_{s\leq\tau_1} B(s)$, et pour tout $0 \leq u \leq v$, définissons une partition Π_u^v comme ci-dessus où le niveau v remplace le niveau 1 (les excursions $(\varepsilon_i)_{i=1}^\infty$ sont donc des excursions au dessus de v et les excursions e_k sont les excursions au dessus de u qui atteignent v).

Pour chaque $t > 0$, posons

$$V(t) := \inf\left\{ s > 0 : \int_0^s 4Z_u^{-1} du > t \right\}$$

où $Z_x = L(\tau_1, x)$ pour chaque $x \geq 0$. (On observe que $V(t) < \sup_{s\leq\tau_1} B(s)$). Fixons $T > 0$ et pour $0 \leq t \leq T$, considérons la partition

$$\hat{\Pi}_t := \Pi_{V(T-t)}^{V(T)}.$$

Proposition 25 $(\hat{\Pi}_s, 0 \leq s \leq T)$ *a la même loi que* $(\kappa_s, 0 \leq s \leq T)$*, le coalescent de Kingman sur l'intervalle de temps* $[0, T]$.

Dans [22] nous avions utilisé le Théorème 19 pour prouver divers résultats concernant le comportement en temps petit des Beta coalescents. De même, ici on peut utiliser le Théorème 24 pour étudier les propriétés fines du coalescent de Kingman lorsqu'il descend de l'infini et que les blocs émergent de la poussière. J'illustre à présent cette stratégie en présentant le résultat sur le spectre multifractal du coalescent de Kingman obtenu dans [16]. Avec les mêmes notations que celles utilisées Section 1.2.5, la question du nombre de blocs qui émergent de la poussière avec un taux atypique t^γ revient à étudier le spectre multifractal local de la mesure de masse η sur l'espace de evans (S, d). Existe-t-il des points $x \in S$ tels que $\eta(B(x,t))$ se comporte comme t^γ quand $t \to 0$, et si oui, quelle est la dimension de Hausdorff de l'ensemble de ces points ? Pour $\gamma > 1$ on définit

$$S_{\text{thin}}(\gamma) = \left\{ x \in S : \limsup_{t \to 0} \frac{\log(\eta(x,t))}{\log t} = \gamma \right\}. \tag{2.14}$$

Cet ensemble correspond à des points de S avec une dimension locale $\eta(x,t)$ atypiquement petite. Pour les points épais, pour chaque $\theta \geq 0$ on considère

$$S_{\text{thick}}(\theta) = \left\{ x \in S : \limsup_{t \to 0} \frac{\eta(x,t)}{t|\log t|} = \theta \right\}. \tag{2.15}$$

Théorème 26 (B., Berestycki [16])

1. *Si* $0 \leq \theta \leq 1$ *Alors*
$$\dim S_{\text{thick}}(\theta) = 1 - \theta, \ \textit{p.s.}$$
Quand $\theta = 1$*,* $S_{\text{thick}}(\theta) \neq \emptyset$ *presque sûrement, mais si* $\theta > 1$ *Alors* $S_{\text{thick}}(\theta)$ *est presque sûrement vide.*

2. *Si* $1 < \gamma \leq 2$ *alors*
$$\dim S_{\text{thin}}(\gamma) = \frac{2}{\gamma} - 1, \ \textit{p.s.}$$
Si $\gamma > 2$ *alors* $S_{\text{thin}}(\gamma) = \emptyset$ *p.s. mais* $S_{\text{thin}}(2) \neq \emptyset$ *p.s..*

En comparant ce résultat au Théorème 13 on voit que seule la partie qui correspond aux "points fins" du spectre s'obtient en prenant une limite quand $\alpha \to 2$ des spectres des Beta coalescents. Sur une échelle puissance, la partie "points épais" du spectre est vide bien que la limite quand $\alpha \to 2$ existe et soit non-dégénérée. Ainsi, ici le plus petit bloc est d'ordre t^2 quand $t \to 0$ et le plus grand bloc est d'ordre $t \log(1/t)$. Dans le cas des Beta coalescent de paramètre $1 < \alpha < 2$, ces quantités sont respectivement $t^{\alpha/(\alpha-1)^2}$ et $t^{1/\alpha}$, qui ne coïncide pas avec $t \log(1/t)$ quand $\alpha \to 2^-$.

Chapitre 3

Formules d'échantillonage

Depuis son introduction en 1982 par Kingman [98, 99], les modèles fondés sur les processus de coalescence sont devenus un outil extrêmement efficace pour résoudre un large spectre de problèmes liés à l'élaboration de modèles aléatoires pour la biologie et à l'analyse statistique des données génétiques toujours plus riches dont nous disposons.

Grâce aux progrès de la biologie moléculaire, ces techniques ont en particulier été appliquées avec succès dans de nombreuses études récentes sur le thème de l'évolution. Parce que les processus de coalescence décrivent les généalogies d'échantillons et non de population entières, ils ont permis de formuler des analyses rigoureuses des données issues de prélèvements.

L'utilisation de la coalescence permet par exemple de formuler un test pour la présence de sélection, d'estimer le taux de mutation, d'étudier l'âge du plus récent ancêtre commun d'une population et elle s'est également révélée utile pour l'analyse des schémas de migrations anciennes dans l'histoire humaine.

Pour ces applications, le coalescent de Kingman est de fait devenu le modèle standard, et les outils statistiques disponibles reposent presque tous sur des hypothèses qui reviennent à supposer que la généalogie des populations considérées est bien approximée par le coalescent de Kingman. Cependant, il existe des situations dans lesquelles l'usage de Λ-coalescents autorisant des collisions multiples comme modèles semble plus adapté. C'est en particulier le cas pour certaines espèces marines pour lesquelles il y a une très forte variance du succès reproductif (voir, e.g., [106]), pour des populations de virus à l'intérieur d'un hôte (ces populations sont soumises à de très fortes pressions sélectives), ou bien lorsque l'on considère l'analyse fine des loci de certaines maladies génétiques (cf. [112]).

L'un des outils clés pour le développement des applications du coalescent est la formule d'échantillonnage de Ewens (ESF) qui décrit entièrement la loi du polymorphisme génétique (i.e. la partition allélique) attendu dans une population stable, neutre et panmictique dont la généalogie est donnée par le coalescent de Kingman. Il y a donc une forte motivation pour comprendre comment cette formule se trouve

modifiée quand on passe à une généalogie Λ-coalescent (see, e.g., [60, 75, 110, 137]). Par exemple, dans un échantillon de n individus, combien d'allèles différents observe-t-on ? Sur combien de sites nucléotide dans l'ADN verra-t-on du polymorphisme ? A quoi ressemble la partition allélique de cet échantillon ? Il semble que seul le coalescent permette d'obtenir des formules fermées et que l'on doive en général se contenter de chercher de bonnes approximations.

L'objet de ce chapitre est de présenter les contributions que mes coauteurs et moi-même avons apporté à ce sujet dans les articles [22, 20]. Je commencerai par rappeler quelques éléments de la théorie classique pour le coalescent de Kingman avant de passer au cas des Λ-coalescents.

3.1 Rappels

3.1.1 Le coalescent de Kingman comme limite d'échelle des généalogie dans les modèles de Cannings

Considérons une population sur laquelle on faut les hypothèses suivantes :

- La taille de la population est constante au cours du temps et fixée à N individus.
- La population évolue au cours du temps par générations successives qui ne se recouvrent pas (on a jamais coexistence d'individus appartenant à différentes générations).
- La population est neutre et non-structurée. Ceci signifie qu'il n'y pas d'individus qui sont avantagés par rapport à d'autres et qu'il n'y a pas d'effets spatiaux (champs moyen).
- La population est haploïde, et en particulier chaque individu a un et un seul parent à la génération précédente.

Le modèle naturel pour l'évolution d'une telle population est alors le *modèle de Cannings* : à chaque génération on tire une copie indépendante d'un vecteur $(d_1, d_2, \ldots, d_N)$ où les $d_i \in \mathbb{N}$ sont des variables échangeables (i.e. la loi du vecteur est invariante par l'action d'une permutation) -ce qui correspond à la neutralité- et telles que $\sum_{i=1}^{N} d_i = N$ presque sûrement. Pour chaque i, d_i est alors le nombre d'enfant de l'individu i dans la génération suivante. Un exemple particulier bien connu de modèle de Cannings est le modèle de Wright-Fisher. À chaque génération, chaque individu i choisit son parent dans la génération précédente uniformément parmi les N individus. Ceci revient à choisir une loi multinomiale pour le vecteur $(d_1, \ldots, d_n)$.

On peut alors définir le processus de la partition ancestrale. On échantillonne $n \leq N$ individus à une génération donnée (disons génération 0) et on définit $\mathcal{R}_n^{(N)}$ comme la partition de $\{1, \ldots, n\}$ telle que i et j sont dans le même bloc de $\mathcal{R}_n^{(N)}$ si et seulement si les individus i et j ont le même ancêtre à la génération $-n$.

Il est bien connu que dans le cas du modèle de Wright-Fisher la restriction à $\{1, \ldots, n\}$ du processus $(\mathcal{R}^{(N)}_{[Nt]}, t \geq 0)$ converge en loi vers le n-coalescent de Kingman (i.e. le coalescent de Kingman restreint au n premiers entiers). Plus généralement, Möhle [109] (voir aussi les notes de cours de Birkner [52]) a donné une condition nécessaire et suffisante pour la convergence des généalogies renormalisées vers le coalescent de Kingman dans un modèle de Cannings.

Théorème 27 *Möhle [109]* $(\mathcal{R}^{(N)}_{[c_n t]}, t \geq 0) \to (\kappa_t, \geq 0)$ *si et seulement si* $c_N \to 0$ *et* $d_N/c_N \to 0$ *quand* $N \to \infty$ *où* $c_N = \mathbb{E}[d_1(d_1 - 1)]/(N-1)$ *est la probabilité que deux individus aient le même parent à la génération précédente et* $d_N = \mathbb{E}[d_1(d_1-1)(d_2-2)]/((N-1)(N-2))$ *est la probabilité pour trois individus d'avoir le même parent.*

Ceci se produit par exemple quand la loi de d_1 a un moment d'ordre 2 uniformément borné (quand N varie). Cette propriété illustre la robustesse du coalescent de Kingman qui apparaît ainsi comme une limite d'échelle universelle pour les généalogies. On verra plus loin que si l'on relâche les hypothèses du Lemme de Möhle les généalogies peuvent être n'importe quel Ξ-coalescent.

3.1.2 Le coalescent de Kingman et la formule d'échantillonnage de Ewens

On introduit à présent des mutations dans la populations. Pour être plus précis on suppose que les individus sont caractérisés par une longue chaîne ADN .... `ACTCAATCGGTA...` avec une faible probabilité de mutation à chaque site. On estime que le taux de mutation par paire de bases chez les eukariotes est de l'ordre de 10^{-10} par événement de réplication et de $10^{-6} - 10^{-10}$ chez les microbes (cf. Drake et al. [61]). C'est-à-dire que chaque lettre de cette chaîne a une très faible probabilité de ne pas être copié correctement entre un individu et ses descendants à la génération suivante. En première approximation, on peut supposer que cette chaîne est assez longue pour considérer que chaque mutation affecte un nouveau site, c'est à dire modifie une lettre du code située à une position différente. Cette approximation, due à Kimura [96] en 1969, porte le nom d'*infinite site model*, ou modèle à nombre de sites infini. On peut par exemple penser que cette séquence ADN est un code pour une couleur avec un propriété d'injectivité (deux séquences différentes donnent deux couleurs distinctes). Ainsi chaque mutation produit une nouvelle couleur jamais vue auparavant. Il est intéressant de remarquer que ce modèle contient également le *infinite allèle model* (ou modèle à nombre infini d'allèles) qui ne diffère que par le niveau d'information disponible : on ne "voit" pas le code ADN mais simplement la couleur. De ce fait on est seulement capable de dire si deux individus portent le même allèle (i.e. la même version du code ADN) ou pas, mais on ne peut pas connaître le nombres de lettre par lesquelles ils diffèrent éventuellement (cf. le très bon livre de Hein, Schierup et Wiuf par exemple [134]).

On suppose que les mutation surviennent avec un taux constant (une hypothèse peu réaliste tant le long du génome qu'au cours du temps, voir à ce propos [9]). Plus

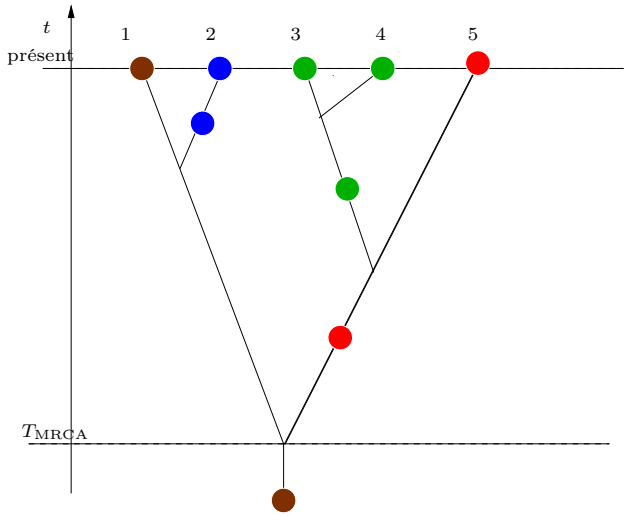

FIGURE 3.1: Un exemple de généalogie marquée pour un échantillon de cinq individus. Le temps s'écoule de bas en haut -la racine est l'ancêtre commun de l'échantillon. La couleur ancestrale est marron, la première mutation crée la couleur rouge et affecte les individus 3,4 et 5, etc...

précisément, si l'on se place dans le cadre d'un modèle de Cannings, on suppose que chaque individu à une probabilité μ_N de muter par rapport à son parent et que dans la limite $N \to \infty$ on a

$$\mu_N / C_N \to \theta \tag{3.1}$$

(donc $N\mu_N \to \theta$ dans le cas du modèle de Wright-Fisher). Avec (3.1), la généalogie d'un échantillon converge à présent vers un arbre de Kingman décoré par un processus de Poisson d'intensité θ le long des branches et donc chaque marque correspond à une mutation.

Afin d'introduire les notions de spectre de fréquence allélique et de spectre de mutations je propose de considérer un échantillon de cinq individus dont la généalogie et l'histoire génétique est résumée par le graphique suivant :

Le tableau suivant montre les séquences génétiques des cinq individus alignées, c'est l'information à laquelle on a accès aujourd'hui.

indiv. 1	A	T	T	C	A	G
indiv. 2	A	T	T	G	A	G
indiv. 3	T	T	T	C	A	C
indiv. 4	T	T	T	C	A	C
indiv. 5	A	T	T	C	A	C
# d'indiv. affectés	1			1		3

Notons que l'on peut reconstruire l'histoire généalogique et des mutations (ou plus

exactement la topologie de cette histoire) de cet échantillon à partir de la donnée du tableau ci-dessus.

- Le nombre S_n (n est la taille de l'échantillon et vaut ici 5) de sites polymorphes (on parle de SNP -single nucleotide polymorphism- ou snip) correspond au nombre de mutations portées par l'arbre et vaut ici 3.
- Le nombre d'allèles (i.e. le nombre de couleurs) différents que l'on voit dans l'échantillon est noté A_n et vaut ici 4 (puisque l'on voit chaque mutation ainsi que le type ancestral).
- Le spectre de mutation est le vecteur $(M_{k,n}, k = 1, \ldots, n)$ où $M_{k,n}$ est le nombre de mutations portées par k individus. Ici $(M_{k,n} k = 1, \ldots, n) = (2, 0, 1, 0, 0)$.
- Le spectre allélique est le vecteur $(F_{k,n}, k = 1, \ldots, n)$ où $F_{k,n}$ est le nombre d'allèles portés par k individus. Ici $(F_{k,n}, k = 1, \ldots, n) = (3, 1, 0, 0, 0)$.
- La partition allélique est la partition de $\{1, \ldots, n\}$ telle que i et j sont dans le même bloc si et seulement si ils ont le même type allélique (i.e. la même couleur).

La loi de la partition allélique (et donc du spectre allélique) a été donnée par Warren Ewens dans un article célèbre de 1972 [79], et l'on appelle aujourd'hui *formule d'échantillonnage de Ewens* la loi de la partition allélique obtenus en jetant des mutations selon un processus de Poisson de taux $\theta/2$ sur l'arbre de Kingman. Soit π une partition de $\{1, \ldots, n\}$ et $\forall j \leq n$ notons a_j le nombre de blocs de taille j de π, alors

$$P(\Pi_n = \pi) = \frac{n!}{\theta(\theta+1)\ldots(\theta+n-1)} \prod_{j=1}^{n} \frac{\theta^{a_j}}{j^{a_j} a_j!}.$$

Une preuve relativement simple de ce résultat peut être obtenu à l'aide du lien avec les urnes de Hoppe (cf. par exemple [132], qui contient également une revue des utilisations de la formule d'échantillonnage de Ewens dans différents champs des mathématiques)

La loi du nombre d'allèles A_n est

$$\mathbb{P}(A_n = k) = S(n,k) \frac{\theta^k}{\theta_{(n)}}$$

où $S(n,k)$ est le nombre de Stirling de la première espèce et $\theta_{(n)} = \theta(\theta+1)\ldots(\theta+n-1)$. On en déduit

$$\mathbb{E}(A_n) = \sum_{j=0}^{n-1} \frac{\theta}{j+\theta} \sim \theta \log n \ , \ \text{Var}\ (A_n) = \sum_{j=0}^{n-1} \frac{\theta j}{(j+\theta)^2} \sim \theta \log n$$

Le nombre de singletons a pour espérance

$$E(F_{1,n}) = n\theta/(n+\theta-1).$$

La loi complète du jeu de données dans le *infinite site model* est un problème difficile qui a été (partiellement) résolu par Griffiths, Ethier et Tavaré dans une série d'articles qui datent de 1987-1995 en utilisant un ensemble de récursions qui permettent de calculer les probabilités pour un jeu de données relativement petit (i.e. $S_n + A_n \leq 30$).

On sait cependant calculer la loi d'un certain nombre de quantités d'intérêt. On a par exemple une formule explicite pour la loi de S_n dont l'espérance et la variance sont données par

$$\mathbb{E}(S_n) = \theta \sum_{j=1}^{n-1} \frac{1}{j}\ ,\ \mathrm{Var}(S_n) = \theta \sum_{j=1}^{n-1} \frac{1}{j} + \theta^2 \sum_{j=1}^{n-1} \frac{1}{j^2}.$$

Pour le spectre de mutation on a

$$\mathbb{E}(M_{j,n}) = \frac{\theta}{j}.$$

3.2 Le concept de taille effective

Le concept de *taille de population effective* joue un rôle important en pratique dans les applications de la théorie du coalescent. Les populations réelles peuvent différer de multiples façons du modèle de Wright-Fisher : lois de reproduction non multinomiales, taille non-constante, structuration géographique ou génétique, présence de sélection, etc...

Lorsque l'on se sert du modèle de Wright-Fisher, ou du coalescent de Kingman pour modéliser une population réelle, la taille de la population ($2N$) dans le modèle haploïde de Wright-Fisher ne peut pas correspondre à la taille de la population réelle. Ainsi pour de nombreux gènes humains, on trouve un ancêtre commun il y a moins de 200.000 ans. Si on compte 20 ans par génération, alors N devrait être inférieur à $200.000/(4 \cdot 20) = 2.500$ (l'espérance de l'âge du dernier ancêtre commun dans une population diploïde est $4N$), ce qui bien sûr est peu réaliste. Mais cela suggère que la population réelle pourrait être bien approximée par un modèle de Wright-Fisher avec $N = 2.500$.

Pour une population réelle, ou pour un modèle de population (e.g. un Wright-Fisher diploïde) la taille de population du modèle de Wright-Fisher haploïde qui fournit la meilleur approximation est appelée taille de population efficace N_e. Ainsi dans l'exemple ci-dessus, $N_e = 2500$. Il y a plusieurs façon de définir N_e (cf. [80]), et je ne présenterais ici qu'un seul exemple appelé taille de population efficace de consanguinité (inbreeding effective population size). On la définit par

$$N_e = \frac{1}{2\mathbb{P}(T_2 = 1)}$$

où T_2 est en générations et désigne le temps qu'il faut à deux individus tirés au hasard parmi la population pour trouver leur ancêtre commun.

On peut illustrer ces notions par l'exemple caricatural suivant. Supposons qu'à chaque génération trois individus sont choisis et chacun produit un tiers de la génération suivante. La probabilité que deux individus pris au hasard ait le même parent vaut alors toujours $1/3$, quel que soit la taille de la population. On a alors $N_e = 1.5$

Si on suppose que ce sont à présent $\sqrt{2N}$ individus qui sont choisit et que chacun a $\sqrt{2N}$ descendants, la probabilité de consanguinité (d'avoir le même parent à la génération précédente) est $1/\sqrt{2N}$ et $N_e = \sqrt{2N}/2$.

Lorsque l'on approxime la généalogie d'une population réelle par un coalescent de Kingman avec taille $2N = 2N_e$, on considère un modèle haploïde de Wright-Fisher qui se comporte de façon réaliste *p*our un aspect particulier de la population en question, ici, le T_{MRCA} de deux individus. Il est important de noter que pour d'autres quantités, par exemple la variance du T_{MRCA} ou le T_{MRCA} d'un échantillon de taille n, le modèle peut avoir un comportement assez éloigné de la réalité.

Typiquement, une fois que l'on a estimé N_e et μ (le taux de mutation par site par génération), on pose $\theta = 2N_e\mu$ (cf. (3.1)) et on utilise la formule de Ewens avec ce θ comme hypothèse nulle pour tester e.g. la présence de sélection.

Pour conclure, il est également intéressant de noter que, ainsi que le montrent les deux exemples que nous avons choisis ci-dessus, plus la variance du succès reproductif est forte et plus on aura un rapport N_e/N petit (voir à ce propos [55] et plus récemment [93]).

3.3 Au-delà du coalescent de Kingman

J'ai déjà mentionné que lorsque les hypothèse du Lemme de Möhle n'étaient pas remplies, la limite d'échelle des généalogies dans un modèle de Cannings pouvait être autre chose qu'un coalescent de Kingman. Möhle et Sagitov [111] on montré qu'on pouvait en fait obtenir tous les Ξ-coalescents (une généralisation des Λ-coalescents qui autorisent les collisions multiples simultanées).

Il y a en particulier toute une classe de modèles de Cannings qui sont des modèles naturels pour des populations dont la taille reste constante en raison de ressources limitées et qui ne permettent d'obtenir que des Beta-coalescent ou le coalescent de Kingman comme généalogies possibles. Plus précisément, soit $\alpha > 1$ fixé et $Y \geq 1$ une variable aléatoire telle que $\mathbb{P}(Y > k) \sim ck^{\alpha}$ pour une certaine constante c. La variable Y a une moyenne finie, mais son moment d'ordre 2 n'est fini que si $\alpha \geq 2$. On suppose que le passage d'une génération à la suivante se fait en deux étapes. Dans un premier temps les individus se reproduisent librement et donnent naissance à un nombre iid d'enfants de distribution Y. Dans un deuxième temps on se rappelle que l'on veut garder la population de taille constante égale à N (par exemple parce que la population vit dans un environnement aux ressources limitées) et l'on échantillonne uniformément sans remise N enfants parmi tous ceux produits pour constituer la

nouvelle génération. Schweinsberg [126] a montré que la limite d'échelle pour les généalogies était alors le Beta $(2-\alpha,\alpha)$-coalescent.

Théorème 28 *Schweinsberg [126] Soit* $(\mathcal{R}_n^{(N)}, n \in \mathbb{N})$ *le processus de la partition ancestrale pour le modèle ci-dessus de paramètre* $\alpha \in (1,2)$. *On a* $(\mathcal{R}_{[c_N t]}^{(N)}, t \geq 0) \to (\Pi_t, \geq 0)$ *où* Π *est un Beta* $(2-\alpha,\alpha)$*-coalescent (où l'on rappelle que* $c_N = \mathbb{E}[d_1(d_1-1)]/(N-1)$ *est la probabilité que deux individus aient le même parent à la génération précédente).*

Ce résultat montre que les Beta-coalescents sont une limite d'échelle naturelle pour les généalogies de populations dans lesquelles certains individus peuvent produire un très grand nombre d'enfants, i.e. pour lesquelles on observe une très grande variance du succès reproductif.

Au delà de ce résultat mathématique, l'idée que des coalescents à collisions multiples sont plus adaptés pour décrire certaines populations émerge également depuis quelques années dans différents champs de la biologie et de la génétique des populations. Le point commun de ces différentes situations est qu'à chaque fois la taille de population efficace est inférieure de plusieurs ordres de grandeur à la taille réelle. Ainsi que je l'ai mentionné plus haut, ceci est souvent considéré être un indice d'une très forte variance du succès reproductif, et comme on vient de le voir, dans ce cas le coalescent de Kingman n'est plus nécessairement la bonne approximation des généalogies.

1) Espèces marines : Récemment différents groupes de chercheurs ont étudié le polymorphisme génétique rencontré dans certaines espèces marines dont le comportement reproductif conduit à une très forte variance du succès reproductif. Parmi ceux-ci on peut citer l'étude de Árnason [7] qui étudie l'ADN mitochondrial de populations de morues de l'atlantique ou encore l'article de Li et Hedgecock [106] où l'étude d'une population de larves d'huître du pacifique conduit les auteurs (dans le titre) à conclure que *l'hétérogénéité génétique dans un échantillon de larves d'huîtres du pacifique [...] est un indice d'une grande variance du succès reproductif.* Pour citer l'abstract de cet article :
Les différences dans la composition génétique d'échantillons de larves produites au cours d'une seule saison de reproduction par une population semi-isolée d'huîtres creuses du Pacifique (Crassostrea gigas) dans la baie Dabob, dans l'État de Washington, confirment de façon spécifique un aspect de l'hypothèse suivant laquelle cette espèce et d'autres animaux marins montrent de grandes variances dans leur succès de reproduction. Sur 666 échantillons les auteurs observent 68 haplotypes (couleurs) distincts dont 49 (72%) ne sont portés que par un seul individu. Ce chiffre est très loin de ce qui serait attendu pour le coalescent de Kingman où, pour $n = 666$, l'ordre de grandeur de la proportion de singletons est $\sim \theta/(\theta \log n) \sim 1/\log n \sim 15\%$. Eldon et Wakeley [75] analysent également l'échantillon décrit dans [106] et proposent une famille de Λ-coalescents à un paramètre (qui ne sont pas les Beta coalescents) qui

peut donc être estimé par une simple statistique (*summary statistic*) : la proportion d'allèles singletons (cf. la discussion ci-dessus). Ils concluent que *For many species, the coalescent with multiple mergers might be a better null model than Kingman's coalescent.* Finalement, Birkner et Blath [51] proposent une méthode qui permet de calculer la vraisemblance complète d'un jeu de données pour un Λ-coalescent général qui permet d'estimer le taux de mutation et la paramètre du modèle de population et les auteurs appliquent leur méthode au cas particulier des Beta coalescent.

2) Populations de virus intra-hôtes et balayages sélectifs répétés : Une population de virus peut être étudiée à deux échelles. Soit l'on considère la population totale du virus, soit l'on s'intéresse à la population des virus qui vivent à l'intérieur d'un individu hôte particulier. C'est ce dernier point de vue auquel Achaz et ses collaborateurs se sont ainsi intéressés récemment dans le cas du virus HIV [1]. Ces populations, (dans la phase "stable" de l'infection) comptent de l'ordre de 10^{10} virus. Un phénomène particulier à ces population et qui est du aux traitements est l'apparition de mutants résistants aux médicaments qui envahissent la population (on parle de balayage sélectif). On dispose de plusieurs études dont le but est l'estimation de la taille de population effective dans un hôte, et toutes concluent à une taille de l'ordre de 10^3 à 10^4 soit un facteur de 10^6 par rapport à la taille réelle. Les auteurs proposent deux explication à ce rapport : soit une structuration de la population, soit l'action de la pression sélective et l'apparition de balayage sélectifs comme dans le modèle de Gillespie [84] dit de "pseudohitchhiking". Les auteurs notent que *The most obvious deviation from the null model, with regard to HIV-1 intra-host evolution, concerns the neutrality of the observed variation. Clearly HIV-1 is under tremendous selective pressure during an infection.* Un balayage sélectif est un événement où un mutant porteur d'une mutation avantageuse apparaît (e.g. résistance à un médicament) et envahit la population. Du fait de la possibilité de recombinaisons, des allèles situés sur des loci proches du site de la mutation peuvent survivre s'ils sont pris en "auto-stop" par le mutant. Dans le modèle de pseudohitchhiking, le rapport N_e/N devient une fonction de la fréquence et de la sévérité des balayages sélectifs. Or c'est justement ce cas précis, une population dont l'évolution est marquée par des balayages sélectifs répété avec recombinaison, que considèrent Durrett et Schweinsberg [71, 72]. Il est montré dans ces articles que la généalogie limite quand $N \to \infty$ est un Λ-coalescent, mais que les Ξ-coalescents (coalescents à collisions multiples simultanées) fournissent une meilleur approximation quand N est fini.

3.4 Formules asymptotiques pour le spectre allélique et le spectre de mutations

On considère une population dont la généalogie est donnée par un Λ-coalescent qui descend de l'infini et telle que les mutations arrivent sur l'arbre généalogique selon un processus de points de Poisson de taux θ par unité de longueur. On s'attend donc à

ce que le nombre de mutations S_n soit asymptotiquement proportionnel à la longueur totale de l'arbre.

Nos résultats étant asymptotiques en la taille de l'échantillon, il est important de souligner que l'on peut construire simultanément pour tous les n les variables aléatoires $M_{k,n}, F_{k,n}, S_n, A_n$, etc ... Ainsi que l'illustre la figure 1.2.2, il s'agit simplement de l'idée classique qu'un Λ-coalescent Π restreint aux entiers $\{1, \ldots, n\}$ est un n-coalescent Markovien $\Pi^{(n)}$. Plus précisément, on appelle $\mathbf{T}_n$ l'arbre de coalescence associé à $\Pi^{(n)}$. Par définition, pour chaque n, $\mathbf{T}_n$ est un sous-arbre de $\mathbf{T}_{n+1}$ et la limite projective que l'on obtient à la limite quand $n \to \infty$ pour $\mathbf{T}_n$ est l'arbre continu $\mathbf{T}$ associé au coalescent Π. Si l'on décore à présent $\mathbf{T}$ avec des marques Poissoniennes d'intensité θ, les variables S_n, A_n etc... sont bien définies pour tous les n simultanément et l'on peut utiliser les propriétés de monotonies simples (e.g., $S_n \leq S_{n+1}$) implicites dans cette construction.

On dira que la mesure Λ est à variations (fortement) régulières d'indice α en zéro si $\Lambda(dx) = f(x)dx$ où $f(x) \sim Ax^{1-\alpha}$ quand $x \to 0$ pour un certain $A > 0$ et $1 < \alpha < 2$.

On rappelle que l'on associe à Λ les fonctions ψ et v définies par

$$\psi(q) := \int_{[0,1]} (e^{-qx} - 1 + qx)\nu(dx),$$

où $\nu(dx) = x^{-2}\Lambda(dx)$, et

$$\int_{v_\psi(t)}^{\infty} \frac{1}{\psi(q)} dq = t,$$

avec la convention que $v_\psi \equiv \infty$ si $\int_a^\infty 1/\psi(q)\, dq = \infty$ (pour un certain, et dans ce cas) pour tout $a > 0$.

Notre premier résultat donne l'asymptotique du nombres de sites de ségrégation et du nombre de familles dans la partition allélique.

Théorème 29 *[20] Soit $X_n = A_n$ ou S_n. Pour chaque $x > 0$ fixé*

$$X_n / \left(\int_x^n \frac{q}{\psi(q)}\, dq \right) \to \theta, \text{ en probabilité.} \tag{3.2}$$

Si de plus Λ est à variations (fortement) régulières d'indice α en zéro, alors les convergences ci-dessus est presque sûre, et donc pour une certaine constante $B \equiv B(A, \alpha)$

$$\frac{S_n}{n^{2-\alpha}} \longrightarrow \theta B, \text{ presque sûrement,}$$
$$\frac{A_n}{n^{2-\alpha}} \longrightarrow \theta B, \text{ presque sûrement.}$$

Le résultat qui concerne S_n est une application relativement aisée des résultats asymptotiques sur la longueur totale de l'arbre du Théorème 7, l'étude du nombre de blocs de la partition allélique A_n demande plus de travail. On a en effet $A_n \leq S_n$ pour

tout n par construction (car chaque allèle non-ancestrale correspond à une unique mutation mais une mutation peut ne pas être "visible" dans la partition allélique finale). Le problème est donc de montrer que la différence $S_n - A_n$ est négligeable (en probabilité ou presque sûrement) devant S_n.

Le caractère presque sûre des convergences de S_n et A_n (nous avions déjà obtenu la convergence en probabilité dans [22]) nous permet d'utiliser des résultats de [86] concernant les partitions échangeables. Le résultat principal de [86] peut en effet s'énoncer ainsi : soit π une partition échangeable telle que presque sûrement $K_n \sim cn^\beta$ où $\beta \in (0,1)$ et K_n est le nombre de blocs de π restreint à $\{1,\dots,n\}$, alors les fréquences asymptotiques des blocs de π sont déterministes et données par une formule qui ne dépend que de c et de β (voir aussi [127]). Ceci nous permet d'obtenir le comportement des spectres comme un corollaire.

Théorème 30 *[20] Supposons que Λ est à variation fortement régulières d'indice α en zéro. On rappelle la constante $B \equiv B(\alpha, A)$ du théorème 29 et on définit $C = B \cdot (2-\alpha)/\Gamma(\alpha-1)$. Soit $X_{k,n} = \begin{cases} F_{k,n} \\ M_{k,n} \end{cases}$*

Pour tout $k \geq 1$, quand $n \to \infty$,

$$\frac{X_{k,n}}{n^{\alpha-2}} \to \theta C \frac{\Gamma(k+\alpha-2)}{k!}, \quad p.s. \tag{3.3}$$

En outre,

$$\left(\frac{F_{j,n} - \mathbb{E}(F_{j,n})}{n^{2-\alpha}\theta B}, j = 1, 2, \dots \right)$$

converge en distribution vers un tableau Gaussien de moyenne zéro et de matrice de covariance

$$\sigma_{r,s} = -\frac{(2-\alpha)\Gamma(r+s+\alpha-2)}{r!s!} 2^{2-r-s-\alpha} \quad , r \neq s$$
$$\sigma_{r,r} = -\frac{(2-\alpha)\Gamma(2r+\alpha-2)}{r!r!} 2^{2(1-r)-\alpha} + \frac{(2-\alpha)\Gamma(r+\alpha-2)}{r!}.$$

Ce dernier résultat prolonge un Théorème que nous avions obtenu dans [22] où les asymptotiques pour les spectres alléliques et de mutation était obtenu en probabilité uniquement dans le cas du Beta-coalescent, et sans théorème central limite. Ce dernier élément est évidemment crucial du point de vue des applications, puisqu'il ouvre la voie à la mise en place de tests statistiques portants sur l'ensemble du spectre de fréquences.

Je conclurai cette section en présentant un résultat qui est l'un des ingrédients principaux des preuves et qui concerne l'âge d'une mutation typique.

Définissons $A_1^n = \{\mathbf{T}_n$ est marqué par au moins une mutation$\}$. Comme la longueur L_n de $\mathbf{T}_n$ diverge, on a que $\mathbb{P}(A_1^n) \to 1$, quand $n \to \infty$. En effet, typiquement $\mathbf{T}_n$ porte un nombre de marques proportionnel à L_n.

Sur l'événement A_1^n, on note M_n l'âge d'une mutation choisie au hasard uniformément parmi les mutations de $\mathbf{T}_n$. On adopte la convention $M_n \equiv \tau_1^n$ sur le complément de A_1^n, où τ_1^n est l'âge du plus récent ancêtre commun de $\{1, \ldots, n\}$.

Définissons

$$f(n) = \begin{cases} n^{1-\alpha}, & \text{si } 1 < \alpha < 3/2\,, \\ n^{-1/2} \log n, & \text{si } \alpha = 3/2\,, \\ n^{\alpha-2}, & \text{si } 3/2 < \alpha < 2\,. \end{cases} \tag{3.4}$$

Proposition 31 *Supposons que Λ est à variation fortement régulières d'indice α en zéro. Alors, pour $c \equiv c(\alpha)$ tel que $t_n \sim cn^{1-\alpha}$ on a*

$$\frac{M_n}{n^{1-\alpha}} \Rightarrow c\,(U^{-(\alpha-1)/(2-\alpha)} - 1), \text{ où } U \stackrel{d}{=} Unif[0,1]. \tag{3.5}$$

En outre, il existe $c_1 \equiv c_1(\alpha) \in (0, \infty)$, tel que pour f donné par (3.4)

$$\lim_{n\to\infty} \frac{\mathbb{E}(M_n)}{f(n)} = c_1. \tag{3.6}$$

Remarque 32 *On peut deviner qu'il y a une transition à $\alpha = 3/2$ au regard de (3.5), le membre de droite étant une variable intégrable si et seulement si $\alpha < 3/2$. Il est intéressant de noter que, $f(n)$ en tant que fonction de α décroît sur $(1, 3/2)$ et croît sur $(3/2, 2)$, mais il parait difficile d'expliquer cela heuristiquement.*

Chapitre 4

Mouvement Brownien avec absorption : généalogie et survie

Ce chapitre est dédié à des résultats plus récents qui sont contenus dans [23] (soumis) et [24] (à soumettre). Je commencerai par une brève présentation de conjectures et de résultats formulés par un groupe de physiciens autour de Bernard Derrida et qui ont motivés notre travail avant d'énoncer les principaux résultats auxquels nous avons abouti.

4.1 Les conjectures de Brunet-Derrida

4.1.1 Les marches aléatoires branchantes avec sélection

L'objet principal des travaux de Brunet, Derrida, Mueler et Munier [43, 44, 46] est l'analyse d'un modèle de population en évolution sous sélection et les liens de ce modèle avec l'équation de Fischer-Kolmogrov Petrovski Piskunov bruitée. Le modèle est le suivant. On considère une population de taille fixée N avec reproduction asexuée. Chaque individu i est caractérisé par un nombre réel x_i qui mesure son avantage sélectif (sa *fitness*) et la population est entièrement caractérisée par la donnée de ces N positions sur $\mathbb{R}$. A chaque génération, l'ensemble de la population est renouvelée selon le mécanisme en deux étapes suivant :

Reproduction-mutation : chaque individu produit k enfants et l'enfant j de l'individu i se trouve en position $x_i + \epsilon_{i,j}$ où les $\epsilon_{i,j}$ sont des variables i.i.d. de distribution commune ρ.

Sélection : On ne garde que les N individus les plus à droites parmi les kN enfants créés. Typiquement on pense à $k = 2$ et ρ est la distribution uniforme sur $[0, 1]$. La population est donc décrite par un système de marches aléatoires branchantes avec sélection.

Si les N individus qui forment la nouvelle génération étaient choisis uniformément et non pas selon leur position, on aurait affaire à un modèle de Wright-Fisher. On sait alors qu'il faut de l'ordre de N générations pour trouver l'ancêtre commun de deux individus et que sur cette échelle de temps la généalogie renormalisée converge vers le coalescent de Kingman.

L'action de la sélection change qualitativement le comportement du modèle. Dans ce cas (quand on sélectionne les N individus les plus à droite pour former la nouvelle génération) les auteurs conjecturent (en s'appuyant sur de puissantes simulations numériques) que la bonne échelle de temps est en $(\log N)^3$ et que sur cette échelle de temps l'arbre généalogique converge vers le coalescent de Bolthausen-Sznitman.

Plus précisément, si T_p désigne l'âge du plus récent ancêtre commun de p individus pris au hasard dans la population, les auteurs montrent que les statistiques $\langle T_p \rangle / \langle T_2 \rangle$ où $\langle \cdot \rangle$ est la notation physicienne pour l'espérance convergent vers les valeurs que l'on observe pour le coalescent de Bolthausen-Sznitman. La donnée de la suite complète $\langle T_p \rangle / \langle T_2 \rangle$ caractérise à un facteur multiplicatif près la loi d'un Λ-coalescent.

Outre cette conjecture qui concerne la généalogie du système, Brunet, Derrida, Mueller et Munier proposent une prédiction extraordinairement précise de la vitesse du système. Plus précisément, l'ensemble de la population de particules se déplace avec une vitesse balistique notée v_N et l'on a $v_N \to v_\infty < \infty$ quand $N \to \infty$ et (cf. équation (76) dans [46])

$$v_\infty - v_n = \frac{c}{2(\log N)^2} - c\frac{3 \log \log N}{(\log N)^3} + \dots \tag{4.1}$$

où c est une constante explicite en fonction de la loi de reproduction déplacement (cf. les résultats de Berard et Gouéré ci-dessou). Cette correction est également précisément celle qui est prédite par Brunet et Derrida pour la vitesse des solutions de l'équation FKPP bruitée (le paramètre N contrôle alors l'intensité du bruit). Plus généralement, Brunet et Derrida montrent informellement que les marches aléatoires branchantes avec sélection sont un exemple de modèle stochastique dont le comportement est décrit par l'équation FKPP bruitée (cf. [40]) (comme par exemple certains modèles de polymères dirigés à température nulle en environnement aléatoire). Je donne quelques détails supplémentaires sur ce point plus bas.

4.1.2 Universalité

Ces résultats sont robustes à la définition précise du modèle. La forme des corrections à la vitesse et la convergence des généalogies vers le coalescent de Bolthausen-Sznitman relèvent donc d'une certaine universalité. Voici quelques uns des changements que l'on peut faire subir au modèle tout en restant dans la même classe d'universalité (i.e. mêmes corrections à la vitesse et, sur une échelle de temps $(\log N)^3$, convergence de la généalogie vers le coalescent de Bolthausen-Sznitman).

- au lieu de prendre les N individus les plus à droite on les tire uniformément parmi les $3N/2$ les plus à droite (cf. [43]).

- Le nombre d'enfants peut être aléatoire (avec certaines hypothèses sur la loi de reproduction) et les variables de déplacement peuvent ne pas être indépendantes, i.e. pour chaque individu on tire de façon i.i.d. un processus de points qui donne les déplacements de ses descendants. Ce processus de points doit vérifier quelques hypothèses que nous ne détaillons pas ici.

- On peut se placer dans le cadre d'un modèle à temps continu. Par exemple on peut considérer le cas d'un mouvement Brownien branchant dans lequel les particules se déplacent indépendamment les unes des autres sur la ligne réelle selon un mouvement Brownien standard, branchent à taux 1 en deux nouvelles particules et le nombre total de particules en vie est maintenu constant égal à N en effaçant la particule la plus à gauche à chaque événement de branchement.

- Certains modèles de polymères dirrigés se rapprochent beaucoup des marches branchantes décrites ci-dessus. Brunet Derrida et Simon [49] ont montré que le coalescent de Bolthausen-Sznitman apparaissait également naturellement comme la généalogie de ces modèles. Je donne quelques détails supplémentaires sur ce point plus loin.

4.1.3 Les résultats rigoureux

Récemment plusieurs auteurs se sont intéressés à ces conjecture et certaines d'entre elles ont été partiellement résolues.

- Bérard et Gouéré [11] considèrent le cas d'une marche aléatoire branchante en temps discret où chaque particules produit 2 enfants dont les déplacements relatifs à leur parents sont i.i.d. de loi ρ avec $\Lambda(t) := \log \int e^{tx}\rho(dx.)$ On suppose que

$$\sigma := \sup\{t \geq 0 \Lambda(-t) < +\infty\} > 0 \; ; \; \zeta := \sup\{t \geq 0 \Lambda(t) < +\infty\} > 0$$

et que $\exists t^* \in]0, \zeta[$ tel que $t^*\Lambda'(t^*) - \Lambda(t^*) = \log 2$. Sous ces hypothèses les auteurs montrent que le déplacement de l'ensemble de la population est balistique de vitesse $v_n(\rho)$ et que $v_N(\rho) \to v_\infty(\rho) = \Lambda'(t^*)$ et en outre

$$v_\infty(\rho) - v_N(\rho) \sim \chi(\rho)(\log N)^{-2}$$

où $\chi(\rho) := \frac{\pi^2}{2} t^* \Lambda''(t^*)$. Ce résultat doit pouvoir s'étendre au cas d'un nombre d'enfants aléatoire et à des modèles à temps continu sans trop de difficultés. Notons que l'un des ingrédients essentiels de la preuve de se résultat repose sur la probabilité de survie de la marche aléatoire tuée sous une barrière linéaire due à Gantert, Hu et Shi [83]. Bérard et Gouéré donnent également une preuve alternative plus simple de ce résultat dans [12]. Ces techniques ne permettent pas pour l'instant d'obtenir le second ordre de la correction de la vitesse (le terme en $\log\log N/(\log N)^3$).

- Durrett et Remenik [70] (voir aussi Durrett et Mayberry [69]) considèrent un système de marches aléatoires branchantes en temps continu, ou chaque particule donne naissance à taux 1 à un enfant situé à une distance tirée indépendamment de tout le reste selon une loi de déplacement ρ. La population est maintenu constante à N en tuant la particule la plus à gauche à chaque événement de naissance. Les auteurs montrent que la suite des processus de la mesure empirique de la population converge vers un processus déterministe à valeur dans les probabilités sur $\mathbb{R}$ qui est solution d'une équation intégro-différentielle à frontière libre. En outre, pour chaque N le système est balistique de vitesse a_N et vu de la particule la plus à gauche, la population converge en variation totale vers un état d'équilibre. Enfin, $a_N \to a$ où a est la vitesse de la particule la plus à droite dans la marche aléatoire branchante sans sélection correspondante.

4.1.4 Polymères dirigés

Dans [39, 42] puis [49], il est conjecturé que le coalescent de Bolthausen Sznitman est également la généalogie de certains modèles de polymères dirigés. L'état du système est de nouveau une population de N points sur la ligne réelle qui évoluent par générations discrètes (on note $x_i(n)$ la position de l'individu i à la génération n). Pour chaque i on tire uniformément avec remise deux individus j_i et j_i' dans la génération n et on pose

$$x_i(n+1) = \max\{x_{j_i} + \alpha_i; x_{j_i'} + \alpha_i'\}$$

où les α_i, α_i' sont des variables iid, et plus précisément Brunet et Derrida prennent des variables de Bernoulli de paramètre p. Dans [39], les auteurs montrent à l'aide de simulations numériques que la correction de la vitesse est encore en $k/(\log N)^2$.

Dans [42], un modèle un peu différent est exploré où

$$x_i(n+1) = \max_{j=1,\dots,N} \{x_j(n) + S_{i,j}(n)\}$$

où les $S_{i,j}(n)$ sont iid de loi ρ. Après avoir montré comment ce modèle est relié à l'équation FKPP bruitée (et est donc dans la même classe d'universalité que les marches aléatoires branchantes avec sélection) un cas particulier est étudié en détail (on prend la distribution de Gumbel pour ρ). Malheureusement ce cas est spécial en ce sens que la vitesse du nuage de particules diverge avec N et ne permet pas de retrouver la correction en $(\log N)^{-2}$.

Finallement, dans [49] le modèle suivant de polymère dirigé en champs moyen et en environement aléatoire est proposé. Chaque génération consiste en N sites qui sont chacun connectés à M site choisis uniformément dans la génération précédente, $2 \leq M \leq N$. Chaque lien (AB) entre deux sites$A$ et B porte une énergie $e_{(A,B)}$. Un polymère dirigé est simplement un chemin dans la structure de graphe ainsi défini qui avance (sans saut) dans les générations. L'énergie E d'un polymère est la somme des énergies des arrêtes qu'il traverse. Le modèle est spécifié par la donnée de M et la loi

des $e_{(AB)}$. On prendra toujours des variables $e_{(A,B)}$ iid, des simulations numériques sont faites dans [49] pour le cas de variables uniformes sur $[0,1]$.

On fixe à présent un site racine dans la génération 0 et pour chaque site A d'une génération $n > 0$ on définit E_A comme l'énergie minimale du polymère sur tous les chemins dirigés possible qui relient la racine à A. Clairement, les chemins d'énergie minimum pour un ensemble de $k \leq N$ sites qui appartiennent à la même génération forment un arbre à k feuilles (au moins quand n est assez grand).

Les simulations effectuées dans [49] suggèrent que, comme pour les marches branchantes avec sélection, cet arbre, sur une échelle de temps d'ordre $(\log N)^3$, converge vers l'arbre de Bolthausen Sznitman.

Les auteurs proposent également la conjecture fascinante suivante. Les modèles de polymères dirigés présentent une transition de phase autour d'une température critique. Les auteurs s'attendent donc à ce que sous cette température critique l'arbre de Bolthausen Sznitman continue à décrire la structure des polymères, tandis qu'au dessus, la situation correspondrait aux marches aléatoires coalescentes et au coalescent de Kingman.

4.1.5 Le lien avec l'équation FKPP bruitée

Ce paragraphe est une tentative de résumer (avec un très faible degré de précision) certains des liens qui relient l'étude de ces marches branchantes avec sélection à l'équation de FKPP bruitée.

En 1937, Fisher [82] d'une part et Kolmogorov, Petrovski et Piskunov [101] d'autre part ont introduit l'équation aux dérivées partielles suivante qui décrit la propagation d'un front dans un milieu instable (dans le cas de Fisher il s'agissait de décrire la propagation d'un gêne mutant dans une population)

$$\partial_t u = \partial^2_{xx} u + u(1-u)$$

(je ne donne pas la version la plus générale de cette équation, en particulier on peut avoir un terme de non-linéarité différent de $u(1-u)$). Cette équation a des solutions de type "onde voyageuse", et si la condition initiale décroît assez rapidement, la vitesse minimum (notée v_∞) est sélectionnée.

L'équation de FKPP bruitée est

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + u - u^2 + \sqrt{\frac{u(1-u)}{N}}\, W(x,t), \tag{4.2}$$

où $W(x,t)$ est un bruit blanc espace-temps. Cette équation est en dualité fonctionnelle avec le modèle de mouvement brownien branchant (à taux 1) et coalescent (les coalescences se produisant le long du temps local d'intersection avec taux $1/N$ (voir [133, 128]). Pour cette équation stochastique, avec de bonnes conditions initiales, on a des solutions (aléatoires) telles que

$$r(t) = \sup\{x \in \mathbb{R} : u(t,x) > 0\} < \infty$$

et vu de $r(t)$ le processus a une unique loi stationnaire non-dégénérée. En d'autre termes, un front se forme et on s'attend à ce qu'il ait une vitesse $v_N = \lim_{t\to\infty} r(t)/t$. Ces résultats ont été prouvés dans [114].

Il est remarquable que la conjecture de Brunet et Derrida concernant les corrections (4.1) pour l'équation (4.2) ait été récemment (partiellement) prouvée par Mueller, mytnik et Quastel [115]. Leur résultat principal est que la vitesse v_N de (4.2) est donnée par

$$2 - \pi^2(\log N)^{-2}$$

avec une erreur d'ordre $\log\log N/(\log N)^3$.

Quel est le lien avec les systèmes de particules du type marches aléatoires branchantes avec sélection ? Je résume de façon très schématique à présent la discussion que l'on peut trouver dans [46]. On peut décrire l'état $(X_1(n), \ldots, X_N(n))$ du système au temps n par un front, c'est à dire une fonction

$$h_n(x) = \frac{1}{N}\#\{i = 1, \ldots, N \text{ t.q. } X_i(N) \geq x\}.$$

Clairement, $h_n(-\infty) = 1$ et $h_n(\infty) = 0$. Soit $Nh^*_{n+1}(x)$ le nombre d'enfants produits par la génération n à droite de x avant l'étape de sélection. On a alors

$$h_{n+1}(x) = \min[1, h^*_{n+1}(x)].$$

On peut calculer aisément la moyenne $\mu(h_n, x)$ et la variance $\sigma(h_n, x)$ de $h^*_{n+1}(x)$ conditionnellement aux positions à la génération n en terme de la distribution de déplacement ρ. En particulier $\sigma(h_n, x)$ est d'ordre $1/\sqrt{N}$. On a alors

$$h_{n+1}(x) = \min[1, 2\int dy\rho(y)h_g(x-y) + \eta_n(x)\sqrt{\sigma(h_n, x)}]$$

où $\eta_n(x, 1)$ est un bruit centré de variance 1 et le premier terme correspond au branchement binaire et au déplacement par ρ. Brunet et Derrida expliquent que cette équation est "similaire" à l'équation FKPP (4.2). La convolution par 2ρ joue le même rôle que l'opérateur de diffusion et étale le front, le fait de prendre le min avec 1 est un mécanisme de saturation comme le terme $-u^2$, et l'amplitude du terme de bruit est la même comme je viens de l'expliquer.

L'objet de ces modèles microscopiques est de comprendre le second ordre de correction de $v_\infty - v_N$ dans (4.1). En effet, le premier terme en $(\log N)^{-2}$ peut s'obtenir par des méthodes purement déterministes. Brunet et Derrida introduisent une équation déterministe avec cutoff

$$\partial_t u = \partial^2_{xx} u + u(1-u)a(u)$$

où $a(u) = 0$ dés que $u \leq 1/N$ et vaut 1 sinon. Ils montrent alors que la correction entre cette nouvelle équation et v_∞ est précisément ce terme $c/(2(\log N)^2)$ vu plus haut.

En revanche, le terme suivant en $\log\log N/(\log N)^3$ est lui dû aux fluctuations aléatoires de la position du front. Dans le langage du système de particules, cette correction (positive) à la vitesse correspond à des événements rares au cours desquels une particule arrive à s'avancer très loin devant ses poursuivants et peut alors "oublier" la sélection. Cette particule génère de nombreux descendant et fait avancer le front plus loin qu'il ne serait allé sinon, créant ainsi une fluctuation. Comme on le verra, cette explication heuristique qui est présentée dans [44] et qui est bien détaillée dans les notes de cours [45] est très proche de notre analyse de la généalogie du mouvement Brownien branchant avec sélection.

4.2 Résultats principaux

4.2.1 Mouvement Brownien branchant avec absorption

L'une des difficultés qui se pose lorsque l'on veut analyser rigoureusement le modèle à taille de population constante est que l'on perd l'indépendance entre les particules. Pour retrouver cette propriété on peut s'intéresser à un modèle voisin dans lequel la taille de la population n'est plus fixée (mais dans lequel la vitesse de sélection est imposée de façon exogène).

Dans ce modèle, on retrouve les deux ingrédients vus plus haut : mutation-reproduction et sélection.

> **Reproduction-sélection** : les individus se déplacent et se reproduisent selon le mécanisme d'un mouvement Brownien branchant. Plus précisément, chaque particule évolue indépendamment des autres en se déplaçant selon un mouvement Brownien standard et branche à taux 1 en donnant naissance à deux nouvelles particules identiques situées au lieu du branchement.
>
> **Sélection** : toutes les particules dont la position au temps t est inférieure ou égale à μt où $\mu \in \mathbb{R}$ sont immédiatement tuées et sortent du système.

Par un simple changement de repère, ce modèle de "mur qui avance" et de Brownien branchant sans dérive est équivalent au modèle dans lequel le mur est fixe en 0 (les particules sont immédiatement tuées en 0) et les particules Browniennes ont une dérive $-\mu$ (i.e. sont poussées vers 0). C'est ce modèle précisément qui va nous intéresser et que nous désignons sous le nom de mouvement Brownien branchant avec absorption en 0 (on note parfois kBBM(μ) pour killed Brownian motion with drift ($-\mu$)). (voir les figures 4.2 et 4.3 qui sont reproduites de la thèse de Damien Simon avec son aimable autorisation).

Ce modèle, dont l'étude remonte au moins à Kesten [95] a fait l'objet de travaux récents de Derrida et Simon [57, 130] dans lesquels une conjecture (étayée par une analyse perturbative de l'équation aux dérivées partielle de FKPP qui peut sans doute être rendue rigoureuse) concernant la probabilité de survie est formulée.

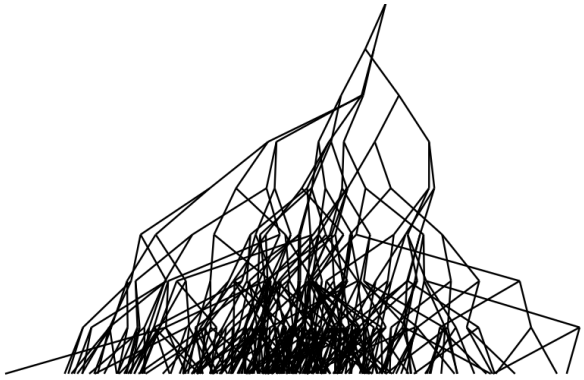

FIGURE 4.1: Mouvement Brownien branchant (ou marche branchante) sans sélection

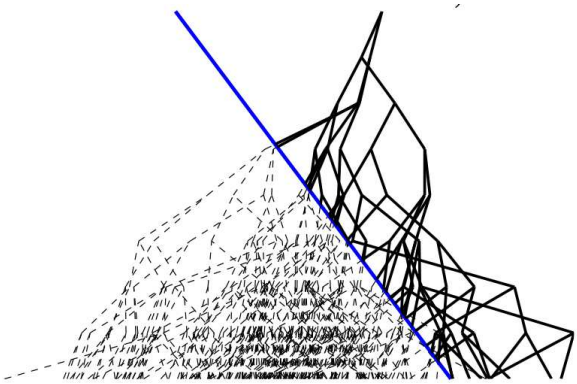

FIGURE 4.2: Mouvement Brownien branchant (ou marche branchante) avec sélection par absorption

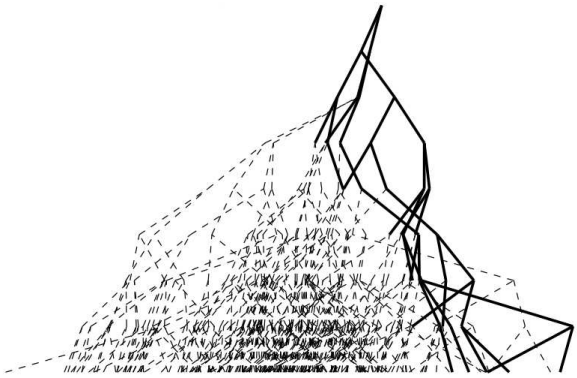

FIGURE 4.3: Mouvement Brownien branchant (ou marche branchante) avec sélection endogène (taille de population fixée à 6).

4.2.2 Survie

Dans la suite on notera $\mathbb{P}_x$ et $\mathbb{E}_x$ pour désigner la loi (et l'espérance sous cette loi) du kBBM qui a pour état initial une particule unique en position $x > 0$. On ne précise pas μ qui est en général fixé. Dans [95], Kesten montre un premier résultat simple sur la survie du processus, qui dans notre cadre peut s'énoncer ainsi :

Théorème 33 (Kesten, [95]) *Si $\mu \geq \sqrt{2}$ le processus s'éteint presque sûrement en temps fini (cas critique et sous-critique).*
Si $\mu < \sqrt{2}$ le processus survit avec probabilité positive et dans ce cas le nombre de particules en vie croît exponentiellement (cas sur-critique). Dans ce cas on note $Q_\mu(x)$ la probabilité sous $\mathbb{P}_x$ que le processus survive.

Plus récemment, quelques résultats supplémentaires ont été obtenus :

- Harris, Harris, et Kyprianou [91] montrent (cf. Théorème 13) que, pour $\mu < \sqrt{2}$, la fonction $x \mapsto Q_\mu(x)$ satisfait l'équation de Kolmogorov

$$\frac{1}{2}Q''_\mu(x) - \mu Q'_\mu(x) = Q_\mu(x)(1 - Q_\mu(x))$$

  avec conditions aux bords $\lim_{x\to 0} Q_\mu(x) = 0$ et $\lim_{x\to\infty} Q_\mu(x) = 1$. Ils montrent également (cf. Théorème 1) que pour chaque $\mu < \sqrt{2}$ fixé il y a une constante K telle que

$$\lim_{x\to\infty} e^{(\sqrt{\mu^2+2}-\mu)x}(1 - Q_\mu(x)) = K.$$

- Dans le cas sous-critique ($\mu > \sqrt{2}$) Harris et Harris [90] utilisent des techniques de martingales et de décomposition en épine dorsale pour calculer les asymptotiques de $\mathbb{P}_x(\zeta > t)$ quand t devient grand (et où ζ est l'instant d'extinction du processus).

- Dans [130] Simon et Derrida obtiennent des estimées pour la probabilité de survie $Q_\mu(x)$. Ils conjecturent qu'il existe K tel que quand $L - x \gg 1$,

$$Q_\mu(x) = KLe^{\sqrt{2}(x-L)}\left(\sin\left(\frac{\pi x}{L}\right) + O\left(\frac{1}{L^2}\right)\right) + O(e^{2\sqrt{2}(x-L)})$$

  (équation (B.16)) et qu'il existe une autre constante c telle que quand $x > L$ ou $L - x$ est $O(1)$,

$$Q_\mu(x) = 1 - \theta(x - L + c) + O\left(\frac{1}{L^2}\right),$$

  où θ résout l'équation différentielle (4.3) ci-dessous (équation (B.17)). Ces estimées sont obtenues à l'aide de méthodes entièrement analytiques qui ne sont cependant pas encore un argument tout à fait rigoureux.

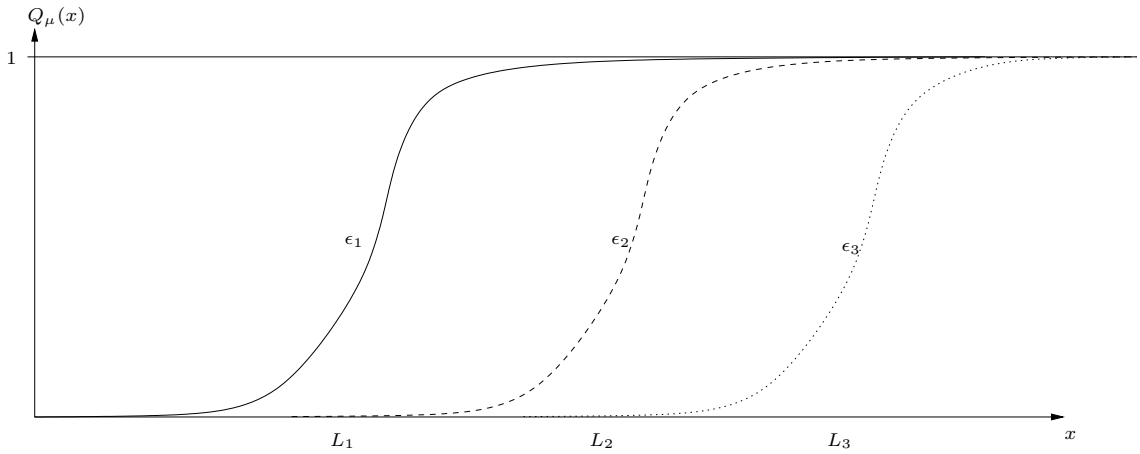

FIGURE 4.4: L'allure de la probabilité de survie $x \mapsto Q_\mu(x)$ pour $\mu - \sqrt{2} - \epsilon_{1,2,3}$ avec $\epsilon_1 > \epsilon_2 > \epsilon_3$. Un front se forme et s'éloigne de 0 (position L).

On pose $\epsilon = \sqrt{2} - \mu > 0$, et on définit L par la relation $1 - \mu^2/2 - \pi^2/2L^2 = 0$, c'est-à-dire $L = \pi/\sqrt{2-\mu^2} = \pi/\sqrt{\epsilon}$. Le théorème suivant que nous obtenons dans [24] valide les prédictions de Derrida et Simon.

Théorème 34 (B.,Berestycki,Schweinsberg [24]) *Soit $\alpha \in \mathbb{R}$ fixé. Alors il existe une fonction $\theta : \mathbb{R} \to \mathbb{R}$ telle que*

$$\lim_{\epsilon \to 0} Q_\mu(L + \alpha) = \theta(\alpha),$$

où $\theta : \mathbb{R} \to (0,1)$ satisfait

$$\frac{1}{2}\theta'' - \sqrt{2}\theta' = \theta(1-\theta). \tag{4.3}$$

Il existe une constante C telle que si $L - x \gg 1$, alors

$$Q_\mu(x) \sim CLe^{-\mu(L-x)} \sin\left(\frac{\pi x}{L}\right),$$

où $\sim$ désigne l'équivalence asymptotique quand $\epsilon \to 0$.

La preuve de ce résultat (dont je donne les grandes lignes plus loin) utilise les résultats que nous avons obtenus dans [23] et que je présente maintenant.

4.2.3 Généalogie et CSBP de Neveu

La présentation de ces résultats est légèrement modifiée par rapport à [24] ou un paramètrage différent est utilisé. Dans toute cette partie, pour chaque ϵ donné on travail sous la loi $\mathbb{P}_x$ où l'on fixe $x = L$. On se fixe $n \geq 1$ et l'on définit le processus de

la partition ancestrale $(\Pi_s^\epsilon, 0 \leq s \leq t)$ à valeur dans $\mathcal{P}_n$ (les partitions de $\{1, \ldots, n\}$) en tirant uniformément n individus dans la populations au temps $t\epsilon^{-3/2}$ et en posant que i et j sont dans le même bloc de Π_s^ϵ si et seulement si les individus i et j ont le même ancêtre au temps $(t-s)\epsilon^{-3/2}$.

Théorème 35 (B.,Berestycki,Schweinsberg [23])

$$(\Pi_s^\epsilon, 0 \leq s \leq t) \underset{\epsilon \searrow 0}{\overset{fdd}{\rightarrow}} (\Pi_{2^{3/2}\pi^3 s}, 0 \leq s \leq t)$$

où Π *est le coalescent de Bolthausen-Sznitman restreint à* $\{1, \ldots, n\}$.

En outre, si l'on appelle M_t^ϵ le nombre de particules en vie au temps t on a

Théorème 36 (B.,Berestycki,Schweinsberg [23])

$$\left(\frac{1}{\epsilon^{3/2} e^{\pi\sqrt{2/\epsilon}}} M_{t\epsilon^{-3/2}}^\epsilon, t \geq 0\right) \underset{\epsilon \searrow 0}{\overset{fdd}{\rightarrow}} (Z_t, t \geq 0)$$

où Z *est un CSBP de mécanisme de branchement* $\psi(u) = au + 2\pi^2 u \log u$, *pour une certaine constante* $a \in \mathbb{R}$, *issu d'une population initiale* Z_0 *aléatoire de loi connue.*

Le Gall et Bertoin avaient déjà observé dans [27] que la généalogie du CSBP de Neveu (de mécanisme de branchement $\psi(u) = u \log u$) était donné par le coalescent de Bolthausen-Sznitman. Ainsi, si l'on peut montrer que la population évolue selon une version du CSBP de Neveu il n'est pas surprenant que sa généalogie soit donnée par le coalescent de Bolthausen-Sznitman.

La renormalisation que l'on utilise dans le résultat ci-dessus montre que si l'on veut avoir de l'ordre de N particules en vie il faut prendre $\epsilon^{3/2} e^{\pi\sqrt{2/\epsilon}} \asymp N$ soit $\epsilon \sim 2\pi^2/(\log N + 3\log\log N)^2$. On peut ainsi décider de paramètrer entièrement le modèle par N en prenant $\mu_N = \sqrt{2 - \frac{2\pi^2}{(\log N + 3\log\log N)^2}}$ et $L = (\log N + 3\log\log N)/\sqrt{2}$.

L'un des résultats clés dans [23] consiste à montrer que l'on peut mesurer la taille de la population en attribuant des poids aux individus tout en conservant la convergence du Théorème 36 vers le CSBP de Neveu (avec une renormalisation différente). Plus précisément, si l'on définit

$$Z_N(t) := \sum_{i=1}^{M_N(t)} e^{\mu X_i(t)} \sin(\pi X_i(t)/L) \mathbf{1}_{\{X_i(t) \leq L\}}, \qquad t \geq 0,$$

on a alors le résultat suivant

Théorème 37 (B.,Berestycki,Schweinsberg [23]) *Sous* $\mathbb{P}_x$ *on a*

$$\left(\frac{Z_N(t)}{N(\log N)^2}, t \geq 0\right) \underset{\epsilon \searrow 0}{\overset{fdd}{\rightarrow}} (Z_t, t \geq 0)$$

où Z *est un CSBP de mécanisme de branchement* $\psi(u) = au + 2\pi^2 u \log u$ *pour une certaine constante* $a \in \mathbb{R}$ *issu d'une population initiale* Z_0 *aléatoire de loi connue.*

Dans [23] nous ne travaillons pas sous $\mathbb{P}_x$, nous formulons le Théorème ci-dessus pour une suite de configuration initiales $X_1(0), \ldots, X_{M_N(0)}(0)$ telles que $Z_N(0)/N(\log N)^2$ converge en distribution vers une certaine variable W non-dégénérée et tels que $Y_N(0) = o(N(\log N)^3)$ avec

$$Y_N(t) := \sum_{i=1}^{M_N(t)} e^{\mu X_i(t)}, \qquad t \geq 0.$$

Dans ce cas W est bien sûr la loi de la population initiale Z_0. Pour prouver les Théorèmes 36 et 37 tels que je les énonce ici, il faut donc montrer que si l'on démarre avec une particule en L, alors en un temps très court (négligeable devant $(\log N)^3$) le système se trouve dans une configuration $X_1, \ldots, X_N$ (où N est aléatoire) telle que pour ces position $Z_N/N(\log N)^2$ converge en distribution et $Y_N = o(N(\log N)^3)$. La loi limite de $Z_N/N(\log N)^2$ donne la loi initiale de Z_0 qui est mentionnée dans les résultats ci-dessus.

4.2.4 Quelques idées de la preuve des Théorèmes 35, 36 et 37

Notre approche dans [23] consiste à traiter séparément les particules qui atteignent une distance d'ordre L du mur en 0. Ces particules sont assez loin pour ne plus "sentir" le mur et ainsi peuvent rapidement produire une grande descendance, ce qui conduit à un saut positif de la taille de la population et à des coalescences multiples de lignées ancestrales lorsque l'on remonte le temps.

Cette approche est à comparer à la description phénoménologique du modèle à population constante que donnent Derrida et ses coauteurs (dans [43, 44] ou dans les notes du mini-cours que Eric Brunet a donné à Marseille en 2007 [45]). Ils montrent comment la correction de deuxième ordre à la vitesse (le terme en $\log\log N/(\log N)^3$) est due à des événements rares durant lesquels une particule se retrouve en avance du front et produit une grande descendance avant d'être rattrapée, créant ainsi une fluctuation dans la position du front.

Pour commencer, on peut considérer un mouvement Brownien branchant dans lequel les particules sont tuées lorsqu'elles atteignent 0 ou L. Si la configuration initiale consiste en une seule particule en position x, alors, pour t assez grand, la densité de particules attendue autour du point y est approximativement $p_t(x,y)\,dy$, où

$$p_t(x,y) = \frac{2}{L} e^{(1-\mu^2/2-\pi^2/2L^2)t} \cdot e^{\mu x} \sin\left(\frac{\pi x}{L}\right) \cdot e^{-\mu y} \sin\left(\frac{\pi y}{L}\right). \tag{4.4}$$

On peut tirer plusieurs informations sur le comportement du mouvement Brownien branchant dans une bande de cette formule. Le paramètre t n'apparaît que dans le premier facteur exponentiel. La taille de la population devrait donc rester stable dés que $1 - \mu^2/2 - \pi^2/2L^2 = 0$ (ce qui correspond en effet à notre choix). Une seconde observation est que la densité est proportionnelle à $e^{\mu x}\sin(\pi x/L)$, c'est-à-dire, en sommant sur toutes les particules au temps t, à $Z_N(t)$. On voit donc que $Z_N(t)$

est un bon prédicteur du nombre de particules que l'on trouvera dans un ensemble à un certain temps futur. Ainsi, $Z_N(t)$ est la bonne façon de mesurer la "taille" de la population. Enfin, la densité est également proportionnelle à $e^{-\mu y}\sin(\pi y/L)$. En conséquence, dés que t est assez grand (d'ordre $(\log N)^2$) le système a "oublié" sa configuration initiale et les particules sont dans une configuration stable dont la densité de particules à y est proportionnelle à $e^{-\mu y}\sin(\pi y/L)$.

L'étape suivante consiste à observer qu'une particule qui atteint L conduit à un "saut" dans la taille de la population. En effet, si l'on démarre avec N particules tirée selon cette densité d'équilibre entre 0 et L, alors $Z_N(0)$ vaut approximativement

$$N\int_0^L e^{\mu y}\sin\left(\frac{\pi y}{L}\right)\cdot CLe^{-\mu y}\sin\left(\frac{\pi y}{L}\right)dy,$$

qui est d'ordre NL^2. D'autre part, si on part avec une particule en L, ses descendants n'iront typiquement pas plus loin qu'une constante à droite de L. Pour estimer la contribution typique d'une particule en L au temps t on utilise (4.4) avec L à la place de x et $L+\alpha$ au lieu de L, avec $\alpha>0$ une constante. La valeur de $Z_N(t)$ doit être du même ordre que

$$\int_0^L e^{\mu y}\sin\left(\frac{\pi y}{L}\right)\cdot\frac{2}{L+\alpha}e^{\mu L}\sin\left(\frac{\pi L}{L+\alpha}\right)e^{-\mu y}\sin\left(\frac{\pi y}{L+\alpha}\right)dy,$$

qui est d'ordre $L^{-1}e^{\mu L}$. En utilisant l'expression de L en fonction de N on obtient des ordres de grandeur identiques. Une particule à L produit donc une augmentation substantielle de la taille de la population.

Pour traiter les particules qui n'atteignent pas L, nous commençons par étudier un mouvement Brownien branchant tué en 0 et en L_A

$$L_A=\frac{1}{\sqrt{2}}\big(\log N+3\log\log N-A\big), \tag{4.5}$$

À l'aide de (4.4), nous obtenons des estimées pour les premiers et deuxièmes moments de différentes quantités ce qui nous permet ensuite de calculer le premier et deuxième moment de $Z_N(t)$ conditionnellement à l'état du processus $\theta(\log N)^3$ unités de temps avant. La borne du second moment est suffisante pour établir une loi des grand nombres quand A est grand qui permet de contrôler la distance entre $Z_N(t)$ et sa moyenne. La troncation à L_A est nécessaire car sinon les moments seraient dominés par des événements rares où une particule arrive très loin à droite et produit une grande descendance qui survit.

L'étape suivante consiste à compter le nombre de particules qui atteignent L_A. On peut montrer que si l'on part de N particules dans leur configuration stable, alors il faut attendre environs $(\log N)^3$ unités de temps pour voir une particule qui atteint L_A. On peut se convaincre qu'il s'agit de la bonne échelle de temps avec l'argument

heuristique suivant. Si $\beta > 0$ est une constante, alors le nombre de particules entre $L - \beta$ et L au temps t est d'ordre

$$N \int_{L-\beta}^{L} CLe^{-\mu y} \sin\left(\frac{\pi y}{L}\right) dy,$$

c'est-à-dire d'ordre $1/(\log N)^3$. Chacune de ces particules a une probabilité positive d'atteindre L entre t et $t+1$ mais on peut montrer que les particules qui sont à une distance de L plus grande qu'une constante ne touche L avant $t+1$ qu'avec une faible probabilité. On a donc $O(1/(\log N)^3)$ particules qui atteignent L par unité de temps.

Lorsqu'une particule atteint L_A, elle pourrait avoir branché juste avant de toucher L_A permettant ainsi à un grand nombre de particules d'atteindre elles aussi L_A. On doit donc contrôler le second moment de ce nombre. Á cette fin on montre que conditionnellement à ce qu'au moins une particule atteigne L_A, l'espérance du nombre de particules qui atteignent L_A dans un intervalle de temps de durée $\theta(\log N)^3$, est bornée par une constante.

La clé pour analyser la contribution des particules qui atteignent L_A se trouve dans un résultat de Neveu [116]. Si une particule démarre à L_A et y est une grande constante, alors le nombre de descendants qui atteignent $L_A - y$ est $y^{-1}e^{\sqrt{2}y}W$, où W est une variable aléatoire. À l'aide de théorèmes Taubériens nous démontrons que pour x grand l'on a $P(W > x) \sim B/x$. C'est ce résultat qui conceptuellement explique pourquoi la généalogie est décrite par un coalescent de Bolthausen-Sznitman. La contribution d'une particule en L_A sera approximativement proportionnelle aux nombre de ses descendants qui atteignent $L_A - y$ pour y assez grand. La probabilité qu'une particule en L_A produise un saut de taille au moins x est ainsi proportionnelle à $1/x$ ce qui entraîne que la mesure de Lévy du CSBP limite aura une densité proportionnelle à x^{-2} comme dans le cas du CSBP de Neveu, conduisant ainsi au résultat de dualité avec le coalescent de Bolthausen-Sznitman. Techniquement, la preuve du Théorème 35 se fait en utilisant les flots de ponts introduits par Bertoin et Le Gall dans [29].

4.2.5 Un bref aperçut de la preuve du Théorème 34

L'essentiel du travail pour montrer le Théorème 34 consiste à montrer que pour α fixé, la quantité $Q_\mu(L+\alpha)$ converge vers une limite appelée $\theta(\alpha)$ quand $\epsilon \to 0$. Le Théorème 34 montre que si l'on démarre avec une particule en L, le processus $M_N(t)$ du nombre de particules, une fois renormalisé, converge vers une CSBP de Neveu $(Z_t, t \geq 0)$ issu d'une population initiale de taille W où W est la variable limite de Neveu que l'on vient de voir. Les trajectoires du CSBP de Neveu issu d'une population initiale x sont de deux types : soient elles croissent sur-exponentiellement, soit elles convergent vers 0 (mais n'atteignent pas 0 en temps fini). Si le mécanisme de branchement est $\psi(u) = au + bu\log u$ La probabilité de ce second événement

est e^{-bx}. Tout le travail consiste donc à montrer que la probabilité d'extinction du processus $(M_N(t), t \geq 0)$ converge bien vers $\mathbb{E}(e^{-bW})$ la probabilité que le processus limite $(Z_t, t \geq 0)$ converge vers 0. Si on démarre avec une particule en $L + \alpha$ à présent, les mêmes arguments montrent que la probabilité d'extinction converge vers $\mathbb{E}(e^{-be^{\sqrt{2}\alpha}W})$ et l'on sait déjà que cette fonction de α résout l'équation KPP (4.3).

La seconde partie du théorème qui concerne le comportement de $Q_\mu(x)$ pour x fixé se fait selon les grandes lignes suivantes. Partant d'une particule à x on compte le nombre R de particules qui touchent $L - \alpha$ si on tue les particules à cet instant. On peut monter que

$$\mathbb{E}_x[R] = \frac{e^{\sqrt{2}\alpha}}{\pi\sqrt{2}\alpha} e^{\mu x} \ \sin\left(\frac{\pi x}{L-\alpha}\right) \frac{1}{N(\log N)^2}(1 + C_{N,\alpha})$$

où pour chaque $\alpha > 0, C_{N,\alpha} \to 0$ quand $N \to \infty$. Une borne sur le second moment de R permet alors de conclure que

$$Q_\mu(x) = \frac{e^{\sqrt{2}\alpha}}{\pi\sqrt{2}\alpha} \cdot e^{\mu x} \sin\left(\frac{\pi x}{L}\right) \cdot \frac{1}{N(\log N)^2} \cdot Q_\mu(L-\alpha)(1 + C_{N,\alpha} + o(\alpha^{-1})),$$

où $o(\alpha^{-1})$ est un terme qui tends vers 0 quand $\alpha \to \infty$ et est uniforme en N. Il suffit maintenant d'appliquer la première partie du théorème pour conclure.

Chapitre 5

Trois aspects des processus de branchements.

Ce chapitre est consacré à une brève description de trois articles dont le premier ([26], en révision) est relié au chapitre précédent puisqu'il traite du lien entre l'équation FKPP et certain processus de branchement (les fragmentations homogènes). Le second porte sur un résultat de grande déviation pour un processus de branchement en environnement aléatoire ([18], paru). Enfin le troisième donne un résultat de décomposition en épine dorsale pour les superprocessus ([25], en révision).

5.1 Analogue de l'équation FKPP pour les fragmentations

Les processus de fragmentations homogènes ont été introduits et étudiés par Bertoin (voir sa monographie [34] à ce sujet). Ce sont des processus qui décrivent comment un objet de masse fini se disloque avec le temps. On dit qu'une fragmentation est homogène quand chaque nouveau fragment se met à se disloquer à son tour, indépendamment du reste, avec la même loi que le processus initial (à la taille initiale près), ce que l'on peut voir comme une forme de la propriété de branchement. Ce type de description peut s'appliquer dans de nombreux contextes et à des échelles variées. Les modèles stochastiques de fragmentations (homogènes ou auto-similaires) ont fait l'objet de nombreux travaux ses dernières années (dont mes travaux de thèse [13, 14, 15]).

Dans un travail en commun avec Simon Harris et Andreas Kyprianou [26] (en révision) nous avons exploré et exploité le lien entre fragmentations homogènes et marches branchantes pour formuler et étudier la notion d'équations de Fisher-Kolmogorov-Petrovskii-Piscounov (FKPP) de fragmentation et leur solutions.

Pour expliquer ce lien entre marches branchantes et fragmentations on peut commencer par se placer dans le cadre simplifié des fragmentations conservatives de

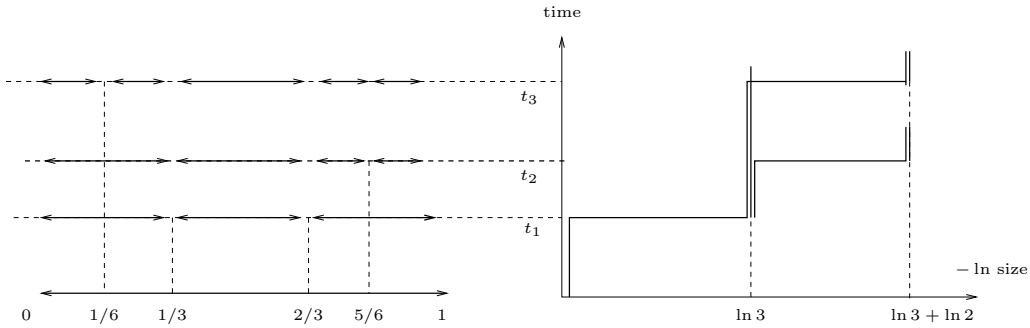

FIGURE 5.1: Le fragment initial $(0,1)$ se divise en trois $(0,1/3),(1/3,2/3),(2/3,1)$ au temps t_1, puis à $t_2,(2/3,1)$ se divise en deux moitiés et finalement au temps $t_3,(0,1/3)$ fait de même.

mesure de dislocation finie. Soit $\nu(\cdot)$ une mesure finie sur $\nabla_1 = \{s_1 \geq s_2 \geq \cdots \geq 0, \sum_i s_i = 1\}$ de masse totale $\nu(\nabla_1) = \gamma$. La fragmentation homogène de mesure de dislocation ν est un processus de Markov $(X(t), t \geq 0)$, à valeur dans ∇_1 (où $X(t) = (X_1(t), X_2(t), \cdots)$). Sa dynamique peut être décrite de la façon suivante. L'état initial est toujours $X(0) = (1, 0, \cdots)$, ce fragment initial de masse 1 attend ensuite un temps exponentiel T_1 de paramètre γ après lequel il se fragmente selon la distribution $X(T_1) \sim \gamma^{-1}\nu(\cdot)$. Chacun de ces nouveaux fragments démarre un nouveau processus de fragmentation de même loi que le processus de départ. C'est-à-dire que chaque fragment $X_i(T_1)$ attend un temps exponentiel-(γ) après lequel il se divise en $(X_i(T_1)s_1, X_i(T_1)s_2, \cdots)$ où $s = (s_1, s_2, \cdots) \sim \gamma^{-1}\nu(\cdot)$ et ainsi de suite. Lorsqu'un fragment se divise, il faut renuméroter tous les fragments puisque les ordres relatifs ont changé.

On peut voir ce processus comme une marche branchante en temps continu en posant $Z(t) = (-\log X_1(t), -\log X_2(t), \cdots)$. Il est immédiat que le processus $Z(t)$ évolue selon la dynamique suivante : si $Z(t) = (z_1, z_2, \cdots)$, alors chaque particule z_i attend un temps exponentiel de paramètre γ puis branche en une descendance dont les distances par rapport à z_i sont données par $(-\log s_1, -\log s_2, \cdots)$ où comme ci-dessus $s \in \nabla_1$ a pour loi $\gamma^{-1}\nu(\cdot)$. La figure 5.1 montre un exemple où $\nu(\cdot) = \delta_{\{1/3,1/3,1/3,0,\cdots\}} + \delta_{\{1/2,1/2,,0,\cdots\}}$.

L'histoire ne s'arrête cependant pas là, et l'on peut définir des fragmentations dont la mesure de dislocation ν est de masse infinie (mais sigma-finie). Dans ce cas les fragmentations arrivent continûment dans le sens où l'on ne peut pas définir de "premier" événement de fragmentation, et le long d'une branche de "l'arbre de fragmentation" les événements de dislocations sont en fait denses dans tout intervalle de temps. La construction de ces processus n'est pas triviale et est principalement due à Bertoin ([28, 34], voir aussi [13]).

Comme pour les coalescence il est pratique de travailler sur l'espace $\mathcal{P}$ des partitions plutôt que sur ∇_1 (l'objet de [13] était justement de montrer que l'on peut passer de l'un à l'autre). On définit donc les fragmentations homogènes comme des

processus de Markov à valeur dans $\mathcal{P}$ que l'on note $\Pi = (\Pi(t), t \geq 0)$.

Définition 38 Soit $\Pi = (\Pi(t), t \geq 0)$ un processus de Markov à valeur dans $\mathcal{P}$ à trajectoires presque sûrement càdlàg [1] On dit que Π est une fragmentation homogène si son semi-groupe vérifie la *propriété de fragmentation* : pour tout $t, t' \geq 0$ la loi conditionnelle de $\Pi(t+t')$ sachant $\Pi(t) = \pi$ est la collection de blocs que l'on obtient en fragmentant chaque bloc $\pi_i, i = 1, \cdots$ de π par une copie indépendant de $\Pi^{(i)}(t')$. On impose en outre que $\Pi(0)$ est la partition triviales constituée d'un seul bloc contenant tout $\mathbb{N}$ (cf. [34] pour plus de détails).

Rappelons qu'une partition $\pi = (\pi_1, \pi_2, \cdots) \in \mathcal{P}$ admet des fréquences asymptotiques si pour chaque i la limite

$$|\pi_i| = \lim_{n\to\infty} \frac{\#\pi_i \cap [n]}{n}$$

existe. La notation $|\pi| = (|\pi_i|, i \in \mathbb{N})$ désigne le réarrangement en ordre décroissant des fréquences des blocs. Il est connu que si Π est une fragmentation homogène alors presque sûrement, pour tout $t \geq 0, \Pi(t)$ admet des fréquences asymptotiques. Le processus $(|\Pi(t)|, t \geq 0)$est ce que l'on appelle une fragmentation de masse. On peut définir directement ces processus (i.e. des processus de Markov vivant dans ∇_1 ayant la propriété de fragmentation) mais toutes les fragmentations de masses peuvent être obtenues comme fréquences d'un fragmentation à valeur dans $\mathcal{P}$ (c'était l'un des principaux résultats que j'avais obtenu dans [13]).

Au vu du caractère "processus à accroissement indépendants et stationnaires" de la définition, il n'est pas surprenant que les fragmentations homogènes admettent une description à la Lévy-Kintchin. En effet, leur loi est complètement caractérisée par une mesure ν sur $\nabla_1^{(-)} := \{s_1 \geq s_2 \geq \ldots \geq 0, \sum_i s_i \leq 1\}$ (*la mesure de dislocation*) telle que

$$\int_{\nabla_1^{(-)}} (1 - s_1)\nu(ds) < \infty \tag{5.1}$$

et un paramètre $\mathtt{c} \geq 0$ (*le taux d'érosion*). On interprète $\nu(ds)$ comme le "taux" auquel un fragment donné se divise en sous-fragments dont les tailles relatives sont données par $s = (s_1, s_2, \ldots)$. Dans [26] on suppose toujours que $c = 0$ (pas d'érosion) et $\mathrm{supp}(\nu) \subset \nabla_1$ (fragmentation conservative).

Une partie de l'information concernant ν est capturée par la fonctionnelle

$$\Phi(q) := \int_{\nabla_1} \left(1 - \sum_{i=1}^{\infty} s_i^{q+1}\right) \nu(ds), \ q > \underline{p}, \tag{5.2}$$

où

$$\underline{p} := \inf\left\{p \in \mathbb{R} : \int_{\nabla_1} \sum_{i=2}^{\infty} s_i^{p+1}\nu(ds) < \infty\right\} \leq 0.$$

[1] la métrique sur $\mathcal{P}$ est donné par $d(\pi, \pi') = 2^{-n(\pi,\pi')}$ où $n(\pi, \pi')$ est le plus grand entier tel que $\pi_{|[n]} = \pi'_{|[n]}$.

On peut sans perte de généralité supposer $\underline{p} < 0$. Il est bien connu que $p \mapsto \Phi(p)$ est l'exposant de Laplace d'un subordinateur naturellement associé à la fragmentation (le log de la taille du fragment typique).

La fonction Φ est donc strictement croissante, concave, analytique et telle que $\Phi(0) = 0$. L'équation

$$(p+1)\Phi'(p) = \Phi(p)$$

sur $p > \underline{p}$ a une une unique solution dans $(0, \infty)$ que l'on note $\overline{p}$ (cf. [30]). En outre $(p+1)\Phi'(p) - \Phi(p) > 0$ où $p \in (\underline{p}, \overline{p})$. La fonction

$$c_p := \frac{\Phi(p)}{p+1} \tag{5.3}$$

atteint donc son unique maximum sur $(\underline{p}, \infty)$ à $\overline{p}$ et ce maximum vaut $\Phi'(\overline{p})$.

Soit donc une fragmentation homogène Π de mesure de dislocation ν. L'équation

$$-c\psi'(x) + \int_{\nabla_1} \{\prod_i \psi(x - \log s_i) - \psi(x)\}\nu(ds) = 0,\ x \in \mathbb{R} \tag{5.4}$$

est l'équation d'*onde voyageuse de fragmentation* de vitesse $c \in \mathbb{R}$. Cette équation (5.4) est l'analogue pour le contexte des fragmentations de la forme "onde voyageuse" de l'équation de FKPP classique. Notre résultat principal concerne l'existence, l'unicité et la représentation des solutions de cette équation.

Introduisons une classe de fonctions dans laquelle nous allons prouver l'unicité.

Définition 39 L'ensemble des fonctions $\psi \in C^1(\mathbb{R})$ qui sont monotones croissantes et telles que $\psi(-\infty) = 0$ et $\psi(\infty) = 1$ est appelé $\mathcal{T}_1$. Pour chaque $p \in (\underline{p}, \overline{p}]$ on définit $\mathcal{T}_2(p) \subset \mathcal{T}_1$ comme étant l'ensemble des fonctions $\psi_p \in \mathcal{T}_1$ telles que $L_p(x) := e^{(p+1)x}(1 - \psi_p(x))$ est monotone croissante.

Le résultat principal est le suivant :

Théorème 40 (B., Harris, Kyprianou [26]) *Soit $p \in (\underline{p}, \overline{p}]$ fixé et supposons que $\psi_p : \mathbb{R} \mapsto (0, 1]$ appartient à $\mathcal{T}_2(p)$. Alors*

$$M(t, p, x) := \prod_i \psi_p(x - \log|\Pi_i(t)| - c_p t),\ t \geq 0 \tag{5.5}$$

est une $\mathbb{F}$-martingale si et seulement si ψ_p résout (5.4) avec $c = c_p$. En outre, à une translation près, il n'existe qu'une seule telle fonction $\psi_p \in \mathcal{T}_2(p)$ qui est donné par

$$\psi_p(x) = \mathbb{E}(\exp\{-e^{-(p+1)x}\Delta_p\}) \tag{5.6}$$

où $\Delta_p > 0$ p.s. est $\mathcal{F}_\infty$-mesurable (il s'agit de la limite de la martingale additive ou de la martingale dérivée associée au paramètre p).

Pour présenter brièvement et de manière informelle le schéma de la preuve de ce résultat il est pratique d'introduire l'opérateur intégro-différentiel

$$\mathcal{A}_p\psi(x) = -c_p\psi'(x) + \int_{\nabla_1} \{\prod_i \psi(x - \log s_i) - \psi(x)\}\nu(ds).$$

Le second terme dans cet expression est un opérateur de saut dirigé par la mesure ν et est donc très proche du générateur de la fragmentation elle-même.

On dit que $\psi_p \in \mathcal{T}_2(p)$ est une *fonction martingale multiplicative* si avec ce choix de ψ (5.5) est une martingale. L'équivalence entre d'une part la propriété analytique $\mathcal{A}_p\psi_p \equiv 0$ et d'autre part le fait que ψ_p est une fonction martingale multiplicative est une conséquence de la représentation classique de Feynman-Kac et de la représentation en semi-martingale de $M(t, p, x)$.

L'unicité de la solution dans $\mathcal{T}_2(p)$ est plus délicate. On commence par étudier l'asymptotique des fonctions martingales multiplicatives. Pour cette discussion on se place dans le cas $p \in (\underline{p}, \overline{p})$.

Soit $\psi_p \in \mathcal{T}_2(p)$ un fonction qui fait de (5.5) une martingale. Comme M est une martingale uniformément intégrable (puisque bornée dans $[0, 1]$) $\psi_p(x) = M(0, p, x) = E(M(\infty, p, x))$. On cherche donc à étudier la limite $M(\infty, p, x)$. En passant au log on voit que

$$-\log M(\infty, p, x) = -\lim_{t\uparrow\infty} \sum_i \log \psi_p(x - \log|\Pi_i(t)| - c_p t).$$

On a défini $L_p(x) = e^{(p+1)x}(1 - \psi_p(x))$, ainsi, si l'on remplace $-\log x$ par son développement $(1 - x)$ (en supposant un argument asymptotiquement proche de 1) et que l'on multiplie par $e^{(p+1)x}$ on voit que

$$-e^{(p+1)x} \log M(\infty, p, x) \approx \sum_i |\Pi_i(t)|^{p+1} e^{\Phi(p)t} L_p(x - \log|\Pi_i(t)| - c_p t) \tag{5.7}$$

quand $t \uparrow \infty$. Or $W(t, p, x) = \sum_i |\Pi_i(t)|^{p+1} e^{\Phi(p)t}$ est ce que l'on appelle la martingale additive qui converge vers $W(\infty, p)$. D'autre part, pour chaque $i \in \mathbb{N}$ fixé on sait que $-\log(|\Pi_i(t)|) - c_p t$ est le subordinateur de Bertoin associé au fragment marqué moins une dérive, et il s'agit donc d'un processus de Lévy sans saut positif. On a donc une loi des grands nombres

$$-\log(|\Pi_i(t)|) - c_p t \sim \alpha_p t$$

quand $t \uparrow \infty$ pour une certaine constante $\alpha_p > 0$. Intuitivement, si l'on peut remplacer $L_p(x - \log|\Pi_i(t)| - c_p t)$ par $L(x + \alpha_p t)$ dans (5.7), on peut factoriser les termes L_p pour obtenir

$$\frac{-e^{(p+1)x} \log M(\infty, p, x)}{W(\infty, p)} \approx L_p(x + \alpha_p t)$$

quand $t \uparrow \infty$. Comme les deux limites de martingales $M(\infty, p, x)$ et $W(\infty, p)$ sont non-triviales on voit que $\lim_{z\to\infty} L_p(z) = k_p$ pour une certaine constante $k_p \in (0, \infty)$.

Une conséquence directe de ceci est que quel que soit la fonction martingale ψ_p dont on est parti, la limite L^1, $M(\infty, p, x)$ est toujours égale (modulo une constante additive en x) à $\exp\{-e^{-(p+1)x}W(\infty, p)\}$ ce qui montre l'unicité.

L'une des difficultés techniques avec ce programme vient du fait que l'on a besoin d'un contrôle uniforme sur les termes $L_p(x - \log|\Pi_i(t)| - c_p t)$ afin de factoriser l'approximation commune $L_p(x+\alpha_p t)$. Une solution pourrait être de contrôler la position de la particule la plus à gauche (c'est par exemple la voie choisie dans [89]). Il semble cependant qu'une autre approche soit préférable dans ce contexte. L'idée consiste à définir les martingales le long de ce que l'on appelle des lignes d'arrêts. Cette technique est aujourd'hui classique dans le contexte des marches branchantes (voir par exemple Biggins et Kyprianou [37, 38]), et du mouvement Brownien branchant (voir par exemple Neveu [116] et Kyprianou [102]).

L'uniforme intégrabilité des martingales $M(t, p, x)$, et $W(t, p, x)$ permet de les voir comme la projection de leurs limites sur $\mathcal{F}_t$, l'information contenue dans l'arbre de fragmentation "coupé" au temps t. Mais on peut aussi projeter ces limites sur d'autres filtrations que l'on génère ne coupant l'arbre différemment, selon des familles croissantes de "lignes d'arrêt".

Une instance d'une telle famille est étudiée par Bertoin et Martinez [32] qui *gèlent* les fragments dés que leur taille devient inférieur à un certain seuil $e^{-z}, z \geq 0$. Au bout d'un temps fini, tous les fragments sont de taille inférieure à ce seuil et le processus est complètement arrêté. On a ainsi "coupé" l'arbre de fragmentation. La collection des fragments $(\Pi_i(\ell^z) : i \in \mathbb{N})$ que l'on obtient par cette procédure génère une autre filtration $\{\mathcal{G}_{\ell^z} : z \geq 0\}$. En projetant les limites $M(\infty, p, x)$ et $W(\infty, p, x)$ sur cette filtration, on définit deux nouvelles martingales uniformément intégrables qui ont toujours la même limite. Ces martingales ressemblent à $M(t, p, x)$ et $W(t, p)$, respectivement, mais le rôle de $(\Pi_i(t) : i \in \mathbb{N})$ est maintenant joué par $(\Pi_i(\ell^z) : i \in \mathbb{N})$. Considérons par exemple le cas $p = 0$, et donc $c_p = 0$, on peut reprendre l'argument heuristique donné au paragraphe précédent pour cette famille de lignes d'arrêt et mettre à profit le contrôle uniforme que l'on a sur les tailles des fragments de $(\Pi_i(\ell^z) : i \in \mathbb{N})$. Une modification pour le cas $p \neq 0$ de ce type d'argument joue un rôle essentiel dans notre preuve.

5.2 Grandes déviations pour les processus de branchement en environnement aléatoire

Les processus de branchement en environnement aléatoire sont une généralisation des processus de Galton Watson où, à chaque génération, la loi de reproduction est tirée au hasard. Avec Vincent Bansaye, nous avons donné plusieurs résultats dans [18] de type *grandes déviations*. Plus précisément, un processus de branchement en environnement aléatoire (BPRE pour faire court) $(Z_n, n \in \mathbb{N})$ de loi d'environnement $\mu \in \mathcal{M}_1(\mathbb{P})$ est un processus de Markov à temps discret qui évolue selon la dynamique suivante : au temps n on tire $\mathbf{p}$ selon μ indépendamment du passé et chaque individu

$i = 1, \ldots, Z_n$ se reproduit indépendamment des autres selon $\mathbf{p}$. On note $\mathbb{P}_{z_0}$ la loi de ce processus issu de z_0 individus initiaux.

On comprend aisément que le comportement de Z_n est lié à celui du processus $M_n = \Pi_1^n m(\mathbf{p}_i)$ et que leurs comportement typiques se correspondent :

$$Z_n \approx M_n, \qquad (n \in \mathbb{N}).$$

Il n'est donc pas surprenant que l'on ait la dichotomie suivante :un BPRE est surcritique (resp. critique, resp. souscritique) si l'espérance de $\log(m(\mathbf{p}))$ (où $m(\mathbf{p})$ est la moyenne de $\mathbf{p}$) par rapport à la loi des environnements μ

$$\mathbb{E}(\log(m(\mathbf{p}))),$$

est positive (resp. zéro, resp. négative). Dans le cas surcritique, le BPRE survit avec une probabilité positive, dans le cas critique ou souscritique il s'éteint presque sûrement.

En outre, dans le cas surcritique, en supposant $\mathbb{E}(\sum_{k\in\mathbb{N}} k^s \mathbf{p}(k)/m(\mathbf{p})) < \infty$ pour un certain $s > 1$, il existe une variable aléatoire$W$ non triviale telle que

$$M_n^{-1} Z_n \overset{n\to\infty}{\longrightarrow} W, \qquad \mathbb{P}(W > 0) = \mathbb{P}(\forall n, Z_n > 0).$$

ce qui assure que conditionnellement à la non-extinction de $(Z_n)_{n\in\mathbb{N}}$

$$\log(Z_n)/n \to \mathbb{E}(\log(m(\mathbf{p}))) \quad \text{p.s.}$$

Les événements de grandes déviations que nous avons considérés sont de la forme

$$A_c(n) = \begin{cases} \{0 < \frac{1}{n}\log Z_n \leq c\} \text{ pour } c < \mathbb{E}(\log(m(\mathbf{p})) \\ \{\frac{1}{n}\log Z_n \geq c\} \text{ pour } c > \mathbb{E}(\log(m(\mathbf{p})) \end{cases}$$

On souhaite connaître le taux de décroissance de la probabilité de $A_c(n)$, i.e. on veut un résultat du type

$$-\frac{1}{n}\log(\mathbb{P}(A_c(n))) \to \chi(c), \qquad \text{with } \chi(c) < \infty.$$

Notre résultat principal concerne le cas *fortement surcritique*

$$\mathbb{P}(\mathbf{p}(0) = 0) = 1$$

(en particulier les environnements sont presque sûrement surcritiques).

On pose

$$\bar{L} := \mathbb{E}(\log(m(\mathbf{p}))).$$

Pour chaque $c \leq \bar{L}$, on pose

$$\chi(c) \;:=\; \inf_{t\in[0,1]} \{-t\log(\mathbb{E}(\mathbf{p}(1))) + (1-t)\psi(c/(1-t))\}.$$

il est aisé de voir que cet infimum est atteint en un point unique t_c par des arguments de convexité. Ainsi

$$\chi(c) = -t_c \log(\mathbb{E}(\mathbf{p}(1))) + (1-t_c)\psi(c/(1-t_c)),\ t_c \in [0, 1-c/\bar{L}].$$

On définit la fonction $f_c : [0,1] \mapsto \mathbb{R}_+$ pour chaque $c < \bar{L}$ par (cf. figure 1) :

$$f_c(t) := \begin{cases} 0, & \text{if } t \leq t_c \\ \frac{c}{1-t_c}(t-t_c), & \text{if } t \geq t_c. \end{cases}$$

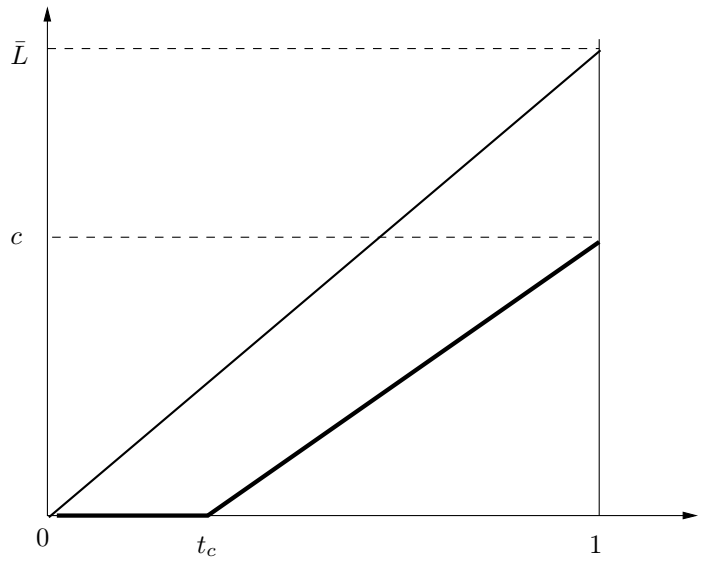

FIGURE 5.2: La fonction $t \mapsto f_c(t)$ pour $c \leq \bar{L}$.

Nous aurons également besoin de l'hypothèse de moment suivante $\mathcal{H}$.

$$\left\{ \begin{array}{c} \exists A > 0 \text{ t.q. } \mu(m(\mathbf{p}) > A) = 0, \\ \exists B > 0 \text{ t.q. } \mu(\sum_{k\in\mathbb{N}} k^2 \mathbf{p}(k) > B) = 0 \end{array} \right\} \tag{$\mathcal{H}$}$$

Le Théorème suivant donne la fonctionnelle de taux pour le coût de grande déviation de $Z_n \leq \exp(cn)$ et la trajectoire asymptotique de Z_n conditionnellement à $Z_n \leq \exp(cn)$.

Théorème 41 (B., Bansaye, [18]) *En supposant* $\mathbb{P}(\mathbf{p}(0) = 0) = 1$ *et* $\mathcal{H}$ *on a*

(a) Si $\mu(\mathbf{p}(1) > 0) > 0$, *alors pour tout* $c < \bar{L}$,

$$-\log(\mathbb{P}(Z_n \leq e^{cn}))/n \overset{n\to\infty}{\longrightarrow} \chi(c),$$

et en outre, conditionnellement à $Z_n \leq e^{cn}$,

$$\sup_{t\in[0,1]} \{ \big| \log(Z_{[tn]})/n - f_c(t) \big| \} \overset{n\to\infty}{\longrightarrow} 0, \qquad \textit{in } \mathbb{P}.$$

(b) Si $\mu(\mathbf{p}(1) > 0) = 0$, alors pour tout $c < \bar{L}$,

$$-\log(\mathbb{P}(Z_n < e^{cn}))/n \overset{n\to\infty}{\longrightarrow} \psi(c),$$

et en outre pour chaque $\inf\{supp \log(m(\mathbf{p}))\} < c < \bar{L}$*, conditionnellement à* $Z_n \leq e^{cn}$*,*

$$\sup_{t\in[0,1]} \{\big| \log(Z_{[tn]})/n - ct\big|\} \overset{n\to\infty}{\longrightarrow} 0, \qquad in\ \mathbb{P}.$$

On peut formuler quelques remarques concernant ce résultat

- Si $\mu(\mathbf{p}(1) > 0) > 0$, alors t_c -le point où la trajectoire "décolle"- peut être soit zéro, soit $1 - c/\bar{L}$, ou encore appartenir à $(0, 1 - c/\bar{L})$.

- Lorsque $m := m(\mathbf{p})$ est déterministe comme dans le cas d'un processus de Galton Watson
  - si $\mu(\mathbf{p}(1) > 0) > 0$ (cas Böttcher), alors $t_c = 1 - c/\log(m)$ et $\chi(c) = t_c \log(\mathbb{E}(\mathbf{p}(1)))$.
  - If $\mu(\mathbf{p}(1) > 0) = 0$ (cas Schröder), alors $\chi(c) = -\infty$.

- Ce résultat a une interprétation heuristique simple. On observe que

$$\mathbb{P}(Z_k = 1, k = 1, \ldots, tn) = \mathbb{E}(\mathbf{p}(1))^{tn} = \exp(\log(\mathbb{E}(\mathbf{p}(1)))tn)$$

  et que

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_{(1-t)n}/n \leq c) = (1-t)\psi(c/(1-t)).$$

  Ceci suggère que

$$\mathbb{P}(Z_k = 1, k = 1, \ldots, tn\,; S_n - S_{tn} \leq cn) \asymp \exp(n[t\log(\mathbb{E}(\mathbf{p}(1))) + (1-t)\psi(c/(1-t))])$$

  et dans ce cas $\chi(c)$ est simplement le coût d'un tel événement avec le choix 'optimal" de t. Il n'est pas dur de voir que $\{Z_k = 1, k = 1, \ldots, tn\,; S_n - S_{tn} \leq cn\}$ est asymptotiquement inclus dans $\{Z_n \leq cn\}$ et que donc $\chi(c)$ est une borne supérieure pour la fonction de taux de Z_n. Si l'on ajoute qu'une fois que $Z_n \gg 1$ est assez grand il ne peut plus dévier de la marche aléatoire S_n associée à la suite des environnements, on voit que χ est un très bon candidat pour la fonction de taux.

- Pour dévier sous c, le processus $(\log(Z_{[nt]})/n)_{t\in[0,1]}$ commence donc par rester borné jusqu'à un temps optimal t_c puis rejoint c en ligne droite en suivant une suite d'environnements adéquats.

5.3 Le squelette des superprocessus

De nombreux auteurs se sont intéressés à des résultats de décomposition pour les superprocessus ou les processus de branchement. Le cas des superprocessus sur-critiques de mécanisme de branchement quadratique a ainsi été étudié par Evans et O'Connell, puis par Engländer et Pinsky. La décomposition qu'ils ont introduite permet d'écrire la distribution du superprocessus à un temps $t \geq 0$ fixé comme la somme de deux processus indépendants. Le premier est une copie du processus originel conditionné à s'éteindre, le second peut être vu comme la somme de copies du processus de départ conditionnées à s'éteindre qui auraient immigré continûment le long de la trajectoire d'un processus de branchement dyadique auxiliaire dont l'état initial est un nombre Poissonien de particules. Ce processus auxiliaire de branchement est ce que l'on peut appeler le squelette du processsus (*backbone* en anglais)[2]. Dans [77] et [76] la décomposition s'obtient à travers l'équation d'évolution du semi-groupe sans construction trajectorielle.

Plus récemment, Duquesne et Winkel [74] ont obtenu une version d'une telle décomposition, sans mouvement spatial, dans un cadre bien plus général. Dans leur cas le branchement quadratique est remplacé par un mécanisme de branchement ψ général qui est l'exposant de Laplace d'un processus de Lévy sans saut négatif et tel que $0 < -\psi'(0+) < \infty$ et $\int^\infty 1/\psi(\xi)\mathrm{d}\xi < \infty$. En outre, cette décomposition est *trajectorielle* et se décrit à travers la croissance des arbres généalogiques inclus dans le CSBP sous-jacent. Dans leur cas, le squelette est un processus de Galton-Watson à temps continu et le caractère général du mécanisme de branchement induit trois ingrédients différents pour décrire l'immigration. On retrouve la copie conditionnée à s'éteindre et l'immigration continue le long du squelette, qui est décrite par un processus ponctuel de Poisson le long du squelette dont l'intensité est une mesure d'excursion qui assigne une masse initiale nulle et un temps de vie fini aux processus immigrés. Le premier nouvel ingrédient est un second processus ponctuel de Poisson le long du squelette dont les atomes sont des réels positifs. Chacun de ces atomes correspond à l'immigration d'un processus de branchement de masse initiale cet atome et dont la loi est celle du processus initial conditionné à s'éteindre. Finalement, aux instants de branchements du squelette, des copies indépendantes du processus originel conditionné à s'éteindre sont greffées sur le processus avec une masse initiale aléatoire qui dépend du nombre d'enfants au point de branchement considéré.

Indépendamment de ce travail, Bertoin et al. [35] ont obtenu des résultats très similaires. Ils identifient le squelette sus-mentionné comme l'arbre des individus prolifiques, c'est à dire les individus dont la descendance devient infinie.

En collaboration avec A. Kyprianou et A. Murillo-Salas, [25] nous avons développé la décomposition de Duquesne et Winkel [66] dans plusieurs directions. Nous autorisons la possibilité $\int^\infty 1/\psi(\xi)\mathrm{d}\xi = \infty$ ce qui inclut le cas des superprocessus dont la masse peut, avec probabilité positive, tendre vers 0 sans jamais s'annuler. Ceci per-

[2]Par opposition à l'épine dorsale (*spine*) ou à la particule immortelle qui apparaît dans une autre décomposition, cf. Roelly-Coppoletta and Rouault (1989) Evans (1993)

met également de traiter le cas de mécanismes de branchement qui correspondent à des processus de Lévy spectralement positifs à variation bornée (cas exclu dans [66]). Finalement notre décomposition est écrite dans le cadre de superprocessus avec déplacement spatial. Les preuves que nous donnons sont également fondamentalement différentes de celles de [66] et reposent sur des techniques de manipulation de semi-groupe dans l'esprit de [77] et sur l'utilisation des $\mathbb{N}$-mesures de Dynkin et Kuznetsov [74].

Bibliographie

[1] Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, Wakeley J. (2004) A robust measure of HIV-1 population turnover within chronically infected individuals. *Mol. Biol. Evol.* **21** 1902–191

[2] Aldous, D. (1985). Exchangeability and related topics. In Hennequin, P., editor, Ecole d'Été de Probabilités de Saint-Flour XIII - 1983, pages 1–198. Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117.

[3] D. Aldous (1991). The continuum random tree I. *Ann. Probab.* **19**, 1-28.

[4] D. Aldous (1993). The continuum random tree III. *Ann. Probab.* **21**, 248-289.

[5] D. J. Aldous (1998). Brownian excusion conditioned on its local time. *Elect. Comm. in Probab.*, 3, 79–90.

[6] Aldous, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation) : a review of the mean-field theory for probabilists. *Bernoulli*, **5**, 3-48.

[7] Árnason, E. (2004). Mitochondrial cytochrome b DNA variation in the high-fecundity atlantic cod : trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**, 1871–1885

[8] O. Angel, N. Berestycki and V. Limic. Global divergence of spatial coalescents. *Soumis.*

[9] Jeffrey E. Barrick, Dong Su Yu, Sung Ho Yoon, Haeyoung Jeong, Tae Kwang Oh, Dominique Schneider, Richard E. Lenski and Jihyun F. Kim (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli *Nature* **461**, 1243-1247

[10] Basdevant, A.L. ; Goldschmidt, C. . Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **13** (2008), no. 17, 486–512.

[11] J. Bérard and J.-B. Gouéré (2008). Brunet-Derrida behavior of branching selection particle systems on the line. À Paraître dans *Communications in Mathematical Physics.* `http://arxiv.org/pdf/0811.2782v3.`

[12] J. Bérard and J.-B. Gouéré (2009). Survival probability of the branching random walk killed below a linear boundary. `http://arxiv.org/pdf/0911.3755v1`

[13] J. Berestycki (2002). Ranked fragmentations. *ESAIM Probab. Statist.* **6**, 157–175 (electronic)

[14] J. Berestycki (2003). Multifractal spectra of fragmentation processes. *J. Statist. Phys.* **113**(3), 411-430.

[15] J. Berestycki (2004). Exchangeable fragmentation-coalescence processes and their equilibrium measures. *Electron. J. Probab.* **9**, no. 25, 770–824 (electronic).

[16] J. Berestycki, N. Berestycki (2009).Kingman's coalescent and Brownian motion. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** , 239–259

[17] N. Berestycki, *Recent progress in coalescent theory.* Ensaios Matemáticos [Mathematical Surveys], 16. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009. 193 pp

[18] V. Bansaye, J. Berestycki (2009). Large deviations for branching processes in random environment. *Markov Process. Related Fields* **15** , no. 4, 493–524

[19] J. Berestycki, N. Berestycki and V. Limic (2010). The Λ-coalescent speed of coming down from infinity. *Ann. Probab.* **38**, no. 1, 207–233

[20] J. Berestycki, N. Berestycki and V. Limic (2010). Sampling formulae for ?-coalescents. *En preparation*

[21] J. Berestycki, N. Berestycki and J. Schweinsberg (2008). Small-time behavior of beta-coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** no. 2, 214–238.

[22] J. Berestycki, N. Berestycki and J. Schweinsberg (2007). Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** , no. 5, 1835–1887.

[23] J. Berestycki, N. Berestycki and J. Schweinsberg (2010). The genealogy of branching Brownian motion with absorption. *Soumis*

[24] J. Berestycki, N. Berestycki and J. Schweinsberg (2010). Survival probability for the branching Borwnian motion with absorption. *En préparation.*

[25] J. Berestycki, A. E. Kyprianou and A. Murillo (2010). The prolific backbone for supercritical superdiffusions. *En révision*

[26] J. Berestycki, S.C Harris and A. E. Kyprianou (2010). Travelling waves and homogeneous fragmentation. *En révision*

[27] J. Bertoin and J.-F. Le Gall (2000). The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields* **117**, 249–266.

[28] Bertoin, J. (2001) Homogeneous fragmentation processes. *Probab. Theory Related Fields* **121**, 301–318.

[29] J. Bertoin and J.-F. Le Gall (2003). Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields* **126**, 261–288.

[30] Bertoin, J. (2003) Asymptotic behaviour of fragmentation processes. J. Europ. Math. Soc. **5**, 395 -416.

[31] J. Bertoin and J.-F. Le Gall (2005). Stochastic flows associated to coalescent processes II : Stochastic differential equations. *Ann. Inst. H. Poincaré Probabilités et Statistiques*, 41, 307–333.

[32] Bertoin, J. and Martinez, S. (2005) Fragmentation energy. *Adv. Appl. Probab.* **37**, 553-570.

[33] J. Bertoin and J.-F. Le Gall (2006). Stochastic flows associated to coalescent processes III : Limit theorems. *Illinois J. Math*, 50, 147–181.

[34] Bertoin, J. (2006) *Random fragmentation and coagulation processes.* Cambridge University Press.

[35] Bertoin, J., Fontbona, J. and Martínez, S. (2008) On prolific individuals in a continuous-state branching process. *J. Appl. Probab.* **45** 714-726.

[36] D. Beysens, X. Campi, and E. Peffekorn, editors. *Proceedings of the workshop : Fragmentation phenomena*, Les Houches Series. World Scientific, 1995.

[37] Biggins, J. and Kyprianou, A.E. (1997) Seneta-Heyde norming in the branching random walk. *Ann. Probab.* **25** 337-360.

[38] Biggins, J. and Kyprianou, A.E. (2005) The smoothing transform : the boundary case. *Electr. J. Probab.* **10**, 609-631.

[39] E. Brunet and B. Derrida. (1997). Shift in the velocity of a front due to a cuto ?. *Phys. Rev. E (3)*, 56(3, part A) :2597 ?2604, 1997.

[40] E. Brunet, and B. Derrida, Microscopic models of traveling wave equations. *Computer Physics Communications* **121-122** (1999), 376-381

[41] Éric Brunet et Bernard Derrida (2001) Effect of Microscopic Noise on Front Propagation. *Journal of Statistical Physics,* **103**, Nos. 1/2,

[42] E. Brunet and B. Derrida (2004). Exactly soluble noisy traveling-wave equation appearing in the problem of directed polymers in a random medium, *Physical Review E* , **70**, 016106.

[43] E. Brunet, B. Derrida, A. H. Mueller, S. Munier (2006). Noisy traveling waves : effect of selection on genealogies. *Europhys.Lett.*, **76** 1-7.

[44] E. Brunet, B. Derrida, A. H. Mueller, S. Munier (2006) A phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys.Rev. E*, **73**, 056126.

[45] E. Brunet. (2007) *L'équation de Fisher-KPP en présence d'un bruit.* Notes de cours non publiées.

[46] E. Brunet, B. Derrida, A. H. Mueller, S. Munier (2007). Effect of selection on ancestry : an exactly soluble case and its phenomenological generalization. *Physical Review E*, **76**, 041104.

[47] Igor M. Rouzine, Éric Brunet and Claus O. Wilke (2008), The traveling wave approach to asexual evolution : Muller's ratchet and speed of adaptation, *Theoretical Population Biology* **73**, 24–46.

[48] Éric Brunet, Igor M. Rouzine and Claus O. Wilke, (2008) The stochastic edge in adaptive evolution, *Genetics*, **179**, 603–620.

[49] Éric Brunet, Bernard Derrida et Damien Simon (2008). Universal tree structures in directed polymers and models of evolving populations, *Physical Review E* **78**, 061102.

[50] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger (2005). Alpha-stable branching and beta-colaescents. *Electron. J. Probab.* **10**, no. 9, 303–325 (electronic).

[51] M. Birkner and J. Blath, (2008). Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. *J. Math. Biol.* **57**, no. 3, 435–465.

[52] Birkner, M. (2005) *Stochastic models from population biology* lecture notes for a course at TU Berlin. Available at `http://joguinf.informatik.uni-mainz.de/~birkner/lehre_archiv/smpb-30.6.05.pdf`.

[53] R. M. Blumenthal and R. K. Getoor (1961). Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.*, 10, 493–516.

[54] E. Bolthausen and A.-S. Sznitman (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197**, 247–276.

[55] J. F. Crow and M. Kimura, (1970) Introduction to Population Genetics Theory. Harper & Row, New York

[56] Darling, R. W. R. and Norris, J. R. Differential equation approximations for Markov chains. *Probab. Surv.* **5** (2008), 37–79.

[57] B. Derrida and D. Simon, The survival probability of a branching random walk in presence of an absorbing wall, Europhys. Lett. 78 (2007), 60006.

[58] P. Donnelly, and T. Kurtz. (1996) A Countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 698–742.

[59] P. Donnelly, and T. Kurtz. (1999) Particle Representations for Measure-Valued Population Models. *Ann. Probab.* **27**, 166–205.

[60] R. Dong, A. Gnedin, and J. Pitman, Exchangeable partitions derived from Markovian coalescents. *Ann. Appl. Probab.* **17** (2007), no. 4, 1172–1201.

[61] Drake, J. W., B. Charlesworth and D. Charlesworth, 1998 Rates of spontaneous mutation. Genetics 148 : 1667–1686.

[62] M.Drmota, *Random trees, An interplay between combinatorics and probability*, Springer Wien New York, Vienna, 2009.

[63] T. Duquesne and J.-F. Le Gall (2002). *Random Trees, Lévy Processes, and Spatial Branching Processes. Astérisque* **281**.

[64] T. Duquesne and J.-F. Le Gall (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theor. Rel. Fields*, **131**, 553–603.

[65] Duquesne, T. (2003). A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.* **31**, 996-1027.

[66] Duquesne, T. and Winkel, M. (2007) Growth of Lévy trees. *Probab. Theory Relat. Fields*, **139**, 313-371.

[67] R. Durrett (2004). *Probability : theory and examples.* Duxbury advanced series, 3rd edition.

[68] R. Durrett (2002). *Probability Models for DNA Sequence Evolution.* Springer Verlag.

[69] Durrett, R. and Mayberry, J., Evolution in predator-prey systems. *Stoch. Proc. Appl.* **120** (2010), 1364-1392

[70] Durrett, R. and Remenik, D., Brunet-Derrida particle systems, free boundary problems and Weiner-Hopf equations. À paraître dans *Ann. Prob*

[71] R. Durrett and J. Schweinsberg (2005). A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stochastic Process. Appl.* **115**, no. 10, 1628–1657.

[72] R. Durrett and J. Schweinsberg (2005). Random partitions approximating the coalescence of lineages during a selective sweep. *Ann. Appl. Probab.* **15** (2005), no. 3, 1591–1651.

[73] E.B. Dynkin (1965). *Markov Processes, Vol. 1.* Springer-Verlag, Berlin.

[74] Dynkin, E.B. and Kuznetsov, S.E. (2004) $\mathbb{N}$-measures for branching exit Markov systems and their applications to differential equations. *Probab. Theory Relat. Fields* **130**, 135-150.

[75] B. Eldon and J. Wakeley (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, **172**, 2621–2633.

[76] Engländer, J. and Pinsky, R.G. (1999) On the construction and support properties of measure-valued diffusions on $D \subseteq R^d$ with spatially dependent branching. *Ann. Probab.* 27(2) 684-730.

[77] Evans, S. N. and O'Connell, N. (1994) Weighted Occupation Time for Branching Particle Systems and a Representation for the Supercritical Superprocess. *Canad. Math. Bull.*, 37 187-196.

[78] S. N. Evans. (2000). Kingman's coalescent as a random metric space. In *Stochastic Models : A Conference in Honour of Professor Donald A. Dawson* (L. G. Gorostiza and B. G. Ivanoff eds.) Canadian Mathematical Society/American Mathematical Society.

[79] Ewens, Warren J. (1972). The sampling theory of selectively neutral alleles, *Theoretical Population Biology*, **3**, 87–112.

[80] Ewens, Warren J. (2004) *M*athematical population genetics. I. Theoretical introduction. Second edition. Interdisciplinary Applied Mathematics, **27**. Springer-Verlag, New York.

[81] A.M. Etheridge. *An introduction to superprocesses.* American mathematical society, University Lecture series (Vol. 20).

[82] Fisher, R. A. The wave of advance of advantageous genes. *Annals of Eugenics* **7** (1937), 355 ?369.

[83] N. Gantert, Yueyun Hu, and Zhan Shi (2008). Asymptotics for the survival probability in a supercritical branching random walk. arXiv :0811.0262, 2008.

[84] J.H. Gillespie (2000), Genetic Drift in an Infinite Population : The Pseudohitchhiking Model. *Genetics*, **. 155**, 909-919.

[85] C. Giraud (2001), Genealogy of shocks in burgers turbulence with white noise initial velocity *Comm. Math. Phys.* **1**, 67.

[86] A. Gnedin, B. Hansen and J. Pitman, (2007) . Notes on the occupancy problem with infinitely many boxes : general asymptotics and power laws. *Probab. Surv.* **4**, 146–171 (electronic)

[87] Goldschmidt, C. and Martin, J. (2005). Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10**, 718-745.

[88] D.R. Grey (1974). Asymptotic behavior of continuous-time continuous-state branching processes. *J. Appl. Prob.*, 11 :669–677.

[89] Harris, S. C. (1999) Travelling-waves for the FKPP equation via probabilistic arguments *Proc. Roy. Soc. Edin.* **129A**, 503-517. ISSN 0308-2105

[90] J. W. Harris and S. C. Harris (2007). Survival probabilities for branching Brownian motion with absorption. *Elect. Comm. Probab.* **12**, 81-92.

[91] J. W. Harris, S. C. Harris, and A. E. Kyprianou (2006). Further probabilistic analysis of the Fisher-Kolmogorov-Petrovskii-Piscounov equation : one-sided traveling waves. *Ann. Inst. H. Poincaré Probab. Statist.* **42**, 125-145.

[92] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).

[93] Hedrick, P. W., 2005 Large variance in reproductive success and the N_e/N ratio. *Evolution* **59**, 1596–1599.

[94] M. Kardar, G. Parisi, and Y.-C. Zhang, Dynamic Scaling of Growing Interfaces *Phys. Rev. Lett.* **56**, 889 (1986)

[95] H. Kesten (1978). Branching Brownian motion with absorption. *Stochastic Process. Appl.* **7**, 9-47.

[96] M. Kimura (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

[97] Kingman, J. F. C. (1978). The representation of partition structures. *J. London Math. Soc.* **18**, 374-380.

[98] J. F. C. Kingman (1982). The coalescent. *Stoch. Proc. Appl.* 13, 235–248.

[99] J. F. C. Kingman (1982), On the genealogies of large populations, *J. Applied Probability*, 19 A : 27–43.

[100] Kesten, J., Ney, P., and Spitzer, F. (1966). The Galton-Watson process with mean one and finite variance. *Theory Probab. Appl.* **11**, 513-540.

[101] A. Kolmogorov, I. Petrovsky, and N. Piscunov, Étude de l'Équation de la di ?usion avec croissance de la quantite de matière et son application à un problème biologique, *Moscow Bull. Math.*,**1** (1937), pp. 1–25.

[102] Kyprianou, A.E. (2004) Travelling wave solutions to the K-P-P equation : alternatives to Simon Harris' probabilistic analysis. *Ann. Inst. H. Poincaré Probab. Statist.* **40**, 53-72.

[103] J. Lamperti (1967). The limit of a sequence of branching processes. *Z. Wahrsch. Verw. Gebiete* **7**, 271–288.

[104] J.-F. Le Gall and Y. Le Jan (1998). Branching processes in Lévy processes : the exploration process. *Ann. Probab.* **26**, 213-252.

[105] V. Limic and A. Sturm (2006). The spatial Lambda-coalescent. *Electron. J. Probab.* **11**, 363–393.

[106] G. Li and D. Hedgecock (1998). Genetic heterogeneity, detected by PCR SSCP, among samples of larval Pacific oysters (Crassostrea gigas) supports the hypothesis of large variance in reproductive success. *Can. J. Fish. Aquat. Sci.* 55 : 1025–1033.

[107] T.M. Liggett (1985). *Interacting particle systems.* Springer, New York.

[108] M. Mézard, G. Parisi, and M. Virasoro, (1987)*S*pin Glass Theory and Beyond World Scientific.

[109] Möhle, M. (1998) Robustness results for the coalescent. *J. Appl. Probab.* **35** , no. 2, 438–447.

[110] M. Möhle (2006). On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli*, 12, 1 :35–53.

[111] M. Möhle and S. Sagitov (2001), A classification of coalescent processes for haploid exchangeable population models. *Ann Probab.* **29**, No. 4, 1547–1562

[112] A.P. Morris,J.C. Whittaker, and D.J. Balding (2002). Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. *Am J Hum Genet*, 70, 686–707.

[113] P. Mörters, N. R. Shieh (2005). Multifractal analysis of branching measure on a Galton-Watson tree. *Third International Congress of Chinese Mathematicians. Part 1, 2*, AMS/IP Stud. Adv. Math., 42, pt. 1, **2**, 655–662, Amer. Math. Soc.,

[114] C. Mueller and R. Sowers. Random traveling waves for the KPP equation with noise. *J. Funct. Anal.*, 128 :439 ?498, 1995.

[115] Mueller, C. ; Mytnik, L. ; Quastel, J. Small noise asymptotics of traveling waves. *Markov Process. Related Fields* **14** (2008), no. 3, 333–342

[116] Neveu, J. (1988) Multiplicative martingales for spatial branching processes. *Seminar on Stochastic Processes, 1987 (Princeton, NJ, 1987)*, 223–242, Progress in Probability and Statistics, Birkhuser Boston, Boston, MA.

[117] J. Pitman (1999). Coalescents with multiple collisions. *Ann Probab.* 27, 1870–1902.

[118] J. Pitman. Combinatorial stochastic processes. *Lecture notes in mathematics, Ecole d'Eté de probabilités de Saint-Flour* XXXII-2002. Vol. 1875, Springer, 2006.

[119] W. Pruitt (1981). The growth of random walks and Lévy processes. *Ann. Probab.*, 9(6) :948–956.

[120] D. Revuz and M. Yor (1999). *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York.

[121] L.C.G. Rogers, D. Williams. (1994) *Diffusions, Markov processes, and martingales.* Vol. 1, Cambridge University Press, 2nd edition.

[122] S. Sagitov (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl Prob.* 36 :4, 1116–1125.

[123] Y. G. Sinai, Comm. Math. Phys., Statistics of shocks in solutions of inviscid Burgers equation **148**, 601 (1992).

[124] J. Schweinsberg (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Comm. Probab.* **5**, 1–11.

[125] Schweinsberg (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, 1–50.

[126] J. Schweinsberg (2003). Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Proc. Appl.*, 106 :107–139.

[127] The number of small blocks in exchangeable random partitions. Preprint arXiv :0911.1793

[128] T. Shiga. Stepping stone models in population genetics and population dynamics. In S. Albeverio et al., editor, *Stochastic Processes in Physics and Engineering*, pages 345 ?355. D. Reidel, 1988.

[129] M.L. Silverstein (1968). A new approach to local times. *J. Math. Mech.* 17, 1023–1054.

[130] D. Simon and B. Derrida (2008). Quasi-stationary regime of a branching random walk in presence of an absorbing wall. *J. Stat. Phys.* **131**, 203-233.

[131] Slack, R. (1968). A Branching process with mean one and possibly infinite variance. *Z. Wahrsch. Verw. Gebiete* **9**, 139-145.

[132] S. Tavaré and W.J. Ewens (1995) The Ewens Sampling Formula. Available at `http://www.cs.cmu.edu/~epxing/CBML/coalescent/esfrep.ps`

[133] Tribe, R. (1996). A traveling wave solution to the Kolmogorov equation with noise. *Stochastics Stochastic Rep.* **56** 317 ?340.

[134] Hein, J., Schierup, M. H. and Wiuf, C. Gene Genealogies, Variation and Evolution (Oxford Univ. Press, New York, 2005).

[135] J. Warren (1999). On a result of David Aldous concerning the trees in the conditioned excursion. *Elect. Comm. in Probab.* 4, 25–29.

[136] Yaglom, A. M. (1947). Certain limit theorems of the theory of branching processes. *Dokl. Acad. Nauk SSSR* **56**, 795-798.

[137] Gnedin, A. and Yakubovich, Y. On the number of collisions in Λ-coalescents. *Electron. J. Probab.* **12** (2007), no. 56, 1547–1567 (electronic)..