École doctorale IAEM Lorraine

Processus de branchements non Markoviens en dynamique et génétique des populations

THÈSE

présentée et soutenue publiquement le 17 Novembre 2016

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention mathématiques)

par

Henry Benoît

Composition du jury

Rapporteurs :	Florian SIMATOS	Professeur, ISAE Supaero Toulouse
	Götz KERSTING	Professeur, Université de Francfort
Examinateurs :	Amaury LAMBERT	Professeur, Université Paris VI
	Jean-Baptiste DURAND	Maître de conférence, Université de Grenoble
Directeurs :	Nicolas CHAMPAGNAT	Chargé de recherche, INRIA Nancy Grand Est
	Dave RITCHIE	Directeur de recherche, INRIA Nancy Grand Est

Institut Élie Cartan de Lorraine, Laboratoire de Mathématiques, BP 239, 54506 VANDOEUVRE-LÈS-NANCY Cedex

*

Résumé

Dans cette thèse nous considérons une population branchante générale où les individus vivent et se reproduisent de manière i.i.d. La durée de vie de chaque individu est distribuée suivant une mesure de probabilité arbitraire et chacun d'eux donne naissance à taux exponentiel. L'arbre décrivant la dynamique de cette population est connu sous le nom de splitting tree. Ces arbres aléatoires ont été introduits par Geiger et Kersting en 1997.
Dans un premier temps nous nous intéressons au processus stochastique qui compte le nombre d'individus vivants à un instant donné. Ce processus est connu sous le nom de processus de Crump-Mode-Jagers binaire homogène, et il est connu que ce processus, quand correctement renormalisé, converge presque sûrement en temps long vers une variable aléatoire (non dégénérée dans le cas surcritique). Grâce à l'étude du splitting tree sous-jacent à la population via les outils introduit par A. Lambert en 2010, nous montrons un théorème central limite pour cette convergence p.s. dans le cas surcritique.
Dans un second temps, nous supposons que les individus subissent des mutations à taux exponentiel sous l'hypothèse d'infinité d'allèles. Cette procédure mène à une partition de la population à un instant donné par familles de même type. Nous nous intéressons alors au spectre de fréquence allélique de la population qui compte la fréquence des tailles de familles dans la population à un instant donnée. A l'aide d'un nouveau théorème permettant de calculer l'espérance de l'intégrale d'un processus stochastique contre un mesure aléatoire quand les deux objets présentent une structure particulière de dépendance, nous obtenons des formules pour calculer tout les moments joints du spectre de fréquence. En utilisant ces formules, et en adaptant la preuve de la première partie, nous obtenons également des théorèmes centraux limites en temps long pour le spectre de fréquence.
Une dernière partie, indépendante des autres, s'intéresse à des questions statistiques sur des arbres de Galton-Watson conditionnés par leurs tailles. L'idée de base est que les processus de contours devraient être utilisés pour faire des statistiques sur des données hiérarchiques dans la mesure où ils ont déjà prouvés leur efficacité dans des cadres plus théoriques. Ce travail est un premier pas dans cette direction. Le but est ici d'estimer la variance de la loi de naissance rendue inaccessible par le conditionnement. On utilise le fait que le processus de contour d'un arbre de Galton-Watson conditionné converge vers une excursion Brownienne quand la taille de l'arbre grandit afin de construire des estimateurs de la variance à partir de forêts. On s'attache ensuite à étudier le comportement asymptotique de ces estimateurs. Dans une dernière partie, on illustre numériquement leurs comportements.

Mots-clés: Splitting trees, processus de branchement, processus de Crump-Mode-Jagers

Abstract

In this thesis we consider a general branching population. The lifetimes of the individuals are supposed to be i.i.d. random variables distributed according to an arbitrary distribution. More-

over, each individual gives birth to new individuals at Poisson rate independently from the other individuals. The tree underlying the dynamics of this population is called a splitting tree. This class of random tree was introduced by Geiger and Kersting in 1997
In a first part we are interested in the population counting process of the tree (i.e. the process which count the number of alive individuals at given times). These processes are known as binary homogeneous Crump-Mode-Jagers processes. Moreover, these processes are known, when properly renormalized, to converge almost surely to some random variable (which is non degenerate in the supercritical case). Thanks to the study of the underlying splitting tree through the tools introduced by A. Lambert in 2010, we show a central limit theorem associated to this a.s. convergence.
In a second part, we suppose that individuals undergo mutation at Poisson rate under the infinitely many alleles assumption. This mechanism leads to a partition of the population by type. We are mainly interested in the so called allelic frequency spectrum which describes the frequency of sizes of families (i.e. sets of individuals carrying the same type) at fixed times. Thanks to a new theorem allowing to compute the expectation of the integral of some random process with respect to a random measure when both objects present a particular dependency structure, we are able to compute every joints moments of the frequency spectrum. These enable us to get central limit theorems for the frequency spectrum by adapting the proof of the first part.
In a last part, we study some statistical problems for size constrained Galton-Watson trees. The idea is that contour processes should be used to perform statistics on tree shaped datas, since such processes proved to be particularly powerful in the theoretical study of trees. Our goal is to estimate the variance of the birth distribution. This is not an easy task since conditioning killed the independence and the homogeneity of the laws of the numbers children of the individuals. Using that the contour process of a size constrained Galton-Watson tree converges to a Brownian excursion as the size of the tree growth, we construct estimators of the variance of the birth distribution. Then, we study the asymptotic behaviour of our estimators. To end, we stress our methods on simulated datas.

Keywords: Splitting trees, branching processes, Crump-Mode-Jagers processes

Table des matières

Chapitre 1

Introduction

Cette thèse porte sur l'étude de certains d'objets aléatoires utilisés en dynamique et génétique des populations.

La dynamique des populations s'intéresse essentiellement à l'étude des variations des effectifs d'individus dans une population au cours du temps. L'utilisation des mathématiques en dynamique des populations remonte au moins à 1826 lorsque Thomas Malthus les utilise dans son livre "Essay on the principle of population" [72]. Pour défendre sa thèse : "I said that a population, when unchecked, increased in a geometrical ratio", il introduit un modèle très simple : $u_{n+1} = 2u_n$. Le terme croissance Matlhusienne viens de là. Quelques années plus tard, en 1838, à la suite des travaux de Malthus, Pierre François Verhulst introduit le modèle de croissance logistique [88] afin de prendre en compte les contraintes environnementales. Depuis, la variété des modèles et leurs complexités n'a cessé de croître, et en faire l'inventaire serait un travail qui dépasse le cadre de cette introduction. Une sophistication naturelle fut dès lors de prendre en compte l'aléa qui influe notamment dans de petites populations. Le modèle probabiliste le plus souvent cité en exemple est bien sûr le processus de Bienaymé-Galton-Watson [39, 4] qui, bien qu'introduit il y a plus de 150 ans, est encore un sujet d'étude aujourd'hui.

La génétique des populations, quant à elle, s'intéresse à l'apparition ou à la variation de la fréquence d'allèles au sein d'une population. Les modèles mathématiques pour la génétique des populations [29, 27] adoptent un point de vue légèrement différent de ceux utilisés en dynamique des populations. En considérant souvent des populations de tailles fixées, ce domaine s'intéresse par exemple aux probabilités de fixation d'un allèle au sein de la population. On pourra penser au modèle de Wright-Fisher [91, 31] dont le pendant en temps rétrograde, le modèle de Kingman [55, 56] a permis d'obtenir une expression explicite pour la loi du spectre de fréquence de la population échantillonnée connu sous le nom de formule d'échantillonage d'Ewens [28]. Ceci permet par exemple de construire des estimateurs pour le taux de mutation de la population. En biologie, le spectre de fréquence a également été utilisé pour détecter une sélection positive d'un gène dans une population en croissance [84, 85].

Dans cette thèse nous considérons un modèle plus sophistiqué que celui de Bienaymé-Galton-Watson afin de prendre en compte le temps et les durées de vies des individus. Nous considérons une population branchante générale. Comme dans le modèle de Galton-Watson, les individus vivent et se reproduisent de manière indépendante les uns des autres. Cependant, nous supposons que leurs durées de vies sont distribuées suivant une loi de probabilité fixée $\mathbb{P}_V$. Ensuite, chaque individu donne naissance à de nouveaux individus à taux fixé b durant sa vie, chaque nouvelle naissance donnant un unique nouvel individu (contrairement à d'autres modèles où des naissances simultanées sont possibles). En fonction des valeurs de b et de $\int_{\mathbb{R}_+} x \, \mathbb{P}_V(dx)$, il est connu [60] que la population montre plusieurs régimes de croissance différents. Si $b \int_{\mathbb{R}_+} x \, \mathbb{P}_V(dx) = 1$ (cas critique) ou $b \int_{\mathbb{R}_+} x \, \mathbb{P}_V(dx) < 1$ (cas sous-critique) alors la population s'éteint presque sûrement. Si $b \int_{\mathbb{R}_+} x \, \mathbb{P}_V(dx) > 1$ (cas surcritique), alors, avec un probabilité positive, la population ne s'éteindra jamais. De plus, en cas de non extinction, la croissance de la population se fait à vitesse exponentielle. Dans ce dernier cas, on peut montrer l'existence d'une constante α strictement positive, appelée paramètre Malthusien, correspondant au taux de croissance exponentiel de la population.

Le modèle décrit plus haut est plus fin que celui de Galton-Watson (ou que son pendant Markovien en temps continu) car il prend en compte le vieillissement éventuel des individus. Il présente cependant deux défauts importants du point de vu biologique :

— Les naissances sont Poissoniennes (donc "sans mémoire").
— Il n'y a pas d'interactions entre les individus.

Une telle population peut naturellement être assimilée à un arbre dans lequel chaque branche représente un individu et dont la longueur représente la durée de vie de l'individu correspondant. Les branchements représentent alors les événements de naissances. L'arbre décrivant la dynamique de la population décrite plus haut est appelé un splitting tree [35].

Comme souvent dans l'étude des arbres, il est commode de construire un opérateur inversible qui transforme l'arbre en une fonction réelle car ce sont des objets bien plus aisés à manipuler que les arbres. Les exemples les plus connus concernent les arbres de Galton-Watson (bien qu'ils soient définis pour n'importe quel arbre discret). On pourra penser au processus de Harris [78] ou encore à la marche de Łukasiewicz [32]. Dans le cas des splitting trees, il a été montré par A. Lambert en 2010 [60] qu'il existe une transformation d'un splitting tree "fini" en un processus càdlàg possédant la très commode propriété d'être un processus de Lévy tué en zéro. Dans le cas d'un arbre infini, l'étude est permise par la troncature de l'arbre en deçà d'une date fixée.

Ce point de vue a donné lieu à de nombreux travaux [20, 67, 62, 64, 66], par exemple sur l'inférence ancestrale sous le modèle des splitting trees [63, 61]. D'autre travaux s'intéressent aux splitting trees avec mutations arrivant soit à la naissance des individus [82, 83, 22, 23] soit de manière Poissonienne durant la vie des individus [13, 14]. En particulier, cet outil s'est avéré utile dans l'étude du comportement asymptotique des processus de branchement construits à partir des splitting trees. Le plus simple d'entre eux est le processus $(N_t, \ t \in \mathbb{R}_+)$ qui compte le nombre d'individus N_t vivants dans l'arbre à l'instant t. Ce processus est connu sous le nom de processus de Crump-Mode-Jagers binaire homogène. Dans le cas surcritique, il a été montré [82] que la quantité

$$\frac{N_t}{\mathbb{E}\left[N_t \mid N_t > 0\right]} \tag{1.1}$$

converge presque sûrement quand t tend vers l'infini. M. Richard [81] a également étudié à l'aide de ces mêmes outils des processus branchement associés à des splitting trees avec immigration. Il montre notamment la convergence presque sûre du processus qui compte la population vivante à un instant t vers une variable aléatoire de loi gamma. De même, il regarde le comportement asymptotique des ratios des populations migrantes par rapport à la population totale sous divers modèles d'immigrations. Là encore, il obtient des lois des grands nombres. Bien que de nombreux résultats de convergence presque sûre aient été montrés, il semble qu'aucun théorème central limite associé à l'une de ces loi des grands nombres n'apparaisse dans la littérature. L'un des apports de cette thèse est un théorème central limite pour la convergence presque sûre du ratio (1.1). Une perspective pourrait être de regarder ces fluctuations dans le cadre des résultats de Mathieu Richard ou dans un cadre plus général. En effet, la preuve de ce théorème central limite semble pouvoir s'étendre à d'autres processus de branchements construit à partir de splitting trees.

Dans cette thèse, nous étudions également un modèle avec mutations. Nous supposons que les individus vivants subissent des mutations à taux exponentiel θ sous l'hypothèse d'infinité d'allèles. Cette hypothèse suppose que chaque mutation remplace le type de l'individu touché par un type entièrement nouveau. Par ailleurs les types sont supposés se transmettre de parents à enfants. Ce mécanisme mène à une partition de la population par types. On note alors $A(k,t)$ le nombre de familles (c'est-à-dire les ensembles d'individus partageant le même type) de taille k à l'instant t. La suite d'entiers $(A(k,t))_{k\geq 1}$ est appelée spectre de fréquence de la population au temps t. Cet objet, bien connu en biologie (c'est celui étudié par la formule d'échantillonnage d'Ewens [28]), a été introduit dans le cadre des splitting trees par N. Champagnat et A. Lambert dans [13] où ils obtiennent une expression explicite pour

$$\mathbb{E}\left[A(k,t)u^{N_t}\right], \quad \forall u \in (0,1).$$

Ils obtiennent ensuite la convergence presque sûre du spectre de fréquence quand correctement renormalisé. Dans un second travail [14], ils s'intéressent à la taille des plus grandes familles et à l'âge des plus anciennes. Mathieu Richard s'est également intéressé à des splitting trees avec mutations mais dans son modèle, il suppose que les mutations ont lieu à la naissance des individus. Il obtient le même type de résultats dans son cadre [83]. Dans cette thèse nous obtenons des formules permettant de calculer tout les moments (joints ou non) du spectre de fréquence dans le modèle à mutations Poissoniennes. Ceci est fait à l'aide d'un nouveau théorème permettant de calculer l'espérance de l'intégrale d'un processus stochastique contre un mesure aléatoire quand les deux objets présentent une structure particulière de dépendance. Par ailleurs, ces formules nous permettent d'étendre la preuve du théorème central limite obtenu pour N_t afin d'en obtenir un pour le spectre de fréquence.

Dans cette thèse nous nous sommes également intéressé à un problème statistique sur des arbres de Galton-Watson conditionnés par leurs tailles. Les problèmes statistiques portant sur des données arborescentes sont en général délicats car l'espace dans lequel vivent ces objets est très grand. Ce travail repose sur l'idée que les processus de contours devraient être utilisés pour faire des statistiques sur ce type de donnée car ces outils ce sont déjà révélé efficace dans l'étude théorique des arbres aléatoires. Un arbre de Galton-Watson peut servir à décrire la généalogie d'une population. On suppose donnée une mesure de probabilité μ sur $\mathbb{N}$ de variance finie et on considère une population démarrant d'un unique individu. Puis on suppose que cet individu donne naissance à un nombre aléatoire d'enfants distribué selon μ. Alors chacun de ces enfants donne lui-même naissance à des nouveaux individus selon le même mécanisme indépendamment des autres. Notre but est d'estimer la variance de μ pour des arbres conditionnés par leurs tailles. Si le problème est simple dans le cadre des arbres de Galton-Watson non conditionnés [47], il est beaucoup plus délicat dans le cadre des arbres conditionnés. En effet, le conditionnement remet en cause l'indépendance et l'homogénéité (des lois) des variables aléatoires correspondants aux nombres d'enfants de chaque individus. On peut par exemple montrer qu'il n'est pas possible d'estimer (à partir d'un seul arbre ou d'une forêt) la moyenne de la loi μ (problème d'identifiabililté). De plus, d'autres résultats suggèrent qu'il n'est pas possible d'estimer σ à partir d'un seul arbre conditionné (par exemple [50]). Dans ce chapitre, nous construisons des estimateurs de σ^{-1} à partir d'une forêt $\mathcal{F} = (\tau_1, \ldots, \tau_N)$ d'arbres indépendants telle que chaque arbre τ_i est un arbre de Galton-Watson conditionné à avoir n_i nœuds. Dans des travaux récents [9], les auteurs cherchent également a estimer σ^{-1} mais sans utiliser les processus de contour pour construire des estimateurs.

Le présent document est découpé en 6 chapitres, chaque chapitre étant centré sur une unité thématique. Les deux premiers chapitres ne contiennent pas de contributions originales. Ils sont dévoués à des introductions aux outils utilisés dans la suite. Le Chapitre 4 introduit des outils originaux utilisés dans la suite. Le chapitre 5 concerne l'étude des processsus de Crump-Mode-Jagers binaires homogènes. Le chapitre 6 s'intéresse au spectre de fréquence allélique d'un splitting tree avec mutations Poissonienne neutre. La première section du Chapitre 4 ainsi que le Chapitre 5 sont issus de la prépublication [40]. Les deux dernières sections du Chapitre 4 ainsi que le Chapitre 6 sont issus de la publication [12] en collaboration avec Nicolas Champagnat. Le dernier chapitre concerne des questions statistiques sur des arbres de Galton-Watson conditionnés. C'est une travail en collaboration avec Romain Azaïs (Nancy) et Alexandre Genadot (Bordeaux).

1.1 Chapitre 2

Ce chapitre ne contient pas de contribution originale. C'est une introduction pédagogique à la théorie des fluctuations des processus de Lévy sans sauts négatifs. Un processus de Lévy est défini comme un processus $(Y_t,\ t \in \mathbb{R}_+)$ càdlàg à valeurs réelles tel que, pour toutes suites de temps $0 \leq t_1 < t_2 < \cdots < t_n$, les accroissements de Y, $Y_{t_2} - Y_{t_1}, Y_{t_3} - Y_{t_2}, \ldots, Y_{t_{n-1}} - Y_{t_n}$, sont des variables aléatoires indépendantes dont les lois ne dépendent respectivement que des écarts $t_2 - t_1, \ldots, t_n - t_{n-1}$. Le principal but du chapitre est d'obtenir les identités de fluctuations pour les processus de Lévy sans sauts négatifs utilisées dans la suite de ce manuscrit de manière aussi simple et directe que possible.

Plus précisément, étant donnés $a < b$ deux nombres réels, on note

$$\tau_b^+ = \inf\{t \geq 0 \mid Y_t > b\} \quad \text{et} \quad \tau_a^- = \inf\{t \geq 0 \mid Y_t < a\}.$$

Le but final est d'arriver à des expressions les plus explicites possibles pour des quantité du type $\mathbb{P}_x\left(\tau_a^- < \tau_b^+\right)$ et pour la loi du couple $(Y_{\tau_b^+ -},\ Y_{\tau_b^+})$ (car Y est susceptible de sortir de l'intervalle (a, b) en sautant).

Finalement, une courte partie à la fin du chapitre rappelle quelques résultats classiques de théorie du renouvellement utilisés dans cette thèse.

1.2 Chapitre 3

Ce chapitre ne contient pas de contribution originale. Dans le même esprit que le chapitre précédent, ce chapitre a pour but d'introduire des outils importants utilisés dans cette thèse. Dans un premier temps, on y présente les *splitting trees*. Comme indiqué dans la section précédente, cette classe d'arbres puise son intérêt dans le fait qu'ils modélisent de manière relativement générale une population biologique.

Pour nous, ces arbres sont très intéressants car les processus de branchement étudiés dans cette thèse peuvent s'écrire comme des fonctionnelles ces arbres. Par exemple, si $\mathbb{T}$ est un splitting tree et N_t est le nombre d'individus présents dans l'arbre au temps t, alors le processus stochastique $(N_t,\ t \in \mathbb{R}_+)$ est un processus de Crump-Mode-Jagers binaire homogène. L'étude de ces processus fait l'objet du Chapitre 5.

Cependant les splitting trees ne sont pas l'outil essentiel introduit dans ce chapitre. Comme souvent avec les arbres, il est plus commode de les transformer en objets plus simples à manipuler mais contenant toute l'information contenue dans l'arbre. Par exemple, c'est le cas du processus de contour pour les arbres de Galton-Watson. Dans le cadre des splitting trees, il existe aussi un processus de contour introduit par A. Lambert [60]. Si la population s'éteint presque sûrement, celui-ci a la particularité d'être un processus de Lévy tué en 0 dont l'exposant de Laplace est donné par

$$\psi(\lambda) = \lambda - \int_{\mathbb{R}_+} (1 - e^{-\lambda x}) b\mathbb{P}_V(dx). \tag{1.2}$$

Dans le cas où la population ne s'éteint pas, l'étude est conduite en considérant la troncature de l'arbre en deçà d'une date fixé.

Un autre objet important introduit dans le Chapitre 4 est le processus ponctuel de coalescence (CPP). La relation de cet objet avec les splitting trees est similaire à celle qu'entretient le coalescent de Kingman avec le modèle de Wright-Fisher dans le sens qu'il décrit les relations généalogiques entre le individus vivants dans l'arbre à un instant donnée.

1.3 Chapitre 5

Le Chapitre 5 est dévolu à des résultats préliminaires qui sont relativement déconnectés thématiquement des chapitres suivants et qui présentent, selon nous, un intérêt particulier qui justifie qu'ils soient mis dans un chapitre différent. Il est découpé en trois parties.

La première partie concerne l'étude du comportement asymptotique de la fonction d'échelle W du contour d'un splitting tree surcritique, c'est à dire d'un processus de Lévy sans sauts négatifs dont l'exposant de Laplace est donné par (1.2). La fonction d'échelle est une fonction intervenant dans l'étude des fluctuations du processus de Lévy [59]. Celle-ci est caractérisée par sa transformée de Laplace :

$$\int_{\mathbb{R}_+} W(s) e^{-\beta s} \, \mathrm{d}s = \frac{1}{\psi(\beta)}.$$

Plus précisément, dans [14], Champagnat N. and Lambert A. montrent, dans le cas surcritique, l'existence d'une constante positive γ telle que

$$e^{-\alpha t}\psi'(\alpha)W(t) - 1 = \mathcal{O}\left(e^{-\gamma t}\right).$$

Le but de cette partie est d'obtenir des estimées plus fines sur ce $\mathcal{O}\left(e^{-\gamma t}\right)$. Plus précisément, on montre le résultat suivant.

Proposition 1.3.1 (Comportement asymptotique de W)**.** *Il existe une fonction positive décroissante càdlàg telle que*

$$W(t) = \frac{e^{\alpha t}}{\psi'(\alpha)} - e^{\alpha t}F(t), \quad t \geq 0,$$

satisfaisant

$$\lim_{t\to\infty} e^{\alpha t}F(t) = \begin{cases} \frac{1}{b\mathbb{E}V-1} & \text{if } \mathbb{E}V < \infty, \\ 0 & \text{sinon.} \end{cases}$$

La preuve de ce résultat est basée sur le fait que la fonction W peut se réécrire en fonction de la mesure de potentiel du subordinateur d'échelle ascendante (ascending ladder process) d'une légère modification de notre processus de Lévy. Dans notre cas particulier, il est possible d'effectuer des calculs explicites concernant la loi de ce subordinateur, ce qui permet d'étudier plus précisément la fonction W. Ce résultat est fondamental pour démontrer les résultats du Chapitre 5.

La seconde partie s'intéresse au calcul de l'espérance d'une intégrale du type

$$\int_{\mathcal{X}} X_s \, \mathcal{N}(ds), \tag{1.3}$$

où $\mathcal{X}$ est un espace polonais, $(X_s,\ s \in \mathcal{X})$ est un processus continu (ou càdlàg quand $\mathcal{X}$ est, par exemple, $\mathbb{R}_+$), et $\mathcal{N}$ est une mesure aléatoire. Bien sûr le résultat est évident quand $\mathcal{N}$ et $(X_s,\ s \in \mathcal{X})$ sont indépendants. Notre but est d'obtenir un théorème permettant de calculer cette espérance lorsque $(X_s,\ s \in \mathcal{X})$ et $\mathcal{N}$ présente une structure particulière de dépendance. Les théorèmes que l'ont obtient ont des applications très importantes dans le Chapitre 6 de cette thèse. On obtient par exemple les résultats suivants.

Theorem 1.3.2. *Soit X un processus stochastique continu de $\mathcal{X}$ dans $\mathbb{R}_+$. Soit $\mathcal{N}$ une mesure aléatoire sur $\mathcal{X}$ d'intensité finie μ. Si X est localement indépendant de $\mathcal{N}$, c'est à dire, pour tout $x \in \mathcal{X}$, il existe un voisinage V_x de x tel que X_x soit indépendant de $\mathcal{N}(V_x \cap \cdot)$. Si on suppose, de plus, qu'il existe une variable aléatoire intégrable Y telle que*

$$|X_x| \leq Y,\ \forall x \in \mathcal{X},\ a.s.$$

et

$$\mathbb{E}\left[Y\mathcal{N}(\mathcal{X})\right] < \infty.$$

Alors

$$\mathbb{E}\int_{\mathcal{X}} X_x \; \mathcal{N}(dx) = \int_{\mathcal{X}} \mathbb{E}\left[X_x\right] \; \mu(dx).$$

Dans le cas càdlàg, on obtient également.

Theorem 1.3.3. *Soit X un processus stochastique de $[0,T] \times \mathcal{X}$ dans $\mathbb{R}_+$ tel que $X_{\cdot,x}$ est càdlàg pour tout x et $X_{s,\cdot}$ est continu pour tout s. Soit $\mathcal{N}$ une mesure aléatoire sur $[0,T] \times \mathcal{X}$ d'intensité μ finie. Si, pour tout s de $[0,T]$, la famille $(X_{s,x}, x \in \mathcal{X})$ est indépendante de la restriction de $\mathcal{N}$ sur $[0,s]$, et s'il existe une variable aléatoire intégrable Y telle que*

$$|X_{s,x}| \leq Y, \ \forall x \in \mathcal{X}, \ \forall s \in [0,t], \ a.s.$$

et

$$\mathbb{E}\left[Y\mathcal{N}\left(\mathcal{X}\right)\right] < \infty.$$

Alors,

$$\mathbb{E}\int_{[0,T]\times\mathcal{X}} X_{s,x}\ \mathcal{N}\left(ds,dx\right) = \int_{[0,T]\times\mathcal{X}} \mathbb{E}\left[X_{s,x}\right]\ \mu\left(ds,dx\right).$$

Ce résultat est utilisé dans le Chapitre 6 où il permet d'étudier les moments du spectre de fréquence.

L'idée de ces théorèmes est la suivante. À une mesure aléatoire $\mathcal{N}$ sur $\mathcal{X}$, on peut associer une famille de mesures de probabilité $(\mathcal{P}_x)_{x\in\mathcal{X}}$. La mesure $\mathcal{P}_x$ est appelé mesure de Palm en x associée à $\mathcal{N}$. La formule de Campbell [18], permet d'exprimer la moyenne de l'intégrale (1.3) en fonction de l'intensité de $\mathcal{N}$ et de la moyenne sous $\mathcal{P}_x$ de X :

$$\mathbb{E}\left[\int_{\mathcal{X}} X_x\ \mathcal{N}(dx)\right] = \int_{\mathcal{X}} \mathbb{E}_{\mathcal{P}_x}\left[X_x\right]\ \mu\left(dx\right),$$

où $\mathbb{E}_{\mathcal{P}_x}$ est l'espérance sous $\mathcal{P}_x$.
Dans la cadre des mesures ponctuelles, on peut penser à $\mathcal{P}_x$ comme à $\mathbb{P}$ conditionné à ce que $\mathcal{N}$ ait un atome en x ($\mathbb{P}(\cdot \mid \mathcal{N}(\{x\}) > 0)$. Dès lors, si X vérifie les hypothèse du Théorème 1.3.2, alors sa loi sous $\mathcal{P}_x$ est la même que sous $\mathbb{P}$ et le théorème est démontré. Cependant, il n'est pas possible de donner un sens au conditionnement $\mathbb{P}(\cdot \mid \mathcal{N}(\{x\}) > 0)$. Par exemple, pour une mesure ponctuelle de Poisson, il faudrait que μ ait un atome en x, ce qui est déjà très restrictif.

La dernière partie concerne l'introduction d'une nouvelle construction du processus ponctuel de coalescence. Le processus ponctuel de coalescence (CPP) est le processus de coalescence associé au modèle des splitting trees, comme le coalescent de Kingman l'est au modèle de Wright-Fisher. Le CPP représente les relations généalogiques entre les lignées des individus vivants à un temps t fixé dans le splitting tree. On parlera de CPP arrêté au temps t. Il a été montré [60] qu'on pouvait le définir comme une suite $(H_i)_{i\geq 0}$ de variables aléatoires telle que $H_0 = t$ et que la famille $(H_i)_{i\geq 1}$ soit i.i.d. de loi donnée par

$$\mathbb{P}\left(H_i > t\right) = \frac{1}{W(t)},$$

arrêtée au premier $H_i > t$. Chaque variable aléatoire H_i est alors associée à un individu et la première coalescence de cette lignée est supposée avoir lieu au bout du temps H_i avec la lignée de l'individu j vérifiant (voir Figure 1.1)

$$j = \max\{k < i \mid H_k > H_i\}.$$

La dernière partie du Chapitre 4 est dévolue au résultat suivant qui donne une construction d'un CPP par recollement de CPP indépendants sur un autre CPP (voir Figure 1.2).

Proposition 1.3.4. *Soit* $\left(\mathcal{P}^{(i)}\right)_{i\geq 1}$ *une suite i.i.d. de CPP de fonction d'échelle* W *au temps* a*, et soit* $\left(N_a^i\right)_{i\geq 1}$ *leurs tailles respectives. Soit* $\hat{\mathcal{P}}$ *un CPP, indépendent de la famille précédente, de fonction d'échelle*

$$\hat{W}(t) := \frac{W(t+a)}{W(a)},$$

au temps $t-a$*, et soit* $\hat{N}_{t-a}$ *sa taille. Posons* $S_0 := 0$ *et*

$$S_i := \sum_{j=1}^{i} N_a^j, \quad \forall i \geq 1.$$

Alors le vecteur aléatoire $\left(H_k,\ 0 \leq k \leq S_{\hat{N}_a - 1}\right)$ *défini, pour tout* $k \geq 0$*, par*

$$H_k = \begin{cases} \mathcal{P}^{(i+1)}_{k-S_i} & \text{si il existe } i \geq 0 \quad \text{tel que } S_i < k < S_{i+1}, \\ \hat{\mathcal{P}}_i + a & \text{si il existe } i \geq 0 \quad \text{tel que } k = S_i, \end{cases}$$

est un CPP de fonction d'échelle W *au temps* t.

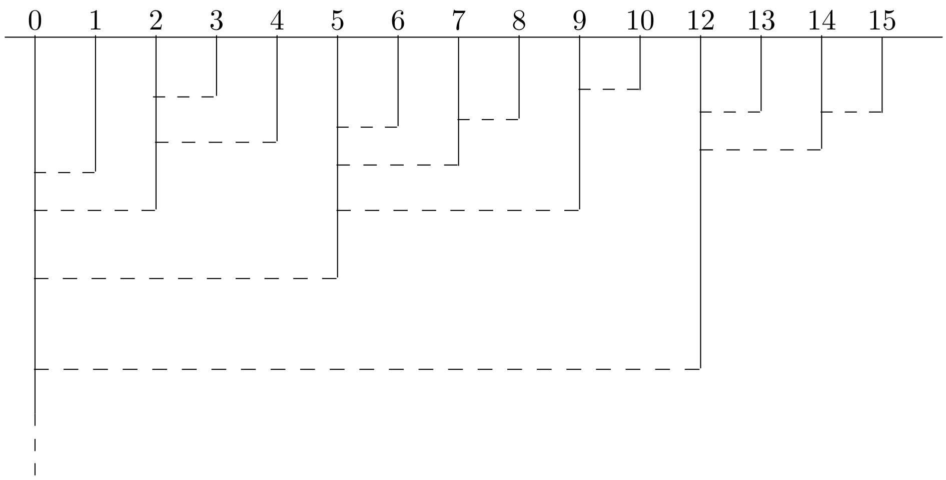

FIGURE 1.1 – Processus ponctuel de coalescence à 16 individus. Les pointillés horizontaux représentent les coalescences.

Ce résultat trouve deux applications. Dans le chapitre 5, il est utile dans la preuve de la loi des grands nombres associée au processus N_t et dans la chapitre 6 il trouve son utilité dans les calculs des moments du spectre de fréquence.

1.4 Chapitre 5

Le Chapitre 5 concerne les processus de Crump-Mode-Jagers binaires homogènes surcritiques. Dans ce chapitre nous nous intéressons au comportement en temps long du processus $(N_t,\ t \in \mathbb{R}_+)$. Le théorème (déjà connu) suivant établit que, correctement renormalisé, le processus N_t converge presque sûrement vers une variable aléatoire dont la loi est exponentielle conditionnellement à la non-extinction de la population.

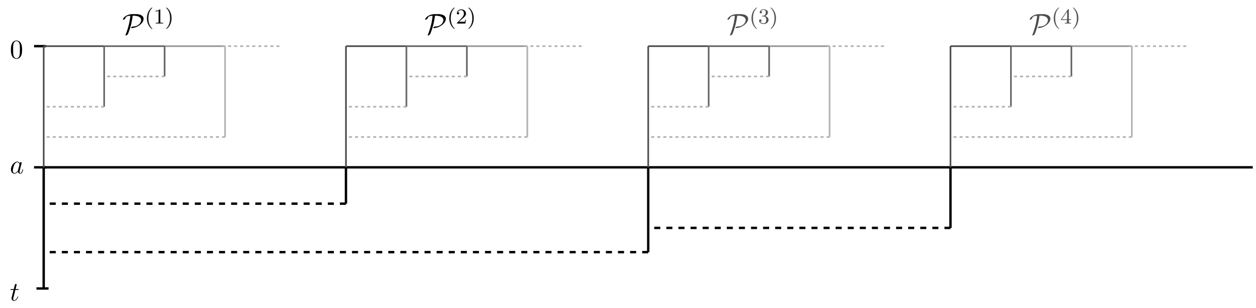

FIGURE 1.2 – Recollement de CPP.

Theorem 1.4.1. *Dans le cas surcritique ($\alpha > 0$), il existe une variable aléatoire $\mathcal{E}$ telle que*

$$\frac{N_t}{W(t)} \underset{t\to\infty}{\longrightarrow} \mathcal{E}, \quad \textit{presque sûrement et dans } L^2.$$

De plus, conditionnellement à la non-extinction, $\mathcal{E}$ suit une loi exponentielle de paramètre 1.

La preuve de ce résultat peut être trouvée dans la thèse de Mathieu Richard [82, Proposition 2.1]. Elle repose sur un critère de convergence presque sûre pour les processus de Crump-Mode-Jagers généraux établis par Nerman, O. [74] dans les années 80. Dans ce chapitre nous donnons une nouvelle preuve élémentaire de la convergence presque sûr de $\frac{N_t}{W(t)}$. Cette preuve a été publié dans [12].

L'apport essentiel de ce chapitre concerne l'étude des fluctuations dans la convergence établie par le théorème précédent. Plus précisément, nous y établissons le théorème central limite suivant pré-publié dans [40].

Theorem 1.4.2. *Dans le cas surcritique ($\alpha > 0$), conditionnellement à la non-extinction, la quantité*

$$\sqrt{W(t)}\left(\frac{N_t}{W(t)} - \mathcal{E}\right)$$

converge en loi, quand t tends vers l'infini, vers une loi de Laplace de moyenne nulle et de variance $2 - \psi'(\alpha)$.

Dans l'état actuel de nos connaissances, c'est la première fois qu'un théorème central limite est établi pour un processus de Crump-Mode-Jagers général alors que des lois de grands nombres pour ces processus sont l'objet de nombreux travaux. La preuve de ce théorème repose sur l'idée suivante :

- Une décomposition de N_t comme la somme des contributions des lignées des différents individus vivants à un instant antérieur.
- Un contrôle des dépendances entre chacune de ces lignées.
- Une expression explicite pour l'erreur quadratique moyenne $\mathbb{E}\left[(\frac{N_t}{W(t)} - \mathcal{E})^2\right]$ grâce à des méthodes de renouvellement.
- Un contrôle fin des erreur du type $\mathbb{E}\left[(\frac{N_t}{W(t)} - \mathcal{E})^n\right]$ ($n = 1, 2, 3$) grâce à des estimées précises sur la fonction $W(t)$.

Par ailleurs, la preuve de ce théorème étant assez souple, elle peut s'étendre à des cas plus complexes de processus de branchements comptés par caractéristiques aléatoires dès lors que l'arbre support est un splitting tree. Dans la Chapitre 6, nous démontrons grâce à cette méthode des théorèmes central limite pour ce type de processus dans le cas particulier du spectre de fréquence.

1.5 Chapitre 6

Dans ce chapitre, nous considérons que notre population subit également des mutations. Les mutations sont supposées arriver de manière Poissonienne à taux θ indépendamment d'un individu à l'autre. On suppose de plus que chaque nouvelle mutation remplace le type de l'individu qu'elle touche par un type totalement nouveau (hypothèse d'infinité d'allèles). Par ailleurs, les types sont supposés se transmettre de parents à enfants. Finalement, on suppose que les mutations n'ont pas d'influence sur la généalogie de la population (mutations neutres).

Ce mécanisme de mutations mène à une partition de la population vivante à un instant t en familles de même type (ou allèle). Notre but est d'étudier la fréquence des tailles de familles. Plus précisément, on note $A(k,t)$ le nombre de familles de taille k au temps t. La suite d'entiers $A(1,t),\ A(2,t),\dots$ est appelée *spectre de fréquence* de la population vivante au temps t.

Dans l'étude de ce spectre de fréquence, un rôle important est joué par la famille *clonale*. Cette famille est définie à un instant t comme l'ensemble des individus vivants à cet instant et possédant le type que possédait l'ancêtre à l'instant 0. On note $Z_0(t)$ le nombre d'individus clonaux à l'instant t. Cet objet a été très étudié par N. Champagnat et A. Lambert dans [13, 14]. Pour étudier cette quantité, une idée est de considérer le splitting tree dit clonal. Dans ce nouvel arbre on considère que les individus sont tués dès qu'ils subissent une mutations. De cette manière, la loi de $(Z_0(t),\ t \in \mathbb{R}_+)$ dans un splitting tree avec mutations est la même que la loi du processus qui compte la population dans un splitting tree clonal. Il facile de se rendre compte que la loi de la durée de vie d'un individu dans le splitting tree clonal est la loi du minimum entre une variable aléatoire exponentielle E de paramètre θ et une variable aléatoire V de loi $\mathbb{P}_V$. On appelle alors W_θ la fonction d'échelle associée au splitting tree clonal. Dans le cas où $\theta > \alpha$, on dit que l'arbre est clonal sous-critique signifiant que le famille ancestrale s'éteindra presque sûrement. Respectivement, si $\alpha = \theta$ on parlera de cas clonal critique et si $\theta < \alpha$ de cas clonal surcritique (dans ce cas la famille clonale ne s'éteint pas avec probabilité positive).
Ce chapitre est découpé en deux grandes parties. La première étudie les moments du spectre de fréquence à l'aide d'une nouvelle représentation du spectre sous forme intégrale. Plus précisément, on formalise l'apparition de mutations à l'aide d'une mesure aléatoire de Poisson $\mathcal{N}$ sur l'arbre. On a alors

$$A(k,t) = \int_{[0,t]\times\mathbb{N}} \mathbb{1}_{B_{i,k}(a)}\ \mathcal{N}(da,di),$$

où $B_{i,k}(a)$ est l'événement : le $i$ème individus (pour un certain ordre) vivant dans l'arbre au temps a à k descendants clonaux (i.e. de même type) au temps t. On remarque alors la chose suivante : si une mutation apparaît dans l'arbre (ou de manière équivalente sur le CPP) à un instant a, le nombre d'individus au temps t portant cette mutation ne dépend que des points/mutations qui arrivent par la suite (et non des mutations passées, voir Figure 1.3). On montre en utilisant

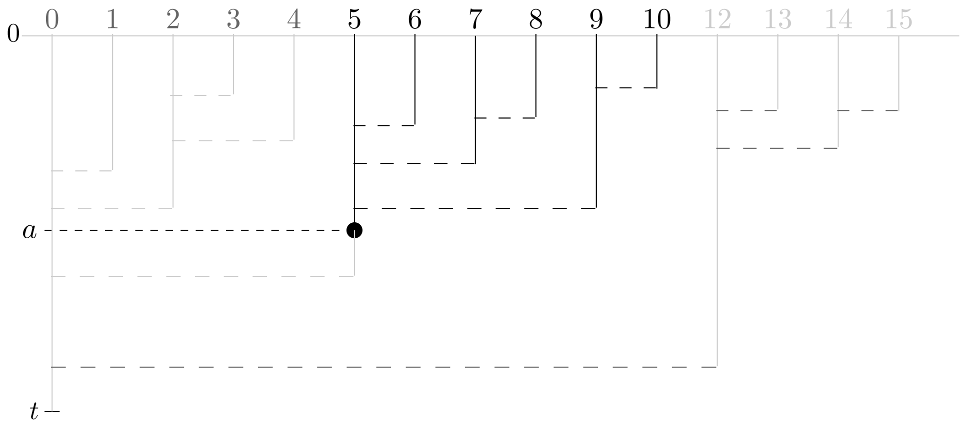

FIGURE 1.3 – Le futur d'une mutation ne dépend que d'un sous-arbre et des autres mutations qui s'y déroulent.

cette idée que le processus $(B_{i,k}(a), (a,i) \in [0,t] \times \mathbb{N})$ satisfait les hypothèse du Théorème 1.3.3 démontré au Chapitre 4, ce qui permet de calculer son espérance.

$$\mathbb{E}[A(k,t) \mid N_t > 0] = W(t) \int_0^t \frac{e^{-\theta a}}{W_\theta(a)^2} \left(1 - \frac{1}{W_\theta(a)}\right)^{k-1} da.$$

Plus encore, cela permet d'obtenir des formules récursives pour tout les moments du spectre de fréquence. Par exemple nous obtenons le théorème suivant.

Theorem 1.5.1. *Pour tout entier positive n et k, on a*

$$\mathbb{E}\left[\binom{A(k,t)}{n} \mid N_t > 0\right]$$
$$= \mathbb{E}\left\{\int_0^t \theta N_{t-a}^{(t)} \sum_{n_1+\cdots+n_{N_{t-a}^{(t)}}=n-1} \mathbb{E}\left[\binom{A(k,a)}{n_1} \mathbb{1}_{Z_0(a)=k} \mid N_a > 0\right] \prod_{m=2}^{N_{t-a}^{(t)}} \mathbb{E}\left[\binom{A(k,a)}{n_m} \mid N_a > 0\right] da\right\},$$

où $N_{t-a}^{(t)}$ suit une loi géométrique de paramètre $\frac{W(a)}{W(t+a)}$.

La preuve de ce résultat repose sur les mêmes idées que le calcul de $\mathbb{E}\left[A(k,t) \mid N_t > 0\right]$. En effet, on peut par exemple montrer que

$$\binom{A(k,t)}{2} = \int_{[0,t]\times\mathbb{N}} \mathbb{1}_{B_{i,k}(a)} \sum_{n=1}^{N^{(t)}t-a} A^{(n)}(k,a)\, \mathcal{N}(da,di),$$

où $A^{(n)}(k,a)$ est le nombre de famille de taille k au temps t dans le sous-arbre induit par le $n$ème individu vivant au temps $t-a$, et $N_{t-a}^{(t)}$ est le nombre d'individus vivant au temps $t-a$ ayant une descendance vivante au temps t. En appliquant à nouveau le Théorème 1.3.3 et en déterminant les lois de $N_{t-a}^{(t)}$ et $A^{(n)}(k,a)$, on arrive alors a obtenir le résultat. De la même manière, nous obtenons des formules pour les moments du type

$$\mathbb{E}\left[\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i} \mid N_t>0\right]$$

et

$$\mathbb{E}\left[\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)}=\ell \mid N_t>0\right],$$

ce qui permet de fermer les formules. L'étude de ces moments est l'objet de la publication [12] en collaboration avec Nicolas Champagnat.

La seconde partie du chapitre s'intéresse au comportement en temps long du spectre fréquence. Premièrement, nos formules sur les moments nous permettent de donner une preuve élémentaire de la loi des grand nombres pour le spectre de fréquence qui est originalement due à N. Champagnat et A. Lambert.

Theorem 1.5.2. *Dans le cas surcritique ($\alpha>0$), pour tout entier strictement positif k,*

$$\frac{A(k,t)}{W(t)} \longrightarrow c_k\mathcal{E} \text{ presque sûrement,}$$

quand t tend vers l'infini, avec

$$c_k=\int_0^t \frac{e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1} da.$$

Pour finir, nos formules sur les moments nous permettent d'étendre la preuve du théorème central limite du précédent chapitre pour obtenir le même type de résultats pour le spectre de fréquence. On obtient par exemple le théorème suivant.

Theorem 1.5.3. *Si $\theta>\alpha>0$ et $\int_{[0,\infty)} e^{(\theta-\alpha)v}\mathbb{P}_V(dv)>1$. Alors, conditionnellement à la non-extinction, on a la convergence en loi suivante :*

$$\left(e^{-\alpha\frac{t}{2}}\left(\psi'(\alpha)A(k,t)-e^{\alpha t}c_k\mathcal{E}\right)\right)_{k\in\mathbb{N}^*} \underset{t\to\infty}{\overset{(d)}{\longrightarrow}} \mathcal{L}(0,K),$$

où $\mathcal{L}(0,K)$ est la loi de Laplace infinie dimensionnelle de covariance K et moyenne nulle.

Ce théorème n'est pas d'un usage très pratique car la variable aléatoire limite $\mathcal{E}$ est inaccessible. De plus, nous n'obtenons pas d'expression explicite pour la matrice covariance K. Les idées de la preuve sont essentiellement les mêmes que pour le Théorème 1.4.2 mais des difficultés supplémentaires apparaissent :

— Il n'est pas possible pour l'instant d'obtenir des expressions explicites pour les moments du type $\mathbb{E}\left[\left(\psi'(\alpha)A(k,t)-e^{\alpha t}c_k\mathcal{E}\right)^2\right]$. Cela complique l'analyse et empêche d'obtenir un expression explicite pour la matrice de covariance K.

— Une mutation peut potentiellement avoir beaucoup de représentant dans l'arbre (surtout si elle est âgé), ce qui complique l'étude des dépendances entre les différentes parties de l'arbre.

Nous obtenons également un autre résultat, potentiellement plus intéressant pour les applications. Il permet d'approcher (en temps long) le spectre de fréquence par une fraction de la population totale.

Theorem 1.5.4. *Si $\theta > \alpha$, alors on a la convergence en loi suivante conditionnellement à la non-extinction,*

$$\psi'(\alpha)\left(e^{-\alpha\frac{t}{2}}\left(A(k,t) - c_k N_t\right)\right)_{k\in\mathbb{N}^*} \xrightarrow[t\to\infty]{(d)} \mathcal{L}\left(0, M\right).$$

L'avantage dans ce contexte est que les méthodes développées pour calculer les moments du spectre de fréquence s'étendent très bien au calcul des erreurs du type $\mathbb{E}\left[(A(k,t) - c_k N_t)^n\right]$, ce qui était un point délicat dans le résultat précédent. Ceci nous permet d'avoir une expression explicite pour la matrice de covariance M.

1.6 Chapitre 7

Le Chapitre 7 est quelque peu déconnecté du reste du manuscrit. Il s'agit d'un travail en collaboration avec Romain Azaïs et Alexandre Genadot. Dans ce chapitre nous nous intéressons à un problème de statistique pour des arbres de Galton-Watson conditionnés par leurs tailles. Note but est d'estimer σ^{-1}, l'inverse de la variance de la loi de naissance μ, à partir d'une forêt $\mathcal{F} = (\tau_1, \dots, \tau_N)$ d'arbres indépendants telle que chaque arbre τ_i est un arbre de Galton-Watson conditionné à avoir n_i nœuds. L'étude statistique d'arbre aléatoire semble être un problème naturel car de nombreuses données peuvent naturellement être représentée par des arbres (systèmes sanguins en biologie, ficher XML en informatiques,....). En particulier, les arbres de Galton-Watson conditionné apparaissent dans le nombreux problèmes [24, 51]. Par exemple, ce modèle particulier à récemment été étudié avec des applications en cancérologie [9].

Nos estimateurs sont basés sur l'adéquation des contours des arbres de la forêt avec leur contour (Harris path) limite moyen. Soit $\tau(n)$ un arbre de Galton-Watson conditionné à avoir n nœuds. Si on note $\mathcal{H}[\tau(n)](t)$ le processus de contour (Harris path) de $\tau(n)$, il est bien connus (voir [1]) que $\mathcal{H}[\tau(n)]$ converge en loi vers une excursion Brownienne $(\frac{1}{\sigma}e_t,\ t \in [0,1])$.

Theorem 1.6.1 (Aldous, 1991)**.** *Quand n tend vers l'infini,*

$$\left(\frac{\mathcal{H}[\tau(n)](2nt)}{\sqrt{n}},\ t \in [0,1]\right) \xrightarrow{(d)} \left(\frac{2}{\sigma}e_t,\ t \in [0,1]\right),$$

dans $\mathcal{C}([0,1],\mathbb{R})$ où $(e_{t,\ t\in\mathbb{R}_+})$ est une excursion Brownienne renormalisée.

Dans ce travail nous introduisons deux estimateurs : le premier $\widehat{\lambda}_{ls}$ est basé sur l'adéquation, au sens L^2, du contour de la forêt (la concaténation des contours de chaque arbres) avec le contour limite moyen. Plus précisément, $\widehat{\lambda}_{ls}$ est définie par

$$\widehat{\lambda}_{ls} = \mathrm{argmin}_{\lambda\in\mathbb{R}_+} \|\mathcal{H}[\mathcal{F}](\cdot) - \lambda H\|^2_{L^2([0,N])},$$

avec

$$\mathcal{H}[\mathcal{F}](t) = \sum_{i=1}^{N} \frac{1}{\sqrt{n_i}} \mathcal{H}[\tau_i](2n_i(t-i+1))\mathbb{1}_{[i-1,i)}(t),\ \forall\, 0 \le t \le N,$$

et

$$H(t) = \mathbb{E}e_{(t-\lfloor t \rfloor)}, \forall\, 0 \leq t \leq N,$$

où $\lfloor \cdot \rfloor$ est la partie entière inférieure.

Un second estimateur est construit de la manière suivante. Pour chaque arbre de la forêt τ_i, on considère la quantité

$$\widehat{\lambda}[\tau_i] = \frac{\langle \mathcal{H}[\tau_n](\cdot), \mathbb{E}\left[e.\right] \rangle_{L^2([0,1])}}{\|\mathbb{E}[e.]\|_{L^2([0,1])}},$$

où $\langle \cdot, \cdot \rangle_{L^2([0,1])}$ est le produit scalaire dans $L^2([0,1])$. De cette manière, $\widehat{\lambda}[\tau_i]$ est la projection de $\mathcal{H}[\tau_i]$ sur le sous espace de $L^2[0,1]$ engendré par $\mathbb{E}[e.]$. Cette quantité mesure l'adéquation du contour de τ_i avec son contour limite moyen. L'estimateur est alors construit comme le paramètre λ qui minimise l'écart (au sens de Wasserstein) entre la loi attendue à la limite, celle de la v.a.

$$\lambda \frac{\langle e., \mathbb{E}\left[e.\right] \rangle_{L^2([0,1])}}{\|\mathbb{E}[e.]\|_{L^2([0,1])}} =: \lambda \Lambda_\infty$$

et la mesure empirique

$$\mathcal{P} = \frac{1}{N} \sum_i \delta_{\widehat{\lambda}[\tau^i_{n_i}]}.$$

Plus précisément,

$$\widehat{\lambda}_W = \mathrm{argmin}_{\lambda > 0} d_W \left(\mathbb{P}_{\lambda \Lambda_\infty}, \mathcal{P} \right),$$

où $\mathbb{P}_{\lambda \Lambda_\infty}$ est la loi de $\lambda \Lambda_\infty$. d_W est la distance de Wasserstein L^2 définie, pour toutes mesures de probabilité μ et ν sur $\mathbb{R}$ par

$$d_W(\mu, \nu) = \inf \left\{ \sqrt{\int_{\mathbb{R}^2} |x-y|^2 \ \gamma(dx,dy)} \mid \gamma \in \mathcal{M}_1(\mathbb{R}^2),\ \pi_1 \gamma = \mu,\ \pi_2 \gamma = \nu \right\},$$

où $\mathcal{M}_1(\mathbb{R}^2)$ est l'ensemble des mesures de probabilité sur $\mathbb{R}^2$, π_i est la projection suivant la $i$ème coordonnée, et $\pi_i \gamma$ est la mesure image de γ par π_i.

Les principaux résultats théoriques sur nos estimateurs sont l'absence de biais asymptotique ainsi qu'une convergence presque sûre du type suivant.

Theorem 1.6.2. *Soit $(u_n)_{n \geq 1}$ une suite d'entier et $\mathcal{F} = (\tau_n)_{n \geq 1}$ une famille infinie d'arbre de Galton-Watson conditionnés de taille respective u_n et de loi de naissance commune μ. Alors,*

$$\forall\, \epsilon > 0,\ \exists\, A \in \mathbb{N}, \left(\min_{n \geq 1} u_n > A \Rightarrow \mathbb{P} \left(\limsup_{N \to \infty} \left| \widehat{\lambda}.[\mathcal{F}_N] - \sigma^{-1} \right| < \epsilon \right) = 1 \right),$$

où $\widehat{\lambda}.[F_N]$ peut être arbitrairement $\widehat{\lambda}_{ls}[F_N]$ ou $\widehat{\lambda}_W[F_N]$.

La première difficulté pour obtenir ces résultats est de démontrer que la variable aléatoire Λ_∞ possède une densité par rapport à la mesure de Lebesgue. Pour ce faire on utilise le calcul de Malliavin. Par la suite, on utilise la théorie des opérateurs de Bernstein-Kantorovich (qui apparaissent naturellement dans les calculs) et des méthodes standards sur les distances de transport.

La dernière partie du chapitre est consacrée à des tests numériques sur nos estimateurs ainsi qu'à la comparaison de nos résultats avec des estimateurs concurrents [9].

Chapitre 2

Preliminaries I : Fluctuation of Lévy processes in a nutshell

The purpose of this chapter is to introduce the fluctuation theory of Lévy processes. Our motivation is that, in Chapter 3, the contour process of a splitting tree (which describes our population dynamics) is almost a Lévy process with no negative jumps. Since this process gives many informations on the underlying tree, it follows that many properties of the splitting trees can be deduce thanks to the tools provided by the theory of Lévy processes.

The theory of Lévy processes is rich and the reader may object that a regular nutshell may not be large enough to contain a complete account on the fluctuations of Lévy processes. It is true, and the present chapter is not designed to be an exhaustive or fully rigorous treatment of this theory. Our goal is rather to give an intuitive treatment of it. It is designed to go as straightforward as possible to the fluctuation identities used in the sequel of this manuscript. That is why, most of the proofs are only sketched and many technical difficulties, which are not of core importance, are evaded.
The following text is based on two excellent references by Jean Bertoin [7] and Andreas E. Kyprianou [59]. We refer the readers interested in a full and rigorous treatment of this theory to these two books.

Section 2.1 is devoted to recall some elementary properties of Poisson random measures. Such measures naturally appear when working with Lévy processes. The results recalled in Section 2.1 play a central role in the other sections of this chapter, in particular the useful compensation formula. Section 2.2 recalls basic facts on Lévy processes which are essential to go further. Section 2.3 explains the link between fluctuations of Lévy processes and excursions of Markov processes. Section 2.4 is an quick introduction to the theory of the excursions of Markov processes. This theory was developed by Itô in his famous work [45, 44]. His approach appeared to be fruitful in many domains of probability. See for instance [77] for applications in the study of Brownian motion and its functional or [68] with applications in the study of scaling limits of random trees. The interested reader can take a look to [89] or [10] for introductions to the subject. Section 2.5 introduces the main tools used in order to study fluctuations of Lévy processes : the so-called ascending and descending ladder processes. Their study is another example of application of Itô's theory. The celebrated Wiener-Hopf factorization is the main result of Section 2.5. It allows to express the law of the ladder processes in terms of the law of the underlying Lévy process. This result comes back to [36]. Section 2.6 presents the main fluctuation identities used in the other chapters of this manuscript and shows how the ladder processes can be used to solve fluctuation problems. The last part, Section 2.7, is quite independent from the rest of the chapter and is devoted to a quick reminder on renewal theory which is used in this thesis.

2.1 Some results on Poisson random measure

In this section we present two important results on Poisson random measure. The first one allows to characterize whether a random measure is Poissonian or not. The second is the celebrated compensation formula which allows computing the expectation of the integral w.r.t. a Poisson random measure.

We recall that a random measure is simply a random variable taking values in some measure space. Our first interest in such object comes from the fact that, in the second part of the manuscript, we use it to model the mutation mechanism in a biological population. In the sequel, $(E, \mathcal{E}, \eta)$ refers to a measured space such that η is σ-finite. For a measurable subset A and a random measure $\mathcal{N}$, $\mathcal{N}(A)$ is a real valued random variable. In the sequel, $\sigma(\mathcal{N}(A \cap \cdot))$ refers to the σ-field generated by the restriction of $\mathcal{N}$ on the subset A. One can easily show that

$$\sigma\left(\mathcal{N}(A \cap \cdot)\right) = \sigma\left(\{\mathcal{N}\left(A \cap B\right) \mid B \in \mathcal{E}\}\right).$$

The interested reader can find a very good introduction to random measure in [18].

Important examples of random measures are Poisson random measures. Such measures naturally appears in the theory of Lévy processes which is the main subject of this chapter. Let us recall the definition.

Definition 2.1.1. *A Poisson random measure on E with intensity η is a random measure satisfying*

- *for any measurable set A of $\mathcal{E}$, $\mathcal{N}(A)$ has Poisson distribution with parameter $\eta(A)$,*
- *for any disjoint and measurable sets A_1 and A_2, the random variables $\mathcal{N}(A_1)$ and $\mathcal{N}(A_2)$ are independent.*

The interested reader can find in [59] (Chapter 2) a quick introduction on Poisson random measure. A more exhaustive reference is [17].

We begin by recalling some useful basic results. These can be found in [59], Theorem 2.7.

Lemma 2.1.2. *Let $\mathcal{N}$ be a Poisson random measure on E with intensity η. Let f be a real-valued measurable function on E. Then, the integral*

$$\int_E f(x)\ \mathcal{N}(dx)$$

is almost surely finite if and only if

$$\int_E 1 \wedge |f(x)|\ \ \eta(dx)$$

is finite. In addition, if f is positive, we have, for any positive real number λ,

$$\mathbb{E}\left[e^{-\lambda \int_E f(x)\ \mathcal{N}(dx)}\right] = \exp\left(-\int_E \left(1 - e^{-\lambda f(x)}\right)\ \eta(dx)\right). \tag{2.1}$$

The next result provides a tool to show that a given random measure is Poissonian. This result plays an important role in the theory of excursions of Markov processes.

Proposition 2.1.3 (Poisson processes characterization of space-time Poisson random measure)**.** *Let $\mathcal{N}$ be a random measure on $\mathbb{R}_+ \times E$. Let $\mathcal{F}_t$ be the σ-field generated by $\mathcal{N}([0,t] \times E \cap \cdot)$. Then, $\mathcal{N}$ is a Poisson random measure if and only if the family of counting processes $(N^A,\ A \in \mathcal{E})$, defined by*

$$N_t^A = \mathcal{N}([0,t] \times A), \quad \forall t \in \mathbb{R}_+, \quad \forall A \in \mathcal{E},$$

satisfies

- *for any measurable set A, $\left(N_t^A,\ t \in \mathbb{R}_+\right)$ is a Poisson process which is Markovian with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.*
- *for any two disjoint and measurable sets B_1 and B_2, the processes $\left(N_t^{B_1},\ t \in \mathbb{R}_+\right)$ and $\left(N_t^{B_2},\ t \in \mathbb{R}_+\right)$ never jump simultaneously (almost surely).*

The proof of this proposition lies on the fact that Poisson processes such as those above are independent. A statement of this fact can be found in [7], Section O.4 and a proof can be found in [80], Section XII.1.

To end this section, we recall, as stated in [59], the celebrated compensation formula for functionals of Poisson random measures. This formula appears to be extremely important in the theory of Lévy processes.

Theorem 2.1.4. *Let $\varphi : \mathbb{R}_+ \times \mathbb{R} \times \Omega \to \mathbb{R}_+$ be a measurable function such that*

- *for any $t \geq 0$, $(\omega, x) \mapsto \varphi(t, x, \omega)$ is measurable with respect to $\sigma(\mathcal{N}([0,t] \times \mathbb{R} \cap \cdot)) \otimes \mathcal{B}(\mathbb{R})$.*
- *for any $x \in \mathbb{R}_+$, $t \mapsto \varphi(t, x, \omega)$ is left continuous for $\mathbb{P}$-almost all ω. Then, for any positive real number t,*

$$\mathbb{E}\left[\int_{[0,t]\times\mathbb{R}} \varphi(t,x)\ \mathcal{N}(dt,dx)\right] = \int_{[0,t]\times\mathbb{R}} \mathbb{E}\left[\varphi(t,x)\right]\ bds\ \eta(dx). \tag{2.2}$$

In Chapter 4, we show Theorem 4.2.2 which might be seen as an extension of the compensation formula for any random measure under, as expected, more restricting hypothesis on φ. The compensation formula can then be obtained as a simple corollary of this result. Unfortunately, in this corollary, a.s. continuity of $x \mapsto \varphi(t,x)$ is required for all fixed t. In that sense, Theorem 4.2.2 is not a generalisation of the compensation formula.

2.2 A quick reminder to Lévy processes

Before going further, let us recall some basic facts about Lévy processes. Here, we say that a process $(X_t,\ t \in \mathbb{R}_+)$ is a Lévy process if it is a càdlàg random process with stationary and independent increments. From this definition, it is easily seen that the law of X_t (for any positive t) is infinitely divisible (i.e. it can be written as the nth convolution power of another probability measure). It is also easy to see that the quantity $\mathbb{E}_0\left[e^{i\lambda X_1}\right]$ characterizes the law of the process.

Remark 2.2.1. *In the sequel, we use the convention that $\mathbb{P}$ refers to the measure associated to the process started from 0. In any other case, it is denoted $\mathbb{P}_x$.*

Since the law of X_1 is infinitely divisible, the Lévy-Khintchine representation theorem for infinitely divisible distribution tells us that there exists a triple (a, σ, Π) where a is a real number, σ a positive real number, and Π is a measure supported by $\mathbb{R}\backslash\{0\}$ satisfying

$$\int_{\mathbb{R}} 1 \wedge x^2 \, \Pi(dx) < \infty,$$

such that

$$\mathbb{E}\left[e^{i\lambda X_t}\right] = e^{-t\Psi(\lambda)}, \quad \forall t \in \mathbb{R}_+,$$

with

$$\Psi(\lambda) = i\lambda a + \sigma^2\lambda^2 + \int_{|x|\geq 1}\left(1 - e^{i\lambda r}\right)\Pi(dr) + \int_{|x|<1}\left(1 - e^{i\lambda r} + i\lambda r\right)\Pi(dr). \tag{2.3}$$

Ψ is called the characteristic exponent of X. The interested reader can find a statement of this theorem in [59] (Theorem 1.3), and a proof in [86].
There exists a more precise and powerful result. Indeed, the celebrated Lévy-Ito decomposition theorem gives an interpretation of the triple (a, σ, Π) in terms of the paths of the process X. More precisely, it can be showed that the law such a Lévy process is the law of the sum of three simpler independent Lévy processes $X^{(1)}, X^{(2)}$ and $X^{(3)}$, where

- $X^{(1)}$ is a drifted Brownian motion with drift a and diffusion coefficient σ,
- $X^{(2)}$ is a compound Poisson process with rate $\Pi(\mathbb{R}\backslash(-1,1))$ and jump law given by the probability measure
$$\frac{\Pi(\cdot \cap \mathbb{R}\backslash(-1,1))}{\Pi(\mathbb{R}\backslash(-1,1))},$$
- $X^{(3)}$ is a square integrable martingale.

Each of the three terms in (2.3) correspond to the characteristic exponents of each of these three processes. The Lévy-Ito decomposition is the subject of Chapter 2 in [59].

In the particular case where X is spectrally positive, meaning that the process never experiences negative jump, the exponential moments,

$$\mathbb{E}\left[e^{-\beta X_1}\right],$$

are finite for all positive β. In such case, one can consider the so-called Laplace exponent of the process, denoted by ψ here, and defined by

$$\psi(\beta) = \log \mathbb{E}\left[e^{-\beta X_1}\right], \ \forall \beta \geq 0.$$

For this, we refer to [59], Section 3.3.

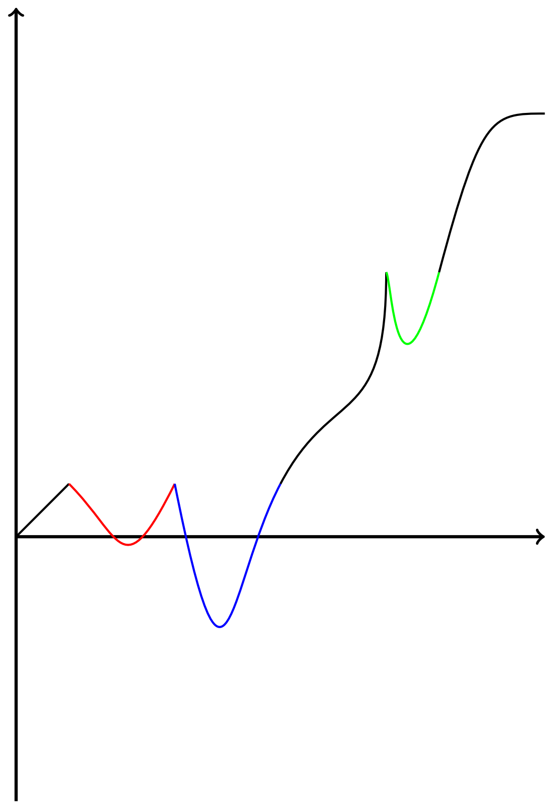

FIGURE 2.1 – The Lévy process X and its excursions (in colours) below its running maxima.

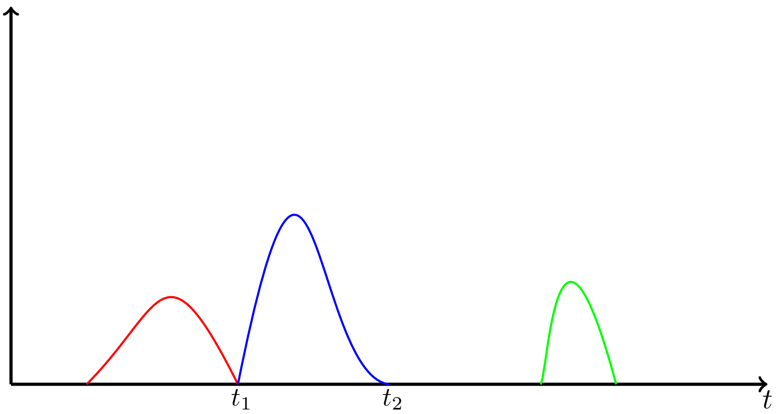

FIGURE 2.2 – The reflected process Y.

2.3 Fluctuations

The idea which allows to handle fluctuation problems is to decompose the path of X in terms its excursions below its running maxima and above its running minima (see Figure 2.1). Let $\overline{X}$ be the running supremum of X, i.e. the process defined by

$$\overline{X}_t = \sup_{s \in [0,t]} X_s, \quad \forall t \in \mathbb{R}_+.$$

It appears that this process remains constant as soon as X experiences an excursion below its maximum. Hence, the process Y defined by

$$Y_t = \overline{X}_t - X_t, \quad \forall t \in \mathbb{R}_+,$$

only contains the informations of these excursions (see Figure 2.2). The key fact is that this process remains Markovian when X is Lévy (see Proposition IV.1 in [7]). From this, it follows that studying the excursions of X below its running maxima boils to study the excursions of Y from 0 (see Figures 2.1 and 2.2).

2.4 Excursions of a Markov process away from zero

In this section, we consider a Markov process Y (which plays the role of our reflected process). We are interested in studying its excursions from 0. Usually, such study begins with a discussion about the behaviour of the process around 0. Since such discussion leads to technicalities which are not central, we completely avoid this question. We refer the reader to [7] to find a treatment of these problems, in particular Section IV.1. In the sequel, we denote by $(\mathcal{G}_t)_{t \in \mathbb{R}_+}$ the natural filtration of Y. θ_t denotes the canonical shift operator for random processes.

The first step is to be able to quantify the time spent by the process in 0 up to a time t. However, in many cases, one cannot be able to quantify it through the Lebesgue measure since the time spent in 0 by the process is likely to be of measure 0 (think about Brownian motion, for instance). It follows that we need to measure the time spent in 0 through a different time scale, which is called "local time". This takes the form of a random process.

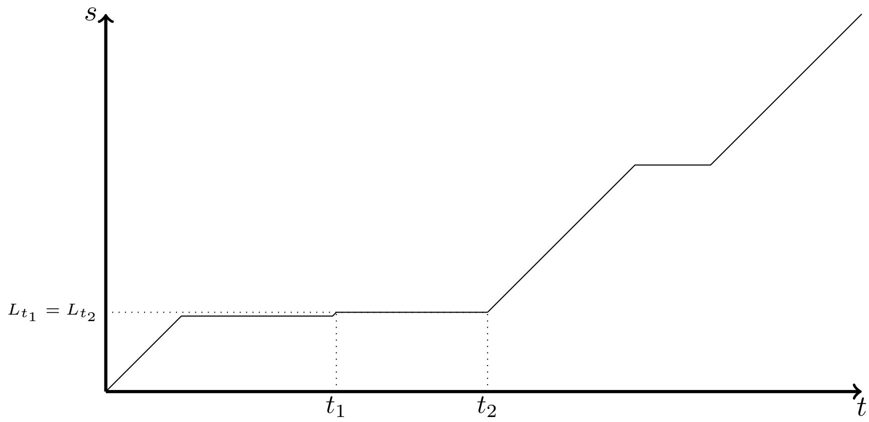

FIGURE 2.3 – The local time L.

Proposition 2.4.1 (local time at 0)**.** *There exists a non-decreasing process L (unique up to a renormalizing constant) with a.s. càdlàg paths, such that*

1. *the Stieljes measure associated to L has support the closure of* $\{t \in \mathbb{R}_+ \mid Y_t = 0\}$*,*
2. *for any stopping time* T *such that* $Y_T = 0$ *on* $\{T < \infty\}$*,*

$$\mathcal{L}_{\mathbb{P}(\cdot|T<\infty)}\Big((Y \circ \theta_T,\ L \circ \theta_T - L_T) \Big) = \mathcal{L}_{\mathbb{P}}\left((Y, L)\right).$$

Moreover, conditionally on $\{T < \infty\}$*,* $(Y \circ \theta_T,\ L \circ \theta_T - L_T)$ *is independent of* $\mathcal{G}_T$*.*

We do not prove this result. The interested reader can find a proof in [7], Section IV.2, using martingale methods.

With this idea in mind, we could roughly say that L_t (for a positive real number t) is the amount of local time spent by Y in 0 up to t regular time. We can now introduce the first quantity of great importance which is the right inverse of L,

$$L_s^{-1} = \inf\left\{t \geq 0 \mid L_t > s\right\}. \tag{2.4}$$

L^{-1} is also called the ascending ladder time process associated to X. In the same manner as L, L_s^{-1} can be interpreted as the quantity of regular time spent up to s local time. In other words, if someone wants to get s local time, he needs to wait L_s^{-1} regular time, that is $L_{L_s^{-1}} = s$.

An important property of this process is the following : let $t_1 < t_2$ two positive real numbers and suppose that Y experiences an excursion between those two times, that is

$$Y_{t_1} = Y_{t_2} = 0 \quad \text{and} \quad \forall t \in (t_1, t_2),\ Y_t > 0.$$

Hence, L_t is constant on (t_1, t_2). Hence, the quantity of regular time one needs to wait to get L_{t_1} local time is t_1, but to get $L_{t_1} + \varepsilon$ (for any positive ε) is at least t_2. Hence, L^{-1} experiences a jump at time L_{t_1} (see Figures 2.3 and 2.4). Moreover, since the local time increases again at time t_2, the size of the jump is equal to $t_2 - t_1$.

It follows that the jumps of the inverse local time L^{-1} correspond exactly to the excursions of the reflected process Y and their amplitudes are the durations of the corresponding excursions (see Figure 2.4).

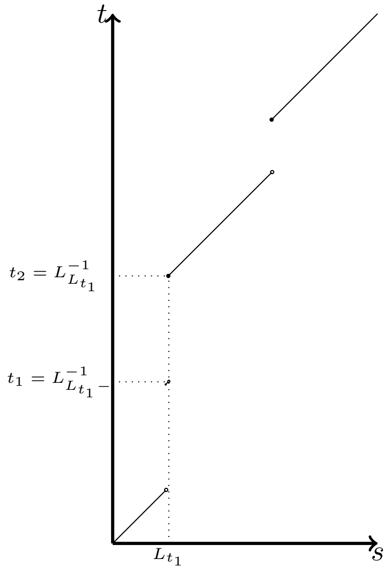

FIGURE 2.4 – The inverse local time L^{-1}.

Of course, such remarks would be pointless if the law of the inverse local time was intractable. But it appears that L^{-1} is a nice process.

Proposition 2.4.2. *The inverse local time L^{-1} is a (possibly killed) non-decreasing Lévy process (i.e. a subordinator). Hence, it can be written as*

$$L_s^{-1} = \begin{cases} \mathbf{d}s + \int_{[0,s]\times\mathbb{R}_+} x\ \mathcal{J}(dv, dx), & \textit{if } s \leq \mathcal{E}, \\ \infty & \textit{else,} \end{cases}$$

where $\mathcal{J}$ is a Poisson random measure with some intensity $J(dx)ds$, $\mathbf{d}$ is a positive real number, and $\mathcal{E}$ is an independent exponential random variable.

Remark 2.4.3 (Killing)**.** *The parameter of the random variable $\mathcal{E}$ is related to the probability of occurrence of an infinite excursion. Indeed, if an infinite excursion occurs, then the local time L remains constant from the beginning of this excursion. This implies that L^{-1} jumps to infinity. In such case L^{-1} is said to be killed. Actually, $\mathcal{E}$ is the time of the first infinite excursion of Y. In the case where the probability of occurrence of an infinite excursion is zero, then L^{-1} is a (unkilled) subordinator which simply writes*

$$L_s^{-1} = \mathbf{d}s + \int_{[0,s]\times\mathbb{R}_+} x\ \mathcal{J}(ds, dx), \quad \forall s \in \mathbb{R}_+.$$

Sketch of proof of Proposition 2.4.2. The proof simply gets back to the definition of a Lévy process in terms of its increments. We need to show that they are homogeneous in law and independents. We assume that $L_s^{-1} < \infty$ a.s. This means that the random variable $\mathcal{E}$ in the above statement equals infinity almost surely. The converse may happen when 0 is transient for Y

which means that the local time never reaches s. In this case, L^{-1} is a killed subordinator and $\{\infty\}$ is the cemetery state.

First note that it follows from (2.4) of the inverse local time that L_u^{-1} is the hitting time of (u,∞) by L. In particular, L_u^{-1} is a stopping time with respect to the natural filtration of Y. Now, using point (ii) of Definition 2.4.1, we have that

$$\left(L_{L_s^{-1}+t} - s,\ t \geq 0\right) \stackrel{d}{=} \left(L_t,\ t \geq 0\right), \quad \forall s \in \mathbb{R}_+.$$

Since, the right inverse of the process in the l.h.s. of the last equality is given by

$$\left(L_{u+s}^{-1} - L_s^{-1},\ u \in \mathbb{R}_+\right),$$

it follows that

$$L_{u+s}^{-1} - L_s^{-1} \stackrel{d}{=} L_u^{-1},$$

which gives the homogeneity of the increments. The independence is deduced from the fact that $\left(L_{L_u^{-1}+t} - s,\ t \geq 0\right)$ is independent from $\mathcal{G}_{L_u^{-1}}$ (Definition 2.4.1, point (ii)). □

Remark 2.4.4 (Laplace exponent of a killed subordinator)**.** *The Laplace exponent of a killed subordinator takes a particular form. Indeed, assume that* $(Z_t,\ t \in \mathbb{R}_+)$ *is a subordinator with triple* $(a, 0, \Pi)$*. This implies that its Laplace exponent is given by*

$$\gamma(\beta) = -a\beta - \int_{\mathbb{R}_+^\star} \left(1 - e^{-\beta x}\right) \Pi(dx), \quad \forall \beta \in \mathbb{R}_+.$$

A first remark is that $\gamma(0)$ *equals* 0 *in any case. Now, let* e_p *be an independent exponential random variable with parameter* p*. The Laplace exponent of* Z *killed at rate* p *is given by*

$$\gamma_\kappa(\beta) = -\log \mathbb{E}\left[e^{-\beta Z_1} \mathbb{1}_{e_p > 1}\right] = -p + \gamma(\beta).$$

Now, $-\gamma_\kappa(0) = p$ *which is the death rate of the killed version of* Z*.*

We can now make some remarks on L^{-1}. First, the drift part corresponds to the case where Y remains in 0 on a set of positive measure, in which case the local time is passing at a proportional speed w.r.t. to the regular time. On another side, the jump measure $\mathcal{J}$ already allows to get some informations about the excursions of Y. For instance, $\mathcal{J}\left([0,s] \times (a,\infty)\right)$ is the number of excursions with duration greater than a and its law is Poissonian (Definition 2.4.2). Hence, using that

$$\mathbb{P}\left(\mathcal{J}\left([0,s] \times (a,\infty)\right) = 0\right) = e^{-sJ((a,\infty))},$$

we have that the time of first excursion longer than a is exponentially distributed (which is not really a surprise). Another interesting example is the following : let S_a be the (local) time of first excursion with duration greater than a. Hence, $L_{S_a-}^{-1}$ is the time (in the usual time scale) where this excursion begins. It follows that the number of excursions with duration greater than $b < a$ before the first excursion with length greater than a is given by $\mathcal{J}((0, S_a) \times (b, a))$. Now, since S_a

is measurable with respect to the σ-field generated by $\mathcal{J}(\{\mathbb{R}_+ \times (a,\infty)\} \cap \cdot)$, S_a is independent from $\mathcal{J}((0,s)\times(b,a))$, for all $s>0$. This implies that

$$\begin{aligned}
\mathbb{P}\left(\mathcal{J}((0,S_a)\times(b,a))=k\right) &= \int_{\mathbb{R}_+} J\left((a,\infty)\right) e^{-sJ((a,\infty))}\mathbb{P}\left(\mathcal{J}((0,s)\times(b,a))=k\right)\, ds \\
&= \frac{J\left((a,\infty)\right)}{J\left((b,\infty)\right)}\left(1-\frac{J\left((a,\infty)\right)}{J\left((b,\infty)\right)}\right)^k. \qquad (2.5)
\end{aligned}$$

The next step is to show that the excursions themselves arrive according to a point process in some function space. First, note that the law of the path $\left(Y_{t+L^{-1}_{S_a-}},\ t\in\mathbb{R}_+\right)$ defines a probability measure, say η_a, on the space $\mathcal{E}^{(a)}$ of the excursions with length greater than a. More precisely,

$$\mathcal{E}^{(a)} = \{f\in\mathbb{D}[0,\infty) \mid f(0)=0,\ \forall t\in(0,a)\ f(t)\neq 0\},$$

endowed with the Skorohod topology of $\mathbb{D}[0,\infty)$. Now, if one wants to define a measure on the whole space of excursions $\cup_{a>0}\mathcal{E}^{(a)}$ using the family $(\eta_a)_{a>0}$, he only needs some compatibility conditions (similar to those of Kolmogorov theorem). More precisely, for each η_a, its restriction to a subspace $\mathcal{E}^{(b)}$ needs to agree with η_b. This family of measure does not satisfy this condition. However, a slight modification of this family,

$$\tilde{\eta}_a := J\left((a,\infty)\right)\eta_a, \quad \forall a\in\mathbb{R}_+,$$

does the trick.
Indeed, let $a>b$ and $A\in\mathcal{B}\left(\mathcal{E}^{(a)}\right)$, then

$$\begin{aligned}
\eta_b(A) &= \mathbb{P}\left(\left(Y_{t+L^{-1}_{S_b-}},\ t\in\mathbb{R}_+\right)\in A,\ L^{-1}_{S_b}-L^{-1}_{S_b-}>a\right) \\
&= \mathbb{P}\left(\left(Y_{t+L^{-1}_{S_a-}},\ t\in\mathbb{R}_+\right)\in A,\ L^{-1}_{s}-L^{-1}_{s-}<b,\ \forall s\in(0,S_a)\right). \\
&= \mathbb{P}\left(\left(Y_{t+L^{-1}_{S_a-}},\ t\in\mathbb{R}_+\right)\in A,\ \mathcal{J}\left((0,S_a)\times(b,\infty)\right)=0\right).
\end{aligned}$$

But the event $\left\{L^{-1}_s-L^{-1}_{s-}<b,\ \forall s\in(0,S_a)\right\}$ belongs to $\mathcal{G}_{L^{-1}_{S_a-}}$. Hence, the two events in the above probability are independent conditionally on $Y_{L^{-1}_{S_a-}}$ which is almost surely equal to 0. Finally, from (2.5),

$$\eta_b(A) = \frac{J\left((a,\infty)\right)}{J\left((b,\infty)\right)}\eta_a(A).$$

Hence, there exists a measure denoted η on $\mathcal{E}:=\cup_{a>0}\mathcal{E}^{(a)}$ such that its restriction on each subspace $\mathcal{E}^{(a)}$ coincides with $\tilde{\eta}_a$.
Now, the main result is the following.

Proposition 2.4.5. *There exists a Poisson random measure $\mathcal{H}$ on $\mathbb{R}_+\times\mathcal{E}$ with intensity $\lambda\otimes\eta$ such that, for all $s>0$, we have*

$$\mathcal{H}\left(\{\{s\}\times\mathcal{E}\}\cap\cdot\right) = \delta_{\theta_{L^{-1}_{s-}}Y}\ \mathbb{1}_{L^{-1}_{s-}-L^{-1}_{s}>0}.$$

Sketch of proof. It is easily seen that $\mathcal{H} : \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}\left(\mathcal{E}^{(a)}\right) \to \mathbb{R}_+$, defined by

$$\mathcal{H}\left([0,t] \times \ A\right) = \text{Card}\left\{ s \in [0,t] \mid L^{-1}_{s-} - L^{-1}_{s} > 0 \text{ and } \left(Y_{L^{-1}_{s-}+t},\ t \in \mathbb{R}_+\right) \in A \right\}, \quad \forall A \in \mathcal{B}\left(\mathcal{E}\right),\ t \in \mathbb{R}_+,$$

defines a random measure on $\mathbb{R}_+ \times \mathcal{E}$. This measure is the number of excursions of Y up to time t that lie in A. In order to show that $\mathcal{H}$ is Poissonian, we use Theorem 2.1.3. For any measurable set A of $\mathcal{B}\left(\mathcal{E}\right)$, the proof of the Poissonian nature of the counting process

$$N^A_t = \mathcal{H}\left([0,t] \times A\right), \ \forall t \in \mathbb{R}_+$$

lies on the same arguments as the proof of Theorem 2.4.2. More precisely, we use that a pure jump Lévy process making only jumps of size 1 is a Poisson process. Moreover, it is quite clear from the construction that $(N^A_t,\ t \in \mathbb{R}_+)$ and $(N^B_t,\ t \in \mathbb{R}_+)$ never jump simultaneously as soon as A is disjoint of B. This implies through the application of Theorem 2.1.3 that $\mathcal{H}$ is a Poissonian random measure.

As usual, the behaviour at 0 of Y leads to technical difficulties which are, as usual, eluded (see [7], theorem IV.10). □

2.5 Ladder processes and their Laplace exponents

In the preceding section, we have seen how the excursions of a Lévy process from its maximum can be described through a Poisson random measure $\mathcal{H}$. This was done using the excursions from 0 of the reflected process. However, the knowledge of $\mathcal{H}$ is not enough to recover the whole trajectory of X. Indeed, it does not describe the behaviour of X when it reaches its maximum. Hence, we need the couple $\left(\mathcal{H}, \overline{X}\right)$ to characterize X. Moreover, it appears that a slight modification of $\overline{X}$ makes it more user-friendly. Indeed, the time changed supremum process $\left(\overline{X}_{L^{-1}_s},\ s \in \mathbb{R}_+\right)$ is also almost a subordinator. The only problem lies in the fact that, if the reflected process experiences an infinite excursion, L^{-1} jumps to infinity. This implies that $\overline{X}_{L^{-1}}$ remains constant which is inconsistent with its Lévy nature. This is the motivation of the definition of the ascending height process H defined by

$$H_s = \begin{cases} \overline{X}_{L^{-1}_s} & \text{if } L^{-1}_s < \infty, \\ \infty & \text{else,} \end{cases}$$

for all s in $\mathbb{R}_+$. Indeed, we have that the 2-dimensional process $((L^{-1}_s, H_s),\ s \in \mathbb{R}_+)$ is also a (eventually killed) subordinator. This process is usually called the ascending ladder process. The proof of this fact lies on the same ideas as the proof Proposition 2.4.2. We do not write this proof here but the interested reader can find it in [7] (Lemma 2, p.157).

Now, the core point is that, in the case of Lévy processes with non negative jumps, the Laplace exponent of the ladder process can be explicitly expressed through the Laplace exponent of X. To prove that, we need to introduce the so-called descending ladder processes which plays a symmetric role as the ascending one but with the idea to decompose the path of X through the excursions from its minimum. More precisely, let $((\widehat{L}^{-1}_s,\ \widehat{H}_s),\ s \in \mathbb{R}_+)$ be the ascending ladder process of $-X$. This bivariate subordinator is called the descending ladder process of X.

We are now able to state the main theorem of this chapter which links the characteristic exponent of X with the Laplace exponent of $((\widehat{L}_s^{-1},\ \widehat{H}_s),\ s \in \mathbb{R}_+)$ and $((L_s^{-1},\ H_s),\ s \in \mathbb{R}_+)$. This is the celebrated Wiener-Hopf factorization.

Theorem 2.5.1. *(Wiener-Hopf factorization) Let Ψ the characteristic exponent of X. Let κ and $\hat{\kappa}$ be the Laplace exponents of $\left(L^{-1}, H\right)$ and $\left(\hat{L}^{-1}, \hat{H}\right)$ (respectively). Then, for all $p \in \mathbb{R}_+$,*

$$\frac{p}{p - i\theta + \psi(\lambda)} = \frac{\kappa(p,0)}{\kappa(p - i\theta, -i\lambda)} \frac{\hat{\kappa}(p,0)}{\hat{\kappa}(p - i\theta, -i\lambda)}. \tag{2.6}$$

Note that in order to give a sense to the above equality, the function κ and $\hat{\kappa}$ needs to be analytically extended to $\{z \mid \Im z \le 0\}$.

Sketch of proof. Let e_p be an exponential random variable with parameter $p \in \mathbb{R}_+$ independent of X. Now, easy calculations show that

$$\mathbb{E}\left[e^{i\theta e_p + i\lambda X_{e_p}}\right] = \frac{p}{p - i\theta + \psi(\lambda)}.$$

On the other hand, let $\overline{G}_t$ be the time of last supremum of X up to time t, that is

$$\overline{G}_t = \sup\left\{s < t \mid \overline{X}_s = X_s\right\}.$$

Now we use that $(\overline{G}_{e_p},\ \overline{X}_{e_p})$ and $(e_p - \overline{G}_{e_p},\ X - \overline{X}_{e_p})$ are independent random variables (see Lemma VI.6 in [7]). This leads to

$$\mathbb{E}\left[e^{i\theta e_p + i\lambda X_{e_p}}\right] = \mathbb{E}\left[e^{i\theta \overline{G}_{e_p} + i\lambda \overline{X}_{e_p}}\right] \mathbb{E}\left[e^{i\theta e_p - \overline{G}_{e_p} + i\lambda X_{e_p} - \overline{X}_{e_p}}\right].$$

However, the duality Lemma for Lévy processes (see [59], Lemma 3.4) tells us that the time reversed process $(X_{T-t} - X_T,\ t \in [0,T])$ has the law of $(-X_t,\ t \in [0,T])$. Hence,

$$\mathbb{E}\left[e^{i\theta e_p + i\lambda X_{e_p}}\right] = \mathbb{E}\left[e^{i\theta \overline{G}_{e_p} + i\lambda \overline{X}_{e_p}}\right] \mathbb{E}\left[e^{i\theta \underline{G}_{e_p} + i\lambda \underline{X}_{e_p}}\right],$$

where $\underline{G}$ and $\underline{X}$ are defined from $-X$ as $\overline{G}$ and $\overline{X}$ are defined from X. It remains to show that

$$\mathbb{E}\left[e^{i\theta \overline{G}_{e_p} + i\lambda \overline{X}_{e_p}}\right] = \frac{\kappa(p,0)}{\kappa(p - i\theta, -i\lambda)}.$$

We work by path decomposition. We have that

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta \overline{G}_{e_p} + i\lambda \overline{X}_{e_p}}\right] &= \mathbb{E}\int_0^\infty q e^{-qt} e^{i\theta \overline{G}_t + i\lambda \overline{X}_t} dt \\
&= \mathbb{E}\int_0^\infty q e^{-qt} e^{i\theta \overline{G}_t + i\lambda \overline{X}_t} \mathbb{1}_{X_t = \overline{X}_t} dt + \mathbb{E}\int_0^\infty q e^{-qt} e^{i\theta \overline{G}_t + i\lambda \overline{X}_t} \mathbb{1}_{X_t \neq \overline{X}_t} dt.
\end{aligned}$$

We treat the two terms in the last equality independently and begin with the second term. Now, assume we have two times t_1 and t_2 satisfying

$$X_{t_1} = \overline{X}_{t_1},\ X_{t_2} = \overline{X}_{t_2}, \quad \text{and} \quad \forall t \in (t_1, t_2),\ X_t \neq \overline{X}_t,$$

meaning that X experiences an excursion from its maximum on the interval (t_1, t_2). Hence, for all t in (t_1, t_2), G_t is constant and equal to $L^{-1}_{L_{t_1}-}$ (see all the Figures above). Similarly, $\overline{X}_t$ equals $\overline{X}_{L^{-1}_{L_{t_1}-}}$. From this, one can see that

$$e^{i\theta\overline{G}_t + i\lambda\overline{X}_t}\mathbb{1}_{X_t\neq\overline{X}_t} = \int_{\mathbb{R}_+\times\mathcal{E}} \mathbb{1}_{L^{-1}_{s-}<t<L^{-1}_s} e^{i\theta L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\mathcal{H}(ds,de).$$

Using this, one has

$$\begin{aligned}
\int_0^\infty qe^{-qt}e^{i\theta\overline{G}_t+i\lambda\overline{X}_t}\mathbb{1}_{X_t\neq\overline{X}_t}dt &= \int_0^\infty qe^{-qt}\int_{\mathbb{R}_+\times\mathcal{E}} \mathbb{1}_{L^{-1}_{s-}<t<L^{-1}_s} e^{i\theta L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\mathcal{H}(ds,de)dt \\
&= \int_{\mathbb{R}_+\times\mathcal{E}} e^{i\theta L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\left(e^{-qL^{-1}_{s-}} - e^{-qL^{-1}_s}\right)\mathcal{H}(ds,de) \\
&= \int_{\mathbb{R}_+\times\mathcal{E}} e^{(i\theta-q)L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\left(1 - e^{-q\mathcal{L}(e)}\right)\mathcal{H}(ds,de),
\end{aligned}$$

where $\mathcal{L}(e)$ denotes the length of the excursion e (that is $L^{-1}_s - L^{-1}_{s-}$). From this, the compensation formula for Poisson functional (2.2) implies that

$$\begin{aligned}
&\mathbb{E}\left[\int_0^\infty qe^{-qt}\int_{\mathbb{R}_+\times\mathcal{E}} \mathbb{1}_{L^{-1}_{s-}<t<L^{-1}_s} e^{i\theta L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\mathcal{H}(ds,de)dt\right] \\
&\qquad = \int_{\mathbb{R}_+\times\mathcal{E}} \mathbb{E}\left[e^{(i\theta-q)L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\right]\left(1 - e^{-q\mathcal{L}(e)}\right) ds\ \eta(de). \qquad (2.7)
\end{aligned}$$

Now, one has on one side,

$$\mathbb{E}\left[e^{-qL^{-1}_s}\right] = e^{s\kappa(q,0)}.$$

But on the other side,

$$\begin{aligned}
\mathbb{E}\left[e^{-qL^{-1}_s}\right] &= \mathbb{E}\left[e^{-qds - q\int_{[0,s]\times\mathbb{R}_+} x\ \mathcal{J}(ds,dx)}\right] \\
&= \mathbb{E}\left[e^{-qds - q\int_{[0,s]\times\mathcal{E}} \mathcal{L}(e)\ \mathcal{H}(ds,de)}\right].
\end{aligned}$$

Hence, using formula (2.1), one has

$$\mathbb{E}\left[e^{-qL^{-1}_s}\right] = e^{-qds}\exp\left(-\int_E\left(1 - e^{-s\mathcal{L}(e)}\right)\ \eta(de)\right).$$

Finally,

$$\int_{\mathcal{E}}(1 - e^{-q\mathcal{L}(e)})\ \eta(de) = -\kappa(q,0) - qd.$$

This, in conjunction with (2.7), entails

$$\begin{aligned}
\mathbb{E}\left[\int_0^\infty qe^{-qt}e^{i\theta\overline{G}_t+i\lambda\overline{X}_t}\mathbb{1}_{X_t\neq\overline{X}_t}dt\right] &= \int_{\mathbb{R}_+}\mathbb{E}\left[e^{(i\theta-q)L^{-1}_{s-}+i\lambda\overline{X}_{L^{-1}_{s-}}}\right] ds\ (\kappa(q,0) - qd) \\
&= \int_{\mathbb{R}_+} e^{t\kappa(q-i\theta,-i\lambda)}ds\ (-\kappa(q,0) - qd) = \frac{-\kappa(q,0) - qd}{\kappa(q-i\theta,-i\lambda)}.
\end{aligned}$$

Concerning the first term, it is easily seen that $G_t = L^{-1}_{L_t-}$ and $G_t = t$ when $X_t = \overline{X}_t$. This implies that

$$\begin{aligned}\mathbb{E}\int_0^\infty qe^{-qt}e^{i\theta\overline{G}_t+i\lambda\overline{X}_t}\mathbb{1}_{X_t=\overline{X}_t}dt &= \mathbb{E}\int_{\mathbb{R}_+} q\exp\Big((i\theta-q)L^{-1}_{L_t}+i\lambda\overline{X}_{L^{-1}_{L_t}}\Big)\mathbb{1}_{L^{-1}_{L_t-}=L^{-1}_{L_t}}dt \\ &= \mathbb{E}\int_{\mathbb{R}_+} q\exp\Big((i\theta-q)L^{-1}_s+i\lambda\overline{X}_{L^{-1}_s}\Big)\mathbb{1}_{L^{-1}_{s-}=L^{-1}_s}\, dL^{-1}_s,\end{aligned}$$

where we use that dL^{-1} is the push-forward measure of the Lesbegue measure by L to obtain the last equality. But on the set $\{s\in\mathbb{R}_+ \mid L^{-1}_{s-}=L^{-1}_s\}$, $dL^{-1}_s = \mathbf{d}ds$, hence

$$\begin{aligned}\mathbb{E}\int_{\mathbb{R}_+} q\exp\Big((i\theta-q)L^{-1}_s+i\lambda\overline{X}_{L^{-1}_s}\Big)\mathbb{1}_{L^{-1}_{s-}=L^{-1}_s}dL^{-1}_s &= \mathbf{d}\mathbb{E}\int_{\mathbb{R}_+} q\exp\Big((i\theta-q)L^{-1}_s+i\lambda\overline{X}_{L^{-1}_s}\Big)ds \\ &= \frac{q\mathbf{d}}{\kappa(q-i\theta,-i\lambda)}.\end{aligned}$$

Finally,

$$\mathbb{E}\left[e^{i\theta\overline{G}_{e_p}+i\lambda\overline{X}_{e_p}}\right] = \frac{-\kappa(p,0)}{\kappa(p-i\theta,-i\lambda)}. \tag{2.8}$$

Using the same computation on $-X$ leads to

$$\mathbb{E}\left[e^{i\theta\underline{G}_{e_p}+i\lambda\underline{X}_{e_p}}\right] = \frac{-\widehat{\kappa}(p,0)}{\widehat{\kappa}(p-i\theta,-i\lambda)}. \tag{2.9}$$

□

The Wiener-Hopf factorization for spectrally positive Lévy processes

Here, we focus on spectrally positive Lévy processes which is the type of processes that appear in the sequel of this manuscript. A spectrally positive Lévy process is supposed to satisfy the condition $\Pi(\mathbb{R}_-)=0$ meaning that the path of the process does not experience negative jumps. In this case, the Wiener-Hopf factorization of Theorem 2.5.1 takes a simpler form. Indeed, first note that when X is a spectrally positive Lévy process, $-\underline{X}$ satisfies the definition of a local time at 0 for the reflected process at the minimum. This implies, since

$$\widehat{L}^{-1}_s = \inf\{t\geq 0 \mid \underline{X}_t < -s\},$$

that the descending ladder process $\widehat{L}^{-1}$ is nothing more than the hitting time of $(-\infty,-s]$ by X. It follows then, using the fact that the process

$$\Big(e^{-\alpha X_t - t\psi(\alpha)},\ t\in\mathbb{R}_+\Big)$$

is a martingale and Doob's optimal stopping theorem at the stopping time $\widehat{L}^{-1}_s$ that

$$\mathbb{E}\left[e^{-\psi(\alpha)\widehat{L}^{-1}_s}\right] = e^{\alpha s}.$$

Finally, since $\underline{X}_{\widehat{L}^{-1}_s} = s$ (using that X decreases only continuously), we have that

$$\mathbb{E}\left[e^{-\alpha\widehat{L}^{-1}_s-\beta\hat{H}_s}\right] = e^{-s(-\phi(\alpha)+\beta)}, \tag{2.10}$$

where ϕ is the right inverse of ψ. Hence, $\hat{\kappa}(\alpha, \beta) = \beta - \phi(\alpha)$. Now, (2.6) entails that

$$\frac{\kappa(p+\alpha,\beta)}{\kappa(p,0)} = \frac{\phi(p)}{p}\frac{\psi(\beta)-(\alpha+p)}{\phi(\alpha+p)-\beta}.$$

Finally,

$$\kappa(\alpha,\beta) = \frac{\psi(\beta)-\alpha}{\phi(\alpha)-\beta}. \tag{2.11}$$

2.6 Fluctuation problems for spectrally positive Lévy processes

The purpose of this section is to use the ladder processes in order to solve fluctuation problems.

2.6.1 Large time behaviour

In this section, we are interested in studying the asymptotic behaviour of X. In particular, does X drifts to $\pm\infty$ or not ? A first necessary condition for our Lévy process to drift to ∞ is that H_t drifts to ∞ as t grows. It is clear from the homogeneity of its increments that a subordinator always drifts to infinity unless it is constant or killed. Hence, it follows that the finiteness of the overall supremum of X depends of the killing of H. This last fact can be seen from the value at 0 of its Laplace exponent. We know from (2.11) that the Laplace exponent of H is given by

$$-\frac{\psi(\beta)}{\beta-\phi(0)}, \tag{2.12}$$

where we recall that ψ is the Laplace exponent of X and ϕ its right-inverse. Using that

$$\psi(\beta) = -a\beta + \frac{1}{2}\sigma^2\beta^2 - \int_{\mathbb{R}_+^\star}\left(1-e^{-\beta x}+\beta x\mathbb{1}_{x<1}\right)\Pi(dx),$$

it is easily seen that ψ is twice differentiable on $\mathbb{R}_+\backslash\{0\}$. One can then use this to show that ψ is convexe. Using its convexity, ψ has a positive zero if and only if $\psi'(0+) < 0$. In that case, according to (2.12), the Laplace exponent of H takes value 0 at 0 (since $\phi(0) > 0$). When, $\psi'(0+) \geq 0$, $\kappa(0,0)$ equals to $\psi'(0+)$. It follows, that H is a killed subordinator only when $\psi'(0+) > 0$.

Suppose, for now, that $\psi'(0+) > 0$. We show that X drifts to $-\infty$. From the remarks above, $\psi'(0+) > 0$ implies that the overall maximum of X, say $\overline{X}_\infty$, is a.s. finite (since H is killed). On the other hand, it follows from (2.10) that $\widehat{\kappa}(0,0)$ equals 0 meaning that $\widehat{H}$ drifts to infinity. Hence,

$$\sup_{t\geq 0} X_t < \infty \text{ a.s. and } \inf_{t\geq 0} X_t = -\infty \text{ a.s.}$$

Now, let x be a positive real number, we have

$$\mathbb{P}\left(\limsup_{t\to\infty} X_t \geq -\frac{x}{2}\right) \leq \mathbb{P}\left(\sup_{t>\tau_{-x}^-} X_t \geq -\frac{x}{2}\right),$$

where

$$\tau_a^- = \inf\{t\geq 0 \mid X_t < a\}, \quad \forall a \in \mathbb{R}.$$

But, by the strong Markov property, we have

$$\mathbb{P}\left(\sup_{t>\tau_{-x}^{-}} X_t \geq -\frac{x}{2}\right) = \mathbb{E}\left\{\mathbb{P}_{X_{\tau_{-x}^{-}}}\left(\overline{X}_\infty \geq -\frac{x}{2}\right)\right\} \leq \mathbb{P}_{-x}\left(\overline{X}_\infty \geq -\frac{x}{2}\right),$$

since $X_{\tau_{-x}^{-}} \leq -x$. Using the properties of the increment of X, the r.h.s. of the last inequality equals

$$\mathbb{P}\left(\overline{X}_\infty \geq \frac{x}{2}\right).$$

To end, note that, since $\overline{X}_\infty$, is a.s. finite, this last probability tends to 0 as x increases. Hence, $\mathbb{P}(\limsup_{t\to\infty} X_t \geq -\frac{x}{2})$ equals 0.

Using a similar method, on can show that X drifts to $-\infty$ if $\psi'(0+) < 0$, and oscillate if $\psi'(0+) = 0$, that is

$$\limsup_{t\to\infty} X_t = \infty \quad \text{and} \quad \liminf_{t\to\infty} X_t = -\infty, \quad \text{almost surely.}$$

Note that this last property gives no information on the recurrence of the process X.

2.6.2 Exit problems for spectrally negative Lévy process

Now, we are interested in the exit time of X from an interval. Let

$$\tau_a^{+} = \inf\{t > 0 \mid X_t > a\} \quad \text{and} \quad \tau_a^{-} = \inf\{t > 0 \mid X_t < a\}$$

the exit times upward and downward.
Let $a > b$, we are interested, for any $x \in (b, a)$, in the evaluation of the probability

$$\mathbb{P}_x\left(\tau_b^{-} < \tau_a^{+}\right).$$

By the properties of homogeneity of the Lévy process, this boils to study the case $b = 0$. Moreover, we have

$$\mathbb{P}_x\left(\tau_0^{-} < \tau_a^{+}\right) = \mathbb{P}_{x-a}\left(\tau_{-a}^{-} < \tau_0^{+}\right).$$

Now we use that the probability that the overall maximum $\overline{X}_\infty$ is lower than 0 starting from $x - a$ is equal to the probability to hit $-a$ before 0 and then that $\overline{X}_\infty$ is greater than 0. More, precisely

$$\begin{aligned}
\mathbb{P}_{x-a}\left(\overline{X}_\infty \leq 0\right) &= \mathbb{E}\left\{\mathbb{P}_{X_{\tau_{-a}^{-}}}\left(\overline{X}_\infty \leq 0\right)\mathbb{1}_{\tau_{-a}^{-}<\tau_0^{+}}\right\} \\
&= \mathbb{P}_{x-a}\left(\tau_{-a}^{-} < \tau_0^{+}\right)\mathbb{P}_{-a}\left(\overline{X}_\infty \leq 0\right).
\end{aligned}$$

Hence, we have that

$$\mathbb{P}_x\left(\tau_0^{-} < \tau_a^{+}\right) = \frac{\mathbb{P}_{x-a}\left(\overline{X}_\infty \leq 0\right)}{\mathbb{P}_{-a}\left(\overline{X}_\infty \leq 0\right)}.$$

On the other hand,

$$\mathbb{E}\left[e^{-\alpha\overline{X}_\infty}\right] = \int_0^\infty \alpha e^{-\alpha s}\mathbb{P}\left(\overline{X}_\infty \leq s\right)\, ds = \int_0^\infty \alpha e^{-\alpha s}\mathbb{P}_{-s}\left(\overline{X}_\infty \leq 0\right)\, ds.$$

Its now time to use what we know about the ascending ladder process. Let $\mathcal{H}$ be a subordinator with Laplace exponent given by $\kappa(0,\beta)-\kappa(0,0)$. This means that $\mathcal{H}$ has the same law as H with the difference that it is not killed. Let also $\mathcal{E}$ be an independent exponential random variable with parameter $\psi'(0+)$. Since $\mathcal{H}$ took at its killing time is equal to the overall supremum of the process X, we have

$$\begin{aligned} \mathbb{E}\left[e^{-\alpha\overline{X}_\infty}\right] &= \mathbb{E}\left[e^{-\alpha\mathcal{H}_{\mathcal{E}}}\right] = \int_0^\infty \psi'(0+)e^{-\psi'(0+)t}\mathbb{E}\left[e^{-\alpha\mathcal{H}_t}\right] dt \\ &= \int_0^\infty \psi'(0+)e^{-\psi'(0+)t}e^{t\kappa(0,\beta)}dt \\ &= \psi'(0+)\frac{\alpha}{\psi(\alpha)}. \end{aligned}$$

Finally, $\mathbb{P}_{-s}\left(\overline{X}_\infty \leq 0\right)$ satisfies

$$\int_0^\infty \frac{e^{-\alpha s}}{\psi'(0+)}\mathbb{P}_{-s}\left(\overline{X}_\infty \leq 0\right)\, ds = \frac{1}{\psi(\alpha)}.$$

The function $x \mapsto \frac{1}{\psi'(0+)}\mathbb{P}_{-x}\left(\overline{X}_\infty \leq 0\right)$ is called the scale function of X and is denoted W.

Now, we want to go a little further. Let τ be the exit time of the interval $(-a,0)$. We are interested in the law of the couple $(X_{\tau-},\ X_\tau)$. As usual, there is no restriction in taking $(-a,0)$, but there are two reasons for this choice. First, it is in that case that the law of the couple is the easier to write. Second, our real interest lies in the overshoot and undershoot of the Lévy process (see Figure 2.5) over a fixed level which has exactly the law of $(-X_{\tau-},\ X_\tau)$ when the chosen interval is $(-a,0)$ (this follows again from the homogeneity of X). Of course, since X is spectrally positive we have $X_{\tau-}=X_\tau=-a$ almost surely on the event $\left\{\tau^-_{-a}<\tau^+_0\right\}$. It is more interesting to see what happens when $\tau^+_0<\tau^-_{-a}$ because X can cross the border by jumping above. Now let A and B be two Borel sets such that $A\subset(-a,0)$ and $B\subset[0,\infty)$. We have, for x in $[0,a)$,

$$\mathbb{P}_{-x}\left(X_{\tau-}\in A,\ X_\tau\in B\right) = \mathbb{E}_{-x}\left[\int_{[0,\infty)\times\mathbb{R}} \mathbb{1}_{\overline{X}_t\leq 0,\ \underline{X}_t\geq -a}\mathbb{1}_{X_{t-}\in A}\mathbb{1}_{X_{t-}+y\in B}\ \mathcal{N}(dt,dy)\right].$$

Now, the compensation formula (2.2) entails that

$$\begin{aligned} \mathbb{P}_{-x}\left(X_{\tau-}\in A,\ X_\tau\in B\right) &= \int_{[0,\infty)\times\mathbb{R}} \mathbb{E}_{-x}\left[\mathbb{1}_{\overline{X}_t\leq 0,\ \underline{X}_t\geq -a}\mathbb{1}_{X_{t-}\in A}\mathbb{1}_{X_{t-}+y\in B}\right]\ dt\ \Pi(dy) \\ &= \int_{[0,\infty)} \mathbb{E}_{-x}\left[\mathbb{1}_{\tau>t}\mathbb{1}_{X_t\in A}\Pi(B-X_t)\right]\ dt \\ &= \int_A \Pi(B-y)\ \mathcal{U}(-x,dy), \end{aligned}$$

where $\mathcal{U}$ is the mean occupation measure of X up to its first exit of $(-a,0)$. That is the measure defined by

$$\mathcal{U}(x,A) = \int_{\mathbb{R}_+} \mathbb{P}_x\left(X_t\in A,\ \tau>t\right) dt, \quad \forall A\in\mathcal{B}\left((-a,0]\right).$$

Moreover, we have, for a positive real number x,

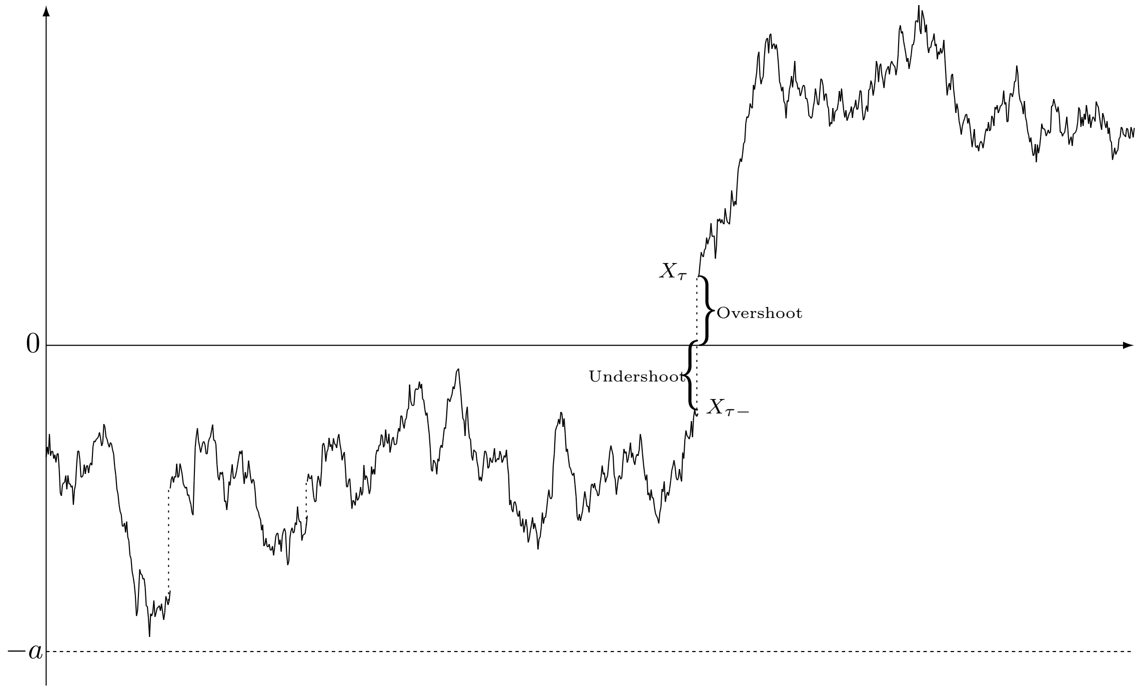

FIGURE 2.5 – Undershoot and overshoot of the process X at the exit time τ of the interval $(-a, 0)$.

$$\int_{\mathbb{R}_+} \mathbb{P}_{-x}\left(X_t \in A,\ \tau > t\right) dt$$
$$= \int_{\mathbb{R}_+} \mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \leq 0\right) - \mathbb{P}_{-x}\left(X_t \in A,\ \underline{X}_t < -a,\ \overline{X}_t \leq 0\right) dt. \quad (2.13)$$

But the probability $\mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \leq 0,\ \underline{X}_t < -a\right)$ rewrites

$$\begin{aligned} \mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \leq 0,\ \underline{X}_t < -a\right) &= \mathbb{P}_{-x}\left(\tau_{-a}^- < \tau_0^+\right) \mathbb{P}_{-a}\left(X_t \in A,\ \overline{X}_t \leq 0\right) \\ &= \frac{W(x)}{W(a)} \mathbb{P}_{-a}\left(X_t \in A,\ \overline{X}_t \leq 0\right). \end{aligned} \quad (2.14)$$

From this point, we focus on the probability in the r.h.s of the last equality. Now, as in the proof of Theorem 2.5.1, let e_p be an exponential random variable with parameter $p > 0$ independent of X. We have

$$\begin{aligned} \int_{\mathbb{R}} e^{-pt} \mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \leq 0\right) dt &= \frac{1}{p} \mathbb{P}_{-x}\left(X_{e_p} \in A,\ \overline{X}_{e_p} \leq 0\right) \\ &= \frac{1}{p} \mathbb{P}_{-x}\left(X_{e_p} - \overline{X}_{e_p} + \overline{X}_{e_p} \in A,\ \overline{X}_{e_p} \leq 0\right). \end{aligned}$$

Now, using again the elements given in the proof of Theorem 2.5.1, $X_{e_p} - \overline{X}_{e_p}$ and $\overline{X}_{e_p}$ are independent. Moreover, $X_{e_p} - \overline{X}_{e_p}$ has the law of $\underline{X}_{e_p}$, which is exponential with parameter $\phi(p)$

according to (2.9) and (2.10). Hence, denoting by $\mathbb{P}_{\overline{X}_{e_p}}$ the law of $\overline{X}_{e_p}$, we get

$$\int_{\mathbb{R}} e^{-pt}\mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \le 0\right) dt \;=\; \int_{\mathbb{R}_-} \frac{\phi(p)}{p} e^{\phi(p)y} \int_{\mathbb{R}} \mathbb{1}_{y+z-x\in A}\mathbb{1}_{z-x\le 0}\mathbb{P}_{\overline{X}_{e_p}}(dz)dy.$$

Now, using (2.8) and (2.11), we have that $\frac{\phi(p)}{p}\mathbb{P}_{\overline{X}_{e_p}}$ converges weakly, as p goes to zero, to a measure whose Laplace transform is given by $\frac{\beta-\phi(0)}{\psi(\beta)}$. This corresponds to the measure

$$W(dz) - \phi(0)W(z)dz,$$

where $W(dz)$ the Stieljes measure associated to W. Consequently,

$$\int_{\mathbb{R}} \mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \le 0\right) dt \;=\; \int_{\mathbb{R}_+} e^{\phi(0)y} \int_{\mathbb{R}} \mathbb{1}_{y+z-x\in A}\mathbb{1}_{z-x\le 0}\left(W(dz) - \phi(0)W(z)dz\right) dy.$$

A simple change of variable leads to

$$\int_A e^{\phi(0)y} \int_{[x+y,x]} e^{-\phi(0)(x+z)}\left(W(dz) - \phi(0)W(z)dz\right) dy.$$

Now, integrating by parts entails

$$\mathbb{P}_{-x}\left(X_t \in A,\ \overline{X}_t \le 0\right) = \int_A e^{\phi(0)y}W(x) - W(x+y)dy.$$

Using this last equality in conjunction with (2.13) and (2.14) leads to

$$\mathcal{U}(-x, A) = \int_A \frac{W(x)W(a+y)}{W(a)} - W(x+y)dy,$$

and to

$$\mathbb{P}_{-x}\left(X_{\tau-} \in A,\ X_\tau \in B\right) \;=\; \int_A \Pi(B-y)\ \left(\frac{W(x)W(a+y)}{W(a)} - W(x+y)\right) dy.$$

We summarize these results in the following Theorem.

Theorem 2.6.1. *Let X be a spectrally positive Lévy process with Laplace exponent given by ψ. Let W be the unique increasing function satisfying*

$$\int_{\mathbb{R}_+} e^{-\beta t}W(t)\ dt = \frac{1}{\psi(\beta)}, \quad \forall \beta \in \mathbb{R}_+.$$

Then,

$$\mathbb{P}_x\left(\tau_a^+ < \tau_0^-\right) = \frac{W(x-a)}{W(a)}, \quad \forall a \in \mathbb{R}_+,\ x \in (0, a).$$

In addition, if $\tau = \tau_a^+ \wedge \tau_0^-$. The law of the overshoot O^+ and the undershoot O^- of the process when crossing level a is given by

$$\mathbb{P}_x\left(O^- \in A,\ O^+ \in B\right) = \int_A \Pi(B+y)\ \left(\frac{W(a-x)W(a-y)}{W(a)} - W(a-x-y)\right) dy, \quad \forall x > 0,$$

for any A in $\mathcal{B}((0, a])$ and B in $\mathcal{B}(\mathbb{R}_+)$.

2.7 Reminder on renewal theory

The purpose of this part is to recall some facts on renewal equations borrowed from [30]. Let $h : \mathbb{R} \to \mathbb{R}$ be a function bounded on finite intervals with support in $\mathbb{R}_+$ and Γ a probability measure on $\mathbb{R}_+$. The equation

$$F(t) = \int_{\mathbb{R}_+} F(t-s)\Gamma(ds) + h(t),$$

called a renewal equation, is known to admit a unique solution finite on bounded interval.
Here, our interest is focused on the asymptotic behavior of F. We said that the function h is DRI (directly Riemann integrable) if for any $\delta > 0$, the quantities

$$\delta \sum_{i=0}^{n} \sup_{t\in[\delta i, \delta(i+1))} f(t)$$

and

$$\delta \sum_{i=0}^{n} \inf_{t\in[\delta i, \delta(i+1))} f(t)$$

converge as n goes to infinity respectively to some real numbers I^{δ}_{sup} and I^{δ}_{inf}, and

$$\lim_{\delta\to 0} I^{\delta}_{sup} = \lim_{\delta\to 0} I^{\delta}_{inf} < \infty.$$

In the sequel, we use the two following criteria for the DRI property :

Lemma 2.7.1. *Let h a function as defined previously. If h satisfies one of the next two conditions, then h is DRI :*

1. *h is non-negative decreasing and classically Riemann integrable on $\mathbb{R}_+$,*
2. *h is càdlàg and bounded by a DRI function.*

We can now state the next result, which is constantly used in the sequel.

Theorem 2.7.2. *Suppose that Γ is non-lattice, and h is DRI, then*

$$\lim_{t\to\infty} F(t) = \gamma \int_{\mathbb{R}_+} h(s)ds,$$

with

$$\gamma := \left(\int_{\mathbb{R}_+} s \; \Gamma(ds) \right)^{-1},$$

if the above integral is finite, and zero otherwise.

Remark 2.7.3. *In particular, if we suppose that Γ is a measure with mass lower than 1, and that there exists a constant $\alpha \geq 0$ such that*

$$\int_{\mathbb{R}_+} e^{\alpha t}\Gamma(dt) = 1,$$

then, one can perform the change a measure

$$\widetilde{\Gamma}(dt) = e^{\alpha t}\Gamma(dt),$$

in order to apply Theorem 2.7.2 to a new renewal equation to obtain the asymptotic behavior of F. (See [30] for details). This method is also used in the sequel.

Chapitre 3

Preliminaries II : Splitting trees

The purpose of this chapter is to present splitting trees in a user-friendly fashion. Almost all the results presented in this chapter come from [60]. A splitting tree is a kind of planar rooted random tree which can be used to describe the dynamics of a biological population. In contrary with the well known Galton-Watson trees which only take into account the genealogical structure of the population, the splitting trees contain informations on the lifetimes of the individuals. For this purpose, individuals are not represented by the nodes of the tree but by its branches. A branch is supposed to have a length equal to the lifetime of the corresponding individual. Hence, the splitting trees describe in a more robust way the dynamics of the population through time.

These trees have been introduced by Geiger and Kersting in [35]. In their work, the authors introduce a contour process which can be roughly described as a height process in a depth-first exploration of the tree. The purpose of their paper is then to study this process and its links to Poisson point processes. Their method allows one of the authors to study splitting trees under various conditioning [34].

Later, in [60], A. Lambert introduced a new contour process which appears to be (almost) a Lévy process. His new method appeared to be fruitful to derive properties of the splitting trees and their functionals (for instance on some particular Crump-Mode-Jagers processes). Many of the results of this thesis were obtained thanks to the tools introduced in [60].

In Section 3.1, we describe a construction of the splitting trees based on the one given in [60]. Section 3.2 is dedicated to the construction of the contour process introduced in [60]. Note that, in contrary with [60], we do not consider trees which can be "locally infinite". This leads to some simplifications. In Section 3.3, the contour process is used to derive basic properties of binary homogeneous Crump-Mode-Jagers processes used in the next chapters. Section 3.4 introduces the backward model associated to splitting trees. In any model of population dynamics, it is interesting to understand the links between the lineages of individuals alive at a fixed time. This is especially true for population genetics. For instance, the Kingman's famous coalescent model is derived from the Wright-Fisher model of population dynamics. The backward model associated to a splitting tree is the so-called *coalescent point process*. Its law is, once again, studied through the contour process.

3.1 Construction

The purpose of this section is to give the mathematical formalism underlying the theory of splitting trees. The first part describes in which space of trees a splitting tree belongs. The second part gives a characterization of the law of a splitting tree. The construction is based on Lambert [60].

The Ulam-Harris-Neveu set and the discrete genealogy
From the mathematical point of view, a splitting tree is a random variable with value in a set of trees with branch length. This set is a subset of $\mathcal{P}(\mathcal{U} \times \mathbb{R})$ where

$$\mathcal{U} := \bigcup_{n \geq 0} \mathbb{N}^n,$$

where $\mathbb{N}^0$ equals $\{\emptyset\}$. $\mathcal{U}$ is the well known Ulam-Harris-Neveu set which means to describe the genealogical structure of the individuals in the tree.

In the sequel, for any $\sigma \in \mathcal{U}$, we denote, for any non-negative integer k, σ^k the kth last ancestor of σ. That is

$$\forall k \in \mathbb{N}, \quad \sigma^k = (\sigma_1, \ldots, \sigma_{n-k}).$$

In this manner, σ^0 equals σ and σ^1 is the parent of σ. By the way, if $\sigma = (\sigma_1, \ldots, \sigma_n)$ belongs to $\mathbb{N}^n$, for some integer n, and if $k \geq n$, we assume that σ^k equals $\emptyset$. $\emptyset$ is called the *ancestor individual*.

Let $P_{\mathcal{U}}$ (resp. $P_{\mathbb{R}}$) be the canonical projection from $\mathcal{U} \times \mathbb{R}$ onto $\mathcal{U}$ (resp. $\mathbb{R}$). For a tree $\mathbb{T}$, $P_{\mathcal{U}}(\mathbb{T})$ can be thought of as the underlying discrete genealogical tree. In the sequel, we denote by $\mathcal{G}$ this discrete genealogy. In order to be admissible as a tree, a subset $\mathbb{T}$ of $\mathcal{U} \times \mathbb{R}$ needs to have a discrete genealogy $\mathcal{G}$ satisfying some compatibility conditions.

Compatiblity conditions on the discrete genealogy $\mathcal{G}$:

- The ancestor belongs to the tree : $\emptyset \in \mathcal{G}$.
- If an individual belongs to the tree, so does its parent :
  $\forall \sigma \in \mathcal{G}, \quad \sigma^1 \in \mathcal{G}$.
- Individuals are well-ordered : $((\sigma_1, \ldots, \sigma_n) \in \mathcal{G}$ and $\sigma_n > 1) \Rightarrow ((\sigma_1, \ldots, \sigma_n - 1) \in \mathcal{G})$.

Now, we introduce the canonical order relation on a discrete tree in order to characterize the relationship between individuals. Let δ and σ be two elements of $\mathcal{U}$, we write $\delta \preceq \sigma$ if δ is an ancestor of σ. That is

$$(\delta \preceq \sigma) \Longleftrightarrow \left(\exists k \in \mathbb{N},\ \sigma^k = \delta\right).$$

This relation defines a partial order on $\mathcal{U}$. We also denote

$$\delta \wedge \sigma = \sup_{\preceq}\left\{\{\eta \in \mathcal{G} \mid \eta \preceq \sigma\} \cap \{\eta \in \mathcal{G} \mid \eta \preceq \delta\}\right\},$$

which is the last common ancestor of δ and σ. Note that this last supremum is well-defined since if η_1 and η_2 are two elements of $\{\eta \in \mathcal{G} \mid \eta \preceq \sigma\}$, then there exist two non-negative integers n_1 and n_2 such that $\eta_1 = \sigma^{n_1}$ and $\eta_2 = \sigma^{n_2}$. Hence $\eta_1 \preceq \eta_2$ if and only if $n_1 \leq n_2$.

Chronological trees

We now describe more accurately the set of admissible trees. We desire to introduce a time structure. In the previous part, we described how the discrete genealogy of a subset $\mathbb{T}$ of $\mathcal{U} \times \mathbb{R}$ must be to make $\mathbb{T}$ admissible as a tree. We, now, describe the "time compatibility conditions".

Time compatibility conditions :

— Individuals are alive for all time between their birthdate and their date of death :

$$\forall \sigma \in \mathcal{G},\ \exists B_\sigma, D_\sigma \in \mathbb{R}_+,\ B_\sigma < D_\sigma \text{ and } (B_\sigma, D_\sigma] = P_{\mathbb{R}}\left(\{t \in \mathbb{R}_+ \mid (\sigma, t) \in \mathbb{T}\}\right).$$

B_σ is its birth date while D_σ is its date of death.

— Individuals are born during the lifetime of their parents : $\forall \sigma \in \mathcal{G}\backslash\{\emptyset\},\ B_\sigma \in (B_{\sigma^1}, D_{\sigma^1})$.

— Individuals are born in the right order :

$$((\sigma_1, \ldots, \sigma_n) \in \mathcal{G} \text{ and } \sigma_n > 1) \Rightarrow \left(B_{(\sigma_1,\ldots,\sigma_n-1)} < B_\sigma\right).$$

— The ancestral individual born at time 0 : $B_\emptyset = 0$.

The set of subsets $\mathbb{T}$ of $\mathcal{U} \times \mathbb{R}_+$ satisfying these compatibility conditions as well as those on the discrete genealogy is called the set of admissible trees and is denoted by $\mathcal{T}$.

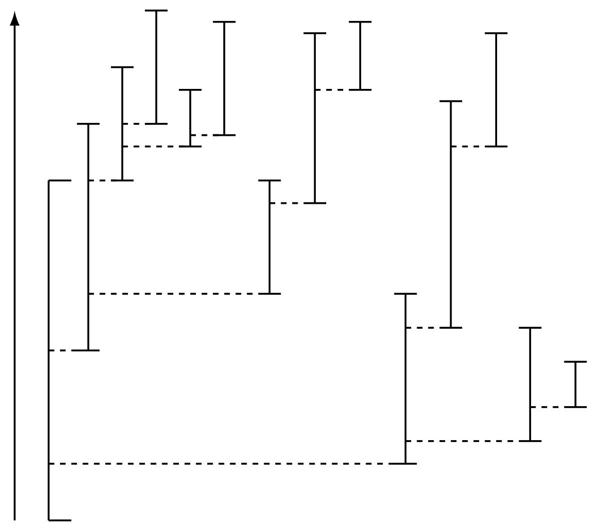

FIGURE 3.1 – Graphical representation of a Splitting tree. The vertical axis represents the biological time and the horizontal axis has no biological meaning. The vertical lines represent the individuals, their lengths correspond to their lifetimes. The dashed lines denote the filiations between individuals.

3.1.1 Chronological trees as measured metric spaces

The purpose of this section is to give more structure to chronological trees. This allows us to define, for instance, Poisson random measures on a tree $\mathbb{T}$.

Topology on $\mathbb{T}$

Here we define a topology on the tree through a metric. The natural distance one may expect between two points of the tree is the length of the "shortest path" between these two points (see Figure 3.2). To do so we need to define the divergence point between points (δ, t) and (σ, s) of $\mathbb{T}$. This point, denoted $(\delta, t) \wedge (\sigma, s)$, is defined by the relations (see Figure 3.1)

$$\begin{cases} P_{\mathcal{U}}\left((\delta, t) \wedge (\sigma, s)\right) = \delta \wedge \sigma, \\ P_{\mathbb{R}}\left((\delta, t) \wedge (\sigma, s)\right) = \inf\left\{B_{\sigma^i},\ B_{\delta^j} \mid i, j \in \mathbb{N},\ \sigma \wedge \delta \prec \sigma^i \text{ and } \sigma \wedge \delta \prec \delta^j\right\}. \end{cases}$$

Now, for two points (δ, t) and (σ, s), we set

$$d\left((\delta, t), (\sigma, s)\right) = t + s - 2P_{\mathbb{R}}\left((\delta, t) \wedge (\sigma, s)\right).$$

The function d defines the desired distance on the tree.

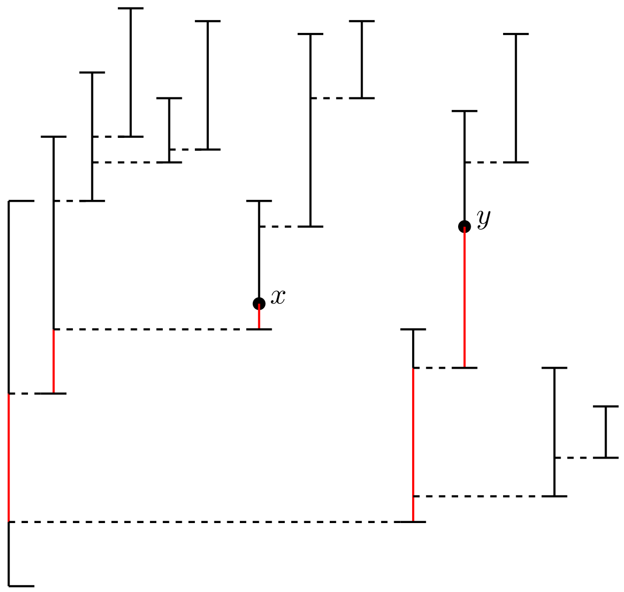

FIGURE 3.2 – Geodesic path connecting x to y

Lebesgue measure on $\mathbb{T}$

Now, we define a Lebesgue measure on $\mathbb{T}$. It is easy to see that, for each individual σ in $P_{\mathcal{U}}(\mathbb{T})$, there is a natural isometry φ_σ from $\{(\delta, t) \in \mathbb{T} \mid \delta = \sigma\}$ to $(B_\sigma, D_\sigma]$. Now, let O be an open set of $\mathbb{T}$, and set

$$\lambda(O) = \sum_{\sigma \in P_{\mathcal{U}}(\mathbb{T})} \lambda(\varphi_\sigma(\mathcal{L}(\sigma) \cap O)),$$

where $\mathcal{L}(\sigma)$ denotes the slice of the tree corresponding to σ, that is

$$\mathcal{L}(\sigma) = \{\sigma\} \times (B_\sigma, D_\sigma].$$

It is easy to see that λ defines a σ-additive functional on the topology of $\mathbb{T}$, which uniquely extend to a measure on $\mathcal{B}(\mathbb{T})$. This defines a Lebesgue measure λ on $\mathbb{T}$.
These two objects (the metric and the measure) are useful to construct the contour process of the tree (see Section 3.2).

3.1.2 The law of a splitting tree

Let b a positive real number and $\mathbb{P}_V$ be a probability measure on $(0, \infty]$. The purpose of this part is to introduce a probability measure $\mathbb{P}_\mathbb{T}$ on $\mathcal{T}$. This measure is called the law of a splitting tree with lifespan measure $b\mathbb{P}_V$. Let us roughly describe this law through the dynamics of the population described by a splitting tree. The population start with a single individual. This individual gives births at exponential rate b. Each child is assumed to have a lifetime distributed according to $\mathbb{P}_V$, independent of its parent or its brotherhood. To end, children give birth according to the same mechanisms (and independently from the other individuals) and so on.

More precisely, let $E : \mathcal{U} \times \mathcal{T} \mapsto \mathcal{T}$ be the application defined by

$$E(i, \mathbb{T}) = \{((\sigma_2, \dots, \sigma_n), t - B_i) \mid ((\sigma_1, \dots, \sigma_n), t) \in \mathbb{T} \text{ and } \sigma_1 = i\}, \quad \forall i \in \mathbb{N}\backslash\{0\}, \quad \forall \mathbb{T} \in \mathcal{T}.$$

This application returns the subtree of $\mathbb{T}$ induced by the ith child of the ancestor individual. Note that if i does not belong to $\mathcal{G}$, $E(i, \mathbb{T})$ equals $\emptyset$.

Now, the law $\mathbb{P}_\mathbb{T}$ of a splitting tree $\mathbb{T}$, with lifetime distribution $\mathbb{P}_V$ and birth rate b, is the unique distribution such that

— $D_\emptyset$ is distributed according to $\mathbb{P}_V$,

— conditionally on $D_\emptyset$, the random measure on $(0, D_\emptyset]$ defined by

$$\sum_{\substack{i \in \mathbb{N} \\ i \in \mathcal{G}}} \delta_{B_i}$$

is a Poisson random measure with rate b,

— for all $i \in \mathbb{N} \cap \mathcal{G}$, the law of $E(i, \mathbb{T})$ is $\mathbb{P}_\mathbb{T}$.

The readers interested in more details should look at [60].

A very important consequence of this last point is that a splitting tree presents a renewal structure. Indeed, any of the subtrees induced by the children of the roots is itself a splitting tree.

3.2 The contour process of a Splitting tree is a Lévy process

A very common method in trees analysis is to transform trees into more convenient objects. In the Galton-Watson case, for instance, one may think of the Harris path or the height process. This allows transforming a tree into a real-valued function which is easier to manipulate. Our purpose in this section is to introduce the same kind of object for spitting trees. This will reveal particularly powerful in the study of the properties of the splitting trees. In particular because the contour process appears to have a nice behaviour. The ideas and results come again from [60].

In order to go further, we need to introduce an exploration process of the tree. To do that, we must choose an order to explore the tree.

3.2.1 The contour process of a finite tree

Let $\mathbb{T}$ be an element of $\mathcal{T}$ which is finite in the sense that $\lambda(\mathbb{T})$ is finite. Assume also that the total number of individuals is finite. This last hypothesis is not a restriction since a splitting tree with finite length must have a finite number of individuals (because lifetimes are i.i.d.).

Exploring chronological trees

We define a total order relation on a chronological tree by setting, for two elements (σ, t) and (δ, s) of $\mathbb{T}$,

$$(\sigma,t) \leq (\delta,s) \Leftrightarrow \begin{cases} \delta \preceq \sigma \text{ and } P_{\mathbb{R}}((\delta,s)\wedge(\sigma,t)) \geq s & (C1) \\ \text{or} & \\ \exists n \in \mathbb{N}^{\star},\ \sigma^n \preceq \delta \text{ and } t > B_{\sigma^n}. & (C2) \end{cases}$$

In a more informal way, the point of birth of the lineage of (δ, s) during the lifetime of the root split the tree in two connected components, then $(\sigma, t) \leq (\delta, s)$ if (σ, t) belongs to the same component as (δ, s) but is not an ancestor of (δ, s) (see Figure 3.3).

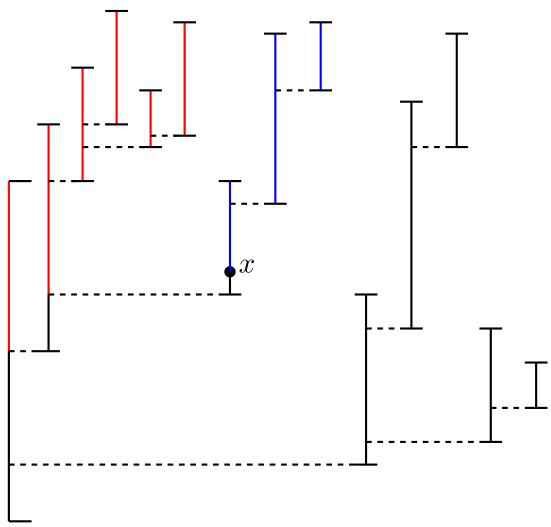

FIGURE 3.3 – In blue and red, the set $\{y \in \mathbb{T} \mid y \leq x\}$. The blue part corresponds to condition (C1) while the red part corresponds to condition (C2).

Now, we have the tools needed to introduce the exploration process. Let φ, be the application defined by

$$\begin{array}{rccc} \varphi: & \mathbb{T} & \to & [0, \lambda(\mathbb{T})), \\ & x & \mapsto & \lambda(\{y \mid\ y \leq x\}). \end{array}$$

The main result is the following.

Proposition 3.2.1 (Lambert, [60]). *φ is an increasing bijection.*

We give a new proof of this result which was hopped to be simpler... but just appears to be different.

Proof. In order to get the result, we prove it for a slight modification of φ. More precisely, in this proof we assume that φ is defined on $\tilde{\mathbb{T}} = \mathbb{T} \cup \{(\emptyset, 0)\}$ as follow :

$$\begin{array}{rccl} \varphi : & \tilde{\mathbb{T}} & \to & [0, \lambda(\mathbb{T})], \\ & x & \mapsto & \begin{cases} \lambda(\{y \mid \ y \leq x\}) & \text{if } y \neq (\emptyset, 0), \\ \lambda(\mathbb{T}) & \text{else.} \end{cases} \end{array}$$

The proof follows these steps :

— φ is strictly increasing (*similar to the proof in [60]*).
— φ is continuous with respect to the order topology on $\tilde{\mathbb{T}}$ induced by $\leq$.
— $\tilde{\mathbb{T}}$ is connected w.r.t. the order topology.
— The range of φ is $[0, \lambda(\mathbb{T})]$.

Let $(\delta, t) < (\sigma, s)$, then there exists $\varepsilon > 0$ such that

$$\{y \in \tilde{\mathbb{T}} \mid \ y \leq (\sigma, s)\} \backslash \{y \in \tilde{\mathbb{T}} \mid \ y \leq (\delta, t)\} \supset B((\sigma, t+\varepsilon), \varepsilon), \quad \text{if } t \neq D_\sigma,$$

and

$$\{y \in \tilde{\mathbb{T}} \mid \ y \leq (\sigma, s)\} \backslash \{y \in \tilde{\mathbb{T}} \mid \ y \leq (\delta, t)\} \supset B((\sigma^1, t+\varepsilon), \varepsilon), \quad \text{if } t = D_\sigma.$$

These imply that φ is strictly increasing (from the definition of λ).

Let us consider from this point (and until the end of this proof) that $\tilde{\mathbb{T}}$ is endowed with the order topology induced by $\leq$. This topology is different from the topology induced by the distance d. We begin by showing that φ is continuous with respect to the order topology (which is trivially not the case with respect to d).

Continuity of φ

Let (σ, t) in $\tilde{\mathbb{T}}$. Assume that (σ, t) is a branching point. That is there exists δ in $\mathcal{G}$ such that δ^1 equals σ and $B_\delta = t$. Let ϵ be a positive real number. Now, consider the segment

$$((\sigma, t-\epsilon), (\delta, B_\delta + \epsilon)) = \{(\delta, s) \mid s \in (B_\delta, B_\delta + \epsilon)\} \cup \{(\sigma, s) \mid s \in (t-\epsilon, t)\}.$$

The last equality holds for ϵ small enough. From this equality, it is easily seen that (for ϵ small enough),

$$\varphi\{((\sigma, t-\epsilon), (\delta, B_\delta + \epsilon))\} = B(\varphi(\sigma, t), 2\epsilon).$$

Since $\{((\sigma, t-\epsilon), (\delta, B_\delta + \epsilon)),\ \epsilon > 0\}$ is a complete neighbourhood system of (σ, t), φ is continuous at the branching points of the tree (we recall that this continuity holds only w.r.t. to the order topology). The other cases (leaf or simple point) are left to the reader.

Connectedness of $\tilde{\mathbb{T}}$

Let A be a subset of $\tilde{\mathbb{T}}$. We want to show that A admits a supremum. Define for all σ in $\mathcal{G}$,

$$M_\sigma = \inf\left(P_{\mathbb{R}}\left(A \cap \{\sigma\} \times (B_\sigma, D_\sigma]\right)\right).$$

Now, according to hypothesis made in the beginning of this section we have that $\mathcal{G}$ is finite. Hence, the set

$$\left\{(\sigma, M_\sigma) \mid \sigma \in P_{\mathcal{U}}(A),\ M_\sigma > B_\sigma\right\} \cup \left\{(\sigma^1, M_\sigma) \mid \sigma \in P_{\mathcal{U}}(A),\ M_\sigma = B_\sigma\right\}$$

is a totally ordered finite set which, hence, has a maximum (σ^*, t^*). We claim that (σ^*, t^*) is a supremum of A. This means that $\tilde{\mathbb{T}}$ is a complete lattice w.r.t. $\leq$.

Moreover, if $(\sigma, t) < (\delta, s)$, it is easily seen (by considering for instance $(\sigma, t - \epsilon)$ or $(\delta, s + \epsilon)$) that there exists a third point x in $\tilde{\mathbb{T}}$ such that : $(\sigma, t) < x < (\delta, s)$ (continuum property). This fact, in conjunction with the fact that $\tilde{\mathbb{T}}$ is a complete lattice, implies that $\tilde{\mathbb{T}}$ is connected w.r.t. the order topology (see [90]).

This is the end

Finally, using that $\varphi((\emptyset, D_\emptyset)) = 0$ and $\varphi((\emptyset, 0)) = \lambda(\mathbb{T})$, we have, since φ is continuous and $\tilde{\mathbb{T}}$ is connected,

$$\varphi(\tilde{\mathbb{T}}) = [0, \lambda(\mathbb{T})].$$

□

Note that the hypothesis that $\mathcal{G}$ is finite is fundamental to make things work. Indeed, as one can see in [60], in the general case the bijection holds only with the local closure of $\mathbb{T}$. A way to adapt this proof to the general case would require to use a compacification of the tree $\mathbb{T}$ (this is more or less what is done in [60]).

The exploration process is now defined as the inverse of φ (see Figure 3.4).

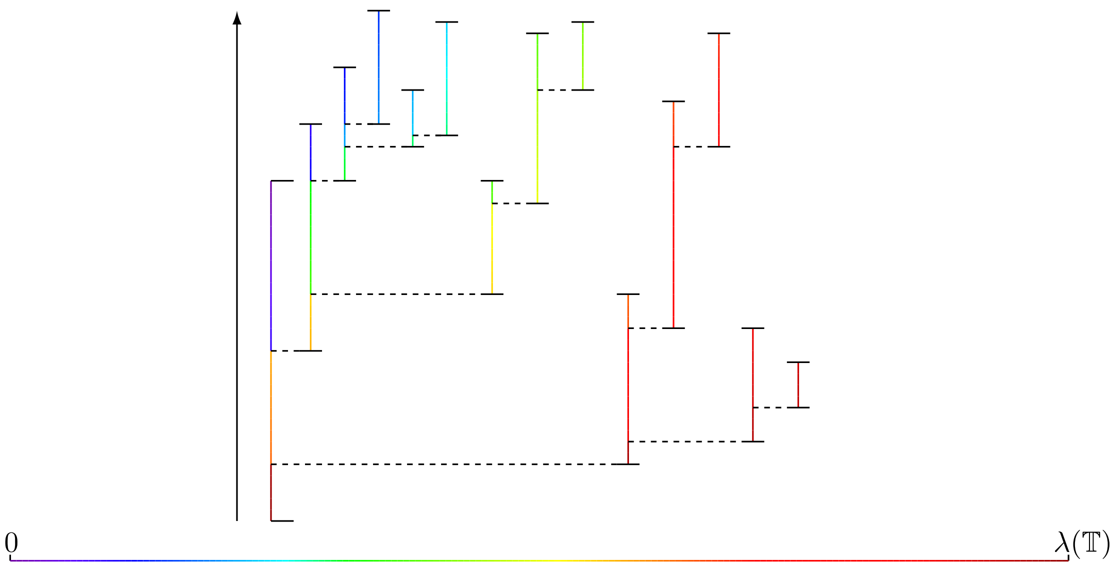

FIGURE 3.4 – A graphical representation of the exploration process. The one-to-one correspondence is represented by corresponding colours.

The contour of a finite tree

We can now define the contour of a finite tree. Informally speaking, the contour process is a real valued process which can be seen as this : it begins at the top of the root and decreases with slope -1 while running back along the life of the root until it meets a birth. The contour process then jumps at the top of the life interval of the child born at this time and continues its exploration as before. If the exploration process does not encounter a birth when exploring the life interval of an individual, it goes back to its parent and continues the exploration from the birth-date of the just left individual (see Figure 3.5).

From the mathematical point of view, the contour process is just the height in the tree (that is the biological time) of the exploration process at time t. More precisely, the contour process Y is defined by

$$Y_s = P_{\mathbb{R}}(\varphi^{-1}(s)), \quad \forall s \in [0, \lambda(\mathbb{T})].$$

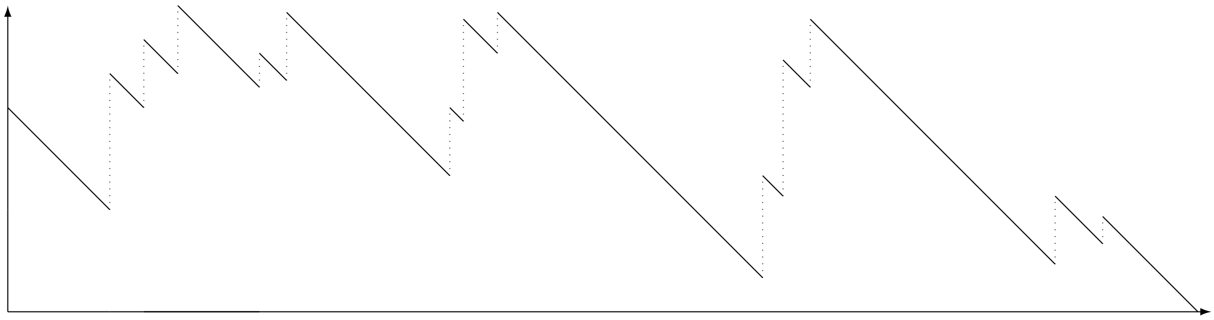

FIGURE 3.5 – The contour process of the finite tree of Figure 3.1.

A useful feature is that the tree $\mathbb{T}$ is in bijection with the graph of the contour process. Indeed, let (s, Y_s) be a point of the graph, then the unique corresponding point in the tree is $\varphi^{-1}(s)$ (see Figure 3.6).

3.2.2 The law of the contour process of a splitting tree

We recall that a splitting tree is a tree describing a population where individuals with (independent) lifetimes distributed according to a distribution $\mathbb{P}_V$ experience birth at rate b. Moreover, the birth processes of different individuals are supposed independents. In this section, we are interested in the law of the contour process of a splitting tree $\mathbb{T}$ (see Section 3.1.2 for a more precise definition).

The first problem is that a splitting tree may not have a finite length $\lambda(\mathbb{T})$. This implies that the above definition does not apply. That is why we define, for all positive real number t, the contour $Y^{(t)}$ of the truncated tree at time t. More precisely, let

$$\mathbb{T}^{(t)} = \mathbb{T} \cap \mathcal{U} \times [0, t],$$

the truncated tree at level t. This means that all parts of the tree above level t are removed. Now, since the number of children of each individual before time t must be finite (because a Poisson random measure with finite rate b is locally finite), then the total length of the tree must be finite. This implies that the contour process $Y^{(t)}$ associated to $\mathbb{T}^{(t)}$ is well-defined. Now, the main result of this chapter is the following.

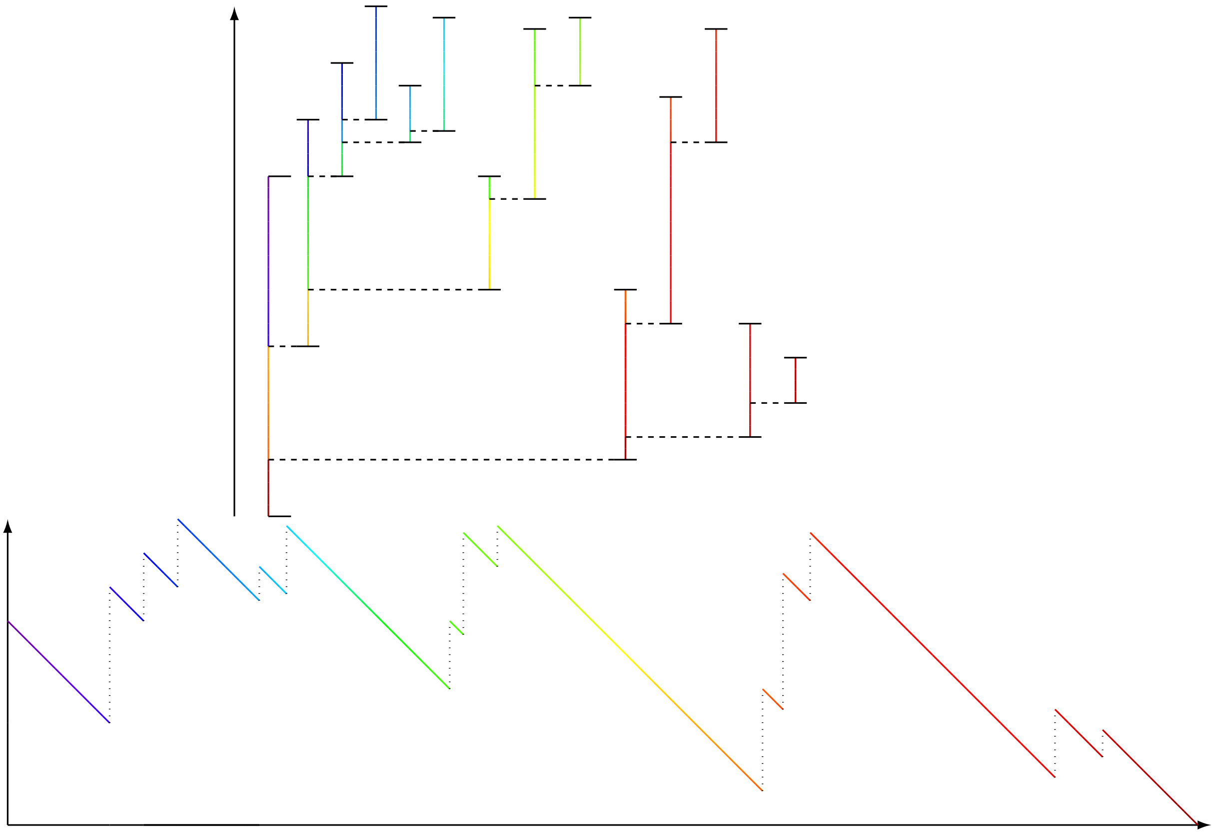

FIGURE 3.6 – One-to-one correspondence between the tree and the graph of the contour represented by corresponding colours.

Theorem 3.2.2 (Lambert [60]). *Let $(X^i)_{i\geq 1}$ be a sequence of i.i.d. Lévy processes with Laplace exponent*

$$\psi(x) = x - \int_{(0,\infty]} \left(1 - e^{-rx}\right) b\mathbb{P}_V(dr), \quad x \in \mathbb{R}_+, \tag{3.1}$$

such that $X_0^1 = t \wedge V$ almost surely and $X_0^i = t$ almost surely, for all $i > 1$. Set

$$\tau_t^i = \inf\{s > 0 \mid X_t^i > t\},$$

and

$$S_i = \sum_{j=1}^{i} \tau_t^i.$$

Then, the process X defined by

$$X_t = \sum_{i\geq 1} X_{t-S_{i-1}} \mathbb{1}_{S_{i-1}\leq t<S_i}, \quad \forall t \in \mathbb{R}_+,$$

killed at its first hitting of 0 has the same distribution as the process $Y^{(t)}$.

We say that $Y^{(t)}$ is a spectrally positive Lévy process started from $V \wedge t$, reflected below t and killed at its first hitting of 0. Moreover, its Laplace exponent is given by (3.1).

3.3 The population counting process

In this section, we introduce the population counting process which is the subject of Chapter 5. We first give its definition. Then, we show how to use the contour process in order to derive its first properties. As in the previous section, the results come from [60].

Definition 3.3.1 (Binary homogenous CMJ process)**.** *Let* $(N_t,\ t \in \mathbb{R}_+)$ *the process defined by*

$$N_t = \sharp \left\{ \mathbb{T} \cap \mathcal{U} \times \{t\} \right\}, \quad \forall t > 0.$$

This process is known as binary homogenous Crump-Mode-Jagers process.

The unidimensional marginal of $(N_t,\ t \in \mathbb{R}_+)$

From Section 3.2.1, we easily see that

$$N_t = \sharp\{Y_s^{(t)} = t \mid s \in \mathbb{R}_+\}.$$

Finally,This remark allows to get, thanks to the theory of Lévy processes, a first information on the process $(N_t,\ t \in \mathbb{R}_+)$. Indeed, let τ_t (resp. τ_0) be the hitting time of t (resp. of 0) by the contour process $Y^{(t)}$. Now, for any positive integer k, the strong Markov property entails that

$$\begin{aligned}\mathbb{P}\left(N_t = k \mid N_t > 0\right) = \mathbb{E}\left\{\mathbb{P}_{t \wedge V}\left(\sharp\{Y_s^{(t)} = t \mid s \in \mathbb{R}_+\} = k \mid \tau_t < \tau_0\right)\right\} \\ = \mathbb{P}_t\left(\sharp\{Y_s^{(t)} = t \mid s > 0\} = k - 1\right).\end{aligned}$$

Once again, the strong Markov property gives

$$\begin{aligned}\mathbb{P}_t\left(\sharp\{Y_s^{(t)} = t \mid s > 0\} = k-1\right) &= \mathbb{P}_t\left(\tau_t < \tau_0\right)\mathbb{P}_t\left(\sharp\{Y_s^{(t)} = t \mid s > 0\} = k-2\right) \\ &= \mathbb{P}_t\left(\tau_t < \tau_0\right)^{k-1}\mathbb{P}_t\left(\sharp\{Y_s^{(t)} = t \mid s > 0\} = 0\right) \\ &= \mathbb{P}_t\left(\tau_t < \tau_0\right)^{k-1}\mathbb{P}_t\left(\tau_0 < \tau_t\right).\end{aligned}$$

Now, by Theorem 2.6.1 (see also Theorem 8.1 in [59]), we have that

$$\mathbb{P}_t\left(\tau_t < \tau_0\right) = 1 - \frac{1}{W(t)},$$

where W is the scale function of the Lévy process whose Laplace exponent is given by (3.1). We recall that W is characterized by its Laplace transform,

$$T_{\mathcal{L}}W(t) = \int_{(0,\infty)} e^{-rt}W(r)dr = \frac{1}{\psi(t)}, \quad \forall t > \alpha, \tag{3.2}$$

where α is the largest root of ψ.

$$\mathbb{P}\left(N_t = k \mid N_t > 0\right) = \frac{1}{W(t)}\left(1 - \frac{1}{W(t)}\right)^{k-1}, \quad \forall k \in \mathbb{N}\backslash\{0\}. \tag{3.3}$$

According to this, N_t is geometrically distributed (conditionally on non-extinction) with parameter $W(t)$. In particular

$$\mathbb{E}\left[N_t \mid N_t > 0\right] = W(t). \tag{3.4}$$

Hence, it would be worth to know the asymptotic behaviour of W in large time in order to get some hints on the behaviour of $(N_t,\ t \in \mathbb{R}_+)$. To this goal, one can use Tauberian theorems for Laplace transform (see [59], Section 7.6). The results are the following

Lemma 3.3.2. *Lambert, [60]*
- *if* $\psi'(0+) > 0$, *then* $W(t) \sim \frac{1}{\psi'(0+)}$,
- *if* $\psi'(0+) = 0$, *then* $W(t) \sim \frac{2t}{\psi''(0+)}$,
- *if* $\psi'(0+) < 0$, *then* $W(t) \sim \frac{e^{\alpha t}}{\psi'(\alpha)}$.

According to Champagnat-Lambert [13], one can even go further and get

Lemma 3.3.3. *Champagnat-Lambert, [13, Thm. 3.21] If* $\psi'(0+) < 0$, *then, there exist a positive constant* γ *such that,*

$$e^{-\alpha t}\psi'(\alpha)W(t) - 1 = \mathcal{O}\left(e^{-\gamma t}\right).$$

In Proposition 4.1.1, we characterize the $\mathcal{O}\left(e^{-\gamma t}\right)$ term of this Lemma.

In the sequel we refer to the supercritical case (resp. critical, subcritical) when $\psi'(0) < 0$ (resp. $\psi'(0) = 0$, $\psi'(0) > 0$). Remark that, differentiating ψ in (3.1), this is equivalent to have $b\mathbb{E}[V] > 1$ (resp. $b\mathbb{E}[V] = 1$, $b\mathbb{E}[V] < 1$).

Extinction

Let t be a positive real number, we have

$$\mathbb{P}\left(N_t = 0\right) = \mathbb{E}\left\{\mathbb{P}_{t\wedge V}\left(\tau_t > \tau_0\right)\right\},$$

which is equal, according to Theorem 2.6.1, to

$$\mathbb{E}\left[\frac{W(t - t\wedge V)}{W(t)}\right] = \int_{[0,t]} \frac{W(t-v)}{W(t)}\,\mathbb{P}_V(dv).$$

Hence,

$$\mathbb{P}\left(N_t > 0\right) = 1 - \frac{W \star \mathbb{P}_V(t)}{W(t)}, \tag{3.5}$$

and

$$\mathbb{E}N_t = W(t) - W \star \mathbb{P}_V(t), \tag{3.6}$$

Moreover, it is easily seen that

$$\mathbb{P}\left(\text{Extinction}\right) = \lim_{t\to\infty} \mathbb{P}\left(N_t = 0\right).$$

Using this with Lemma 3.3.3, one can get in the critical and subcritical cases,

$$\mathbb{P}\left(\text{Extinction}\right) = 1.$$

Similarly, using again Lemma 3.3.3, we have, in the supercritical case,

$$\lim_{t\to\infty} \int_{[0,t]} \frac{W(t-v)}{W(t)}\,\mathbb{P}_V(dv) = \int_{\mathbb{R}_+} e^{-\alpha t}\mathbb{P}_V(dv).$$

But according to (3.1), one has

$$\int_{\mathbb{R}_+} e^{-\lambda v}\mathbb{P}_V(dv) = 1 + \frac{\psi(\lambda) - \lambda}{b}. \tag{3.7}$$

Finally,

$$\mathbb{P}\,(\text{Extinction}) = 1 - \frac{\alpha}{b}, \tag{3.8}$$

and

$$\mathbb{P}\,(\text{NonEx}) = \frac{\alpha}{b}. \tag{3.9}$$

Using the convexity of ψ, one can easily see that $\alpha > 0$ if and only if $\psi'(0) < 0$. In short, the only case where the population does not almost surely extinct is the supercritical one.

3.4 Backward model : coalescent point process

The purpose of this section is to analyse the genealogical model associated to splitting trees. Some previous works (for instance [13, 14, 15]) show that some properties of a splitting tree are easier to study using the tree describing the genealogical relation between the lineages of the individuals alive at a time t. This is true in particular when one wants to study the genotype of individuals in the population (if we add mutations to the model). Indeed, the difference between two individuals in terms of genotype should depend only on the time past since their lineages has diverged. Hence, this particular genealogical tree, known as *coalescent point processes* (CPP), contains the essential informations to study, for instance, the allelic partition. In order to derive the law of that genealogical tree, we need to characterize the joint law of the *times of coalescence* between pairs of individuals in the population, which are the times since their lineages have split.

In the sequel, let $\mathcal{I}_t = \{\mathbb{T} \cap \mathcal{U} \times \{t\}\}$ denotes the set of individuals alive at fixed time t. This set is naturally ordered through the total order on $\mathbb{T}$. We may refer to the ith individual in this order as the ith individual alive at time t (provided $i \leq N_t$). This individual is denoted $\mathcal{I}_t(i)$.

Before defining the divergence time between the lineages of two individuals, let us define what a lineage is.

Definition 3.4.1 (Lineage)**.** *The lineage of an individual alive at time t, or equivalently of a point (σ, t) in $\mathbb{T}$, is defined by the set (see Figure 3.7)*

$$\mathcal{L}in((\sigma, t)) = \left(\cup_{n \geq 1} \left\{(\sigma^n, s) \mid s \leq B_{\sigma^{n-1}}\right\}\right) \cup \left\{(\sigma, s) \mid s \leq t\right\}. \tag{3.10}$$

We refer the reader to Section 3.1 for the definition of σ^n. The set $\mathcal{L}in((\sigma, t))$ corresponds to the set of successive (in the intuitive sense) points linking (σ, t) to $(\emptyset, 0)$(see Figure 3.7). The time of coalescence $\mathcal{C}_{i,j}$ between individuals i and j is the amount of time spent since their lineages have diverged (see Figure 3.8). It can be defined by

$$C_{i,j} = t - \sup P_{\mathbb{R}} \left\{\mathcal{L}in((\mathcal{I}_t(i), t)) \cap \mathcal{L}in((\mathcal{I}_t(j), t))\right\}.$$

But one can show, for two individuals i and j (such that $i \leq j$), that (see also Figure 3.9)

$$C_{i,j} = \sup \left\{C_{k,k+1} \mid k \in]\!]i, j]\!]\right\}.$$

Hence, all the coalescence times are characterized by the coalescence times of adjacent individuals. In the sequel, $C_{i,i+1}$ is denoted H_i. The so-called *coalescent point process* (CPP) is defined as the sequence $(H_i)_{0 \leq i \leq N_t - 1}$. The CPP of the population is indeed described by this sequence, saying that a lineage coalesces with the first deeper branch on its left (see Figure 3.9).

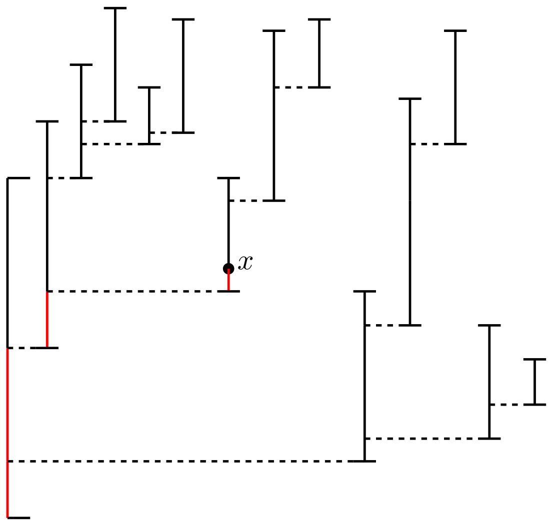

FIGURE 3.7 – The lineage of point x.

3.4.1 The law of the CPP

In order to derive the law of the sequence $(H_i)_{0\leq i\leq N_t-1}$, we use once again the contour process of the splitting tree. The first step is to reword the coalescence times in terms of the contour. Its appears that the coalescence time H_i between two adjacent individuals i and $i+1$ is equal in distribution to the *depth of the excursion of the contour below t between the visit of these two individuals.*

Let (σ, t) and (δ, t) be, respectively, the ith and $(i+1)$th individuals alive at time t (we assume that they exist). Now, the branching point $(\sigma, t) \wedge (\delta, t)$ between the lineages of these two individuals can be obtain as

$$\sup \left\{ \mathcal{L}in((\sigma, t)) \backslash \mathcal{L}in((\delta, t)) \right\}.$$

We are interested in the points explored by the exploration process between (σ, t) and (δ, t). These points are given by

$$\begin{aligned} &\varphi^{-1}\left([\varphi((\sigma,t)),\varphi((\delta,t))]\right) \\ &\qquad = \mathcal{L}in((\sigma, t)) \backslash \mathcal{L}in((\delta, t)) \biguplus \left\{ x \in \mathbb{T} \mid \sup \left\{ \mathcal{L}in((\sigma, t)) \backslash \mathcal{L}in((\delta, t)) \right\} < x \leq (\delta, t) \right\}, \end{aligned}$$

where φ denotes the exploration process defined in Subsection 3.2.1 and $\uplus$ denotes the union of disjoint sets.

Now let x in $\mathbb{T}$ such that

$$\sup \left\{ \mathcal{L}in((\sigma, t)) \backslash \mathcal{L}in((\delta, t)) \right\} < x < (\delta, t).$$

Hence, we must have $\delta \wedge \sigma \preceq P_{\mathcal{U}}(x)$ and

$$P_{\mathbb{R}}(x) > P_{\mathbb{R}}\left(\sup \left\{ \mathcal{L}in((\sigma, t)) \backslash \mathcal{L}in((\delta, t)) \right\}\right).$$

Otherwise, we would have $x \geq (\delta, t)$ (see the definition of the order relation).

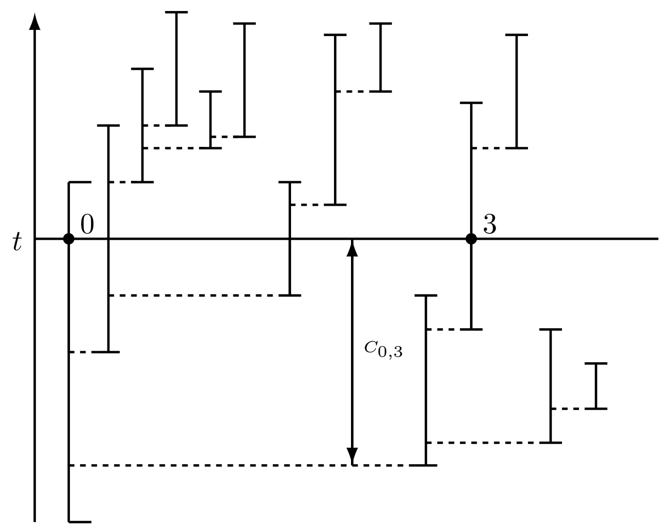

FIGURE 3.8 – Coalescence time between individuals 0 and 3 represented as an arrow.

This implies that

$$P_{\mathbb{R}}\left(\sup\left\{\mathcal{L}in((\sigma,t))\backslash\mathcal{L}in((\delta,t))\right\}\right) \leq P_R(\varphi^{-1}(s)), \quad \forall s \in [\varphi((\sigma,t)), \varphi((\delta,t))].$$

But the right hand side of the last inequality is the contour process. Finally, we get that the diverging time of the lineages of (σ, t) and (σ, t) is given by

$$\min\left\{Y_s^{(t)} \mid s \in [\varphi((\sigma,t)), \varphi((\delta,t))]\right\}.$$

This also implies that the coalescence time between those two individuals is the depth of the excursion of the contour process on the time interval $[\varphi((\sigma,t)), \varphi((\delta,t))]$.

Now, let H be the depth of an excursion below t of a Lévy process with Laplace exponent given by (3.1). It is easily seen that

$$\mathbb{P}(H > s) = \mathbb{P}_t\left(\tau_s^- < \tau_t^+\right), \quad \forall s \in \mathbb{R}_+,$$

where τ_s^- and τ_t^+ were defined in the beginning of Section 2.6.2. But Theorem 2.6.1 (see also [59, 7]) gives

$$\mathbb{P}_t\left(\tau_s^- < \tau_t^+\right) = \frac{1}{W(s)},$$

where W is the scale function of our Lévy process. Finally, we have

Proposition 3.4.2. *Let $(X_i)_{i\geq 1}$ be an i.i.d. family of random variables with law given by*

$$\mathbb{P}(X_1 > s) = \frac{1}{W(s)}, \quad \forall s \geq 0.$$

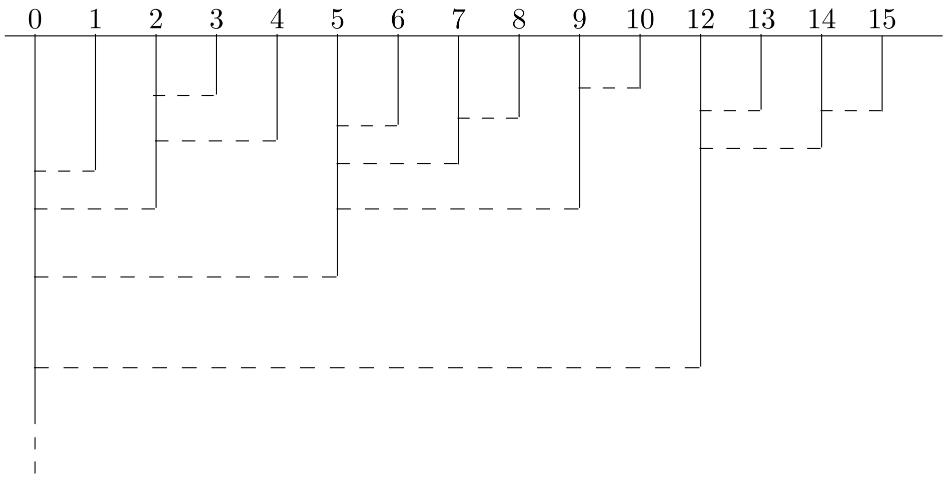

FIGURE 3.9 – A coalescent point process for 16 individuals, hence 15 branches.

Let

$$N = \inf\{i \geq 1 \mid X_i > t\}.$$

Finally, set X_0 *equals* t*. Then,*

$$(H_i)_{0 \leq i \leq N_t - 1} \stackrel{d}{=} (X_i)_{0 \leq i \leq N-1} \,.$$

Chapitre 4

On some auxiliary results

The purpose of this chapter is to state and prove three preliminary results which are crucial for the two forthcoming chapters. Since they are not only related to splitting trees and that they have their own mathematical interest, we decided to dedicate a chapter to these results.

Section 4.1 gives precise asymptotic estimates on the scale function of the contour of a splitting tree. Some weaker results were already stated in [60, 13]. This original result can be found in [40]. Section 4.2 is devoted to the proof of an extension of the Campbell formula concerning the expectation of the integral of a random process with respect to a random measure when both objects present some local independence properties. These results can also be seen as extension of the well known compensation formula for Poisson functional (see 2.1.4). These two formulas enable us to use a very elegant formalism to model the frequency spectrum of a splitting with neutral Poissonian mutations. This also allows us to obtain formulas for moments for the frequency spectrum in Chapter 6. The results of Section 4.2 can be found in [12]. Finally, Section 4.3 is devoted to an alternative construction of the CPP which plays an important role in the computation of the moments of the frequency spectrum (see Chapter 6).

4.1 Asymptotic behavior of the scale function of the contour process

Before stating and proving the result of this section, we make some reminders from Chapter 2 adapted in the context of our particular Lévy process. First, we recall that the law of a spectrally positive Lévy process $(Y_t,\ t \in \mathbb{R}_+)$ is uniquely characterized by its Laplace exponent ψ,

$$\psi_Y(\lambda) = \log \mathbb{E}\left[e^{-\lambda Y_1}\right],\ \lambda \in \mathbb{R}_+,$$

which in our case take the form of (3.1) :

$$\psi_Y(\lambda) = x - \int_{(0,\infty]} \left(1 - e^{-rx}\right) b\mathbb{P}_V(dr), \quad \lambda \in \mathbb{R}_+.$$

We also assume that $\psi'(0+) < 0$, so that the greatest zero of ψ is positive. Let α be this zero. This corresponds to the supercritical case for the splitting tree. In this section, we suppose that $Y_0 = 0$.

For a such Lévy process, the local time at the reflected process (see Chapter 2) $(L_t,\ t \in \mathbb{R})$ can be chosen as

$$L_t = \sum_{i=0}^{n_t} e^i, \quad t \in \mathbb{R}_+,$$

where $\left(e^i\right)_{i\geq 0}$ is a family of i.i.d. exponential random variables with parameter 1, and

$$n_t := \mathrm{Card}\{0 < s \leq t \mid Y_s = \sup_{u \leq s} Y_u\},$$

is the number of times Y reaches its running maximum up to time t. We recall that the ascending ladder process associated to Y is defined as

$$H_t = Y_{L_t^{-1}}, \quad t \in \mathbb{R}_+,$$

where $\left(L_t^{-1}, t \in \mathbb{R}_+\right)$ is the right-inverse of L. It is easily seen that H is a subordinator whose values are the successive new maxima of Y.
Conversely, in our case, the process $(\inf_{s\leq t} Y_s,\ t \in \mathbb{R}_+)$ can be chosen as descending ladder time process $\left(\hat{L}_t,\ t \in \mathbb{R}_+\right)$. The descending ladder process $\hat{H}$ is then defined from $\hat{L}$ as H was defined from L.
The Wiener-Hopf factorization, given in Theorem 2.5.1, allows us to connect the characteristic exponent ψ_Y of Y with the characteristic exponents of the bivariate Lévy processes $((L_t, H_t),\ t \in \mathbb{R}_+)$ and $((\hat{L}_t, \hat{H}_t),\ t \in \mathbb{R}_+)$, respectively denoted by κ and $\hat{\kappa}$. In our particular case, where Y is spectrally negative, we have

$$\begin{cases} \kappa(\gamma, \beta) = \frac{\gamma - \psi_Y(\beta)}{\phi_Y(\gamma) - \beta}, & \gamma, \beta \in \mathbb{R}_+, \\ \hat{\kappa}(\gamma, \beta) = \phi_Y(\gamma) + \beta, & \gamma, \beta \in \mathbb{R}_+, \end{cases}$$

where ϕ_Y is the right-inverse of ψ_Y. Taking $\gamma = 0$ allows us to recover the Laplace exponent ψ_H of H from which we obtain the relation,

$$\psi_Y(\lambda) = (\lambda - \phi_Y(0))\, \psi_H(\lambda). \tag{4.1}$$

We have now all the notation to state and prove the main result of this section.

Proposition 4.1.1 (Behavior of W)**.** *In the supercritical case ($\alpha > 0$), there exists a positive non-increasing càdlàg function F such that*

$$W(t) = \frac{e^{\alpha t}}{\psi'(\alpha)} - e^{\alpha t} F(t), \quad t \geq 0,$$

and

$$\lim_{t\to\infty} e^{\alpha t} F(t) = \begin{cases} \frac{1}{b\mathbb{E}V - 1} & \text{if } \mathbb{E}V < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $Y^\sharp$ be a spectrally negative Lévy process with Laplace exponent given by

$$\psi^\sharp(\lambda) = \lambda - \int_{\mathbb{R}_+} \left(1 - e^{-\lambda x}\right) e^{-\alpha x} b\, \mathbb{P}_V(dx).$$

It is known that $Y^\sharp$ has the law of the contour process of the supercritical splitting tree with lifespan measure $\mathbb{P}_V$ conditioned to extinction (see [60]). In this case the largest root of $\psi^\sharp$ is zero, meaning that the process $Y^\sharp$ does not go to infinity and that $\phi_{Y^\sharp}(0) = 0$. Elementary manipulations on Laplace transform show that the scale function $W^\sharp$ of $Y^\sharp$ is related to W by

$$W^\sharp(t) = e^{-\alpha t} W(t), \quad t \in \mathbb{R}_+.$$

Let $H^\sharp$ be the ascending ladder subordinator associated to the Lévy process $Y^\sharp$. In the case where $\phi_{Y^\sharp}(0) = 0$, and in this case only, the scale function $W^\sharp$ can be rewritten as (see [59] or use Laplace transform),

$$W^\sharp(t) = \int_0^\infty \mathbb{P}\left(H_x^\sharp \le t\right) dx. \tag{4.2}$$

In other words, if we denote by U the potential measure of $H^\sharp$,

$$W^\sharp(t) = U[0,t].$$

Now, it is easily seen from (4.1) that the Laplace exponent $\psi_{H^\sharp}$ of $H^\sharp$ takes the form,

$$\psi_{H^\sharp}(\lambda) = \psi'(\alpha) - \int_{[0,\infty]} \left(1 - e^{-\lambda r}\right) \Upsilon(dr),$$

where

$$\Upsilon(dr) = \int_{(r,\infty)} e^{-\alpha v} b \mathbb{P}_V(dv) dr = \mathbb{E}\left[e^{-\alpha V} \mathbb{1}_{V>r}\right] b dr.$$

Moreover,

$$\Upsilon(\mathbb{R}_+) = 1 - \psi'(\alpha),$$

which means that $H^\sharp$ is a compound Poisson process with jump rate $1-\psi'(\alpha)$, jump distribution $J(dr) := \frac{\mathbb{E}\left[e^{-\alpha V}\mathbb{1}_{V>r}\right]}{1-\psi'(\alpha)} dr$, and killed at rate $\psi'(\alpha)$. It is well known (or elementary by conditioning on the number of jumps at time x), that the law $\mathbb{P}_{H_x^\sharp}$ of $H_x^\sharp$ $(x \in \mathbb{R}_+)$ is given

$$\mathbb{P}_{H_x^\sharp}(dt) = e^{-\psi'(\alpha)x} \sum_{k\ge 0} e^{-(1-\psi'(\alpha))x} \frac{\left((1-\psi'(\alpha))\, x\right)^k}{k!} J^{\star k}(dt).$$

Some calculations now lead to,

$$U(dx) = \sum_{k \ge 0} \Upsilon^{\star k}(dx).$$

From this point, since Υ is a sub-probability, $U(x) := U[0,x]$ satisfies the following defective renewal equation,

$$U(x) = \int_{\mathbb{R}_+} U(x-u)\Upsilon(du) + \mathbb{1}_{\mathbb{R}_+}(x).$$

Finally, since

$$\int_{\mathbb{R}_+} e^{\alpha x} \Upsilon(dx) = 1,$$

and since, from Lemma 2.7.1,

$$t \to U(t,\infty),$$

is clearly a directly Riemann integrable function as a positive decreasing integrable function. Hence, as suggested in Remark 2.7.3,

$$e^{\alpha x}\left(U(\mathbb{R}_+) - U(x)\right) \underset{x\to\infty}{\longrightarrow} \frac{1}{\alpha\mu},$$

with

$$\mu = \int_{\mathbb{R}_+} r e^{\alpha r}\Upsilon(dr) = \frac{1}{\alpha}\left(b\mathbb{E}V - 1\right),$$

if V is integrable. In the case where V is not integrable, the limit is 0.
To end the proof, note using relation (4.2) and the fact that $H^\sharp$ is killed at rate $\psi'(\alpha)$ that,

$$W^\sharp(t) = \frac{1}{\psi'(\alpha)} - U(t,\infty).$$

□

4.2 A formula to compute the expectation of an integral with respect to a random measure

In this section, we use notation and vocabulary from [18].
Let $\mathcal{X}$ a be Polish space. We recall that a random measure is a measurable mapping from a probability space to the space $\mathcal{M}_b\left(\mathcal{X}\right)$ of all boundedly finite measures on $\mathcal{X}$, i.e. such that each bounded set has finite mass.
The purpose of this section is to prove an extension of the Campbell formula (see Proposition 13.1.IV in [18]), giving the expectation of an integral with respect to a random measure when the integrand has specific "local" independence properties w.r.t. to the measure.
For this purpose, we need to introduce the notion of Palm measure related to a random measure $\mathcal{N}$. So let $\mathcal{N}$ be a random measure on $\mathcal{X}$ with intensity measure μ. Let also $(X_x,\ x\in\mathcal{X})$ be a continuous random process with value in $\mathbb{R}_+$. Since this section is devoted to prove relations concerning only the distributions of $\mathcal{N}$ and X, we can assume without loss of generality that our random elements X and $\mathcal{N}$ are defined (in the canonical way) on the space

$$\mathcal{C}\left(\mathcal{X}\right)\times\mathcal{M}_b\left(\mathcal{X}\right),$$

where $\mathcal{C}(\mathcal{X})$ denotes the space of continuous function on $\mathcal{X}$. This space is Polish as a product of Polish spaces. We denote by $\mathcal{F}$ the corresponding product Borel σ-field.
For the random measure $\mathcal{N}$, the corresponding Campbell measure $\mathcal{C}_{\mathcal{N}}$ is the measure defined on $\sigma\left(\mathcal{F}\times\mathcal{B}\left(\mathcal{X}\right)\right)$ by extension of the following relation on the semi-ring $\mathcal{F}\times\mathcal{B}\left(\mathcal{X}\right)$,

$$\mathcal{C}_{\mathcal{N}}\left(F\times B\right) = \mathbb{E}\left[\mathbb{1}_F\mathcal{N}\left(B\right)\right],\quad F\in\mathcal{F},\quad B\in\mathcal{B}\left(\mathcal{X}\right).$$

It is straightforward to see that $\mathcal{C}_{\mathcal{N}}$ is σ-finite and for each F in $\mathcal{F}$ the measure $\mathcal{C}_{\mathcal{N}}\left(F\times\cdot\right)$ is absolutely continuous with respect to μ. Then, from Radon-Nikodym's theorem, for each $F\in\mathcal{F}$, there exist $y\in\mathcal{X}\mapsto P_y\left(F\right)$ in $L^1\left(\mu\right)$ such that,

$$\mathcal{C}_{\mathcal{N}}\left(F\times B\right) = \int_B P_y\left(F\right)\ \mu\left(dy\right),$$

uniquely defined up to its values on μ-null sets.
Since our probability space is Polish, P can be chosen to be a probabilistic kernel, i.e. for all F in $\mathcal{F}$,

$$y \in \mathcal{X} \mapsto P_y(F) \text{ is mesurable,}$$

and for all y in $\mathcal{X}$,

$$F \in \mathcal{F} \mapsto P_y(F) \text{ is a probability measure.}$$

The probability measure P_y is called the Palm measure of $\mathcal{N}$ at point y. Since X is continuous, it is $\mathcal{B}(\mathcal{X}) \otimes \mathcal{F}$ measurable, and it is easily deduced from this point that

$$\mathbb{E} \int_{\mathcal{X}} X_x \, \mathcal{N}(dx) = \int_{\mathcal{X}} \mathbb{E}_{P_x}[X_x] \, \mu(dx), \tag{4.3}$$

where $\mathbb{E}_{P_x}$ denotes the expectation w.r.t. P_x. Formula (4.3) is the so-called Campbell formula.
We can now state, the main results of this section which are the aforementioned extensions of the above formula.

Theorem 4.2.1. *Let X be a continuous process from $\mathcal{X}$ to $\mathbb{R}_+$. Let $\mathcal{N}$ be a random measure on $\mathcal{X}$ with finite intensity measure μ. Assume that X is locally independent from $\mathcal{N}$, that is, for all $x \in \mathcal{X}$, there exists a neighbourhood V_x of x such that X_x is independent from $\mathcal{N}(V_x \cap \cdot)$. Suppose moreover that there exists an integrable random variable Y such that*

$$|X_x| \le Y, \ \forall x \in \mathcal{X}, \ a.s.$$

and

$$\mathbb{E}[Y\mathcal{N}(\mathcal{X})] < \infty.$$

Then we have

$$\mathbb{E} \int_{\mathcal{X}} X_x \, \mathcal{N}(dx) = \int_{\mathcal{X}} \mathbb{E}[X_x] \, \mu(dx). \tag{4.4}$$

However, the continuity condition of the preceding theorem is too restrictive for our purposes. We need a more specific result.

Theorem 4.2.2. *Let X be a process from $[0,T] \times \mathcal{X}$ to $\mathbb{R}_+$ such that $X_{\cdot,x}$ is càdlàg for all x and $X_{s,\cdot}$ is continuous for all s. Let $\mathcal{N}$ be a random measure on $[0,T] \times \mathcal{X}$ with finite intensity measure μ. Assume that, for each s in $[0,T]$, the family $(X_{s,x}, x \in \mathcal{X})$ is independent from the restriction of $\mathcal{N}$ on $[0,s]$, that there exists an integrable random variable Y such that*

$$|X_{s,x}| \le Y, \ \forall x \in \mathcal{X}, \ \forall s \in [0,t], \ a.s.$$

and that

$$\mathbb{E}[Y\mathcal{N}(\mathcal{X})] < \infty.$$

Then we have

$$\mathbb{E} \int_{[0,T]\times\mathcal{X}} X_{s,x} \, \mathcal{N}(ds,dx) = \int_{[0,T]\times\mathcal{X}} \mathbb{E}[X_{s,x}] \, \mu(ds,dx). \tag{4.5}$$

Let $[\![1,n]\!]$ denotes the set $\mathbb{N}\cap[1,n]$. Before going further, we recall that a dissecting system is a sequence $\{A_{n,j}, j\in[\![1,K_n]\!]\}_{n\geq 0}$ of nested partitions of $\mathcal{X}$, where $(K_n)_{n\geq 0}$ is an increasing sequence of integers, such that

$$\lim_{n\to\infty}\max_{j\in[\![1,K_n]\!]}\text{diam } A_{n,j}=0.$$

In the spirit of the works of Kallenberg on the approximation of simple point processes, the proof of Theorems 4.2.1 is based on the following Theorem which can be find in [53] or in [73] (Section WIII.9).

Theorem 4.2.3 (Kallenberg [53])**.** *Let μ and ν be two finite measures on the Polish space $\mathcal{X}$, such that μ is absolutely continuous with respect to ν. Let f be the Radon-Nikodym derivative of μ w.r.t. ν. Then, for any dissecting system $\{A_{n,j}, j\in[\![1,K_n]\!]\}_{n\geq 0}$ of $\mathcal{X}$, we have*

$$\lim_{n\to\infty}\sum_{j=1}^{K_n}\frac{\mu\left(A_{n,j}\right)}{\nu\left(A_{n,j}\right)}\mathbb{1}_{s\in A_{n,j}}=f(s),\quad \textit{for } \mu\textit{-almost all } s\in\mathcal{X}.$$

We can now prove our results.

Proof of Theorem 4.2.1. Let $\{A_{n,j}, j\in[\![1,K_n]\!]\}_{n\geq 0}$ be a dissecting system of $\mathcal{X}$. We denote by $A_n(x)$ the element of the partition $(A_{n,j})_{1\leq j\leq K_n}$ which contain x. Let also T be a denumerable dense subset of $\mathcal{X}$. We use lower and upper approximations of X. More precisely, let for all positive integer k and for all a un $\mathcal{X}$,

$$\begin{aligned}\underline{X}_x^{(k)} &:= \inf\left\{X_s|s\in T\cap A_k(x)\right\}=\sum_{j=1}^{K_k}\underline{\chi}_j^{(k)}\mathbb{1}_{x\in A_{j,k}},\\ \overline{X}_x^{(k)} &:= \sup\left\{X_s|s\in T\cap A_k(x)\right\}=\sum_{j=1}^{K_k}\overline{\chi}_j^{(k)}\mathbb{1}_{x\in A_{j,k}},\end{aligned}$$

with

$$\overline{\chi}_j^{(k)}=\sup\left\{X_s|s\in A_{j,k}\cap T\right\}\text{ and }\underline{\chi}_j^{(k)}=\inf\left\{X_s|s\in A_{j,k}\cap T\right\}.$$

Note that the supremum and infinimum are taken on $T\cap A_k(a)$ to ensure that $\underline{X}_j^{(k)}$ and $\bar{X}_j^{(k)}$ are measurable, but the set T could be removed by continuity of X. We remark that, for any j, k, the measure

$$\mathbb{E}\left[\overline{\chi}_j^{(k)}\mathcal{N}\left(\bullet\right)\right]$$

is absolutely continuous with respect to μ and it follows from Campbell's formula (4.3) that the Radon-Nikodym derivative is

$$\mathbb{E}_{P_x}\left[\overline{\chi}_j^{(k)}\right].$$

Thus, it follows from Theorem 4.2.3 that, μ-a.e.,

$$\mathbb{E}_{P_x}\left[\overline{\chi}_j^{(k)}\right]=\lim_{n\to\infty}\frac{\mathbb{E}\left[\overline{\chi}_j^{(k)}\mathcal{N}\left(A_n(x)\right)\right]}{\mu\left(A_n(x)\right)}.$$

Then, since $\underline{X}^{(k)}$ and $\overline{X}^{(k)}$ are finite sums of such random variables,

$$\mathbb{E}_{P_x}\left[\overline{X}_x^{(k)}\right] = \lim_{n\to\infty} \frac{\mathbb{E}\left[\overline{X}_x^{(k)}\mathcal{N}\left(A_n(x)\right)\right]}{\mu\left(A_n(x)\right)},$$

and

$$\mathbb{E}_{P_x}\left[\underline{X}_x^{(k)}\right] = \lim_{n\to\infty} \frac{\mathbb{E}\left[\underline{X}_x^{(k)}\mathcal{N}\left(A_n(x)\right)\right]}{\mu\left(A_n(x)\right)},$$

outside a μ-null set which can be chosen independent of k by countability. Now, since

$$\underline{X}_x^{(k)} \leq X_x \leq \overline{X}_x^{(k)},$$

it follows that

$$\mathbb{E}_{P_x}\left[\underline{X}_x^{(k)}\right] \leq \liminf_{n\to\infty} \frac{\mathbb{E}\left[X_x\mathcal{N}(A_n(x))\right]}{\mathbb{E}\left[\mathcal{N}(A_n(x))\right]} \leq \limsup_{n\to\infty} \frac{\mathbb{E}\left[X_x\mathcal{N}(A_n(x))\right]}{\mathbb{E}\left[\mathcal{N}(A_n(x))\right]} \leq \mathbb{E}_{P_x}\left[\overline{X}_x^{(k)}\right], \ \mu - a.e..$$

Now, since X is continuous,

$$\overline{X}_x^{(k)} \underset{k\to\infty}{\longrightarrow} X_x \qquad \text{and} \qquad \underline{X}_x^{(k)} \underset{k\to\infty}{\longrightarrow} X_x,$$

it follows, from Lebesgue's Theorem, that

$$\mathbb{E}_{P_x}\left[X_x\right] = \lim_{n\to\infty} \frac{\mathbb{E}\left[X_x\mathcal{N}(A_n(x))\right]}{\mathbb{E}\left[\mathcal{N}(A_n(x))\right]}, \quad \mu - a.e..$$

Now, since $A_{n,j}$ is a dissecting system, there exists an integer N such that, for all $n > N$, $A_n(x) \subset V_x$. That is, for n large enough,

$$\frac{\mathbb{E}\left[X_x\mathcal{N}(A_n(x))\right]}{\mathbb{E}\left[\mathcal{N}(A_n(x))\right]} = \mathbb{E}X_x.$$

Finally,

$$\mathbb{E}_{P_x}\left[X_x\right] = \mathbb{E}\left[X_x\right], \quad \mu - a.e..$$

And the conclusion comes from (4.3). □

Proof of Theorem 4.2.2. Clearly, we may assume without loss of generality that $T = 1$. Define, for all integer M,

$$X_{s,x}^M = \sum_{k=0}^{M-1} X_{\frac{k+1}{M},x}\mathbb{1}_{s\in\left[\frac{k}{M},\frac{k+1}{M}\right)}.$$

Since $X_{.,x}$ is càdlàg, this sequence of processes converges pointwise to $(X_{s,x},\ s \in [0,1])$ for all ω. Then, by Lebesgue's theorem,

$$\begin{aligned}\mathbb{E}\left[\int_{[0,1]\times\mathcal{X}} X_{s,x}\ \mathcal{N}(ds,dx)\right] &= \int_{[0,1]\times\mathcal{X}} \mathbb{E}_{P_{s,x}}\left[X_{s,x}\right]\ \mu(ds,dx), \\ &= \lim_{M\to\infty}\sum_{k=0}^{M-1}\int_{[0,1]} \mathbb{1}_{s\in\left[\frac{k}{M},\frac{k+1}{M}\right)\times\mathcal{X}}\ \mathbb{E}_{P_{s,x}}\left[X_{\frac{k+1}{M},x}\right]\mu(ds,dx).\end{aligned}$$

Clearly, for fixed k, $(s,x) \mapsto X_{\frac{k+1}{M},x}$ is continuous on $[\frac{k}{M}, \frac{k+1}{M}] \times \mathcal{X}$. Hence, Theorem 4.2.1 can be applied to

$$\begin{array}{ccc} \left[\frac{k}{M}, \frac{k+1}{M}\right] \times \mathcal{X} & \to & \mathbb{R}_+, \\ (s,x) & \mapsto & X_{\frac{k+1}{M},x}, \end{array}$$

to conclude the proof.

□

4.3 A recursive construction of the CPP

The purpose of this section is to given an alternative construction of the CPP. This construction comes from the joint work with N. Champagnat [12]. We recall that a CPP at time t can be seen as sequence $(H_i)_{0\leq i\leq N_t-1}$ where $(H_i)_{i\geq 1}$ is an i.i.d. sequence of random variables with distribution given by

$$\mathbb{P}(H_i > s) = \frac{1}{W(s)},$$

stopped at its first value N_t greater than t, and H_0 equals to t. We also recall that W is the scale function of the contour process of a splitting tree (see Section 3.2).
The motivation of the construction given above comes from the fact that if a mutation (in a model with mutation) occurs on an individuals at some time, the future of the family carrying this mutation does not depend on the whole tree but only on the subtree induced by this individual. This fact can be equivalently studied through the CPP rather than in the tree directly. Here, we consider the CPP at some time t and we introduce a construction of this CPP which underlines this independence. Suppose we are given a sequence $\left(\mathcal{P}^{(i)}\right)_{i\geq 1}$ of coalescent point processes stopped at time a with scale function W. Then, take an independent CPP $\hat{\mathcal{P}}$, where the law of the branches corresponds to the excess over a of a branch with scale function W conditioned to be higher than a. As stated in the next proposition, the tree build from the grafting of the $\mathcal{P}^{(i)}$ above each branch of $\hat{\mathcal{P}}$ is also a CPP with scale function W stopped at time t (see Figure 4.1).

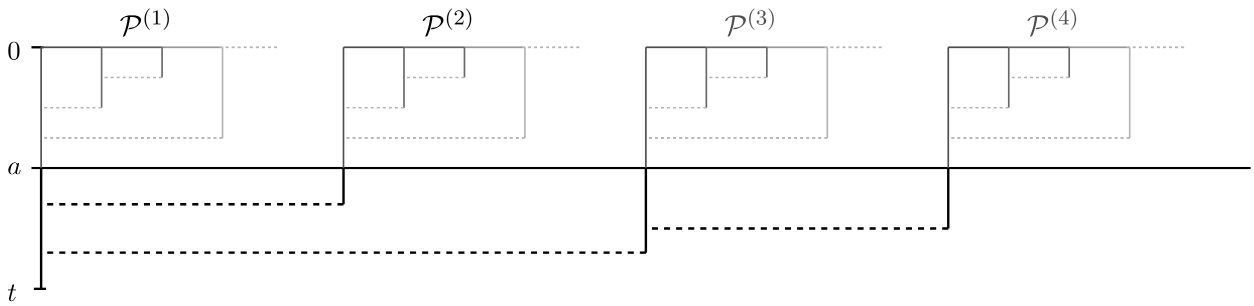

Figure 4.1 – Grafting of trees.

Proposition 4.3.1. *Let $\left(\mathcal{P}^{(i)}\right)_{i\geq 1}$ be an i.i.d. sequence of coalescent point processes with scale function W at time a, and let $\left(N_a^i\right)_{i\geq 1}$ be their respective population sizes. Let $\hat{\mathcal{P}}$ be a coalescent* *point process, independent of the previous family, with scale function*

$$\hat{W}(t) := \frac{W(t+a)}{W(a)},$$

at time $t-a$*, and let* $\hat{N}_{t-a}$ *denotes its population size. Let* $S_0 := 0$ *and*

$$S_i := \sum_{j=1}^{i} N_a^j, \quad \forall i \geq 1.$$

Then the random vector $\left(H_k,\ 0 \leq k \leq S_{\hat{N}_a - 1}\right)$ *defined, for all* $k \geq 0$*, by*

$$H_k = \begin{cases} \mathcal{P}^{(i+1)}_{k-S_i} & \text{if } S_i < k < S_{i+1}, \quad \text{for some } i \geq 0, \\ \hat{\mathcal{P}}_i + a & \text{if } k = S_i, \quad \text{for some } i \geq 0, \end{cases}$$

is a CPP with scale function W *at time* t*.*

Proof. Note that $H_0 = \hat{\mathcal{P}}_0 + a$. To prove the result, it is enough to show that the sequence $(H_k)_{k\geq 1}$ is an i.i.d. sequence with the same law as H, given by

$$\mathbb{P}(H > s) = \frac{1}{W(s)}, \quad \forall s > 0.$$

The independence follows from the construction. We details the computation for the joint law of (H_l, H_k) and leave the easy extension to the general case to the reader. Let $k > l$ be two positive integers, and let also s_1, s_2 be two positive real numbers. We denote by $\mathcal{S}$ the random set $\{S_i,\ i \geq 1\}$. Hence,

$$\begin{aligned} \mathbb{P}(H_k < s_1,\ H_l < s_2) =& \mathbb{P}(H < s_1 \mid H < a)\, \mathbb{P}(H < s_2 \mid H < a)\, \mathbb{P}(l \notin \mathcal{S},\ k \notin \mathcal{S}) \\ &+ \mathbb{P}\left(a + \hat{H} < s_1\right) \mathbb{P}(H < s_2 \mid H < a)\, \mathbb{P}(l \notin \mathcal{S},\ k \in \mathcal{S}) \\ &+ \mathbb{P}(H < s_1 \mid H < a)\, \mathbb{P}\left(a + \hat{H} < s_2\right) \mathbb{P}(l \in \mathcal{S},\ k \notin \mathcal{S}) \\ &+ \mathbb{P}\left(a + \hat{H} < s_1\right) \mathbb{P}\left(a + \hat{H} < s_2\right) \mathbb{P}(l \in \mathcal{S},\ k \in \mathcal{S}), \end{aligned}$$

where $\hat{H}$ denotes a random variable with the law of the branches of $\hat{\mathcal{P}}$, i.e. such that

$$\mathbb{P}\left(\hat{H} > s\right) = \frac{W(s)}{W(s+a)}, \quad \forall s > 0.$$

Now, since the random variables S_i are sums of geometric random variables, we get

$$\begin{aligned} \mathbb{P}(H_k < s_1,\ H_l < s_2) = & \left(p\mathbb{P}(H < s_1 \mid H < a) + (1-p)\mathbb{P}\left(a + \hat{H} < s_1\right)\right) \\ & \times \left(p\mathbb{P}(H < s_2 \mid H < a) + (1-p)\mathbb{P}\left(a + \hat{H} < s_2\right)\right), \end{aligned}$$

with $p = \mathbb{P}(k \in \mathcal{S})$. Moreover we have,

$$\begin{aligned}\mathbb{P}(H_k \leq s) =& \sum_{i\geq 1} \Big\{ \mathbb{P}(H_k \leq s \mid k \in]\!]S_{i-1}, S_i[\![) \, \mathbb{P}(k \in]\!]S_{i-1}, S_i[\![) \\ &+ \mathbb{P}(H_k \leq s \mid k = S_i) \, \mathbb{P}(k = S_i) \Big\} \\ =& \mathbb{P}(H \leq s \mid H < a) \, \mathbb{P}\left(\bigcup_{i\geq 1} \{k \in]\!]S_{i-1}, S_i[\![\} \right) \\ &+ \mathbb{P}(H \leq s \mid H > a) \, \mathbb{P}\left(\bigcup_{i\geq 1} \{k = S_i\} \right).\end{aligned}$$

Since the S_i's are sums of geometric random variables of parameters $\hat{W}(t-a)^{-1}$, they follow binomial negative distributions with parameters i and $\hat{W}(t-a)^{-1}$. Hence, since

$$\mathbb{P}(S_i = k) = \begin{cases} 0, & \text{if } k < i, \\ \binom{i-1}{k-1} \hat{W}(t-a)^{-i} \left(1 - \hat{W}(t-a)^{-1}\right)^{k-i}, & \text{else,} \end{cases}$$

some elementary calculus leads to

$$\mathbb{P}\left(\bigcup_{i\geq 1} \{k = S_i\} \right) = \mathbb{P}(H > a), \quad \forall k \in \mathbb{N}.$$

which ends the proof. $\square$

Chapitre 5

On the population counting process (a.k.a. binary homogeneous CMJ processes)

5.1 Introduction

In this chapter, we consider the population counting process N_t (giving the number of living individuals at time t) of a splitting tree. This process is a binary homogeneous Crump-Mode-Jagers (CMJ) process. Crump-Mode-Jagers processes are very general branching processes. Such processes are known to have many applications. For instance, in biology, they have recently been used to model spreading diseases (see [76, 6]) or for questions in population genetics ([13, 14]). Another example of application appears in queuing theory (see [65] and [38]).
In the supercritical case, it is known that the quantity $e^{-\alpha t}N_t$, where α is the Malthusian parameter of the population, converges almost surely. This result has been proved in [82] using Jagers-Nerman's theory of general branching processes counted by random characteristics. One of our goals in this chapter is to give a new proof of this result using only elementary probabilistic tools and relying on fluctuation analysis of the process. This proof comes from a joint work with Nicolas Champagnat on the frequency spectrum of a splitting tree [12] (see also Chapter 6).
The other goal of this chapter is to investigate the behaviour of the error in the aforementioned convergence. This study comes from the preprint [40]. Many papers studied the second order behaviour of converging branching processes. Early works investigate the Galton-Watson case. In [41] and [42], Heyde obtains rates of convergence and gets central limit theorems in the case of supercritical Galton-Watson when the limit has finite variance. Later, in [3], Asmussen obtained the polynomial convergence rates in the general case. In our model, the particular case when the individuals never die (i.e. $\mathbb{P}_V = \delta_\infty$, implying that the population counting process is a Markovian Yule process) has already been studied. More precisely, Athreya showed in [5], for a Markovian branching process Z with appropriate conditions, and such that $e^{-\alpha t}Z_t$ converges to some random variable W a.s., that the error

$$\frac{Z_t - e^{\alpha t}W}{\sqrt{Z_t}},$$

converges in distribution to some Gaussian random variable.
In the case of general CMJ processes, there was no similar result except a very recent work of Iksanov and Meiners [43] giving sufficient conditions for the error terms in the convergence of supercritical general branching processes to be $o(t^{\delta})$ in a very general background (arbitrary birth point process). Although our model is more specific, we give more precise results. Indeed, we give an exact rate of convergence, $e^{\frac{\alpha}{2}t}$, and characterize the limit. Moreover, we believe that our method can also apply to other general branching processes counted by random characteristics, as soon as the birth point process is Poissonian.

Section 5.2 is devoted to the statement of the law of large numbers for N_t and the associated central limit theorem. Section 5.3 is devoted to the new proof of the law of large numbers. Finally, the central limit theorem is proved in Section 5.4.

5.2 Statement of the limit theorems

We recall that we consider a general branching population where individuals live and reproduce independently. The lifetimes of the individuals are i.i.d. random variables distributed as a random variable V with law $\mathbb{P}_V$. Moreover, individuals give birth at a Poissonian rate b. We refer the reader to Chapter 3 for the details about the model. Let us also recall that the Laplace distribution with zero mean and variance σ^2 is the probability distribution whose characteristic function is given by

$$\lambda \in \mathbb{R} \mapsto \frac{1}{1+\frac{1}{2}\sigma^2\lambda^2}.$$

It particular, it has a density given by

$$x \in \mathbb{R} \mapsto \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}.$$

We denote this law by $\mathcal{L}\left(0,\sigma^2\right)$. We also recall that, if G is a Gaussian random variable with zero mean and variance σ^2 and $\mathcal{E}$ is an exponential random variable with parameter 1 independent of G, then $\sqrt{\mathcal{E}}G$ is Laplace $\mathcal{L}\left(0,\sigma^2\right)$.

In the sequel of this chapter (as well as in Chapter 6) we denote by $\mathbb{P}_t$ the probability measure $\mathbb{P}(\cdot \mid N_t > 0)$. Similarly, we denotes by $\mathbb{P}_\infty$ the measure $\mathbb{P}(\cdot \mid \text{ Non-extinction})$.

We can now state the main results of the chapter. First, let us recall the law of large number for N_t.

Theorem 5.2.1. *In the supercritical case, that is $b\mathbb{E}[V] > 1$, there exists a random variable $\mathcal{E}$, such that*

$$e^{-\alpha t} N_t \underset{t\to\infty}{\to} \frac{\mathcal{E}}{\psi'(\alpha)}, \quad \textit{a.s. and in } L^2.$$

In addition, under $\mathbb{P}_\infty$, $\mathcal{E}$ is exponentially distributed with parameter one.

Section 5.3 is devoted to a new proof of this theorem.

In this chapter, we also prove the following theorem on the second order properties of the above convergence.

Theorem 5.2.2. *In the supercritical case, we have, under $\mathbb{P}_\infty$,*

$$e^{-\frac{\alpha}{2}t}\left(\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}\right) \xrightarrow[t\to\infty]{(d)} \mathcal{L}\left(0, 2-\psi'(\alpha)\right).$$

The proof of this theorem is the subject of Section 5.4. Note that, according to (3.2), we have

$$\psi'(x) = 1 - \int_{\mathbb{R}_+} xe^{-xv}\, b\mathbb{P}_V(dv), \quad \forall x \in \mathbb{R}_+, \tag{5.1}$$

which implies that

$$2 - \psi'(\alpha) = 1 + \int_{\mathbb{R}_+} ve^{-\alpha v}\, b\mathbb{P}_V(dv) > 0.$$

Note that one can also see using (3.2) and the fact that we are in the supercritical case that

$$\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}_V(dv) = 1 - \frac{\alpha}{b}. \tag{5.2}$$

5.3 An alternative proof of the law of large numbers

The purpose of this section is to show the law of large numbers for N_t. We recall once again that we are in the supercritical case ($\alpha > 0$). This last hypothesis implies that $W(t) \sim \frac{e^{\alpha t}}{\psi'(\alpha)}$. The goal of this section is to prove the almost sure convergence of the population counting process. We first show that the convergence holds in probability, using the convergence of the process which counts at time t the number N_t^∞ of individuals having infinite descent. More formally, recalling that a splitting tree is a subset of $\left(\cup_{k\geq 0}\mathbb{N}^k\right) \times \mathbb{R}_+$ (see Section 3.1), an individual (u,t) in the tree $\mathbb{T}$ is said to have infinite descent at time t if for any $T > t$ there exist $\tilde{u}$ in $\bigcup_{n\geq 0}\mathbb{N}^n$ such that $(u\tilde{u}, T)$ belong to $\mathbb{T}$.
Finally, to obtain the almost sure convergence, we show in Theorem 5.2.1 that N_t can not fluctuate faster than a Yule process.

Proposition 5.3.1. *Let $(N_t^\infty,\ t \in \mathbb{R}_+)$ be the number of alive individuals at time t having infinite descent. Then, under $\mathbb{P}_\infty$, N^∞ is a Yule process with parameter α.*

Proof. Let $T, t \in \mathbb{R}_+$. Let, for $T < t$, $N_t^{(T)}$ be the number of individuals at time t who have alive children at time T. We extend this notation to $t > T$ by setting $N_t^{(T)} = 0$ in this case. Fix S a positive real number, we consider the quantity,

$$\sup_{t\leq S}\left|N_t^{(T)} - N_t^\infty\right|.$$

There exists a (random) finite time T^S such that $N_S^{(T^S)} = N_S^\infty$. This means that the progeny of all the individuals alive at time S who have finite descent are extinct at time T^S. Moreover, $N_t^{(T^S)} = N_t^\infty$ for all $t < S$, since, otherwise, there would exist an individual at time t who has alive descent at time T^S but which does not have an infinite descent.
Hence, for all $T > T^S$, $\sup_{t\leq S}\left|N_t^{(T)} - N_t^\infty\right| = 0$. In particular, as $T \to \infty$, $N^{(T)}$ converges to N^∞ a.s. for the Skorokhod topology of $\mathbb{D}\left[0,\infty\right)$ and N^∞ is a.s. càdlàg.

Now, it remains to derive from $N^{(T)}$ the law of the process N^{∞}. A first remark is that $N_t^{(T)}$ is the number of alive individuals in the upper CPP in the construction given in Chapter 4, Section 4.3. Hence, applying Proposition 4.3.1, with $a = T-t$, gives that $N_t^{(T)}$ is the number of individuals in the CPP $\hat{\mathcal{P}}$ (according to the notation of Proposition 4.3.1). Hence, it is geometrically distributed with parameter $\frac{W(T)}{W(T-t)}$.

Now, we recursively apply this property on a sequence $0 < s_1 < s_2 < \cdots < s_n < T$. By a recursive use of Proposition 4.3.1, we see that, under $\mathbb{P}_T$, the process $\left(N_{s_l}^{(T)},\ 1 \le l \le n\right)$ is a time inhomogeneous Markov chain with geometric initial distribution with parameter

$$\mathbb{P}_t\left(H > T \mid H > T - s_1\right),$$

and the law of $N_{s_l}^{(T)}$ given $N_{s_{l-1}}^{(T)}$ is the law of a sum of $N_{s_{l-1}}^{(T)}$ i.i.d. geometric random variable with parameter

$$p_l = \mathbb{P}\left(H > T - s_{l-1} \mid H > T - s_l\right),$$

i.e. a binomial negative with parameters $N_{s_{l-1}}^{(T)}$ and $1 - p_l$. Hence,

$$\mathbb{P}_t\left(N_{s_1}^{(T)} = m_1, \ldots, N_{s_n}^{(T)} = m_n\right) = p_1\left(1 - p_1\right)^{m_1 - 1} \prod_{i=2}^{n} \binom{m_i + m_{i-1} - 1}{m_i} p_i^{m_{i-1}} \left(1 - p_i\right)^{m_i - 1}.$$

Moreover, we have, by Lemma 3.3.3,

$$p_1 = \frac{W(T - s_1)}{W(T)} \underset{t\to\infty}{\longrightarrow} e^{-\alpha s_1},$$

and

$$p_l = \frac{W(T - s_l)}{W(T - s_{l-1})} \underset{t\to\infty}{\longrightarrow} e^{-\alpha(s_l - s_{l-1})}.$$

This leads to,

$$\mathbb{P}_t\left(N_{s_1}^{(T)} = m_1, \ldots, N_{s_n}^{(T)} = m_n\right)$$
$$\underset{t\to\infty}{\longrightarrow} e^{-\alpha s_1}\left(1 - e^{-\alpha s_1}\right)^{m_1 - 1} \prod_{i=2}^{n} \binom{m_i + m_{i-1} - 1}{m_i} e^{-\alpha m_{i-1}(s_l - s_{l-1})} \left(1 - e^{-\alpha(s_l - s_{l-1})}\right)^{m_i - 1}.$$

Since the right hand side term corresponds to the finite dimensional distribution of a Yule process with parameter α, this concludes the proof. $\square$

Because N^{∞} is a Yule process, $e^{-\alpha t} N_t^{\infty}$ converges a.s. (under $\mathbb{P}_{\infty}$) to an exponential random variable of parameter 1, denoted $\mathcal{E}$ hereafter, when t goes to infinity (see for instance [4]).

Remark 5.3.2. *Let N be a integer valued random variable. In the sequel we say that a random vector with random size $(X_i)_{1 \le i \le N}$ form an i.i.d. family of random variables independent of N, if and only if*

$$\left(X_1, \ldots, X_N\right) \overset{d}{=} \left(\tilde{X}_1, \ldots, \tilde{X}_N\right),$$

where $\left(\tilde{X}_i\right)_{i \ge 1}$ is a sequence of i.i.d. random variables distributed as X_1 independent of N.

We are now able to prove the law of large numbers for N_t.

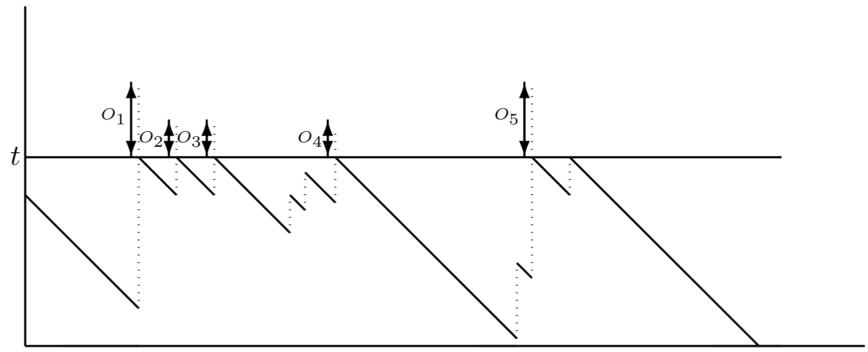

FIGURE 5.1 – Reflected (below t) contour process with overshoot over t.

Proof of Theorem 5.2.1. We first look at the quantity,

$$\mathbb{E}_t\left[e^{-2\alpha t}\left(N_t^\infty - \psi'(\alpha)N_t\right)^2\right].$$

First note that N_t^∞ can always be written as a sum of Bernoulli trials,

$$N_t^\infty = \sum_{i=1}^{N_t} B_i^{(t)}, \tag{5.3}$$

corresponding to the fact that the ith individual has infinite descent or not.
Now, by construction of the splitting tree, the descent of each individual alive at time t can be seen as a (sub-)splitting tree where the lifetime of the root follows a particular distribution (that is the law of the residual lifetime of the corresponding individual). We denote by O_i the residual lifetime of the ith individual which correspond to the ith overshoot of the contour process above t (see Figure 5.1). In particular, these subtrees are dependent only through the residual lifetimes $(O_i)_{1\leq i\leq N_t}$ of the individuals. Hence, the random variables $\left(B_i^{(t)}\right)_{i\geq 2}$ are independent conditionally on the family $(O_i)_{1\leq i\leq N_t}$. In addition, the family $(O_i)_{1\leq i\leq N_t}$ has independence properties under $\mathbb{P}_t$. This is the subject of the following lemma which is proved at the end of this section.

Lemma 5.3.3. *Under $\mathbb{P}_t$, the family $(O_i,\ i\in [\![1,N_t]\!])$ forms a family of independent random variables, independent of N_t, and, except O_1, having the same distribution.*

The proof of this lemma is postponed at the end of this section. Hence, it follows that, under $\mathbb{P}_t$, the random variables $(B_i^{(t)})_{1\geq i\geq N_t}$ are independent and identically distributed for $i\geq 2$ (in the sense of Remark 5.3.2). Let us denote by $\hat{p}_t$ the parameter of $B_1^{(t)}$, and by p_t the common parameter of the others i.i.d. Bernoulli random variables. It follows from (5.3) that

$$\mathbb{E}_t\left[N_t^\infty\right] = p_t\left(W(t)-1\right) + \hat{p}_t$$

and from the Yule nature of N^∞ under $\mathbb{P}_\infty$ (Proposition 5.3.1) that $\mathbb{E}_\infty\left[N_t^\infty\right] = e^{\alpha t}$.

Now, since

$$\mathbb{E}_\infty\left[N_t^\infty\right] = \mathbb{E}_t\left[N_t^\infty\right]\frac{\mathbb{P}\left(N_t > 0\right)}{\mathbb{P}\left(\text{Non-ex}\right)},$$

we have

$$e^{\alpha t} = \left(p_t\left(W(t)-1\right)+\hat{p}_t\right)\frac{\mathbb{P}\left(N_t > 0\right)}{\mathbb{P}\left(\text{Non-ex}\right)}.$$

We recall from Section 3.3 (see also [60]) that,

$$\mathbb{P}\left(\text{Non-ex}\right) = \mathbb{E}\left[e^{-\alpha V}\right],$$

and

$$\mathbb{P}\left(N_t > 0\right) = \mathbb{E}\left[\frac{W(t-V)}{W(t)}\right],$$

where V is a random variable with law $\mathbb{P}_V$ (i.e. the lifetime of a typical individual). It then follows, from Lesbegue's Theorem that,

$$\frac{\mathbb{P}\left(N_t > 0\right)}{\mathbb{P}\left(\text{Non-ex}\right)} - 1 = \mathcal{O}\left(e^{-\beta t}\right), \tag{5.4}$$

with $\beta = \alpha \wedge \gamma$ where the constant γ is given by Lemma 3.3.3. Hence,

$$p_t e^{-\alpha t} W(t) = 1 + \mathcal{O}\left(e^{-\beta t}\right). \tag{5.5}$$

Now, using (5.3), we have

$$\mathbb{E}_t\left[N_t^\infty N_t\right] = \mathbb{E}_t\left[N_t\left(N_t-1\right)\right]p_t + \hat{p}_t\mathbb{E}_t N_t = 2W(t)^2 p_t + \mathcal{O}\left(e^{\alpha t}\right), \tag{5.6}$$

where the second equality comes from the fact that N_t is geometrically distributed with parameter $W(t)^{-1}$ under $\mathbb{P}_t$.

Recalling also that N_t^∞ is geometrically distributed with parameter $e^{-\alpha t}$ under $\mathbb{P}_\infty$, it follows that

$$\mathbb{E}_t\left[\left(N_t^\infty - \psi'(\alpha)N_t\right)^2\right] = 2e^{2\alpha t}\frac{\mathbb{P}\left(\text{Non-ex}\right)}{\mathbb{P}\left(N_t > 0\right)} - 4\psi'(\alpha)W(t)^2 p_t + 2\psi'(\alpha)^2 W(t)^2 + \mathcal{O}\left(e^{\alpha t}\right).$$

Hence, it follows from (5.5), (5.6), (5.4) and Lemma 3.3.3, that

$$\mathbb{E}_t\left[e^{-2\alpha t}\left(N_t^\infty - \psi'(\alpha)N_t\right)^2\right] = \mathcal{O}\left(e^{-\beta t}\right). \tag{5.7}$$

Let us define now, for all integer n, $t_n = \frac{2}{\beta}\log n$. Then, by the previous estimation, it follows from Borel-Cantelli lemma and a Markov-type inequality that,

$$\lim_{n\to\infty} e^{-\alpha t_n} N_{t_n} = \psi'(\alpha)\mathcal{E}, \quad a.s., \tag{5.8}$$

on the survival event.

From this point, we need to control the fluctuation of N between the times $(t_n)_{n\geq 1}$. The births can be controlled by comparisons with a Yule process, but the deaths are harder to control. For this, we use that, by (5.8), $e^{-\alpha t_{n+1}} N_{t_{n+1}} - e^{-\alpha t_n} N_{t_n}$ is small, for n large. It then follows that if the quantity

$$\inf_{s\in[t_n,t_{n+1}]} e^{-\alpha t_n} N_{t_n} - e^{-\alpha s} N_s,$$

takes very low negative values, then

$$\sup_{s\in[t_n,t_{n+1}]} e^{-\alpha s} N_s - e^{-\alpha t_{n+1}} N_{t_{n+1}},$$

must take very high positive value. More precisely,

$$\begin{aligned}
\mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} \left|e^{-\alpha t_n} N_{t_n} - e^{-\alpha s} N_s\right| > \epsilon\right) &\leq \mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} e^{-\alpha s} N_s - e^{-\alpha t_n} N_{t_n} > \epsilon\right) \\
&\quad + \mathbb{P}_{t_n}\left(e^{-\alpha t_n} N_{t_n} - e^{-\alpha t_{n+1}} N_{t_{n+1}} + \sup_{s\in[t_n,t_{n+1}]} e^{-\alpha t_{n+1}} N_{t_{n+1}} - e^{-\alpha s} N_s > \epsilon\right) \\
&\leq \mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} e^{-\alpha s} N_s - e^{-\alpha t_n} N_{t_n} > \epsilon\right) + \mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} e^{-\alpha t_{n+1}} N_{t_{n+1}} - e^{-\alpha s} N_s > \epsilon\right) \\
&\quad + \mathbb{P}_{t_n}\left(e^{-\alpha t_n} N_{t_n} - e^{-\alpha t_{n+1}} N_{t_{n+1}} > \epsilon\right)
\end{aligned}$$

Now, there exists a Yule process Y with parameter b such that $Y_0 = N_{t_n}$ and for all s in $[0, t_{n+1} - t_n]$,

$$N_{t_n} - N_s \leq Y_{s-t_n} - Y_0, \quad a.s. \tag{5.9}$$

This Yule process can be constructed from the population at time t_n by extending the lifetimes of all individuals to infinity, and constructing births from the same Poisson process as in the splitting tree. This leads to

$$\begin{aligned}
\mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} \left|e^{-\alpha t_n} N_{t_n} - e^{-\alpha s} N_s\right| > \epsilon\right) &\leq \mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} Y_{s-t_n} - Y_0 > \epsilon\, e^{\alpha t_n}\right) \\
&\quad + \mathbb{P}_{t_n}\left(\sup_{s\in[t_n,t_{n+1}]} Y_{t_{n+1}-t_n} - Y_{s-t_n} > \epsilon\, e^{\alpha t_n}\right) + \mathbb{P}_{t_n}\left(e^{-\alpha t_n} N_{t_n} - e^{-\alpha t_{n+1}} N_{t_{n+1}} > \epsilon\right) \\
&\leq 2\, \mathbb{P}_{t_n}\left(Y_{t_{n+1}} - Y_{t_n} > \epsilon\, e^{\alpha t_n}\right) + \mathbb{P}_{t_n}\left(e^{-\alpha t_n} N_{t_n} - e^{-\alpha t_{n+1}} N_{t_{n+1}} > \epsilon\right).
\end{aligned}$$

Since Markov inequalities are not precise enough to go further, we need to compute exactly the probability,

$$\mathbb{P}_{t_n}\left(Y_{t_{n+1}-t_n} - Y_0 > \epsilon\, e^{\alpha t_n}\right).$$

From the branching and Markov properties, $Y_{t_{n+1}-t_n} - Y_0$ is a sum of a geometric number, with parameter $W(t_n)^{-1}$, of independent and i.i.d. geometric random variables supported on $\mathbb{Z}_+$ with parameter $e^{-b(t_{n+1}-t_n)}$. Hence, $Y_{t_{n+1}-t_n} - Y_0$ is geometric supported on $\mathbb{Z}_+$ with parameter

$$\frac{e^{-b(t_{n+1}-t_n)}}{W(t_n)\left(1 - e^{-b(t_{n+1}-t_n)}\left(1 - \frac{1}{W(t_n)}\right)\right)},$$

and, we have

$$\mathbb{P}_{t_n}\left(Y_{t_{n+1}-t_n} - Y_0 \geq k\right) = \left(1 - \frac{1}{W(t_n)\left(e^{b(t_{n+1}-t_n)} - 1\right) + 1}\right)^k.$$

Using

$$W(t_n) = \mathcal{O}\left(e^{\alpha t_n}\right) = \mathcal{O}\left(n^{\frac{2\alpha}{\beta}}\right),$$

we have

$$W(t_n)\left(e^{b(t_{n+1}-t_n)} - 1\right) = \mathcal{O}\left(n^{\frac{\alpha}{2\beta}-1}\right).$$

Finally,

$$\mathbb{P}_{t_n}\left(Y_{t_{n+1}-t_n} - Y_0 > \epsilon e^{\alpha t_n}\right) \leq \left(1 - \frac{1}{1 + \mathcal{C} n^{\frac{\alpha}{2\beta}-1}}\right)^{n^{\frac{\alpha}{2\beta}}},$$

for some positive real constant $\mathcal{C}$. Borel-Cantelli's Lemma then entails

$$\lim_{n\to\infty} \sup_{s\in[t_n,t_{n+1}]} \left|e^{-\alpha t_n} N_{t_n} - e^{-\alpha s} N_s\right| = 0, \quad \text{almost surely,}$$

which ends the proof of the almost sure convergence.
Now, for the convergence in L^2, we have that

$$\mathbb{E}_t\left[\left(\psi'(\alpha)e^{-\alpha t}N_t - \mathcal{E}\right)^2\right] \leq 2\mathbb{E}_t\left[e^{-2\alpha t}\left(N_t^\infty - \psi'(\alpha)N_t\right)^2\right] + 2\mathbb{E}_t\left[\left(e^{-\alpha t}N_t^\infty - \mathcal{E}\right)^2\right].$$

The first term in the right hand side of the last inequality converges to 0 according to (5.7). For the second term, since N_t^∞ and $\mathcal{E}$ vanish on the extinction event, we have

$$\lim_{t\to\infty} \mathbb{E}_t\left[\left(e^{-\alpha t}N_t^\infty - \mathcal{E}\right)^2\right] = \lim_{t\to\infty} \mathbb{E}_\infty\left[\left(e^{-\alpha t}N_t^\infty - \mathcal{E}\right)^2\right].$$

The conclusion comes from the fact that $\left(e^{-\alpha t}N_t^\infty,\ t \in \mathbb{R}_+\right)$ is a martingale uniformly bounded in L^2. □

In the preceding proof, we postponed the demonstration of the independence of the residual lifetimes of the alive individuals at time t. We give its proof now, which is quite similar to the Proposition 5.5 of [60].

Proof of Lemma 5.3.3. Let $\left(Y^{(i)}\right)_{0\leq i\leq N_t}$ be a family of independent Lévy processes with Laplace exponent

$$\psi(x) = x - \int_{(0,\infty]} \left(1 - e^{-rx}\right)\Lambda(dr), \quad x \in \mathbb{R}_+,$$

conditioned to hit (t,∞) before 0, for $i \in \{0,\ldots,N_t - 1\}$, and conditioned to hit 0 before (t,∞) for $i = N_t$. We also assume that,

$$Y_0^{(0)} = t \wedge V,$$

and

$$Y_0^{(i)} = t, \quad i \in \{1,\ldots,N_t\}.$$

Now, denote by τ_i the exit time of the ith process of $(0,t)$ and

$$T_n = \sum_{i=0}^{n-1} \tau_i, \quad n \in \{0, \dots, N_t + 1\}.$$

Then, the process defined for all $s \in [0, T_{N_t}]$ by

$$Y_s = \sum_{i=0}^{N_t} Y^{(i)}_{s-T_i} \mathbb{1}_{T_i \le s < T_{i+1}},$$

has the law of the contour process of a splitting tree cut under t. Moreover, the quantity $Y^{(i)}_{\tau_i} - Y^{(i)}_{\tau_i -}$ is the lifetime of the ith alive individual at time t. The family of residual lifetime $(O_i)_{1 \le i \le N_t}$ has then the same distribution as the sequence of the overshoots of the contour above u. Thus, the independence of the Lévy processes $Y^{(i)}$ ensures us that $(O_i,\ i \in [\![2, N_t]\!])$ is an i.i.d family of random variables, and that O_1 is independent of the other O_i's. $\square$

5.4 Proof of Theorem 5.2.2

In this section, we prove the central limit theorem associated to the law of large numbers for N_t. The first step of the method is to obtain informations on the moments of the error in the a.s. convergence of the process. Using the renewal structure of the tree and formulae on the expectation of a random integral, we are able to express the moments of the error in terms of the scale function of a Lévy process. This process is known to be the contour process of the splitting tree as constructed in [60]. The asymptotic behaviours of the moments are then precisely studied thanks to the precise asymptotic results obtained on the scale function W in Proposition 4.1.1. The second ingredient is a decomposition of the splitting tree into subtrees whose laws are characterized by the overshoots of the contour process over a fixed level. Finally, the error term can be decomposed as the sum of the error made in each subtrees. Our controls on the moments ensure that the error in each subtree decreases fast enough compared to the growth of the population (see Section 5.4.2).
The first section is devoted to the introduction of a useful lemma used in this work on the expectation of a random integral. Section 5.4.2 details the main lines of the method. Theorem 5.2.2 is finally proved in Section 5.4.3.

5.4.1 Preliminaries : A lemma on the expectation of a random integral with respect to a Poisson random measure

The purpose of this part is to state and prove a lemma concerning the expectation of a random integral.

Lemma 5.4.1. *Let ξ be a Poisson random measure on $\mathbb{R}_+$ with intensity $\theta\lambda(da)$ where θ is a positive real number and λ the Lebesgue measure. Let also $\left(X^{(i)}_u,\ u \in \mathbb{R}_+\right)_{i \ge 1}$ be an i.i.d. sequence of non-negative càdlàg random processes independent of ξ. Let also Y be a random* *variable independent of* ξ *and of the family* $\left(X_u^{(i)},\ u \in \mathbb{R}_+\right)_{i\geq 1}$*. If* ξ_u *denotes* $\xi\left([0,u]\right)$*, then, for any* $t \geq 0$*,*

$$\mathbb{E}\int_{[0,t]} X_u^{(\xi_u)} \mathbb{1}_{Y>u}\ \xi(du) = \int_0^t \mathbb{P}\left(Y > u\right) \theta \mathbb{E} X_u du,$$

where $(X_u,\ u \in \mathbb{R}_+) = \left(X_u^{(1)},\ u \in \mathbb{R}_+\right)$*. In addition, for any* $t \leq s$*, we have*

$$\mathbb{E}\left[\int_{[0,t]} X_v^{(\xi_v)} \mathbb{1}_{Y>v}\ \xi(dv) \int_{[0,s]} X_u^{(\xi_u)} \mathbb{1}_{Y>u}\ \xi(du)\right] = \int_0^t \theta \mathbb{E}\left[X_u^2\right] \mathbb{P}\left(Y > u\right)\ du$$
$$+ \int_0^t \int_0^s \theta^2 \mathbb{E} X_u \mathbb{E} X_v \mathbb{P}\left(Y > u, Y > v\right)\ dudv.$$

Proof. Since the proof of the two formulas lies on the same ideas, we only give the proof of the second formula.
First of all, let $f : \mathbb{R}_+^2 \to \mathbb{R}_+$ be a positive measurable deterministic function. We recall that, for a Poisson random measure, the measures of two disjoint measurable sets are independent random variables. That is, for A, B in the Borel σ-field of $\mathbb{R}_+$, $\xi(A \cap B^c)$ is independent from $\xi(B)$, which leads to

$$\mathbb{E}\left[\xi(A)\xi(B)\right] = \mathbb{E}\xi(A)\mathbb{E}\xi(B) + \mathrm{Var}\xi(A \cap B).$$

Using the approximation of f by an increasing sequence of simple function, as in the construction of Lebesgue's integral, it follows from the Fubini-Tonelli theorem and the monotone convergence theorem, that

$$\mathbb{E}\int_{[0,t]\times[0,s]} f(u,v)\ \xi(du)\xi(dv) = \int_0^t \theta f(u,u)\ du + \int_0^t \int_0^s \theta^2 f(u,v)\ dudv.$$

Since the desired relation only depends on the law of our random objects, we can assume without loss of generality that ξ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the family $\left(X_s^{(i)},\ s \in \mathbb{R}_+\right)_{i\geq 1}$ is defined on an other probability space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$. Then, using a slight abuse of notation, we define ξ on $\Omega \times \tilde{\Omega}$ by $\xi_{(\omega,\tilde{\omega})} = \xi_\omega$, and similarly for the family X.
Then, by Fubini-Tonneli Theorem, with the notation $\xi_\omega^v = \xi_\omega\left([0,v]\right)$,

$$\begin{aligned}
&\mathbb{E}\left[\int_{[0,t]\times[0,s]} X_v^{(\xi_v)} X_u^{(\xi_u)}\ \xi(du)\xi(dv)\right] \\
&= \int_{\Omega\times\tilde{\Omega}} \int_{[0,t]\times[0,s]} X_v^{(\xi_\omega^v)}(\tilde{\omega}) X_u^{(\xi_\omega^u)}(\tilde{\omega})\ \xi_\omega(du)\xi_\omega(dv)\ \mathbb{P}\otimes\tilde{\mathbb{P}}\left(d\omega, d\tilde{\omega}\right) \\
&= \int_{\Omega} \int_{[0,t]\times[0,s]} \left[\int_{\tilde{\Omega}} X_v^{(\xi_\omega^v)}(\tilde{\omega}) X_u^{(\xi_\omega^u)}(\tilde{\omega}) \tilde{\mathbb{P}}\left(d\tilde{\omega}\right)\right]\ \xi_\omega(du)\xi_\omega(dv)\ \mathbb{P}(d\omega).
\end{aligned}$$

But since the $X^{(i)}$ are identically distributed and ξ is a simple measure (purely atomic with mass one for each atom) we deduce that, if u and v are two atoms of ξ_ω, $\xi_\omega^v = \xi_\omega^u$ if and only if $u = v$, which implies that

$$\int_{\tilde{\Omega}} X_v^{(\xi_\omega^v)}(\tilde{\omega}) X_u^{(\xi_\omega^u)}(\tilde{\omega}) \tilde{\mathbb{P}}\left(d\tilde{\omega}\right) = \begin{cases} \mathbb{E}X_u\mathbb{E}X_v, & u \neq v, \\ \mathbb{E}X_u^2, & u = v, \end{cases} \qquad \xi_\omega - a.e.$$

The result follows readily, and the case with the indicator function of Y is left to the reader. □

5.4.2 Strategy of proof

Now, we detail the main lines of the proof. Let $(G_n)_{n\geq 1}$ be a sequence of geometric random variables with respective parameter $\frac{1}{n}$, and $(X_i)_{i\geq 1}$ a L^2 family of i.i.d. random variables with zero mean independent of $(G_n)_{n\geq 1}$. It is easy to show that the characteristic function of

$$Z_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{G_n} X_i, \tag{5.10}$$

is given by

$$\mathbb{E}e^{i\lambda Z_n} = \frac{1 + o_n(1)}{1 + \lambda^2 \mathbb{E}X_1^2 + o_n(1)}, \tag{5.11}$$

from which we deduce that Z_n converges in distribution to $\mathcal{L}(0, \mathbb{E}X_1^2)$.
If we suppose that the population counting process N is a Yule Markov process, it clearly follows from the branching and Markov properties that, for $s < t$,

$$N_t = \sum_{i=1}^{N_s} N_{t-s}^i, \tag{5.12}$$

where the family $\left(N_{t-s}^i\right)_{i\geq 1}$ is an i.i.d. sequence of random variables distributed as N_{t-s} and independent of N_s. Moreover, since N_s is geometrically distributed with parameter $e^{-\alpha s}$, taking the renormalized limit leads to,

$$\lim_{t\to\infty} e^{-\alpha t} N_t =: \mathcal{E} = e^{-\alpha s} \sum_{i=1}^{N_s} \mathcal{E}_i,$$

where $\mathcal{E}_1, \dots, \mathcal{E}_{N_s}$ is an i.i.d. family of exponential random variables with parameter one, and independent of N_s. Hence,

$$N_t - e^{\alpha t}\mathcal{E} = \sum_{i=1}^{N_s} \left(N_{t-s}^i - e^{\alpha(t-s)} \mathcal{E}_i \right),$$

is a geometric sum of centered i.i.d. random variables. This remark and (5.10) suggest the desired CLT in the Yule case.

However, in the general case, we need to overcome some important difficulties. First of all, equation (5.12) is wrong in general. Nevertheless, a much weaker version of (5.12) can be obtained in the general case. To make this clear, if $u < t$ are two positive real numbers, then the number of alive individual at time t is the sum of the contributions of each subtrees $\mathbb{T}\left(O_i\right)$ induced by each alive individuals at time u (see Figure 5.2). Provided there are individuals alive at time u, we denote by $(O_i)_{1\leq i\leq N_u}$ the residual lifetimes (see Figure 5.2) of the alive individuals at time u indexed using that the ith individual is the ith individual visited by the contour process. Hence,

$$N_t = \sum_{i=1}^{N_u} N_{t-u}^i \left(O_i\right), \tag{5.13}$$

where $\left(N_{t-u}^i \left(O_i\right)\right)_{i\leq N_u}$ denote the population counting processes of the subtrees $\mathbb{T}(O_i)$ induced by each individual. The notation refers to the fact that each subtree has the law of a standard

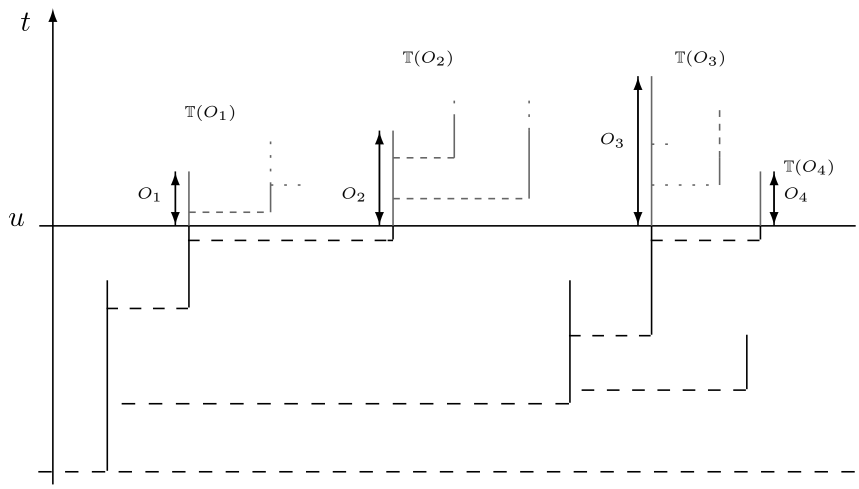

FIGURE 5.2 – Residual lifetimes with subtrees associated to living individuals at time u.

splitting tree with the only difference that the lifelength of the root is given by O_i. More precisely, we define, for all $i \geq 1$ and $o \in \mathbb{R}_+$, $N^i_{t-u}(o)$ the population counting process of the splitting tree constructed from the same random objects as the ith subtree of Figure 5.2, where the life duration of the first individual is equal to o. Hence, from the independence properties between each individuals, $\left(N^i_{t-u}(o),\ t \geq u, o \geq 0\right)_{i\geq 1}$ is a family of independent processes, independent of $(O_i)_{1\leq i\leq N_u}$, and $\left(N^i_{t-u}(o), t \geq u\right)$ has the law of the population counting process of a splitting tree but where the lifespan of the ancestor is o. Note that the lifespans of the other individuals are still distributed as V. From the discussion above, it follows that the family of processes $\left(N^i_{t-u}(O_i),\ t \geq u\right)_{1\leq i\leq N_u}$ are dependent only through the residual lifetimes $(O_i)_{1\leq i\leq N_u}$ and the law of $(N_t(O_i),\ t \in \mathbb{R}_+)$ under $\mathbb{P}_u$ is the law of standard population counting process of splitting tree where the lifespan of the root is distributed as O_i under $\mathbb{P}_u$.
Unfortunately, the computation of (5.11) does not apply to (5.13). This issue is solved by the following lemma which is an improvement of Lemma 5.3.3 and whose proof is similar to one of Proposition 5.5 of [60].

Lemma 5.4.2. *Let u in $\mathbb{R}_+$, we denote by O_i for i an integer between 1 and N_u the residual lifetime of the ith individuals alive at time u. Then under $\mathbb{P}_u$, the family $(O_i,\ i \in [\![1, N_u]\!])$ form a family of independent random variables, independent of N_u, and, expect O_1, having the same distribution, given by, for $2 \leq i \leq N_t$,*

$$\mathbb{P}_u(O_i \in dx) = \int_{\mathbb{R}_+} \frac{W(u-y)}{W(u)-1} b\mathbb{P}\left(V - y \in dx\right)\ dy. \tag{5.14}$$

Moreover, it follows that the family $(N_s(O_i), s \in \mathbb{R}_+)_{1\leq i\leq N_u}$ is an independent family of process, i.i.d. for $i \geq 2$, and independent of N_u.

Proof. Let $\left(Y^{(i)}\right)_{0\leq i\leq N_u}$ a family of independent Lévy processes with Laplace exponent

$$\psi(x) = x - \int_{(0,\infty]} \left(1 - e^{-rx}\right) \Lambda(dr), \quad x \in \mathbb{R}_+,$$

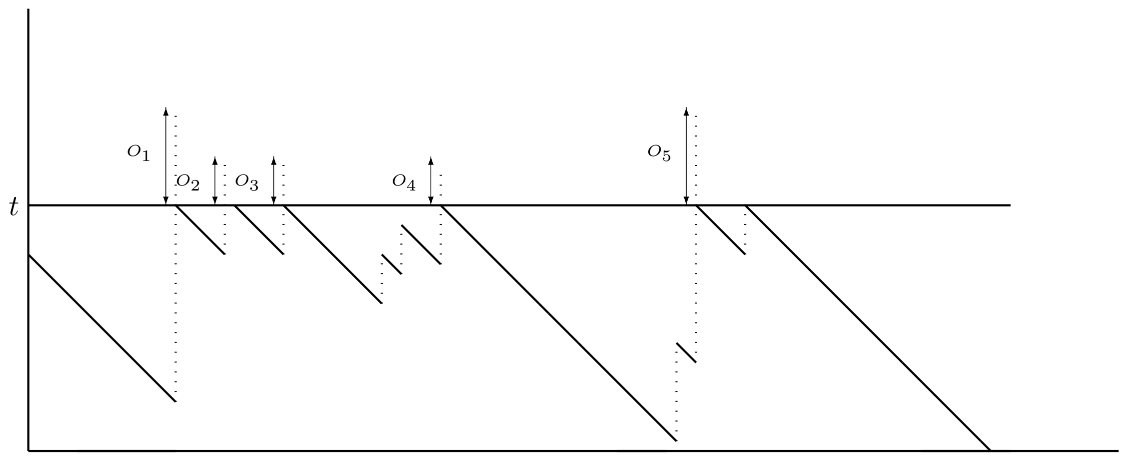

FIGURE 5.3 – Reflected JCCP with overshoot over t. Independence is provided by the Markov property.

conditioned to hit (u, ∞) before hitting 0, for $i \in \{0, \dots, N_u - 1\}$, and conditioned to hit 0 first for $i = N_u$. We also assume that,

$$Y_0^{(0)} = u \wedge V,$$

and

$$Y_0^{(i)} = u, \quad i \in \{1, \dots, N_u\}.$$

Now, denote by τ_i the exit time of the ith process out of $(0, u)$ and

$$T_n = \sum_{i=0}^{n-1} \tau_i, \quad n \in \{0, \dots, N_u + 1\}.$$

Then, the process defined, for all s, by

$$Y_s = \sum_{i=0}^{N_u} Y_{s-T_i}^{(i)} \mathbb{1}_{T_i \leq s < T_{i+1}},$$

has the law of the contour process of a splitting tree cut above u. Moreover, the quantity $Y_{\tau_i} - Y_{\tau_i -}$ is the lifetime of the ith alive individual at time t. The family of residual lifetimes $(O_i)_{1 \leq i \leq N_u}$ has then the same distribution as the sequence of the overshoots of the Y above u. Thus, the Markov property ensures us that $(O_i,\ i \in [\![2, N_u]\!])$ is an i.i.d. family of random variables. The Markov property also ensures that O_1 is independent of the other O_i's.
It remains to derive the law of O_i. Let Y be a Lévy process with Laplace exponent ψ. We denote by τ_u^+ the time of first passage of $-Y$ above u and τ_0^- the time of first passage of $-Y$ below 0. Then, for all $i \geq 2$,

$$\mathbb{P}_u (O_i \in dx) = \mathbb{P}_0 \left(-Y_{\tau_0^-} \in dx \mid \tau_0^- < \tau_u^+ \right).$$

On the other hand, Theorem 2.6.1 gives for any measurable subsets $A \subset [0, u]$, $B \subset (0, -\infty)$,

$$\mathbb{P}_0 \left(-Y_{\tau_0^-} \in B, -Y_{\tau_0^- -} \in A \right) = \int_A \mathbb{P}_{-V} (B - y) \frac{W(u-y)}{W(u)} dy.$$

The result follows easily from

$$\mathbb{P} \left(\tau_0^- < \tau_u^+ \right) = 1 - \frac{1}{W(u)}.$$

□

Remark 5.4.3. *It is important to note that the law of the residual lifetimes of the individuals considered above depends on the particular time u we choose to cut the tree. That is why, in the sequel, we may denote $O_i^{(u)}$ for O_i when we want to underline the dependence in time of the law of the residual lifetimes.*

In addition, as suggested by (5.11), we need to compute the expected quadratic error in the convergence of N_t,

$$\mathbb{E} \left[\left(\psi'(\alpha) N_t - e^{\alpha t} \mathcal{E} \right)^2 \right],$$

which implies to compute $\mathbb{E} N_t \mathcal{E}$.

Although this moment is easy to obtain in the Markovian case, the method does not extend easily to the general case. One idea is to characterize it as a solution of a renewal equation in the spirit of the theory of general CMJ processes.

To make this, we use the renewal structure of a splitting tree : the splitting trees can be constructed (see Chapter 3) by grafting i.i.d. splitting tree on a branch (a tree with a single individual) of length $V_\emptyset$ distributed as V. Therefore, there exists a family $\left(N_t^{(i)},\ t \in \mathbb{R}_+ \right)_{i \geq 1}$ of i.i.d. population counting processes with the same law as $(N_t,\ t \in \mathbb{R}_+)$, and a Poisson random measure ξ on $\mathbb{R}_+$ with intensity $b\, da$ such that

$$N_t = \int_{[0,t]} N_{t-u}^{(\xi_u)} \mathbb{1}_{V_\emptyset > u}\ \xi(du) + \mathbb{1}_{V_\emptyset > t}, \quad a.s., \tag{5.15}$$

where $\xi_u = \xi \left([0, u] \right)$.

Another difficulty comes from the fact that, unlike (5.10), the quantities summed in (5.13) are time-dependent, which requires a careful analysis of the asymptotic behaviour of their moments. The calculus and the asymptotic analysis of these moments is made in Section 5.4.3 : In Lemma 5.4.4, we compute $\mathbb{E} N_t \mathcal{E}$, and then with Lemmas 5.4.5 and 5.4.7, we study the asymptotic behaviour of the error of order 2 and 3 respectively. The second part of Section 5.4.3 is devoted to the study of the same questions for the population counting processes of the subtrees described in Figure 5.2 (when the lifetime of the root is not distributed as V). Finally, Section 5.4.4 is devoted to the proof of Theorem 5.2.2.

One of the difficulties in studying the behaviour of the moments is to get better estimates on the scale function W than those of Lemma 3.3.3. This is the subject of the next section.

5.4.3 Preliminary moments estimates

We begin the proof of Theorem 5.2.2 by computing moments, and analysing their asymptotic behaviours. A first part is devoted to the case of a splitting tree where the lifetime of the root is distributed as V whereas a second part study the case where the lifespan of the root is arbitrary (for instance, as the subtrees described by Figure 5.2).
This section is devoted to the calculus of the expectation of $\left(N_t - e^{\alpha t}\mathcal{E}\right)^2$. We start with the simple case where the initial individual has life-length distributed as V. Secondly, we study the asymptotic behavior of these moments. In the second part of this section, we prove similar result for arbitrary initial distributions.
The expectations above are given with respect to $\mathbb{P}$, however since N_t and $\mathcal{E}$ vanish on the event $\{N_t = 0\}$, we can easily recover the results with respect to $\mathbb{P}_t$ by using (3.9) and (3.5) (see Corollary 5.4.6).

Case $V_\emptyset \overset{d}{=} V$

We start with the computation of $\mathbb{E}N_t\mathcal{E}$.

Lemma 5.4.4 (Join moment of $\mathcal{E}$ and N_t)**.** *The function $t \to \mathbb{E}\left[N_t\mathcal{E}\right]$ is the unique solution bounded on finite intervals of the renewal equation,*

$$\begin{aligned} f(t) = &\int_{\mathbb{R}_+} f(t-u)be^{-\alpha u}\mathbb{P}\left(V > u\right) du \\ &+ \alpha b\mathbb{E}\left[N_.\right] \star \left(\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V > \cdot, V > v\right) dv\right)(t) \\ &+ \alpha \int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V > t, V > v\right) dv, \end{aligned} \tag{5.16}$$

and its solution is given by

$$\left(1 + \frac{\alpha}{b} - e^{-\alpha t}\right) W(t) - \left(1 - e^{-\alpha t}\right) W \star \mathbb{P}_V(t).$$

Proof. As explained in Section 5.4.2,

$$N_t = \int_{[0,t]} N_{t-u}^{(\xi_u)} \mathbb{1}_{V_\emptyset > u}\ \xi(du) + \mathbb{1}_{V_\emptyset > t},$$

where ξ a Poisson point process with rate b on the real line, $\left(N^{(i)}\right)_{i\geq 1}$ is a family of independent CMJ processes with the same law as N and $V_\emptyset$ is the lifespan of the root. Moreover, the three objects $N^{(u)}$, ξ and $V_\emptyset$ are independent.
It follows that, for $s > t$

$$\begin{aligned} N_t N_s = &\int_{[0,t]\times[0,s]} N_{t-u}^{(\xi_u)} N_{s-v}^{(\xi_v)} \mathbb{1}_{V_\emptyset > u} \mathbb{1}_{V_\emptyset > v}\ \xi(du)\xi(dv) \\ &+ \int_{[0,t]} N_{t-u}^{(\xi_u)} \mathbb{1}_{V_\emptyset > u}\ \xi(du) \mathbb{1}_{V_\emptyset > s} + \int_{[0,s]} N_{s-u}^{(\xi_u)} \mathbb{1}_{V_\emptyset > u}\ \xi(du) \mathbb{1}_{V_\emptyset > t} + \mathbb{1}_{V_\emptyset > t}\mathbb{1}_{V_\emptyset > s}, \end{aligned}$$

and, using Lemma 5.4.1,

$$\begin{aligned}\mathbb{E}N_tN_s = &\int_{[0,t]} b\mathbb{E}\left[N_{t-u}N_{s-u}\right]\ \mathbb{P}\left(V > u\right)\ du \\ &+ \int_{[0,t]\times[0,s]} b^2\mathbb{E}\left[N_{t-u}\right]\mathbb{E}\left[N_{s-v}\right]\mathbb{P}\left(V > u, V > v\right)\ du\ dv \\ &+ \mathbb{P}\left(V > s\right)\int_{[0,t]} b\mathbb{E}\left[N_{t-u}\right]\ du + \int_{[0,s]} b\mathbb{E}\left[N_{s-u}\right]\mathbb{P}\left(V > u, V > t\right)\ du + \mathbb{P}\left(V > s\right).\end{aligned}$$

Then, thanks to the estimate $W(t) = \mathcal{O}\left(e^{\alpha t}\right)$ (see Lemma 3.3.3 or 4.1.1) and the L^1 convergence of $W(s)^{-1}N_tN_s$ to $N_t\mathcal{E}$ as s goes to infinity (since, by Theorem 5.2.1, $\frac{N_s}{W(s)}$ converge in L^2 and using Cauchy-Schwarz inequality), we can exchange limit and integrals to obtain,

$$\begin{aligned}\lim_{s\to\infty}\mathbb{E}N_t\frac{N_s}{W(s)} = \underbrace{\mathbb{E}N_t\mathcal{E}}_{:=f(t)} = &\underbrace{\int_{[0,t]} \mathbb{E}\left[N_{t-u}\mathcal{E}\right]e^{-\alpha u}\ \mathbb{P}\left(V > u\right)\ b\ du}_{=:f\star G(t)} \\ &+ \underbrace{\int_{[0,t]\times[0,\infty)} \alpha b\mathbb{E}\left[N_{t-u}\right]e^{-\alpha v}\mathbb{P}\left(V > u, V > v\right)\ du\ dv}_{=:\zeta_1(t)} \\ &+ \underbrace{\int_{[0,\infty]} \alpha e^{-\alpha v}\mathbb{P}\left(V > v, V > t\right)\ dv,}_{=:\zeta_2(t)}\end{aligned}$$

where we used that $\lim_{t\to\infty} W(t)^{-1}\mathbb{E}N_t = \frac{\alpha}{b}$.

Now, we need to solve the last equation to obtain the last part of the lemma. To do that, we compute the Laplace transform of each part of the equation. Note that, since $W(t) = \mathcal{O}\left(e^{\alpha t}\right)$, it is easy to see that the Laplace transform of each term of (5.16) is well-defined as soon as $\lambda > \alpha$ (using Cauchy-Schwarz inequality for the first term). Now, using (3.7),

$$\begin{aligned}T_{\mathcal{L}}e^{\alpha\cdot}G(\lambda) &= b\int_{\mathbb{R}_+} e^{-\lambda t}\mathbb{P}\left(V > t\right)dt = b\int_{\mathbb{R}_+} e^{-\lambda t}\int_{(t,\infty)} \mathbb{P}_V\left(dv\right)dt \\ &= \frac{1}{\lambda}\int_{\mathbb{R}_+}\left(1 - e^{-\lambda v}\right)\ b\mathbb{P}_V\left(dv\right) = 1 - \frac{\psi(\lambda)}{\lambda}. \end{aligned} \tag{5.17}$$

So,

$$T_{\mathcal{L}}G(\lambda) = 1 - \frac{\psi(\lambda+\alpha)}{\lambda+\alpha}.$$

Then,

$$\begin{aligned}T_{\mathcal{L}}\zeta_1(\lambda) = \alpha T_{\mathcal{L}}\mathbb{E}N_\cdot(\lambda)T_{\mathcal{L}}&\left(b\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V > \cdot, V > v\right)dv\right)(\lambda) \\ &= \left(\frac{\lambda}{\psi(\lambda)} - 1\right)\underbrace{T_{\mathcal{L}}\left(\alpha\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V > \cdot, V > v\right)dv\right)(\lambda)}_{=\mathcal{L}\zeta_2(\lambda)}.\end{aligned}$$

and, using (5.17), we get

$$T_{\mathcal{L}}\zeta_2(\lambda) = \alpha \int_{\mathbb{R}_+} e^{-\lambda t} \int_{\mathbb{R}_+} e^{-\alpha v} \mathbb{P}\left(V > t, V > v\right) dv \; dt = \frac{1}{b}\left(\frac{\psi\left(\lambda+\alpha\right)}{\lambda} - \frac{\psi(\lambda)}{\lambda}\right).$$

Finally, we obtain,

$$\begin{aligned} &T_{\mathcal{L}} f(\lambda) \\ &= T_{\mathcal{L}} f(\lambda)\left(1 - \frac{\psi(\lambda+\alpha)}{\lambda+\alpha}\right) + \left(\frac{\lambda}{\psi(\lambda)} - 1\right)\frac{1}{b}\left(\frac{\psi\left(\lambda+\alpha\right)}{\lambda} - \frac{\psi(\lambda)}{\lambda}\right) + \frac{1}{b}\left(\frac{\psi\left(\lambda+\alpha\right)}{\lambda} - \frac{\psi(\lambda)}{\lambda}\right). \end{aligned}$$

Hence,

$$T_{\mathcal{L}} f(\lambda) = \frac{\lambda}{b}\left(\frac{1}{\psi(\lambda)} - \frac{1}{\psi(\lambda+\alpha)}\right).$$

Finally, using (3.2) and

$$bT_{\mathcal{L}}\left(W \star \mathbb{P}_V\right)(\lambda) = \frac{\left(\psi(\lambda) - b + \lambda\right)}{\psi(\lambda)},$$

allows to inverse the Laplace transform of f and get the result. □

Lemma 5.4.4 allows us to compute the expected quadratic error.

Lemma 5.4.5 (Quadratic error in the convergence of N_t)**.** *Let $\mathcal{E}$ the a.s. limit of $\psi'(\alpha)e^{-\alpha t}N_t$. Then,*

$$\lim_{t\to\infty} e^{-\alpha t}\mathbb{E}\left(\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}\right)^2 = \frac{\alpha}{b}\left(2 - \psi'(\alpha)\right).$$

Proof. Let

$$\mu := \lim_{t\to\infty} e^{\alpha t} F(t),$$

where F is defined in Proposition 4.1.1. We have, using Proposition 4.1.1 and (5.2),

$$\begin{aligned} &\int_{[0,t]} W(t-u)\mathbb{P}_V(du) \\ &\quad = \frac{e^{\alpha t}}{\psi'(\alpha)}\left(1 - \frac{\alpha}{b}\right) - \mu - \frac{e^{\alpha t}}{\psi'(\alpha)}\int_{(t,\infty)} e^{-\alpha u}\mathbb{P}_V(du) + \int_{[0,t]}\left(\mu - e^{\alpha(t-u)}F(t-u)\right)\mathbb{P}_V(du) \\ &\quad = \frac{e^{\alpha t}}{\psi'(\alpha)}\left(1 - \frac{\alpha}{b}\right) - \mu + o(1). \end{aligned}$$

Hence, the expression of $\mathbb{E}N_t\mathcal{E}$ given by Lemma 5.4.4 can be rewritten, thanks to Lemmas 4.1.1, as

$$\mathbb{E}N_t\mathcal{E} = \frac{2\alpha e^{\alpha t}}{b\psi'(\alpha)} - \frac{\alpha}{b}\left(\frac{1}{\psi'(\alpha)} + \mu\right) + o(1), \tag{5.18}$$

Using (3.3) and (3.5) in conjunction with Proposition 4.1.1, we also have

$$e^{-\alpha t}\mathbb{E}N_t^2 = 2\frac{\alpha e^{\alpha t}}{b\psi'(\alpha)^2} - \frac{2\alpha\mu}{b\psi'(\alpha)} - \frac{\alpha}{b\psi'(\alpha)} + o(1). \tag{5.19}$$

Hence, it finally follows from (5.18) and (5.19) that

$$\begin{aligned} e^{-\alpha t}\mathbb{E}\left(\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}\right)^2 &= \psi'(\alpha)^2 e^{-\alpha t}\mathbb{E}N_t^2 - 2\psi'(\alpha)\mathbb{E}N_t\mathcal{E} + \frac{2\alpha e^{\alpha t}}{b} \\ &= -2\frac{\alpha\mu}{b}\psi'(\alpha) - \frac{\alpha\psi'(\alpha)}{b} + 2\frac{\alpha}{b}\left(1 + \psi'(\alpha)\mu\right) + o(1) \\ &= \frac{\alpha}{b}\left(2 - \psi'(\alpha)\right) + o(1). \end{aligned}$$

□

It is worth noting that, using (3.5) and the method above, we have the following result.

Corollary 5.4.6. *We have*

$$\frac{1}{\mathbb{P}\left(N_t > 0\right)} = \frac{b}{\alpha} - \frac{b\mu\psi'(\alpha)}{\alpha}e^{-\alpha t} + o(e^{-\alpha t}), \tag{5.20}$$

which leads to

$$\mathbb{E}_t N_t\mathcal{E} = \frac{2e^{\alpha t}}{\psi'(\alpha)} - \frac{1}{\psi'(\alpha)} - 3\mu + o(1). \tag{5.21}$$

Our last estimate is the boundedness of the third moments.

Lemma 5.4.7 (Boundedness of the third moment)**.** *The third moment of the error is asymptotically bounded, that is*

$$\mathbb{E}\left[\left|e^{-\frac{\alpha}{2}t}\left(\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}\right)\right|^3\right] = \mathcal{O}\left(1\right).$$

Proof. We define for all $t \geq 0$, N_t^∞ as the number of individuals alive at time t which have an infinite descent. According to Proposition 5.3.1, N^∞ is a Yule process under $\mathbb{P}_\infty$.
We have

$$\mathbb{E}\left[\left|\frac{\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right|^3\right] \leq 8\mathbb{E}\left[\left|\frac{\psi'(\alpha)N_t - N_t^\infty}{e^{\frac{\alpha}{2}t}}\right|^3\right] + 8\mathbb{E}\left[\left|\frac{N_t^\infty - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right|^3\right].$$

Now, we know according to the proof of Theorem 5.2.1 (and this is easy to prove using the decomposition of Figure 5.2) that N^∞ can be decomposed as

$$N_t^\infty = \sum_{i=1}^{N_t} B_i^{(t)},$$

where $\left(B_i^{(t)}\right)_{i\geq 1}$ is a family of independent Bernoulli random variables, which is i.i.d. for $i \geq 2$, under $\mathbb{P}_t$. Hence,

$$\mathbb{E}_t\left[\left|\frac{\psi'(\alpha)N_t - N_t^\infty}{e^{\frac{\alpha}{2}t}}\right|^3\right] \leq e^{-\frac{3}{2}\alpha t}\mathbb{E}_t\left[\left(\sum_{i=1}^{N_t}\left(\psi'(\alpha) - B_i^{(t)}\right)\right)^4\right]^{\frac{3}{4}}.$$

Since, it is known from the proof of Theorem 5.2.1 that

$$\mathbb{E}B_2^{(t)} = \psi'(\alpha) + \mathcal{O}\left(e^{-\alpha t}\right),$$

it is straightforward that

$$\mathbb{E}_t\left[\left|\frac{\psi'(\alpha)N_t - N_t^\infty}{e^{\frac{\alpha}{2}t}}\right|^3\right]$$

is bounded.

On the other hand, we know that a Yule process is a time-changed Poisson process (see for instance [4], Theorem III.11.2), that is, if P_t is a Poisson process independent of $\mathcal{E}$ under $\mathbb{P}_\infty$,

$$\mathbb{E}\left[\left|\frac{N_t^\infty - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right|^3\right] = \mathbb{E}_\infty\left[\left|\frac{P_{\mathcal{E}(e^{\alpha t}-1)} - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right|^3\right]\mathbb{P}(\mathrm{NonEx}).$$

Now, using Hölder inequality, it remains to bound

$$\mathbb{E}_\infty\left[\left(\frac{P_{\mathcal{E}(e^{\alpha t}-1)} - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)^4\right] = e^{-2\alpha t}\int_{\mathbb{R}_+}\mathbb{E}_\infty\left[\left(P_{x(e^{\alpha t}-1)} - e^{\alpha t}x\right)^4\right]e^{-x}dx.$$

Finally, for a Poissonian random variable X with parameter ν, straightforward computations give that $\mathbb{E}\left[(X-\nu)^4\right] = 3\nu^2 + \nu$, which allows us to end the proof. □

Case with arbitrary initial distribution $\mathbb{P}_{V_\emptyset}$

In order to study the behavior of the sub-splitting trees involved in the decomposition described in Figure 5.2, we investigate the behaviour of a splitting tree where the ancestor lifelength is not distributed as V, but follows an arbitrary distribution. Let Ξ be a random variable in $(0,\infty]$, giving to the life-length of the ancestor and by $N(\Xi)$ the associated population counting process. Using the decomposition of $N(\Xi)$ over the lifespan of the ancestor, as described in Section 5.4.2, we have

$$N_t(\Xi) = \int_{\mathbb{R}_+} N_{t-u}^{(\xi_u)}\mathbb{1}_{\Xi>u}\,\xi(du) + \mathbb{1}_{\Xi>t}, \tag{5.22}$$

where $\left(N^i\right)_{i\geq 1}$ is a family of i.i.d. CMJ processes with the same law as N independent of Ξ and ξ, as described in section 5.4.2. Let, for all $i \geq 1$, $\mathcal{E}_i$ be

$$\mathcal{E}_i := \lim_{t\to\infty}\psi'(\alpha)e^{-\alpha t}N_t^i, \quad a.s, \tag{5.23}$$

and, let $\mathcal{E}(\Xi)$ be the random variable defined by

$$\mathcal{E}(\Xi) := \int_{[0,\infty]}\mathcal{E}_{(\xi_u)}e^{-\alpha u}\mathbb{1}_{\Xi>u}\,\xi(du). \tag{5.24}$$

Lemma 5.4.8 (First moment)**.** *The first moment is asymptotically bounded, that is*

$$\mathbb{E}\left(\psi'(\alpha)N_t(\Xi) - e^{\alpha t}\mathcal{E}(\Xi)\right) = \mathcal{O}(1),$$

uniformly with respect to the random variable Ξ.

Proof. Using Lemma 5.4.1, (5.22) and (5.24) with have

$$\mathbb{E}\left(\psi'(\alpha)N_t(\Xi) - e^{\alpha t}\mathcal{E}(\Xi)\right) = \int_{[0,t]} \left(\psi'(\alpha)\mathbb{E}N_{t-u} - e^{\alpha(t-u)}\mathbb{E}\mathcal{E}\right) e^{-\alpha u}\mathbb{P}\left(\Xi > u\right) bdu,$$

which leads using (3.6) and (3.9) to

$$\begin{aligned}&\mathbb{E}\left(\psi'(\alpha)N_t(\Xi) - e^{\alpha t}\mathcal{E}(\Xi)\right)\\ &\quad= \int_{[0,t]} \underbrace{\left(\psi'(\alpha)W(t-u) - \psi'(\alpha)W \star \mathbb{P}_V(t-u) - \frac{\alpha}{b}e^{\alpha(t-u)}\right)}_{=:I_{t-u}} e^{-\alpha u}\mathbb{P}\left(\Xi > u\right) bdu. \quad (5.25)\end{aligned}$$

We get using Proposition 4.1.1 and (5.2),

$$\begin{aligned}I_s =& e^{\alpha s} - \psi'(\alpha)e^{\alpha s}F(s) - e^{\alpha s}\left(1 - \frac{\alpha}{b}\right)\\ &+ \psi'(\alpha)\int_{[0,s]} e^{\alpha(s-v)}F(s-v)\mathbb{P}_V(dv) + e^{\alpha s}\int_{(s,\infty)} e^{-\alpha v}\mathbb{P}_V(dv) - \frac{\alpha}{b}e^{\alpha s}\\ =& e^{\alpha s}\int_{(s,\infty)} e^{-\alpha v}\mathbb{P}_V(dv) + o(1).\end{aligned}$$

Hence, $(I_s)_{s\geq 0}$ is bounded. The result, now, follows from (5.25). □

Lemma 5.4.9 (L^2 convergence in the general case)**.** *$\psi'(\alpha)e^{-\alpha t}N_t(\Xi)$ converge a.s. and in L^2 to $\mathcal{E}(\Xi)$, and*

$$\lim_{t\to\infty} e^{-\alpha t}\mathbb{E}\left(\psi'(\alpha)N_t(\Xi) - e^{\alpha t}\mathcal{E}(\Xi)\right)^2 = \frac{\alpha}{b}\left(2 - \psi'(\alpha)\right)\int_{\mathbb{R}_+} e^{-\alpha s}\mathbb{P}\left(\Xi > s\right) bds,$$

where the convergence is uniform with respect to Ξ in $(0,\infty]$. In the particular case when Ξ follows the distribution of $O_2^{(\beta t)}$ given by (5.14), we have, for $0 < \beta < \frac{1}{2}$,

$$\lim_{t\to\infty} e^{\alpha t}\mathbb{E}_{\beta t}\left(e^{-\alpha t}\psi'(\alpha)N_t(O_2^{(\beta t)}) - \mathcal{E}(O_2^{(\beta t)})\right)^2 = \left(2 - \psi'(\alpha)\right)\psi'(\alpha).$$

Proof. From (5.22) and (5.24), we have

$$\left(e^{-\alpha t}\psi'(\alpha)N_t(\Xi) - \mathcal{E}(\Xi)\right)^2 = \left[\int_{\mathbb{R}_+}\left(e^{-\alpha(t-u)}\psi'(\alpha)N_{t-u}^{(\xi_u)} - \mathcal{E}_{(u)}\right) e^{-\alpha u}\mathbb{1}_{\Xi > u}\ \xi(du) + e^{-\alpha t}\mathbb{1}_{\Xi > t}\right]^2 \quad (5.26)$$

and, using Lemma 5.4.1,

$$\begin{aligned}
&\mathbb{E}\left(\psi'(\alpha)e^{-\alpha t}N_t(\Xi)-\mathcal{E}(\Xi)\right)^2\\
=&\mathbb{E}\left(\int_{\mathbb{R}_+}\left(\psi'(\alpha)e^{-\alpha(t-u)}N_{t-u}^{(\xi_u)}-\mathcal{E}_{(u)}\right)e^{-\alpha u}\mathbb{1}_{\Xi>u}\ \xi(du)\right)^2\\
&+e^{-2\alpha t}\mathbb{P}\left(\Xi>t\right)+2e^{-\alpha t}\mathbb{E}\mathbb{1}_{\Xi>t}\int_{\mathbb{R}_+}\left(\psi'(\alpha)e^{-\alpha(t-u)}N_{t-u}^{(\xi_u)}-\mathcal{E}_{(u)}\right)e^{-\alpha u}\mathbb{1}_{\Xi>u}\ \xi(du),\\
=&\int_{\mathbb{R}_+}\mathbb{E}\left[\left(\psi'(\alpha)e^{-\alpha(t-u)}N_{t-u}^{(\xi_u)}-\mathcal{E}_{(u)}\right)^2\right]e^{-2\alpha u}\mathbb{P}\left(\Xi>u\right)\ bdu\\
&+\int_{\mathbb{R}_+}\mathbb{E}\left(\psi'(\alpha)e^{-\alpha(t-u)}N_{t-u}^{(\xi_u)}-\mathcal{E}_{(u)}\right)\mathbb{E}\left(\psi'(\alpha)e^{-\alpha(t-v)}N_{t-v}^{(\xi_v)}-\mathcal{E}_{(v)}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\times e^{-\alpha(u+v)}\mathbb{P}\left(\Xi>u,\Xi>v\right)\ bdu\ dv\\
&+e^{-2\alpha t}\mathbb{P}\left(\Xi>t\right)+2e^{-\alpha t}\int_{\mathbb{R}_+}\mathbb{E}\left(\psi'(\alpha)e^{-\alpha(t-u)}N_{t-u}^{(\xi_u)}-\mathcal{E}_{(u)}\right)e^{-\alpha u}\mathbb{P}\left(\Xi>u,\Xi>t\right)\ bdu.
\end{aligned}$$

Moreover, since,

$$\psi'(\alpha)\mathbb{E}e^{-\alpha t}N_t-\mathcal{E}=\mathcal{O}\left(e^{-\alpha t}\right),$$

this leads, using Lemma 5.4.8, to

$$\lim_{t\to\infty}e^{\alpha t}\mathbb{E}\left(e^{-\alpha t}\psi'(\alpha)N_t(\Xi)-\mathcal{E}(\Xi)\right)^2=\frac{\alpha}{b}\left(2-\psi'(\alpha)\right)\int_{\mathbb{R}_+}e^{-\alpha u}\mathbb{P}\left(\Xi>u\right)bdu.$$

Now, we have from (5.14) and Lemma 3.3.3,

$$\lim_{u\to\infty}\mathbb{P}_u\left(O_2>s\right)=\lim_{u\to\infty}\int_{\mathbb{R}_+}\frac{W(u-y)}{W(u)-1}\mathbb{P}\left(V>s+y\right)bdy=\int_{\mathbb{R}_+}e^{-\alpha y}\mathbb{P}\left(V>s+y\right)bdy.$$

It follows then from Lebesgue theorem that,

$$\lim_{t\to\infty}\int_{\mathbb{R}_+}e^{-\alpha s}\mathbb{P}_{\beta t}\left(O_2>s\right)bds=\frac{b\psi'(\alpha)}{\alpha}.$$

□

Lemma 5.4.10 (Boundedness in the general case.)**.** *The error of order* 3 *in asymptotically bounded, that is*

$$e^{-\frac{3}{2}\alpha t}\mathbb{E}\left|\psi'(\alpha)N_t(\Xi)-e^{\alpha t}\mathcal{E}(\Xi)\right|^3=\mathcal{O}\left(1\right),$$

uniformly w.r.t. Ξ.

Proof. Rewriting $N(\Xi)$ and $\mathcal{E}\left(\Xi\right)$ as in the proof of Lemma 5.4.9, we see that,

$$\begin{aligned}
&e^{-\frac{3}{2}t}\ \mathbb{E}\left|\psi'(\alpha)N_t(\Xi)-e^{\alpha t}\mathcal{E}(\Xi)\right|^3\\
&\quad=e^{-\frac{3}{2}t}\ \mathbb{E}\left[\left|\int_{[0,t]}\left(\psi'(\alpha)N_{t-u}^{(\xi_u)}-e^{\alpha(t-u)}\mathcal{E}_{(u)}\right)\mathbb{1}_{\Xi>u}\ \xi(du)+\psi'(\alpha)\mathbb{1}_{\Xi>t}\right|^3\right]\\
&\quad\leq 8\mathbb{E}\left|\int_{[0,t]}e^{-\frac{3}{2}(t-u)}\left(\psi'(\alpha)N_{t-u}^{(\xi_u)}-e^{\alpha(t-u)}\mathcal{E}_{(u)}\right)\ e^{-\frac{1}{2}u}\mathbb{1}_{\Xi>u}\xi(du)\right|^3+8\psi'(\alpha)e^{-\frac{1}{2}t}\mathbb{P}\left(\Xi>t\right)^3
\end{aligned}$$

We denote by I the first term of the r.h.s. of the last inequality, leading to

$$\begin{aligned}
I &\leq 8\mathbb{E}\int_{[0,t]^3} \prod_{i=1}^{3} \left| e^{-\frac{1}{2}(t-s_i)} \left(\psi'(\alpha) N_{t-s_i}^{(\xi_{s_i})} - e^{\alpha(t-s_i)} \mathcal{E}_{(s_i)} \right) \right| \; e^{-\frac{1}{2}s_i} \mathbb{1}_{\Xi > s_i} \xi(ds_1)\xi(ds_2)\xi(ds_3) \\
&\leq 8\mathbb{E}\int_{[0,t]^3} \sum_{j=1}^{3} \left| e^{-\frac{1}{2}(t-s_j)} \left(\psi'(\alpha) N_{t-s_j}^{(\xi_{s_j})} - e^{\alpha(t-s_j)} \mathcal{E}_{(s_j)} \right) \right|^3 \prod_{i=1}^{3} e^{-\frac{1}{2}s_i} \mathbb{1}_{\Xi > s_i} \xi(ds_1)\xi(ds_2)\xi(ds_3) \\
&\leq 24\mathbb{E}\int_{[0,t]} \left| e^{-\frac{1}{2}(t-u)} \left(\psi'(\alpha) N_{t-u}^{(\xi_u)} - e^{\alpha(t-u)} \mathcal{E}_{(u)} \right) \right|^3 \; e^{-\frac{1}{2}u} \mathbb{1}_{\Xi > u} \xi(du) \left(\int_{[0,t]} e^{-\frac{1}{2}u} \xi(du) \right)^2 \\
&\leq 24\mathbb{E}\int_{[0,t]} \left| e^{-\frac{1}{2}(t-u)} \left(\psi'(\alpha) N_{t-u}^{(\xi_u)} - e^{\alpha(t-u)} \mathcal{E}_{(u)} \right) \right|^3 \; e^{-\frac{1}{2}u} \mathbb{1}_{\Xi > u} \; \mu(du),
\end{aligned}$$

with

$$\mu(du) = \left(\int_{[0,t]} e^{-\frac{1}{2}s} \xi(ds) \right)^2 \xi(du).$$

Now, since μ is independent from the family $\left(N^{(i)}\right)$ and $\left(\mathcal{E}_{(i)}\right)$, an easy adaptation of the proof of Lemma 5.4.1, leads to

$$\begin{aligned}
&e^{-\frac{3}{2}t} \; \mathbb{E}\left| \psi'(\alpha) N_t(\Xi) - e^{\alpha t} \mathcal{E}(\Xi) \right|^3 \\
&\leq 24\mathbb{E}\int_{[0,t]} \mathbb{E}\left[\left| e^{-\frac{1}{2}(t-u)} \left(\psi'(\alpha) N_{t-u} - e^{\alpha(t-u)} \mathcal{E} \right) \right|^3 \right] \; e^{-\frac{1}{2}u} \mathbb{1}_{\Xi > u} \; \mu(du) + 8\psi'(\alpha) e^{-\frac{1}{2}t} \mathbb{P}\left(\Xi > t\right)
\end{aligned}$$

Using Lemma 5.4.7 to bound

$$\mathbb{E}\left| e^{-\frac{3}{2}(t-u)} \left(N_{t-u} - e^{\alpha(t-u)} \mathcal{E} \right) \right|^3,$$

in the previous expression, finally leads to

$$e^{-\frac{3}{2}t} \; \mathbb{E}\left| \psi'(\alpha) N_t(\Xi) - e^{\alpha t} \mathcal{E}(\Xi) \right|^3 \leq \mathcal{C} \; \left(\mathbb{E}\left(\int_{\mathbb{R}_+} e^{-\frac{1}{2}u} \xi(du) \right)^3 + 1 \right),$$

for some real positive constant $\mathcal{C}$. □

5.4.4 Proof of Theorem 5.2.2

We fix a positive real number u. From this point, we recall the decomposition of the splitting tree as described in Section 5.4.2 (see also Figure 5.2). We also recall that, for all i in $\{1, \ldots, N_u\}$, the process $\left(N_s^i\left(O_i\right), \; s \in \mathbb{R}_+\right)$ is the population counting process of the (sub-)splitting tree $\mathbb{T}\left(O_i\right)$. As explained in Section 5.4.2, it follows from the construction of the splitting tree, that, for all i in $\{1, \ldots, N_u\}$, there exists an i.i.d. family of processes $\left(N^{i,j}\right)_{j \geq 1}$ independent from N_u with the same law as $(N_t, \; t \in \mathbb{R}_+)$, and an i.i.d. family $\left(\xi^{(i)}\right)_{1 \leq i \leq N_u}$ of random measure independent from N_u and from $\left(N^{i,j}\right)_{j \geq 1}$ the family with same law as ξ, such that

$$N_t^i\left(O_i\right) = \int_{[0,t]} N_{t-u}^{i,j} \mathbb{1}_{O_i > u} \; \xi^{(i)}(du) + \mathbb{1}_{O_i > t}, \quad \forall t \in \mathbb{R}_+, \quad \forall i \in \{1, \ldots, N_u\}. \tag{5.27}$$

As in (5.24), we define, for all i in $\{1,\dots,N_u\}$,

$$\mathcal{E}\left(O_i\right) := \int_{[0,t]} \mathcal{E}_{i,\xi_u^{(i)}} e^{-\alpha u} \mathbb{1}_{O_i>u}\ \xi^{(i)}(du), \tag{5.28}$$

where $\mathcal{E}_{i,j} := \lim_{t\to\infty} \psi'(\alpha) e^{-\alpha t} N_t^{i,j}$.
Hence, it follows from Lemma 5.4.9, that $e^{-\alpha t} N_t^i\left(O_i\right)$ converges to $\mathcal{E}\left(O_i\right)$ in L^2.
Note also that, from Lemma 5.4.2, the family $\left(N_t^i\left(O_i\right),\ t\in\mathbb{R}_+\right)_{2\le i\le N_u}$ is i.i.d. and independent from N_u under $\mathbb{P}_u$, as well as the family $\left(\mathcal{E}\left(O_i\right)\right)_{2\le i\le N_u}$ (in the sense of Remark 5.3.2). Note that the law under $\mathbb{P}_u$ of the processes of the family $\left(N_t^i\left(O_i\right),\ t\in\mathbb{R}_+\right)_{2\le i\le N_u}$ is the law of standard population counting processes where the lifespan of the root is distributed as O_2 under $\mathbb{P}_u$ (except for the first one).

Lemma 5.4.11 (Decomposition of $\mathcal{E}$)**.** *We have the following decomposition of $\mathcal{E}$,*

$$\mathcal{E} = e^{-\alpha u} \sum_{i=1}^{N_u} \mathcal{E}_i\left(O_i\right), \quad a.s.$$

Moreover, under $\mathbb{P}_u$, the random variables $\left(\mathcal{E}_i\left(O_i\right)\right)_{i\ge1}$ (defined by (5.28)) are independent, independent of N_u, and identically distributed for $i\ge 2$.

Proof. **Step 1 : Decomposition of $\mathcal{E}$.**
For all t in $\mathbb{R}_+$, we denote by N_t^∞ the number of individuals alive at time t which have an infinite descent. For all i, we define, for all $t\ge0$, $N_t^\infty\left(O_i\right)$ from $\mathbb{T}\left(O_i\right)$ as N_t^∞ was defined from the whole tree. Now, it is easily seen that

$$N_t^\infty = \sum_{i=1}^{N_u} N_{t-u}^\infty\left(O_i\right).$$

Hence, if $e^{-\alpha t} N_t^\infty\left(O_i\right)$ converges a.s. to $\mathcal{E}\left(O_i\right)$, then

$$\lim_{t\to\infty} e^{-\alpha t} N_t^\infty = \lim_{t\to\infty} e^{-\alpha u} \sum_{i=1}^{N_u} e^{-\alpha(t-u)} N_{t-u}^\infty\left(O_i\right) = e^{-\alpha u}\sum_{i=1}^{N_u} \mathcal{E}\left(O_i\right).$$

So, it just remains to prove the a.s. convergence to get the desired result.
Step 2 : a.s. convergence of $N^\infty\left(O_i\right)$ to $\mathcal{E}\left(O_i\right)$.
For this step, we fix $i\in\{1,\dots,N_u\}$.
In the same spirit as (5.27) (see also Section 5.4.2), it follows from the construction of the splitting tree $\mathbb{T}\left(O_i\right)$, that there exists, an i.i.d. (and independent of N_u) sequence of processes $\left(N_s^{j,\infty},\ s\in\mathbb{R}_+\right)_{j\ge1}$ with the same law as $\left(N_t^\infty,\ t\in\mathbb{R}_+\right)$ (under $\mathbb{P}$), such that

$$N_t^\infty\left(O_i\right) = \int_{[0,t]} N_{t-u}^{\xi_u^{(i)},\infty}\ \mathbb{1}_{O_i>u}\ \xi^{(i)}(du) + \mathbb{1}_{O_i=\infty},\ \forall t\ge0.$$

Now, it follows from Theorem 5.2.1, that for all j,

$$\lim_{t\to\infty} e^{-\alpha t} N_t^{j,\infty} = \mathcal{E}_{i,j},\ a.s.,$$

where $\mathcal{E}_{i,j}$ was defined in the beginning of this section. Let

$$\mathcal{C}_j := \sup_{t \in \mathbb{R}_+} e^{-\alpha t} N_t^{j,\infty}, \quad \forall j \geq 1,$$

and

$$\mathcal{C} := \sup_{t \in \mathbb{R}_+} e^{-\alpha t} N_t^{\infty}.$$

Then, the family $(\mathcal{C}_j)_{j \geq 1}$ is i.i.d., since the processes $\left(N^{j,\infty}\right)_{j \geq 1}$ are i.i.d, with the same law as $\mathcal{C}$. Hence,

$$\int_{[0,t]} e^{-\alpha(t-u)} N_{t-u}^{\xi_u^{(i)},\infty} \, e^{-\alpha u} \mathbb{1}_{O_i > u} \, \xi^{(i)}(du) \leq \int_{[0,t]} \mathcal{C}_{\xi_u^{(i)}} e^{-\alpha u} \mathbb{1}_{O_i > u} \, \xi^{(i)}(du). \tag{5.29}$$

It is easily seen that $\mathbb{E}\left[\mathcal{C}\right] = \mathbb{P}\left(\text{NonEx}\right) \mathbb{E}_\infty \left[C\right]$. Now, since, from Proposition 5.3.1, N_t^{∞} is a Yule process under $\mathbb{P}_\infty$ (and hence $e^{-\alpha t} N_t^{\infty}$ is a martingale), Doobs's inequalities entails that the random variable $\mathcal{C}$ is integrable. Hence, the right hand side of the (5.29) is a.s. finite, and we can apply Lesbegue Theorem to get

$$\lim_{t \to \infty} e^{-\alpha t} N_t^{\infty}\left(O_i\right) = \int_{[0,t]} \mathcal{E}_{i,\xi_u^{(i)}} \, e^{-\alpha u} \mathbb{1}_{O_i > u} \, \Gamma(du) = \mathcal{E}\left(O_i\right), \quad a.s.,$$

where the right hand side of the last equality is just the definition of $\mathcal{E}\left(O_i\right)$.

□

We have now all the tools needed to prove the central limit theorem for N_t.

Proof of Theorem 5.2.2. Let $u < t$, two positive real numbers. From Lemma 5.4.11 and Section 5.4.2, we have

$$N_t = \sum_{i=1}^{N_u} N_{t-u}^{(i)}\left(O_i\right)$$

and

$$e^{\alpha t} \mathcal{E} = \sum_{i=1}^{N_u} e^{\alpha(t-u)} \mathcal{E}_i\left(O_i\right).$$

Then,

$$\frac{\psi'(\alpha) N_t - e^{\alpha t} \mathcal{E}}{e^{\frac{\alpha}{2} t}} = \sum_{i=1}^{N_u} \frac{\psi'(\alpha) N_{t-u}^{(i)}\left(O_i\right) - e^{\alpha(t-u)} \mathcal{E}_i\left(O_i\right)}{e^{\frac{\alpha}{2}(t-u)} e^{\frac{\alpha}{2} u}}. \tag{5.30}$$

Using Lemma 5.4.2, we know that, under $\mathbb{P}_u$, $\left(N_{t-u}^i(O_i), t > u\right)_{1 \leq i \leq N_u}$ are independent processes, i.i.d. for $i \geq 2$ and independent of N_u. Let us denote by φ and $\tilde{\varphi}$ the characteristic functions

$$\varphi(\lambda) := \mathbb{E}\left[\exp\left(i\lambda\left(\frac{\psi'(\alpha) N_{t-u}^2\left(O_2\right) - e^{\alpha(t-u)} \mathcal{E}_2\left(O_2\right)}{e^{\frac{\alpha}{2}(t-u)}}\right)\right)\right], \quad \lambda \in \mathbb{R}$$

and

$$\tilde{\varphi}(\lambda) := \mathbb{E}\left[\exp\left(i\lambda\left(\frac{\psi'(\alpha) N_{t-u}^1\left(O_1\right) - e^{\alpha(t-u)} \mathcal{E}_1\left(O_1\right)}{e^{\frac{\alpha}{2}(t-u)}}\right)\right)\right], \quad \lambda \in \mathbb{R}.$$

It follows from (5.30) and Lemma 5.4.2 that,

$$\mathbb{E}_u\left[\exp\left(i\lambda\frac{\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right] = \frac{\tilde{\varphi}\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}{\varphi\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}\mathbb{E}_u\left[\varphi\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)^{N_u}\right]$$

Since N_u is geometric with parameter $W(u)^{-1}$ under $\mathbb{P}_u$,

$$\mathbb{E}_u\left[\exp\left(i\lambda\frac{\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right] = \frac{\tilde{\varphi}\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}{\varphi\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}\frac{W(u)^{-1}\varphi\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}{1-\left(1-W(u)^{-1}\right)\varphi\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)}$$

Using Taylor formula for φ, we obtain,

$$\mathbb{E}_u\left[\exp\left(i\lambda\frac{\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right] = \tilde{\varphi}\left(\frac{\lambda}{e^{\frac{\alpha}{2}u}}\right)\frac{1}{D(\lambda,t,u)}$$

where,

$$\begin{aligned}
D(\lambda,t,u) &= W(u) \\
&- (W(u)-1)\left(1 + i\lambda\mathbb{E}\left[\frac{\psi'(\alpha)N^i_{t-u}(O_2) - e^{\alpha(t-u)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-u)}e^{\frac{\alpha}{2}u}}\right]\right. \\
&\qquad\left. - \frac{\lambda^2}{2}\mathbb{E}\left[\left(\frac{\psi'(\alpha)N^i_{t-u}(O_2) - e^{\alpha(t-u)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-u)}e^{\frac{\alpha}{2}u}}\right)^2\right] + R(\lambda,t,u)\right) \\
&= 1 - i\lambda\frac{W(u)-1}{e^{\frac{\alpha}{2}u}}\mathbb{E}\left[\frac{\psi'(\alpha)N^i_{t-u}(O_2) - e^{\alpha(t-u)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-u)}}\right] \\
&\qquad + \frac{\lambda^2}{2}\frac{W(u)-1}{e^{\alpha u}}\mathbb{E}\left[\left(\psi'(\alpha)\frac{N^i_{t-u}(O_2) - e^{\alpha(t-u)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-u)}}\right)^2\right] \\
&\qquad - (W(u)-1)R(\lambda,t,u),
\end{aligned}$$

with, for all $\epsilon > 0$ and all λ in $(-\epsilon,\epsilon)$,

$$\begin{aligned}
|R(\lambda,t,u)| &\leq \sup_{\lambda\in(-\epsilon,\epsilon)}\left|\frac{\partial^3}{\partial\lambda^3}\varphi(\lambda)\right| \\
&\leq \mathbb{E}\left[\left|\left(\frac{\psi'(\alpha)N^i_{t-u}(O_2) - e^{\alpha(t-u)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-u)}}\right)\right|^3\right]\frac{\epsilon^3 e^{-\frac{3}{2}\alpha u}}{6} \leq C\epsilon^3 e^{-\frac{3}{2}u}, \quad (5.31)
\end{aligned}$$

for some real positive constant C obtained using Lemma 5.4.10.
From this point, we set $u = \beta t$ with $0 < \beta < \frac{1}{2}$. It follows then from the Lemmas 5.4.9 and 5.4.2, that

$$\lim_{t\to\infty}\mathbb{E}_{\beta t}\left[\left(\frac{\psi'(\alpha)N^i_{t-\beta t}(O_2) - e^{\alpha(t-\beta t)}\mathcal{E}_2(O_2)}{e^{\frac{\alpha}{2}(t-\beta t)}}\right)^2\right] = \psi'(\alpha)\left(2-\psi'(\alpha)\right). \quad (5.32)$$

Moreover, we have from Lemma 5.4.8, and since $\beta < \frac{1}{2}$,

$$\lim_{t\to\infty}W(\beta t)e^{-\frac{\alpha}{2}t}\mathbb{E}\left[\psi'(\alpha)N^i_t(O_2) - e^{\alpha t}\mathcal{E}_2(O_2)\right] = 0. \quad (5.33)$$

Finally, the relations (5.31), (5.32) and (5.33) lead to

$$\lim_{t\to\infty} \mathbb{E}_{\beta t}\left[\exp\left(i\lambda\frac{N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right] = \frac{1}{1+\frac{\lambda^2}{2}\left(2-\psi'(\alpha)\right)}.$$

To conclude, note that,

$$\begin{aligned}&\left|\mathbb{E}_{\beta t}\left[\exp\left(i\lambda\frac{N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right] - \mathbb{E}_{\infty}\left[\exp\left(i\lambda\frac{N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right)\right]\right| \\ &\quad = \left|\mathbb{E}\left[e^{i\lambda\frac{\psi'(\alpha)N_t - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}}\left(\frac{\mathbb{1}_{N_{\beta t}>0}}{\mathbb{P}\left(N_{\beta t}>0\right)} - \frac{\mathbb{1}_{\mathrm{NonEx}}}{\mathbb{P}\left(\mathrm{NonEx}\right)}\right)\right]\right| \leq \mathbb{E}\left[\left|\frac{\mathbb{1}_{N_{\beta t}>0}}{\mathbb{P}\left(N_{\beta t}>0\right)} - \frac{\mathbb{1}_{\mathrm{NonEx}}}{\mathbb{P}\left(\mathrm{NonEx}\right)}\right|\right]\end{aligned}$$

goes to 0 as t goes to infinity. This ends the proof of Theorem 5.2.2. $\square$

Chapitre 6

On the frequency spectrum of a splitting tree with neutral Poissonian mutations

6.1 Introduction

The purpose of this chapter is to study splitting trees with neutral Poissonian mutations. We consider the same model as in the previous chapter, but we assume that individuals also experience mutations at Poisson rate. Each mutation leads to a totally new type replacing the previous type of the individual, this is the *infinitely-many alleles* assumption. Every time an individual gives birth to new individual, it transmits its type to his child. This mutation process is a way to model the occurrence of a new type in a population (such as a new species or a new phenotype in a given species). Our study concerns the allelic partition of the living population at a fixed time t, which is characterized by the frequency spectrum $(A(k,t))_{k\geq 1}$ of the population, where each integer $A(k,t)$ is the number of families represented by k alive individuals at time t. A famous example is the Ewens sampling formula which gives the distribution of the frequency spectrum when the genealogy is given by the Kingman coalescent [29]. Other works studied similar quantities in the case of Galton-Waston branching processes (see [8] or [37]). The purpose of this chapter is to obtain explicit formulas for the moments of the frequency spectrum and then to use this formulas in order to extend the central limit theorem proved in Chapter 5 to the frequency spectrum.

The model with Poissonian mutations was studied in Champagnat and Lambert [13, 14], where many properties of the frequency spectrum and the clonal family (the family who carries the type of the first individuals at time 0) were obtained. The population counting process $(N_t,\ t \in \mathbb{R}_+)$ and the frequency spectrum $(A(k,t))_{k\geq 1}$ belong to the class of general branching processes counted by random characteristics. This class of processes has been deeply studied by Jagers and Nerman, who give, for instance, criteria for the long time convergence of such processes [46, 74, 48, 49, 87]. Using these tools, Richard and Lambert [60, 82] shown the almost sure convergence of N_t, properly renormalized, to an exponential random variable in the supercritical case. The almost sure convergence of the ratios $\frac{A(k,t)}{N_t}$ was proved in [13] using similar tools. From this, one can easily deduce the a.s. convergence of $\frac{A(k,t)}{W(t)}$ where we recall that $W(t)$ is the average number of individuals at time t conditionally on $N_t > 0$. This result was stated without proof in [15].

An important tool is the so called *coalescent point process* (CPP) : given the individuals alive at a fixed time, the coalescent point process at time t is the tree describing the relation between the lineages of all individuals alive at time t. Here, the term lineage of an individual refers to the succession of individuals, from child to parent, backward in time until the ancestor of the population. Roughly speaking, the CPP is the genealogical tree of the lineages of the individuals. This tool goes back to Aldous and Popovic [2] who introduced it for a Markovian model. Later in [60], Lambert showed the general link between coaslescent point processes and the splitting trees.

In this work, we use the representation of the CPP of a splitting tree as an i.i.d. sequence of random variables $(H_i)_{i\geq 1}$. More precisely, we use the new construction of the coalescent point process given in Chapter 4, and thanks to Theorem 4.2.2, this allows us to obtain explicit recursive formulas for the moments of the frequency spectrum, valid for any parameter of the model. As an application, we prove the almost sure convergence of the frequency spectrum avoiding the use of the theory of general branching processes counted by random characteristics in the supercritical case. Of course, these moment formulas can also provide many valuable informations. For instance, on the error in the aforementioned convergence. Another application is then to prove central limit theorems for the frequency spectrum (such as the one of Chapter 5).

Section 6.2 is dedicated to the description of the models and the introduction of anterior results (essentially from [60, 13]) used in the sequel. In Section 6.3, we state results (Theorems 6.3.1 and 6.3.2) giving explicit formulas for the factorial moments of the frequency spectrum $(A(k,t))_{k\geq 1}$ expressed in terms of the lower order moments. A first example of the method in Subsection 6.3.1 focusing on the expectation of $A(k,t)$. Although, the computation of this expectation was already known from [13], we give here a much more simple proof. Subsection 6.3.2 is dedicated to the proofs of Theorems 6.3.1 and 6.3.2. We give the asymptotic behaviour of higher moments in Subsection 6.3.5. All these sections come from a joint work with Nicolas Champagnat published in [12]. In Section 6.4, we state the same kind of limit theorems as those for N_t stated in Section 5.2. The following sections are devoted to the proofs of these results. Section 6.5 gives a new proof of the already known law of large numbers for the frequency spectrum (originally obtained in [13]). Sections 6.6, 6.7 and 6.8 give the proof of the various CLT stated in Section 6.4.

6.2 Splitting trees with neutral Poissonian mutations

Here we define what we call a splitting tree with neutral mutation. Since a splitting tree $\mathbb{T}$ is a measured space (with a σ-finite measure λ), one can define on $\mathbb{T}$ a Poisson random measure with intensity λ. Hereafter we call mutations every atoms of this measure on $\mathbb{T}$. However, since the only observable mutation at time t are the one which occurred on the lineage of the individuals alive at this time, we can define the occurrence of mutations directly on the CPP. So, let $\mathcal{P}$ be a Poisson random measure on $(0,t)\times\mathbb{N}$ with intensity measure $\theta\lambda\otimes C$ where λ is the Lebesgue measure on $(0,t)$ and C is the counting measure on $\mathbb{N}$. The mutation random measure on the CPP is then defined by

$$\mathcal{N}\left(da,di\right) = \mathbb{1}_{H_i > t-a} \mathbb{1}_{i<N_t} \mathcal{P}\left(di,da\right), \tag{6.1}$$

where an atom at (a,i) means that the ith branch of the CPP experiences a mutation at time $t-a$.

We assume that each mutation gives a totally new type to its holder (infinitly-many alleles model) and that the types are transmitted to offspring. This rule yields a partition of the population by type at a given time t. The distribution of the frequency of types in the population is called the frequency spectrum and is defined as the sequence $(A(k,t))_{k\geq 1}$ where $A(k,t)$ is the number of types carried by exactly k individuals in the alive population at time t (or, for short, the number of families of size k at this time) excluding the family holding the original type of the root. In the study of the frequency spectrum, an important role is played by the family carrying the type of the root.The type of the ancestor individual at time 0 is said *clonal*. Moreover, at any time t, the set of individuals carrying this type is called the clonal family. We denote by $Z_0(t)$ the size of the clonal family at time t.

To study this family it is easier to consider the clonal splitting tree constructed from the original splitting tree by cutting every branches beyond mutations. This clonal splitting tree is a standard splitting tree without mutations where individuals are killed as soon as they die or experience a mutation. The new lifespan law $\mathbb{P}_{V_\theta}$ is then the minimum between an exponential random variable of parameter θ and an independent copy of V. As a splitting tree, one can study its contour process whose Laplace exponent is given, using simple manipulations on Laplace transforms, by

$$\psi_\theta(x) = x - \int_{(0,\infty]} \left(1 - e^{-rx}\right) b\mathbb{P}_{V_\theta}(dr) = \frac{x\psi(x+\theta)}{x+\theta}.$$

In the case where $\alpha - \theta > 0$ (resp. $\alpha - \theta < 0$, $\alpha - \theta = 0$) the clonal population is supercritical (resp. sub-critical, critical), and we talk about clonal supercritical (resp. sub-critical, critical) case.
We denote by W_θ the scale function of the Lévy process induced by this new tree, related to ψ_θ as in (3.1). This leads to

$$\mathbb{P}\left(Z_0(t) = k \mid Z_0(t) > 0\right) = \frac{1}{W_\theta(t)} \left(1 - \frac{1}{W_\theta(t)}\right)^{k-1}.$$

Moreover, $\mathbb{E}\left[N_t\right]$ satisfies the renewal equation

$$f(t) = \mathbb{P}\left(V > t\right) + b\int_0^t f(t-s)\mathbb{P}\left(V > s\right) ds,$$

which, applied to the clonal splitting tree, allows obtaining after some easy calculations,

$$\frac{\mathbb{P}\left(Z_0(t) > 0\right)}{\mathbb{P}\left(N_t > 0\right)} = \frac{e^{-\theta t} W(t)}{W_\theta(t)},$$

from which one can deduce

$$\mathbb{P}\left(Z_0(t) = k \mid N_t > 0\right) = \frac{e^{-\theta t} W(t)}{W_\theta(t)^2} \left(1 - \frac{1}{W_\theta(t)}\right)^{k-1}, \ \forall k \geq 1, \tag{6.2}$$

and

$$\mathbb{P}\left(Z_0(t)=0 \mid N_t>0\right)=1-\frac{e^{-\theta t}W(t)}{W_\theta(t)}.$$

The main idea underlying our study is that the behaviour of any family in the CPP is the same as the clonal one but on a smaller time scale.
For the rest of this chapter, unless otherwise stated, the notation $\mathbb{P}_t$ refers to $\mathbb{P}\left(\cdot \mid N_t>0\right)$ and $\mathbb{P}_\infty$ refers to the probability measure conditioned on the non-extinction event, denoted *Non-Ex* in the sequel.
Finally, we recall the asymptotic behaviour of the scale functions $W(t)$ and $W_\theta(t)$, which is widely used in the sequel.

Lemma 6.2.1. *(Champagnat-Lambert [14]) Assume $\alpha>0$, there exists a positive constant γ such that*

$$e^{-\alpha t}\psi'(\alpha)W(t)-1=\mathcal{O}\left(e^{-\gamma t}\right).$$

In the case that $\theta<\alpha$ (clonal supercritical case),

$$W_\theta(t) \underset{t\to\infty}{\sim} \psi_\theta'(\alpha-\theta)^{-1}e^{(\alpha-\theta)t}.$$

In the case that $\theta>\alpha$ (clonal sub-critical case),

$$W_\theta(t)=\frac{\theta}{\psi(\theta)}+\mathcal{O}\left(e^{-(\theta-\alpha)t}\right).$$

In the case where $\theta=\alpha$ (clonal critical case),

$$W_\alpha(t) \underset{t\to\infty}{\sim} \frac{\alpha t}{\psi'(\alpha)}.$$

From this lemma, one can obtain that the probability that the clonal family reaches a fixed size at time t decreases exponentially fast with t.

Corollary 6.2.2. *In the supercritical case ($\alpha>0$), for any positive integer k,*

$$\mathbb{P}_t\left(Z_0(t)=k\right)=\mathcal{O}\left(e^{-\delta t}\right),$$

where δ is equal to θ (resp. $2\alpha-\theta$) in the clonal critical and sub-critical cases (resp. supercritical case).

Remark 6.2.3. *Note that Lemma 6.2.1 implies in particular that, for any positive integer k,*

$$tW(t)^{k-1}=o\left(W(t)^k\right).$$

6.3 Moments formulas of the frequency spectrum

For two positive real numbers $a<t$, we denote by $N_{t-a}^{(t)}$ the number of individuals alive at time $t-a$ who have descent alive at time t. In the CPP of the individuals alive at time t, $N_{t-a}^{(t)}$ corresponds to the number of branches higher than $t-a$, that is $\sharp\left\{H_i \mid i\in\{0,\dots,N_t-1\},\ H_i>t-a\right\}$.

In the sequel, we use the following notation for multi-indexed sums : let K, N be two positive integers and $\ell_1, \ldots, \ell_K$ some non-negative integers, then the notation

$$\sum_{n_1^{1:K}+\cdots+n_N^{1:K}=\ell_{1:K}}$$

refers to the sum

$$\sum_{\substack{n_1^1+\cdots+n_N^1=\ell_1 \\ \cdots \\ n_1^K+\cdots+n_N^K=\ell_K}} .$$

In order to lighten notation, we also use the convention that for any integer n and any negative integer k,

$$\binom{n}{k} = 0.$$

We recall that $\mathbb{P}_t$ is the conditional probability on the event $\{N_t > 0\}$ and that $\mathbb{E}_t$ is the corresponding expectation. In the both following theorem, we know, according to Proposition 4.3.1, that the random variable $N_{t-a}^{(t)}$ is geometrically distributed with parameter $\frac{W(t+a)}{W(a)}$ under $\mathbb{P}_t$. We can now state the main theorems of this section.

Theorem 6.3.1. *For any positive integers n and k, we have,*

$$\mathbb{E}_t\left[\binom{A(k,t)}{n}\right] = \mathbb{E}_t\left\{\int_0^t \theta N_{t-a}^{(t)} \sum_{n_1+\cdots+n_{N_{t-a}^{(t)}}=n-1} \mathbb{E}_a\left[\binom{A(k,a)}{n_1}\mathbb{1}_{Z_0(a)=k}\right] \prod_{m=2}^{N_{t-a}^{(t)}} \mathbb{E}_a\left[\binom{A(k,a)}{n_m}\right] da\right\}.$$

We also have a similar result for the joint moments of the frequency spectrum.

Theorem 6.3.2. *Let $n_1, \ldots, n_N$ and $k_1, \ldots, k_N$ be positive integers. We have*

$$\begin{aligned}\mathbb{E}_t\left[\prod_{i=1}^N \binom{A(k_i,t)}{n_i}\right] &= \sum_{l=1}^N \mathbb{E}_t\Bigg\{\int_0^t \theta N_{t-a}^{(t)} \sum_{n_1^{1:N}+\cdots+n_{N_{t-a}^{(t)}}^{1:N}=n_{1:N}-\delta_{1:N,l}} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A(k_i,a)}{n_1^i}\mathbb{1}_{Z_0(a)=k_l}\right] \\ &\qquad \times \prod_{m=2}^{N_{t-a}^{(t)}} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A(k_i,a)}{n_m^i}\right] da\Bigg\},\end{aligned} \tag{6.3}$$

where δ refers to the Kronecker symbol.

In Subsection 6.3.3, we also give formulas for moments like $\mathbb{E}_t\left[\binom{A(k,t)}{n}\mathbb{1}_{Z_0(t)=\ell}\right]$.

6.3.1 An example : the expectation of $A(k,t)$

Before going further, we point out that this section uses the recursive construction of the CPP given in Section 4.3. A nice application of this construction is the derivation of the expectation of $A(k,t)$. Indeed, suppose that a mutation occurs on branch i at a time a. Then, by construction of the CPP, the future of this family depends only on what happens on the branches $(H_j, i \le j < \tau)$ (see Figure 6.1), where

$$\tau = \inf \{ j > i \mid \ H_j \ge a \} .$$

In fact, this set of branches is also a CPP with scale function W stopped at a (we talk about sub-CPP), and the number of individuals carrying the mutation at time t is the number of clonal individuals in this sub-CPP. We recall that this expectation was first calculated in [13], with a

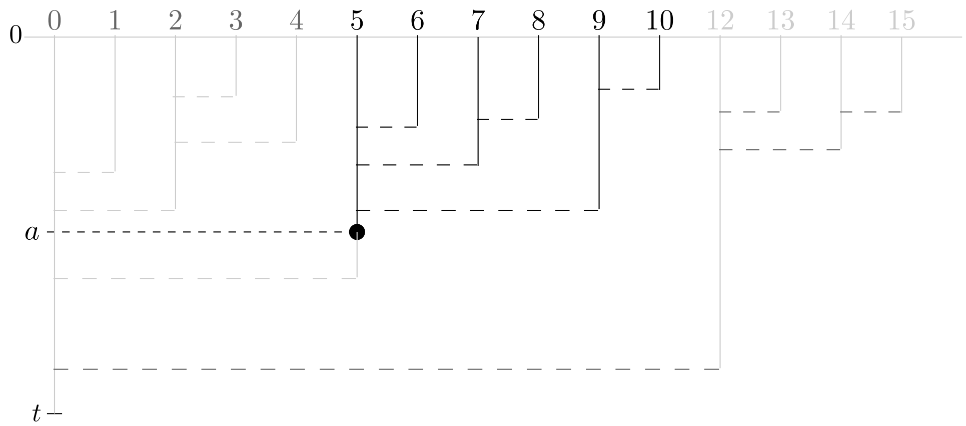

FIGURE 6.1 – The future of a mutation only depends on a sub-tree of the genealogical tree.

much more complicated proof.

Theorem 6.3.3. *([13, Cor. 3.4]) For any positive integer k, we have*

$$\mathbb{E}_t \left[A(k,t) \right] = W(t) \int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2} \left(1 - \frac{1}{W_\theta(a)} \right)^{k-1} da.$$

Proof. Since $A(k,t)$ is the number of types represented at time t by k individuals, it is equivalent to enumerate all the mutations and ask if they have exactly k clonal children at time t. This remark leads to the following integral representation of $A(k,t)$:

$$A\left(k,t\right) = \int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \, \mathcal{N}\left(da, di\right), \tag{6.4}$$

where $\mathcal{N}$ is defined in (6.1), and $Z_0^i(a)$ denotes the number of alive individuals at time t carrying the same type as the type carried at time $t-a$ on the ith branch of the CPP of the individuals alive at time t (the notation comes from the fact that $Z_0^i(a)$ corresponds to the size of the clonal family in the sub-CPP induced by the ith individual at time $t-a$, see Figure 4.1). From

Proposition 4.3.1, it follows that $\mathbb{1}_{Z_0^i(a)=k}$ satisfies the conditions of Theorem 4.2.2, so

$$\mathbb{E}_t\left[A(k,t)\right] = \int_0^t \theta\, \mathbb{P}_a\left(Z_0(a)=k\right) \mathbb{E}_t N_{t-a}^{(t)}\, da = W(t) \int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2} \left(1 - \frac{1}{W_\theta(a)}\right)^{k-1} da,$$

using (6.2). □

6.3.2 Proof of Theorems 6.3.1 and 6.3.2

Let a and t be two positive real numbers such that $a < t$, and n a positive integer. We call k-mutation, a mutation represented by k alive individuals at time t in the splitting tree. Let $\left(A^{(i)}(k,a)\right)_{k\geq 1}$ be the frequency spectrum in the i-th subtree of construction provided by Proposition 4.3.1.
To count the number of n-tuples in the set of k-mutations, we look along the tree and seek for mutations in the CPP. For each k-mutation encountered, we count the number of $(n-1)$-tuples made of younger k-mutations. The $(n-1)$-tuples should be enumerated by decomposition in each subtree in order to exploit the independence property of the subtrees of Proposition 4.3.1. Suppose that a mutation is encountered at a time a, then the number of $(n-1)$-tuples made of younger mutations is given by

$$\sum_{n_1+\dots+n_{N_{t-a}^{(t)}}=n-1} \prod_{m=1}^{N_{t-a}^{(t)}} \binom{A^{(m)}(k,a)}{n_m}.$$

So the number $\binom{A(k,t)}{n}$ of n-tuples of k-mutations is given by

$$\binom{A(k,t)}{n} = \int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \sum_{n_1+\dots+n_{N_{t-a}^{(t)}}=n-1} \prod_{m=1}^{N_{t-a}^{(t)}} \binom{A^{(m)}(k,a)}{n_m} \mathcal{N}(da,di), \tag{6.5}$$

$$= \sum_{\ell\geq 1} \int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \sum_{n_1+\dots+n_l=n-1} \prod_{m=1}^{\ell} \binom{A^{(m)}(k,a)}{n_m} \mathbb{1}_{N_{t-a}^{(t)}=\ell}\, \mathcal{N}(da,di),$$

where $Z_0^i(a)$ was defined in the proof of Theorem 6.3.3. Finally, using the independence provided by Proposition 4.3.1, it follows from Theorem 4.2.2 applied to all the integrals with respect to the random measures $\mathbb{1}_{N_{t-a}^{(t)}=k}\mathcal{N}(da,di)$, that

$$\mathbb{E}_t\left[\binom{A(k,t)}{n}\right]$$

$$= \mathbb{E}_t \int_{[0,t]\times\mathbb{N}} \sum_{n_1+\dots+n_{N_{t-a}^{(t)}}=n-1} \mathbb{E}_a\left[\binom{A(k,a)}{n_1} \mathbb{1}_{Z_0(a)=k}\right] \prod_{m=2}^{N_{t-a}^{(t)}} \mathbb{E}_a\left[\binom{A(k,a)}{n_m}\right] \mathcal{N}\left(da,di\right).$$

Finally, using that the $\mathcal{N}(da,di) = \mathbb{1}_{H_i>t-a}\mathbb{1}_{i<N_t}\mathcal{P}(di,da)$ where $\mathcal{P}$ independent from the CPP (and, hence, from $N_{t-a}^{(t)}$), it follows that

$$\begin{aligned}\mathbb{E}_t\left[\binom{A(k,t)}{n}\right] &= \mathbb{E}_t\int_{[0,t]}\sum_{n_1+\dots+n_{N_{t-a}^{(t)}}=n-1}\mathbb{E}_a\left[\binom{A(k,a)}{n_1}\mathbb{1}_{Z_0(a)=k}\right] \\ &\quad\times\prod_{m=2}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\binom{A(k,a)}{n_m}\right]\int_{\mathbb{N}}\mathbb{1}_{H_i>t-a}\mathbb{1}_{i<N_t}\ C(di)\theta da, \\ &= \mathbb{E}_t\int_{[0,t]}\theta N_{t-a}^{(t)}\sum_{n_1+\dots+n_{N_{t-a}^{(t)}}=n-1}\mathbb{E}_a\left[\binom{A(k,a)}{n_1}\mathbb{1}_{Z_0(a)=k}\right]\prod_{m=2}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\binom{A(k,a)}{n_m}\right]da, \end{aligned} \tag{6.6}$$

which ends the proof of Theorem 6.3.1.
The proof of Theorem 6.3.2 follows exactly the same lines, and we leave it to the reader.

6.3.3 Joint moments of the frequency spectrum and $\mathbb{1}_{Z_0(t)=\ell}$

In order to compute the terms of the form

$$\mathbb{E}_t\left[\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)=\ell}\right]$$

involved in (6.3), we need to extend the representation (6.5) of $\binom{A(k,t)}{n}$ to take into account the indicator function of $\{Z_0(t)=\ell\}$. To do this, when integrating w.r.t. $\mathcal{N}(da,di)$, we need to ask that the sum of the number of clonal individuals in each subtree for which the type at time $t-a$ is the ancestral type, is equal to k. We begin with the case

$$\mathbb{E}\left[A(k,t)\mathbb{1}_{Z_0(t)=\ell}\right]$$

in order to highlight the ideas. In this case, we have the following result.

Proposition 6.3.4.

$$\begin{aligned}&\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(t)=\ell}\right] \\ =&\mathbb{E}_t\int_0^t\left(N_{t-a}^{(t)}-Z_0^{(t)}(a)\right)\mathbb{P}_a\left(Z_0(a)=k\right)\sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)}=\ell}\ \prod_{i=1}^{Z_0^{(t)}(a)}\mathbb{P}_a\left(Z_0(a)=\ell_i\right)\ \theta da \\ &+\mathbb{E}_t\int_0^t Z_0^{(t)}(a)\mathbb{P}_a\left(Z_0(a)=k\right)\sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-1}=\ell}\ \prod_{i=1}^{Z_0^{(t)}(a)-1}\mathbb{P}_a\left(Z_0(a)=\ell_i\right)\ \theta da.\end{aligned} \tag{6.7}$$

Proof. Recalling that $N_{t-a}^{(t)}$ refers to the size whole population in the lower tree $\hat{\mathcal{P}}$ of the construction of Proposition 4.3.1, we similarly define $Z_0^{(t)}(a)$ as the size of the clonal population in the same tree (with the convention that mutations that occur at time $t-a$, i.e. on the leaves of the tree $\hat{\mathcal{P}}$, do not affect $Z_0^{(t)}(a)$). It follows that

$$A(k,t)\mathbb{1}_{Z_0(t)=\ell}$$
$$= \int_{[0,t]\times\mathbb{N}} \frac{\mathbb{1}_{Z_0^j(a)=k}}{\left(Z_0^{(t)}(a)-B_j\right)!} \sum_{\sigma\in I} \mathbb{1}_{\sigma \text{ is ancestral}} \sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell} \prod_{i=1}^{Z_0^{(t)}(a)-B_j} \mathbb{1}_{Z_0^{\sigma_i}(a)=\ell_i}\, \mathcal{N}(da,dj), \tag{6.8}$$

where I is the set of injections from $\left\{1,\dots,Z_0^{(t)}(a)-B_j\right\}$ to $\left\{1,\dots,N_{t-a}^{(t)}\right\}$, B_j is the indicator function of the event

$$\{\text{the } j\text{th individual at time } t-a \text{ is clonal}\},$$

and "σ is ancestral" denotes the event that the individuals $\sigma_1,\dots,\sigma_{Z_0^{(t)}(a)-B_j}$ at time $t-a$ have the ancestral type. Now, using the same method as in the proof of Theorem 6.3.1 leads to

$$\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(t)=\ell}\right]$$
$$= \mathbb{E}_t \int_{[0,t]\times\mathbb{N}} \mathbb{P}_a\left(Z_0(a)=k\right) \sum_{\sigma\in I} \mathbb{1}_{\sigma \text{ is ancestral}} \sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell} \prod_{i=1}^{Z_0^{(t)}(a)-B_j} \mathbb{P}_a\left(Z_0(a)=\ell_i\right) \frac{\mathcal{N}(da,dj)}{\left(Z_0^{(t)}(a)-B_j\right)!}$$
$$= \mathbb{E}_t \int_{[0,t]\times\mathbb{N}} \mathbb{P}_a\left(Z_0(a)=k\right) \sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell} \prod_{i=1}^{Z_0^{(t)}(a)-B_j} \mathbb{P}_a\left(Z_0(a)=\ell_i\right) \mathcal{N}(da,dj)$$
$$= \mathbb{E}_t \int_{[0,t]\times\mathbb{N}} \mathbb{P}_a\left(Z_0(a)=k\right) \sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell} \prod_{i=1}^{Z_0^{(t)}(a)-B_j} \mathbb{P}_a\left(Z_0(a)=\ell_i\right) \mathbb{1}_{H_j>t-a}\mathbb{1}_{j<N_t}\mathcal{P}(da,dj).$$

Now, $Z_0^{(t)}(a)$ is not independent from $\mathcal{P}$, but we have that $Z_0^{(t)}(a)$ is independent from $\mathcal{P}\left([a,T]\cap\cdot\right)$ for all $a<T$. Hence, Theorem 4.2.2 applies to $\widetilde{X}_a := Z_0^{(t)}(t-a)$ and $\widetilde{\mathcal{P}}$ defined for all measurable set $A\subset[0,t]$ by

$$\widetilde{\mathcal{P}}(A) = \mathcal{P}(t-A),$$

and, as in (6.6),

$$\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(t)=\ell}\right]$$
$$= \mathbb{E}_t \int_{[0,t]\times\mathbb{N}} \mathbb{P}_a\left(Z_0(a)=k\right) \sum_{\ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell} \prod_{i=1}^{Z_0^{(t)}(a)-B_j} \mathbb{P}_a\left(Z_0(a)=\ell_i\right) \mathbb{1}_{H_j>t-a}\mathbb{1}_{j<N_t}\theta da\, C(dj).$$

Finally, integrating with respect to $C(dj)$ leads to the result. □

This last proposition in not exactly a closed formula since its involves the law of the couple $(N_{t-a}^{(t)}, Z_0^{(t)}(a))$. To close the formula, we need an explicit formula for the joint generating function of $N_{t-a}^{(t)}$ and $Z_0^{(t)}(a)$. Let

$$F(u,v) = \mathbb{E}_t\left[u^{N_{t-a}^{(t)}} v^{Z_0^{(t)}(a)}\right],\ u,v \in [0,1],$$

which is given, thanks to Proposition 4.1 of [13], by

$$F(u,v) = u\frac{\hat{W}(t-a,u)}{\hat{W}(t-a)}\left(1 - \frac{e^{-\theta(t-a)}\hat{W}(t-a,u)}{\frac{v}{1-v} + \hat{W}_\theta(t-a,u)}\right), \tag{6.9}$$

where $\hat{W}$ is the scale function of the lower CPP, $\hat{\mathcal{P}}$, defined in Proposition 4.3.1,

$$\hat{W}(t,u) := \frac{\hat{W}(t)}{\hat{W}(t) - u\left(\hat{W}(t) - 1\right)},$$

and

$$\hat{W}_\theta(t,u) := e^{-\theta t}\hat{W}(t,u) + \theta\int_0^t \hat{W}(s,u)e^{-\theta s}\ ds.$$

Proposition 6.3.5. *For all $k \geq 1$ and $l \geq 0$,*

$$\begin{aligned}
&\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(t)=\ell}\right] \\
&= \int_0^t \mathbb{P}_a\left(Z_0(a)=k\right)\sum_{j=1}^{l}\binom{l-1}{j-1}\frac{1}{j!}\left(1-\frac{1}{W_\theta(a)}\right)^{l-j}\left(\frac{e^{-\theta a}W(a)}{W_\theta(a)^2\mathbb{P}\left(Z_0(a)=0\right)}\right)^j H_j\left(1, 1-\frac{e^{-\theta a}W(a)}{W_\theta(a)}\right)\theta da \\
&\quad + \int_0^t \mathbb{P}_a\left(Z_0(a)=k\right)\sum_{j=1}^{l}\binom{l-1}{j-1}\frac{1}{j!}\left(1-\frac{1}{W_\theta(a)}\right)^{l-j}\left(\frac{e^{-\theta a}W(a)}{W_\theta(a)^2\mathbb{P}\left(Z_0(a)=0\right)}\right)^j G_j\left(1-\frac{e^{-\theta a}W(a)}{W_\theta(a)}\right)\theta da,
\end{aligned}$$

where

$$H_j(u,v) := v^j\partial_v^j\partial_u uF(u,v) - v^{j+1}\partial_v^{j+1}\left\{v\mathbb{E}\left[v^{Z_0^{(t)}(a)}\right]\right\},$$

and

$$G_j := v^{j-1}\partial_v^j\mathbb{E}_t\left[v^{Z_0^{(t)}(a)}\right].$$

Proof. Let A_1 and A_2 denote the two terms of the r.h.s. of (6.7). We detail the computations of A_1. The case A_2 is similar.

$$\begin{aligned}
A_1 = \mathbb{E}_t\int_0^t &\left(N_{t-a}^{(t)} - Z_0^{(t)}(a)\right)\mathbb{P}_a\left(Z_0(a)=k\right) \\
&\times \sum_{j=1}^{Z_0^{(t)}(a)\wedge l}\binom{Z_0^{(t)}(a)}{j}\sum_{\substack{\ell_1+\dots+\ell_j=\ell \\ \ell_j>0}}\prod_{i=1}^{j}\mathbb{P}_a\left(Z_0(a)=\ell_i\right)\mathbb{P}_a\left(Z_0(a)=0\right)^{Z_0(a)-j}\ \theta da.
\end{aligned}$$

Since, from (6.2),

$$\prod_{i=1}^{j}\mathbb{P}_a\left(Z_0(a)=\ell_i\right) = \prod_{i=1}^{j}\frac{e^{-\theta a}W(a)}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{\ell_i-1} = \left(\frac{e^{-\theta a}W(a)}{W_\theta(a)^2}\right)^j\left(1-\frac{1}{W_\theta(a)}\right)^{l-j},$$

we get

$$
\begin{aligned}
A_1 =& \mathbb{E}_t \int_0^t \left(N_{t-a}^{(t)} - Z_0^{(t)}(a)\right) \mathbb{P}_a\left(Z_0(a) = k\right) \sum_{j=1}^{Z_0^{(t)}(a) \wedge l} \binom{Z_0^{(t)}(a)}{j} \binom{l-1}{j-1} \\
& \times \left(\frac{e^{-\theta a} W(a)}{W_\theta(a)^2}\right)^j \left(1 - \frac{1}{W_\theta(a)}\right)^{l-j} \mathbb{P}_a\left(Z_0(a) = 0\right)^{Z_0(a)-j} \ \theta da \\
=& \int_0^t \mathbb{P}_a\left(Z_0(a) = k\right) \sum_{j=1}^{l} \binom{l-1}{j-1} \frac{1}{j!} \left(1 - \frac{1}{W_\theta(a)}\right)^{l-j} \left(\frac{e^{-\theta a} W(a)}{W_\theta(a)^2 \mathbb{P}_a\left(Z_0(a) = 0\right)}\right)^j \\
& \times \mathbb{E}_t\left[\left(N_{t-a}^{(t)} - Z_0^{(t)}(a)\right)\left(Z_0^{(t)}(a)\right)_{(j)} \mathbb{P}_a\left(Z_0(a) = 0\right)^{Z_0^{(t)}(a)}\right] \ \theta da
\end{aligned}
$$

Finally, if we define, for all integer j,

$$
H_j(u,v) := v^j \partial_v^j \partial_u u F(u,v) - v^{j+1} \partial_v^{j+1} \left\{ v \mathbb{E}\left[v^{Z_0^{(t)}(a)}\right]\right\},
$$

and

$$
G_j := v^{j-1} \partial_v^j \mathbb{E}_t\left[v^{Z_0^{(t)}(a)}\right],
$$

we get

$$
\begin{aligned}
& \mathbb{E}_t\left[A(k,t) \mathbb{1}_{Z_0(t)=\ell}\right] \\
=& \int_0^t \mathbb{P}_a\left(Z_0(a) = k\right) \sum_{j=1}^{l} \binom{l-1}{j-1} \frac{1}{j!} \left(1 - \frac{1}{W_\theta(a)}\right)^{l-j} \left(\frac{e^{-\theta a} W(a)}{W_\theta(a)^2 \mathbb{P}\left(Z_0(a) = 0\right)}\right)^j H_j\left(1, 1 - \frac{e^{-\theta a} W(a)}{W_\theta(a)}\right) \ \theta da \\
&+ \int_0^t \mathbb{P}_a\left(Z_0(a) = k\right) \sum_{j=1}^{l} \binom{l-1}{j-1} \frac{1}{j!} \left(1 - \frac{1}{W_\theta(a)}\right)^{l-j} \left(\frac{e^{-\theta a} W(a)}{W_\theta(a)^2 \mathbb{P}\left(Z_0(a) = 0\right)}\right)^j G_j\left(1 - \frac{e^{-\theta a} W(a)}{W_\theta(a)}\right) \theta da.
\end{aligned}
$$

□

These ideas also lead to the following formula, which is proved similarly.

Corollary 6.3.6. *Let $n_1, \ldots, n_N$ and $k_1, \ldots, k_N$ be positive integers. Let ℓ be a positive integer. We have*

$$
\begin{aligned}
&\mathbb{E}_t\left[\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)=\ell}\right]\\
&=\sum_{\kappa=1}^{N}\mathbb{E}_t\int_{[0,t]}\left(N_{t-a}^{(t)}-Z_0^{(t)}(a)\right)\sum_{\substack{n_1^{1:N}+\cdots+n^{1:N}_{N_{t-a}^{(t)}}=n_{1:N}-\delta_{1:N,l}\\ \ell_2+\cdots+\ell_{Z_0^{(t)}(a)+1}=\ell}}\prod_{m=Z_0^{(t)}(a)+2}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\right]\\
&\quad\times\prod_{m=2}^{Z_0^{(t)}(a)+1}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\mathbb{1}_{Z_0(a)=\ell_m}\right]\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\mathbb{1}_{Z_0(a)=k_\kappa}\right]\ \theta da\\
&\quad+\sum_{\kappa=1}^{N}\mathbb{E}_t\int_{[0,t]}Z_0^{(t)}(a)\sum_{\substack{n_1^{1:N}+\cdots+n^{1:N}_{N_{t-a}^{(t)}}=n_{1:N}-\delta_{1:N,l}\\ \ell_2+\cdots+\ell_{Z_0^{(t)}(a)+1}=\ell}}\prod_{m=Z_0^{(t)}(a)+1}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\right]\\
&\quad\times\prod_{m=2}^{Z_0^{(t)}(a)}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\mathbb{1}_{Z_0(a)=\ell_{m_2}}\right]\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\mathbb{1}_{Z_0(a)=k_\kappa}\right]\ \theta da.
\end{aligned}
$$

Proof. According to Section 6.3.2, we have the following integral representation.

$$
\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i}=\sum_{l=1}^{N}\int_{[0,t]\times\mathbb{N}}\mathbb{1}_{Z_0^j(a)=k_l}\sum_{n_1^{1:N}+\cdots+n^{1:N}_{N_{t-a}^{(t)}}=n_{1:N}-\delta_{1:N,l}}\prod_{m=1}^{N_{t-a}^{(t)}}\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^j}\mathcal{N}(da,dj).
$$

Now, using this equation in conjunction with the decomposition of $\mathbb{1}_{Z_0(t)=\ell}$ used in Section 6.3.3, we have

$$
\begin{aligned}
\prod_{i=1}^{N}\binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)=\ell}&=\sum_{l=1}^{N}\int_{[0,t]\times\mathbb{N}}\mathbb{1}_{Z_0^j(a)=k_l}\sum_{\sigma\in I}\mathbb{1}_{\sigma\text{ is ancestral}}\\
\times\sum_{\substack{n_1^{1:N}+\cdots+n^{1:N}_{N_{t-a}^{(t)}}=n_{1:N}-\delta_{1:N,l}\\ \ell_1+\cdots+\ell_{Z_0^{(t)}(a)-B_j}=\ell}}&\prod_{m_1=1}^{N_{t-a}^{(t)}}\prod_{i=1}^{N}\binom{A^{m_1}(k_i,a)}{n_{m_1}^i}\prod_{m_2=1}^{Z_0^{(t)}(a)-B_j}\mathbb{1}_{Z_0^{\sigma m_2}(a)=\ell_{m_2}}\frac{\mathcal{N}(da,dj)}{\left(Z_0^{(t)}(a)-B_j\right)!}.
\end{aligned}
$$

We refer the reader to the proof of Proposition 6.3.4 for the definitions of I,B_j, and the event $\{\sigma$ is ancestral$\}$. The definitions of $A^{(m)}(k,a)$ and $Z_0^{(m)}(a)$ can be found in the beginning of this section.

Now, we take the expectation in the last equality. Thanks to the method used in the proof of Proposition 6.3.4, we have

$$\mathbb{E}_t\left[\prod_{i=1}^N \binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)=\ell}\right]$$

$$= \sum_{\kappa=1}^N \mathbb{E}_t\Bigg\{ \int_{[0,t]\times\mathbb{N}} \sum_{\sigma\in I} \mathbb{1}_{\sigma \text{ is ancestral}} \sum_{\substack{n_1^{1:N}+\dots+n^{1:N}_{N^{(t)}_{t-a}}=n_{1:N}-\delta_{1:N,l}\\ \ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell}} \prod_{\substack{m_1=1\\ m_1\neq\sigma, m_1\neq i}}^{N^{(t)}_{t-a}} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A^{m_1}(k_i,a)}{n^i_{m_1}}\right]$$

$$\times \prod_{m_2=1}^{Z_0^{(t)}(a)-B_j} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A^{\sigma_{m_2}}(k_i,a)}{n^i_{\sigma_{m_2}}}\mathbb{1}_{Z_0^{\sigma_{m_2}}(a)=\ell_{m_2}}\right] \mathbb{E}_a\left[\prod_{i=1}^N \binom{A^i(k_j,a)}{n_i^j}\mathbb{1}_{Z_0^j(a)=k_\kappa}\right] \frac{\mathcal{N}(da,dj)}{\left(Z_0^{(t)}(a)-B_j\right)!}\Bigg\},$$

where $m_1 \neq \sigma$ means that $m_1 \notin \sigma\left(\left\{1,\dots,Z_0^{(t)(a)}-B_j\right\}\right)$. Now, following, as above, we get

$$\mathbb{E}_t\left[\prod_{i=1}^N \binom{A(k_i,t)}{n_i}\mathbb{1}_{Z_0(t)=\ell}\right]$$

$$= \sum_{\kappa=1}^N \mathbb{E}_t\Bigg\{ \int_{[0,t]\times\mathbb{N}} \sum_{\sigma\in I} \mathbb{1}_{\sigma \text{ is ancestral}} \sum_{\substack{n_1^{1:N}+\dots+n^{1:N}_{N^{(t)}_{t-a}}=n_{1:N}-\delta_{1:N,l}\\ \ell_1+\dots+\ell_{Z_0^{(t)}(a)-B_j}=\ell}} \prod_{m_1=Z_0^{(t)}(a)-B_i+1}^{N^{(t)}_{t-a}} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A(k_i,a)}{n^i_{m_1}}\right]$$

$$\times \prod_{m_2=2}^{Z_0^{(t)}(a)-B_i+1} \mathbb{E}_a\left[\prod_{i=1}^N \binom{A(k_i,a)}{n^i_{m_2}}\mathbb{1}_{Z_0(a)=\ell_{m_2}}\right] \mathbb{E}_a\left[\prod_{i=1}^N \binom{A(k_1,a)}{n_i^1}\mathbb{1}_{Z_0(a)=k_\kappa}\right] \frac{\mathbb{1}_{H_i>t-a}\mathbb{1}_{j<N_t}\theta da\ C(di)}{\left(Z_0^{(t)}(a)-B_j\right)!}\Bigg\}.$$

Then, the sum with σ can be removed since there is no term depending on σ. Finally, integrating with respect to $C(di)$ leads to the result. □

Together with Theorems 6.3.1 and 6.3.2 and using the joint law of $N^{(t)}_{t-a}$ and $Z_0^{(t)}(a)$ given in (6.9), these formulas give explicit recursion to compute each factorial moment of the frequency spectrum.

Remark 6.3.7. *Although, these formulas are quite heavy, an important interest lies in the method used to compute them. Indeed, this method should work to obtain the joint moments of $A(k,t)$ with any quantity which can be expressed, at any time a, as the sum of contributions of each subtrees. For instance, since*

$$N_t = \sum_{i=1}^{N^{(t)}_{t-a}} N_a^i, \quad \forall a\in[0,t],$$

where N_a^i is the number of individuals of the i-th subtrees at time a, we are able to compute the joint moments of N_t and $(A(k,t))_{k\geq 1}$. For example, using the integral representation (6.4) of

$A(k,t)$ and following the proof of Theorem 6.3.3 , we have that

$$
\begin{aligned}
&\mathbb{E}_t\left[A(k,t)N_t\right]\\
&=\mathbb{E}_t\int_{[0,t]\times\mathbb{N}}\sum_{j=1}^{N_{t-a}^{(t)}} N_a^j \mathbb{1}_{Z_0^{(i)}(a)=k}\mathcal{N}\left(da,di\right)\\
&=\int_{[0,t]}\theta\mathbb{E}_t\left[N_{t-a}^{(t)}\left(N_{t-a}^{(t)}-1\right)\right]\mathbb{E}_a\left[N_a\right]\mathbb{P}_a\left(Z_0(a)=k\right)da+\int_{[0,t]}\theta\mathbb{E}_t\left[N_{t-a}^{(t)}\right]\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]\theta da\\
&=\int_{[0,t]}W(t)^2\left(1-\frac{W(a)}{W(t)}\right)\frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1}da+W(t)\int_{[0,t]}\theta\frac{\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]}{W(a)}\theta da.
\end{aligned}
\tag{6.10}
$$

6.3.4 Application to the computation of the covariances of the frequency spectrum

A quantity of particular interest is the limit covariance between two terms of the frequency spectrum.

Proposition 6.3.8. *Suppose that $\alpha > 0$. Let k and l two positive integers, then,*

$$
Cov_t\left(A\left(k,t\right),A\left(l,t\right)\right)=W(t)^2c_kc_l+o\left(W(t)^2\right),
$$

where

$$
c_k:=\int_0^\infty\frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1}ds,\ \forall k\in\mathbb{N}\backslash\{0\}.
$$

Proof. In order to show how quantities in Theorem 6.3.2 can be manipulated, we detail the proof. Using Theorem 6.3.2, we obtain

$$
\begin{aligned}
\mathbb{E}_t\left[A(k,t)A(l,t)\right]=&\int_0^t\theta\mathbb{E}_t\left[N_{t-a}^{(t)}\left(N_{t-a}^{(t)}-1\right)\right]\left(\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[A(l,a)\right]+\mathbb{P}_a\left(Z_0(a)=\ell\right)\mathbb{E}_a\left[A(k,a)\right]\right)da\\
&+\int_0^t\theta\mathbb{E}_tN_{t-a}^{(t)}\left(\mathbb{E}_a\left[A(l,a)\mathbb{1}_{Z_0(a)=k}\right]+\mathbb{E}_a\left[A(k,a)\mathbb{1}_{Z_0(a)=\ell}\right]\right)da.
\end{aligned}
$$

Recalling, from Proposition 4.3.1, that $N_{t-a}^{(t)}$ is geometrically distributed with parameter $\frac{W(a)}{W(t)}$ under $\mathbb{P}_t$,

$$
\mathbb{E}_t\left[N_{t-a}^{(t)}\right]=\frac{W(t)}{W(a)}\quad\text{and}\quad\mathbb{E}_t\left[N_{t-a}^{(t)}\left(N_{t-a}^{(t)}-1\right)\right]=2\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right).
$$

Since

$$
\mathbb{E}\left[A(k,a)\mathbb{1}_{Z_0(a)=\ell}\right]\leq\mathbb{E}\left[A(k,a)\right]=\mathcal{O}(W(a)),
$$

it follows by Lemma 3.3.3 and Theorem 6.3.3, that

$$
\mathbb{E}_t\left[A(k,t)A(l,t)\right]=2\int_0^t\theta\frac{W(t)^2}{W(a)^2}\Big\{\mathbb{P}_a(Z_0(a)=\ell)\mathbb{E}_a\left[A(k,a)\right]+\mathbb{P}_a(Z_0(a)=k)\mathbb{E}_a\left[A(l,a)\right]\Big\}da+\mathcal{O}\left(tW(t)\right).
$$

By Theorem 6.3.3 and (6.2), the r.h.s. is equal to

$$\begin{aligned}
&2W(t)^2\int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(\left(1-\frac{1}{W_\theta(a)}\right)^{k-1}\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{l-1} ds\right.\\
&\left.+\left(1-\frac{1}{W_\theta(a)}\right)^{l-1}\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1} ds\right)da+\mathcal{O}\left(tW(t)\right)\\
=&2W(t)^2\int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1} da\int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{l-1} da+\mathcal{O}\left(tW(t)\right).
\end{aligned}$$

The proof ends thanks to Remark 6.2.3. □

6.3.5 Asymptotic behaviour of the moments of the frequency spectrum

In this part, we study the long time behaviour of the moments of the frequency spectrum. From this point and until the end of this chapter, we suppose that the tree is supercritical, that is $\alpha > 0$.

Proposition 6.3.9. *For any positive multi-integers n and k in $\mathbb{N}^N$,*

$$\mathbb{E}_t\left[\prod_{i=1}^N \binom{A(k_i,t)}{n_i}\right]=\frac{W(t)^{|n|}|n|!}{\prod_{i=1}^N n_i!}\prod_{i=1}^N c_{k_i}^{n_i}+\mathcal{O}\left(tW(t)^{|n|-1}\right), \tag{6.11}$$

where the c_{k_i}'s are as defined in Proposition 6.3.8.

Proof. **Step 1 : Preliminaries and ideas.**
The proposition is proved by induction.
Using the symmetry of the formula provided by Theorem 6.3.2, we may restrict to the study of the term $l = 1$ in (6.3). Hence, we want to study

$$\mathbb{E}_t\int_0^t \theta N_{t-a}^{(t)}\sum_{n_1^{1:N}+\cdots+n_{N_{t-a}^{(t)}}^{1:N}=n_{1:N}-\delta_{1:N,1}}\mathbb{E}_a\left[\prod_{i=1}^N\binom{A(k_i,t)}{n_1^i}\mathbb{1}_{Z_0(a)=k_1}\right]\prod_{m=2}^{N_{t-a}^{(t)}-1}\mathbb{E}_a\left[\prod_{i=1}^N\binom{A(k_i,a)}{n_m^i}\right]da. \tag{6.12}$$

We recall that the terms of the multi-sum in the above formula correspond to the ways of allocating the mutations in the subtrees. The analysis relies on the fact that the growth of each term depends on the repartition of the mutations. In particular, the main term correspond to the case where all mutations are allocated to different subtrees.
To capitalize on this fact, let $\mathcal{M}_{N_{t-a}^{(t)}}$ the subset of $\mathcal{M}_{(N_{t-a}^t-1)\times N}(\mathbb{N})$ (the space of matrices of size $\left(N_{t-a}^{(t)}-1\right)\times N$ with coefficients in $\mathbb{N}$), such that each $\mathbf{n}$ in $\mathcal{M}_{N_{t-a}^{(t)}}$ satisfies the relation

$$\sum_{m=1}^{N_{t-a}^{(t)}-1} n_m^i = n_i-\delta_{i,1},\quad \forall i\in\mathbb{N}. \tag{6.13}$$

The notations $\mathbf{n}_m$ and $\mathbf{n}^i$ refer to the multi-integers $\left(n_m^1, \ldots, n_m^N\right)$ and $\left(n_1^i, \ldots, n_{N_{t-a}^{(t)}}^i\right)$ respectively. To simplify the analysis, we highlight three cases of interest :

$$C_1 := \left\{\mathbf{n} \in \mathcal{M}_{N_{t-a}^{(t)}} \mid \forall i, n_1^i = 0,\ \forall i \geq 1, \forall m \geq 2,\ n_m^i \leq 1,\ \text{and} \left(n_m^i = 1 \Rightarrow n_m^k = 0, \forall k \neq i\right)\right\}.$$

This set corresponds to the case where all the mutations are taken in different subtrees and are not taken in the tree where a mutation just occurs. In fact, this corresponds to the dominant term of (6.12) because as $N_{t-a}^{(t)}$ tends to be large, the mutations tend to occur in different subtrees. Let also

$$C_2 := \left\{\mathbf{n} \in \mathcal{M}_{N_{t-a}^{(t)}} \mid \forall i, n_1^i = 0\right\} \backslash C_1.$$

Finally, let

$$C_3 := \left\{\mathbf{n} \in \mathcal{M}_{N_{t-a}^{(t)}} \mid \sum_{i=1}^{N} n_1^i > 0\right\}.$$

Step 2 : Uniform bound on the number of tuple of mutations in the subtrees.
Assuming that the relation of Lemma 6.3.9 is true for any multi-integer $n^\star$ such that $|n^\star| = |n|-1$, we have

$$\prod_{m=1}^{N_{t-a}^{(t)}} \mathbb{E}_a\left[\prod_{i=1}^{N} \binom{A(k_i, a)}{n_m^i}\right] = \prod_{m=1}^{N_{t-a}^{(t)}} \left(\frac{W(a)^{|\mathbf{n}_m|}|\mathbf{n}_m|!}{\prod_{i=1}^{N} n_m^i!} \prod_{i=1}^{N} c_{k_i}^{n_m^i} + \mathcal{O}\left(aW(a)^{|\mathbf{n}_m|-1}\right)\right). \tag{6.14}$$

$$\tag{6.15}$$

Since there are at most $|n|-1$ multi-integers $\mathbf{n}_m$ such that $|\mathbf{n}_m| > 0$ (because of the condition (6.13)), we can assume without loss of generality, up to reordering the indices, that $n_m^i = 0$, for all $m \geq |n|$, and so all the terms with $m > |n|$ in the product of (6.14) are equal to one. Hence,

$$\prod_{m=1}^{N_{t-a}^{(t)}-1} \mathbb{E}_a\left[\prod_{i=1}^{N} \binom{A(k_i, a)}{n_m^i}\right] \leq \mathcal{C}_{\mathbf{n}} W(a)^{|n|-1}, \tag{6.16}$$

for some constant $\mathcal{C}_{\mathbf{n}}$ depending only on the choice of $\mathbf{n}$ in $\mathcal{M}_{|n|}$.
Moreover, since $\mathcal{M}_{|n|}$ is finite, then

$$\prod_{m=1}^{N_{t-a}^{(t)}} \mathbb{E}_a\left[\prod_{i=1}^{N} \binom{A(k_i, a)}{n_m^i}\right] \leq \mathcal{C} W(a)^{|n|-1}. \tag{6.17}$$

Step 3 : Analysis of C_1.
For $\mathbf{n} \in C_1$, and in this case only, the product

$$\prod_{i=1}^{N} \binom{A(k_i, a)}{n_m^i}$$

has only one term different from 1, and it follows from Theorem 6.3.3, that

$$\prod_{m=1}^{N_{t-a}^{(t)}-1} \mathbb{E}\left[\prod_{i=1}^{N} \binom{A(k_i, a)}{n_m^i}\right] = W(a)^{|n|-1} \prod_{i=1}^{N} \left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2} \left(1 - \frac{1}{W_\theta(s)}\right)^{k_i} ds\right)^{n_i - \delta_{i,1}}.$$

The corresponding contribution in (6.12) is

$$I_1 := \int_0^t \theta W(a)^{|n|-1} \mathbb{P}_a\left(Z_0(t)=k_1\right) \prod_{i=1}^N \left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2} \left(1-\frac{1}{W_\theta(s)}\right)^{k_i} ds \right)^{n_i-\delta_{i,1}} \mathbb{E}_a \left[N_{t-a}^{(t)} \mathrm{Card}(C_1) \right] da.$$

Now, $\mathrm{Card}(C_1)$ is the number of way we can choose $|n|-1$ subtrees among the $N_{t-a}^{(t)}-1$ possible subtrees and choosing a way to allocate to each chosen subtree a mutation sizes $k_1,\ldots,k_N$, i.e.

$$\mathrm{Card}(C_1) = \binom{N_{t-a}^{(t)}-1}{|n|-1} \frac{(|n|-1)!}{\prod_{i=1}^N (n_i-\delta_{i,1})!}.$$

Finally,

$$I_1 = \int_0^t \theta W(a)^{|n|-1} \mathbb{P}_a\left(Z_0(t)=k_1\right) \prod_{i=1}^N \left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2} \left(1-\frac{1}{W_\theta(s)}\right)^{k_i} ds \right)^{n_i-\delta_{i,1}} \frac{\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(|n|)} \right]}{\prod_{i=1}^N (n_i-\delta_{i,1})!} da,$$

where $(x)_{(|n|)}$ is the falling factorial of order $|n|$. Since, $N_{t-a}^{(t)}$ is geometrically distributed under $\mathbb{P}_t$ with parameter $\frac{W(t)}{W(a)}$, it follows that

$$I_1 = \frac{|n|!W(t)^{|n|}}{\prod_{i=1}^N (n_i-\delta_{i,1})!} \int_0^t \theta \frac{e^{-\theta a}}{W_\theta(a)^2} \left(1-\frac{1}{W_\theta(a)}\right)^{k_1-1} \prod_{m=1}^N \left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2} \left(1-\frac{1}{W_\theta(s)}\right)^{k_i} ds \right)^{n_i-\delta_{i,1}} da$$
$$+ \mathcal{O}\left(tW(t)^{|n|-1}\right)$$

Step 4 : Analysis of C_2.
We denote

$$I_2 := \mathbb{E}_t \int_0^t N_{t-a}^{(t)} \sum_{\mathbf{n}\in C_2} \mathbb{P}_a\left(Z_0(a)=k_1\right) \prod_{m=1}^{N_{t-a}^{(t)}-1} \mathbb{E}_a \left[\prod_{i=1}^N \binom{A(k_i,a)}{n_m^i} \right] da. \tag{6.18}$$

Now, since

$$\mathrm{Card}(C_2) = \mathcal{O}\left(\left(N_{t-a}^{(t)}\right)^{|n|-2} \right),$$

we have using estimation (6.17),

$$\begin{aligned} I_2 &\le \int_0^t N_{t-a}^{(t)} \sum_{\mathbf{n}\in C_2} \mathcal{C} W(a)^{|n|-1} da \\ &\le \widetilde{\mathcal{C}} \int_0^t \left(N_{t-a}^{(t)}\right)^{|n|-1} W(a)^{|n|-1} da, \end{aligned}$$

for some positive real constant $\widetilde{\mathcal{C}}$. Using that $N_{t-a}^{(t)}$ is geometrically distributed with parameter $\frac{W(t)}{W(a)}$, it follows that there exists a positive real number $\hat{C}$ such that

$$I_2 \le \hat{C} \int_0^t \left(\frac{W(t)}{W(a)}\right)^{|n|-1} W(a)^{|n|-1} da.$$

Which imply that,

$$I_2 = \mathcal{O}\left(tW(t)^{|n|-1}\right).$$

Step 5 : Analysis of C_3.
In the case where there is a positive n_1^i (C_3 case), using that

$$\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\mathbb{1}_{Z_0(a)=k_l}\right] \le \mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\right],$$

we have,

$$\int_0^t N_{t-a}^{(t)}\sum_{\mathbf{n}\in C_3}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\mathbb{1}_{Z_0(a)=k_l}\right]\prod_{m=2}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\right]da,$$
$$\le \int_0^t N_{t-a}^{(t)}\sum_{\mathbf{n}\in C_3}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_1^i}\right]\prod_{m=2}^{N_{t-a}^{(t)}}\mathbb{E}_a\left[\prod_{i=1}^{N}\binom{A(k_i,a)}{n_m^i}\right]da,$$

which is very similar to the the other steps. This term is $\mathcal{O}\left(tW(t)^{|n|-1}\right)$ because the condition $\sum_i n_1^i > 0$ reduces the number of terms in the multi-sum. Indeed,

$$\begin{aligned}
\mathrm{Card}(C_3) &= \sum_{j_{1:N}=0 \ s.t.\ \sum_i j_i>0}^{n_{1:N}-\delta_{1:N,1}} \ \sum_{n_2^i+\dots+n^i_{N_{t-a}^{(t)}}=n_i-\delta_{i,l}-j_i} 1\\
&= \sum_{j_{1:N}=0 \ s.t.\ \sum_i j_i>0}^{n_{1:N}-\delta_{1:N,1}} \ \prod_{i=1}^{N}\frac{\prod_{i=1}^{N}\left(N_{t-a}^{(t)}-1+n_i-\delta_{i,1}-j_i\right)_{(n_i-\delta_{i,1}-j_i)}}{\prod_{i=1}^{N}(n_i-\delta_{i,1}-j_i)!}\\
&\le \mathcal{C}\sum_{j_{1:N}=0 \ s.t.\ \sum_i j_i>0}^{n_{1:N}-\delta_{1:N,1}} \left(N_{t-a}^{(t)}\right)^{|n|-1-\sum j_i}.
\end{aligned}$$

Then, the expectation of the last quantity gives a polynomial of degree $|n|-1$ in $\frac{W(t)}{W(a)}$. Using the same study as I_2 shows that this part is of order $\mathcal{O}\left(tW(t)^{|n|-1}\right)$.
Finally, summing over l ends the proof since the leading term is

$$\sum_{l=1}^{N}\frac{|n|!W(t)^{|n|}}{\prod_{i=1}^{N}(n_i-\delta_{i,1})!}\int_0^t \theta\frac{e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k_l-1}\prod_{m=1}^{N}\left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k_m}ds\right)^{n_m-\delta_{m,1}} da,$$

while the rest is a finite sum of $\mathcal{O}\left(tW(t)^{|n|-1}\right)$-terms. By Lemma 6.2.1,

$$c_k = \int_0^t \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1}ds + \mathcal{O}\left(e^{-\gamma t}\right),$$

where γ is equal to θ (resp. $2\alpha-\theta$) in the clonal critical and subcritical cases (resp. supercritical case). Hence, we deduce (6.11). □

Remark 6.3.10. *Taking the behavior of $\mathbb{P}(Z_0(a)=k)$ into account and using the Cauchy-Schwartz inequality for $\mathbb{E}\left[A(k,a)\mathbb{1}_{Z_0(a)=\ell}\right]$ one could actually prove that the error term in* (6.11) *is of order $\mathcal{O}\left(W(t)^{|n|-1}\right)$ in the clonal sub-critical and super-critical cases, and $\mathcal{O}\left(\log t\ W(t)^{|n|-1}\right)$ in the clonal critical case.*

Corollary 6.3.11. *We have, conditionally on the nonextinction,*

$$\lim_{t\to\infty}\left(\frac{A(k,t)}{W(t)}\right)_{k\geq 1}=\mathcal{E}\left(c_k\right)_{k\geq 1}\qquad \textit{in distribution},$$

where $\mathcal{E}$ is an exponential random variable with parameter 1.

Proof. From Lemma 6.3.9, we have

$$\lim_{t\to\infty} W(t)^{-|n|}\ \mathbb{E}_t\left[\prod_{i=1}^{K} A\left(k_i,t\right)^{n_i}\right]=|n|!\prod_{i=1}^{N} c_{k_i}^{n_i}.$$

Since the finite dimensional law of a process with form $\mathcal{E}\left(c_k\right)_{k\geq 1}$ is fully determined by its moments, it follows from the multidimensional moment problem (see [57]) and from the fact the events $\{N_t>0\}$ increase to the event of nonextinction, that we have the claimed convergence. □

6.4 Limit theorems for the frequency spectrum

The purpose of this section is to state the same kind of limit theorem as those obtained for N_t in Chapter 5. We begin by the law of large number. This result was proved in [13].

Theorem 6.4.1. *We have,*

$$e^{-\alpha t}\left(A\left(k,t\right)\right)_{k\geq 1}\underset{t\to\infty}{\longrightarrow}\frac{\mathcal{E}}{\psi'(\alpha)}\left(c_k\right)_{k\geq 1},\quad \textit{a.s. and in } L^2,$$

where $\mathcal{E}$ is the same random variable as in Theorem 5.2.1, and c_k was defined in Proposition 6.3.8.

Now, we can state central limit theorems related to this convergence. It can take several forms by before we recall that the Laplace distribution with zero mean and covariance matrix K is the probability distribution whose characteristic function is given, for all $\lambda\in\mathbb{R}^n$ by

$$\frac{1}{1+\frac{1}{2}\lambda' K\lambda}$$

We denote this law by $\mathcal{L}\left(\mu,K\right)$. We also recall that, if G is a Gaussian random vector with zero mean and covariance matrix K and $\mathcal{E}$ is an exponential random variable with parameter 1 independent of G, then $\sqrt{\mathcal{E}}G$ is Laplace $\mathcal{L}\left(\mu,K\right)$.
We can now state the first CLT.

Theorem 6.4.2. *Suppose that $\theta>\alpha$ and $\int_{[0,\infty)} e^{(\theta-\alpha)v}\mathbb{P}_V(dv)>1$. Then, we have, under $\mathbb{P}_\infty$,*

$$\lim_{t\to\infty}\left(e^{-\alpha\frac{t}{2}}\left(\psi'(\alpha)A(k,t)-e^{\alpha t}c_k\mathcal{E}\right)\right)_{k\in\mathbb{N}}\overset{d}{=}\mathcal{L}\left(0,K\right),$$

where K is some covariance matrix and the constants c_k are defined in Proposition 6.3.8.

The proof of this result can be found in Section 6.6.

Remark 6.4.3. *We are not able to compute explicitly the covariance matrix K in the general case due to our method of demonstration. However, all our other results give explicit formulas. In particular, the case where $\mathbb{P}_V$ is exponential is given by the next theorem. The Yule case is also covered in the following theorem for $d = 0$ although it does not satisfy the hypothesis of Theorem 6.4.2.*

Theorem 6.4.4. *Suppose that V is exponentially distributed with parameter $d \in [0, b)$. In this case, $\alpha = b - d$. We still suppose that $\alpha < \theta$, then*

$$\lim_{t\to\infty} \left(e^{-\alpha\frac{t}{2}} \left(\psi'(\alpha)A(k,t) - e^{\alpha t} c_k \mathcal{E}\right)\right)_{k\in N} \stackrel{d}{=} \mathcal{L}\left(0, K\right), \ \text{w.r.t. } \mathbb{P}_\infty,$$

where K is given by

$$K_{l,k} = M_{l,k} + c_k c_l \frac{\alpha}{b}\left(1 - 6\frac{d}{\alpha}\right),$$

and

$$\begin{aligned} M_{l,k} = \\ 2\psi'(\alpha)\int_0^\infty \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(\left(1-\frac{1}{W_\theta(a)}\right)^{l-1}\left(\mathbb{E}_a\left[A(k,a)\right]-c_k W(a)\right)+\left(1-\frac{1}{W_\theta(a)}\right)^{k-1}\left(\mathbb{E}_a\left[A(l,a)\right]-c_l W(a)\right)\right)da \\ -\psi'(\alpha)\int_0^\infty \theta W(a)^{-1}\mathbb{E}_a\left[\left(A(k,a)-c_k N_a\right)\mathbb{1}_{Z_0(a)=l}+\left(A(l,a)-c_l N_a\right)\mathbb{1}_{Z_0(a)=k}\right], \quad (6.19) \end{aligned}$$

where W, W_θ, $\psi'(\alpha)$ are defined in the Section 6.2.

The proof of this result can be found in Section 6.8. Note that an explicit formula for $\mathbb{E}_t A(k,t)$ is given by 6.3.3. Explicit formulas for $\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(t)=l}\right]$ are given by 6.3.5, and $\mathbb{E}_t\left[N_a \mathbb{1}_{Z_0(t)=k}\right]$ by 6.10.

Remark 6.4.5. *The condition on V in Theorem 6.4.2 is required only to ensure controls of the moments of the considered quantities. However, although the Yule case does not satisfy this condition ($V = \infty$ p.s.) it is included in this last theorem (d=0). This suggests that the condition on V may not be needed.*

The next theorem concerns the error between $A(k,t)$ and $c_k N_t$. This case is easier to treat and we have an explicit expression of the covariance matrix of the limit.

Theorem 6.4.6. *Suppose that $\theta > \alpha$, then*

$$\lim_{t\to\infty} \psi'(\alpha)\left(e^{-\alpha\frac{t}{2}}\left(A(k,t) - c_k N_t\right)\right)_{k\in\mathbb{N}} \stackrel{d}{=} \mathcal{L}\left(0, M\right), \ \text{w.r.t. } \mathbb{P}_\infty,$$

where M is defined in relation (6.19).

The proof of this result can be found in Section 6.7.

6.5 Proof of Theorem 6.4.1

Proof. Using (6.10) and the bound $\mathbb{E}\left[N_a \mathbb{1}_{Z_0(a)=k}\right] \leq \mathbb{E}\left[N_a\right]$, it follows that

$$\mathbb{E}_t\left[\left(c_k N_t - A(k,t)\right)^2\right] = 2W(t)^2 \left(\int_t^\infty \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{k-1} da\right)^2 + \mathcal{O}\left(W(t)\right).$$

Finally, it follows from Lemma 6.2.1 that

$$\mathbb{E}_t\left[e^{-2\alpha t}\left(c_k N_t - A(k,t)\right)^2\right] \underset{t\to\infty}{\sim} \mathcal{C}e^{-\gamma t},$$

where γ is equal to θ (resp. $2\alpha - \theta$) in the clonal critical and sub-critical cases (resp. supercritical case).

From this point we follow the proof of Theorem 5.2.1, except that the Yule process used in (5.9) must be replaced by another Yule process corresponding to the a binary fission every time an individual experiences a *birth or a mutation*, i.e. the new Yule process has parameter $b + \theta$. Indeed, the process $A(k,t)$ can make a positive jump only in two cases : the first corresponding to the birth of an individual in a family of size $k-1$, the other one correspond to a mutation occurring on an individual in a family of size $k+1$. □

6.6 Proof of Theorem 6.4.2

The proof of this theorem follows the same structure as Section 5.4. We refer the reader to this section for the details. It begins by some estimate on moments.

6.6.1 Preliminary moments estimates

We start by computing the moment in the case of a standard splitting tree.

Case $V_\emptyset \overset{\mathcal{L}}{=} V$

One of the main difficulties to extend the preceding proof to the frequency spectrum is to get estimates on

$$\mathbb{E}\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}c_k\mathcal{E}\right)^n\right], \text{ for } n = 2 \text{ or } 3.$$

We first study the renewal equation satisfied by $\mathbb{E}A(k,t)\mathcal{E}$ similarly as in Lemma 5.4.4.

Lemma 6.6.1 (Joint moment of $\mathcal{E}$ and $A(k,t)$). *$\mathbb{E}\left[A(k,t)\mathcal{E}\right]$ is the unique solution bounded on finite intervals of the renewal equation,*

$$\begin{aligned} f(t) = &\int_{\mathbb{R}_+} f(t-u)be^{-\alpha u}\mathbb{P}\left(V > u\right)du \\ &+ \alpha\mathbb{E}\left[A(k,.)\right] \star b\left(\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V > ., V > v\right)dv\right)(t) \\ &+ \alpha\mathbb{E}\left[\mathcal{E}X_t\right], \end{aligned} \tag{6.20}$$

with X_t the number of families of size k alive at time t whose original mutation has taken place during the lifetime of the ancestor individual.

Proof. We recall that $A(k,t)$ is the number of non-ancestral families of size k at time t. Similarly as for N_t, $A(k,t)$ can be obtained as the sum of the contributions of all the trees grafted on the lifetime of the ancestor individual in addition to the mutations which take place on the ancestral branch, that is,

$$A(k,t) = \int_{[0,t]} A(k,t-u,\xi_u)\mathbb{1}_{V_\emptyset>u}\xi(du) + X_t,$$

where $(A(k,t,i), t \in \mathbb{R}_+)_{i\geq 1}$ is a family of independent processes having the same law as $A(k,t)$. Now, taking the product $A(k,t)N_s$ and using the same arguments as in the proof of lemma 5.4.4 to take the limit in s leads to the result. In particular, the last term is obtained using that

$$\lim_{s\to\infty} \mathbb{E}\left[X_t \frac{N_s}{W(s)}\right] = \mathbb{E}\left[X_t\mathcal{E}\right].$$

□

The result of Lemma 6.6.1 is quite disappointing since the presence of the mysterious process X_t prevents any explicit resolution of equation (6.20). However, one may note that equation (6.20) is quite similar to equation (5.16) driving $\mathbb{E}N_t\mathcal{E}$, so if the contribution of X_t in the renewal structure of the process is small enough, one can expect the same asymptotic behaviour for $\mathbb{E}A(k,t)\mathcal{E}$ as for $\mathbb{E}N_t\mathcal{E}$. Moreover, we clearly have on X_t the following a.s. estimate,

$$X_t \leq \int_{[0,t]} \mathbb{1}_{Z_0^{(u)}(t-u)>0}\mathbb{1}_{V>u}\xi(du), \tag{6.21}$$

where $Z_0^{(i)}$ denote for the ancestral families on the ith trees grafted on the ancestral branch. Hence, if we take $\theta > \alpha$ and we suppose $V < \infty$ a.s., one can expect that X_t decreases very fast. These are the ideas the following Lemma is based on. Moreover, as it is seen in the proof of the following lemma, the hypothesis $V < \infty$ a.s. can be weakened.

Lemma 6.6.2. *Under the hypothesis of Theorem 6.4.2, for all $k \geq 1$, there exists a constant $\gamma_k \in \mathbb{R}$ such that,*

$$\lim_{t\to\infty} \mathbb{E}N_t\mathcal{E}c_k - \mathbb{E}A(k,t)\mathcal{E} = \gamma_k. \tag{6.22}$$

Proof. Combining equations (5.16) and (6.20), we get that,

$$\begin{aligned}\mathbb{E}N_t\mathcal{E}c_k - \mathbb{E}A(k,t)\mathcal{E} &= \int_{\mathbb{R}_+} \left(\mathbb{E}N_{t-u}\mathcal{E}c_k - \mathbb{E}A(k,t-u)\mathcal{E}\right) be^{-\alpha u}\mathbb{P}\left(V>u\right) du \\ &\quad \underbrace{+\, \alpha b \left(c_k\mathbb{E}N_. - \mathbb{E}\left[A(k,.)\right]\right) \star \left(\int_{\mathbb{R}_+} e^{-\alpha v}\mathbb{P}\left(V>.,V>v\right) dv\right)(t)}_{:=\xi_1^{(k)}(t)} \\ &\quad + \underbrace{c_k\mathbb{P}\left(V>t\right) - \alpha\mathbb{E}\left[X_t\mathcal{E}\right]}_{:=\xi_2^{(k)}(t)},\end{aligned}$$

which is also a renewal equation. On one hand, using equations (3.4) and Theorem 6.3.3 imply that

$$\mathbb{E}_t\left[c_kN_t - A(k,t)\right] = W(t)\int_t^\infty \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1} ds,$$

which leads using Lemma 6.2.1, to

$$\begin{aligned}\xi_1(t) =&\alpha \int_{\mathbb{R}_+} \left(c_k \mathbb{E} N_{t-u} - \mathbb{E}\left[A(k,t-u)\right]\right) \int_{\mathbb{R}_+} e^{-\alpha v} \mathbb{P}\left(V > u, V > v\right) dv du \\ \leq& \, \mathcal{C} \int_{[0,t]} e^{(\alpha-\theta)t-u} \mathbb{P}\left(V > u\right) du \int_{[0,\infty)} e^{-\alpha u} du \\ \leq& \, \frac{\mathcal{C}}{\alpha} e^{-(\theta-\alpha)t} \int_0^t e^{(\theta-\alpha)u} \mathbb{P}\left(V > u\right) du, \end{aligned} \tag{6.23}$$

for some positive real constant $\mathcal{C}$.
The derivative of the r.h.s. of (6.23) is given by

$$\frac{\mathcal{C}}{\alpha} e^{-(\theta-\alpha)t} \left(e^{(\theta-\alpha)t} \mathbb{P}\left(V > t\right) - (\alpha - \theta) \int_0^t e^{(\theta-\alpha)u} \mathbb{P}\left(V > u\right) du \right), \quad t > 0, \tag{6.24}$$

which is equal to

$$\frac{\mathcal{C}}{\alpha} e^{-(\theta-\alpha)t} \left(1 - \int_{[0,t]} e^{(\theta-\alpha)s} \mathbb{P}_V(ds) \right), \quad t > 0,$$

using Stieljes integration by parts. Now, since,

$$\int_{[0,\infty)} e^{(\theta-\alpha)s} \mathbb{P}_V(ds) > 1,$$

this shows that the right hand side of (6.23) is decreasing for t large enough. Moreover, it is straightforward to shows that the r.h.s. of (6.23) is also integrable. This implies that $\xi_1^{(k)}$ is DRI (see Section 2.7 for the definition of DRI) from Lemma 2.7.1. On the other hand, it follows from (6.21) that

$$X_t \mathcal{E} \leq \mathcal{E} \int_{[0,t]} \mathbb{1}_{Z_0^{(u)}(t-u)>0} \mathbb{1}_{V>t} \xi(du). \tag{6.25}$$

Then, we obtain using Cauchy-Schwarz inequality, that

$$\mathbb{E}\left[X_t \mathcal{E}\right] \leq \sqrt{\frac{2\alpha}{b}} \mathbb{E}\left[\left(\int_{[0,t]} \mathbb{1}_{Z_0^{(u)}(t-u)>0} \mathbb{1}_{V>t} \xi(du) \right)^2 \right]^{1/2}.$$

It follows that we need to investigate the behavior of

$$\mathbb{E}\left[\left(\int_{(0,t)} \mathbb{1}_{Z_0^{(u)}(t-u)>0} \mathbb{1}_{V>t} \xi(du) \right)^2 \right],$$

which is equal to

$$\int_0^t \mathbb{P}\left(Z_0(t-u) > 0\right) \mathbb{P}\left(V > t\right) b du + \int_{[0,t]^2} \mathbb{P}\left(Z_0(t-v) > 0\right) \mathbb{P}\left(Z_0(t-u) > 0\right) \mathbb{P}\left(V > u, V > v\right) b^2 du\, dv,$$

using Lemma 5.4.1. Then, since, from (6.2) and Lemma 6.2.1,

$$\mathbb{P}_{t-u}\left(Z_0(t-u) > 0\right) = \frac{e^{-\theta(t-u)} W(t-u)}{W_\theta(t-u)} = \mathcal{O}(e^{-(\theta-\alpha)(t-u)}),$$

it follows, using that the right hand side of (6.23) is DRI and Lemma 2.7.1, that $\xi_2^{(k)}$ is DRI. Finally, it comes from Theorem 2.7.2, that

$$\lim_{t\to\infty} \mathbb{E}N_t\mathcal{E}c_k - \mathbb{E}A(k,t)\mathcal{E} = \frac{\alpha}{\psi'(\alpha)} \int_{\mathbb{R}_+} \xi_1^{(k)}(s) + \xi_2^{(k)}(s)ds. \tag{6.26}$$

□

Using the preceding lemma, we can now get the quadratic error in the convergence of the frequency spectrum.

Lemma 6.6.3 (Quadratic error for the convergence of $A(k,t)$.). *Let k and l two positive integers. Then under the hypothesis of Theorem 6.4.2, there exists a family of real numbers $(a_{k,l})_{l,k\geq 1}$ such that,*

$$\lim_{t\to\infty} e^{-\alpha t}\mathbb{E}\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}\mathcal{E}c_k\right)\left(\psi'(\alpha)A(l,t) - e^{\alpha t}\mathcal{E}c_l\right)\right] = \frac{\alpha}{b}a_{k,l},$$

where the sequence $(c_k)_{k\geq 1}$ is defined in Proposition 6.3.8.

Proof. Now, noting

$$c_k(t) := \int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{k-1} da, \tag{6.27}$$

we have, from the proof of Proposition 6.3.8 and Lemma 4.1.1,

$$\psi'(\alpha)^2\mathbb{E}_t\left[A(k,t)A(l,t)\right] = 2e^{2\alpha t}c_k(t)c_l(t) + e^{\alpha t}\left(4\psi'(\alpha)e^{\alpha t}F(t)c_k(t)c_l(t) + \frac{R}{\psi'(\alpha)}\right) + \mathcal{O}\left(1\right), \tag{6.28}$$

with

$$\begin{aligned}
R := &-\psi'(\alpha)\int_0^\infty 2\theta\, \frac{e^{-\theta a}W(a)}{W_\theta(a)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{l-1}\int_0^a \frac{e^{-\theta s}}{W_\theta(s)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{k-1} dsda\\
&-\psi'(\alpha)\int_0^\infty 2\theta\, \frac{e^{-\theta a}W(a)}{W_\theta(a)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{k-1}\int_0^a \frac{e^{-\theta s}}{W_\theta(s)^2}\left(1 - \frac{1}{W_\theta(a)}\right)^{l-1} dsda\\
&+\psi'(\alpha)\int_0^\infty \theta W(a)^{-1}\left(\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(a)=l}\right] + \mathbb{E}_t\left[A(l,t)\mathbb{1}_{Z_0(a)=k}\right]\right) da,
\end{aligned}$$

and F, μ are defined in Lemma 4.1.1. Now, using (5.20), we have

$$\mathbb{E}_t\mathcal{E}^2 - 2 = -2\mu\psi'(\alpha)e^{-\alpha t} + o(e^{-\alpha t}),$$

which leads to

$$\begin{aligned}
&\mathbb{E}_t\left[\left(e^{-\alpha t}\psi'(\alpha)A(k,t) - \mathcal{E}c_k\right)\left(e^{-\alpha t}\psi'(\alpha)A(l,t) - \mathcal{E}c_l\right)\right]\\
=&\mathbb{E}_t\left[e^{-2\alpha t}\psi'(\alpha)^2A(k,t)A(l,t)\right] - c_l\mathbb{E}_t\left[e^{-\alpha t}\psi'(\alpha)A(k,t)\mathcal{E}\right] - c_k\mathbb{E}_t\left[e^{-\alpha t}\psi'(\alpha)A(l,t)\mathcal{E}\right]\\
&+ 2c_kc_l - 2c_kc_l\mu\psi'(\alpha)e^{-\alpha t} + o(e^{-\alpha t}),\\
=&2\left(c_k(t) - c_k\right)\left(c_l(t) - c_l\right) - 4\mu\psi'(\alpha)c_kc_le^{-\alpha t} + Re^{-\alpha t}\\
&-\left(2c_k(t)c_l + 2c_l(t)c_k - 2c_kc_l\psi'(\alpha)e^{-\alpha t}\mathbb{E}_tN_t\mathcal{E}\right)\\
&+\psi'(\alpha)c_le^{-\alpha t}\mathbb{E}_t\left[\left(c_kN_t - A(k,t)\right)\mathcal{E}\right] + \psi'(\alpha)c_ke^{-\alpha t}\mathbb{E}_t\left[\left(c_lN_t - A(l,t)\right)\mathcal{E}\right] + o(e^{-\alpha t}),
\end{aligned}$$

Since, by Lemma 6.2.1

$$c_k(t) = c_k + \mathcal{O}(e^{-\theta t}) = c_k + o(e^{-\alpha t}),$$

it follows, combining (6.26), (5.21), and Lemma 6.6.2, that

$$\begin{aligned}
&e^{\alpha t}\mathbb{E}_t\left[\left(e^{-\alpha t}\psi'(\alpha)A(k,t) - \mathcal{E}c_k\right)\left(e^{-\alpha t}\psi'(\alpha)A(l,t) - \mathcal{E}c_l\right)\right] \\
&\quad=\psi'(\alpha)\left(c_k\gamma_l + c_l\gamma_k\right) + c_kc_l\left(2e^{\alpha t} - 2\psi'(\alpha)\mathbb{E}_tN_t\mathcal{E}\right) + R - 4\mu\psi'(\alpha)c_kc_l + o(1) \\
&\quad=\psi'(\alpha)\left(c_k\gamma_l + c_l\gamma_k\right) + c_kc_l\left(\frac{1}{\psi'(\alpha)} + 3\mu\right) + R - 4\mu\psi'(\alpha)c_kc_l + o(1).
\end{aligned}$$

The result follows readily from the fact that $\mathbb{P}\left(N_t > 0\right) \sim \frac{\alpha}{b}$.

□

Lemma 6.6.4 (Boundedness of the third moment)**.** *Let k_1, k_2, k_3 three positive integers, then*

$$\mathbb{E}\left[\prod_{i=1}^{3}\left|e^{-\frac{\alpha}{2}t}\left(\psi'(\alpha)A(k_i,t) - e^{\alpha t}\mathcal{E}c_{k_i}\right)\right|\right] = \mathcal{O}\left(1\right).$$

Proof. We have,

$$\mathbb{E}\left[\left|\prod_{i=1}^{3}\frac{\left(\psi'(\alpha)A(k_i,t) - e^{\alpha t}\mathcal{E}c_{k_i}\right)}{e^{\frac{\alpha}{2}t}}\right|\right] \leq \prod_{i=1}^{3}\left(\mathbb{E}\left[\left|\frac{\left(\psi'(\alpha)A(k_i,t) - e^{\alpha t}\mathcal{E}c_{k_i}\right)}{e^{\frac{\alpha}{2}t}}\right|^3\right]\right)^{\frac{1}{3}}.$$

Hence, we only have to prove the Lemma for $k_1 = k_2 = k_3 = k$. Hence,

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{\left(\psi'(\alpha)A(k,t) - e^{\alpha t}\mathcal{E}c_k\right)}{e^{\frac{\alpha}{2}t}}\right|^3\right] \leq 8\mathbb{E}\left[\left|\frac{\psi'(\alpha)A(k,t) - c_kN_t}{e^{\frac{\alpha}{2}t}}\right|^3\right] + 8c_k\mathbb{E}\left[\left|\frac{\psi'(\alpha)N_t - N_t^\infty}{e^{\frac{\alpha}{2}t}}\right|^3\right] \\
+ 8c_k\mathbb{E}\left[\left|\frac{N_t^\infty - e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}\right|^3\right].
\end{aligned}$$

The last two terms have been treated in the proof of Lemma 5.4.7, and the boundedness of

$$\mathbb{E}\left[\left|\frac{\psi'(\alpha)A(k,t) - c_kN_t}{e^{\frac{\alpha}{2}t}}\right|^3\right],$$

follows from the following Lemma 6.6.5 and Hölder's inequality.

□

Lemma 6.6.5. *For all $k \geq 1$,*

$$\mathbb{E}\left[\left(\frac{A(k,t) - c_kN_t}{e^{-\frac{\alpha}{2}t}}\right)^4\right],$$

is bounded.

Due to technicality, the proof of this lemma is postponed to the end of this chapter.

Arbitrary initial distribution case

The following Lemmas are the counter part of Lemmas 5.4.8, 5.4.9, and 5.4.10. They play the same role in the proof of Theorem 6.4.2. In the sequel, we denote by $(A(k,t,\Xi))_{k\geq 1}$, the frequency spectrum of the splitting tree where the lifetime of the ancestral individual is Ξ, in the same manner as for $N_t(\Xi)$ in the previous section.

Lemma 6.6.6 (L^2 convergence in the general case)**.** *Consider the general frequency spectrum $(A(k,t,\Xi))_{k\geq 1}$, then, for all k, $\psi'(\alpha)e^{-\alpha t}A(k,t,\Xi)$ converge to $\mathcal{E}(\Xi)$ (see 5.24) in L^2 as t goes to infinity and*

$$\lim_{t\to\infty} e^{-\alpha t}\mathbb{E}\left[\left(\psi'(\alpha)A(k,t,\Xi)-e^{\alpha t}\mathcal{E}(\Xi)c_k\right)\left(\psi'(\alpha)A(l,t,\Xi)-e^{\alpha t}\mathcal{E}(\Xi)c_k\right)\right]=\frac{\alpha}{b}a_{k,l}\int_{\mathbb{R}_+}e^{-\alpha u}\mathbb{P}\left(\Xi>u\right)bdu,$$

where the convergence is uniform w.r.t. the random variable Ξ. In the case where Ξ is distributed as $O_2^{(\beta t)}$, for $0<\beta<\frac{1}{2}$ (see section 5.4.2), we get

$$\lim_{t\to\infty} e^{-\alpha t}\mathbb{E}\left[\left(\psi'(\alpha)A(k,t,O_2^{(\beta t)})-e^{\alpha t}\mathcal{E}(O_2^{(\beta t)})c_k\right)\left(\psi'(\alpha)A(l,t,O_2^{\beta t})-e^{\alpha t}\mathcal{E}(O_2^{(\beta t)})c_k\right)\right]=\psi'(\alpha)a_{k,l}.$$

Lemma 6.6.7 (First moment)**.** *The first moments are asymptotically bounded, that is, for all $k\geq 1$,*

$$\mathbb{E}\left(\psi'(\alpha)A(k,t)(\Xi)-e^{\alpha t}c_k\mathcal{E}(\Xi)\right)\leq\mathcal{O}(1),$$

uniformly with respect to the random variable Ξ.

Lemma 6.6.8 (Boundedness in the general case.)**.** *Let k_1,k_2,k_3 three positive integers, then*

$$\mathbb{E}\left[\left|\prod_{i=1}^{3}\frac{\left(\psi'(\alpha)A(k_i,t)-e^{\alpha t}\mathcal{E}c_{k_i}\right)}{e^{\frac{\alpha}{2}t}}\right|\right]=\mathcal{O}\left(1\right),$$

uniformly with respect to the random variable Ξ.

We do not detail the proofs of these results since they are direct adaptations of the proofs of Lemmas 5.4.8, 5.4.9 and 5.4.10.

6.6.2 Proof of the theorem

The following result is based on the fact that, in the clonal sub-critical case, the lifetime of a family is expected to be small. It follows that, in the decomposition of Figure 5.2, one can expect that all the family of size k live in different subtrees as soon as $t>>u$. This is the point of the following lemma.

Lemma 6.6.9. *Suppose that $\alpha<\theta$. If we denote by $\Gamma_{u,t}$ the event,*

$$\Gamma_{u,t}=\left\{\text{"there is no family in the population at time } t \text{ which is older than } u\text{"}\right\},$$

then, for all β in $(0,1-\frac{\alpha}{\theta})$, we have

$$\lim_{t\to\infty}\mathbb{P}_{\beta t}\left(\Gamma_{\beta t,t}\right)=1.$$

Proof. The proof of this Lemma, as the calculation of the moments of $A(k,t)$ relies on the representation of the genealogy of the living population at time t as a coalescent point process (see Section 3.4). Moreover, we denote by $\widetilde{N}_u^{(t)}$ the number of living individuals at time u who have alive descent at time t. According to Proposition 4.3.1, under $\mathbb{P}_t$, $\widetilde{N}_u^{(t)}$ is geometrically distributed with parameter $\frac{W(t-u)}{W(t)}$.
Now, $\mathbb{1}_{\Gamma_{u,t}}$ can be rewritten as

$$\mathbb{1}_{\Gamma_{u,t}} = \prod_{i=1}^{\widetilde{N}_u^{(t)}} \mathbb{1}_{\left\{Z_0^i(t-u)=0\right\}},$$

where $Z_0^i(t-u)$ denotes the number of individuals alive at time t descending from the ith individual alive at time u and carrying its type (the clonal type of the sub-CPP). Moreover, using again Proposition 4.3.1, we know that that under $\mathbb{P}_t$, the family $Z_0^{(i)}(t-u)$ is an i.i.d. family of random variables distributed as $Z_0(t-u)$ under $\mathbb{P}_{t-u}$, and $\widetilde{N}_u^{(t)}$ is independent of $Z_0^{(i)}(t-u)$ (still under $\mathbb{P}_t$).
Then,

$$\mathbb{P}_t\left(\Gamma_{t,u}\right) = \mathbb{E}_t\left[\mathbb{P}_{t-u}\left(Z_0(t-u)=0\right)^{\widetilde{N}_u^{(t)}}\right] = \frac{\mathbb{P}_{t-u}\left(Z_0(t-u)=0\right)\frac{W(t-u)}{W(t)}}{1-\mathbb{P}_{t-u}\left(Z_0(t-u)=0\right)\left(1-\frac{W(t-u)}{W(t)}\right)}.$$

Using (6.2), some calculus leads to,

$$\mathbb{P}_t\left(\Gamma_{t,u}\right) = 1 - \frac{1}{1+\frac{W_\theta(t-u)}{e^{-\theta(t-u)}W(t)}\left(1-\frac{e^{-\theta(t-u)}W(t-u)}{W_\theta(t-u)}\right)}.$$

Now, since,

$$\mathbb{P}_t\left(\Gamma_{t,u}\right) = \mathbb{P}_u\left(\Gamma_{t,u}\right)\frac{\mathbb{P}\left(N_u>0\right)}{\mathbb{P}\left(N_t>0\right)} + \frac{\mathbb{P}\left(\Gamma_{t,u}, N_t=0, N_u>0\right)}{\mathbb{P}\left(N_t>0\right)},$$

taking $u=\beta t$, we obtain, using Lemma 6.2.1 and

$$\mathbb{P}\left(N_t=0, N_{\beta t}>0\right) = \mathbb{P}\left(N_{\beta t}>0\right) - \mathbb{P}\left(N_t>0\right) \underset{t\to\infty}{\to} 0,$$

the desired result. □

Proof of Theorem 6.4.2. Fix $0<u<t$. Note that the event $\Gamma_{u,t}$ of Lemma 6.6.9 can be rewritten as

$$\mathbb{1}_{\Gamma_{u,t}} = \prod_{i=1}^{N_u} \mathbb{1}_{\left\{Z_0^i(t-u,O_i)=0\right\}}, \tag{6.29}$$

where $Z_0^i(t-u,O_i)$ denote the number of individuals alive at time t carrying the same type as the ith alive individual at time u, that is the ancestral family of the splitting constructed from the residual lifetime of the ith individual (see Section 5.4.2).
Let K be a multi-integer, we denote by $\mathcal{L}^{(K)}$ (resp. $A(K,t)$) the random vector $\left(\mathcal{L}^{k_1},\dots,\mathcal{L}^{k_N}\right)$ (resp. $(A(k_1,t),\dots,A(k_N,t))$) with

$$\mathcal{L}_t^{k_i} = \frac{\psi'(\alpha)A(k,t)-c_k e^{\alpha t}\mathcal{E}}{e^{\frac{\alpha}{2}t}}.$$

On the event $\Gamma_{u,t}$, we have a.s.,

$$A(k_l,t) = \sum_{i=1}^{N_u} A^{(i)}(k_l, t-u, O_i), \quad \forall l = 1, \dots, N,$$

where the family $\left(A^{(i)}\left(k_l, t-u, O_i\right)\right)_{i\geq 1}$ stand for the frequency spectrum for each subtree, which are independent from Lemma 5.3.3 (see also Section 5.4.2 and Figure 5.2). Hence, using Lemma 5.4.11,

$$\mathcal{L}_t^{k_l} = \sum_{i=1}^{N_u} \frac{\psi'(\alpha) A^{(i)}(k_l, t-u, O_i) - e^{\alpha(t-u)} \mathcal{E}_i(O_i) c_{k_l}}{e^{\frac{\alpha}{2}u} e^{\frac{\alpha}{2}(t-u)}}.$$

By Lemma 5.3.3, that the family $\left(A^i(k_l, t-u, O_i)\right)_{2\leq i\leq N_u}$ is i.i.d. under $\mathbb{P}_u$.
In the sequel, we denote, for all l and $i \geq 1$,

$$\tilde{A}^{(i)}\left(k_l, t-u, O_i\right) = \frac{\psi'(\alpha) A^{(i)}\left(k_l, t-u, O_i\right) - e^{\alpha(t-u)} \mathcal{E}_i(O_i) c_{k_l}}{e^{\frac{\alpha}{2}(t-u)}}.$$

As in the proof of Theorem 5.2.2, let

$$\varphi_K\left(\xi\right) := \mathbb{E}\left[\exp\left(i < \tilde{A}\left(K, t-u, O_2\right), \xi >\right) \mathbb{1}_{Z_0^2(t-u,O_2)=0}\right],$$
$$\tilde{\varphi}_K\left(\xi\right) := \mathbb{E}\left[\exp\left(i < \tilde{A}\left(K, t-u, O_1\right), \xi >\right) \mathbb{1}_{Z_0^1(t-u,O_1)=0}\right].$$

From this point, following closely the proof of Theorem 5.2.2, with β in $\left(0, \frac{1}{2} \wedge (1-\frac{\alpha}{\theta})\right)$, the only difficulty is to handle the indicator function $\mathbb{1}_{Z_0(t-u,O_i)>0}$ in the Taylor development of φ_K. We show how it can be done for one of the second order terms, and leave the rest of the details to the reader.
It follows from Hölder's inequality that

$$\mathbb{E}\left[\left(\frac{\psi'(\alpha) A^{(i)}\left(k_l, (1-\beta)t, O_i\right) - e^{\alpha((1-\beta)t)} \mathcal{E}_i(O_i) c_{k_l}}{e^{\frac{\alpha}{2}((1-\beta)t)}}\right)^2 \mathbb{1}_{Z_0^2((1-\beta)t,O_2)>0}\right]$$
$$\leq \mathbb{E}\left[\left(\frac{\psi'(\alpha) A^{(i)}\left(k_l, (1-\beta)t, O_i\right) - e^{\alpha(1-\beta)t} \mathcal{E}_i(O_i) c_{k_l}}{e^{\frac{\alpha}{2}(1-\beta)t}}\right)^3\right]^{\frac{2}{3}} \mathbb{P}\left(Z_0^2((1-\beta)t, O_2) > 0\right)^{\frac{1}{3}}, \quad (6.30)$$

from which it follows, using Lemma 6.6.8, that the r.h.s. of this last inequality is

$$\mathcal{O}\left(\mathbb{P}\left(Z_0^2(t-u, O_2) > 0\right)^{\frac{1}{3}}\right).$$

Now, using (6.29) and Lemma 6.6.9, it is easily seen that

$$\lim_{t\to\infty} \mathbb{P}\left(Z_0^2((1-\beta)t, O_2) > 0\right) = 0.$$

Finally, using Lemma 6.6.3, we get

$$\lim_{t\to\infty} \mathbb{E}\left[\left(\frac{\psi'(\alpha) A^{(i)}\left(k_l, t-u, O_i\right) - e^{\alpha(t-u)} \mathcal{E}_i(O_i) c_{k_l}}{e^{\frac{\alpha}{2}(t-u)}}\right)^2 \mathbb{1}_{Z_0^2(t-u,O_2)=0}\right] = \psi'(\alpha) a_{k,k}.$$

These allow us to conclude that

$$\lim_{t\to\infty} \mathbb{E}_{\beta t}\left[e^{i<\mathcal{L}_t^{(K)},\xi>}\mathbb{1}_{\Gamma_t}\right] = \frac{1}{1+\sum_{i,j=1}^N \mathcal{M}_{i,j}\ \xi_i\xi_j},$$

where $K_{i,j}$ is given by

$$\mathcal{M}_{i,j} := \psi'(\alpha)a_{K_i,K_j},$$

with K is the multi-integer $(k_1,\dots,k_N)$, and the $a_{l,k}$s are defined in Lemma 6.6.3.
To end the proof, note that,

$$\left|\mathbb{E}_\infty\left[e^{i<\mathcal{L}_t^{(K)},\xi>}\right] - \mathbb{E}_{\beta t}\left[e^{i<\mathcal{L}_t^{(K)},\xi>}\mathbb{1}_{\Gamma_{\beta t,t}}\right]\right| \leq \mathbb{E}\left[\left|\frac{\mathbb{1}_{\text{NonEx}}}{\mathbb{P}\left(\text{NonEx}\right)} - \frac{\mathbb{1}_{N_{\beta t}>0}\mathbb{1}_{\Gamma_{\beta t,t}}}{\mathbb{P}\left(N_{\beta t}>0\right)}\right|\right] \underset{t\to\infty}{\to} 0,$$

thanks to Lemma 6.6.9. □

6.7 Proof of Theorem 6.4.6

Since all the ideas of the proof of this theorem have been developed in the last two section, we do not detail all the proof. The only step which needs clarification is the computation of the covariance matrix of the Laplace limit law $\mathcal{M}$. According to the proof of Theorem 6.4.2, it is given by

$$\begin{aligned}\mathcal{M}_{i,j} := \lim_{t\to\infty}\frac{W(\beta t)}{e^{\alpha\beta t}}\mathbb{E}\Bigg[&\left(\frac{\psi'(\alpha)A^{(i)}\left(k_i,(1-\beta)t,O_i\right)-\psi'(\alpha)c_{k_i}N_{(1-\beta)t}}{e^{\frac{\alpha}{2}((1-\beta)t)}}\right)\\ &\times\left(\frac{\psi'(\alpha)A^{(i)}\left(k_j,(1-\beta)t,O_i\right)-c_{k_j}N_{(1-\beta)t}}{e^{\frac{\alpha}{2}((1-\beta)t)}}\right)\mathbb{1}_{Z_0^2((1-\beta)t,O_2)>0}\Bigg],\end{aligned}$$

which is equal, thanks to (6.30) and an easy adaptation of Lemma 5.4.9, to

$$\mathcal{M}_{i,j} = \lim_{t\to\infty}\frac{b\psi'(\alpha)}{\alpha}\frac{W(\beta t)}{e^{\alpha\beta t}}e^{\alpha t}\mathbb{E}\left[\left(e^{-\alpha t}A(k_i,t)-c_{k_i}e^{-\alpha t}N_t\right)\left(e^{-\alpha t}A(k_j,t)-c_{k_j}e^{-\alpha t}N_t\right)\right].$$

So it remains to get the limit of

$$e^{\alpha t}\mathbb{E}\left[\left(e^{-\alpha t}\psi'(\alpha)A(k,t)-\psi'(\alpha)c_ke^{-\alpha t}N_t\right)\left(e^{-\alpha t}\psi'(\alpha)A(l,t)-c_le^{-\alpha t}\psi'(\alpha)N_t\right)\right],$$

as t goes to infinity. We recall that using the calculus made in the proof of Theorem 6.4.2, we have

$$\begin{aligned}&\mathbb{E}_tA(k,t)N_t\\ &\quad= 2W(t)^2c_k(t) - 2W(t)\int_{[0,t]}\theta\mathbb{P}_a\left(Z_0(a)=k\right)da + W(t)\int_{[0,t]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]da.\end{aligned}\tag{6.31}$$

Moreover, (6.28) entails

$$\psi'(\alpha)^2\mathbb{E}_tA(k,t)A(l,t) = 2W(t)^2c_k(t)c_l(t) + RW(t) + o(e^{-\alpha t}),$$

with

$$\begin{aligned}R :=& -\psi'(\alpha)\int_0^\infty 2\theta W(a)^{-1}\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[A(l,a)\right]da\\&+\psi'(\alpha)\int_0^\infty 2\theta W(a)^{-1}\mathbb{P}_a\left(Z_0(a)=l\right)\mathbb{E}_a\left[A(k,a)\right]da\\&+\psi'(\alpha)\int_0^\infty \theta W(a)^{-1}\left(\mathbb{E}_t\left[A(k,t)\mathbb{1}_{Z_0(a)=l}\right]+\mathbb{E}_t\left[A(l,t)\mathbb{1}_{Z_0(a)=k}\right]\right)da.\end{aligned}$$

These identities allow us to obtain

$$\begin{aligned}&\mathbb{E}_t\left[\left(A(k,t)-c_kN_t\right)\left(A(l,t)-c_lN_t\right)\right]=2W(t)^2c_k(t)c_l(t)+e^{-\alpha t}R+o(e^{-\alpha t}),\\&-2c_lc_k(t)W(t)^2+2c_lW(t)\int_{[0,t]}\theta\mathbb{P}_a\left(Z_0(a)=k\right)da-c_lW(t)\int_{[0,t]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]da\\&-2c_kc_l(t)W(t)^2+2c_lW(t)\int_{[0,t]}\theta\mathbb{P}_a\left(Z_0(a)=l\right)da-c_kW(t)\int_{[0,t]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=l}\right]da\\&+c_kc_lW(t)^2\left(2-\frac{1}{W(t)}\right)\\&=2W(t)^2\left(c_k(t)-c_l\right)\left(c_l(t)-c_k\right)+e^{-\alpha t}\frac{R}{\psi'(\alpha)}+o(e^{-\alpha t}),\\&+2c_lW(t)\int_{[0,t]}\theta\mathbb{P}_a\left(Z_0(a)=k\right)da-c_lW(t)\int_{[0,t]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]da\\&+2c_lW(t)\int_{[0,t]}\theta\mathbb{P}_a\left(Z_0(a)=l\right)da-c_kW(t)\int_{[0,t]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=l}\right]da\\&-c_kc_lW(t).\end{aligned}$$

Taking the limit as t goes to infinity leads to

$$\begin{aligned}M_{k,l}:=\lim_{t\to\infty}&\psi'(\alpha)^2e^{-\alpha t}\mathbb{E}_t\left[\left(A(k,t)-c_kN_t\right)\left(A(l,t)-c_lN_t\right)\right]=R\\&+2\psi'(\alpha)c_l\int_{[0,\infty]}\theta\mathbb{P}_a\left(Z_0(a)=k\right)da-\psi'(\alpha)c_l\int_{[0,\infty]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=k}\right]da\\&+2\psi'(\alpha)c_l\int_{[0,\infty]}\theta\mathbb{P}_a\left(Z_0(a)=l\right)da-\psi'(\alpha)c_k\int_{[0,\infty]}\theta W(a)^{-1}\mathbb{E}_a\left[N_a\mathbb{1}_{Z_0(a)=l}\right]da\\&-\psi'(\alpha)c_kc_l.\end{aligned}\tag{6.32}$$

Finally, since $\mathbb{P}\left(N_t>0\right)\sim\frac{\alpha}{b}$,

$$\mathcal{M}_{i,j}=M_{k_i,k_j}.$$

6.8 Markovian cases

Theorem 5.2.2 for the Markovian case is already well known (see [4]), however the allelic partition for such model has not been studied. We can get more information on the unknown covariance matrix K in the case where the life duration distribution is exponential. Our study also cover the case $\mathbb{P}_V=\delta_\infty$ (Yule case), although it does not fit the conditions required by the Theorem 6.4.2. The reason comes from our method of calculation for $\mathbb{E}\left[A(k,t)\mathcal{E}\right]$. Let us consider the filtration

$(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, where $\mathcal{F}_t$ is the σ-field generated by the tree truncated above t and the restriction of the mutation measure on $[0,t)$.
Then N_t is Markovian with respect to $\mathcal{F}_t$ and for all positive real numbers $t \le s$,

$$\mathbb{E}\left[A(k,t)N_s \mid \mathcal{F}_t\right] = A(k,t)N_t\mathbb{E}\left[N_{s-t}\right].$$

So that,

$$\mathbb{E}\left[A(k,t)N_s\right] = \mathbb{E}\left[A(k,t)N_t\right]\left(W(s-t) - \mathbb{P}_V \star W(s-t)\right).$$

By making a renormalization by $e^{-\alpha s}$ and taking the limit as s goes to infinity, we get,

$$\mathbb{E}\left[A(k,t)\mathcal{E}\right] = \psi'(\alpha)e^{-\alpha t}\mathbb{E}\left[A(k,t)N_t\right],$$

since, in the Markovian case, it is known from [15] that

$$\frac{\alpha}{b} = \psi'(\alpha).$$

Suppose first that $d > 0$. It follows that,

$$\begin{aligned}\mathbb{E}&\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}c_k\mathcal{E}\right)\left(\psi'(\alpha)A(l,t) - e^{\alpha t}c_l\mathcal{E}\right)\right] = \psi'(\alpha)^2\mathbb{E}_t\left[A(k,t)A(l,t)\right]\mathbb{P}\left(N_t > 0\right)\\ &- c_k\psi'(\alpha)^2\mathbb{E}_t\left[A(l,t)N_t\right]\mathbb{P}\left(N_t > 0\right) - c_l\psi'(\alpha)^2\mathbb{E}_t\left[A(k,t)N_t\right]\mathbb{P}\left(N_t > 0\right)\\ &+ 2\psi'(\alpha)e^{2\alpha t}c_kc_l\end{aligned}$$

By (5.20),

$$\mathbb{P}\left(N_t > 0\right) = \psi'(\alpha) + \psi'(\alpha)^2\mu e^{-\alpha t} + o(e^{-\alpha t}),$$

so

$$\begin{aligned}&\mathbb{E}\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}c_k\mathcal{E}\right)\left(\psi'(\alpha)A(l,t) - e^{\alpha t}c_l\mathcal{E}\right)\right]\\ &= \mathbb{P}\left(N_t > 0\right)\psi'(\alpha)^2\mathbb{E}_t\left[\left(A(k,t) - c_kN_t\right)\left(A(l,t) - c_lN_t\right)\right] + c_kc_l\psi'(\alpha)\left(2e^{2\alpha t} - \psi'(\alpha)\mathbb{E}_t\left[N_t^2\right]\mathbb{P}\left(N_t > 0\right)\right).\end{aligned}$$

Finally, since, using Proposition 4.1.1,

$$\lim_{t\to\infty} e^{-\alpha t}\left(2e^{2\alpha t} - \psi'(\alpha)\mathbb{E}_t\left[N_t^2\right]\mathbb{P}\left(N_t > 0\right)\right) = \psi'(\alpha)\left(1 - 6\mu\right),$$

it follows from (6.32),

$$\begin{aligned}\lim_{t\to\infty}&\mathbb{E}\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}c_k\mathcal{E}\right)\left(\psi'(\alpha)A(l,t) - e^{\alpha t}c_l\mathcal{E}\right)\right]\\ &= \psi'(\alpha)M_{k,l} + c_kc_l\psi'(\alpha)^2\left(1 - 6\mu\right) = \psi'(\alpha)M_{k,l} + c_kc_l\psi'(\alpha)^2\left(1 - 6\frac{d}{\alpha}\right),\end{aligned}$$

using that $\mu = \frac{1}{b\mathbb{E}V - 1}$. In the Yule case, an easy adaptation of the preceding proof leads to

$$\lim_{t\to\infty}\mathbb{E}\left[\left(\psi'(\alpha)A(k,t) - e^{\alpha t}c_k\mathcal{E}\right)\left(\psi'(\alpha)A(l,t) - e^{\alpha t}c_l\mathcal{E}\right)\right] = M_{k,l} + c_kc_l.$$

6.9 Postponed estimates

6.9.1 Formula for the fourth moment of the error

Lemma 6.9.1.

$$\begin{aligned}
\mathbb{E}_t\left[(A(k,t)-c_kN_t)^4\right] &= 4\int_{[0,t]}\theta\frac{W(t)}{W(a)}\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_kN_a\right)^3\right]da\\
&+48\int_{[0,t]}\theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}N_aA(k,a)\right]\mathbb{E}_a\left[(c_kN_a-A(k,a))\right]da\\
&+24\int_{[0,t]}\theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}N_a^2\right]\mathbb{E}_a\left[(A(k,a)-c_kN_a)\right]da\\
&+24\int_{[0,t]}\theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)^2\right]\mathbb{E}_a\left[(A(k,a)-c_kN_a)\right]da\\
&+8\int_{[0,t]}\theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[(A(k,a)-c_kN_a)^3\right]da\\
&+48\int_{[0,t]}\theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)\right]\mathbb{E}_a\left[(A(k,a)-c_kN_a)^2\right]da\\
&+72\int_{[0,t]}\theta\frac{W(t)^3}{W(a)^3}\left(1-\frac{W(a)}{W(t)}\right)^2\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_kN_a\right)\right]\mathbb{E}_a\left[(A(k,a)-c_kN_a)\right]^2da\\
&+72\int_{[0,t]}\theta\frac{W(t)^3}{W(a)^3}\left(1-\frac{W(a)}{W(t)}\right)^2\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[(A(k,a)-c_kN_a)^2\right]\mathbb{E}_a\left[A(k,a)-N_ac_k\right]da\\
&+96\int_{[0,t]}\theta\frac{W(t)^4}{W(a)^4}\left(1-\frac{W(a)}{W(t)}\right)^3\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[(A(k,a)-c_kN_a)\right]^3da+c_k^4\mathbb{E}_tN_t^4
\end{aligned}$$

Proof. The proof of this Lemma lies on the calculation of the expectation of each term in the development of

$$(A(k,t)-c_kN_t)^4\,.$$

We begin by computing

$$\mathbb{E}_t\left[A(k,t)^4\right].$$

Using the formulas for the moments, we have

$$\begin{aligned}
A(k,t)^4 =& 4\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k_i} \sum_{u_{1:3}=1}^{N_{t-a}^{(t)}} \prod_{\substack{j=1\\ i\neq j}}^{3} A^{(u_j)}(k,a)\mathcal{N}(da,di) \\
=& 4\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} A^i(k,a)A^i(k,a)A^i(k,a)\mathcal{N}(da,di) \\
&+ 4\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \sum_{\substack{j_1,j_2,j_3=1\\ j_1\neq j_2\neq j_3\neq i}}^{N_{t-a}^{(t)}} A^{j_1}(k,a)A^{j_2}(k,a)A^{j_3}(k,a)\mathcal{N}(da,di) \\
&+ 12\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} A^i(k,a)A^i(k,a) \sum_{j=1,j\neq i}^{N_{t-a}^{(t)}} A^j(k,a)\mathcal{N}(da,di) \\
&+ 4\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \sum_{j=1,j\neq i}^{N_{t-a}^{(t)}} A^j(k,a)^3\mathcal{N}(da,di) \\
&+ 12\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} A^i(k,a) \sum_{j_1,j_2=1,j_1\neq j_2\neq i}^{N_{t-a}^{(t)}} A^{j_1}(k,a)A^{j_2}(k,a)\mathcal{N}(da,di) \\
&+ 24\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} A^i(k,a) \sum_{j_1=1,j_1\neq i}^{N_{t-a}^{(t)}} A^{j_1}(k,a)A^{j_1}(k,a)\mathcal{N}(da,di) \\
&+ 12\int_{[0,t]\times\mathbb{N}} \mathbb{1}_{Z_0^i(a)=k} \sum_{j_1,j_2=1,j_1\neq j_2\neq i}^{N_{t-a}^{(t)}} A^{j_1}(k,a)^2A^{j_2}(k,a)\mathcal{N}(da,di).
\end{aligned} \tag{6.33}$$

The decomposition of the sum in form

$$\sum_{u_{1:3}=1}^{N_{t-a}^{(t)}},$$

has then been made to distinguish independence properties in our calculation. Actually, as soon as, $i \neq j$, $A^i(k,a)$ is independent from $A^j(k,a)$. It is essential to note that the expectation of these integrals with respect to the random measure $\mathcal{N}$ are all calculated thanks to Theorem 4.2.2.

So, taking the expectation now leads to,

$$\begin{aligned}
\mathbb{E}_t\left[A(k,t)^4\right] =& 4\int_{[0,t]} \theta\mathbb{E}_a\left[N_{t-a}^{(t)}\right]\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)^3\right]\theta da \\
&+ 4\int_{[0,t]} \theta\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(4)}\right]\mathbb{E}_a\left[A(k,a)\right]^3 da \\
&+ 12\int_{[0,t]} \theta\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)^2\right]\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(2)}\right]\mathbb{E}_a\left[A(k,a)\right] da \\
&+ 4\int_{[0,t]} \theta\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(2)}\right]\mathbb{E}_a\left[A(k,a)^3\right] da \\
&+ 12\int_{[0,t]} \theta\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)\right]\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(3)}\right]\mathbb{E}_a\left[A(k,a)\right]^2 da \\
&+ 24\int_{[0,t]} \theta\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}A(k,a)\right]\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(2)}\right]\mathbb{E}_a\left[A(k,a)^2\right] da \\
&+ 12\int_{[0,t]} \theta\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(N_{t-a}^{(t)}\right)_{(3)}\right]\mathbb{E}_a\left[A(k,a)^2\right]\mathbb{E}_a\left[A(k,a)\right] da.
\end{aligned}$$

Using the same method for all the other terms and that, for any positive real number a lower than t,

$$N_t = \sum_{i=1}^{N_{t-a}^{(t)}} N_a^{(i)},$$

we get Lemma 6.9.1 by reassembling similar terms together. The last term is obtained using the geometric distribution of N_t under $\mathbb{P}_t$. □

6.9.2 Boundedness of the fourth moment

Lemma 6.9.2. *We begin the proof of the boundedness of the fourth moment by some estimates.*

$$\mathbb{E}_t\left[(A(k,t)-c_kN_t)\right] = \mathcal{O}\left(e^{-(\theta-\alpha)t}\right), \tag{i}$$

$$\mathbb{E}_t\left[(A(k,t)-c_kN_t)^3\right] = \mathcal{O}\left(W(t)^2\right), \tag{ii}$$

$$\mathbb{E}_t\left[(A(k,t)-c_kN_t)^2\right] = \mathcal{O}\left(W(t)\right), \tag{iii}$$

$$\mathbb{E}_tN_t^n = \mathcal{O}(e^{n\alpha t}), \quad n\in\mathbb{N}^*, \tag{iv}$$

$$\mathbb{P}_t\left(Z_0(t)=k\right) = \mathcal{O}(e^{(\alpha-\theta)t}). \tag{v}$$

Proof. Relation (i) is easily obtained using the expectation of N_t and $A(k,t)$ and the behaviour of W provided by Proposition 4.1.1. The relation (iii) has been obtained in the proof of Theorem

5.2.1. The two last relations are easily obtained from (3.3), (6.2) and Lemma 6.2.1. The relation (ii) is obtained using the following estimation,

$$\left|\mathbb{E}_t\left[\left(A(k,a)-c_kN_a\right)^3\right]\right| \leq \mathbb{E}_t\left[N_a\left(A(k,a)-c_kN_a\right)^2\right].$$

We begin the proof by computing the r.h.s. of the previous inequality using the same techniques as before.

$$\begin{aligned}
\mathbb{E}\left[A(k,t)^2N_t\right] = 2\int_0^t \theta\frac{W(t)}{W(a)}\mathbb{E}\left[N_aA(k,a)\mathbb{1}_{Z_0(a)=k}\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}\left[N_a\mathbb{1}_{Z_0(a)=k}\right]\mathbb{E}\left[A(k,a)\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}\left[A(k,a)\mathbb{1}_{Z_0(a)=k}\right]\mathbb{E}\left[N_a\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[A(k,a)N_a\right]da \\
+12\int_0^t \theta\frac{W(t)^3}{W(a)^3}\left(1-\frac{W(a)}{W(t)}\right)^2\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[A(k,a)\right]\mathbb{E}\left[N_a\right]da.
\end{aligned}$$

$$\begin{aligned}
2\mathbb{E}\left[A(k,t)N_t^2\right] = 2\int_0^t \theta\frac{W(t)}{W(a)}\mathbb{E}\left[N_a^2\mathbb{1}_{Z_0(a)=k}\right]da \\
+8\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}\left[N_a\mathbb{1}_{Z_0(a)=k}\right]\mathbb{E}\left[N_a\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[N_a^2\right]da \\
+12\int_0^t \theta\frac{W(t)^3}{W(a)^3}\left(1-\frac{W(a)}{W(t)}\right)^2\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[N_a\right]^2da.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{E}\left[N_t\left(A(k,t)-c_kN_t\right)^2\right] = 2\int_0^t \theta\frac{W(t)}{W(a)}\mathbb{E}\left[N_a\left(A(k,a)-c_kN_a\right)\mathbb{1}_{Z_0(a)=k}\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}\left[N_a\mathbb{1}_{Z_0(a)=k}\right]\mathbb{E}\left[A(k,a)-c_kN_a\right]da \\
\\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{E}\left[\left(A(k,a)-c_kN_a\right)\mathbb{1}_{Z_0(a)=k}\right]\mathbb{E}\left[N_a\right]da \\
+4\int_0^t \theta\frac{W(t)^2}{W(a)^2}\left(1-\frac{W(a)}{W(t)}\right)\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[N_a\left(A(k,a)-c_kN_a\right)\right]da \\
+12\int_0^t \theta\frac{W(t)^3}{W(a)^3}\left(1-\frac{W(a)}{W(t)}\right)^2\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}\left[N_a\right]\mathbb{E}\left[A(k,a)-c_kN_a\right]da \\
+c_k^2\mathbb{E}_tN_t^3.
\end{aligned}$$

Now, an analysis similar to the one of Lemma 6.6.5 leads to the result. □

Proof of Lemma 6.6.5. The ideas of the proof, is to analyses one to one every terms of the expression of

$$\mathbb{E}_t\left[(A(k,t)-c_k N_t)^4\right],$$

given by Lemma 6.9.1 using Lemma 6.9.2 to show that they behave as $\mathcal{O}\left(W(t)^2\right)$. Since the ideas are the same for every terms, we just give a few examples.
First of all, we consider

$$\int_{[0,t]} \frac{W(t)}{W(a)} \mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_k N_a\right)^3\right] da.$$

Using Lemma 6.9.2 (ii), we have

$$\int_{[0,t]} \frac{W(t)}{W(a)} \mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_k N_a\right)^3\right] da = \mathcal{O}\left(W(t)^2\right).$$

Now take the term

$$\int_{[0,t]} \frac{W(t)^2}{W(a)^2} \mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k} N_a^2\right] \mathbb{E}_a\left[(A(k,a)-c_k N_a)\right] da,$$

we have from Lemma 6.9.2 (i) and (iv),

$$\int_{[0,t]} \frac{W(t)^2}{W(a)^2} \mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k} N_a^2\right] \mathbb{E}_a\left[(A(k,a)-c_k N_a)\right] da \leq \int_{[0,t]} \frac{W(t)^2}{W(a)^2} \mathbb{E}_a\left[N_a^2\right] e^{-(\theta-\alpha)a} da = \mathcal{O}\left(W(t)^2\right).$$

Every term in $W(t)$ or $W(t)^2$ are treated this way. Now, we consider the term in $W(t)^4$ which is

$$I := 96 \int_{[0,t]} \frac{W(t)^4}{W(a)^4} \mathbb{P}_a\left(Z_0(a)=k\right) \mathbb{E}_a\left[(A(k,a)-c_k N_a)\right]^3 da + 24 W(t)^4 c_k^4,$$

since N_t is geometrically distributed under $\mathbb{P}_t$, and that

$$\mathbb{E}_t N_t^4 = 24 W(t)^4 - 36 W(t)^3 + \mathcal{O}(W(t)^2). \tag{6.34}$$

On the other hand, using the law of $Z_0(t)$ given by (6.2) and the expectation of $A(k,t)$ given by Theorem 6.3.3 (under $\mathbb{P}_t$), we have,

$$\begin{aligned}
&96 \int_{[0,t]} \frac{W(t)^4}{W(a)^4} \mathbb{P}_a\left(Z_0(a)=k\right) \mathbb{E}_a\left[(A(k,a)-c_k N_a)\right]^3 da \\
&= -96 W(t)^4 \int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1}\left(\int_0^a \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1} ds\right)^3 da \\
&= -24 W(t)^4 \left(\int_0^t \frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1} da\right)^4.
\end{aligned}$$

Finally,

$$I = 24W(t)^4\left(\int_t^{\infty}\frac{\theta e^{-\theta a}}{W_\theta(a)^2}\left(1-\frac{1}{W_\theta(a)}\right)^{k-1}da\right)^4 = \mathcal{O}\left(W(t)^4e^{-4\theta t}\right) = o(1).$$

The last example is the most technical and relies with the term in $W(t)^3$, which is, using (6.34) and Lemma 6.9.1,

$$\begin{aligned}J :=&72\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_kN_a\right)\right]\mathbb{E}_a\left[\left(A(k,a)-c_kN_a\right)\right]^2da\\ &+72\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(A(k,a)-c_kN_a\right)^2\right]\mathbb{E}_a\left[A(k,a)-N_ac_k\right]da\\ &-288\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(A(k,a)-c_kN_a\right)\right]^3da-36c_k^4W(t)^3.\end{aligned}$$

On the other hand, using the calculus made in the proof of Theorem 6.4.2, we have

$$\begin{aligned}&\mathbb{E}_a\left[\left(A(k,a)-c_kN_a\right)^2\right]\\ =&4\int_{[0,a]}\frac{W(a)^2}{W(s)^2}\left(1-\frac{W(s)}{W(a)}\right)\mathbb{P}_s\left(Z_0(s)=k\right)\mathbb{E}_a\left(A(k,s)-c_kN_s\right)ds\\ &+2\int_{[0,a]}\frac{W(s)}{W(a)}\;\mathbb{E}_a\left[\mathbb{1}_{Z_0(s)=k}\left(A(k,s)-c_kN_s\right)\right]ds+c_k^2W(a)^2\left(2-\frac{1}{W(a)}\right).\end{aligned}$$

Substituting this last expression in J leads to

$$\begin{aligned}J =& -144\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_kN_a\right)\right]\int_{[a,\infty]}\frac{\mathbb{P}\left(Z_0(a)=k\right)}{W(s)^2}\mathbb{E}_a\left[\left(A(k,s)-c_kN_s\right)\right]dsda\\ &+144W(t)^3\int_{[0,t]}\frac{1}{W(a)}\;\mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_kN_a\right)\right]\int_{[a,t]}\frac{1}{W(s)^2}\mathbb{P}_s\left(Z_0(s)=k\right)\mathbb{E}_a\left[A(k,s)-N_sc_k\right]da\\ &-144c_k^2\int_{[0,t]}\frac{W(t)^3}{W(a)}\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[A(k,a)-N_ac_k\right]da\\ &+144\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{P}\left(Z_0(a)=k\right)\mathbb{E}_a\left[A(k,a)-N_ac_k\right]^3da\\ &-288\int_{[0,t]}\frac{W(t)^3}{W(a)^2}\mathbb{P}_a\left(Z_0(a)=k\right)\int_{[0,a]}\frac{1}{W(s)}\mathbb{P}_s\left(Z_0(s)=k\right)\mathbb{E}_a\left(A(k,s)-c_kN_s\right)ds\mathbb{E}_a\left[A(k,a)-N_ac_k\right]da\\ &+72\int_{[0,t]}\frac{W(t)^3}{W(a)}\mathbb{P}_a\left(Z_0(a)=k\right)c_k^2\left(2-\frac{1}{W(a)}\right)\mathbb{E}_a\left[A(k,a)-N_ac_k\right]da\\ &-288\int_{[0,t]}\frac{W(t)^3}{W(a)^3}\mathbb{P}_a\left(Z_0(a)=k\right)\mathbb{E}_a\left[\left(A(k,a)-c_kN_a\right)\right]^3da-36c_k^4W(t)^3.\end{aligned}$$

Using many times that,

$$\begin{aligned}
&\int_{[0,t]} \frac{\theta \mathbb{P}\left(Z_0(a)=k\right)}{W(s)^2} \mathbb{E}_a\left[(A(k,s)-c_k N_s)\right] ds \\
&\quad = -\int_{[0,t]} \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1} \int_{[s,\infty]} \frac{\theta e^{-\theta u}}{W_\theta(u)^2}\left(1-\frac{1}{W_\theta(u)}\right)^{k-1} du ds \\
&\quad = \frac{c_k^2}{2} - \frac{1}{2}\left(\int_{[t,\infty]} \frac{\theta e^{-\theta s}}{W_\theta(s)^2}\left(1-\frac{1}{W_\theta(s)}\right)^{k-1} ds\right)^2,
\end{aligned}$$

thanks to (6.2), Theorem 6.3.3, and (3.6), we finally get

$$\begin{aligned}
J = &-144\left(c_k^2 - c_k(t)^2\right) \int_{[0,t]} \frac{W(t)^3}{W(a)^3} \mathbb{E}_a\left[\mathbb{1}_{Z_0(a)=k}\left(A(k,a)-c_k N_a\right)\right] da \\
&+ 36 W(t)^3 \left(c_k^2 \left(\int_{[t,\infty]} \frac{W(t)^3}{W(a)^3} \mathbb{E}_a\left[A(k,a)-N_a c_k\right]^3 da \right)^2 - \left(\int_{[t,\infty]} \frac{W(t)^3}{W(a)^3} \mathbb{E}_a\left[A(k,a)-N_a c_k\right]^3 da \right)^4 \right) \\
&+ 144\left(c_k - c_k(t)\right)^2 \int_{[0,t]} \frac{W(t)^3}{W(a)} \mathbb{E}_a\left[A(k,a)-N_a c_k\right] da \\
&+ 36 W(t)^3 \left(c_k - c_k(t)\right)^4.
\end{aligned}$$

This shows that J is $\mathcal{O}\left(W(t)^2\right)$. □

Chapitre 7

On the inference for size constrained Galton-Watson trees

This chapter is dedicated to a joint work with Romain Azais from team Bigs (Inria Nancy) and Alexandre Genadot from team CQFD (Inria Bordeaux). It originally arose from the idea that contour processes should be used in order to perform statistics on tree shaped data. Indeed, such objects proved to be powerful in the theoretical study of trees, and are often more convenient to manipulate than trees.

Many data are naturally modelled by an ordered tree structure : from blood vessels in biology to XML files in computer science [92] through the secondary structure of RNA in biochemistry. The statistical analysis of a dataset of hierarchical records is thus of a great interest. In particular, detecting differences in large tree structures is a complex and challenging issue. This question may be tackled via an editing distance that is the minimum number of elementary operations (insert or delete a node for example) that must be done to transform a tree into another. As a consequence, one may compute the distance matrix of a given tree dataset and thus apply any appropriate clustering method which should solve the initial statistical problem. Nevertheless, this kind of strategy is not well-adapted to exhibit the evolution of a tree structure. The space of ordered trees can not be represented in a Euclidean state space and visualizing the main differences appearing in the history of a tree data over time is often difficult. Of course, there are classical and easily-computable indicators to at least partially sum up the dynamic of a tree data : number of nodes, height, outdegree, number of leaves, etc. Each of them may be adapted to a given application. However, they all increase with the size of the tree and not really model the main structure of the data. The aim of this work is to introduce a real-valued quantity that describes the key structure of an ordered tree independently of its size.

In probability theory, we often encounter trees that have been generated from independent and identically distributed numbers of offsprings, which leads to the so-called Galton-Watson trees (sometimes also referred to as Bienaymé-Galton-Watson trees). Of course, these stochastic trees have random sizes. However, in many practical applications we are faced to random trees with a given size. We thus consider the class of Galton-Watson trees conditioned on having a certain number of nodes. This class is referred to as conditioned Galton-Watson trees. It is well-known that several classes of random trees can be seen as conditioned Galton-Watson trees [24, 51] :

Motzkin trees from the uniform offspring distribution on the set $\{0, 1, 2\}$, Catalan trees from the offspring distribution $(0.25, 0.5, 0.25)$ on $\{0, 1, 2\}$, Cayley trees from the Poisson offspring distribution, etc. We also refer the reader to [9, 3.1 Galton-Watson trees] for an enumeration of some specific parameterizations. In other words, conditioned Galton-Watson trees model a large variety of random hierarchical structures. In this paper we focus on conditioned critical Galton-Watson trees, that is to say that the expectation of the offspring distribution is 1.

Conditioned Galton-Watson trees are simple critical Galton-Watson trees which are conditioned to have a fixed size. Our main goal is to estimate the variance of the birth distribution. As for the discrete genealogy of a splitting tree, the node of a Galton-Watson tree can be labelled according to the Ulam-Harris-Neveu notation (see Chapter 3). In standard Galton-Watson trees, the number of children of each node is distributed according to a probability measure μ on $\mathbb{N}$. Moreover, if we denote by ζ_u the number of children of individual u in $\cup_{n\geq 0}\mathbb{N}^n$, then the family $(\zeta_u)_{u\in\cup_n\mathbb{N}^n}$ is assumed to be i.i.d. in the case of classical Galton-Watson trees. Hence, inferring the variance of the birth distribution is easy using for instance the empirical variance. However, in the size constrained case, one cannot expect to make estimations through standard statistic methods since the independence and homogeneity properties of the family ζ_u have been broken up by the conditioning. Such problem has already been studied in a recent work by Bharath et al. [9], in which the authors use the knowledge of the asymptotic distribution of the height an uniformly sampled node in order to make inferences from a forest of trees. Here we introduce two new estimators based on the contour processes of a forest of Galton-Watson trees which appear to have a better behaviour.

Section 7.1 is devoted to an introduction about Galton-Watson trees conditioned on having a fixed size and its contours. For discrete trees, there are many different contour processes which can be constructed from a tree. In this work we focus on the well known Harris path. Subsection 7.1.2 gives the definition of the Harris path. Subsection 7.1.3 recalls the well known limit theorem which describes the asymptotic behaviour of the Harris path as the number of nodes in the tree increases. Section 7.2 is devoted to the introduction and the study of our estimators. The estimators are introduced in Subsection 7.2.2. Subsection 7.3.2 and 7.3.3 are dedicated to the theoretical study of our estimators. Finally, in Section 7.4 we apply our methods on simulations of conditioned Galton-Watson trees.

7.1 Basics on size-constrained Galton-Watson trees

This section is devoted to an introduction to size constrained Galton-Watson trees. The next subsection simply recalls the definition. Subsection 7.1.2 gives the definition of the Harris path. Subsection 7.1.3 recalls the well known limit theorem which describes the asymptotic behaviour of the Harris path as the number of nodes in the tree increases.

7.1.1 Definition

Intuitively, a Galton-Watson tree can be seen as a tree encoding the dynamic of a population generated from some offspring distribution μ on $\mathbb{N}$. A Galton-Watson tree τ with offspring distribution μ is a random rooted tree constructed recursively as follows.

— The number of children $\zeta_\emptyset$ emanating from the root is a random variable with law μ. The first generation consists thus in $\zeta_\emptyset$ vertices.
— Assume that the n^{th} generation of children has been constructed and consists in a set of vertices $\mathcal{V}_n \subset \mathbb{N}^n$ (with the Ulam-Harris-Neveu labelling). Then, the generation $n+1$ is constructed such that $\{\zeta_v : v \in \mathcal{V}_n\}$ is a collection of independent random variables with law μ.

In the sequel, we use the notation $\mathrm{GW}(\mu)$ for the law of a Galton-Watson trees with offspring distribution μ. The asymptotic behavior of Galton-Watson trees may exhibit different regimes depending on the average number of offsprings per capita,

$$\overline{\mu} = \sum_{k\geq 0} k\mu(k).$$

— The subcritical case : $\overline{\mu} < 1$. In this case, the number of vertices is almost-surely finite with finite expectation. This means that the population almost-surely extincts and has a finite expected number of individuals.
— The critical case : $\overline{\mu} = 1$. The fact that the offspring distribution μ is critical also ensures the almost-sure finiteness of a critical Galton-Watson tree, except when $\mu(1) = 1$. When $\mu(1) < 1$, in contrary to the sub-critical case, the expected number of individuals is infinite.
— The supercritical case : $\overline{\mu} > 1$. In this case, the number of vertices explodes with positive probability.

Then, we write $\mathrm{GW}_n(\mu)$ for the law of Galton-Watson trees conditioned on having n vertices. We always state our results assuming critical Galton-Watson processes. However, this is not really a restriction since, as noted in [78, 6.3 Brownian asymptotics for conditioned Galton-Watson trees], the measure $\mathrm{GW}_n(\mu)$ is the same as $\mathrm{GW}_n(\mu_\theta)$ where, for an arbitrary $\theta > 0$ such that $g(\theta) = \sum \mu(k)\theta^k$ is finite, $\mu_\theta(k) = \mu(k)\theta^k/g(\theta)$. Therefore, in some sense, conditioned non-critical Galton-Watson trees are critical ones.

Remark 7.1.1. *An important consequence of this last remark is that we cannot estimate the mean of the distribution μ from a conditioned Galton-Watson tree (or from a forest of such trees).*

7.1.2 From ordered trees to Harris paths

In graph theory, a tree τ is a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that satisfies these two conditions : $\mathcal{G}$ is connected and has no cycles. In addition, a rooted tree is a tree in which one node has been distinguished as the root, denoted here by $r(\tau)$ (always drawn at the bottom of the tree in this chapter). In this case, the edges are assigned a natural orientation, away from the root towards the leaves. One obtains a directed rooted tree in which there exists a parent-child relationship : the parent of a node v is the first vertex met on the path to the root starting from v. The length of this path (in number of nodes) is called the height $h(v)$ of v. The set $c(v)$ of children of a vertex v is the set of nodes that have v as parent. An ordered or plane tree is a rooted tree in which an ordering has been specified for the set of children of each node, conventionally drawn from left to right. We recall from Chapter 3 that the Ulam-Harris-Neveu labelling provides a natural order.

Tree structures can be traversed in many different ways, for example using the order introduced in Chapter 3. A depth-first search algorithm traversing a tree in this order is given in Algorithm

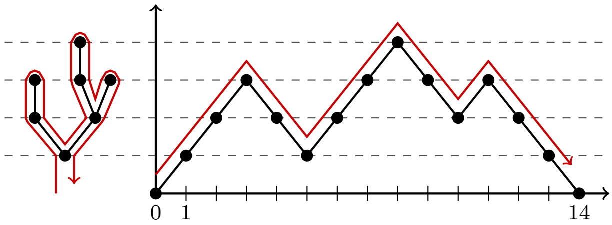

FIGURE 7.1 – Construction of the Harris path (right) from 0 to $2n = 14$ as the contour of an ordered tree (left) with $n = 7$ nodes.

1 below (where for a node v of τ, $t[v]$ denotes the subtree containing v and its children). This algorithm is a particular feature of the classical depth-first search algorithm. Indeed it is a version that returns to parent when all the descendants of a given child of the node have been visited. In this case, each node v appears $\zeta_v + 1$ times. The result is thus a sequence of length

$$\sum_{v\in\mathcal{V}}(\zeta_v + 1) = \#\mathcal{V} + \sum_{v\in\mathcal{V}}\zeta_v = 2\#\mathcal{V} - 1,$$

because the root is the only vertex not to be counted.

The Harris walk $\mathcal{H}[\tau]$ of an ordered rooted tree τ is defined from both the depth-first search returning to parent algorithm and the notion of height of nodes. $\mathcal{H}[\tau]$ is defined as a sequence of integers indexed by the set $\{0, \ldots, 2\#\mathcal{V}\}$ as follows :
— $\mathcal{H}[\tau](0) = \mathcal{H}[\tau](2\#\mathcal{V}) = 0$,
— for $1 \leq k \leq 2\#\mathcal{V}$, $\mathcal{H}[\tau](k) = h(v) + 1$ where v is the k^{th} node in pre-order (returning to parent) traversal of τ.

The Harris process is then defined as the linear interpolation of the Harris walk (see example in Figure 7.1). Note that, as displayed in Figure 7.2, the tree can be recovered from the contour such that the correspondence is one to one. In the sequel, we denote by $(\mathcal{H}[\tau](t),\ t \in [0, 2\sharp\mathcal{V}])$ the linear interpolation of the Harris walk.

Function `DFS`(τ , $l = \emptyset$)**:**
 Data: an ordered tree τ
 Result: vertices of τ in depth-first order
 add $r(\tau)$ to l
 for v **in** $c(r(\tau))$ **do**
  if *r(t[v])* **is not in** l **then**
   call `DFS` ($t[v]$,l)
   add again $r(\tau)$ to l
 return l

Algorithm 1: Recursive depth-first search.

7.1.3 Asymptotic behaviour of the Harris path

We consider a tree τ_n with distribution $\mathrm{GW}_n(\mu)$ where μ is some critical offspring distribution whose variance is denoted by σ^2. We focus on the asymptotic behavior of the Harris process

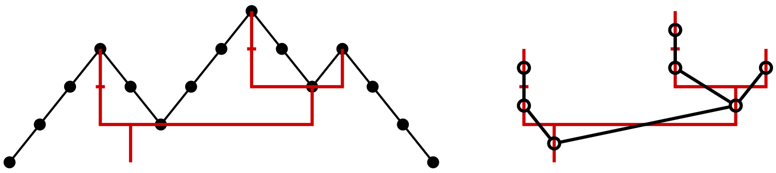

FIGURE 7.2 – The ordered tree of Figure 7.1 in its Harris path (left) : each vertical axis represents a node of the original structure (right), (See [79]). A common picture helping to see how to recover the tree from the contour is to imagine putting glue under the contour and then crushing the contour horizontally such that the inner parts of the contour which face each others are glued.

$\mathcal{H}[\tau_n](2n\cdot)$ when n tends to infinity. The convergence in distribution has been stated and proved by Aldous [1, Theorem 23].

Theorem 7.1.2. *When n goes to infinity, we have*

$$\left(\frac{\mathcal{H}[\tau_n](2nt)}{\sqrt{n}}, t \in [0,1]\right) \xrightarrow[n\to\infty]{(d)} \left(\frac{2}{\sigma} e_t, t \in [0,1]\right),$$

where e is a standard Brownian excursion, the convergence holding in law in the space $\mathcal{C}([0,1],\mathbb{R})$.

Let us simply recall that a standard Brownian excursion is a Brownian motion conditioned (for instance in the sense of h-transform) on being positive and to taking the value 0 at time 1. The density of e_t, for $0 \le t \le 1$, is given in [80, XI. 3. Bessel Bridges] and writes

$$f_{e_t}(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\sqrt{t(1-t)}^3} \exp\left(-\frac{x^2}{2t(1-t)}\right) \mathbf{1}_{\mathbb{R}_+}(x). \tag{7.1}$$

From this, we can compute some simple functionals of the excursion. For instance, we have,

$$\forall\, 0 \le t \le 1, \quad \mathbb{E}[e_t] = \frac{4}{\sqrt{2\pi}} \sqrt{t(1-t)} \quad \text{and} \quad \mathbb{E}\left[e_t^2\right] = 3t(1-t). \tag{7.2}$$

In the sequel, we denote by $(E_t,\ t \in [0,1])$ the expectation of a normalized Brownian excursion, that is

$$E_t = \mathbb{E}[e_t],\ \forall t \in [0,1].$$

The easiest way to simulate a Brownian excursion certainly is from its identity in law with a three-dimensional Bessel bridge, which simply is the Euclidean norm of a three-dimensional Brownian bridge,

$$(e_t,\ t \in [0,1]) \overset{(d)}{=} \left(\sqrt{\sum_{i=1}^{3} \left(B_t^{(i)} - tB_1^{(i)}\right)^2},\ t \in [0,1]\right) \tag{7.3}$$

where $B^{(1)}$, $B^{(2)}$ and $B^{(3)}$ are three independent Brownian motions.

The convergence presented in Theorem 7.1.2 also holds in expectation [25, Theorem 1].

Theorem 7.1.3. *When n goes to infinity, we have,*

$$\forall\, 0 \le t \le 1, \quad \mathbb{E}\left[\frac{\mathcal{H}[\tau_n](2nt)}{\sqrt{n}}\right] \xrightarrow[n\to\infty]{} \frac{2}{\sigma} E_t.$$

Note that the quantity appearing in this theorem is σ^{-1}. For this practical reason, we decided to estimate σ^{-1}.

7.2 Inferring σ^{-1} from a forest

In this section we propose two methods in order to get estimations on σ^{-1}. After what, we study the behaviour of our estimators. Let τ_n be a size-constrained Galton-Watson tree, and $\mathcal{H}[\tau_n]$ its Harris path. As usual, the idea should be to construct some operator $T : \mathcal{C}([0,1]) \mapsto \mathbb{R}_+$ such that, as n increase, $T\mathcal{H}[\tau_n]$ becomes close to σ^{-1} in some sense. However, the weak convergence given by Theorem 7.1.2 does not enabled us to expect a strong convergence. Worst, it appears according to [50] that one cannot construct on a same probability space a couple (τ_n, τ_{n+1}) such that τ_n is a subtree of τ_{n+1}. This suggest that we cannot hope to construct an estimator of σ^{-1} from only one tree. Our purpose is then to construct an efficient estimator of σ^{-1} from a forest.

7.2.1 Adequacy of the Harris path with the expected contour

Let $\tau_n \sim \mathrm{GW}_n(\mu)$ with $\overline{\mu} = 1$. We assume that the offspring distribution μ is unknown. By virtue of Theorem 7.1.3, the asymptotic average behavior of the normalized Harris process $(n^{-1/2}\mathcal{H}[\tau_n](2nt),\ 0 \leq t \leq 1)$ is given by $(2\sigma^{-1}E_t,\ 0 \leq t \leq 1)$, where σ^{-1} is obviously also unknown. We propose to estimate σ^{-1} by minimizing the $\mathbb{L}^2$-error defined by

$$\lambda \mapsto \left\| \frac{\mathcal{H}[\tau_n](2n\cdot)}{\sqrt{n}} - 2\lambda E \right\|_2^2 .$$

The solution of this least-square problem is well-known and is given by

$$\widehat{\lambda}[\tau_n] = \frac{\langle \mathcal{H}[\tau_n](2n\cdot), E\rangle}{2\sqrt{n}\|E\|_2^2}. \tag{7.4}$$

Corollary 7.2.1. *When n goes to infinity, we have*

$$\widehat{\lambda}[\tau_n] \xrightarrow{(d)} \sigma^{-1}\Lambda_\infty,$$

where the real random variable Λ_∞ is defined by

$$\Lambda_\infty = \frac{\langle e, E\rangle}{\|E\|_2^2}.$$

Proof. The result directly follows from Theorem 7.1.2 because the functional $x \mapsto \langle x, E\rangle$ is continuous on $\mathcal{C}([0,1])$. □

Remark 7.2.2. *The convergence in distribution stated in Corollary 7.2.1 seems quite unsatisfactory because this means that $\widehat{\lambda}[\tau_n]$ is not a consistent estimator of σ^{-1} and the least-square strategy thus looks like inadequate. Nevertheless, one can not expect a stronger convergence from the observation of only one stochastic process within a finite window of time. This is why one may only focus on the estimation of the parameter of interest σ^{-1} from a forest of conditioned Galton-Watson trees. This statistical framework is also considered in [9].*

Computing $\widehat{\lambda}[\tau_n]$ is only a first step in the estimation of the inverse standard deviation from a large number of conditioned Galton-Watson trees. As a consequence, the distribution of the limit variable Λ_∞ is of first importance.

Proposition 7.2.3. *The random variable Λ_∞ admits a density f_{Λ_∞} with respect to the Lebesgue measure. Furthermore,*

$$\mathbb{E}[\Lambda_\infty] = 1. \tag{7.5}$$

Proof. The existence of a density was already known [70, 71] for the random variable $\int_0^1 e_s \mathrm{d}s$. In these papers the study is performed thanks to the analysis of the double Laplace transform

$$\lambda \mapsto \int_0^\infty \exp(-\lambda t)\mathbb{E}\left[\exp\left(-t\int_0^1 e_s \mathrm{d}s\right)\right]\mathrm{d}t.$$

Thanks to the Feynmann-Kac formula, the authors express this quantity in terms of Airy functions. Then, they inverse the Laplace transform via analytical methods. Unfortunately, their method does not extend to our case. Indeed, in their case, an expression of the double Laplace transform given above is derived from the Feynmann-Kac formula for standard Brownian motion which tells us that the function

$$u(t,x) = \mathbb{E}_x\left[f(B_t)\exp\left(\int_0^t B_s ds\right)\right], \quad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{R},$$

is the solution of the PDE

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\Delta u(t,x) + xu(t,x) & \forall x \in \mathbb{R},\ t \in \mathbb{R}_+, \\ u(0,x) = f(x) & \forall x \in \mathbb{R}. \end{cases}$$

In this case, taking the Laplace transform in time of u leads to an ODE whose solution can be express in term of Airy functions (see [52]). In our case, the PDE becomes inhomogeneous in time which makes such transformation useless. As a consequence, one cannot obtain informations by this method.

That is why we propose a new method using Malliavin calculus and the representation of the Brownian excursion as a three-dimensional Bessel bridge (7.3) to show that Λ_∞ admits a density. We consider the probability space $(\mathcal{C}([0,1],\mathbb{R}^3), \mathcal{F}, \mathbb{W})$, where $\mathcal{C}([0,1],\mathbb{R}^3)$ is endowed with the topology of uniform convergence, $\mathcal{F}$ is the corresponding Borel σ-field and $\mathbb{W}$ is the Wiener measure. Let T be the continuous linear operator defined by

$$\begin{array}{rccc} T: & \mathcal{C}([0,1],\mathbb{R}^3) & \to & (\mathcal{C}([0,1],\mathbb{R}^3), \\ & \varphi & \mapsto & (T\varphi(s) = \varphi_s - s\varphi_1)\,. \end{array}$$

Let also Γ be the following function,

$$\Gamma : \varphi \mapsto \int_0^1 \|\varphi(s)\|_3 E_s \mathrm{d}s.$$

where $\|x\|$ denotes the Euclidian norm on $\mathbb{R}^3$. With these notations and (7.3), we have that the pushforward measure of $\mathbb{W}$ through the application

$$F : \varphi \mapsto \Gamma(T\varphi),$$

is the law of $\|E\|_2^2\Lambda_\infty$. In other words, the random variable F is equal in distribution to $\|E\|_2^2\Lambda_\infty$. Now for every φ in $\mathcal{C}([0,1],\mathbb{R}^3)$ such that $Leb\left(\overline{\{t\in\mathbb{R}_+\,:\,\varphi(t)=0\}}\right)=0$, we have that Γ is Frechet differentiable at point φ (where Leb denotes the Lebesgue measure). Indeed, set

$$\begin{array}{cccc} D_\varphi\Gamma: & (\mathcal{C}([0,1],\mathbb{R}^3) & \to & \mathbb{R}, \\ & h & \mapsto & \int_0^1 \frac{\langle\varphi(s),h(s)\rangle}{\|\varphi(s)\|}E_s\,\mathrm{d}s. \end{array}$$

Then, some straightforward manipulations give

$$\int_0^1\left[\|\varphi(s)+h(s)\|-\|\varphi(s)\|-\frac{\langle\varphi(s),h(s)\rangle}{\|\varphi(s)\|}\right]\mathrm{d}s=\int_0^1\left[\frac{\|h(s)\|^2+\langle\varphi(s),h(s)\rangle\left(1-\frac{\|\varphi(s)+h(s)\|}{\|\varphi(s)\|}\right)}{\|\varphi(s)+h(s)\|+\|\varphi(s)\|}\right]\mathrm{d}s.$$

Now, Cauchy-Schwarz inequality entails

$$\begin{aligned}\left|\int_0^1\left[\|\varphi(s)+h(s)\|-\|\varphi(s)\|-\frac{\langle\varphi(s),h(s)\rangle}{\|\varphi(s)\|}\right]\mathrm{d}s\right| &\le \int_0^1\left[\frac{\|h(s)\|^2+\|h(s)\|\Big|\|\varphi(s)\|-\|\varphi(s)+h(s)\|\Big|}{\|\varphi(s)+h(s)\|+\|\varphi(s)\|}\right]\mathrm{d}s\\ &\le \|h\|_\infty\int_0^1\left[\frac{\|h(s)\|+\Big|\|\varphi(s)\|-\|\varphi(s)+h(s)\|\Big|}{\|\varphi(s)+h(s)\|+\|\varphi(s)\|}\right]\mathrm{d}s.\end{aligned}$$

Now, since

$$\int_0^1\left[\frac{\|h(s)\|+\Big|\|\varphi(s)\|-\|\varphi(s)+h(s)\|\Big|}{\|\varphi(s)+h(s)\|+\|\varphi(s)\|}\right]\mathrm{d}s$$

is well-defined (because the integrand is bounded by 2) and goes to zero as $\|h\|_\infty$ goes to zero, this prove that $D_\varphi\Gamma$ is the Frechet derivative of Γ at point φ. Now, since T is linear, we have that F is Frechet differentiable at every φ such that $Leb\left(\overline{\{t\in\mathbb{R}_+\,:\,\varphi(t)=0\}}\right)=0$ and $D_\varphi F=D_{T\varphi}\Gamma\circ T$.

We now show that F belongs to the Malliavin-Sobolev space $\mathbb{D}^{1,2}$ (see [75, p. 25-27] for the definition of this space). Let h be an element of $\mathbb{L}^2([0,1],\mathbb{R}^3)$, it is easily seen that

$$\left|F(\omega+\int_0^\cdot h_s ds)-F(\omega)\right|\le\int_0^1\left\{\left\|\int_0^t h_s\mathrm{d}s\right\|+t\left\|\int_0^1 h_s\mathrm{d}s\right\|\right\}E_t\mathrm{d}t.$$

But in the right hand side of the last inequality, we have, using Jensen's inequality,

$$\begin{aligned}\int_0^1\left\{\sqrt{\sum_{i=1}^3\left(\int_0^t h_s^{(i)}\mathrm{d}s\right)^2}+t\sqrt{\sum_{i=1}^3\left(\int_0^1 h_s^{(i)}\mathrm{d}s\right)^2}\right\}E_t\mathrm{d}t &\le \int_0^1\sqrt{\sum_{i=1}^3\left(\int_0^1(h_s^{(i)})^2\mathrm{d}s\right)}(1+t)E_t\mathrm{d}t\\ &=\int_0^1\|h\|_{\mathbb{L}^2([0,1],\mathbb{R}^3)}(1+s)E_s\mathrm{d}s.\end{aligned}$$

From this, using the results of [75, p. 35], we have that F belongs to the space $\mathbb{D}^{1,2}$.

Before going further let us recall some facts on Malliavin derivative. When, working with the probability space $(\mathcal{C}([0,1],\mathbb{R}^3),\mathcal{F},\mathbb{W})$, its is known (see Section 1.2.1 in [75]) that there exists strong connexions between Malliavin derivative and Frechet derivative for a random variable G of $\mathbb{D}^{1,2}$ defined from $(\mathcal{C}([0,1],\mathbb{R}^3),\mathcal{F},\mathbb{W})$ to $\mathbb{R}$. Since, the Frechet derivative $D_\omega G$ at point ω of G is a continuous linear form from $\mathcal{C}([0,1],\mathbb{R}^3)$ into $\mathbb{R}$, it can be identified to a triple $(\mu_1^\omega,\mu_2^\omega,\mu_3^\omega)$ of σ-finite measures on $\mathbb{R}$ such that

$$D_\varphi Gh = \sum_{i=1}^{3}\int_{[0,1]} h_s^i\ \mu_i^\omega(ds), \quad \forall h \in \mathcal{C}([0,1],\mathbb{R}^3).$$

In such case, the Malliavin derivative of G is random process belonging to $\mathbb{L}^2([0,1],\mathbb{R}^3)$ given by

$$\left\{(\mu_1^\omega(u,1],\mu_2^\omega(u,1],\mu_3^\omega(u,1])\,, \quad u\in[0,1]\right\}.$$

In our case, since

$$D_\varphi Fh = \int_0^1 h_s\left\{\frac{\varphi_s - s\varphi_1}{\|\varphi_s - s\varphi_1\|}E_s\ ds - \left[\int_0^1 \frac{v(\varphi_v - v\varphi_1)}{\|\varphi_v - v\varphi_1\|}E_v\ dv\right]\ \delta_1(ds)\right\},$$

it follows that the Malliavin derivative of F is given by

$$DF = \left(\int_0^1 \frac{(\omega_s - s\omega_1)E_s}{\|\omega_s - s\omega_1\|}(\mathbb{1}_{s>u} - s)\mathrm{d}s, \quad u\in[0,1]\right) \in \mathbb{L}^2([0,1],\mathbb{R}^3).$$

Now, since DF is $\mathbb{W}$-almost everywhere not zero (in $\mathbb{L}^2([0,1],\mathbb{R}^3)$), we have using [75, Theorem 2.1.2] the existence of a density for the push-forward measure of $\mathbb{W}$ by F with respect to the Lebesgue measure.

□

It should be noted that the weak limit of $\widehat{\lambda}[\tau_n]$ has mean equal to σ^{-1} by (7.5). Moreover, it can be showed that the random variable Λ_∞ is square integrable. Indeed, since the function E is bounded, we have

$$0 \le \Lambda_\infty \le \mathcal{C}\int_0^1 e_t\mathrm{d}t,$$

for some positive constant $\mathcal{C}$. Now, its is known that the random variable $\int_0^1 e_t\mathrm{d}t$ admit moments at all order (see for instance [71]).

The variance of Λ_∞ can then be evaluated numerically in order to compare our methods with other estimators. We use Monte-Carlo simulations to produce a sample with same law as Λ_∞ to achieve this task. This lead to

$$\mathbf{V}\mathrm{ar}(\Lambda_\infty) \simeq 0.0690785.$$

At this point, it is quite interesting to compare our approach to the one developed in [9]. As in the present paper, the authors of [9] construct estimators for the inverse standard deviation of the offspring distribution of a forest of conditioned critical Galton-Watson trees. Their strategy relies on the distance to the root of a uniformly sampled vertex v of the considered tree $\tau_n \sim \mathrm{GW}_n(\mu)$,

$$\widehat{\delta}[\tau_n] = \frac{h(v)}{\sqrt{n}},$$

where we recall that $h(v)$ is the height of v in the tree. Using Theorem 7.1.2, it has been shown that $\widehat{\delta}[\tau_n]$ converges in law, when the number of nodes n goes to infinity, towards $\sigma^{-1}\Delta_\infty$ where the random variable Δ_∞ follows the Rayleigh distribution with parameter scale 1 [9, Proposition 4] with density,

$$\forall\, x \in \mathbb{R}_+,\; f_{\Delta_\infty}(x) = x \exp\left(-\frac{1}{2}x^2\right).$$

This was not noticed in [9], but we emphasize that $\widehat{\delta}[\tau_n]$ is somehow biased because $\mathbb{E}[\Delta_\infty] = \sqrt{\frac{\pi}{2}} \neq 1$. Nevertheless, one may avoid this issue by considering the quantity

$$\widehat{\delta}[\tau_n] = \sqrt{\frac{2}{\pi}}\delta[\tau_n]$$

which converges to $\sigma^{-1}\sqrt{\frac{2}{\pi}}\Delta_\infty$ which is σ^{-1} on average. As a consequence, $\widehat{\lambda}[\tau_n]$ and $\widehat{\delta}[\tau_n]$ are two quantities directly computable from the tree τ_n and that may be used to estimate the inverse standard deviation. We propose to compare them from their respective asymptotic dispersion. A first comparison may be done by computing the variances of Λ_∞ and $\sqrt{\frac{2}{\pi}}\Delta_\infty$. One has

$$\mathbf{V}\text{ar}\left(\sqrt{\frac{2}{\pi}}\Delta_\infty\right) \simeq 0.2732395 \qquad \text{and} \qquad \mathbf{V}\text{ar}(\Lambda_\infty) \simeq 0.0690785.$$

This difference in the dispersions is quite apparent in Figure 7.3 where the densities of $\sqrt{\frac{2}{\pi}}\Delta_\infty$ and Λ_∞ have been displayed. Consequently, one may expect better results in terms of dispersion from our strategy.

7.2.2 Estimation strategies

In this section, we detail two ideas in order to estimate σ^{-1} from a forest of conditioned Galton-Watson trees. A forest is define as a tuple a trees. Let N be a positive integer. In this section, we consider a forest $\mathcal{F}$ made of N independent trees $\tau^1, \dots, \tau^N$ with respective sizes $n_1, \dots, n_N$ and respective laws $GW_{n_1}(\mu), \dots, GW_{n_N}(\mu)$.

Least-square estimation

This first strategy lies on the goodness of fit between the Harris path of the forest with the expected limiting contour. This adequacy is measured thanks to an $\mathbb{L}^2$ norm.
More precisely, we denote by $(\mathcal{H}[\mathcal{F}](t),\; t \in [0, N])$ the Harris path of the forest $\mathcal{F}$. This process is defined by

$$\forall\, 0 \leq t \leq N,\; \mathcal{H}[\mathcal{F}](t) = \sum_{i=1}^{N} \frac{1}{\sqrt{n_i}} \mathcal{H}[\tau^i](2n_i(t-i+1))\mathbb{1}_{[i-1,i)}(t),$$

the Harris path of a forest being the concatenation of the Harris path of each tree, in the natural order. We propose to estimate σ^{-1} by $\widehat{\lambda}_{ls}[\mathcal{F}]$ that minimizes the $\mathbb{L}^2$ error

$$\|\mathcal{H}[\mathcal{F}](\cdot) - \lambda H(\cdot - \lfloor \cdot \rfloor)\|^2_{\mathbb{L}^2}.$$

That is

$$\widehat{\lambda}_{ls}[\mathcal{F}] = \mathrm{argmin}_{\lambda \in \mathbb{R}_+} \|\mathcal{H}[\mathcal{F}](\cdot) - \lambda H(\cdot - \lfloor \cdot \rfloor)\|^2_{\mathbb{L}^2}.$$

As aforementioned in (7.4), $\widehat{\lambda}_{ls}[\mathcal{F}]$ can be explicitly computed. Indeed, on can check that

$$\widehat{\lambda}_{ls}[\mathcal{F}] = \frac{\langle \mathcal{H}[\mathcal{F}](\cdot), H(\cdot - \lfloor \cdot \rfloor)\rangle}{\|H(\cdot - \lfloor \cdot \rfloor)\|^2_{\mathbb{L}^2}}. \tag{7.6}$$

Interestingly, $\widehat{\lambda}_{ls}[\mathcal{F}]$ is only the average of the quantities $\widehat{\lambda}[\tau^i]$ (defined in (7.4)),

$$\widehat{\lambda}_{ls}[\mathcal{F}] = \frac{1}{N} \sum_{i=1}^{N} \widehat{\lambda}[\tau^i].$$

Thus, according to Theorem 7.2.1 and 7.1.3, one can expect that $\widehat{\lambda}_{ls}[\mathcal{F}]$ tends to σ^{-1} in some sense, when both N and n_i go to infinity, by virtue of the law of large numbers.

Estimation by minimal Wasserstein distance

In the preceding method, we did not use our knowledge of the limiting distribution of the random variable of type $\lambda[\tau^n]$. In order to take this into account, one may want to test the goodness of fit between the empirical measure $\widehat{\mathcal{P}}$ defined by

$$\widehat{\mathcal{P}} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\widehat{\lambda}[\tau^i]}. \tag{7.7}$$

and the law of Λ_∞. Using Wasserstein metrics to align distributions is rather natural since it corresponds to the transportation cost between two probability laws. In particular, this feature appears to be useful in a statistical framework [16, 33]. In our case, $\widehat{\mathcal{P}}$ is expected to look like (in some sense) $\sigma^{-1}\Lambda_\infty$ in the limit of an infinite forest of infinite trees. That is why, we propose to estimate σ^{-1} with the real number λ which minimizes the distance between $\widehat{\mathcal{P}}$ and the law of $\lambda\Lambda_\infty$, denoted $\mathbb{P}_{\lambda\Lambda_\infty}$. More precisely, our estimator $\widehat{\lambda}_W[\mathcal{F}]$ is defined by

$$\widehat{\lambda}_W[\mathcal{F}] = \mathrm{argmin}_{\lambda>0} d_W\left(\widehat{\mathcal{P}}\,,\, \mathbb{P}_{\lambda\Lambda_\infty}\right), \tag{7.8}$$

where d_W denotes the Wasserstein distance of order 2.

The Wasserstein distance of order 2, denoted $d_W(\nu_1, \nu_2)$, between two probability measures ν_1 and ν_2 can be defined (see for instance [21]) from their cumulative distribution functions F_1 and F_2 as follows,

$$d_W(\nu_1, \nu_2) = \sqrt{\int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t)\right)^2 \mathrm{d}t}. \tag{7.9}$$

Let $\widehat{F}$ be the cumulative function of the empirical measure $\widehat{\mathcal{P}}$, while $F_{\lambda\Lambda_\infty}$ stands for the cumulative function of the random variable $\lambda\Lambda_\infty$. As a consequence of (7.9), one has

$$\begin{aligned} d_W\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{\widehat{\lambda}[\tau^i]}\,,\, \mathbb{P}_{\lambda\Lambda_\infty}\right)^2 &= \int_0^1 \left(\widehat{F}^{-1}(s) - F^{-1}_{\lambda\Lambda_\infty}(s)\right)^2 \mathrm{d}s \\ &= \int_0^1 \left(\widehat{F}^{-1}(s) - \lambda F^{-1}_{\Lambda_\infty}(s)\right)^2 \mathrm{d}s, \end{aligned}$$

thanks to the fact that $F^{-1}_{\lambda\Lambda_\infty} = \lambda F^{-1}_{\Lambda_\infty}$. It follows that minimizing the Wasserstein distance boils down to solve a least-square minimization problem. Hence, it comes that

$$\begin{aligned}\widehat{\lambda}_W[\mathcal{F}] &= \frac{\langle \widehat{F}^{-1}, F^{-1}_{\Lambda_\infty}\rangle}{\|F^{-1}_{\Lambda_\infty}\|^2_{\mathbb{L}^2}} \\ &= \frac{1}{\|F^{-1}_{\Lambda_\infty}\|^2_{\mathbb{L}^2}} \sum_{i=1}^{N} \widehat{\lambda}[\tau^{(i)}] \int_{\frac{i-1}{N}}^{\frac{i}{N}} F^{-1}_{\Lambda_\infty}(s)\mathrm{d}s,\end{aligned}$$

where $(\widehat{\lambda}[\tau^{(i)}])_{1\leq i\leq N}$ denotes the order statistic associated to the family $(\widehat{\lambda}[\tau^i])_{1\leq i\leq N}$.

Remark 7.2.4. *We point out the fact that there is no problem of definition in the above quantities because both $\widehat{F}^{-1}$ and $F^{-1}_{\Lambda_\infty}$ belong to $\mathbb{L}^2([0,1])$. In the first case, this follows from the fact that $\widehat{F}^{-1}$ is bounded (because $\widehat{\mathcal{P}}$ has compact support). For $F^{-1}_{\Lambda_\infty}$, this comes from the uniform sampling principle which entails that*

$$\int_0^1 F^{-1}_{\Lambda_\infty}(u)^2 \,\mathrm{d}u = \mathbb{E}[\Lambda^2_\infty].$$

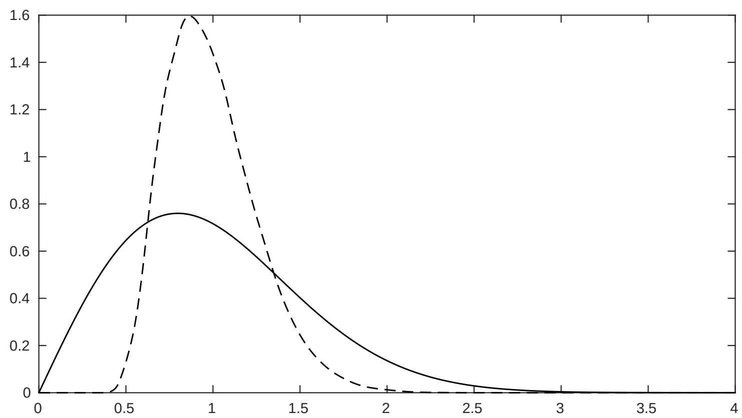

FIGURE 7.3 – Densities of $\sqrt{\frac{2}{\pi}}\Delta_\infty$ (full line) where Δ_∞ follows the Rayleigh distribution given by $f(x) = \frac{\pi}{2}x\exp\left(-\frac{\pi x^2}{4}\right)$ for $x \in \mathbb{R}_+$ and of Λ_∞ (dashed line) estimated from 100 000 simulated Brownian excursions.

7.3 Main results

7.3.1 Asymptotic regimes

In this section, we study the asymptotic properties of our estimators. Before going further, let us introduce some notations. In the sequel, the set of integer sequences is denoted $\mathbb{S}$. For any positive real number A, we denote by $\mathbb{S}_A$ the subset of $\mathbb{S}$ defined by

$$\mathbb{S}_A = \left\{ u \in \mathbb{S} \mid \min_{i\geq 1} u_i \geq A \right\}$$

In addition, for any sequence u in $\mathbb{S}$ and any positive integer N, $\vec{u}_N$ is the multi-integer made of the N first components of u, that is

$$\vec{u}_N = (u_1, \ldots, u_N)\,.$$

Moreover, for any multi-integer $\mathbf{n}$ in $\cup_{n\geq 1}\mathbb{N}^n$, we denote by $\ell(\mathbf{n})$ its number of components and by $\mathfrak{m}(\mathbf{n})$ its minimal value, that is

$$\mathfrak{m}(\mathbf{n}) = \min_{1\leq i\leq \ell(\mathbf{n})} \mathbf{n}_i.$$

Somehow, in the forests we are about to consider, $\mathfrak{m}(\mathbf{n})$ refers to the size of the smallest tree whereas $\ell(\mathbf{n})$ refers to the size of the forest.

Now, let us introduce our probabilistic framework. Let $(\tau_n^k)_{n,k\geq 1}$ be a family of conditioned Galton-Watson trees such that, for a given n, the family $(\tau_n^k)_{k\geq 1}$ is i.i.d. $GW_n(\mu)$. From this family, we define, for any mutli-integer $\mathbf{n} = (n_1, \ldots, n_N)$, the random forest $\mathcal{F}_\mathbf{n}$ made of the trees $(\tau_{n_1}^1, \ldots, \tau_{n_N}^N)$.

The idea of this construction is to consider increasing (in the sense of inclusion) sequences of random forests. Indeed, assume we are given a sequence $(u_n)_{n\geq 1}$ of integer (corresponding to the size of our trees), then the N first trees of the forest $\mathcal{F}_{\vec{u}_{N+1}}$ are the same as the trees of the forest $\mathcal{F}_{\vec{u}_N}$.

To be crystal clear, let us precise what we mean by saying that something converges as $\mathfrak{m}(\mathbf{n})$ (or $\ell(\mathbf{n})$) goes to infinity. Let f be an application from $\cup_{n\geq 1}\mathbb{N}^n$ into some metric space $(\mathcal{E}, d)$ (of course, what we are about to say trivially extend to any topological space). We say that f converges to some element e of $\mathcal{E}$ as $\mathfrak{m}(\mathbf{n})$ ($\ell(\mathbf{n})$, respectively) goes to infinity if

$$\forall\varepsilon > 0,\ \exists A \in \mathbb{R}_+,\ \forall \mathbf{n} \in \cup_{n\geq 1}\mathbb{N}^n, \quad \mathfrak{m}(\mathbf{n}) > A\ (\ell(\mathbf{n}) > A, \text{respectively})\ \Rightarrow\ d(f(\mathbf{n}), \mathbf{e}) < \varepsilon.$$

In this section, two asymptotic regimes are considered : when $\ell(\mathbf{n})$ goes to infinity (infinite forest regime) and when $\mathfrak{m}(\mathbf{n})$ goes to infinity (infinite trees regime). In the following section Corollary 7.10 and Proposition 7.3.5 are concerned with the infinite trees regime ($\mathfrak{m}(\mathbf{n}) \to \infty$) whereas Proposition 7.3.3, Lemma 7.3.6, and Proposition 7.3.7 are concerned with the infinite forest regime ($\ell(\mathbf{n}) \to \infty$).

7.3.2 Least square estimation

This first result focuses on the regime of large trees.

Corollary 7.3.1. *When* $\mathfrak{m}(\mathbf{n})$ *goes to infinity, we have*

$$\widehat{\lambda}_{ls}[\mathcal{F}_\mathbf{n}] \xrightarrow{(d)} \sigma^{-1} \frac{1}{\ell(\mathbf{n})} \sum_{i=1}^{\ell(\mathbf{n})} \Lambda_{\infty,i}, \tag{7.10}$$

where the $\Lambda_{\infty,i}$*'s are* N *independent copies of* Λ_∞. *Furthermore, when* $\ell(\mathbf{n})$ *is fixed and* $\mathfrak{m}(\mathbf{n})$ *goes to infinity, we have*

$$\mathbb{E}\left[\widehat{\lambda}_{ls}[\mathcal{F}_\mathbf{n}]\right] \longrightarrow \sigma^{-1}.$$

Proof. The first convergence is a direct consequence of the independence properties of the family $(\tau^i_{\mathbf{n}_i})_{1\leq i\leq l(\mathbf{n})}$ and the fact that each one converges to a random variable with law Λ_∞ by Corollary 7.2.1. Now, its remains to prove the second statement. Since the family $(\tau^i_{\mathbf{n}_i})_{1\leq i\leq \ell(\mathbf{n})}$ is made of independent random variables its follows from Theorem 7.1.3 and the definition (7.4) of $\widehat{\lambda}[\tau_{n_i}]$ that the proof of this last statement boils down to prove that

$$\int_0^1 \mathbb{E}\left[\frac{\mathcal{H}[\tau_n](2ns)}{\sqrt{n}}\right] E_s \,\mathrm{d}s \underset{n\to\infty}{\longrightarrow} \frac{2}{\sigma}\int_0^1 E_s^2 \,\mathrm{d}s,$$

where τ_n is some tree with law $GW_n(\mu)$. Its is known from [25, Lemma 4] that, for any positive integer n and real number $0<t<1$,

$$\forall x\in\mathbb{R}_+,\ \mathbb{P}\left(\frac{\mathcal{H}[\tau^i_n](2nt)}{\sqrt{n}}>x\right)\leq \frac{C}{t}\exp\left(\frac{-Dx}{\sqrt{t}}\right). \tag{7.11}$$

From this last estimate, one can easily shows that $\mathbb{E}\left[\frac{\mathcal{H}[\tau_n](2ns)}{\sqrt{n}}\right]$ is uniformly bounded (w.r.t. n) by $t\mapsto \frac{C}{D\sqrt{t}}$ which is integrable on $[0,1]$. Finally, the result follows from Theorem 7.1.3 and the dominated convergence Theorem. □

Remark 7.3.2. *It is worth noting that the limit appearing in the right hand side of* (7.10) *is unbiased, that is*

$$\mathbb{E}\left[\sigma^{-1}\frac{1}{\ell(\mathbf{n})}\sum_{i=1}^{\ell(\mathbf{n})}\Lambda_{\infty,i}\right]=\sigma^{-1}.$$

Moreover, (7.5) *and the law of large numbers entails that this same limit converges a.s. to* σ^{-1} *as* $\ell(\mathbf{n})$ *goes to infinity.*

The following result states a stronger convergence when $\ell(\mathbf{n})$ goes to infinity before $\mathfrak{m}(\mathbf{n})$. The spirit of this result is that, given an increasing sequence of random forest, the least square estimator cannot be too far from σ^{-1} as soon as the size of the trees are large enough. It particular, due to the results of [50], one cannot expect a stronger convergence.

Proposition 7.3.3. *We have,*

$$\forall\epsilon>0,\ \exists A\in\mathbb{N},\ \forall u\in\mathbb{S}_A,\ \mathbb{P}\left(\limsup_{N\to\infty}\left|\widehat{\lambda}_{ls}[\mathcal{F}_{\vec{u}_N}]-\sigma^{-1}\right|<\epsilon\right)=1.$$

Proof. We begin the proof by showing that the family $\left(\widehat{\lambda}[\tau^k_n]\right)_{n,k\in\mathbb{N}}$ has uniformly bounded fourth moments. On one hand, we know from [25, Lemma 4] that, for any positive integers i,n and real number $0<t<1$,

$$\forall x\in\mathbb{R}_+,\ \mathbb{P}\left(\frac{\mathcal{H}[\tau^i_n](2nt)}{\sqrt{n}}>x\right)\leq \frac{C}{t}\exp\left(\frac{-Dx}{\sqrt{t}}\right). \tag{7.12}$$

On the other hand, by Jensen's inequality, there exists a positive constant c such that

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{\lambda}[\tau^i_n]\right)^4\right] &\leq c\int_0^1 \mathbb{E}\left[\left(\frac{\mathcal{H}[\tau^i_n](2ns)}{\sqrt{n}}\right)^4\right]\mathrm{d}s\\ &= 4c\int_0^1\int_{\mathbb{R}_+} x^3\,\mathbb{P}\left(\frac{\mathcal{H}[\tau^i_n](2ns)}{\sqrt{n}}>x\right)\mathrm{d}x\ \mathrm{d}s.\end{aligned}$$

Finally, using (7.12) gives the desired bound,

$$\mathbb{E}\left[\widehat{\lambda}[\tau_n^i]^4\right] \leq \frac{12cC}{D^4}. \tag{7.13}$$

From this point we consider a sequence u of integers. This sequence corresponds to the sizes of the trees in our increasing sequence of random forests $(\mathcal{F}_{\vec{u}_N})_{N\geq 1}$. We recall according to the definitions given in the beginning of this section that the random forest $\mathcal{F}_{\vec{u}_N}$ is composed of the trees $(\tau_{u_1}^1, \ldots, \tau_{u_N}^N)$.

Let $m_{u_i}^i$ be the expectation of $\widehat{\lambda}[\tau_{u_i}^i]$. It is worth noting that this expectation depends only on the integer u_i. Now, using the uniform bound on the fourth moment, it is easy to show using standard methods that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\lambda}[\tau_{u_i}^i] - m_{u_i}^i\right) \xrightarrow{a.s.} 0, \tag{7.14}$$

when N goes to infinity. Moreover, using Theorem 7.1.3, we have that $m_{u_i}^i$ converges to σ^{-1} as u_i goes to infinity, form which it follows that for any $\epsilon > 0$, there exists an integer A such that

$$\left|m_{u_i}^i - \sigma^{-1}\right| < \epsilon \tag{7.15}$$

whenever $u_i > A$. Finally, letting all the u_i's be greater than A, we have that there exists a measurable set Ω_u , with mass 1, such that, using (7.14) and (7.15), for all ω in this set,

$$\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{i=1}^{N}\widehat{\lambda}[\tau_{u_i}^i](\omega) - \sigma^{-1}\right| \leq \limsup_{N\to\infty}\frac{1}{N}\left|\sum_{i=1}^{N}\widehat{\lambda}[\tau_{u_i}^i](\omega) - m_{u_i}^i\right| + \limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left|m_{u_i}^i - \sigma^{-1}\right| \leq \epsilon.$$

□

Remark 7.3.4. *According to the proof the preceding theorem, it would be very interesting to have estimates on the rate of convergence in the result stated in Theorem 7.1.3. Indeed, this would enable us to have estimate on the error in the convergence stated in Proposition 7.3.3 given in terms of the smallest trees in the increasing sequence of random forests.*

7.3.3 Estimation by minimal Wasserstein distance

As in the preceding section, we begin by looking at the convergence in $\mathfrak{m}(\mathbf{n})$.

Proposition 7.3.5. *When $\mathfrak{m}(\mathbf{n})$ goes to infinity (and $\ell(\mathbf{n})$ is fixed), we have*

$$\widehat{\lambda}_W[\mathcal{F}_{\mathbf{n}}] \xrightarrow{(d)} \frac{1}{\sigma\|F_{\Lambda_\infty}^{-1}\|_2^2}\sum_{i=1}^{\ell(\mathbf{n})}\Lambda_{\infty,(i)}\int_{\frac{i-1}{\ell(\mathbf{n})}}^{\frac{i}{\ell(\mathbf{n})}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s,$$

where the $\Lambda_{\infty,(i)}$'s are N independent copies of Λ_∞ sorted in increasing order. In addition, the limit is asymptotically unbiased, in the sense that, when $\ell(\mathbf{n})$ goes to infinity,

$$\frac{1}{\sigma\|F_{\Lambda_\infty}^{-1}\|_2^2}\mathbb{E}\left[\sum_{i=1}^{\ell(\mathbf{n})}\Lambda_{\infty,(i)}\int_{\frac{i-1}{\ell(\mathbf{n})}}^{\frac{i}{\ell(\mathbf{n})}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s\right] \longrightarrow \frac{1}{\sigma}.$$

Proof. The convergence in distribution is straightforward from Corollary 7.2.1 and standard methods on order statistics. We now prove that the estimator is asymptotically unbiased. In order to lighten the notation, let us set

$$N = \ell(\mathbf{n}).$$

It is well known, since Λ_∞ has a density, that, for any $1 \leq i \leq N$, one has (see for instance [19])

$$\mathbb{E}\left[\Lambda_{\infty,(i)}\right] = N\binom{N-1}{i-1}\int_0^\infty xF_{\Lambda_\infty}(x)^{i-1}(1-F_{\Lambda_\infty}(x))^{N-i}f_{\Lambda_\infty}(x)\mathrm{d}x.$$

Hence,

$$\mathbb{E}\left[\sum_{i=1}^N \Lambda_{\infty,(i)}\int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s\right]$$
$$= N\int_0^\infty xf_{\Lambda_\infty}(x)\sum_{i=1}^N\binom{N-1}{i-1}F_{\Lambda_\infty}(x)^{i-1}(1-F_{\Lambda_\infty}(x))^{N-i}\int_0^{\frac{1}{N}} F_{\Lambda_\infty}^{-1}\left(s+\frac{i-1}{N}\right)\mathrm{d}s\ \mathrm{d}x.$$

This rewrites thanks to the right inverse sampling principle as

$$\mathbb{E}\left[\sum_{i=1}^N \Lambda_{\infty,(i)}\int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s\right] = \int_0^1 F_{\Lambda_\infty}^{-1}(x)K_n\left(F_{\Lambda_\infty}^{-1}\right)(y)\ \mathrm{d}y,$$

where K_n is defined for all function f in $L^2([0,1])$ by

$$K_n(f)(y) = \sum_{i=1}^N\binom{N-1}{i-1}y^{i-1}(1-y)^{N-i}\int_0^{\frac{1}{N}} f\left(s+\frac{i-1}{N}\right)\mathrm{d}s, \quad \forall y\in[0,1].$$

The operators K_n are known as Bernstein-Kantorovich operators which were introduce in 1930 by Kantorovich in order to extend the properties of Bernstein polynomials to non-continuous functions (see [54]). In particular, it is known that, for all f in $L^2([0,1])$, $K_n(f)$ converges strongly to f in $L^2([0,1])$ (see [69] for an old but practical reference).

Now, according to Cauchy-Schwarz inequality we have that

$$\left|\int_0^1 F_{\Lambda_\infty}^{-1}(x)K_n\left(F_{\Lambda_\infty}^{-1}\right)(y)\ \mathrm{d}y - \int_0^1 F_{\Lambda_\infty}^{-1}(y)^2\ \mathrm{d}y\right| \leq \|F_{\Lambda_\infty}^{-1}\|_{L^2}^2\int_0^1\left|K_n(F_{\Lambda_\infty}^{-1})(y)-F_{\Lambda_\infty}^{-1}(y)\right|^2\mathrm{d}y.$$

But since, $K_n(F_{\Lambda_\infty}^{-1})$ converges to $F_{\Lambda_\infty}^{-1}$ in $L^2([0,1])$, we finally obtain

$$\mathbb{E}\left[\sum_{i=1}^N \Lambda_{\infty,(i)}\int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s\right]\underset{N\to\infty}{\longrightarrow}\|F_{\Lambda_\infty}^{-1}\|_{L^2}^2,$$

leading to the result. □

In addition, we have the same kind of strong convergence result for this estimator. Its lies on the fact that the empirical measure $\widehat{\mathcal{P}}$ defined in (7.7) must be close (in Wasserstein distance) to the law of $\sigma^{-1}\Lambda_\infty$ as soon as the trees are large enough. More precisely, we have the following lemma.

Lemma 7.3.6. *Let $\mathcal{P}$ be the law of $\sigma^{-1}\Lambda_\infty$. Let also $\mathcal{P}_{\mathbf{n}}$ be the empirical distribution defined for any multi-integer $\mathbf{n}$ by*

$$\mathcal{P}_{\mathbf{n}} = \frac{1}{\ell(\mathbf{n})} \sum_{i=1}^{\ell(\mathbf{n})} \delta_{\widehat{\lambda}[\tau^i_{\mathbf{n}_i}]}.$$

Then, the following statement holds,

$$\forall\, \epsilon > 0,\ \exists\, A \in \mathbb{N},\ \forall\, u \in \mathbb{S}_A, \quad \mathbb{P}\left(\limsup_{N\to\infty} d_W\left(\mathcal{P}_{\vec{u}_N}, \mathcal{P}\right) < \epsilon \right) = 1.$$

Proof. Let $\mathbf{n}$ be a multi-integer. Let Π_δ being the canonical projection of $\mathbb{R}$ on $[-\delta, \delta]$, for a positive real number δ. We have

$$d_W\left(\mathcal{P}_{\mathbf{n}}(\omega), \mathcal{P}\right) \leq d_W\left(\mathcal{P}_{\mathbf{n}}(\omega), \Pi_\delta \mathcal{P}_{\mathbf{n}}(\omega)\right) + d_W\left(\Pi_\delta \mathcal{P}_{\mathbf{n}}(\omega), \Pi_\delta \mathcal{P}\right) + d_W\left(\mathcal{P}, \Pi_\delta \mathcal{P}\right), \tag{7.16}$$

where $\Pi_\delta \mu$ denotes the image measure of μ by Π_δ. To obtain the desired result, we need to control each of the three terms in the right hand side of (7.16).

— **Third term.** First, it is clear, for any probability measure μ, that Π_δ is a transport of μ on $\Pi_\delta\mu$ which needs not to be optimal [11, 2. Generalities on Kantorovich transport distances]. Hence,

$$d_W\left(\mu, \Pi_\delta\mu\right) \leq \sqrt{\int_{\mathbb{R}} |x - \Pi_\delta(x)|^2 \,\mu(dx)}.$$

It follows, since $x \to x^2$ is integrable with respect to $\mathcal{P}$, that δ can be chosen in order to have

$$d_W\left(\mathcal{P}, \Pi_\delta\mathcal{P}\right) \leq \sqrt{\mathbb{E}\left[(\Lambda_\infty)^2\, \mathbf{1}_{|\Lambda_\infty|>\delta}\right]} < \frac{\varepsilon}{3}. \tag{7.17}$$

— **First term.** On the other hand, following the same lines as in the proof of Proposition 7.3.3, one can shows that

$$\forall\, \epsilon > 0,\ \exists\, A \in \mathbb{N},\ \forall\, u \in \mathbb{S}_A, \quad \mathbb{P}\left(\limsup_{N\to\infty} \left| \frac{1}{N} \sum_{i=1}^{N} \widehat{\lambda}[\tau^i_{u_i}]^2 \mathbf{1}_{|\widehat{\lambda}[\tau^i_{u_i}]|>\delta} - \mathbb{E}\left[\Lambda_\infty^2 \mathbf{1}_{|\Lambda_\infty|>\delta}\right] \right| < \epsilon \right) = 1. \tag{7.18}$$

This bound allows us to control the first term in the right hand side of (7.16) since

$$\begin{aligned} d_W\left(\mathcal{P}_{\mathbf{n}}(\omega), \Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega)\right) &\leq \sqrt{\int_{\mathbb{R}} |x - \Pi_\delta(x)|^2\, \mathcal{P}_{\mathbf{n}}(\omega)(dx)} \\ &\leq \sqrt{\frac{1}{N} \sum_{i=1}^{N} \widehat{\lambda}[\tau^i_{n_i}](\omega)^2 \mathbf{1}_{|\widehat{\lambda}[\tau^i_{n_i}](\omega)|>\delta}}. \end{aligned}$$

Hence, it remains to control the second term.

— **Second term.** Since $\Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega)$ and $\Pi_\delta\mathcal{P}$ are compactly supported measure, for any multi-integer $\mathbf{n}$, we have the following duality formula for the first order Wasserstein distance (which we denote W_1),

$$W_1\left(\Pi_\delta\mathcal{P}_{\vec{n}}(\omega), \Pi_\delta\mathcal{P}\right) = \sup_{\phi\in \mathrm{Lip}_1([-\delta,\delta])} \left\{ \left| \int_{\mathbb{R}} \phi(x)(\Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega)(\mathrm{d}x) - \Pi_\delta\mathcal{P}(\mathrm{d}x)) \right| \right\},$$

where $Lip_1([-\delta,\delta])$ denotes the set of 1-Lipschitz continuous function on $[-\delta,\delta]$. Since, $[-\delta,\delta]$ is compact, $Lip_1([-\delta,\delta])$ is separable endowed with the uniform topology. This implies the existence of a countable family $(f_k)_{k\geq 1}$ which is dense. Using again the method of the proof of Proposition 7.3.3, one can show

$$\forall\,\epsilon>0,\ \exists\,A\in\mathbb{N},\ \forall\,u\in\mathbb{S}_A,\quad \mathbb{P}\left(\limsup_{N\to\infty}|\mathcal{P}_{\vec{u}_N}f_k-\mathcal{P}f_k|<\epsilon\right)=1, \tag{7.19}$$

where $\mathcal{P}f$ denotes $\int_{\mathbb{R}} f(x)\mathcal{P}(dx)$. Now, the density of $(f_k)_{k\geq 1}$ entails that for any function f in $Lip_1([-\delta,\delta])$, one can finds a function f_k such that $\|f_k-f\|_\infty<\varepsilon$, for any positive ε. Hence, (7.19) holds for any function in $\mathcal{C}_K$ on the same event. Moreover, since $\Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega)$ and $\Pi_\delta\mathcal{P}$ are compactly supported measures,

$$d_W\left(\Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega),\Pi_\delta\mathcal{P}\right)\leq C\sqrt{W_1\left(\Pi_\delta\mathcal{P}_{\mathbf{n}}(\omega),\Pi_\delta\mathcal{P}\right)},$$

which implies

$$\forall\,\epsilon>0,\ \exists\,A\in\mathbb{N},\ \forall\,u\in\mathbb{S}_A,\quad \mathbb{P}\left(\limsup_{N\to\infty}d_W\left(\Pi_\delta\mathcal{P}_{\vec{u}_N}(\omega),\Pi_\delta\mathcal{P}\right)<\epsilon\right)=1. \tag{7.20}$$

Finally, using (7.17), (7.18) and (7.20) in (7.16) leads to the result. □

Proposition 7.3.7. *We have,*

$$\forall\,\epsilon>0,\ \exists\,A\in\mathbb{N},\ \forall\,u\in\mathbb{S}_A,\quad \mathbb{P}\left(\limsup_{N\to\infty}\left|\widehat{\lambda}_W[\mathcal{F}_{\vec{u}_N}]-\frac{1}{\sigma}\right|<\epsilon\right)=1.$$

Proof. By the Cauchy-Schwarz inequality, the convergence of this estimator follows from the convergence of the Wasserstein distance in the following manner,

$$\begin{aligned}\left|\widehat{\lambda}_W[\mathcal{F}_{\mathbf{n}}]-\frac{1}{\sigma}\right| &= \frac{\left|\langle\widehat{F}[\mathcal{F}_{\mathbf{n}}]^{-1}-\sigma^{-1}F_{\Lambda_\infty}^{-1},F_{\Lambda_\infty}^{-1}\rangle\right|}{\|F_{\Lambda_\infty}^{-1}\|_2^2}\\ &\leq \frac{\left\|\widehat{F}[\mathcal{F}_{\mathbf{n}}]^{-1}-\sigma^{-1}F_{\Lambda_\infty}^{-1}\right\|_2\|F_{\Lambda_\infty}^{-1}\|_2}{\|F_{\Lambda_\infty}^{-1}\|_2^2}\\ &= \frac{d_W\left(\mathcal{P}_{\mathbf{n}},\mathcal{P}\right)}{\|F_{\Lambda_\infty}^{-1}\|_2}.\end{aligned}$$

The result finally arises from the preceding Lemma. □

7.4 Numerical simulations

7.4.1 Simulation of conditioned Galton-Watson trees

In order to test our estimation techniques on Galton-Watson forests, we need to make some numerical experiment. However, simulation of conditioned Galton-Watson tree is a difficult problem of independent importance. In this section, we briefly present an algorithm due to Devroye [24] allowing to achieve this aim. Note that, it is (with a direct rejection method) the only known (in the best of our knowledge) algorithm allowing to simulate size constrained Galton-Watson trees.

The main idea of the algorithm is the following : assume that μ is supported on $\{0, \dots, K\}$, for an integer K. If N_0 denotes the number of individuals with no children, N_1 the number of individuals with 1 children and so on... Then, the sequence $(N_0, \dots, N_K)$ is distributed following a multinomial distribution with parameter n and $(\mu_k)_{0\leq k\leq K}$ conditioned to have

$$\sum_{i=0}^{k} iN_i = n-1.$$

— **Simulation of numbers of children.** The multinomial distribution of parameters $(\mu_k)_{0\leq k\leq K}$ and n may be defined by its probability mass function,

$$\mathbb{P}(N_0 = n_0, \dots, N_K = n_K) = \begin{cases} \dfrac{n!}{n_0! \dots n_K!}\mu_0^{n_0}\dots\mu_K^{n_K} & \text{if} \quad \displaystyle\sum_{k=0}^{K} n_k = n, \\ 0 & \text{else.} \end{cases}$$

Simulation of the multinomial distribution presents no difficulty. By rejection sampling, we simulate multinomial random variables until obtaining a sequence $(N_0)_{0\leq k\leq K}$ satisfying

$$\sum_{k=0}^{K} kN_k = n-1.$$

We define the sequence $(\zeta_i)_{1\leq i\leq n}$ from

$$(\zeta_i)_{1\leq i\leq n} = (\underbrace{0,\dots,0}_{N_0}, \underbrace{1,\dots,1}_{N_1}, \; \dots \; , \underbrace{K,\dots,K}_{N_K}).$$

Let $(\xi_i)_{1\leq i\leq n}$ be a sequence obtained as a random permutation of $(\zeta_i)_{1\leq i\leq n}$. A suitable technique for random shuffling is presented in [58, Algorithm P (p.139)]. The sequence $(\xi_i)_{1\leq i\leq n}$ represents the vertices's numbers of children in the depth-first search order.

— **Computation of Łukasciewicz walk.** Let L be the process defined by $L(0) = 0$ and,

$$\forall\, 0 \leq k \leq n-2, \quad L(k+1) = L(k) + \xi_{k+1} - 1.$$

Set $l = 1 + \operatorname{argmin}\{L(k) \ : \ 0 \leq k \leq n-1\}$. Then there exists a tree τ_n with n nodes whose Łukasciecwicz walk is defined by

$$\mathcal{L}[\tau_n](k) = \begin{cases} L(l+k) + \min L - 1 & \text{if} \quad 0 \leq k \leq n-1-l, \\ L(k-n+l) + \min L - 1 & \text{if} \quad n-l \leq k \leq n-1. \end{cases}$$

The computation of $\mathcal{L}[\tau_n]$ from L is illustrated in Figure 7.4.

— **From Łukasiewicz walk to height process.** Now, we compute the corresponding height process [26, eq.(2)],

$$\forall\, 0 \leq k \leq n-1, \quad \mathfrak{H}[\tau_n](k) = \#\left\{0 \leq j \leq k-1 \ : \ \mathcal{L}[\tau_n](j) = \min_{j\leq l\leq n} \mathcal{L}[\tau_n](l)\right\}.$$

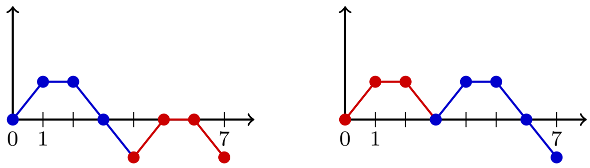

FIGURE 7.4 – Illustration of the re-arrangement procedure.

— **From height process to contour process.** Let $(b_k)_{0\leq k\leq n-1}$ be the sequence defined from $b_k = 2k - \mathfrak{H}[\tau_n](k)$ if $0 \leq k \leq n-1$ and $b_n = 2(n-1)$. Then the b_i's are sorted in increasing order. The contour process $\mathcal{C}[\tau_n](k)$ is defined for any $0 \leq k \leq 2n-2$ by [26, eq.(1)]

$$\mathcal{C}[\tau_n](k) = \begin{cases} \mathfrak{H}[\tau_n](i) - (k - b_i) & \text{if } \ \exists\, 0 \leq i \leq n-2, \quad b_i \leq k < b_{i+1} - 1, \\ k - b_{i+1} + \mathfrak{H}[\tau_n](i+1) & \text{if } \ \exists\, 0 \leq i \leq n-2, \quad b_{i+1} - 1 \leq k < b_{i+1}, \\ \mathfrak{H}[\tau_n](b_{n-1}) - (k - b_{n-1}) & \text{if } \qquad\qquad\qquad\qquad\quad b_{n-1} \leq k \leq b_n. \end{cases}$$

— **From contour process to Harris path.** The Harris path is only a small modification of the contour process, defined by $\mathcal{H}[\tau_n](0) = \mathcal{H}[\tau_n](2n) = 0$ and

$$\forall\, 1 \leq k \leq 2n-1, \ \mathcal{H}[\tau_n](k) = \mathcal{C}[\tau_n](k-1) + 1.$$

7.4.2 Inference for a forest of binary size-constrained Galton-Watson trees

The aim of this section is to analyze the finite-sample behavior of both estimators introduced in this chapter by means of numerical experiments. The theoretical study achieved in Section 7.3 shows that we can expect to obtain good numerical results, at least for large trees and/or a large forest. To this goal, we consider a forest of independent conditioned Galton-Watson trees with common critical birth distribution μ such that $\mu(k) = 0$ for $k \geq 3$. In such case, μ is entirely characterized by its variance σ^2. Simulations of Galton-Watson trees $\mathrm{GW}_n(\mu)$ are performed with the method provided in Subsection 7.4.1.

Let $\mathcal{F} = (\tau^i)_{1\leq i\leq N}$ be a forest of N independent trees such that, for any $1 \leq i \leq N$, $\tau^i \sim \mathrm{GW}_{n_i}(\mu)$ for some integer n_i. From the Harris process of each tree τ^i, one first computes the quantity

$$\widehat{\lambda}\left[\tau^i\right] \quad = \quad \frac{\langle \mathcal{H}[\tau^i](2n_i \cdot), E\rangle}{2\sqrt{n_i}\|E\|_2^2},$$

where E is known and defined in (7.2). Then, we propose to estimate σ^{-1} in the two following ways.

Least Squares	Wasserstein
$\widehat{\lambda}_{ls}[\mathcal{F}] = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\widehat{\lambda}\left[\tau^i\right]$	$\widehat{\lambda}_W[\mathcal{F}] = \dfrac{1}{\|F_{\Lambda_\infty}^{-1}\|_2^2}\displaystyle\sum_{i=1}^{N}\widehat{\lambda}\left[\tau^{(i)}\right]\int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s$

Remark 7.4.1. *In order to compute $\widehat{\lambda}_W[\mathcal{F}]$, we need to be able to perform computations using the function $F_{\Lambda_\infty}^{-1}$. Unfortunately, in view of the theoretical study of Λ_∞ made in Subsection 7.2.1, one cannot expect to have an explicit expression for this function. In the following of this section, we use a numerical estimation of $F_{\Lambda_\infty}^{-1}$ by Monte Carlo simulations. To achieve this goal, we perform simulations of Λ_∞ by simulating Brownian excursion thanks to* (7.3). *In order to ensure that the error made on $F_{\Lambda_\infty}^{-1}$ does not propagate too much in our results, F_{Λ}^{-1} is estimated with an important sample of simulations of Λ_∞ (exactly 10^6 simulations).*

The theoretical investigations of Section 7.3 establish that our estimators are unbiased in the "infinite trees" regime $\mathfrak{m}(\mathbf{n}) \to \infty$. Nevertheless, the problem is not as simple when working with finite trees. A clear illustration of this comes from the numerical evaluations of the average Harris processes of finite trees. Indeed, the numerical study of Figure 7.5 shows that the average Harris processes of small trees seem to be lower than the limiting Harris process. Hence, the quantities $\widehat{\lambda}[\tau^i]$ are expected to underestimate the target σ^{-1}. But any estimator based on the asymptotic behavior of conditioned Galton-Watson trees is expected to present such a bias. In particular, we state in our numerical experiments that the estimator proposed in [9] presents the same bias.

The natural question arising from the preceding comments is : how is the bias of a conditioned Galton-Watson tree related to its size and/or the unknown parameter σ ? The numerical study presented in Figure 7.6 shows that the quantity $\eta(n) = \sigma^{-1}\mathbb{E}[\widehat{\lambda}[\tau_n]]^{-1}$, where $\tau_n \sim \mathrm{GW}_n(\mu)$, seems close to be uncorrelated to σ at least when σ is large enough. This allows us to construct a bias corrector independent on the unknown standard deviation σ. In addition, the dependency on n may be modeled by the relation $\eta(n) = 1 - (a\sqrt{n} + b)^{-1}$. The coefficients appearing in η may be estimated from simulated data,

$$\widehat{\eta}(n) = 1 - (0.504273\sqrt{n} + 0.9754839)^{-1}$$

(see Figure 7.6 again). The correction is obviously expected to be better for large values of σ. Finally, we construct the following corrected versions of the estimators $\widehat{\lambda}_{ls}[\mathcal{F}]$ and $\widehat{\lambda}_W[\mathcal{F}]$.

Corrected Least Squares	Corrected Wasserstein
$\widehat{\lambda}^c_{ls}[\mathcal{F}] = \frac{1}{N}\sum_{i=1}^{N} \widehat{\eta}(\#\tau^i)\widehat{\lambda}\left[\tau^i\right]$	$\widehat{\lambda}^c_W[\mathcal{F}] = \frac{1}{\Vert F_{\Lambda_\infty}^{-1}\Vert_2^2}\sum_{i=1}^{N} \widehat{\eta}\left(\#\tau^{(i)}\right)\widehat{\lambda}\left[\tau^{(i)}\right]\int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\Lambda_\infty}^{-1}(s)\mathrm{d}s$

Computing the estimators proposed in this chapter is not an easy task. According to Remark 7.4.1, this needs to perform an important number of simulations of Λ_∞ in order to get an accurate approximation of $F_{\Lambda_\infty}^{-1}$. Moreover, to be able to correct the bias highlighted above, one needs to perform many simulations of finite trees. Together with this work, we propose a Matlab toolbox which already includes these preliminary computations and allows to directly compute our estimators for a forest. This toolbox as well as its documentation and the scripts used in this chapter are available at the page : *http ://agh.gforge.inria.fr.*

The study of Figure 7.7 shows that for values of σ greater than 0.5, the bias correction works properly. Moreover, it also shows that the estimator developed in [9] present the same kind of bias as ours, which can also be corrected. In the case of small parameter σ, the bias correction

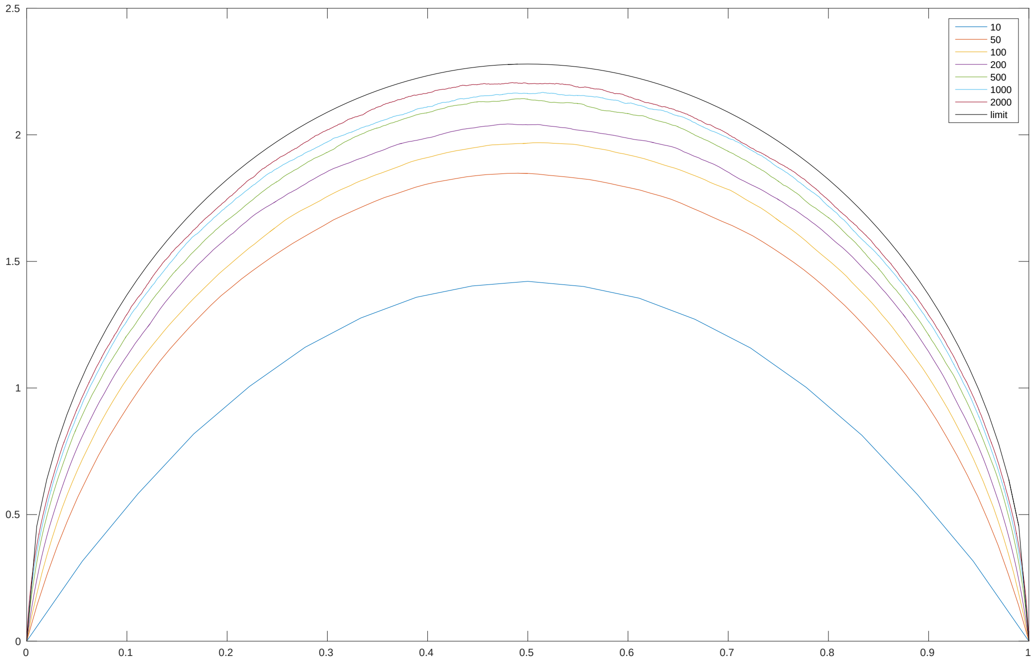

FIGURE 7.5 – Mean contours of binary conditioned Galton-Watson trees with size n and $\sigma = 0.7$ calculated from 2000 trees for each values of n.

is not as accurate. This was expected because the bias corrector does not fit as well to the bias curve for small small values of sigma as its does for greater values of σ.

Since we have an estimation procedure which seems to work, the natural further study is to see how the quality of our estimators vary as the characteristics of the forest change. We begin by looking at the variations when the size of the trees increase. A priori, the sizes of the trees in the considered forest should not have influence on the dispersion of the estimators. Indeed, our estimation strategy is based on the approximation of the Harris path of a finite tree by its limit. As a consequence, the size parameter only governs the quality of this approximation. Whatever the sizes of the trees, the dispersion will be given by the variance of the limit distribution Λ_∞. As expected Figure 7.8 shows that the dispersion of the estimators does not change as the sizes of the trees change when σ takes great values. Similarly, as shown in Figure 7.9, for small values of σ, the sizes of the trees do not influence the dispersion of the estimator. However, Figure 7.9 also shows that the sizes of the trees have a positive influence of the bias of the estimators.
Finally, Figure 7.10 shows the variation of the quality of the Least-square estimator as the size of the forest changes. It appears to be consistent with the theoretical fluctuation intervals given by the central limit theorem.

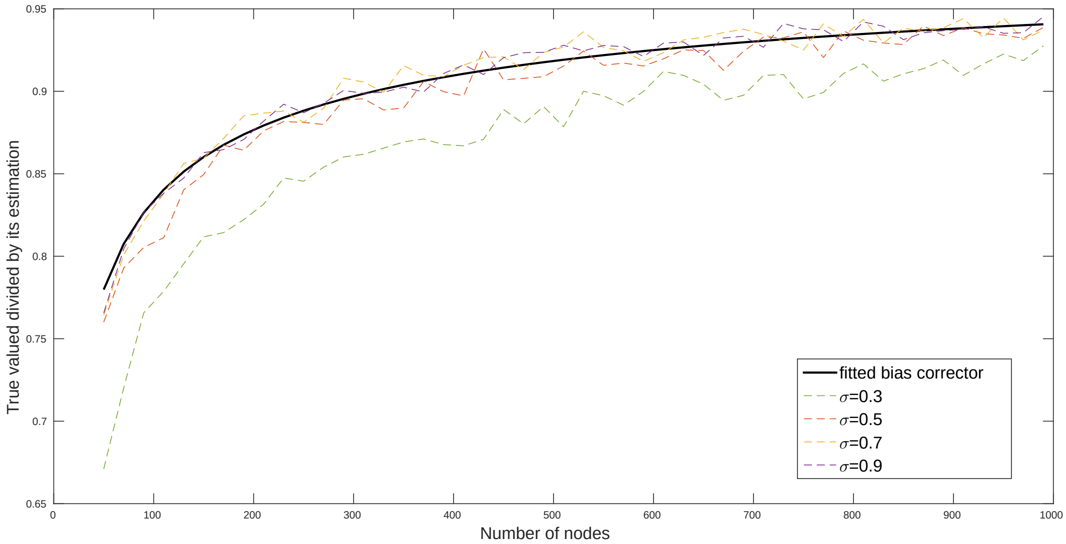

FIGURE 7.6 – Bias of the least-square estimator for different values of σ with a fitted bias corrector function.

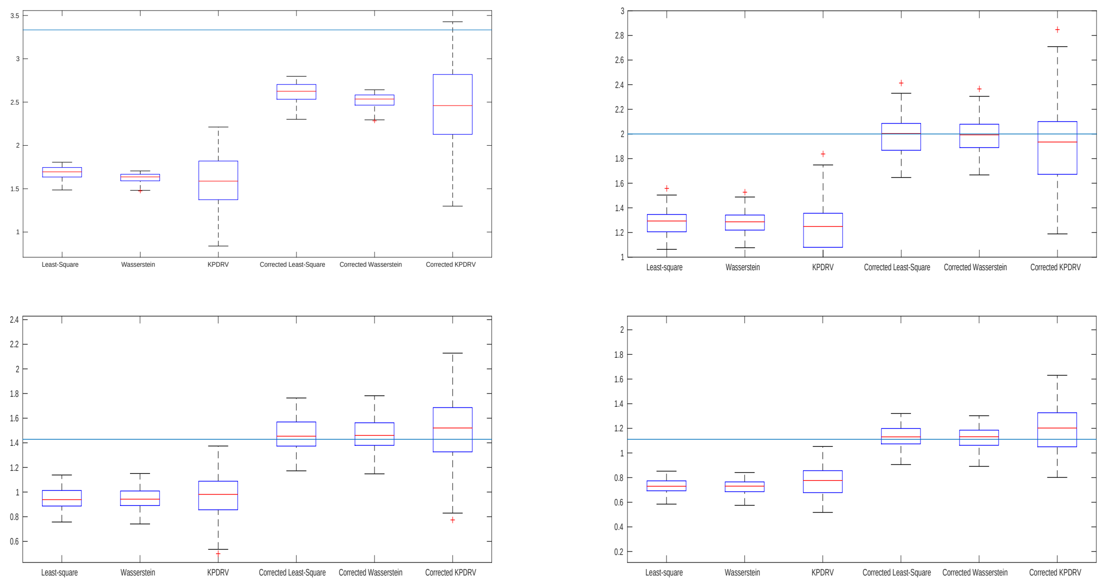

FIGURE 7.7 – Estimation and bias correction for forests of 10 trees with 20 nodes for σ equals to 0.3 (top, left) 0.5 (top, right), 0.7 (bottom, left) and 0.9 (bottom right).

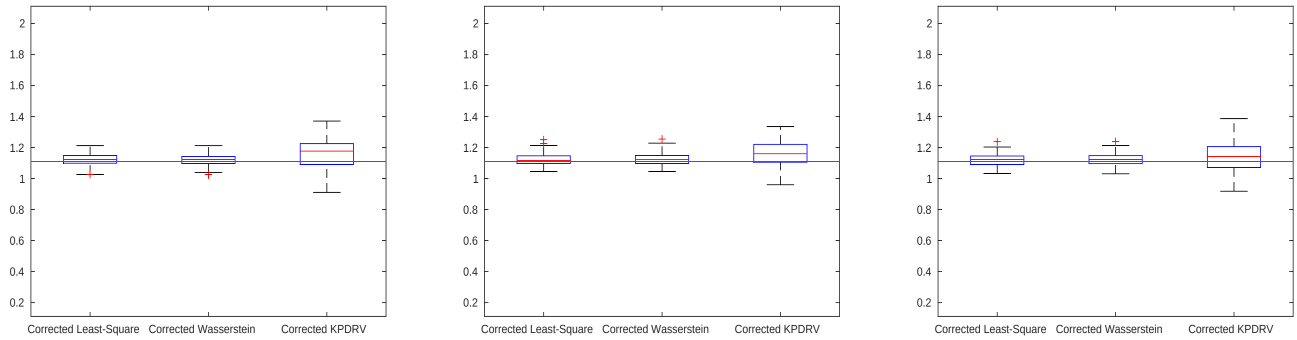

FIGURE 7.8 – Variation of the size of the tree for small σ (equal to 0.9) : tree sizes varying from 20 nodes (left), 50 nodes (center), to 100 nodes (right). Forests of 50 trees.

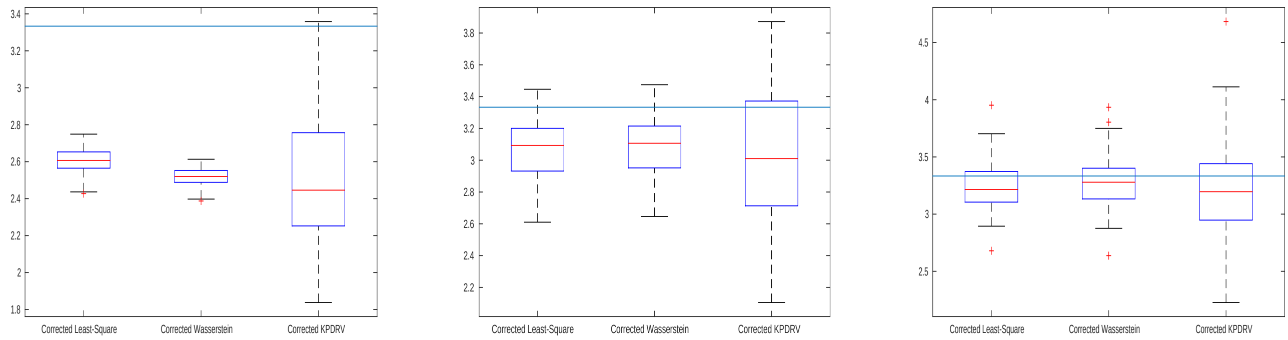

FIGURE 7.9 – Variation of the size of the tree for large σ (equal to 0.3) : tree sizes varying from 20 nodes (left), 50 nodes (center), to 100 nodes (right). Forests of 50 trees.

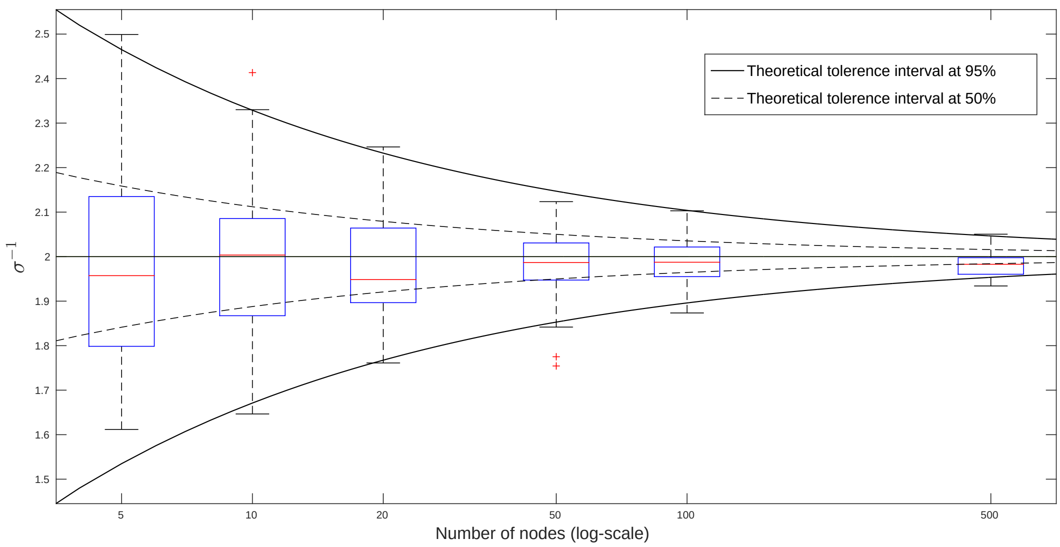

FIGURE 7.10 – Least-square estimation of σ^{-1} for different sizes of forests.

Bibliographie

[1] David Aldous. The Continuum Random Tree III. *Ann. Probab.*, 21(1) :248–289, 01 1993.

[2] David Aldous and Lea Popovic. A critical branching process model for biodiversity. *Adv. in Appl. Probab.*, 37(4) :1094–1115, 2005.

[3] Søren Asmussen. Convergence rates for branching processes. *Ann. Probability*, 4(1) :139–146, 1976.

[4] K. B. Athreya and P. E. Ney. *Branching processes.* Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York ; MR0373040].

[5] Krishna Balasundaram Athreya. Limit theorems for multitype continuous time Markov branching processes. II. The case of an arbitrary linear functional. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 13 :204–214, 1969.

[6] Frank Ball, Miguel González, Rodrigo Martínez, and Maroussia Slavtchova-Bojkova. Stochastic monotonicity and continuity properties of functions defined on Crump-Mode-Jagers branching processes, with application to vaccination in epidemic modelling. *Bernoulli*, 20(4) :2076–2101, 2014.

[7] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1996.

[8] Jean Bertoin. The structure of the allelic partition of the total population for Galton-Watson processes with neutral mutations. *Ann. Probab.*, 37(4) :1502–1523, 2009.

[9] Karthik Bharath, Prabhanjan Kambadur, Dipak Dey, Rao Arvin, and Veerabhadran Baladandayuthapani. Inference for large tree-structured data. *Preprint arXiv :1404.2910*, 2014.

[10] Robert M. Blumenthal. *Excursions of Markov processes.* Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.

[11] Sergey Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics and kantorovich transport distances. *preprint*, 2014.

[12] Nicolas Champagnat and Benoît Henry. Moments of a splitting tree with neutral poissonian mutations. 2016. To appear in EJP.

[13] Nicolas Champagnat and Amaury Lambert. Splitting trees with neutral Poissonian mutations I : Small families. *Stochastic Process. Appl.*, 122(3) :1003–1033, 2012.

[14] Nicolas Champagnat and Amaury Lambert. Splitting trees with neutral Poissonian mutations II : Largest and oldest families. *Stochastic Process. Appl.*, 123(4) :1368–1414, 2013.

[15] Nicolas Champagnat, Amaury Lambert, and Mathieu Richard. Birth and death processes with neutral mutations. *Int. J. Stoch. Anal.*, pages Art. ID 569081, 20, 2012.

[16] C. Czado and A. Munk. Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society : Series B*, 60(1) :223–241, 1998.

[17] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.

[18] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

[19] H. A. David and H. N. Nagaraja. *Order statistics.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

[20] Miraine Dávila Felipe and Amaury Lambert. Time reversal dualities for some random forests. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(1) :399–426, 2015.

[21] Eustasio del Barrio, Evarist Giné, and Carlos Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.*, 27(2) :1009–1071, 1999.

[22] C. Delaporte. Lévy processes with marked jumps I : Limit theorems. *ArXiv e-prints*, May 2013.

[23] C. Delaporte. Lévy processes with marked jumps II : Application to a population model with mutations at birth. *ArXiv e-prints*, May 2013.

[24] Luc Devroye. Simulating size-constrained galton-watson trees. *SIAM Journal on Computing*, 41(1) :1–11, 2012.

[25] Michael Drmota and Jean-François Marckert. Reinforced weak convergence of stochastic processes. *Statist. Probab. Lett.*, 71(3) :283–294, 2005.

[26] Thomas Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2) :996–1027, 04 2003.

[27] Alison Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009.

[28] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3 :87–112 ; erratum, ibid. 3 (1972), 240 ; erratum, ibid. 3 (1972), 376, 1972.

[29] Warren J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. Theoretical introduction.

[30] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[31] Ronald Aylmer Fisher. *The genetical theory of natural selection : a complete variorum edition.* Oxford University Press, 1930.

[32] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[33] Santiago Gallón, Jean-Michel Loubes, and Elie Maza. Statistical properties of the quantile normalization method for density curve alignment. *Mathematical Biosciences*, vol. 242(2) :pp. 129–142, April 2013.

[34] J. Geiger. Contour processes of random trees. In *Stochastic partial differential equations (Edinburgh, 1994)*, volume 216 of *London Math. Soc. Lecture Note Ser.*, pages 72–96. Cambridge Univ. Press, Cambridge, 1995.

[35] J. Geiger and G. Kersting. Depth-first search of random trees, and Poisson point processes. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 111–126. Springer, New York, 1997.

[36] Priscilla Greenwood and Jim Pitman. Fluctuation identities for Lévy processes and splitting at the maximum. *Adv. in Appl. Probab.*, 12(4) :893–902, 1980.

[37] R. C. Griffiths and Anthony G. Pakes. An infinite-alleles version of the simple branching process. *Adv. in Appl. Probab.*, 20(3) :489–524, 1988.

[38] Sergei Grishechkin. On a relationship between processor-sharing queues and Crump-Mode-Jagers branching processes. *Adv. in Appl. Probab.*, 24(3) :653–698, 1992.

[39] Theodore E. Harris. *The theory of branching processes*. Dover Phoenix Editions. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin ; MR0163361 (29 #664)].

[40] Benoît Henry. Clts for general branching processes related to splitting trees. Preprint available at http ://arxiv.org/abs/1509.06583.

[41] C. C. Heyde. A rate of convergence result for the super-critical Galton-Watson process. *J. Appl. Probability*, 7 :451–454, 1970.

[42] C. C. Heyde. Some central limit analogues for supercritical Galton-Watson processes. *J. Appl. Probability*, 8 :52–59, 1971.

[43] Alexander Iksanov and Matthias Meiners. Rate of convergence in the law of large numbers for supercritical general multi-type branching processes. *Stochastic Process. Appl.*, 125(2) :708–738, 2015.

[44] Kiyosi Itô. Poisson point processes attached to Markov processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III : Probability theory*, pages 225–239. Univ. California Press, Berkeley, Calif., 1972.

[45] Kiyosi Itô. *Poisson point processes and their application to Markov processes*. SpringerBriefs in Probability and Mathematical Statistics. Springer, Singapore, 2015. With a foreword by Shinzo Watanabe and Ichiro Shigekawa.

[46] Peter Jagers. Convergence of general branching processes and functionals thereof. *J. Appl. Probability*, 11 :471–478, 1974.

[47] Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London-New York-Sydney, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.

[48] Peter Jagers and Olle Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, 16(2) :221–259, 1984.

[49] Peter Jagers and Olle Nerman. Limit theorems for sums determined by branching and other exponentially growing processes. *Stochastic Process. Appl.*, 17(1) :47–71, 1984.

[50] Svante Janson. Conditioned Galton-Watson trees do not grow. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Discrete Math. Theor. Comput. Sci. Proc., AG, pages 331–334. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006.

[51] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surveys*, 9 :103–252, 2012.

[52] Svante Janson et al. Brownian excursion area, wright's constants in graph enumeration, and other brownian areas. *Probability Surveys*, 4 :80–145, 2007.

[53] Olav Kallenberg. *Random measures*. Akademie-Verlag, Berlin ; Academic Press, Inc., London, fourth edition, 1986.

[54] L. V. Kantorovitch. Sur certains développements suivant les polynômes de la forme de S. Bernstein.

[55] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3) :235–248, 1982.

[56] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, (Special Vol. 19A) :27–43, 1982. Essays in statistical science.

[57] Christian Kleiber and Jordan Stoyanov. Multivariate distributions and the moment problem. *J. Multivariate Anal.*, 113 :7–18, 2013.

[58] D Knuth. Seminumerical algorithms, 2nd edn, vol. 2 of the art of computer programming, 1981.

[59] Andreas E. Kyprianou. *Fluctuations of Lévy processes with applications*. Universitext. Springer, Heidelberg, second edition, 2014. Introductory lectures.

[60] Amaury Lambert. The contour of splitting trees is a Lévy process. *Ann. Probab.*, 38(1) :348–395, 2010.

[61] Amaury Lambert, Helen K. Alexander, and Tanja Stadler. Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *J. Theoret. Biol.*, 352 :60–70, 2014.

[62] Amaury Lambert and Chunhua Ma. The coalescent in peripatric metapopulations. *J. Appl. Probab.*, 52(2) :538–557, 2015.

[63] Amaury Lambert, Hélène Morlon, and Rampal S. Etienne. The reconstructed tree in the lineage-based model of protracted speciation. *J. Math. Biol.*, 70(1-2) :367–397, 2015.

[64] Amaury Lambert and Lea Popovic. The coalescent point process of branching trees. *Ann. Appl. Probab.*, 23(1) :99–144, 2013.

[65] Amaury Lambert, Florian Simatos, and Bert Zwart. Scaling limits via excursion theory : interplay between Crump-Mode-Jagers branching processes and processor-sharing queues. *Ann. Appl. Probab.*, 23(6) :2357–2381, 2013.

[66] Amaury Lambert and Mike Steel. Predicting the loss of phylogenetic diversity under non-stationary diversification models. *J. Theoret. Biol.*, 337 :111–124, 2013.

[67] Amaury Lambert and Pieter Trapman. Splitting trees stopped when the first clock rings and Vervaat's transformation. *J. Appl. Probab.*, 50(1) :208–227, 2013.

[68] Jean-François Le Gall. Itô's excursion theory and random trees. *Stochastic Process. Appl.*, 120(5) :721–749, 2010.

[69] G. G. Lorentz. *Bernstein polynomials.* Mathematical Expositions, no. 8. University of Toronto Press, Toronto, 1953.

[70] G. Louchard. Kac's formula, Levy's local time and Brownian excursion. *J. Appl. Probab.*, 21(3) :479–499, 1984.

[71] Guy Louchard and Svante Janson. Tail estimates for the Brownian excursion area and other Brownian areas. *Electron. J. Probab.*, 12 :no. 58, 1600–1632, 2007.

[72] Thomas Robert Malthus. An essay on the principle of population. 1798.

[73] Paul-A. Meyer. *Probability and potentials.* Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1966.

[74] Olle Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete*, 57(3) :365–395, 1981.

[75] David Nualart. *The Malliavin calculus and related topics.* Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[76] Peter Olofsson and Suzanne S. Sindi. A Crump-Mode-Jagers branching process model of prion loss in yeast. *J. Appl. Probab.*, 51(2) :453–465, 2014.

[77] J. Pitman and M. Yor. Itô's excursion theory and its applications. *Jpn. J. Math.*, 2(1) :83–96, 2007.

[78] Jim Pitman. *Combinatorial stochastic processes*, volume 32. Springer Science & Business Media, 2006.

[79] Jim Pitman and Matthias Winkel. Growth of the Brownian forest. *Ann. Probab.*, 33(6) :2188–2211, 2005.

[80] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion.* Grundlehren der mathematischen Wissenchaften A series of comprehensive studies in mathematics. Springer, 1999.

[81] M. Richard. Limit theorems for supercritical age-dependent branching processes with neutral immigration. *Adv. in Appl. Probab.*, 43(1) :276–300, 2011.

[82] Mathieu Richard. *Arbres, Processus de branchement non Markoviens et processus de Lévy.* Thèse de doctorat, Université Pierre et Marie Curie, Paris 6.

[83] Mathieu Richard. Splitting trees with neutral mutations at birth. *Stochastic Process. Appl.*, 124(10) :3206–3230, 2014.

[84] Pardis C Sabeti, David E Reich, John M Higgins, Haninah ZP Levine, Daniel J Richter, Stephen F Schaffner, Stacey B Gabriel, Jill V Platko, Nick J Patterson, Gavin J McDonald, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909) :832–837, 2002.

[85] Pardis C Sabeti, Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, Elizabeth H Byrne, Steven A McCarroll, Rachelle Gaudet, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164) :913–918, 2007.

[86] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

[87] Z. Taïb. Branching processes and neutral mutations. In *Stochastic modelling in biology (Heidelberg, 1988)*, pages 293–306. World Sci. Publ., Teaneck, NJ, 1990.

[88] Pierre François Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18 :14–54, 1845.

[89] Shinzo Watanabe. Itô's theory of excursion point processes and its developments. *Stochastic Process. Appl.*, 120(5) :653–677, 2010.

[90] Stephen Willard. *General topology*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1970 original [Addison-Wesley, Reading, MA ; MR0264581].

[91] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2) :97–159, 1931.

[92] Rui Yang, Panos Kalnis, and Anthony K. H. Tung. Similarity evaluation on tree-structured data. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, SIGMOD '05, pages 754–765, New York, NY, USA, 2005. ACM.