Population Dynamics and Random Genealogies[1]

AMAURY LAMBERT[2]

November 2006

[1]This document is intended primarily to supporting attendants to the 'Simposio de Probabilidad y Procesos Estocásticos' held on November 20–24, in Guanajuato, Mexico (CIMAT). Later on, it should serve as the basis for a graduate textbook to be written in collaboration with Sylvie Méléard (Université Paris 10, Nanterre and École Polytechnique, Palaiseau). Please do not hesitate to write and point out errors/ ask questions/ send (positive ;-) comments (email `amaury.lambert -at- ens.fr`).

[2]Université Paris 6 Pierre et Marie Curie and École Normale Supérieure, Paris.

Foreword

A short story of great men

The first quantitative study of populations came up with the book of Reverend Thomas Robert Malthus in 1798 [83], in which the growth of human populations was computed to be geometric, while that of resources only arithmetic. This idea of *limited growth* was one of the first notches against the prevailing dogma of pre-established harmony from divine origin. Unfortunately, his computation made Malthus predict the collapse of humanity and by the way, urge for birth control policies and... suppression of welfare !

Shortly after, this novel idea that resources are in limited amount inspired the young Charles Darwin. More wisely than Malthus, Darwin was not willing to apply this idea to human populations, but to all other living species, and these first thoughts are at the origin of the theory of natural selection published in 1859 [28].

A few years later, Lord Francis Galton, who was proudly half-cousin of Darwin, was wondering whether the observed demographic decay of British noble families was due to their diminished fertility or to the law of chances. Shamefully, Galton was unaware of the earlier writings of Irénée–Jules Bienaymé on that subject [54], and, in that matter, Malthus and Darwin's works on populations were useless to him. It was his friend Reverend Henry William Watson who reinvented what is now known as the Bienaymé–Galton–Watson (BGW) process [100], and solved (erroneously) his question.

Surprisingly, it was not until the 1920's that the BGW process made its third appearance, in the early works of Ronald Aylmer Fisher [41, 42] and John Burdon Sanderson Haldane [49]. At that point, it is important to recall that Fisher was not only the famous statistician, but also one of the three great pioneers of population genetics, along with Haldane and Wright. For further details on the historical aspects of branching processes and their applications to biology, see [50, 63].

More recent progress

If any, there are two things we can learn from this naïve story. First, although one should admit that later applications of branching processes will also include physical models, it is not anecdotic that one of the most popular objects in probability theory –for the phrase 'branching processes', a multidisciplinary academic research engine provides about 187 000 grouped citations– has its roots in population biology. Indeed, the *complexity* of biological interactions is responsible for numerous phenomena that have *no obvious visible causes*, like fluctuations of population sizes and allele frequencies, mutant fixations, species invasions, epidemics, extinctions, to name but a few, and, as Henri Poincaré would say, 'we say they are due to chance'. This makes me believe that population biology is a natural great source of inspiration for probabilists.

Second, it is known (and unfortunate) that human intuition poorly handles stochasticity [43, 59], so it is no wonder that our little story has begun with a deterministic object –a geometric sequence. And indeed, this scenario has repeated itself various times in the last century, with the mathematical theories of epidemics (Kermack–McKendrick), ecological interactions (Lotka–Volterra), spatial diffusion (Kolmogorov–Petrovsky–Piscounov), morphogenesis (Turing),...all making extensive use of *deterministic tools* like dynamical systems and partial differential equations.

Therefore, an interesting feature of recent mathematical research in biology is its propensity to request help from probability theory more and more spontaneously. Without even speaking of tools from the 'omics' era, there are at least two examples of recent, active fields of probability theory related to population biology: *genealogy*, or *phylogeny*, *modelling and reconstruction* [12, 33], whose toy model is the *Kingman coalescent* [68], and *social networks* [34], whose toy model is the *Erdös–Rényi random graph* [35]. In other words, the interplay between probability theory and population biology is not only still in motion, it is *accelerating*.

Outline

This document comprises four chapters.

In the first chapter, we define and study various neutral models of discrete populations. By 'neutral', we mean that all individuals in the population are *exchangeable*, that is, they all have the same chances of transmitting their genes to the next generation (no Darwinian selection). We distinguish genealogical models where the total population size is *fixed*, from those where the population size *flucutates randomly*. Former models are used in *mathematical population genetics* [27, 39], whereas latter models, usually called 'branching models', are more popular in *mathematical ecology* [70, 89] and *adaptive dynamics* [21, 22, 23].

Fixed-size models include those of *Cannings*, *Wright–Fisher* (discrete time) and *Moran* (continuous time); branching models include *Bienaymé–Galton–Watson processes* (discrete time) and *birth–death processes* (continuous time).

In the second chapter, we define scaling limits of these models, to be seen as genealogies of continuous populations. In discrete time, these continuous genealogies can be modelled by successive *compositions of bridges* (fixed population size) or by successive *compositions of subordinators* (branching population size).

In continuous time, the continuous analogue of models with fixed size is the *stochastic flow of bridges* of Bertoin and Le Gall [14], also called *generalized Fleming–Viot process*, or GFV-process. That of branching models is the *continuous-state branching process*, or *CB-process*. Both processes have diffusion limits, called respectively the *Fisher–Wright diffusion* and the *Feller diffusion*. Connections between the two kinds of models are also studied, and special attention is given to *extinction* (probability, expected time, conditioning).

When (sub)populations are bound to extinction, there may exist distributions that are invariant *conditional* on extinction not to have occurred. These distributions are called *quasi-stationary distributions* (QSD) and provide a rigorous notion for the dynamics of populations that seem stable, at least at the human timescale. The third chapter is devoted to the study of quasi-stationarity.

Special examples are as follows. The limiting distribution, as $t \to \infty$, of Z_t conditional on $\{Z_t \neq 0\}$, is (erroneously) called the *Yaglom distribution*, whereas the limit, as $s \to \infty$, of $(Z_u; 0 \leq u \leq t)$ conditional on $\{Z_{t+s} \neq 0\}$, is called the *Q-process*. For all models introduced in the first two chapters, we display QSD's (including Yaglom distributions) and Q-processes. We compare QSD's with the invariant distributions of the Q-process, and briefly tackle the question of uniqueness of the QSD.

In the fourth chapter, we consider *splitting trees*, which are those real trees (trees with edge lengths) where all individuals have i.i.d. *lifespans* (the edge lengths) during which they give birth at constant rate to copies of themselves. The number of individuals alive at time t is a (generally not Markovian) branching process called the homogeneous binary *Crump–Mode–Jagers process*.

We prove that the *contour process* of splitting trees with general lifetime is a killed Lévy process, that we call the *jumping chronological contour process* (JCCP). We use this observation to derive a certain number of properties of these trees, including *spine decompositions*, a generalization of the *coalescent point process* of Aldous and Popovic [2, 88], and connections between Lévy processes and branching processes previously discovered by Le Gall and Le Jan [80] and by Bertoin and Le Gall [13]. These connections can be used to provide various generalizations of the well-known results, called *Ray–Knight theorems*, that give the law of the local time process of killed Brownian motions.

A fifth chapter was originally planned, to deal with *coalescent processes* and *allele partitions*. It will hopefully see the light at some point. It should include key objects such as the Kingman coalescent, Λ-coalescents, Fisher's log-series of species abundance, the Chinese restaurant process, Ewens' sampling formula, and the GEM distribution. References for this chapter include [1, 12, 18, 33].

Notation. In models where the total population size is *fixed*, the letter Y will denote either the absolute or the relative *random* size of a *subpopulation*. In models where the *total* population size is *random*, we will denote this size by Z.

Secondary references. Books on fixed size models and mathematical population genetics include [27, 33, 39, 45]. On branching models and random trees, secondary references include [6, 7, 31, 38, 48, 50, 57].

Contents

Chapter 1

Neutral models of discrete populations

1.1 Discrete populations in discrete time

1.1.1 Fixed size models

To stick to standard notation, we set the constant population size to $2N$, where N is a positive integer. This is due to the fact that in population genetics, it is handy to think of a population as a basket of gametes, or 'gametic urn' (one gamete per homologous gene per individual, so that each diploid individual theoretically yields *two* –possibly different– gametes).

Cannings model

At each time step, the $2N$ individuals are randomly labelled $i = 1, \dots, 2N$. The dynamics of the *Cannings model* are given by the following rules.

- generation $n + 1$ is made up of the *offspring* of individuals from generation n
- for any i, individual i from generation n begets a number η_i of offspring, so that $\sum_i \eta_i = 2N$
- the law of the $2N$-tuple $(\eta_1, \dots, \eta_{2N})$ is *exchangeable*, that is, invariant under permutation of labels
$$(\eta_1, \dots, \eta_{2N}) \stackrel{\mathcal{L}}{=} (\eta_{\pi(1)}, \dots, \eta_{\pi(2N)}),$$
for any permutation π of $\{1, \dots, 2N\}$.

Start with a subpopulation of size $Y_0 = y$, and let Y_n denote the number of *descendants* of this subpopulation at time n. In the Cannings model, $(Y_n; n \geq 0)$ is a discrete-time Markov chain, with two absorbing states 0 and $2N$. For any integer $0 \leq y \leq 2N$, we write $\mathbb{P}_y$ for the conditional probability measure $\mathbb{P}(\cdot \mid Y_0 = y)$. Let τ denote the *absorption time*

$$\tau = \inf\{n : Y_n = 0 \text{ or } 2N\}.$$

If we exclude the trivial case where each individual begets exactly one child ($\eta_1 = 1$ a.s.), then it is easily seen that $\tau < \infty$ a.s.

Definition 1.1.1 *In total generality, the event* $\{Y_\tau = 0\}$ *is called* extinction *and denoted* $\{\text{Ext}\}$*, whereas* $\{Y_\tau = 2N\}$ *is called* fixation*, and denoted* $\{\text{Fix}\}$*.*

To understand what is meant by 'fixation', imagine that the subpopulation under focus is the set of individuals of a certain type which is transmitted faithfully from parent to offspring.

Proposition 1.1.2 *The Markov chain* $(Y_n; n \geq 0)$ *is a martingale, and the fixation probability* $\mathbb{P}_y(\text{Fix})$ *equals* $y/2N$*.*

Proof. By exchangeability, the η_i's are equally distributed and since $\sum_i \eta_i = 2N$, we must have $\mathbb{E}(\eta_1) = 1$. Then conditional on Y_n, and thanks to exchangeability, $Y_{n+1} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{Y_n} \eta_i$. Taking (conditional) expectations, we get $\mathbb{E}(Y_{n+1} \mid Y_n, Y_{n-1}, \dots, Y_0) = Y_n$. Now observe that τ is a stopping time and apply the stopping theorem to the bounded martingale Y

$$y = \mathbb{E}_y(Y_\tau) = \mathbb{E}_y(0 \cdot \mathbf{1}_{Y_\tau=0} + 2N \cdot \mathbf{1}_{Y_\tau=2N}) = 2N\mathbb{P}_y(\text{Fix}).$$

Another way of proving that $\mathbb{P}_y(\text{Fix}) = y/2N$ consists in giving a different type to each individual in the initial population. By exchangeability, each type has the same probability to become fixed, which thus has to be equal to $1/2N$. □

Now let P denote the transition matrix of the Markov chain Y. In other words, P is the square matrix of order $2N+1$ and generic element

$$P_{yz} = \mathbb{P}_y(Y_1 = z) \qquad 0 \le y, z \le 2N.$$

Since $\mathbb{P}_y(Y_n = z) = e_y' \, P^n \, e_z$ (where $'$ denotes transposition and e_x is the vector with zeros everywhere except a 1 at level x), it can be useful to get some information about the eigenvalues of P. Recall that the dominant eigenvalue (eigenvalue with maximum modulus) of a transition matrix is always 1, and notice that here e_0 and e_{2N} are left eigenvectors for P associated to 1, so that the eigenvalue 1 has at least multiplicity 2.

Theorem 1.1.3 (Cannings [19]) *The eigenvalues of P ranked in nonincreasing order are $\lambda_0 = \lambda_1 = 1$ and*

$$\lambda_j = \mathbb{E}(\eta_1 \eta_2 \cdots \eta_j) \qquad 1 \le j \le 2N.$$

Remark 1.1 *In particular, writing $\sigma^2 := \mathrm{Var}(\eta_1)$, it is elementary to compute*

$$\lambda_2 = \mathbb{E}(\eta_1 \eta_2) = 1 - \frac{\sigma^2}{2N-1}.$$

As a consequence, the multiplicity of the eigenvalue 1 is exactly 2 and the probabilities $\mathbb{P}(\tau > n) = \mathbb{P}(Y_n \notin \{0, 2N\})$ decrease with n like a geometric sequence with reason $1 - \sigma^2/(2N-1)$. For large populations, this can be quite slow.

Proof. Let Z be the following square matrix of order $2N+1$

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1^2 & 1^3 & \dots & 1^{2N} \\ 1 & 2 & 2^2 & 2^3 & \dots & 2^{2N} \\ 1 & 3 & 3^2 & 3^3 & \dots & 3^{2N} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 2N & (2N)^2 & (2N)^3 & \dots & (2N)^{2N} \end{pmatrix}$$

It is standard that Z is invertible. We are going to show that there is a triangular square matrix A such that $P = ZAZ^{-1}$, and whose diagonal elements are $a_{00} = 1$ and

$$a_{jj} = \mathbb{E}(\eta_1 \eta_2 \cdots \eta_j) \qquad 1 \le j \le 2N.$$

Since the diagonal elements of the (triangular) matrix A are the common eigenvalues of P and A, the proof will be over. Rewriting $P = ZAZ^{-1}$ as $PZ = ZA$, we are looking for coefficients $(a_{ij})_{i \le j}$ such that

$$\sum_{k=0}^{2N} p_{ik} k^j = \sum_{k=0}^{j} i^k a_{kj}.$$

Recalling that P is the matrix transition of Y, we can rephrase the question as finding polynomials $R_j(X) = \sum_{k=0}^{j} a_{kj} X^k$ of degree $j = 0, 1, \dots, 2N$, such that

$$\mathbb{E}(Y_1^j \mid Y_0 = i) = R_j(i).$$

Now since

$$\mathbb{E}(Y_1^j \mid Y_0 = i) = \mathbb{E}((\eta_1 + \cdots + \eta_i)^j),$$

we deduce from the exchangeability property that indeed $\mathbb{E}(Y_1^j \mid Y_0 = i) = R_j(i)$, where R_j is the polynomial of degree j given by

$$R_j(X) = \sum_{k=1}^{j} X(X-1)\cdots(X-k+1) \sum_{\alpha_\cdot \geq 1 \,:\, \alpha_1+\cdots+\alpha_k = j} \mathbb{E}(\eta_1^{\alpha_1} \cdots \eta_k^{\alpha_k}).$$

The result follows, since a_{jj} is the dominant coefficient of R_j, which is easily seen to equal $\mathbb{E}(\eta_1\eta_2\cdots\eta_j)$. □

Exercise 1.1 *For any $1 \leq z \leq 2N-1$, define N_z as the total number of visits of z*

$$N_z := \mathrm{Card}\{n : Y_n = z\}.$$

Establish a relation between $\mathbb{P}_y(Y_n = z \mid Fix)$ and $\mathbb{P}_y(Y_n = z)$, and deduce a relation between $\mathbb{E}_y(N_z \mid Fix)$ and $\mathbb{E}_y(N_z)$. Conclude that

$$\mathbb{E}_y(\tau \mid Fix) = y^{-1} \sum_{z=1}^{2N-1} z \mathbb{E}_y(N_z).$$

Wright–Fisher model

The *Wright–Fisher model* (WF) is a particular instance of the Cannings model, where $(\eta_1, \dots, \eta_{2N})$ follows the *multinomial* distribution with parameters $(2N; 1/2N, \dots, 1/2N)$. As for the associated Markov chain Y,

$$P_{yz} = \binom{2N}{z} \left(\frac{y}{2N}\right)^z \left(1 - \frac{y}{2N}\right)^{2N-z}.$$

In other words, conditional on $Y_n = y$,

$$Y_{n+1} \stackrel{\mathcal{L}}{=} \mathrm{Bin}(2N, y/2N),$$

that is, Y_{n+1} follows the binomial distribution with number of trials $2N$ and success probability $y/2N$. Yet another way of defining the model is to say that

- each individual from generation $n+1$ picks its one parent at random, *uniformly* among the individuals of generation n
- these $2N$ samplings are *independent*.

Results for the Cannings model apply, in particular $\lambda_0 = \lambda_1 = 1$, and [40]

$$\lambda_j = \prod_{i=1}^{j-1} \left(1 - \frac{i}{2N}\right) \qquad 2 \le j \le 2N.$$

In classical textbooks, treatment of the WF model usually goes through two further steps. The first one is computing equivalents, as N grows, of such quantities as the mean time to absorption $\mathbb{E}(\tau)$, thanks to *diffusion approximations*. This will be done in Chapter 2. The second step is to introduce the *Kingman coalescent*, which will be done in Chapter 5...

1.1.2 Branching models

Bienaymé–Galton–Watson model

First, assume we are given the law of a random integer ξ

$$p_k := \mathbb{P}(\xi = k) \qquad k \ge 0,$$

where p_0 and p_1 will always be assumed to be both different from 0 and 1. The population size at time n will be denoted by Z_n. Assume that at each time n, individuals in the population are randomly labelled $i = 1, \dots, Z_n$. The dynamics of the BGW model are given by the following rules

- generation $n + 1$ is made up of the *offspring* of individuals from generation n
- conditional on Z_n, for any $1 \le i \le Z_n$, individual i from generation n begets a number ξ_i of offspring
- the ξ_i's are *independent* and all distributed as ξ.

The Markov chain $(Z_n; n \ge 0)$ is called the *BGW process*. It contains less information than the whole *BGW model*, which provides a complete genealogical information. If $Z(x)$ denotes a BGW process started with $Z_0 = x$ individuals, then it is straightforward to check that the following *branching property* holds

$$Z(x + y) \stackrel{\mathcal{L}}{=} Z(x) + Z(y), \tag{1.1}$$

where the two terms of the r.h.s. are meant to be independent. In general, stochastic processes that satisfy (1.1) are called *branching processes*.

It is convenient to consider the generating function f of ξ

$$f(s) := \mathbb{E}(s^\xi) = \sum_{k\ge 0} p_k s^k \qquad s \in [0, 1],$$

as well as its expectation

$$m := \mathbb{E}(\xi) = f'(1-) \in (0, +\infty].$$

Similarly as in the previous subsection, we write $\mathbb{P}_z$ for the conditional probability measure $\mathbb{P}(\cdot \mid Z_0 = z)$. Unless otherwise specified, $\mathbb{P}$ will denote $\mathbb{P}_1$.

Proposition 1.1.4 *The generating function of Z_n is given by*

$$\mathbb{E}_z(s^{Z_n}) = f_n(s)^z \qquad s \in [0, 1], \tag{1.2}$$

where f_n is the n-th iterate of f with itself. In particular, $\mathbb{E}(Z_n \mid Z_0 = z) = m^n z$.

Proof. One can compute the generating function of Z_{n+1} conditional on $Z_n = z$ as

$$\mathbb{E}(s^{Z_{n+1}} \mid Z_n = z) = \mathbb{E}(s^{\sum_{i=1}^{z} \xi_i}) = \prod_{i=1}^{z} \mathbb{E}(s^{\xi_i}) = f(s)^z.$$

Iterating the last displayed equation then yields (1.2). Differentiating (1.2) w.r.t. s and letting s tend to 1, gives the formula for the expectation. □

As in the previous subsection, we say that *extinction* occurs if Z hits 0, and denote {Ext} this event. Before stating the next result, recall that f is an increasing, convex function such that $f(1) = 1$. As a consequence, f has at most 2 fixed points in $[0, 1]$. More specifically, 1 is the only fixed point of f in [0,1] if $m \leq 1$, and if $m > 1$, f has another distinct fixed point traditionally denoted by q.

Theorem 1.1.5 *If extinction does not occur, then* $\lim_{n\to\infty} Z_n = +\infty$ *a.s. In addition,*

$$\mathbb{P}_z(\text{Ext}) = q^z.$$

Proof. Notice that Z is an irreducible Markov chain with two classes. Since $\{0\}$ is an accessible, absorbing state, the class $\{1, 2, 3, \ldots\}$ is transient, and the first part of the theorem is proved.

To get the second part, observe that {Ext} is the increasing union, as $n \uparrow \infty$, of the events $\{Z_n = 0\}$, so that

$$\mathbb{P}(\text{Ext}) = \lim_{n\to\infty} \uparrow \mathbb{P}(Z_n = 0).$$

Thanks to (1.2), $\mathbb{P}_z(Z_n = 0) = f_n(0)^z$, so that $\mathbb{P}(\text{Ext})$ is the limit of the sequence $(q_n)_n$ defined recursively as $q_0 = 0$ and $q_{n+1} = f(q_n)$. By continuity of f, this limit is a fixed point of f, so it belongs to $\{q, 1\}$. But $0 = q_0 < q$ so taking images by the increasing function f and iterating, one gets the double inequality $q_n < q \leq 1$, which ends the proof. □

Definition 1.1.6 *A BGW model is said* subcritical *if* $m < 1$, critical *if* $m = 1$, *and* supercritical *if* $m > 1$.

Exercise 1.2 *Assuming that* $\sigma^2 := Var(\xi)$ *is finite, prove that*

$$\text{Var}(Z_n \mid Z_0 = 1) = \begin{cases} \sigma^2 m^{n-1} \frac{m^n - 1}{m-1} & \textit{if} \quad m \neq 1 \\ n\sigma^2 & \textit{if} \quad m = 1. \end{cases}$$

Exercise 1.3 *Assume* $m > 1$. *Show that conditional on* {Ext}, Z *has the same law as the subcritical branching process* $Z^\star$ *with offspring distribution* $p_k^\star = q^{k-1} p_k$, *whose generating function is*

$$f^\star(s) = q^{-1} f(qs) \qquad s \in [0, 1].$$

This subcritical branching process is called the *dual* process. Note that a similar result holds for the subtree of *infinite lines of descent* conditional on $\{\text{Ext}^c\}$.

Examples–exercises

Binary. Assume $p_k = 0$ for all $k \geq 3$, and call this model $Binary(p_0, p_1, p_2)$. The process is supercritical iff $p_2 > p_0$, and in that case, the extinction probability is $q = p_0/p_2$. The dual process is $Binary(p_2, p_1, p_0)$.

Geometric. Assume $p_k = (1-a)a^k$, and call this model $Geometric(a)$. We have $m = a/(1-a)$. The process is supercritical iff $a > 1/2$, and in that case, the extinction probability is $q = (1-a)/a$. The dual process is $Geometric(1-a)$.

Poisson. Assume $p_k = e^{-a}a^k/k!$, and call this model $Poisson(a)$. The process is supercritical iff $a > 1$, and in that case, we only get the inequality $q < 1/a$. The dual process is $Poisson(qa)$.

Linear fractional. Assume $p_0 = b$ and $p_k = (1-b)(1-a)a^{k-1}$ for $k \geq 1$. Call this model $LF(a, b)$. We have $m = (1-b)/(1-a)$. The process is supercritical iff $a > b$, and in that case, the extinction probability is $q = b/a$. The dual process is $LF(b, a)$. This example has very interesting features, see e.g. section I.4 of [6].

BGW process with immigration

Assume that in addition to the law of a random integer ξ (offspring distribution) with generating function f, we are also given the law of a random integer ζ (immigration law) with generating function g. The dynamics of the *BGW model with immigration* is given by the following rules

- generation $n+1$ is made up of the *offspring* of individuals from generation n *and* of a random number ζ_{n+1} of *immigrants*, where the ζ_i's are *independent* and all distributed as ζ
- conditional on Z_n, for any $1 \leq i \leq Z_n$, individual i from generation n begets a number ξ_i of offspring
- the ξ_i's are *independent* and all distributed as ξ.

It is important to understand that to each immigrant is given an independent BGW descendant tree with the same offspring distribution. The population size process $(Z_n; n \geq 0)$ of this model is a discrete-time Markov chain called *BGW process with immigration*. It is straightforward that

$$\mathbb{E}_z(s^{Z_1}) = g(s)f(s)^z.$$

Iterating this last equation yields

$$\mathbb{E}_z(s^{Z_n}) = f_n(s)^z \prod_{k=0}^{n-1} g \circ f_k(s) \qquad s \in [0,1]. \tag{1.3}$$

The following theorems concern the asymptotic growth of BGW processes with immigration.

Theorem 1.1.7 (Heathcote [51]) *Assume $m < 1$. Then the following dichotomy holds*

$$\begin{aligned} \mathbb{E}(\log^+ \zeta) < \infty &\Rightarrow (Z_n) \text{ converges in distribution} \\ \mathbb{E}(\log^+ \zeta) = \infty &\Rightarrow (Z_n) \text{ converges in probability to } +\infty. \end{aligned}$$

Theorem 1.1.8 (Seneta [93]) *Assume $m > 1$. Then the following dichotomy holds*

$$\begin{array}{lll} \mathbb{E}(\log^+\zeta) < \infty & \Rightarrow & \lim_{n\to\infty} m^{-n} Z_n \textit{ exists and is finite a.s.} \\ \mathbb{E}(\log^+\zeta) = \infty & \Rightarrow & \limsup_{n\to\infty} c^{-n} Z_n = \infty \textit{ for any positive } c \textit{ a.s.} \end{array}$$

To prove the last two theorems, we will need the following

Lemma 1.1.9 *Let $\zeta_1, \zeta_2, \dots$ be i.i.d. random variables distributed as ζ. Then for any $c > 1$,*

$$\begin{array}{lll} \mathbb{E}(\log^+\zeta) < \infty & \Rightarrow & \sum_{n\geq 1} c^{-n}\zeta_n < \infty \textit{ a.s.} \\ \mathbb{E}(\log^+\zeta) = \infty & \Rightarrow & \limsup_{n\to\infty} c^{-n}\zeta_n = \infty \textit{ a.s.} \end{array}$$

Proof. By the Borel–Cantelli lemma, one can prove that for any sequence of i.i.d. nonnegative r.v.'s $X, X_1, X_2, \dots$, $\limsup_{n\to\infty} X_n/n = 0$ or ∞ according whether $\mathbb{E}(X)$ is finite or not. Then the lemma follows from taking $X = \log^+(\zeta)$. Indeed, writing $\log c =: a > 0$, one then gets $c^{-n}\zeta_n = \exp -n(a - X_n/n)$ (as soon as $\zeta_n \neq 0$), which ends the proof. $\square$

Proof of Theorem 1.1.7. (inspired from [82], where it was inspired from [4]) Let Y_k have the law of the pure BGW process (without immigration) at generation k, when started at a r.v. distributed as a pack of immigrants, that we will denote by ζ_k. Then observe that the BGW process with immigration Z started at 0 and evaluated at generation n satisfies the following equality in distribution

$$Z_n \overset{\mathcal{L}}{=} \sum_{k=0}^{n} Y_k,$$

where the Y_k's are independent, and for each k, Y_k stands for the contribution to generation n from the immigrants of generation $n-k$. As a consequence, it suffices to determine whether

$$Z_\infty := \sum_{k=0}^{\infty} Y_k,$$

where the Y_k's are all independent, is finite or infinite. Notice that thanks to Kolmogorov's zero–one law, Z_∞ is finite a.s. or infinite a.s. Let $\mathcal{G}$ be the σ-field generated by *all* the r.v.'s $\zeta_0, \zeta_1, \zeta_2, \dots$

First assume that $\mathbb{E}(\log^+\zeta) < \infty$. Then

$$\mathbb{E}(Z_\infty \mid \mathcal{G}) = \sum_{k=0}^{\infty} \zeta_k m^k,$$

so by Lemma 1.1.9, this conditional expectation is finite a.s., which entails that Z_∞ itself is finite a.s.

Now, assume that Z_∞ is finite a.s. Since the Y_k's are integer r.v.'s, only finitely many of them can be different from 0. Thanks to the Borel–Cantelli lemma conditional on $\mathcal{G}$,

$$\sum_{k=0}^{\infty} \mathbb{P}(Y_k \neq 0 \mid \zeta_k) < \infty \qquad \text{a.s.}$$

But writing Z' for the pure BGW process (without immigration),

$$\mathbb{P}(Y_k \neq 0 \mid \zeta_k) \geq \zeta_k \mathbb{P}_1(Z'_k \neq 0) \geq \zeta_k \mathbb{P}_1(Z'_1 \neq 0)^k.$$

This proves that conditional on $\mathcal{G}$,

$$\sum_{k=0}^{\infty} \zeta_k \mathbb{P}(\xi \neq 0)^k < \infty \qquad \text{a.s.}$$

Again, conclude with Lemma 1.1.9. □

Proof of Theorem 1.1.8. (also inspired from [82], where it was also inspired from [4]) First, consider the case when $\mathbb{E}(\log^+ \zeta) = \infty$. Thanks to Lemma 1.1.9, $\limsup_{n\to\infty} c^{-n}\zeta_n = \infty$. Since $Z_n \geq \zeta_n$, the result follows.

Now consider the case when $\mathbb{E}(\log^+ \zeta) < \infty$. Let $\mathcal{F}_n$ stand for the σ-algebra generated by the r.v.'s $Z_0, Z_1, \ldots, Z_n$ as well as *all* the r.v.'s $\zeta_0, \zeta_1, \zeta_2, \ldots$ Then

$$\begin{aligned} \mathbb{E}(Z_{n+1}/m^{n+1} \mid \mathcal{F}_n) &= m^{-n-1}\mathbb{E}(\sum_{i=1}^{Z_n} \xi_i + \zeta_{n+1} \mid \mathcal{F}_n) \\ &= \frac{Z_n}{m^n} + \frac{\zeta_{n+1}}{m^{n+1}} . \end{aligned}$$

Two consequences stem from this last equation. The first one is that (Z_n/m^n) is a submartingale w.r.t. the filtration $(\mathcal{F}_n)$. The second one is that, by an immediate induction,

$$\mathbb{E}(Z_n/m^n \mid \mathcal{F}_0) = Z_0 + \sum_{k=0}^{n} \frac{\zeta_k}{m^k} \qquad n \geq 1.$$

Now thanks to Lemma 1.1.9, $(Z_n/m^n)_n$ is a submartingale with *bounded* expectations, so it converges a.s. to a finite r.v. □

Kesten–Stigum theorem

Assume that $1 < m < \infty$ and set $W_n := m^{-n} Z_n$. It is elementary to check that $(W_n; n \geq 0)$ is a nonnegative martingale (see e.g. proof of Theorem 1.1.8), so it converges a.s. and in L^1 to a nonnegative random variable W

$$W := \lim_{n\to\infty} \frac{Z_n}{m^n} .$$

To be sure that the geometric growth at rate m is the correct asymptotic growth for the BGW process, one has to make sure that $W = 0$ (if and) only if extinction occurs.

Theorem 1.1.10 (Kesten–Stigum [65]) *Either* $\mathbb{P}(W = 0) = q$ *or* $\mathbb{P}(W = 0) = 1$*. The following are equivalent*

(i) $\mathbb{P}(W = 0) = q$

(ii) $\mathbb{E}(W) = 1$

(iii) $\mathbb{E}(\xi \log^+ \xi) < \infty$.

For a recent generalization of this theorem to more dimensions, see [5].

Remark 1.2 *It is worth noting with C.C. Heyde [53], that even when* $\mathbb{E}(\xi \log^+ \xi) = \infty$*, there is a deterministic sequence* (C_n) *such that* $\lim_n C_{n+1}/C_n = m$, $\lim_n Z_n/C_n = W$ *exists and is finite a.s., and* $\mathbb{P}(W = 0) = q$.

On the other hand, if $\mathbb{E}(\xi^2) < \infty$*, the fact that* $\mathbb{E}(W) = 1$ *comes directly from the fact that* (W_n) *is bounded in* L^2*. This was proved by Kolmogorov in 1938 [69].*

Proof. (following [81]) Let us prove that $\mathbb{P}(W = 0) \in \{q, 1\}$. Conditional on Z_1, the descendant trees of all individuals from generation 1 are independent BGW trees, so that, with obvious notation,

$$W = \lim_{n\to\infty} Z_{n+1}/m^{n+1} = m^{-1} \lim_{n\to\infty} \sum_{i=1}^{Z_1} (Z_n^{(i)}/m^n) = m^{-1} \sum_{i=1}^{Z_1} W_i,$$

where the W_i's are independent r.v.'s, independent of Z_1, all distributed as W. As a consequence,

$$\mathbb{P}(W = 0) = \mathbb{E}(\mathbb{P}(\sum_{i=1}^{Z_1} W_i = 0) \mid Z_1) = \mathbb{E}(\mathbb{P}(W = 0)^{Z_1}) = f(\mathbb{P}(W = 0)).$$

In conclusion, $\mathbb{P}(W = 0)$ is a fixed point of the generating function f of ξ, so it is either q or 1.

As $(W_n; n \geq 0)$ is a martingale, we can define a new probability on chains, say $\mathbb{P}^\uparrow$, with the following martingale change of measure (Doob's harmonic transform)

$$\mathbb{P}_x^\uparrow(A) = x^{-1}\mathbb{E}_x(W_n, A) \qquad A \in \mathcal{F}_n,$$

where $\mathcal{F}_n$ is $\sigma(Z_0, Z_1, \dots, Z_n)$. Taking $x = 1$, this absolute continuity relationship becomes for a general event $A \in \mathcal{F}_\infty$

$$\mathbb{P}^\uparrow(A) = \mathbb{E}(W, A) + \mathbb{P}^\uparrow(A, \limsup_{n\to\infty} W_n = \infty). \tag{1.4}$$

Indeed, let P denote the probability measure defined by $P := (\mathbb{P} + \mathbb{P}^\uparrow)/2$. Since both $\mathbb{P}$ and $\mathbb{P}^\uparrow$ are absolutely continuous w.r.t. P, the associated Radon–Nikodym derivatives (U_n) and (V_n) are P-martingales satisfying

$$\mathbb{P}(A) = E(U_n, A) \quad \text{and} \quad \mathbb{P}^\uparrow(A) = E(V_n, A), \qquad A \in \mathcal{F}_n. \tag{1.5}$$

These two nonnegative P-martingales converge P-a.s. and in $L^1(P)$ to r.v.'s U and V, respectively. Then adding the two equalities in (1.5) yields $P(U_n + V_n = 2) = 1$ so that $P(U_n = V_n = 0) = 0$. In addition, for any $A \in \mathcal{F}_n$, get two different expressions for $E(U_n W_n, A)$, to find $P(V_n = U_n W_n) = 1$. Letting $n \to \infty$ in the last two results, we find that

$$P(U = 0, V \neq 0, \lim_{n\to\infty} W_n = \infty) + P(U \neq 0, \lim_{n\to\infty} W_n = V/U) = 1.$$

By absolute continuity, this equality still holds if replacing P with $\mathbb{P}$ or $\mathbb{P}^\uparrow$. Therefore, for any $A \in \mathcal{F}_\infty$,

$$\mathbb{P}^\uparrow(A, \limsup_{n\to\infty} W_n < \infty) = \mathbb{P}^\uparrow(A, U \neq 0) = E(V, A, U \neq 0) = E(UW, A, U \neq 0) = \mathbb{E}(W, A),$$

where the second equality follows from taking limits in (1.5). Thus, we have proved (1.4).

Next, since $\mathbb{E}_z^{\uparrow}(s^{Z_n-1}) = \mathbb{E}_z(Z_n s^{Z_n-1})/zm^n = f_n(s)^{z-1}f_n'(s)/m^n$, it is easy to check by induction that

$$\mathbb{E}_z^{\uparrow}(s^{Z_n-1}) = f_n(s)^{z-1}\prod_{k=0}^{n-1}(f'\circ f_k(s)/m).$$

Identifying this with (1.3) proves that under $\mathbb{P}^{\uparrow}$, $(Z_n - 1; n \geq 0)$ is a BGW process with *immigration*, where the immigration law is given by

$$\mathbb{P}(\zeta = k-1) = kp_k/m \qquad k \geq 1,$$

which has generating function $g(s) = f'(s)/m$. For deeper insight into this *size-biased* immigration, see [81], as well as forthcoming chapters on quasi-stationary distributions and branching processes conditioned to be never extinct (Chapter 3), as well as spine decompositions of splitting trees (Chapter 4).

Notice that

$$\mathbb{E}(\xi\log^+\xi) < \infty \Leftrightarrow \mathbb{E}(\log^+\zeta) < \infty.$$

Then we know, thanks to Theorem 1.1.8, that

$$\begin{array}{lcl}\mathbb{E}(\xi\log^+\xi) < \infty & \Rightarrow & \mathbb{P}^{\uparrow}(\lim_{n\to\infty} W_n \text{ exists and is finite}) = 1\\ \mathbb{E}(\xi\log^+\xi) = \infty & \Rightarrow & \mathbb{P}^{\uparrow}(\limsup_{n\to\infty} W_n = \infty) = 1.\end{array}$$

Conclude using (1.4). □

Results for (sub)critical processes on the asymptotic decay of $\mathbb{P}(Z_n \neq 0)$, as $n \to \infty$, will be displayed in Chapter 3. They can (and will) be proved thanks to a similar comparison with branching processes with immigration.

1.1.3 A relation between fixed size and branching

By the branching property, a BGW model which is *conditioned* to have constant size $2N$ is a particular instance of the Cannings model. A further result in that direction is the following.

Proposition 1.1.11 *A BGW model with Poisson offspring distribution conditioned to have constant size has the same law as the Wright–Fisher model.*

Proof. Let η_i be the offspring number of individual i in the BGW model conditioned to have constant size $2N$. Thus for any $2N$-tuple of integers $(k_1,\dots,k_{2N})$ *summing up to* $2N$,

$$\begin{aligned}\mathbb{P}(\eta_1 = k_1,\dots,\eta_{2N} = k_{2N}) &= \mathbb{P}(\xi_1 = k_1,\dots,\xi_{2N} = k_{2N} \mid \sum_{i=1}^{2N}\xi_i = 2N)\\ &= \frac{\prod_{i=1}^{2N}\mathbb{P}(\xi_i = k_i)}{\mathbb{P}_{2N}(Z_1 = 2N)}.\end{aligned}$$

Now let f be the generating function of the Poisson distribution with mean, say, m. Since the generating function of Z_1 under $\mathbb{P}_{2N}$ is $f(s)^{2N}$, one gets

$$\mathbb{E}_{2N}(s^{Z_1}) = (\exp(-m(1-s)))^{2N} = \exp(-2Nm(1-s)),$$

so that under $\mathbb{P}_{2N}$, Z_1 follows a Poisson distribution with mean $2Nm$. As a consequence,

$$\begin{aligned}\mathbb{P}(\eta_1 = k_1, \ldots, \eta_{2N} = k_{2N}) &= \frac{\prod_{i=1}^{2N} e^{-m} m^{k_i}/k_i!}{e^{-2Nm}(2Nm)^{2N}/(2N)!} \\ &= \frac{(2N)!}{k_1! \cdots k_{2N}!}\left(\frac{1}{2N}\right)^{2N},\end{aligned}$$

which is the multinomial distribution with parameters $(2N; 1/2N, \ldots, 1/2N)$. □

An alternative relationship is mentioned in [92], where an exchangeable genealogy is obtained by repeating the following scheme at each generation. Generate N i.i.d. positive integers (branching scheme) and sample exactly N individuals uniformly among these.

1.2 Discrete populations in continuous time

1.2.1 Basic reminders on generators

We will like to define continous-time Markov chains with values in $\{0, 1, 2, \ldots\}$ from their *infinitesimal generator*, that is, a *rate matrix* Q with nonnegative elements, except the diagonal ones, which satisfy

$$-q_{ii} =: q_i \geq \sum_{k\neq i} q_{ik} \qquad i \geq 0.$$

If the above inequality is strict, then the chain starting in state i can reach (and remain in) a cemetery state ∂ after an exponential time of parameter $q_i - \sum_{k\neq i} q_{ik}$. If instead of this inequality, equality holds for all i's, the rate matrix Q is said *conservative*.

It is always possible to build a (right-continuous) Markov chain $(F_t; t \geq 0)$ with a given rate matrix Q, iterating the following rule.

- starting from $F_0 = i$, wait an exponential time S_1 with parameter q_i
- given $S_1 = t$, make a jump to state j with probability q_{ij}/q_i ($j = \partial$ with probability $1 - \sum_{k\neq i} q_{ik}/q_i$)
- given $F_t = j$, start afresh from state j.

The chain thus obtained is called the *minimal* process. The exponential durations $S_1, S_2, \ldots$ are called *holding times*, or *sojourn times*.

The minimal process can have a finite lifetime in two cases. The first case occurs when the cemetery state is reached, which can only happen if the rate matrix is not conservative. Then we say that the chain is *killed*. The killing time will be denoted T_∂. The second case occurs when

$$T_\infty := \sum_{k\geq 1} S_k < \infty.$$

Since $T_\infty < \infty$ obviously implies that $\lim_{t\uparrow T_\infty} F_t = +\infty$, we then say that the chain *blows up*.

Transition functions $(P_{ij}(t); i, j \in \mathbb{N}, t \geq 0)$ are nonnegative functions satisfying $P_{ij}(0) = \delta_{ij}$, $\sum_j P_{ij}(t) \leq 1$, as well as the Chapman–Kolmogorov equations (semigroup property) $P_{ij}(t+$

$s) = \sum_k P_{ik}(t)P_{kj}(s)$. They are said *Q-functions* iff $P'(0+) = Q$. There is some ambiguity about the definition of the Markov chain from its rate matrix, since there may be other Markov chains than the minimal process whose transition functions are Q-functions. Actually, these other Markov chains can be obtained by resurrecting the minimal process after $T_\partial \wedge T_\infty$ (see forthcoming Theorem 1.2.1).

Recall the *backward Kolmogorov equations* (associated with Q)

$$P_{ij}(t) = \delta_{ij}e^{-q_it} + \int_0^t ds\ e^{-q_is} \sum_{k\neq j} q_{ik}P_{kj}(t-s) \qquad i,j \geq 0, t > 0,$$

as well as the *forward Kolmogorov equations*

$$P_{ij}(t) = \delta_{ij}e^{-q_jt} + \int_0^t ds\ e^{-q_js} \sum_{k\neq j} P_{ik}(t-s)q_{kj} \qquad i,j \geq 0, t > 0.$$

In terms of the paths of a Markov chain $(X_t, t \geq 0)$, the former equation can equivalently be read as

$$\mathbb{P}_i(X(t) = j) = \delta_{ij}\mathbb{P}(S > t) + \int_0^t \mathbb{P}(S \in ds) \sum_{k\neq j} \frac{q_{ik}}{q_i} \mathbb{P}_k(X_{t-s} = j),$$

where S is the *first* jump time of X. Considering the *last* jump before t provides a similar interpretation to the forward equation. In the case when the minimal process blows up, notice that there can fail to be a last jump, so the forward equation may not always hold, even in the conservative case.

The following theorem fixes the ideas as far as uniqueness is concerned. The statement is relatively clear without entering the details, and we refer the reader to standard textbooks such as [6] for more precise results and proofs.

Theorem 1.2.1 *If $T_\infty = \infty$ a.s., then the minimal process is the unique solution to the backward equations. If in addition $T_\partial < \infty$, the minimal process is the unique Markov chain with rate matrix Q.*

If $T_\infty < \infty$ with positive probability, then for any i.i.d. sequence of random integers $U_1, U_2, \ldots$, the minimal process resurrected in state U_i at each blow up time $T_\infty^{(i)}$ satisfies the backward equations.

If $T_\partial < \infty$ with positive probability and the minimal process has an entrance law at ∞, that is, $\lim_{i\to\infty} \mathbb{P}_i(F_t = j)$ exists for some j, then the minimal process resurrected at ∞ at each killing time satisfies the forward equations.

In what follows, it will be implicit that, whenever specifying a rate matrix, the Markov chain we consider is the minimal process associated with it. Anyway, we will only deal with conservative rate matrices, so that $T_\partial = \infty$, and most of the time, the minimal process will have $T_\infty = \infty$ a.s., so that there will remain no ambiguity about which process we consider.

1.2.2 Fixed size models

Start with a population of fixed size $2N$. Let η be a r.v. with values in $\{1, \ldots, 2N\}$. A general version of the Cannings model in continuous time can be defined as follows. Each individual gives birth at constant rate to a random number of children distributed as η, so that no two birth events can occur simultaneously. Then at each birth time S, conditional on $\eta = k$, draw randomly k individuals from the population at time $S-$, and replace them with the newborns. A continuous-state space version of this general model is studied in the next chapter. In what follows, we focus on the *Moran model*, which is the particular case of this model associated to $\eta = 1$ a.s.

Moran model

Label randomly the individuals of the population. To each ordered pair (i, j) of individuals, attach an exponential clock with parameter c. If the first clock that rings is that of, say, (i_0, j_0), then individual i_0 gives birth to a new individual who takes the place of individual j_0. Then relabel the individuals and start afresh new exponential clocks.

Since the mean number of birth–death events per unit time is $4N^2c$, and we want to measure time in *generations* as in the Cannings model, that is, we would like to see a mean number of $2N$ birth–death events per unit time, we fix $c = 1/2N$.

Start with a subpopulation of size $Y_0 = y$, and let Y_t denote the number of *descendants* of this subpopulation at time t. In the *Moran model*, $(Y_t; t \geq 0)$ is a continuous-time Markov chain, again with two absorbing states 0 and $2N$. The transition rates of this chain are given *after* the following lemma.

Lemma 1.2.2 *If* $X_1, X_2, \ldots,$ *are independent exponential variables with parameters* $a_1, a_2, \ldots,$ *then* $\min_i(X_i)$ *is an exponential variable with parameter* $\sum_i a_i$.

The first time at which a birth of an individual from a subpopulation of size y occurs is the minimum of all exponential variables attached to pairs (i, j), where $i \leq y$ and $j > y$. There are $y(2N - y)$ such pairs, so thanks to the previous lemma, this random time is an exponential variable with parameter $y(2N - y)/2N$. As a consequence, the transition rates for the Moran model are

$$\begin{cases} y \to y+1 & \text{at rate} \quad y(2N-y)/2N \\ y \to y-1 & \text{at rate} \quad y(2N-y)/2N. \end{cases}$$

In particular, notice that the embedded Markov chain associated with the Moran model is the *simple random walk stopped* upon hitting 0 or $2N$.

Recall that τ is the absorption time, that is, the first time Y hits $\{0, 2N\}$. Then let

$$T_x := \inf\{t : Y_t = x\} \qquad 0 \leq x \leq 2N.$$

As usual $\{\tau = T_0\}$ is the extinction event, and $\{\tau = T_{2N}\}$ the fixation event. As in the Cannings model, the exchangeability property assures that Y is a martingale so that $\mathbb{P}_y(\text{Fix}) = y/2N$.

Expectations of hitting times

Now let us compute the expectation of the absorption time. With

$$S_y := \int_0^\infty \mathbf{1}_{Y_t = y} \, dt,$$

we have $\tau = \sum_{y=1}^{2N-1} S_y$. In addition, since each sojourn time in state y is exponential with parameter $y(2N-y)/2N$,

$$\mathbb{E}(S_y) = 2N\mathbb{E}(N_y)/y(2N-y),$$

where N_y is the number of visits to y. By the Markov property, for any $k \geq 1$,

$$\mathbb{P}(N_y = k) = \mathbb{P}(T_y < \tau)(1-\rho(y))^{k-1}\rho(y),$$

where $\rho(y)$ denote the probability of no returning to y starting from y. In particular,

$$\mathbb{E}(N_y) = \frac{\mathbb{P}(T_y < \tau)}{\rho(y)} \ .$$

Applying the stopping theorem to the martingale Y at time $T_0 \wedge T_y$, we get

$$\mathbb{P}_x(T_y < T_0) = \frac{x}{y} \qquad x \leq y,$$

and by symmetry

$$\mathbb{P}_x(T_y < T_{2N}) = \frac{2N-x}{2N-y} \qquad x \geq y.$$

Then, since the embedded Markov chain is a simple random walk,

$$\rho(y) = \frac{1}{2}\mathbb{P}_{y+1}(T_{2N} < T_y) + \frac{1}{2}\mathbb{P}_{y-1}(T_0 < T_y),$$

and using the two preceding equations, we get

$$\rho(y) = \frac{2N}{y(2N-y)} \ .$$

Putting this together,

$$\mathbb{E}_x(N_y) = \begin{cases} x(2N-y)/2N & \text{if} \quad x \leq y \\ y(2N-x)/2N & \text{if} \quad x \geq y, \end{cases}$$

which finally yields

$$\mathbb{E}_x(S_y) = \begin{cases} x/y & \text{if} \quad x \leq y \\ (2N-x)/(2N-y) & \text{if} \quad x \geq y, \end{cases}$$

Adding up all these expressions gives the result for $\mathbb{E}_x(\tau)$.

Theorem 1.2.3 *The absorption time τ has expectation*

$$\mathbb{E}_x(\tau) = \sum_{y=1}^{x-1} \frac{2N-x}{2N-y} + \sum_{y=x}^{2N-1} \frac{x}{y},$$

and conditional on fixation

$$\mathbb{E}_x(\tau \mid \text{Fix}) = 2N - x + \sum_{y=1}^{x-1} \frac{(2N-x)y}{x(2N-y)} \ .$$

Proof. It only remains to prove the conditional expectation. The proof is the same as in the exercise ending the subsection on the Cannings model, and relies on the following application of Fubini's theorem and the Markov property

$$\mathbb{E}_x(S_y, \text{Fix}) = \mathbb{E}_x \int_0^\infty \mathbf{1}_{Y_t=y}\mathbf{1}_{\text{Fix}}\, dt = \int_0^\infty dt\ \mathbb{P}_x(Y_t = y)\mathbb{P}_y(\text{Fix}) = \mathbb{E}_x(S_y)\mathbb{P}_y(\text{Fix}).$$

Recalling that $\mathbb{P}_y(\text{Fix}) = y/2N$, one gets

$$\mathbb{E}_x(\tau \mid \text{Fix}) = \sum_{y=1}^{2N-1} \mathbb{E}_x(S_y)\frac{y}{x}\ ,$$

which finishes the proof. □

Elementary computations provide the following

Corollary 1.2.4 *For a subpopulation starting from one single individual, the absorption time has*

$$\mathbb{E}_1(\tau) = \sum_{y=1}^{2N-1} y^{-1} \sim \log(N) \qquad \textit{and} \qquad \mathbb{E}_1(\tau \mid \text{Fix}) = 2N.$$

For a subpopulation starting from $2N-1$ *individuals,*

$$\mathbb{E}_{2N-1}(\tau \mid \text{Fix}) = \frac{2N}{2N-1}\sum_{y=2}^{2N} y^{-1} \sim \log(N)$$

For an initial subpopulation representing a proportion p *of a large population (*$N \to \infty$ *and* $2Nx \to p$*),*

$$\mathbb{E}(\tau) \sim -2N(p\log(p) + (1-p)\log(1-p))$$
$$\mathbb{E}(\tau \mid \text{Fix}) \sim -2N(1-p)\log(1-p)/p.$$

A typical result to remember is that the time to fixation, when starting from *half* the total population, has mean $2\log(2)N \approx 1.4N$ (generations).

The asymptotic expressions given in the last corollary can be obtained directly by diffusion approximation methods, which will be done in detail in the next chapter.

1.2.3 Branching models

Markov branching process

Consider a discrete population where individuals give birth independently at constant rate b, to $k \geq 1$ new offspring individuals with probability p_k, and die independently at constant rate $d > 0$. Let $\pi_k := bp_k$ stand for the birth rate per individual of k-sized clutches, and $a := b + d$ the total birth–death rate per individual. Since each individual is replaced by 0 individual with probability d/a and by $k+1$ individuals with probability π_k/a, the genealogy associated with this birth–death scheme is that of a BGW process with offspring generating function

$$f(s) = \frac{1}{a}(d + \sum_{k\geq 1} \pi_k s^{k+1}) \qquad s \in [0,1].$$

Check that the process is critical if $b = d$ and supercritical (resp. subcritical) if $b > d$ (resp. $b < d$). Set $m := f'(1)$ as well as

$$u(s) := a(f(s) - s) \qquad s \in [0, 1].$$

Then recall that the extinction probability q is the smallest root of u. The Markov chain $(Z_t; t \geq 0)$ counting the number of individuals at time t trivially satisfies the branching property, and thus, is called a *Markov branching process*. Its extinction probability q is the smallest root of u and, thanks to Lemma 1.2.2, its transition rates are given by

$$\begin{cases} n & \to & n+k & \text{at rate} & n\pi_k \\ n & \to & n-1 & \text{at rate} & nd. \end{cases}$$

Remark 1.3 *When $\pi_1 = b$, Z is called* binary *branching process, or* linear birth–death process*. If in addition $d = 0$, it is called* binary fission *process, or* linear birth process*, or* Yule–Furry process*, or* Yule process*.*

Remark 1.4 *Another interpretation is the following. Each birth event of a clutch of size k can be considered alternatively as the birth of $k + 1$ individuals simultaneously with the death of the mother. In other words, each individual lives an exponential* lifespan *with parameter a at the end of which she gives birth to k individuals with probability π_k/a, with $\pi_0 := d$. In that case, nothing prevents us from giving the lifespans a more general distribution than the mere exponential. At this degree of generality, the process Z is called a* Bellman–Harris process*.*

Remark 1.5 *Processes that satisfy the branching property, but are not necessarily Markovian are called* general branching processes*, or* Crump–Mode–Jagers processes*.*

Since we assume conservative rate matrix, recall that either Z blows up in finite time T_∞ with positive probability, or it has infinite lifetime a.s.

Theorem 1.2.5 *The branching process has infinite lifetime a.s. iff $m < \infty$, or $m = \infty$ and*

$$\int^1 \frac{ds}{u(s)} = -\infty.$$

Proof. We exclude the trivial case when $q = 1$. Let $h_i(t) := \mathbb{P}_i(T_\infty > t) = \sum_{j\geq 0} \mathbb{P}_i(Z_t = j)$. Notice that $h(t) := h_1(t)$ is nonincreasing in t, that $h(0^+) = 1$ and that $h(t) \in (q, 1]$. Next, by the branching property, observe that $h_i(t) = h(t)^i$. Summing up the Kolmogorov backward equations yields

$$h(t) = e^{-at} + \int_0^t ds\ e^{-as} \sum_{k\neq 1} q_{1k} h(t-s)^k \qquad t > 0.$$

Differentiating this last equation and integrating by parts, this becomes

$$h'(t) = u(h(t)) \qquad t > 0.$$

Assume that the branching process has finite lifetime with positive probability, so that $h(t_0) < 1$ for some t_0. Fixing $\varepsilon \in (q, 1)$ and setting $F(x) := \int_\varepsilon^x ds/u(s)$, we get that $t - F(h(t)) =$

$t_0 - F(h(t_0))$, so letting $t \to 0^+$ entails that $F(1-)$ has a finite value. As a consequence, if h is not identically 1, then $\int^1 ds/u(s)$ is finite (which also forces $m = \infty$).

Conversely, assume that $m = \infty$ and that $\int^1 ds/u(s)$ converges, which allows to define $F(x) := \int_1^x ds/u(s)$ for $x \in (q, 1]$. Integrating $h' = u(h)$ with this new F, entails $t - F(h(t)) = 0$, which implies $h(t) < 1$ as soon as $t > 0$. □

Next, set

$$q(t) := \mathbb{P}_1(T < t) \qquad t \in (0, +\infty],$$

so that $q(0+) = 0$ and $q(\infty) = q$, the extinction probability.

Theorem 1.2.6 *The law of the extinction time is given implicitly by*

$$\int_0^{q(t)} \frac{ds}{u(s)} = t \qquad t \geq 0.$$

Proof. The idea is the same as previously. Since $\mathbb{P}_i(Z_t = 0) = q(t)^i$, the Kolmogorov backward equations yield

$$q(t) = \int_0^t ds\ e^{-as} (\sum_{k\geq 1} \pi_k q(t-s)^{k+1} + d) \qquad t > 0,$$

which becomes after differentiation and integration by parts

$$q'(t) = u(q(t)) \qquad t > 0.$$

Recalling that $q(t) \in [0, q)$, we set $F(x) := \int_0^x ds/u(s)$, and integrate the last displayed equation, to finally get $F(q(t)) = t$. □

Binary case. In the binary case ($\pi_1 = b$), $u(s) = d - (b+d)s + bs^2$, the extinction probability is $q = \min(d/b, 1)$, and with $r := b - d$,

$$q(t) = \begin{cases} d(e^{rt} - 1)/(be^{rt} - d) & \text{if} \quad b \neq d \\ bt/(1 + bt) & \text{if} \quad b = d. \end{cases}$$

In the subsection on birth–death processes, we will see in addition that the expected time to extinction equals

$$\mathbb{E}_1(T) = \frac{1}{b} \log \left(\frac{1}{1 - b/d} \right)$$

if $b < d$, and is infinite if $b = d$.

Time change

From the transition rates of the branching process, notice that the holding time in state n is exponential with parameter an, which means that birth–death events occur at a rate which is linear w.r.t. the population size. Then 'decelerating the clock' proportionally to the population size will give a Markov chain whose rates are independent from its current state, that is, a *random walk*. In other words, we define implicitly X as the solution to

$$Z_t = X\left(\int_0^t Z_s\,ds\right) \qquad t > 0.$$

More rigorously, recall that T is the (possibly infinite) extinction time of the branching process and set

$$\theta_t := \int_0^t Z_s\,ds \qquad t > 0.$$

Since θ is strictly increasing on $[0, T)$, we let κ be its inverse on $[0, \theta_T)$. Next define

$$X_t := Z \circ \kappa_t \qquad t < \theta_T.$$

The following theorem is a classical result in random time changes (see e.g [37, chapter 6], or [71, pp.25–26])

Theorem 1.2.7 *The process $(X_t; t \geq 0)$ is a random walk killed upon hitting 0. Its transition rates are for $n \geq 1$*

$$\begin{cases} n & \to & n+k & \text{at rate} & \pi_k \\ n & \to & n-1 & \text{at rate} & d. \end{cases}$$

Remark 1.6 *Because jumps of this random walk to the left are of absolute size at most 1, it is sometimes called a* left-continuous random walk.

Of course, the converse can easily be stated. In other words, one can recover the branching process Z from the random walk X killed at T_0, its (possibly infinite) first hitting time of 0. Indeed, for $t < T_0$,

$$\int_0^t \frac{ds}{X_s} = \int_0^{\kappa_t} \frac{d\theta_u}{X(\theta_u)} = \int_0^{\kappa_t} \frac{Z_u\,du}{Z_u} = \kappa_t$$

Check that if θ' is the inverse of κ on $[0, \kappa(T_0-))$, then

$$\theta_t = \begin{cases} \theta'_t & \text{if} & t < \kappa(T_0-) \\ T_0 & \text{if} & t \geq \kappa(T_0-). \end{cases}$$

As a conclusion, $Z = X \circ \theta$. Notice that this procedure can be achieved without reference to the initial process Z, by considering a left-continuous random walk X straight from the beginning, defining κ and θ as previously, and proving that $X \circ \theta$ has the same transition rates as Z.

Applications of this time change are numerous, and will be studied in more detail for branching processes with continuous-state space. In that setting, the time change is usually referred to as the *Lamperti transform*.

Extensions

Here, we want to give the definitions of two useful additional models: the branching process with *immigration*, and the branching process with *logistic growth*. Both models are simple extensions of the branching process, with the interesting feature that the former *never becomes extinct*, and the latter *never goes to* ∞.

Immigration. Let ν be a positive, finite measure on the nonnegative integers, and set $\rho := \sum_{k\geq 0} \nu_k$. In a branching model with immigration,

- at rate ρ, packs of immigrating individuals enter the population
- each pack comprises k individuals with probability ν_k/ρ
- all individuals present in the population die and reproduce independently according to the branching scheme.

Then the branching process with immigration has transition rates

$$\begin{cases} n & \to & n+k & \text{at rate} & n\pi_k + \nu_k \\ n & \to & n-1 & \text{at rate} & nd. \end{cases}$$

Exercise 1.4 *Prove that the branching process with immigration has infinite lifetime a.s. iff the pure branching process associated with it also has.*

Logistic growth. Let c be a positive real number, called the *competition intensity.* In a branching model with logistic growth,

- all individuals present in the population reproduce independently according to the branching scheme
- in addition to natural deaths at constant rate d, each individual kills and replaces every other individual at rate c.

Then the branching process with logistic growth, or *logistic branching process*, has transition rates

$$\begin{cases} n & \to & n+k & \text{at rate} & n\pi_k \\ n & \to & n-1 & \text{at rate} & nd + cn(n-1). \end{cases}$$

Recall that T stands for the extinction time.

Theorem 1.2.8 ([74]) *If $d = 0$, then the logistic branching process $(Z_t; t \geq 0)$ converges in distribution, and if $d \neq 0$, it becomes extinct with probability 1.*

Provided that $\sum \pi_k \log(k) < \infty$, the logistic branching process comes down from infinity, that is,

$$\lim_{i\uparrow\infty} \mathbb{P}(Z_t = j \mid Z_0 = i) \text{ exists for all } j \geq 0, t > 0.$$

In addition, $\mathbb{E}_\infty(T) < \infty$.

In the binary case, a proof can be found in the next subsection.

1.2.4 Birth–death processes

Examples

A *birth–death process* (BDP) is a continuous-time Markov chain whose jumps can only equal ± 1. Its transition rates are written

$$\begin{cases} n \to n+1 & \text{at rate} \quad \lambda_n \\ n \to n-1 & \text{at rate} \quad \mu_n, \end{cases}$$

with $\mu_0 = 0$. The Yule process corresponds to $\lambda_n = \lambda n$, $\mu_n = 0$, the linear BDP to $\lambda_n = \lambda n$, $\mu_n = \mu n$, the BDP with immigration to $\lambda_n = \lambda n + \rho$, $\mu_n = \mu n$, and the logistic BDP to $\lambda_n = \lambda n$, $\mu_n = \mu n + cn(n-1)$.

Boundaries

Theorems II.2.2. and II.2.3. in [3] on uniqueness of the Kolmogorov equations associated with BDP rate matrices can be rephrased in terms of *blowing up* (backward equation) and *coming down from infinity* (forward equation).

Theorem 1.2.9 *Assume that $\lambda_n > 0$ for all $n \geq 1$. Then the (minimal) BDP has infinite lifetime a.s. iff*

$$R := \sum_{n\geq 1} \left(\frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n \lambda_{n-1}} + \cdots + \frac{\mu_n \cdots \mu_2}{\lambda_n \cdots \lambda_2 \lambda_1} \right)$$

is infinite.

Exercise 1.5 *Prove that all four BDP's mentioned as examples have infinite lifetime a.s.*

Theorem 1.2.10 *Assume that $\mu_n > 0$ for all $n \geq 1$. Then the BDP comes down from infinity iff*

$$S := \sum_{n\geq 1} \frac{1}{\mu_{n+1}} \left(1 + \frac{\lambda_n}{\mu_n} + \cdots + \frac{\lambda_n \cdots \lambda_1}{\mu_n \cdots \mu_1} \right)$$

is finite.

Exercise 1.6 *Let $\alpha \in (1, +\infty)$ and define the generalized logistic BDP as the Markov chain with rates*

$$\begin{cases} n \to n+1 & \textit{at rate} \quad \lambda n \\ n \to n-1 & \textit{at rate} \quad \mu n^{\alpha}. \end{cases}$$

Prove that any generalized logistic BDP comes down from infinity (for $\alpha = 2$, this is Theorem 1.2.8 in the binary case).

Extinction

Assume that $\lambda_0 = 0$, so that 0 is absorbing for the BDP and let

$$u_n := \mathbb{P}(\text{Ext} \mid Z_0 = n)$$

be the probability that 0 is hit in finite time. Further, recall that T is the extinction time and let

$$\theta_n := \mathbb{E}(T, \mathrm{Ext} \mid Z_0 = n)$$

be the mean time to extinction on $\{\mathrm{Ext}\}$. Then the following recursions can be obtained, that rely on the fact that jumps of a BDP are of amplitude at most 1

$$\begin{cases} \lambda_n u_{n+1} - (\lambda_n + \mu_n) u_n + \mu_n u_{n-1} &= \quad 0 \\ \lambda_n \theta_{n+1} - (\lambda_n + \mu_n)\theta_n + \mu_n \theta_{n-1} &= -u_n \end{cases} \qquad n \geq 1.$$

The only non-trivial cases occur either when all rates λ_n, μ_n are nonzero ($n \geq 1$), or when they are nonzero up to a certain level N for which $\lambda_N = \mu_N = 0$. But this amounts to considering $u_n^{(N)} := \mathbb{P}_n(T < T_N)$ in the general case. Then set

$$U_N := \sum_{k=1}^{N-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} .$$

Elementary computations show that

$$u_n^{(N)} = (1 + U_N)^{-1} \sum_{k=n}^{N-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} \qquad n = \{1, \dots, N-1\}.$$

In particular, $u_1^{(N)} = U_N / 1 + U_N$.

Theorem 1.2.11 *If U_N grows to infinity as $N \to \infty$, then all extinction probabilities are equal to 1. If it converges to a finite limit U_∞, then*

$$u_n = (1 + U_\infty)^{-1} \sum_{k=n}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} \qquad n \geq 1.$$

Application to the branching process (linear BDP). when $\lambda \leq \mu$, the process is (sub)critical so that extinction occurs with probability 1, and if $\lambda > \mu$, then $u_n = (\mu/\lambda)^n$.

Next turn to the expected time to extinction. First, set

$$\theta_n^{(N)} := \mathbb{E}_n(T, T < T_N) \qquad n \leq N.$$

By Beppo Levi's theorem, $\theta_n^{(N)}$ converges to $\theta_n = \mathbb{E}_n(T, \mathrm{Ext})$ as $N \to \infty$. Set

$$\rho_k := \frac{\lambda_1 \cdots \lambda_{k-1}}{\mu_1 \cdots \mu_k} .$$

Theorem 1.2.12 *On $\{\mathrm{Ext}\}$, the expected time to extinction is finite iff $\sum \rho_k u_k^2 < \infty$. Then*

$$\theta_n = u_n \sum_{k=1}^{n-1} (1 + U_k) \rho_k u_k + (1 + U_n) \sum_{k \geq n} \rho_k u_k^2 \qquad n \geq 1,$$

and in particular,

$$\mathbb{E}_1(T, \mathrm{Ext}) = \sum_{k \geq 1} \rho_k u_k^2.$$

Corollary 1.2.13 *When* $\mathbb{P}(\mathrm{Ext}) = 1$*, the expected time to extinction is finite iff* $\sum \rho_k < \infty$*. Then*

$$\theta_n = \sum_{k=1}^{n-1}(1+U_k)\rho_k + (1+U_n)\sum_{k\geq n}\rho_k \qquad n \geq 1,$$

and in particular,

$$\mathbb{E}_1(T) = \sum_{k\geq 1}\rho_k.$$

Application to the branching process (linear BDP). It can be shown that when the branching process with birth rate λ and death rate μ is supercritical ($\lambda > \mu$) and *conditioned to extinction*, it has the same law as a branching process with birth rate μ and death rate λ. This allows us to concentrate on the case when extinction occurs a.s. Since in that case ($\lambda \leq \mu$), $\rho_k = \lambda^{k-1}/k\mu^k$,

$$\mathbb{E}_1(T) = \lambda^{-1}\sum_{k\geq 1}k^{-1}(\lambda/\mu)^k = -\lambda^{-1}\log(1-\lambda/\mu),$$

if $\lambda < \mu$, and $\mathbb{E}_1(T) = \infty$ if $\lambda = \mu$ (critical case).

Proof of the theorem. The recursion satisfied by $\theta_n^{(N)}$ is

$$\lambda_n\theta_{n+1}^{(N)} - (\lambda_n+\mu_n)\theta_n^{(N)} + \mu_n\theta_{n-1}^{(N)} = -u_n^{(N)} \qquad n \leq N.$$

Elementary manipulations show that

$$\theta_n^{(N)} = (1+U_n)\theta_1^{(N)} - \sum_{k=1}^{n-1}\sigma_k^{(N)},$$

where

$$\sigma_k^{(N)} := \sum_{i=1}^{k}\frac{u_i^{(N)}}{\lambda_i}\prod_{j=i+1}^{k}\frac{\mu_j}{\lambda_j},$$

an empty product being equal to 1 by convention. It is not harder to get

$$\sum_{k=1}^{n-1}\sigma_k^{(N)} = \sum_{k=1}^{n-1}\rho_k(U_n-U_k)u_k^{(N)} \qquad n \leq N.$$

In addition, because $\theta_N^{(N)} = 0$, we get

$$\theta_1^{(N)} = (1+U_N)^{-1}\sum_{k=1}^{N-1}\rho_k(U_N-U_k)u_k^{(N)} = \sum_{k=1}^{N-1}\rho_k\left(u_k^{(N)}\right)^2,$$

which, by Beppo Levi's theorem, yields

$$\theta_1 = \sum_{k\geq 1}\rho_k u_k^2,$$

where both sides might be infinite. Taking the same limit for $\theta_n^{(N)}$ gives

$$\theta_n = \mathbb{E}_n(T, \mathrm{Ext}) = (1 + U_n)\theta_1 - \sum_{k=1}^{n-1} \rho_k (U_n - U_k) u_k.$$

Replacing with the expression we just got for θ_1, and using the fact that when $U_\infty < \infty$, then $u_k = (U_\infty - U_k)/(1 + U_\infty)$, we get

$$\theta_n = u_n \sum_{k=1}^{n-1} \rho_k u_k (1 + U_k) + (1 + U_n) \sum_{k \geq n} \rho_k u_k^2,$$

and this expression also holds when the extinction probabilities are equal to 1. □

Chapter 2

Scaling Limits

In this chapter, we will consider real-valued stochastic processes in continuous time that arise as scaling limits from the models seen in the last chapter.

2.1 Discrete time

2.1.1 Fixed size : composition of bridges

Before dealing with continuous populations, recall the *Cannings model* for a population of fixed size $2N$, and label its individuals randomly as $i = 1, 2 \ldots, 2N$. For each i, let $\eta_n(i)$ be the number of descendants of individual i at generation n, and set

$$Y_n(i) := \sum_{j=1}^{i} \eta_n(j),$$

so that $Y_n(i)$ is the number of descendants, at generation n, of the subpopulation formed of the first i individuals. Observe that in the Cannings model, the $(\eta_n(i); i = 1, \ldots, 2N)$ are exchangeable, so that for each fixed n, the process $(Y_n(i); i = 0, 1, \ldots, 2N)$ is a *nondecreasing process with exchangeable increments* satisfying $Y_n(0) = 0$, and $Y_n(2N) = 2N$. Such a process is called a *discrete bridge.*

Consistent labelling assumption. The labelling of individuals at each generation is supposed to be *consistent* with the genealogy, that is, for any $1 \le i < j \le 2N$, the offspring of individual i at the next generation all have *smaller* labels than those of j.

Under the consistent labelling assumption, and by time homogeneity of the Cannings model, we can iterate the sampling of the first generation and assert that there are i.i.d. random applications $(B_n)_{n\ge1}$ from $\{1, 2, \ldots, 2N\}$ into itself, all distributed as Y_1, such that $Y_{n+1} = B_{n+1} \circ Y_n$. In conclusion, we got a representation of the genealogy of a Cannings model by *successive compositions of i.i.d. discrete bridges.*

To define the Cannings model in continuous-state space, we can analogously consider a chain $(Y_n(\cdot)\,; n \ge 0)$ of mappings from $[0, 1]$ to $[0, 1]$ with the following interpretation. For any $0 \le x \le 1$, $Y_n(x)$ is the *relative size*, at generation n, of the subpopulation *descending* from an initial subpopulation that had relative size x at generation 0. By analogy with the discrete case, the law of the chain $(Y_n(\cdot)\,; n \ge 0)$ is given as follows. There are i.i.d. *bridges* $(B_n)_{n\ge1}$ such that

$$Y_{n+1} = B_{n+1} \circ Y_n,$$

where

Definition 2.1.1 *A* bridge B *is a right-continuous process from* $[0, 1]$ *to* $[0, 1]$ *such that*
(i) $B(0) = 0$ *and* $B(1) = 1$
(ii) B *has a.s. nondecreasing paths and exchangeable increments.*

It is known since [60] that any bridge B can be represented as follows

$$B(x) = (1 - D)x + \sum_i \Delta_i \mathbf{1}_{\{s_i \le x\}} \qquad x \in [0, 1],$$

where the *jump times* $(s_i)_i$ form a sequence of i.i.d. *uniform* random variables, and the *jump sizes* $(\Delta_i)_i$ form an *exchangeable* sequence of positive random variables, independent from the latter sequence, such that $D := \sum_i \Delta_i \leq 1$ a.s.

In the next subsection, a similar construction of branching models is given. Namely, branching processes in continous state-space and discrete time, called *Jirina processes*, can be defined thanks to *compositions of subordinators*. Properties of both models, as well as relations between them, will be developed in the next section, concerning continuous populations in continuous time. The interested reader will consult [13, 14, 15, 16].

2.1.2 Stochastic size : Jirina process

Definition

Recall the BGW process $(Z_n(x); n \geq 0)$ where $x \in \mathbb{N}$ stands for the initial condition Z_0. From the branching property (1.1), we can construct a doubly indexed process $(Z_n(x)\,; n, x \geq 0)$ such that, for each fixed integer n, $(Z_n(x)\,; x = 0, 1, 2, \ldots)$ is a Markov chain whose increments are i.i.d. with common law that of $Z_n(1)$. Indeed, $Z_n(x + y) - Z_n(x)$ is the descendance at generation n of individuals $x + 1, \ldots, y$, so it is independent of $Z_n(x)$ and has he law of $Z_n(y)$. Therefore, it is an integer-valued *increasing random walk* starting from 0. Recall that the equivalent in continuous state-space of an increasing random walk is a *subordinator*, that is, an increasing *Lévy process* (see Appendix).

The *Jirina process* is a branching process in discrete time but continuous state-space. Precisely, the Jirina process is a time-homogeneous Markov chain $(Z_n; n \geq 0)$ with values in $[0, +\infty)$ satisfying the branching property (1.1). As in the discrete case, writing $Z_0 = x \in [0, +\infty)$, then for each integer n, $(Z_n(x)\,; x = 0, 1, 2, \ldots)$ has i.i.d. nonnegative increments. In particular, $(Z_1(x)\,; x \geq 0)$ is a subordinator, that we prefer to denote S. Let F be its Laplace exponent

$$\mathbb{E}(\exp(-\lambda S(x))) = \exp(-xF(\lambda)) \qquad \lambda, x \geq 0.$$

By time homogeneity, the descendance at generation n of the individuals $1, \ldots, x$, is the descendance at generation 1, of their descendance at generation n. Rigorously, there are i.i.d. subordinators $(S_n)_{n\geq 1}$ distributed as S, such that, conditional on $Z_0, Z_1, \ldots, Z_n$,

$$Z_{n+1} = S_{n+1} \circ Z_n.$$

In particular, by Bochner's subordination, the process $x \mapsto Z_n(x)$ is a subordinator with Laplace exponent F_n the n-th iterate of F, so that

$$\mathbb{E}_x(\exp(-\lambda Z_n)) = \exp(-xF_n(\lambda)) \qquad \lambda \geq 0.$$

Definition 2.1.2 *We say that Z is a Jirina process with* branching mechanism *F.*

Interpretation

In what follows, we use the word *lifetime* for a time interval whose width is called the *lifespan*. Assume that the subordinator S is a *compound Poisson process*, that is, it has no drift and its Lévy measure Λ is finite. In other words, $S_t = \sum_{s\leq t} \Delta_s$, where (s, Δ_s) are the atoms of a Poisson point process with intensity measure $dx\Lambda(dy)$. Let b denote the mass of the Lévy measure Λ. Now we consider a random tree where each individual is given a *lifespan* in $(0, +\infty)$

- generation $n+1$ is made up of the *offspring* of individuals from generation n
- each individual from generation n gives birth at *rate* b during her *lifetime*, to one offspring at a time, and the *lifespans* of her offspring are i.i.d. with distribution $\Lambda(\cdot)/b$
- conditional on the lifespans and the number $\mathcal{Z}_n$ of individuals from generation n, the birth processes are *independent*.

The *sum* Z_n *of all lifespans* of individuals from generation n is the value of the Jirina process at time n. The number $\mathcal{Z}_n$ of individuals from generation n is equal to the number of jumps of the compound Poisson process S_n on the interval $[0, Z_{n-1}]$. The chain $(\mathcal{Z}_n; n \geq 0)$ is a BGW process whose offspring distribution is a mixed Poisson law

$$p_k = \int_0^\infty \Lambda(dx) b^{-1} e^{-bx} \frac{(bx)^k}{k!} \ .$$

The Jirina process Z and the BGW process $\mathcal{Z}$ coincide if $\Lambda = b\delta_1$, where δ_1 is the Dirac measure at 1. Trees constructed as previously are called *splitting trees*, and Chapter 4 will be devoted to their study.

Immigration

Thanks to the previous interpretation, the natural construction of a Jirina process with immigration is as follows. Let $(I_n)_{n\geq 1}$ be a sequence of i.i.d. random nonnegative real numbers. The variable I_n embodies the immigration at generation n, and in the interpretation given earlier, I_n is to be seen as the lifespan of a single immigrating individual. Consequently, the *Jirina process with immigration* can be defined as the Markov process $(Z_n; n \geq 0)$ such that, conditional on $Z_0, Z_1, \dots, Z_n$,

$$Z_{n+1} = I_{n+1} + S_{n+1} \circ Z_n,$$

where $(S_n)_{n\geq 1}$ are i.i.d. subordinators. Let F be the common Laplace exponent of these subordinators, and G be the common Laplace transform of the I's

$$G(\lambda) := \mathbb{E}(\exp(-\lambda I_1)) \qquad \lambda \geq 0.$$

Then if F_n is the n-th iterate of F, it is easy to show by induction that

$$\mathbb{E}_x(\exp(-\lambda Z_n)) = \exp(-xF_n(\lambda)) \prod_{k=0}^{n-1} G \circ F_k(\lambda) \qquad \lambda \geq 0. \tag{2.1}$$

Note the similarity with (1.3)

Definition 2.1.3 *We say that Z is a Jirina process with* branching mechanism F *and* immigration mechanism G.

2.2 Continuous time

In this section, we consider real-valued stochastic processes that are the analogues in continuous time of the models seen earlier. Except for a whole subsection on diffusions, we define these Markov processes directly from their transition semigroup, so we will not need to say too much about *infinitesimal generators*. In the first subsection, though, we say a word about the effect on generators of conditionings on extinction or fixation.

2.2.1 Generators, absorption times and conditionings

Consider a continuous time Markov process $(X_t; t \geq 0)$ with values in $I \subseteq [0, \infty)$ and transition semigroup $(P_t; t \geq 0)$ defined for any bounded measurable f by

$$P_t f(x) := \mathbb{E}_x(f(X_t)) \qquad x \in I.$$

Roughly speaking, the infinitesimal generator of X is a linear operator L which is the analogue of the rate matrix in the discrete setting. It is defined on a vector space of sufficiently smooth functions called its *domain*. For each f in the domain of L, Lf is a function satisfying

$$Lf(x) = \lim_{t\downarrow 0} \frac{1}{t}(P_t f(x) - f(x)) \qquad x \in I.$$

In the continuous setting, the *Kolmogorov backward equations* read

$$P_t f(x) = f(x) + \int_0^t ds\, LP_s f(x),$$

whereas the *Kolmogorov forward equations* read

$$P_t f(x) = f(x) + \int_0^t ds\, P_s L f(x).$$

Next set τ the first hitting time T_o of some *absorbing* point o in I such that for any $x \in I$,

$$u(x) := \mathbb{P}_x(\tau < \infty) > 0,$$

that is, o is also *accessible*. Set also

$$f(x) := \mathbb{E}_x(\tau, \tau < \infty).$$

To have a sense of what we are doing, think of

- populations with fixed size, where $o = 1$, so that $\{\tau < \infty\}$ is the event of *fixation*
- general populations with no immigration, where $o = 0$, so that $\{\tau < \infty\}$ is the event of *extinction*, or more precisely, extinction *with absorption*, since for continuous populations, extinction occurs as soon as $X_t \to 0$.

The next statement provides harmonic equations satisfied by

$$U(t, x) := \mathbb{P}_x(\tau > t) \quad \text{and} \quad F(t, x) := \mathbb{E}_x(\tau, \tau < t).$$

The proof relies on the use of Kolmogorov equations applied to the function

$$g(x) := \mathbf{1}_{\{x \neq o\}},$$

which is assumed to be in the domain of the generator, which means in particular that τ has a density w.r.t. the Lebesgue measure.

Theorem 2.2.1 *Assume that g is in the domain of L. Then for any $x \neq o$ and $t \geq 0$,*

$$LU(t, x) = \frac{\partial U}{\partial t}(t, x) \quad \text{and} \quad Lu(x) = 0.$$

In addition,

$$LF(t, x) = U(t, x) - 1 \quad \text{and} \quad Lf(x) = -u(x).$$

Proof. Observe that $U(t,x) = P_t g(x)$ and, by Fubini's theorem

$$F(t,x) = \mathbb{E}_x \int_0^t \mathbf{1}_{\{X_s \neq o\}} ds = \int_0^t P_s g(x)\, ds.$$

An application of the backward equations to g yields

$$U(t,x) = P_t g(x) = g(x) + \int_0^t ds\, L P_s g(x)\, ds,$$

so that $U(t,x) = 1 + \int_0^t ds\, LU(x,s)\, ds$, since $g(x) = 1$ when $x \neq o$. Differentiating this last equation yields the first equation of the theorem. For the second one, notice that by the Markov property, $u(X_t) = \mathbb{P}(\tau < \infty \mid \mathcal{F}_t)$, so that

$$P_t u(x) = \mathbb{E}_x(\mathbb{P}(\tau < \infty \mid \mathcal{F}_t)) = \mathbb{P}_x(\tau < \infty) = u(x),$$

which implies $Lu(x) = 0$.

For the second part of the theorem, apply L to F, which yields

$$LF(t,x) = L \int_0^t P_s g(x)\, ds = \int_0^t L P_s g(x)\, ds = \int_0^t LU(x,s)\, ds = U(t,x) - 1,$$

where swapping L and the integral can be justified as follows. First, thanks to Fubini's theorem and the semigroup property, we get

$$P_u \int_0^t P_s g(x)\, ds = \mathbb{E}_x \int_0^t P_s g(X_u)\, ds = \int_0^t ds\, \mathbb{E}_x P_s g(X_u) = \int_0^t ds\, P_u P_s g(x) = \int_0^t ds\, P_{u+s} g(x),$$

so that

$$\begin{aligned}
\frac{1}{u}\left(P_u \int_0^t P_s g(x)\, ds - \int_0^t P_s g(x)\, ds\right) &= \frac{1}{u}\int_0^t (P_{u+s} g(x) - P_s g(x))\, ds \\
&= \frac{1}{u}\int_0^t (\mathbb{P}_x(\tau > s+u) - \mathbb{P}_x(\tau > s))\, ds \\
&= -\frac{1}{u}\int_0^t ds \int_u^{u+s} dv\, w(v),
\end{aligned}$$

where w stands for the density $-\partial U/\partial t$ of τ. By Fubini's theorem again, we finally get

$$LF(t,x) = \lim_{u \downarrow 0} -\frac{1}{u}\int_0^{t+u} dv\, w(v)(v \wedge u),$$

and the result follows by dominated convergence. As for the last equation $Lf = -u$, the same proof rigorously applies with t set to $+\infty$. □

In the next statement, we characterize the law of the process X conditioned by absorption at o via a harmonic change of measure.

Theorem 2.2.2 *The process* $(u(X_t);t\geq 0)$ *is a positive martingale and* $\mathbb{P}^\star := \mathbb{P}(\cdot \mid \tau<\infty)$ *is obtained by the following h-transform*

$$\mathbb{P}^\star_x(\Theta) = \mathbb{E}_x\left(\mathbf{1}_\Theta \frac{u(X_t)}{u(x)}\right) \qquad x\in I,$$

for any event Θ *in the* σ*-field* $\mathcal{F}_t$ *generated by* $(X_s;s\leq t)$*. As a consequence, the generator* $L^\star$ *of the process* X *conditioned by* $\{\tau<\infty\}$ *is given by*

$$L^\star f(x) = \frac{L(uf)(x)}{u(x)} \qquad x\in I.$$

Proof. Since $u(X_t)=\mathbb{P}(\tau<\infty\mid\mathcal{F}_t)$, it is trivially a martingale. Then for any $\Theta\in\mathcal{F}_t$,

$$\mathbb{P}^\star_x(\Theta) = \frac{1}{u(x)}\mathbb{E}_x(\mathbf{1}_\Theta\mathbf{1}_{\{\tau<\infty\}}) = \frac{1}{u(x)}\mathbb{E}_x(\mathbf{1}_\Theta\mathbb{P}(\tau<\infty\mid\mathcal{F}_t)),$$

which shows the first part of the theorem. As a consequence,

$$P^\star_t f(x) = \mathbb{E}_x\left(f(X_t)\frac{u(X_t)}{u(x)}\right) = \frac{1}{u(x)}P_t(uf)(x),$$

so that

$$\frac{1}{t}(P^\star_t f(x) - f(x)) = \frac{1}{tu(x)}(P_t(uf)(x) - (uf)(x)),$$

and the result follows letting $t\downarrow 0$. □

2.2.2 Fixed size : generalized Fleming–Viot processes

In the very beginning of the present chapter, we have shown that the genealogy of the Cannings model can be represented as follows. Define B a *discrete bridge* as an increasing sequence $(B(j)\,;0\leq j\leq 2N)$ that has *exchangeable increments* and satisfies $B(0)=0$ and $B(2N)=2N$. The interpretation is that $B(j)$ is the number of descendants, after one generation, of the first j individuals of the initial population. Now let B_n be i.i.d. discrete bridges, and for any integers $m<n$, set $B_{m,m}:=\mathrm{Id}$ and

$$B_{m,n} := B_n\circ\cdots\circ B_{m+1}.$$

Then $B_{m,n}(j)$ is the size, at generation n, of the subpopulation descending from the first j individuals of the population at generation m (this interpretation requires a consistent labelling of individuals that was mentioned in the first subsection). Equivalently, $B_{m,n}(j)-B_{m,n}(j-1)$ is the descendance at generation n of individual j belonging to generation m. In particular,

$$Y_n := B_{0,n}$$

gives the genealogical structure of the initial population, at generation n.

Then we have shown that this construction can be generalized to continuous populations by considering *bridges* as in Definition 2.1.1, that is, nondecreasing right-continuous processes B with exchangeable increments from $[0,1]$ to $[0,1]$ such that $B(0)=0$ and $B(1)=1$. All interpretations given above carry over to continuous populations provided that the phrase 'population size' is replaced with 'relative size' (i.e., proportion, frequency). A slight difference is that the descendance at generation n of individual x belonging to generation m is given by $B_{m,n}(x) - B_{m,n}(x-)$.

Thus, a *discrete flow of* (non-discrete) bridges can be constructed directly by generating i.i.d. bridges and composing them as previously, which yields a collection of bridges $(B_{m,n}\,; 0 \leq m \leq n)$ such that

$$B_{m,n} \circ B_{\ell,m} = B_{\ell,n} \qquad \ell \leq m \leq n,$$

where the law of a bridge $B_{m,n}$ depends solely on $n-m$ and for any $n_1 \leq \cdots \leq n_k$, the bridges $B_{n_1,n_2}, \ldots, B_{n_{k-1},n_k}$ are independent. One of the tasks of [14] was to construct a flow of bridges in continuous time.

Definition 2.2.3 *A flow of bridges is a collection of bridges $(B_{s,t}\,; 0 \leq s \leq t)$ such that*

(i) for any $s < t < u$,

$$B_{t,u} \circ B_{s,t} = B_{s,u}$$

(ii) the law of a bridge $B_{s,t}$ solely depends on $t-s$ and for any $t_1 \leq \cdots \leq t_k$, the bridges $B_{t_1,t_2}, \ldots, B_{t_{k-1},t_k}$ are independent

(iii) the bridge $B_{0,0}$ is the identity, and for every $x \in [0,1]$, $B_{0,t}(x)$ converges to x in probability as $t \downarrow 0$.

Such a construction can be achieved via a Poissonian construction that relies on *simple bridges*. More precisely, to every $u \in [0,1]$ and $r \in [0,1]$, we can associate the *simple bridge* $b_{u,r}$ for which u is the unique jump time and r the size of this jump, as

$$b_{u,r}(x) := (1-r)x + r\mathbf{1}_{\{u \leq x\}} \qquad x \in [0,1].$$

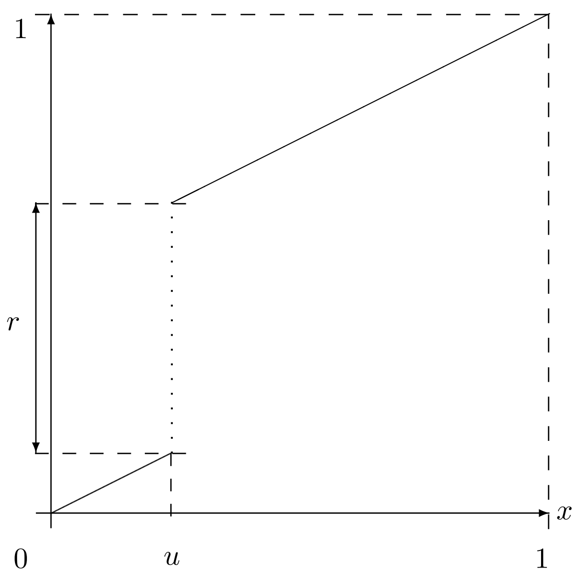

Figure 2.1: Graph of the simple bridge $x \mapsto b_{u,r}(x)$.

Here comes the Poissonian construction of the flow. Let ν be a positive measure on $(0, 1]$ such that $\int_0^1 x^2\nu(dx) < \infty$, and $(t_i, u_i, r_i; i \in \mathbb{N})$ the atoms of a Poisson measure with intensity $dt \otimes du \otimes \nu_n(dr)$ on $[0, \infty] \times (0, 1) \times (0, 1]$, where

$$\nu_n(dr) := \mathbf{1}_{\{r>1/n\}}\nu(dr).$$

Since we have truncated the measure ν to $(1/n, 1]$, this Poisson measure is finite on any bounded time interval, and we can assume that its atoms are ranked in the increasing order of their first component $t_1 < t_2 < \cdots$ Next define the simple bridge $b^{(i)} := b_{u_i,r_i}$, and for any $0 \le s < t$, set

$$B^n_{s,t} := b^{(j)} \circ \cdots \circ b^{(i)},$$

whenever there are $i < j$ such that $t_i \le s < t_{i+1}$ and $t_j \le t < t_{j+1}$, otherwise $B^n_{s,t} := \mathrm{Id}$.

Theorem 2.2.4 (Bertoin–Le Gall [14]) *The collection $(B^n_{s,t}\,; 0 \le s \le t)$ is a flow of bridges, called ν_n-simple flow. It converges weakly, as $n \to \infty$, to a flow of bridges $(B_{s,t}\,; 0 \le s \le t)$ called ν-simple flow.*

Actually, the primary highlight in [14] is not the previous nice construction of continuous flows of bridges but a one-to-one correspondence between these flows and Λ-*coalescents*. Despite the complexity of the subject, we think it might be useful to the reader to get a flavour of this area.

A Λ-coalescent is a Markov process $(\Pi_t; t \ge 0)$ with values in the partitions of the integers, associated with a finite measure Λ on $(0, 1]$. The construction is roughly as follows. First, generate a Poisson point process on $[0, \infty) \times (0, 1]$ with intensity $dt \otimes r^{-2}\Lambda(dr)$. Then at each atom (t, r), mark each block of the current partition independently and with probability r, and merge together all marked blocks (to get a coarser partition).

Now go back to the ν-simple flow $(B_{s,t}\,; 0 \le s \le t)$. Consider a collection $(V_i; i \in \mathbb{N})$ of independent uniform r.v. on $(0, 1)$, and fix $t > 0$. Then the equivalence relation

$$i \sim_s j \Leftrightarrow B^{-1}_{s,t}(V_i) = B^{-1}_{s,t}(V_j)$$

induces a random partition $\tilde{\Pi}_{s,t}$ on the integers. In words, $i \sim_s j$ if V_i and V_j 'fall into' the same jump of $B_{s,t}$. By construction, the partition $\tilde{\Pi}_{s,t}$ gets finer as s grows, ending at the all-singleton partition $\tilde{\Pi}_{t,t}$. The main theorem in [14] states that $(\tilde{\Pi}_{t-s,t}; 0 \le s \le t)$ is Markovian and has the same transitions as the Λ-coalescent $(\Pi_s; 0 \le s \le t)$ started at the all-singleton partition, with

$$\Lambda(dx) = x^2\nu(dx).$$

A beautiful consequence of this correspondence is the following formula (equation (8) in [15])

$$\mathbb{E}(Y_t(x)^n) = \mathbb{E}(x^{\#\Pi^n_t}), \tag{2.2}$$

where $\#\Pi^n_t$ is the number of blocks at time t of the Λ-coalescent restricted to $\{1, \dots, n\}$. The two sides of this equation can be seen as two different ways of expressing the probability that n individuals drawn at random in the population at time t have ancestors in the same subpopulation of relative size x. This is a generalization of a formerly known duality between the Kingman coalescent and Fisher–Wright diffusion (see forthcoming Theorem 2.3.2).

Proof of the theorem. The fact that $(B^n_{s,t}\,;0\le s\le t)$ is a flow follows essentially from the fact that the composition of independent bridges is a bridge. The weak convergence relies on the one-to-one correspondence with Λ-coalescents mentioned in the foregoing paragraph, and the associated weak convergence proved by J. Pitman [87]. □

Now set

$$Y_t := B_{0,t}.$$

Definition 2.2.5 *The real number $Y_t(x)\in[0,1]$ is the relative size, at time t, of the subpopulation descending from an initial subpopulation of relative size x.*

The process $(Y_t(x)\,;t\ge 0, x\in[0,1])$ is called a generalized Fleming–Viot process*, or* GFV-process*. This terminology is justified by the fact that $Y_t(\cdot)$ is a random* distribution function *on $[0,1]$, and hence $dY_t(x)$ can be viewed as a random* population density *on $[0,1]$.*

Another result of Bertoin and Le Gall is that for any p-tuple $(x_1,\dots,x_p)$ of $[0,1]$, the càdlàg process $(Y_t(x_1),\dots Y_t(x_p)\,;t\ge 0)$ is a Markov process. By exchangeability, $\mathbb{E}(Y_t(x)) = x$, so $(Y_t\,;t\ge 0)$ is a martingale.

Theorem 2.2.6 (Bertoin–Le Gall [14]) *The limit of the martingale $(Y_t\,;t\ge 0)$ is 0 or 1 a.s. and for any $x\in(0,1)$,*

$$\mathbb{P}(\lim_{t\to\infty} Y_t(x) = 1) = x.$$

By monotonicity of bridges, there is a random uniform point $\mathbf{e}\in(0,1)$ called the primitive eve*, such that $\lim_{t\to\infty} Y_t(x)$ equals 0 for any $x<\mathbf{e}$ and 1 for any $x\ge\mathbf{e}$. In addition,*

$$\lim_{t\to\infty} Y_t(\mathbf{e}) - Y_t(\mathbf{e}-) = 1.$$

Other interesting results in this direction are displayed in Subsection 2.2.4.

2.2.3 Stochastic size : continuous-state branching process

In this subsection, we introduce and study the *continuous-state branching process*, or *CSBP*, or *CB-process*, which is a strong Markov process $(Z_t;t\ge 0)$ with values in $[0,\infty]$ and càdlàg paths satisfying the branching property. In particular, 0 is an absorbing state for Z. CB-processes were first considered in [58]. Fundamental properties of the CB-process were discovered in [17, 47, 78]. One might like to consult [31], [71, chapter 10], or [61].

Definition 2.2.7 *We will say that* extinction *occurs if $\lim_{t\to\infty} Z_t = 0$, and that* absorption *occurs if there is some t such that $Z_t = 0$. The first event will be denoted $\{\mathrm{Ext}\}$ as usual, and the second one by $\{\mathrm{Abs}\}$.*

Lamperti transform

By analogy with the continuous-time, discrete state-space setting, we will see that there is a one-to-one correspondence between CB-processes and the continous analogue of left-continuous random walks, namely *Lévy processes with no negative jumps*. This correspondence can be seen thanks to various different bijections, but the simplest one is the *Lamperti transform*, which is exactly the same time-change as that given in the previous chapter. Similarly as in the discrete setting, define implicitly X as the solution to

$$Z_t = X\left(\int_0^t Z_s\,ds\right) \qquad t \geq 0.$$

More rigorously, let T be the (possibly infinite) absorption time of the branching process and set

$$\theta_t := \int_0^t Z_s\,ds \qquad t > 0.$$

Since θ is continuous increasing on $[0, T)$, we let κ be its inverse on $[0, \theta_T)$. Next define

$$X_t := Z \circ \kappa_t \qquad t < \theta_T.$$

To contrast with the law $\mathbb{P}_x$ of the CB-process Z started at $x \geq 0$, we will denote by P_x the law of X started at $x \in \mathbb{R}$. For the same reason, T_0 will denote the lifetime of X, that is, its (possibly infinite) first hitting time of 0.

Theorem 2.2.8 (Lamperti [78]) *The process $(X_t; t \geq 0)$ is a Lévy process with no negative jumps killed upon hitting 0. Let ψ be its Laplace exponent (when not killed), defined for any x as*

$$E_x(\exp -\lambda X_t) = E_0(\exp -\lambda(X_t + x)) = \exp(-\lambda x + t\psi(\lambda)). \tag{2.3}$$

Then the Laplace transform of the one-dimensional marginal of Z is given by

$$\mathbb{E}_x(\exp -\lambda Z_t) = \exp(-x u_t(\lambda)) \qquad t, x, \lambda \geq 0,$$

where $t \mapsto u_t(\lambda)$ is the unique nonnegative solution of the integral equation

$$v(t) + \int_0^t \psi(v(s))ds = \lambda. \tag{2.4}$$

Remark 2.1 *Exactly as in the discrete setting, the Lamperti transform can be performed in both directions, starting with a Lévy process X with no negative jumps, noticing that for $t < T_0$,*

$$\kappa_t = \int_0^t \frac{ds}{X_s},$$

and that θ is the inverse of κ, stopped at $\kappa(T_0-)$. Conclude with $Z = X \circ \theta$. In particular, this procedure can be achieved without reference to the initial process Z, so it proves at the same time the existence of CB-processes, and their one-to-one correspondence with Lévy processes with no negative jumps.

Surprisingly, the proof of this theorem is more difficult than the discrete version would let think, and to my knowledge, it has not been published anywhere. The following proof is part of a work in progress with Maria-Emilia Caballero and Gerónimo Uribe (UNAM, Mexico). Actually, this proof is about to be shortened at the moment of sending the file...

Proof. As a first step, we show that Z has no negative jumps, so X will not either. For any $\delta > 0$, let

$$S_\delta := \inf\{t : Z_t - Z_{t-} < -\delta\},$$

and assume that $\mathbb{P}_x(S_\delta < \infty) > 0$ for some δ. Then there is some δ such that for any $t > 0$, we can find $x \geq \delta$ for which

$$\mathbb{P}_x(S_\delta < t) > 0.$$

Indeed, if this were not to hold, a recursive application of the Markov property would show that $S_\delta = \infty$ a.s. for any δ. Now for any integer n, by the branching property, $Z(x)$ is the sum of n independent copies of $Z(x/n)$. Then since Z has no fixed time discontinuity, if none of these copies has a negative jump of amplitude greater than δ in the interval $[0, t)$, neither will their sum. As a consequence, $\mathbb{P}_{x/n}(S_\delta < t) > 0$ for any n. This entails that for all $t > 0$ and all $0 < \varepsilon \leq \delta$, $\mathbb{P}_\varepsilon(S_\delta < t) > 0$. This contradicts the fact that Z is right-continuous and takes only nonnegative values.

The second step of the proof focuses on the generators of Z and X. To account for the fact that these processes can have finite lifetime, we consider the one-point compactification of $[0, \infty)$ as $[0, \infty) \cup \{\partial\}$, where ∂ is an additional point usually called *cemetery state*. Then set $Z_t = \partial$ for all $t \geq D$, where $D := T_\infty = \sup\{t : Z_t \in [0, \infty)\}$ is the lifetime of the CB-process and $X_t = \partial$ for all $t \geq \tau$, where

$$\tau := T_0 \wedge T_\infty = \theta(T \wedge D).$$

Recall from standard theory that all functions vanish at ∂, so that continuous functions on $[0, \infty) \cup \{\partial\}$ vanish at ∞. In addition, we let (P_t) be the semigroup of Z, and L its infinitesimal generator. For any f in the extended domain of L, in particular f is *continuous* on $[0, \infty) \cup \{\partial\}$, define

$$M_t^f := f(Z_t) - \int_0^t Lf(Z_s)\, ds,$$

which is a martingale by definition, and define also

$$N_t^f := f(X_t) - \int_0^t Af(X_s)\, ds,$$

where A is the linear operator defined for any f in the extended domain of L by $(Af(\partial) = 0$ and)

$$Af(x) := x^{-1} Lf(x) \qquad x > 0.$$

Changing variable $s = \theta_u$ in the integral, we get

$$\begin{aligned}
N_{t\wedge\tau}^f &= f(Z(\kappa_t \wedge T \wedge D)) - \int_0^{\kappa_t \wedge T \wedge D} Af(Z_u)\, Z_u\, du \\
&= f(Z(\kappa_t \wedge T \wedge D)) - \int_0^{\kappa_t \wedge T \wedge D} Lf(Z_u)\, du \\
&= M^f(\kappa_t \wedge T \wedge D).
\end{aligned}$$

Now since κ_t, T and D are all stopping times, for any $s > 0$ and $0 < u < t$, by the optional stopping theorem,

$$\mathbb{E}_x(M^f(s \wedge \kappa_t \wedge T \wedge D) \mid \mathcal{F}_{\kappa_u}) = M^f(s \wedge \kappa_u \wedge T \wedge D).$$

As a consequence, for any fixed s, $(N^f(\theta_s \wedge t \wedge \tau); t \geq 0)$ is a martingale w.r.t. the natural filtration of X. Next, observe that

$$\theta_s < t \wedge \tau \Leftrightarrow \int_0^{t\wedge\tau} \frac{dr}{X_r} > s,$$

so that we have to distinguish two cases. If $\int_0^{t\wedge\tau} dr/X_r < \infty$, then $\lim_{s\to\infty} \theta_s \geq t \wedge \tau$. If $\int_0^{t\wedge\tau} dr/X_r = \infty$, then clearly $T_0 \leq t$ and $\lim_{s\to\infty} \theta_s = T_0$. In all cases,

$$\lim_{s\to\infty} \theta_s \wedge t \wedge \tau = t \wedge \tau,$$

but one has to be careful that on $\{T_0 \leq t\}$, $\lim_{s\to\infty} X(\theta_s \wedge t \wedge \tau) = 0$, so that for any continuous f vanishing at ∞,

$$\lim_{s\to\infty} f(X(\theta_s \wedge t \wedge \tau)) = f(\tilde{X}(t \wedge \tau)),$$

where $\tilde{X}$ is the process *stopped, not killed*, when it hits zero. Precisely, $\tilde{X}_t := X_t \mathbf{1}_{t<T_0}$. Now the martingale property of $(N^f(\theta_s \wedge t \wedge \tau); t \geq 0)$ entails, for any $s > 0$,

$$\mathbb{E}_x(f(X(\theta_s \wedge t \wedge \tau)) = f(x) + \mathbb{E}_x \int_0^t Af(X_u)\mathbf{1}_{\{u<\theta_s\wedge t\wedge\tau\}}\, du.$$

Then for any f in the extended domain of L such that Lf *is of constant sign*, letting $s \to \infty$ in the last displayed equation and applying respectively the dominated convergence theorem to the l.h.s. (f is bounded because it is continuous and vanishes at ∞) and the monotone convergence theorem to the r.h.s., we get

$$\mathbb{E}_x(f(\tilde{X}(t \wedge \tau)) = f(x) + \mathbb{E}_x \int_0^{t\wedge\tau} Af(X_u)\, du. \tag{2.5}$$

As a third step, we study the Laplace transforms of Z. By the branching property, the law of Z_t is infinitely divisible so for any $\lambda, t \geq 0$, there is a nonnegative real number $u_t(\lambda)$ such that

$$\mathbb{E}_x(\exp -\lambda Z_t) = \exp(-xu_t(\lambda)),$$

where it is easily seen that $\lambda \mapsto u_t(\lambda)$ is a nondecreasing, infinitely differentiable function. Setting $e_\lambda(x) := \exp(-\lambda x)$, the last relation can be expressed as

$$P_t e_\lambda = e_{u_t(\lambda)}.$$

The *semigroup property* of P_t carries over to u_t, since

$$e_{u_{t+s}(\lambda)} = P_{t+s}e_\lambda = P_t P_s e_\lambda = P_t e_{u_s(\lambda)} = e_{u_t \circ u_s(\lambda)}$$

which reads

$$u_{t+s} = u_t \circ u_s. \tag{2.6}$$

A less formal way to write the previous sequence of equations is to apply the Markov property as follows

$$\mathbb{E}_x(\exp -\lambda Z_{t+s}) = \mathbb{E}_x(\mathbb{E}(\exp -\lambda Z_{t+s} \mid \mathcal{F}_t)) = \mathbb{E}_x(\exp(-Z_t u_s(\lambda))) = \exp(-x u_t \circ u_s(\lambda)).$$

On the other hand, by dominated convergence,

$$\lim_{t\downarrow 0} \mathbb{E}_x(\exp -\lambda Z_t) = \exp(-\lambda x),$$

so that $t \mapsto u_t(\lambda)$ is continuous at 0^+, and

$$\lim_{t\downarrow 0} u_t(\lambda) = \lambda \qquad \lambda \geq 0.$$

Note that the semigroup property (2.6) entails the continuity of the function $t \mapsto u_t(\lambda)$ everywhere. Actually, this function is even differentiable, since e_λ is always in the extended domain of L, and by the Kolmogorov backward equation

$$e_{u_t(\lambda)}(x) = e_\lambda(x) + \int_0^t ds\, Le_{u_s(\lambda)}(x) \qquad t, x \geq 0,$$

which yields after differentiation

$$-x\frac{\partial u_t(\lambda)}{\partial t} e_{u_t(\lambda)}(x) = Le_{u_t(\lambda)}(x). \tag{2.7}$$

In particular, setting

$$F(\lambda) := \frac{\partial u_t(\lambda)}{\partial t} \,|_{t=0},$$

and $t = 0$ in the last equation, we get

$$Le_\lambda(x) = -xF(\lambda)e_\lambda(x) \qquad \lambda, x \geq 0. \tag{2.8}$$

Recasting this in (2.7), we get the fundamental relationship

$$\frac{\partial u_t(\lambda)}{\partial t} = F(u_t(\lambda)) \qquad \lambda, t \geq 0. \tag{2.9}$$

Next, let $\varphi(t) := \lim_{\lambda\to\infty} \uparrow u_t(\lambda)$, so that

$$\mathbb{P}_x(Z_t = 0) = \exp(-x\varphi(t)) \qquad t, x \geq 0.$$

In the case when 0 is accessible by Z, $\varphi(t) < \infty$ for any t and $\lim_{t\downarrow 0} \varphi(t) = +\infty$. Also, because 0 is absorbing, φ is decreasing. On the other hand, the semigroup property of (u_t) and their continuity, entail that $\varphi(t+\varepsilon) = u_\varepsilon(\varphi(t))$, so letting $\varepsilon \downarrow 0$, we get

$$\varphi'(t) = F \circ \varphi(t) \qquad t \geq 0.$$

Since φ is decreasing, setting $\eta := \lim_{t\to\infty} \downarrow \varphi(t)$, we get $F(\lambda) < 0$ for any $\lambda > \eta$.

In the case when 0 is not accessible by Z, $\varphi \equiv \infty$, but still, η can be defined by $\exp(-x\eta) = \lim_{\varepsilon\downarrow 0} \mathbb{P}_x(T_\varepsilon < \infty)$. Actually, it is possible to prove that if $\eta < \infty$, then one also has $F(\lambda) < 0$

when $\lambda > \eta$, and that $\eta = \infty$ if and only if Z has a.s. increasing paths.

As a fourth step, we characterize the law of X, applying equation (2.5) to exponential functions. Recall (2.8). Since the function e_λ is in the extended domain of L and Le_λ is of constant sign, we can apply (2.5) to this function and obtain

$$\mathbb{E}_x(\exp(-\lambda\tilde{X}(t\wedge\tau))) = e^{-\lambda x} - F(\lambda)\,\mathbb{E}_x\int_0^{t\wedge\tau}\exp(-\lambda X_s)\,ds.$$

Then, observing that $\mathbb{E}_x(\exp(-\lambda\tilde{X}(t\wedge\tau))) = \mathbb{E}_x(\exp(-\lambda X_t), t<\tau) + \mathbb{P}_x(T_0 \le t)$, elementary methods for the resolution of ordinary differential equations yield

$$\mathbb{E}_x(\exp(-\lambda X_t + tF(\lambda)), t<\tau) = e^{-\lambda x} - \mathbb{E}_x(\exp(F(\lambda)T_0), T_0 \le t) \qquad t,\lambda \ge 0. \tag{2.10}$$

In the case when $T_0 = \infty$ a.s.,

$$\mathbb{E}_x(\exp(-\lambda X_t), t < T_\infty) = \mathbb{E}_x(\exp(-\lambda X_t)) = \exp(-\lambda x - tF(\lambda)) \qquad t,\lambda \ge 0,$$

which proves that X is a Lévy process with Laplace exponent $\psi = -F$ given by (2.3). Actually, since $T_0 = \infty$, X must be a subordinator.

In all other cases we have $\eta < \infty$, and for any $\lambda > \eta$, $F(\lambda) < 0$, so that the l.h.s. in (2.10) vanishes as $t \to \infty$, and

$$\mathbb{E}_x(\exp(F(\lambda)T_0), T_0 < \infty) = \exp(-\lambda x) \qquad \lambda > \eta.$$

Then recasting this into (2.10) provides

$$\mathbb{E}_x(\exp(-\lambda X_t + tF(\lambda)), t<\tau) = \mathbb{E}_x(\exp(F(\lambda)T_0), t < T_0 < \infty) \qquad t \ge 0, \lambda > \eta.$$

From this, we get, for any $t \ge 0$ and $\lambda > \eta$,

$$\begin{aligned}
\mathbb{E}_x(\exp(-\lambda\tilde{X}_t + (t\wedge T_0)F(\lambda))) &= \mathbb{E}_x(\exp(-\lambda X_t + tF(\lambda)), t < T_0) + \mathbb{E}_x(\exp(F(\lambda)T_0), T_0 \le t)\\
&= \mathbb{E}_x(\exp(-\lambda X_t + tF(\lambda)), t < \tau) + \mathbb{E}_x(\exp(F(\lambda)T_0), T_0 \le t)\\
&= \mathbb{E}_x(\exp(F(\lambda)T_0), t < T_0 < \infty) + \mathbb{E}_x(\exp(F(\lambda)T_0), T_0 \le t)\\
&= \mathbb{E}_x(\exp(F(\lambda)T_0), T_0 < \infty)\\
&= e^{-\lambda x},
\end{aligned}$$

which proves that $(\exp(-\lambda\tilde{X}_t + (t\wedge T_0)F(\lambda)); t\ge 0)$ is a martingale. This proves that $\tilde{X}$ is a Lévy process with Laplace exponent $-F$ *stopped* when it hits 0, so that X is a Lévy process with Laplace exponent $-F$ *killed* when it hits 0. Again, this yields (2.3) with $F = -\psi$. As a conclusion, we get (2.4), thanks to (2.9) and the fact that $u_t(\lambda) \to \lambda$ as $t \downarrow 0$.

It only remains to show that (2.4) has a unique solution, which stems from the fact that when $\psi'(0+) < \infty$, ψ is Lipschitz on compact subsets of $[0,\infty)$. In the case when $\psi'(0+) = \infty$, the proof is a little bit more technical and can be found in [96]. □

Main properties

The Laplace exponent ψ of X is also called *branching mechanism* of Z, which in turn is sometimes denoted CB(ψ). The branching mechanism can be specified [11, 17, 78, 96] by the Lévy–Khinchin formula (5.1) displayed in the Appendix. Recall that ψ is a convex function such that $\psi(0) = 0$ and

$$\rho := \psi'(0+) \in [-\infty, +\infty).$$

From now on, we also discard the case when X is a subordinator, which corresponds to nondecreasing paths of both X and Z. This is consistent with the assumption made in the discrete setting that $p_0 \neq 0$ (discrete time), or $d \neq 0$ (continuous time).

Recall that η is the largest root of ψ.

Lemma 2.2.9 *For $\lambda < \eta$ (resp. $> \eta$, $= \eta$), $t \mapsto u_t(\lambda)$ increases (resp. decreases, remains constant equal) to η. The following equation gives an implicit characterization of $u_t(\lambda)$*

$$\int_{u_t(\lambda)}^{\lambda} \frac{ds}{\psi(s)} = t \qquad t, \lambda \geq 0. \tag{2.11}$$

Proof. For clarity, write $y(t)$ instead of $u_t(\lambda)$ and recall from Theorem 2.2.8 that $y' = -\psi(y)$, with $y(0) = \lambda$. Since $\psi(\eta) = 0$ and $\psi'(\eta) \geq 0$, η is globally attractive for y (recall that if $\eta \neq 0$, then $\psi'(0) < 0$). As a consequence, integrating the differential equation as in the theorem is possible because $y(t)$ and $y(0)$, i.e. λ and $u_t(\lambda)$, are always in the same connected component of $[0, \infty)\backslash\{\eta\}$. Specifically, if we set

$$G_\lambda(v) = \int_\lambda^v \frac{ds}{\psi(s)},$$

then $y' = -\psi(y)$, with boundary condition $y(0) = \lambda$, integrates as $G_\lambda \circ y(t) = -t$. □

Corollary 2.2.10 *For any $x \geq 0$, $\mathbb{P}_x(\mathrm{Ext}) = \exp(-x\eta)$.*

Proof. It is easy to deduce from Theorem 2.2.8 that either Z_t goes to infinity (including blow-up) or it goes to 0. Since for any positive λ,

$$\lim_{t\to\infty} \mathbb{E}_x(\exp -\lambda Z_t) = \mathbb{P}_x(\lim_{t\to\infty} Z_t = 0),$$

and $u_\infty(\lambda) = \eta$, the result follows. □

Theorem 2.2.11 (Grey [47]) *The CB-process Z blows up with positive probability iff $\rho = -\infty$ and*

$$\int_0 \frac{ds}{\psi(s)} > -\infty.$$

Absorption at 0 occurs with positive probability iff

$$\int^\infty \frac{ds}{\psi(s)} < \infty.$$

Proof. Thanks to Theorem 2.2.8, it is elementary to show that if T_∞ denotes the blow-up time,

$$\mathbb{P}_x(T_\infty > t) = \exp(-xu_t(0)).$$

If blow-up occurs with positive probability, then $u_t(0) > 0$ for all $t > 0$ and thanks to (2.11),

$$\int_{u_t(0)}^{0+} \frac{ds}{\psi(s)} = t,$$

which proves that $\int_0 ds/\psi(s) > -\infty$ (and actually implies that $\rho = -\infty$). Conversely, assume that this integral converges. Then we can take limits in (2.11) as $\lambda \downarrow 0$, and the last displayed equation holds, showing that $u_t(0)$ is positive as soon as t is.

For the absorption time, again thanks to Theorem 2.2.8,

$$\mathbb{P}_x(T < t) = \exp(-xu_t(\infty)),$$

so that absorption occurs with positive probability iff $u_t(\infty) > 0$. The rest of the proof is identical as that for the blow-up time. $\square$

Corollary 2.2.12 *If $\rho > -\infty$, then Z has integrable marginals and*

$$\mathbb{E}_x(Z_t) = x\exp(-\rho t) \qquad t \geq 0.$$

Definition 2.2.13 *A CB-process is said* subcritical *if $\rho > 0$,* critical *if $\rho = 0$, and* supercritical *if $\rho < 0$.*

Proof of the corollary. Set

$$f(t) := \frac{\partial u_t(\lambda)}{\partial \lambda} \,|_{\lambda=0} .$$

Since there is no blow-up (ρ is finite), $u_t(0) = 0$, and differentiating $\mathbb{E}_x(\exp -\lambda Z_t) = \exp(-xu_t(\lambda))$, we get $\mathbb{E}_x(Z_t) = xf(t)$. Then differentiating (2.4) yields

$$f(t) + \rho \int_0^t f(s)\,ds = 1,$$

which implies that $f(t) = \exp(-\rho t)$. $\square$

Corollary 2.2.14 *Assume $\int^\infty 1/\psi$ converges, and put*

$$\phi(t) := \int_t^\infty \frac{ds}{\psi(s)} \qquad t > \eta.$$

The mapping $\phi : (\eta, \infty) \to (0, \infty)$ is bijective decreasing, and we write φ for its inverse mapping. Then

$$u_t(\lambda) = \varphi(t + \phi(\lambda)) \qquad \lambda > \eta,$$

and furthermore,

$$\mathbb{P}_x(T < t) = \exp(-x\varphi(t)).$$

In particular $\mathbb{P}_x(\mathrm{Abs}) = \exp(-\eta x) = \mathbb{P}_x(\mathrm{Ext})$.

Proof. Because $\int^\infty ds/\psi(s)$ converges, (2.11) entails

$$\int_{u_t(\lambda)}^\infty \frac{ds}{\psi(s)} - \int_\lambda^\infty \frac{ds}{\psi(s)} = t,$$

which reads $\phi(u_t(\lambda)) - \phi(\lambda) = t$, or equivalently $u_t(\lambda) = \varphi(t + \phi(\lambda))$. Now from the proof of the previous theorem,

$$\mathbb{P}_x(T < t) = \exp(-xu_t(\infty)),$$

and because ϕ vanishes at $+\infty$, we get $u_t(\infty) = \varphi(t)$. The last line of the statement comes from the fact that $\mathbb{P}_x(\text{Abs}) = \mathbb{P}_x(T < \infty)$ and Corollary 2.2.10. □

Remark 2.2 *When $\int^\infty 1/\psi$ converges, the sample-paths of Z have* infinite variation *a.s. (see Appendix). When $\int^\infty 1/\psi$ diverges, absorption is impossible but with positive probability* $\lim_{t\to\infty} Z_t = 0$.

Proposition 2.2.15 *Assume $\eta > 0$. Then the supercritical CB-process with branching mechanism ψ conditioned on* $\{\text{Ext}\}$*, is the subcritical CB-process with branching mechanism $\psi^\natural$, where $\psi^\natural(\lambda) = \psi(\lambda + \eta)$.*

Exercise 2.1 *Prove the previous statement.*

Extensions

Here, we give the definitions of the CB-process with *immigration*, and CB-process with *logistic growth*.

Immigration. Recall from the previous chapter that for discrete branching processes with immigration, the total number of immigrants up until time t is a compound Poisson process with intensity measure ν only charging nonnegative integers, that is, a *renewal process*. In the continuous setting, this role is played by a *subordinator*, which is characterized by its Laplace exponent, denoted by χ. Then the continuous-state branching process with immigration, denoted CBI(ψ, χ), is a strong Markov process characterized by its Laplace transform

$$\mathbb{E}_x(\exp -\lambda Z_t) = \exp\left(-xu_t(\lambda) - \int_0^t \chi(u_s(\lambda))ds\right) \qquad \lambda \geq 0, \tag{2.12}$$

where $u_t(\lambda)$ is given by (2.11). The last equation is the analogue of (1.3) and (2.1).

As a consequence, a CBI(ψ, χ) has infinitesimal generator B whose action on the exponential functions $x \mapsto e_\lambda(x) = \exp(-\lambda x)$ is given by

$$Be_\lambda(x) = (x\psi(\lambda) - \chi(\lambda))e_\lambda(x) \qquad x \geq 0.$$

The seminal papers on CBI-processes are [62, 86].

Logistic growth. For simplicity, we can say that the logistic branching process in continuous-state space [74], or *LB-process*, is the Markov process with generator U given by

$$Uf(x) = xAf(x) - cx^2 f'(x) \qquad x \geq 0,$$

where $c \geq 0$ and A is the generator of a Lévy process X with no negative jumps. When $c = 0$, note that we recover the generator of a standard CB-process.

Actually, we prefer not to define a process from its generator, and we can rigorously define Z as

$$R_t = Z\left(\int_0^t ds/R_s\right) \qquad t < T_0,$$

where R is the *Ornstein–Uhlenbeck type process* strong solution to

$$dR_t = dX_t - cR_t dt \qquad t > 0.$$

In the discrete setting, the process R corresponds to the Markov chain with rates

$$\begin{cases} n & \to & n+k & \text{at rate} & \pi_k \\ n & \to & n-1 & \text{at rate} & d + c(n-1), \end{cases}$$

that we would then time-change (accelerate) to multiply the rates by n and get the logistic branching process.

Since the properties of the LB-process reviewed hereafter have rather technical proofs, we merely state them and refer the reader to [74] for details. From now on, we assume that $\mathbb{E}(\log(X_1)) < \infty$, which is equivalent to $\int^\infty \log(r)\Lambda(dr) < \infty$.

In the first statement, we consider the case when X is a subordinator. We then denote by $\delta \geq 0$ its drift coefficient, so that (see Appendix)

$$\psi(\lambda) = -\delta\lambda - \int_0^\infty \Lambda(dr)(1 - e^{-\lambda r}) \qquad \lambda \geq 0.$$

We introduce Condition (∂), where ρ is defined as

$$\rho := \int_0^\infty \Lambda(dr) \leq \infty.$$

We say that (∂) holds iff (at least) one of the following holds

- $\delta \neq 0$
- $\rho = \infty$
- $c < \rho < \infty$.

Theorem 2.2.16 ([74]) *Assume X is a subordinator. Then the LB-process Z oscillates in $(\delta/c, \infty)$ and*

(i) If (∂) holds, then it is positive-recurrent in $(\delta/c, \infty)$.

(ii) If (∂) does not hold, then it is null-recurrent in $(0, \infty)$ and converges to 0 in probability.

From now on, X is assumed not to be a subordinator. In the next theorem, note that the criterion for absorption does not depend on c and is the same as for the CB-process ($c = 0$).

Theorem 2.2.17 ([74]) *Assume X is not a subordinator. Then the LB-process goes to 0 a.s., and if T denotes the absorption time, then $\mathbb{P}(T < \infty) = 1$ or 0 according to whether $\int^{\infty} 1/\psi$ converges or diverges.*

The next statement assures that the LB-process comes down from infinity (still under the integrability condition given earlier).

Theorem 2.2.18 ([74]) *The probabilities $(\mathbb{P}_x, x \geq 0)$ converge weakly, as $x \to \infty$, to the law $\mathbb{P}_\infty$ of the so-called* logistic branching process starting from infinity. *In addition, if $\int^{\infty} 1/\psi$ converges, then under $\mathbb{P}_\infty$ the absorption time T is a.s. finite and has finite expectation.*

2.2.4 A relation between GFV-processes and CB-processes

If $(Z_t(x); t \geq 0)$ stands for a CB-process starting from x, the branching property entails that for all t

$$Z_t(x + y) = Z_t(x) + \tilde{Z}_t(y),$$

where $\tilde{Z}_t(y)$ is an independent copy of $Z_t(y)$. Then Kolmogorov's existence theorem ensures that one can build on a same probability space a doubly indexed process $(Z_t(x); t, x \geq 0)$ such that, for each fixed time t, $Z_t(\cdot)$ is a *subordinator*.

Thanks to Theorem 2.2.8, the subordinator $x \mapsto Z_t(x)$ has Laplace exponent $u_t(\cdot)$ which satisfies the semigroup property $u_t = u_{t-s} \circ u_s$. This is due to the following relationship in terms of Bochner subordination

$$Z_t(x) = \tilde{Z}_{t-s} \circ Z_s(x),$$

where $\tilde{Z}_{t-s}$ is an independent copy of Z_{t-s}. Thanks to this last equation, we can invoke once again Kolmogorov's existence theorem and show that on a same probability space there is a *flow of subordinators* $S_{s,t}$ satisfying

(i) for any $s < t < u$,

$$S_{t,u} \circ S_{s,t} = S_{s,u}$$

(ii) the subordinator $S_{s,t}$ has Laplace exponent $u_{t-s}(\cdot)$ and for any $t_1 \leq \cdots \leq t_k$, the subordinators $S_{t_1,t_2}, \ldots, S_{t_{k-1},t_k}$ are independent.

In particular, $(S_{0,t}(x); t, x \geq 0)$ has the same law as the CB-process $(Z_t(x); t, x \geq 0)$. This representation in terms of subordinators originated in [13]. It may seem a little bit complex, but can be understood more easily having a quick look at the subsection on Jirina processes in the beginning of the present chapter, or adopting the viewpoint of splitting trees developed in Chapter 4.

This representation is extremely similar to that given in Definition 2.2.3, in terms of *flows of bridges*, for exchangeable genealogies of continuous populations in continuous time with *fixed population size*. Indeed, recall from Subsection 2.2.2 that a flow of bridges can be constructed thanks to a Poisson point process with intensity $dt \otimes du \otimes \nu(dr)$ on $[0, \infty] \times (0, 1) \times (0, 1]$, where ν is a positive measure on $(0, 1]$ such that $\int_0^1 x^2 \nu(dx) < \infty$. To each atom (t, u, r) is associated a simple bridge with one single jump of size r occurring at u. Composing these bridges gives rise to a flow of bridges $(B_{s,t}\,; 0 \le s \le t)$ called a ν-simple flow.

Then the generalized Fleming–Viot process $(Y_t(x)\,; t, x) := (B_{0,t}(x)\,; t, x)$ is to be interpreted as the relative size at time t of an initial subpopulation of size x. It is thus tempting to establish a link between the GFV-process Y and the CB-process Z via a relation between bridges and subordinators. The idea is that the growth of a very small subpopulation $Y_t(\varepsilon x)$ is blind to the constraint of constant population size, and so must resemble a CB-process with branching mechanism involving the measure ν.

Specifically, for $\varepsilon > 0$, let ν^ε be a positive measure on $(0, 1]$ such that $\int_0^1 x^2 \nu^\varepsilon(dx) < \infty$. Then let Y_t be the GFV-process associated to ν^ε and *starting from* εx, and set

$$Y_t^\varepsilon := \varepsilon^{-1} Y_{t/\varepsilon} \qquad t \ge 0,$$

so that in particular Y^ε starts from x.

Theorem 2.2.19 (Bertoin–Le Gall [16]) *Let $\tilde\nu^\varepsilon$ denote the image of ν^ε under the dilation $r \mapsto r/\varepsilon$. If the measures $(r^2 \wedge r)\tilde\nu^\varepsilon(dr)$ converge weakly as $\varepsilon \downarrow 0$ to a finite measure on $(0, \infty)$, which we may write in the form $(r^2 \wedge r)\pi(dr)$, then the rescaled GFV-process Y_t^ε converges in distribution to the CB-process Z starting from x, with Laplace exponent ψ given by*

$$\psi(\lambda) = \int_0^\infty (e^{-\lambda r} - 1 + \lambda r)\pi(dr) \qquad \lambda \ge 0.$$

In particular, Z is a critical *CB-process.*

2.3 Diffusions

In this section, we study the processes defined in the last section in the special case when their sample paths are continuous a.s. The process under focus will be the *Fisher–Wright diffusion* in the case of fixed size populations, and the *Feller diffusion* in the branching case.

2.3.1 Fisher–Wright diffusion

Definition

A slightly different definition of the Moran model from that given in the previous chapter is as follows. In a population of constant size $2N$, each individual has a constant birth rate equal to N (instead of 1 in the earlier definition), and at each birth event, an individual is simultaneously killed, who is chosen uniformly among the old ones. Then the Markov chain $Y_t^{(N)}$ counting the *proportion*, at time t, of descendants from some fixed initial subpopulation, has transition rates

$$\begin{cases} y \;\to\; y + 1/2N & \text{at rate} \quad 2N^2 y(1-y) \\ y \;\to\; y - 1/2N & \text{at rate} \quad 2N^2 y(1-y). \end{cases}$$

The process $Y^{(N)}$ takes its values in $\{0, 1/2N, \dots, (2N-1)/2N, 1\}$ and its generator L_N is given by

$$L_N f(x) = 2N^2 x(1-x)\left(f\big(x + \frac{1}{2N}\big) - f(x)\right) + 2N^2 x(1-x)\left(f\big(x - \frac{1}{2N}\big) - f(x)\right).$$

As a consequence, if f is of class C^2 on $[0,1]$, the last quantity converges

$$\lim_{N\to\infty} L_N f(x) = \frac{1}{2}x(1-x)f''(x)x \in [0,1].$$

Theorem 2.3.1 *The sequence of Markov processes* $(Y^{(N)})$ *converges weakly on the Skorokhod space of càdlàg processes with values in* $[0,1]$ *to the diffusion* Y*, strong solution to the following SDE*

$$dY_t = \sqrt{Y_t(1-Y_t)}dB_t \qquad t \geq 0,$$

where B *is the standard Brownian motion. This diffusion is the so-called* Fisher–Wright diffusion*.*

In this document, we do not want to spend time proving convergence theorems, but a rigorous proof of the previous statement can be found in [18]. In passing, this proof uses a well-known *duality relationship*, that has very interesting consequences, and is worth being stated as follows.

Theorem 2.3.2 *The* n*-th moment of the Fisher–Wright diffusion is given by*

$$\mathbb{E}_y(Y_t^n) = \mathbb{E}_n(y^{N_t}),$$

where N_t *is a pure death process with transition rate from* k *to* $k-1$ *equal to* $k(k-1)/2$*.*

The pure death process mentioned in the theorem is the number of blocks in the Kingman coalescent. Since it comes down from infinity (check this thanks to Theorem 1.2.10), the Lebesgue dominated convergence theorem yields

$$\mathbb{P}_y(T_1 < t) = \mathbb{E}_\infty(y^{N_t}),$$

where T_1 is the total coalescence time (first hitting time of 1). Note that a generalization of this duality relationship to Λ-coalescents was obtained recently by Bertoin and Le Gall [15], as was mentioned in Subsection 2.2.2, equation (2.2).

Approximations

Note that we might like to speed up time otherwise than at a rate exactly equal to the population size, so that for example, each individual has a birth rate of $N\sigma$. Then of course this new scaling will result in a generator equal to σ times the one we have obtained previously, that is, $Lf(x) = (\sigma/2)x(1-x)f''(x)$.

Actually, if we come back to the initial model where each individual gives birth at rate 1, we can approximate the dynamics of the proportion of descendants from a fixed initial subpopulation in a total population of constant size $2N$, as N gets large, as the Fisher–Wright diffusion slown down at rate $\sigma = 1/N$. This approximation is very convenient, since it allows to use the corresponding generator, say A_N, given by

$$A_N f(x) = \frac{1}{2N}x(1-x)f''(x) \qquad x \in [0,1],$$

and the associated Kolmogorov equations to derive such quantities as the expected time to absorption. Indeed, set

$$f_N(x) = \mathbb{E}_x(\tau, \text{Fix}),$$

where τ is the absorption time, that is, the first hitting time of the boundary $\{0, 1\}$ by the diffusion with generator A_N. Now, recall that in Subsection 2.2.1, we gave a few applications of generators, in particular thanks to Theorem 2.2.1, we have $A_N f_N(x) = -u_N(x)$, where $u_N(x) = x$ is the fixation probability and $f_N(0) = f_N(1) = 0$. It takes a straightforward calculation to deduce that

$$\mathbb{E}_x(\tau, \mathrm{Fix}) \approx -2N(1-x)\log(1-x),$$

whereas

$$\mathbb{E}_x(\tau) \approx -2N(1-x)\log(1-x) - 2Nx\log(x).$$

Since absorption is certain, the last quantity, say g_N, was computed thanks to $A_N g_N = -1$. Note that we have recovered very quickly the results obtained fastidiously for the Moran model in Subsection 1.2.2.

Selection (and mutation)

A natural way of modifying this model is to add *selection*. Specifically, we can assume that the individuals of the *subpopulation* we are following through time are of a special type w.r.t. the rest of the population, that confers them an increased (*positive* selection) or decreased (*negative* selection) birth rate, say $N\sigma_N(1+r_N)$. Then $Y^{(N)}$ has the following transition rates

$$\begin{cases} y \;\rightarrow\; y + 1/2N & \text{at rate} \quad 2N^2\sigma_N(1+r_N)y(1-y) \\ y \;\rightarrow\; y - 1/2N & \text{at rate} \quad 2N^2\sigma_N y(1-y). \end{cases}$$

The new generator L_N of $Y^{(N)}$ is given by

$$L_N f(x) = 2N^2\sigma_N(1+r_N)x(1-x)\left(f\big(x+\frac{1}{2N}\big) - f(x)\right) + 2N^2\sigma_N x(1-x)\left(f\big(x-\frac{1}{2N}\big) - f(x)\right).$$

Assuming again that f is of class C^2, we see that

$$L_N f(x) = 2N^2\sigma_N x(1-x)\left(\frac{r_N}{2N}f'(x) + \frac{2+r_N}{8N^2}f''(x) + o(N^{-2})\right).$$

Now there are three possibilities

- if $r_N = O(1)$, then the diffusive motion, called *genetic drift*, vanishes in the limit compared to the action of selection and the correct timescaling is given by $\sigma_N = 1/2N$. The limiting motion is driven by the ordinary differential equation $\dot{y} = ry(1-y)$ (whether fixation or extinction occurs depends solely on the sign of r)
- if $r_N = O(1/N)$, we can keep the initial timescaling by setting $\sigma_N = \sigma$, and then with $r := \lim_N \sigma N r_N$, the limiting generator is

$$Lf(x) = rx(1-x) + \frac{\sigma}{2}x(1-x)f''(x) \qquad x \in [0, 1].$$

- $r_N = o(1/N)$, then selection has negligible effect compared to genetic drift, and we are back to the situation of the last theorem.

As a conclusion, we give

Definition 2.3.3 *The solution to the following SDE*

$$dY_t = rY_t(1 - Y_t)dt + \sqrt{\sigma Y_t(1 - Y_t)}dB_t$$

is called Fisher–Wright diffusion with selection*.*

Remark 2.3 *Most species are diploid. An individual has two copies of each gene, and her phenotype depends on the type (allele) of each of these copies and their interaction. So the growth rate of a subpopulation of individuals bearing a certain allele, say A, will not depend solely on the frequencies of each allele $A, B, C, \ldots$ in the population, but also on how pairs of alleles interact to confer a certain fitness (propensity to propagate one's genes, i.e. reproduce) to their bearer. A classical way of accounting for diploidy is to replace the drift function $a(y) = ry(1 - y)$ with $a(y) = ry(1 - y)(y + h(1 - 2y))$, where h is the so-called* dominance coefficient*. This drift function pops up assuming that, conditional on being bearer of the allele A, a newborn*

1. *is* homozygote *AA with probability y and* heterozygote *$AB, AC, \ldots$ with probability $1 - y$*
2. *has marginal birth rate, or* fitness*, $r(1 - h)$ if she is homozygote, and rh if she is heterozygote (the fitness of all others is 0 by definition).*

Thinking of the deterministic scaling limit ($\sigma = 0$), the following terminology appears natural

- *when $h \in [0, 1]$, selection is said* directional *(no deterministic equilibrium). If $h = 0$ or 1, dominance is* complete *(a heterozygote has the same fitness as a homozygote); if $h = 1/2$, dominance is absent; if $h \in (0, 1)$, dominance is* incomplete
- *when $h < 0$, selection is said* disruptive *(one deterministic unstable equilibrium). One speaks of* underdominance *(the heterozygote is less fit than the homozygote and all others)*
- *when $h > 1$, selection is said* stabilizing *(one deterministic stable equilibrium). One speaks of* overdominance *(the heterozygote is fitter than the homozygote and all others).*

Remark 2.4 *Even if we will not treat that subject in detail, we want to point out that it is standard to assume that mutations from the focal type to other types (and possibly conversely) occur with a certain probability $\theta/2N$ at each birth event. This adds to the drift function $a(y)$ an extra term equal to*

$$\theta_0(1 - y) - \theta_1 y,$$

where θ_0 (resp. θ_1) is the (rescaled) mutation rate towards (resp. from) the focal type from (resp. towards) any other type.

If both θ_0 and θ_1 are nonzero, the fixation probability is zero. On the other hand, the same tricks can be used to compute the stationary distribution*, say $\pi(x)$. In particular $L\pi(x) = 0$.*

Conditioning

Recall again Subsection 2.2.1 and apply Theorem 2.2.2 to the Fisher–Wright diffusion Y (with or without selection) with generator L. Then the Fisher–Wright diffusion *conditioned* on fixation has generator $L^\star$ given by

$$L^\star f(x) = \frac{L(uf)(x)}{u(x)} \qquad x \in [0,1],$$

where $u(x) := \mathbb{P}_x(\text{Fix})$. Recall that $Lu(x) = 0$, so it is easy to get, in the selection case ($r \neq 0$)

$$u(x) = \frac{1-\exp(-2rx/\sigma)}{1-\exp(-2r/\sigma)} \qquad x \in [0,1],$$

while as usual $u(x) = x$ in the neutral case ($r = 0$).

Exercise 2.2 *Prove the following statement.*

Proposition 2.3.4 *The Fisher–Wright diffusion with selection ($r \neq 0$) conditioned on fixation satisfies the SDE*

$$dY_t = rY_t(1-Y_t)\coth\left(\frac{rY_t}{\sigma}\right)dt + \sqrt{\sigma Y_t(1-Y_t)}dB_t,$$

which becomes in the absence of selection ($r = 0$)

$$dY_t = \sigma(1-Y_t)dt + \sqrt{\sigma Y_t(1-Y_t)}dB_t.$$

2.3.2 CB-diffusions

Observe that if a CB-process Z has continuous paths, then by Lamperti's time-change, it is also the case of the associated Lévy process, so that the branching mechanism must be of the form $\psi(\lambda) = \sigma\lambda^2/2 - r\lambda$. Using again Lamperti's time-change $\theta_t := \int_0^t Z_s ds$ and its right-inverse κ, $X := Z \circ \kappa$ is a (killed) Lévy process with continuous paths, namely $X_t = \sqrt{\sigma}\beta_t + rt$, where β is a standard Brownian motion. As a consequence, $X(\theta_t) - r\theta_t$ is a local martingale with increasing process $\sigma\theta_t$, or equivalently, $Z_t - r\int_0^t Z_s ds$ is a local martingale with increasing process $\sigma\int_0^t Z_s ds$. This entails

Definition 2.3.5 *The CB-diffusion with branching mechanism $\psi(\lambda) = \sigma\lambda^2/2 - r\lambda$ satisfies*

$$dZ_t = rZ_t dt + \sqrt{\sigma Z_t}dB_t \qquad t > 0,$$

where B is a standard Brownian motion. Such a diffusion is generally called Feller diffusion *(denomination sometimes exclusive to $r = 0$), and when $r = 0$ and $\sigma = 4$,* squared Bessel process with dimension 0.

Since $\int^\infty 1/\psi$ converges, we can define ϕ, and by elementary calculus, check that if $r = 0$, then for any $t > 0$,

$$\phi(t) = \varphi(t) = 2/\sigma t,$$

so that

$$u_t(\lambda) = \frac{\lambda}{1 + \sigma\lambda t/2} ,$$

whereas if $r \neq 0$,

$$\phi(t) = -r^{-1} \log(1 - 2r/\sigma t) \quad \text{and} \quad \varphi(t) = (2r/\sigma)e^{rt}/(e^{rt} - 1),$$

so that

$$u_t(\lambda) = \frac{2re^{rt}/\sigma}{e^{rt} - 1 + 2/\sigma\lambda} .$$

Note that $\rho = \psi'(0+) = -r$. A CB-diffusion is subcritical if $r < 0$, critical if $r = 0$, and supercritical if $r > 0$.

In the supercritical case, the probability of extinction in t units of time is

$$\mathbb{P}_x(T < t) = \exp - \left(\frac{2rx/\sigma}{1 - e^{-rt}}\right) ,$$

so that in particular

$$\mathbb{P}_x(\text{Ext}) = \exp -(2rx/\sigma).$$

Last but not least, a supercritical CB-diffusion with parameters (r, σ) conditioned on its ultimate extinction is distributed as a subcritical CB-diffusion with parameters $(-r, \sigma)$. This property can be derived either from Theorem 2.2.2 or from Proposition 2.2.15.

2.3.3 A relation between Fisher–Wright and Feller diffusions

We want to display the same kind of relation that was shown in the first chapter, namely, that a BGW model (with Poisson offspring) conditioned to have constant size, has the same law as the Wright–Fisher model. In contrast with the discrete case, a CB-process does not provide us with a genealogy, but only with a random population size. As a consequence, we will not display a relationship between microscopic genealogies, but only between genealogies associated to macroscopic subpopulations. Specifically, we have

Theorem 2.3.6 *Let $Z^{(1)}, Z^{(2)}$ be two independent CB-diffusions with parameters (r_1, σ_1) and (r_2, σ_2). Then conditional on $Z^{(1)} + Z^{(2)} = z$ at all times, the frequency $Y := Z^{(1)}/(Z^{(1)} + Z^{(2)})$ is a diffusion on $[0, 1]$ satisfying the following SDE*

$$dY_t = \bar{s}\, Y_t(1 - Y_t)\, dt + \sqrt{Y_t(1 - Y_t)}\sqrt{\frac{\sigma_2}{z} Y_t + \frac{\sigma_1}{z}(1 - Y_t)}\, dB_t$$

where the resulting selection coefficient $\bar{s}$ *equals*

$$\bar{s} = r_1 - r_2 + \frac{1}{z}(\sigma_2 - \sigma_1).$$

In particular if $\sigma_1 = \sigma_2 =: \sigma$,

$$dY_t = (r_1 - r_2)\, Y_t(1 - Y_t)\, dt + \sqrt{\frac{\sigma}{z} Y_t(1 - Y_t)}\, dB_t.$$

The last displayed equation in the theorem is a Fisher–Wright diffusion with selection coefficient $r_1 - r_2$. If one compares this diffusion with that given in the subsection on large population approximations in the last section, one can identify the quantity σ/z with the quantity N, which is usually termed in population genetics the *effective population size* (as far as *stochasticity* is concerned), that is, the constant size that a population would have if its demographic stochasticity was to be compared to that of a Wright–Fisher population. The theorem gives a more rigorous interpretation to this effective population size, as the ratio of the *offspring variance* σ to the *census size* z (the real size). For applications to the last theorem, see e.g. [75].

Exercise 2.3 *Prove the last theorem applying Itô's formula to the bivariate mapping* $(x, y) \mapsto x/(x + y)$*, and a classical representation of Brownian martingales [90, Proposition V.3.8].*

A somewhat reverse relation between CB-diffusions and Fisher–Wright diffusions can be derived from the ideas developed in Section 2.2.2 on stochastic flows of bridges. Indeed, if one looks at the growth of a very small subpopulation of a Fisher–Wright diffusion model, it is likely that this growth will be blind to the constraint of constant population size, so that it will resemble that of a Feller diffusion.

Proposition 2.3.7 *Let* Y *be a* Fisher–Wright diffusion *solving*

$$dY_t = \sqrt{Y_t(1 - Y_t)}\, dB_t$$

and starting from εx*. Rescale this subpopulation as* $Z_t^\varepsilon := \varepsilon^{-1} Y_{\varepsilon t}$*. Then* Z^ε *converges in distribution to the* Feller diffusion

$$dZ_t = \sqrt{Z_t}\, dB_t$$

starting from x*.*

Chapter 3

Quasi–stationary distributions and the Q-process

3.1 What is quasi–stationarity ?

Let X be a Markov chain on the integers or a Markov process on $[0, \infty)$ for which 0 is one (and the only) *accessible absorbing state*. Then the only *stationary* probability is the Dirac mass at 0. A *quasi-stationary distribution* (QSD) is a probability measure ν satisfying

$$\mathbb{P}_\nu(X_t \in A \mid X_t \neq 0) = \nu(A) \qquad t \geq 0. \tag{3.1}$$

More generally, a QSD is a subinvariant distribution for a killed Markov process. Hereafter, we will only consider finite QSDs. Such a quasi-stationary distribution may not be unique, but a specific one is defined (if it exists) as the law of Υ, where

$$\mathbb{P}(\Upsilon \in A) := \lim_{t\to\infty} \mathbb{P}_x(X_t \in A \mid X_t \neq 0),$$

for any Dirac initial condition $x \neq 0$. The r.v. Υ is sometimes called the *Yaglom limit*, in reference to the proof of this result for BGW processes, erroneously attributed to A.M. Yaglom[1]. Then by application of the simple Markov property,

$$\mathbb{P}_\nu(T > t + s) = \mathbb{P}_\nu(T > s)\mathbb{P}_\nu(T > t),$$

so that the extinction time T under $\mathbb{P}_\nu$ has a *geometric* distribution in discrete-time models and an *exponential* distribution in continuous-time models.

Other asymptotic conditional distributions include

$$\lim_{t\to\infty} \mathbb{P}_x(X_t \in A \mid X_{t+s} \neq 0),$$

but we shall not consider such conditionings here, and rather focus on

$$\mathbb{P}^\uparrow_x(\Theta) := \lim_{s\to\infty} \mathbb{P}_x(\Theta \mid X_{t+s} \neq 0),$$

defined, if it exists, for any $\Theta \in \mathcal{F}_t$. Thus resulting law $\mathbb{P}^\uparrow$ is that of a (possibly dishonest) Markov process $X^\uparrow$, that we call the *Q-process*.

A natural question is to compare the stationary distributions of $X^\uparrow$ (if any) with the quasi-stationary distribution Υ, that is, compare the asymptotic distribution of X_t *conditioned on not yet being absorbed*, with the asymptotic distribution of X_t *conditioned on not being absorbed in the distant future*. Intuitively, the second conditioning is more stringent than the first one, and should thus charge more heavily the paths that stay away from 0, than the first conditioning. The following statement gives a mathematical formulation of this intuition, in terms of stochastic domination.

Proposition 3.1.1 *If for any $t > 0$ the mapping $x \mapsto \mathbb{P}_x(X_t \neq 0)$ is nondecreasing, then for any starting point $x \neq 0$, and for any $t, s > 0$,*

$$\mathbb{P}_x(X_t > a \mid X_{t+s} \neq 0) \geq \mathbb{P}_x(X_t > a \mid X_t \neq 0) \qquad a > 0.$$

[1] according to Pr V.A. Vatutin (private communication), this result is due to B.A. Sevast'yanov. The result of A.M. Yaglom concerns the convergence of the rescaled conditional distribution to the exponential distribution in the critical case.

Then, if there exists a Q-process $X^\uparrow$, by letting $s \to \infty$,

$$\mathbb{P}_x(X_t^\uparrow > a) \geq \mathbb{P}_x(X_t > a \mid X_t \neq 0) \qquad a > 0.$$

If there is a Yaglom limit Υ, and in addition the Q-process converges in distribution to a r.v. $X_\infty^\uparrow$, then by letting $t \to \infty$,

$$X_\infty^\uparrow \overset{stoch}{\geq} \Upsilon.$$

Remark 3.1 *By a standard coupling argument, the monotonicity condition for the probabilities $x \mapsto \mathbb{P}_x(X_t \neq 0)$ is satisfied for any strong Markov process with no negative jumps.*

Proof. It takes a standard application of Bayes' theorem to get that for any $a, x, s, t > 0$,

$$\begin{array}{rcl} \mathbb{P}_x(X_t > a \mid X_{t+s} \neq 0) & \geq & \mathbb{P}_x(X_t > a \mid X_t \neq 0) \\ & \Updownarrow & \\ \mathbb{P}_x(X_{t+s} \neq 0 \mid X_t > a) & \geq & \mathbb{P}_x(X_{t+s} \neq 0 \mid X_t \neq 0), \end{array}$$

so it remains to prove the last displayed inequality. Next for any $m \geq 0$, set W_m the r.v. defined as

$$\mathbb{P}(W_m \in dr) = \mathbb{P}_x(X_t \in dr \mid X_t > m) \qquad r > 0.$$

For any $m \leq m'$, and $u \geq 0$, check that

$$\mathbb{P}(W_{m'} > u) \geq \mathbb{P}(W_m > u),$$

which means $W_{m'} \overset{stoch}{\geq} W_m$, so in particular $W_a \overset{stoch}{\geq} W_0$. Finally, observe that

$$\mathbb{P}_x(X_{t+s} \neq 0 \mid X_t > a) \geq \mathbb{P}_x(X_{t+s} \neq 0 \mid X_t \neq 0) \Leftrightarrow \mathbb{E}(f(W_a)) \geq \mathbb{E}(f(W_0)),$$

where $f(x) := \mathbb{P}_x(X_s \neq 0)$. Since f is nondecreasing, the proof is complete. □

3.2 Markov chains with finite state-space

3.2.1 Perron–Frobenius theory

Let X be a Markov chain on $\{0, 1, \dots, 2N\}$ that has two communication classes, namely $\{0\}$ and $\{1, \dots, 2N\}$. We assume further that 0 is absorbing and accessible. Next, let P be the transition matrix, that is, the matrix with generic element $p_{ij} := \mathbb{P}_i(X_1 = j)$ (row i and column j), and let Q be the square matrix of order $2N$ obtained from P by deleting its first row and its first column. In particular,

$$\mathbb{P}_i(X_n = j) = q_{ij}(n) \qquad i, j \geq 1,$$

where $q_{ij}(n)$ is the generic element of the matrix Q^n (row i and column j).

Everything that follows still holds if there are two absorbing states, $\{0\}$ and $\{2N\}$, instead of one, if one deletes also the last column and row of P, and replacing the event $\{X_n = 0\}$ with the event $\{X_n = 0 \text{ or } 2N\}$.

Recall that the eigenvalue with maximal modulus of a matrix with nonnegative entries is real and nonnegative, and is called the *dominant eigenvalue*. The dominant eigenvalue of P is 1, but that of Q is strictly less than 1. Now because we have assumed that all nonzero states communicate, Q is regular, so thanks to the Perron–Frobenius theorem, its dominant eigenvalue, say $\lambda \in (0,1)$ has multiplicity 1. We write v for its right eigenvector (column vector with positive entries) and u for its left eigenvector (row vector with positive entries), normalized so that

$$\sum_{i\geq 1} u_i = 1 \quad \text{and} \quad \sum_{i\geq 1} u_i v_i = 1.$$

Theorem 3.2.1 *Let $(X_n; n \geq 0)$ be a Markov chain in $\{0,1,\dots,2N\}$ absorbed at 0, such that 0 is accessible and all nonzero states communicate. Then X has a Yaglom limit Υ given by*

$$\mathbb{P}(\Upsilon = j) = u_j \qquad j \geq 1,$$

and there is a Q-process $X^\uparrow$ whose transition probabilities are given by

$$\mathbb{P}_i(X_n^\uparrow = j) = \frac{v_j}{v_i}\lambda^{-n}\mathbb{P}_i(X_n = j) \qquad i,j \geq 1.$$

In addition, the Q-process converges in distribution to the r.v. $X_\infty^\uparrow$ with law

$$\mathbb{P}(X_\infty^\uparrow = j) = u_j v_j \qquad j \geq 1.$$

Exercise 3.1 *Prove the previous statement using the following key result in the Perron–Frobenius theorem*

$$\lim_{n\to\infty} \lambda^{-n} q_{ij}(n) = u_j v_i \qquad i,j \geq 1.$$

3.2.2 Application to population genetics

It would be straightforward to apply the last theorem to the Wright–Fisher model if we knew how to express the eigenvectors associated to the dominant eigenvalue of Q, which is $\lambda_2 = (2N-1)/2N$. Unfortunately, this is not the case, but nice expressions are available for the Moran model.

To stick to the preceding framework, we could consider the discrete-time Markov chain associated to the Moran model, namely, the chain with transition probabilities

$$p_{i,i-1} = p_{i,i+1} = \left(\frac{i}{2N}\right)\left(1 - \frac{i}{2N}\right) \quad \text{and} \quad p_{i,i} = \left(\frac{i}{2N}\right)^2 + \left(1 - \frac{i}{2N}\right)^2.$$

This model is studied in detail in [39], so we prefer to provide the result in continuous time, which is extremely similar to that in discrete time. In addition, it will give a first flavour of what happens in the other models we consider.

Recall the transition rates for the Moran model

$$\begin{cases} i \;\to\; i+1 & \text{at rate} \quad ai(2N-i) \\ i \;\to\; i-1 & \text{at rate} \quad ai(2N-i), \end{cases}$$

where the individual birth rate a was taken equal to $1/2N$ in Chapter 1, and to 1 in Chapter 2. Let (Q_t) be the semigroup of the Moran model killed upon absorption (i.e. the analogue of the matrix Q^n in the discrete setting), and R its rate matrix, that is, the square matrix of order $2N-1$ with generic element

$$r_{ij} = \left\{ \begin{array}{rll} ai(2N-i) & \text{if} & j = i \pm 1 \\ -2ai(2N-i) & \text{if} & j = i. \end{array} \right. \qquad i,j \in \{1, \dots, 2N-1\}.$$

It is known that $Q_t = \exp(tR)$, and it is easy to check that

$$u = \left(\begin{array}{ccccc} 1 & \cdots & 1 & \cdots & 1 \end{array} \right) \qquad \text{and} \qquad v = \left(\begin{array}{c} 1(2N-1) \\ \vdots \\ j(2N-j) \\ \vdots \\ (2N-1)1 \end{array} \right)$$

are resp. left and right eigenvectors of R for the eigenvalue $-2a$, and hence of Q_t for the eigenvalue $\exp(-2at)$. The latter value can be proved to be the dominant eigenvalue of Q_t, so the following result is straightforward.

Theorem 3.2.2 *Let $(Y_t; t \geq 0)$ denote the Moran model with individual birth rate a. It has a Yaglom limit Υ which is* uniform *on $\{1, \dots, 2N-1\}$*

$$\mathbb{P}(\Upsilon = j) := \lim_{t\to\infty} \mathbb{P}(Y_t = j \mid Y_t \notin \{0, 2N\}) = \frac{1}{2N-1} \qquad j \in \{1, \dots, 2N-1\}.$$

*The Moran model conditioned on being never absorbed, or Q-*process *$Y^{\uparrow}$, has transitions*

$$\mathbb{P}_i(Y_t^{\uparrow} = j) := \lim_{s\to\infty} \mathbb{P}_i(Y_t = j \mid Y_{t+s} \notin \{0, 2N\}) = e^{2at}\, \frac{j(2N-j)}{i(2N-i)}\, \mathbb{P}_i(Y_t = j).$$

In addition, the Q-process converges in distribution to $Y_\infty^{\uparrow}$, given by

$$\mathbb{P}(Y_\infty^{\uparrow} = j) = c_N^{-1} j(2N-j),$$

where c_N is a normalizing constant equal to $(2N-1)(2N)(2N+1)/6$.

3.3 Markov chains with countable state-space

In this section, we will consider the case when X is a Markov chain with integer values, having 0 as an accessible absorbing state, and all nonzero states communicating. A more general setting is studied in [46], but it roughly amounts to the situation just described.

3.3.1 R-theory

In his pioneering papers [97, 98], D. Vere-Jones extended the Perron–Frobenius theory to infinite matrices Q with generic element q_{ij}, $i,j \geq 1$, where it is implicit that

$$q_{ij} = \mathbb{P}_i(X_1 = j) \qquad i,j \geq 1.$$

The results on the applications of R-theory to Markov chains with one absorbing state can be found in [94]. Let us also mention that a similar theory has been set up for processes with uncountable state-space by P. Tuominen and R.L. Tweedie [99].

Under the assumptions on X made above, the following result holds, where R^{-1} is the analogue of the dominant eigenvalue in the finite case.

Theorem 3.3.1 (Vere-Jones [97]) *The entire series $z \mapsto \sum_{n\geq 1} q_{ij}(n)z^n$ all have the same radius of convergence, say R, and the following dichotomy holds.*

*Either the series $z \mapsto \sum_{n\geq 1} q_{ij}(n)R^n$ all converge, and Q is said R-*transient*, or they all diverge, and Q is said R-*recurrent*.*

*In the latter case, either the sequences $(q_{ij}(n)R^n)$ all vanish, and Q is said R-*null*, or they converge to a positive limit, and Q is said R-*positive*.*

Remark 3.2 *This theorem also applies to stochastic matrices corresponding to recurrent chains (here we consider a killed chain or a transient chain). Then the transition matrix is either 1-null (when the chain is null-recurrent), or 1-positive (when the chain is positive-recurrent).*

Theorem 3.3.2 (Vere-Jones [97]) *The value R is the greatest value of r for which there exist nonzero r-subinvariant vectors (i.e. $ruQ \leq u$, or $rQv \leq v$). The infinite matrix Q is R-recurrent iff there is a unique pair (u, v) (up to a constant factor) such that the row vector u and the column vector v both have positive entries and*

$$RuQ = u \quad \text{and} \quad RQv = v.$$

In addition, if Q is R-recurrent, then it is R-positive iff $uv < \infty$. In that case, $R^n q_{ij}(n) \to u_j v_i / uv$.

The link with quasi-stationarity is the following

Theorem 3.3.3 (Seneta–Vere-Jones [94]) *Let a_i be the probability that X hits 0 starting from i. Then the three following limits*

$$\lim_{n\to\infty} \mathbb{P}_i(Z_n = j \mid Z_n \neq 0), \qquad \lim_{n\to\infty} \mathbb{P}_i(Z_n = j \mid Z_{n+k} \neq 0) \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathbb{P}_j(Z_n \neq 0)}{\mathbb{P}_i(Z_n \neq 0)}$$

all *exist and are honest, iff Q is R-positive with $R > 1$ and the left eigenvector u satisfies $\sum_{i\geq 1} a_i u_i < \infty$.*

3.3.2 Application to the BGW model

R-theory and quasi-stationarity

The results from the previous subsection can be applied to the BGW process $(Z_n, n \geq 0)$ with offspring distribution $(p_k, k \geq 0)$ and associated p.g.f. f. Recall that m stands for the mean offspring and that we always assume $p_0 p_1 \neq 0$. From now on, $m \leq 1$. Note that everything that follows can also be stated for the supercritical process provided it is *conditioned on extinction*, since such conditioned process has the same law as a subcritical BGW process. A slight difference is that the condition $L \log L$ that will appear below is always fulfilled in the latter case.

To compute the radius of convergence R defined in the previous subsection for the BGW model, note that

$$q_{11}(n) = \mathbb{P}_1(Z_n = 1) = f_n'(0),$$

where f_n is the n-th iterate of f, so that

$$\frac{q_{11}(n+1)}{q_{11}(n)} = \frac{f_{n+1}'(0)}{f_n'(0)} = f' \circ f_n(0),$$

which converges to m (since $f_n(0)$ converges to the extinction probability 1). This proves that $R = m^{-1}$. Also, one can prove that $v_j = j$ (up to a positive constant factor), since then

$$(Qv)_j = \sum_{k\geq 1} k\mathbb{P}_j(Z_1 = k) = \frac{d}{ds} f(s)^j \mid_{s=1} = mj = mv_j.$$

One of the contributions of [94] is to prove that Q is m^{-1}–positive iff $\sum_k p_k(k \log k) < \infty$. This ensures that this condition is equivalent to the *joint* existence of the Yaglom limit and the Q-process. Indeed, when Q is m^{-1}–positive, $\sum_k ku_k < \infty$, so in particular $\sum_k u_k < \infty$, which proves that the condition $\sum_k a_k u_k < \infty$ in Theorem 3.3.3 is fulfilled (recall $a_k = 1$ in the subcritical case).

The computation of u is not possible in general; the knowledge of v will be exploited for the Q-process.

The most refined statement for the Yaglom limit is the following

Theorem 3.3.4 (Sevast'yanov [95], Heathcote–Seneta–Vere-Jones [52]) *In the subcritical case, there is a r.v. Υ with probability distribution $(u_j, j \geq 1)$ such that $uQ = mu$ and*

$$\lim_{n\to\infty} \mathbb{P}(Z_n = j \mid Z_n \neq 0) = u_j \qquad j \geq 1.$$

The following dichotomy holds.

If $\sum_k p_k(k \log k) = \infty$, then $m^{-n}\mathbb{P}(Z_n \neq 0)$ goes to 0 and Υ has infinite expectation.

If $\sum_k p_k(k \log k) < \infty$, then $m^{-n}\mathbb{P}(Z_n \neq 0)$ has a positive limit, and Υ has finite expectation such that

$$\lim_{n\to\infty} \mathbb{E}_x(Z_n \mid Z_n \neq 0) = \mathbb{E}(\Upsilon).$$

Remark 3.3 *The functional version of the equation $uQ = mu$ is $g(f(s)) - g(f(0)) = mg(s)$, where g is the probability generating function of Υ. Taking $s = 1$ implies that $g(f(0)) = 1 - m$, so that*

$$1 - g(f(s)) = m(1 - g(s)) \qquad s \in [0, 1],$$

which can be read as

$$\mathbb{E}(1 - s^{Z_1} \mid Z_0 = \Upsilon, Z_1 \neq 0) = \mathbb{E}(1 - s^{Z_1} \mid Z_0 = \Upsilon)/m = \mathbb{E}(1 - s^{\Upsilon}).$$

Remark 3.4 *As pointed out in [94], for any $\alpha \in (0, 1)$,*

$$g_\alpha(s) := 1 - (1 - g(s))^\alpha \qquad s \in [0, 1]$$

is the generating function of an honest probability distribution which is a QSD associated to the speed of mass loss m^α*, since*

$$1 - g_\alpha(f(s)) = m^\alpha(1 - g_\alpha(s)) \qquad s \in [0,1].$$

This statement is to be related to forthcoming Theorem 3.5.2.

We wish to provide here the so-called 'conceptual proof' of [81]. To do this, we will need the following lemma.

Lemma 3.3.5 *Let* (ν_n) *be a set of probability measures on the positive integers, with finite means* a_n*. Let* $\hat{\nu}_n(k) := k\nu_n(k)/a_n$*. If* $(\hat{\nu}_n)$ *is tight, then* (a_n) *is bounded, while if* $\hat{\nu}_n \to \infty$ *in distribution, then* $a_n \to \infty$*.*

Proof (Lyons–Pemantle–Peres). Let μ_n be the law of Z_n conditioned on $Z_n \neq 0$. For any planar embedding of the tree, we let u_n be the leftmost child of the root that has descendants at generation n, and H_n the number of such descendants. If $Z_n = 0$, we put $H_n = 0$. Then check that

$$\mathbb{P}(H_n = k) = \mathbb{P}(Z_n = k \mid Z_n \neq 0, Z_1 = 1) = \mathbb{P}(Z_{n-1} = k \mid Z_{n-1} = 0).$$

Since $H_n \leq Z_n$, (μ_n) *increases stochastically.* Now since $\mathbb{E}(Z_n) = \mathbb{E}(Z_n, Z_n \neq 0)$,

$$\mathbb{P}(Z_n \neq 0) = \frac{\mathbb{E}(Z_n)}{\mathbb{E}(Z_n \mid Z_n \neq 0)} = a_n^{-1} m^n,$$

where $a_n := \sum_{k\geq 1} k\mu_n(k)$. This shows that $(m^{-n}\mathbb{P}(Z_n \neq 0))$ *decreases* and that its limit is nonzero iff the means of μ_n are bounded, that is (a_n), are bounded.

Now consider the BGW process with immigration $Z^\uparrow$, where the immigrants come in packs of $\zeta = k$ individuals with probability $(k+1)p_{k+1}/m$. The associated generating function is $g(s) = f'(s)/m$. Then the law of $Z_n^\uparrow$ is given by (1.3)

$$\mathbb{E}_0(s^{Z_n^\uparrow}) = \prod_{k=0}^{n-1} (f' \circ f_k(s)/m) \qquad s \in [0,1].$$

An immediate recursion shows that

$$\mathbb{E}_0(s^{Z_n^\uparrow + 1}) = m^{-n} s f_n'(s) \qquad s \in [0,1].$$

Now a straightforward calculation provides

$$a_n^{-1} \sum_{k\geq 1} k\mu_n(k)s^k = m^{-n} s f_n'(s) \qquad s \in [0,1].$$

We deduce that the size-biased distribution $\hat{\mu}_n$ is the law of the BGW process with immigration $Z^\uparrow$ (+1) started at 0. Now thanks to Theorem 1.1.7, this distribution converges to a proper distribution or to $+\infty$, according whether $\sum_{k\geq 1} \mathbb{P}(\zeta = k)\log k$ is finite or infinite, that is, according whether $\sum_k p_k(k\log k)$ is finite or infinite. Conclude thanks to the previous lemma. □

Exercise 3.2 *In the linear-fractional case, where* $p_0 = b$ *and* $p_k = (1-b)(1-a)a^{k-1}$ *for* $k \geq 1$*, prove that the Yaglom limit is geometric with parameter* $\min(a/b, b/a)$.

By the previous theorem, we know that in the subcritical case, under the $L \log L$ condition, the probabilities $\mathbb{P}(Z_n \neq 0)$ decrease geometrically with reason m. The following statement gives their rate of decay in the critical case.

Theorem 3.3.6 (Kesten–Ney–Spitzer [64]) *Assume* $\sigma^2 := \mathrm{Var}(Z_1) < \infty$. *Then we have*
(i) Kolmogorov's estimate [69]

$$\lim_{n\to\infty} n\mathbb{P}(Z_n \neq 0) = \frac{2}{\sigma^2}$$

(ii) Yaglom's limit law [102]

$$\lim_{n\to\infty} \mathbb{P}(Z_n/n \geq x \mid Z_n \neq 0) = \exp(-2x/\sigma^2) \qquad x > 0.$$

For a modern proof, see [82].

Q-process

Now we state without proof the results concerning the conditioning of Z on being non extinct in the distant future (they can be found in [6, pp.56–59]).

Theorem 3.3.7 *The* Q*-process* $Z^\uparrow$ *can be properly defined as*

$$\mathbb{P}_i(Z_n^\uparrow = j) := \lim_{k\to\infty} \mathbb{P}_i(Z_n = j \mid Z_{n+k} \neq 0) = \frac{j}{i} m^{-n} \mathbb{P}_i(Z_n = j) \qquad i, j \geq 1. \tag{3.2}$$

It is transient if $m = 1$*, and if in addition* $\sigma^2 := \mathrm{Var}(Z_1) < \infty$*, then*

$$\lim_{n\to\infty} \mathbb{P}(2Z_n^\uparrow/\sigma^2 \geq x) = \int_x^\infty y \exp(-y)\, dy \qquad x > 0.$$

If $m < 1$*, it is positive-recurrent iff*

$$\sum_{k\geq 1} p_k (k \log k) < \infty.$$

In the latter case, the stationary law is the size-biased distribution (ku_k/μ) *of the Yaglom limit* u *from the previous theorem.*

Observe that the generating function of the transition probabilities of the Q-process is given by

$$\sum_{j\geq 0} \mathbb{P}_i(Z_1^\uparrow = j) s^j = \sum_{j\geq 0} \mathbb{P}_i(Z_1 = j) \frac{j}{i} m^{-1} s^j = \frac{sf'(s)}{m} f(s)^{i-1},$$

where the foregoing equality is reminiscent of the BGW process with immigration. More precisely, it provides a useful recursive construction for a Q-process tree, called size-biased tree (see also [81], as well as the section in Chapter 1 dedicated to BGW trees and Chapter 4). At each generation, a particle is marked. Give to the others independent BGW descendant trees with offspring distribution p, which is (sub)critical. Give to the marked particle k children with probability μ_k, where

$$\mu_k = \frac{kp_k}{m} \qquad k \geq 1,$$

and mark one of these children at random.

This construction shows that the Q-process tree contains one infinite branch and one only, that of the marked particles, and that $(Z_n^{\uparrow} - 1, n \geq 0)$ is a BGW process with immigration f'/m (by construction, $Z_n^{\uparrow}$ is the total number of particles belonging to generation n; just remove the marked particle at each generation to recover the process with immigration).

3.4 Kimmel's branching model

We consider a deterministic binary tree, interpreted as the genealogical tree of a cell population. Each cell is represented by a finite sequence $\mathbf{i} = (i_1, \dots, i_n)$, where n is the depth, or generation, of the cell in the tree, and the $i's$ are elements of $\{0, 1\}$. Each cell $\mathbf{i}$ contains a number $Z_{\mathbf{i}}$ (integer) of parasites. The dynamics is given by the following rules

- When the cell $\mathbf{i}$ divides, it gives birth to two daughter cells $\mathbf{i}0$ and $\mathbf{i}1$
- Upon division, each parasite that was contained in the cell at its birth has proliferated into $Z^{(0)} + Z^{(1)}$ parasites oriented to one of the two daughter cells, $\mathbf{i}0$ and $\mathbf{i}1$, respectively
- all parasites proliferate independently and with the same law.

As a consequence, conditional on $Z_{\mathbf{i}} = n$, the pair $(Z_{\mathbf{i}1}, Z_{\mathbf{i}2})$ is distributed as

$$\sum_{k=1}^{n} (Y_k^{(0)}, Y_k^{(1)}),$$

where the r.v.'s $(Y_k^{(0)}, Y_k^{(1)})$ are i.i.d. with common distribution $(Z^{(0)}, Z^{(1)})$. This model is due to M. Kimmel and was first studied in [66]. Note that here, time is discrete (generations), and the law of $(Z^{(0)}, Z^{(1)})$ is *not necessarily symmetric*.

Observe that in this model, the total number of parasites $(\mathcal{Z}_n; n \geq 0)$ is a BGW process with offspring distribution $\xi = Z^{(0)} + Z^{(1)}$. In addition, the number of parasites $(Z_n; n \geq 0)$ along a *random line of descent* is a BGW process in *random environment*, where there are two environments (law $Z^{(0)}$ and law $Z^{(1)}$) chosen with equal probabilities.

We use the following notation

$$\begin{aligned} m_0, m_1 &= \mathbb{E}(Z^{(0)}), \mathbb{E}(Z^{(1)}) \\ \mathbb{G}_n &= \text{generation } n \\ \mathbb{G}_n^\star &= \text{set of \textit{contaminated} cells in } \mathbb{G}_n \\ \partial\mathbb{T} &= \text{set of infinite lines of descent} \\ \partial\mathbb{T}^\star &= \text{set of infinite \textit{contaminated} lines of descent} \end{aligned}$$

Theorem 3.4.1 (Bansaye [8]) *The following limit exists a.s.*

$$L := \lim_{n\to\infty} 2^{-n} \# \mathbb{G}_n^\star.$$

If $m_0 m_1 \leq 1$*, then* $\mathbb{P}(L = 0) = 1$ *(dilution).*
If $m_0 m_1 > 1$*, then* $\{L = 0\} = \{\mathrm{Ext}\}$*.*

In words, L is the *asymptotic proportion of contaminated cells*. In the case $m_0 m_1 > 1$, the only way for the organism to heal up, that is, for this asymptotic proportion to be zero, is that all parasites die out.

Now assume $m_0 + m_1 > \max(m_0^2 + m_1^2, 1)$, so that in particular

- $m_0 m_1 \leq 1$: the organism heals up ($L = 0$) and parasites become extinct along any random line of descent (the BGW process with random environment $(Z_n; n \geq 0)$ is subcritical)
- $m_0 + m_1 > 1$: parasites (may) grow overall (the BGW process $(\mathcal{Z}_n; n \geq 0)$ is supercritical)

Theorem 3.4.2 (Bansaye [8]) *Let*

$$F_k(n) := \frac{\#\{\mathbf{i} \in \mathbb{G}_n^\star : Z_\mathbf{i} = k\}}{\#\mathbb{G}_n^\star} \ .$$

Then conditionally on $\{\mathrm{Ext}\}^c$*,*

$$\lim_{n\to\infty} F_k(n) \stackrel{P}{=} \mathbb{P}(\Upsilon = k),$$

where Υ *is the Yaglom limit of the number of parasites along a random line of descent.*

In addition, let

$$F_k(n,p) := \frac{\#\{\mathbf{i} \in \mathbb{G}_{n+p}^\star : Z_{\mathbf{i}|n} = k\}}{\#\mathbb{G}_{n+p}^\star} \ .$$

Then conditionally on $\{\mathrm{Ext}\}^c$*,*

$$\lim_{n\to\infty}\lim_{p\to\infty} F_k(n,p) \stackrel{P}{=} \frac{k\mathbb{P}(\Upsilon = k)}{\mathbb{E}(\Upsilon)},$$

which is the stationary probability of the Q-process along a random line of descent.

Remark 3.5 *The QSD* Υ *is a distribution conditional on* asymptotically evanescent *events. This is not the case for the limit of* $F_k(n)$*, which provides therefore a* realization *of the QSD. This has consequences in terms of numeric computations of QSD's.*

Remark 3.6 *As* $p \to \infty$*, one roughly has*

$F_k(n,p) \approx \mathbb{P}($ *a line of descent uniformly picked in* $\partial\mathbb{T}^\star$ *contained* k *parasites at generation* n $)$.

Because descendances of parasites separate into disjoint lines of descent *with high probability, a* uniform pick *in* $\partial\mathbb{T}^\star$ *roughly amounts to a* size-biased pick *at generation* n*. Roughly speaking, if there are two cells at generation* n*, the first one containing 1 parasite and the second one containing* k *parasites, then the probability that a uniform line in* $\partial\mathbb{T}^\star$ *descends from a parasite (this makes sense because of separation) in the second cell is* k *times greater than its complementary (size-biasing of* Υ*).*

This provides a conceptual explanation for the link between the Yaglom limit and the stationary measure of the Q-process (size-biasing).

3.5 The CB-process

In this section, we consider a CB-process $(Z_t; t \geq 0)$ with branching mechanism ψ. Everything that follows is done under the assumptions that

$$\rho := \psi'(0+) \geq 0 \quad \text{and} \quad \phi(t) := \int_t^\infty \frac{ds}{\psi(s)} < \infty,$$

that is, Z is (sub)critical and absorbed at 0 with probability 1. If we dropped the first assumption and assumed $\rho < 0$, the rest of the section would remain unchanged provided that we conditioned Z on eventual absorption. Indeed, thanks to Proposition 2.2.15, this amounts to considering a subcritical CB-process.

Actually, one could also drop the second assumption [85] by conditioning Z_t on events of the type $\{Z_{t+s} > \epsilon\}$ or $\{S(\epsilon) > t + s\}$, where $S(\epsilon)$ is the last hitting time of ϵ.

Before continuing further, we state a technical lemma, whose proof, as well as those of all other statements can readily be found in [76]. Recall that φ is the inverse mapping of ϕ.

Lemma 3.5.1 *Assume $\rho \geq 0$ and let $G(\lambda) := \exp(-\rho\phi(\lambda))$. Then for any positive λ*

$$\lim_{t\to\infty} \frac{u_t(\lambda)}{\varphi(t)} = G(\lambda),$$

and for any nonnegative s

$$\lim_{t\to\infty} \frac{\varphi(t+s)}{\varphi(t)} = e^{-\rho s}.$$

When $\rho > 0$, the following identities are equivalent

(i) $G'(0+) < \infty$

(ii) $\int^\infty r \log r \, \Lambda(dr) < +\infty$

(iii) There is a positive constant c such that $\varphi(t) \sim c \exp(-\rho t)$, as $t \to \infty$.

In that case, $G'(0+) = c^{-1}$.

3.5.1 Quasi-stationary distributions

Recall that a QSD ν is defined by (3.1). Then by application of the simple Markov property,

$$\mathbb{P}_\nu(T > t + s) = \mathbb{P}_\nu(T > s)\mathbb{P}_\nu(T > t),$$

so that the extinction time T under $\mathbb{P}_\nu$ has an exponential distribution with parameter, say, γ. Then γ can be seen as the constant speed of mass loss of $(0, \infty)$ under $\mathbb{P}_\nu$. It is a natural question to characterize all the quasi-stationary probabilities associated to a given speed of mass loss γ.

Theorem 3.5.2 *Assume $\rho > 0$ (subcritical case). For any $\gamma \in (0, \rho]$ there is a unique QSD ν_γ associated to the speed of mass loss γ. It is characterized by its Laplace transform*

$$\int_{(0,\infty)} \nu_\gamma(dr)\, e^{-\lambda r} = 1 - e^{-\gamma\phi(\lambda)} \qquad \lambda \geq 0.$$

There is no QSD associated to $\gamma > \rho$.

In addition, the minimal QSD ν_ρ *is the so-called* Yaglom distribution*, in the sense that for any starting point* $x \geq 0$*, and any Borel set* A

$$\lim_{t\to\infty} \mathbb{P}_x(Z_t \in A \mid T > t) = \nu_\rho(A).$$

From now on, we will denote by Υ the r.v. with distribution ν_ρ. Since the Laplace transform of Υ is $1 - G$, Υ is integrable iff $\int^\infty r \log r \, \Lambda(dr) < \infty$, and then

$$\mathbb{E}(\Upsilon) = c^{-1},$$

where c is defined in Lemma 3.5.1.

Proof. There are multiple ways of proving this theorem. The most straightforward way is the following

$$1 - e^{-\gamma t} = \mathbb{P}_{\nu_\gamma}(T < t) = \int_{(0,\infty)} \nu_\gamma(dr)\, e^{-r\varphi(t)},$$

so that, writing $t = \phi(\lambda)$, one gets

$$1 - e^{-\gamma\phi(\lambda)} = \int_{(0,\infty)} \nu_\gamma(dr)\, e^{-\lambda r} \qquad \lambda \geq 0.$$

Another way of getting this consists in proving that $\nu_\gamma Q = -\gamma Q + \gamma\delta_0$, where Q is the infinitesimal generator of the Feller process Z and δ_0 is the Dirac measure at 0. Taking Laplace transforms then leads to the differential equation

$$\gamma(1 - \chi_\gamma(\lambda)) = -\psi(\lambda)\chi_\gamma'(\lambda) \qquad \lambda \geq 0,$$

where χ_γ stands for the Laplace transform of ν_γ. Solving this equation with the boundary condition $\chi(0) = 1$ yields the same result as given above.

Next recall that $\phi(\lambda) = \int_\lambda^\infty du/\psi(u)$, so that $\phi'(\lambda) \sim -1/\rho\lambda$ and $\phi(\lambda) \sim -\rho^{-1}\log(\lambda)$, as $\lambda \downarrow 0$. This entails

$$\int_{(0,\infty)} r\nu_\gamma(dr)\, e^{-\lambda r} \sim C(\lambda)\lambda^{\gamma/\rho - 1} \qquad \text{as } \lambda \downarrow 0,$$

where C is slowly varying at 0^+, which would yield a contradiction if $\gamma > \rho$.

Before proving that $1 - G^{\gamma/\rho}$ is indeed a Laplace transform, we display the Yaglom distribution of Z. Observe that

$$\mathbb{E}_x(1 - e^{-\lambda Z_t} \mid T > t) = \frac{\mathbb{E}_x(1 - e^{-\lambda Z_t})}{\mathbb{P}_x(T > t)} = \frac{1 - e^{-xu_t(\lambda)}}{1 - e^{-x\varphi(t)}}$$

so that, by Lemma 3.5.1,

$$\lim_{t\to\infty} \mathbb{E}_x(e^{-\lambda Z_t} \mid T > t) = 1 - G(\lambda) \qquad \lambda > 0.$$

Since $G(0^+) = 0$, this proves indeed that $1 - G$ is the Laplace transform of some probability measure ν_ρ on $(0, \infty)$. It just remains to show that when $\gamma \in (0, \rho)$, $1 - G^{\gamma/\rho}$ is indeed the Laplace transform of some probability measure ν_γ on $(0, \infty)$. Actually this stems from the following result applied to $g = G$ and $\alpha = \gamma/\rho$: if $1 - g$ is the Laplace transform of some probability measure on $(0,\infty)$, then so is $1-g^\alpha$, for any $\alpha\in(0,1)$. □

It is not difficult to get a similar result as the last theorem in the critical case. Assume $\rho = 0$ and $\sigma := \psi''(0+) < +\infty$ (Z has second-order moments). Variations on the arguments of the proof of Lemma 3.5.1 then show that $\varphi(t)\sim 2/\sigma t$ as $t\to\infty$, and

$$\lim_{t\to\infty} u_t(\lambda/t)/\varphi(t) = \frac{1}{1+2/\sigma\lambda} \qquad \lambda>0.$$

Since

$$\mathbb{E}_x(1-e^{-\lambda Z_t/t} \mid T>t) = \frac{1-e^{-xu_t(\lambda/t)}}{1-e^{-x\varphi(t)}} \qquad \lambda>0,$$

the following statement follows, which displays the usual 'universal' exponential limiting distribution of the rescaled conditioned critical process.

Theorem 3.5.3 *Assume $\rho=0$ and $\sigma:=\psi''(0+)<+\infty$. Then*

$$\lim_{t\to\infty}\mathbb{P}_x(Z_t/t>z\mid T>t) = \exp(-2z/\sigma) \qquad z\ge 0.$$

3.5.2 The Q-process

The next theorem states the existence in some special sense of the branching process conditioned to be never extinct, or Q-process.

Theorem 3.5.4 *Let $x>0$.*

(i) The conditional laws $\mathbb{P}_x(\cdot\mid T>t)$ converge as $t\to\infty$ to a limit denoted by $\mathbb{P}_x^\uparrow$, in the sense that for any $t\ge 0$ and $\Theta\in\mathcal{F}_t$,

$$\lim_{s\to\infty}\mathbb{P}_x(\Theta\mid T>s) = \mathbb{P}_x^\uparrow(\Theta).$$

(ii) The probability measures $\mathbb{P}^\uparrow$ can be expressed as h-transforms of $\mathbb{P}$ based on the $(\mathbb{P},(\mathcal{F}_t))$-martingale

$$D_t = Z_t e^{\rho t},$$

that is

$$d\mathbb{P}^\uparrow_{x|\mathcal{F}_t} = \frac{D_t}{x}\,.\,d\mathbb{P}_{x|\mathcal{F}_t}$$

(iii) The process $Z^\uparrow$ which has law $\mathbb{P}_x^\uparrow$ is a CBI(ψ, χ) started at x, where χ is (the Laplace transform of a subordinator) defined by

$$\chi(\lambda)=\psi'(\lambda)-\psi'(0+), \qquad \lambda\ge 0.$$

Now we investigate the asymptotic properties of this Q-process $Z^\uparrow$. The symbol $\mathbb{P}^\uparrow$ denotes the law of the Q-process, whereas $P^\uparrow$ is that of the Lévy process *conditioned to stay positive* (see Appendix).

Also recall that the Yaglom r.v. Υ displayed in Theorem 3.5.2 is integrable as soon as $\int^\infty r\log r\,\Lambda(dr)<\infty$.

Theorem 3.5.5 *(i) (Lamperti transform) If* $\rho = 0$*, then*

$$\lim_{t\to\infty} Z_t^{\uparrow} = +\infty \qquad a.s.$$

Moreover, set

$$\theta_t = \int_0^t Z_s^{\uparrow}\, ds, \qquad t \geq 0,$$

and let κ *be its right inverse, then for* $x > 0$*, the process* $Z^{\uparrow} \circ \kappa$ *under* $\mathbb{P}_x$ *has law* $P_x^{\uparrow}$*.*

(ii) If $\rho > 0$*, the following dichotomy holds.*
(a) If $\int^{\infty} r \log r\, \Lambda(dr) = \infty$*, then*

$$\lim_{t\to\infty} Z_t^{\uparrow} \stackrel{P}{=} +\infty.$$

(b) If $\int^{\infty} r \log r\, \Lambda(dr) < \infty$*, then* $Z_t^{\uparrow}$ *converges in distribution as* $t \to \infty$ *to a positive r.v.* $Z_{\infty}^{\uparrow}$ *which has the distribution of the size-biased Yaglom distribution*

$$\mathbb{P}(Z_{\infty}^{\uparrow} \in dr) = \frac{r\mathbb{P}(\Upsilon \in dr)}{\mathbb{E}(\Upsilon)} \qquad r > 0.$$

3.6 Diffusions

3.6.1 Fisher–Wright diffusion

Recall the Fisher–Wright diffusion Y, which is the scaling limit of the Moran model

$$dY_t = \sqrt{Y_t(1 - Y_t)}\, dB_t.$$

Works of M. Kimura ([67], but see also [39, chapter 5] and the references therein) are at the origin of the following theorem.

Theorem 3.6.1 *The Yaglom limit* Υ *of the Fisher–Wright diffusion exists and is* uniform *on the open interval* $(0, 1)$*.*

Its Q-process $Y^{\uparrow}$ *can be obtained via the following h-transform*

$$\mathbb{P}_x(Y_t^{\uparrow} \in dy) := \lim_{s\to\infty} \mathbb{P}_x(Y_t \in dy \mid Y_{t+s} \notin \{0, 1\}) = e^t\, \frac{y(1-y)}{x(1-x)}\, \mathbb{P}_x(Y_t \in dy).$$

In addition, the Q-process converges in distribution to $Y_{\infty}^{\uparrow}$*, given by*

$$\mathbb{P}(Y_{\infty}^{\uparrow} \in dy) = \frac{y(1-y)}{6}\, dy \qquad y \in (0, 1).$$

3.6.2 CB-diffusion

In this subsection, we briefly translate our results in the case when the CB-process is a diffusion. Recall that Z is absorbed with positive probability, and is solution to

$$dZ_t = rZ_t dt + \sqrt{\sigma Z_t} dB_t \qquad t > 0,$$

where B is a standard Brownian motion. Recall that if $r = 0$, then for any $t > 0$,

$$\phi(t) = \varphi(t) = 2/\sigma t,$$

whereas if $r \neq 0$,

$$\phi(t) = -r^{-1} \log(1 - 2r/\sigma t) \quad \text{ and } \quad \varphi(t) = (2r/\sigma)e^{rt}/(e^{rt} - 1).$$

Note that $\rho = \psi'(0+) = -r$.

The quasi-stationary distributions

Here we assume that $r < 0$ (subcritical case), so that $\rho = -r > 0$. Then from Theorem 3.5.2, for any $\gamma \in (0, \rho]$, the Laplace transform of the QSD ν_γ is

$$\int_0^\infty \nu_\gamma(dx)\, e^{-\lambda x} = 1 - \left(\frac{\lambda}{\lambda + 2\rho/\sigma}\right)^{\gamma/\rho}$$

In particular, whenever $\gamma < \rho$, ν_γ has infinite expectation, and it takes only elementary calculations to check that for any $\gamma < \rho$, ν_γ has a density f_γ given by

$$f_\gamma(t) = \frac{2\rho/\sigma}{\Gamma(1 - \gamma/\rho)\Gamma(\gamma/\rho)} \int_0^1 ds\ s^{\gamma/\rho}(1 - s)^{-\gamma/\rho}\ e^{-2\rho t s/\sigma} \qquad t > 0.$$

This can also be expressed as

$$\nu_\gamma((t, \infty)) = \mathbb{E}(\exp(-2\rho B t/\sigma)) \qquad t > 0,$$

where B is a random variable with law $\mathrm{Beta}(\gamma/\rho, 1 - \gamma/\rho)$.

Finally, for $\gamma = \rho$, the Laplace transform is easier to invert and provides the Yaglom distribution. The Yaglom r.v. Υ with distribution ν_ρ is an exponential variable with parameter $2\rho/\sigma$

$$\mathbb{P}(\Upsilon \in dx) = (2\rho/\sigma)e^{-2\rho x/\sigma} \qquad x \geq 0.$$

The Q-process

Here we assume that $r \leq 0$. From Theorem 3.5.4, the Q-process is a CBI-process with branching mechanism ψ and immigration mechanism $\chi = \psi' - \rho$. Now recall that the infinitesimal generator B of a CBI(ψ, χ) acts on the exponential functions $e_\lambda(x) := \exp(-\lambda x)$ as

$$Be_\lambda(x) = (x\psi(\lambda) - \chi(\lambda))e_\lambda(x) \qquad x \geq 0.$$

In the present case, $\psi(\lambda) = \sigma\lambda^2/2 - r\lambda$ and $\chi(\lambda) = \sigma\lambda$, so that

$$Be_\lambda(x) = (x\frac{\sigma\lambda^2}{2} - xr\lambda - \sigma\lambda)e_\lambda(x) \qquad x \geq 0,$$

which yields, for any twice differentiable function f,

$$Af(z) = \frac{\sigma}{2}xf''(x) + rxf'(x) + \sigma f'(x) \qquad x \geq 0,$$

and this can equivalently be read as

$$dZ_t^\uparrow = rZ_t^\uparrow dt + \sqrt{\sigma Z_t^\uparrow} dB_t + \sigma dt,$$

where $Z^\uparrow$ stands for the Q-process. Note that the immigration can readily be seen in the additional deterministic term σdt.

Now if $r < 0$, according to Theorem 3.5.5, the Q-process converges in distribution to the r.v. $Z_\infty^\uparrow$ which is the size-biased Υ. But Υ is an exponential r.v. with parameter $2\rho/\sigma$, so that

$$\mathbb{P}(Z_\infty^\uparrow \in dx) = (2\rho/\sigma)^2 \; xe^{-2\rho x/\sigma} \qquad x \geq 0,$$

or equivalently,

$$Z_\infty^\uparrow \overset{(d)}{=} \Upsilon_1 + \Upsilon_2,$$

where Υ_1 and Υ_2 are two independent copies of the Yaglom r.v. Υ.

3.6.3 More general diffusions

Let $(Z^N)_N$ be a sequence of birth–death processes on $\{0, N^{-1}, 2N^{-1}, \ldots\}$, with birth and death rates from state x respectively equal to $\lambda_N(x)$ and $\mu_N(x)$. More visually,

$$\begin{cases} x & \rightarrow & x + 1/N & \text{at rate} & \lambda_N(x) \\ x & \rightarrow & x - 1/N & \text{at rate} & \mu_N(x), \end{cases}$$

where we assume that $\lambda_N(0) = \mu_N(0) = 0$, ensuring that the state 0 is absorbing. Let us further assume that $\sup_N \lambda_N(x) \leq B(x),\ x \geq 0$ and that there exist a positive constant γ and a function h of class C^1 on $[0, +\infty)$ such that $h(0) = 0$, and for any $x \in [0, +\infty)$,

$$\lim_{N\to\infty} \frac{1}{N}(\lambda_N(x) - \mu_N(x)) = h(x) \quad \text{and} \quad \lim_{N\to\infty} \frac{1}{2N^2}(\lambda_N(x) + \mu_N(x)) = \gamma. \tag{3.3}$$

If $\gamma = 0$, the asymptotic behaviour of Z^N (as N gets large) is close to the dynamical system $\dot{z} = h(z)$. If γ is strictly positive, the sequence $(Z^N)_N$ converges in law to a generalized Feller diffusion defined as solution to the stochastic differential equation

$$dZ_t = \sqrt{\gamma Z_t} dB_t + h(Z_t)dt. \tag{3.4}$$

In particular, this is the case if $\lambda_N(x) = (\gamma N^2 + \mu N)x$ and $\mu_N(x) = (\gamma N^2 + \mu N)x + cx^2/N$. The limiting diffusion is then the *Feller diffusion with logistic growth* [36, 74], already defined in the last chapter, that is, $h(z) = rz - cz^2$, where we have set $r = \lambda - \mu$.

The goal of [20] is to display sufficient conditions on h to get the existence of QSDs. Thus, we make the following additional assumptions on h (recall h is of class C^1 on $[0,+\infty)$ and $h(0)=0$) denoted Assumption (HH)

$$(i)\lim_{x\to\infty}\frac{h(x)}{\sqrt{x}}=-\infty, \qquad (ii)\lim_{x\to\infty}\frac{xh'(x)}{h(x)^2}=0.$$

Notice that Assumption (ii) holds for most classical functions (polynomial, exponential, logarithmic,...). In particular (HH) holds for any subcritical branching diffusion ($h(x)=-rx$, with $r>0$), and any logistic branching diffusion ($h(x)=rx-cx^2$). Assumption (i) ensures at least that the process is absorbed at 0 with probability 1. However, we can modify it so as to have a limit equal to $+\infty$ instead of $-\infty$, provided the process is further *conditioned on extinction.* Indeed, the following statement ensures that conditioning on extinction a diffusion satisfying the last assumption, roughly amounts to *replacing h with* $-h$.

Proposition 3.6.2 *Assume that Z is given by (3.4), where h satisfies (HH)(ii), along with*

$$\lim_{x\to\infty}h(x)=+\infty.$$

Let $u(x):=\mathbb{P}_x(\mathrm{Ext})$, and X the diffusion Z conditioned on extinction. Then X is given by

$$dX_t=\sqrt{\gamma X_t}dB_t+\left(h(X_t)+\gamma X_t\frac{u'(X_t)}{u(X_t)}\right)dt.$$

In addition,

$$h(x)+\gamma x\frac{u'(x)}{u(x)}\sim_{x\to\infty}-h(x).$$

Remark 3.7 *Recall that when $h(x)=rx$ (pure branching diffusion), the conditioning on extinction* exactly *turns h into $-h$.*

Proof. Recall from Subsection 2.2.1 that the generator $L^\star$ of X is given by

$$L^\star f(x)=\frac{1}{u(x)}L(uf)(x)\qquad x\geq 0,$$

where L is the generator of Z. Because $Lf(x)=(\gamma x/2)f''(x)+h(x)f'(x)$ and $Lu=0$, it is easy to get

$$L^\star f(x)=\frac{\gamma}{2}xf''(x)+\left(h(x)+\gamma x\frac{u'(x)}{u(x)}\right)\qquad x\geq 0,$$

which ends the first part of the theorem. The second part relies on technical tricks using $Lu=0$ and Assumption (HH)(ii). □

Theorem 3.6.3 (Cattiaux et al. [20]) *Under Assumption (HH), the diffusion Z given by (3.4) has a Yaglom limit. If in addition $\int^\infty 1/h>-\infty$, then*
(i) Z comes down from infinity
(ii) the Yaglom limit is the unique *QSD*
(iii) the Q-process is well defined.

Chapter 4

Random trees and Ray–Knight type theorems

In this chapter, we start with introducing *splitting trees*, namely those random trees where individuals give birth at constant rate during a lifetime with *general distribution*, to i.i.d. copies of themselves. We show that the *contour process* of such trees is a Lévy process, which allows to derive new properties of splitting trees, including the coalescent point process, and to recover well-known connections between Lévy processes and branching processes. These connections open up to more general problems such as Ray–Knight type theorems.

Unless otherwise specified, the results mentioned in this chapter are to be found in [77].

4.1 Preliminaries on trees

4.1.1 Discrete trees

We consider (locally finite) rooted trees [31, 84]. Let $\mathcal{U}$ be the set of finite sequences of integers.

A *discrete tree* $\mathcal{T}$ is a subset of $\mathcal{U}$, such that each vertex of $\mathcal{T}$ is represented thanks to the so-called Ulam–Harris–Neveu labelling as follows. The root of the tree is $\emptyset$, and the j-th child of $u = (u_1, \ldots, u_n) \in \mathbb{N}^n$, is uj, where vw stands for the concatenation of the sequences v and w. Then we define $|u| = n$ the *generation*, or *genealogical height* or *depth*, of u.

In addition, we write $u \prec v$ if u is an *ancestor* of v, that is, there is a sequence w such that $v = uw$. For any $u = (u_1, \ldots, u_n)$, $u|k$ denotes the ancestor $(u_1, \ldots, u_k)$ of u at generation k. Finally, we denote by $u \wedge v$ the *most recent common ancestor*, in short *mrca*, of u and v, that is, the sequence w with highest generation such that $w \prec u$ and $w \prec v$.

4.1.2 Chronological trees

We consider particular instances of real trees, as defined e.g. in [36, 79]. The real trees we consider here can roughly be seen as the set of full edges embedded in the plane of some discrete tree, where each edge length is a *lifespan*.

Specifically, each individual of the underlying discrete tree possesses a *birth level* α and a *death level* ω, both nonnegative real numbers such that $\alpha < \omega$, and (possibly zero) offspring whose birth times are distinct from one another and belong to the interval (α, ω). We think of a *chronological tree* as the set of all so-called *existence points* of individuals (vertices) of the discrete tree.

Definition. More rigorously, let

$$\mathbb{U} = \mathcal{U} \times [0, +\infty),$$

and let p_1 and p_2 stand respectively for the canonical projections on $\mathcal{U}$ and $[0, +\infty)$.

A *chronological tree* $\mathbb{T}$ is a subset of $\mathbb{U}$ such that $\mathcal{T} := p_1(\mathbb{T})$ is a discrete tree and for any $u \in \mathcal{T}$, there are $0 \leq \alpha(u) < \omega(u) \leq \infty$ such that $(u, \sigma) \in \mathbb{T}$ if and only if $\sigma \in (\alpha(u), \omega(u)]$. We write $\rho := (\emptyset, 0)$ for the root of $\mathbb{T}$.

For any $u \in \mathcal{T}$, $\alpha(u)$ is the birth level of u, $\omega(u)$ her death level, and we denote by $\zeta(u)$ its *lifespan* $\zeta(u) := \omega(u) - \alpha(u)$. We will always assume that $\alpha(\emptyset) = 0$. Also for any $u \in \mathcal{T}$ and $j \in \mathbb{N}$ such that $uj \in \mathcal{T}$, the birth time $\alpha(uj)$ of u's daughter uj is in $(\alpha(u), \omega(u))$. Points of the type $(u, \alpha(uj))$ are called *branching points*.

Because the construction is rather obvious and for the sake of conciseness, we do not mention all other requirements needed to properly define a chronological tree.

The number of individuals alive at the chronological level τ is denoted by Ξ_τ

$$\Xi_\tau = \text{Card}\{v \in \mathcal{T} : \alpha(v) < \tau \leq \omega(v)\} = \text{Card}\{x \in \mathbb{T} : p_2(x) = \tau\},$$

and $(\Xi_\tau; \tau \geq 0)$ is usually called the *width process*.

Genealogical and metric structures. A chronological tree can naturally be equipped with the following genealogical structure. For any $x, y \in \mathbb{T}$ such that $x = (u, \sigma)$ and $y = (v, \tau)$, we will say that x is an *ancestor* of y, and write $x \prec y$ as for discrete trees, if $u \prec v$ and

- if $u = v$, then $\sigma \leq \tau$
- if $u \neq v$, then $\sigma \leq \alpha(uj)$, where j is the unique integer such that $uj \prec v$.

For $y = (v, \tau)$, the *segment* $[\rho, y]$ is the set of ancestors of x, that is

$$[\rho, y] := \{(v, \sigma) : \alpha(v) < \sigma \leq \tau\} \cup \{(u, \sigma) : \exists k, u = v|k, \alpha(v|k) < \sigma \leq \alpha(v|k+1)\}.$$

For any $x, y \in \mathbb{T}$, it is not difficult to see that there is a unique $z \in \mathbb{T}$ such that $[\rho, x] \cap [\rho, y] = [\rho, z]$. The existence point z is the point of highest level in $\mathbb{T}$ such that $z \prec x$ and $z \prec y$. In particular, notice that $p_1(z) = p_1(x) \wedge p_1(y)$ (i.e. $p_1(z)$ is the mrca of $p_1(x)$ and $p_1(y)$). The level $p_2(z)$ is called the *coalescence level* of x and y, and z the *coalescence point* (or most recent common ancestor) of x and y, denoted as for discrete trees by $z = x \wedge y$.

Total order '$\leq$'. There is a total order $\leq$ on $\mathbb{T}$, and for any $x, y \in \mathbb{T}$, we say that x is to the *left-hand side* of y (and y is to the *right-hand side* of x) if $x \leq y$. This order is defined as follows

- if $y \prec x$ then $x \leq y$
- if neither $x \prec y$ nor $x \prec y$, then $x \wedge y = (u, \sigma)$ is a branching point and there is an integer j such that $\sigma = \alpha(uj)$. Notice that $u \prec p_1(x)$ and $u \prec p_1(y)$, but either $uj \prec p_1(x)$ or $uj \prec p_1(y)$. If $uj \prec p_1(y)$, then $x \leq y$, otherwise $y \leq x$.

It is important to notice that if $\mathbb{T}$ is not reduced to the root, then for any $x \in \mathbb{T}$,

$$(\emptyset, \omega(\emptyset)) \leq x \leq (\emptyset, \alpha(\emptyset)) = \rho.$$

Miscellaneous. For any $u \in \mathcal{T}$, we define the *total length* of the tree $\lambda(\mathbb{T})$ as the sum of all lifespans

$$\lambda(\mathbb{T}) = \sum_{u \in \mathcal{T}} \zeta(u) \leq \infty.$$

Actually the Borel σ-field of $\mathbb{U}$ can be defined as the σ-field generated by sets of the type $\{u\} \times A$, for $u \in \mathcal{U}$ and A any Borel set of $[0, \infty]$. Then λ can be seen as the Lebesgue measure on $\mathbb{T}$, and for any Borel subset $\mathbb{S}$ of $\mathbb{T}$, it makes sense to define its length $\lambda(\mathbb{S})$. We will abusively say that $\mathbb{S}$ is *finite* if it has finite length and finite discrete part $p_1(\mathbb{S})$.

4.2 The exploration process

4.2.1 Definition

In most works on real trees, the latter are defined from their contour, which is a real function coding the genealogy. Here we do the opposite.
Hereafter, $\mathbb{T}$ denotes a finite chronological tree, with total length $\ell = \lambda(\mathbb{T})$.

Definition 4.2.1 *For any $x \in \mathbb{T}$, the set*

$$\mathbb{S}(x) := \{y \in \mathbb{T} : y \leq x\}$$

is measurable, so we can define the mapping φ from $\mathbb{T}$ to the real interval $[0, \ell]$ as

$$\varphi(x) := \lambda(\mathbb{S}(x)).$$

Then in particular, $\varphi(\emptyset, \omega(\emptyset)) = 0$ and $\varphi(\rho) = \ell$. The process $(\varphi^{-1}(t); t \in [0, \ell])$ is called the exploration process*. Its second projection will be denoted by $(X_t; t \in [0, \ell])$ and called JCCP, standing for* jumping chronological contour process.

The JCCP is a càdlàg function taking the values of all levels of all points in $\mathbb{T}$, once and once only, starting at the death level of the ancestor and following this rule : when the visit of an individual v with lifespan $(\alpha(v), \omega(v)]$ begins, the value of the JCCP is $\omega(v)$. The JCCP then visits lower chronological levels of v's lifespan at linear speed -1. If v has no child, then this visit lasts exactly the lifespan $\zeta(v)$ of v; if v has at least one child, then the visit is interrupted each time a birth level of one of v's daughters, say w, is encountered (youngest child first since the visit started at the death level). At this point, the JCCP jumps from $\alpha(w)$ to $\omega(w)$ and starts the visit of the existence levels of w. Since the tree is finite, the visit of v has to terminate: it does so at the chronological level $\alpha(v)$ and continues the exploration of the existence levels of v's mother, at the level where it had been interrupted. This procedure then goes on recursively as soon as level 0 is encountered ($0 = \alpha(\emptyset)$ = birth level of the root).

A figure representing a splitting tree and the associated JCCP is moved to page 97 (Fig. 4.3).

Remark 4.1 *The JCCP has another interpretation [80, Fig.1, p.230] in terms of queues. Each jump Δ_t is interpreted as a customer of a one-server queue arrived at time t with a load Δ_t. This server treats the customers' loads at constant speed 1 and has priority LIFO (last in – first out). The tree structure is derived from the following rule: each customer is the mother of all customers who interrupted her while she was being served. Then the value X_t of the JCCP is the remaining load in the server at time t.*

4.2.2 Properties of the JCCP

Actually, the chronological tree itself can be recovered from the JCCP, modulo labelling of sisters. In the next statement, we give useful applications of this correspondence.

For each $t \in [0, \ell]$, set

$$\hat{t} := \sup\{s \leq t : X_s < X_t\} \vee 0 \qquad 0 \leq t \leq \ell.$$

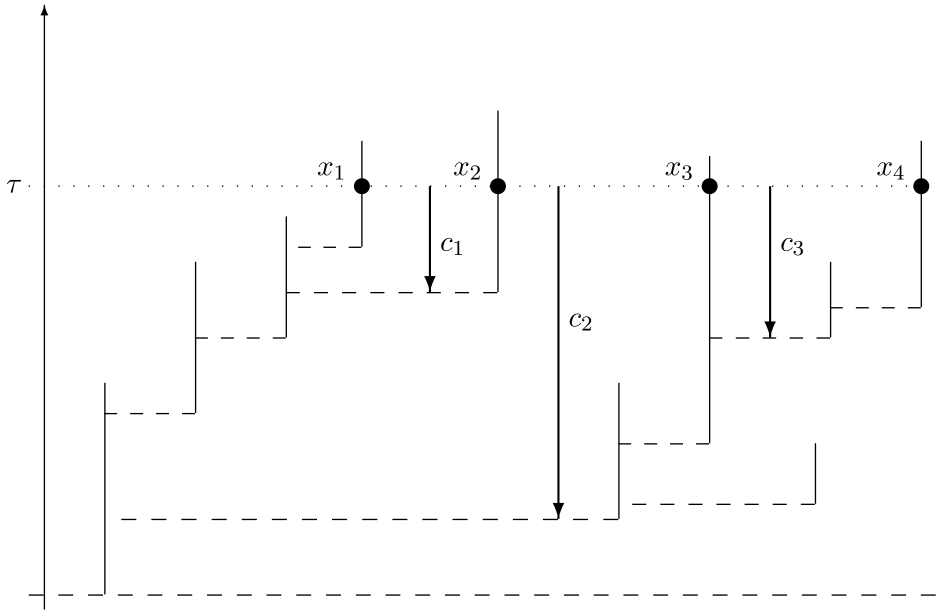

Figure 4.1: Illustration of a splitting tree showing the durations c_1, c_2, c_3 elapsed since *coalescence* for each of the three consecutive pairs $(x_1, x_2), (x_2, x_3)$ and (x_3, x_4) of the $\Xi_\tau = 4$ individuals alive at level τ. The *heights* (generations in the discrete tree) of points x_1, x_2, x_3, x_4 are respectively 3, 3, 2, 4.

Theorem 4.2.2 *Let $x = (u, \sigma)$ and $y = (v, \tau)$ denote any two points in $\mathbb{T}$, and set $s = \varphi(x)$ and $t = \varphi(y)$. Then the following hold*

(i) The first visit to v is $\hat{t}$

$$\varphi(v, \omega(v)) = \hat{t}.$$

If t is a jump time of X, then $t = \hat{t}$, and the first visit to the mother u of v in $\mathcal{T}$ is given by

$$\varphi(u, \omega(u)) = \sup\{s \leq t : X_s < X_{t-}\}.$$

(ii) Ancestry between x and y.

$$y \prec x \ \Leftrightarrow \ \hat{t} \leq s \leq t$$

(iii) Coalescence level between x and y (assume e.g. $s \leq t$).

$$p_2(x \wedge y) = \inf_{s \leq r \leq t} X_r.$$

For any $t \in [0, \ell]$, we define the process $X^{(t)}$ on $[0, t]$ as

$$X^{(t)}_r := X_{t-} - X_{(t-r)-} \qquad r \in [0, t],$$

with the convention that $X_{0-} = 0$. The following corollary states that the record times of $X^{(t)}$ are the first visits to v's ancestors, and displays the process $(H_t; t \geq 0)$ of genealogical heights, or *height process*, as a functional of the path of the JCCP.

Corollary 4.2.3 *For $v \in \mathcal{T}$, set $n = |v|$ and t_k the first visit to $v_k = v|k$ (ancestor of v belonging to generation k), that is,*

$$t_k := \varphi(v_k, \omega(v_k)) \qquad 0 \leq k \leq n.$$

(i) Define recursively the record times of $X^{(t)}$ by $s_1 = t - \hat{t}$ and

$$s_{k+1} = \inf\{s \geq s_k : X_s^{(t)} > X_{s_k}^{(t)}\} \qquad 1 \leq k \leq n.$$

Then

$$t_k = t - s_{n-k+1} \qquad 0 \leq k \leq n.$$

(ii) Define H_t as the genealogical height of $\varphi^{-1}(t)$ in $\mathcal{T}$. Then H_t is given by

$$\begin{aligned} H_t := |p_1(\varphi^{-1}(t))| &= \mathrm{Card}\{0 \leq s \leq t : X_s^{(t)} = \sup_{0 \leq r \leq s} X_r^{(t)}\} \\ &= \mathrm{Card}\{0 \leq s \leq t : X_{s-} < \inf_{s \leq r \leq t} X_r\}. \end{aligned} \tag{4.1}$$

4.3 Random splitting trees

In this section, we consider random chronological trees, called *splitting trees*, and corresponding to (binary homogeneous) Crump–Mode–Jagers processes.

4.3.1 Definition

The *Crump–Mode–Jagers process*, or *CMJ process*, denotes the general, non Markovian, branching process. The random genealogy associated to it satisfies very loose rules, that can be described as follows.

- given her lifetime $(\alpha, \omega]$, each individual gives birth at times $\alpha + t_1 < \alpha + t_2 < \cdots$, to clutches of respective sizes $\xi_1, \xi_2, \ldots$
- the reproduction scheme is then characterized by the joint distribution of the lifespan $\zeta := \omega - \alpha$ and the point process (t_i, ξ_i) on $(0, \zeta] \times \mathbb{N}$
- reproduction schemes of all individuals are i.i.d.

However, in what follows, we will always make the following additional assumptions

1. all clutch sizes (the ξ_i's) are a.s. equal to 1 (*binary splitting*)
2. conditional on the lifetime ζ, the point process (t_i) is a *Poisson point process* on $(0, \zeta)$ with intensity b (*homogeneous* reproduction scheme)
3. the common distribution of lifetimes is $\Lambda(\cdot)/b$, where Λ is some positive measure on $(0, \infty]$ with mass b.

Trees satisfying the previous assumptions fit in the framework given in the preliminaries on trees, so they are random chronological trees. They could also be called general binary trees with constant birth rate, or homogeneous binary Crump–Mode–Jagers trees, but we will prefer to use the terminology from the literature as '*splitting trees*' [44]. On the other hand, this terminology is unfortunate because it evokes binary fission.

We assume further that the tree starts with one ancestor individual $\emptyset$, with deterministic lifetime $(0, \chi]$.

Observe that for any $v \in \mathbb{T}$, in agreement with the definition, the pairs $(\alpha(vi), \zeta(vi))_{i\geq 1}$ made of the birth levels and lifespans of v's offspring are the atoms of a Poisson measure on $(\alpha(v), \omega(v)) \times (0, +\infty)$ with intensity measure $\mathrm{Leb} \otimes \Lambda$, where Leb stands for the Lebesgue measure. In words, each individual gives birth at rate b during her lifetime (α, ω), to independent copies of herself whose lifespan common distribution is $\Lambda(\cdot)/b$. In particular, the total offspring number of v, conditional on $\zeta(v) = z$, is a Poisson random variable with parameter bz.

Two branching processes. Recall that Ξ_τ denotes the number of individuals alive at τ (width of $\mathbb{T}$ at level τ)

$$\Xi_\tau = \mathrm{Card}\{v \in \mathcal{T} : \alpha(v) < \tau \leq \omega(v)\},$$

and set $\mathcal{Z}_n$ the number of individuals of generation n

$$\mathcal{Z}_n = \mathrm{Card}\{v \in \mathcal{T} : |v| = n\}.$$

From the definition, it is easy to see that $(\mathcal{Z}_n; n \geq 0)$ is a BGW process started at 1, with offspring generating function f

$$f(s) = \int_{(0,\infty)} b^{-1}\Lambda(dz)\ e^{-bz(1-s)} \qquad s \in [0, 1],$$

so the mean number of offspring per individual is

$$m := \int_{(0,\infty)} z\Lambda(dz).$$

As for the width process $(\Xi_\tau; \tau \geq 0)$, it is a *homogeneous binary Crump–Mode–Jagers process*. Unless Λ is exponential, this branching process is *not* Markovian.

Remark for modeling purpose. This birth–death scheme can be seen alternatively as a constant birth intensity measure $b\,\mathrm{Leb}$ combined with an *age-dependent death intensity measure* μ, that is

$$\mathbb{P}(\zeta > z) = \exp\left(-\int_0^z \mu(d\sigma)\right) \qquad z > 0,$$

which forces the equality

$$\exp\left(-\int_0^z \mu(d\sigma)\right) = \bar{\Lambda}(z)/b \qquad z > 0,$$

where $\bar{\Lambda}(z) = \int_{(z,\infty)} \Lambda(ds)$, and yields the following equation for μ

$$\mu(dz) = \frac{\Lambda(dz)}{\bar{\Lambda}(z)} \qquad z > 0.$$

If for some reason one has to proceed the other way round (the birth rate b and some death intensity measure μ are given), notice that the lifespan measure Λ then has

$$\Lambda(dz) = b\mu(dz) \exp\left(-\int_0^z \mu(d\sigma)\right) \qquad z > 0,$$

so that the only requirement that μ has to fulfill in order to fit the present framework is

$$\int_0^1 \mu(d\sigma) < \infty \text{ and } \int_1^\infty \mu(d\sigma) = \infty.$$

Note that the second condition is equivalent to a.s. finiteness of lifespans.

4.3.2 Law of the JCCP

From now on, we consider a splitting tree $\mathbb{T}$ with lifespan measure Λ, and we condition the lifespan $\zeta(\emptyset)$ of the ancestor to equal $\chi > 0$. To keep this in mind, we denote the law of the tree by $\mathbb{P}_\chi$.

As usual we say that extinction occurs when $\mathbb{T}$ is finite, and denote this event by $\{\text{Ext}\}$. Recall that the jumping chronological contour process (JCCP) of a chronological tree is well defined as soon as this tree is finite. For any positive real number τ, we define $C_\tau(\mathbb{T})$ as the tree obtained after cutting all branches of $\mathbb{T}$ above level τ

$$C_\tau(\mathbb{T}) := \{x \in \mathbb{T} : p_2(x) \leq \tau\}.$$

Observe that $C_\tau(\mathbb{T})$ is a finite chronological tree.

We denote by Y the compensated compound Poisson process $t \mapsto Y_t := -t + \sum_{s \leq t} \Delta_s$, where $(\Delta_t, t \geq 0)$ is a Poisson point process with intensity measure $\text{Leb} \otimes \Lambda$. Then Y is a Lévy process with no negative jumps, whose Laplace exponent (see Appendix) will be denoted by ψ

$$\psi(\lambda) := \lambda - \int_0^\infty (1 - \exp(-\lambda r))\, \Lambda(dr) \qquad \lambda \geq 0.$$

The following statement is the fundamental result of this section. It is a little bit surprising at first sight, in the sense that, eventhough $(\varphi^{-1}(t); t \geq 0)$ is not Markovian, its second projection is. Recall that T_A is the first hitting time of A.

Theorem 4.3.1 *Denote by $(X_t, t \geq 0)$ the JCCP of $\mathbb{T}$ whenever $\mathbb{T}$ is finite, and by $(X_t^{(\tau)}, t \geq 0)$ the JCCP of $C_\tau(\mathbb{T})$.*

(i) Define recursively $t_0 = 0$, and $t_{i+1} = \inf\{t > t_i : X_t^{(\tau)} \in \{0, \tau\}\}$. Then under $\mathbb{P}_\chi$, the killed paths $e_i := (X_{t_i+t}^{(\tau)}, 0 \leq t < t_{i+1} - t_i)$, $i \geq 0$, form a sequence of i.i.d. excursions, distributed as the Lévy process Y killed at $T_0 \wedge T_{(\tau,+\infty)}$, ending at the first excursion hitting 0 before $(\tau, +\infty)$. These excursions all start at τ, but the first one, which starts at $\min(\chi, \tau)$.

(ii) Under $\mathbb{P}_\chi(\cdot \mid \text{Ext})$, X has the law of the Lévy process Y, started at χ, conditioned on, and killed upon, hitting 0.

Proof. We only provide a hand-waving proof, in the (sub)critical case, which is sufficient to understand the main point.

The JCCP visits every individual starting from its date of death, going backwards through chronological levels (at unit speed), and interrupting the visit by the visit of its children (youngest child first, since the visit started at the death level). The JCCP X is then the sum of a pure-jump process and the linear process $t \mapsto -t$, started at χ and killed when it hits 0. Since the sizes of jumps are exactly the lifespans of the individuals of the CMJ tree, the jumps of X have i.i.d. sizes. Since individuals give birth at constant rate b during their lifetime, interarrivals of births are exponential with parameter b and independent. Because the exploration is done at unit speed, and by the lack-of-memory property of the exponential distribution, this also holds for the interarrivals of the jumps of X. As a consequence, X has the law of Y killed when it hits 0. □

4.3.3 New properties of the Crump–Mode–Jagers process

Recall that $\mathbb{P}_\chi$ denotes the law of the splitting tree with lifespan measure Λ, conditional on $\zeta(\emptyset) = \chi$. From now on, we will denote by P_χ the law of the Lévy process Y with Laplace exponent ψ, when it is started at χ. The scale function W is the positive function with Laplace transform $1/\psi$, and η is the largest root of ψ (see Appendix). Check that

$$\psi'(0+) = 1 - m,$$

so that $\eta > 0$ corresponds to the supercritical case, and $\eta = 0$, either to the subcritical case (if $\psi'(0+) > 0$) or the critical case (if $\psi'(0+) = 0$). Also note that because ψ is convex, $\psi(0) = 0$, and $\psi(\lambda) \geq \lambda - b$, one has

$$\eta < b.$$

Corollary 4.3.2 *The probability of extinction is $\mathbb{P}_\chi(\mathrm{Ext}) = e^{-\eta\chi}$. In addition, the law of the one-dimensional marginal of the homogeneous CMJ process $\Xi_\tau = \mathrm{Card}\{t : X_t = \tau\}$ (number of individuals alive at level τ) is given by*

$$\mathbb{P}_\chi(\Xi_\tau = 0) = P_\chi(T_0 < T_{(\tau,+\infty)}) = W(\tau - \chi)/W(\tau),$$

and conditional on being nonzero, Ξ_τ has a geometric distribution *with success probability*

$$P_\tau(T_0 > T_{(\tau,+\infty)}) = 1 - 1/W(\tau).$$

In particular, $\mathbb{E}_\chi(\Xi_\tau \mid \Xi_\tau \neq 0) = W(\tau)$.

Corollary 4.3.3 *In the supercritical case, set $\mathbb{P}^\natural := \mathbb{P}(\cdot \mid \mathrm{Ext})$. Then the JCCP under $\mathbb{P}^\natural$ is a Lévy process with Laplace exponent $\psi^\natural$*

$$\psi^\natural(\lambda) = \psi(\lambda + \eta) = \lambda - \int_0^\infty e^{-\eta r}(1 - e^{-\lambda r})\Lambda(dr) \qquad \lambda \geq 0.$$

As a consequence, the supercritical splitting tree conditioned on its extinction has the same law as the subcritical splitting tree with lifespan measure $e^{-\eta r}\Lambda(dr)$. In particular its birth rate equals $b - \eta$.

Corollary 4.3.4 *The asymptotic behaviour of the CMJ process is as follows.*
(i) (Yaglom's distribution) In the subcritical case,

$$\lim_{\tau\to\infty} \mathbb{P}(\Xi_\tau = n \mid \Xi_\tau \neq 0) = m^{n-1}(1-m) \qquad n \geq 1.$$

(ii) In the critical case, provided that $\int^\infty r^2\Lambda(dr) < \infty$,

$$\lim_{\tau\to\infty} \mathbb{P}(\Xi_\tau/\tau > x \mid \Xi_\tau \neq 0) = \exp(-\psi''(0+)\ x/2) \qquad x \geq 0.$$

(iii) In the supercritical case, provided that $\int^\infty r\log(r)\Lambda(dr) < \infty$, *conditional on* $\{\mathrm{Ext}^c\}$,

$$\lim_{\tau\to\infty} e^{-\eta\tau}\ \Xi_\tau = \xi \qquad a.s.,$$

where ξ *is an exponential variable with parameter* $1/\psi'(\eta)$.

Becaue of the last result, η is called the *Malthusian parameter.*

4.3.4 The coalescent point process

Fix $\tau > 0$. For any chronological tree $\mathbb{T}$, we let $(x_i(\tau); 1 \leq i \leq \Xi_\tau)$ denote the ranked points $x_1 \leq x_2 \leq \cdots$ of $\mathbb{T}$ such that $p_2(x_i) = \tau$. In particular, the vertices $p_1(x_i)$ of $\mathcal{T}$ are exactly the individuals alive at level τ.

Theorem 4.3.5 *Under* $\mathbb{P}$, *the coalescence level between* $x_j(\tau)$ *and* $x_k(\tau)$ $(j \leq k)$ *is given by*

$$p_2(x_j(\tau) \wedge x_k(\tau)) = \min\{a_i : j \leq i < k\},$$

where $a_i := p_2(x_i \wedge x_{i+1})$, $1 \leq i \leq \Xi_\tau$, *form a sequence of i.i.d. r.v., killed at the first negative one, and distributed as* $\tau - \inf Y_t$, *where* Y *is the Lévy process with Laplace exponent* ψ *started at* τ *and killed upon exiting* $(0, \tau]$. *As a consequence, the duration* $c_i = \tau - a_i$ *elapsed since the coalescence of points* $x_i(\tau)$ *and* $x_{i+1}(\tau)$ *has*

$$\mathbb{P}(c_i \leq \sigma) = \frac{1 - 1/W(\sigma)}{1 - 1/W(\tau)} \qquad \sigma \leq \tau.$$

Please go back to Fig. 4.1 page 79 for a picture of these coalescence levels.

Remark 4.2 *Taking* $\Lambda(dz) = be^{-bz}dz$, *one can recover Lemma 3 in [85]. Namely, since an elementary calculation yields* $W(x) = 1 + bx$,

$$\mathbb{P}(c_i \in d\sigma) = \frac{1}{(1+b\sigma)^2}\frac{1+b\tau}{\tau} \qquad \sigma \leq \tau.$$

4.4 Spine decomposition of infinite trees

In this section, we rely on the properties of the JCCP to give a decomposition of infinite trees into an infinite skeleton together with the finite trees grafted on it. We treat two cases.

The first case is that of the supercritical splitting tree, where the infinite skeleton is a *Yule tree* and the finite trees are distributed according to $\mathbb{P}^\natural = \mathbb{P}(\cdot \mid \mathrm{Ext})$.

The second case is that of (sub)critical splitting trees conditioned on not being extinct at generation n. As $n \to \infty$, the limiting tree has *one infinite spine*, and the finite trees are distributed according to $\mathbb{P}$.

Because those trees are infinite, we cannot define the JCCP of the whole tree as previously, but only the JCCP of the *first infinite subtree* $\mathbb{F}_1$ defined as follows. Rigorously, the exploration process is always defined for the truncation at level τ of an infinite tree. We denote its inverse by φ^τ. Then $\mathbb{F}_1$ is the set of existence points x such that $\varphi^\tau(x)$ converges to a finite limit as $\tau \to \infty$. In other terms, $\mathbb{F}_1$ is the subtree coming just before the leftmost infinite branch, that is,

$$\mathbb{F}_1 := \{x \in \mathbb{T} : \lambda(\mathbb{S}(x)) < \infty\},$$

where we remind the reader that $\mathbb{S}(x) = \{y \in \mathbb{T} : y \le x\}$. By construction, $\mathbb{F}_1$ admits an exploration process and hence a JCCP, both having infinite lifetime on $\{\mathrm{Ext}^c\}$ (the subtree $\mathbb{F}_1$ has infinite length).

See Fig. 4.2 page 86.

4.4.1 The supercritical infinite tree

Recall that $\mathbb{P}_\chi(\mathrm{Ext}^c) = 1 - e^{-\eta\chi}$, where η is the Malthusian parameter. Then set

$$p := \mathbb{P}(\mathrm{Ext}^c) = \int_0^\infty b^{-1}\Lambda(dr)\mathbb{P}_r(\mathrm{Ext}^c) = \eta/b.$$

Under $\mathbb{P}(\cdot \mid \mathrm{Ext}^c)$, define recursively $u_0 = \emptyset$, and for $i \ge 0$

$u_0 \cdots u_i u_{i+1}$ is the *youngest* daughter with *infinite descendance* of $u_0 \cdots u_i$.

Then $u := u_0 u_1 u_2 \cdots$ is the *first infinite branch* of the discrete tree $\mathcal{T}$ in the lexicographical order associated to the 'youngest-first' labelling. For $k \ge 0$, let A_k be the *age* at which the individual $u|k$ gives birth to $u|k+1$, and let R_k be her *residual lifetime* at that level

$$A_k := \alpha(u|k+1) - \alpha(u|k) \quad \text{and} \quad R_k := \omega(u|k) - \alpha(u|k+1).$$

Modulo the labelling of individuals, the sequence $(A_k, R_k)_{k\ge 0}$ characterizes the *leftmost backbone* or *spine* $\mathbb{B}_1$ of the real tree

$$\mathbb{B}_1 := \{x \in \mathbb{T} : p_1(x) = u|n \text{ for some } n\}.$$

Finally, let $\mathbb{A}_1$ be the *leftmost infinite branch*

$$\mathbb{A}_1 := \{x \in \mathbb{T} : p_1(x) = u|n \text{ for some } n, p_2(x) \le \alpha(u|n+1)\},$$

and $\mathbb{R}_1$ the *thorns* of the backbone

$$\mathbb{R}_1 := \mathbb{B}_1 \backslash \mathbb{A}_1.$$

Notice that the ages characterize $\mathbb{A}_1$ and the residual lifetimes characterize $\mathbb{R}_1$. For a graphical representation of the infinite tree, see Fig. 4.2.

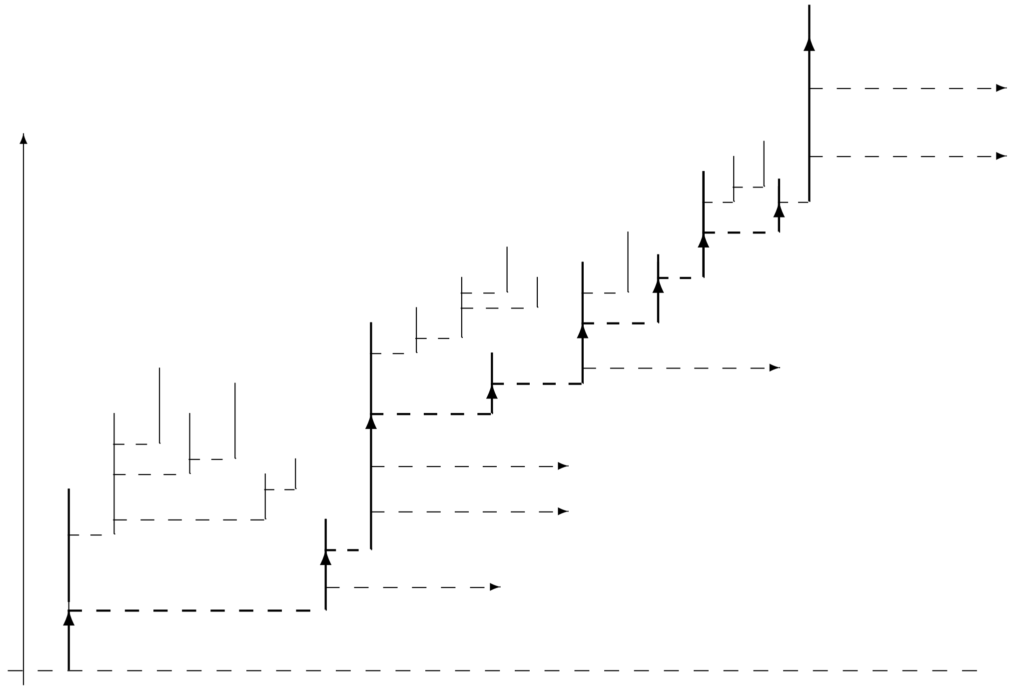

Figure 4.2: An 'infinite' splitting tree. The *backbone* $\mathbb{B}_1$ is the subtree made up of the vertical edges in bold, decomposed into the *leftmost infinite branch* $\mathbb{A}_1$ (bold arrows whose sizes are the *ages*) and the *thorns* $\mathbb{R}_1$ (superior part of the bold edges, whose sizes are the *residual lifetimes*). The *first infinite subtree* $\mathbb{F}_1$ is the collection of finite subtrees grafted on $\mathbb{R}_1$ together with $\mathbb{R}_1$ itself. The horizontal dashed lines indicate the points where trees on $\mathbb{A}_1$ are grafted.

Theorem 4.4.1 *Assume that $\eta > 0$ (supercritical case) and recall that $\mathbb{P}^\natural = \mathbb{P}(\cdot \mid \text{Ext})$.*

Under $\mathbb{P}(\cdot \mid \text{Ext}^c)$, the pairs (A_k, R_k), $k \geq 0$, are independent copies of (A, R), where

$$\mathbb{P}(A + R \in dz, R \in dr) = e^{-\eta r}\, dr\, \Lambda(dz) \qquad 0 < r < z.$$

Next let $\mathbb{B}' = \mathbb{A}' \cup \mathbb{R}'$ be a copy of $\mathbb{B}_1 = \mathbb{A}_1 \cup \mathbb{R}_1$ (characterized by the pairs (A_k, R_k)), and let

- *$(x_i)_{i\in\mathbb{N}}$ be the atoms of a Poisson measure with intensity $b(1-p)$ on $\mathbb{B}'$*
- *$(y_i)_{i\in\mathbb{N}}$ be the atoms of a Poisson measure with intensity bp on $\mathbb{A}'$.*

Conditionally on $\mathbb{B}'$ and the foregoing Poisson measures, graft i.i.d. finite subtrees with common law $\mathbb{P}^\natural$ at all points (x_i), and define recursively $\mathbb{X}$ as the tree thus obtained with independent copies of itself grafted at all points (y_i). Then $\mathbb{X}$ has the law of $\mathbb{T}$ under $\mathbb{P}(\cdot \mid \text{Ext}^c)$. In particular, the subtree of points with infinite descendance is a Yule tree *with birth rate $bp = \eta$.*

Proof. By the branching property, it suffices to characterize the joint law of the first infinite subtree $\mathbb{F}_1$ and that of the leftmost infinite branch $\mathbb{B}_1$. From Theorem 4.3.1, we can deduce that the law under $\mathbb{P}(\cdot \mid \text{Ext}^c)$ of the JCCP of $\mathbb{F}_1$ is that of a Lévy process with Laplace exponent

ψ starting from a r.v. with law $\Lambda(\cdot)/b$ and conditioned not to hit 0. As a consequence, setting J the future infimum of X

$$J_t := \inf\{X_s : s \geq t\},$$

and U_k the k-th jump time of J, we get

$$(A_k, R_k) = (J(U_k) - X(U_k-), X(U_k) - J(U_k)).$$

Moreover, conditional on these ages and residual lifetimes, excursions $(X_{U_k+t} - J_{U_k}; 0 \leq t < U_{k+1} - U_k)$ of X above its future infimum are independent and distributed as X started at R_k under $P^\natural$ and killed upon hitting zero. Since $\mathbb{F}_1$ is a collection of finite subtrees whose JCCPs are exactly the excursions of X above its future infimum, we have proved that, conditionally on $\mathbb{B}_1$, all these finite subtrees are i.i.d. with law $\mathbb{P}^\natural$.

It only remains to prove that the ages and residual lifetimes (A_k, R_k) have the law displayed in the theorem. To that end, consider any jump of X occurring at time, say, S. By the Markov property, conditional on $X(S-) = x$ and $X(S) - X(S-) = z$, the event that this jump is a jump of the future infimum and the value of that jump on this event, are independent from the past. This yields recursively the independence of the pairs (A_k, R_k). Now let I be the infimum of $\{X_t : t \geq S\}$. It is known that $x + z - I$ is exponential with parameter η, so that, since $p = \mathbb{P}(S$ is a jump time of $J)$, we get

$$\begin{aligned}
\mathbb{P}(A + R \in dz, R \in dr) &= p^{-1} \int_0^\infty \mathbb{P}(X(S-) \in dx, X(S) - X(S-) \in dz, x + z - I \in dr) \\
&= p^{-1} b^{-1} \Lambda(dz)\, \eta\, e^{-\eta r}\, dr.
\end{aligned}$$

The proof ends recalling that $p = \eta/b$. □

4.4.2 Conditioned (sub)critical trees

As in the previous subsection, we want to give a spine decomposition of the splitting tree, but here we are interested in the (sub)critical case, when the tree is conditioned to reach large *heights*. This is the exact analogue of the Q-process applied to splitting trees.

We define $\mathbb{P}^n$ as the conditional law $\mathbb{P}$ of a splitting tree on the event of reaching height n

$$\mathbb{P}^n := \mathbb{P}(\cdot \mid \mathcal{Z}_n \neq 0).$$

Under $\mathbb{P}^n$, define recursively $u_0 = \emptyset$, and for $i \geq 0$,

$u_0 \cdots u_i u_{i+1}$ is the *youngest* daughter of $u_0 \cdots u_i$ who has living descendants at generation n.

Then $u := u_0 \cdots u_n$ is the first lineage of the discrete tree $\mathcal{T}$ reaching generation n in the lexicographical order associated to the 'youngest-first' labelling. Note that u depends on n.

The ages and residual lifetimes (A_k, R_k) are defined as previously, as well as the *backbone* or *spine* $\mathbb{B}^n$

$$\mathbb{B}^n := \{x \in \mathbb{T} : p_1(x) = u|k \text{ for some } k\}.$$

In the following statement, when we say that the laws $\mathbb{P}^n$ converge for finite truncations as $n \to \infty$, we mean convergence for events that are measurable w.r.t. the truncation $C_\sigma(\mathbb{T}) = \{x \in \mathbb{T} : p_2(x) \leq \sigma\}$ defined earlier, for some fixed σ.

Theorem 4.4.2 *Assume that $\int^\infty \Lambda(dr)(r\log r) < \infty$. As $n \to \infty$, the law of $\mathbb{B}^n$ under $\mathbb{P}^n$ converges for finite truncations to an infinite spine $\mathbb{B}'$ characterized by the pairs (A_k, R_k), $k \geq 0$, which form a sequence of independent copies of $(UD, (1-U)D)$, where*

$$\mathbb{P}(D \in dz) = m^{-1} z \Lambda(dz) \qquad z > 0,$$

and U is an independent r.v. uniform on $(0,1)$.

Next let $(x_i)_{i\in\mathbb{N}}$ be the atoms of a Poisson measure with intensity b on $\mathbb{B}'$. Then as $n \to \infty$, the law of $\mathbb{T}$ under $\mathbb{P}^n$ converges for finite truncations to the infinite tree with one infinite branch $\mathbb{B}'$ obtained by grafting, at all points (x_i) on $\mathbb{B}'$, i.i.d. finite subtrees with common law $\mathbb{P}$.

Remark 4.3 *One could think that conditioning on reaching large* heights *amounts asymptotically to conditioning on reaching high* levels*. Actually, in the latter case, the limiting process would be a Lévy process (killed upon hitting zero) conditioned to drift to $+\infty$. In the critical case, such a process exists and is known as a Lévy process conditioned to stay positive (see Appendix). But in the subcritical case, such a process does not exist without additional assumptions on the existence of exponential moments [55]. In a work in progress, we will show that the former conditioning leads to a new definition of the Lévy process conditioned to stay positive, which coincides, in the critical case, with the usual definition.*

Sketch of proof. Let $\sigma > 0$, and set $N_n(\sigma)$ the number of individuals at level τ having living descendants at generation n. Since $\mathbb{P}(\mathcal{Z}_n \neq 0)$ vanishes as $n \to \infty$, $\mathbb{P}(N_n(\sigma) = 1 \mid \mathcal{Z}_n \neq 0) \to 1$ as $n \to \infty$. This ensures that the limiting tree under $\mathbb{P}^n$ has only one infinite branch, which must therefore have the law of $\mathbb{B}'$. To get the law of the spine $\mathbb{B}'$, first observe that

$$\mathbb{P}^n(R_k < r, A_k + R_k \in dz) = \rho_{n-k}^{-1} b^{-1} \Lambda(dz) \mathbb{P}_r(\mathcal{Z}_{n-k} \neq 0),$$

where

$$\rho_j := b^{-1} \int_0^\infty \Lambda(dz) \mathbb{P}_z(\mathcal{Z}_j \neq 0) \qquad j \in \mathbb{N}.$$

Now under $\mathbb{P}_z$, $(\mathcal{Z}_n; n \geq 0)$ is a BGW process with mean m, starting from a Poisson variable with mean z. More precisely, its offspring distribution is given by

$$p_k := b^{-1} \int_0^\infty \Lambda(dz) \frac{(bz)^k}{k!}$$

It is well-known that since $\int^\infty \Lambda(dr)(r\log r) < \infty$, we have $\sum_k p_k(k\log k) < \infty$. Then thanks to Theorem 3.3.4, there is a positive constant c such that

$$\lim_{n\to\infty} m^{-n} \mathbb{P}(\mathcal{Z}_n \neq 0 \mid \mathcal{Z}_0 = j) = cj,$$

and since $\mathbb{P}(\mathcal{Z}_n \neq 0 \mid \mathcal{Z}_0 = j) \leq j\, \mathbb{P}(\mathcal{Z}_n \neq 0 \mid \mathcal{Z}_0 = 1)$, the dominated convergence theorem ensures that $m^{-n}\rho_n \to cm/b$, and

$$\lim_{n\to\infty} \mathbb{P}^n(R_k < r, A_k + R_k \in dz) = m^{-1}\, r\, \Lambda(dz) \qquad 0 < r < z.$$

The theorem follows from a conditional (on the spine) application of the branching property. We leave to the reader the task of concluding rigorously. □

4.5 Ray–Knight type theorems

4.5.1 Ray–Knight theorems

Let H be any real stochastic process whose occupation measure at time t admits a density $(L_t^a; a \in \mathbb{R})$ called the *local time*. This means that for any bounded measurable f

$$\int_0^t f(X_s)\, ds = \int_{-\infty}^{+\infty} f(a)\, L_t^a\, da.$$

When H is transient or killed at some a.s. finite time, one can consider the total local time $(L_\infty^a; a \in \mathbb{R})$, that we will call *local time process*, as indexed by a.

Theorems displaying stochastic processes whose local time process is Markovian are usually called Ray–Knight type theorems since the pioneering works of D. Ray and F.B. Knight on Brownian motion. There are two such theorems [90].

1. The first Ray–Knight theorem states that the local time process of a *reflecting Brownian motion* starting from 0 and *killed upon summing up x units of local time at 0*, is a *squared Bessel process of dimension 0*, starting from x. Note that this Markov process is the *CB-process with branching mechanism $\psi(\lambda) = \lambda^2/2$*

2. The second Ray–Knight theorem, also called Ray–Knight–Williams theorem, states that the local time process on $[0, a]$ of a Brownian motion starting from a and killed upon hitting 0, is a squared Bessel process of dimension 2 starting from 0. Note that this Markov process is the *CBI-process with branching mechanism $\psi(\lambda) = \lambda^2/2$ and immigration mechanism $\chi(\lambda) = \lambda/2$*.

   Another version of this theorem uses the theorem due to J. Pitman stating that after time-reversal, the Brownian motion starting from a and killed upon hitting 0 becomes a Bessel process of dimension 3 starting from 0 and killed at its last hitting time of a. Since local times are invariant under time reversal, the Ray–Knight–Williams theorem can even be stated for $a = \infty$ as follows: *the local time process of a Bessel process of dimension 3 starting from 0 is a squared Bessel process of dimension 2 starting from 0.*

4.5.2 A new viewpoint on Le Gall–Le Jan theorem

The finite variation case

The next statement is a slight refinement of Proposition 3.2 in [80] and sheds light on the genealogy defined in [13] by composing subordinators.

For background on Jirina processes, see Subsection 2.1.2.

Theorem 4.5.1 *Let X be a Lévy process with Laplace exponent $\psi(\lambda) = \lambda - \int_0^\infty (1 - e^{-\lambda r})\Lambda(dr)$ such that $\psi'(0+) \geq 0$, started at χ and killed when it hits 0. Then define the height process*

$$H_t = \mathrm{Card}\{0 \leq s \leq t : X_{s-} < \inf_{s \leq r \leq t} X_r\} \qquad t < T_0,$$

and Z the local time process of H

$$Z_n := \int_0^{T_0} dt\, \mathbf{1}_{\{H_t = n\}} \qquad n \geq 0.$$

Then $(Z_n; n \geq 0)$ *is a* Jirina process *with branching mechanism* F *starting from* $Z_0 = \chi$*, where*

$$F(\lambda) := \int_0^\infty (1 - e^{-\lambda r})\, \Lambda(dr) \qquad \lambda \geq 0.$$

Remark 4.4 *Recall that a Jirina process with branching mechanism* F *is such that for any integer* $n \geq 1$*,*

$$Z_n = S_n \circ \cdots \circ S_1(\chi),$$

where the S_i *are i.i.d. subordinators with Laplace exponent* F*. It satisfies*

$$E_\chi(\exp(-\lambda Z_n)) = \exp(-\chi F_n(\lambda)) \qquad \lambda \geq 0,$$

where F_n *is the* n*-th iterate of* F *(see Subsection 2.1.2). Note that, after* Ξ *and* $\mathcal{Z}$*, we got with* Z *a third branching process.*

Proof. Recall from Theorem 4.2.2 and equation (4.1) that $H_t = |p_1 \circ \varphi^{-1}(t)|$, that is, H_t is the genealogical height of the individual visited at time t by the exploration process φ^{-1}, so that Z_n is the sum of lifespans of all individuals of generation n, that is,

$$Z_n = \sum_{v:|v|=n} \zeta(v).$$

Next recall that conditional on $Z_{n-1} = z$, the number of individuals of generation n is a Poisson variable with mean bz, and all their lifespans are independent with law $\Lambda(\cdot)/b$. This is not different from saying that $Z_n = S(z)$, where S is a compound Poisson process with intensity measure Λ. The theorem follows from a recursive application of this observation. □

Remark 4.5 *Actually, the previous theorem even holds if* Λ *is not finite any longer, provided that* $\int_0 r\Lambda(dr) < \infty$ *(and* $\int^\infty \Lambda(dr) < \infty$*). Indeed, this assumption ensures that the corresponding Lévy process* X *has* finite variation *(see Appendix), so that, even if birth events 'rain down' on each lifetime, the jumps of the past supremum of* X *are* discrete*. Indeed, recall that* $X^{(t)}$ *is the time-reversed of* X *at time* t

$$X^{(t)}_s = X_{t-} - X_{(t-s)-} \qquad s \in [0, t],$$

and set $S^{(t)}$ *its past supremum*

$$S^{(t)}_s := \sup\{X^{(t)}_r : 0 \leq r \leq s\}.$$

Since X *and* $X^{(t)}$ *have the same law, the jumps of* $S^{(t)}$ *are also discrete and we can define the height process* H *of* X *exactly as in (4.1), that is,*

$$H_t := \mathrm{Card}\{0 \leq s \leq t : X^{(t)}_s = S^{(t)}_s\}.$$

The infinite variation case

Note that the last theorem can be stated in the form of a Ray–Knight theorem as follows.

> 'The local time process of the height process H of X is Markovian, where X is a Lévy process with no negative jumps and finite variation, which does not drift to $+\infty$ and is killed upon hitting zero. In addition the law of this Markov process can be specified as that of a (sub)critical Jirina process'.

Actually, most of what has been done in this chapter associating Lévy processes with finite variation to splitting trees can be adapted to general Lévy processes with no negative jumps, associated to continuous trees called *Lévy trees* [30, 32]. We do not define Lévy trees, but they will be underlying the discussion, since we want to take advantage of the genealogies encoded by Lévy processes thanks to the intuition we get from the study of splitting trees with finite variation.

The generalization of Theorem 4.5.1 to the infinite variation case is due to J.F. Le Gall and Y. Le Jan [80]. Let X be any Lévy process with no negative jumps. When X has infinite variation, $S^{(t)} - X^{(t)}$ is a strong Markov process for whom 0 is regular, so H cannot be defined as the cardinal of a finite set, but as the *local time* of $S^{(t)} - X^{(t)}$ at 0.

Theorem 4.5.2 (Le Gall–Le Jan [80]) *Assume that X has infinite variation and does not drift to $+\infty$. Then the local time process of the height process H of the Lévy process X started at x and killed upon hitting 0, is a* CB-process *with branching mechanism ψ started at x.*

Remark 4.6 *This last theorem is not only a generalization of Theorem 4.5.1 to the infinite variation case, but also includes as a special case, and hence can be seen as* an extension of, the first Ray–Knight theorem. *Indeed,* if X is a standard Brownian motion *started at x and killed upon hitting 0, then $\psi(\lambda) = \lambda^2/2$, and by continuity of the path, $H_t = X_t - I_t$ where I is the past infimum of X.*

Now thanks to Lévy's equivalence theorem, H has the law of a reflecting Brownian motion *killed upon summing up x units of local time at 0. According to the first Ray–Knight theorem, the local time process of H is indeed the CB-process with branching mechanism ψ, that is, the squared Bessel process with dimension 0 (see Subsection 4.5.1).*

If we try to push further the analysis of the relations between Lévy processes and trees to get new Ray–Knight type theorems, we can follow two tracks, having in common the focus on infinite trees.

1. the first possibility is to consider the JCCP of the *first infinite subtree* defined in the previous section, either in the supercritical case, or in the (sub)critical case when the tree is conditioned on reaching large heights

2. the second track relies on the JCCP of the *truncated tree.*

4.5.3 JCCP of the first infinite subtree

The supercritical case

In Subsection 4.4.1, we have seen that in the supercritical case under $\mathbb{P}(\cdot \mid \mathrm{Ext}^c)$, the JCCP of the first infinite subtree $\mathbb{F}_1$ is a Lévy process conditioned to stay positive, and Theorem 4.4.1

has provided us with a genealogical characterization of this tree. Indeed, as pointed out in the proof of this theorem, $\mathbb{F}_1$ is a collection of finite subtrees starting from R_k at generation k (see Fig. 4.2), and, conditional on R_k, these subtrees are i.i.d. with law $\mathbb{P}^\natural$. Moreover, the (R_k) are i.i.d. with common distribution

$$\mathbb{P}(R \in dr) = \bar{\Lambda}(r)e^{-\eta r}\, dr,$$

where $\bar{\Lambda}(r) = \int_{(r,\infty)} \Lambda(dz)$.

This can be recorded in the following statement, where X is assumed to be a Lévy process with finite variation and Laplace exponent $\psi(\lambda) = \lambda - \int_0^\infty (1-e^{-\lambda r})\, \Lambda(dr)$ such that $\psi'(0+) < 0$ (i.e., X drifts to $+\infty$), started at χ and *conditioned not to hit 0.*

The height process of X is defined as usual, and its local time process Z as in the beginning of this section, that is,

$$Z_n := \int_0^\infty dt\, \mathbf{1}_{\{H_t=n\}} \qquad n \geq 0,$$

since we work under $P(\cdot \mid T_0 = \infty)$. Recall from Theorem 4.5.1 that F is the Laplace exponent of a subordinator, given by $F(\lambda) = \int_0^\infty (1 - e^{-\lambda r})\, \Lambda(dr)$.

Theorem 4.5.3 *Under $P_\chi(\cdot \mid T_0 = \infty)$, the local time process $(Z_n; n \geq 0)$ of H is a* Jirina process with immigration. *It has branching mechanism $F^\natural$*

$$F^\natural(\lambda) := F(\lambda + \eta) \qquad \lambda \geq 0,$$

and immigration mechanism G

$$G(\lambda) := \frac{F(\lambda+\eta)}{\lambda+\eta} \qquad \lambda \geq 0.$$

It is started at the first residual lifetime R_0, which is an exponential variable with parameter η conditioned to be smaller than χ.

Proof. We prove the theorem in the case when Λ is finite. Set $\mathbb{I}_n$ the n-th thorn of $\mathbb{R}_1$, which has length R_n. From Theorem 4.4.1, for any $x \in \mathbb{I}_n$, the height of x is $|p_1(x)| = n$. Then adapting the proof of Theorem 4.5.1, we get that

$$Z_n = R_n + \sum_{v:\exists\sigma,(v,\sigma)\in\mathbb{F}_1\backslash\mathbb{R}_1,|v|=n} \zeta(v).$$

Next, thanks to Theorem 4.4.1 again, all vertices v with height n such that $(v,\sigma) \in \mathbb{F}_1\backslash\mathbb{R}_1$ for some σ, descend from the thorns with indices strictly lower than n and have lifespan measure $e^{-\eta r}\,\Lambda(dr)$ (thanks to Corollary 4.3.3), so that, conditional on $Z_{n-1} = z$, wr have $Z_n = R_n + S(z)$, where S is a subordinator independent from R_n and with Laplace exponent

$$F^\natural(\lambda) = \int_0^\infty e^{-\eta r}(1 - e^{-\lambda r})\,\Lambda(dr) = F(\lambda+\eta) \qquad \lambda \geq 0.$$

As a consequence, Z is a Jirina process with immigration, where the branching mechanism is $F^\natural$ and the immigration is given by the common distribution of (R_n), so that

$$G(\lambda) = \mathbb{E}(\exp(-\lambda R)) = \int_0^\infty \bar{\Lambda}(r)\, e^{-(\lambda+\eta)r}\, dr = \frac{F(\lambda+\eta)}{\lambda+\eta} \qquad \lambda \geq 0,$$

where the last equality follows from an elementary integration by parts. □

Actually, a generalization of this theorem can be found in [72]. Indeed, let X be a general Lévy process with no negative jumps and infinite variation, and Laplace exponent ψ. As in the previous subsection, it is possible to define its height H_t as the local time at 0 of $S^{(t)} - X^{(t)}$.

Theorem 4.5.4 ([72]) *Assume that X has infinite variation and drifts to $+\infty$ ($\psi'(0+) < 0$). Then the local time process of the height process H of the Lévy process X started at x and conditioned not to hit 0, is a CBI-process with branching mechanism $\psi^\natural$*

$$\psi^\natural(\lambda) = \psi(\lambda + \eta) \qquad \lambda \geq 0,$$

and immigration mechanism χ

$$\chi(\lambda) := \frac{\psi(\lambda + \eta)}{\lambda + \eta} \qquad \lambda \geq 0.$$

It is started from an exponential r.v. with parameter η conditioned to be smaller than x.

Remark 4.7 *The similarity between both theorems is obvious. However, note that in the finite case, $F^\natural$ is the Laplace exponent of a subordinator and G is the Laplace transform of a positive r.v., whereas in the infinite case, $\psi^\natural$ is the Laplace exponent of a general Lévy process with no negative jumps and χ is the Laplace exponent of a subordinator. Of course, the biggest difference lies in the proofs of these theorems. In the infinite variation case, the proof relies heavily on* excursion theory.

The critical case

Exactly as in the supercritical case, we will state a result that uses the spine decomposition given in the previous section for a splitting tree conditioned to reach large heights. We have seen that the limiting JCCP of the first infinite subtree is a Lévy process conditioned to stay positive. In the critical case, this process is defined by harmonic transform of the initial law of the Lévy process X via the martingale $X_t \mathbf{1}_{\{t<T_0\}}$ (see Appendix). The proof of the next statement relies on Theorem 4.4.2. Since this proof mimicks that for the supercritical case, we will not display it.

Let X be a Lévy process with finite variation and Laplace exponent $\psi(\lambda) = \lambda - \int_0^\infty (1 - e^{-\lambda r})\,\Lambda(dr)$ such that $\psi'(0+) = 0$, started at χ and *conditioned to stay positive.* Its law is denoted $P_\chi^\uparrow$. As usual, H denotes the height process of X.

Theorem 4.5.5 *Under $P_\chi^\uparrow$, the local time process $(Z_n; n \geq 0)$ of H is a* Jirina process with immigration. *It has branching mechanism F and immigration mechanism G such that $G(0) = 0$ and*

$$G(\lambda) := \frac{F(\lambda)}{\lambda} \qquad \lambda > 0.$$

It is started at the r.v. R_0, which is uniform on $(0, \chi)$.

Remark 4.8 *One might like to take better advantage of Theorem 4.4.2, since the previous statement only takes into account the first infinite subtree, which is the subtree to the* left-hand side of the spine. *It would indeed be possible to state a bivariate version of this theorem using the knowledge we have of the joint law of left-hand and right-hand sides of the spine, but this would require defining first a bivariate version of the Lévy process conditioned to stay positive. The interested reader will consult [29].*

As in the supercritical case, we state the infinite variation counterpart of the previous theorem. Let X be a Lévy process with no negative jumps and with infinite variation such that $\psi'(0+) = 0$, started at x and conditioned to stay positive, in the aforementioned sense (h-transform). Its law is denoted by $P_x^{\uparrow}$. Note that by local absolute continuity, it is still possible to define its height H_t as the local time at 0 of $S^{(t)} - X^{(t)}$.

Theorem 4.5.6 ([72]) *Under $P_x^{\uparrow}$, the local time process of the height process H of X, is a* CBI-process *with branching mechanism ψ and immigration mechanism χ such that $\chi(0) = 0$ and*

$$\chi(\lambda) := \frac{\psi(\lambda)}{\lambda} \qquad \lambda > 0.$$

It is started from a uniform r.v. on $(0, x)$ (from 0 if $x = 0$).

Remark 4.9 *This last theorem is not only a generalization of Theorem 4.5.5 to the infinite case, but also includes as a special case, and hence can be seen as* an extension of, the second Ray–Knight theorem. *Indeed, if X is a standard Brownian motion started at* 0 *and conditioned to stay positive, then $\psi(\lambda) = \lambda^2/2$, and it is known [101] that X is a* Bessel process of dimension 3 *started at 0. By continuity of the path, $H = X$ so by the second Ray–Knight theorem, the local time process of H is indeed a CBI-process with branching mechanism ψ and immigration mechanism $\chi(\lambda) = \psi(\lambda)/\lambda = \lambda/2$, that is, the squared Bessel process of dimension 2 (see Subsection 4.5.1).*

4.5.4 JCCP of the truncated tree

Truncations

Let $\mathbb{T}$ be a (possibly infinite in the supercritical case) splitting tree associated to the finite lifespan measure Λ and recall that $C_\tau(\mathbb{T})$ denotes the tree truncated at level τ

$$C_\tau(\mathbb{T}) = \{x \in \mathbb{T} : p_2(x) \leq \tau\}.$$

Recall from Section 4.2 that the JCCP $(X_t^{(\tau)}; t \geq 0)$ of the truncated splitting tree $C_\tau(\mathbb{T})$ is well defined. Next, if we define $H^{(\tau)}$ as the height process of $X^{(\tau)}$, and $Z^{(\tau)}$ as the local time process of $H^{(\tau)}$, that is, as usual,

$$Z_n^{(\tau)} := \int_0^{T_0} dt\, \mathbf{1}_{\{H_t^{(\tau)} = n\}} \qquad n \geq 0,$$

then, by a similar proof as that of Theorem 4.5.1,

$$Z_n^{(\tau)} = \sum_{v:|v|=n, \omega(v) \leq \tau} \zeta(v) + \sum_{v:|v|=n, \alpha(v) \leq \tau < \omega(v)} (\tau - \alpha(v)).$$

The integers n (depending on τ) for which the second sum in the last displayed equation is not empty are called *incomplete generations.*

The presence of incomplete generations impedes us from getting a new Ray–Knight type theorem for a fixed τ, but only as $\tau \to \infty$. Indeed, it is easily seen that the lowest incomplete generation tends to ∞ as $\tau \to \infty$, so that $(Z_n^{(\tau)}; n \geq 0)$ converges for finite-dimensional distributions, to the Jirina process with branching mechanism $F(\lambda) = \int_0^\infty (1 - e^{-\lambda r})\, \Lambda(dr)$. To state this in a useful way, we construct the process $(X_t^{(\tau)}; t, \tau \geq 0)$ without further reference to the tree.

Construction of doubly indexed JCCP

Let ψ be defined as the usual Laplace exponent $\lambda - F(\lambda)$ and χ be the lifespan of the ancestor. As seen in Theorem 4.3.1, the JCCP $(X_t^{(\tau)}; t \geq 0)$ of the truncated splitting tree $C_\tau(\mathbb{T})$ is the concatenation of a sequence of i.i.d. excursions, distributed as the Lévy process with Laplace exponent ψ killed at $T_0 \wedge T_{(\tau,+\infty)}$, and ending at the first excursion hitting 0 before $(\tau, +\infty)$. These excursions all start at τ, but the first one, which starts at $\min(\chi, \tau)$.

One can see $X^{(\tau)}$ as a Lévy process *reflected below* τ. Actually, there are various ways of reflecting a real-valued path below or above a constant barrier. One of the possibilities is as follows.

We will use also the notation C_τ for the map which associates to a real-valued path ϵ the path $C_\tau(\epsilon)$ obtained after cutting the excursions of ϵ above τ. Rigorously, set

$$A_t^\tau(\epsilon) = \int_0^t ds\, \mathbf{1}_{\{\epsilon_s < \tau\}} \qquad t \geq 0,$$

and let $a^\tau(\epsilon)$ be its right inverse

$$a_t^\tau(\epsilon) = \inf\{s : A_s^\tau(\epsilon) > t\} \qquad t \geq 0.$$

Then the truncation of ϵ at level τ is defined as

$$C_\tau(\epsilon) = \epsilon \circ a^\tau(\epsilon).$$

A problem with this procedure is for *transient* paths. Indeed, if ϵ_t goes to ∞ as $t \to \infty$, then $a^\tau(\epsilon)$ is only defined on $[0, S_\tau)$, where S_τ is the last hitting time of τ, and so is $C_\tau(\epsilon)$. On the contrary, here, we aim at constructing paths which end up at 0, and are reflected as many times as necessary to hit 0.

The right way to do this is exactly as in the theorem, by concatenating independent excursions below τ, of the Lévy process X with Laplace exponent ψ. In the infinite variation case, one has to use Itô's synthesis theorem [56]. But in reality, we can do that in the same fashion for both finite and infinite variation cases, by considering S the past supremum of a general Lévy process X with no negative jumps after it has overshot level τ

$$S_t^{(\tau)} := \sup\{X_s : s \leq t\} \vee \tau.$$

Definition 4.5.7 *The process $\tilde{X}^{(\tau)}$ defined as*

$$\tilde{X}_t^{(\tau)} := X_t - S_t^{(\tau)} + \tau \qquad t \geq 0,$$

will be called the Lévy process X reflected below *τ.*

Actually, since we want to let $\tau \to \infty$, we need a joint construction of the Lévy process reflected below τ for all positive τ. In particular, notice that by construction of the truncated trees in the beginning of this subsection, we had for any $\tau < \tau'$,

$$C_\tau(X^{(\tau')}) = X^{(\tau)}.$$

Proposition 4.5.8 *There is a doubly indexed process* $(\tilde{X}_t^{(\tau)}; t, \tau \geq 0)$ *such that for each* τ, $\tilde{X}^{(\tau)}$ *has the law of the Lévy process* X *reflected below* τ *and for any* $\tau < \tau'$,

$$C_\tau(\tilde{X}^{(\tau')}) = \tilde{X}^{(\tau)}.$$

Proof. By standard results on Lévy processes, for any $\tau < \tau'$,

$$C_\tau(\tilde{X}^{(\tau')}) \stackrel{\mathcal{L}}{=} \tilde{X}^{(\tau)}.$$

Then the proposition stems from Kolmogorov's existence theorem. □

Theorem 4.5.9 *In the finite variation case, the family, indexed by* τ, *of the processes* $(\tilde{X}_t^{(\tau)}; t \geq 0)$ *all killed upon hitting 0, have the same law as the family of JCCPs* $(X^{(\tau)})_\tau$ *of the splitting tree with lifespan measure* Λ *and truncated at level* τ.

If $Z_n^{(\tau)}$ *stands for the time spent at level* n *by the height process* $H^{(\tau)}$ *of* $\tilde{X}^{(\tau)}$, *then as* $\tau \to \infty$, $(Z_n^{(\tau)}; n \geq 0)$ *converges for finite-dimensional distributions to the* Jirina process *with branching mechanism* F.

The infinite variation case

In the infinite variation case, Proposition 4.5.8 still holds, and it should be possible to state a similar theorem as the previous one (work in progress).

In addition, there is a case when there are no incomplete generations, so that it is not necessary to let $\tau \to \infty$ to get a proper genealogy. This occurs when the path of the Lévy process is continuous, since then there are no individuals who were born before τ and will die after τ. *The following theorem is currently only a conjecture.* In this statement, reflection above 0 means reflection on the infimum, and reflection below τ means reflection on the supremum (as in Definition 4.5.7).

Theorem 4.5.10 *Let* B *be a Brownian motion with drift reflected* above *0* and below τ, *and killed upon summing up* x *units of local time at 0. Then the local time process* $(Z_t; t \in [0, \tau])$ *of* B *is a CB-process with branching mechanism* ψ *started at* x.

Remark 4.10 *This theorem should be viewed as a* third Ray–Knight theorem.

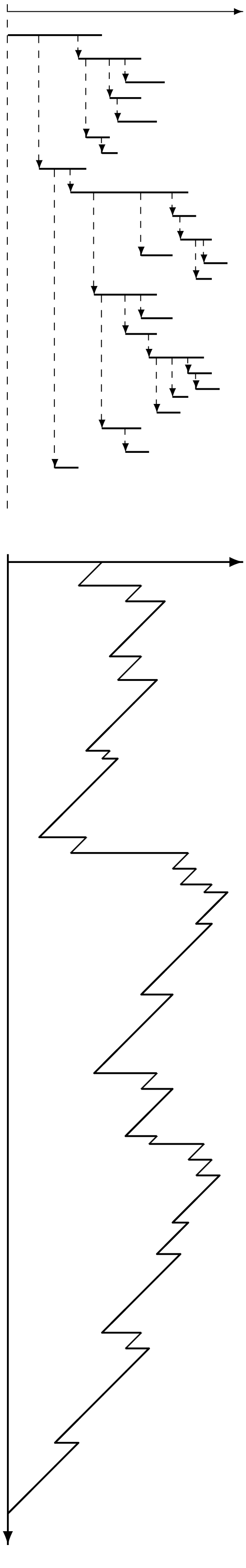

Figure 4.3: A splitting tree and the associated jumping chronological contour process (JCCP).

Chapter 5

Coalescence and Structured Populations

Appendix

We denote by $(X_t; t \geq 0)$ a Lévy process (i.e. a process with independent and homogeneous increments with a.s. càdlàg paths) with no negative jumps, and by P_x its distribution conditional on $X_0 = x$. The following facts can all be found in [11, 71, 91].

The (convex) *Laplace exponent* ψ of X is defined by

$$E_0(\exp(-\lambda X_t)) = \exp(t\psi(\lambda)) \qquad t, \lambda \geq 0,$$

and is specified by the Lévy–Khinchin formula

$$\psi(\lambda) = \alpha\lambda + \beta\lambda^2 + \int_0^\infty (e^{-\lambda r} - 1 + \lambda r \mathbf{1}_{r<1})\, \Lambda(dr) \qquad \lambda \geq 0, \tag{5.1}$$

where $\alpha \in \mathbb{R}$, $\beta \geq 0$ denotes the *Gaussian coefficient*, and Λ is a *Lévy measure* on $(0, \infty)$, that is, a positive measure such that $\int_0^\infty (r^2 \wedge 1)\Lambda(dr) < \infty$.

The paths of X have *finite variation* a.s. if and only if $\beta = 0$ and $\int_0^1 r\Lambda(dr) < \infty$. Otherwise its paths have *infinite variation* a.s.

When X has increasing paths a.s., it is called a *subordinator*. In that case, $\psi(\lambda) < 0$ for any positive λ, and we will prefer to define its Laplace exponent as $-\psi$. Indeed, since a subordinator has finite variation, its Laplace exponent can be written as

$$-\psi(\lambda) = \mathrm{d}\lambda + \int_0^\infty (1 - e^{-\lambda r})\, \Lambda(dr) \qquad \lambda \geq 0,$$

where $\mathrm{d} \geq 0$ is called the *drift coefficient*. If X is not a subordinator, since ψ is convex,

$$\lim_{\lambda\to\infty} \psi(\lambda) = +\infty.$$

Set $S_t := \sup_{s \leq t} X_s$ the past supremum of X. If X is not a subordinator but has finite variation, the zero set of $S - X$ is discrete. In addition, $\psi(\lambda) = a\lambda - \int_0^\infty (1 - e^{-\lambda r})\, \Lambda(dr)$ for some positive a, so that $\psi(\lambda) \leq a\lambda$, and $\int^\infty 1/\psi$ diverges (from this observation stems the fact that CB-processes that hit 0 have infinite variation).

When X has infinite variation, $S - X$ is a strong Markov process for which 0 is regular, that is,

$$\inf\{t > 0 : S_t - X_t = 0\} = 0 \qquad P - a.s.$$

When X is not a subordinator, denote by η the largest root of ψ

$$\eta := \sup\{\lambda : \psi(\lambda) = 0\}.$$

If $\eta > 0$, ψ has exactly two roots (0 and η), otherwise ψ has a unique root $\eta = 0$. The right-inverse of ψ is denoted by $\phi : [0,\infty) \to [\eta,\infty)$ and has $\phi(0) = \eta$.

We write

$$T_A = \inf\{t \geq 0 : X_t \in A\},$$

for the first entrance time in a Borel set A. It is known that

$$E_0(e^{-qT_{\{x\}}}) = e^{\phi(q)x} \qquad q \geq 0, x \leq 0.$$

There exists a unique continuous function $W : [0,+\infty) \to [0,+\infty)$, with Laplace transform

$$\int_0^\infty e^{-\lambda x} W(x)dx = \frac{1}{\psi(\lambda)} \qquad \lambda > \eta,$$

such that for any $0 \leq x \leq a$,

$$P_x(T_0 < T_{(a,+\infty)}) = W(a-x)/W(a). \tag{5.2}$$

The function W is strictly increasing and called the *scale function*.

We introduce two laws connected with P.

1. The probability measure $P_x^\uparrow$ is the law of the Lévy process started at $x > 0$ and *conditioned to stay positive*. When X drifts to $+\infty$, the conditioning is taken in the usual sense, since X stays positive with positive probability. When X oscillates, as X reaches 0 continuously, the process $(X_t \mathbf{1}_{\{t<T_0\}}, t \geq 0)$ is a martingale. Then $P_x^\uparrow$ is defined by local absolute continuity w.r.t. P_x with density $x^{-1}X_t\mathbf{1}_{\{t<T_0\}}$ on $\mathcal{F}_t$ $(t \geq 0)$.

   It is known that the probability measures $P_x^\uparrow$ converge weakly as $x \to 0+$ to a Markovian law $P_0^\uparrow$. For details, see [9, 10, 24, 25, 26].

2. The law $P^\natural$ is that of a spectrally positive Lévy process with Laplace exponent $\psi^\natural : \lambda \mapsto \psi(\lambda + \eta)$. When $\eta > 0$, the path of X a.s. drifts to $+\infty$, and if I_∞ denotes its overall infimum (Lemma VII.7(i) in [11]), then for any $t \geq 0$,

   $$\lim_{x\to\infty} P(\Theta \mid I_\infty > -x) = P^\natural(\Theta) \qquad \Theta \in \mathcal{F}_t.$$

   This process is thus called the Lévy process *conditioned to drift to* $-\infty$, and for every real number x, $P_x^\natural$ is defined by local absolute continuity w.r.t. P_x with density $\exp(-\eta(X_t - x))$ on $\mathcal{F}_t$ $(t \geq 0)$.

Bibliography

[1] Aldous, D.J. (1999)
Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean–field theory for probabilists. *Bernoulli* **5**(1) 3–48.

[2] Aldous, D.J., Popovic, L. (2004)
A critical branching process model for biodiversity *Adv. Appl. Prob.* **37**(4) 1094–1115.

[3] Anderson, W.J. (1991)
Continuous time Markov chains – an application oriented approach. Springer–Verlag, New York.

[4] Asmussen, S., Hering, H. (1983)
Branching processes. Birkäuser, Boston.

[5] Athreya, K.B. (2000)
Change of measures for Markov chains and the $L \log L$ theorem for branching processes. *Bernoulli* **6** 323–338.

[6] Athreya, K.B., Ney P.E. (1972)
Branching processes. Springer–Verlag, New York.

[7] Axelrod, D.E., Kimmel, M. (2002)
Branching processes in biology. Springer–Verlag, New York.

[8] Bansaye, V. (2006)
Proliferating parasites in dividing cells : Kimmel's branching model revisited. Preprint.

[9] Bertoin, J. (1991)
Sur la décomposition de la trajectoire d'un processus de Lévy spectralement positif en son infimum. *Ann. Inst. H. Poincaré* **27** 537–547.

[10] Bertoin, J. (1992)
An extension of Pitman's theorem for spectrally positive Lévy processes. *Ann. Probab.* **20** 1464–1483.

[11] Bertoin, J. (1996)
Lévy Processes. Cambridge U. Press.

[12] Bertoin, J. (2006)
Random Fragmentation and Coagulation Processes. Cambridge U. Press.

[13] Bertoin, J., Le Gall, J.F. (2000)
The Bolthausen–Sznitman coalescent and the genealogy of continuous–state branching processes. *Probab. Theory Relat. Fields* **117**(2) 249–266.

[14] Bertoin, J., Le Gall, J.F. (2003)
Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126**(2) 261–288.

[15] Bertoin, J., Le Gall, J.F. (2005)
Stochastic flows associated to coalescent processes II: Stochastic differential equations. *Ann. Inst. Henri Poincaré* **41** 307–333.

[16] Bertoin, J., Le Gall, J.F. (2006)
Stochastic flows associated to coalescent processes II: Limit theorems. *Illinois J. Math.* In press.

[17] Bingham, N.H. (1976)
Continuous branching processes and spectral positivity. *Stoch. Proc. Appl.* **4** 217–242.

[18] Birkner, M. (2005)
Stochastic models from population biology. Lecture notes for a course at TU Berlin. http://www.wias-berlin.de/people/birkner/smpb-30.6.05.pdf

[19] Cannings, C. (1974)
The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. Appl. Prob.* **6** 260–290.

[20] Cattiaux, P., Collet, P., Lambert, A., Martinez, S., Méléard, S., San Martin, J. (2006)
Quasi–stationary distributions and diffusion models in population dynamics. In preparation.

[21] Champagnat, N. (2006)
A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.* **116** 1127–1160.

[22] Champagnat, N., Ferrière, R., Méléard S. (2006)
Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theor. Popul. Biol.* **69** 297–321.

[23] Champagnat, N., Lambert, A. (2006)
Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Prob.* (in press)

[24] Chaumont, L. (1994)
Sur certains processus de Lévy conditionnés à rester positifs. *Stochastics Stoch. Reports* **47** 1–20.

[25] Chaumont, L. (1996)
Conditionings and path decompositions for Lévy processes. *Stoch. Proc. Appl.* **64** 39–54.

[26] Chaumont, L., Doney, R.A. (2005)
On Lévy processes conditioned to stay positive. *Electron. J. Probab.* **10** 948–961.

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

[28] Darwin, C. (1859)
On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.

[29] Duquesne, T. (2006)
Continuum random trees and branching processes with immigration. *Stoch. Proc. Appl.* In press.

[30] Duquesne, T. (2006)
The coding of compact real trees by real valued functions. Preprint.

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

[32] Duquesne, T., Winkel, M. (2006)
Growth of Lévy trees. *Probab. Theory Relat. Fields.* In press.

[33] Durrett, R. (2002)
Probability Models for DNA Sequence Evolution. Springer–Verlag, Berlin.

[34] Durrett, R. (2006)
Random Graph Dynamics. Cambridge U. Press.

[35] Erdös, P., Rényi, A. (1960)
On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5** 17—61.

[36] Etheridge, A.M. (2004)
Survival and extinction in a locally regulated population. *Ann. Appl. Prob.* **14** 188–214.

[37] Ethier, S. N., Kurtz, T. G. (1986)
Markov Processes, Characterization And Convergence. John Wiley & Sons, New York.

[38] Evans, S.N. (2005)
Probability and real trees. École de Probabilités de St–Flour.

[39] Ewens, W.J. (2005)
Mathematical Population Genetics. 2nd edition, Springer–Verlag, Berlin.

[40] Feller, W. (1951)
Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 227–246.

[41] Fisher, R.A. (1922)
On the dominance ratio. *Proc. Roy. Soc. Edin.* **42** 321–341.

[42] Fisher, R.A. (1930)
The distribution of gene ratios for rare mutations. *Proc. Roy. Soc. Edin.* **50** 205–220.

[43] Galton, F. (1894)
A plausible paradox in chances. *Nature* **49** 365–366.

[44] Geiger, J., Kersting, G. (1997)
Depth–first search of random trees, and Poisson point processes *in Classical and modern branching processes* (Minneapolis, 1994) IMA Math. Appl. Vol. 84. Springer–Verlag, New York.

[45] Gillespie, J.H. (1998)
Population Genetics: A Concise Guide. Johns Hopkins U. Press, Baltimore, MD.

[46] Gosselin, F. (2001)
Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *Ann. Appl. Prob.* **11** 261–284.

[47] Grey, D.R. (1974)
Asymptotic behaviour of continuous–time, continuous state–space branching processes. *J. Appl. Prob.* **11** 669–677.

[48] Haccou, P., Jagers, P., Vatutin, V.A. (2005)
Branching Processes. Variation, Growth, and Extinction of Populations. Series: Cambridge Studies in Adaptive Dynamics (No. 5). Cambridge U. Press.

[49] Haldane, J.B.S. (1927)
A mathematical theory of natural and artificial selection. Part V : Selection and mutation. *Proc. Camb. Phil. Soc.* **23** 838–844.

[50] Harris, T.E. (1963)
The theory of branching processes. Springer–Verlag, Berlin.

[51] Heathcote, C. R. (1966)
Corrections and comments on the paper "A branching process allowing immigration". *J. Roy. Statist. Soc. Ser. B* **28** 213–217.

[52] Heathcote, C. R., Seneta, E., Vere-Jones, D. (1967)
A refinement of two theorems in the theory of branching processes. *Teor. Verojatnost. i Primenen.* **12** 341–346.

[53] Heyde, C.C. (1970)
Extension of a result of Seneta for the super–critical Galton–Watson process. *Ann. Math. Statist.* **41** 739–742.

[54] Heyde, C.C., Seneta, E. (1972)
The simple branching process, a turning point test and a fundamental inequality: a historical note on I.J. Bienaymé. *Biometrika* **59** 680–683.

[55] Hirano, K. (2001)
Lévy processes with negative drift conditioned to stay positive. *Tokyo J. Math.* **24** 291–308.

[56] Itô, K. (1971)
Poisson point processes attached to Markov processes. *Proc. Sixth Berkeley Symp.* vol.III 225–240. U. of California Press.

[57] Jagers, P. (1975)
Branching processes with biological applications. John Wiley & Sons, London New York Sydney.

[58] Jirina, M. (1958)
Stochastic branching processes with continuous state space. *Czech. Math. J.* **8** 292–312.

[59] Kahneman, D., Slovic, P., Tversky, A. (1993)
Judgment Under Uncertainty– Heuristics and Biases. Cambridge U. Press.

[60] Kallenberg, O. (1973)
Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **27** 23–36.

[61] Kallenberg, P.J.M. (1979)
Branching processes with continuous state space. Mathematisch Centrum, Amsterdam.

[62] Kawazu, K., Watanabe, S. (1971)
Branching processes with immigration and related limit theorems. *Teor. Verojanost i Primenen* **16** 34–51.

[63] Kendall, D.G. (1975)
The genalogy of genealogy: branching processes before (and after) 1873. *Bull. London Math. Soc.* **7** part 3 (21) 225–253.

[64] Kesten, H., Ney, P., Spitzer, F. (1966)
The Galton–Watson process with mean one and finite variance. *Teor. Verojatnost. i Primenen.* **11** 579–611.

[65] Kesten, H., Stigum, B.P. (1966)
A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** part 3 (21) 1211–1223.

[66] Kimmel, M. (1997)
Quasistationarity in a branching model of division–within–division. In *Classical and modern branching processes* (Minneapolis, 1994) IMA Math. Appl. Vol. 84. Springer–Verlag, New York.

[67] Kimura, M. (1957)
Some problems of stochastic processes in genetics. *Ann. Math.* **28** 882–901.

[68] Kingman, J.F.C. (1982)
The coalescent. *Stochastic Process. Appl.* **13**(3) 235–248.

[69] Kolmogorov, A.N. (1938)
Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ.*, Tomsk **2** 1–6.

[70] Kot, M. (2001)
Elements of Mathematical Ecology. Cambridge U. Press.

[71] Kyprianou, A.E. (2006)
Introductory lectures on fluctuations of Lévy processes with applications. Springer–Verlag, Berlin Heidelberg.

[72] Lambert, A. (2002)
The genealogy of continuous–state branching processes with immigration. *Probab. Theory Relat. Fields* **122** (1): 42–70.

[73] Lambert, A. (2003)
Coalescence times for the branching process. *Adv. Appl. Prob.* **35** (4): 1071–1089.

[74] Lambert, A. (2005)
The branching process with logistic growth. *Ann. Appl. Prob.* **15** 1506–1535.

[75] Lambert, A. (2006
Probability of fixation under weak selection: a branching process unifying approach. *Theor. Popul. Biol.* **69** 419–441.

[76] Lambert, A. (2006)
Quasi–stationary distributions and the continuous–state branching process conditioned to be never extinct. Submitted.

[77] Lambert, A. (2006)
Contour of splitting trees and new properties of homogeneous Crump–Mode–Jagers processes. In preparation.

[78] Lamperti, J. (1967)
Continuous–state branching processes. *Bull. Amer. Math. Soc.* **73** 382–386.

[79] Le Gall, J.F. (2005)
Random trees and applications. *Probab. Surv.* **2** 245–311.

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

[81] Lyons, R., Pemantle, R., Peres, Y. (1995)
Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** 1125–1138.

[82] Lyons, R., Peres, Y. (2006)
Probability on trees and networks, a book in progress.
http://mypage.iu.edu/ rdlyons/prbtree/prbtree.html

[83] Malthus, T.R. (1798)
An Essay on the Principle of Population. J. Johnson, London.

[84] Neveu, J. (1986)
Arbres et processus de Galton–Watson. *Ann. Inst. H. Poincaré* **22** 199–207.

[85] Pakes, A.G. (1998)
Conditional limit theorems for continuous time and state branching processes. Technical report.

[86] Pinsky, M.A. (1972)
Limit theorems for continuous state branching processes with immigration. *Bull. Amer. Math. Soc.* **78** 242–244.

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

[88] Popovic, L. (2004)
Asymptotic genealogy of a critical branching process. *Ann. Appl. Prob.* **14** (4) 2120—2148.

[89] Renshaw, E. (1991)
Modelling biological populations in space and time. Cambridge Studies in Mathematical Biology, 11. Cambridge U. Press.

[90] Revuz, D., Yor, M. (1999)
Continuous Martingales and Brownian Motion. (3rd revised ed.) Springer–Verlag, Berlin Heidelberg New York.

[91] Sato, K.I. (1999)
Lévy processes and infinitely divisible distributions. Cambridge studies in advanced mathematics Vol. 68. Cambridge U. Press.

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

[93] Seneta, E. (1970)
On the supercritical Galton–Watson process with immigration. *Math. Biosci.* **7** 9–14.

[94] Seneta, E., Vere-Jones, D. (1966)
On quasi–stationary distributions in discrete–time Markov chains with a denumerable infinity of states. *J. Appl. Prob.* **3** 403–434.

[95] Sevast'yanov, B.A. (1971)
Branching processes. Nauka, Moscow.

[96] Silverstein, M. L. (1967–68)
A new approach to local times. *J. Math. Mech.* **17** 1023–1054.

[97] Vere-Jones, D. (1967)
Ergodic properties of nonnegative matrices. I. *Pacific J. Math.* **22** 361–386.

[98] Vere-Jones, D. (1968)
Ergodic properties of nonnegative matrices. II. *Pacific J. Math.* **26** 601–620.

[99] Tuominen, P., Tweedie, R.L. (1979)
Exponential decay and ergodicity of general Markov processes and their discrete skeletons. *Adv. Appl. Probab.* **11**(4) 784—803.

[100] Watson, H.W., Galton, F. (1874)
On the probability of the extinction of families. *J. Anthropol. Inst. Great Britain and Ireland* **4** 138–144.

[101] Williams, D. (1974)
Path decomposition and continuity of local time for one dimensional diffusions. *Proc. London Math. Soc.* **28**(4) 738—768.

[102] Yaglom, A.M. (1947)
Certain limit theorems of the theory of branching stochastic processes (in Russian). *Dokl. Akad. Nauk. SSSR* (n.s.) **56** 795–798.