

Some aspects of discrete branching processes *

AMAURY LAMBERT[†]

April 2010

Abstract

We first define Markovian discrete branching processes in discrete time, known as Bienaymé-Galton-Watson (BGW) processes. We review some basic results like the extinction-expansion dichotomy, computation of the one-dimensional marginals using generating functions, as well as of the extinction probability. We introduce the BGW process with immigration and state three limit theorems.

We define the notion of quasi-stationarity for Markov chains and provide the basic results in the case of a finite-state space. In the case of BGW processes, we characterize Yaglom quasi-stationary limit (one-dimensional distribution conditional on non extinction) and the Q -process (process conditioned on non-extinction in the distant future).

Then we show how to code the genealogy of a BGW tree thanks to a killed random walk. The law of the total progeny of the tree is studied thanks to this correspondence. Alternative proofs are given via Dwass-Kemperman identity and the ballot theorem.

Last, we introduce the coalescent point process of branching trees : on a representation of the quasi-stationary genealogy of an infinite population, which is also doubly infinite in time; on splitting trees, which are those trees whose width process is a (generally not Markovian) binary, homogeneous Crump-Mode-Jagers process.

For the interested reader, the following books are classical references on branching models and random trees : [4, 5, 8, 11, 15, 16, 21].

Contents

| | | |
|----------|---|----------|
| 1 | The Bienaymé-Galton-Watson model | 2 |
| 1.1 | Definitions | 2 |
| 1.2 | Some computations | 5 |
| 1.2.1 | Exercises | 5 |
| 1.2.2 | Examples | 5 |
| 1.3 | BGW process with immigration | 5 |
| 1.4 | Kesten-Stigum theorem | 6 |

*This document will serve as a support for 4 lectures given at the CIMPA school in April 2010 in Saint-Louis, Senegal. Sections 1 and 2 contain previously published material [28].

[†]Université Paris 6 Pierre et Marie Curie, France.

| | | |
|----------|--|-----------|
| 2 | Quasi-stationarity | 7 |
| 2.1 | Quasi-stationarity in Markov chains | 7 |
| 2.2 | Finite state-space and Perron–Frobenius theory | 7 |
| 2.3 | Quasi-stationarity in BGW processes | 8 |
| 2.3.1 | Yaglom limit | 8 |
| 2.3.2 | The Q -process | 10 |
| 3 | Random walks and BGW trees | 11 |
| 3.1 | The Lukasiewicz path | 11 |
| 3.2 | Kemperman’s formula | 13 |
| 3.2.1 | Cyclic shifts | 13 |
| 3.2.2 | Lagrange inversion formula | 15 |
| 3.2.3 | Dwass proof | 15 |
| 4 | Coalescent point processes | 17 |
| 4.1 | The coalescent point process of quasi-stationary BGW trees | 17 |
| 4.1.1 | A doubly infinite population | 17 |
| 4.1.2 | The coalescent point process | 18 |
| 4.1.3 | Disintegration of the quasi-stationary distribution | 19 |
| 4.2 | Splitting trees | 20 |

1 The Bienaymé–Galton–Watson model

Historically, the Bienaymé–Galton–Watson model was the first stochastic model of population dynamics [20, 43].

1.1 Definitions

Assume we are given the law of a random integer ξ

$$p_k := \mathbb{P}(\xi = k) \quad k \geq 0,$$

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 tree are given by the following rules (see Fig. 1).

- generation $n + 1$ is made up of the *offspring* of individuals from generation n
- 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 .

The Markov chain $(Z_n; n \geq 0)$ is called the *BGW process*. It contains less information than the whole *BGW tree*, which provides the whole 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

$$Z(x + y) \stackrel{\mathcal{L}}{=} Z(x) + \tilde{Z}(y), \quad (1)$$

where \tilde{Z} is an independent copy of Z . In general, stochastic processes that satisfy (1) are called *branching processes*.

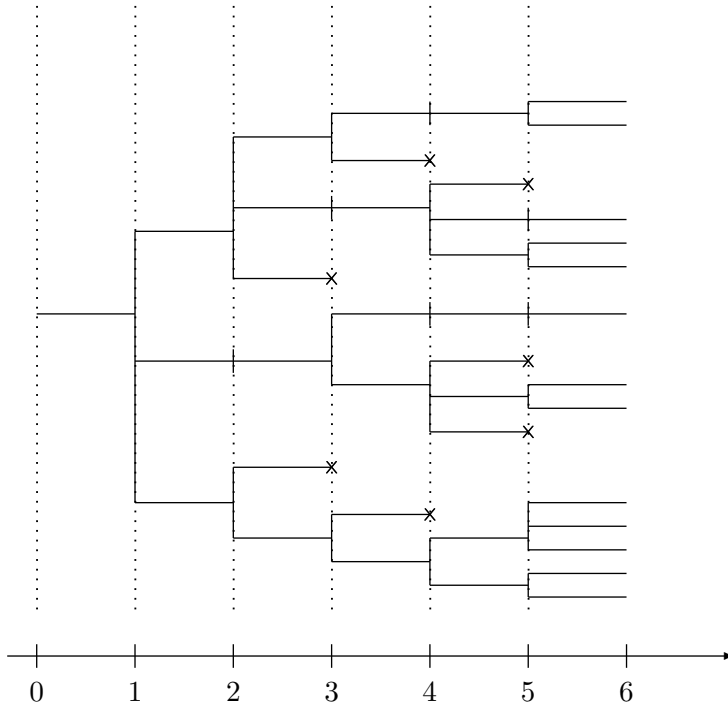


Figure 1: A Bienaymé–Galton–Watson tree through 6 generations, starting from one ancestor (crosses mean death with zero offspring; vertical axis has no meaning).

It is convenient to consider the generating function f of ξ (see Fig. 2)

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

as well as its expectation

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

We write \mathbb{P}_z for the conditional probability measure $\mathbb{P}(\cdot \mid Z_0 = z)$. Unless otherwise specified, \mathbb{P} will stand for \mathbb{P}_1 .

Proposition 1.1 *The generating function of Z_n is given by*

$$\mathbb{E}_z(s^{Z_n}) = f_n(s)^z \quad s \in [0, 1], \quad (2)$$

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

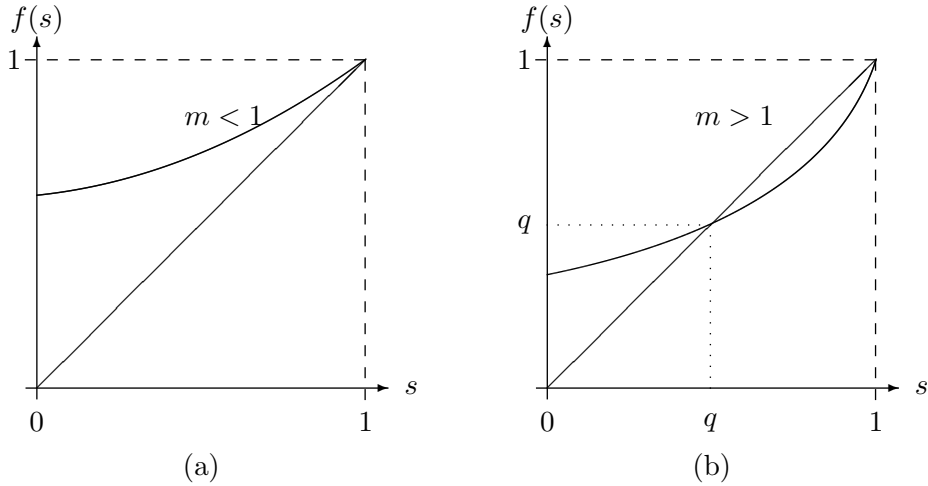


Figure 2: Graph of the probability generating function f (a) for a subcritical BGW tree (b) for a supercritical BGW tree, with extinction probability q shown.

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

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

Iterating the last displayed equation then yields (2). Differentiating (2) w.r.t. s and letting s tend to 1, gives the formula for the expectation. \square

We say that *extinction* occurs if Z hits 0, and denote $\{\text{Ext}\}$ 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 (see Fig. 2).

Definition 1.2 A BGW tree is said subcritical if $m < 1$, critical if $m = 1$, and supercritical if $m > 1$.

Theorem 1.3 We have

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

which equals 1 in the subcritical and critical cases. In the supercritical case, $\lim_{n \rightarrow \infty} Z_n = +\infty$ conditionally on non-extinction a.s.

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

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

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

Thanks to (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. \square

1.2 Some computations

1.2.1 Exercises

Exercise 1 Assuming that $\sigma^2 := \text{Var}(\xi)$ is finite, prove that

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

Exercise 2 Assume $m > 1$. Show that conditional on $\{\text{Ext}\}$, Z has the same law as the subcritical branching process Z^* with offspring distribution $p_k^* = q^{k-1}p_k$, whose generating function is

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

This subcritical branching process is called the dual process.

Exercise 3 State and prove a similar result for the subtree of infinite lines of descent conditional on $\{\text{Ext}^c\}$.

1.2.2 Examples

Binary. Assume $p_k = 0$ for all $k \geq 3$, and call this model $\text{Binary}(p_0, p_1, p_2)$. 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 $\text{Binary}(p_2, p_1, p_0)$.

Geometric. Assume $p_k = (1-a)a^k$, and call this model $\text{Geometric}(a)$. 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 $\text{Geometric}(1-a)$.

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

Linear fractional. Assume $p_0 = b$ and $p_k = (1-b)(1-a)a^{k-1}$ for $k \geq 1$. Call this model $\text{LF}(a, b)$. 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 $\text{LF}(b, a)$. This example has very interesting features, see e.g. section I.4 of [4].

1.3 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 tree 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

$$\mathbb{E}_z(s^{Z_n}) = f_n(s)^z \prod_{k=0}^{n-1} g \circ f_k(s) \quad s \in [0, 1]. \quad (3)$$

The following theorem concerns the asymptotic growth of subcritical BGW processes with immigration.

Theorem 1.4 (Heathcote [17]) *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 1.5 (Seneta [39]) *Assume $m > 1$. Then the following dichotomy holds*

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

For recent proofs of these theorems, see [34], where they are inspired from [2]. For yet another formulation, see [28].

1.4 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, so it converges a.s. to a nonnegative random variable W

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

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 1.6 (Kesten–Stigum [25]) *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) (W_n) converges in L^1
- (iv) $\mathbb{E}(\xi \log^+ \xi) < \infty$.

For a recent generalization of this theorem to more dimensions, see [3]. For a recent proof, see [33]. Results for (sub)critical processes on the asymptotic decay of $\mathbb{P}(Z_n \neq 0)$, as $n \rightarrow \infty$, will be displayed in the following section.

Exercise 4 Prove that $(W_n; n \geq 0)$ indeed is a martingale and that $\mathbb{P}(W = 0)$ is a fixed point of f . Under the assumption $\mathbb{E}(\xi^2) < \infty$, show that (W_n) is bounded in L^2 , and deduce (iii), (ii) and (i) (this was done by Kolmogorov in 1938 [27]).

2 Quasi-stationarity

2.1 Quasi-stationarity in Markov chains

Let X be a Markov chain on the non-negative integers for which 0 is one (and the only) *accessible absorbing state*. Then the only *stationary* probability of X is the Dirac mass at 0. A *quasi-stationary distribution* (QSD) is a probability measure ν satisfying

$$\mathbb{P}_\nu(X_n \in A \mid X_n \neq 0) = \nu(A) \quad n \geq 0. \quad (4)$$

Hereafter, we will only consider quasi-stationary *probabilities*. A quasi-stationary distribution may not be unique, but a specific candidate is defined (if it exists) as the law of Υ , where

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

for some 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, attributed to A.M. Yaglom [44]. Then by application of the simple Markov property,

$$\mathbb{P}_\nu(T \geq n + k) = \mathbb{P}_\nu(T \geq n)\mathbb{P}_\nu(T \geq k),$$

so that the extinction time T under \mathbb{P}_ν has a *geometric* distribution. Further set

$$\mathbb{P}_x^\uparrow(\Theta) := \lim_{k \rightarrow \infty} \mathbb{P}_x(\Theta \mid X_{n+k} \neq 0),$$

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

2.2 Finite state-space and Perron–Frobenius theory

A comprehensive account on the applications of Perron–Frobenius theory to Markov chains is the book by E. Seneta [40].

Let X be a Markov chain on $\{0, 1, \dots, N\}$ that has two communication classes, namely $\{0\}$ and $\{1, \dots, N\}$. We assume further that 0 is accessible from the other class. 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 N obtained from P by deleting its first row and its first column. In particular,

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

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

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

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

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

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

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

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

Exercise 5 *Prove the previous statement using the following key result in the Perron–Frobenius theorem*

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

2.3 Quasi-stationarity in BGW processes

2.3.1 Yaglom limit

In the BGW case, the most refined result for the Yaglom limit is the following, where we recall that m stands for the offspring mean. Here Q refers to the truncated transition matrix of the BGW process, i.e., indexed by the positive integers.

Theorem 2.2 (Yaglom [44], Sevast’yanov [42], Heathcote–Seneta–Vere-Jones [18]) *In the subcritical case, there is a r.v. Υ with probability distribution $(u_j, j \geq 1)$ such that $uQ = mu$ and*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = j \mid Z_n \neq 0) = u_j \quad j \geq 1.$$

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

$$\lim_{n \rightarrow \infty} \mathbb{E}_x(Z_n \mid Z_n \neq 0) = \mathbb{E}(\Upsilon).$$

Exercise 6 Prove that the equation $uQ = mu$ also reads

$$1 - g(f(s)) = m(1 - g(s)) \quad s \in [0, 1],$$

where g is the probability generating function of Υ . Then define for any $\alpha \in (0, 1)$,

$$g_\alpha(s) := 1 - (1 - g(s))^\alpha \quad s \in [0, 1].$$

Show as in [41] that g_α is the generating function of an honest probability distribution which is a QSD associated to the rate of mass decay m^α (non-uniqueness of QSDs).

We wish to provide here the so-called ‘conceptual proof’ of [33]. To do this, we will need the following lemma.

Lemma 2.3 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 $\hat{\nu}_n \rightarrow \infty$ in distribution, then $a_n \rightarrow \infty$.

Proof of Theorem 2.2 (Joffe [22], Lyons–Pemantle–Peres [33]). 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. Then consider the functions G_n defined by

$$G_n(s) := \mathbb{E}_1(1 - s^{Z_n} \mid Z_n \neq 0) = \frac{1 - f_n(s)}{1 - f_n(0)} \quad s \in [0, 1].$$

The previously mentioned stochastic monotonicity implies that the sequence $(G_n(s))_n$ is nondecreasing. Let $G(s)$ be its limit. Then G is nonincreasing, and $-G$ is convex, so it is continuous on $(0, 1)$. Then notice that $G_n(f(s)) = \Gamma(f_n(0))G_{n+1}(s)$, where

$$\Gamma(s) := \frac{1 - f(s)}{1 - s} \quad s \in [0, 1].$$

Since $\Gamma(s)$ goes to m as $s \rightarrow 1$ and $f_n(0)$ goes to 1 as $n \rightarrow \infty$, we get that $G(f(s)) = mG(s)$. This entails $G(1-) = mG(1-)$, and since $m < 1$, $G(1-) = 0$, which ensures that $1 - G$ is the generating function of a proper random variable.

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 (3)

$$\mathbb{E}_0(s^{Z_n^\uparrow}) = \prod_{k=0}^{n-1} (f' \circ f_k(s)/m) \quad s \in [0, 1].$$

An immediate recursion shows that

$$\mathbb{E}_0(s^{Z_n^\uparrow+1}) = m^{-n} s f'_n(s) \quad s \in [0, 1].$$

Now a straightforward calculation provides

$$a_n^{-1} \sum_{k \geq 1} k \mu_n(k) s^k = m^{-n} s f'_n(s) \quad s \in [0, 1].$$

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 1.4, 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. \square

Exercise 7 *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 2.4 (Kesten–Ney–Spitzer [24]) *Assume $\sigma := \text{Var}(Z_1) < \infty$. Then we have*

(i) *Kolmogorov's estimate [27]*

$$\lim_{n \rightarrow \infty} n \mathbb{P}(Z_n \neq 0) = \frac{2}{\sigma}$$

(ii) *Yaglom's universal limit law [44]*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n/n \geq x \mid Z_n \neq 0) = \exp(-2x/\sigma) \quad x > 0.$$

For a recent proof, see [13, 34].

2.3.2 The Q -process

We finish this section with results on the Q -process of BGW processes (they can be found in [4, pp.56–59]).

Theorem 2.5 *The Q -process Z^\uparrow can be properly defined as*

$$\mathbb{P}_i(Z_n^\uparrow = j) := \lim_{k \rightarrow \infty} \mathbb{P}_i(Z_n = j \mid Z_{n+k} \neq 0) = \frac{j}{i} m^{-n} \mathbb{P}_i(Z_n = j) \quad i, j \geq 1. \quad (5)$$

It is transient if $m = 1$, and if in addition $\sigma := \text{Var}(Z_1) < \infty$, then

$$\lim_{n \rightarrow \infty} \mathbb{P}(2Z_n^\uparrow/\sigma \geq x) = \int_x^\infty y \exp(-y) dy \quad x > 0.$$

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

$$\sum_{j \geq 0} \mathbb{P}_i(Z_1^\uparrow = j) s^j = \sum_{j \geq 0} \mathbb{P}_i(Z_1 = j) \frac{j}{i} m^{-1} s^j = \frac{s f'(s)}{m} f(s)^{i-1},$$

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 [33]). 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

$$\mu_k = \frac{k p_k}{m} \quad k \geq 1,$$

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 branching mechanism f and immigration mechanism 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 Random walks and BGW trees

3.1 The Lukasiewicz path

The so-called Ulam–Harris–Neveu labelling of a plane tree assumes that each individual (vertex) of the tree is labelled by a finite word of positive integers whose length is the generation of the individual. The set of finite words of integers is denoted \mathcal{U}

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

where \mathbb{N} is the set of positive integers and \mathbb{N}^0 is reduced to the empty set. More specifically, the root of the tree is denoted \emptyset and her offspring are labelled $1, 2, 3, \dots$ from left to right. The recursive rule is that the offspring of any individual labelled u are labelled $u1, u2, u3, \dots$ from left to right, where ui is the mere concatenation of the word u and the number i . From

now on, we only consider finite plane trees. There are two ways of exploring a finite plane tree that can be explained simply thanks to the example on Fig. 3. The *breadth-first search* consists in visiting all individuals of the first generation from left to right, then all individual of the second generation from left to right, etc. In this example, breadth-first search gives $\emptyset, 1, 2, 3, 11, 21, 22, 31, 32, 111, 112, 113$. This exploration algorithm is still valid for trees with (finite breadths but) possibly infinite depths. The *depth-first search* is the lexicographical order associated with the Ulam–Harris–Neveu labelling. In this example, depth-first search gives $\emptyset, 1, 11, 111, 112, 113, 2, 21, 22, 3, 31, 32$. This exploration algorithm is still valid for trees with (finite depths but) possibly infinite breadths.

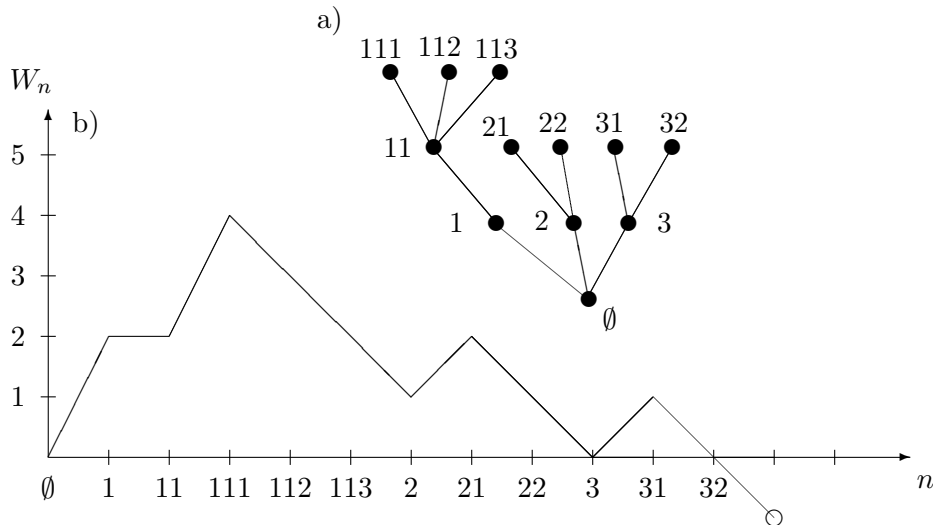


Figure 3: a) A Bienaymé–Galton–Watson tree and b) the associated random walk W .

From now on, we stick to the depth-first search procedure, and we denote by v_n the word labelling the n -th individual of the tree in this order. For any integers $i < j$ and any finite word u , we agree to say that ui is ‘younger’ than uj . Also we will write $u \prec v$ if u is an ancestor of v , that is, there is a sequence w such that $v = uw$ (in particular $u \prec u$). Last, for any individual u of the tree, we let $\rho(u)$ denote the number of younger sisters of u , and for any integer n , we put $W_n = 0$ if $v_n = \emptyset$, and if v_n is any individual of the tree different from the root,

$$W_n := \sum_{u \prec v_n} \rho(u).$$

An example of this finite chain of non-negative integers is represented on Fig. 3.

Proposition 3.1 *Assume that the tree is a subcritical or critical BGW tree, and denote by ξ a r.v. with the offspring distribution. Then the process $(W_n; n \geq 0)$ is a random walk killed upon hitting -1 , with steps in $\{-1, 0, 1, 2, \dots\}$ distributed as $\xi - 1$.*

We omit the proof of this proposition, and we refer the reader to [31] or [32]. The idea is that $W_{n+1} - W_n + 1$ is exactly the number of offspring of v_n . The independence of jumps is obvious but the proof is a little tedious because of the random labelling. The following corollary is straightforward.

Corollary 3.2 *The total progeny of a (sub)critical BGW tree with offspring r.v. ξ has the same law as the first hitting time of -1 of the random walk with steps distributed as $\xi - 1$ and started at 0 .*

3.2 Kemperman's formula

Of the following three paragraphs, the first two are inspired from [35, Section 6.1].

3.2.1 Cyclic shifts

We start with a combinatorial formula, which is a variant of the *ballot theorem*. Let (x_1, \dots, x_n) be a sequence of integers with values in $\{-1, 0, 1, \dots\}$ such that $\sum_{i=1}^n x_i = -k$. Let $(s_j; 0 \leq j \leq n)$ be the walk with steps $x = (x_1, \dots, x_n)$, that is, $s_0 := 0$ and $s_j := \sum_{i=1}^j x_i$, for $j = 1, \dots, n$. Let $\sigma_i(x)$ be the i -th cyclic shift of x , that is, the sequence of length n whose j -th term is x_{i+j} where $i + j$ is understood modulo n .

Lemma 3.3 *There are exactly k distinct $i \in \{1, \dots, n\}$ such that the walk with steps $\sigma_i(x)$ first hits $-k$ at time n .*

Proof. Consider $i = m$ the least i such that $s_i = \min_{1 \leq j \leq n} s_j$ to see there is at least one such i . By replacing x with $\sigma_m(x)$, it may be assumed that the original walk first hits $-k$ at time n . Now for $j = 1, \dots, k$, set $t_j := \min\{i \geq 1 : s_i = -j\}$, the first hitting time of $-j$ by the walk. Then it is easy to check that the walk with steps $\sigma_i(x)$ first hits $-k$ at time n iff i is one of the t_j , $j = 1, \dots, k$. \square

Now assume that $X := (X_1, X_2, \dots, X_n)$ is a *cyclically exchangeable* sequence of random variables with values in $\{-1, 0, 1, \dots\}$, that is, for any $i = 1, \dots, n$, X and $\sigma_i(X)$ have the same law. Set $S_j := \sum_{i=1}^j X_i$ the walk with steps X . Then the sequence (X_1, X_2, \dots, X_n) conditioned by $\{S_n = -k\}$ is still cyclically exchangeable, so that each walk with steps $\sigma_i(X)$ has the same law as S conditioned by $\{S_n = -k\}$. As a consequence, if T_{-k} denotes the first hitting time of $-k$ by the walk S , then

$$\begin{aligned} n\mathbb{P}(T_{-k} = n \mid S_n = -k) &= \sum_{i=1}^n \mathbb{E}(\mathbf{1}_{T_{-k}=n} \mid S_n = -k) \\ &= \sum_{i=1}^n \mathbb{E}(\mathbf{1}_{T_{-k}^{(i)}=n} \mid S_n = -k) \\ &= \mathbb{E} \left(\sum_{i=1}^n \mathbf{1}_{T_{-k}^{(i)}=n} \mid S_n = -k \right) \end{aligned}$$

where the exponent (i) refers to the walk with steps $\sigma_i(X)$. Then thanks to the previous lemma, the sum inside the expectation in the last display equals k . This can be recorded in the next equality, known as *Kemperman's formula*

$$\mathbb{P}(T_{-k} = n) = \frac{k}{n} \mathbb{P}(S_n = -k). \quad (6)$$

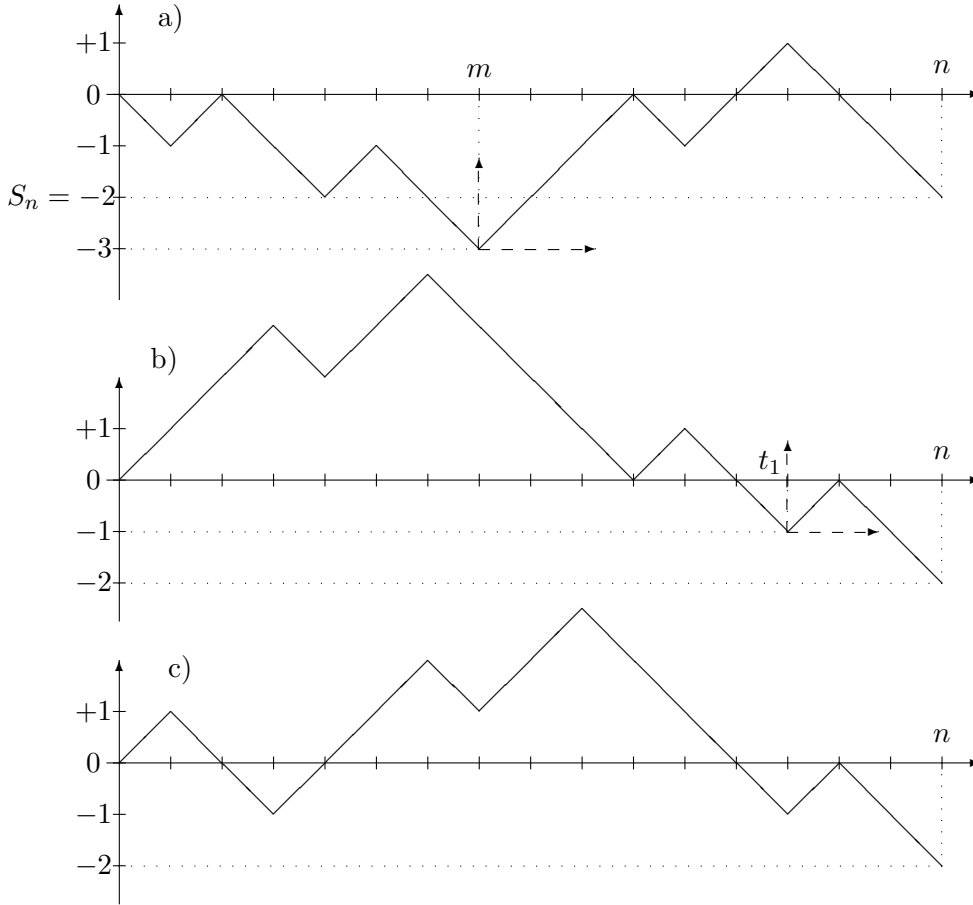


Figure 4: a) A walk S hitting its minimum -3 for the first time at time m , with $S_n = -2$; b) the walk obtained from a) by cyclic permutation of steps starting at time m ; c) the walk obtained from b) by cyclic permutation of steps starting at the first hitting time t_1 of -1 . For smaller values $-k$ of S_n , further cyclic permutations at the first hitting time t_j of $-j$ yields the other walks hitting $-k$ for the first time at n .

Actually, Kemperman [23] showed this formula for i.i.d. r.v. See also [6] for an alternative proof.

In the i.i.d. setting, if $Z = (Z_n; n \geq 0)$ denotes a (sub)critical BGW process with offspring distributed as some r.v. ξ , then (6) reads as in the following statement, where we have substituted $\xi - 1$ to the generic step X , using Corollary 3.2 along with the branching property.

Proposition 3.4 *If $Y := \sum_{i \geq 1} Z_i$ denotes the total progeny of the BGW process Z , then*

$$\mathbb{P}(Y = n \mid Z_0 = k) = \frac{k}{n} \mathbb{P}(\xi_1 + \dots + \xi_n = n - k) = \frac{k}{n} \mathbb{P}(Z_1 = n - k \mid Z_0 = n), \quad (7)$$

where ξ_1, ξ_2, \dots denote i.i.d. r.v. distributed as ξ .

3.2.2 Lagrange inversion formula

An alternative way of proving (6) in the i.i.d. setting is as follows. Consider the random walk S with (i.i.d.) steps X_1, X_2, \dots with values in $\{-1, 0, 1, \dots\}$ and let f denote the probability generating function of $X + 1$ (which is the offspring p.g.f. of the associated BGW tree), as well as h_k the p.g.f. of the first hitting time T_{-k} of $-k$ by S . We assume that $\mathbb{E}X \leq 0$ so that in particular $f(0) \neq 0$. We start with the following two exercises.

Exercise 8 *By an appeal to the last subsection or a recursive application of the strong Markov property, prove that T_{-k} is the sum of k i.i.d. r.v. distributed as T_{-1} , so that $h_k = h^k$, where h stands for h_1 .*

Exercise 9 *Conditioning on X_1 , prove that for all $z \in [0, 1]$*

$$h(z) = zf(h(z)). \quad (8)$$

Actually, since $f(0) \neq 0$, *Lagrange inversion formula* ensures that there is a unique analytic function h satisfying (8) for z in a neighbourhood of 0. In addition, the expansion in powers of z is such that

$$[z^n]h(z)^k = \frac{k}{n}[z^{n-k}]f(z)^n.$$

Since $h^k = h_k$ is the p.g.f. of T_{-k} and f^n is that of a sum of i.i.d. r.v. distributed as $X_1 + 1$, the previous lemma entails

$$\mathbb{P}(T_{-k} = n) = \frac{k}{n} \mathbb{P}\left(\sum_{j=1}^n (X_j + 1) = n - k\right),$$

where the event in the r.h.s. is also $\{S_n = -k\}$, so we finally get (6).

3.2.3 Dwass proof

Proposition 3.4 has yet another proof due to M. Dwass [9, 10], to which it owes the name of *Dwass identity*. Actually, the argument of [10] relies on the previous knowledge that the function $z \mapsto z/f(z)$ is a one-to-one mapping of $[0, 1]$ onto itself with inverse denoted h (hence the same h as in the previous subsection) and that h is the p.g.f. of the total progeny of the BGW tree [9, 12]. Actually, this last assertion can be proved as an exercise.

Exercise 10 *By conditioning on Z_1 , prove that the p.g.f. of the total progeny Y of the BGW tree satisfies (8).*

The contribution of the first cited reference [9] is left as another exercise.

Exercise 11 *If S is a random walk with steps in $\{-1, 0, 1, 2, \dots\}$, then $\mathbb{P}_0(S_n < 0 \forall n \geq 1) = -\mathbb{E}_0(S_1)$.*

Let us start with those facts in mind.

For any $z \in (0, 1]$, let $\xi^{(z)}$ denote the r.v. with p.g.f. $f^{(z)} : x \mapsto f(zx)/f(z)$, set $X^{(z)} := \xi^{(z)} - 1$ and let $S^{(z)}$ be the random walk with (i.i.d.) steps distributed as $X^{(z)}$. Thanks to the result of the last exercise,

$$\mathbb{P}(S_n^{(z)} < 0 \forall n \geq 1) = -\mathbb{E}X^{(z)} = 1 - \frac{zf'(z)}{f(z)}.$$

Then if L_0 denotes the last passage time at 0 by $S^{(z)}$, we have for any $k \geq 0$,

$$\begin{aligned} \mathbb{P}(X^{(z)} = k) &= \sum_{n \geq k} \mathbb{P}(S_1^{(z)} = k, L_0 = n + 1) \\ &= \sum_{n \geq k} \mathbb{P}(S_1^{(z)} = k, S_{n+1}^{(z)} = 0, S_{n+1+i}^{(z)} < 0 \forall i \geq 1) \\ &= \mathbb{P}(X^{(z)} = k) \sum_{n \geq k} \mathbb{P}_0(S_n^{(z)} = -k) \left(1 - \frac{zf'(z)}{f(z)}\right). \end{aligned} \quad (9)$$

Now

$$\left(f^{(z)}\right)^n(x) = \frac{f(zx)^n}{f(z)^n} = \sum_{j \geq 1} \mathbb{P}(\xi_1 + \dots + \xi_n = j) \frac{z^j}{f(z)^n} x^j,$$

so that for any $j \geq 0$,

$$\mathbb{P}(\xi_1^{(z)} + \dots + \xi_n^{(z)} = j) = \frac{z^j}{f(z)^n} \mathbb{P}(\xi_1 + \dots + \xi_n = j),$$

which also reads for any $j \geq -n$,

$$\mathbb{P}(X_1^{(z)} + \dots + X_n^{(z)} = j) = \frac{z^{j+n}}{f(z)^n} \mathbb{P}(X_1 + \dots + X_n = j).$$

As a consequence, we get from (9) (although the case when $\mathbb{P}(X^{(z)} = k) = 0$ is not treated in [10])

$$\sum_{n \geq k} \mathbb{P}_0(S_n = -k) \frac{z^{n-k}}{f(z)^n} = \left(1 - \frac{zf'(z)}{f(z)}\right)^{-1}, \quad (10)$$

Now recalling the function h and setting $z = h(u)$, differentiating the equality $h(u) = uf(h(u))$ yields

$$\frac{uh'(u)}{h(u)} = \left(1 - \frac{zf'(z)}{f(z)}\right)^{-1},$$

which transforms (10) into

$$\sum_{n \geq k} \mathbb{P}_0(S_n = -k) u^{n-1} = h'(u) h(u)^{k-1}.$$

Integrating the last equation in u yields

$$\sum_{n \geq k} \frac{k}{n} \mathbb{P}_0(S_n = -k) u^n = h(u)^k.$$

The remark in the beginning of the present subsection provides the result.

4 Coalescent point processes

4.1 The coalescent point process of quasi-stationary BGW trees

The present subsection is extracted from [30].

4.1.1 A doubly infinite population

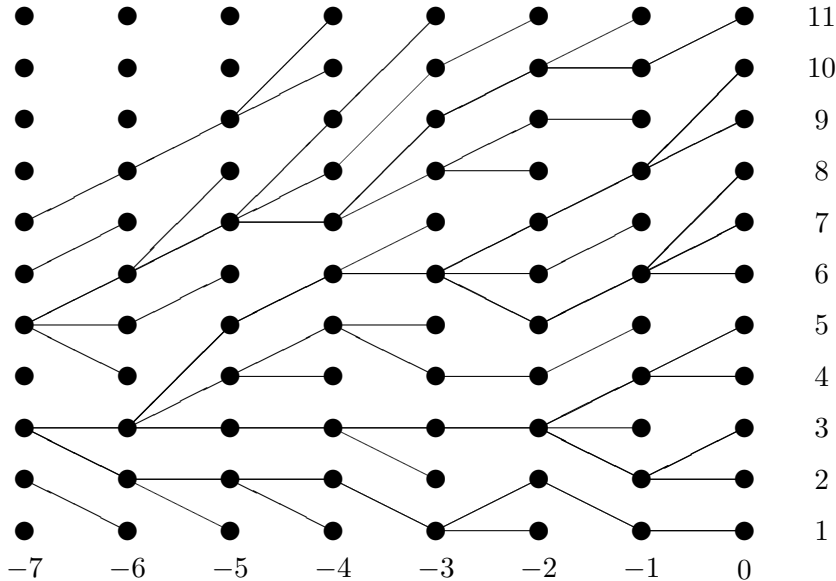


Figure 5: A pathwise representation of the quasi-stationary genealogy. Horizontal labels are generations and vertical labels are individuals.

Let $(\xi(n, i))$ be a doubly indexed sequence of independent integers distributed as ξ , where n is indexed by \mathbb{Z} and i is indexed by $\mathbb{N} = \{1, 2, \dots\}$. For all integers $n \in \mathbb{Z}$ and $i \in \mathbb{N}$, we think of a pair (n, i) as the i -th individual in generation n . We endow these populations with the following asexual genealogy. Individual (n, i) has mother $(n - 1, j)$ if

$$\sum_{k=1}^{j-1} \xi(n-1, k) < i \leq \sum_{k=1}^j \xi(n-1, k).$$

Such a genealogy is represented on Figure 5. In particular, if $Z^{(n, i)}(k)$ denotes the number of descendants of (n, i) at generation $n + k$, then the processes $(Z^{(n, i)}(k); k \geq 0)$ are identically distributed Bienaymé–Galton–Watson (BGW) processes starting from 1. We will refer to Z as some generic BGW process with this distribution.

From now on, we seek to focus on the ancestry of the population at time 0, and we write $(-n, \sigma(n, i))$ for the ancestor of individual $(0, i)$ at generation $-n$. We are interested in the time of coalescence $C_{i, j}$ of individuals $(0, i)$ and $(0, j)$, that is,

$$C_{i, j} := \min\{n \geq 1 : \sigma(n, i) = \sigma(n, j)\},$$

where it is understood that $\min \emptyset = +\infty$. Defining

$$A_i := C_{i,i+1},$$

it is easily seen that by construction, for any $i \leq j$,

$$C_{i,j} = \max\{A_i, A_{i+1}, \dots, A_{j-1}\}.$$

Thus, the sequence A_1, A_2, \dots contains all the information about the (unlabelled) genealogy of the standing population. It is called the *coalescent point process*.

Definition 4.1 For the BGW process Z , we define ζ_n as the number of individuals at generation 1 having alive descendance at generation n . In particular, ζ_n has the law of

$$\sum_{k=1}^{\xi} \epsilon_k,$$

where the ϵ_k are i.i.d. Bernoulli r.v. with success probability $p_{n-1} := \mathbb{P}(Z_{n-1} \neq 0)$, independent of ξ . We also define ζ'_n as the r.v. distributed as $\zeta_n - 1$ conditional on $\zeta_n \neq 0$.

Exercise 12 Prove that

$$\mathbb{P}(\zeta'_n = 0) = \frac{1 - f_{n-1}(0)}{1 - f_n(0)} f'(f_{n-1}(0)).$$

This is the probability that the descendance at generation n , conditioned to be nonempty, of a BGW tree started from one individual, has only one common ancestor at generation 1.

4.1.2 The coalescent point process

Recall that A_i denotes the time of coalescence of individuals $(0, i)$ and $(0, i + 1)$. We will sometimes call it the i -th branch length of the coalescent. Set

$$\mathcal{D}(n, i) := \{\text{daughters of } (-n, \sigma(n, i)) \text{ with descendants in } \{(0, j); j \geq i\}\},$$

and further

$$B(n, i) := \#\mathcal{D}(n, i) - 1.$$

We also set $B(n, 0) := 0$ for all n .

Theorem 4.2 The i -th branch length is a simple functional of $(B(\cdot, i); i \geq 0)$

$$A_i = \min\{n \geq 1 : B(n, i) \neq 0\}.$$

In addition, the sequence-valued chain $(B(\cdot, i); i \geq 0)$ is a Markov chain started at the null sequence. For any sequence of nonnegative integers $(b_n)_{n \geq 0}$, the law of $B(\cdot, i + 1)$ conditionally given $B(n, i) = b_n$ for all n , is given by the following transition. We have $B(n, i + 1) = b_n$ for all $n > A_i$, $B(A_i, i + 1) = b_{A_i} - 1$ and the r.v.'s $B(n, i + 1)$, $1 \leq n < A_i$, are independent, distributed as ζ'_n . In particular, the law of A_1 is given by

$$\mathbb{P}(A_1 > n) = \prod_{k=1}^n \mathbb{P}(\zeta'_k = 0) = \frac{1}{1 - f_n(0)} \prod_{k=1}^n f'(f_{k-1}(0)) = \frac{f'_n(0)}{1 - f_n(0)} = \mathbb{P}(Z_n = 1 \mid Z_n \neq 0).$$

Exercise 13 Compute the first terms of the sequence $(B(\cdot, i); i \geq 0)$ associated to the realization of Figure 5.

In general, the coalescent point process $(A_i, i \geq 1)$ is not itself a Markov process, but in the linear-fractional case, it actually is a sequence of i.i.d. r.v. See [38] for a related observation.

Proposition 4.3 *In the linear fractional case with parameters (a, b) , the branch lengths $(A_i, i \geq 1)$ are i.i.d. with common distribution given by*

$$\mathbb{P}(A_1 > n) = \frac{b - a}{bm^n - a}$$

when $a \neq b$, and when $a = b$ (critical case), by

$$\mathbb{P}(A_1 > n) = \frac{1 - a}{na + 1 - a}.$$

4.1.3 Disintegration of the quasi-stationary distribution

In this subsection, we assume that $f'(1) < 1$ (subcritical case) and we denote by $(u_k)_{k \geq 1}$ the Yaglom limit with generating function, say g . Set

$$U := \min\{i \geq 1 : A_i = +\infty\}.$$

Then for any $i \leq U < j$, the coalescence time between individuals $(0, i)$ and $(0, j)$ is $\max\{A_k : i \leq k < j\} = +\infty$, so that $(0, i)$ and $(0, j)$ do not have a common ancestor. Now set

$$V := \max\{A_k : 1 \leq k < U\},$$

the coalescence time of the subpopulation $\{(0, i) : 1 \leq i \leq U\}$ (where it is understood that $\max \emptyset = 0$), that is, $-V$ is the generation of the most recent common ancestor of this subpopulation. We provide the joint law of (U, V) in the next proposition.

Proposition 4.4 *The law of V is given by*

$$\mathbb{P}(V \leq n) = u_1 \frac{1 - f_n(0)}{f_n'(0)} = \frac{u_1}{\mathbb{P}(Z_n = 1 \mid Z_n \neq 0)} \quad n \geq 0.$$

Conditional on $V = n \geq 1$, U has the law of Z_n conditional on $\zeta_n \geq 2$. In addition, U follows Yaglom's quasi-stationary distribution, which entails the following disintegration formula

$$g(s) = \mathbb{P}(V = 0)s + \sum_{n \geq 1} \mathbb{P}(V = n)\mathbb{E}(s^{Z_n} \mid \zeta_n \geq 2) \quad s \in [0, 1].$$

In the linear fractional case with parameters (a, b) ($a > b$, subcritical case), recall from Section 2 that the Yaglom quasi-stationary distribution is geometric with failure probability b/a . Thanks to Proposition 4.3, the branch lengths are i.i.d. and are infinite with probability $1 - (b/a)$. Then thanks to the previous proposition, the quasi-stationary size U is the first i such that A_i is infinite, which indeed follows the aforementioned geometric law.

Proof. From the transition probabilities of the Markov chain $(B(\cdot, i); i \geq 1)$ given in Theorem 4.2, we deduce that

$$V = \max\{n : B(n, 1) \neq 0\},$$

so that, thanks to Theorem 4.2,

$$\mathbb{P}(V \leq n) = \prod_{k \geq n+1} \mathbb{P}(\zeta'_k = 0) = \frac{\prod_{k \geq 1} \mathbb{P}(\zeta'_k = 0)}{\prod_{k=1}^n \mathbb{P}(\zeta'_k = 0)} = \frac{\mathbb{P}(A_1 = +\infty)}{\mathbb{P}(A_1 \geq n)}.$$

Now since

$$\mathbb{P}(A_1 \geq n) = \frac{f'_n(0)}{1 - f_n(0)} = \mathbb{P}(Z_n = 1 \mid Z_n \neq 0),$$

and because this last quantity converges to u_1 as $n \rightarrow \infty$, we get the result for the law of V .

Recall that $(-n, \sigma(n, 1))$ is the ancestor, at generation $-n$, of $(0, 1)$, so that the total descendance $\Upsilon_n := Z^{(-n, \sigma(n, 1))}(n)$ of $(-n, \sigma(n, 1))$ at generation 0 has the law of Z_n conditional on $Z_n \neq 0$. Now since no individual $(0, j)$ with $j > U$ has a common ancestor with $(0, 1)$, we have the inequality $\Upsilon_n \leq U$. On the other hand, $(0, U)$ and $(0, 1)$ do have a common ancestor (all coalescence times $(A_i; 1 \leq i \leq U - 1)$ are finite), so that there is n such that $\Upsilon_n = U$. Since the sequence (Υ_n) is obviously nondecreasing, it is stationary at U , that is,

$$\lim_{n \rightarrow \infty} \Upsilon_n = U \quad \text{a.s.}$$

Since Υ_n has the law of Z_n conditional on $Z_n \neq 0$, U follows the Yaglom quasi-stationary distribution.

Actually, since $-V$ is the generation of the most recent common ancestor of $(0, 1)$ and $(0, U)$, we have the following equality

$$U = \Upsilon_V.$$

Now recall that $V = \max\{n : B(n, 1) \neq 0\}$. By definition of $B(\cdot, 1)$, we can write $\{V = n\} = E_n \cap F_n$, where $E_n := \{B(k, 1) = 0, \forall k > n\}$ and

$$F_n := \{B(n, 1) \geq 1\} = \{\#\mathcal{D}(n, 1) \geq 2\}.$$

Now observe that E_n is independent of all events of the form $F_n \cap \{\Upsilon_n = k\}$ (it concerns the future of parallel branches), so that $\mathbb{P}(\Upsilon_n = k \mid V = n) = \mathbb{P}(\Upsilon_n = k \mid F_n)$. In other words, Υ_n has the law of Z'_n conditional on $\zeta_n \geq 2$, where Z'_n is Z_n conditional on $Z_n \neq 0$. Since $\{Z_n \neq 0\} = \{\zeta_n \geq 1\}$, we finally get that conditional on $V = n$, U has the law of Z_n conditional on $\zeta_n \geq 2$. \square

4.2 Splitting trees

In this subsection, we consider random trees with random edge lengths called splitting trees [14]. Splitting trees are those random trees where individuals give birth at constant rate b during a lifetime with general distribution $\Lambda(\cdot)/b$, to i.i.d. copies of themselves, where Λ is a positive measure on $(0, \infty]$ with total mass b called the *lifespan measure*. We assume that they are started with one unique progenitor. We denote by \mathbb{P} their law, and the subscript s in \mathbb{P}_s means conditioning on the lifetime of the progenitor being s . Of course if \mathbb{P} bears no subscript, this means that the lifetime of the progenitor follows the usual distribution $\Lambda(\cdot)/b$.

Two branching processes. Denote by N_t the number of individuals alive at time t (breadth of the tree at level t) and by Z_n the number of individuals of generation n . The breadth process $(N_t; t \geq 0)$, is a homogeneous, binary *Crump–Mode–Jagers process* (CMJ). Unless Λ is exponential, this branching process is *not* Markovian. On the other hand, it is easy to see that $(Z_n; n \geq 0)$ is a BGW process started at 1, with offspring generating function f

$$f(s) = \int_{(0,\infty)} b^{-1} \Lambda(dz) e^{-bz(1-s)} \quad s \in [0, 1],$$

so the mean number of offspring per individual is

$$m := \int_{(0,\infty)} z \Lambda(dz).$$

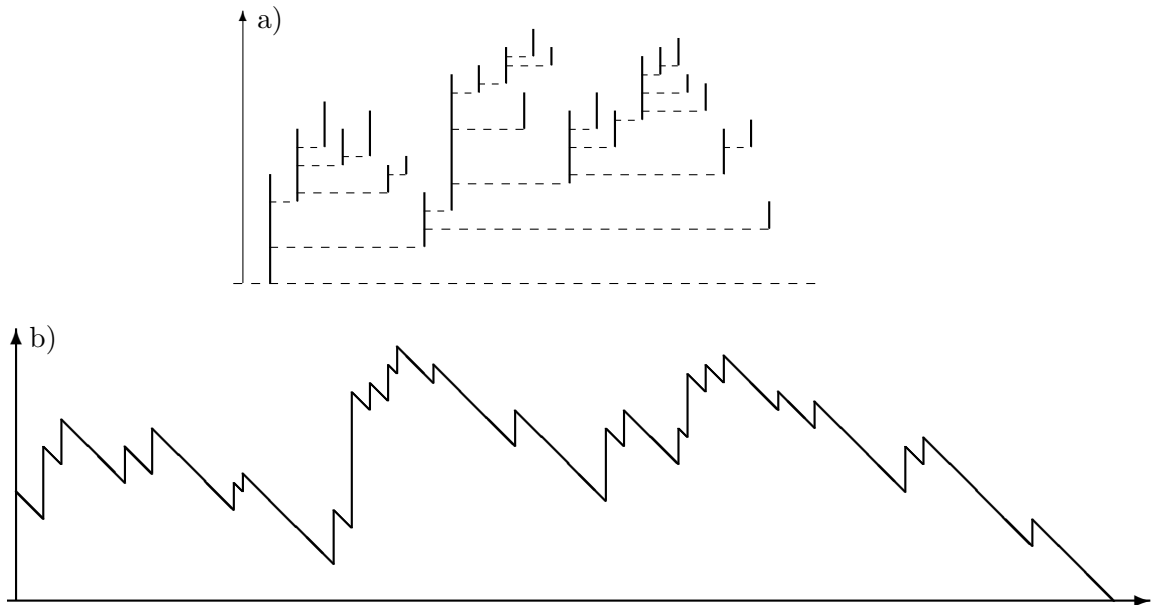


Figure 6: a) A realization of a *splitting tree* with finite extinction time. Horizontal axis has no interpretation, but horizontal arrows indicate filiation; vertical axis indicates real time; b) The associated jumping chronological *contour process* with jumps in solid line.

In [29], we have considered the so-called jumping chronological contour process (JCCP) of the splitting tree truncated up to height (time) t , which starts at the death time of the progenitor, visits all existence times (smaller than t) of all individuals exactly once and terminates at 0. We have shown the following result.

Theorem 4.5 *The JCCP is a compound Poisson process X with jump measure Λ , compensated at rate -1 , reflected below t , and killed upon hitting 0.*

We denote the law of X by P , to make the difference with the law \mathbb{P} of the CMJ process. As seen previously, we record the lifetime duration, say s , of the progenitor, by writing P_s for its conditional law on $X_0 = s$.

Let us be a little more specific about the JCCP. Recall this process visits all existence times of all individuals of the truncated tree. For any individual of the tree, we denote by α its birth time and by ω its death time. 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 all the existence times of v 's lifespan at constant speed -1 . If v has no child, then this visit lasts exactly the lifespan of v ; if v has at least one child, then the visit is interrupted each time a birth time 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 $\omega(w) \wedge t$ and starts the visit of the existence times of w . Since the tree has finite length, the visit of v has to terminate: it does so at the chronological level $\alpha(v)$ and continues the exploration of the existence times of v 's mother, at the height (time) where it had been interrupted. This procedure then goes on recursively as soon as 0 is encountered (birth time of the progenitor). See Figure 6 for an example.

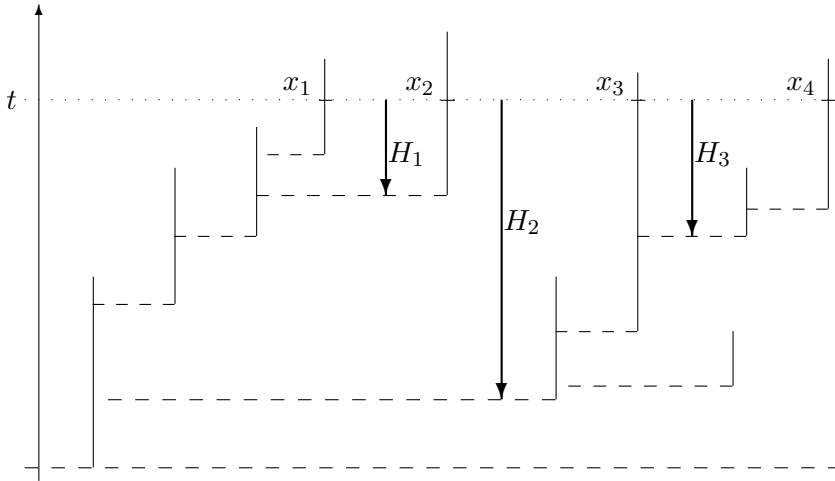


Figure 7: Illustration of a splitting tree showing the durations H_1, H_2, H_3 elapsed since *coalescence* for each of the three consecutive pairs $(x_1, x_2), (x_2, x_3)$ and (x_3, x_4) of the $N_t = 4$ individuals alive at time t .

Since the contour process is Markovian, its excursions between consecutive visits of points at height t are i.i.d. excursions of X . Observe in particular that the number of visits of t by X is exactly the number N_t of individuals alive at time t , where N is the aforementioned CMJ process. See Figure 7.

A first consequence is the computation of the one-dimensional marginals of N . Let T_A denote the first hitting time of the Borel set A by X . Conditional on the initial progenitor to have lived s units of time, we have

$$\mathbb{P}_s(N_t = 0) = P_s(T_0 < T_{(t, +\infty)}), \quad (11)$$

and, applying recursively the strong Markov property,

$$\mathbb{P}_s(N_t = k \mid N_t \neq 0) = P_t(T_{(t, +\infty)} < T_0)^{k-1} P_t(T_0 < T_{(t, +\infty)}). \quad (12)$$

Note that the subscript s in the last display is useless.

The second consequence is that the depths of the excursions of X away from t are i.i.d., distributed as some random variable $H := t - \inf_s X_s$, where X is started at t and killed upon hitting $\{0\} \cup (t, +\infty)$. But in the splitting tree, the depth H_i of the excursion between the i -th visit of t and its $(i + 1)$ -th one is also the *coalescence time* (or *divergence time*) between individual i and individual $i + 1$, that is, the time elapsed since the lineages of individual i and $i + 1$ have diverged. Further, it can actually be shown that the coalescence time $C_{i,i+k}$ between individual i and individual $i + k$ is given by

$$C_{i,i+k} = \max\{H_{i+1}, \dots, H_{i+k}\}, \quad (13)$$

so that the genealogical structure of the alive population of a splitting tree is entirely given by the knowledge of a sequence of independent random variables H_1, H_2, \dots that we will call *branch lengths*, all distributed as H . As in the previous subsection, we call the whole sequence the *coalescent point process*. This point process is a generalization of that defined in [1, 36]. Those processes can be seen as analogues to the coalescent associated to the genealogy of populations with constant size through time [26].

Here, exact formulae can be deduced from the fact that the contour process is not only Markovian, but a Lévy process with no negative jumps. In particular, it can be convenient to handle its Laplace exponent ψ (see Figure 8) instead of its jump measure Λ , that is,

$$\psi(a) := a - \int_0^\infty \Lambda(dx)(1 - e^{-ax}) \quad a \geq 0. \quad (14)$$

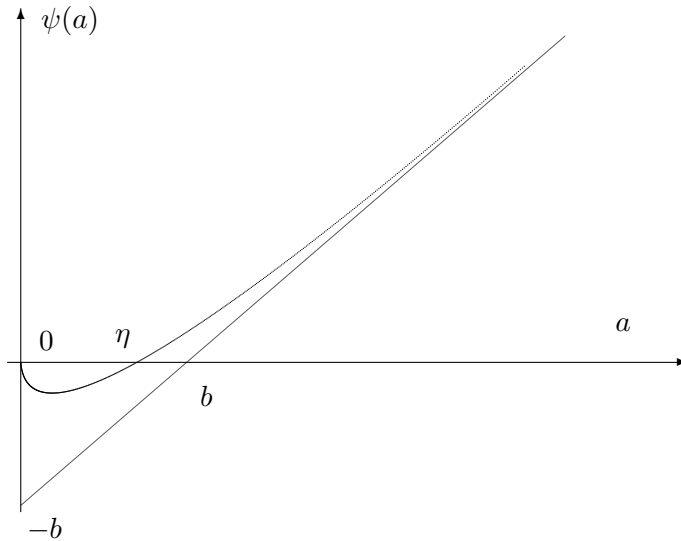


Figure 8: Graph of the Laplace exponent ψ of the JCCP in the supercritical case.

We know that the process is subcritical, critical or supercritical, according whether $m = \int_{(0,\infty]} r\Lambda(dr) < 1, = 1$ or > 1 . In the latter case, the rate η at which $(N_t; t \geq 0)$ grows exponentially on the event of non-extinction, called the *Malthusian parameter*, is the only nonzero root of the convex function ψ . Furthermore, the probabilities of exit from an interval by X have a nice solution [7] in the form

$$P_s(T_0 < T_{(t,+\infty)}) = \frac{W(t-s)}{W(t)}, \quad (15)$$

where the so-called *scale function* W is the nonnegative, nondecreasing, differentiable function such that $W(0) = 1$, characterized by its Laplace transform

$$\int_0^\infty dx e^{-ax} W(x) = \frac{1}{\psi(a)} \quad a > \eta. \quad (16)$$

As a consequence, the typical branch length H between two consecutive individuals alive at time t has the following distribution (conditional on there being at least two extant individuals at time t)

$$\mathbb{P}(H < s) = P_t(T_{(t,+\infty)} < T_s \mid T_{(t,+\infty)} < T_0) = \frac{1 - \frac{1}{W(s)}}{1 - \frac{1}{W(t)}} \quad 0 \leq s \leq t. \quad (17)$$

Let us stress that in some particular cases, (16) can be inverted. When Λ has an exponential density, $(N_t; t \geq 0)$ is a Markovian birth–death process with (birth rate b and) death rate, say d . If $b \neq d$, then (see [29] for example)

$$W(x) = \frac{d - be^{(b-d)x}}{d - b} \quad x \geq 0,$$

whereas if $b = d$,

$$W(x) = 1 + bx \quad x \geq 0.$$

When Λ is a point mass at ∞ , $(N_t; t \geq 0)$ is a pure-birth process, called Yule process, with birth rate b . Then (let $d \rightarrow 0$)

$$W(x) = e^{bx} \quad x \geq 0.$$

In the Markovian case (constant birth and death rates), it had already been noticed by B. Rannala [37] that the coalescence times of a population whose genealogy is given by a (linear) birth–death process started (singly) t units of time ago and whose size is conditioned to be n , are identical to those of the order statistics of n i.i.d. random variables with density

$$h(s) = \frac{(1 - p_0(s))(d - bp_0(s))}{p_0(t)} \quad 0 < s < t,$$

where d is the death rate and

$$p_0(t) := \frac{d(e^{rt} - 1)}{be^{rt} - d},$$

where $r := b - d$. Now (17) applied to the expression of the scale function given previously for the Markovian case ($b \neq d$) agrees with the findings of B. Rannala under the form

$$h(s)ds = \mathbb{P}(H \in ds) = \frac{r^2 e^{rs}}{(be^{rs} - d)^2} \cdot \frac{be^{rt} - d}{e^{rt} - 1} ds \quad 0 < s < t.$$

It is remarkable that in this case, exchanging b and d leaves the distribution of H unchanged. No extension of this fact is known in the general case.

References

- [1] Aldous, D.J., Popovic, L. (2004) A critical branching process model for biodiversity *Adv. Appl. Prob.* **37**(4) 1094–1115.
- [2] Asmussen, S., Hering, H. (1983) *Branching processes*. Birkäuser, Boston.
- [3] Athreya, K.B. (2000) Change of measures for Markov chains and the $L \log L$ theorem for branching processes. *Bernoulli* **6** 323–338.
- [4] Athreya, K.B., Ney P.E. (1972) *Branching processes*. Springer–Verlag, New York.
- [5] Axelrod, D.E., Kimmel, M. (2002) *Branching processes in biology*. Springer–Verlag, New York.
- [6] Bennies, J., Kersting, G. (2000) A random walk approach to Galton-Watson trees. *J. Theoret. Probab.* **13** 777–803.
- [7] Bertoin, J. (1996) *Lévy Processes*. Cambridge U. Press.
- [8] Duquesne, T., Le Gall, J.F. (2002) *Random Trees, Lévy Processes and Spatial Branching Processes*. Astérisque 281.
- [9] Dwass, M. (1968) A theorem about infinitely divisible distributions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **9** 287–289.
- [10] Dwass, M. (1969) The total progeny in a branching process and a related random walk. *J. Appl. Probability* **6** 682–686.
- [11] Evans, S.N. (2008) *Probability and real trees*. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Lecture Notes in Mathematics, 1920. Springer, Berlin.
- [12] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*. Wiley, New York.
- [13] Geiger, J. (1999) Elementary new proofs of classical limit theorems for Galton–Watson processes. *J. Appl. Prob.* **36** 301–309.
- [14] 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.
- [15] 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.
- [16] Harris, T.E. (1963) *The theory of branching processes*. Springer–Verlag, Berlin.
- [17] Heathcote, C. R. (1966) Corrections and comments on the paper “A branching process allowing immigration”. *J. Roy. Statist. Soc. Ser. B* **28** 213–217.

- [18] 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.
- [19] Heyde, C.C. (1970) Extension of a result of Seneta for the super-critical Galton–Watson process. *Ann. Math. Statist.* **41** 739–742.
- [20] 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.
- [21] Jagers, P. (1975) *Branching processes with biological applications*. John Wiley & Sons, London New York Sydney.
- [22] Joffe, A. (1967) On the Galton–Watson branching processes with mean less than one. *Ann. Math. Stat.* **38** 264–266.
- [23] Kemperman, J. H. B. (1961) *The passage problem for a stationary Markov chain*. Statistical Research Monographs, U. of Chicago Press, Chicago, Ill.
- [24] Kesten, H., Ney, P., Spitzer, F. (1966) The Galton–Watson process with mean one and finite variance. *Teor. Verojatnost. i Primenen.* **11** 579–611.
- [25] Kesten, H., Stigum, B.P. (1966) A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** part 3 (21) 1211–1223.
- [26] Kingman, J.F.C. (1982) The coalescent. *Stochastic Process. Appl.* **13**(3) 235–248.
- [27] Kolmogorov, A.N. (1938) Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk* **2** 1–6.
- [28] Lambert, A. (2008) Population Dynamics and Random Genealogies. *Stoch. Models* **24** 45–163.
- [29] Lambert, A. (2010) The contour of splitting trees is a Lévy process. *Ann. Probab.* **38** 348–395.
- [30] Lambert, A., Popovic, L. (2010) The coalescent point process for heights of branching trees. In preparation.
- [31] Le Gall, J.F. (2005) Random trees and applications. *Probab. Surv.* **2** 245–311.
- [32] Le Gall, J.F., Le Jan, Y. (1998) Branching processes in Lévy processes : the exploration process. *Ann. Probab.* **26** 213–252.
- [33] 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.
- [34] Lyons, R., with Peres, Y. Probability on trees and networks. Cambridge U. Press. In preparation. Current version available at <http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html>
- [35] Pitman, J. (2006) *Combinatorial stochastic processes*. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics, 1875. Springer-Verlag, Berlin.

- [36] Popovic, L. (2004) Asymptotic genealogy of a critical branching process. *Ann. Appl. Prob.* **14** (4) 2120–2148.
- [37] Rannala, B. (1997) Gene genealogy in a population of variable size. *Heredity* **78** 417–423.
- [38] Rannala, B. (1997) On the genealogy of a rare allele. *Theoret. Popul. Biol.* **52** 216–223.
- [39] Seneta, E. (1970) On the supercritical Galton–Watson process with immigration. *Math. Biosci.* **7** 9–14.
- [40] Seneta, E. (2006) *Nonnegative matrices and Markov chains*. (3rd revised ed.) Springer Series in Statistics, Springer.
- [41] 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.
- [42] Sevast'yanov, B.A. (1971) *Branching processes*. Nauka, Moscow.
- [43] Watson, H.W., Galton, F. (1874) On the probability of the extinction of families. *J. Anthropol. Inst. Great Britain and Ireland* **4** 138–144.
- [44] 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.