



# Pathwise decompositions of Lévy processes: applications to epidemiological modeling

Miraine Dávila Felipe

## ► To cite this version:

Miraine Dávila Felipe. Pathwise decompositions of Lévy processes: applications to epidemiological modeling. Probability [math.PR]. Université Pierre et Marie Curie, 2016. English. NNT: . tel-01439330

**HAL Id: tel-01439330**

**<https://theses.hal.science/tel-01439330>**

Submitted on 18 Jan 2017

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

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



Université Pierre et Marie Curie

École Doctorale de Sciences Mathématiques de Paris Centre

# THÈSE DE DOCTORAT

Discipline : Mathématiques

présentée par

**Miraine DÁVILA FELIPE**

---

**Pathwise decompositions of Lévy processes:  
applications to epidemiological modeling**

---

dirigée par Amaury LAMBERT  
Bernard CAZELLES

Après avis des rapporteurs :

M. Jean-François DELMAS École des Ponts ParisTech  
M. Joaquín FONTBONA Universidad de Chile

Soutenue le 14 décembre 2016 devant le jury composé de :

M. Jean-François DELMAS	École des Ponts ParisTech	rapporteur
M <sup>me</sup> Irina KOURKOVA	Université Paris 6	examinatrice
M. Amaury LAMBERT	Université Paris 6	directeur
M <sup>me</sup> Sylvie MÉLÉARD	École Polytechnique	présidente
M. Florian SIMATOS	ISAE Supaero	examineur

LPMA – UPMC  
Laboratoire de Probabilités  
et Modèles Aléatoires  
4 Place Jussieu  
75005 Paris

ED 386  
Ecole Doctorale de Sciences  
Mathématiques de Paris Centre  
4 place Jussieu  
75252 Paris Cedex 05

*A mi familia.*

*Los locos encontraron cerrada la puerta del jardín.  
Los cuerdos también la encontraron cerrada.  
Los cuerdos se tendieron allí sin llaves  
y sus cuerpos se llenaron de hormigas y hojas secas.  
Los locos –los locos– rieron, mirando con fijeza  
y pasaron todos a través de la puerta.*

*Teresa Melo (Santiago de Cuba, 1961)*



# Résumé

## Décompositions trajectorielles de processus de Lévy: application à la modélisation de dynamiques épidémiologiques

### Résumé

Résumé: Cette thèse est consacrée à l'étude de décompositions trajectorielles de processus de Lévy spectralement positifs et des relations de dualité pour des processus de ramification, motivée par l'utilisation de ces derniers comme modèles probabilistes d'une dynamique épidémiologique. Nous modélisons l'arbre de transmission d'une maladie comme un arbre de ramification, où les individus évoluent indépendamment les uns des autres, ont des durées de vie i.i.d. (périodes d'infectiosité) et donnent naissance (infections secondaires) à un taux constant durant leur vie. Le processus d'incidence dans ce modèle est un processus de Crump-Mode-Jagers (CMJ) et le but principal des deux premiers chapitres est d'en caractériser la loi conjointement avec l'arbre de transmission partiellement observé, inféré à partir des données de séquences. Dans le Chapitre I, nous obtenons une description en termes de fonctions génératrices de la loi du nombre d'individus infectieux, conditionnellement à l'arbre de transmission reliant les individus actuellement infectés. Une version plus élégante de cette caractérisation est donnée dans le Chapitre II, en passant par un résultat général d'invariance par retournement du temps pour une classe de processus de ramification. Finalement, dans le Chapitre III nous nous intéressons à la loi d'un processus de ramification (sous)critique vu depuis son temps d'extinction. Nous obtenons un résultat de dualité qui implique en particulier l'invariance par retournement du temps depuis leur temps d'extinction des processus CMJ (sous)critiques et de l'excursion hors de 0 de la diffusion de Feller critique (le processus de largeur de l'arbre brownien d'Aldous).

### Mots-clefs

Processus de Lévy, processus de branchement, dualité, modélisation de maladies infectieuses, méthodes phylodynamiques

---

## Abstract

This dissertation is devoted to the study of some pathwise decompositions of spectrally positive Lévy processes, and duality relationships for certain (possibly non-Markovian) branching processes, driven by the use of the latter as probabilistic models of epidemiological dynamics. More precisely, we model the transmission tree of a disease as a splitting tree, i.e. individuals evolve independently from one another, have i.i.d. lifetimes (periods of infectiousness) that are not necessarily exponential, and give birth (secondary infections) at a constant rate during their lifetime. The incidence of the disease under this model is a Crump-Mode-Jagers process (CMJ); the overarching goal of the two first chapters is to characterize the law of this incidence process through time, jointly with the partially observed (inferred from sequence data) transmission tree. In Chapter I we obtain a description, in terms of probability generating functions, of the conditional likelihood of the number of infectious individuals at multiple times, given the transmission network linking individuals that are currently infected. In the second chapter, a more elegant version of this characterization is given, passing by a general result of invariance under time reversal for a class of branching processes. Finally, in Chapter III we are interested in the law of the (sub)critical branching process seen from its extinction time. We obtain a duality result that implies in particular the invariance under time reversal from their extinction time of the (sub)critical CMJ processes and the excursion away from 0 of the critical Feller diffusion (the width process of the continuum random tree).

## Keywords

Lévy processes, branching processes, duality, epidemic modeling, phylodynamics

# Contents

<b>Introduction</b>	<b>9</b>
1 Preliminaries . . . . .	10
2 Statement of results . . . . .	18
<b>I Inferring population dynamics from virus phylogenies</b>	<b>31</b>
Introduction . . . . .	31
1 Preliminaries . . . . .	33
2 Likelihood computation . . . . .	37
Appendix . . . . .	59
I.A Remaining proofs . . . . .	59
I.B Some useful formulas and calculations . . . . .	60
<b>II Time reversal dualities for some random forests</b>	<b>61</b>
1 Introduction . . . . .	61
2 Preliminaries . . . . .	63
3 Results . . . . .	72
4 Epidemiology . . . . .	81
Appendix . . . . .	84
<b>III Branching processes seen from their extinction time</b>	<b>87</b>
1 Introduction . . . . .	87
2 Preliminaries . . . . .	91
3 Main results . . . . .	93
4 Applications . . . . .	105
5 Remaining proofs . . . . .	108
<b>Bibliography</b>	<b>111</b>





# Introduction

*“In branching processes, they all meet: pure mathematical development, biology, physics, and demography, and the concoction is spiced to perfection by the social and cultural context in which it is formed.”*

– Peter Jagers, Oberwolfach, January 2009

The mathematical modeling of epidemics and in particular the emerging field of *phylogenetics* [GPG<sup>+</sup>04], which seeks to understand how pathogen genetic variation and phylogenies are shaped by epidemiological processes, is the core motivation for this work. On that account, this dissertation is devoted to the study of some pathwise decompositions of spectrally positive Lévy processes (SPLP), and duality relationships for certain (possibly non-Markovian) branching processes, driven by the use of the latter as probabilistic models of epidemiological dynamics.

The growing availability of pathogen genetic data and the increase in computer processing power provide an unprecedented opportunity to shed light on the role of evolution in infectious diseases processes [PR09]. Consequently, it has become important to develop mathematical models and statistical methods that can cope with this fast trend and fully exploit the potential of these data. Here we focus on the question of linking sequence data to more traditional sources of information on disease processes, in particular incidence data, to better understand the dynamics of an epidemic [FPG<sup>+</sup>15, Sca16]. In this direction, we have obtained several results characterizing the properties of a class of stochastic processes that are suitable models for transmission dynamics. In particular when the population of susceptible individuals is sufficiently large (e.g. outbreaks), one can be placed in the context of a simplified model without density dependence. A suitable class of these models are birth and death (BD) processes with constant rates [SKvW<sup>+</sup>12, LGBS13, SKBD13], or more generally Crump-Mode-Jagers process (CMJ), where the lifetimes of individuals (periods of infectiousness) are not necessarily exponential, and individuals might have infinitely many offspring (secondary infections). More precisely, the transmission network is modeled as a splitting tree and the incidence of a disease is a CMJ [LS13, LAS14], so we make no restrictive assumption on the distribution of the duration of infection. These results have a number of potential applications to phylodynamics and molecular epidemiology that are outlined in Chapter I.

The overarching goal of the two first chapters is to characterize the law of the incidence process through time, jointly with the partially observed (inferred from sequence data) transmission tree. In Chapter I we obtain a first description in terms of probability generating functions of the conditional likelihood of the number of infectious individuals at multiple deterministic times (including present time) given the transmission network linking individuals that are currently infected. This quantity was previously unknown for

this model and is required for applications involving data. However, the inversion of the Laplace transform of these joint distributions is a difficult problem and poses some obstacles so the result can be used in terms of statistical data analysis, which are discussed in this chapter.

Then, in Chapter II a more elegant version of this characterization is given, passing by a general result of invariance under time reversal for a class of branching process which is of independent interest. We consider a random forest  $\mathcal{F}^*$ , defined as a sequence of i.i.d. BD trees, each started at time 0 from a single ancestor, stopped at the first tree having survived up to a fixed time  $T$ . We denote by  $(\xi_t^*, 0 \leq t \leq T)$  the population size process associated to this forest, and we prove that if the BD trees are supercritical, then the time-reversed process  $(\xi_{T-t}^*, 0 \leq t \leq T)$ , has the same distribution as  $(\tilde{\xi}_t^*, 0 \leq t \leq T)$ , the corresponding population size process of an equally defined forest  $\tilde{\mathcal{F}}^*$ , but where the underlying BD trees are subcritical, obtained by swapping birth and death rates, or equivalently, conditioning on ultimate extinction. We generalize this result to splitting trees (i.e. life durations of individuals are not necessarily exponential), provided that the i.i.d. lifetimes of the ancestors have a specific explicit distribution, different from that of their descendants. The results are based on an identity between the contour of these random forests truncated up to  $T$  and the duality property of Lévy processes. An important consequence is that it allows to derive other useful properties of the forests with potential applications in epidemiology. In particular, this result paves the way for the development and implementation of effective methods for epidemiological inference and their application to sequence and incidence data for pathogens such as HIV-1.

In Chapter III, we are interested in the law of the of a (sub)critical branching process seen from its extinction time, so we consider a spectrally positive Lévy process  $X$  that does not drift to  $+\infty$ , coding for its genealogical structure (in the sense of a contour or exploration process [LGLJ98, Lam10]). Denote by  $I$  the past infimum process defined for each  $t \geq 0$  by  $I_t := \inf_{[0,t]} X$ . Let  $\gamma$  be the unique time at which the excursion of the reflected process  $X - I$  away from 0 attains its supremum. Then, we prove that the pre- $\gamma$  and the post- $\gamma$  subpaths of this excursion are invariant under space-time reversal. This implies in particular that the local time process of this excursion is also invariant when seen backward from its height. These results show that some (sub)critical branching processes such as the (sub)critical Crump-Mode-Jagers (CMJ) processes and the excursion away from 0 of the critical Feller diffusion, which is the width process of the continuum random tree, are invariant under time reversal from their extinction time.

This introductory chapter aims at giving to the reader the general mathematical framework we work with and a brief description of the main results obtained.

## 1 Preliminaries

### 1.1 Basic notation and definitions

Let  $E = \mathbb{R} \cup \{\partial\}$  where  $\partial$  is a topologically isolated point, so-called *cemetery point*. Let  $\mathcal{B}(E)$  denote the Borel  $\sigma$ -field of  $E$ . Consider the space  $\mathcal{D}(\mathbb{R}_+, E)$  (or simply  $\mathcal{D}$ ) of càdlàg functions  $\omega$  from  $\mathbb{R}_+$  into the measurable space  $(E, \mathcal{B}(E))$  endowed with Skorokhod topology [JS03], stopped upon hitting  $\partial$  and denote the corresponding Borel  $\sigma$ -field by  $\mathcal{B}(\mathcal{D})$ . Define the *lifetime* of a path  $\omega \in \mathcal{D}$  as  $\zeta = \zeta(\omega)$ , the unique value in  $\mathbb{R}_+ \cup \{+\infty\}$  such that  $\omega(t) \in \mathbb{R}$  for  $0 \leq t \leq \zeta$ , and  $\omega(t) = \partial$  for every  $t > \zeta$ . Here  $\omega(t-)$  stands for the left limit of  $\omega$  at  $t \in \mathbb{R}_+$ ,  $\Delta\omega(t) = \omega(t) - \omega(t-)$  for the size of the (possible) jump at  $t \leq \zeta$

and we make the usual convention  $\omega(0-) = \omega(0)$ . Notice that if  $\Delta\omega(\zeta) \neq 0$ , the path is not right-continuous at  $\zeta$ . This unusual convention is made for commodity regarding the notation and the results here are still valid under the usual convention of keeping the path constant right after  $\zeta$  (stopped trajectory), which has also the property of recording the possible jump of the path at its lifetime.

We consider stochastic processes, on the probability space  $(\mathcal{D}, \mathcal{B}(\mathcal{D}), P)$ , say  $X = (X_t, t \geq 0)$ , also called the coordinate process, having  $X_t = X_t(\omega) = \omega(t)$ . In particular, we consider only processes with no-negative jumps, that is such that  $\Delta X_t \in \mathbb{R}_+$  for every  $t \geq 0$ . The canonical filtration is denoted by  $(\mathcal{F}_t)_{t \geq 0}$ .

Let  $\mathcal{P}(E)$  be the collection of all probability measures on  $E$ . We use the notation  $P_x(X \in \cdot) = P(X \in \cdot | X_0 = x)$  and for  $\mu \in \mathcal{P}(E)$ ,

$$P_\mu(X \in \cdot) := \int_E P_x(X \in \cdot) \mu(dx).$$

In the sequel, the absence of subscript indicates that the process starts at 0. For any measure  $\mu$  on  $[0, \infty]$ , we denote by  $\bar{\mu}$  its tail, that is

$$\bar{\mu}(x) := \mu([x, +\infty]).$$

Define by  $\tau_A := \inf\{t > 0 : X_t \in A\}$ , the first hitting time of the set  $A \in \mathcal{B}(E)$ , with the conventions  $\tau_x = \tau_{\{x\}}$ , and  $\tau_x^- = \tau_{(-\infty, x)}$ ,  $\tau_x^+ = \tau_{(x, +\infty)}$  for any  $x \in \mathbb{R}$ .

### Some path transformations of càdlàg functions

In this subsection we will define some families of operators on the space of càdlàg functions  $\omega \in \mathcal{D}$ :

- the classical *shift operators*,  $\theta_s, s \in \mathbb{R}_+$ , defined by

$$[\theta_s(\omega)]_t := \omega_{s+t}, \quad \forall t \in \mathbb{R}_+$$

- the non-standard shift operators,  $\theta'_s, s \in \mathbb{R}_+$ , defined by

$$[\theta'_s(\omega)]_t := \omega_{s+t} - \omega_s, \quad \forall t \in \mathbb{R}_+$$

- the *killing operators*,  $k_s, s \in \mathbb{R}_+$ , defined by

$$[k_s(\omega)]_t := \begin{cases} \omega_t & \text{if } t < s \\ \partial & \text{otherwise} \end{cases}$$

the killing operator can be generalized to killing at random times, for instance  $k_{T_A}(X) = k_{T_A(X)}(X)$ , denotes the process  $X$ , *killed at the first passage into  $A$* . It is easy to see that if  $X$  is a Markov process, so is  $k_{T_A}(X)$ . We set

$$k_0(\omega) = \partial.$$

## 1.2 Spectrally positive Lévy processes (SPLP)

The aim of this section is to recall and establish some results within the theory of *Lévy processes* (i.e. processes with stationary and independent increments, and almost sure right continuous with left limits paths), that will be used later to prove our results. We refer to [Ber96] or [Kyp06] for a detailed review on the subject.

Let  $X = (X_t, t \geq 0)$  be a real-valued Lévy process with no negative jumps, also called *spectrally positive*. Its Laplace exponent  $\psi$ , is defined  $\forall \lambda \geq 0$  by,

$$E_0 \left[ e^{-\lambda X_t} \right] = e^{t\psi(\lambda)},$$

and is expressed as follows thanks to the Lévy-Kintchine formula,

$$\psi(\lambda) = \alpha\lambda + \beta\lambda^2 + \int_{(0,\infty]} \left( e^{-\lambda r} - 1 + \lambda r \mathbb{1}_{\{r < 1\}} \right) \Pi(dr), \quad (1)$$

where  $\alpha \in \mathbb{R}$ ,  $\beta \geq 0$  is called the *Gaussian coefficient* and  $\Pi$  is a  $\sigma$ -finite measure on  $(0, \infty]$ , called the *Lévy measure*, satisfying  $\int_{(0,\infty)} (r^2 \wedge 1) \Pi(dr) < \infty$ . Notice we allow  $\Pi$  to charge  $+\infty$ , which amounts to killing the process at rate  $\Pi(\{+\infty\})$ .

The asymptotic behavior of the paths of  $X$  is deduced from the characteristics of its Laplace exponent around the origin. Indeed, this behavior can be separated in three exhaustive cases

- subcritical = drifts to  $-\infty$ , iff  $\psi'(0+) > 0$ ,
- critical = oscillates, iff  $\psi'(0+) = 0$ ,
- supercritical = drifts to  $+\infty$ , iff  $\psi'(0+) < 0$ .

The paths of  $X$  have finite variation (on every compact time interval) a.s. if and only if  $\beta = 0$  and  $\int_{(0,1]} r \Pi(dr) < \infty$  (otherwise they have infinite variation a.s.), in which case the Laplace exponent adopts the simplest form

$$\psi(\lambda) = -\mathbf{d}\lambda - \int_{(0,\infty]} \left( 1 - e^{-\lambda r} \right) \Pi(dr), \quad (2)$$

where  $\mathbf{d}$  is called the *drift coefficient*. If  $X$  has increasing paths a.s. it is called a *subordinator*, then  $\mathbf{d} \geq 0$ . When  $\mathbf{d} < 0$  and  $\Pi$  is finite, is called a compensated compound Poisson process and in this case the Laplace exponent is infinitely differentiable, strictly convex (when  $\Pi \neq 0$ ),  $\psi(0) = 0$  (except when  $\Pi$  charges  $+\infty$ , in which case  $\psi(0) = \psi(0+) = -\Pi(\{+\infty\})$ ) and  $\psi(+\infty) = +\infty$ . Suppose from now that  $\Pi(+\infty) = 0$ . Define in this case

$$\eta := \sup\{\lambda \geq 0 : \psi(\lambda) = 0\}, \quad (3)$$

the largest root of  $\psi$ . We have that  $\eta = 0$  is the unique root of  $\psi$  when  $\psi'(0+) \geq 0$ . Otherwise, the Laplace exponent has two roots, 0 and  $\eta > 0$ . It is also known that for any  $x > 0$ ,

$$P_x(\tau_0 < +\infty) = e^{-\eta x}.$$

We are interested in particular in the case where  $X$  has Laplace exponent 2 with  $\mathbf{d} = -1$ . In this case, letting  $m := \int_{(0,\infty)} r \Pi(dr)$ , we have that  $\eta = 0$  if and only if  $m = 1 - \psi'(0+) \leq 1$ .

More generally, to solve exit problems for completely asymmetric Lévy processes (that make no positive or no negative jumps) there exists a very useful tool, called the *scale function*, which can be defined as the unique continuous increasing function  $W : [0, +\infty) \rightarrow [0, +\infty)$ , characterized by its Laplace transform,

$$\int_0^{+\infty} e^{-\lambda x} W(x) dx = \frac{1}{\psi(\lambda)}, \quad \lambda > \eta,$$

such that for any  $0 < x < a$ ,

$$P_x(\tau_0 < \tau_a^+) = \frac{W(a-x)}{W(x)}. \quad (4)$$

### Local times and excursion theory

Suppose  $\Pi(+\infty) = 0$ . Denote the past infimum process of  $X$  by  $I_t := \inf_{0 \leq s \leq t} X_s$ . We will now give some elements on excursion theory for Lévy processes reflected at their infimum  $X - I$ , which are also a Markov processes in the canonical filtration  $(\mathcal{F}_t)_{t \geq 0}$  of  $X$  (and also in their own natural filtration). One can construct a local time at 0 and develop an excursion theory for the reflected process, as we explain hereafter. For further details about the following results, we refer to [Ber96, Chapter IV].

For a Markov process, a point  $x$  of its state space is said to be *regular* or *irregular* for itself, if  $P_x(\tau_x = 0)$  is 1 or 0. In a similar way, when the process is real-valued, we can say it is *regular downwards* or *upwards* if we replace  $\tau_x$  by  $\tau_x^-$  or  $\tau_x^+$  respectively. The following result due to Rogozin [Rog66] states that in the case of a spectrally positive Lévy process,

(A.1) 0 is regular for  $(-\infty, 0)$  and

(A.2) 0 is regular for  $(0, \infty)$  (or equivalently for itself) if and only if  $\beta > 0$  or  $\int_0^1 r \Pi(dr) = +\infty$ , that is when the paths of  $X$  have unbounded variation.

The proofs can be found in [Rog66] and also in [Cha94, Ber96].

A consequence of (A.1) is that  $\{0\}$  is always regular for itself for the reflected process  $X - I$ . Then, according to the general theory for Markov processes, there exists a local time at 0 for  $X - I$ , here denoted by  $(L_t, t \geq 0)$  that can be defined as the unique (up to a multiplicative constant) adapted additive functional that grows only on the zeros of  $X - I$ . Furthermore, the fact that  $X$  has no negative jumps entails that  $-I$  satisfies these conditions, so it is an explicit local time for the reflected process. Then, we consider in the sequel that  $L = -I$ . Its right-continuous inverse

$$L_t^{-1} := \inf \{s > 0 : L_s > t\},$$

is the same as  $\tau_{-t} = \inf \{s \geq 0 : X_s < -t\}$ , the first hitting time of  $(-\infty, -t)$  under  $P_0$ . Notice that we write  $\tau_{-t}$  for  $\tau_{-t}^-$  since in the spectrally positive case they are a.s. identical. Moreover,  $\tau_{-t}$  is a (possibly killed) subordinator whose jumps coincide exactly with the excursion intervals of  $X - I$ , so it represents the appropriate time scale for the so-called *excursion process*, that we will now describe.

Let  $\mathcal{E}$  be the space of real-valued càdlàg functions with finite lifetime  $\zeta \in [0, \infty)$ . Define the excursion process  $\epsilon = (\epsilon_t, 0 < t \leq L_\infty)$ , taking values in  $\mathcal{E} \cup \partial$  as follows

$$\epsilon_t := \begin{cases} (X_{L_{t-}^{-1}+s} - I_{L_{t-}^{-1}}, 0 \leq s \leq L_t^{-1} - L_{t-}^{-1}), & \text{if } L_t^{-1} - L_{t-}^{-1} > 0, \\ \partial, & \text{if } L_t^{-1} - L_{t-}^{-1} = 0, \text{ or } t = \infty \end{cases} \quad \text{for } t \geq 0.$$

Then according to Itô's excursion theory,  $(t, \epsilon_t)_{t \geq 0}$  is a Poisson point process, possibly stopped at the first excursion with infinite lifetime (which arrives only and a.s. when  $X$  drifts to  $+\infty$ ). Its intensity is  $dt \, \underline{n}(d\epsilon)$ , where  $\underline{n}$  is a measure on  $\mathcal{E}$  called the *excursion measure*.

Let  $\varepsilon$  be the generic excursion of  $X - I$  away from 0, define the local time process  $(\Gamma(\varepsilon, r), r \geq 0)$  for the canonical excursion  $\varepsilon \in \mathcal{E}$  as a Borel function satisfying

$$\int_0^{\zeta(\varepsilon)} \phi(\varepsilon_s) ds = \int_0^\infty \Gamma(\varepsilon, r) \phi(r) dr, \quad (5)$$

for any continuous function  $\phi$  with compact support in  $[0, \infty)$ . This local time processes are known to exist in the infinite variation case, see for instance [Ber96]. When  $X$  has

finite variation we can define an equivalent process, taking values in  $\mathbb{N} \cup \{+\infty\}$ , as the number of times the excursion hits level  $r$ , i.e.

$$\Gamma(\varepsilon, r) = \sum_{0 \leq s \leq \zeta} \mathbb{1}_{\{\varepsilon_s = r\}}. \quad (6)$$

### Time-reversal duality for Lévy processes

One of the key ingredients of our results is the duality property under time-reversal of Lévy processes (see [Ber96, Chapter II]). Roughly speaking, it states that if a path is space-time-reversed at a finite time horizon, the new path has the same distribution as the original process. We will use the following formulation subsequently: for every fixed  $t > 0$  and every non-negative measurable function  $F$  we have that

$$E[F(k_t \circ X)] = E[F(\rho \circ (k_t \circ X))], \quad (7)$$

where  $k_t$  denotes the killing operator at  $t \in \mathbb{R}_+$  and  $\rho$  denotes the space-time-reversal operator, that can be defined for any  $\omega \in \mathcal{E}$  as

$$\rho \circ \omega := (\omega_{\zeta-} - \omega_{(\zeta-s)-}, 0 \leq s \leq V).$$

By integrating over  $t$ , this result is still valid if the process is killed at an independently distributed finite time.

### 1.3 Branching processes and trees

We consider *branching processes* as stochastic processes with non-negative values satisfying the *branching property*, meaning that for any  $x, y > 0$ , the process started at  $x + y$  has the same distribution as a sum of two independent copies of itself, starting respectively at  $x$  and  $y$ . The simplest branching processes are those in discrete time and state space, the well-known Bienaymé-Galton-Watson (GW) processes [AN72]. In the case of discrete time and continuous state-space, we use the term *Jirina processes* as in [Lam10]. For continuous time and discrete state space we speak of Crump-Mode-Jagers (CMJ) processes and finally the so-called *continuous state branching process* (CSBP) for continuous time and state spaces. Our results concern mainly the latter two, so we will spend more time specifying their characteristics. In particular CMJ processes are the only branching processes to be possibly non-Markovian. We refer to [AN72, Jag75] for the general theory of branching processes.

### Splitting trees and Crump-Mode-Jagers process

*Splitting trees* were introduced in [Gei96] and [GK97] as random trees that arise from a model where individuals behave independently from one another, have i.i.d. life durations, and give birth at constant rate during their lives. A more general definition is given in [Lam10] including the case where the birth rate is infinite. They represent the genealogical structure of a population (or particle system) originating at time 0 with one single progenitor and evolving under such model. These models are a generalization of the BD model, in which the lifetimes of individuals do not necessarily follow exponential distributions, and individuals might have infinitely many offspring. We give here an intuitive description of splitting trees, and we refer to [Lam10, Sec. 4.1] for a rigorous definition.

A splitting tree is a random variable  $\mathcal{T}$  in the space of chronological trees (see [Lam10]), characterized by a  $\sigma$ -finite measure  $\Pi$  on  $(0, \infty]$  called the *lifespan measure*, satisfying

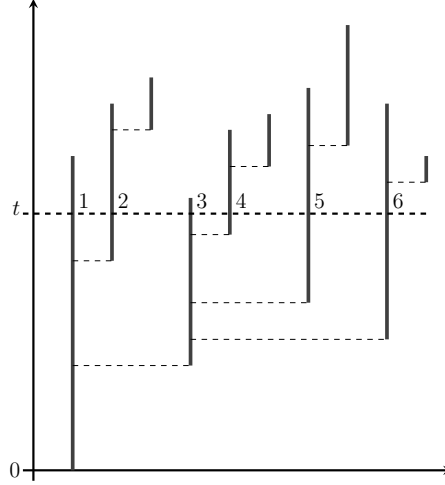


Figure 1 – An example of splitting tree with finite length (sum of all edges). The vertical axis indicates time; the horizontal axis has no meaning, but the dotted horizontal lines show filiation. Six individuals are alive at time  $t$  (bold lines), labeled from 1 to 6.

$\int_{(0,\infty]} (r \wedge 1) \Pi(dr) < \infty$ . When  $\Pi$  is finite with mass  $b$ , the tree corresponds to a population where individuals have i.i.d. lifetimes distributed as  $\Pi(\cdot)/b$  and give birth to single descendants throughout their lives at constant rate  $b$ , all having the same independent behavior. In the general definition individuals may have infinitely many offspring and most of the forthcoming results remain valid if  $\Pi$  is infinite, as in [Lam10]. However, we will treat only the finite case and assume  $\Pi$  finite from now on, for simplicity on the proofs.

We represent splitting trees in the half-plane as in Fig. 1, with time running from bottom to top, and dotted lines representing filiations between individuals: the one on the left is the parent, and that on the right its descendant. All edges are parallel and the children of a given individual are placed to the right of this individual, from the youngest one (to the left) to the eldest one (to the right). This allows to consider a way of *labeling* the (countable) set of individuals alive at a given time  $t$ , as it is shown in Fig. 1. Hereafter we always refer to that order when labeling extant individuals at a given time.

We can define the *width* process of a splitting tree, which counts the number of particles alive in the tree over time, as a functional  $\Xi$  that maps a splitting tree  $\mathcal{T}$  to the function  $\xi : \mathbb{R}_+ \rightarrow \mathbb{N}$  counting the number of extant individuals at time  $t \geq 0$

$$\Xi(\mathcal{T}) := (\xi_t(\mathcal{T}), t \geq 0).$$

These functions are càdlàg, piecewise constant, from  $\mathbb{R}_+$  into  $\mathbb{N} \cup \{+\infty\}$ , and are absorbed at 0. Then we can define the extinction event  $\text{Ext} := \{\lim_{t \rightarrow \infty} \xi_t(\mathcal{T}) = 0\}$  and the time of extinction of the population in a tree as

$$T_{\text{Ext}} := \inf\{t \geq 0 : \xi_t(\mathcal{T}) = 0\},$$

with the usual convention  $\inf \emptyset = \infty$ . A tree, or its width process  $\Xi$ , is said to be subcritical, critical or supercritical if

$$m := \int_{(0,+\infty]} r \Pi(dr).$$



is less than, equal to or greater than 1.

The width process  $\Xi(\mathcal{T}) = (\xi_t(\mathcal{T}), t \geq 0)$  of a splitting tree with finite lifespan measure is known to be a *binary homogeneous Crump-Mode-Jagers process* (CMJ). This process is not Markovian, unless  $\Pi$  is exponential (*birth-death process*) or a Dirac mass at  $\{+\infty\}$  (*Yule process*).

#### 1.4 Links between SPLP and BP

There are a few recent but now well known results that establish different forms of bijections between random trees and certain real-valued stochastic processes, usually called *exploration processes*. In particular, the study of the genealogy associated to branching processes is an important aspect towards their application in multiple fields such as population dynamics, population genetics and evolutionary biology. The idea of coding this genealogical structure through a continuous or jumping stochastic process represents a powerful tool to analyze the properties of the trees generated by the branching mechanism and it has been widely exploited with diverse purposes. See for instance [GK97, LGLJ98, DLG02, Pop04, Lam10, BPS12].

Another link between branching processes and Lévy processes that is worth mentioning for the sake of completeness, is the so-called Lamperti's transform. It is a bijection, via a *random time change* between CSBP and SPLP killed upon hitting 0. We refer the reader to [Lam67] and a more recent study [CLUB09].

#### The continuum random tree and Feller's branching diffusion

Real trees can be defined as the continuous limiting object of rescaled discrete trees and can be coded by a continuous function in a way similar to the coding of discrete trees by their contour functions. *Aldous' Continuum Random Tree* (the so-called CRT) can be defined as the random real tree coded by a normalized Brownian excursion  $\mathbf{e}$ , i.e. the positive Brownian excursion conditioned to have lifetime 1. More generally, the tree coded by Brownian motion (possibly with drift) reflected at 0, is called Brownian forest. We refer to [Ald93, LG05] for the formalism on real trees.

The second one of the two classical Ray-Knight theorem [RY91] establishes that the local time process of a reflected Brownian motion is Feller's branching diffusion. More precisely, let  $B$  be a Brownian motion reflected at 0 and  $(L_s^a, s, a \geq 0)$  the family of its local times, where the index  $s$  corresponds to the *time* of the original process  $B$  and  $a$  is the level variable moving in the state-space of  $B$ . Consider, for  $x > 0$ ,

$$\varsigma_x = \inf\{s : L_s^0 > x\}.$$

Then, the process  $(L_{\varsigma_x}^t, t \geq 0)$ , viewed as a process indexed by  $t$ , which corresponds to the (total) local time accumulated by the resulting path at *height*  $t$ , is equal in distribution to the square of a 0-dimensional Bessel process started at  $x$ , that is, a *standard* Feller branching diffusion  $(Z_t^x, t \geq 0)$ . The latter is defined as the unique strong solution of the SDE

$$dZ_t^x = 2\sqrt{Z_t^x}dW_t^x, \text{ with } Z_0^x = x.$$

This may be understood as a description of the genealogy encoded in Feller's branching diffusion. The exploration path is a concatenation of Brownian excursions, with each excursion corresponding to a CRT and the reflected Brownian motion coding (in the sense of Aldous) the *real tree* which describes the genealogy of the population that evolves

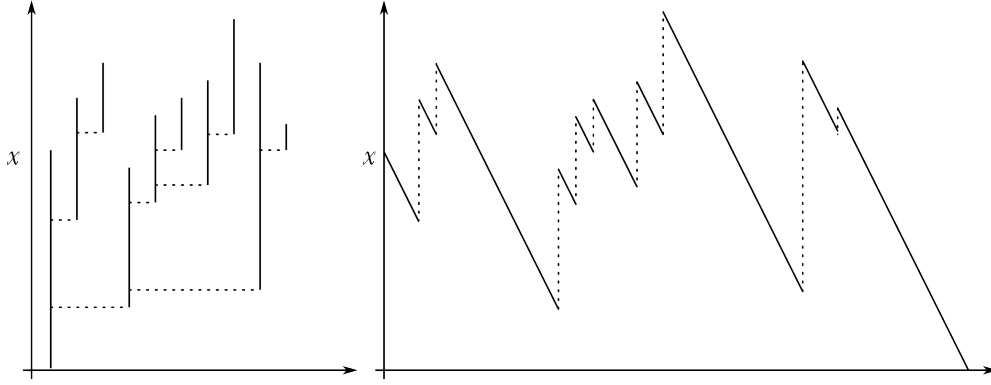


Figure 2 – An example of chronological tree with finite length (left) and its JCCP (right).

according to Feller's diffusion [LG05]. The local time of the exploration path at height  $t$  measures the mass (or size) of the population that is alive at this level in the genealogical forest and this mass is  $Z_t$ , the state of the branching process at time  $t$ .

### The contour process

There exist different versions of the contour process of a splitting tree. In this work we will focus our attention in a particular way of exploring a splitting tree, a non-continuous version of the so-called *contour processes*, introduced in [Lam10] and more recently generalized to *totally ordered measured* (TOM) trees satisfying the splitting property, which are the continuum analogue of splitting trees in the setting of real trees, as it is shown in [LB16]. Here we give only an intuitive description of the process, and we refer to these papers for the rigorous definition. In the sequel, we will often prefer to use the more accurate denomination given by [Lam10], that of *jumping chronological contour process* (JCCP) of the splitting tree. Given a tree  $\mathcal{T}$ , embedded in the plane, its contour or JCCP, denoted by  $\mathcal{C}(\mathcal{T})$  is a continuous time, real-valued process, that starts at the lifespan of the ancestor and then rolls backward along the right-hand side of this first branch at speed  $-1$  until it encounters a birth event, when it jumps up of a height of the lifespan of this new individual, getting to the next tip, and then repeating this algorithm until it eventually hits 0, as shown in Fig. 2.

The JCCP visits all the existence times of each individual exactly once and the number of times it hits a time level, say  $s \geq 0$ , is equal to the number of individuals in the population at time  $s$ . More precisely, for any finite tree  $\mathcal{T}$ , the local time of its contour process is the population size process of the tree, that is

$$(\Gamma(\mathcal{C}(\mathcal{T}), r), 0 \leq r \leq T_{\text{Ext}}) = \Xi(\mathcal{T}),$$

where  $\Gamma$  is defined as in Equation (6).

One of the main results in [Lam10], which is key to our study, is that the JCCP of a splitting tree is a spectrally positive Lévy process  $Y$ , with Laplace exponent given by (2) with  $\mathbf{d} = -1$  and then the binary homogeneous CMJ is its local time. More precisely, the result is the following.

**Theorem** [Lam10, Th.4.3]. *Let  $T$  be a fixed positive time and  $\mathcal{T}^{(T)}$  be a splitting tree with lifespan measure  $\Pi$  in  $(0, +\infty)$ , truncated below height  $T$ . The process  $\mathcal{C}(\mathcal{T}^{(T)})$ , conditional*

on the lifespan of the root individual to be  $x$ , the JCCP of the tree, is distributed as  $Y$ , started at  $x$ , reflected below  $T$  and killed upon hitting 0. Moreover, conditional on  $\text{Ext}$  and on the lifespan of the root individual to be  $x$ , the JCCP is distributed as  $Y$ , started at  $x$ , conditioned on and killed upon hitting 0.

An important consequence of this result is that in the (sub)critical regime, under  $P_x$

$$(\Gamma(k_{\tau_0} \circ Y, r), r \geq 0) \quad (8)$$

is a CMJ with lifespan measure  $\Pi$ , starting with one progenitor with lifespan  $x$ . This result also holds in general (including supercritical) for the process  $Y$  and the splitting tree truncated below  $T$  or conditioned on extinction.

Other interesting consequences of this theorem, that we can also find in [Lam10], are the following. First, for any  $x > 0$ , conditional on being nonzero, the size of the population in the tree a time  $T$ , that is  $\xi_T$ , follows the geometric distribution with success probability  $1/W(T)$ . Secondly, in the supercritical case ( $m > 1$ ), conditional on  $\text{Ext}^c$ ,

$$e^{-\eta t} \xi_t \xrightarrow[t \rightarrow \infty]{} \mathcal{E}, \quad P - a.s.,$$

where  $\mathcal{E}$  is an exponential random variable with parameter  $\psi'(\eta)$ . This result justifies that  $\eta$  is seen as the so-called *Malthusian parameter*.

## The coalescent point process

Let us now explain how the reconstructed genealogies of splitting trees can be studied using the JCCP. Fix a time level  $T > 0$  and consider a splitting tree  $\mathcal{T}$  with lifespan measure  $\Pi$  that survives up until time  $T$ . Let  $(x_i(T), 1 \leq i \leq \xi_T(\mathcal{T}))$  denote the individuals on the tree at time  $T$ , ranked according to the order described before (see Fig. 1). Conditional on the population size  $\xi_T(\mathcal{T})$  at time  $T$ , denote by  $H_i, 1 \leq i \leq \xi_T(\mathcal{T}) - 1$ , the coalescence times between the  $i$ -th and the  $(i+1)$ -th individual, that is the time until the most recent ancestor (abbreviated TMRCA in the literature) of individuals  $x_i(T)$  and  $x_{i+1}(T)$ . We have the following result characterizing the law of this sequence with the help of the scale function previously introduced and Equation (4).

**Theorem** [Lam10, Th.5.4]. *Conditional on  $\xi_T(\mathcal{T}) \geq 1$ , the sequence of coalescence times  $(H_i, 1 \leq i \leq \xi_T(\mathcal{T}) - 1)$  is distributed as a sequence of i.i.d. random variables stopped at its first value greater than  $T$ . The common distribution is that of  $H = \inf_{[0,t]} X$ , where  $X$  is a Lévy process with Laplace exponent given by (2) (with  $\mathbf{d} = -1$ ), started at  $T$  and killed when exiting  $(0, T]$ . Moreover, conditionally on  $\xi_T(\mathcal{T}) > 1$ , the common distribution of the random variables  $(H_i, 1 \leq i \leq \xi_T(\mathcal{T}) - 1)$  is given by*

$$P(H \leq x \mid H < T) = \frac{1 - 1/W(x)}{1 - 1/W(T)}, \quad \text{for } 0 \leq x \leq T.$$

The sequence  $(H_i, 1 \leq i \leq \xi_T(\mathcal{T}) - 1)$  is called *coalescent point process* (CPP) and fully characterizes the genealogy of the extant population at  $T$ , since the coalescence time between any pair of individuals  $i$  and  $i+k$ , can be defined as  $\max_{i < j \leq i+k} H_j$ .

## 2 Statement of results

### 2.1 Main results of Chapter I

In this chapter we consider the first model described in Subsection 1.3, that of splitting trees. The population of infected individuals follows this model from time 0, where the

epidemic originated, until present time  $T_0$ . More precisely we suppose the transmission tree of an epidemic between 0 and  $T_0$  is distributed as a splitting tree, denoted by  $\mathcal{T}$ , with lifespan measure  $\Pi$  of mass  $b$ . Then the infected population size process, denoted by  $(I_t)_{0 \leq t \leq T_0}$ , has the law of the associated CMJ, that is, for each  $0 \leq t \leq T_0$ ,

$$I_t = \xi_t(\mathcal{T}).$$

In the case where susceptible individuals are abundant, such as in outbreaks, this model without density dependence is suitable for the dynamics of the infected population and has been widely used in this context. See for instance [Bec74, Bec77, TFLS06] and more recently [Sta09, SKvW<sup>+</sup>12]. The main motivation of this chapter is to pave the way for the statistical inference on the parameters of this model on the basis of data consisting in

- The reconstructed transmission tree (i.e. the information about non sampled hosts is erased from the original process). These trees are indeed considered to be estimated from pathogen sequences from hosts sampled at present time  $T_0$ , so in our model we make no distinction between the phylogeny of sequences and the transmission tree (and no uncertainty on the branch lengths is considered). This hypothesis makes sense when the epidemiological and evolutionary timescales can be supposed to be similar [VKB13, PFR13].
- Incidence time data, sparsed between 0 and  $T_0$ . That is, the number of new cases registered through time, typically daily, weekly or monthly. This information may come from hospital records, surveillance programs (local or national), and is not necessarily collected at regular intervals.

We also take into account the sampling intensity of individuals that are currently infected through an additional parameter  $p$ , whose role is the following: each host at  $T_0$  is sampled independently with probability  $p$ .

There is an increasing number of works on modeling and inferring population dynamics from phylogenetic data, e.g. [VPW<sup>+</sup>09, Sta11, RBK13, SKBD13, SKRdP14]. However, important aspects remain to be explored and there are still many challenges so that phylo-dynamic inference methods can be extended to more complex systems and data [FPG<sup>+</sup>15]. In particular, few models exist linking sequences and incidence data, and most of them make further hypothesis about their independence, which is not always justified (e.g. in a context of dense sampling such as HIV epidemic). Therefore, in our approach, these observed statistics are assumed to be generated from a unique forward in time process and they are not independent in general. This assumption makes the computation of the likelihood as their joint distribution a delicate and complex issue, even in the linear BD model, since it requires to integrate over all the possible extinct (unobserved) subtrees between 0 and  $T_0$ .

We know from [Lam10] that the reconstructed phylogenetic tree of extant (sampled) individuals under this model, conditionally on the population to be non extinct at time  $T_0$  (i.e.  $I_{T_0} \neq 0$ ), is a coalescent point process, that is, a sequence of i.i.d. random variables  $H_i, i \geq 1$ , killed at the first value greater than  $T_0$ , as described in Subsection 1.4. We recall that at any level, individuals are labeled from left to right as we have explained in in Subsection 1.3, also depicted in Fig. 3, that is by the sequence  $(i)_{1 \leq i \leq I_{T_0}}$ . Additionally, for  $1 \leq i < j \leq X_{T_0} - 1$ , the  $i$ -th and  $j$ -th infected individuals coalesce (find the most recent common origin of their infection) at a time distributed as the maximum of  $j - i$  i.i.d. random variables with same law  $H$ , for instance  $H_{i+1}, \dots, H_j$ . In particular, conditional on

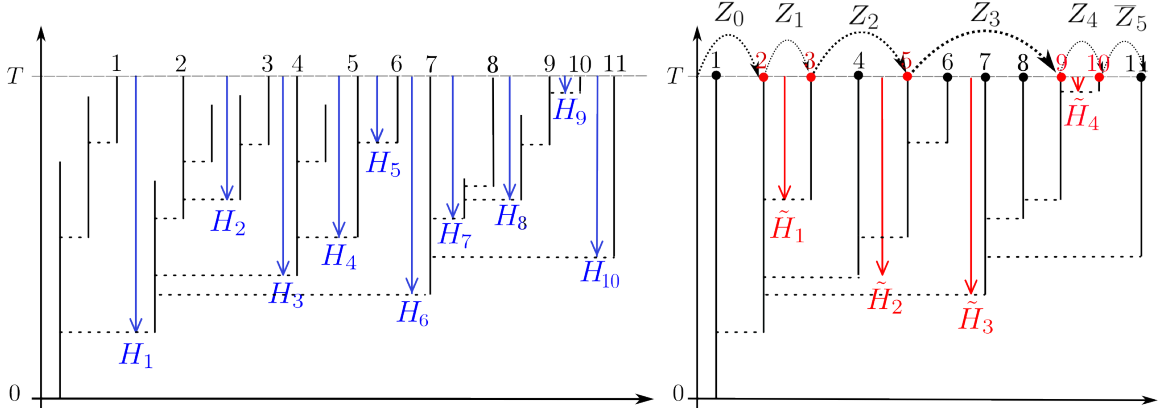


Figure 3 – An example of splitting tree between 0 and  $T_0$  (bold lines, left). The CPP of the 11 extant individuals at  $T_0$  (blue) and the CPP of the 5 sampled individuals (red). The variables  $H_i$  (in blue) represent the TMRCA between individuals  $i$  and  $i + 1$ , for  $1 \leq i \leq 10$ . The dashed curved arrows represent the sequence of variables  $(Z_i, 0 \leq i \leq \vartheta - 1, \bar{Z}_\vartheta)$ , counting the number of individuals between two sampled ones.

$X_{T_0} \neq 0$ ,  $X_{T_0}$  follows a geometric distribution with parameter  $\mathbb{P}(H < T_0)$ . The common law of these so-called branch lengths distributed as  $H$  is

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

where  $W$  is the scale function of a Lévy process  $X$  with Lévy measure  $\Pi$  and Laplace exponent defined by (2). In everything that follows, we suppose that  $H$  has a distribution that is absolutely continuous w.r.t. Lebesgue measure and its probability density function will be denoted by  $f$ . See Fig 3 (left) for an example of coalescent point process.

Furthermore, we include the sampling scheme described above, where each infected individual at  $T_0$  is sampled with probability  $p$ . Denote by  $\vartheta = \vartheta(p, T_0)$  the number of sampled individuals at any  $T > 0$ . Then, conditionally on  $\vartheta \geq 1$ , the resulting  $p$ -reconstructed tree is also a coalescent point process  $(\tilde{H}_i, 1 \leq i \leq \vartheta)$ , with a common law that we will denote by  $\tilde{H}$ , corresponding to the maximum of a sequence of  $G$  independent random variables distributed as  $H$ , where  $G \sim \mathcal{G}(p)$ , that is

$$\tilde{H} := \max_{1 \leq i \leq G} H_i.$$

Conditional on  $\vartheta$ , we let  $(\tilde{x}_i, 1 \leq i \leq \vartheta)$  be the labels of sampled individuals at  $T$ , that is, a subsequence of  $[1, I_T]$ . As well as in [Lam10, Theorem5.4], we have that conditional on  $\vartheta \geq 1$ ,  $(\tilde{H}_i, 1 \leq i \leq \vartheta)$  is distributed as a sequence of i.i.d. r.v., stopped at its first value greater than  $T$ . The common law of these branch lengths  $\tilde{H}$  is, for any  $x \geq 0$ ,

$$P(\tilde{H} \leq x) = \frac{pP(H \leq x)}{1 - pP(H \leq x)}.$$

Then, the typical distribution of the TMRCA between two consecutive sampled individuals at  $T$  (if any) is

$$P(\tilde{H} \leq x \mid \tilde{H} < T) = P(H \leq x \mid H < T) \frac{1 - pP(H \leq T)}{1 - pP(H \leq x)}, \quad \text{for } 0 \leq x \leq T.$$

See Fig 3 (right) for an example of  $p$ -coalescent point process.

### Random sampling at present time

Under this incomplete sampling model presented above, our goal is, in a first stage, to calculate the probability function of a vector of times corresponding to the coalescence times between the  $\vartheta$  sampled individuals, jointly with the population size at the present time,  $I_{T_0}$ . Since we cannot access to this likelihood directly as was pointed out before, we will rather look at the probability generating function of  $I_{T_0}$  under the event of observing the coalescence times vector  $\mathbb{T}_{\vartheta-1}$  and  $\vartheta = K$ , where  $\vartheta$  is the random number of sampled infected individuals at  $T_0$ . Let us define the function  $G : [0, 1] \rightarrow \mathbb{R}_+$  as follows

$$G(u) = \mathbb{E} \left[ u^{I_{T_0}} 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} \mid I_{T_0} \neq 0 \right]. \quad (9)$$

We have the following description (see Fig. 3),

- $Z_0 = \tilde{x}_1$ , the number of unsampled individuals until the first sampled one,
- $Z_i = \tilde{x}_{i+1} - \tilde{x}_i$ , the number of unsampled individuals between the  $i$ -th and the  $(i+1)$ -th sampled individuals for  $i = 1, \dots, \vartheta - 1$ ,
- $\bar{Z}_{\vartheta} = I_T - \tilde{x}_{\vartheta}$ , the number of unsampled individuals with label greater than the last sampled one,
- $H_j^{(i)}$ ,  $i, j \geq 1$  are i.i.d. random variables with same law as  $H$ ,
- $\tilde{H}_i = \max \{H_1^{(i)}, \dots, H_{Z_i}^{(i)}\}$  for  $i = 1, \dots, \vartheta - 1$ , i.e. the TMRCA between individuals  $\tilde{x}_i$  and  $\tilde{x}_{i+1}$ .

According to this notation, the total population extant at  $T_0$  can be expressed as (see Fig. I.5)

$$I_{T_0} = Z_0 + \dots + Z_{\vartheta-1} + \bar{Z}_{\vartheta}.$$

Additionally, it is not hard to see that individuals with a label smaller than  $\tilde{x}_1$  or greater than  $\tilde{x}_{\vartheta}$ , which are counted in variables  $Z_0$  and  $\bar{Z}_{\vartheta}$ , do not take part in any coalescent event of interest. Hence, we can establish the following result.

**Theorem.** *Let  $G(u)$  denote the probability generating function of  $I_{T_0}$  defined by (9). We have the following identity for all  $u \in [0, 1]$ ,*

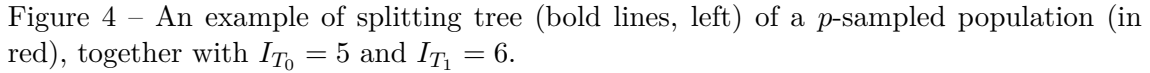
$$G(u) = \left( \prod_{i=1}^{K-1} \frac{p\mathbb{P}(H < t_i)u}{1 - (1-p)\mathbb{P}(H < t_i)u} \right) \times \frac{p\mathbb{P}(H > T_0)u}{(1 - \mathbb{P}(H < T_0)(1-p)u)^2}. \quad (10)$$

### Likelihood based on data at two times

Now we consider that we have, additionally to the data in the previous section, the incidence at a time  $T_1 < T_0$  (see Fig. 4). We want to calculate an analogous generating-type function, now in two variables  $u, v \in [0, 1]$ , that we will define below. First, to simplify, the vector  $\mathbb{T}_{K-1} = (t_1, \dots, t_{K-1})$  is supposed ordered, that is

$$0 < t_1 < \dots < t_{k-1} < T_0 - T_1 < \dots < t_{K-1},$$

so that the first  $k$  sampled individuals coalesce between present time  $T_0$  and  $T_1$ , and the rest of them between  $T_1$  and the more distant past time 0. Then, the first  $k-1$  coalescent events can be modeled as in the previous theorem, and those depths that might be greater than  $T_0 - T_1$  are studied with the help of some results on the JCCP of a splitting tree. What we should notice is that, when we include coalescence events that take place between


$$G(u, v) = \mathbb{E} \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{k-1} < t_{k-1}, \hat{H}_k < t_k, \dots, \hat{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} u^{I_{T_1}} v^{I_{T_0}} \mid I_{T_0} \neq 0 \right]. \quad (11)$$
$$\mathbb{P}(B = 0) = \frac{\mathbb{P}(H \leq T_0 - T_1)}{\mathbb{P}(H < T_0)}, \quad \mathbb{P}(B = 1) = \frac{\mathbb{P}(T_0 - T_1 < H < T_0)}{\mathbb{P}(H < T_0)} \quad (12)$$
$$N(x, T_0) := \sum_{i \geq 1} \mathbb{1} \left( \tau_x^i < \tau_0 \wedge \tau_{T_0}^+ \right),$$
$$\mathbb{P}_{T_1} \left( \tau_T^+ < \tau_s, N(T_1, T_0) = n \right) = (\rho_1(s))^n \rho_2(s)$$

with

$$\begin{aligned}\rho_1(s) &= 1 - \frac{W(T_0 - s)}{W(T_1 - s)W(T_0 - T_1)}, \\ \rho_2(s) &= \frac{W(T_0 - s) - W(T_0 - T_1)}{W(t - s)W(T_0 - T_1)}.\end{aligned}$$

To find an explicit expression for  $G(u, v)$  we also need to use the function  $h$  defined as follows for a random variable  $B$  with distribution given by (12), independent of  $X$ ,

$$h(v, s) := \mathbb{E} \left[ \left( 1_{\{X_{\tau_0 \wedge \tau_T^+} > s\}} v^{N+1} \right)^B \right] = \frac{\mathbb{P}(H \leq T_0 - T_1)}{\mathbb{P}(H < T_0)} + \frac{\rho_2(s)v}{1 - v\rho_1(s)} \frac{\mathbb{P}(T_0 - T_1 < H < T_0)}{\mathbb{P}(H < T_0)}. \quad (13)$$

Using (I.14) and the strong Markov property we also have that

$$\begin{aligned}\tilde{h}(v, x) &:= \mathbb{E}_x \left[ 1_{\{\tau_T^+ < \tau_0\}} v^{N(t, T)} \right] = \mathbb{P}_x \left( \tau_t \geq \tau_T^+ \right) + \mathbb{E}_t \left[ 1_{\{\tau_T^+ < \tau_0\}} v^{N(t, T)} \right] \mathbb{P}_x \left( \tau_t < \tau_T^+ \right) \\ &= \frac{\mathbb{P}(T - x < H \leq T - t)}{\mathbb{P}(H > T - x)} + \frac{\mathbb{P}(H > T - t)}{\mathbb{P}(H > T - x)} \frac{\rho_2(0)v}{1 - v\rho_1(0)}.\end{aligned} \quad (14)$$

We can finally state the result.

**Theorem.** *The function  $G(u, v)$  satisfies the following identity for  $h$  defined by (13),*

$$\begin{aligned}G(u, v) &= \left( \prod_{i=1}^{k-1} \frac{p\mathbb{P}(H < t_i)u}{1 - (1-p)\mathbb{P}(H < t_i)u} \right) \times \frac{pu}{(1 - \mathbb{P}(H < T)(1-p)u)} \times \frac{\rho_2(0)}{1 - v\rho_1(0)} \\ &\quad \prod_{i=k}^{K-1} \frac{p\mathbb{P}(H \leq T)h(v, T - t_i)u}{1 - (1-p)\mathbb{P}(H \leq T)h(v, T - t_i)u} \times \frac{h(v, 0)(1 - (1-p)\mathbb{P}(H < T))}{1 - (1-p)\mathbb{P}(H < T)uh(v, 0)} \times \tilde{h}(v, x),\end{aligned}$$

for a function  $h$  defined by (13) and  $\tilde{h}$  by (14).

Finally, a generalization of the previous theorems is obtained for  $n$  variables  $u, u_1, \dots, u_n \in [0, 1]$ , and the probability generating function defined as follows,

$$G(u, u_1, \dots, u_n) = \mathbb{E} \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} u^{I_T} u_1^{I_{T_1}} \dots u_n^{I_{T_n}} \mid I_T \neq 0 \right]$$

where  $T = T_0 > T_1 > \dots T_n > 0$  are fixed times between present time  $T$  and time 0, is assumed again to be the origin of the epidemic. This last result will be proved only in the birth and death case, and involves a description of the population size process, backward in time, as a sum of inhomogeneous branching processes with immigration.

## 2.2 Main results of Chapter II

The results obtained in this chapter appear in [DFL15], and are, in some sense, a generalization of those obtained in Chapter I. Indeed, here we consider the same model as in the previous chapter, denoting present time by  $T$ . Then, we obtain a characterization of the law of the infected population size process  $I := (I_t, 0 \leq t \leq T)$  as a continuous time process between 0 and  $T$ , conditionally on the reconstructed phylogeny of hosts at time  $T$ , when  $I$  is distributed as a CMJ and then the total phylogeny is a splitting tree, as depicted before. This is accomplished with the help of a more general result for branching processes that we will subsequently describe. The results are based in the identity in law between the JCCP of a splitting tree and a SPLP as it was stated in the preliminaries, and the duality property that enjoy the latter. Also, a few elements on excursion theory for Lévy processes are necessary.



## Definitions

Define a *forest* as any finite sequence of independent splitting trees, each starting from one individual. More specifically, consider a finite measure  $\Pi$  with mass  $b$  and for any positive integer  $m$ , let a  $m$ -forest be defined as  $\mathcal{F} = (\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m)$ , where  $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m$  are i.i.d. splitting trees with lifespan measure  $\Pi$ . It is straightforward to extend the notion of width process to a forest as the sum of the widths of every tree of the sequence, i.e.,

$$\Xi(\mathcal{F}) := \sum_{i=1}^m \Xi(\mathcal{T}_i).$$

As well, we can extend the notion of contour process to a forest  $\mathcal{F} = (\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m)$  of finite trees, similarly to the way it is done in [DLG02], by concatenating the contour functions,

$$\mathcal{C}(\mathcal{F}) := [\mathcal{C}_t(\mathcal{T}_1), \mathcal{C}_t(\mathcal{T}_2), \dots, \mathcal{C}_t(\mathcal{T}_m)].$$

It will be denoted simply  $\mathcal{C}$  when there is no risk of confusion. We should notice that the function thus obtained determines a unique sequence of chronological trees since they all start with one single ancestor.

Fix a finite horizon  $T$ . We call a forest stopped at 1st surviving tree, denoted by  $\mathcal{F}^*$ , a sequence of i.i.d. splitting trees stopped at the first tree that survives up until  $T$ .

Also define for  $p \in (0, 1)$  a forest  $\mathcal{F}^p$  as a sequence of independent splitting trees  $(\mathcal{T}_1, \dots, \mathcal{T}_{N_p}, \mathcal{T}_{N_p+1}^{(T)})$ , where,

- $\mathcal{T}_1, \dots, \mathcal{T}_{N_p}$ : are i.i.d. conditioned on extinction before  $T$
- $\mathcal{T}_{N_p+1}$ : is conditionned on survival up until time  $T$
- $N_p$ : is an independent geometric random variable with  $\mathbb{P}(N_p = k) = (1-p)^k p$ ,  $k \geq 0$

Consider now a finite variation spectrally positive Lévy process  $Y$  with Lévy measure  $\Pi$  and Laplace exponent  $\psi$  given by (2) (with  $\mathbf{d} = -1$ ). Consider  $\eta$  as in (II.2) and define an associated measure  $\tilde{\Pi}$  as follows

$$\tilde{\Pi}(dr) := e^{-\eta r} \Pi(dr).$$

Consider also an analogous process  $\tilde{Y}$  with Laplace exponent  $\tilde{\psi}$  defined as  $\psi$  but with Lévy measure  $\tilde{\Pi}$ . Let  $W$  and  $\tilde{W}$  denote the respective scale functions of  $\psi$  and  $\tilde{\psi}$ . From now on  $P$  denotes the probability measure of  $Y$  and  $\tilde{P}$  that of  $\tilde{Y}$ .

For a finite variation SPLP, the undershoot and overshoot at 0 of an excursion starting at 0 and conditional on  $\tau_0^+ < +\infty$  are defined as the amounts of the jump  $\Delta Y_{\tau_0^+}$  that are respectively below and above level 0. See [Kyp06] for the details and see Fig. 5 for an example. We denote respectively by  $\mu_{\top}$  and  $\mu_{\perp}(dv)$  their probability measures on  $[0, \infty)$  under  $P_0(\cdot | \tau_0^+ < +\infty)$ , defined as follows,

$$\mu_{\top}(du) := P_0(-Y_{\tau_0^+} \in du | \tau_0^+ < +\infty) = \frac{e^{-\eta u} \bar{\Pi}(u)}{m \wedge 1} du \quad (15)$$

$$\mu_{\perp}(dv) := P_0(Y_{\tau_0^+} \in dv | \tau_0^+ < +\infty) = \frac{e^{\eta v} \tilde{\Pi}(v)}{m \wedge 1} dv \quad (16)$$

We define forests  $\mathcal{F}_{\top}^p, \mathcal{F}_{\perp}^p$  distributed as  $\mathcal{F}^p$ , but with the lifetimes of the ancestors having the distribution  $\mu_{\top}$  and  $\mu_{\perp}$ , which differs from that of the rest of individuals in

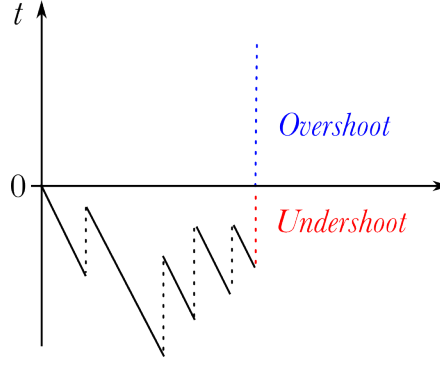


Figure 5 – Example of a finite variation SPLP process  $Y$  until it crosses level 0.

the trees, i.e.  $\Pi(\cdot)/b$ . In the sequel, the subscripts  $\perp$  and  $\top$  are used to refer to trees and forests where the ancestors have these particular distributions.

Finally define the following parameters  $\gamma$  and  $\tilde{\gamma}$  corresponding to the probabilities of non-extinction at time  $T$  of the populations in the trees  $\tilde{\mathcal{T}}_\top$  and  $\mathcal{T}_\perp$  respectively. Thanks to [Lam10, Th.4.3 and Th.5.4] these probabilities can also be expressed in terms of the scale functions  $W$  and  $\tilde{W}$ , and correspond to hitting time inequalities for the Lévy proceses  $Y$  and  $\tilde{Y}$ . More precisely we have the following,

$$\begin{aligned}\gamma &= \frac{1}{W(T)} = \mathbb{P}(\xi_T(\tilde{\mathcal{T}}_\top) \neq 0) = \tilde{P}_\top(\tau_T^+ < \tau_0) = P_T(\tau_0 < \tau_T^+), \\ \tilde{\gamma} &= \frac{1}{\tilde{W}(T)} = \mathbb{P}(\xi_T(\mathcal{T}_\perp) \neq 0) = P_\perp(\tau_T^+ < \tau_0) = \tilde{P}_T(\tau_0 < \tau_T^+).\end{aligned}$$

We prove that, in the supercritical case ( $m \geq 1$ ,  $\eta > 0$ ), a forest  $\tilde{\mathcal{F}}_\perp^\gamma$  is the same as a forest stopped at first surviving tree, where the splitting trees are distributed following the couple of measures  $(\perp, \Pi)$ , respectively for the ancestor and the descendants. In the same way,  $\tilde{\mathcal{F}}_\top^\gamma$  has the law of a forest stopped at first surviving tree with measures  $(\top, \tilde{\Pi})$ . This allows to obtain the following result of duality under time-reversal.

**Theorem.** *We have the following identity in distribution,*

$$(\xi_{T-t}(\tilde{\mathcal{F}}_\perp^\gamma), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\tilde{\mathcal{F}}_\top^\gamma), 0 \leq t \leq T)$$

*In the subcritical and critical cases (i.e.  $m \leq 1$ ),*

$$(\xi_{T-t}(\mathcal{F}_\perp^\gamma), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\mathcal{F}_\top^\gamma), 0 \leq t \leq T)$$

*and actually in this case  $\mu_\perp(dr) = \mu_\top(dr) = \frac{\bar{\Pi}(r)}{m}dr$ .*

In the exponential case, that is when  $\Pi(dr) = bde^{-dr}dr$  (for some  $d > 0$ ), the population size process of the splitting tree is a linear BD process. Moreover, for this choice of  $\Pi$ , the overshoot (16) and undershoot (15) are also exponentially distributed with parameter  $d$  and  $b \vee d$  respectively. Accordingly, in this case the previous theorem adopts a simpler form and can be stated as follows. Fix the parameters  $b$  and  $d$  such that  $b \geq d$  (supercritical case) and define two analogous forests stopped at 1st surviving tree:  $\mathcal{F}^*$  with birth and death parameters  $(b, d)$  and  $\tilde{\mathcal{F}}^*$  with birth and death parameters swapped, that is  $(d, b)$ ,

which is the same as the supercritical process conditioned on extinction [AN72]. Then, we have the next identity in distribution,

$$(\xi_{T-t}(\mathcal{F}^*), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\tilde{\mathcal{F}}^*), 0 \leq t \leq T).$$

Furthermore, the result obtained in [DFL15] is more general, since it concerns not only a duality between the population size processes forward and backward in time, but also on the genealogies of the forests. In other words, we establish that we can construct a dual forest  $\tilde{\mathcal{F}}^*$ , from the forest  $\mathcal{F}^*$ , by setting up different filiations between individuals, but letting the edges of the initial trees unchanged, as depicted in Fig 6. This new genealogy has no interpretation in terms of the original family to our knowledge, and can be seen as the tool to reveal the intrinsic branching structure of the backward-in-time process.

### Conditional on the reduced tree: applications to epidemiology

As a consequence of the previous results we are able to propose now a description of the population size process  $I := (I_t, 0 \leq t \leq T)$ , conditional on the coalescence times of hosts at time  $T$  (i.e. the reconstructed phylogeny) to be  $t_1, \dots, t_n$ . We consider that

$$I_t = \xi_t(\mathcal{F}_\perp^{\tilde{\gamma}}).$$

We state, under these conditions, that the process  $I$ , backward in time, is the sum of the width processes of  $n$  independent splitting trees, each conditioned on dying out at  $t_i$  for  $1 \leq i \leq n$ , plus an additional i.i.d. tree conditioned on surviving up until time  $T$ , see Fig. 7.

**Theorem.** Consider a forest  $\mathcal{F}_\perp^{\tilde{\gamma}}$  and let  $(H_i)_{i \geq 1}$  be the coalescence times from individuals at  $T$ . Define

$$\tilde{\mathbb{P}}_{T,i} := \tilde{\mathbb{P}}_T(\cdot | T_{Ext} = t_i), \quad \forall i \geq 1 \quad \text{and} \quad \tilde{\mathbb{P}}_{T,*} := \tilde{\mathbb{P}}_T(\cdot | T_{Ext} > T).$$

Then

$$(\xi_{T-t}(\mathcal{F}_\perp^{\tilde{\gamma}}), \mathbb{P}(\cdot | H_i = t_i, 1 \leq i \leq n)) = \left( \sum_{i=1}^{n+1} \xi_t(\mathcal{T}_i), (\tilde{\mathbb{P}}_{T,1} * \dots * \tilde{\mathbb{P}}_{T,*})(\cdot) \right).$$

The fact that in our model the sampled epidemic comes from one single ancestor at time 0 corresponds to different strains of the pathogen in their attempts to invade the population, but where only one succeeds (at time  $T$ ). However, if various invading strains succeed, analogous results can be deduced by concatenating (summing) an equal number of forests. The general assumption will be then, that for each *successful* strain, there is a geometric (random) number of other strains of the pathogen that become extinct before time  $T$ . The probability of success of these geometric r.v. depends on the recovery and transmission parameters. Finally, estimating these parameters from molecular and epidemiological data using this branching processes model, can be addressed through MLE or Bayesian inference. These statistical questions are not directly treated here.

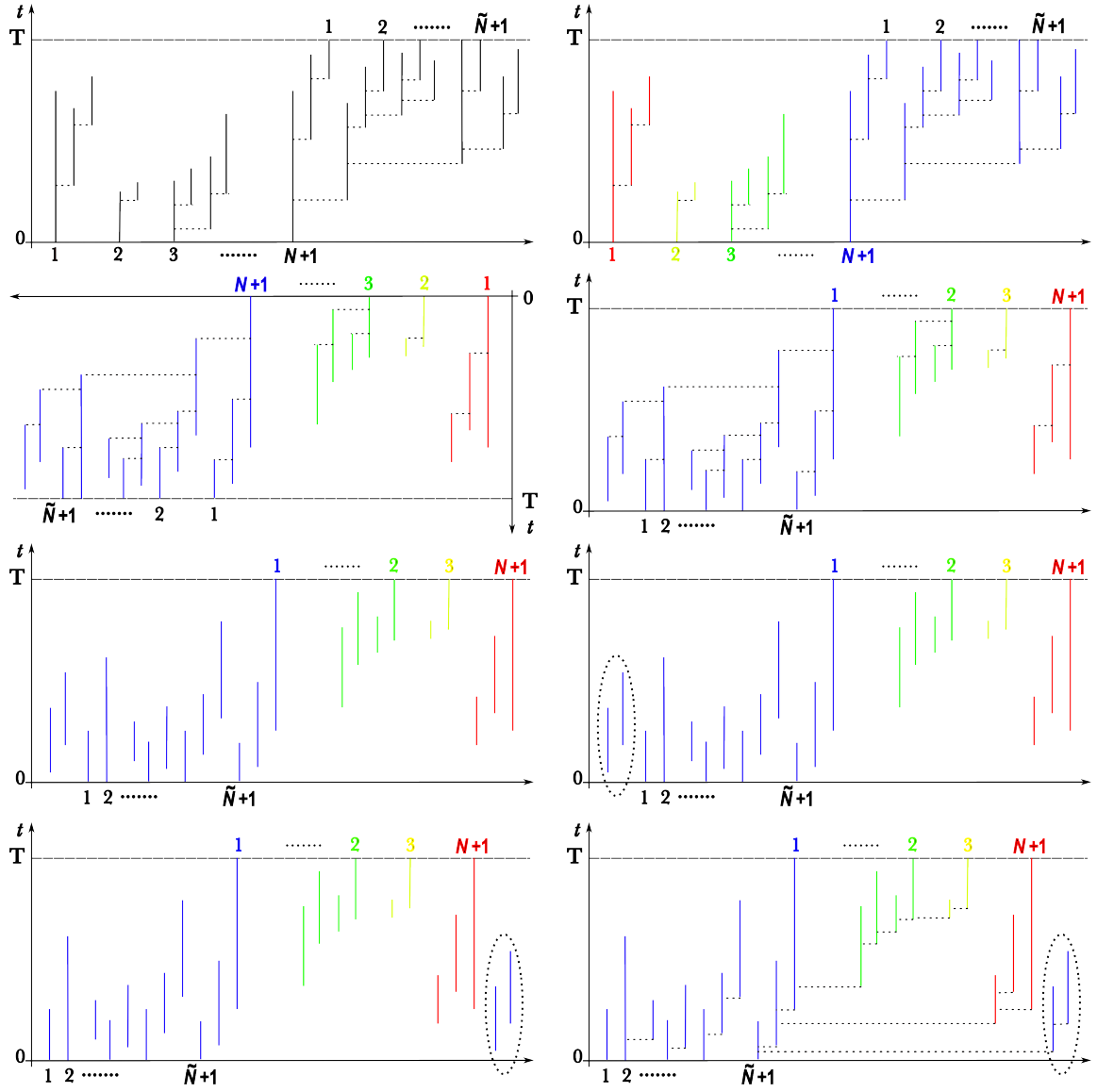


Figure 6 – Example of splitting forest stopped at first surviving tree,  $\mathcal{F}^*$  (top left) and the sequence of transformations leading to the dual forest  $\tilde{\mathcal{F}}^*$  (bottom right). Individuals belonging to the same family in  $\mathcal{F}^*$  are given the same color (top right), so that we can notice how they are related in the dual forest  $\tilde{\mathcal{F}}^*$  (bottom right).

### 2.3 Main results of Chapter III

This work is a follow-up to the previous chapter, where we have obtained a property of invariance under time-reversal, from a deterministic time  $T$ , for the branching process corresponding to the population size process of the forests stopped at 1st surviving tree. Now we study a similar property of duality, not from a deterministic time, but from the (random) extinction time of the process. It is worth stressing that, besides the implications concerning branching processes, some of our lemmas are interesting in their own right since they give some invariance results for subpaths of SPLP. The results are exposed in the paper [DFL16].

Let  $X = (X_t, t \geq 0)$  be a SPLP with Laplace exponent  $\psi$  given by (1). We suppose

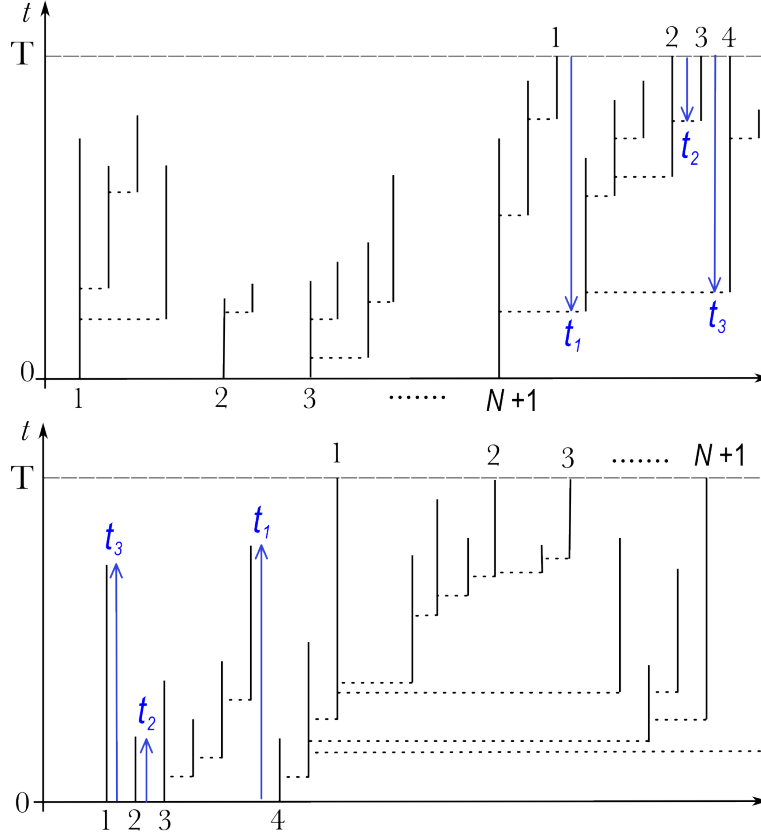


Figure 7 – Top: example of splitting forest stopped at first surviving tree,  $\mathcal{F}^*$  with 4 surviving individuals (hosts) at  $T$  and respective coalescence times between them equal to  $t_1, t_2, t_3$ . Bottom: the same forest after space-time-reversal.

that  $X$  is (sub)critical, meaning that it does not drift to  $+\infty$ , which is equivalent to  $\psi'(0+) \geq 0$ . We consider the process reflected at its infimum  $X - I$  as defined in the preliminaries, with the excursion measure  $\underline{n}$  normalized so that  $-I$  is the associated local time and  $\tau_{-t} = \inf\{s \geq 0 : X_s < -t\}$  the inverse local time.

Let  $\varepsilon$  be the generic excursion of  $X - I$  away from 0 and  $\gamma$  the last instant at which this excursion attains its supremum, that is

$$\gamma = \gamma(\varepsilon) = \sup\{s > 0 : \varepsilon_s = \bar{\varepsilon}_s\},$$

where  $\bar{\varepsilon}_s = \sup_{[0,s]} \varepsilon$ . We are interested in the decomposition of  $\varepsilon$  at  $\gamma$ .

Consider the following space-time-reversal transformation of the paths in  $\mathcal{E}$ , that we call *rotation*, and can be defined for any  $\omega \in \mathcal{E}$  as

$$\rho \circ \omega := (\omega_{V-} - \omega_{(V-s)-}, 0 \leq s \leq V).$$

We prove that the pre-supremum and the post-supremum processes, denoted respectively  $\overleftarrow{\varepsilon}_\gamma = (\varepsilon_s, 0 \leq s \leq \gamma)$  and  $\overrightarrow{\varepsilon}_\gamma = (\varepsilon_{\gamma+s} - \varepsilon_\gamma, \gamma \leq s \leq V)$ , are invariant under rotation. We summarize this result in the following proposition.

**Proposition.** *For any measurable functional  $h : \mathcal{E} \rightarrow \mathbb{R}_+$  we have the following identities*

under the excursion measure of  $X - I$

$$\begin{aligned}\underline{n}\left(h\left(\overleftarrow{\varepsilon}_\gamma\right)\right) &= \underline{n}\left(h \circ \rho\left(\overleftarrow{\varepsilon}_\gamma\right)\right), \\ \underline{n}\left(h\left(\overrightarrow{\varepsilon}_\gamma\right)\right) &= \underline{n}\left(h \circ \rho\left(\overrightarrow{\varepsilon}_\gamma\right)\right).\end{aligned}$$

Moreover, these results imply the following theorem, for which we need first to define the functional  $\chi : \mathcal{E} \rightarrow \mathcal{E}$  as

$$\chi(\varepsilon) := \left[\rho\left(\overleftarrow{\varepsilon}_\gamma\right), \rho\left(\overrightarrow{\varepsilon}_\gamma\right) + \varepsilon_\gamma\right],$$

where for any two elements  $\omega_1, \omega_2 \in \mathcal{E}$ ,  $[\omega_1, \omega_2]$  stands for their concatenation.

**Theorem.** *For every bounded measurable functionals  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  we have*

$$\underline{n}(F) = \underline{n}(F \circ \chi)$$

A first consequence of this theorem is the invariance under time reversal of the local time process of the excursion of  $X - I$  away from 0.

**Corollary.** *The local time process of the excursions of  $X - I$  away from 0, is invariant under time reversal, that is*

$$(\Gamma(\varepsilon, r), 0 \leq r \leq \varepsilon_\gamma) \stackrel{d}{=} (\Gamma(\varepsilon, \varepsilon_\gamma - r), 0 \leq r \leq \varepsilon_\gamma).$$

Moreover, the (sub)critical CMJ's branching process and excursion away from 0 of the critical Feller's branching diffusion, are invariant under time reversal from their extinction time.

The idea behind the second part of this Corollary, is again, the one-to-one relationship that can be established between a SPLP and branching process, in particular through the JCCP. Similar results concerning the duality by time-reversal of branching processes have been given in the literature. In particular, in [AP05] we can find a time-reversal invariance principle for the linear birth and death process in the critical case, when the process is conditioned on the number of individuals at the time of reversal to be equal to  $n$ . As suggested by the authors, the rescaled limit of the time-reversed process when  $n \rightarrow \infty$ , is the Feller branching diffusion. This suggests an alternative way of obtaining the second result on the previous Corollary. See also [Est75] and more recently [AR02, KRS07, DH13] for the treatment of the reverse of Galton-Watson processes and specifically the *Esty time reversal*, which is the limit obtained by conditioning a GW process in negative time upon entering the state 0 (extinction) at time 0 and starting in the state 1 at time  $-n$ , when  $n$  tends to  $+\infty$ . We also refer to [BD16] for a time reversal property for the number of ancestors process of a stationary CSBP with sub-critical quadratic branching mechanism.



# Chapter I

## Inferring population dynamics from virus phylogenies

*“The lack of real contact between mathematics and biology is either a tragedy, a scandal, or a challenge, it is hard to decide which.”*

– Gian-Carlo Rota, [Kac08]

### Introduction

Phyldynamic models seek to characterize the relationship between evolutionary and epidemiological processes by combining concepts from phylogenetics, epidemiology, population genetics and immunology into a mathematically explicit and statistically rigorous framework [WDD07]. This approach has helped to illuminate many aspects of infectious disease biology such as historical patterns of incidence, transmission between populations, cryptic structure due to heterogeneity in host susceptibility and behavior, and host-mediated selection on parasite genomes [GPG<sup>+</sup>04, PR09]. It has also been used as theoretical support for the evaluation of control strategies for the spread of infectious diseases [MSM<sup>+</sup>13]. In recent years, this field has witnessed the development of many new mathematical models and statistical tools, but these have mostly relied exclusively on pathogen sequences. Integrating multiple sources of data promises to provide deeper insights into infectious disease dynamics and better opportunities for reconstructing the past and predicting the future [VPW<sup>+</sup>09, RRK11].

When the population of susceptible individuals is sufficiently large (e.g. outbreaks), one can be placed in the context of a simplified model without density dependence, a process of birth and death at constant rates. A birth event should be interpreted as the transmission of the infection to a healthy individual and a death event as an infective individual that stops being contagious (recovery, death, etc.). The purpose of this work is to infer the parameters that control the dynamics of this simplified model by taking advantage of the information brought by the reconstructed phylogenetic tree (or coalescent) from living individuals (or only sampled), in the present time, and given the number of total cases at some other earlier times. On that account, we consider the first model described in Subsection 1.3, that of splitting trees and Crump-Mode-Jagers (CMJ) processes. Then, the population of infected individuals follows this model from time 0, where the epidemic originated, until present time  $T$ . More precisely we suppose the transmission tree of an epidemic between 0 and  $T$  is distributed as a splitting tree, denoted by  $\mathcal{T}$ , with lifespan



measure  $\Pi$  of mass  $b$ . Then the infected population size process, denoted by  $(I_t)_{0 \leq t \leq T}$ , has the law of the associated CMJ process, that is, for each  $0 \leq t \leq T$ ,

$$I_t = \xi_t(\mathcal{T}),$$

where  $(\xi_t(\mathcal{T}), t \geq 0)$  denotes the *width* process of a splitting tree, which counts the number of particles alive in the tree over time as defined in the Introduction. Recall that birth and death processes are a particular case of this model. See Fig I.1.

In recent years, interest in using phylogenies of extant taxa to infer the patterns of diversification has considerably grown. In [NMH94] authors apply the birth-death process to phylogenies of extant taxa by introducing the reconstructed birth-death process. It was the first likelihood based-method presented to infer speciation and extinction rates on the basis of *reconstructed* phylogenies. This method has been widely used for estimating speciation and extinction rates and several likelihood-based approaches now exist that infer these rates under different scenarios (see for instance [AP05]). In forthcoming Subsection 2.1 we give a brief review of some of these different approaches, where there is a common basic idea: speciation and extinction rates are determined that maximize the likelihood of the reconstructed tree. However, even in the linear birth-death model, one of the simplest one might consider, the quantification of the likelihood on the basis of available data can be a delicate and complex issue. Especially when one considers that the state of the population at the present is only partially observed, i.e. the population is incompletely sampled. If additionally, we have some extra information about the past, such as the total extant population at earlier moments, the likelihood function can be even more complicated to characterize.

Here, we obtain expressions for the probability generating function of the number of cases in the present and/or at prior times, conditional on the coalescent or reconstructed tree from present time hosts, assuming a splitting tree as underlying model. We exploit the fact that the jumping chronological contour process (JCCP) of these trees with finite length is a Lévy process, that is a càdlàg process with stationary and independent increments (see Subsection 1.4 for more details). We recall that the genealogy of a splitting tree conditioned to be extant at a fixed time  $T$  ( $I_T \neq 0$ ) is given by a *coalescent point process* (CPP), that is, a sequence of i.i.d. random variables  $H_i, i \geq 1$ , killed at its first value greater than  $T$ . More specifically, for  $1 \leq i < j \leq I_T - 1$  the time elapsed into the past until the  $i$ -th and  $j$ -th individuals alive at time  $T$  coalesce, i.e. find their *most recent ancestor* (TMRCA), is distributed as the maximum of  $j - i$  i.i.d. random variables with same law as  $H$ , for instance  $H_{i+1}, \dots, H_j$ . In particular, conditional on  $I_T \neq 0$ ,  $I_T$  follows a geometric distribution with parameter  $\mathbb{P}(H < T)$ . The common law of these so-called node depths is characterized with the help of the scale function of the JCCP of the splitting tree [Lam10].

Additionally, we consider that each host at time  $T$  is sampled independently with probability  $p \in (0, 1)$ , so the genealogy of the sample is still that of a coalescent point process with a typical node depth distributed as

$$\tilde{H} := \max_{1 \leq i \leq \xi} H_i$$

where  $\xi \sim \mathcal{G}(p)$  (geometric) is independent from the i.i.d. sequence  $\{H_i\}_{i \geq 1}$ .

We denote by  $\vartheta$  the number of sampled individuals at time  $T$ , and we consider that the observed data consists in the sequence  $(\tilde{H}_i)_{1 \leq i \leq \vartheta-1}$  and the vector  $(I_T, I_{T_1}, \dots, I_{T_N})$ , where  $T > T_1 > \dots > T_N > 0$  are deterministic times. Then, our main results consist

in characterizations of the probability generating function of these variables, defined as follows, for any  $u, u_1, \dots, u_N \in [0, 1]$ ,

$$G(u, u_1, \dots, u_N) = \mathbb{E} \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} u^{I_T} u_1^{I_{T_1}} \dots u_N^{I_{T_N}} \mid I_T \neq 0 \right].$$

We give an explicit expression for this function in the cases of  $N = 1, 2$  considering the general CMJ process as underlying model. A characterization is also obtained for  $N \geq 3$ , but only in the birth and death case, where the function  $G$  is expressed as the composition and product of several probability generating functions (*pgf*), each of them well characterized.

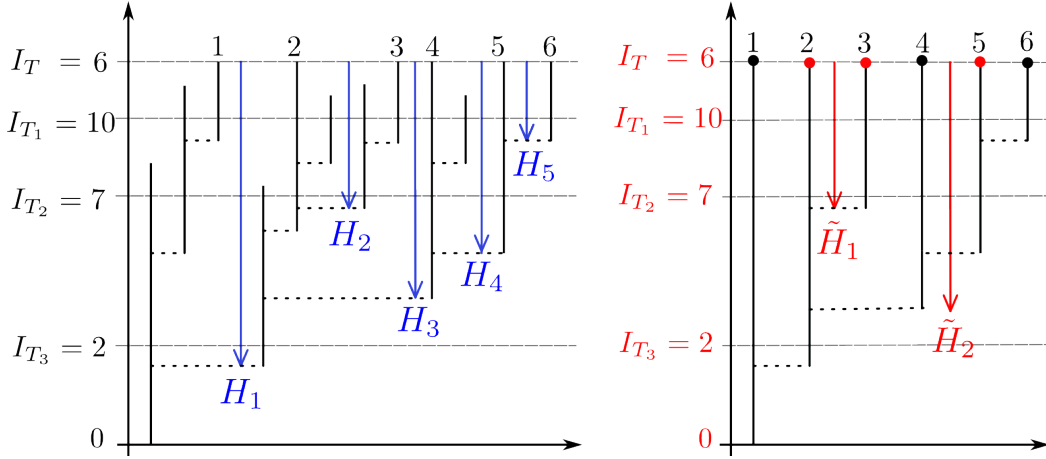


Figure I.1 – An example of splitting tree up to time  $T$  (left, bold lines), the coalescent point process  $(H_i)_{1 \leq i \leq 5}$  from extant individuals at  $T$  (left, blue arrows) and the reduced tree (right, bold lines). The population size at times  $T, T_1, T_2, T_3$  is represented at the respective levels. The red tips in the right panel represent sampled individuals at  $t$ , with respective coalescence times  $(\tilde{H}_i)_{i=1,2}$

The chapter is organized as follows. There is a section of preliminaries, where we give some basic notions about reconstructed trees and coalescent processes. A section containing our main results follows, preceded by a brief review on some previous works concerning the problem of likelihood computation and a remind on spectrally positive Lévy processes [Ber96], which are indispensable in our approach [Lam10]. At the end we present some conclusions and perspectives of this work.

## 1 Preliminaries

The evolutionary relationships between a set of organisms or groups of organisms, called taxa, are most often represented as phylogenetic trees. Then, for our purposes, we need to consider the trees associated to the processes, from the beginning of the epidemic at time 0 with a single host (*origin* of the population), up to present time  $T$ . We will first consider a linear birth and death model and the associated family tree. We call *reduced* or *reconstructed tree*, and also reconstructed phylogeny or coalescent from individuals at a fixed time  $T$ , the subtree spanned by infected individuals at  $T$ , consisting in an oriented (to the right), ultrametric binary tree characterized by the coalescence times (or node

depths) between extant hosts, i.e. the extinct lineages are removed from the complete tree Fig. I.2 (left).

We will describe now a non-homogeneous Markov process associated to the birth and death tree, which is related to the reconstructed tree. This new process allows us to formulate the main questions we intend to answer in this work. Afterwards, we recall some results concerning the coalescent process derived from a more general class of branching processes, the Crump-Mode-Jagers (CMJ) processes [Lam10, Lam11].

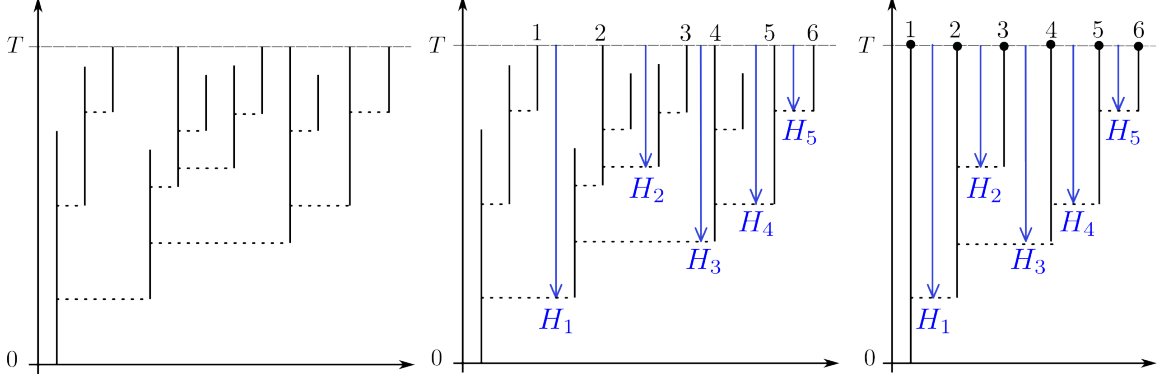


Figure I.2 – An example of total tree up to time  $T$  (left), the coalescence times between extant individuals at  $T$  (blue arrows) and the reduced tree (left)

### 1.1 Model: non-homogeneous Markov processes associated to a linear birth-death model

On a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  let us consider a continuous time birth and death process  $(I_t)_{t \geq 0}$ , with constant rates  $\lambda$  and  $\mu$  starting from  $I_0 = 1$ . From the biological point of view this model represents a population where individuals evolve independently from each other and reproduce or die at exponential times with rates  $\lambda$  and  $\mu$  respectively. Then,  $I_t$  represents the number of individuals alive at  $t \geq 0$ . As we mentioned in the introduction, we have in mind applications of this model to the field of epidemiology. Hence, the birth and death process represents the transmission process of a pathogen in a sufficiently large susceptible population (no density dependence is considered here).

Consider a fixed time  $T$  corresponding to present time. We will now define a *non-homogeneous Markov process* on  $\mathbb{N}^2$ , denoted  $\{Y_t, Z_t\}_{t \geq 0}$ , associated to  $I$ , that can be defined for  $0 \leq t \leq T$  as follows (for abbreviation the dependence on  $T$  is omitted)

- $Y_t$  is the number of individuals at  $t$  having extant descendants at time  $T$ ,
- $Z_t = I_t - Y_t$ .

According to this definition  $(Y_0, Z_0) = (1, 0)$  conditional on  $I_T > 0$ . If we denote by  $p_0(t) = \mathbb{P}_1(I_t = 0)$ ,  $0 \leq t \leq T$ , that is the probability of extinction of the population before time  $t$ , and by  $q(t)$  the probability of surviving (clearly  $p_0(t) + q(t) = 1 \ \forall 0 \leq t \leq T$ ), the process  $(Y, Z)$  will have the following transition dynamics, conditional on  $(Y_t, Z_t) = (n, m)$ :

$$(Y_t, Z_t) \begin{cases} (n, m) \rightarrow (n-1, m) & \text{with rate } 0, \\ (n, m) \rightarrow (n+1, m) & \text{with rate } \lambda n q(T-t), \\ (n, m) \rightarrow (n, m-1) & \text{with rate } \mu m / (1 - q(T-t)), \\ (n, m) \rightarrow (n, m+1) & \text{with rate } \lambda(m+n)(1 - q(T-t)). \end{cases}$$

Kolmogorov equations lead to the following ordinary differential equation satisfied by  $p_0$  (see for instance [AN72])

$$\begin{cases} p'_0(t) &= \mu - (\lambda + \mu)p_0(t) + \lambda p_0^2(t) \\ p_0(0) &= 0 \end{cases},$$

which admits an explicit solution

$$p_0(t) = \frac{\mu - \mu e^{(\lambda - \mu)t}}{\mu - \lambda e^{(\lambda - \mu)t}}.$$

More generally, if we denote by  $p_n(t)$  the probability that a lineage leaves  $n$  descendants at time  $t$ , a result from [Ken49] states that for  $\lambda > \mu \geq 0$ , setting  $a = \frac{\mu}{\lambda}$  and  $r = \lambda - \mu$ , we have

$$\begin{aligned} p_0(t) &= \mathbb{P}_1(I_t = 0) = \frac{\mu(e^{rt} - 1)}{\lambda e^{rt} - \mu}, \\ p_1(t) &= \mathbb{P}_1(I_t = 1) = \frac{r^2 e^{-rt}}{(\lambda - \mu e^{-rt})^2}, \\ p_n(t) &= a^{n-1} p_1(t) (p_0(t))^{n-1} \quad n \geq 2. \end{aligned}$$

With the above description we can set the problem and describe the available data we focus on here. Consider a finite sequence of times  $0 \leq s_1 \leq s_2 \leq \dots \leq s_k \leq T$ . Our goal is to describe the joint law of  $(Y_s)_{0 \leq s \leq T}$  and  $I_{s_1}, I_{s_2}, \dots, I_{s_k}$ , so we can estimate the parameters of the model (for instance  $\lambda, \mu$ ) based on the observation of these variables. Additionally, we will incorporate later a sampling scheme to this model.

## 1.2 Reconstructed phylogenetic tree

The situation depicted in the previous section concerns the dynamics of birth and death processes, but indeed, some of the results we obtain here are established in a more general branching model, a *binary homogeneous Crump-Mode-Jagers process* (CMJ) [Lam10, Lam11]. A CMJ process describes a population where individuals reproduce independently of each other, have i.i.d. lifetime durations with arbitrary distribution (not necessarily exponential), and give birth at constant rate during their lifetime, giving rise to a single offspring at each birth event. A particular aspect of this model is that, since no assumption is made on the distribution of the lifetime durations of individuals, the process is not necessarily Markovian (unless the lifetime distribution is exponential or a Dirac mass at  $\{\infty\}$ ). The tree associated to the genealogy of a CMJ process is a *splitting tree*. We refer to the introduction and to [Lam10] for a more complete description of these processes.

As pointed out before, our aim is to derive an analytic formula for the likelihood of the reconstructed transmission tree under these dynamics, jointly with the population size process. To be more precise, we are interested in computing the likelihood of the reconstructed tree from  $N$  individuals alive at time  $T$  which coalesced at times  $t_1, t_2, \dots, t_{N-1}$ , derived from a CMJ process that started with a single individual at time 0 ( $I_0 = 1$ ), as in Fig. I.3. It should be noted that we are not interested in the topology of the coalescence process since the likelihood does not depend on the topology of the reconstructed phylogeny, see for instance [Tho75] or [EHS<sup>+</sup>11, LS13], we will come back to this fact later.

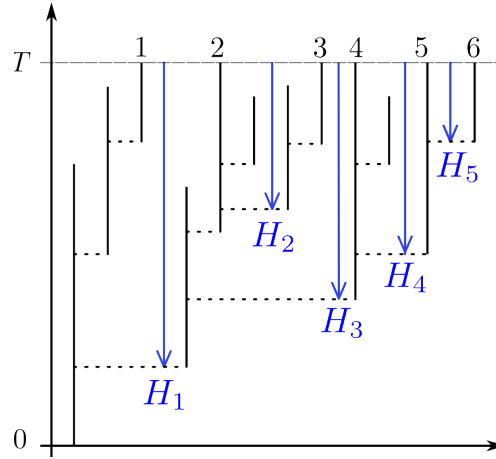


Figure I.3 – A representation of a splitting tree (bold lines), with the descendants placed to the right of their parent and the reconstructed tree from 6 extant individuals at  $T$  (blue arrows), the coalescent point process  $(H_i)_{1 \leq i \leq 5}$ .

In a first step we assume that the total population extant at time  $T$  is observed,  $I_T = N$  and their respective coalescence times are available. We start by presenting different approaches found in the literature to solve this problem, where the constraints on the calculation of the likelihood may vary depending on the available data and other considerations. They all lead to slightly different, but equivalent expressions. We will focus our attention on the calculation through the coalescent point process, which is the approach we will use to extend the result to the case of incomplete sampling at the present, with possibly available data at present and prior times.

### 1.3 Some results about the coalescent point processes

As we have explained in the introduction, in [Lam10] the author establishes that the genealogy of a splitting tree conditioned to be extant at a fixed time  $T$  ( $I_T \neq 0$ ) is given by a *coalescent point process*, that is, a sequence of i.i.d. random variables  $H_i, i \geq 1$ , killed at its first value greater than  $T$ . In other words, for  $1 \leq i < j \leq I_T - 1$  the time elapsed into the past until the  $i$ -th and  $j$ -th individuals alive at time  $T$  find their TMRCa, is distributed as the maximum of  $j - i$  i.i.d. random variables with same law as  $H$ , for instance  $H_{i+1}, \dots, H_j$ . In particular, conditional on  $I_T \neq 0$ ,  $I_T$  follows a geometric distribution with parameter  $\mathbb{P}(H < T)$ .

In everything that follows, we suppose that  $H$  has a distribution that is absolutely continuous w.r.t. Lebesgue measure and its probability density function will be denoted by  $f$ . The common law of these so-called node depths distributed as  $H$  is [Lam10]:

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

where  $W$  is a non-negative, nondecreasing, differentiable function, which is characterized by its Laplace transform and is called *scale function* (of a Lévy process). Again, we refer to the introduction for a precise definition of this function.

Even if most of the results presented here are valid for a general CMJ processes, we will often specify the implications in the Markovian case governed by the exponential distribution, which is the most widely exploited case in the literature. In this case  $I$  is a

linear birth-death process and the scale function takes the form

$$W(x) = \frac{\lambda e^{rx} - \mu}{r}, \quad \text{for } \lambda \neq \mu, \quad x \geq 0, \quad \text{and} \\ W(x) = 1 + \lambda x, \quad \text{for } \lambda = \mu \quad x \geq 0,$$

and the probability density function of  $H$  is

$$f(s) = \frac{-W'(s)}{W^2(s)} = \frac{\lambda r^2 e^{rs}}{(\mu - \lambda e^{rs})^2} = \lambda p_1(s). \quad (\text{I.1})$$

If we denote by  $f_T$  the probability density function of the variable  $H$  conditioned on  $\{H < T\}$  [Lam11], we have for every  $0 < s \leq T$ ,

$$f_T(s)ds = \mathbb{P}(H \in ds \mid H < T) = \frac{f(s)ds}{1 - \frac{1}{W(T)}} = \left( \frac{\mu - \lambda e^{rT}}{\lambda(1 - e^{rT})} \right) \left( \frac{\lambda r^2 e^{rs}}{(\mu - \lambda e^{rs})^2} \right) ds = \frac{\mu p_1(s)}{p_0(T)} ds.$$

## 2 Likelihood computation

### 2.1 Completely sampled population: review of previous results

As we said in the introduction, Nee et al. [NMH94], were the first to present a likelihood-based method to infer speciation and extinction rates on the basis of *reconstructed* phylogenies [Sta11]. Before them, some authors have already been interested in the derivation of an analytic formula for the likelihood of the reconstructed tree. Let us briefly review some previous solutions to this problem, included that of [NMH94].

As we defined at the end of Subsection 1.1,  $p_n(t)$  is the probability that a lineage leaves  $n$  descendants at time  $t$  under a birth-death process. Different expressions have been computed for the density of coalescence times, i.e. of observing  $N$  lineages at present time  $T$  that coalesce at times  $t_1 > t_2 \dots > t_{N-1}$  in the past. We denote by  $\mathbb{T}_{N-1} = (t_1, \dots, t_{N-1})$  the vector of these coalescence times and we set again  $a = \frac{\mu}{\lambda}$  and  $r = \lambda - \mu$ .

[Tho75] (pp. 54-58) In the Theorem on page 56, the author states that the density of the *evolutionary tree* satisfies that:

$$\text{Lik}(\mathbb{T}_{N-2}, N, F \mid \lambda, \mu, T) \text{ is proportional to } \prod_{j=1}^{N-2} (p_1(t_j))$$

Where  $F$  is the topology of the tree. A remarkable fact about this result is that it shows the likelihood to be independent of  $F$ , as we pointed out before.

[NMH94] (pp. 308, eq. (21)) They obtain the likelihood for a phylogenetic tree as follows, considering  $t_1$  (the coalescence time of the last two groups of individuals, also called the crown age) as the origin of the tree

$$\text{Lik}(\mathbb{T}_{N-1}) = (N-1)! r^{N-2} \exp \left( r \sum_{n=1}^{N-2} t_{n+1} \right) (1-a)^N \\ \times \prod_{n=1}^{N-1} \frac{1}{(\exp(rt_n) - a)^2}$$

[Ran97] (pp. 419, eq. (8)) He calculates the joint density of coalescence times of  $N$  individuals that are descended from a single ancestral lineage that appeared at time 0, conditioned on the present population size to be  $N$ :

$$\text{Lik}(\mathbb{T}_{N-1} \mid I_T = N, T, \lambda, \mu) = (N-1)! \prod_{i=1}^{N-1} \left( \mu \frac{p_1(t_i)}{p_0(T)} \right) \quad (\text{I.2})$$

In our approach the likelihood of coalescence times  $\{t_1, \dots, t_{N-1}\}$  jointly with  $I_T = N$  can be seen as a sequence of i.i.d. random variables  $H_i$ ,  $i \geq 1$ , where the first  $N-1$  are equal to  $t_i$ ,  $1 \leq i \leq N-1$  and are killed at the first value greater than  $T$ :

$$\text{Lik}(\mathbb{T}_{N-1} \mid I_T \neq 0, T, \lambda, \mu) = (N-1)! \prod_{i=1}^{N-1} f(t_i) \mathbb{P}(H > T)$$

All the above expressions are equivalent and can be obtained from each other if we take into account the different considerations. Let us just see briefly how our formula agrees with Equation (I.2) by [Ran97] :

$$\begin{aligned} \text{Lik}(\mathbb{T}_{N-1}, I_T = N \mid I_T \neq 0) &= \text{Lik}(\mathbb{T}_{N-1} \mid I_T = N) \mathbb{P}(I_T = N \mid I_T \neq 0) \\ &= (N-1)! \left( \prod_{i=1}^{N-1} \left( \mu \frac{p_1(t_i)}{p_0(T)} \right) \right) (1 - \mathbb{P}(H > T))^{N-1} \mathbb{P}(H > T) \\ &= (N-1)! \left( \prod_{i=1}^{N-1} f_T(t_i) \mathbb{P}(H < T) \right) \mathbb{P}(H > T) \end{aligned}$$

which corresponds to our expression as we announced. Note that conditional on  $I_T = N$ , the node depths are independent copies of  $H$  conditioned on  $H < T$ .

Then, using Equation (I.1) for the density on the Markovian case we have:

$$= (N-1)! r^{N-1} (1-a)^N \frac{1}{\lambda e^{rT} - a} \exp \left( r \sum_{i=1}^{N-1} t_i \right) \prod_{i=1}^{N-1} \frac{1}{(\lambda e^{rt_i} - a)^2}$$

## 2.2 Incomplete sampling: model description and some results

Now will be assumed that at present time  $T$  we do not observe all the extant individuals but only a *fraction* of them. Indeed, we will do as follows: each individual alive at time  $T$  will be sampled independently with probability  $p \in (0, 1)$ , so that the sampling action can be considered as a sequence  $B_i$ ,  $i \geq 1$  of i.i.d. Bernoulli trials. As it is shown in [LT13], if a coalescent point process is sampled in this way, then the genealogy of the sample is still that of a coalescent point process with a typical node depth distributed as

$$\tilde{H} := \max_{1 \leq i \leq \xi} H_i$$

where  $\xi \sim \mathcal{G}(p)$  is independent from the i.i.d. sequence  $\{H_i\}_{i \geq 1}$ .

We see clearly that the law of  $\tilde{H}$  is totally characterized by that of  $H$  and parameter  $p$ , which can be summed up by the following statement (proof in the Appendices).

**Lemma 1.** *Let  $f$  denote the probability density function of  $H$ . Then, the probability density function of  $\tilde{H}$ , denoted by  $\tilde{f}$ , is*

$$\tilde{f}(t) = \frac{f(t)p}{(1 - \mathbb{P}(H \leq t)(1-p))^2}, \quad \text{for } t \geq 0.$$

We will establish and prove some results concerning the reconstructed tree (Fig. I.4) obtained by this sampling scheme, which will be used later for the calculation of the likelihood.

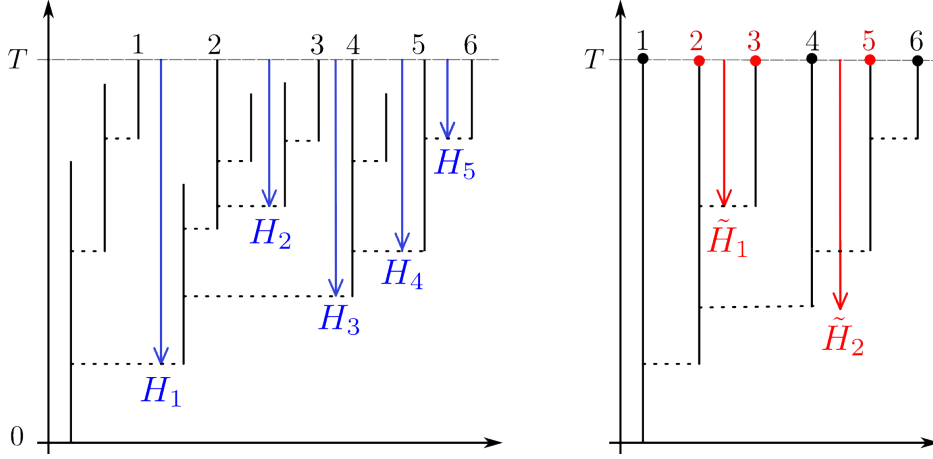


Figure I.4 – A coalescent point process where extant individuals are sampled with probability  $p$ . From a total population of 6 individuals at  $T$ , only 3 were sampled (in red). In this case  $\vartheta = 3$  and the labels of sampled individuals are  $\tilde{x}_1 = 2, \tilde{x}_2 = 3, \tilde{x}_3 = 5$ . Red lines represent the coalescence times between two consecutive sampled individuals.

### Computing the law of the number of sampled individuals

We are now interested in the total number of sampled individuals at  $T$ , that we denote by  $\vartheta := \vartheta(p, T)$ , see Fig. I.4. Conditional on  $I_T = N$ ,  $\vartheta$  follows a binomial distribution with parameters  $(p, N)$ . Nevertheless, we are more interested in determining its distribution without conditioning on the total population size, but just on the fact that  $I_T \neq 0$ . The result is the following.

**Proposition 2.** *Conditionally on  $I_T \neq 0$ , the r.v.  $\vartheta$ , counting the number of sampled individuals at  $T$  has the following probability generating function*

$$\mathbb{E}[u^\vartheta] = \frac{\mathbb{P}(H > T)(1 - p) + p\mathbb{P}(H > T)u}{1 - (1 - p)\mathbb{P}(H \leq T) - (1 - p)\mathbb{P}(H > T)u}, \quad \text{for } 0 \leq u \leq 1.$$

**Proof.** We know that conditional on  $I_T \neq 0$ , the number of individuals alive at  $T$  follows a geometric distribution, hence, we can express  $\vartheta$  as follows,

$$\vartheta = \sum_{i=1}^G B_i,$$

where  $B_i \sim \text{Bernoulli}(p)$ ,  $i = 1, 2, \dots$  and  $G \sim \mathcal{G}(\mathbb{P}(H > T))$  are independent random variables. This represents the number of sampled individuals from a total of  $G$  extant individuals at  $T$ , conditional on  $I_T$  to be nonzero. In order to simplify the notation we define  $\alpha = p$ ,  $\beta = \mathbb{P}(H > T)$  and  $\gamma = 1 - (1 - \alpha)(1 - \beta)$ . We work with these parameters



to make the calculations and we replace them at the end. We have that

$$\begin{aligned}\mathbb{E}[u^\vartheta] &= \mathbb{E}\left[u^{\sum_{i=1}^G B_i}\right] = \sum_{n \geq 1} \mathbb{E}\left[u^{\sum_{i=1}^n B_i}\right] \mathbb{P}(G = n) = \sum_{n \geq 1} \left(\mathbb{E}[u^{B_1}]\right)^n \mathbb{P}(G = n) \\ &= \sum_{n \geq 1} (u\alpha + (1 - \alpha))^n (1 - \beta)^{n-1} \beta = \frac{\beta(1 - \alpha) + \alpha\beta u}{\gamma - (1 - \alpha)\beta u}\end{aligned}$$

Replacing  $\alpha = p$ ,  $\beta = \mathbb{P}(H > T)$ , and  $\gamma$  as defined before leads to the result.  $\square$

### Comparing two geometric variables

In what follows we will often need to evaluate expressions involving the comparison between two geometric random variables with different parameters. Thus, we give the following result that can be easily shown through very simple calculations, which are included in the Appendix for the sake of completeness.

**Lemma 3.** *Let  $X_1$  and  $X_2$  be two independent geometric random variables with parameters  $\alpha$  and  $\beta$  respectively,  $\alpha, \beta \in (0, 1)$ . Then, conditionally on  $\{X_1 \leq X_2\}$  or  $\{X_1 < X_2\}$ ,  $X_1$  also follows the geometric distribution with parameter  $\gamma = 1 - (1 - \alpha)(1 - \beta)$ . Moreover, the probability generating functions of  $X_1$  under  $\{X_1 \leq X_2\}$  and  $X_1$  under  $\{X_1 < X_2\}$  are, for  $|u| \leq 1$ ,*

$$\mathbb{E}[u^{X_1} 1_{\{X_1 \leq X_2\}}] =: g(u, \alpha, \beta) = \frac{\alpha u}{1 - u(1 - \gamma)}, \quad (\text{I.3})$$

$$\mathbb{E}[u^{X_2} 1_{\{X_2 < X_1\}}] = g(u, \beta, \alpha)(1 - \alpha) = \frac{\beta(1 - \alpha)u}{1 - u(1 - \gamma)}. \quad (\text{I.4})$$

We keep the definition of  $\gamma$  for the rest of the work, depending on  $\alpha$  and  $\beta$ . We also state the following (trivial) result, which despite being very simple will be repeatedly used in the sequel.

**Lemma 4.** *For any sequence of i.i.d. random variables  $\{X_i\}_{i \geq 1}$  and an independent discrete r.v.  $N$  taking its values in  $\mathbb{N}$ . For every  $x, u \in \mathbb{R}$ ,  $|u| \leq 1$ , it holds that:*

$$\mathbb{E}\left[1_{\{X_1 \leq x, X_2 \leq x, \dots, X_N \leq x\}} u^N\right] = \mathbb{E}\left[(\mathbb{P}(X_1 \leq x)u)^N\right]$$

### 2.3 Sampled reconstructed tree from the total population at present time

Our goal in this section is to characterize the probability distribution of the total population size at  $T$ , conditionally on the coalescence times of sampled individuals, following the sampling model presented before. This is a delicate and complex issue, even in the linear birth-death model, since it requires to integrate over all the possible extinct (unobserved) subtrees between 0 and  $T$ , so it is not possible to compute the likelihood directly. Therefore, we will rather look at the probability generating function of  $I_T$  on the event of observing the TMRCA between sampled individuals at  $T$ , to be smaller than the times in vector  $\mathbb{T}_{\vartheta-1} = (t_1, \dots, t_{\vartheta-1})$  and the number of sampled individuals to be  $\vartheta = K$ . We will exploit the previously stated fact that the genealogy of a splitting tree conditioned to be extant at a fixed time  $T$  is given by a coalescent point process.

We remind the labeling order introduced in Subsection 1.3 for the total population of the splitting tree at a given level, say  $T$ . Individual labels are denoted by  $(x_i, 1 \leq i \leq I_T)$

and we set  $x_i = i$  for  $i \geq 1$ . Now, let  $(\tilde{x}_i, 1 \leq i \leq \vartheta)$  be the labels of sampled individuals at  $T$ , that is, a subsequence of  $[1, I_T]$ . Let us define the function  $G : [0, 1] \rightarrow \mathbb{R}_+$  as follows

$$G(u) = \mathbb{E} \left[ u^{I_T} 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} \mid I_T \neq 0 \right], \quad (\text{I.5})$$

where, consistently with the notation used in previous sections, we let

- $\tilde{H}_i = \max \{H_1^{(i)}, \dots, H_{Z_i}^{(i)}\}$  for  $i = 1, \dots, \vartheta - 1$ , i.e. the TMRCA between individuals  $\tilde{x}_i$  and  $\tilde{x}_{i+1}$ ,
- $Z_0 = \tilde{x}_1$ , the number of unsampled individuals until the first sampled one,
- $Z_i = \tilde{x}_{i+1} - \tilde{x}_i$ , the number of unsampled individuals between the  $i$ -th and the  $(i+1)$ -th sampled individuals for  $i = 1, \dots, \vartheta - 1$ ,
- $\bar{Z}_\vartheta = I_T - \tilde{x}_\vartheta$ , the number of unsampled individuals with label greater than the last sampled one,
- $H_j^{(i)}$ ,  $i, j \geq 1$  are i.i.d. random variables with same law as  $H$ .

According to this notation, the total population extant at  $T$  can be expressed as (see Fig. I.5)

$$I_T = Z_0 + \dots + Z_{\vartheta-1} + \bar{Z}_\vartheta.$$

Additionally, it is not hard to see that individuals with a label smaller than  $\tilde{x}_1$  or greater than  $\tilde{x}_\vartheta$ , which are counted in variables  $Z_0$  and  $\bar{Z}_\vartheta$ , do not take part in any coalescent event of interest. Notice also that  $G$  depends also on  $T_0, K$  the vector  $\mathbb{T}_{K-1}$  and of course on the parameters of the model, i.e.  $p$  and  $\lambda, \mu$  in the Markovian case. These dependencies are not made explicit here in order to avoid heavy notation.

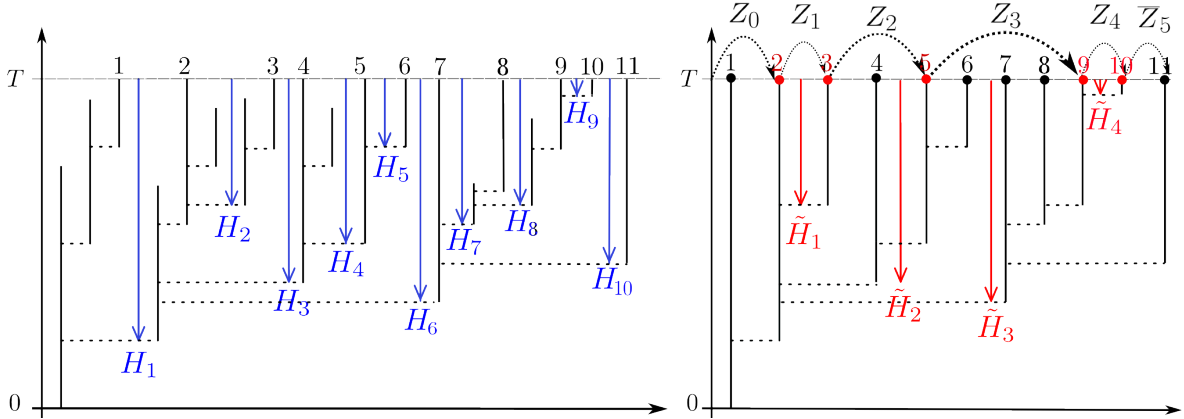


Figure I.5 – An example of splitting tree (bold lines, left), of CPP (blue) and of sampled coalescent (red). The dashed curved arrows represent the sequence of variables  $(Z_i, 0 \leq i \leq \vartheta - 1, \bar{Z}_\vartheta)$ , counting the number of individuals between two sampled ones.

**Theorem 5.** Let  $G$  denote the probability generating function of  $I_T$  as defined by (I.5). We have the following identity for all  $u \in [0, 1]$  and  $K \geq 2$ ,

$$G(u) = \left( \prod_{i=1}^{K-1} \frac{p\mathbb{P}(H < t_i)u}{1 - (1-p)\mathbb{P}(H < t_i)u} \right) \times \frac{p\mathbb{P}(H > T)u}{(1 - \mathbb{P}(H < T)(1-p)u)^2}. \quad (\text{I.6})$$

**Proof.** Following the above description of variables  $Z_i$  and Proposition 3, and thanks to the lack of memory of the geometric distribution, we can establish that, conditionally on  $\vartheta$

$$\begin{aligned} Z_0 &= G'_0 \mathbb{1}_{\{G'_0 \leq G''_0\}}, \\ Z_i &= G'_i \mathbb{1}_{\{G'_i < G''_i\}}, \quad \text{for } i = 1, \dots, \vartheta - 1, \\ \bar{Z}_\vartheta &= (G''_\vartheta - 1) \mathbb{1}_{\{G''_\vartheta \leq G'_\vartheta\}}, \end{aligned}$$

where  $G'_i \sim \mathcal{G}(p)$  and  $G''_i := \inf\{j \geq 1 : H_j^{(i)} > T\} \sim \mathcal{G}(\mathbb{P}(H > T))$  are two sequences of i.i.d. random variables independent of each other. We should note that conditional on the number of sampled individuals  $\vartheta$ , the pairs of random variables  $(\tilde{H}_i, Z_i)$  are independent of each other and their distributions are totally characterized by  $G'_i, G''_i$ .

Then the generating function  $G$  satisfies,

$$\begin{aligned} G(u) &= \mathbb{E} \left[ \left( \prod_{i=1}^{\vartheta-1} \mathbb{1}_{\{\tilde{H}_i < t_i\}} \right) \mathbb{1}_{\{\vartheta=K\}} u^{Z_0 + \dots + Z_{\vartheta-1} + \bar{Z}_\vartheta} \mid I_T \neq 0 \right] \\ &= \mathbb{E} \left[ \left( \prod_{i=1}^{K-1} \mathbb{1}_{\{\tilde{H}_i < t_i\}} u^{G'_i \mathbb{1}_{\{G'_i < G''_i\}}} \right) u^{G'_0 \mathbb{1}_{\{G'_0 \leq G''_0\}}} u^{G''_K - 1 \mathbb{1}_{\{G''_K \leq G'_K\}}} \right] \end{aligned}$$

then thanks to the independence and the i.i.d. condition of  $H_j^{(i)}$ ,

$$\begin{aligned} &= \prod_{i=1}^{K-1} \mathbb{E} \left[ \mathbb{1}_{\{\tilde{H}_i < t_i\}} u^{G'_i \mathbb{1}_{\{G'_i < G''_i\}}} \right] \mathbb{E} \left[ u^{G'_0 \mathbb{1}_{\{G'_0 \leq G''_0\}}} \right] \mathbb{E} \left[ u^{G''_K - 1 \mathbb{1}_{\{G''_K \leq G'_K\}}} \right] \\ &= \prod_{i=1}^{K-1} \mathbb{E} \left[ \mathbb{1}_{\{H_1^{(i)} < t_i, \dots, H_{G'_i}^{(i)} < t_i\}} u^{G'_i \mathbb{1}_{\{G'_i < G''_i\}}} \right] \mathbb{E} \left[ u^{G'_0 \mathbb{1}_{\{G'_0 \leq G''_0\}}} \right] \mathbb{E} \left[ u^{G''_K - 1 \mathbb{1}_{\{G''_K \leq G'_K\}}} \right]. \end{aligned}$$

It is clear from the definition of  $G''_i$  that  $\{H_1^{(i)} < t_i, \dots, H_{G'_i}^{(i)} < t_i\} \subset \{G'_i < G''_i\}$ . Then, applying Lemma 4 and the function  $g$  defined as in Equation (I.3), we get,

$$\begin{aligned} &= \prod_{i=1}^{K-1} \mathbb{E} \left[ (\mathbb{P}(H < t_i) u)^{G'_i} \right] \mathbb{E} \left[ u^{G'_0 \mathbb{1}_{\{G'_0 \leq G''_0\}}} \right] \mathbb{E} \left[ u^{G''_K - 1 \mathbb{1}_{\{G''_K \leq G'_K\}}} \right] \\ &= \left( \prod_{i=1}^{K-1} \frac{p \mathbb{P}(H < t_i) u}{1 - (1-p) \mathbb{P}(H < t_i) u} \right) g(u, p, \mathbb{P}(H > T)) \frac{g(u, \mathbb{P}(H > T), p)}{u} \\ &= \left( \prod_{i=1}^{K-1} \frac{p \mathbb{P}(H < t_i) u}{1 - (1-p) \mathbb{P}(H < t_i) u} \right) \times \frac{p \mathbb{P}(H > T) u}{(1 - \mathbb{P}(H < T)(1-p)u)^2}, \end{aligned}$$

which completes the proof.  $\square$

### MLE algorithm

It is possible to implement a maximum likelihood estimation (MLE) algorithm starting from a generating function such as the one presented in the previous section. In fact it is not hard to see that,

$$\frac{1}{n!} \frac{\partial^n G}{\partial u^n}(0) = \mathbb{P}(I_T = n, \mathbb{T}_{K-1} \mid I_T \neq 0)$$

then to find the MLE within a set of parameters  $\Theta \subset \mathbb{R}^d$  we just need to maximize the right-hand side on the above equation with respect to the parameter  $\theta \in \Theta$ . For instance, in the Markovian case  $\theta$  might be  $(p, \lambda, \mu)$  or just  $(\lambda, \mu)$  if we considered known the sampling rate.

### A special case

Let us review a simple example which is not covered by Theorem 5:  $\vartheta = 0$ .

In the case that no individual is sampled at present time  $T$  we have,

$$G(u) = \mathbb{E} \left[ u^{I_T} 1_{\{\vartheta=0\}} \mid I_T \neq 0 \right] = \mathbb{E} \left[ u^{G''} 1_{\{G'' < G'\}} \right] = \frac{p\mathbb{P}(H < T)u}{1 - (1-p)\mathbb{P}(H < T)u}$$

and

$$\frac{\partial G}{\partial u} = \frac{(1-p)\mathbb{P}(H > T)}{(1 - (1-p)\mathbb{P}(H < T)u)^2}$$

then

$$\frac{\partial G}{\partial u}(0) = (1-p)\mathbb{P}(H > T)$$

which corresponds with the model, since it means the first depth  $H$  was larger than  $T$  so there is only one individual alive at  $T$  and it is not sampled.

If we compute the second derivative we have

$$\frac{1}{2} \frac{\partial^2 G}{\partial u^2} = \frac{(1-p)^2 \mathbb{P}(H > T) \mathbb{P}(H < T)}{(1 - (1-p)\mathbb{P}(H < T)u)^3}$$

then

$$\frac{\partial^2 G}{\partial u^2}(0) = (1-p)^2 \mathbb{P}(H > T) \mathbb{P}(H < T)$$

which means there are two individuals alive at  $T$  and none of them is being sampled.

And in general if there are  $n$  extant individuals at present, all of them unsampled the probability function is,

$$\mathbb{P}(I_T = n, \vartheta = 0 \mid I_T \neq 0) = (1-p)^n (\mathbb{P}(H < T))^{n-1} \mathbb{P}(H > T)$$

## 2.4 Likelihood based on incidence data at two times

In long time studies of populations and their phylogenies, we might have not only the observed reconstructed tree from extant and sampled individuals at present time, but also the total population at different time points in the past. We are interested to incorporate this information to the model, which shall improve the estimations. To do this we need to compute a new likelihood function, resulting from adding the information of the population size at a past time, say  $t$ . This section is devoted to give a characterization of the law of  $(\tilde{H}_i, 1 \leq i \leq \vartheta - 1)$ , jointly with  $(I_t, I_T)$  for  $t < T$ , analogous to that of Theorem 5. To this end, we first need to establish some results concerning the theory of Lévy processes that will be used later.

### Coding splitting trees through their contour

We refer to Subsection 1.4 for the description of the Jumping Chronological Contour Process (JCCP) of a splitting tree, which is one of the one-to-one mappings existing that link branching processes to SPLP. We recall that, given a tree embedded in the plane, its contour or JCCP is a continuous time, real-valued process, that starts at the lifespan of the ancestor and then rolls backward along the right-hand side of this first branch at speed  $-1$  until it encounters a birth event, when it jumps up of a height of the lifespan of this new individual, getting to the next tip, and then repeating this algorithm until it eventually hits 0, as in Fig. 2. The JCCP visits all the existence times of each individual exactly once and the number of times it hits a time level, say  $s \geq 0$ , is equal to the number of individuals in the population at time  $s$ .

One of the main results in [Lam10], which is key to our study, is that the JCCP of a splitting tree is a SPLP, with Laplace exponent given by (2) with  $\mathbf{d} = -1$  and then the binary homogeneous CMJ is its local time. We let  $X = \{X_t, t \geq 0\}$  denote such a SPLP in the sequel and we will calculate some expressions related to it hereafter. Define also the running infimum and the running supremum of the process  $X$  as follows for any  $t \geq 0$ ,

$$\underline{X}_t := \inf_{0 \leq s \leq t} X_s, \quad \text{and} \quad \overline{X}_t := \sup_{0 \leq s \leq t} X_s.$$

### Some excursion calculations for spectrally positive Lévy processes

We are interested in calculating the probability that a single excursion of  $X$  remains within a certain interval. As introduced before, we denote by  $\mathbb{P}_x$  its distribution conditional on  $X_0 = x$ . We refer to [Ber96, Kyp06] and the introductory chapter for the basic definitions concerning Lévy processes.

Let us first recall [Ber96, Theorem VII.2.8], concerning the two-sided-exit problem (exit of an interval from the bottom or from the top by  $X$ ), which adapted to the case of SPLP, states that for  $0 \leq x \leq a$ :

$$\mathbb{P}_x(\tau_0 > \tau_a^+) = 1 - \frac{W(a-x)}{W(a)}, \quad (\text{I.7})$$

where  $\tau_0 = \inf\{t > 0 : X_t = 0\}$ ,  $\tau_a^+ = \inf\{t > 0 : X_t > a\}$  are stopping times with respect to  $X$  and  $W$  is the scale function introduced before, a non-negative, nondecreasing, differentiable function, such that  $W(0) = 1$  (finite variation), which is characterized by its Laplace transform. More generally, in the context of finite variation processes, we can define the following stopping times

$$\begin{aligned} \tau_x^1 &= \tau_x = \inf\{t > 0 : X_t = x\}, \\ \tau_x^n &= \inf\{t > \tau_x^{n-1} : X_t = x\}, \quad n \geq 2. \end{aligned}$$

We know from the general theory for Markov processes, that the entire path of a SPLP can be decomposed into subpaths called excursions, with the help of these stopping times [Kyp06], as we will now describe. Notice in particular that for each  $n \geq 1$ ,  $\tau_x^{n+1} - \tau_x^n$  is equal in distribution to  $\tau_x$ . Thanks to the strong Markov property and the fact that the process has stationary independent increments, it holds that for every  $n \geq 2$ , in the event  $\{\tau_x^{n-1} < +\infty\}$  the path

$$\epsilon_n = \{X_t : \tau_x^{n-1} < t \leq \tau_x^n\}$$

is independent of  $\mathcal{F}_{\tau_x^{n-1}}$  and has the same law as  $\epsilon_1$  (with  $\tau_x^0 := 0$ ). The sequence of independent and identically distributed sections of path  $(\epsilon_n)_{n \geq 1}$  are called excursions of  $X$  from  $x$ .

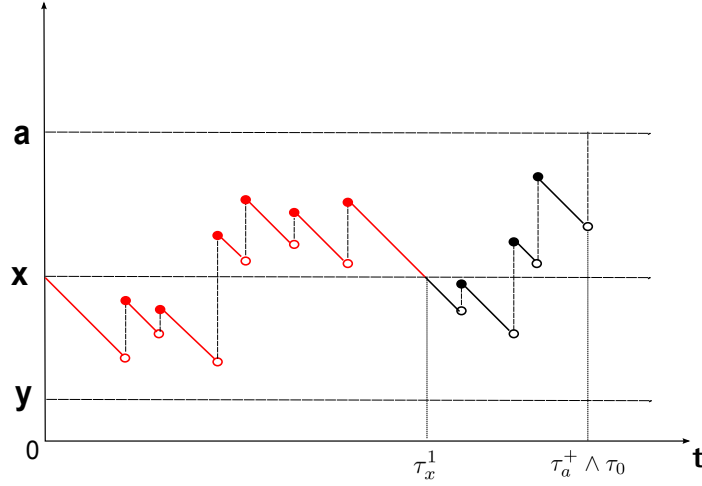


Figure I.6 – A trajectory of the process  $X$  until the exit time of the interval  $(0, a]$ , that is  $\tau_a^+ \wedge \tau_0$ , in this case from the top. In red: the first excursion of  $X$  from  $x$ , i.e.  $\epsilon_1$ .

We focus now our attention on the behaviour of  $X$  before leaving the interval  $[0, a]$  at a random time  $\tau_0 \wedge \tau_a^+$  which is a.s. finite under  $\mathbb{P}_x$  for any  $x \in [0, a]$ . We can define a random variable with values in  $\mathbb{N}$ , counting the number of excursions away from  $x$  without leaving an interval  $A$  such that  $x \in A$  as follows

$$N_x(A) := \sum_{n \geq 1} 1_{\{\tau_x^n < \tau_{A^c}\}}. \quad (\text{I.8})$$

where  $\tau_A$  is the hitting time of interval  $A$ . We set  $N_x(y, a) = N_x((y, a))$ , and in general we will omit the dependence on the set  $A$  when there is no risk of confusion. Then, we are looking for the probability that  $X$  makes  $n$  excursions from  $x$ , before it makes its first exit from the interval  $(y, a)$  at  $a$ , which is

$$\mathbb{P}_x \left( \tau_a^+ < \tau_y, N_x(y, a) = n \right), \quad \text{for } 0 \leq y \leq x, n \in \mathbb{N}. \quad (\text{I.9})$$

By applying recursively the strong Markov property, it is not hard to see that for  $n \geq 0$  we have the following relationship,

$$\mathbb{P}_x \left( \tau_a^+ < \tau_y, N_x(y, a) = n \right) = \left( \mathbb{P}_x \left( \tau_x < \tau_y \wedge \tau_a^+ \right) \right)^n \mathbb{P}_x \left( \tau_a^+ < \tau_y \wedge \tau_x \right).$$

Finally, the next lemma allows us to calculate these probabilities.

**Lemma 6.** *Let  $X$  be a SPLP with the previous assumptions and notations. Then it holds that:*

$$\mathbb{P}_x \left( \tau_x < \tau_0 \wedge \tau_a^+ \right) = 1 - \frac{W(a)}{W(x)W(a-x)}$$

**Proof of Lemma 6.** First, we denote by  $S$  the desired quantity, that is  $S := \mathbb{P}_x(\tau_x < \tau)$ , where  $\tau = \tau_0 \wedge \tau_a^+$ . Using the properties of independence and equal distribution of the excursions, the probability that  $X$  exits the interval  $(0, a]$  from the top can be expressed

as follows

$$\begin{aligned}
\mathbb{P}_x(\tau_a^+ < \tau_0) &= \mathbb{E} \left[ \mathbb{E}_x[1_{\{\tau_a^+ < \tau_0\}} \mid N_x(0, a)] \right] = \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0 \mid N_x(0, a) = n) \mathbb{P}_x(N_x(0, a) = n) \\
&= \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0 \mid \tau_x^n < \tau < \tau_x^{n+1}) \mathbb{P}_x \left( \sum_{k \geq 1} 1_{\{\tau_x^k < \tau\}} = n \right) \\
&= \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0 \mid \tau_x > \tau) \mathbb{P}_x(\tau_x^1 < \tau, \dots, \tau_x^n < \tau, \tau_x^{n+1} > \tau) \\
&= \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0 \mid \tau_x > \tau) \mathbb{P}_x \left( \sum_{i=0}^n \tilde{\tau}_x^i < \tau < \sum_{i=0}^{n+1} \tilde{\tau}_x^i \right),
\end{aligned}$$

where  $\tilde{\tau}_x^i = \tau_x^i - \tau_x^{i-1}$ ,  $i \geq 1$ . We can expand the second factor using strong Markov property and the i.i.d. condition of these excursions,

$$\begin{aligned}
\mathbb{P}_x \left( \sum_{i=0}^n \tilde{\tau}_x^i < \tau < \sum_{i=0}^{n+1} \tilde{\tau}_x^i \right) &= \mathbb{P}_x \left( \tau < \sum_{i=0}^{n+1} \tilde{\tau}_x^i \mid \sum_{i=0}^n \tilde{\tau}_x^i < \tau \right) \mathbb{P}_x \left( \sum_{i=0}^n \tilde{\tau}_x^i < \tau \right) \\
&= \mathbb{P}_x(\tau < \tau_x) \mathbb{P}_x \left( \bigcap_{i=0}^{n-1} \left\{ \sup_{\tau_x^i < t \leq \tau_x^{i+1}} X_t < a, \inf_{\tau_x^i < t \leq \tau_x^{i+1}} X_t > 0 \right\} \right) \\
&= \mathbb{P}_x(\tau < \tau_x) (\mathbb{P}_x(\tau_x < \tau))^n
\end{aligned}$$

then going back to the sum we get a geometric series and since  $\mathbb{P}_x(\tau_x < \tau) \neq 1$

$$\begin{aligned}
&= \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0 \mid \tau_x > \tau) \mathbb{P}_x(\tau_x > \tau) (\mathbb{P}_x(\tau_x < \tau))^n \\
&= \sum_{n \geq 0} \mathbb{P}_x(\tau_a^+ < \tau_0, \tau_x > \tau) (\mathbb{P}_x(\tau_x < \tau))^n \\
&= \frac{1}{1 - \mathbb{P}_x(\tau_x < \tau)} \mathbb{P}_x(\tau_a^+ < \tau_0, \tau_x > \tau),
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{P}_x(\tau_a^+ < \tau_0, \tau_x > \tau) &= \mathbb{P}_x(\tau_a^+ < \tau_0, \tau_a^+ < \tau_x) = \mathbb{P}_x(\tau_x < \tau_0, \tau_a^+ < \tau_x) \\
&= \mathbb{P}_x(\tau_x < \tau_0) - \mathbb{P}_x(\tau_x < \tau_0, \tau_a^+ > \tau_x) = \mathbb{P}_x(\tau_x < \tau_0) - \mathbb{P}_x(\tau_x < \tau).
\end{aligned}$$

The first term in the above difference can be seen as the union of increasing events,  $\{\tau_x < \tau_0\} = \bigcup_{a > x, a \in \mathbb{Q}} \{\tau_a^+ < \tau_0\}$ . Then, (I.7) and the continuity of the function  $W$  we get that

$$\mathbb{P}_x(\tau_x < \tau_0) = \lim_{a \downarrow x} 1 - \frac{W(a - x)}{W(a)} = 1 - \frac{1}{W(x)}$$

and hence the following equation on  $S$  is satisfied,

$$1 - \frac{W(a - x)}{W(a)} = \frac{1 - \frac{1}{W(x)} - S}{1 - S}, \quad (\text{I.10})$$

which proves the lemma.  $\square$

We can apply the lemma to calculate the probabilities in (I.9). By a simple translation of the origin to  $y$  we obtain that

$$\begin{aligned}\mathbb{P}_x(\tau_x < \tau_y \wedge \tau_a^+) &= 1 - \frac{W(a-y)}{W(x-y)W(a-x)} \\ \mathbb{P}_x(\tau_a^+ < \tau_y \wedge \tau_x) &= \frac{W(a-y) - W(a-x)}{W(x-y)W(a-x)}\end{aligned}$$

Then for  $n \geq 0$  and  $0 \leq y \leq x \leq a$ ,

$$\mathbb{P}_x(\tau_a^+ < \tau_y, N_x(y, a) = n) = \left(1 - \frac{W(a-y)}{W(x-y)W(a-x)}\right)^n \frac{W(a-y) - W(a-x)}{W(x-y)W(a-x)} \quad (\text{I.11})$$

### Likelihood computation

We have gathered the elements needed to calculate a generating-type function as in the previous section, now in two variables  $u, v \leq 1$ , defined as follows,

$$G(u, v) = \mathbb{E}_x \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} u^{I_t} v^{I_T} \mid I_T \neq 0 \right], \quad (\text{I.12})$$

where we condition on the life duration of the ancestor (individual originating the epidemic) to be equal to  $x \geq t$ . In order to simplify the notation and without loss of generality, the vector  $\mathbb{T}_{K-1} = (t_1, \dots, t_{K-1})$  is taken ordered, that is

$$0 < t_1 < \dots < t_{k-1} < T - t < \dots < t_{K-1}$$

so that the first  $k$  sampled individuals coalesce between present time  $T$  and  $t$ , and the rest of them might find their TMRCA between  $t$  and the more distant past time 0. Therefore, the first  $k-1$  coalescence events can be modeled as in Theorem 5. Those that might be greater than  $T-t$  need a more delicate study, that will be done with the help of the results on Lévy processes we have just seen. What we should notice is that, when we include coalescence events that take place between 0 and  $t$ , we also add information about extant individuals at  $t$ , which do not survive up to time  $T$ . Hence, for each of these coalescence events, there is an input in the total population at  $t$  that needs to be quantified in (I.12). In that sense, for these coalescence times that might be larger than  $T-t$  we change the notation to  $\hat{H}_i$ ,  $i \geq k$ . Consequently, the function  $G$  takes the form

$$G(u, v) = \mathbb{E}_t \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{k-1} < t_{k-1}, \hat{H}_k < t_k, \dots, \hat{H}_{\vartheta-1} < t_{\vartheta-1}\}} 1_{\{\vartheta=K\}} u^{I_t} v^{I_T} \mid I_T \neq 0 \right].$$

This notation is set just to highlight the fact that the events of the type  $\{\tilde{H}_i < t_i\}$  differ from  $\{\hat{H}_i < t_i\}$  when considered within the above expectation. Then, to describe  $I_t$  we need to know which of these  $\hat{H}_i$  node depths is larger than  $T-t$ . Accordingly, we associate to each  $\hat{H}_i$  an independent Bernoulli r.v.  $B_i$ , with common distribution defined as follows,

$$\mathbb{P}(B = 0) = \frac{\mathbb{P}(H < T - t)}{\mathbb{P}(H < T)} = \mathbb{P}_T(\tau_T^+ < \tau_t \mid \tau_T^+ < \tau_0), \quad (\text{I.13a})$$

$$\mathbb{P}(B = 1) = \frac{\mathbb{P}(T - t \leq H < T)}{\mathbb{P}(H < T)} = \mathbb{P}_T(\tau_t \leq \tau_T^+ \mid \tau_T^+ < \tau_0). \quad (\text{I.13b})$$

Suppose that for the individual labeled  $i$  at  $T$ , we have the situation of  $\{\hat{H}_i > T - t\}$ , or equivalently  $B_i = 1$ . This coalescence time between individuals  $i$  and  $i+1$  corresponds to



the minimum of an excursion from  $T$  of the JCCP reflected at  $T$ . Moreover, the number of hitting times of the level  $s$  by this excursion, for any  $0 \leq s \leq T$  is equal to the number of individuals that are placed to the right of  $i$  and to the left of  $i - 1$  (following the order we specified in the introduction). Then, conditional on  $\hat{H}_i > T - t$ , the probability of  $\hat{H}_i > T - t_i$ , jointly with having  $n$  individuals that were extant at  $t$ , that are to the right of  $i$  and the left of  $i + 1$  in the tree, and from which only one has alive descendants at time  $T$ , can be expressed as follows in terms of the JCCP, thanks to Lemma 6 and the formulae before it, for any  $0 \leq s < t$ ,

$$\mathbb{P}_t \left( \tau_T^+ < \tau_s, N_t(t, T) = n \right) = (\rho_1(s))^n \rho_2(s), \quad (\text{I.14})$$

where

$$\begin{aligned} \rho_1(s) &= 1 - \frac{W(T-s)}{W(t-s)W(T-t)}, \\ \rho_2(s) &= \frac{W(T-s) - W(T-t)}{W(t-s)W(T-t)}. \end{aligned}$$

To compute the likelihood we will also need to use the functions  $h$  and  $r$  defined as follows for a random variable  $B$  with distribution (I.13)

$$\begin{aligned} h(v, s) &:= \mathbb{E}_t \left[ \left( 1_{\{\tau_T^+ < \tau_s\}} v^{N_t(t, T)+1} \right)^B \right] \\ &= \frac{\mathbb{P}(H < T-t)}{\mathbb{P}(H < T)} + \mathbb{E}_t \left[ 1_{\{\tau_T^+ < \tau_s\}} v^{N_t(t, T)+1} \right] \frac{\mathbb{P}(T-t \leq H < T)}{\mathbb{P}(H < T)} \\ &= \frac{\mathbb{P}(H < T-t)}{\mathbb{P}(H < T)} + \frac{\rho_2(s)v}{1-v\rho_1(s)} \frac{\mathbb{P}(T-t \leq H < T)}{\mathbb{P}(H < T)}. \end{aligned} \quad (\text{I.15})$$

Using (I.14) and the strong Markov property we also have that

$$\begin{aligned} \tilde{h}(v, x) &:= \mathbb{E}_x \left[ 1_{\{\tau_T^+ < \tau_0\}} v^{N(t, T)} \right] = \mathbb{P}_x \left( \tau_t \geq \tau_T^+ \right) + \mathbb{E}_t \left[ 1_{\{\tau_T^+ < \tau_0\}} v^{N(t, T)} \right] \mathbb{P}_x \left( \tau_t < \tau_T^+ \right) \\ &= 1 - \frac{W(T-x)}{W(T-t)} + \frac{\rho_2(0)v}{1-v\rho_1(0)} \frac{W(T-x)}{W(T-t)} \\ &= \frac{\mathbb{P}(T-x < H \leq T-t)}{\mathbb{P}(H > T-x)} + \frac{\rho_2(0)v}{1-v\rho_1(0)} \frac{\mathbb{P}(H > T-t)}{\mathbb{P}(H > T-x)}. \end{aligned} \quad (\text{I.16})$$

Then we are ready to state the next result,

**Theorem 7.** *Let  $G(u, v)$  denote the probability generating function of  $(I_t, I_T)$  as defined by (I.12). We have the following identity for all  $u \in [0, 1]$  and  $K \geq 2$ ,*

$$\begin{aligned} G(u, v) &= \left( \prod_{i=1}^{k-1} \frac{p\mathbb{P}(H < t_i)u}{1 - (1-p)\mathbb{P}(H < t_i)u} \right) \times \frac{pu}{(1 - \mathbb{P}(H < T)(1-p)u)} \times \frac{\rho_2(0)}{1 - v\rho_1(0)} \\ &\quad \prod_{i=k}^{K-1} \frac{p\mathbb{P}(H \leq T)h(v, T-t_i)u}{1 - (1-p)\mathbb{P}(H \leq T)h(v, T-t_i)u} \times \frac{h(v, 0)(1 - (1-p)\mathbb{P}(H < T))}{1 - (1-p)\mathbb{P}(H < T)uh(v, 0)} \times \tilde{h}(v, x), \end{aligned}$$

for a function  $h$  defined by (I.15) and  $\tilde{h}$  by (I.16).

*Proof.* As in Theorem 5, consider two independent sequences of random variables  $G'_l \sim \mathcal{G}(p)$  and  $G''_l \sim \mathcal{G}(\mathbb{P}(H > T))$ ,  $1 \leq l \leq K$ . For every  $l \geq k$  we need to compute the law of the coalescence time being smaller than the corresponding  $t_l$ , jointly with the generating

function of the number of individuals involved in this coalescence event. Set  $s_l = T - t_l$  for every  $k \leq l \leq K$  and define

$$\vartheta_l := \vartheta(G'_l) = \sum_{i=1}^{G'_l} B_i,$$

where we take a sequence of i.i.d. random variables  $(B_i)_{i \geq 1}$  distributed as  $B$  (I.13) and independent from  $G'_l, G''_l$ . Each term  $B_i$  in this sequence, for every  $1 \leq i \leq G'_l$ , stands for the event that the node depth of the individual  $\tilde{x}_l + i$  (label at  $T$ ), reaches or not level  $t$ . We let the pairs  $(D_i^l, N_i^l)_{i \geq 1}$  denote i.i.d. random variables distributed as  $(\underline{X}_{\tau_0 \wedge \tau_T^+}, N(t, T))$  under  $\mathbb{P}_t$ . Then we can compute for each  $s_l, G', G''$  the following functions,

$$\begin{aligned} g(u, v, s_l) &= \mathbb{E} \left[ 1_{\{\widehat{H}_l < t_l\}} v^{\sum_{i=1}^{\vartheta_l} (N_i^l + 1)} u^{G'_l} 1_{G'_l < G''_l} \right] \\ &= \mathbb{E} \left[ 1_{\{\min_{1 \leq i \leq \vartheta_l} D_i^l > s_l\}} v^{\sum_{i=1}^{\vartheta_l} (N_i^l + 1)} u^{G'_l} 1_{G'_l < G''_l} \right] \\ &= \mathbb{E} \left[ 1_{\{D_i^l > s_l, 1 \leq i \leq \vartheta_l\}} v^{\sum_{i=1}^{\vartheta_l} (N_i^l + 1)} u^{G'_l} 1_{G'_l < G''_l} \right] \\ &= \mathbb{E} \left[ \prod_{i=1}^{\vartheta_l} \left( 1_{\{D_i^l > s_l\}} v^{(N_i^l + 1)} \right) u^{G'_l} 1_{G'_l < G''_l} \right] \\ &= \mathbb{E} \left[ \prod_{i=1}^{G'_l} \left( \left( 1_{\{D_i^l > s_l\}} v^{(N_i^l + 1)} \right)^{B_i} u \right) 1_{G'_l < G''_l} \right] \\ &= \sum_{j \geq 1} \mathbb{E} \left[ \prod_{i=1}^j \left( \left( 1_{\{D_i^l > s_l\}} v^{(N_i^l + 1)} \right)^{B_i} u \right) \right] \mathbb{P}(G'_l = j, G'_l < G''_l) \\ &= \sum_{j \geq 1} \left( \mathbb{E}_t \left[ \left( 1_{\{\underline{X}_{\tau_0 \wedge \tau_T^+} > s_l\}} v^{N(t, T) + 1} \right)^B \right] \right)^j u^j \mathbb{P}(G'_l = j, G'_l < G''_l) \end{aligned}$$

which is the generating function of  $G'_l \mathbb{1}_{\{G'_l < G''_l\}}$ , (I.4), evaluated in  $\mathbb{E}_t \left[ \left( 1_{\{\underline{X}_{\tau_0 \wedge \tau_T^+} > s_l\}} v^{N(t, T) + 1} \right)^B \right] u$ , which is the same as  $h(v, s_l)u$ . Then

$$g(u, v, s_l) = \frac{p \mathbb{P}(H \leq T) h(v, s_l) u}{1 - (1 - p) \mathbb{P}(H \leq T) h(v, s_l) u}.$$

Hence, as in Theorem 5, the function  $G$  can be expressed as a product

$$\begin{aligned} G(u, v) &= \mathbb{E} \left[ u^{G'_0} 1_{\{G'_0 \leq G''_0\}} \right] \prod_{i=1}^{k-1} \mathbb{E} \left[ (\mathbb{P}(H < t_i) u)^{G'_i} 1_{G'_i < G''_i} \right] \prod_{i=k}^{K-1} g(u, v, T - t_i) \times \\ &\quad \mathbb{E} \left[ 1_{\{D_i > 0, i=1, \dots, \vartheta(G''_K)\}} u^{G''_K - 1} 1_{\{G''_K \leq G'_K\}} v^{\sum_{i=1}^{\vartheta(G''_K)} N_i^K} v^{\overline{N}_K} \right] \mathbb{E}_x \left[ v^{N_0} \mathbb{1}_{\{D_0 > 0\}} \right] \end{aligned}$$

where  $\overline{N}_K$  represents the *last* individuals alive at  $t$  which die before  $T$ , that is those placed to the right of the last sampled individual at  $T$ , in the order defined by the JCCP. This variable  $\overline{N}_K$  has the same law as  $N(t, T) \mathbb{1}_{\{\underline{X}_{\tau_0 \wedge \tau_T^+} > 0\}}$  under  $P_T$ , in the same way as  $N_0 \mathbb{1}_{D_0 > 0}$ , which stand for the individuals alive at  $t$  before there is one that survives up to time  $T$ .

Then, in order to complete the proof we only need to compute the last two expected values. The last one can be evaluated using (I.16) since

$$\mathbb{E}_x \left[ v^{N_0} \mathbb{1}_{\{D_0 > 0\}} \right] = \mathbb{E}_x \left[ v^{N(t,T)} \mathbb{1}_{\{\tau_T^+ < \tau_0\}} \right] = \tilde{h}(v, x).$$

The first one is

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{1}_{\{D_i > 0, i=1, \dots, \vartheta(G''_K)\}} u^{G''_K-1} \mathbb{1}_{\{G''_K \leq G'_K\}} v^{\sum_{i=1}^{\vartheta(G''_K)} N_i^K + 1} \right] \mathbb{E} \left[ v^{\bar{N}_K} \right] \\ &= \frac{1}{u} \mathbb{E} \left[ \prod_{i=1}^{G''_K} \left( v^{N_i^K + 1} \right)^{B_i} u \mathbb{1}_{\{G''_K \leq G'_K\}} \right] \mathbb{E}_t \left[ v^{N(t,T)} \mathbb{1}_{\{\underline{X}_{\tau_0 \wedge \tau_T^+} > 0\}} \right] \end{aligned}$$

remember that  $G''$ , conditional on  $G'' \leq G'$  is geometric with parameter  $\gamma = 1 - (1 - p)\mathbb{P}(H < T)$ , and that the second expectation can be evaluated again thanks to (I.14), then

$$\begin{aligned} &= \frac{1}{u} \frac{u \mathbb{E} \left[ \left( \mathbb{1}_{\{D > 0\}} v^{N+1} \right)^B \right] \gamma}{1 - (1 - \gamma) u \mathbb{E} \left[ \left( \mathbb{1}_{\{D > 0\}} v^{N+1} \right)^B \right]} \mathbb{E} \left[ v^N \mathbb{1}_{\{D > 0\}} \right] \\ &= \frac{h(v, 0) \gamma}{1 - (1 - \gamma) u h(v, 0)} \frac{\rho_2(0)}{1 - v \rho_1(0)}, \end{aligned}$$

which ends the proof.  $\square$

## 2.5 Generalization: a backward inhomogeneous branching process in the exponential case

In this subsection we are interested in computing a generating-type function as in the previous section, but now in  $N$  variables  $u, u_1, \dots, u_N \in [0, 1]$ , defined as follows,

$$G(u, u_1, \dots, u_N) = \mathbb{E} \left[ \mathbb{1}_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{\vartheta-1} < t_{\vartheta-1}\}} \mathbb{1}_{\{\vartheta=K\}} u^{I_T} u_1^{I_{T_1}} \dots u_N^{I_{T_N}} \mid I_T \neq 0 \right]$$

where  $T = T_0 > T_1 > \dots > T_N > 0$  are fixed times between present time  $T$  and time 0, which is assumed to be the origin of the epidemic. Again, variables  $\tilde{H}$  represent the coalescence times between two consecutive individuals, among the  $\vartheta$  sampled individuals at  $T$ , and we are given a sequence of real variables  $\{t_i\}_{i \in \mathbb{N}} \subset [0, T]$ . We define the complementary sequence  $s$  as

$$s_i = T - t_i, \quad \forall i \in \mathbb{N}.$$

Throughout this subsection we consider only the case where  $I$  is a birth and death process, that is, the periods of infectiousness are exponentially distributed. This hypothesis comes from the fact that in the general case the CMJ process is non-Markovian and this property turns out to be essential for the techniques we use to prove our results, as we will now see. We also suppose that the life duration of the individual at the origin of the epidemic is equal to  $x$  and we work with  $\mathbb{P}_x$ .

Thanks to a path-wise decomposition of the JCCP of a birth and death tree, relying on the strong Markov property and the independence of sampled individuals at  $T$ , it is possible to describe the law of  $(I_T, I_{T_1}, \dots, I_{T_N})$ , as a sum of independent subpopulations, as it is shown in Fig. I.7. This subpopulations, when seen backward in time, are *inhomogeneous*

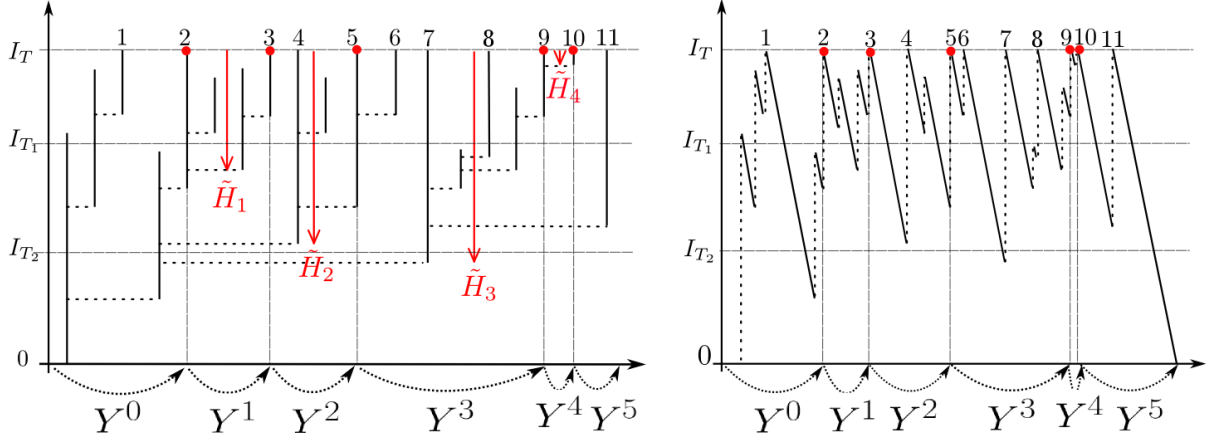


Figure I.7 – A splitting tree (left, bold lines), the CPP of sampled individuals at  $T$  (red) and the population size of the process at times  $0 < T_2 < T_1 < T$ . The corresponding JCCP is represented in the right panel. The curved arrows represent the subpopulations described by variables  $(Y^i)_{0 \leq i \leq 5}$  (left) and the corresponding excursions of the reflected JCCP (right).

*branching processes* (IBP). The reproduction laws vary from one subpopulation to another, and two of them (first and last, following the forthcoming notation) allow immigration.

The idea behind this decomposition is very simple. First, thanks to the independence of the excursions of the JCCP below level  $T$ , conditional on  $I_T$ , the law of  $I_s$  at any prior time  $s < T$ , can be split into a sum of  $I_T + 1$  random variables  $(I_s^i, 0 \leq i \leq I_T)$  (that we call subpopulations), corresponding to the excursions of the contour process between any pair of these individuals, plus the excursion before hitting  $T$  for the first time, as depicted in Fig. I.7. To clarify, conditionally on  $I_T$ , we have for any  $0 \leq s \leq T$  that

$$I_s = \sum_{i=0}^{I_T} I_s^i, \quad (\text{I.17})$$

where the law of each  $I_s^i$  can be specified with the help of excursion theory for Lévy processes. On this account, we can identify in (I.17) a branching structure, where for  $1 \leq i \leq I_T$ , the  $i$ -th individual at  $T$  has  $I_s^i$  descendants at time  $T - s$ , and  $I_s^0$  are the number of immigrants arrived to the population at that time.

Adding the sampling scheme presented previously, where individuals are sampled independently at time  $T$  with probability  $p$ , amounts to consider  $I_T$ , conditional on the number of sampled individuals  $\vartheta$ , as a sum of  $\vartheta + 1$  random variables as it was done in Section 2. More precisely, the joint law of  $(I_T, I_{T_1}, \dots, I_{T_N})$ , under  $\mathbb{P}_x$  is that of the sum of IBP's with different reproduction laws, i.e.

$$(I_T, I_{T_1}, \dots, I_{T_N}) = \sum_{l=0}^{K+1} (Y_0^l, Y_1^l, \dots, Y_N^l),$$

where for each  $0 \leq l \leq K + 1$ ,  $Y^l$  is an inhomogeneous branching processes (with immigration for  $l = 0, K + 1$ ), more precisely, for every  $1 \leq l \leq K$ ,  $1 \leq k \leq N$ , conditioned on

$Y_{k-1}^l$ ,

$$Y_k^l = \sum_{i=1}^{Y_{k-1}^l} \xi_i^{l,k}$$

and

$$Y_k^0 = \sum_{i=1}^{Y_{k-1}^0} \xi_i^{0,k} + \phi_k, \quad Y_k^{K+1} = \sum_{i=1}^{Y_{k-1}^{K+1}} \xi_i^{K+1,k} + \psi_k.$$

The index  $l$  (sub-population index) indicates the dependence on  $s_l$ , the coalescence time between the  $l$ -th and  $(l+1)$ -th sampled individuals (we set  $s_0, s_{K+1} = 0$ ); and  $k$  (generation index) denotes the dependence on levels  $T_k, T_{k-1}$ .

### Initial sub-populations:

As in the previous sections, conditional on  $\vartheta$ , the number of sampled individuals at  $T$ , we can describe the population extant at  $T$  as a sum of geometric variables in competition, that is

- $Y_0^0 = G'_0 \mathbb{1}_{G'_0 < G''_0}$ , where  $G'_0 \sim \mathcal{G}(p)$  starting from 0, and  $G''_0 \sim \mathcal{G}(\beta_T)$  starting from 1 are independent r.v, with  $\beta_T = \mathbb{P}_T(\tau_T^+ < \tau_0)$ , i.e. for  $n \geq 0$ ,

$$\mathbb{P}(Y_0^0 = n) = \mathbb{P}(G'_0 = n, G'_0 < G''_0) = (1-p)^n p(\beta_T)^n$$

Then, its probability generating function is given by,

$$g_{0,0}(u) = \mathbb{E}[u^{Y_0^0}] = \frac{p}{1 - (1-p)\beta_T u} \quad (\text{I.18})$$

- $Y_0^l = G'_l \mathbb{1}_{G'_l < G''_l}$ , for  $1 \leq l \leq K$ . Both are geometric independent variables starting from 1:  $G'_l \sim \mathcal{G}(p)$ ,  $G''_l \sim \mathcal{G}(\beta_T)$ , then for  $n \geq 1$ ,

$$\mathbb{P}(Y_0^l = n) = \mathbb{P}(G'_l = n, G'_l < G''_l) = (1-p)^{n-1} p(\beta_T)^n$$

Again, with  $pgf$ ,

$$g_{0,l}(u) = \mathbb{E}[u^{Y_0^l}] = \frac{p\beta_T u}{1 - (1-p)\beta_T u} \quad (\text{I.19})$$

- $Y_0^{K+1} = G''_{K+1} \mathbb{1}_{G''_{K+1} \leq G'_{K+1}}$ , and  $G'_{K+1} \sim \mathcal{G}(p)$ ,  $G''_{K+1} \sim \mathcal{G}(\beta_T)$ , then for  $n \geq 1$ ,

$$\mathbb{P}(Y_0^{K+1} = n) = \mathbb{P}(G''_{K+1} = n, G''_{K+1} \leq G'_{K+1}) = (1-p)^{n-1} (\beta_T)^{n-1} (1-\beta_T)$$

With,

$$g_{0,K+1}(u) = \mathbb{E}[u^{Y_0^{K+1}}] = \frac{(1-\beta_T)u}{1 - (1-p)\beta_T u} \quad (\text{I.20})$$

### Reproduction law:

For each  $l$  and  $k$  ( $0 \leq l \leq K+1$ ,  $1 \leq k \leq N$ ), such that  $s_l < T_k$ , variables  $(\xi_i^{l,k})_{i \geq 1}$  are i.i.d., taking values in  $\mathbb{N}$  and counting the number of hits of the contour process of level  $T_k$ , starting at  $T_{k-1}$  before it exits the interval  $[s_l, T_{k-1}]$  by the top. Then

$$\mathbb{P}(\xi^{l,k} = 0) = \mathbb{P}_{T_{k-1}}(\tau_{T_{k-1}}^+ < \tau_{T_k}) = 1 - \frac{1}{W(T_{k-1} - T_k)}$$

and for  $n \geq 1$

$$\begin{aligned} \mathbb{P}(\xi^{l,k} = n) &= \mathbb{P}_{T_{k-1}}(\tau_{T_{k-1}}^+ > \tau_{T_k}) \left( \mathbb{P}_{T_k}(\tau_{T_k} < \tau_{T_{k-1}}^+ \wedge \tau_{s_l}) \right)^{n-1} \mathbb{P}_{T_k}(\tau_{T_{k-1}}^+ < \tau_{T_k} \wedge \tau_{s_l}) \\ &= \left( 1 - \frac{W(T_{k-1} - s_l)}{W(T_k - s_l)W(T_{k-1} - T_k)} \right)^{n-1} \frac{W(T_{k-1} - s_l) - W(T_{k-1} - T_k)}{W(T_k - s_l)W^2(T_{k-1} - T_k)}. \end{aligned}$$

Notice that  $\xi^{l,k} = 0$ ,  $\mathbb{P}_x$  a.s.  $\forall k \in \mathbb{N}$ ,  $k \geq \inf\{n \in \mathbb{N} : T_n < s_l\}$ .

In order to simplify the computation of the *pgf* for these variables we will introduce some notation:

- $\Delta_k = T_{k-1} - T_k$ , for every  $1 \leq k \leq N$
- $\beta_s = 1 - \frac{1}{W(s)}$ , for  $s \geq 0$
- $\alpha_{k,l} = 1 - \frac{W(T_{k-1} - s_l)}{W(T_k - s_l)W(T_{k-1} - T_k)}$ , for every  $l, k$  such that  $T_k - s_l \geq 0$

Then, the *pgf* for these variables are,

$$g_{k,l}(u) = \mathbb{E}[u^{\xi^{l,k}}] = \beta_{\Delta_k} + \frac{(\beta_{T_k - s_l} - \alpha_{k,l})(1 - \beta_{\Delta_k})u}{1 - \alpha_{k,l}u} \quad (\text{I.21})$$

### Immigration in $Y^{K+1}$ :

Variables  $\psi_k$  are analogous to  $\xi^{0,k}$  but exit will be from the bottom of the interval  $[0, T_{k-1}]$ , they describe the last visits of the process before it is killed, i.e. the visits to level  $T_k$ , starting at  $T_{k-1}$ , before leaving the interval  $[0, T_{k-1}]$  by 0. So they are distributed as follows, for  $n \geq 0$

$$\begin{aligned} \mathbb{P}(\psi_k = n + 1) &= \mathbb{P}_{T_{k-1}}(\tau_{T_{k-1}}^+ > \tau_{T_k}) \left( \mathbb{P}_{T_k}(\tau_{T_k} < \tau_{T_{k-1}}^+ \wedge \tau_{s_l}) \right)^{n-1} \mathbb{P}_{T_k}(\tau_0 < \tau_{T_k}^+) \\ &= \left( 1 - \frac{W(T_{k-1})}{W(T_k)W(T_{k-1} - T_k)} \right)^n \frac{1}{W(T_k)W(T_{k-1} - T_k)} \end{aligned}$$

In this case we can easily see that,

$$h_{k,K+1}(u) = \mathbb{E}[u^{\psi_k}] = \frac{(1 - \beta_{T_k})(1 - \beta_{\Delta_k})u}{1 - \alpha_{k,l}u} \quad (\text{I.22})$$

### Immigration in $Y^0$ :

Variables  $\phi_k$ , on the other hand, will count the first visits of level  $k$ , before it visits for the first time the level above,  $k-1$ , conditional on the fact that the process leaves  $[0, T]$  by the top. This is due to the condition  $I_T \neq 0$  in function  $G$ . Besides, since these variables describe the first excursions of the contour process, which starts at  $x$  (the lifespan of the first individual), before it hits  $T$  by the first time, we need to compute their law under  $\mathbb{P}_x$ . There will be several cases to consider depending on the position of other levels with respect to  $x$ . See Fig. I.8. Notice this result is proved only in the exponential case, with birth and death parameters  $\lambda$  and  $\mu$ , and where we set  $r = \lambda - \mu$ . In this case the scale function  $W$  is known and it was specified before, however to avoid a heavier notation we keep using  $W$  in the sequel.

**Lemma 8.** *For each  $x \in [0, T]$ , there exists a unique  $k^*$  between 1 and  $N$  such that:  $T_{k^*} \leq x < T_{k^*-1}$ . Then, the laws of the immigrants in the process  $Y^0$ , under  $\tilde{\mathbb{P}} = \mathbb{P}_x(\cdot \mid X_T \neq 0)$  can be described as follows,*

$$\begin{aligned}\tilde{\mathbb{P}}(\phi^{k^*} = 0) &= \frac{W(T_{k^*-1})}{W(T_{k^*-1} - T_{k^*})} \left( \frac{W(T_{k^*-1} - T_{k^*}) - W(T_{k^*-1} - x)}{W(T) - W(T - x)} \right) e^{r(T - T_{k^*-1})}, \\ \tilde{\mathbb{P}}(\phi^{k^*} = n) &= \frac{W(T_{k^*-1})}{W(T_{k^*-1} - T_{k^*})} \left( \frac{W(T_{k^*-1} - T_{k^*}) - W(T_{k^*-1} - x)}{W(T) - W(T - x)} \right) e^{r(T - T_{k^*-1})}, \\ &\quad \times \left( 1 - \frac{W(T_{k^*-1})}{W(T_{k^*})W(T_{k^*-1} - T_{k^*})} \right)^{n-1} \frac{1}{W(T_{k^*-1})}, \quad n \geq 1\end{aligned}$$

with  $h_{k^*,0}(u) = \tilde{\mathbb{E}} \left[ u^{\phi^{k^*}} \right] =$

$$\frac{W(T_{k^*-1})}{W(T_{k^*-1} - T_{k^*})} \left( \frac{W(T_{k^*-1} - T_{k^*}) - W(T_{k^*-1} - x)}{W(T) - W(T - x)} \right) e^{r(T - T_{k^*-1})} \left( 1 + \frac{(1 - \beta_{T_{k^*-1}})u}{1 - \alpha_{k^*,0}u} \right)$$

For every  $0 \leq k < k^*$ ,

$$\begin{aligned}\tilde{\mathbb{P}}(\phi^k = 0) &= \frac{W(T_{k-1})}{W(T_{k-1} - T_k)} \left( \frac{1}{W(T) - W(T - x)} \right) e^{r(T - T_k)}, \\ \tilde{\mathbb{P}}(\phi^k = n) &= \frac{W(T_{k-1})}{W(T_{k-1} - T_k)} \left( \frac{W(T_{k-1} - x)}{W(T) - W(T - x)} \right) e^{r(T - T_{k-1})}, \\ &\quad \times \left( 1 - \frac{W(T_{k-1})}{W(T_k)W(T_{k-1} - T_k)} \right)^{n-1} \frac{W(T_{k-1}) - W(T_{k-1} - T_k)}{W(T_{k-1} - T_k)},\end{aligned}$$

with  $h_{k,0}(u) = \tilde{\mathbb{E}} \left[ u^{\phi^k} \right] =$

$$\frac{W(T_{k-1})}{W(T_{k-1} - T_k)} \left( \frac{1}{W(T) - W(T - x)} \right) e^{r(T - T_k)} \left( W(T_{k-1} - x) + \frac{\left( \frac{W(T_{k-1})}{W(T_{k-1} - T_k)} - 1 \right) u}{1 - \alpha_{k^*,0}u} \right).$$

And for every  $k > k^*$ ,  $\phi_k = 0$ , then  $h_{k,0}(u) = 1$ .

*Proof.* We will detail the proof for  $k^*$ , since we can proceed in a very similar way for the other values of  $k$ . Notice that the variable  $\phi^{k^*}$  will count the number of times the JCCP hits level  $T_{k^*}$ , starting at  $x$ , before it leaves the interval  $[0, T_{k^*-1}]$ , conditional on leaving  $[0, T]$  by the top. Then, applying the strong Markov property, Lemma 12 and

Equation (A-I.5) we obtain,

$$\begin{aligned}
\tilde{\mathbb{P}}(\phi^{k^*} = 0) &= \mathbb{P}_x \left( \tau_{T_{k^*-1}}^+ < \tau_{T_{k^*}} \mid \tau_T^+ < \tau_0 \right) = \frac{\mathbb{P}_x \left( \tau_{T_{k^*-1}}^+ < \tau_{T_{k^*}}, \tau_T^+ < \tau_0 \right)}{\mathbb{P}_x \left( \tau_T^+ < \tau_0 \right)} \\
&= \frac{\mathbb{P}_x \left( \tau_{T_{k^*-1}}^+ < \tau_{T_{k^*}} \right)}{\mathbb{P}_x \left( \tau_T^+ < \tau_0 \right)} \left( \int_0^{T-T_{k^*-1}} \mathbb{P}_{T_{k^*-1}+v} \left( \tau_T^+ < \tau_0 \right) \mathbb{P} \left( (X_{\tau_{T_{k^*-1}}^+} - T_{k^*-1}) \in dv \right) \right. \\
&\quad \left. + \mathbb{P} \left( (X_{\tau_{T_{k^*-1}}^+} - T_{k^*-1}) > T - T_{k^*-1} \right) \right) \\
&= \frac{\mathbb{P}_x \left( \tau_{T_{k^*-1}}^+ < \tau_{T_{k^*}} \right)}{\mathbb{P}_x \left( \tau_T^+ < \tau_0 \right)} \left( \int_0^{T-T_{k^*-1}} \mathbb{P}_{T_{k^*-1}+v} \left( \tau_T^+ < \tau_0 \right) \mu e^{-\mu v} dv + e^{-\mu(T-T_{k^*-1})} \right) \\
&= \frac{1 - \frac{W(T_{k^*-1}-x)}{W(T_{k^*}-T_{k^*-1})}}{1 - \frac{W(T-x)}{W(T)}} \left( \frac{W(T_{k^*}-1)}{W(T)} e^{r(T-T_{k^*-1})} - e^{-\mu(T-T_{k^*-1})} + e^{-\mu(T-T_{k^*-1})} \right) \\
&= \frac{W(T_{k^*-1})}{W(T_{k^*-1}-T_{k^*})} \left( \frac{W(T_{k^*-1}-T_{k^*}) - W(T_{k^*-1}-x)}{W(T) - W(T-x)} \right) e^{r(T-T_{k^*-1})}
\end{aligned}$$

And similarly, for  $n \geq 1$ ,

$$\begin{aligned}
\tilde{\mathbb{P}}(\phi^{k^*} = n) &= \mathbb{P}_x \left( \underbrace{\tau_{T_{k^*}} < \tau_{T_{k^*-1}}^+, \left\{ \tau_{T_{k^*}}^{(i)} < \tau_{T_{k^*-1}}^+ \wedge \tau_0 \right\}_{i=1, \dots, n-1}, \tau_{T_{k^*-1}}^+ \wedge \tau_0 < \tau_{T_{k^*}}^{(n)}}_{\text{event } \varepsilon_n} \mid \tau_T^+ < \tau_0 \right) \\
&= \frac{\mathbb{P}_x(\varepsilon_n)}{\mathbb{P}_x(\tau_T^+ < \tau_0)} \mathbb{P}_x \left( \tau_T^+ < \tau_0 \mid \varepsilon_n \right)
\end{aligned}$$

where, by strong Markov property, we have,

$$\mathbb{P}_x \left( \tau_T^+ < \tau_0 \mid \varepsilon_n \right) = \frac{W(T_{k^*-1})}{W(T)} e^{r(T-T_{k^*-1})}$$

and

$$\begin{aligned}
\mathbb{P}_x(\varepsilon_n) &= \mathbb{P}_x \left( \tau_{T_{k^*}} < \tau_{T_{k^*-1}}^+ \right) \left( \mathbb{P}_{T_{k^*}} \left( \tau_{T_{k^*}} < \tau_{T_{k^*-1}}^+ \wedge \tau_0 \right) \right)^{n-1} \mathbb{P}_{T_{k^*}} \left( \tau_{T_{k^*-1}}^+ < \tau_{T_{k^*}} \wedge \tau_0 \right) \\
&= \frac{W(T_{k^*-1}-x)}{W(T_{k^*-1}-T_{k^*})} \left( 1 - \frac{W(T_{k^*-1})}{W(T_{k^*})W(T_{k^*-1}-T_{k^*})} \right)^{n-1} \left( 1 - \frac{W(T_{k^*-1}-T_{k^*})}{W(T_{k^*-1})} \right)
\end{aligned}$$

leading to the announced expression.  $\square$

### Some results on inhomogeneous branching processes

**Lemma 9.** *Let  $(X_n)_{n \geq 0}$  be an inhomogeneous branching process such that the reproductive law on generation  $n$  has probability generating function  $g_n$ , i.e.*

$$g_n(u) := \mathbb{E} \left[ u^{X_n} \mid X_{n-1} = 1 \right], \quad n \geq 1, \quad u \in [0, 1],$$



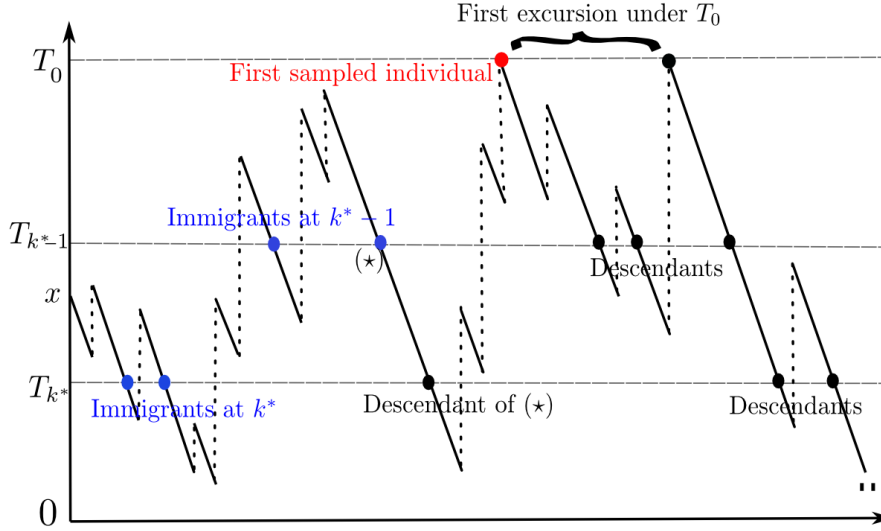


Figure I.8 – The excursions of the JCCP of a splitting tree and the corresponding subpopulations. In blue, the immigrants, i.e. the number of visits of level  $T_{k^*}$ , before the process goes above  $T_{k^*-1}$  (upper level), without hitting 0. Black dots represent the descendants, i.e. number of visits of  $T_{k^*}$ , starting from  $T_{k^*-1}$ , before it goes again above  $T_{k^*-1}$ , without hitting 0. In red, sampled individuals, i.e. the visits of level  $T_0$  which are sampled (with probability  $p$ ).

with initial population with pgf  $g_0$ . Then for every  $N \geq 0$ ,

$$\mathbb{E} \left[ u_0^{X_0} u_1^{X_1} \cdots u_N^{X_N} \right] = g_0 (g_1 (\cdots (g_N(u_N) u_{N-1})) u_0)$$

*Proof.* It follows from a direct application of branching property, that we have for any  $k \geq 0$  that  $\mathbb{E}_k \left[ u_1^{X_1} \right] = (g_1(u_1))^k$ , then

$$\mathbb{E} \left[ u_0^{X_0} u_1^{X_1} \right] = \sum_{k \geq 0} \mathbb{E}_k \left[ u_1^{X_1} \right] u_0^k \mathbb{P}(X_0 = k) = g_0(g_1(u_1)u_0)$$

and by induction on  $N$  we obtain the formula in the lemma.  $\square$

**Lemma 10.** Let  $(X_n)_{n \geq 0}$  be an inhomogeneous branching process with immigration characterized by the probability generating functions of: its initial population ( $g_0$ ), its reproductive law at generation  $n$  ( $g_n$ ) and the law of the immigrants at generation  $n$  ( $h_n$ ). More precisely for each  $n \geq 0$ ,

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,i} + \psi_n$$

where  $\xi_{n,i}$ ,  $i \geq 1$  are i.i.d., independent of  $\psi_n$ , characterized by their respective pgf  $g_n$  and  $h_n$ . Then for every  $N \geq 0$ ,

$$\mathbb{E} \left[ u_0^{X_0} u_1^{X_1} \cdots u_N^{X_N} \right] = g_0 (g_1 (\cdots g_N(u_N) \cdots) u_0) \prod_{i=1}^{N-1} h_i (g_{i+1} (\cdots g_N(u_N) \cdots u_{i+1}) u_i)$$

*Proof.* Again, an application of the branching property together with the independence of the immigrant population, and recursion on  $N$ , leads to the result. We define for each

$k \geq 0$  the function  $\tilde{g}_1^k(u) = \mathbb{E} \left[ u^{X_1} | X_0 = k \right]$  and it is not hard to see that,

$$\tilde{g}_1^k(u) = \mathbb{E} \left[ u^{\sum_{i=1}^k \xi_{1,i}} u^{\psi_1} \right] = (g_1(u))^k h_1(u)$$

The formula holds trivially for  $N = 0$ , and supposing it is also satisfied for  $N$ , by adding another generation we get,

$$\begin{aligned} \mathbb{E} \left[ u_0^{X_0} \cdots u_N^{X_N} u_{N+1}^{X_{N+1}} \right] &= \sum_{k \geq 0} \mathbb{E} \left[ u_1^{X_1} \cdots u_{N+1}^{X_{N+1}} | X_0 = k \right] u_0^k \mathbb{P}(X_0 = k) \\ &= \sum_{k \geq 0} \mathbb{E} \left[ u_1^{\sum_{i=1}^k \xi_{1,i} + \psi_1} u_2^{X_2} \cdots u_{N+1}^{X_{N+1}} \right] u_0^k \mathbb{P}(X_0 = k) \end{aligned}$$

The expectation in the sum can be computed using the induction hypothesis, by changing the initial population law and by translating the reproductive laws of the branching process at each generation in the following way,

$$\begin{aligned} g_0 &\longrightarrow \tilde{g}_1^k \\ g_i, h_i &\longrightarrow g_{i+1}, h_{i+1} \text{ for all } 1 \leq i \leq N \end{aligned}$$

Then

$$\begin{aligned} &= \sum_{k \geq 0} \tilde{g}_1^k (g_2(\cdots g_{N+1}(u_{N+1}) \cdots) u_1) \prod_{i=1}^{N-1} h_{i+1} (g_{i+2}(\cdots g_{N+1}(u_{N+1}) \cdots u_{i+1}) \\ &\quad \times u_0^k \mathbb{P}(X_0 = k) \\ &= \prod_{i=1}^{N-1} h_{i+1} (g_{i+2}(\cdots g_{N+1}(u_{N+1}) \cdots u_{i+1}) h_1 (g_2(\cdots g_{N+1}(u_{N+1}) \cdots u_1) \\ &\quad \times \sum_{k \geq 0} (g_1 (g_2(\cdots g_{N+1}(u_{N+1}) \cdots) u_1))^k u_0^k \mathbb{P}(X_0 = k) \\ &= \prod_{i=1}^N h_i (g_{i+1}(\cdots g_{N+1}(u_{N+1}) \cdots u_i) \times g_0 (g_1(\cdots g_{N+1}(u_{N+1}) \cdots) u_0) \end{aligned}$$

Which concludes the proof of the lemma.  $\square$

All the elements exposed above lead to the following result about the function  $G$  defined at the begining of this section.

**Theorem 11.** *The pgf  $G : [0, 1]^{N+1} \rightarrow \mathbb{R}_+$ , which is defined as*

$$G(u_0, u_1, \dots, u_N) = \tilde{\mathbb{E}} \left[ 1_{\{\tilde{H}_1 < t_1, \dots, \tilde{H}_{K-1} < t_{K-1}\}} u_0^{X_{T_0}} u_1^{X_{T_1}} \cdots u_N^{X_{T_N}} \right],$$

*satisfies the following identity under the framework described in this section and the corresponding notation,*

$$\begin{aligned} G(u_0, u_1, \dots, u_N) &= \prod_{l=1}^K g_{0,l} (g_{1,l}(\cdots (g_{N,l}(u_N) u_{N-1})) u_0) \times \\ &\quad \prod_{l=0, K+1} \left( g_{0,l} (g_{1,l}(\cdots g_{N,l}(u_N) \cdots) u_0) \prod_{i=1}^{N-1} h_{i,l} (g_{i+1,l}(\cdots g_{N,l}(u_N) \cdots u_{i+1}) u_i) \right), \end{aligned}$$

where the functions  $g$ 's and  $h$ 's are defined by (I.18), (I.19), (I.20), (I.21) and Lemma 8.

## 2.6 Conclusion

We have briefly reviewed some of the solutions given so far to the problem of calculating the likelihood of a reconstructed tree when the underlying model is a birth and death process. We have subsequently extended some of these approaches by characterizing the conditional likelihood of the number of infectious individuals through time given the transmission tree linking individuals that are currently infected. This quantity was previously unknown for this model and is required for applications involving data. However, the inversion of the generating functions we have obtained here is a difficult problem, computationally and mathematically, so the expressions obtained are difficult to use in terms of statistical data analysis.

Beyond the practical aspects of the application of our results, according to our understanding, the most important contribution of this work is hidden in (I.17), which allowed us to unveil a nice time-reversal duality for CMJ processes, that was unknown even in the case of birth and death process and is the target of the forthcoming Chapter II.

## Appendix

### I.A Remaining proofs

*Proof of Lemma 1.*

$$\begin{aligned}
\mathbb{P}(\tilde{H} \leq t) &= \mathbb{E}[1_{\{\tilde{H} \leq t\}}] = \mathbb{E}[\mathbb{E}[1_{\tilde{H} \leq t} \mid \xi]] = \sum_{j \geq 1} \mathbb{E}[1_{\{\tilde{H} \leq t\}} \mid \xi = j] \mathbb{P}(\xi = j) \\
&= \sum_{j \geq 1} \mathbb{P}(\max_{1 \leq i \leq j} H_i \leq t) \mathbb{P}(\xi = j) = \sum_{j \geq 1} \mathbb{P}(H_1 \leq t, \dots, H_j \leq t) \mathbb{P}(\xi = j) \\
&= \sum_{j \geq 1} (\mathbb{P}(H \leq t))^j \mathbb{P}(\xi = j) = \mathbb{E}[(F(t))^\xi] = \frac{F(t)p}{1 - F(t)(1 - p)}
\end{aligned}$$

Where  $F(t) = \mathbb{P}(H \leq t)$ . Note that since  $p < 1$  the numerator of the last expression is nonzero. Finally under the assumption that  $H$  has probability density function  $f(t) = F'(t)$  it holds:

$$\tilde{f}(t) = \frac{f(t)p}{(1 - F(t)(1 - p))^2}$$

□

*Proof of Lemma 3.* With an elementary calculation we can compute the law of  $X_1$  conditional on  $X_1 \leq X_2$ . For every  $k \geq 1$ ,

$$\begin{aligned}
\mathbb{P}(X_1 = k \mid X_1 \leq X_2) &= \frac{\mathbb{P}(X_1 = k, X_1 \leq X_2)}{\mathbb{P}(X_1 \leq X_2)} = \frac{\mathbb{P}(X_1 = k, X_2 \geq k)}{\sum_{j \geq 1} \mathbb{P}(X_1 = j, X_2 \geq j)} \\
&= \frac{\mathbb{P}(X_1 = k) \mathbb{P}(X_2 \geq k)}{\sum_{j \geq 1} \mathbb{P}(X_1 = j) \mathbb{P}(X_2 \geq j)} = \frac{(1 - \alpha)^{k-1} \alpha (1 - \beta)^{k-1}}{\sum_{j \geq 1} (1 - \alpha)^{j-1} \alpha (1 - \beta)^{j-1}} \\
&= ((1 - \alpha)(1 - \beta))^{k-1} (1 - (1 - \alpha)(1 - \beta)) \\
&= (1 - \gamma)^{k-1} \gamma,
\end{aligned} \tag{A-I.1}$$

which is a geometric distribution with parameter  $\gamma$ .

The same holds for  $X_1$  conditional on  $X_1 < X_2$ , which is also  $\mathcal{G}(\gamma)$ . Notice that the symmetry on the parameter  $\gamma$  implies the result also holds if we swap  $X_1$  and  $X_2$ .

Now we are interested to calculate the probability generating function of  $X_1$ , when  $X_1$  is smaller than  $X_2$ , where  $X_1 \sim \mathcal{G}(\alpha)$  and  $X_2 \sim \mathcal{G}(\beta)$  are independent. Using Equation (A-I.1) and the fact that  $\mathbb{P}(X_1 \leq X_2) = \frac{\alpha}{\gamma}$  we obtain that

$$\begin{aligned}
\mathbb{E}[u^{X_1} 1_{\{X_1 \leq X_2\}}] &= \sum_{k \geq 1} u^k \mathbb{P}(X_1 = k, X_1 \leq X_2) = \sum_{k \geq 1} u^k \mathbb{P}(X_1 = k \mid X_1 \leq X_2) \mathbb{P}(X_1 \leq X_2) \\
&= \sum_{k \geq 1} u^k (1 - \gamma)^{k-1} \gamma \frac{\alpha}{\gamma} = \frac{\alpha u}{1 - u(1 - \gamma)}
\end{aligned}$$

A similar procedure yields the second generating function in the lemma. □

## I.B Some useful formulas and calculations

Formula (8.29) in [Kyp06] can be easily adapted to spectrally positive Lévy processes, then for  $x \in (0, t)$ , and every  $u \in (0, t]$ ,  $v \in (0, +\infty)$ :

$$\begin{aligned} & \mathbb{P}_x \left( (X_{\tau_t^+} - t) \in dv, (t - X_{\tau_t^+}) \in du, \tau_t^+ < \tau_0 \right) \\ &= \left[ \frac{W(t-x)W(t-u) - W(t)W(t-x-u)}{W(t)} \right] du \Lambda(dv+u) \end{aligned} \quad (\text{A-I.2})$$

**Lemma 12.** *In the Markovian (finite) case, the law of the overshoot of the contour process above a level  $t$ , is still exponential with same parameter as the life span of individuals, and it is independent of the fact that the process reaches  $(t, +\infty)$  before 0, under  $\mathbb{P}_x$  (with  $x < t$ ). More precisely:*

$$\mathbb{P}_x \left( (X_{\tau_t^+} - t) \in dv, \tau_t^+ < \tau_0 \right) = \mathbb{P}_x(\tau_t^+ < \tau_0) \mu e^{-\mu v} dv \quad (\text{A-I.3})$$

*Proof.* We can directly compute this probability by integrating (A-I.2) with respect to variable  $u$ , from 0 to  $t$ , and replacing  $W$  by its expression in the exponential case where  $\Lambda(ds) = \lambda \mu e^{-\mu s} ds$ , which is known to be

$$W(s) = \frac{\lambda e^{rs} - \mu}{r}, \quad \text{if } s \geq 0, \quad W(s) = 0, \quad \text{otherwise.}$$

This result can also be admitted without any computation, from the strong Markov property and the lack of memory of the exponential distribution.  $\square$

An elementary calculation using (I.7) leads to the following two formulas which will be used several times to prove our main results,

$$\int_0^t W(t-u) e^{-\mu u} du = \frac{e^{rt} - 1}{r} = \frac{W(t) - 1}{\lambda}, \quad (\text{A-I.4})$$

and for every  $t \in [0, T]$  we have

$$\int_0^{T-t} \mathbb{P}_{t+u} \left( \tau_T^+ < \tau_0 \right) e^{-\mu u} du = e^{r(T-t)} \frac{W(t)}{W(T)} - e^{-\mu(T-t)}. \quad (\text{A-I.5})$$

## Chapter II

# Time reversal dualities for some random forests

*This chapter is based in the article [DFL15] appeared in ALEA Lat. Am. J. Probab. Math. Stat.*

### 1 Introduction

We consider a model of branching population in continuous time, where individuals behave independently from one another. They give birth to identically distributed copies of themselves at some positive rate throughout their lives, and have generally distributed lifetime durations. A splitting tree [GK97, Lam10] describes the genealogical structure under this model and the associated population size process is a so-called (binary, homogeneous) Crump-Mode-Jagers (CMJ) process. When the lifetime durations are exponential or infinite (and only in this case) this is a Markov process, more precisely, a linear birth-death (BD) process.

Here, trees are assumed to originate at time 0 from one single ancestor. For a fixed time  $T > 0$ , we define a forest  $\mathcal{F}^*$  as a sequence of i.i.d. splitting trees, stopped at the first one having survived up to  $T$ , and we consider the associated population size or width process  $(\xi_t^*, 0 \leq t \leq T)$ . In the case of birth-death processes we have the following identity in distribution .

**Theorem 13.** *Let  $\mathcal{F}^*$  be a forest as defined previously, of supercritical birth-death trees with parameters  $b > d > 0$ , then the time-reversed process satisfies*

$$(\xi_{T-t}^*, 0 \leq t \leq T) \stackrel{d}{=} (\tilde{\xi}_t^*, 0 \leq t \leq T)$$

*where the right-hand side is the width process of an equally defined forest, but where the underlying trees are subcritical, obtained by swapping birth and death rates (or equivalently, by conditioning on ultimate extinction AN72).*

We further generalize this result to splitting trees, provided that the (i.i.d.) lifetimes of the ancestors have a specific distribution, explicitly known and different from that of their descendants. This additional condition comes from the *memory* in the distribution of the lifespans when they are not exponential, that imposes a distinction between ancestors and their descendants as we will see in Section 3. To our knowledge, this is the first time a result is established that reveals this kind of duality in branching processes: provided the

initial population is structured as described before, the width process seen backward in time is still the population size process of a similarly defined forest.

Furthermore, these dualities through an identity in distribution are established not only for the population size processes, but for the forests themselves. In other words, we give here the construction of the dual forest  $\tilde{\mathcal{F}}^*$ , from the forest  $\mathcal{F}^*$ , by setting up different filiations between them, but where the edges of the initial trees remain unchanged. This new genealogy has no interpretation, so far, in terms of the original family, and can be seen as the tool to reveal the intrinsic branching structure of the backward-in-time process.

The results are obtained via tree contour techniques and some properties of Lévy processes. The idea of coding the genealogical structure generated by the branching mechanism through a continuous or jumping stochastic process has been widely exploited with diverse purposes by several authors, see for instance [Pop04, GK97, LGLJ98, DLG02, Lam10, BPS12].

Here we make use of a particular way of exploring a splitting tree, called *jumping chronological contour process* (JCCP). We know from [Lam10] that this process has the law of a spectrally positive Lévy process properly reflected and killed. The notion of JCCP can be naturally extended to a forest by concatenation. Then our results are proved via a pathwise decomposition of the contour process of a forest and space-time-reversal dualities for Lévy processes [Ber92]. We define first some path transformations of the contour of a forest  $\mathcal{F}^*$ , after which, the reversed process will have the law of the contour of a forest  $\tilde{\mathcal{F}}^*$ . The invariance of the *local time* (defined here as the number of times the process hits a fixed value of its state-space  $\mathbb{R}_+$ ) of the contour after these transformations, allows us to deduce the aforementioned identity in distribution between the population size processes.

Branching processes are commonly used in biology to represent, for instance, the evolution of individuals with asexual reproduction [Jag91, KA02], or a group of species [NMH94, Sta09], as well as the spread of an epidemic outbreak in a sufficiently large susceptible population [Bec74, Bec77, TFLS06]. We are particularly interested in the last application, which was the primary motivation for this work, and where our duality result has some interesting consequences.

When modeling epidemics, we specify that what was called so far a *birth event* should be thought of as a transmission event of the disease from one (infectious) individual to another (susceptible, assumed to be in excess). In the same way a *death event* will correspond to an infectious individual becoming non-infectious (will no longer transmit the pathogen, e.g. recovery, death, emigration, etc.). Then the branching process describes the dynamics of the size of the infected population, and the splitting tree encodes the history of the epidemic.

In the last few decades, branching processes have found many applications as stochastic individual-based models for the transmission of diseases, especially the Markovian case (notice however, that the assumption of exponentially distributed periods of infectiousness is mainly motivated by mathematical tractability, rather than biological realism). In recent years, the possibility of sequencing pathogens from patients has been constantly increasing, and with it, the interest in using the phylogenetic trees of pathogen strains to infer the parameters controlling the epidemiological mechanisms, leading to a new approach in the field, the so-called phylodynamic methods [GPG<sup>+</sup>04]. A very short review on the subject is given later in Section 4.

Here we consider the situation where the data consists in incidence time series (number of new cases registered through time) and the *reconstructed transmission tree* (i.e. the

information about non sampled hosts is erased from the original process). These trees are considered to be estimated from pathogen sequences from present-time hosts. These observed statistics are assumed to be generated from a unique forward in time process and since no further hypothesis is made, they are not independent in general. Hence, the computation of the likelihood as their joint distribution quickly becomes a delicate and complex issue, even in the linear birth-death model, since it requires to integrate over all the possible extinct (unobserved) subtrees between 0 and  $T$ .

Therefore, to solve this problem, we propose a description of the population size process  $I := (I_t, 0 \leq t \leq T)$ , conditional on the reconstructed genealogy of individuals that survive up until time  $T$  (i.e. the reconstructed phylogeny). This result is a consequence of the aforementioned duality between random forests  $\mathcal{F}^*$  and  $\tilde{\mathcal{F}}^*$ . We state that, under these conditions, the process  $I$ , backward in time, is the sum of the width processes of independent birth-death trees, each conditioned on its height to be the corresponding time of coalescence, plus an additional tree conditioned on surviving up until time  $T$ .

The structure in the population, that is the definition of our forests, and the fact that in our model the sampled epidemic comes all from one single ancestor at time 0, can be thought of as a group of strains of a pathogen in their attempts to invade the population, but where only one succeeds (at time  $T$ ). However, if various invading strains succeed, analogous results can be deduced by concatenating (summing) an equal number of forests. The general assumption will be then, that for each *successful* strain, there is a geometric (random) number of other strains of the pathogen that become extinct before time  $T$ . The probability of success of these geometric r.v. depends on the recovery and transmission parameters. Finally, estimating these parameters from molecular and epidemiological data using this branching processes model, can be addressed through MLE or Bayesian inference. However, these statistical questions are not directly treated here.

For the sake of clarity, and in order to be consistent with most of the literature on branching processes, we will prefer to use the terms of birth and death events throughout the document, except when we present the outcomes of our results in the specific context of epidemiology in Section 4. The rest of the paper is organized as follows: Section 2 is dedicated to some preliminaries on Lévy processes, trees, forests and their respective contours. Finally, in Section 3, we state our main results and give most of their proofs, although some of them are left to the Appendix.

## 2 Preliminaries

### Basic notation

Let  $E = \mathbb{R} \cup \{\partial\}$  where  $\partial$  is a topologically isolated point, so-called *cemetery point*. Let  $\mathcal{B}(E)$  denote the Borel  $\sigma$ -field on  $E$ . Consider the space  $\mathcal{D}(\mathbb{R}_+, E)$  (or simply  $\mathcal{D}$ ) of càdlàg functions  $\omega$  from  $\mathbb{R}_+$  into the measurable space  $(E, \mathcal{B}(E))$  endowed with Skorokhod topology [JS03], stopped upon hitting  $\partial$  and denote the corresponding Borel  $\sigma$ -field by  $\mathcal{B}(\mathcal{D})$ . Define the *lifetime* of a path  $\omega \in \mathcal{D}$  as  $\zeta = \zeta(\omega)$ , the unique value in  $\mathbb{R}_+ \cup \{+\infty\}$  such that  $\omega(t) \in \mathbb{R}$  for  $0 \leq t < \zeta$ , and  $\omega(t) = \partial$  for every  $t \geq \zeta$ . Here  $\omega(t-)$  stands for the left limit of  $\omega$  at  $t \in \mathbb{R}_+$ ,  $\Delta\omega(t) = \omega(t) - \omega(t-)$  for the size of the (possible) jump at  $t \neq \zeta$  and we make the usual convention  $\omega(0-) = \omega(0)$ .

We consider stochastic processes, on the probability space  $(\mathcal{D}, \mathcal{B}(\mathcal{D}), \mathbf{P})$ , say  $X = (X_t, t \geq 0)$ , also called the coordinate process, having  $X_t = X_t(\omega) = \omega(t)$ . In particular, we consider only processes with no-negative jumps, that is such that  $\Delta X_t \in \mathbb{R}_+$  for every  $t \geq 0$ . The canonical filtration is denoted by  $(\mathcal{F}_t)_{t \geq 0}$ .



Let  $\mathcal{P}(E)$  be the collection of all probability measures on  $E$ . We use the notation  $\mathbf{P}_x(X \in \cdot) = \mathbf{P}(X \in \cdot | X_0 = x)$  and for  $\mu \in \mathcal{P}(E)$ ,

$$\mathbf{P}_\mu(X \in \cdot) := \int_E \mathbf{P}_x(X \in \cdot) \mu(dx).$$

For any measure  $\mu$  on  $[0, \infty]$ , we denote by  $\bar{\mu}$  its tail, that is

$$\bar{\mu}(x) := \mu([x, +\infty])$$

Define by  $\tau_A := \inf\{t > 0 : X_t \in A\}$ , the first hitting time of the set  $A \in \mathcal{B}(E)$ , with the conventions  $\tau_x = \tau_{\{x\}}$ , and  $\tau_x^- = \tau_{(-\infty, x)}$ ,  $\tau_x^+ = \tau_{(x, +\infty)}$  for any  $x \in \mathbb{R}$ .

### Some path transformations of càdlàg functions

In this subsection we will define some deterministic path transformations and functionals of càdlàg stochastic processes. Define first the following classical families of operators acting on the paths of  $X$ :

- the *shift operators*,  $\theta_s, s \in \mathbb{R}_+$ , defined by

$$\theta_s(X_t) := X_{s+t}, \quad \forall t \in \mathbb{R}_+$$

- the *killing operators*,  $k_s, s \in \mathbb{R}_+$ , defined by

$$k_s(X) := \begin{cases} X_t & \text{if } s < t \\ \partial & \text{otherwise} \end{cases}$$

the killing operator can be generalized to killing at random times, for instance  $X \circ k_{\tau_A} = k_{\tau_A}(X) = k_{\tau_A(X)}(X)$ , denotes the process *killed at the first passage into A*. It is easy to see that if  $X$  is a Markov process, so is  $X \circ k_{\tau_A}$ .

- the *space-time-reversal operator*  $\rho_s, s \in \mathbb{R}_+^*$ , as

$$\rho_s(X)_t := X_0 - X_{(s-t)-} \quad \forall t \in [0, s)$$

and we denote simply by  $\rho$  the space-time-reversal operator at the lifetime of the process, when  $\zeta < +\infty$ , that is

$$\rho(X)_t := X_0 - X_{(\zeta-t)-} \quad \forall t \in [0, \zeta)$$

The notations  $\mathbf{P} \circ \theta_s^{-1}$ ,  $\mathbf{P} \circ k_s^{-1}$  and  $\mathbf{P} \circ \rho^{-1}$  stand for the law of the shifted, killed and space-time-reversed processes when  $\mathbf{P}$  is the law of  $X$ .

When  $X$  has finite variation we can define its local time process, taking values in  $\mathbb{N} \cup \{+\infty\}$  as follows,

$$\Gamma_r(X) = \text{Card}\{t \geq 0 : X_t = r\}, \quad r \in \mathbb{R},$$

that is the number of times the process hits  $r$  (before being sent to  $\partial$ ).

For a sequence of processes in the same state space, say  $(X_i)_{i \geq 1}$  with lifetimes  $(\zeta_i, i \geq 1)$ , we define a new process by the concatenation of the terms of the sequence, denoted by

$$[X_1, X_2, \dots]$$

where the juxtaposition of terms is considered to stop at the first element with infinite lifetime. For instance, if  $\zeta_1 < +\infty$  and  $\zeta_2 = +\infty$ , then for every  $n \geq 2$

$$[X_1, X_2, \dots, X_n]_t = \begin{cases} X_{1,t} & \text{if } 0 \leq t < \zeta_1 \\ X_{2,t-\zeta_1} & t \geq \zeta_1 \end{cases}$$

We consider now a *clock* or time change that was introduced in [Ber96, Don07] in order to construct the probability measure of a Lévy process conditioned to stay positive, that will have the effect of erasing all the subpaths of  $X$  taking non-positive values and closing up the gaps. We define it here for a càdlàg function  $X$ , for which we introduce the time it spends in  $(0, +\infty)$ , during a fixed interval  $[0, t]$ , for every  $t \leq \zeta(X)$ ,

$$A_t := \int_0^t \mathbb{1}_{\{X_u > 0\}} du$$

and its right-continuous inverse,  $\alpha(t) := \inf \{u \geq 0 : A_u > t\}$ , such that a time substitution by  $\alpha$ , gives a function with values in  $[0, +\infty) \cup \{\partial\}$ , in the following sense,

$$(X \circ \alpha)_t = \begin{cases} X_{\alpha(t)} & \text{if } \alpha(t) < +\infty \\ \partial & \text{otherwise} \end{cases}$$

*Remark 14.* Analogous time changes  $\alpha^s$  (or  $\alpha_s$ ) can be defined for any  $s \in \mathbb{R}$ , removing the excursions of the function above (or below) the level  $s$ , that is by time-changing  $X$  via the right-continuous inverse of  $t \mapsto \int_0^t \mathbb{1}_{\{X_u < s\}} du$  (or  $t \mapsto \int_0^t \mathbb{1}_{\{X_u > s\}} du$ ).

**Last passage from  $T$  to 0:** Fix  $T > 0$ . Define the following special points for  $X$

$$\begin{aligned} g_T &:= \inf\{t > 0 : X_t = T\} \\ g_0 &:= \inf\{t > g_T : X_{t-} = 0\} \\ \bar{g}_T &:= \sup\{t < g_0 : X_t = T\} \end{aligned}$$

and suppose  $g_0 < +\infty$ . We want to define a transformation of  $X$  so that the subpath in the interval  $[\bar{g}_T, g_0]$  will be placed at the beginning, shifting to the right the rest of the path. Therefore, we define the function  $\vartheta(X) = \vartheta : [0, g_0] \rightarrow [0, g_0]$  as,

$$\vartheta(t) := \begin{cases} \bar{g}_T + t & \text{if } 0 \leq t < g_0 - \bar{g}_T \\ t - (g_0 - \bar{g}_T) & \text{if } g_0 - \bar{g}_T \leq t \leq g_0 \end{cases}$$

and then we consider the transformed path  $\chi(X)$ , defined for each  $s \geq 0$  as,

$$\chi(X)_s := \begin{cases} X_{\vartheta(s)} & \text{if } s \in [0, g_0) \\ \partial & \text{if } s \geq g_0 \end{cases}$$

## 2.1 Spectrally positive Lévy processes

*Lévy processes* are those stochastic processes with stationary and independent increments, and almost sure right continuous with left limits paths. During this subsection we recall some classic results from this theory and establish some others that will be used

later. We refer to [Ber96] and [Kyp06] for a detailed review on the subject.

We consider a real-valued Lévy process  $Y = (Y_t, t \geq 0)$  and we denote by  $P_x$  its law conditional on  $Y_0 = x$ . We assume  $Y$  is *spectrally positive*, meaning it has no negative jumps. This process is characterized by its Laplace exponent  $\psi$ , defined for any  $\lambda \geq 0$  by

$$E_0 \left[ e^{-\lambda Y_t} \right] = e^{t\psi(\lambda)}.$$

We assume, furthermore, that  $Y$  has finite variation. Then  $\psi$  can be expressed, thanks to the Lévy-Khintchine formula as

$$\psi(\lambda) = -\mathbf{d}\lambda - \int_0^\infty (1 - e^{-\lambda r}) \Pi(dr), \quad (\text{II.1})$$

where  $\mathbf{d} \in \mathbb{R}$  is called *drift coefficient* and  $\Pi$  is a  $\sigma$ -finite measure on  $(0, \infty]$ , called the *Lévy measure*, satisfying  $\int_0^\infty (r \wedge 1) \Pi(dr) < \infty$ . Notice we allow  $\Pi$  to charge  $+\infty$ , which amounts to killing the process at rate  $\Pi(\{+\infty\})$ .

The Laplace exponent is infinitely differentiable, strictly convex (when  $\Pi \not\equiv 0$ ),  $\psi(0) = 0$  (except when  $\Pi$  charges  $+\infty$ , in which case  $\psi(0) = \psi(0+) = -\Pi(\{+\infty\})$ ) and  $\psi(+\infty) = +\infty$ . Suppose  $\Pi(\{+\infty\}) = 0$  and  $\mathbf{d} = -1$  and define

$$\eta := \sup\{\lambda \geq 0 : \psi(\lambda) = 0\}, \quad (\text{II.2})$$

the largest root of  $\psi$  and

$$m := \int_{(0, \infty)} r \Pi(dr).$$

Then we have that  $\eta = 0$  is the unique root of  $\psi$ , when  $\psi'(0+) = 1 - m \geq 0$ . Otherwise the Laplace exponent has two roots, 0 and  $\eta > 0$ . It is known that for any  $x > 0$ ,

$$P_x(\tau_0 < +\infty) = e^{-\eta x}.$$

More generally, we know from Theorem VII.8 in [Ber96] that for completely asymmetric Lévy processes (that make no positive or no negative jumps) there exists a unique continuous increasing function  $W : [0, +\infty) \rightarrow [0, +\infty)$ , called the *scale function*, characterized by its Laplace transform,

$$\int_0^{+\infty} e^{-\lambda x} W(x) dx = \frac{1}{\psi(\lambda)}, \quad \lambda > \eta,$$

such that for any  $0 < x \leq a$ ,

$$P_x(\tau_0 < \tau_a^+) = \frac{W(a-x)}{W(a)}. \quad (\text{II.3})$$

### Pathwise decomposition

For a Markov process, a point  $x$  of its state space is said to be *regular* or *irregular* for itself, if  $\mathbf{P}_x(\tau_x = 0)$  is 1 or 0. In a similar way, when the process is real-valued, we can say it is *regular downwards* or *upwards* if we replace  $\tau_x$  by  $\tau_x^-$  or  $\tau_x^+$  respectively. For a spectrally positive Lévy processes we know from [Ber96] that  $Y$  has bounded variation if and only if

0 is irregular (and irregular upwards). In this case there is a natural way of decomposing the process into excursions from any point  $x$  on its state space. For simplicity, here we depict the situation for  $x = 0$ , the generalization being straightforward from the Markov property.

The process  $Y$  under  $P_0$ , can be described as a sequence of independent and identically distributed excursions from 0, stopped at the first one with infinite lifetime. Define the sequence of the successive hitting times of 0, say  $(\tau_0^{(i)}, i \geq 0)$  with  $\tau_0^{(0)} = 0$ . For  $i \geq 0$ , on  $\{\tau_0^{(i)} < +\infty\}$ , define the shifted process,

$$\epsilon_i := \left( Y_{\tau_0^{(i)}+t}, 0 \leq t < \tau_0^{(i+1)} - \tau_0^{(i)} \right).$$

The strong Markov property and the stationarity of the increments imply that  $(\epsilon_i, i \geq 0)$  is a sequence of i.i.d. excursions, all distributed as  $(Y_t, 0 \leq t \leq \tau_0)$  under  $P_0$ , with a possibly finite number of elements, say  $N + 1$ , which is geometric with parameter  $P_0(\tau_0 = +\infty)$ , corresponding to the time until the occurrence of an infinite excursion. Hence, the paths of  $Y$  are structured as the juxtaposition of these i.i.d. excursions,  $N$  with finite lifetime, followed by a final infinite excursion.

We now introduce, for any  $a > 0$ , the process  $Y$  reflected below  $a$ , that is the process being immediately restarted at  $a$  when it enters  $(a, +\infty)$ . This process is also naturally decomposed in its excursions below  $a$ , in the same way described before. More precisely, let  $(\epsilon_i, i \geq 1)$  be a sequence of i.i.d. excursions distributed as  $Y$  under  $P_a$ , but killed when they hit  $(a, +\infty)$ , that is, with common law  $P_a \circ k_{\tau_a^+}^{-1}$ . Notice that since the process is irregular upwards, then necessarily these excursions have a strictly positive lifetime  $P_a$ -a.s.. Define the reflected process  $Y^{(a)}$  as their concatenation, that is

$$Y^{(a)} := [\epsilon_1, \epsilon_2, \dots] \quad (\text{II.4})$$

### Overshoot and undershoot

Formula (8.29) from [Kyp06] adapted to spectrally positive Lévy processes gives the joint distribution of the overshoot and undershoot of  $Y$  when it first enters the interval  $[a, +\infty]$ , without hitting 0. Let  $x \in (0, a)$ , then for  $u \in (0, a]$  and  $v \in (0, +\infty)$ ,

$$\begin{aligned} P_x \left( (a - Y_{\tau_a^+}^-) \in du, (Y_{\tau_a^+} - a) \in dv, \tau_a^+ < \tau_0 \right) \\ = \left( \frac{W(a-x)W(a-u)}{W(a)} - W(a-x-u) \right) du \Pi(dv+u) \end{aligned} \quad (\text{II.5})$$

In the same way, if the restriction on the minimum of the process before hitting  $[a, +\infty)$  is removed, choosing  $x = a = 0$  for simplicity, we have for  $u, v \in (0, +\infty)$

$$P_0 \left( -Y_{\tau_0^+}^- \in du, Y_{\tau_0^+} \in dv, \tau_0^+ < +\infty \right) = e^{-\eta u} du \Pi(dv+u) \quad (\text{II.6})$$

Then we can compute the distribution of the undershoot and overshoot of an excursion away from 0, of the process  $Y$  starting at 0 on the event  $\tau_0^+ < +\infty$ . By integrating (II.6) we get,

$$\begin{aligned} P_0 \left( -Y_{\tau_0^+}^- \in du, \tau_0^+ < +\infty \right) &= du \int_{[0, +\infty]} e^{-\eta u} \Pi(dv+u) = e^{-\eta u} \bar{\Pi}(u) du \\ P_0 \left( Y_{\tau_0^+} \in dv, \tau_0^+ < +\infty \right) &= e^{\eta v} dv \int_{[v, +\infty]} e^{-\eta y} \Pi(dy) = e^{\eta v} \bar{\Pi}(v) dv \end{aligned}$$

where we define the measure

$$\tilde{\Pi}(dy) := e^{-\eta y} \Pi(dy)$$

on  $(0, +\infty)$  of mass  $\tilde{b} := b - \eta$  when  $\Pi$  has mass  $b$ .

Further, we can deduce again from Theorem VII.8 in [Ber96] that

$$P_0(\tau_0^+ < +\infty) = 1 - \frac{W(0)}{W(\infty)} = 1 \wedge m \quad (\text{II.7})$$

as a consequence of (II.3) and

$$\lim_{t \rightarrow +\infty} W(t) = \begin{cases} +\infty & \text{if } m \geq 1 \\ \frac{1}{1-m} & \text{if } m < 1 \end{cases}, \quad W(0) = 1.$$

These two formulas come from the analysis of the behavior at 0 and  $+\infty$  of  $W$ 's Laplace transform,  $1/\psi$ , followed by the application of a Tauberian theorem. We refer to Propositions 5.4 and 5.8 from [Lam10] for the details.

We denote by  $\mu_{\top}$  the probability measure on  $[0, \infty)$  of the undershoot away from 0 under  $P_0(\cdot | \tau_0^+ < +\infty)$ , defined as follows,

$$\mu_{\top}(du) := P_0(-Y_{\tau_0^+} \in du | \tau_0^+ < +\infty) = \frac{e^{-\eta u} \tilde{\Pi}(u)}{m \wedge 1} du \quad (\text{II.8})$$

Analogously the probability distribution of the corresponding overshoot is,

$$\mu_{\perp}(dv) := P_0(Y_{\tau_0^+} \in dv | \tau_0^+ < +\infty) = \frac{e^{\eta v} \tilde{\Pi}(v)}{m \wedge 1} dv \quad (\text{II.9})$$

*Remark 15.* In the exponential case with rates  $b, d > 0$ , that is when  $\Pi(dr) = bde^{-dr}dr$  and

$$\psi(\lambda) = \lambda - \frac{b\lambda}{d + \lambda}, \quad \lambda \geq 0,$$

it is not hard to see that the overshoot is also exponentially distributed with parameter  $d$ , and the undershoot with parameter  $b \vee d$ .

### Process conditioned on not drifting to $+\infty$

The following statements are direct consequences of Corollary VII.2 and Lemma VII.7 from [Ber96]. The process  $Y$  drifts to  $-\infty$ , oscillates or drifts to  $+\infty$ , if  $\psi'(0+)$  is respectively positive, zero, or negative. As we mentioned before, only in the last case the Laplace exponent  $\psi$  has a strictly positive root  $\eta$ , which leads to considering a new family of probability measure  $\tilde{P}_x$  via the exponential martingale  $(e^{-\eta(Y_t - x)}, t \geq 0)$ , that is with Radon-Nikodym density

$$\left. \frac{d\tilde{P}_x}{dP_x} \right|_{\mathcal{F}_t} = e^{-\eta(Y_t - x)}.$$

As we will show later, this can be thought of as the law of the initial Lévy process conditioned to not drift to  $+\infty$ , and in fact,  $Y$  under  $\tilde{P}$  is still a spectrally positive Lévy process with Laplace exponent

$$\tilde{\psi}(\lambda) = \psi(\eta + \lambda) = -d\lambda - \int_0^{\infty} (1 - e^{-\lambda r}) \tilde{\Pi}(dr),$$

which is the Laplace exponent of a finite variation, spectrally positive Lévy process with drift  $d$  and Lévy measure  $\tilde{\Pi}$ . Furthermore, it is established that for every  $x > 0$ , the law of the process until the first hitting time of 0, that is  $(Y_t, 0 \leq t < \tau_0)$ , is the same under  $P_x(\cdot | \tau_0 < +\infty)$  as under  $\tilde{P}_x(\cdot)$ , that is

$$P_x(\cdot | \tau_0 < +\infty) \circ k_{\tau_0}^{-1} = \tilde{P}_x \circ k_{\tau_0}^{-1} \quad (\text{II.10})$$

Finally, we compute the following convolution products, which will be used hereafter, obtained from a direct inversion of the Laplace transform for  $W$  and  $\tilde{W}$  (see p. 204-205 in [Ber96]), where  $\tilde{W}$  is the scale function defined with respect to the Laplace exponent  $\tilde{\psi}$ ,

$$\int_0^T W(T-v) \tilde{\Pi}(v) dv = W(T) - 1 \quad (\text{II.11})$$

$$\int_0^T \tilde{W}(T-v) \tilde{\Pi}(v) dv = \tilde{W}(T) - 1 \quad (\text{II.12})$$

### Time-reversal for Lévy processes

Another classical property of Lévy processes is their duality under time-reversal in the following sense: if a path is space-time-reversed at a finite time horizon, the new path has the same distribution as the original process. More precisely, in the case of spectrally positive Lévy processes we have the following results from [Ber92]:

**Proposition 16** *Duality. The process  $Y$  has the following properties:*

- (i) *under  $P_0(\cdot | \tau_0 < +\infty)$ ,  $(Y_t, 0 \leq t < \tau_0)$  and  $(-Y_{(\tau_0-t)-}, 0 \leq t < \tau_0)$  have the same law*
- (ii) *under  $P_0(\cdot | -Y_{\tau_0^+} = u)$  the reversed excursion,  $(-Y_{(\tau_0^+-t)-}, 0 \leq t < \tau_0^+)$  has the same distribution as  $(Y_t, 0 \leq t < \tau_0)$  under  $P_u(\cdot | \tau_0 < +\infty)$*
- (iii) *under  $P_0(\cdot | -Y_{\tau_0^+} = y, \Delta Y(\tau_0^+) = z)$ , the processes  $(-Y_{(\tau_0^+-t)-}, 0 \leq t < \tau_0^+)$  and  $(Y_{\tau_0^++t}, 0 \leq t < \tau_0 - \tau_0^+)$  are independent. The first one has law  $P_y(\cdot | \tau_0 < +\infty) \circ k_{\tau_0}^{-1}$  and the second one has law  $P_{z-y} \circ k_{\tau_0}^{-1}$*

### 2.2 Trees and forests

We refer to [Lam10] for the rigorous definition and properties of discrete and chronological trees. The notation here may differ from that used by this author, so it will be specified in the following, as well as the main features that will be subsequently required.

A *discrete tree*, denoted by  $\mathcal{U}$ , is a subset of  $\mathbb{U} := \bigcup_{n \geq 0} \mathbb{N}^n$ , satisfying some specific well known properties. From a discrete tree  $\mathcal{U}$  we can obtain an  $\mathbb{R}$ -tree by adding birth levels to the vertices and lengths (*lifespans*) to the edges, getting what is called a *chronological tree*  $\mathcal{T}$ . For each individual  $u$  of a discrete tree  $\mathcal{U}$ , the associated *birth level* is denoted by  $\alpha(u)$  and the *death level* by  $\omega(u)$  ( $\alpha(u) \in \mathbb{R}_+$ ,  $\omega(u) \in \mathbb{R}_+ \cup \{+\infty\}$  and such that  $\alpha(u) < \omega(u)$ ).

Then  $\mathcal{T}$  can be seen as the subset of  $\mathbb{U} \times [0, +\infty)$  containing all the *existence points* of individuals (vertices of the discrete tree): for every  $u \in \mathcal{U}$ ,  $s \in [0, +\infty)$ , then  $(u, s) \in \mathcal{T}$  if and only if  $\alpha(u) < s \leq \omega(u)$ . The root will be denoted by  $\rho := (\emptyset, 0)$  and  $\pi_1$  and  $\pi_2$  stand respectively for the canonical projections on  $\mathbb{U}$  and  $[0, +\infty)$ .  $\mathbb{T}$  denotes the set of all chronological trees.

For any individual  $u$  in the discrete tree  $\mathcal{U} = \pi_1(\mathcal{T})$ , we denote by  $\zeta(u)$  its lifespan, i.e.  $\zeta(u) := \omega(u) - \alpha(u)$ . Then the *total length* of the chronological tree is the sum of the lifespans of all the individuals, that is,

$$\ell(\mathcal{T}) := \sum_{v \in \pi_1(\mathcal{T})} \zeta(v) \leq +\infty.$$

We will also refer to the *truncated tree* up to level  $s$ , denoted by  $\mathcal{T}^{(s)}$  for the chronological tree formed by the existence points  $(u, t)$  such that  $t \leq s$ . A chronological tree is said to be *locally finite* if for every level  $s \in [0, +\infty)$  the total length of the truncated tree is finite,  $\ell(\mathcal{T}^{(s)}) < +\infty$ .

We can define the *width* or *population size process* of locally finite chronological trees as a mapping  $\Xi$  that maps a chronological tree  $\mathcal{T}$  to the function  $\xi : \mathbb{R}_+ \rightarrow \mathbb{N}$  counting the number of extant individuals at time  $t \geq 0$

$$\Xi(\mathcal{T}) := (\xi_t(\mathcal{T}), t \geq 0) \tag{II.13}$$

where for every  $t \geq 0$ ,

$$\xi_t(\mathcal{T}) = \text{Card}\{v \in \pi_1(\mathcal{T}) : \alpha(v) < t \leq \omega(v)\}$$

These functions are càdlàg, piecewise constant, from  $\mathbb{R}_+$  into  $\mathbb{N} \cup \{+\infty\}$ , and are absorbed at 0. Then we can define the time of extinction of the population in a tree as  $T_{\text{Ext}} := \inf\{t \geq 0 : \xi_t(\mathcal{T}) = 0\}$ .

Chronological trees are assumed to be embedded in the plane, as on Fig. II.1 (right), with time running from bottom to top, dotted lines representing filiations between individuals: the one on the left is the parent, and that on the right its descendant.

We will call a *forest* every finite sequence of chronological trees, and we will denote the set of all forests by  $\mathbb{F}$ . More specifically for every positive integer  $m$ , let us define the set of *m-forests* as follows,

$$\mathbb{F}_m := \{(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m) : \mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m \in \mathbb{T}\}$$

then,

$$\mathbb{F} = \bigcup_{m \in \mathbb{N}} \mathbb{F}_m$$

It is straightforward to extend the notion of width process to a forest, say  $\mathcal{F} = (\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m) \in \mathbb{F}$ , as the sum of the widths of every tree of the sequence, i.e.,

$$\Xi(\mathcal{F}) := \sum_{i=1}^m \Xi(\mathcal{T}_i)$$

### 2.3 The contour process

As mentioned before, the genealogical structure of a chronological tree can be coded via continuous or càdlàg functions. They are usually called contour or exploration processes, since they refer mostly to deterministic functions of a (randomly generated) tree. In this case, the contour is a  $\mathbb{R}$ -valued stochastic process containing all the information about the tree, so that the latter can always be recovered from its contour. Among the different ways of exploring a tree we will exploit the jumping chronological contour process (JCCP) from [Lam10].

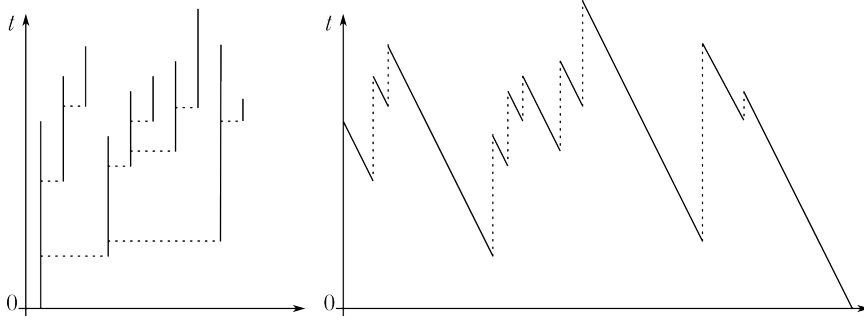


Figure II.1 – An example of chronographical tree with finite length (left) and its contour process (right).

The JCCP of a chronographical tree  $\mathcal{T}$  with finite length  $\ell = \ell(\mathcal{T})$ , denoted by  $\mathcal{C}(\mathcal{T})$ , is a function from  $[0, \ell]$  into  $\mathbb{R}_+$ , that starts at the lifespan of the ancestor and then walks backward along the right-hand side of this first branch at speed  $-1$  until it encounters a birth event, when it jumps up of a height of the lifespan of this new individual, getting to the next tip, and then repeating this procedure until it eventually hits 0, as we can see in Fig. II.1 (see [Lam10] for a formal definition).

Then it visits all the existence times of each individual exactly once and satisfies that the number of times it hits a time level, say  $s \geq 0$ , is equal to the number of individuals in the population at time  $s$ . More precisely, for any finite tree  $\mathcal{T}$

$$\Gamma \circ \mathcal{C}(\mathcal{T}) = \Xi(\mathcal{T}),$$

and more generally, if  $\mathcal{T}$  is locally finite, this is also satisfied for the truncated tree at any level  $s > 0$ , that is

$$\Gamma \circ \mathcal{C}(\mathcal{T}^{(s)}) = \Xi(\mathcal{T}^{(s)}).$$

We can extend the notion of contour process to a forest  $\mathcal{F} = (\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m)$  of finite total length  $\bar{\ell} := \sum_{i=1}^m \ell(\mathcal{T}_i)$ , similarly to the way it is done in [DLG02], by concatenating the contour functions,

$$[(\mathcal{C}_t(\mathcal{T}_1), t \in [0, \ell(\mathcal{T}_1))), (\mathcal{C}_t(\mathcal{T}_2), t \in [0, \ell(\mathcal{T}_2))), \dots, (\mathcal{C}_t(\mathcal{T}_m), t \in [0, \ell(\mathcal{T}_m))].$$

It will be denoted as well by  $\mathcal{C}(\mathcal{F})$  or simply  $\mathcal{C}$  when there is no risk of confusion. We notice that the function thus obtained determines a unique sequence of chronographical trees since they all start with one single ancestor.

## 2.4 Stochastic model

We consider a population (or particle system) that originates at time 0 with one single progenitor. Then individuals (particles) evolve independently of each other, giving birth to i.i.d. copies of themselves at constant rate, while alive, and having a life duration with general distribution. The family tree under this stochastic model will be represented by a *splitting tree*, that can be formally defined as an element  $\mathcal{T}$  randomly chosen from the set of chronographical trees, characterized by a  $\sigma$ -finite measure  $\Pi$  on  $(0, \infty]$  called the *lifespan measure*, satisfying  $\int_{(0, \infty]} (r \wedge 1) \Pi(dr) < \infty$ .

In the general definition individuals may have infinitely many offspring. However, for simplicity we will assume that  $\Pi$  has mass  $b$ , corresponding to a population where



individuals have i.i.d. lifetimes distributed as  $\Pi(\cdot)/b$  and give birth to single descendants throughout their lives at constant rate  $b$ , all having the same independent behavior. In most of the following results this hypothesis is not necessary, and they remain valid if  $\Pi$  is infinite (see [Lam10]).

Under this model the width process  $\Xi(\mathcal{T}) = (\xi_t(\mathcal{T}), t \geq 0)$  counting the population size through time is a *binary homogeneous Crump-Mode-Jagers process* (CMJ). This process is not Markovian, unless  $\Pi$  is exponential (*birth-death process*) or a Dirac mass at  $\{+\infty\}$  (*Yule process*).

A tree, or its width process  $\Xi$ , is said to be subcritical, critical or supercritical if

$$m := \int_0^{+\infty} r\Pi(dr)$$

is respectively less than, equal to or greater than 1, and we define the extinction event  $\text{Ext} := \{\lim_{t \rightarrow \infty} \xi_t(\mathcal{T}) = 0\}$ .

For a splitting tree we can define, as well as for its deterministic analogue, the JCCP. Actually, the starting point of the present work, is one of the key results in [Lam10], where the law of the JCCP of a splitting tree truncated up to  $T$  or conditional on having finite length is characterized by a Lévy process.

**Theorem 17** [Lam10]. *If  $X^{(T)}$  is the JCCP of a splitting tree with lifespan measure  $\Pi$  truncated up to  $T \in (0, +\infty)$  and  $Y$  is a spectrally positive Lévy process with finite variation and Laplace exponent  $\psi(\lambda) = \lambda - \int_0^\infty (1 - \exp(-\lambda r))\Pi(dr)$ ,  $\lambda \geq 0$ , then conditional on the lifespan of the ancestor to be  $x$ ,  $X^{(T)}$  has the law of  $Y$ , started at  $x \wedge T$ , reflected below  $T$  and killed upon hitting 0. Furthermore, conditional on extinction,  $X$  has the law of  $Y$  started at  $x$ , conditioned on, and killed upon hitting 0.*

We state here without proof the following elementary lemma that is repeatedly used in the proofs.

**Lemma 18.** *Let  $Z$  be a r.v. in a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ , taking values in a measurable space  $(E, \mathcal{A})$ . Let  $A \in \mathcal{A}$  be such that  $p := \mathbf{P}(Z \in A) \neq 0$ . Let  $Z_1, Z_2, \dots$  be a sequence of i.i.d. r.v. distributed as  $Z$ , and set  $N := \inf\{n : Z_n \in A\}$ . Then we have the following identity in distribution,*

$$(Z_1, \dots, Z_N) \stackrel{d}{=} (Z'_1, \dots, Z'_G, Z''_{G+1}),$$

where

- $G$  is a r.v. independent of  $Z'_1, \dots, Z'_G, Z''_{G+1}$  that has geometric distribution with parameter  $p$ , i.e.  $\mathbf{P}(G = k) = (1 - p)^k p$ , for  $k \geq 0$ ,
- $Z'_1, \dots, Z'_G$  are i.i.d. r.v. with probability distribution  $\mathbf{P}(Z \in \cdot | Z \notin A)$ ,
- $Z''_{G+1}$  is an independent r.v. distributed as  $\mathbf{P}(Z \in \cdot | Z \in A)$ .

### 3 Results

From now on we consider a finite measure  $\Pi$ , on  $(0, +\infty)$ , of mass  $b$  and  $m := \int_0^{+\infty} r\Pi(dr)$  and  $Y$  a spectrally positive Lévy process with drift coefficient  $\mathbf{d} = -1$ , Lévy measure  $\Pi$ , and Laplace exponent denoted by  $\psi$ , i.e.,

$$\psi(\lambda) = \lambda - \int_0^\infty (1 - e^{-\lambda r})\Pi(dr), \quad \lambda \geq 0 \tag{II.14}$$

As in the preliminaries, we denote by  $P_x$  the law of the process conditional on  $Y_0 = x$ , by  $W$  the corresponding scale function and by  $\eta$  the largest root of  $\psi$ . For any  $s \in \mathbb{R}_+$ , denote by  $Y^{(s)}$  the process reflected below  $s$ , as defined in the preliminaries.

We also recall the definition of the measure  $\tilde{\Pi}(dy) := e^{-\eta y} \Pi(dy)$  on  $(0, +\infty)$ . Then  $\tilde{\psi}, \tilde{m}, \tilde{W}, \tilde{P}, \tilde{Y}$ , will denote the Laplace exponent, the mean value, the scale function, the law, and the process itself with Lévy measure  $\tilde{\Pi}$ . As we have seen before, this spectrally positive Lévy process with Laplace exponent  $\tilde{\psi}$ , killed when it hits 0, has the same law as  $(Y_t, 0 \leq t < \tau_0)$  conditioned on hitting 0, when starting from any  $x > 0$ , that is  $\tilde{P}_x \circ k_{\tau_0}^{-1} = P_x(\cdot | \tau_0 < +\infty) \circ k_{\tau_0}^{-1}$ .

Fix  $p \in (0, 1)$  and  $T \in (0, +\infty)$ . We define a forest  $\mathcal{F}^p$  consisting in  $N_p + 1$  splitting trees with lifespan measure  $\Pi$  as follows,

$$\mathcal{F}^p := (\mathcal{T}_1, \dots, \mathcal{T}_{N_p}, \mathcal{T}_{N_p+1})$$

where,

- $\mathcal{T}_1, \dots, \mathcal{T}_{N_p}$  are i.i.d. splitting trees conditioned on extinction before  $T$
- $\mathcal{T}_{N_p+1}$  is an independent splitting tree conditioned on survival up until time  $T$
- $N_p$  is a geometric random variable with parameter  $p$ , independent of all trees in the sequence, i.e.  $\mathbb{P}(N_p = k) = (1 - p)^k p$ , for  $k \geq 0$ .

Notice that, if  $p = \mathbb{P}(\xi_T(\mathcal{T}) \neq 0)$ , then  $\mathcal{F}^p$  has the same distribution as a sequence of i.i.d. splitting trees, stopped at its first element surviving up until time  $T$  (see Lemma 18). Hereafter we will frequently make use of this identity in law.

We will add a subscript to denote equally constructed forests, but where the i.i.d. lifetimes of the ancestors on the splitting trees are different from that of their descendants. More precisely we will refer to  $\mathcal{F}_\perp^p$  or  $\mathcal{F}_\top^p$  if the ancestors are distributed respectively as the overshoot and undershoot defined by (II.9) and (II.8), conditional on  $\{\tau_0^+ < +\infty\}$ , i.e.,

$$\perp : \zeta(\emptyset) \sim \mu_\perp \qquad \top : \zeta(\emptyset) \sim \mu_\top$$

We follow the same convention for splitting trees, that is,  $\mathcal{T}_\top$  and  $\mathcal{T}_\perp$  denote trees starting from one ancestor with these distributions, as well as for their probability laws, denoted by  $\mathbb{P}_\top, \mathbb{P}_\perp$ .

Finally, we use the notation  $\tilde{\mathcal{F}}^p$  when the lifespan measure of all individuals is  $\tilde{\Pi}$ , instead of  $\Pi$ . The addition of a subscript is assumed to affect the lifetime distribution of the ancestors in the exact same way as described before.

We start with the following result, which is the extension of Propositions 16 and Theorem 17 to the case of these splitting trees with size-biased ancestors. Its proof is given later in the Appendix.

**Lemma 19.** *Let  $\mathcal{T}_\perp$  be a splitting tree and  $Y$  a spectrally positive Lévy process, as defined in Section 3, then the contour process  $\mathcal{C} = \mathcal{C}(\mathcal{T}_\perp)$  has the following properties*

- (i) *Under  $\mathbb{P}_\perp(\cdot | \text{Ext})$ ,  $\mathcal{C}$  has the same distribution as  $(Y_{\tau_0^+ + t}, 0 \leq t \leq \tau_0 - \tau_0^+)$  under  $P_0(\cdot | \tau_0 < +\infty)$ .*
- (ii) *Under  $\mathbb{P}_\perp(\cdot | \xi_T = 0)$ ,  $\mathcal{C}$  has the distribution of  $(Y_{\tau_0^+ + t}, 0 \leq t \leq \tau_0 - \tau_0^+)$  under  $P_0(\cdot | \tau_0 < \tau_T^+, \tau_0^+ < +\infty)$ .*

- (iii) Under  $\mathbb{P}_\perp$ , the contour of the truncated tree  $\mathcal{T}_\perp^{(T)}$ , is distributed as  $(Y_{\tau_0^+ + t}, 0 \leq t \leq \tau_0 - \tau_0^+)$  under  $P_0(\cdot | \tau_0^+ < +\infty)$ , reflected at  $T$  and killed upon hitting 0.

Define now the two parameters,

$$\gamma := \frac{1}{W(T)} = P_T(\tau_0 < \tau_T^+) \quad \text{and} \quad \tilde{\gamma} := \frac{1}{\tilde{W}(T)} = \tilde{P}_T(\tau_0 < \tau_T^+).$$

We have the following two results on forests,

**Lemma 20.** *In the supercritical and critical cases ( $m \geq 1$ ) we have*

$$\tilde{\gamma} = \mathbb{P}_\perp(\xi_T \neq 0) \quad \text{and} \quad \gamma = \tilde{\mathbb{P}}_\top(\xi_T \neq 0).$$

*Proof.* See Appendix.

Define the following forests stopped at first surviving tree,

- $\mathcal{F}^*$  := a sequence of i.i.d. splitting trees with law  $\mathbb{P}_\perp$  stopped at the first tree that survives up to time  $T$ .
- $\tilde{\mathcal{F}}^*$  := a sequence of i.i.d. splitting trees with law  $\tilde{\mathbb{P}}_\top$  stopped at the first tree that survives up to time  $T$ .

**Lemma 21.** *In the supercritical and critical cases ( $m \geq 1$ ) we have*

$$\mathcal{F}_\perp^{\tilde{\gamma}} \stackrel{d}{=} \mathcal{F}^* \quad \text{and} \quad \tilde{\mathcal{F}}_\top^{\gamma} \stackrel{d}{=} \tilde{\mathcal{F}}^*.$$

*Proof.* By definition, a forest  $\mathcal{F}^*$  consists in a number of trees, say  $\tilde{N} + 1$ , where  $\tilde{N}$  is a geometric random variable with probability of success  $\mathbb{P}_\perp(\xi_T \neq 0)$ , counting the trees that die out before  $T$ , until there is one that survives. Hence, thanks to Lemma 18, the only thing that remains to prove is that  $\tilde{\gamma}$  is exactly this probability of success for the forest  $\mathcal{F}^*$ , in the same way that  $\gamma$  for the forest  $\tilde{\mathcal{F}}^*$ , which is the statement in Lemma 20.  $\square$

Then we are ready to state our first result concerning the population size processes of these forests,

**Theorem 22.** *We have the following identity in distribution,*

$$(\xi_{T-t}(\mathcal{F}_\perp^{\tilde{\gamma}}), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\tilde{\mathcal{F}}_\top^{\gamma}), 0 \leq t \leq T).$$

*In the subcritical and critical cases ( $m \leq 1$ ),*

$$(\xi_{T-t}(\mathcal{F}_\perp^{\gamma}), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\mathcal{F}_\top^{\gamma}), 0 \leq t \leq T)$$

*and actually in this case  $\zeta(\emptyset)$  has density  $\mu_\perp(dr) = \mu_\top(dr) = \frac{\bar{\Pi}(r)}{m}dr$ .*

*In the supercritical and critical cases ( $m \geq 1$ ) we have*

$$(\xi_{T-t}(\mathcal{F}^*), 0 \leq t \leq T) \stackrel{d}{=} (\xi_t(\tilde{\mathcal{F}}^*), 0 \leq t \leq T).$$

**Remark 23.** Theorem 13 from the Introduction is a particular case of this theorem when  $\Pi$  is exponential.

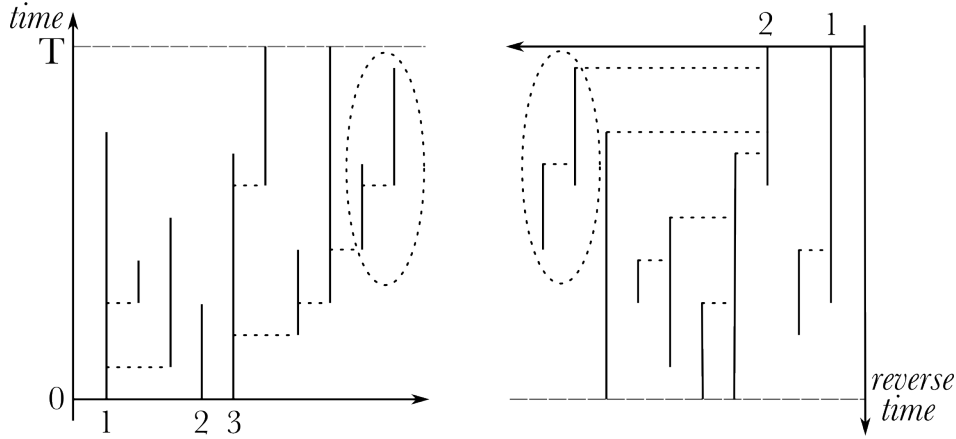


Figure II.2 – An example of forest  $\mathcal{F}^*$  consisting in three chronological trees (left) and its dual (right).

*Remark 24.* When  $m < 1$ ,  $\tilde{\mathcal{F}}_\perp^\gamma$  has no interpretation as a stopped sequence of i.i.d. splitting trees as in Lemma 21, because  $\gamma \neq \tilde{\mathbb{P}}_\top (\xi_T \neq 0)$ . Indeed, in this case

$$\begin{aligned} P_0 \left( \tau_0 < \tau_T^+ \mid \tau_0^+ < +\infty \right) &= \int_0^T P_v \left( \tau_0 < \tau_T^+ \right) P_0 \left( Y_{\tau_0^+} \in dv \mid \tau_0^+ < +\infty \right) \\ &= \int_0^T \frac{W(T-v)}{W(T)} \frac{\bar{\Pi}(v)}{m} dv = \frac{1}{m} \left( 1 - \frac{1}{W(T)} \right) \end{aligned}$$

where the last equality comes from (II.11). This entails that in this case the number of trees on the forest  $\tilde{\mathcal{F}}^*$  is geometric with parameter

$$1 - \frac{1 - \frac{1}{W(T)}}{1 - \frac{1}{W(\infty)}} \neq \gamma.$$

Actually, we will state a more general equality in distribution in the subsequent Theorem 25, concerning not only the underlying population size processes of the forests, but the two *dual* forests themselves (see Fig. II.2). For a forest  $\mathcal{F}$  consisting of  $N$  chronological trees that go extinct before  $T$ , and an  $(N+1)$ -st tree  $\mathcal{T}_{N+1}$  that reaches time  $T$ , truncated up to this time, its *dual* forest (in reverse time) can be defined as follows: its roots are the individuals of  $\mathcal{F}$  extant at  $T$ , birth events become death events and vice versa, and the parental relations are re-drawn from the top of edges to the right (when looking in the original time direction), such that daughters are now to the left of their mothers. This rule is applied to all edges, except for those which are to the right of the last individual that survives up until time  $T$ , that are translated to the left of the first ancestor before re-drawing the parental relations, as it is shown in Fig. II.2. Hence, we postpone the proof of Theorem 22, that will be established later as a consequence of this more general result.

Hereafter, we will consider the contour process of a forest  $\mathcal{F}$  truncated at  $T$  (i.e. each tree is truncated at  $T$ ) for which we use the following notation  $\mathcal{C}^{(T)}(\mathcal{F}) = \mathcal{C}(\mathcal{F}^{(T)})$ . Also define for a càdlàg process  $X$ , a deterministic transformation of its path by

$$\mathcal{K}(X) := \rho \circ \chi(X)$$

which exists when  $g_0(X) < +\infty$  (see preliminaries). This operator has the effect of shifting the last excursion from  $T$  to 0 to the left and then apply the space-time-reversal of the process as defined in the preliminaries.

**Theorem 25.** *Let  $\mathcal{C}^{(T)}(\mathcal{F}_\perp^\gamma)$  be the JCCP of a random forest  $\mathcal{F}_\perp^\gamma$  truncated at  $T$ . Then, after applying the operator  $\mathcal{K}$ , the process obtained has the law of the contour of a forest  $\tilde{\mathcal{F}}_\perp^\gamma$ , also truncated at  $T$ . More precisely,*

$$\mathcal{K}\left(\mathcal{C}^{(T)}(\mathcal{F}_\perp^\gamma)\right) \stackrel{d}{=} \mathcal{C}^{(T)}\left(\tilde{\mathcal{F}}_\perp^\gamma\right)$$

Now the proof of Theorem 22 can be achieved as a quite immediate consequence of this second theorem.

*Proof of Theorem 22.* Recall our definition of the local time  $\Gamma$  of a process  $Y$  with finite variation and finite lifetime. We only need to notice that, for any càdlàg function  $X$  such that  $g_0(X) < +\infty$ ,

$$\left(\Gamma_{T-t}(X \circ k_{g_0(X)}), 0 \leq t \leq T\right) = \left(\Gamma_t \circ \mathcal{K}(X), 0 \leq t \leq T\right).$$

This is true, in particular, when  $X$  is the contour of a truncated forest  $\mathcal{C}^{(T)}(\tilde{\mathcal{F}}_\perp^\gamma)$ , which lifetime is precisely  $g_0(\mathcal{C})$ . Since the local time process of the contour of a tree,  $\Gamma \circ \mathcal{C}$ , is the same as its population size process  $\Xi$ , the first result is established.

The second statement about the subcritical and critical cases is immediate from the first one, and the fact that measures  $\mu_\top$  and  $\mu_\perp$  are the same, as well as  $\Pi$  and  $\tilde{\Pi}$ , when  $m \leq 1$ .

Finally, the third identity is also a consequence of the first one and Lemma 21.  $\square$

To demonstrate Theorem 25 we will consider first two independent sequences, with also independent elements, distributed as excursions of the process  $Y$  starting at 0 or  $T$  and killed upon hitting 0 or  $T$ . Notice that,  $P_0$ -a.s., we have  $\{\tau_0 < \tau_T^+, \tau_0^+ < +\infty\} = \{\tau_0 < \tau_T^+\}$  and  $\{\tau_T^+ < \tau_0, \tau_0^+ < +\infty\} = \{\tau_T^+ < \tau_0\}$ , however we choose to use the left-hand-side events in the definitions below, to emphasize the fact that the process is conditioned on hitting  $(0, +\infty)$ . More precisely define the sequence  $(\epsilon_i, 1 \leq i \leq \tilde{N} + 1)$  as follows

- $(\epsilon_i)$  are i.i.d. with law  $P_0(\cdot | \tau_0 < \tau_T^+, \tau_0^+ < +\infty) \circ k_{\tau_0}^{-1}$  for  $1 \leq i \leq \tilde{N}$ ,
- $\epsilon_{\tilde{N}+1}$  has law  $P_0(\cdot | \tau_T^+ < \tau_0, \tau_0^+ < +\infty) \circ k_{\tau_T^+}^{-1}$
- $\tilde{N}$  is an independent geometric random variable with probability of success  $\tilde{\gamma}$ .

Also define the sequence  $(\tilde{\epsilon}_i, 1 \leq i \leq N + 1)$  as

- $\tilde{\epsilon}_i$  are i.i.d. with the law  $P_T(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_T^+}^{-1}$  for  $1 \leq i \leq N$
- $\tilde{\epsilon}_{N+1}$  has law  $P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$
- $N$  is an independent geometric random variable with probability of success  $\gamma$ .

We denote by  $Z$  the process obtained by the concatenation of these two sequences of excursions in the same order they were defined, that is  $Z := [\epsilon, \tilde{\epsilon}]$  (see Fig. II.3). We will prove first that, after a time change erasing the negative values of these excursions and closing up the gaps, the process thus obtained has the same law as the contour of the forest  $\mathcal{F}_\perp^\gamma$  truncated at  $T$ .

**Claim 1:** We have the following identity in law:

$$Z \circ \alpha \stackrel{d}{=} \mathcal{C}^{(T)}(\mathcal{F}_\perp^\gamma)$$

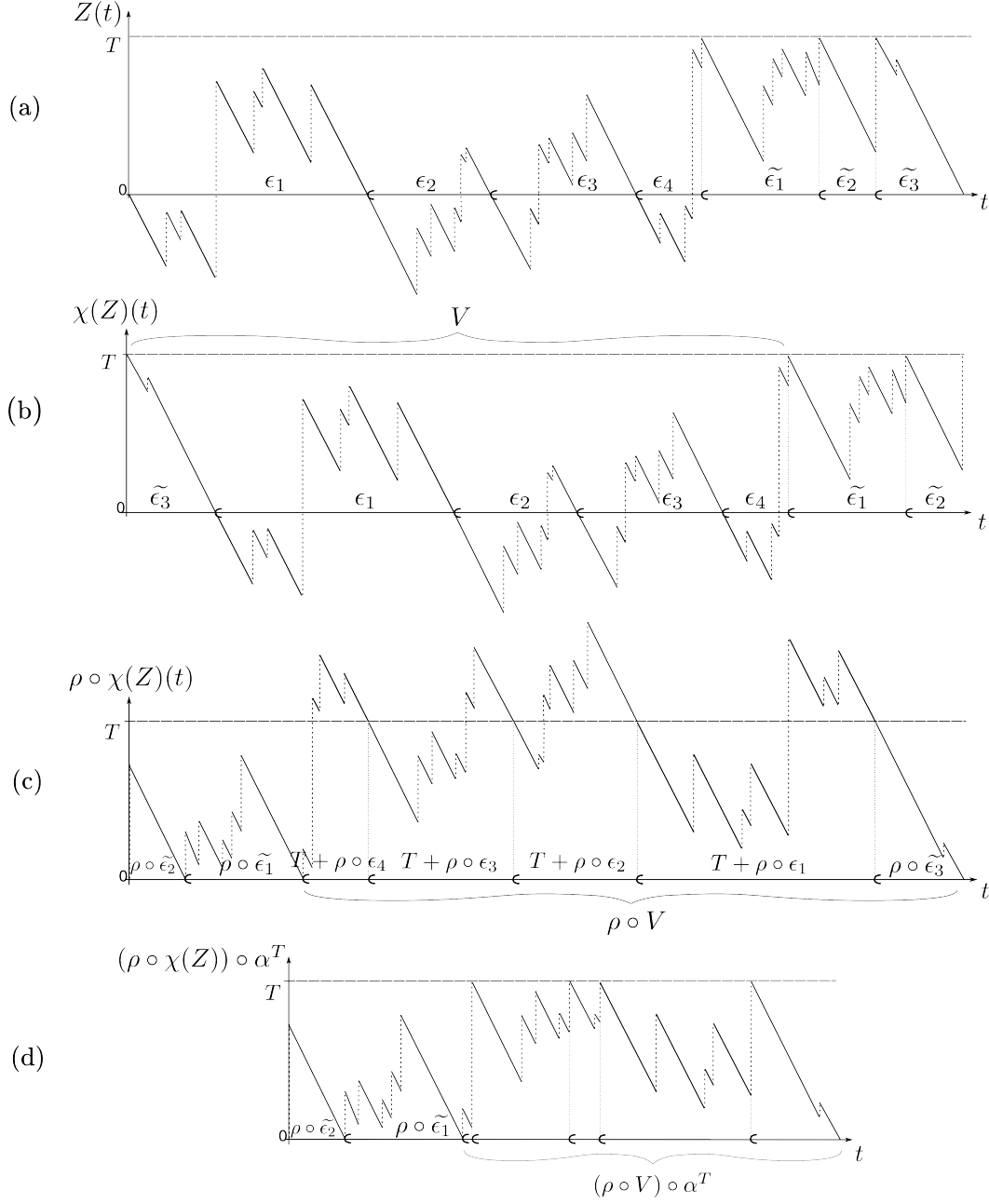


Figure II.3 – (a) An example of the Lévy process  $Z$  with  $\tilde{N} = 3$  excursions before hitting  $T$ , and  $N = 2$  excursions before hitting  $0$  again, and some path transformations of  $Z$ : (b)  $\chi(Z)$  places the last excursion  $\tilde{\epsilon}_3$  at the origin and shifts to the right the rest of the path; (c)  $\rho \circ \chi(Z)$  is the space-time-reversal of the path at (b); and (d)  $(\rho \circ \chi(Z)) \circ \alpha^T$  erases the sub-paths of (c) taking values greater than  $T$  and closes up the gaps, hence the shorter length of the path.

*Proof.* Since a forest  $\mathcal{F}_\perp^{\tilde{\gamma}}$  is a finite sequence of independent trees and, when truncated, its contour process, say  $\mathcal{C}^{(T)}(\mathcal{F}_\perp^{\tilde{\gamma}}) := (\mathcal{C}_t, 0 \leq t \leq \tilde{\ell})$ , is defined as the concatenation of the contour of the trees of this sequence, its law will be characterized by the law of the sequence of the killed paths  $e_i := (\mathcal{C}_{t_i+t}, 0 \leq t < t_{i+1} - t_i)$ , where  $t_0 = 0$ ,  $t_i = \sum_{j=1}^i \ell(\mathcal{T}_{\perp,i})$  for  $1 \leq i \leq N_{\tilde{\gamma}}$ ,  $t_{N_{\tilde{\gamma}}+1} = t_{N_{\tilde{\gamma}}} + \ell(\mathcal{T}_{\perp, N_{\tilde{\gamma}}+1}^{(T)}) = \ell(\mathcal{F}_\perp^{\tilde{\gamma}})$ . Here  $N_{\tilde{\gamma}} + 1$  is the number of trees in  $\mathcal{F}_\perp^{\tilde{\gamma}}$ . These killed paths are then the JCCP's of each of the trees in the forest, that is  $e_i = \mathcal{C}(\mathcal{T}_{\perp,i}^{(T)})$  for every  $1 \leq i \leq N_{\tilde{\gamma}} + 1$ , which are by definition independent, and their number  $N_{\tilde{\gamma}} + 1$  is geometric with parameter  $\tilde{\gamma}$ . The first  $N_{\tilde{\gamma}}$  are identically distributed and conditioned on extinction before  $T$ , and the last one conditioned on surviving up until time  $T$ .

On the other hand, define for the sequence of excursions  $\epsilon$ , the lifetime of its terms,  $\zeta_i = \zeta(\epsilon_i)$  and the first time they hit  $[0, +\infty)$ ,  $\zeta_i^+ = \zeta^+(\epsilon_i)$ . These excursions start at 0 and always visit first  $(-\infty, 0)$ . Under  $\{\tau_0^+ < +\infty\}$ , 0 is recurrent for the reflected process, so we have  $0 < \zeta_i^+ < \zeta_i < +\infty$ ,  $P_0$ -a.s. for every  $i \geq 1$ . If now we apply to each of these excursions the time change  $\alpha$ , removing the non positive values and closing up the gaps, we obtain for  $1 \leq i \leq \tilde{N} + 1$ ,

$$\epsilon_i \circ \alpha = (\epsilon_i(\zeta_i^+ + t), 0 \leq t < \zeta_i - \zeta_i^+)$$

so, for  $1 \leq i \leq \tilde{N}$ ,  $\epsilon_i \circ \alpha$  has the law of  $(Y_{\tau_0^++t}, 0 \leq t \leq \tau_0 - \tau_0^+)$  under  $P_0(\cdot | \tau_0 < \tau_T^+, \tau_0^+ < +\infty)$ . Thanks to Lemma 19 (ii), this is the same as the law of the contour of each of the first  $N_{\tilde{\gamma}}$  trees on the forest  $\mathcal{F}_\perp^{\tilde{\gamma}}$ .

The last excursion,  $\epsilon_{\tilde{N}+1} \circ \alpha$  has the law of  $(Y_{\tau_0^++t}, 0 \leq t \leq \tau_T^+ - \tau_0^+)$  under  $P_0(\cdot | \tau_T^+ < \tau_0, \tau_0^+ < +\infty)$ , that is, an excursion of  $Y$  starting from an initial value distributed according to  $\mu_\perp$ , conditioned on hitting  $T$  before 0.

If we look now at the second sequence  $\tilde{\epsilon}$ , we notice that, since  $\gamma = P_T(\tau_0 < \tau_T^+)$ , it is distributed as a sequence of i.i.d. excursions with law  $P_T \circ k_{\tau_0 \wedge \tau_T^+}(\cdot)$ , stopped at the first one hitting 0 before  $(T, +\infty)$ . Thus, Theorem 4.3 in [Lam10] guarantees that the concatenation  $[\epsilon_{\tilde{N}+1} \circ \alpha, \tilde{\epsilon}]$  has the law of the contour of the last tree in the forest  $\mathcal{F}_\perp^{\tilde{\gamma}}$  truncated at  $T$ , say  $\mathcal{T}_{\perp, N_{\tilde{\gamma}}+1}^{(T)}$ .

Finally, since  $\tilde{N}$  has the same distribution as  $N_{\tilde{\gamma}}$ , we have,

$$[e_i, 1 \leq i \leq N_{\tilde{\gamma}} + 1] \stackrel{d}{=} [\epsilon_i \circ \alpha, 1 \leq i \leq \tilde{N} + 1, \tilde{\epsilon}_i, 1 \leq i \leq N + 1]$$

Now the right-hand-side equals  $[\epsilon, \tilde{\epsilon}] \circ \alpha$ , because the time change  $\alpha$  is the inverse of an additive functional, so it commutes with the concatenation, and the elements of the sequence  $\tilde{\epsilon}$  do not take negative values, so the time change  $\alpha$  has no effect on them. This ends the proof of the claim.  $\square$

Now we will look at the process  $Z$  after relocating the last excursion of the second sequence,  $\tilde{\epsilon}_{N+1}$  to the beginning and shifting the rest of the path to the right. More precisely, consider the process

$$V = [\tilde{\epsilon}_{N+1}, \epsilon_1, \dots, \epsilon_{\tilde{N}+1}],$$

and consider also the space-time-reversed process  $\rho \circ V$  (see Fig. II.3). It is not hard to see that

$$\rho \circ V = [T + \rho \circ \epsilon_{\tilde{N}+1}, T + \rho \circ \epsilon_{\tilde{N}}, \dots, T + \rho \circ \epsilon_1, \rho \circ \tilde{\epsilon}_{N+1}],$$

since  $V(0) = \epsilon_{\tilde{N}+1}(0) = T$  and all the other excursions in  $V$  take the value 0 at 0. Then we have the following result on the law of this process, reflected at  $T$ .

**Claim 2:** We have the following identity in law:

$$(\rho \circ V) \circ \alpha^T \stackrel{d}{=} \mathcal{C}^{(T)}(\tilde{\mathcal{T}}_+')$$

where  $\tilde{\mathcal{T}}_+' is a splitting tree conditioned on surviving up until time  $T$ .$

For the proof of Claim 2 we will need the following two results that are proved in the Appendix.

**Lemma 26.** *The probability measure  $P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$  is invariant under space-time-reversal.*

**Lemma 27.** *For any  $a > 0$ ,  $x \in (0, a)$  and  $\Lambda \in \mathcal{F}_{\tau_a}$  the following identity holds*

$$\tilde{P}_x(\Lambda, \tau_a^+ < \tau_0) = P_x(\Lambda, \tau_a^+ < \tau_0, \tau_a < +\infty) e^{-\eta(a-x)}$$

In particular,  $\tilde{P}_x(\tau_a^+ < \tau_0) = P_x(\tau_a^+ < \tau_0, \tau_a < +\infty) e^{-\eta(a-x)}$ , hence

$$\tilde{P}_x(\Lambda | \tau_a^+ < \tau_0) = P_x(\Lambda | \tau_a^+ < \tau_0, \tau_a < +\infty)$$

*Proof of Claim 2.* We can deduce from Proposition 16 the following observations about the laws of the space-time-reversed excursions:

- (1) conditional on  $\epsilon_{\tilde{N}+1}(\zeta-) = T - u$ , the reversed excursion  $T + \rho \circ \epsilon_{\tilde{N}+1}$  has law  $P_u(\cdot | \tau_T^+ < \tau_0, \tau_T < +\infty) \circ k_{\tau_T}^{-1}$
- (2) for  $1 \leq i \leq \tilde{N}$ , the excursions  $T + \rho \circ \epsilon_i$  have law  $P_T(\cdot | \tau_T^+ < \tau_0, \tau_T < +\infty) \circ k_{\tau_T}^{-1}$

and thanks to Lemma 26, we also have

- (3)  $\rho \circ \tilde{\epsilon}_{N+1} \stackrel{d}{=} \tilde{\epsilon}_{N+1}$ , with common law  $P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$ .

Now we would like to express the laws of the excursions in (1), (2) and (3) in terms of the probability measure  $\tilde{P}$ , the probability of  $Y$  conditioned on not drifting to  $+\infty$  (see preliminaries). For (3) we easily have, from (II.10), that

$$P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1} = P_T(\cdot | \tau_0 < \tau_T^+, \tau_0 < +\infty) \circ k_{\tau_0}^{-1} = \tilde{P}_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$$

The result in Lemma 27 entails in particular that excursions in (1) and (2), killed at the time they hit  $(T, +\infty)$ , have distribution  $\tilde{P}_u(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_T^+}^{-1}$  and  $\tilde{P}_T(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_T^+}^{-1}$  respectively. Besides, notice that to kill these excursions at  $\tau_T^+$  is the same as applying the time change  $\alpha^T$ , i.e. removing the part of the path taking values in  $(T, +\infty)$ .

Then, conditional on  $V(\zeta-) = T - u$ , the reversed process after the time change  $\alpha^T$ , that is  $(\rho \circ V) \circ \alpha^T$  consists in a sequence of independent excursions distributed as the Lévy process  $\tilde{Y}$  killed at  $\tau_0 \wedge \tau_T^+$ , all starting at  $T$ , but the first one, starting at  $u$ . There are  $\tilde{N}$  excursions from  $T$ , conditioned on hitting  $T$  before 0 and a last excursion from  $T$  conditioned on hitting 0 before returning to  $T$ , and killed upon hitting 0. Observe that  $\tilde{N}$  has geometric distribution with parameter  $\tilde{\gamma}$ , which is exactly the probability that an excursion of  $\tilde{Y}$ , starting at  $T$ , exits the interval  $(0, T)$  from the bottom, that is  $\tilde{P}_T(\tau_0 < \tau_T^+)$  (Equation (II.3)). This implies that excursions in (2) and (3) after the time change  $\alpha^T$ , form a sequence of i.i.d. excursions of  $\tilde{Y}$  starting at  $T$ , killed at  $\tau_0 \wedge \tau_T^+$ , ending at the first one that hits 0 before  $[T, +\infty)$  (Lemma 18).



The fact that the time change  $\alpha^T$  commutes with the concatenation, allows to conclude that  $(\rho \circ V) \circ \alpha^T$ , conditional on  $V(\zeta-) = T - u$ , has the law of the process  $\tilde{Y}$  starting at  $u$ , conditioned on hitting  $T$  before 0, reflected below  $T$  and killed upon hitting 0.

From the definition of the sequence  $\epsilon$  we deduce that  $V(\zeta-) = \epsilon_{\tilde{N}+1}(\zeta-)$  has the law of the undershoot of  $Y$  at  $T$  of an excursion starting at  $T$  and conditioned on hitting 0 before  $(T, +\infty)$ . As usual, the strong Markov property and the stationary increments of the Lévy process entail that this excursion is invariant under translation of the space, hence this undershoot has the distribution  $P_0(-Y_{\tau_0^+} \in \cdot | \tau_{-T} < \tau_0^+ < +\infty)$ . This implies that the law of  $(\rho \circ V) \circ \alpha^T$  is

$$\int_0^T P_0(-Y_{\tau_0^+} \in du | \tau_{-T} < \tau_0^+ < +\infty) \tilde{P}_u(Y^{(T)} \in \cdot | \tau_T^+ < \tau_0) \circ k_{\tau_0}^{-1}, \quad (\text{II.15})$$

and we will show this is the same as

$$\tilde{P}_T(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_0}^{-1} = \frac{\int_0^T \mu_T(du) \tilde{P}_u(Y^{(T)} \in \cdot, \tau_T^+ < \tau_0) \circ k_{\tau_0}^{-1}}{\int_0^T \mu_T(du) \tilde{P}_u(\tau_T^+ < \tau_0)}. \quad (\text{II.16})$$

The strong Markov property and Proposition 16 imply that

$$P_0(-Y_{\tau_0^+} \in du | \tau_{-T} < \tau_0^+ < +\infty) = \mu_T(du) \frac{P_u(\tau_T^+ < \tau_0 | \tau_0 < +\infty)}{P_0(\tau_{-T} < \tau_0^+ | \tau_0^+ < +\infty)}.$$

Then by using this identity and (II.10), we have that (II.15) equals

$$\begin{aligned} & \int_0^T \mu_T(du) \frac{P_u(\tau_T^+ < \tau_0 | \tau_0 < +\infty)}{P_0(\tau_{-T} < \tau_0^+ | \tau_0^+ < +\infty)} \frac{\tilde{P}_0(Y^{(T)} \in \cdot, \tau_T^+ < \tau_0) \circ k_{\tau_0}^{-1}}{\tilde{P}_u(\tau_T^+ < \tau_0)} \\ &= \frac{\int_0^T \mu_T(du) \tilde{P}_0(Y^{(T)} \in \cdot, \tau_T^+ < \tau_0) \circ k_{\tau_0}^{-1}}{P_0(\tau_{-T} < \tau_0^+ | \tau_0^+ < +\infty)}. \end{aligned}$$

Finally, the numerator in the last term is the same as the one in the right-hand-side of (II.16), due again to (II.10) and the dualities from Proposition 16.

Then, as announced,  $(\rho \circ V) \circ \alpha^T$  has the law (II.16), which is, thanks to Lemma 19, the contour of a splitting tree of lifespan measure  $\tilde{\Pi}$  and ancestor distributed as  $\mu_T$ , conditioned on surviving up until time  $T$  and truncated at  $T$ , say  $\tilde{\mathcal{T}}_T^T$ .  $\square$

It remains to understand the effect on the i.i.d. excursions  $(\tilde{\epsilon}_i, 1 \leq i \leq N)$ , of the reversal operator  $\rho$ , which is given in the following statement.

**Claim 3:** For  $1 \leq i \leq N$ ,

$$\rho \circ \tilde{\epsilon}_i \stackrel{d}{=} \mathcal{C}(\tilde{\mathcal{T}}_{T,i})$$

where  $\tilde{\mathcal{T}}_{T,i}$  is a sequence of i.i.d. splitting trees conditioned on dying out before  $T$ .

*Proof.* We know from Proposition 16 that, conditional on  $\tilde{\epsilon}_i(\zeta-) = T - u$ , the reversed excursions  $\rho \circ \tilde{\epsilon}_i$ , has the law of  $\tilde{Y}$  starting from  $u$ , conditioned on hitting 0 before  $T$  and killed upon hitting 0, that is  $\tilde{P}_u(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$ .

The same reasoning as for Claim 2 yields that  $T - \tilde{\epsilon}_i(\zeta -)$  has distribution

$$P_0 \left( -Y_{\tau_0^+ -} \in \cdot \mid \tau_0^+ < \tau_{-T} < +\infty \right),$$

for every  $1 \leq i \leq N$ , and then, each of the reversed excursions  $\rho \circ \tilde{\epsilon}_i$  has the law  $\tilde{P}_\top(\cdot \mid \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$ . This is the same as the contour process of splitting trees, say  $\tilde{\mathcal{T}}_{\top, i}$ , all i.i.d. for  $1 \leq i \leq N$  and conditioned on dying out before  $T$ .  $\square$

*Proof of Theorem 25.* We have now all the elements to complete the proof of this result. Notice the transformations we have done to the trajectories of the process  $Z = [\epsilon, \tilde{\epsilon}]$  in terms of its excursions, can also be expressed in terms of the time-changes  $\alpha, \alpha^T$  and the path transformations  $\rho, \chi$  (Fig. II.3), as follows

$$\chi(Z) = [V, \tilde{\epsilon}_1, \dots, \tilde{\epsilon}_N]$$

and stressing that all these paths start by taking the value  $T$ , we have

$$\rho \circ \chi(Z) = [\rho \circ \tilde{\epsilon}_N, \dots, \rho \circ \tilde{\epsilon}_1, \rho \circ V].$$

On the other hand, after Claim 1, we have  $\mathcal{C}^{(T)}(\mathcal{F}_\perp^{\tilde{\gamma}}) = Z \circ \alpha$ , so

$$\begin{aligned} \mathcal{K}(\mathcal{C}^{(T)}(\mathcal{F}_\perp^{\tilde{\gamma}})) &\stackrel{d}{=} \mathcal{K} \circ Z \circ \alpha = (\rho \circ \chi) \circ (Z \circ \alpha) = \\ &= [\rho \circ \tilde{\epsilon}_N, \dots, \rho \circ \tilde{\epsilon}_1, (\rho \circ V) \circ \alpha^T] \end{aligned}$$

We have proved that the right-hand term in the last equation has the law of the contour of a sequence of independent splitting trees  $\tilde{\mathcal{T}}_{\top, i}$ , which are i.i.d. for  $1 \leq i \leq N$ , conditioned on dying out before  $T$ , and a last tree  $\tilde{\mathcal{T}}_\top$  conditioned on surviving up until time  $T$ . Since  $N$  is geometric with parameter  $\gamma$ , this is the same as the process  $\mathcal{C}^{(T)}(\tilde{\mathcal{F}}_\top^\gamma)$ , which concludes the proof.  $\square$

## 4 Epidemiology

In general, in the context of epidemiology, phylogenetic trees are considered to be estimated from genetic sequences obtained at a single time point, or sampled sequentially through time since the beginning of the epidemic. There exists several methods allowing this estimation, which are not addressed here. We assume the estimated *reconstructed trees* (i.e. the information about non sampled hosts is erased from the original process) are the transmission trees from sampled individuals and no uncertainty on the branch lengths is considered. This hypothesis makes sense when the epidemiological and evolutionary timescales can be supposed to be similar [VKB13]. These reconstructed phylogenies can provide information on the underlying population dynamic process [Tho75, NMH94, DPR03] and there is an increasing amount of work on this relatively new field of phylodynamics.

Most of phylodynamic models are based on Kingman's coalescent, but its poor realism in the context of epidemics (rapid growth, rapid fluctuations, dense sampling) has motivated other authors to use birth-death or SIR processes for the dynamics of the epidemics [VPW<sup>+</sup>09, RRK11, SKvW<sup>+</sup>12, LT13, LAS14]. A common feature for most of these works is that they use likelihood-based methods that intend to infer the model parameters on the basis of available data, via maximum likelihood estimation (MLE), or in a Bayesian framework.

Usually, not only the reconstructed phylogenies described above are available, but also incidence time-series, that is, the number of new cases registered through time (typically daily, weekly or monthly). This information may come from hospital records, surveillance programs (local or national), and is not necessarily collected at regular intervals. As we mentioned before, here we are interested in the scenario where both types of data are available: phylogenetic trees (reconstructed from pathogen sequences) and incidence time series.

From the probabilistic point of view, we are interested in the distribution of the size process of the host population, denoted by  $I := (I_t, 0 \leq t \leq T)$ , conditional on the *reconstructed transmission tree* from infected individuals at time  $T$ . More precisely, we want to characterize the law of  $I$ , conditional on  $\sigma = (\sigma_i, 1 \leq i \leq n)$  to be the coalescence times in the reconstructed tree of transmission from extant hosts at  $T$ .

We suppose that the host population has the structure of a forest  $\mathcal{F}_\perp^{\tilde{\gamma}}$ , so we consider there are a geometric number  $\tilde{N}$  (with parameter  $\tilde{\gamma}$ ) of infected individuals at time 0, for which the corresponding transmission tree dies out before  $T$ , and a last one which is at the origin of all the present-time infectives. Let  $(H_i)_{1 \leq i \leq N}$  be the coalescence times between these infected individuals at  $T$ , where as before,  $N$  is an independent geometric r.v. with parameter  $\gamma$ . Here we are interested in characterizing the distribution of  $\Xi(\mathcal{F}_\perp^{\tilde{\gamma}})$ , the total population size process of infected individuals on  $[0, T]$ , conditional on  $H_i = \sigma_i, 1 \leq i \leq N$ .

Since in our model, all infected individuals at  $T$  belong to the last tree in the forest, which we know from the definition, is conditioned on surviving up until time  $T$ , a result from [Lam10] and the pathwise decomposition of the contour process of a forest from the previous section, tells us that these coalescence times are precisely the depths of the excursions away from  $T$  of  $\mathcal{C}(\mathcal{F}_\perp^{\tilde{\gamma}})$ , that is,  $H_i \stackrel{d}{=} \inf \tilde{\epsilon}_i$  for  $1 \leq i \leq N$ . We recall that  $\tilde{\epsilon}_i$  are i.i.d. with law  $P_T(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_T^+}^{-1}$ .

Then, conditioning the population size process on the coalescence times is the same as these excursions of the Lévy process  $Y$  conditioned on their infimum, which becomes, after the corresponding space-time-reversal, their supremum, since,  $P_T(\cdot | \tau_T^+ < \tau_0) \circ k_{\tau_T^+}^{-1}$ -a.s., we have  $\sup_{[0, \tau_T^+)} \rho \circ \tilde{\epsilon} = \sup_{t \in [0, \tau_T^+)} T - \tilde{\epsilon}((\tau_T^+ - t) -) = \inf_{[0, \tau_T^+)} \tilde{\epsilon}$ . Besides, thanks to Theorem 25, when we reverse the time, these excursions themselves are distributed as the contour of independent subcritical (with measure  $\tilde{\Pi}$ ) splitting trees conditioned on hitting 0 before  $T$ , starting from a value distributed as  $\mu_\top$  in  $[0, T]$ . Therefore, conditioning these excursions on their supremum is the same as conditioning the corresponding trees on their height, that is, conditioning on the time of extinction of each tree with lifespan measure  $\tilde{\Pi}$  to equal the corresponding time of coalescence  $\sigma_i$ . We notice that conditioning on an event as  $\{T_{\text{Ext}} = s\}$ , is possible here since the time of extinction of a tree  $\tilde{\mathcal{T}}_\top$  always has a density (because  $\mu_\top$  has a density).

We also know from the proof of Theorem 25, that the total population process of the forest, also takes into account the width process of the excursion  $V \circ \alpha$ , which is independent of  $\tilde{\epsilon}_i, 1 \leq i \leq N$ , and when reversed, has the law of the contour of a splitting tree truncated up to  $T$ , and conditioned on surviving up until time  $T$ , that is  $(\rho \circ V) \circ \alpha^T \stackrel{d}{=} \mathcal{C}^{(T)}(\tilde{\mathcal{T}}_\top^T)$  as stated in Claim 2.

More precisely we have the following result,

**Theorem 28.** *Let  $\mathcal{F}_\perp^{\tilde{\gamma}}$  and  $(H_i)_{i \geq 1}$  as defined before. Then, under  $P(\cdot | H_i = \sigma_i, 1 \leq i \leq N)$ , the population size process backward in time,  $(\xi_{T-t}(\mathcal{F}_\perp^{\tilde{\gamma}}), 0 \leq t \leq T)$ , is the sum of the width processes of  $N + 1$  independent splitting trees  $(\tilde{\mathcal{T}}_{\top, i})_{1 \leq i \leq N+1}$ , where,*

- for  $1 \leq i \leq N$ ,  $\tilde{\mathcal{T}}_{\top, i}$  are subcritical splitting trees with lifespan measure  $\tilde{\Pi}$ , starting

with an ancestor with lifespan distributed as  $\mu_{\top}$ , and conditioned on its time of extinction to be  $\sigma_i$ , that is with law  $\tilde{\mathbb{P}}_{\top}(\cdot | T_{Ext} = \sigma_i)$

- the last tree  $\tilde{\mathcal{T}}_{\top, N+1}$  is a subcritical splitting tree with lifespan measure  $\tilde{\Pi}$ , starting with an ancestor with lifespan distributed as  $\mu_{\top}$ , and conditioned on surviving up until time  $T$ , that is with law  $\tilde{\mathbb{P}}_{\top}(\cdot | T_{Ext} > T)$ .

## Appendix

### A-II.1 Remaining proofs

In this section we proceed to the proofs of Lemma 19, Lemma 20, Lemma 26 and Lemma 27.

*Proof of Lemma 19.* These statements are quite immediate consequences of Theorem 4.3 from [Lam10], the strong Markov property and stationarity of the increments of  $Y$ . We will expand the arguments for (i) and the other statements can be proved similarly. We know from 17, that conditional on  $\text{Ext}$  and  $\zeta(\emptyset) = x$ , the contour  $\mathcal{C} = \mathcal{C}(\mathcal{T}_\perp)$  has the law of  $(Y_t, 0 \leq t \leq \tau_0)$  under  $P_x(\cdot | \tau_0 < +\infty)$ . Hence, it follows from the definition of  $\mathcal{T}_\perp$  and by conditioning on its ancestor lifespan that,

$$\begin{aligned} \mathbb{P}_\perp(\mathcal{C} \in \cdot | \text{Ext}) &= \int_0^\infty \mathbb{P}(\mathcal{C} \in \cdot | \text{Ext}, \zeta(\emptyset) = x) \mathbb{P}_\perp(\zeta(\emptyset) \in dx) \\ &= \int_0^\infty P_x(Y \circ k_{\tau_0} \in \cdot | \tau_0 < +\infty) P_0(Y_{\tau_0^+} \in dx | \tau_0^+ < +\infty) \\ &= \int_0^\infty P_0(Y \circ \theta_{\tau_0^+} \circ k_{\tau_0} \in \cdot | Y_{\tau_0^+} = x, \tau_0 < +\infty) P_0(Y_{\tau_0^+} \in dx | \tau_0^+ < +\infty) \\ &= P_0(Y \circ \theta_{\tau_0^+} \circ k_{\tau_0} \in \cdot | \tau_0 < +\infty), \end{aligned}$$

which ends the proof of (i).  $\square$

*Proof of Lemma 20.* First we want to prove that  $\tilde{\gamma} = \mathbb{P}_\perp(\xi_T \neq 0)$ . To do that, we can express this probability in terms of the contour process, thanks to Lemma 19 (ii), we have

$$\mathbb{P}_\perp(\xi_T = 0) = P_0(\tau_0 < \tau_T^+ | \tau_0^+ < +\infty)$$

According to (II.7), in the supercritical case, we have  $P_0(\tau_0^+ < +\infty) = 1$ . The probability that the process  $Y$ , starting from 0, returns to 0 before it reaches the interval  $[T, +\infty)$ , can be computed by integrating with respect to the measure of the overshoot at 0, of an excursion starting from 0. Then, by applying the strong Markov property, equations (II.3) and (II.12), we prove the first identity,

$$\begin{aligned} P_0(\tau_0 < \tau_T^+) &= \int_0^T P_v(\tau_0 < \tau_T^+) P_0(Y_{\tau_0^+} \in dv) \\ &= \int_0^T P_v(\tau_0 < \tau_T^+) e^{\eta v} \widetilde{\Pi}(v) dv = \int_0^T \frac{W(T-v)}{W(T)} e^{\eta v} \widetilde{\Pi}(v) dv \\ &= \frac{1}{W(T)e^{-\eta T}} \int_0^T W(T-v) e^{-\eta(T-v)} \widetilde{\Pi}(v) dv = \frac{1}{\widetilde{W}(T)} \int_0^T \widetilde{W}(T-v) \widetilde{\Pi}(v) dv \\ &= 1 - \frac{1}{\widetilde{W}(T)} \end{aligned}$$

The second statement is that  $\gamma = \tilde{\mathbb{P}}_T (\xi_T \neq 0)$ , which follows from,

$$\begin{aligned} \tilde{P}_T (\tau_0 < \tau_T^+) &= \int_0^T \tilde{P}_u (\tau_0 < \tau_T^+) P_0 (-Y_{\tau_0^+ -} \in du) \\ &= \int_0^T \frac{\tilde{W}(T-u)}{\tilde{W}(T)} e^{-\eta u} \bar{\Pi}(u) du = \int_0^T \frac{e^{\eta(T-u)} \tilde{W}(T-u)}{e^{\eta T} \tilde{W}(T)} \bar{\Pi}(u) du \\ &= \frac{1}{W(T)} \int_0^T W(T-u) \bar{\Pi}(u) du = 1 - \frac{1}{W(T)} \end{aligned}$$

The second statement can also be established thanks to Proposition 16 (ii). This duality property, together with Equations (II.3) and (II.10), imply that

$$\tilde{\mathbb{P}}_T (\xi_T = 0) = \tilde{P}_T (\tau_0 < \tau_T^+) = P_T (\tau_T^+ < \tau_0) = 1 - \frac{1}{W(T)}$$

which is the expected result.  $\square$

*Proof of Lemma 26.* Let  $\varepsilon := Y \circ k_{\tau_0} = (Y_t, 0 \leq t < \tau_0)$ , we want to prove that  $\varepsilon$  and  $\rho \circ \varepsilon$  have the same law under the probability measure  $P_T (\cdot | \tau_0 < \tau_T^+)$ .

Define

$$\varsigma_T := \sup\{t \in [0, \tau_0) : Y_t = T\}.$$

First, since the process makes no negative jumps, we have the following identities  $P_0$ -a.s.

$$\{\varsigma_T < +\infty\} = \{\exists t < \tau_0 : Y_t = T\} \cap \{\tau_0 < +\infty\} = \{\tau_T < \tau_0 < +\infty\} = \{\tau_T^+ < \tau_0 < +\infty\}.$$

Consider  $\varepsilon$  under  $P_0(\cdot | \tau_T^+ < \tau_0 < +\infty)$ . Under this probability,  $\tau_T < +\infty$  a.s., so we can apply the strong Markov property at  $\tau_T$  and we have that

$$P_0 (\varepsilon \circ \theta_{\tau_T} \in \cdot | \tau_T^+ < \tau_0 < +\infty) = P_T (\cdot | \tau_0 < +\infty) \circ k_{\tau_0}^{-1}.$$

Now consider the excursions away from  $T$  of the process  $Y$  under  $P_T(\cdot | \tau_0 < +\infty) \circ k_{\tau_0}^{-1}$ , which consist in a sequence of i.i.d. subpaths distributed as  $P_T(\cdot | \tau_T^+ < +\infty) \circ k_{\tau_T}^{-1}$ , stopped at the first one that hits 0 before  $T$  and killed at  $\tau_0$  (see the introduction of the chapter for a description of the excursion process away from a point in the irregular case). This “last” excursion in the sequence has distribution  $P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$ . Thanks to the above identity, this last excursion can also be defined as the excursion shifted to the last passage at  $T$ , that is  $\varepsilon \circ \theta_{\varsigma_T}$  under  $P_0(\cdot | \tau_T^+ < \tau_0 < +\infty)$ . Hence we have that

$$P_0(\varepsilon \circ \theta_{\varsigma_T} \in \cdot | \tau_T^+ < \tau_0 < +\infty) = P_T (\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}. \quad (\text{A-II.1})$$

On the other hand, we know from Proposition 16 that the probability measure of  $\varepsilon$  starting at 0 and conditional on  $\{\tau_0 < +\infty\}$  is invariant under space-time-reversal, meaning that for every bounded measurable function  $F$  we have

$$E_0 [F(\varepsilon) | \tau_0 < +\infty] = E_0 [F(\rho \circ \varepsilon) | \tau_0 < +\infty]$$

In particular, for any  $f$  also bounded and measurable, take

$$F(\varepsilon) = f(\varepsilon \circ \theta_{\varsigma_T(\varepsilon)}) 1_{\{\sup \varepsilon \geq T\}}.$$

If we now apply this function to the reversed excursion, it is not hard to see that,  $P_0(\cdot | \tau_0 < +\infty)$  a.s. we have,  $\{\sup_{[0, \tau_0]} \rho \circ \varepsilon \geq T\} = \{\inf_{[0, \tau_0]} \varepsilon \leq -T\}$  and  $\varsigma_T(\rho \circ \varepsilon) = \tau_0(\varepsilon) - \tau_{-T}(\varepsilon)$ . From the definition of space-time-reversal and shifting operators it also follows that

$$(\rho \circ \varepsilon) \circ \theta_{\varsigma_T(\rho \circ \varepsilon)} = (-\varepsilon((\tau_0 - t)-)) \circ \theta_{\tau_0 - \tau_{-T}}(\varepsilon) = -\varepsilon((\tau_{-T} - t)-) = \rho \circ (\varepsilon \circ k_{\tau_{-T}}).$$

Hence for any bounded measurable  $f$  it holds that

$$E_0 \left[ f(\varepsilon \circ \theta_{\varsigma_T}) 1_{\{\sup \varepsilon \geq T\}} \middle| \tau_0 < +\infty \right] = E_0 \left[ f(\rho \circ (\varepsilon \circ k_{\tau_{-T}})) 1_{\{\inf \varepsilon \leq -T\}} \middle| \tau_0 < +\infty \right],$$

and in particular for  $f \equiv 1$  we have

$$P_0(\sup \varepsilon \geq T | \tau_0 < +\infty) = P_0(\inf \varepsilon \leq -T | \tau_0 < +\infty).$$

Combining these two equations and using that  $\{\sup \varepsilon \geq T\} = \{\tau_T^+ < \tau_0\}$  and  $\{\inf \varepsilon \leq -T\} = \{\tau_{-T} < \tau_0^+\}$  a.s. under  $P_0(\cdot | \tau_0 < +\infty)$ , we have

$$P_0(\varepsilon \circ \theta_{\varsigma_T} \in \cdot | \tau_T^+ < \tau_0) = P_0(\rho \circ (\varepsilon \circ k_{\tau_{-T}}) \in \cdot | \tau_{-T} < \tau_0^+). \quad (\text{A-II.2})$$

According to (A-II.1), the left-hand-side in (A-II.2) equals  $P_T(\cdot | \tau_0 < \tau_T^+) \circ k_{\tau_0}^{-1}$ . Finally, the conclusion comes from the following consequences of the strong Markov property. First, the excursion  $\varepsilon \circ k_{\tau_{-T}}$  has the same law under  $P_0(\cdot | \tau_{-T} < \tau_0, \tau_0 < +\infty)$  as under  $P_0(\cdot | \tau_{-T} < \tau_0^+)$ . And  $\rho \circ (\varepsilon \circ k_{\tau_{-T}})$  under  $P_0(\cdot | \tau_{-T} < \tau_0^+)$  has the same law as  $\rho \circ (\varepsilon \circ k_{\tau_0})$  under  $P_T(\cdot | \tau_0 < \tau_T^+)$ .  $\square$

*Proof of Lemma 27.* We need to prove that for any  $a > 0$ ,  $x \in [0, a]$  and  $\Lambda \in \mathcal{F}_{\tau_a}$  the following identity holds

$$\tilde{P}_x(\Lambda, \tau_a^+ < \tau_0) = P_x(\Lambda, \tau_a^+ < \tau_0, \tau_a < +\infty) e^{-\eta(a-x)}.$$

Recall that  $\eta = 0 \iff m \leq 1$ , in which case  $\tilde{P} = P$  and the process drifts to  $-\infty$  so this identity is trivial. Hence we will suppose from now  $\eta > 0$ . Since the process makes only positive jumps, we have  $\{\tau_a < \tau_0\} = \{\tau_a^+ < \tau_0\}$ ,  $\tilde{P}_x$  a.s., so we can look instead at  $\tilde{P}_x(\Lambda, \tau_a < \tau_0)$ .

Indeed, we can prove the following more general identity, for any  $\Theta \in \mathcal{F}_{\tau_a}$ ,

$$\tilde{P}_x(\Theta, \tau_a < \infty) = E_x \left[ \mathbb{1}_{\{\Theta, \tau_a < \infty\}} e^{-\eta(a-x)} \right]. \quad (\text{A-II.3})$$

Let  $t > 0$ . First notice that  $\Theta \cap \{\tau_a < t\} \in \mathcal{F}_{\tau_a}$ . This, together with the definition of  $\tilde{P}$ , the strong Markov property and the martingale property of  $e^{-\eta(Y_t - x)}$ , allow to state that

$$\tilde{P}_x(\Theta, \tau_a < t) = E_x \left[ \mathbb{1}_{\{\Theta, \tau_a < t\}} \frac{e^{-\eta Y_t}}{e^{-\eta x}} \right] = E_x \left[ \mathbb{1}_{\{\Theta, \tau_a < t\}} E_x \left[ \frac{e^{-\eta Y_t}}{e^{-\eta x}} \middle| \mathcal{F}_{\tau_a} \right] \right] = E_x \left[ \mathbb{1}_{\{\Theta, \tau_a < t\}} \frac{e^{-\eta a}}{e^{-\eta x}} \right].$$

Now, with the help of the Monotone Convergence Theorem, we can take the limit when  $t \uparrow \infty$ , which leads to (A-II.3). Finally, by choosing  $\Theta = \Lambda \cap \{\tau_a < \tau_0\}$  we obtain the desired result.  $\square$

## Chapter III

# Branching processes seen from their extinction time via path decompositions of reflected Lévy processes

*This chapter is based in the article [DFL16].*

### 1 Introduction

There exist several links between spectrally positive Lévy processes (SPLP) and branching processes that have been known and exploited for a few decades already. We can find their origin in the seminal works of Lamperti [Lam67], where it is shown that there exists a one-to-one correspondence, via a random time change (the so-called Lamperti transformation) between continuous state branching processes (CSBP) and possibly killed SPLP [CLUB09, LSZ13]. Lévy processes also provide a suitable way of coding the genealogical structure of branching processes, through exploration or contour processes [LGLJ98, Lam10, LB16]. Additionally, Ray-Knight type theorems link the local time processes of SPLP to the width processes of branching populations [PW11, LB16].

Here we consider a Lévy process with no negative jumps  $X = (X_t, t \geq 0)$ , with probability distribution  $P_x = P(\cdot | X_0 = x)$  and with Laplace exponent  $\psi$ , defined by

$$E_0[\exp(-\lambda X_t)] = \exp(t\psi(\lambda)).$$

Thanks to the Lévy-Kinchin formula,  $\psi$  can be expressed as follows for any  $\lambda \geq 0$ ,

$$\psi(\lambda) = \alpha\lambda + \beta\lambda^2 + \int_0^\infty (e^{-\lambda r} - 1 + \lambda r \mathbb{1}_{r < 1}) \Pi(dr), \quad (\text{III.1})$$

where  $\alpha \in \mathbb{R}$ ,  $\beta \geq 0$  is called the *Gaussian coefficient* and  $\Pi$  is a  $\sigma$ -finite measure on  $(0, \infty)$ , called the *Lévy measure*, satisfying  $\int_{(0, \infty)} (r^2 \wedge 1) \Pi(dr) < \infty$ . The paths of  $X$  have finite variation a.s. if and only if  $\beta = 0$  and  $\int_{(0, 1]} r \Pi(dr) < \infty$ . Otherwise they have infinite variation a.s. We assume that  $X$  is (sub)critical, meaning it does not drift to  $+\infty$ , which is equivalent to  $\psi'(0+) \geq 0$ .

We consider the process reflected at its infimum  $X - I$ , where for each  $t \geq 0$  we denote  $I_t := \inf_{[0, t]} X$ . A result due to Rogozin [Rog66] states that for SPLP processes 0 is always regular for  $(-\infty, 0)$ , so is also regular for itself for the reflected process (and it is regular



for  $(0, +\infty)$  if and only if  $X$  has infinite variation paths a.s.). We know from general theory for Markov process that there exists a local time at 0 for  $X - I$ , here denoted by  $(L_t, t \geq 0)$  that can be defined as the unique (up to a multiplicative constant) adapted additive functional that grows exactly on the zeros of  $X - I$ . Furthermore, the fact that  $X$  has no negative jumps entails that  $-I$  satisfies these conditions, so it is an explicit local time for the reflected process. Then, its right-continuous inverse

$$\tau_t := \inf \{s > 0 : -I_s > t\}$$

is the same as  $T_{-t} = \inf \{s \geq 0 : X_s < -t\}$ , the first hitting time of  $(-\infty, -t)$ . It is a (possibly killed) subordinator whose jumps coincide exactly with the excursion intervals of  $X - I$  so it represents the appropriate time scale for the so-called *excursion process*, that we will now describe.

Let  $\mathcal{E}$  be the space of real-valued càdlàg functions with finite lifetime  $V \in [0, \infty)$ , and we denote by  $\partial$  a topologically isolated state, so-called *cemetery point*. Define the excursion process  $\epsilon = (\epsilon_t, 0 < t \leq -I_\infty)$ , taking values in  $\mathcal{E} \cup \partial$  as follows

$$\epsilon_t := \begin{cases} ((X_{\tau_{t-}+s} - I_{\tau_{t-}}, 0 \leq s \leq \tau_t - \tau_{t-}), & \text{if } \tau_t - \tau_{t-} > 0, \\ \partial, & \text{if } \tau_t - \tau_{t-} = 0, \text{ or } t = \infty, \end{cases} \quad \text{for } t \geq 0.$$

Then according to Itô's excursion theory,  $(t, \epsilon_t)_{t \geq 0}$  is a Poisson point process, possibly stopped at the first excursion with infinite lifetime (which arrives only and a.s. when  $X$  drifts to  $+\infty$ ). Its intensity is  $dt \, \underline{n}(d\epsilon)$ , where  $\underline{n}$  is a measure on  $\mathcal{E}$  called the *excursion measure*. We refer to [Ber96, Chapter IV] for further details.

Let  $\varepsilon$  be the generic excursion of  $X - I$  away from 0 and  $\gamma$  the last instant at which this excursion attains its supremum, that is

$$\gamma = \gamma(\varepsilon) = \sup \{s > 0 : \varepsilon_s = \bar{\varepsilon}_s\},$$

where  $\bar{\varepsilon}_s = \sup_{[0,s]} \varepsilon$ . We are interested in the disintegration of  $\varepsilon$  at  $\gamma$ . Define the space-time-reversal transformation  $\rho$  for any excursion  $\omega \in \mathcal{E}$ , as  $\rho \circ \omega := (\omega_{V-} - \omega_{(V-s)-}, 0 \leq s \leq V)$ . We also call it *rotation*. We prove that the pre-supremum and the post-supremum processes, denoted respectively  $\overleftarrow{\varepsilon} = k_\gamma \circ \varepsilon = (\varepsilon_s, 0 \leq s \leq \gamma)$  and  $\overrightarrow{\varepsilon} = \theta'_\gamma \circ \varepsilon = (\varepsilon_{\gamma+s} - \varepsilon_\gamma, 0 \leq s \leq V - \gamma)$ , are invariant for this space-time-reversal transformation under the measure  $\underline{n}$ . Moreover, this results implies the following theorem, for which we need first to define the functional  $\chi : \mathcal{E} \rightarrow \mathcal{E}$  as

$$\chi(\varepsilon) := \left[ \rho(k_\gamma \circ \varepsilon), \rho(\theta'_\gamma \circ \varepsilon) + \varepsilon_\gamma \right],$$

where for any two elements  $\omega_1, \omega_2 \in \mathcal{E}$ ,  $[\omega_1, \omega_2]$  stands for their concatenation.

**Theorem 29.** *For every bounded measurable functionals  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  we have*

$$\underline{n}(F) = \underline{n}(F \circ \chi)$$

Williams [Wil74] studied the decomposition of the generic excursion  $X - I$  at its maximum for Brownian motion, showing it consists in two Bessel processes of dimension 3, started at 0, running to encounter each other, and killed upon hitting the same independent random variable (see e.g. [RY91, Chapter XII] or [Cha94, Section 5]). This result has been generalized to SPLP by Chaumont [Cha96], and for general Lévy processes by Duquesne [Duc03]. The law of each, the pre- $\gamma$  and post- $\gamma$  subpaths is characterized in [Cha96] and

[Duq03] in terms of the law of  $X$  *conditioned to stay positive*, denoted  $P^\uparrow$ . Our study differs from these works since here we show that this distributions are invariant under *space-time-reversal*. Rather our results provide in passing properties of reversal invariance under the law  $P^\uparrow$  that we state later in Section 3. For further details on path decomposition theorems for Lévy processes at the overall maximum, minimum and other random times, we refer to [Mil77b, Mil77a, GP80].

More recently, in [AD09, DH13] the authors also establish Williams decompositions under the excursion measure for the exploration process associated with the Lévy continuum random tree and super-processes with a spatially non-homogeneous quadratic branching mechanism. Several properties of these branching processes are then derived from these decompositions, such as a closed formula for the probability of hitting zero for a CSBP with immigration. In [DH13], the Q-process is obtained by looking at the super-process from the root and letting the extinction time tend to infinity. Moreover, an equivalent of the Esty time reversal from [KRS07] is given in a continuous setting.

Another result is obtained by Miermont in [Mie01] concerning a similar decomposition via the Vervaat's transform. Proposition 1 in [Mie01] applied to SPLP of infinite variation not drifting to  $+\infty$  and having bi-continuous marginal densities w.r.t. to Lebesgue measure, has the following implications: the excursion above the infimum conditioned to have a duration equal to  $l$  is well defined, and if we cut this excursion at a uniform point  $v$  and we concatenate the post- $v$  and the pre- $v$  subpaths (in this order), we get a Lévy bridge going from 0 to 0 in  $l$  units of time (which is well defined under these hypotheses). Such a bridge is clearly invariant by rotation and hence, by conditioning respectively on the events  $\{v < \gamma\}$  and  $\{v > \gamma\}$ , we could have obtained that the laws of  $\overleftarrow{\varepsilon}$  and  $\overrightarrow{\varepsilon}$  are also invariant by rotation. This approach provides an alternative way of obtaining some of our results under the technical hypothesis of continuity on the marginal densities.

A first consequence of Theorem 29 is the invariance under time reversal of the local time process of the excursion  $X - I$  away from 0. To be more specific, define the local time process  $(\Gamma(\varepsilon, r), r \geq 0)$  for the canonical excursion  $\varepsilon \in \mathcal{E}$  as a Borel function satisfying

$$\int_0^{V(\varepsilon)} \phi(\varepsilon_s) ds = \int_0^\infty \Gamma(\varepsilon, r) \phi(r) dr, \quad (\text{III.2})$$

for any continuous function  $\phi$  with compact support in  $[0, \infty)$ . This local time processes are known to exist in the infinite variation case, see for instance [Ber96]. When  $X$  has finite variation we can define an equivalent process, taking values in  $\mathbb{N} \cup \{+\infty\}$ , as the number of times the excursion hits level  $r$ , i.e.

$$\Gamma(\varepsilon, r) = \sum_{0 \leq s \leq V} \mathbb{1}_{\{\varepsilon_s = r\}}. \quad (\text{III.3})$$

Then we can state the following result.

**Corollary 30.** *The local time process of the excursions of  $X - I$  away from 0, is invariant under time reversal, that is*

$$(\Gamma(\varepsilon, r), 0 \leq r \leq \varepsilon_\gamma) \stackrel{d}{=} (\Gamma(\varepsilon, \varepsilon_\gamma - r), 0 \leq r \leq \varepsilon_\gamma). \quad (\text{III.4})$$

As we previously pointed out, the main motivation that led us to look into this pathwise decomposition comes from *branching processes*, that is stochastic processes with non-negative values satisfying the *branching property*. This means that for any  $x, y > 0$ , the

process started at  $x + y$  has the same distribution as a sum of two independent copies of itself, starting respectively at  $x$  and  $y$ . The simplest branching processes are those in discrete time and state space, the well-known Bienaymé-Galton-Watson (GW) processes [AN72]. In the case of discrete time and continuous state-space, we use the term *Jirina processes* as in [Lam10]. For continuous time and discrete state space we speak of Crump-Mode-Jagers (CMJ) processes and finally the so-called *continuous state branching process* (CSBP) for continuous time and state spaces. Our results concern mainly the latter two, so we will spend more time specifying their characteristics in Section 4, but we refer to [AN72, Jag75] for the general theory of branching processes.

*Splitting trees* are random trees formed by individuals that behave independently from each other, have i.i.d. lifetime durations (possibly infinite), and give birth to i.i.d. copies of themselves during their lives (single births). The excursion of  $X - I$  can be viewed, in the finite variation case, as the contour process of a (sub)critical splitting tree and then its local time process is a CMJ [Lam10]. In the infinite variation case, under some mild assumptions, this excursion codes the genealogy of a *totally ordered measured* (TOM) tree satisfying the splitting property, which are the continuum analogue of chronological trees in the setting of real trees, as it is shown in [LB16]. Hence, Theorem 29 leads in particular to the invariance under time reversal from its extinction time of the (sub)-critical CMJ process which is the local time of the SPLP characterized by (III.1). The same holds for the excursion away from 0 of the critical Feller diffusion, which is the width process of the continuum random tree, [Ald91]. This can be summarized in the following corollary of Theorem 29.

**Corollary 31.** *The (sub)critical CMJ's branching process and excursion away from 0 of the critical Feller's branching diffusion, are invariant under time reversal from their extinction time.*

Similar results concerning the duality by time-reversal of branching processes have been given in the literature. In particular, in [AP05] we can find a time-reversal invariance principle for the linear birth and death process in the critical case, when the process is conditioned on the number of individuals at the time of reversal to be equal to  $n$ . As suggested by the authors, the rescaled limit of the time-reversed process when  $n \rightarrow \infty$ , is the Feller branching diffusion. This suggests an alternative way of obtaining the second result on the previous Corollary. See also [Est75] and more recently [AR02, KRS07, DH13] for the treatment of the reverse of Galton-Watson processes and specifically the *Esty time reversal*, which is the limit obtained by conditioning a GW process in negative time upon entering the state 0 (extinction) at time 0 and starting in the state 1 at time  $-n$ , when  $n$  tends to  $+\infty$ . We also refer to [BD16] for a time reversal property for the number of ancestors process of a stationary CSBP with sub-critical quadratic branching mechanism.

This paper is a follow-up to [DFL15], where we have obtained a property of invariance under time-reversal, from a deterministic time  $T$ , for the population size process of certain random forests. The latter are defined as a sequence of splitting trees stopped at the first tree surviving up to time  $T$ . In [DFL15] we focused on the time-reversal from a deterministic time  $T$  whereas here we are interested in the same property from the extinction time of the process. It is worth stressing that, besides the implications concerning branching processes, some of our lemmas are interesting in their own right since they provide some invariance results for subpaths of SPLP.

The paper is organized as follows. In a short Section 2 we introduce some preliminary notions and notations. It is followed up by Section 3, which contains our main results on

the path decompositions of SPLP reflected at their infimum under the excursion measure. In Section 4 we recall some notions linking SPLP to branching processes and give the main implications of our results in the context of the latter. Finally Section 5 is devoted to completing the remaining demonstrations.

## 2 Preliminaries

### Basic notation

Let  $\mathcal{B}(\mathbb{R})$  denote the Borel  $\sigma$ -field of  $\mathbb{R}$ . Consider the space  $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$  (or simply  $\mathcal{D}$ ) of càdlàg functions  $\omega$  from  $\mathbb{R}_+$  into the measurable space  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  endowed with Skorokhod topology [JS03]. Denote the corresponding Borel  $\sigma$ -field by  $\mathcal{B}(\mathcal{D})$ . Define the *lifetime* of a path  $\omega \in \mathcal{D}$  as  $V = V(\omega) := \inf\{t \geq 0 : \omega(s) = \omega(t), \forall s \geq t\}$ , with the convention  $\inf \emptyset = \infty$ . Here  $\omega(t-)$  stands for the left limit of  $\omega$  at  $t \in \mathbb{R}_+$ ,  $\Delta\omega(t) = \omega(t) - \omega(t-)$  for the size of the (possible) jump at  $t \leq V$  and we make the usual assumption  $\omega(0-) = \omega(0)$ . The subspace of functions in  $\mathcal{D}$  with finite lifetime is denoted  $\mathcal{E}$  as in the introduction.

We consider stochastic processes, on the probability space  $(\mathcal{D}, \mathcal{B}(\mathcal{D}), P)$ , say  $X = (X_t, t \geq 0)$ , also called the coordinate process, having  $X_t = X_t(\omega) = \omega(t)$ . In particular, we will consider only processes with no negative jumps, that is such that  $\Delta X_t \in \mathbb{R}_+$  for every  $t \geq 0$ . The canonical filtration is denoted by  $(\mathcal{F}_t)_{t \geq 0}$  and we let  $\mathcal{P}(E)$  be the collection of all probability measures on any space  $E$ . We use the notation  $P_x(X \in \cdot) = P(X \in \cdot | X_0 = x)$ . In the absence of subscript the process is considered to start at 0 a.s.

Define by  $T_A := \inf\{t > 0 : X_t \in A\}$ , the first hitting time of the set  $A \in \mathcal{B}(\mathbb{R})$ , with the conventions  $T_0 = T_{\{0\}}$ , and for any  $x > 0$ ,  $T_{-x} = T_{(-\infty, -x)}$ ,  $T_x = T_{(x, +\infty)}$ . Note that in general  $T_{\{\pm x\}} \neq T_{\pm x}$ , for  $x > 0$ . However, since  $X$  has no negative jumps,  $X$  is a.s. continuous at  $T_{-x}$  for all  $x > 0$ , and then it holds that  $T_{\{-x\}} = T_{-x} = T_{(-\infty, -x)}$  a.s.

As usual, for real valued functions,  $\|\cdot\|_\infty$  stands for the uniform norm on the corresponding space, and  $\|\omega\|_S := \sup_{[0, S]} |\omega|$  for the supremum up to a finite value.

### Some path transformations of càdlàg functions

In this subsection we will define some families of operators on the space of càdlàg functions  $\omega \in \mathcal{D}$ :

- the classical *shift operators*,  $\theta_s, s \in \mathbb{R}_+$ , defined by

$$[\theta_s(\omega)]_t := \omega_{s+t}, \quad \forall t \in \mathbb{R}_+$$

- the non-standard shift operators,  $\theta'_s, s \in \mathbb{R}_+$ , defined by

$$[\theta'_s(\omega)]_t := \omega_{s+t} - \omega_s, \quad \forall t \in \mathbb{R}_+$$

- the *killing operators*,  $k_s, s \in \mathbb{R}_+$ , defined by

$$[k_s(\omega)]_t := \begin{cases} \omega_t, & \text{if } t < s \\ \omega_s, & \text{otherwise} \end{cases}$$

the killing operator can be generalized to killing at random times, for instance  $k_{T_A}(X) = k_{T_A(X)}(X)$ , denotes the process  $X$ , *killed at the first passage into  $A$* . It is easy to see that if  $X$  is a Markov process, so is  $k_{T_A}(X)$ . We set

$$k_0(\omega) = \partial.$$

What we call killed path here, is more commonly denominated *stopped* path. The difference is that killing usually refers to the path being sent to an isolated state after the killing time, whereas here they remain constant with a real value. We highlight our interest in keeping the track of the final jump of the functions we study, which justifies this choice.

- the *space-time-reversal* mapping  $\rho_s$ ,  $s \in \mathbb{R}_+^*$ , as

$$[\rho_s(\omega)]_t := \begin{cases} \omega_s - \omega_{(s-t)-} & \forall t \in [0, s] \\ \omega_s - \omega_0 & \text{if } t > s \end{cases}$$

and when  $V < +\infty$ , we call *rotation*, denoted simply by  $\rho$ , the space-time-reversal operator at the lifetime of a path, that is  $\rho = \rho_V$ . Notice that  $[\rho_s(\omega)]_0 = \Delta\omega_s$  (possibly  $\neq 0$ ).

The notations  $P \circ \theta_s^{-1}$ ,  $P \circ k_s^{-1}$  and  $P \circ \rho^{-1}$  stand for the law of the shifted, killed and space-time-reversed processes when  $P$  is the law of  $X$ .

For a sequence of functions in the same state space, say  $(\omega_i)_{i \geq 1}$  with lifetimes  $(V_i, i \geq 1)$ , we define a new process by the concatenation of the terms of the sequence, denoted by

$$[\omega_1, \omega_2, \dots]$$

where the juxtaposition of terms is considered to stop at the first element with infinite lifetime. For instance, if  $V_1 < +\infty$  and  $V_2 = +\infty$ , then for every  $n \geq 2$

$$[\omega_1, \omega_2, \dots, \omega_n]_t = \begin{cases} \omega_{1,t} & \text{if } 0 \leq t \leq V_1 \\ \omega_{2,t-V_1} & t > V_1 \end{cases}.$$

Notice that a concatenation of càdlàg functions thus defined, might not be a càdlàg function, since, for instance, in the case where  $n = 2$ , the first function might end with a jump, so the new function  $[\omega_1, \omega_2]$  will be càdlàg only if  $\omega_{1,V_1} = \omega_{2,0}$ . This is always the case in our applications, that is why we choose to concatenate functions in this less usual way, which has the property of recording the final jump of each path.

### Skhorokhod topology

As mentioned before, we consider the space of càdlàg functions  $\mathcal{D}$  and the subset of excursions  $\mathcal{E}$  (paths with finite lifetime), to be endowed with the topology induced by Skorokhod's topology, which makes  $\mathcal{D}$  a Polish space. We refer to [JS03] for further details on this topology, which can be characterized as follows: a sequence  $(\varepsilon_n)$  on  $\mathcal{E}$  converges to  $\varepsilon$  when  $n \rightarrow \infty$ , if and only if there exists a sequence  $(\lambda_n)$  of *changes of time* (continuous, strictly increasing functions, with  $\lambda_n(0) = 0$  and  $\lambda_n(t) \uparrow \infty$  when  $t \uparrow \infty$ ), such that  $\|\lambda_n - \text{Id}\|_\infty \rightarrow 0$  and  $\|\varepsilon_n \circ \lambda_n - \varepsilon\|_T \rightarrow 0$  for all  $T \geq 0$ . The space of continuous bounded functions from  $\mathcal{E}$  into  $\mathbb{R}_+$  with respect to the Skorokhod topology, will be denoted by  $\mathcal{C}_b(\mathcal{E}, \mathbb{R}_+)$ .

### Properties of the Laplace exponent and scale function of a SPLP

The Laplace exponent given by III.1 is infinitely differentiable, strictly convex (when  $\Pi \neq 0$  or  $\beta \neq 0$ ),  $\psi(0) = 0$  and  $\psi(+\infty) = +\infty$ . Let  $\eta := \sup\{\lambda \geq 0 : \psi(\lambda) = 0\}$ . Then we have that  $\eta = 0$  is the unique root of  $\psi$ , when  $\psi'(0+) \geq 0$ . Otherwise the Laplace exponent has two roots, 0 and  $\eta > 0$ . It is known that for any  $x > 0$ ,

$$P_x(T_0 < +\infty) = e^{-\eta x}.$$

More generally, there exists a unique continuous increasing function  $W : [0, +\infty) \rightarrow [0, +\infty)$ , called the *scale function*, characterized by its Laplace transform,

$$\int_0^{+\infty} e^{-\lambda x} W(x) dx = \frac{1}{\psi(\lambda)}, \quad \lambda > \eta,$$

such that for any  $0 < x < a$ ,

$$P_x(T_0 < T_a) = \frac{W(a-x)}{W(x)}. \quad (\text{III.5})$$

### Time-reversal duality for Lévy processes

One of the key ingredients of our results is the duality property under time-reversal of Lévy processes (see [Ber96, Chapter II] for details). Roughly speaking, it states that if a path is space-time-reversed at a finite time horizon, the new path has the same distribution as the original process. We will use the following formulation subsequently: for every fixed  $t > 0$  and every non-negative measurable function  $F$  we have that

$$E[F(k_t \circ X)] = E[F(\rho \circ (k_t \circ X))]. \quad (\text{III.6})$$

By integrating over  $t$ , this result is still valid if the process is killed at an independently random finite time.

## 3 Main results

Throughout this section  $X$  denotes a SPLP with Lévy measure  $\Pi$  on  $(0, +\infty)$ , whose Laplace exponent denoted by  $\psi$  is defined by (III.1). As in the preliminaries, we let  $P_x$  denote the law of the process conditioned on  $X_0 = x$ . We assume  $\psi'(0+) \geq 0$ , meaning we are in the (sub)critical regime. Let  $S_t := \sup\{X_s, 0 \leq s \leq t\}$  and  $I_t := \inf\{X_s, 0 \leq s \leq t\}$  be the running supremum and the running infimum of the Lévy process  $X$ .

### Pre-supremum processes

We recall that  $\underline{n}$  denotes the excursion measure of the process  $X - I$  away from 0. Let  $g_t$  and  $d_t$  be the left and right-end points of the excursion straddling  $t$ , denoted  $e_t$ , that is,

$$e_t := (X_{g_t+s} - I_t, 0 \leq s \leq d_t - g_t).$$

For any excursion  $\varepsilon$  and any  $s \in \mathbb{R}_+$  define its supremum  $\bar{\varepsilon}_s := \sup_{[0,s]} \varepsilon$  and the last instant where the supremum is attained on the interval  $[0, s]$ , that is

$$\gamma(s) = \gamma(s, \varepsilon) := \operatorname{argmax}_{[0,s]} \varepsilon = \sup \{s' \in [0, s] : \varepsilon(s') = \bar{\varepsilon}_s\}.$$

A result from [Mil77b, Duq03] ensures this instant is unique a.s. thanks to the regularity of 0 for  $(-\infty, 0)$ . Let also  $V = V(\varepsilon)$  stand for the lifetime of the excursion (in particular  $V(e_t) = d_t - g_t$ ). In general when using  $\gamma$  and  $V$ , the dependence on the excursion under focus will be omitted unless there is a risk of confusion. Notice that under  $\underline{n}$ ,  $\gamma(s, \varepsilon) = 0$  if and only if  $\varepsilon = \partial$  or  $s = 0$ .

Define in a similar way, for the process  $X$ ,

$$\begin{aligned}\bar{\sigma}_t(X) &:= \operatorname{argmax}_{[0,t]} X, \quad \text{and} \\ \underline{\sigma}_t(X) &:= \operatorname{arginf}_{[0,t]} X = \sup \{s' \in [0,t] : I_{s'} = X_{s'-}\}.\end{aligned}$$

Let  $\omega \in \mathcal{D}$  be any path with finite lifetime. We are interested in the trajectories where the infimum is attained before the maximum, so let us define the event

$$\mathcal{A}(\omega) := \{\operatorname{arginf} \omega < \operatorname{argmax} \omega\}.$$

We will be interested in particular in  $\mathcal{A}(k_t \circ X)$  where

$$\mathcal{A}(k_t \circ X) = \{\underline{\sigma}_t(X) < \bar{\sigma}_t(X)\} = \left\{ S_{g_t} - X_{g_t} < \sup_{(0,t-g_t)} e_t \right\}.$$

We can now state the next result.

**Proposition 32.** *The pre-supremum process of the excursion of  $X - I$  away from zero is invariant under time reversal, that is, for any measurable functional  $h : \mathcal{E} \rightarrow \mathbb{R}_+$ ,*

$$\underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) \right) = \underline{n} \left( h \circ \rho \left( k_{\gamma(V)} \circ \varepsilon \right) \right). \quad (\text{III.7})$$

This result is based on the following lemmas, for which we need first to define the set  $\mathcal{H}$  of functions of *exponential type* in the lifetime of excursions. That is, measurable functions  $f : \mathcal{E} \rightarrow \mathbb{R}_+$ , satisfying  $f(\partial) = 0$ , such that there exist two non-negative constants  $c$  and  $C$  such that  $f(\varepsilon)e^{-cV(\varepsilon)} \leq C$ , for every  $\varepsilon \in \mathcal{E}$ .

**Lemma 33.** *For any functional  $f \in \mathcal{H}$ , the function*

$$\underline{n} \left( f \left( k_{\gamma(s)} \circ \varepsilon \right) \mathbb{1}_{\{s < V\}} \right)$$

*is right-continuous for every  $s > 0$ .*

*Proof.* See Section 5. □

**Lemma 34.** *For any functional  $f \in \mathcal{H}$  and every  $s \geq 0$ , we have the following identities*

(i)

$$\underline{n} \left( f \left( k_{\gamma(s)} \circ \varepsilon \right) \mathbb{1}_{\{s < V\}} \right) = \underline{n} \left( f \circ \rho \left( k_{\gamma(s)} \circ \varepsilon \right) \mathbb{1}_{\{s < V\}} \right), \quad (\text{III.8})$$

(ii)

$$\underline{n} \left( f(k_{\gamma(V)} \circ \varepsilon) \mathbb{1}_{\{\gamma(V) < s < V\}} \right) = \underline{n} \left( f \circ \rho(k_{\gamma(V)} \circ \varepsilon) \mathbb{1}_{\{\gamma(V) < s < V\}} \right). \quad (\text{III.9})$$

*Proof.* Let us define the following functional

$$F_1(k_t \circ X) = f \left( k_{\gamma(V)} \circ \theta'_{\underline{\sigma}_t(X)} \circ k_t(X) \right) \mathbb{1}_{\{(\bar{\sigma}_t - \underline{\sigma}_t)(X) > 0\}} g \left( X_{\bar{\sigma}_t} - X_{\underline{\sigma}_t-}, X_t - X_{\underline{\sigma}_t-} \right),$$

where  $f, g$  are also non-negative measurable functions, such that  $f(\partial) = 0$ . It is not hard to see that a.s.  $\mathcal{A}(k_t \circ X) = \{(\bar{\sigma}_t - \underline{\sigma}_t)(X) > 0\} = \mathcal{A}(\rho(k_t \circ X))$  since  $\underline{\sigma}_t(\rho(X \circ k_t)) =$

$t - \bar{\sigma}_t(k_t \circ X)$  and  $\bar{\sigma}_t(\rho(k_t \circ X)) = t - \underline{\sigma}_t(k_t \circ X)$ . Hence, the duality (III.6) applied to  $F_1$  gives that

$$\begin{aligned} E \left[ f \left( X_{\underline{\sigma}_t+s} - X_{\underline{\sigma}_t-}, 0 \leq s \leq \bar{\sigma}_t - \underline{\sigma}_t \right) \mathbb{1}_{\{\underline{\sigma}_t < \bar{\sigma}_t\}} g \left( X_{\bar{\sigma}_t} - X_{\underline{\sigma}_t-}, X_t - X_{\underline{\sigma}_t-} \right) \right] \\ = E \left[ f \left( X_{\bar{\sigma}_t} - X_{(\bar{\sigma}_t-s)-}, 0 \leq s \leq \bar{\sigma}_t - \underline{\sigma}_t \right) \mathbb{1}_{\{\underline{\sigma}_t < \bar{\sigma}_t\}} g \left( X_{\bar{\sigma}_t} - X_{\underline{\sigma}_t-}, X_{\bar{\sigma}_t} \right) \right]. \end{aligned}$$

Notice that  $\underline{\sigma}_t$  is the left-end point of  $e_t$ , the excursion straddling  $t$ , this point is denoted  $g_t$ . Similarly,  $\bar{\sigma}_t$  is the point where this excursion attains its maximum before  $t$ . This implies that  $P$ -a.s. on  $\mathcal{A}(k_t \circ X)$ , we have  $k_{\gamma(t-g_t)} \circ e_t = (X_{\underline{\sigma}_t+s} - X_{\underline{\sigma}_t-}, 0 \leq s \leq \bar{\sigma}_t - \underline{\sigma}_t)$  and  $\rho(k_{\gamma(t-g_t)} \circ e_t) = (X_{\bar{\sigma}_t} - X_{(\bar{\sigma}_t-s)-}, 0 \leq s \leq \bar{\sigma}_t - \underline{\sigma}_t)$ , which allows us to write the preceding identity as follows

$$E \left[ f \left( k_{\gamma(t-g_t)} \circ e_t \right) \mathbb{1}_{\mathcal{A}(k_t \circ X)} g(e_t(\gamma(t-g_t)), e_t(t-g_t)) \right] \quad (\text{III.10a})$$

$$= E \left[ f \left( \rho \left( k_{\gamma(t-g_t)} \circ e_t \right) \right) \mathbb{1}_{\mathcal{A}(\rho \circ (k_t \circ X))} g(e_t(\gamma(t-g_t)), e_t(\gamma(t-g_t)) + I_t) \right]. \quad (\text{III.10b})$$

We will first develop the left-hand side of this equation. First, rewrite  $\mathcal{A}(k_t \circ X)$  as  $\{S_{g_t} - X_{g_t} < \max_{(0, t-g_t)}(e_t)\}$  and instead of stopping the process at  $t$ , we kill it at an exponential independent rate, or equivalently, we integrate (III.10a) against  $qe^{-qt}$ , giving

$$\int_0^{+\infty} qe^{-qt} dt E \left[ f \left( k_{\gamma(t-g_t)} \circ e_t \right) \mathbb{1}_{\{S_{g_t} - X_{g_t} < e_t(\gamma(t-g_t))\}} g(e_t(\gamma(t-g_t)), e_t(t-g_t)) \right].$$

As in the introduction, we let  $(\tau_u)_{u \geq 0}$  denote the inverse of the local time at 0 of the process  $X - I$ , and by  $\epsilon_u$  the excursion starting at  $\tau_{u-}$ . If we exchange the expectation and the integral in the preceding equation (Fubini's theorem), we can express the quantity inside the expectation as a sum taken over all the excursion intervals of  $X - I$  away from 0

$$\begin{aligned} \int_0^{+\infty} qe^{-qt} dt E \left[ \sum_{u: \Delta\tau_u > 0} \mathbb{1}_{\{\tau_{u-} < t \leq \tau_u\}} f \left( k_{\gamma(t-\tau_{u-})} \circ \epsilon_u \right) \mathbb{1}_{\{S_{\tau_{u-}} - X_{\tau_{u-}} < \epsilon_u(\gamma(t-\tau_{u-}))\}} \right. \\ \left. \times g(\epsilon_u(\gamma(t-\tau_{u-})), \epsilon_u(t-\tau_{u-})) \right] \\ = E \left[ \sum_{u: \Delta\tau_u > 0} \int_0^{\infty} qe^{-qt} dt \mathbb{1}_{\{\tau_{u-} < t \leq \tau_u\}} f \left( k_{\gamma(t-\tau_{u-})} \circ \epsilon_u \right) \mathbb{1}_{\{S_{\tau_{u-}} - X_{\tau_{u-}} < \epsilon_u(\gamma(t-\tau_{u-}))\}} \right. \\ \left. \times g(\epsilon_u(\gamma(t-\tau_{u-})), \epsilon_u(t-\tau_{u-})) \right]. \end{aligned}$$

We have applied Fubini's theorem once more for the last step. By the change of variable



$s = t - \tau_{u-}$ , taking again the expectation and applying the compensation formula we get

$$E \left[ \sum_{u: \Delta \tau_u > 0} \int_0^{\Delta \tau_u} q e^{-q \tau_{u-}} e^{-q s} ds f(k_{\gamma(s)} \circ \epsilon_u) \mathbb{1}_{\{S_{\tau_{u-}} - X_{\tau_{u-}} < \epsilon_u(\gamma(s))\}} g(\epsilon_u(\gamma(s)), \epsilon_u(s)) \right] \quad (\text{III.11a})$$

$$= E \left[ \int_0^{+\infty} du e^{-q \tau_u} \int \underline{n}(d\varepsilon) \int_0^{V(\varepsilon)} ds q e^{-q s} f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{S_{\tau_u} - X_{\tau_u} < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right] \\ = E \left[ \int_0^{+\infty} du e^{-q \tau_u} \varphi(S_{\tau_u} - X_{\tau_u}) \right], \quad (\text{III.11b})$$

where  $\varphi(x) = \underline{n} \left( \int_0^{V(\varepsilon)} ds q e^{-q s} f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{x < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right)$ .

Define  $G_q := \arginf_{(0, \mathbb{e}_q)} X$ , where  $\mathbb{e}_q$  is exponentially distributed with parameter  $q$  and is independent of  $X$ . Then we can apply again the compensation formula to expand the last expression as follows

$$E[\varphi(S_{G_q} - X_{G_q})] = E \left[ \sum_{u: \Delta \tau_u > 0} \mathbb{1}_{\{\tau_{u-} < \mathbb{e}_q < \tau_u\}} \varphi(S_{\tau_{u-}} - X_{\tau_{u-}}) \right] \\ = E \left[ \sum_{u: \Delta \tau_u > 0} (e^{-q \tau_{u-}} - e^{-q \tau_u}) \varphi(S_{\tau_{u-}} - X_{\tau_{u-}}) \right] \\ = E \left[ \int_0^{+\infty} du e^{-q \tau_u} \varphi(S_{\tau_u} - X_{\tau_u}) \right] \underline{n}(1 - e^{-qV}).$$

So we get that (III.11a) is equal to

$$\frac{E[\varphi(S_{G_q} - X_{G_q})]}{\underline{n}(1 - e^{-qV})}, \quad (\text{III.12})$$

and by applying Fubini's theorem one more time, and replacing  $\varphi$  by its expression this is equal again to

$$\frac{q}{\underline{n}(1 - e^{-qV})} E \left[ \underline{n} \left( \int_0^V ds e^{-q s} f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{Y_q < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right) \right] \\ = \frac{q}{\underline{n}(1 - e^{-qV})} \underline{n} \left( \int_0^V ds e^{-q s} f(k_{\gamma(s)} \circ \varepsilon) P(Y_q < \varepsilon_{\gamma(s)}) g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right) \\ = \frac{q}{\underline{n}(1 - e^{-qV})} \int_0^\infty ds e^{-q s} \underline{n} \left( f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{s < V\}} P(Y_q < \varepsilon_{\gamma(s)}) g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right).$$

where  $Y_q$  stands for an independent random variable, distributed as  $S_{G_q} - X_{G_q}$ .

We now go back to Equation (III.10) and apply all the above arguments to (III.10b). By choosing the same function  $f$  and observing that  $\varepsilon_{\gamma(s)} = \max_{[0, s]} \varepsilon = \max_{[0, s]} \rho(k_{\gamma(s)} \circ \varepsilon)$ , we obtain that this is equal to the following expressions, analogous to (III.11) and (III.12),

$$E \left[ \int_0^{+\infty} du e^{-q \tau_u} \tilde{\varphi}(S_{\tau_u} - X_{\tau_u}, I_{\tau_u}) \right] = \frac{E[\tilde{\varphi}(S_{G_q} - X_{G_q}, I_{G_q})]}{\underline{n}(1 - e^{-qV})},$$

where  $\tilde{\varphi}(x, y) = \underline{n} \left( \int_0^{V(\varepsilon)} ds q e^{-qs} f \circ \rho(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{x < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_{\gamma(s)} + y) \right)$ . Using the same arguments as before, and denoting by  $(Y_q, Z_q)$  a pair distributed as  $(S_{G_q} - X_{G_q}, I_{G_q})$ , this is also equal to

$$\begin{aligned} & \frac{q}{\underline{n}(1 - e^{-qV})} E \left[ \underline{n} \left( \int_0^V ds e^{-qs} f \circ \rho(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{Y_q < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_{\gamma(s)} + Z_q) \right) \right] \\ &= \frac{q}{\underline{n}(1 - e^{-qV})} \int_0^\infty ds e^{-qs} \underline{n} \left( f \circ \rho(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{s < V\}} E \left[ \mathbb{1}_{\{Y_q < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_{\gamma(s)} + Z_q) \right] \right), \end{aligned}$$

where it should be noted that the expectation is taken with respect to the law of  $(Y_q, Z_q)$ . Finally, since the first term in the above product is the same for (III.10a) and (III.10b), Equation (III.10) is equivalent to

$$\begin{aligned} & \int_0^\infty ds e^{-qs} \underline{n} \left( f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{s < V\}} P(Y_q < \varepsilon_{\gamma(s)}) g(\varepsilon_{\gamma(s)}, \varepsilon_s) \right) \\ &= \int_0^\infty ds e^{-qs} \underline{n} \left( f \circ \rho(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{s < V\}} E \left[ \mathbb{1}_{\{Y_q < \varepsilon_{\gamma(s)}\}} g(\varepsilon_{\gamma(s)}, \varepsilon_{\gamma(s)} + Z_q) \right] \right). \quad (\text{III.13}) \end{aligned}$$

- (i) In order to prove the first identity in the lemma we start by taking  $g \equiv 1$ . The probability  $P(Y_q < \varepsilon_{\gamma(s)})$  is also a function of  $k_{\gamma(s)} \circ \varepsilon$ , that is in addition strictly positive since  $\varepsilon_{\gamma(s)} = 0$  only if  $\varepsilon = \partial$ . Therefore we can consider

$$f(k_{\gamma(s)} \circ \varepsilon) = \frac{h(k_{\gamma(s)} \circ \varepsilon) e^{\alpha \gamma(s)}}{P(Y_q < \varepsilon_{\gamma(s)})}, \quad \forall s > 0, \quad h(\partial) = f(\partial) = 0,$$

where  $h$  is a non-negative bounded function and  $\alpha$  a non-negative constant. Then, (III.13) entails

$$\int_0^\infty ds e^{-qs} \underline{n} \left( h(k_{\gamma(s)} \circ \varepsilon) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right) = \int_0^\infty ds e^{-qs} \underline{n} \left( h \circ \rho(k_{\gamma(s)} \circ \varepsilon) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right).$$

Under regularity conditions guaranteeing the existence and injectivity of the Laplace transform, this identity implies (III.8). This is true in particular if both sides in (III.8) are right-continuous functions of  $s$ , for every  $s \in (0, +\infty)$ , grow at most exponentially and are locally integrable on  $[0, +\infty)$  [Zay96]. Lemma 33 (i) ensures the right-continuity on  $(0, +\infty)$ , so we will now focus on showing that the r.h.s. in (III.8) satisfies the other two conditions. Notice all the arguments we use below also apply when changing  $h$  by  $h \circ \rho$ .

Let us first show local integrability. For  $q > \alpha$ ,  $s > 0$  and any constant  $K > 0$ , we have that

$$\begin{aligned} \int_0^K ds \underline{n} \left( h(k_{\gamma(s)} \circ \varepsilon) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right) &\leq \|h\|_\infty e^{\alpha K} \int_0^K ds \underline{n}(s < V) \\ &= \|h\|_\infty e^{\alpha K} \underline{n}(V \wedge K) \\ &\leq C_{K, h, \alpha} \underline{n}(V \wedge 1), \end{aligned}$$

where as usual,  $\|\cdot\|_\infty$  stands for the uniform norm (on the space  $\mathcal{E}$  in this case),  $C_{K,h,\alpha}$  is a non-negative constant and  $\underline{n}(V \wedge 1)$  is always finite.

The exponential growth condition is also straightforward

$$\begin{aligned} \underline{n}\left(h\left(k_{\gamma(s)} \circ \varepsilon\right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}}\right) &\leq \|h\|_\infty \int \underline{n}(d\varepsilon) e^{\alpha s} \mathbb{1}_{\{s < V\}} \\ &\leq \|h\|_\infty \underline{n}(V > s) e^{\alpha s} \leq C'' e^{\alpha s}, \end{aligned}$$

for every  $s > 1$ , since as we have mentioned before,  $\underline{n}(V > 1)$  is finite.

Thus, we can conclude that the Laplace transform of both functions in (III.8) exist, so this identity holds for every  $s \geq 0$  (is trivial for  $s = 0$ ).

- (ii) In order to prove the second identity we will follow a similar path. Go back to Equation (III.13), on its r.h.s., we disintegrate  $\underline{n}$  with respect to  $\varepsilon_{\gamma(s)}$ , getting

$$\begin{aligned} &\int_0^\infty ds e^{-qs} \int_{x \in (0, +\infty)} \underline{n}\left(f \circ \rho\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{\varepsilon_{\gamma(s)} \in dx, s < V\}}\right) E\left[\mathbb{1}_{\{x > Y_q\}} g(x, x + Z_q)\right] \\ &= \int_0^\infty ds e^{-qs} \int_{x \in (0, +\infty)} \underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{\varepsilon_{\gamma(s)} \in dx, s < V\}}\right) E\left[\mathbb{1}_{\{x > Y_q\}} g(x, x + Z_q)\right] \\ &= \int_0^\infty ds e^{-qs} \underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} E\left[\mathbb{1}_{\{\varepsilon_{\gamma(s)} > Y_q\}} g\left(\varepsilon_{\gamma(s)}, \varepsilon_{\gamma(s)} + Z_q\right)\right]\right) \\ &= \int_0^\infty ds e^{-qs} \underline{n}\left(f \circ \rho\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} P\left(\varepsilon_{\gamma(s)} > Y_q\right) g\left(\varepsilon_{\gamma(s)}, \varepsilon_s\right)\right). \end{aligned}$$

We have applied the identity (III.8) for the first step, which implies in particular that for every  $f \in \mathcal{H}$  and  $s > 0$ ,

$$\underline{n}\left(f \circ \rho\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{\varepsilon_{\gamma(s)} \in dx, s < V\}}\right) = \underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{\varepsilon_{\gamma(s)} \in dx, s < V\}}\right),$$

since  $\varepsilon_{\gamma(s)}$  is invariant under time reversal of the excursion  $\varepsilon \circ k_{\gamma(s)}$  and  $f \circ \rho \circ \rho \equiv f$ . The latter argument also justifies the last equality, when applied directly to (III.13).

Moreover, since this is also equal to the l.h.s. in (III.13) we have

$$\begin{aligned} &\int_0^\infty ds e^{-qs} \underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} P\left(\varepsilon_{\gamma(s)} > Y_q\right) g\left(\varepsilon_{\gamma(s)}, \varepsilon_s\right)\right) \\ &= \int_0^\infty ds e^{-qs} \underline{n}\left(f \circ \rho\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} P\left(\varepsilon_{\gamma(s)} > Y_q\right) g\left(\varepsilon_{\gamma(s)}, \varepsilon_s\right)\right). \quad (\text{III.14}) \end{aligned}$$

We choose the function  $g$  as follows, for  $0 < z < x$ ,

$$g(x, z) = \frac{E_z[H(k_{T_0} \circ X, x)]}{P(Y_q < x)},$$

where  $H : \mathcal{E} \times (0, +\infty) \rightarrow \mathbb{R}_+$  is a measurable function.

Then, the Markov property of  $\underline{n}$  yields, for every  $s \geq 0$ ,

$$\begin{aligned} &\underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} E_{\varepsilon_s}\left[H\left(k_{T_0} \circ X, \varepsilon_{\gamma(s)}\right)\right]\right) \\ &= \underline{n}\left(f\left(k_{\gamma(s)} \circ \varepsilon\right) \mathbb{1}_{\{s < V\}} H\left(\theta_s \circ \varepsilon, \varepsilon_{\gamma(s)}\right)\right). \end{aligned}$$

In particular, for  $H(\varepsilon, x) = \mathbb{1}_{\{\sup \varepsilon < x\}}$  this is equal to

$$\underline{n} \left( f \left( k_{\gamma(s)} \circ \varepsilon \right) \mathbb{1}_{\{s < V\}} \mathbb{1}_{\{\sup \theta_s \circ \varepsilon < \varepsilon_{\gamma(s)}\}} \right).$$

Consequently, for this choice of  $g$  and  $H$ , Equation (III.14) becomes,

$$\begin{aligned} & \int_0^\infty ds \, e^{-qs} \underline{n} \left( f \left( k_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}_{\{\gamma(V) < s < V\}} \right) \\ &= \int_0^\infty ds \, e^{-qs} \underline{n} \left( f \circ \rho \left( k_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}_{\{\gamma(V) < s < V\}} \right). \end{aligned} \quad (\text{III.15})$$

Here again we have the identity between the Laplace transform of two functions of  $s$ . When  $f(\varepsilon) = h(\varepsilon)e^{\alpha V}$  for some  $h$  bounded and  $\alpha > 0$ , both integrands are locally integrable on  $[0, +\infty)$  by the same arguments used previously for (i). Right-continuity also holds thanks to Lemma 33 (ii), so we can inverse the Laplace transform in (III.15), leading to the desired identity for  $s > 0$ . The condition  $h(\mathcal{K}) = 0$  allows to conclude.  $\square$

*Proof of Proposition 32.* Let us go back to Equation (III.9). We can replace  $f(\cdot)$  by  $h(\cdot)e^{q\gamma(V(\cdot))}$ , for  $h$  a bounded measurable function. Then we integrate both sides with respect to  $qe^{-qs}ds$  and apply Fubini's theorem, which leads to

$$\underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) \left( 1 - e^{-q(V-\gamma(V))} \right) \right) = \underline{n} \left( h \circ \rho \left( k_{\gamma(V)} \circ \varepsilon \right) \left( 1 - e^{-q(V-\gamma(V))} \right) \right),$$

or equivalently,

$$\begin{aligned} & \underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) \left( 1 - e^{-q(V-\gamma(V))} \right) \mathbb{1}_{\{V-\gamma(V) > 0\}} \right) \\ &= \underline{n} \left( h \circ \rho \left( k_{\gamma(V)} \circ \varepsilon \right) \left( 1 - e^{-q(V-\gamma(V))} \right) \mathbb{1}_{\{V-\gamma(V) > 0\}} \right). \end{aligned}$$

Since  $h$  is non-negative, monotone convergence applies when  $q \rightarrow +\infty$ , leading to

$$\underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}_{\{V-\gamma(V) > 0\}} \right) = \underline{n} \left( h \circ \rho \left( k_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}_{\{V-\gamma(V) > 0\}} \right),$$

Finally, notice that  $\{\gamma(V) = V\} = \{V = 0\}$ , so the condition  $h(\partial) = 0$  ensures that the identity in the proposition holds for any  $h$  bounded. This is still true for any non-negative function, again by a monotone convergence argument.  $\square$

### Post-supremum process

We now give a result analogous to Proposition 32 for the post-supremum process of the excursions of  $X - I$  away from 0.

**Proposition 35.** *The post-supremum process of the excursion of  $X - I$  away from zero is invariant under time reversal, that is, for any measurable functional  $h : \mathcal{E} \rightarrow \mathbb{R}_+$ ,*

$$\underline{n} \left( h \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \right) = \underline{n} \left( h \circ \rho \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \right). \quad (\text{III.16})$$

Before we proceed to prove this result, we need to establish some lemmas, and the following proposition, which is interesting in its own right, since besides serving to prove our main results, it gives the invariance under time-reversal of parts of the trajectory of a killed SPLP.

**Proposition 36.** *Let  $x > 0$  and  $X$  be a SPLP starting at 0 and killed upon hitting  $(-\infty, -x)$ . This process shifted to the largest value where it attains its supremum before  $T_{-x}$  is invariant by rotation. More precisely, for every  $x > 0$ ,  $P$ -a.s.*

$$\theta'_{\gamma(V)} \circ k_{T_{-x}} \circ X \stackrel{d}{=} \rho \left( \theta'_{\gamma(V)} \circ k_{T_{-x}} \circ X \right).$$

We also need the following lemma, whose proof can be found in Section 5.

**Lemma 37.** *For every  $x > 0$  and every functional  $h \in \mathcal{C}_b(\mathcal{E}, \mathbb{R}_+)$ , the function  $z : [0, +\infty) \rightarrow [0, +\infty)$  defined as*

$$z(x) := E \left[ h \left( \theta'_{\gamma(V)} \circ k_{T_{-x}} \circ X \right) \right]$$

*is right-continuous on  $(0, +\infty)$ .*

*Proof of Proposition 36.* To demonstrate this result we follow a similar path to that of the proof of Lemma 34. We start by considering the complementary of the event  $\mathcal{A}$  defined at the beginning of this section, that is

$$\mathcal{A}^c(k_t \circ X) = \{\bar{\sigma}_t(X) < \underline{\sigma}_t(X)\}.$$

Notice that we may have  $\Delta X_{\underline{\sigma}_t} \neq 0$ , in particular in the finite variation case in which the excursions of  $X - I$  away from 0 starts by a jump [Cha96, CD05]. In order to make the notation less heavy we develop the proof only for the infinite variation case. Just a few modifications are needed to treat the general case. Likewise, the bounded variation case is a straightforward consequence of Lemma 3.8 in [DFL15], where, conditionally on  $\varepsilon_\gamma = x$ , the post-supremum process is shown to have the law  $P_x(\cdot | T_0 < T_x) \circ k_{T_0}^{-1}$ , which is invariant under time-reversal.

We define the functional  $F_2$  as follows

$$F_2(k_t \circ X) = f \left( \theta'_{\gamma(V)} \circ k_{\underline{\sigma}_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t),$$

where  $f$  and  $g$  are non-negative measurable functions and such that  $f(\partial) = 0$ . In order to apply the duality property (III.6) to this function, let us look at  $F_2 \circ \rho(k_t \circ X)$  or equivalently  $F_2(\rho \circ k_t \circ X)$ . Notice first that  $\mathcal{A}^c$  and  $X_t$  are invariant under time-reversal at  $t$ , and additionally, under  $\mathcal{A}^c$ , we also have that  $\bar{\sigma}_t = \bar{\sigma}_{\underline{\sigma}_t}$ , so it holds that

$$E \left[ f \left( \theta'_{\gamma(V)} \circ k_{\underline{\sigma}_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t) \right] = E \left[ f \circ \rho \left( \theta'_{\gamma(V)} \circ k_{\underline{\sigma}_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t) \right].$$

Let us integrate this equality in  $t$  against the Lebesgue measure,

$$\int_0^{+\infty} dt \, E \left[ f \left( \theta'_{\gamma(V)} \circ k_{\underline{\sigma}_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t) \right] \tag{III.17a}$$

$$= \int_0^{+\infty} dt \, E \left[ f \circ \rho \left( \theta'_{\gamma(V)} \circ k_{\underline{\sigma}_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t) \right]. \tag{III.17b}$$

Using the same strategy as before, we can express some quantities in this equation in terms of the excursion straddling  $t$  of the process reflected at its infimum. We recall that

$(\tau_u)_{u \geq 0}$  denotes the inverse of the local time at 0 of the process  $X - I$ , and  $\epsilon_u$  the excursion starting at  $\tau_{u-}$ . Recall also that  $-I$  is the local time at 0 for this excursion process and its inverse is  $\tau_u = T_{-u} = T_{(-\infty, -u)}$ . Thus, we can expand (III.17b) as follows

$$\begin{aligned} & \int_0^\infty dt \, E \left[ f \left( \theta'_{\gamma(V)} \circ k_{\sigma_t} \circ X \right) \mathbb{1}_{\mathcal{A}^c(k_t \circ X)} g(X_t) \right] \\ &= \int_0^\infty dt \, E \left[ \sum_{u: \Delta\tau_u > 0} \mathbb{1}_{\{\tau_{u-} < t \leq \tau_u\}} f \left( \theta'_{\gamma(V)} \circ k_{\tau_{u-}} \circ X \right) \mathbb{1}_{\{\sup_{(0, t-\tau_{u-})} \epsilon_u < S_{\tau_{u-}} - I_{\tau_{u-}}\}} g(X_t) \right]. \end{aligned}$$

Thanks to Fubini's theorem and the change of variable  $s = t - \tau_{u-}$ , followed by the application of the compensation formula, we obtain that this is equal to

$$\begin{aligned} &= E \left[ \sum_{u: \Delta\tau_u > 0} \int_0^{\Delta\tau_u} ds \, f \left( \theta'_{\gamma(V)} \circ k_{\tau_{u-}} \circ X \right) \mathbb{1}_{\{\sup_{(0, s)} \epsilon_u < S_{\tau_{u-}} - I_{\tau_{u-}}\}} g(X_{\tau_{u-}+s}) \right] \\ &= E \left[ \int_0^\infty du \, f \left( \theta'_{\gamma(V)} \circ k_{\tau_u} \circ X \right) \int \underline{n}(d\varepsilon) \int_0^V ds \, \mathbb{1}_{\{\sup_{(0, s)} \varepsilon < S_{\tau_u} - I_{\tau_u}\}} g(X_{\tau_u} + \varepsilon_s) \right] \\ &= \int_0^\infty du \, e^{-qu} E \left[ f \left( \theta'_{\gamma(V)} \circ k_{T_{-u}} \circ X \right) C_q(S_{T_{-u}} - I_{T_{-u}}) \right], \end{aligned}$$

where for any  $y \geq 0$  we set  $C_q(y) = \underline{n} \left( \int_0^V e^{q\varepsilon_s} ds \, \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y\}} \right)$  and we have taken the function  $g$  of the form  $g(x) = e^{qx}$ , with  $q > 0$ . Fubini's theorem was applied again in the last step.

Now notice that  $S_{T_{-u}} - I_{T_{-u}}$  is precisely the height of the excursion  $\theta'_{\gamma(V)} \circ k_{T_{-u}} \circ X$ , which allows us to choose the function  $f$  as follows,

$$f(\omega) = \frac{h(\omega)}{C_q(\sup \omega - \inf \omega)},$$

for any  $\omega \in \mathcal{E}$  and  $f(\partial) = h(\partial) = 0$ , where  $h$  is bounded, measurable and positive function. This is a valid choice for  $f$  as long as, for any  $y > 0$ , conditionally on the height of the excursion  $\omega$  to be equal to  $y$ ,  $C_q(y)$  is not 0 nor  $+\infty$ . By Fubini's theorem we have that

$$\underline{n} \left( \int_0^V e^{q\varepsilon_s} ds \, \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y\}} \right) = \int_0^{+\infty} ds \, \underline{n} \left( e^{q\varepsilon_s} \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y, s < V\}} \right),$$

which is equal to 0 if and only if  $\underline{n} \left( e^{q\varepsilon_s} \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y\}} \right) = 0$  for almost every  $s > 0$ , but this is not possible since by monotone convergence we have that

$$\underline{n} \left( e^{q\varepsilon_s} \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y, s < V\}} \right) \geq \underline{n} \left( \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y, s < V\}} \right) \nearrow \underline{n}(V > 0) = +\infty, \text{ when } s \searrow 0.$$

On the other hand,  $C_q(y)$  is also finite since

$$\int_0^{+\infty} ds \, \underline{n} \left( e^{q\varepsilon_s} \mathbb{1}_{\{\varepsilon_{\gamma(s)} < y, s < V\}} \right) \leq e^{qy} \int_0^1 ds \, \underline{n}(s < V) + e^{qy} \int_1^{+\infty} ds \, \underline{n}(\varepsilon_{\gamma(s)} < y, s < V).$$

The first term in the sum is equal to  $e^{qy} \underline{n}(V \wedge 1)$  which is always finite. For the integral in the second term we have that

$$\begin{aligned} \int_1^{+\infty} ds \, \underline{n}(\varepsilon_{\gamma(s)} < y, s < V) &\leq \int_1^{+\infty} ds \int_{(0, y)} \underline{n}(\varepsilon_1 \in dx, 1 < V) P_x(s-1 < T_0 < T_y) \\ &= \int_{(0, y)} \underline{n}(\varepsilon_1 \in dx, 1 < V) \int_1^{+\infty} ds \, P_x(s-1 < T_0 < T_y) \end{aligned}$$

and we have that  $\int_1^{+\infty} ds P_x(s-1 < T_0 < T_0 < T_y) \leq E_x [T_0 \mathbb{1}_{\{T_0 < T_y\}}] \leq E_x [T_0 \wedge T_y]$  which is finite for any  $x \in (0, y)$ , whether the process oscillates or drifts to  $\pm\infty$ , and is even more continuous for every  $x \in [0, y]$  as we will now see. Consider the  $q$ -resolvent kernel for the two-sided exit problem from the interval  $(0, y)$ , that is for any borel set  $A \subseteq (0, y)$  let

$$U^q(x, A) = E_x \left[ \int_0^{T_0 \wedge T_y} e^{-qt} \mathbb{1}_{\{X_t \in A\}} dt \right], \quad q \leq 0.$$

We know from [Ber97] that since the Lévy process has absolutely continuous resolvent kernels, there exists a  $q$ -resolvent density for the killed process that is given by [Ber97, Theorem 1] and that leads to the following

$$E_x [T_0 \wedge T_y] = U^0(x, (0, y)) = \frac{W(x)}{W(y)} \int_0^y W(y-r) dr - \int_0^x W(x-r) dr,$$

which is finite and continuous on every  $x \in [0, y]$  thanks to the properties of the scale function (defined by (III.5)). Thanks to the continuity on  $x$ , by integrating  $E_x [T_0 \wedge T_y]$  against the finite measure  $\underline{n}(\varepsilon_1 \in dx, 1 < V)$  over the bounded set  $(0, y)$ , we get a finite value, so we can conclude that  $C_q(y)$  is also finite.

Once we know we can choose  $f$  this way, and taking into account that all the preceding steps are valid for both sides in (III.17), in particular that the height of the excursion  $\theta'_{\gamma(V)} \circ k_{T-u} \circ X$  is invariant by the transformation  $\rho$ , Equation (III.17) becomes

$$\int_0^\infty du e^{-qu} E \left[ h \left( \theta'_{\gamma(V)} \circ k_{T-u} \circ X \right) \right] = \int_0^\infty du e^{-qu} E \left[ h \circ \rho \left( \theta'_{\gamma(V)} \circ k_{T-u} \circ X \right) \right].$$

We have one more time an identity between Laplace transforms of two functions. In virtue of Lemma 37, for any  $h \in \mathcal{C}_b(\mathcal{E}, \mathbb{R}_+)$ , this functions are right-continuous on  $(0, +\infty)$ . Besides, they are both bounded since  $h$  is, hence we have for any  $h \in \mathcal{C}_b(\mathcal{E}, \mathbb{R}_+)$  that

$$E \left[ h \left( \theta'_{\gamma(V)} \circ k_{T-u} \circ X \right) \right] = E \left[ h \circ \rho \left( \theta'_{\gamma(V)} \circ k_{T-u} \circ X \right) \right],$$

which completes the proof.  $\square$

*Proof of Proposition 35.* We can expand the l.h.s. in (III.16) by applying the Markov property of  $\underline{n}$  for any  $0 < s < \gamma(V)$  in the following manner,

$$\begin{aligned} \underline{n} \left( h \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}(s < \gamma(V)) \right) &= \int_{x \in (0, +\infty)} \int_{y \geq x} \underline{n} \left( \mathbb{1}(s < \gamma(V), \varepsilon_s \in dx, \varepsilon_{\gamma(V)} \in dy) h \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \right) \\ &= \int_{x \in (0, +\infty)} \int_{y \geq x} \underline{n}(\varepsilon_s \in dx, \bar{\varepsilon}_s < y, s < V) E_x \left[ h \left( \theta'_{\gamma(V)} \circ k_{T_0} \circ X \right) \mathbb{1} \left( \sup_{[0, T_0]} X \in dy \right) \right] \\ &= \int_{x \in (0, +\infty)} \int_{y \geq x} \underline{n}(\varepsilon_s \in dx, \bar{\varepsilon}_s < y, s < V) E_x \left[ h \circ \rho \left( \theta'_{\gamma(V)} \circ k_{T_0} \circ X \right) \mathbb{1} \left( \sup_{[0, T_0]} X \in dy \right) \right] \\ &= \underline{n} \left( h \circ \rho \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \mathbb{1}(s < \gamma(V)) \right), \end{aligned} \quad (\text{III.18})$$

where we have used Proposition 36 in the third line, which is possible since  $P_x$ -a.s.

$$\sup_{[0, T_0]} X = \sup_{[0, T_0]} \theta'_{\gamma(V)} \circ k_{T_0} \circ X = \sup_{[0, T_0]} \rho \left( \theta'_{\gamma(V)} \circ k_{T_0} \circ X \right).$$

Finally, since  $h$  is non-negative, the monotone convergence theorem can be applied to III.18 when  $s \downarrow 0$  and allow us to conclude.  $\square$

We are now ready to prove our main result, stated in the introduction, that we recall now. We first recall that the functional  $\chi : \mathcal{E} \rightarrow \mathcal{E}$  is defined as

$$\chi(\varepsilon) := \left[ \rho(k_\gamma \circ \varepsilon), \rho(\theta'_\gamma \circ \varepsilon) + \varepsilon_\gamma \right].$$

**Theorem 29.** *For every bounded measurable functionals  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  we have*

$$\underline{n}(F) = \underline{n}(F \circ \chi)$$

*Proof.* In the unbounded variation case, 0 is regular for both half-lines, so we can apply Theorem 4.10 from [Duq03], which ensures that the supremum of the excursion of  $X - I$  away from zero, i.e.  $\varepsilon_\gamma(V)$ , admits a density w.r.t. to Lebesgue measure under  $\underline{n}$ . What is more, this theorem states that for every  $x > 0$ , the pre and post-supremum subpaths are independent under  $\underline{n}(\cdot | \varepsilon_\gamma = x)$ . When the trajectories have finite variation, the conditional independence also holds, this result is due to [Mil73, GP80] and also [Cha94, Cha96]. Hence, we can disintegrate by the law of  $\varepsilon_\gamma$  and use this independence property, which together with Propositions 32 and 35, lead to the following identities

$$\begin{aligned} \underline{n}(F(\varepsilon)) &= \underline{n}(F([k_\gamma \circ \varepsilon, \theta'_\gamma \circ \varepsilon])) = \int_{x \in (0, +\infty)} \underline{n}(F([k_\gamma \circ \varepsilon, \theta'_\gamma \circ \varepsilon]) \mathbb{1}_{\{\varepsilon_\gamma \in dx\}}) \\ &= \int_{x \in (0, +\infty)} \underline{n}(F([k_\gamma \circ \varepsilon, \theta'_\gamma \circ \varepsilon]) | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \int_{x \in (0, +\infty)} \underline{n}(F([k_\gamma \circ \varepsilon, \theta'_\gamma \circ \varepsilon + x]) | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \int_{x \in (0, +\infty)} \int F([\eta, \eta']) \underline{n}(k_\gamma \circ \varepsilon \in d\eta, \theta'_\gamma \circ \varepsilon + x \in d\eta' | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \int_{x \in (0, +\infty)} \int \int F([\eta, \eta']) \underline{n}(k_\gamma \circ \varepsilon \in d\eta | \varepsilon_\gamma = x) \underline{n}(\theta'_\gamma \circ \varepsilon + x \in d\eta' | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \int_{x \in (0, +\infty)} \int \int F([\eta, \eta']) \underline{n}(\rho(k_\gamma \circ \varepsilon) \in d\eta | \varepsilon_\gamma = x) \underline{n}(\rho(\theta'_\gamma \circ \varepsilon) + x \in d\eta' | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \int_{x \in (0, +\infty)} \underline{n}(F([\rho(k_\gamma \circ \varepsilon), \rho(\theta'_\gamma \circ \varepsilon) + x]) | \varepsilon_\gamma = x) \underline{n}(\varepsilon_\gamma \in dx) \\ &= \underline{n}(F([\rho(k_\gamma \circ \varepsilon), \rho(\theta'_\gamma \circ \varepsilon) + \varepsilon_\gamma])) = \underline{n}(F(\chi \circ \varepsilon)). \end{aligned}$$

□

**Corollary 30.** *The local time process of the excursions of  $X - I$  away from 0, is invariant under time reversal, that is*

$$(\Gamma(\varepsilon, r), 0 \leq r \leq \varepsilon_\gamma) \stackrel{d}{=} (\Gamma(\varepsilon, \varepsilon_\gamma - r), 0 \leq r \leq \varepsilon_\gamma). \quad (\text{III.4})$$

*Proof.* For a fixed path  $\varepsilon \in \mathcal{E}$  corresponding to an excursion of  $X - I$  away from 0, let us identify the local time process of  $\chi(\varepsilon)$ , defined by the occupation density formula (III.2), as the measurable function  $(\Gamma(\chi(\varepsilon), r), r \geq 0)$  satisfying

$$\int_0^{V(\varepsilon)} \phi([\chi(\varepsilon)]_s) ds = \int_0^\infty \Gamma(\chi(\varepsilon), r) \phi(r) dr, \quad (\text{III.19})$$



for any continuous function  $\phi$  with compact support in  $[0, \infty)$ . Its existence is guaranteed by Theorem 29 since  $\chi(\varepsilon) \stackrel{d}{=} \varepsilon$ . The l.h.s. in this equation can be expanded in the following way (writing  $\gamma = \gamma(V)$ )

$$\begin{aligned} \int_0^{V(\varepsilon)} \phi([\chi(\varepsilon)]_s) ds &= \int_0^{\gamma(V)} \phi([\chi(\varepsilon)]_s) ds + \int_{\gamma(V)}^V \phi([\chi(\varepsilon)]_s) ds \\ &= \int_0^{\gamma(V)} \phi(\varepsilon_\gamma - \varepsilon_{(\gamma-s)-}) ds + \int_{\gamma(V)}^V \phi(\varepsilon_\gamma - \varepsilon_{(\gamma+V-s)-}) ds \\ &= \int_0^{V(\varepsilon)} \phi(\varepsilon_\gamma - \varepsilon_s) ds = \int_0^\infty \Gamma(\chi'(\varepsilon), r) \phi(r) dr, \end{aligned} \quad (\text{III.20})$$

where  $[\chi'(\varepsilon)]_s = \varepsilon_\gamma - \varepsilon_s$ , for any  $s \geq 0$ . On the other hand, we know from Theorem 29 that for any function  $\phi$  satisfying the conditions mentioned before, we have that

$$\int_0^\infty \Gamma(\chi(\varepsilon), r) \phi(r) dr \stackrel{d}{=} \int_0^\infty \Gamma(\varepsilon, r) \phi(r) dr.$$

Finally, this identity, together with (III.19) and (III.20), imply that

$$\Gamma(\varepsilon, \cdot) \stackrel{d}{=} \Gamma(\chi'(\varepsilon), \cdot),$$

which terminates the proof.  $\square$

*Remark 38.* Beyond the independence between the pre and post-supremum subpaths, Theorem 4.10 in [Duq03] gives the following characterization of the law of the pre and post-supremum processes under  $\underline{n}(\cdot | \varepsilon_\gamma = x)$  in terms of the laws  $P^\uparrow$  and  $P^\downarrow$ , corresponding to the process  $X$  conditioned respectively to stay positive or negative. We refer to [Ber93, Ber96] and also [Duq03] for the details on the construction of these laws. The result in the aforementioned theorem has the following implications in our setting in the case of infinite variation. For every positive measurable functional  $h : \mathcal{E} \rightarrow \mathbb{R}_+$  we have

$$\begin{aligned} \underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) \middle| \varepsilon_\gamma = x \right) &= E^\uparrow [h(k_{T_x} \circ X) | X_{T_x} = x], \\ \underline{n} \left( h \left( \theta'_{\gamma(V)} \circ \varepsilon \right) \middle| \varepsilon_\gamma = x \right) &= E^\downarrow [h(k_{T_{-x}} \circ X)]. \end{aligned}$$

Thus, these identities combined with Propositions 35 and 32 imply that for any  $x > 0$ , the laws  $P^\uparrow \circ k_{T_x}^{-1}(\cdot | X_{T_x} = x)$  and  $P^\downarrow \circ k_{T_{-x}}^{-1}$  are also invariant by rotation. Additionally, we know from [Ber96, Chapter VII] that when the process drifts to  $-\infty$ , the law  $P^\downarrow$  can be viewed as the conditional law  $\bar{n}(\cdot | V = \infty)$ , or equivalently, as the law of  $X - S$  shifted at its last passage time at the origin. Here  $\bar{n}$  denotes the excursion measure of  $X - S$  away from 0, defined as in [Duq03] such that it records the final jump of the excursion. Hence, we also have for any positive measurable function  $h$  that

$$\bar{n} (h(k_{T_{-x}} \circ \varepsilon) | V = \infty) = \bar{n} (h \circ \rho(k_{T_{-x}} \circ \varepsilon) | V = \infty).$$

## 4 Applications

The study of the genealogical structure of branching processes is an essential aspect when it comes to their applications in the fields of population dynamics, population genetics and evolutionary biology. In the case of discrete state-space, the genealogy comes naturally from discrete trees, while for continuous-state processes their definition is a more delicate issue and is done via a non-Markovian process called the height process, which was introduced by Le Gall and Le Jan [LGLJ98] and is a functional of a SPLP. We will now briefly outline a few connections between random trees, branching processes and Lévy processes.

### 4.1 The continuum random tree

Real trees can be defined as the continuous limiting object of rescaled discrete trees and can be coded by a continuous function in a way similar to the coding of discrete trees by their contour functions. *Aldous' Continuum Random Tree* (the so-called CRT) can be defined as the random real tree coded by a normalized Brownian excursion  $e$ , i.e. the positive Brownian excursion conditioned to have lifetime 1. More generally, the tree coded by Brownian motion (possibly with drift) reflected at 0, is called Brownian forest. We refer to [Ald93, LG05] for the formalism on real trees.

#### Ray-Knight theorems

The second Ray-Knight theorem [RY91] establishes that the local time process of a reflected Brownian motion is Feller's branching diffusion. More precisely, let  $B$  be a Brownian motion reflected at 0 and  $(L_s^a, s, a \geq 0)$  the family of its local times, where the index  $s$  corresponds to the *time* of the original process  $B$  and  $a$  is the level variable moving in the state-space of  $B$ . Consider, for  $x > 0$ ,

$$\varsigma_x = \inf\{s : L_s^0 > x\}.$$

Then, the process  $(L_{\varsigma_x}^t, t \geq 0)$  is equal in distribution to the square of a 0-dimensional Bessel process started at  $x$ , that is, a *standard* Feller branching diffusion  $(Z_t^x, t \geq 0)$ . The latter is defined as the unique strong solution of the SDE

$$dZ_t^x = 2\sqrt{Z_t^x}dW_t^x, \text{ with } Z_0^x = x.$$

This may be understood as a description of the genealogy encoded in Feller's branching diffusion, meaning that reflected Brownian motion codes (in the sense of Aldous) the *real tree* which describes the genealogy of the population which evolves according to Feller's diffusion [LG05].

### 4.2 Splitting trees, CMJ's and contour process

A chronological tree is the subset of  $\bigcup_{n \geq 0} \mathbb{N}^n \times [0, +\infty)$  containing all the *existence points* of individuals living for a certain amount of time and giving birth to other during their lifetime. They are represented in the plane, as in Fig. III.1 (right), with time running from bottom to top, dotted lines representing filiations between individuals: the one on the left is the parent, and that on the right its descendant. We refer to [Lam10] for the details.

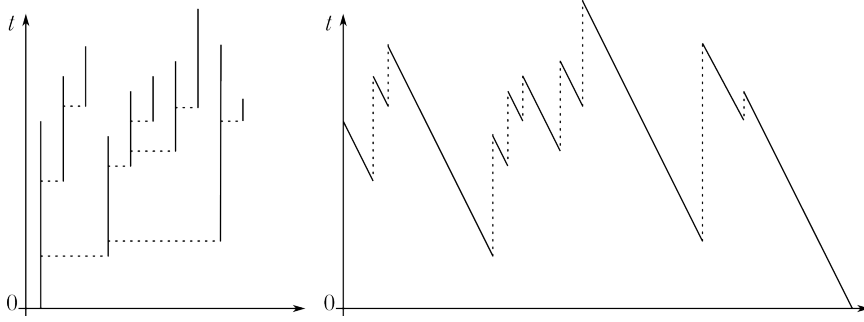


Figure III.1 – An example of chronographical tree with finite length (left) and its contour process (right).

Consider a population (or particle system) that originates at time 0 with one single progenitor, where individuals (particles) evolve independently of each other, giving birth to i.i.d. copies of themselves at constant rate, while alive, and having a lifetime duration with general distribution. The family tree under this stochastic model is a *splitting tree*, that can be formally defined as an element  $\mathcal{T}$  randomly chosen from the set of chronographical trees, characterized by a  $\sigma$ -finite measure  $\Pi$  on  $(0, \infty]$  called the *lifespan measure*, satisfying  $\int_{(0, \infty]} (r \wedge 1) \Pi(dr) < \infty$ . This means that if  $\Pi$  has mass  $b$ , the tree corresponds to a population where individuals have i.i.d. lifetimes distributed as  $\Pi(\cdot)/b$  and give birth to single descendants throughout their lives at constant rate  $b$ , all having the same independent behavior. In the general definition individuals may have infinitely many offspring and most of the following results remain valid if  $\Pi$  is infinite.

We can define the *width* or *population size process* of locally finite chronographical trees as a mapping  $\Xi$  that maps a chronographical tree  $\mathcal{T}$  to the function  $\xi : \mathbb{R}_+ \rightarrow \mathbb{N}$  counting the number of extant individuals at time  $t \geq 0$

$$\Xi(\mathcal{T}) := (\xi_t(\mathcal{T}), t \geq 0).$$

These functions are càdlàg, piecewise constant, from  $\mathbb{R}_+$  into  $\mathbb{N}$ , and are absorbed at 0. Then we can define the extinction event  $\text{Ext} := \{\lim_{t \rightarrow \infty} \xi_t(\mathcal{T}) = 0\}$  and the time of extinction of the population in a tree as

$$T_{\text{Ext}} := \inf\{t \geq 0 : \xi_t(\mathcal{T}) = 0\},$$

with the usual convention  $\inf \emptyset = \infty$ . A tree, or its width process  $\Xi$ , is said to be subcritical, critical or supercritical if

$$m := \int_{(0, +\infty]} r \Pi(dr).$$

is less than, equal to or greater than 0.

The width process  $\Xi(\mathcal{T}) = (\xi_t(\mathcal{T}), t \geq 0)$  of a splitting tree is known to be a *binary homogeneous Crump-Mode-Jagers process* (CMJ). This process is not Markovian, unless  $\Pi$  is exponential (*birth-death process*) or a Dirac mass at  $\{+\infty\}$  (*Yule process*).

### The contour of a splitting tree

As mentioned before, the genealogical structure of a chronographical tree can be coded via continuous or càdlàg functions. We focus in particular in the jumping chronographical

contour process (JCCP) from [Lam10]. The JCCP of a chronological tree  $\mathcal{T}$  with finite length  $\ell = \ell(\mathcal{T})$  (the sum of lifespans of all individuals), denoted by  $\mathcal{C}(\mathcal{T})$ , is a function from  $[0, \ell]$  into  $\mathbb{R}_+$ , that starts at the lifespan of the ancestor and then runs backward along the right-hand side of this first branch at speed  $-1$  until it encounters a birth event, when it jumps up of a height of the lifespan of this new individual, getting to the next tip, and then repeating this procedure until it eventually hits 0, as we can see in Fig. III.1 (see [Lam10] for a formal definition).

The JCCP visits all the existence times of each individual exactly once and the number of times it hits a time level, say  $s \geq 0$ , is equal to the number of individuals in the population at time  $s$ . More precisely, for any finite tree  $\mathcal{T}$ , the local time of its contour process is the population size process, that is

$$(\Gamma(\mathcal{C}(\mathcal{T}), r), 0 \leq r \leq T_{\text{Ext}}) = \Xi(\mathcal{T}),$$

where  $\Gamma$  is defined as in Equation (III.3).

One of the main results in [Lam10] states that the law of  $\mathcal{C}(\mathcal{T})$  when the tree has lifespan measure  $\Pi$ , conditional on Ext and on the lifespan of the root individual to be  $x$ , is a spectrally positive Lévy process  $Y$ , with Laplace exponent  $\psi(\lambda) = \lambda - \int_0^\infty (1 - \exp(-\lambda r))\Pi(dr)$ ,  $\lambda \geq 0$ , started at  $x$ , conditioned and killed upon hitting 0. A consequence of this result is that, under  $P_x$

$$(\Gamma(k_{T_0} \circ Y, r), r \geq 0) \tag{III.21}$$

is a CMJ with lifespan measure  $\Pi$ , starting with one progenitor with lifespan  $x$ .

These arguments together with Theorem 29 lead to Corollary 31.

### 4.3 Other results

In the same way as we did in Section 3, we consider the excursion process of  $X - S$  away from 0, the canonical excursion is denoted by  $\varepsilon$ , and  $\bar{n}$  is the excursion measure of this process, defined as in [Duq03] such that it records the final jump of the excursion. Define for any  $s \in \mathbb{R}_+$ , the largest instant at which the excursion attains its minimum on the interval  $[0, s]$ , that is

$$\nu(s) = \nu(s, \varepsilon) := \operatorname{arginf}_{[0, s]} \varepsilon = \sup \{s' \in [0, s] : \varepsilon(s' -) = \underline{\varepsilon}_s\},$$

where  $\underline{\varepsilon}_s := \inf_{[0, s]} \varepsilon$ . We write  $\nu = \nu(V)$  for the infimum up to the lifetime of the excursion. Then, the following results can be derived from those obtained in Section 3.

**Lemma 39.** *The post-supremum of the excursion of  $X - I$ , conditioned to have height  $x$  has the same distribution as the pre-infimum of the excursion of  $X - S$  conditioned to have depth greater than  $x$  killed upon hitting  $x$  by the first time. That is, for every bounded measurable functional  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  we have*

$$\underline{n} \left( F \left( \theta'_\gamma \circ \varepsilon \right) \mid \varepsilon_\gamma = x \right) = \bar{n} \left( F \left( k_{T-x} \circ \varepsilon \right) \mid T-x < \infty \right) \tag{III.22}$$

**Lemma 40.** *The conditional measure  $\bar{n} \circ k_{\nu(V)}^{-1} (\cdot \mid V < \infty)$  is invariant under time reversal. That is, for every bounded measurable functional  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  we have*

$$\bar{n} (F(k_\nu \circ \varepsilon) \mid V < \infty) = \bar{n} (F \circ \rho(k_\nu \circ \varepsilon) \mid V < \infty).$$

Moreover, conditionally on  $\varepsilon_\nu = x$ , they are also equal to  $\bar{n}(F(k_{T-x} \circ \varepsilon) \mid T-x < \infty)$ .

## 5 Remaining proofs

*Proof of Lemma 33.* Every function  $f \in \mathcal{H}$  can be expressed as  $f(\varepsilon) = h(\varepsilon)e^{\alpha V(\varepsilon)}$ , for a non-negative bounded function  $h$  satisfying  $h(\partial) = 0$  and a non-negative constant  $\alpha$ . Hence, here we want to prove that for every non-negative bounded function  $h$  with  $h(\mathcal{K}) = 0$  and any non-negative constant  $\alpha$ , the functions

$$\begin{aligned} & \underline{n} \left( h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right) \\ & \underline{n} \left( h \left( k_{\gamma(V)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{\gamma(V) < s < V\}} \right) \end{aligned}$$

are right-continuous for every  $s > 0$ .

Let us start by (i). Fix  $s > 0$ , and a sequence  $(s_n) \subset \mathbb{R}_+$  such that  $s_n \downarrow s$ . For  $\delta > 0$ , define the following subsets of  $\mathcal{E}$ :

$$\Upsilon_s(\delta) := \{\varepsilon \in \mathcal{E} : \bar{\varepsilon}(s - \delta) = \bar{\varepsilon}(s + \delta)\},$$

Then, we can analyze the continuity of  $\underline{n} \left( h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right)$  at  $s$  by splitting the space  $\mathcal{E}$  as follows for any  $\delta' > 0$

$$\begin{aligned} & \left| \underline{n} \left( h \left( k_{\gamma(s_n)} \circ \varepsilon \right) e^{\alpha \gamma(s_n)} \mathbb{1}_{\{s_n < V\}} \right) - \underline{n} \left( h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right) \right| \\ & \leq \int \left| h \left( k_{\gamma(s_n)} \circ \varepsilon \right) e^{\alpha \gamma(s_n)} \mathbb{1}_{\{s_n < V\}} - h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right| \underline{n}(d\varepsilon) \\ & = \underbrace{\int_{V \leq s + \delta'} |\cdot| \underline{n}(d\varepsilon)}_{(1)} + \underbrace{\int_{(\Upsilon_s(\delta))^c, V > s + \delta'} |\cdot| \underline{n}(d\varepsilon)}_{(2)} + \underbrace{\int_{\Upsilon_s(\delta), V > s + \delta'} |\cdot| \underline{n}(d\varepsilon)}_{(3)}. \end{aligned}$$

Notice the subtractions below make sens since  $\underline{n}(V > s) < +\infty$  for any  $s > 0$ . Now let us see what happens with each of the terms in this sum:

(1) Since  $s_n \geq s$ ,  $0 \leq \gamma(s) \leq s$  and  $h$  is bounded, we have

$$\begin{aligned} & \int_{V \leq s + \delta'} \left| h \left( k_{\gamma(s_n)} \circ \varepsilon \right) e^{\alpha \gamma(s_n)} \mathbb{1}_{\{s_n < V\}} - h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right| \underline{n}(d\varepsilon) \\ & \leq 2\|h\|_\infty e^{\alpha(s + \delta')} \underline{n}(s < V \leq s + \delta'). \end{aligned}$$

For every  $s, \delta' > 0$  it holds that  $\underline{n}(s < V \leq s + \delta') < +\infty$ . Therefore, downward monotone convergence applies and it implies that

$$\underline{n}(s < V \leq s + \delta') \longrightarrow 0, \text{ when } \delta' \rightarrow 0.$$

This allows to choose, for every  $\eta > 0$ , a suitable  $\delta'$  such that the term (1) is smaller than  $\frac{\eta}{2}$ .

(2) Again,  $h$  bounded implies that

$$\begin{aligned} & \int_{\substack{(\Upsilon_s(\delta))^c \\ V > s + \delta'}} \left| h \left( k_{\gamma(s_n)} \circ \varepsilon \right) e^{\alpha \gamma(s_n)} \mathbb{1}_{\{s_n < V\}} - h \left( k_{\gamma(s)} \circ \varepsilon \right) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right| \underline{n}(d\varepsilon) \\ & \leq 2\|h\|_\infty e^{\alpha(s + \delta')} \underline{n}((\Upsilon_s(\delta))^c, V > s + \delta'). \end{aligned}$$

On the other hand, from the definition of  $\Upsilon_s(\delta)$  and since the supremum is attained at a unique point  $\underline{n}$ -a.s., it follows from the dominated convergence theorem that

$$\lim_{\delta \rightarrow 0} \underline{n}((\Upsilon_s(\delta))^c, V > s + \delta') = \underline{n}(\varepsilon_s = \bar{\varepsilon}_s, V > s + \delta'). \quad (\text{III.23})$$

We now show that the r.h.s. of this limit is 0 for every fixed  $s > 0$ . For any  $u \in (0, s)$ ,

$$\begin{aligned} \underline{n}(\varepsilon_s = \bar{\varepsilon}_s, V > s + \delta') &= \int_{x \in (0, +\infty)} \int_{y \geq x} \underline{n}(\varepsilon_s = \bar{\varepsilon}_s, V > s + \delta', \varepsilon_u \in dx, \bar{\varepsilon}_u \in dy) \\ &= \int_{x \in (0, +\infty)} \int_{y \geq x} \underline{n}(\varepsilon_u \in dx, \bar{\varepsilon}_u \in dy) P_x(T_0^- > s + \delta' - u, S_{s-u} = X_{s-u} \geq y), \end{aligned}$$

where the last line comes from the Markov property. Besides, for  $y \geq x$

$$\begin{aligned} P_x(T_0^- > s + \delta' - u, S_{s-u} = X_{s-u} \geq y) &= P_0(T_{-x}^- > s + \delta' - u, S_{s-u} = X_{s-u} \geq y - x) \\ &\leq P_0(S_{s-u} = X_{s-u}) = P(\exists t > 0 : \mathcal{L}^{-1}(t) = s - u) = P(\mathcal{L}^{-1}(\mathcal{T}_{s-u}) = s - u), \end{aligned}$$

where  $\mathcal{L}^{-1}$  is the so-called *ladder time process*, which is the inverse of the local time at 0 of the process reflected at its supremum,  $S - X$ ; and  $\mathcal{T}_v := \inf\{t : \mathcal{L}^{-1}(t) > v\}$  for any  $v \geq 0$ . We know from [Ber96] that  $\mathcal{L}^{-1}$  is a subordinator, with drift equal to 0 when 0 is regular for  $(-\infty, 0)$ , which is always the case in absence of negative jumps (see for instance [Cha13]). Another result from [Ber96, Chapter III.2] tells us that any subordinator  $Y$  with drift 0 never creeps over any level  $x > 0$ , that is  $P(Y_{T_x^+} = x) = 0$ . Hence, we can conclude that

$$\underline{n}(\varepsilon_s = \bar{\varepsilon}_s, V > s + \delta') = 0, \quad (\text{III.24})$$

which guarantees, together with (III.23), that for any  $\eta > 0$ , we can choose  $\delta < \delta'$  sufficiently small that

$$\int_{\substack{(\Upsilon_s(\delta))^c \\ V > s + \delta'}} \left| h(k_{\gamma(s_n)} \circ \varepsilon) e^{\alpha \gamma(s_n)} \mathbb{1}_{\{s_n < V\}} - h(k_{\gamma(s)} \circ \varepsilon) e^{\alpha \gamma(s)} \mathbb{1}_{\{s < V\}} \right| \underline{n}(d\varepsilon) < \frac{\eta}{2}.$$

- (3) Let  $N_\delta$  be such that for  $n \geq N_\delta$ ,  $|s_n - s| < \delta$ , then  $\forall n \geq N_\delta, \forall \varepsilon \in \Upsilon_s(\delta)$ , such that  $V(\varepsilon) > s + \delta'$ , we have  $\gamma(s_n, \varepsilon) = \gamma(s, \varepsilon)$  and  $\mathbb{1}_{\{s_n < V\}} = \mathbb{1}_{\{s < V\}} = 1$ . Hence the third term is 0 for  $n \geq N_\delta$ .

Finally, we can conclude that the function  $\underline{n}(f(k_{\gamma(s)} \circ \varepsilon) \mathbb{1}_{\{s < V\}})$  is right-continuous for every  $s > 0$ .

For (ii) take as well  $s > 0$  and  $s_n \downarrow s$ . Fix  $\delta > 0$ , then there exists  $N_\delta$  such that for every  $n \geq N_\delta$ ,  $|s_n - s| < \delta$  and also

$$\begin{aligned} &\left| \underline{n}(f(k_{\gamma(V)} \circ \varepsilon) \mathbb{1}_{\{\gamma(V) < s_n < V\}} - \underline{n}(f(k_{\gamma(V)} \circ \varepsilon) \mathbb{1}_{\{\gamma(V) < s < V\}})) \right| \\ &= \int f(k_{\gamma(V)} \circ \varepsilon) \left| \mathbb{1}_{\{s \leq \gamma(V) < s_n < V\}} - \mathbb{1}_{\{\gamma(V) < s < V \leq s_n\}} \right| \underline{n}(d\varepsilon) \\ &\leq \|h\|_\infty e^{\alpha(s+\delta)} (\underline{n}(s \leq \gamma(V) < s_n < V) + \underline{n}(\gamma(V) < s < V \leq s_n)). \end{aligned}$$

Moreover, by dominated convergence, when  $n \rightarrow \infty$ ,

$$\underline{n}(\gamma(V) < s < V \leq s_n) \longrightarrow \underline{n}(\emptyset) = 0$$

and

$$\underline{n}(s \leq \gamma(V) < s_n < V) \longrightarrow \underline{n}(s = \gamma(V), s < V) \leq \underline{n}(\varepsilon_s = \bar{\varepsilon}_s, s < V) = 0,$$

as we has just proved in (III.24). So the function in (ii) is also right-continuous.  $\square$

*Proof of Lemma 37.* Notice first that for all  $x > 0$ , since  $X$  has no negative jumps,  $X$  is a.s. continuous at  $T_{-x}$ , that is  $P(\Delta X_{T_{-x}} = 0) = 1$ . This allow us to apply [JS03, Proposition VI.2.11 and VI.2.12], which ensure in this context that if we have  $x_n \downarrow x$ , then a.s.  $T_{-x_n} \downarrow T_{-x}$ , and moreover, the killed paths  $k_{T_{-x_n}} \circ X$  also converge to  $k_{T_{-x}} \circ X$  when  $n \rightarrow \infty$  in Skorokhod topology. Hence, it exists a sequence  $(\lambda_n)$  of changes of time (see Section 2) such that  $\|\lambda_n - \text{Id}\|_\infty \rightarrow 0$  and  $\|k_{T_{-x_n}} \circ X \circ \lambda_n - k_{T_{-x}} \circ X\|_M \rightarrow 0$  for all  $M \geq 0$ .

Additionally, since the sequence  $(x_n)$  is decreasing, we deduce from the definition of  $\gamma$  that  $(\gamma(T_{-x_n}, X))$  is also a decreasing sequence, and that for all  $n \geq 0$  we have

$$\gamma(T_{-x_n}, X) \geq \gamma(T_{-x}, X).$$

Hence  $\gamma(T_{-x_n}, X) \downarrow \ell$  for some  $\ell \geq 0$ . Suppose that  $\ell > \gamma(T_{-x}, X)$ , this implies that for every  $n \geq 0$ ,  $T_{-x} < \gamma(T_{-x_n}, X)$ , so we have

$$T_{-x} < \gamma(T_{-x_n}, X) \leq T_{-x_n}.$$

Then, the convergence of  $(T_{-x_n})$  entail that  $\gamma(T_{-x_n}, X) \downarrow T_{-x}$ , which in turn, since  $X$  is continuous at  $T_{-x}$ , implies that  $X_{\gamma(T_{-x_n})} \downarrow X_{T_{-x}} = -x$ . The latter is not possible since  $P$ -a.s.,  $\sup_{[0, T_{-x_n}]} X \geq 0$  for every  $n$ . Hence, we can conclude that  $\ell = \gamma(T_{-x}, X)$ , i.e.

$$\gamma(T_{-x_n}, X) \downarrow \gamma(T_{-x}, X).$$

Moreover, since for  $T_{-x}$  we also have that  $P$ -a.s.,  $\sup_{[0, T_{-x}]} X \geq 0$ , we can ensure that  $\gamma(T_{-x}, X) < T_{-x}$ , so the sequence  $(\gamma(T_{-x_n}, X))$  is not only convergent, but it is constant from some  $N \geq 0$ . As a consequence, we have that  $\forall n \geq N$

$$\theta'_{\gamma(V)} \circ k_{T_{-x_n}} \circ X = \theta'_{\gamma(T_{-x})} \circ k_{T_{-x_n}} \circ X,$$

whereby

$$\|\theta'_{\gamma(V)} \circ k_{T_{-x_n}} \circ X \circ \lambda_n - \theta'_{\gamma(V)} \circ k_{T_{-x}} \circ X\|_M \leq \|k_{T_{-x_n}} \circ X \circ \lambda_n - k_{T_{-x}} \circ X\|_M \rightarrow 0,$$

for all  $M \geq 0$ .

These arguments, together with the continuous mapping theorem applied to  $h \in \mathcal{C}_b(\mathcal{E}, \mathbb{R}_+)$ , lead to the convergence of  $h(\theta'_{\gamma(V)} \circ k_{T_{-x_n}} \circ X)$  to  $h(\theta'_{\gamma(V)} \circ k_{T_{-x}} \circ X)$ . Finally, since  $h$  is bounded, the dominated convergence theorem applies, and we can conclude that

$$\lim_n z(x_n) = z(x),$$

that is,  $z$  is right-continuous at  $x > 0$ . Since  $x$  is arbitrary, the result is proved.  $\square$

# Bibliography

- [AD09] Romain Abraham and Jean-François Delmas. Williams' decomposition of the Lévy continuum random tree and simultaneous extinction probability for populations with neutral mutations. *Stochastic Processes and their Applications*, 119(4):1124 – 1143, 2009.
- [Ald91] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.
- [Ald93] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [AN72] Krishna B. Athreya and Peter E. Ney. *Branching processes*. Springer-Verlag, New York-Heidelberg, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.
- [AP05] David Aldous and Lea Popovic. A critical branching process model for biodiversity. *Adv. in Appl. Probab.*, 37(4):1094–1115, 2005.
- [AR02] Gerold Alsmeyer and Uwe Rösler. Asexual versus promiscuous bisexual Galton-Watson processes: the extinction probability ratio. *Ann. Appl. Probab.*, 12(1):125–142, 2002.
- [BD16] Hongwei Bi and Jean-François Delmas. Total length of the genealogical tree for quadratic stationary continuous-state branching processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(3):1321–1350, 2016.
- [Bec74] Niels Becker. On parametric estimation for mortal branching processes. *Biometrika*, 61:393–399, 1974.
- [Bec77] Niels Becker. Estimation for discrete time branching processes with application to epidemics. *Biometrics*, 33(3):515–522, 1977.
- [Ber92] Jean Bertoin. An extension of Pitman's theorem for spectrally positive Lévy processes. *Ann. Probab.*, 20(3):1464–1483, 1992.
- [Ber93] Jean Bertoin. Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35, 1993.
- [Ber96] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
- [Ber97] Jean Bertoin. Exponential decay and ergodicity of completely asymmetric Lévy processes in a finite interval. *Ann. Appl. Probab.*, 7(1):156–169, 1997.



- [BPS12] Mamadou Ba, Etienne Pardoux, and Ahmadou Bamba Sow. Binary trees, exploration processes, and an extended Ray-Knight theorem. *J. Appl. Probab.*, 49(1):210–225, 2012.
- [CD05] Loïc Chaumont and Ronald A. Doney. On Lévy processes conditioned to stay positive. *Electron. J. Probab.*, 10:no. 28, 948–961, 2005.
- [Cha94] Loïc Chaumont. Sur certains processus de Lévy conditionnés à rester positifs. *Stochastics Stochastics Rep.*, 47(1-2):1–20, 1994.
- [Cha96] Loïc Chaumont. Conditionings and path decompositions for Lévy processes. *Stochastic Process. Appl.*, 64(1):39–54, 1996.
- [Cha13] Loïc Chaumont. On the law of the supremum of Lévy processes. *Ann. Probab.*, 41(3A):1191–1217, 2013.
- [CLUB09] Ma. Emilia Caballero, Amaury Lambert, and Gerónimo Uribe Bravo. Proof(s) of the lamperti representation of continuous-state branching processes. *Probab. Surveys*, 6(0):62–89, 2009.
- [DFL15] Miraine Dávila Felipe and Amaury Lambert. Time reversal dualities for some random forests. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(1):399–426, 2015.
- [DFL16] Miraine Dávila Felipe and Amaury Lambert. Branching processes seen from their extinction time via path decompositions of reflected Lévy processes. *ArXiv e-prints*, 1610.09195, 2016.
- [DH13] Jean-François Delmas and Olivier Hénard. A Williams decomposition for spatially dependent superprocesses. *Electron. J. Probab.*, 18(0), 2013.
- [DLG02] Thomas Duquesne and Jean-François Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002.
- [Don07] Ronald A. Doney. *Fluctuation theory for Lévy processes*, volume 1897 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [DPR03] Alexei Drummond, Oliver G. Pybus, and Andrew Rambaut. Inference of viral evolutionary rates from molecular sequences. *Adv Parasitol*, 54:331–358, 2003.
- [Duq03] Thomas Duquesne. Path decompositions for real Levy processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(2):339–370, 2003.
- [EHS<sup>+</sup>11] Rampal S. Etienne, Bart Haegeman, Tanja Stadler, Tracy Aze, Paul N. Pearson, Andy Purvis, and Albert B. Phillimore. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, October 2011.
- [Est75] Warren W. Esty. The reverse Galton-Watson process. *J. Appl. Probability*, 12(3):574–580, 1975.
- [FPG<sup>+</sup>15] Simon D.W. Frost, Oliver G. Pybus, Julia R. Gog, Cecile Viboud, Sebastian Bonhoeffer, and Trevor Bedford. Eight challenges in phylodynamic inference. *Epidemics*, 10(0):88 – 92, 2015. Challenges in Modelling Infectious {Disease} Dynamics.

- [Gei96] Jochen Geiger. Size-biased and conditioned random splitting trees. *Stochastic Process. Appl.*, 65(2):187–207, 1996.
- [GK97] J. Geiger and G. Kersting. Depth-first search of random trees, and Poisson point processes. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 111–126. Springer, New York, 1997.
- [GP80] Priscilla Greenwood and Jim Pitman. Fluctuation identities for Lévy processes and splitting at the maximum. *Adv. in Appl. Probab.*, 12(4):893–902, 1980.
- [GPG<sup>+</sup>04] Bryan T. Grenfell, Oliver G. Pybus, Julia R. Gog, James L. N. Wood, Janet M. Daly, Jenny A. Mumford, and Edward C. Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327–332, 2004.
- [Jag75] Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London-New York-Sydney, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.
- [Jag91] Peter Jagers. The growth and stabilization of populations. *Statistical Science*, 6(3):pp. 269–274, 1991.
- [JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [KA02] Marek Kimmel and David E. Axelrod. *Branching processes in biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, 2002.
- [Kac08] Mark Kac. *Discrete thoughts : essays on mathematics, science, and philosophy*. Birkhäuser, Boston, 2008.
- [Ken49] David G. Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):pp. 230–282, 1949.
- [KRS07] Fima C. Klebaner, Uwe Rösler, and Serik Sagitov. Transformations of galton-watson processes and linear fractional reproduction. *Advances in Applied Probability*, 39(4):1036–1053, 2007.
- [Kyp06] Andreas E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Universitext. Springer-Verlag, Berlin, 2006.
- [Lam67] John Lamperti. Continuous state branching processes. *Bull. Amer. Math. Soc.*, 73:382–386, 1967.
- [Lam10] Amaury Lambert. The contour of splitting trees is a Lévy process. *Ann. Probab.*, 38(1):348–395, 2010.

- [Lam11] Amaury Lambert. Species abundance distributions in neutral models with immigration or mutation and general lifetimes. *J. Math. Biol.*, 63(1):57–72, 2011.
- [LAS14] Amaury Lambert, Helen K. Alexander, and Tanja Stadler. Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *Journal of Theoretical Biology*, 352(0):60 – 70, 2014.
- [LB16] Amaury Lambert and Gerónimo Uribe Bravo. Totally ordered measured trees and splitting trees with infinite variation. *ArXiv e-prints*, 1607.02114v1, 2016.
- [LG05] Jean-François Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.
- [LGBS13] Gabriel E. Leventhal, Huldrych F. Günthard, Sebastian Bonhoeffer, and Tanja Stadler. Using an epidemiological model for phylogenetic inference reveals density-dependence in hiv transmission. *Molecular Biology and Evolution*, 2013.
- [LGLJ98] Jean-Francois Le Gall and Yves Le Jan. Branching processes in Lévy processes: the exploration process. *Ann. Probab.*, 26(1):213–252, 1998.
- [LS13] Amaury Lambert and Tanja Stadler. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology*, 90:113–128, dec 2013.
- [LSZ13] Amaury Lambert, Florian Simatos, and Bert Zwart. Scaling limits via excursion theory: interplay between Crump-Mode-Jagers branching processes and processor-sharing queues. *Ann. Appl. Probab.*, 23(6):2357–2381, 2013.
- [LT13] Amaury Lambert and Pieter Trapman. Splitting trees stopped when the first clock rings and Vervaat’s transformation. *J. Appl. Probab.*, 50(1):208–227, 2013.
- [Mie01] Grégory Miermont. Ordered additive coalescent and fragmentations associated to Levy processes with no positive jumps. *Electron. J. Probab.*, 6:no. 14, 33 pp. (electronic), 2001.
- [Mil73] Pressley Warwick Millar. Exit properties of stochastic processes with stationary independent increments. *Trans. Amer. Math. Soc.*, 178:459–479, 1973.
- [Mil77a] Pressley Warwick Millar. Random times and decomposition theorems. In *Probability (Proc. Sympos. Pure Math., Vol. XXXI, Univ. Illinois, Urbana, Ill., 1976)*, pages 91–103. Amer. Math. Soc., Providence, R. I., 1977.
- [Mil77b] Pressley Warwick Millar. Zero-one laws and the minimum of a Markov process. *Trans. Amer. Math. Soc.*, 226:365–391, 1977.
- [MSM<sup>+</sup>13] Gkikas Magiorkinis, Vana Sypsa, Emmanouil Magiorkinis, Dimitrios Paraskevis, Antigoni Katsoulidou, Robert Belshaw, Christophe Fraser, Oliver George Pybus, and Angelos Hatzakis. Integrating phylodynamics and epidemiology to estimate transmission diversity in viral epidemics. *PLoS Comput Biol*, 9(1):e1002876, jan 2013.

- [NMH94] Sean Nee, Robert M. May, and Paul H. Harvey. The Reconstructed Evolutionary Process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344(1309):305–311, May 1994.
- [PFR13] O. G. Pybus, C. Fraser, and A. Rambaut. Evolutionary epidemiology: preparing for an age of genomic plenty. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1614), 2013.
- [Pop04] Lea Popovic. Asymptotic genealogy of a critical branching process. *Ann. Appl. Probab.*, 14(4):2120–2148, 2004.
- [PR09] Oliver G. Pybus and Andrew Rambaut. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*, 10(8):540–550, aug 2009.
- [PW11] Etienne Pardoux and Anton Wakolbinger. From Brownian motion with a local time drift to Feller’s branching diffusion with logistic growth. *Electronic Communications in Probability*, 16(0):720–731, 2011.
- [Ran97] Bruce Rannala. Gene genealogy in a population of variable size. *Heredity*, 78(1309):417–423, April 1997.
- [RBK13] David A. Rasmussen, Maciej F. Boni, and Katia Koelle. Reconciling phylogenetics with epidemiology: The case of dengue virus in southern vietnam. *Molecular Biology and Evolution*, 2013.
- [Rog66] Boris Alekseevich Rogozin. On distributions of functionals related to boundary problems for processes with independent increments. *Theory of Probability & Its Applications*, 11(4):580–591, jan 1966.
- [RRK11] David A. Rasmussen, Oliver Ratmann, and Katia Koelle. Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput. Biol.*, 7(8):e1002136, 11, 2011.
- [RY91] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1991.
- [Sca16] S.V. Scarpino. Evolutionary medicine IV. evolution and emergence of novel pathogens. In *Encyclopedia of Evolutionary Biology*, pages 77–82. Elsevier BV, 2016.
- [SKBD13] Tanja Stadler, Denise Kühnert, Sebastian Bonhoeffer, and Alexei J. Drummond. Birth-death skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proceedings of the National Academy of Sciences*, 110(1):228–233, 2013.
- [SKRdP14] Tanja Stadler, Denise Kühnert, David A. Rasmussen, and Louis du Plessis. Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data. *PLoS Currents*, 2014.
- [SKvW<sup>+</sup>12] Tanja Stadler, Roger Kouyos, Viktor von Wyl, Sabine Yerly, Jürg Böni, Philippe Bürgisser, Thomas Klimkait, Beda Joos, Philip Rieder, Dong Xie, Huldrych F. Günthard, Alexei J. Drummond, Sebastian Bonhoeffer, and the Swiss HIV Cohort Study. Estimating the Basic Reproductive Number from

- Viral Sequence Data. *Molecular Biology and Evolution*, 29(1):347–357, January 2012.
- [Sta09] Tanja Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58 – 66, 2009.
- [Sta11] Tanja Stadler. Inferring speciation and extinction processes from extant species data. *Proceedings of the National Academy of Sciences*, 108(39):16145–16146, September 2011.
- [TFLS06] Mark M. Tanaka, Andrew R. Francis, Fabio Luciani, and S. A. Sisson. Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, 173(3):1511–1520, 2006.
- [Tho75] Elizabeth A. Thompson. *Human Evolutionary Trees*. Cambridge University Press, 1975.
- [VKB13] Erik M. Volz, Katia Koelle, and Trevor Bedford. Viral phylodynamics. *PLoS Comput Biol*, 9(3):e1002947, 03 2013.
- [VPW<sup>+</sup>09] Erik M. Volz, Sergei L. Kosakovsky Pond, M. J. Ward, Andrew J. Leigh Brown, and Simon D. W. Frost. Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430, 2009.
- [WDD07] Nathan D. Wolfe, Claire P. Dunavan, and Jared Diamond. Origins of major human infectious diseases. *Nature*, 447(7142):279–83, May 2007.
- [Wil74] David Williams. Path decomposition and continuity of local time for one-dimensional diffusions. I. *Proc. London Math. Soc. (3)*, 28:738–768, 1974.
- [Zay96] Ahmed I. Zayed. *Handbook of function and generalized function transformations*. Mathematical Sciences Reference Series. CRC Press, Boca Raton, FL, 1996.