

# On branching processes with rare neutral mutations

Airam Blancas Benítez

Advisor: Víctor Rivero

CIMAT, Guanajuato, México - SMILE, Paris 6

April 7, 2014

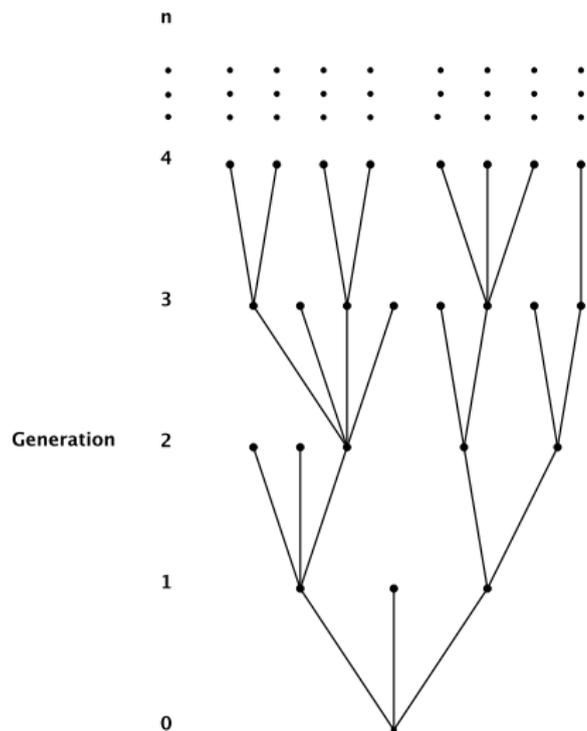
## 1 Preliminaries

- 1 Preliminaries
- 2 The process conditioned to non-extinction

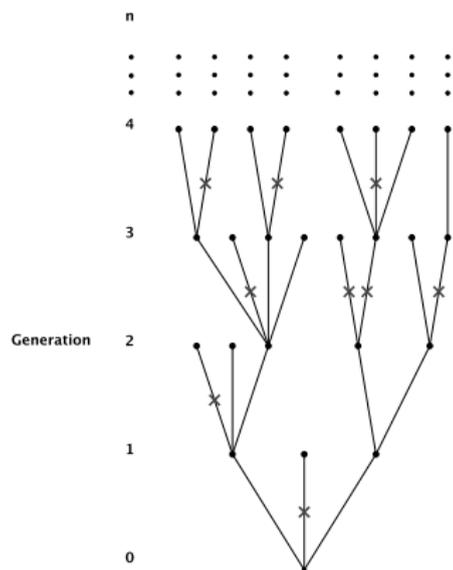
- 1 Preliminaries
- 2 The process conditioned to non-extinction
- 3 Asymptotic behavior

- 1 Preliminaries
- 2 The process conditioned to non-extinction
- 3 Asymptotic behavior

# Galton-Watson process

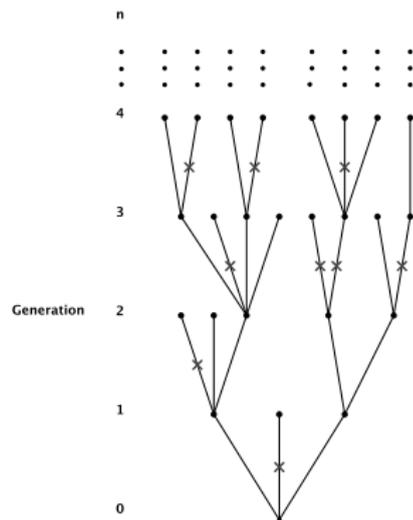


# Galton-Watson process with neutral mutations



- The offspring distribution is  $\xi^{(+)} := \xi^{(c)} + \xi^{(m)}$ .
- The population has infinitely many alleles.

# Galton-Watson process with neutral mutations

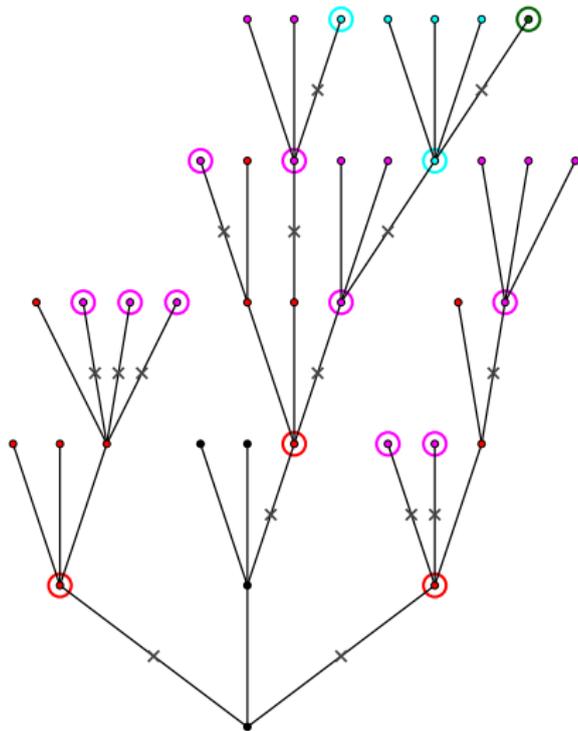


## Notation

$T_n$  the total population of individuals of the  $n$ -th type.

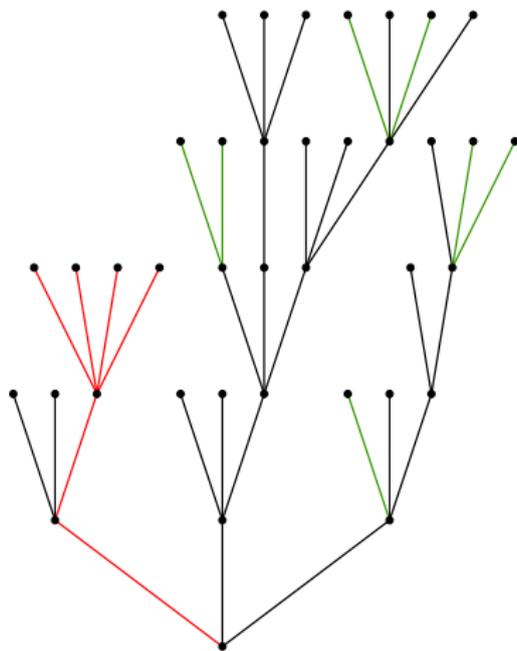
$M_n$  the total number of mutants of  $n$ -th type.

# An example

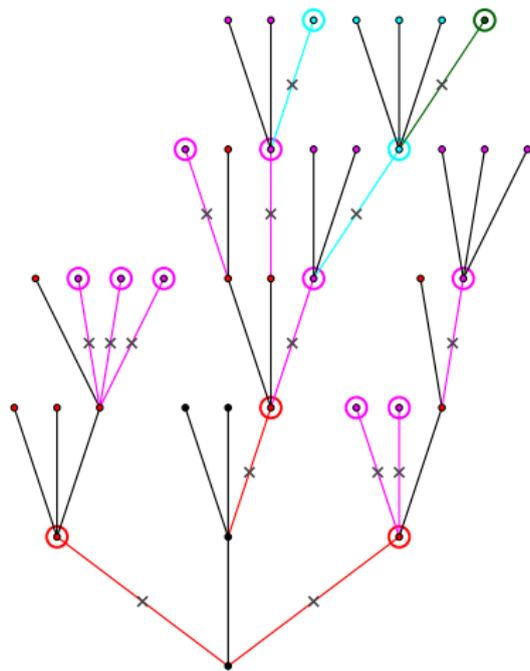


## Basic definitions

*Line* is a family of edges such that every branch from the root contains at most one edge in that family.



*Stopping line* is a random line such that for every edge in the tree, the event that this edge is part of the line only depends on the marks found on the path from the root to that edge.



## The general branching property

"Conditionally on the set of children of a stopping line, the families that those beget are independent copies of the initial tree".

## The general branching property

"Conditionally on the set of children of a stopping line, the families that those beget are independent copies of the initial tree".

## Lemma (Bertoin [3])

Under  $\mathbb{P}_a$

$$\{M_n : n \in \mathbb{Z}\}$$

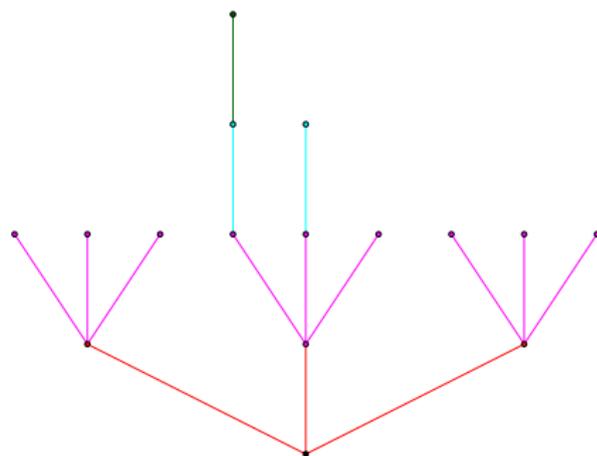
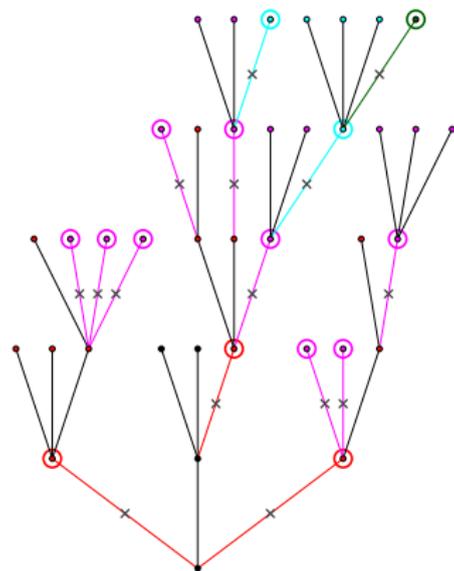
is a Galton-Watson process with reproduction law  $\mathbb{P}_1(M_1 \in \cdot)$ . More generally,

$$\{(T_n, M_{n+1}) : n \in \mathbb{Z}\}$$

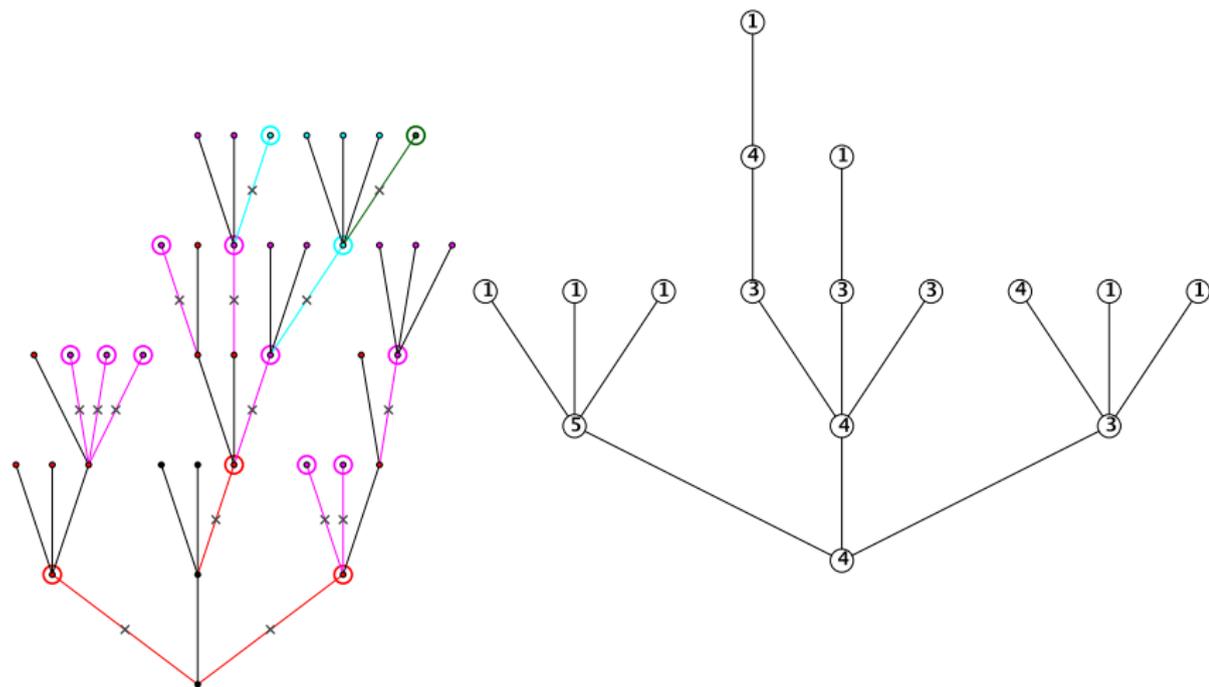
is a Markov chain, with transition probabilities

$$\mathbb{P}_a(T_n = k, M_{n+1} = l \mid T_{n-1} = i, M_n = j) = \mathbb{P}_j(T_0 = k, M_1 = l).$$

# Tree of alleles



# Tree of alleles



The *tree of alleles* is a process  $\mathcal{A} = (\mathcal{A}_u : u \in \mathbb{U})$  indexed in

$$\mathbb{U} := \bigcup_{k \in \mathbb{Z}_+} \mathbb{N}^k,$$

where  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}^0 = \{\emptyset\}$ .

### Construction

- $\mathcal{A}_\emptyset = T_0$ ,
- $\mathcal{A}_{uj}$  = The size of the  $j$ -th allelic sub-population of the type  $|u| + 1$  which descend from the allelic sub-family indexed by the vertex  $u$ , where  $|\cdot|$  denotes the level of the vertex.

### Remark

$$T_k = \sum_{|u|=k} \mathcal{A}_u \quad \text{and} \quad M_{k+1} = \sum_{|u|=k} d_u,$$

where  $d_u := \max\{j \geq 1 : \mathcal{A}_{uj} > 0\}$  agreeing that  $\max \emptyset = 0$ .

## Lemma (Bertoin [3])

For every integer  $a \geq 1$  and  $k \geq 0$ , the tree of alleles fulfills the following properties under  $\mathbb{P}_{a(n)}^{(n)}$  conditionally on  $((\mathcal{A}_u, d_u) : |u| \leq k)$

- i)  $((\mathcal{A}_{uj}, d_{uj}) : 1 \leq j \leq d_u)$ ,  $u$  vertex at level  $k$  such that  $\mathcal{A}_u > 0$ , are independent.
- ii) for each vertex  $u$  at level  $k$  with  $\mathcal{A}_u > 0$ , the  $d_u$ -tuple  $((\mathcal{A}_{uj}, d_{uj}) : 1 \leq j \leq d_u)$  is distributed as  $(T_0, M_1)^{(d_u \downarrow)}$  under  $\mathbb{P}_1$ .

The notation  $(d_u \downarrow)$  means that we rearranged the  $d_u$ -tuple in the decreasing order of the first coordinate, with the convention that in the case of ties, the coordinates are ranked uniformly at random.

- 1 Preliminaries
- 2 The process conditioned to non-extinction
- 3 Asymptotic behavior

## Assumptions

- $T = \inf\{n \geq 1 : M_n = 0\} < \infty, \quad \mathbb{P}_a\text{-c.s.}$

## Assumptions

- $T = \inf\{n \geq 1 : M_n = 0\} < \infty$ ,  $\mathbb{P}_a$ -c.s.
- $\mathbb{P}(M_1 = 1) > 0$ ,
- $\mathbb{P}(M_1 = 0) + \mathbb{P}(M_1 = 1) < 1$ ,
- $\mathbb{P}(M_1 = j) \neq 1$ , for any  $j$ .

## Assumptions

- $T = \inf\{n \geq 1 : M_n = 0\} < \infty$ ,  $\mathbb{P}_a$ -c.s.
- $\mathbb{P}(M_1 = 1) > 0$ ,
- $\mathbb{P}(M_1 = 0) + \mathbb{P}(M_1 = 1) < 1$ ,
- $\mathbb{P}(M_1 = j) \neq 1$ , for any  $j$ .

## Generating function

$$\varphi_n(x, y) := \mathbb{E}_1(x^{T_{n-1}}y^{M_n}), \quad x, y \in [0, 1].$$

$$f_n(y) := \varphi_n(1, y), \quad y \in [0, 1].$$

## Assumptions

- $T = \inf\{n \geq 1 : M_n = 0\} < \infty$ ,  $\mathbb{P}_a$ -c.s.
- $\mathbb{P}(M_1 = 1) > 0$ ,
- $\mathbb{P}(M_1 = 0) + \mathbb{P}(M_1 = 1) < 1$ ,
- $\mathbb{P}(M_1 = j) \neq 1$ , for any  $j$ .

## Generating function

$$\varphi_n(x, y) := \mathbb{E}_1(x^{T_{n-1}} y^{M_n}), \quad x, y \in [0, 1].$$

$$f_n(y) := \varphi_n(1, y), \quad y \in [0, 1].$$

## Remark

$$\varphi_n(x, y) = f_{n-1}(\varphi(x, y)), \quad x, y \in [0, 1],$$

$$\sum_{k, l=0}^{\infty} P_{(i,j),(k,l)}^n x^k y^l = (\varphi_n(x, y))^j, \quad i, j \geq 1.$$

## Theorem

Let  $a \in \mathbb{Z}_+$  and  $\mathcal{F}_n$  the natural filtration of the process  $\{(T_{n-1}, M_n) : n \in \mathbb{N}\}$ . There exists a probability measure  $\mathbb{P}_a^\uparrow$  that can be expressed as a  $h$ -transform of  $\mathbb{P}_a$  using the  $(\mathcal{F}_n)$ -martingale

$$Y_n = \frac{M_n q^{M_n - a}}{(f'(q))^n}.$$

where  $f(y) = \mathbb{E}_1(y^{M_1})$  and  $q$  denotes the extinction probability of  $\{M_n : n \in \mathbb{Z}\}$ . That is

$$d\mathbb{P}_a^\uparrow|_{\mathcal{F}_n} = \frac{Y_n}{a} d\mathbb{P}_a|_{\mathcal{F}_n}, \quad n \in \mathbb{N}.$$

Furthermore,  $\mathbb{P}_a^\uparrow$  is the law of a Markov chain  $\{(T_n^\uparrow, M_{n+1}^\uparrow), n \in \mathbb{Z}_+\}$  with  $n$ -step transition probabilities,

$$Q_{(i,j),(k,l)}^n = \frac{lq^{l-j}}{j(f'(q))^n} P_{(i,j),(k,l)}^n, \quad j, l \geq 1,$$

where  $\{P_{(i,j),(k,l)}^n : i, j, k, l \in \mathbb{Z}_+\}$  denotes the  $n$ -step transition probabilities of  $\{(T_n, M_{n+1}), n \in \mathbb{Z}_+\}$ .

## Theorem

Suppose that  $\mathbb{E}(\xi^{(+)}) \leq 1$ .

i) *The Yaglom limit*

$$\lim_{n \rightarrow \infty} \mathbb{P}(T_{n-1} = i, M_n = j | n < T < \infty)$$

exists and has a generating function  $\widehat{\varphi}(x, y)$  such that for all  $n \in \mathbb{N}$ ,

$$m^n \widehat{\varphi}(x, y) = \widehat{f}(\varphi_n(x, y)) - \widehat{f}(\varphi_n(x, 0)), \quad x, y \in [0, 1].$$

ii) *Let  $a \in \mathbb{Z}_+$  and  $n$  fixed. The conditional laws of the process*

*$\{(T_k, M_{k+1}) : 0 \leq k \leq n-1\}$  under  $\mathbb{P}_a(\cdot | n+k < T < \infty)$  converge to the  $k \rightarrow \infty$  to a limit probability measure  $\mathbb{P}_a^\uparrow$ , i.e. for any  $n \geq 0$*

$$\lim_{k \rightarrow \infty} \mathbb{P}_a(A | n+k < T < \infty) = \mathbb{P}_a^\uparrow(A), \quad \forall A \in \mathcal{F}_n.$$

## Proposition

The generating function of the  $n$ -step transition probabilities for the process

$$\{(T_n^\uparrow, M_{n+1}^\uparrow), n \in \mathbb{Z}_+\}$$

is given by

$$\sum_{k,l=1}^{\infty} Q_{(i,j),(k,l)}^n x^k y^l = \frac{yq^{1-j}}{[f'(q)]^n} [\varphi_n(x, qy)]^{j-1} \frac{\partial}{\partial y} \varphi(x, qy) \prod_{i=1}^{n-1} f'(\varphi_i(x, qy)).$$

## Proposition

The generating function of the  $n$ -step transition probabilities for the process

$$\{(T_n^\uparrow, M_{n+1}^\uparrow), n \in \mathbb{Z}_+\}$$

is given by

$$\sum_{k,l=1}^{\infty} Q_{(i,j),(k,l)}^n x^k y^l = \frac{yq^{1-j}}{[f'(q)]^n} [\varphi_n(x, qy)]^{j-1} \frac{\partial}{\partial y} \varphi(x, qy) \prod_{i=1}^{n-1} f'(\varphi_i(x, qy)).$$

## Corollary

If  $\{M_n, n \in \mathbb{Z}_+\}$  is critical or subcritical, then  $\{M_n^\uparrow - 1, n \in \mathbb{Z}_+\}$  is a Galton-Watson process with immigration  $[f, \frac{f'}{m}]$ .

- 1 Preliminaries
- 2 The process conditioned to non-extinction
- 3 Asymptotic behavior**

- For every  $n \in \mathbb{N}$ , let  $\{Z_k^{(n)} : k \in \mathbb{Z}_+\}$  a Galton Watson process such that
  - $Z_0^{(n)} = a(n)$  ancestors.
  - Reproduction law

$$\pi_k^+ = \mathbb{P}(\xi^{(+)} = k), \quad k \in \mathbb{Z}_+.$$

is critical and

$$\bar{\pi}^+(j) := \mathbb{P}(\xi^{(+)} > j) \in RV_{\infty}^{-\alpha}, \quad \alpha \in (1, 2), \quad (1)$$

where  $RV_{\infty}^{-\alpha}$  denotes the class of functions which are regularly varying with index  $-\alpha$  at  $\infty$ .

- For every  $n \in \mathbb{N}$ , let  $\{Z_k^{(n)} : k \in \mathbb{Z}_+\}$  a Galton Watson process such that
  - $Z_0^{(n)} = a(n)$  ancestors.
  - Reproduction law

$$\pi_k^+ = \mathbb{P}(\xi^{(+)} = k), \quad k \in \mathbb{Z}_+.$$

is critical and

$$\bar{\pi}^+(j) := \mathbb{P}(\xi^{(+)} > j) \in RV_{\infty}^{-\alpha}, \quad \alpha \in (1, 2), \quad (1)$$

where  $RV_{\infty}^{-\alpha}$  denotes the class of functions which are regularly varying with index  $-\alpha$  at  $\infty$ .

- Mutations affect each child according to a fixed probability  $p(n) \in (0, 1)$  and independently of the other children.

## Proposition

If condition (1) holds, then there exists  $r(n) \in RV_{\infty}^{-\alpha}$  such that

$$r(n)\pi^+(ndy) \xrightarrow{n \rightarrow \infty} c_{\alpha} \frac{dy}{y^{1+\alpha}},$$

where  $c_{\alpha}$  is a constant that depends on  $\alpha$ . In particular

$$\exp \left\{ -t \int_{[0, \infty)} (1 - e^{-\lambda y} - \lambda y) r(n)\pi^+(ndy) \right\} \xrightarrow{n \rightarrow \infty} e^{-t\lambda^{\alpha}}.$$

For every  $n \in \mathbb{N}$ , let  $\{Z_k^{(n)} : k \in \mathbb{Z}_+\}$  a Galton Watson process.

- $Z_0^{(n)} = a(n)$  ancestors.
- Reproduction law  $\rho^{(n)}$ .

Assume that

$$\lim_{n \rightarrow \infty} n^{-1}a(n) = x \quad \text{and} \quad \lim_{n \rightarrow \infty} n\bar{\rho}^{(n)}(ny) = \bar{\nu}(y), \quad (2)$$

where  $\nu$  is a measure on  $(0, \infty)$  with  $\int(1 \wedge y)\nu(dy) < \infty$ . Then

$$n^{-1}Z_1^{(n)} \Longrightarrow Z_1,$$

with  $Z_1$  is a random variable with values in  $[0, \infty)$  which is infinitely divisible.

More generally, an application of the Markov property shows

$$\{n^{-1}Z_k^{(n)} : k \in \mathbb{Z}_+\} \Longrightarrow \{Z_k : k \in \mathbb{Z}_+\},$$

where  $\{Z_k : k \in \mathbb{Z}_+\}$  is a (discrete time) continuous state branching process, in short, CSBP, with reproduction measure  $\nu$  and started from  $x$ .

Thanks to the Lévy Itô decomposition,

$$Z_1 = \sum_{i=1}^{\infty} a_i$$

where  $a_1 \geq a_2 \geq \dots$  are the atoms ranked in decreasing order of a Poisson random measure on  $(0, \infty)$  with intensity  $x\nu$ , with the convention that atoms are repeated according to their multiplicity.

Thanks to the Lévy Itô decomposition,

$$Z_1 = \sum_{i=1}^{\infty} a_i$$

where  $a_1 \geq a_2 \geq \dots$  are the atoms ranked in decreasing order of a Poisson random measure on  $(0, \infty)$  with intensity  $x\nu$ , with the convention that atoms are repeated according to their multiplicity. Besides,

- $\{\xi_i^n : 1 \leq i \leq a(n)\}$  i.i.d. variables with law  $\rho^{(n)}$ .
- $\{a_i^{(n)} : 1 \leq i \leq a(n)\}$  the decreasing reordering of the rescaled variables  $\{n^{-1}\xi_i^n : 1 \leq i \leq a(n)\}$ .

Assuming (2),

$$(a_1^{(n)}, a_2^{(n)}, \dots, a_{a(n)}^{(n)}) \implies (a_1, a_2, \dots).$$

in the sense of finite dimensional distributions.

## Definition (Bertoin [3])

Fix  $x > 0$  and  $\nu$  a measure on  $(0, \infty)$  with  $\int (1 \wedge y)\nu(dy) < \infty$ . A tree-indexed CSBP with reproduction measure  $\nu$  and initial population size  $x$  is a process  $\{\mathcal{Z}_u : u \in \mathbb{U}\}$  with values in  $\mathbb{R}_+$  whose distribution is characterized by induction on the levels as follows:

- i)  $\mathcal{Z}_\emptyset = x$ , a.s.;
- ii) for every  $k \in \mathbb{Z}_+$ , conditionally on  $\{\mathcal{Z}_v : v \in \mathbb{U}, |v| \leq k\}$ , the sequences  $\{\mathcal{Z}_{uj} : j \in \mathbb{N}\}$  for the vertices  $u \in \mathbb{U}$  at generation  $|u| = k$  are independent, and each sequence  $\{\mathcal{Z}_{uj} : j \in \mathbb{N}\}$  is distributed as the family of the atoms of a Poisson random measure on  $(0, \infty)$  with intensity  $\mathcal{Z}_u\nu$ , where the atoms are repeated according to their multiplicity, ranked in the decreasing order, and completed by an infinite sequence of 0 if the Poisson measure is finite.

## Theorem

Assuming (1),

$$a(n) \sim nx \quad \text{and} \quad p(n) \sim cn^{-1}, \quad \text{as } n \rightarrow \infty. \quad (3)$$

The rescaled tree of alleles  $(r(n))^{-1}\mathcal{A}^{(n)}$  converges in the sense of finite dimensional distributions towards a process  $\{\mathcal{Z}_u^{1/\alpha} : u \in \mathbb{U}\}$  called tree-indexed CSBP with reproduction measure

$$\nu^\alpha(dx) = c_\alpha x^{-1-1/\alpha} dx, \quad x > 0, \alpha \in (1, 2). \quad (4)$$

More precisely, we have the joint convergence in the sense of finite dimensional distributions

$$\mathcal{L} \left( \left( (r(n))^{-1}\mathcal{A}_u, (r(n)p(n))^{-1}d_u : k \in \mathbb{U} \right), \mathbb{P}_{a(n)}^{p(n)} \right) \Longrightarrow \left( (\mathcal{Z}_u^{1/\alpha}, \mathcal{Z}_u^{1/\alpha}) : u \in \mathbb{U} \right).$$



A. Lambert. Quasi-stationary distributions and continuous-state branching process conditioned to be never extinct. Elec. J. Prob. 12 420-446 (2007).



C.M. Goldie, J.L. Teugeles and N.H. Bingham. Regular Variation. Cambridge, Cambridge University Press, 1987.



J. Bertoin. A limit theorem for trees of alleles in branching processes with rare neutral mutations. Stochastic Processes and their Applications 120 (2010) 678-697.



K.B. Athreya and P.E. Ney. Branching Processes. Springer-Verlag, New York-Heidelberg, 1972.



K. Kawazu and S. Watanabe. Branching processes with immigration and related limit theorems. (1971)