Electron. Commun. Probab. **23** (2018), no. 42, 1–12.
https://doi.org/10.1214/18-ECP143
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A 2-spine decomposition of the critical Galton-Watson tree and a probabilistic proof of Yaglom's theorem

Yan-Xia Ren * Renming Song † Zhenyao Sun ‡

Abstract

In this note we propose a two-spine decomposition of the critical Galton-Watson tree and use this decomposition to give a probabilistic proof of Yaglom's theorem.

Keywords: Galton-Watson process; Galton-Watson tree; spine decomposition; Yaglom's theorem; martingale change of measure.
AMS MSC 2010: 60J80; 60F05.
Submitted to ECP on March 7, 2018, final version accepted on June 17, 2018.
Supersedes arXiv:1706.07125.

1 Introduction

1.1 Model

Consider a critical Galton-Watson process $(Z_n)_{n\geq 0}$ with $Z_0 = 1$ and offspring distribution μ on $\mathbb{N}_0 := \{0, 1, \dots\}$ which has mean 1 and finite variance $\sigma^2 > 0$, i.e.,

$$\sum_{k=0}^{\infty} k\mu(k) = 1 \tag{1.1}$$

and

$$0 < \sigma^2 := \sum_{k=0}^{\infty}(k-1)^2\mu(k) = \sum_{k=0}^{\infty} k(k-1)\mu(k) < \infty. \tag{1.2}$$

For simplicity, we will refer to $(Z_n)_{n\geq 0}$ as a *μ-Galton-Watson process*. It is well known that

Theorem 1.1 ([5]). *For a μ-Galton-Watson process $(Z_n)_{n\geq 0}$ satisfying* (1.1) *and* (1.2)*, we have*

1. $nP(Z_n > 0) \xrightarrow[n\to\infty]{} 2/\sigma^2$;
2. $\{n^{-1}Z_n; P(\cdot|Z_n > 0)\} \xrightarrow[n\to\infty]{d} Y$,

where Y is an exponential random variable with mean $\sigma^2/2$.

*Peking University, China. E-mail: yxren@math.pku.edu.cn. The research of Yan-Xia Ren is supported in part by NSFC (Grant Nos. 11671017 and 11731009), and LMEQF.
†University of Illinois at Urbana-Champaign, USA. E-mail: rsong@illinois.edu. The research of Renming Song is supported in part by the Simons Foundation (#429343, Renming Song).
‡Peking University, China. E-mail: zhenyao.sun@pku.edu.cn. Zhenyao Sun is supported by the China Scholarship Council. Corresponding author.

Under a third moment assumption, assertions (1) and (2) of Theorem 1.1 are due to [6] and [10] respectively. Theorem 1.1(2) is usually called Yaglom's theorem. For probabilistic proofs of the above results, we refer our readers to [2], [3] and [7].

In [7], Lyons, Pemantle and Peres gave a probabilistic proof of Theorem 1.1 using the so-called size-biased μ-Galton-Watson tree. In this note, by *size-biased transform* we mean the following: Let X be a random variable and $g(X)$ be a Borel function of X with $P(g(X) \geq 0) = 1$ and $E[g(X)] \in (0, \infty)$. We say a random variable W is a $g(X)$-size-biased transform (or simply $g(X)$-transform) of X if

$$E[f(W)] = \frac{E[g(X)f(X)]}{E[g(X)]}$$

for each positive Borel function f. An X-transform of X is sometimes called a size-biased transform of X.

We now recall the size-biased μ-Galton-Watson tree introduced in [7]. Let L be a random variable with distribution μ. Denote by $\dot{L}$ an L*-transform* of L. The celebrated *size-biased* μ*-Galton-Watson tree* is then constructed as follows:

- There is an initial particle which is marked.
- Any marked particle gives independent birth to a random number of children according to $\dot{L}$. Pick one of those children randomly as the new marked particle while leaving the other children as unmarked particles.
- Any unmarked particle gives birth independently to a random number of unmarked children according to L.
- The evolution goes on.

Notice that the marked particles form a descending family line which will be referred to as the *spine*. Define $\dot{Z}_n$ as the population of the nth generation in the size-biased tree. It is proved in [7] that the process $(\dot{Z}_n)_{n\geq 0}$ is a martingale transform of the process $(Z_n)_{n\geq 0}$ via the martingale $(Z_n)_{n\geq 0}$. That is, for any generation number n and any bounded Borel function g on $\mathbb{N}_0^n$,

$$E[g(\dot{Z}_1, \dots, \dot{Z}_n)] = \frac{E[Z_n g(Z_1, \dots, Z_n)]}{E[Z_n]}. \tag{1.3}$$

It is natural to consider probabilistic proofs of analogous results of Theorem 1.1 for more general critical branching processes. Vatutin and Dyakonova [9] gave a probabilistic proof of Theorem 1.1(1) for multitype critical branching processes. As far as we know, there is no probabilistic proof of Yaglom's theorem for multitype critical branching processes. It seems that it is difficult to adapt the probabilistic proofs in [3] and [7] for monotype branching processes to more general models, such as multitype branching processes, branching Hunt processes and superprocesses.

In this note, we propose a $k(k-1)$-type size-biased μ-Galton-Watson tree equipped with a two-spine skeleton, which serves as a change-of-measure of the original μ-Galton-Watson tree; and with the help of this two-spine technique, we give a new probabilistic proof of Theorem 1.1(2), i.e. Yaglom's theorem. The main motivation for developing this new proof for the classical Yaglom's theorem is that this new method is generic, in the sense that it can be generalized to more complicated critical branching systems. In fact, in our follow-up paper [8], we show that, in a similar spirit, a two-spine structure can be constructed for a class of critical superprocesses, and a probabilistic proof of a Yaglom type theorem can be obtained for those processes.

Another aspect of our new proof is that we take advantage of a fact that the exponential distribution can be characterized by a particular x^2-type size-biased distributional equation. An intuitive explanation of our method, and a comparison with the methods of [3] and [7], are made in the next subsection. We think this new point of view of convergence to the exponential law provides an alternative insight on the classical Yaglom's theorem.

We now give a formal construction of our $k(k-1)$-type size-biased μ-Galton-Watson tree. Denote by $\dot{L}$ an L-transform of L, and by $\ddot{L}$ an $L(L-1)$-transform of L. Fix a generation number n and pick a random generation number K_n uniformly among $\{0,\dots,n-1\}$. The *$k(k-1)$-type size-biased μ-Galton-Watson tree with height n* is then defined as a particle system such that:

- There is an initial particle which is marked.
- Before or after generation K_n, any marked particle gives birth independently to a random number of children according to $\dot{L}$. Pick one of those children randomly as the new marked particle while leaving the other children as unmarked particles.
- The marked particle at generation K_n, however, gives birth, independent of other particles, to a random number of children according to $\ddot{L}$. Pick two different particles randomly among those children as the new marked particles while leaving the other children as unmarked particles.
- Any unmarked particle gives birth independently to a random number of unmarked children according to L.
- The system stops at generation n.

If we track all the marked particles, it is clear that they form a *two-spine skeleton* with K_n being the last generation where those two spines are together. It would be helpful to consider this skeleton as two disjoint spines, where *the longer spine* is a family line from generation 0 to n and *the shorter spine* is a family line from generation K_n+1 to n.

For any $0\le m\le n$, denote by $\ddot{Z}_m^{(n)}$ the population of the mth generation in the $k(k-1)$-type size-biased μ-Galton-Watson tree with height n. The main reason for proposing such a model is that the process $(\ddot{Z}_m^{(n)})_{0\le m\le n}$ can be viewed as a $Z_n(Z_n-1)$-transform of the process $(Z_m)_{0\le m\le n}$. This is made precise in the result below which will be proved in Section 2.1.

Theorem 1.2. *Let $(Z_m)_{m\ge 0}$ be a μ-Galton-Watson process and $(\ddot{Z}_m^{(n)})_{0\le m\le n}$ be the population of a $k(k-1)$-type size-biased μ-Galton-Watson tree with height n. Suppose that μ satisfies* (1.1) *and* (1.2)*. Then, for any bounded Borel function g on $\mathbb{N}_0^n$,*

$$E[g(\ddot{Z}_1^{(n)},\dots,\ddot{Z}_n^{(n)})]=\frac{E[Z_n(Z_n-1)g(Z_1,\dots,Z_n)]}{E[Z_n(Z_n-1)]}.$$

The idea of considering a branching particle system with more than one spine is not new. A particle system with k spines was constructed in [4] and used in the many-to-few formula for branching Markov processes and branching random walks. Inspired by [4], we use a two-spine model to characterize the $k(k-1)$-type size-biased branching process.

1.2 Methods

Suppose that X is a non-negative random variable with $E[X]\in(0,\infty)$. Then its distribution conditioned on $\{X>0\}$ can be characterized by its conditional expectation $E[X|X>0]$ and its size-biased transform $\dot{X}$. In fact, for each $\lambda\ge 0$,

$$\begin{aligned}E[1-e^{-\lambda X}|X>0]&=\frac{E[1-e^{-\lambda X}]}{P(X>0)}\\&=\frac{1}{P(X>0)}\int_0^\lambda E[Xe^{-sX}]ds=E[X|X>0]\int_0^\lambda E[e^{-s\dot{X}}]ds.\end{aligned}\tag{1.4}$$

As a consequence, Theorem 1.1 is equivalent to

$$E\Big[\frac{Z_n}{n}|Z_n>0\Big] \xrightarrow[n\to\infty]{} \frac{\sigma^2}{2} \tag{1.5}$$

and

$$E[e^{-s\frac{\dot Z_n}{n}}] \xrightarrow[n\to\infty]{} E[e^{-s\dot Y}], \tag{1.6}$$

where $\dot Y$ is a Y-transform of the exponential random variable Y. Indeed, since $E[Z_n]=1$, (1.5) is equivalent to Theorem 1.1(1); and assuming (1.5), according to (1.4), we can see that (1.6) is equivalent to Theorem 1.1(2). In Section 3, for completeness, we will simplify the argument of [2] and [9], and give a proof of Theorem 1.1(1).

Our method of proving (1.6) takes advantage of a fact that the exponential distribution is characterized by an x^2-type size-biased distributional equation. This is made precise in the next lemma, which will be proved in Section 3:

Lemma 1.3. *Let Y be a strictly positive random variable with finite second moment. Then Y is exponentially distributed if and only if*

$$\ddot Y \overset{d}{=} \dot Y + U\cdot \dot Y', \tag{1.7}$$

where $\dot Y$ and $\dot Y'$ are both Y-transforms of Y, $\ddot Y$ is a Y^2-transform of Y, U is a uniform random variable on $[0,1]$, and $\dot Y$, $\dot Y'$ and U are independent.

With this lemma and Theorem 1.2, we can give an intuitive explanation of the exponential convergence in Yaglom's Theorem. From the construction of the $k(k-1)$-type size-biased μ-Galton-Watson tree $(\ddot Z^{(n)}_m)_{0\le m\le n}$, we see that the population $\ddot Z^{(n)}_n$ in the nth generation can be separated into two parts: descendants from the longer spine and descendants from the shorter spine. Due to their construction, the first part, the descendants from the longer spine at generation n, is distributed approximately like $\dot Z_n$, while the second part, the descendants from the shorter spine at generation n, is distributed approximately like $\dot Z_{\lfloor U\cdot n\rfloor}$. Those two parts are approximately independent of each other. So, after a renormalization, we have roughly that

$$\frac{\ddot Z^{(n)}_n}{n} \overset{d}{\approx} \frac{\dot Z_n}{n} + U\cdot \frac{\dot Z'_{\lfloor Un\rfloor}}{Un}, \tag{1.8}$$

where the process $(\dot Z'_m)$ is an independent copy of $(\dot Z_m)$. Suppose that $\dot Z_n/n$ converges weakly to a random variable $\dot Y$, and $\ddot Z_n/n$ converges weakly to a random variable $\ddot Y$. Then, according to [7, Lemma 4.3], $\ddot Y$ is a size-biased transform of $\dot Y$. Therefore, letting $n\to\infty$ in (1.8), $\dot Y$ should satisfy (1.7), which, by Lemma 1.3, suggests that (1.6) is true.

It is interesting to compare this method of proving exponential convergence with the methods used in [3] and [7]. In [7], Lyons, Pemantle and Peres characterize the exponential distribution by a different but well-known x-type size-biased distributional equation: A nonnegative random variable Y with positive finite mean is exponentially distributed if and only if it satisfies that

$$Y \overset{d}{=} U\cdot \dot Y \tag{1.9}$$

where $\dot Y$ is a Y-transform of Y, and U is a uniform random variable on $[0,1]$, which is independent of $\dot Y$. With the help of the size-biased tree, they then show that $\lceil U\cdot \dot Z_n\rceil$ is distributed approximately like Z_n conditioned on $\{Z_n>0\}$. So, after a renormalization, they have roughly that

$$\Big\{\frac{Z_n}{n}; P(\cdot|Z_n>0)\Big\} \overset{d}{\approx} U\cdot \frac{\dot Z_n}{n}. \tag{1.10}$$

Suppose that $\{Z_n/n; P(\cdot|Z_n > 0)\}$ converge weakly to a random variable Y, and $\dot{Z}_n/n$ converge weakly to a random variable $\dot{Y}$. Then, according to [7, Lemma 4.3], $\dot{Y}$ is the size-biased transform of Y. Therefore, letting $n \to \infty$ in (1.10), Y should satisfy (1.9), which suggests that Y is exponentially distributed.

In [3], Geiger characterizes the exponential distribution by another distributional equation: If $Y^{(1)}$ and $Y^{(2)}$ are independent copies of a random variable Y with positive finite variance, and U is an independent uniform random variable on $[0, 1]$, then Y is exponentially distributed if and only if

$$Y \overset{d}{=} U(Y^{(1)} + Y^{(2)}). \tag{1.11}$$

Geiger then shows that for (Z_n), conditioned on non-extinction at generation n, the distribution of the generation of the most recent common ancestor (MRCA) of the particles at generation n is asymptotically uniform among $\{0, 1, \dots, n\}$ (a result due to [11], see also [2]), and there are asymptotically two children of the MRCA, each with at least 1 descendant in generation n. After a renormalization, roughly speaking, Geiger has that

$$\Big\{\frac{Z_n}{n}; P(\cdot|Z_n > 0)\Big\} \overset{d}{\approx} U \cdot \frac{Z^{(1)}_{\lfloor Un \rfloor}}{Un} + U \cdot \frac{Z^{(2)}_{\lfloor Un \rfloor}}{Un}, \tag{1.12}$$

where for each m, $Z^{(1)}_m$ and $Z^{(2)}_m$ are independent copies of $\{Z_m; P(\cdot|Z_m > 0)\}$. Therefore, if $\{Z_n/n; P(\cdot|Z_n > 0)\}$ converges weakly to a random variable Y, then Y should satisfy (1.11), which suggests that Y is exponentially distributed.

From this comparison, we see that all the methods mentioned above share one similarity: They all establish the exponential convergence via some particular distributional equation. However, since the equations (1.7), (1.9) and (1.11) are different, the actual way of proving the convergence varies. In [7], an elegant tightness argument is made along with (1.10). However, it seems that this tightness argument is not suitable for (1.12), due to a property that the conditional convergence for some subsequence Z_{n_k}/n_k implies the convergence of $U \cdot \dot{Z}_{n_k}/n_k$, but does not imply the convergence of $Z^{(i)}_{\lfloor Un_k \rfloor}/Un_k, i = 1, 2$. Instead, a contraction type argument in the L^2-Wasserstein metric is used in [3].

For similar reasons, in this note, to actually prove the exponential convergence using (1.8) and (1.7), some efforts also must be made. We observe that the distributional equation (1.8) admits a so-called size-biased add-on structure, which is related to Lèvy's theory of infinitely divisible distributions: Suppose that X is a nonnegative random variable with $a := E[X] \in (0, \infty)$; then X is infinitely divisible if and only if there exists a nonnegative random variable A independent of X such that $\dot{X} \overset{d}{=} X + A$. In fact, the Laplace exponent of X can be expressed as

$$-\ln E[e^{-\lambda X}] = a\alpha(\{0\})\lambda + a \int_{(0,\infty)} \frac{1 - e^{-\lambda y}}{y} \alpha(dy),$$

where α is the distribution of A. Moreover, if A is strictly positive, then

$$-\ln E[e^{-\lambda X}] = a \int_0^\lambda E[e^{-sA}] ds.$$

From this point of view, after considering the Laplace transforms of (1.8) and (1.7), we can establish the convergence of $E[e^{-\lambda \dot{Z}_n/n}]$ to $E[e^{-\lambda \dot{Y}}]$, which will eventually lead us to Yaglom's theorem. This is made precise in Section 3. A similar type of argument is also used in our follow-up paper [8] for critical superprocesses.

2 Trees and their decompositions

2.1 Spaces and measures

In this subsection, we give a proof of Theorem 1.2. Consider *particles* as elements in the space

$$\mathcal{U} := \{\emptyset\} \cup \bigcup_{k=1}^{\infty} \mathbb{N}^k,$$

where $\mathbb{N} := \{1, 2, \dots\}$. Therefore elements in $\mathcal{U}$ are of the form 213, which we read as the individual being the 3rd child of the 1st child of the 2nd child of the initial ancestor $\emptyset$. For two particles $u = u_1 \dots u_n, v = v_1 \dots v_m \in \mathcal{U}$, uv denotes the concatenated particle $uv := u_1 \dots u_n v_1 \dots v_m$. We use the convention $u\emptyset = \emptyset u = u$ and $u_1 \dots u_n = \emptyset$ if $n = 0$. For any particle $u := u_1 \dots u_{n-1} u_n$, we define its *generation* as $|u| := n$ and its *parent particle* as $\overleftarrow{u} := u_1 \dots u_{n-1}$. For any particle $u \in \mathcal{U}$ and any subset $\mathbf{a} \subset \mathcal{U}$, we define the *number of children of* u *in* $\mathbf{a}$ as $l_u(\mathbf{a}) := \#\{\alpha \in \mathbf{a} : \overleftarrow{\alpha} = u\}$. We also define the *height* of $\mathbf{a}$ as $|\mathbf{a}| := \sup_{\alpha \in \mathbf{a}} |\alpha|$ and its *population in the* n*th generation* as $X_n(\mathbf{a}) := \#\{u \in \mathbf{a} : |u| = n\}$. A *tree* $\mathbf{t}$ is defined as a subset of $\mathcal{U}$ such that there exists an $\mathbb{N}_0$-valued sequence $(l_u)_{u \in \mathcal{U}}$, indexed by $\mathcal{U}$, satisfying

$$\mathbf{t} = \{u_1 \dots u_m \in \mathcal{U} : m \geq 0, u_j \leq l_{u_1 \dots u_{j-1}}, \forall j = 1, \dots, m\}.$$

A *spine* $\mathbf{v}$ on a tree $\mathbf{t}$ is defined as a sequence of particles $\{v^{(k)} : k = 0, 1, \dots, |\mathbf{t}|\} \subset \mathbf{t}$ such that $v^{(0)} = \emptyset$ and $\overleftarrow{v^{(k)}} = v^{(k-1)}$ for any $k = 1, \dots, |\mathbf{t}|$. In the case that $|\mathbf{t}| = \infty$, we simply write $k = 0, 1, \dots$ as $k = 0, 1, \dots, |\mathbf{t}|$.

Fix a generation number $n \in \mathbb{N}$. Define the following spaces:

- *The space of trees with height no more than* n,

$$\mathbb{T}_{\leq n} := \{\mathbf{t} : \mathbf{t} \text{ is a tree with } |\mathbf{t}| \leq n\}.$$

- *The space of* n*-height trees with one distinguishable spine*,

$$\dot{\mathbb{T}}_n := \{(\mathbf{t}, \mathbf{v}) : \mathbf{t} \text{ is a tree with } |\mathbf{t}| = n, \mathbf{v} \text{ is a spine on } \mathbf{t}\}.$$

- *The space of* n*-height trees with two different distinguishable spines*,

$$\ddot{\mathbb{T}}_n := \{(\mathbf{t}, \mathbf{v}, \mathbf{v}') : (\mathbf{t}, \mathbf{v}) \in \dot{\mathbb{T}}_n, (\mathbf{t}, \mathbf{v}') \in \dot{\mathbb{T}}_n, \mathbf{v} \neq \mathbf{v}'\}.$$

Let $(L_u)_{u \in \mathcal{U}}$ be a collection of independent random variables with law μ, indexed by $\mathcal{U}$. Denote by T the random tree defined by

$$T := \{u_1 \dots u_m \in \mathcal{U} : 0 \leq m \leq n, u_j \leq L_{u_1 \dots u_{j-1}}, \forall j = 1, \dots, m\}.$$

We refer to T as a *μ-Galton-Watson tree with height no more than n* since its population $(X_m(T))_{0 \leq m \leq n}$ is a μ-Galton-Watson process stopped at generation n. Define the *μ-Galton-Watson measure* $\mathbf{G}_n$ on $\mathbb{T}_{\leq n}$ as the law of the random tree T. That is, for any $\mathbf{t} \in \mathbb{T}_{\leq n}$,

$$\mathbf{G}_n(\mathbf{t}) := P(T = \mathbf{t}) = P(L_u = l_u(\mathbf{t}) \text{ for any } u \in \mathbf{t} \text{ with } |u| < n) = \prod_{u \in \mathbf{t} : |u| < n} \mu(l_u(\mathbf{t})).$$

Recall that $\dot{L}$ is an L-transform of L. Define $\dot{C}$ as a random number which, conditioned on $\dot{L}$, is uniformly distributed on $\{1, \dots, \dot{L}\}$. Independent of $(L_u)_{u \in \mathcal{U}}$, let $(\dot{L}_u, \dot{C}_u)_{u \in \mathcal{U}}$ be a collection of independent copies of $(\dot{L}, \dot{C})$, indexed by $\mathcal{U}$. We then use $(L_u)_{u \in \mathcal{U}}$ and

$(\dot{L}_u, \dot{C}_u)_{u\in\mathcal{U}}$ as the building blocks to construct the size-biased μ-Galton-Watson tree $\dot{T}$ and its distinguishable spine $\dot{V}$ following the steps described in Section 1.1. We use L_u as the number of children of particle u if u is unmarked and use $\dot{L}_u$ if u is marked. In the latter case, we always set the $\dot{C}_u$-th child of u, i.e. particle $u\dot{C}_u$, as the new marked particle. For convenience, we stop the system at generation n. To be precise, the random spine $\dot{V}$ is defined by

$$\dot{V} := \{v_1 \dots v_m \in \mathcal{U} : 0 \le m \le n, v_j = \dot{C}_{v_1 \dots v_{j-1}}, \forall j = 1, \dots, m\},$$

and the random tree $\dot{T}$ is defined by

$$\dot{T} := \{u_1 \dots u_m \in \mathcal{U} : 0 \le m \le n, u_j \le \tilde{L}_{u_1 \dots u_{j-1}}, \forall j = 1, \dots, m\},$$

where, for any $u \in \mathcal{U}$, $\tilde{L}_u := L_u \mathbf{1}_{u \notin \dot{V}} + \dot{L}_u \mathbf{1}_{u \in \dot{V}}$.

We now consider the distribution of the $\dot{\mathbb{T}}_n$-valued random element $(\dot{T}, \dot{V})$. For any $(\mathbf{t}, \mathbf{v}) \in \dot{\mathbb{T}}_n$, the event $\{(\dot{T}, \dot{V}) = (\mathbf{t}, \mathbf{v})\}$ occurs if and only if:

- $L_u = l_u(\mathbf{t})$ for each $u \in \mathbf{t} \setminus \mathbf{v}$ with $|u| < n$ and
- $(\dot{L}_{v_1 \dots v_m}, \dot{C}_{v_1 \dots v_m}) = (l_{v_1 \dots v_m}(\mathbf{t}), v_{m+1})$ for each $v_1 \dots v_{m+1} \in \mathbf{v}$ with $0 \le m \le n-1$.

Therefore, the distribution of $(\dot{T}, \dot{V})$ can be determined by

$$P((\dot{T}, \dot{V}) = (\mathbf{t}, \mathbf{v})) = \prod_{u \in \mathbf{t}\setminus\mathbf{v} : |u| < n} \mu(l_u(\mathbf{t})) \cdot \prod_{u \in \mathbf{v} : |u| < n} l_u(\mathbf{t}) \mu(l_u(\mathbf{t})) \frac{1}{l_u(\mathbf{t})} = \mathbf{G}_n(\mathbf{t}). \tag{2.1}$$

The *size-biased μ-Galton-Watson measure* $\dot{\mathbf{G}}_n$ on $\mathbb{T}_{\le n}$ is then defined as the law of the $\mathbb{T}_{\le n}$-valued random element $\dot{T}$. That is, for any $\mathbf{t} \in \mathbb{T}_{\le n}$,

$$\begin{aligned} \dot{\mathbf{G}}_n(\mathbf{t}) &:= P(\dot{T} = \mathbf{t}) = \sum_{\mathbf{v} : (\mathbf{t}, \mathbf{v}) \in \dot{\mathbb{T}}_n} P((\dot{T}, \dot{V}) = (\mathbf{t}, \mathbf{v})) \\ &= \#\{\mathbf{v} : (\mathbf{t}, \mathbf{v}) \in \dot{\mathbb{T}}_n\} \cdot \mathbf{G}_n(\mathbf{t}) = X_n(\mathbf{t}) \cdot \mathbf{G}_n(\mathbf{t}). \end{aligned} \tag{2.2}$$

Equations (2.1), (2.2) and their consequence (1.3) were first obtained in [7]. We use these equations to help us to understand how the $k(k-1)$-type size-biased μ-Galton-Watson tree can be represented.

Recall that K_n is a random generation number uniformly distributed on $\{0, \dots, n-1\}$, and $\ddot{L}$ is an $L(L-1)$-transform of L. Define $(\ddot{C}, \ddot{C}')$ as a random vector which, conditioned on $\ddot{L}$, is uniformly distributed on $\{(i,j) \in \mathbb{N}^2 : 1 \le i \ne j \le \ddot{L}\}$. Suppose that $(L_u)_{u\in\mathcal{U}}, (\dot{L}_u, \dot{C}_u)_{u\in\mathcal{U}}$, $(\ddot{L}, \ddot{C}, \ddot{C}')$ and K_n are independent of each other. We now use these elements to build the $k(k-1)$-type size-biased μ-Galton-Watson tree $\ddot{T}$ and its two different distinguishable spines $\ddot{V}$ and $\ddot{V}'$ following the steps described in Section 1.1. Write $C_u := \dot{C}_u \mathbf{1}_{|u| \ne K_n} + \ddot{C} \mathbf{1}_{|u| = K_n}$ and $C'_u := \dot{C}_u \mathbf{1}_{|u| \ne K_n} + \ddot{C}' \mathbf{1}_{|u| = K_n}$. We define the random spines $\ddot{V}$ and $\ddot{V}'$ as

$$\begin{aligned} \ddot{V} &:= \{v_1 \dots v_m \in \mathcal{U} : 0 \le m \le n, v_j = C_{v_1 \dots v_{j-1}}, \forall j = 1, \dots, m\}, \\ \ddot{V}' &:= \{v_1 \dots v_m \in \mathcal{U} : 0 \le m \le n, v_j = C'_{v_1 \dots v_{j-1}}, \forall j = 1, \dots, m\}, \end{aligned}$$

and the random tree $\ddot{T}$ as

$$\ddot{T} := \{u_1 \dots u_m \in \mathcal{U} : 0 \le m \le n, u_j \le L''_{u_1 \dots u_{j-1}}, \forall j = 1, \dots, m\},$$

where, for any $u \in \mathcal{U}$, $L''_u := L_u \mathbf{1}_{u \notin \ddot{V} \cup \ddot{V}'} + \dot{L}_u \mathbf{1}_{u \in \ddot{V} \cup \ddot{V}', |u| \ne K_n} + \ddot{L} \mathbf{1}_{u \in \ddot{V} \cup \ddot{V}', |u| = K_n}$.

We now consider the distribution of $(\ddot{T}, \ddot{V}, \ddot{V}')$. For any $(\mathbf{t}, \mathbf{v}, \mathbf{v}') \in \ddot{\mathbb{T}}_n$, the event $\{(\ddot{T}, \ddot{V}, \ddot{V}') = (\mathbf{t}, \mathbf{v}, \mathbf{v}')\}$ occurs if and only if:

- $K_n = k_n := |\mathbf{v} \cap \mathbf{v}'|$,
- $L_u = l_u(\mathbf{t})$ for each $u \in \mathbf{t} \setminus (\mathbf{v} \cup \mathbf{v}')$ with $|u| < n$,
- $(\dot{L}_{v_1 \dots v_m}, \dot{C}_{v_1 \dots v_m}) = (l_{v_1 \dots v_m}(\mathbf{t}), v_{m+1})$ for each $v_1 \dots v_m v_{m+1} \in \mathbf{v} \cup \mathbf{v}'$ with $k_n \neq m < n$ and
- $(\ddot{L}, \ddot{C}, \ddot{C}') = (l_{v_1 \dots v_{k_n}}(\mathbf{t}), v_{k_n+1}, v'_{k_n+1})$ for $v_1 \dots v_{k_n} v_{k_n+1} \in \mathbf{v}$ and $v_1 \dots v_{k_n} v'_{k_n+1} \in \mathbf{v}'$.

Using this analysis, we get that

$$\begin{aligned} P\big((\ddot{T}, \ddot{V}, \ddot{V}') = (\mathbf{t}, \mathbf{v}, \mathbf{v}')\big) &= \frac{1}{n} \cdot \prod_{u \in \mathbf{t} \setminus (\mathbf{v} \cup \mathbf{v}'): |u| < n} \mu(l_u(\mathbf{t})) \cdot \prod_{u \in \mathbf{v} \cup \mathbf{v}': k_n \neq |u| < n} l_u(\mathbf{t}) \mu(l_u(\mathbf{t})) \frac{1}{l_u(\mathbf{t})} \\ &\quad \cdot \prod_{u \in \mathbf{v} \cup \mathbf{v}': |u| = k_n} \frac{l_u(\mathbf{t})(l_u(\mathbf{t}) - 1)\mu(l_u(\mathbf{t}))}{\sigma^2} \frac{1}{l_u(\mathbf{t})(l_u(\mathbf{t}) - 1)} \\ &= \frac{1}{n\sigma^2} \mathbf{G}_n(\mathbf{t}). \end{aligned}$$

The $k(k-1)$*-type size-biased* μ*-Galton-Watson measure* $\ddot{\mathbf{G}}_n$ on $\mathbb{T}_{\le n}$ is then defined as the law of the random element $\ddot{T}$. That is, for any $\mathbf{t} \in \mathbb{T}_{\le n}$,

$$\begin{aligned} \ddot{\mathbf{G}}_n(\mathbf{t}) := P(\ddot{T} = \mathbf{t}) &= \sum_{(\mathbf{v}, \mathbf{v}'): (\mathbf{t}, \mathbf{v}, \mathbf{v}') \in \ddot{\mathbb{T}}_n} P\big((\ddot{T}, \ddot{V}, \ddot{V}') = (\mathbf{t}, \mathbf{v}, \mathbf{v}')\big) \\ &= \#\{(\mathbf{v}, \mathbf{v}') : (\mathbf{t}, \mathbf{v}, \mathbf{v}') \in \ddot{\mathbb{T}}_n\} \cdot \frac{\mathbf{G}_n(\mathbf{t})}{n\sigma^2} = \frac{X_n(\mathbf{t})(X_n(\mathbf{t}) - 1)}{n\sigma^2} \cdot \mathbf{G}_n(\mathbf{t}). \end{aligned} \tag{2.3}$$

We note in passing that, because of the way they are constructed, the measures $(\ddot{\mathbf{G}}_n)_{n \ge 1}$ are not consistent, that is, the measure $\ddot{\mathbf{G}}_n$ is not the restriction of $\ddot{\mathbf{G}}_{n+1}$. This implies that the change of measure in Theorem 1.2 is not a martingale change of measure.

Proof of Theorem 1.2. Note that

$$\{(X_m(\mathbf{t}))_{0 \le m \le n}; \mathbf{G}_n\} \overset{d}{=} (Z_m)_{0 \le m \le n} \quad \text{and} \quad \{(X_m(\mathbf{t}))_{0 \le m \le n}; \ddot{\mathbf{G}}_n\} \overset{d}{=} (\ddot{Z}_m)_{0 \le m \le n}.$$

According to (2.3), for any bounded Borel function g on $\mathbb{N}_0^n$, we can verify that

$$\begin{aligned} E[g(\ddot{Z}_1^{(n)}, \dots, \ddot{Z}_n^{(n)})] &= \ddot{\mathbf{G}}_n[g(X_1(\mathbf{t}), \dots, X_n(\mathbf{t}))] \\ &= \mathbf{G}_n\Big[\frac{X_n(\mathbf{t})(X_n(\mathbf{t}) - 1)}{n\sigma^2} g(X_1(\mathbf{t}), \dots, X_n(\mathbf{t}))\Big] \\ &= \frac{1}{n\sigma^2} E[Z_n(Z_n - 1) g(Z_1, \dots, Z_n)]. \end{aligned} \tag{2.4}$$

Taking $g \equiv 1$ in equation (2.4), we get that

$$E[Z_n(Z_n - 1)] = E[\dot{Z}_n - 1] = n\sigma^2. \tag{2.5}$$

□

2.2 Spine decompositions.

Using the notation introduced in the previous section, we are now ready to give a precise meaning to (1.8):

Proposition 2.1. *Let* $(\dot{Z}_m)_{0 \le m \le n}$ *be the population of a size-biased* μ*-Galton watson tree and* $(\ddot{Z}_m^{(n)})_{0 \le m \le n}$ *be the population of a* $k(k-1)$*-type size-biased* μ*-Galton-Watson tree with height* n*. Suppose that* μ *satisfies* (1.1) *and* (1.2)*. Then*

$$E[e^{-\lambda \ddot{Z}_n^{(n)}}] = E[e^{-\lambda \dot{Z}_n}] E[g(\lambda, \lfloor Un \rfloor) e^{-\lambda \dot{Z}_{\lfloor Un \rfloor}}],$$

where U *is a uniform random variable on* $[0, 1]$ *independent of* $\{\dot{Z}_m : 0 \le m \le n\}$*; and* $g(\lambda, m)$ *is a function on* $[0, \infty) \times \mathbb{N}_0$ *such that* $g(\lambda, m) \to 1$*, uniformly in* λ *as* $m \to \infty$*.*

Proof. For any particle $u = u_1 \dots u_n$, we define $[\emptyset, u] := \{u_1 \dots u_j : j = 0, \dots, n\}$ as the *descending family line from* $\emptyset$ *to* u. The particles in $\dot{T}$ can be separated according to their nearest spine ancestor. For each $k = 0, \dots, n$, we write $\dot{A}_k := \{u \in \dot{T} : |[\emptyset, u] \cap \dot{V}| = k\}$. Then

$$X_n(\dot{T}) = \sum_{k=0}^{n} X_n(\dot{A}_k). \tag{2.6}$$

Notice that the right side of the above equation is a sum of independent random variables; and from their construction, we see that $X_n(\dot{A}_k) \overset{d}{=} Z^{(\dot{L}-1)}_{n-k-1}$. Here, $Z^{(\dot{L}-1)}_{(-1)} := 1$ and $(Z^{(\dot{L}-1)}_m)_{m\in\mathbb{N}_0}$ denotes a μ-Galton-Watson process with $Z^{(\dot{L}-1)}_0$ distributed according to $\dot{L}-1$. Taking Laplace transforms on both sides of (2.6) we get

$$E[e^{-\lambda \dot{Z}_n}] = \prod_{k=0}^{n} E[e^{-\lambda Z^{(\dot{L}-1)}_{n-k-1}}]. \tag{2.7}$$

Similarly, we consider the $k(k-1)$-type size-biased μ-Galton-Watson tree $(\ddot{T}, \ddot{V}, \ddot{V}')$. Write

$$\ddot{A}^l_k := \{u \in \ddot{T} : |[\emptyset, u] \cap \ddot{V}| = k, [\emptyset, u] \cap (\ddot{V}' \setminus \ddot{V}) = \emptyset\}$$

and

$$\ddot{A}^s_k := \{u \in \ddot{T} : |[\emptyset, u] \cap \ddot{V}'| = k, [\emptyset, u] \cap (\ddot{V}' \setminus \ddot{V}) \neq \emptyset\}.$$

Then,

$$X_n(\ddot{T}) = \sum_{k=0}^{n} X_n(\ddot{A}^l_k) + \sum_{k=K_n+1}^{n} X_n(\ddot{A}^s_k). \tag{2.8}$$

Notice that, conditioning on $K_n = m$ with $m \in \{0, \dots, n-1\}$, the right side of the above equation is a sum of independent random variables; and from their construction, we see that $X_n(\ddot{A}^l_k) \overset{d}{=} Z^{(\dot{L}-1)}_{n-k-1}$ for each $k \neq m$; $X_n(\ddot{A}^l_m) \overset{d}{=} Z^{(\ddot{L}-2)}_{n-m-1}$; and $X_n(\ddot{A}^s_k) \overset{d}{=} Z^{(\dot{L}-1)}_{n-k-1}$ for each $k \geq m+1$. Here, $Z^{(\ddot{L}-2)}_{(-1)} := 1$ and $(Z^{(\ddot{L}-2)}_k)_{k\in\mathbb{N}_0}$ is a μ-Galton-Watson process with initial population distributed according to $\ddot{L}-2$.

Taking Laplace transform on both sides of (2.8) and using (2.7), we get

$$\begin{aligned}
E[e^{-\lambda \ddot{Z}^{(n)}_n}] &= \frac{1}{n}\sum_{m=0}^{n-1}\Big(\prod_{k=0,k\neq m}^{n} E[e^{-\lambda Z^{(\dot{L}-1)}_{n-k-1}}]\Big)\cdot E[e^{-\lambda Z^{(\ddot{L}-2)}_{n-m-1}}]\cdot\Big(\prod_{k=m+1}^{n} E[e^{-\lambda Z^{(\dot{L}-1)}_{n-k-1}}]\Big) \\
&= E[e^{-\lambda \dot{Z}_n}]\frac{1}{n}\sum_{m=0}^{n-1}\frac{E[e^{-\lambda Z^{(\ddot{L}-2)}_{n-m-1}}]}{E[e^{-\lambda Z^{(\dot{L}-1)}_{n-m-1}}]}\cdot E[e^{-\lambda \dot{Z}_{n-m-1}}] \\
&= E[e^{-\lambda \dot{Z}_n}]\frac{1}{n}\sum_{m=0}^{n-1}\frac{E[e^{-\lambda Z^{(\ddot{L}-2)}_{m}}]}{E[e^{-\lambda Z^{(\dot{L}-1)}_{m}}]}\cdot E[e^{-\lambda \dot{Z}_m}] = E[e^{-\lambda \dot{Z}_n}]E[g(\lambda, \lfloor Un\rfloor)e^{-\lambda \dot{Z}_{\lfloor Un\rfloor}}],
\end{aligned}$$

where

$$P(Z^{(\ddot{L}-2)}_m = 0) \leq g(\lambda, m) := \frac{E[e^{-\lambda Z^{(\ddot{L}-2)}_m}]}{E[e^{-\lambda Z^{(\dot{L}-1)}_m}]} \leq P(Z^{(\dot{L}-1)}_m = 0)^{-1}.$$

Notice that, from the criticality, $P(Z^{(\ddot{L}-2)}_m = 0)$ and $P(Z^{(\dot{L}-1)}_m = 0)^{-1}$ converge to 1. ☐

3 Proofs

Proof of Theorem 1.1(1). Denote by B^j_n the event that the Galton-Watson process $(Z_n)_{n\geq 0}$ survives up to generation n, and the left-most particle in the n-th generation is a descendant of the jth particle of the first generation. Write $q_n = P[Z_n = 0] = f^{(n)}(0)$ and

$p_n = 1 - q_n$ where f is the probability generating function of the offspring distribution μ. Then

$$\begin{aligned} E[Z_n|Z_n>0] &= \sum_{k=1}^{\infty} E[Z_n; Z_1=k|Z_n>0] = p_n^{-1}\sum_{k=1}^{\infty} E[Z_n; Z_1=k; Z_n>0] \\ &= p_n^{-1}\sum_{k=1}^{\infty}\sum_{j=1}^{k} E[Z_n; Z_1=k; B_n^j] = p_n^{-1}\sum_{k=1}^{\infty}\sum_{j=1}^{k} P[Z_1=k; B_n^j]E[Z_n|Z_1=k, B_n^j] \\ &= p_n^{-1}\sum_{k=1}^{\infty}\sum_{j=1}^{k} P[Z_1=k; B_n^j]\Big(E[Z_{n-1}|Z_{n-1}>0]+k-j\Big) \\ &= E[Z_{n-1}|Z_{n-1}>0] + \frac{p_{n-1}}{p_n}\sum_{k=1}^{\infty}\sum_{j=1}^{k}\mu(k)q_{n-1}^{j-1}(k-j). \end{aligned} \tag{3.1}$$

The criticality implies that $q_n \uparrow 1$ as $n \to \infty$, and that

$$\frac{p_n}{p_{n-1}} = \frac{1-f^{(n)}(0)}{1-f^{(n-1)}(0)} = \frac{1-f(q_{n-1})}{1-q_{n-1}} \xrightarrow[n\to\infty]{} f'(1) = 1.$$

By the monotone convergence theorem,

$$\frac{p_{n-1}}{p_n}\sum_{k=1}^{\infty}\sum_{j=1}^{k}\mu(k)q_{n-1}^{j-1}(k-j) \xrightarrow[n\to\infty]{} \sum_{k=1}^{\infty}\sum_{j=1}^{k}\mu(k)(k-j) = \sum_{k=1}^{\infty}\mu(k)k(k-1)/2 = \frac{\sigma^2}{2}.$$

Now combining (3.1) with the above, we get

$$\begin{aligned} \frac{1}{nP(Z_n>0)} &= \frac{1}{n}E[Z_n|Z_n>0] \\ &= \frac{1}{n}E[Z_0|Z_0>0] + \frac{1}{n}\sum_{m=1}^{n}\frac{p_{m-1}}{p_m}\sum_{k=1}^{\infty}\sum_{j=1}^{k}\mu(k)q_{m-1}^{j-1}(k-j) \\ &\xrightarrow[n\to\infty]{} \frac{\sigma^2}{2}. \end{aligned}$$

□

In order to compare distributions using their size-biased add-on structures, we need the following lemma:

Lemma 3.1. *Let X_0 and X_1 be two non-negative random variables with the same mean $a = E[X_0] = E[X_1] \in (0,\infty)$. Let F_0 be defined by $E[e^{-\lambda \dot{X}_0}] = E[e^{-\lambda X_0}]F_0(\lambda)$, where $\dot{X}_0$ is an X_0-transform of X_0, and F_1 be defined by $E[e^{-\lambda \dot{X}_1}] = E[e^{-\lambda X_1}]F_1(\lambda)$, where $\dot{X}_1$ is an X_1-transform of X_1. Then,*

$$\big|E[e^{-\lambda X_0}] - E[e^{-\lambda X_1}]\big| \le a\int_0^\lambda |F_0(s)-F_1(s)|ds, \quad \lambda \ge 0.$$

Proof. Since $\dot{X}_0$ is an X_0-transform of X_0, we have

$$\partial_\lambda(-\ln E[e^{-\lambda X_0}]) = \frac{E[X_0e^{-\lambda X_0}]}{E[e^{-\lambda X_0}]} = \frac{aE[e^{-\lambda \dot{X}_0}]}{E[e^{-\lambda X_0}]} = aF_0(\lambda).$$

Similarly, $\partial_\lambda(-\ln E[e^{-\lambda X_1}]) = aF_1(\lambda)$. Therefore, since $x - \ln x$ is decreasing on $[0,1]$,

$$\begin{aligned} \big|E[e^{-\lambda X_0}] - E[e^{-\lambda X_1}]\big| &\le \big|\ln E[e^{-\lambda X_0}] - \ln E[e^{-\lambda X_1}]\big| = a\Big|\int_0^\lambda F_0(s)ds - \int_0^\lambda F_1(s)ds\Big| \\ &\le a\int_0^\lambda |F_0(s)-F_1(s)|ds \end{aligned}$$

as desired. □

We are now ready to prove Lemma 1.3. It is elementary to verify that if Y is exponentially distributed, then it satisfies (1.7). So we only need to show that if Y is a strictly positive random variable with finite second moment, then (1.7) implies that it is exponentially distributed. The following lemma will be used to prove this.

Lemma 3.2. *Suppose that $c > 0$ is a constant, and F is a non-negative bounded function on $[0, \infty)$ satisfying that, for any $\lambda \geq 0$,*

$$F(\lambda) \leq \frac{1}{c} \int_0^1 du \int_0^\lambda F(us) ds. \tag{3.2}$$

Then $F \equiv 0$.

Proof. By dividing both sides of (3.2) by $\|F\|_\infty$, without loss of any generality, we can assume F is bounded by 1. We prove this lemma by contradiction. Assume that

$$\rho := \inf\{x \geq 0 : F(x) \neq 0\} < \infty, \tag{3.3}$$

with the convention $\inf \emptyset = \infty$. Then, for each $\lambda \geq 0$,

$$F(\rho + \lambda) = \frac{1}{c} \int_0^1 du \int_0^{\rho+\lambda} F(us) ds = \frac{1}{c} \int_0^1 du \int_\rho^{\rho+\lambda} F(us) ds \leq \frac{\lambda}{c}.$$

Using this new upper bound, we have

$$F(\rho + \lambda) = \frac{1}{c} \int_0^1 du \int_\rho^{\rho+\lambda} F(us) ds \leq \frac{1}{c} \int_0^1 du \int_\rho^{\rho+\lambda} \frac{\lambda}{c} ds \leq \frac{\lambda^2}{c^2}.$$

Repeating this process, we have $F(\rho+\lambda) \leq \frac{\lambda^m}{c^m}$ for each $m \in \mathbb{N}$, which implies that $F = 0$ on $[\rho, \rho + c)$. This, however, contradicts (3.3). □

Proof of Lemma 1.3. Suppose that Y is a strictly positive random variable with finite second moment, and (1.7) is true. Define $a := E[\dot{Y}] \in (0, \infty)$. Consider an exponential random variable $\mathbf{e}$ with mean $a/2$. It is elementary to verify that $\mathbf{e}$ satisfies (1.7), in the sense that $\ddot{\mathbf{e}} \overset{d}{=} \dot{\mathbf{e}} + U\dot{\mathbf{e}}'$, where $\dot{\mathbf{e}}$ and $\dot{\mathbf{e}}'$ are both $\mathbf{e}$-transforms of $\mathbf{e}$, $\ddot{\mathbf{e}}$ is an $\mathbf{e}^2$-transform of $\mathbf{e}$, U is a uniform random variable on $[0, 1]$, and $\dot{\mathbf{e}}$, $\dot{\mathbf{e}}'$ and U are independent. Notice that $E[\dot{\mathbf{e}}] = a$, therefore we can compare the distribution of $\dot{Y}$ with that of $\dot{\mathbf{e}}$ using Lemma 3.1. This gives that

$$\big|E[e^{-\lambda \dot{Y}}] - E[e^{-\lambda \dot{\mathbf{e}}}]\big| \leq a \int_0^\lambda \int_0^1 \big|E[e^{-su\dot{Y}}] - E[e^{-su\dot{\mathbf{e}}}]\big| du ds, \quad \lambda \geq 0,$$

which, according to Lemma 3.2, says that $\dot{Y} \overset{d}{=} \dot{\mathbf{e}}$. Since Y and $\mathbf{e}$ are strictly positive, according to (1.4), we have

$$E[1 - e^{-\lambda Y}]/E[Y] = E[1 - e^{-\lambda \mathbf{e}}]/E[\mathbf{e}], \quad \lambda \geq 0.$$

Letting $\lambda \to \infty$, we get $E[Y] = E[\mathbf{e}]$. Therefore, $Y \overset{d}{=} \mathbf{e}$ as desired. □

Proof of Theorem 1.1(2). Consider an exponential random variable Y with mean $\sigma^2/2$. Let $\dot{Y}$ be a Y-transform of Y. As in Section 1.2, we only need to prove that $\dot{Z}_n/n$ converge weakly to $\dot{Y}$. From Proposition 2.1, we know that

$$E[e^{-\lambda \ddot{Z}_n^{(n)}}] = E[e^{-\lambda \dot{Z}_n}] E[g(\lambda, \lfloor Un \rfloor) e^{-\lambda \dot{Z}_{\lfloor Un \rfloor}}],$$

where U is a uniform random variable on $[0,1]$ independent of $\{\dot{Z}_m : 0 \le m \le n\}$; and $g(\lambda, m)$ is a function on $[0,\infty) \times \mathbb{N}_0$ such that $g(\lambda, m) \to 1$, uniformly in λ as $m \to \infty$. After a renormalization, we have that

$$E[e^{-\lambda \frac{\ddot{Z}_n^{(n)}-1}{n}}] = E[e^{-\lambda \frac{\dot{Z}_n - 1}{n}}] E\big[g(\frac{\lambda}{n}, \lfloor Un \rfloor) e^{-\lambda U \frac{\dot{Z}_{\lfloor Un \rfloor}}{Un}}\big], \quad \lambda \ge 0.$$

According to Theorem 1.2, one can verify that $(\ddot{Z}_n^{(n)} - 1)/n$ is a $(\dot{Z}_n - 1)/n$ transform of $(\dot{Z}_n - 1)/n$. Therefore, the above equation can be viewed as the size-biased add-on structure for the random variable $(\dot{Z}_n - 1)/n$. It is easy to see that the mean of $\dot{Y}$ is σ^2. According to (2.5), the mean of $(\dot{Z}_n - 1)/n$ is also σ^2. Then comparing the distribution of $(\dot{Z}_n - 1)/n$ with that of $\dot{Y}$, and using Lemma 3.1, we get that

$$\big|E[e^{-\lambda \frac{\dot{Z}_n - 1}{n}}] - E[e^{-\lambda \dot{Y}}]\big| \le \sigma^2 \int_0^\lambda ds \int_0^1 \big|g(\frac{s}{n}, \lfloor un \rfloor) E[e^{-su \frac{\dot{Z}_{\lfloor un \rfloor}}{un}}] - E[e^{-su\dot{Y}}]\big| du.$$

Taking $n \to \infty$ and using the reverse Fatou's lemma, we arrive at

$$M(\lambda) \le \sigma^2 \int_0^1 du \int_0^\lambda M(us) ds, \quad \lambda \ge 0,$$

where $M(\lambda) := \limsup_{n\to\infty} |E[e^{-\lambda \frac{\dot{Z}_n}{n}}] - E[e^{-\lambda \dot{Y}}]|$. Thus by Lemma 3.2, we have $M \equiv 0$, which says that $\dot{Z}_n/n$ converges weakly to $\dot{Y}$. □

References

[1] Athreya, K. and Ney, P.: Functionals of critical multitype branching processes. *Ann. Probab.* **2** (1974), 339–343. MR-0356264

[2] Geiger, J.: Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Probab.* **36** (1999), 310–309. MR-1724856

[3] Geiger, J.: A new proof of Yaglom's exponential limit law. *Mathematics and Computer Science (Versailles, 2000).* 245–249, Trends Math., Birkhäuser, Basel, 2000. MR-1798303

[4] Harris, S. C. and Roberts, M. I.: The many-to-few lemma and multiple spines. *Ann. Inst. Henri Poincaré, Probab. Stat.* **53** (2017), 226–242. MR-3606740

[5] Kesten, H., Ney, P and Spitzer, F.: The Galton-Watson process with mean one and finite variance. *Teor. Verojatnost. i Primenen.* **11** (1966), 579–611. MR-0207052

[6] Kolmogorov, A. N.: Zur lösung einer biologischen aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk* **2** (1938), 1–12.

[7] Lyons, R., Pemantle, R. and Peres, Y.: Conceptual Proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** (1995), 1125–1138. MR-1349164

[8] Ren, Y.-X., Song, R. and Sun, Z.: Spine decompositions and limit theorems for a class of critical superprocesses. arXiv:1711.09188

[9] Vatutin, V. A. and Dyakonova, E. E.: The survival probability of a critical multitype Galton-Watson branching process. Proceedings of the Seminar on Stability Problems for Stochastic Models, Part II (Nałęczow, 1999). *J. Math. Sci. (New York)* **106** (2001), 2752–2759. MR-1878742

[10] Yaglom, A. M.: Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* **56** (1947), 795–798. MR-0022045

[11] Zubkov, A. M.: Limit distributions of the distance to the nearest common ancestor. *Teor. Verojatnost. i Primenen.* **20** (1975), 614–623. MR-0397915

Acknowledgments. We thank the referee for helpful comments on the first version of this paper.