# Chapter 5
# Size-biased Galton-Watson trees with a spine

LYONS, PEMANTLE & PERES [24] developed a conceptual tool that allows to give new and totally different probabilistic proofs of some classical limit theorems for GWP's including Theorem 2.2 of KESTEN & STIGUM. It is based on the comparison of a GWT $\boldsymbol{GW}$ with a certain *size-biased* version of it, denoted as $\widehat{\boldsymbol{GW}}$, and essentially amounts to a change of measure argument involving a harmonic transform. In order to elaborate this further, we first provide a short introduction of the concept of size-biasing before proceeding with the definition of a size-biased GWT and a subsequent study of its relevant properties.

## 5.1 Size-biased distributions and random variables

As size-biased distributions and random variables form an important ingredient to the construction of a size-biased GWT in the next section, we start by defining these objects properly.

**Definition 5.1.** Let $\nu$ be a distribution on $(\mathbb{R}_{\ge},\mathscr{B}(\mathbb{R}_{\ge}))$ with finite and positive mean $\gamma=\int x\,\nu(dx)$. Then the distribution $\widehat{\nu}$, defined by

$$\widehat{\nu}(B) \;=\; \frac{1}{\gamma}\int_B x\,\nu(dx),\quad B\in\mathscr{B}(\mathbb{R}_{\ge}),$$

is called the *size-biasing of* $\nu$ or *size-biased distribution associated with* $\nu$. Given a random variable $X$ with distribution $\nu$, any random variable $\widehat{X}$ with distribution $\widehat{\nu}$ is called a *size-biasing of* $X$.

In the following, we will mainly consider size-biasings of distributions $\nu=(\nu_n)_{n\ge 0}$ on $\mathbb{N}_0$. In this case, we obviously have $\widehat{\nu}=(\widehat{\nu}_n)_{n\ge 0}$ with

$$\widehat{\nu}_n \;=\; \frac{n\nu_n}{\gamma}, \quad n \in \mathbb{N}_0,$$

where $\gamma = \sum_{n\geq 1} n\nu_n$.

If $\widehat{X}$ denotes the size-biasing of a random variable $X$ with distribution $\nu$, then $\widehat{\nu} = \mathbb{P}(\widehat{X} \in \cdot)$ has the $\nu$-density

$$\frac{d\widehat{\nu}}{d\nu}(x) \;=\; \frac{x}{\gamma}\mathbf{1}_{\mathbb{R}_\geq}(x) \;=\; \frac{x}{\mathbb{E}X}\mathbf{1}_{\mathbb{R}_\geq}(x) \tag{5.1}$$

which immediately implies that, for any measurable function $\varphi : \mathbb{R}_\geq \to \mathbb{R}_\geq$, the identity

$$\mathbb{E}\varphi(\widehat{X}) \;=\; \int \varphi(x)\,\widehat{\nu}(dx) \;=\; \int \frac{x\varphi(x)}{\mathbb{E}X}\,\nu(dx) \;=\; \frac{\mathbb{E}X\varphi(X)}{\mathbb{E}X} \tag{5.2}$$

holds true. We further note that $\widehat{X}$ is almost surely positive.

If $\nu$ has a $\lambda\!\!\lambda$-density $g$, where $\lambda\!\!\lambda$ denotes Lebesgue measure, then the same holds naturally true for $\widehat{\nu}$, and we infer with the help of the product rule for Radon-Nikodym derivatives that

$$\widehat{g}(x) \;:=\; \frac{d\widehat{\nu}}{d\lambda\!\!\lambda}(x) \;=\; \frac{d\widehat{\nu}}{d\nu}(x)\frac{d\nu}{d\lambda\!\!\lambda}(x) \;=\; \frac{xg(x)}{\gamma}\mathbf{1}_{\mathbb{R}_\geq}(x). \tag{5.3}$$

## Problems

**Problem 5.2.** Let $\nu = (\nu_n)_{n\geq 0}$ be a distribution on $\mathbb{N}_0$ with gf $f$ and finite positive mean. Find the gf $\widehat{f}$ of the size-biasing $\widehat{\nu}$.

**Problem 5.3.** Let $\nu$ be a distribution on $\mathbb{R}_\geq$ with finite positive mean. Show that $\widehat{\nu}$ is stochastically larger than $\nu$, i.e. $\nu((x,\infty)) \leq \widehat{\nu}((x,\infty))$ for all $x \in \mathbb{R}_\geq$.

## 5.2 Size-biased Galton-Watson trees: construction and properties

The concept of size-biasing a nonnegative random variable shall now be extended in an appropriate manner so as to construct a size-biasing $\widehat{\boldsymbol{GW}}$ of a given GWT $\boldsymbol{GW}$. In connection with the Kesten-Stigum theorem, we will then deal with the question under which condition on the underlying offspring distribution the distribution of $\widehat{\boldsymbol{GW}}$, denoted as $\widehat{\mathsf{GW}}$, is dominated by $\mathsf{GW}$, which is the analog of (5.1). As it further turns out, $\mathsf{GW}$ and $\widehat{\mathsf{GW}}$ are mutually singular [☞ Lemma 5.24 and Theorem 5.25] whenever domination fails to hold.

***Construction of $\widehat{\boldsymbol{GW}}$.*** Given any offspring distribution $(p_n)_{n\ge 0}$ with finite positive mean $\mathsf{m}$, the reader is reminded that an associated GWT $\boldsymbol{GW}$ is constructed from a family $\{X_{\mathsf{v}}:\mathsf{v}\in\mathbb{V}\}$ of iid random variables with this distribution, where $X_{\mathsf{v}}$ denotes the number of children of the (potential) individual $\mathsf{v}$. Let now $\{(\widehat{X}_n,U_n):n\ge 0\}$ be an additional family of iid random vectors, defined on the same probability space $(\Omega,\mathfrak{A},\mathbb{P})$, with generic copy $(\widehat{X},U)$, and satisfying the following conditions:

(SB1) $\{(\widehat{X}_n,U_n):n\ge 0\}$ and $\{X_{\mathsf{v}}:\mathsf{v}\in\mathbb{V}\}$ are independent.

(SB2) The distribution of $\widehat{X}$ is $(\widehat{p}_n)_{n\ge 0}$, that is

$$\mathbb{P}(\widehat{X}=n) \;=\; \widehat{p}_n \;=\; \frac{np_n}{\mathsf{m}}$$

for $n\in\mathbb{N}_0$.

(SB3) For any $n$ with $p_n>0$, the conditional distribution of $U$ given $\widehat{X}=n$ is a discrete uniform distribution on $\{1,...,n\}$, that is

$$\mathbb{P}(U=k|\widehat{X}=n) \;=\; \frac{1}{n}$$

for $k=1,...,n$.

With the help of these random vectors, the construction of the *size-biased GWT* $\widehat{\boldsymbol{GW}}$ with distinguished path $\boldsymbol{V}=(\upsilon_n)_{n\ge 0}$, called *spine*, can now be accomplished as follows: As usual, start at the root $\varnothing$ which is viewed as the ancestor of a given population and is also the first vertex of the spine. It produces $\widehat{X}_0\ge 1$ children of which $\upsilon_1$, the next vertex of the spine, is picked at random with the help of $U_0$. This vertex (individual) has $\widehat{X}_1\ge 1$ descendants (children), while any other individual $\mathsf{v}$ of the first generation produces offspring in accordance with $(p_n)_{n\ge 0}$ (using the random variable $X_{\mathsf{v}}$). Continuing in this manner, each $\upsilon_n$ of the spine is picked at random from the $\widehat{X}_{n-1}\ge 1$ children of its predecessor (mother) $\upsilon_{n-1}$ (using $U_{n-1}$) and reproduces in accordance with the size-biased distribution $(\widehat{p}_n)_{n\ge 0}$. All other individuals not belonging to the spine reproduce in the usual way in accordance with the offspring distribution $(p_n)_{n\ge 0}$. As a result, we obtain an infinite, but locally finite random tree $\widehat{\boldsymbol{GW}}$ with a distinguished infinite path $(\upsilon_n)_{n\ge 0}$ as depicted in Figure 5.2.

Here are the formal details: Put $\widehat{\boldsymbol{GW}}:=\bigcup_{n\ge 0}\widehat{\boldsymbol{GW}}_n$, where $\widehat{\boldsymbol{GW}}_0:=\{\varnothing\}$ and the $\widehat{\boldsymbol{GW}}_n$ for $n\ge 1$ are recursively defined by

$$\widehat{\boldsymbol{GW}}_n \;=\; \boldsymbol{A}_n\cup\boldsymbol{B}_n,$$

with

$$\begin{aligned}\boldsymbol{A}_n \;&:=\; \left\{\mathsf{v}i\in\mathbb{N}^n:\mathsf{v}\in\widehat{\boldsymbol{GW}}_{n-1}\backslash\{\upsilon_{n-1}\},\, i\le X_{\mathsf{v}}\right\},\\ \boldsymbol{B}_n \;&:=\; \left\{\upsilon_{n-1}i:1\le i\le\widehat{X}_{n-1}\right\}.\end{aligned}$$

Put further $\boldsymbol{V}:=(\upsilon_n)_{n\ge 0}$, where $\upsilon_0:=\varnothing$ and

$$\upsilon_n := \upsilon_{n-1}U_{n-1} = U_0...U_{n-1}.$$

for $n \geq 1$. We have thus defined a mapping

$$(\widehat{\boldsymbol{GW}}, \boldsymbol{V}) : \Omega \to \mathbb{T} \times \partial\mathbb{V},$$

where

$$\partial\mathbb{V} := \{(\mathsf{v}^n)_{n\geq 0} : \mathsf{v}^n \in \mathbb{N}^n \text{ and } \mathsf{v}^{n+1} \succeq \mathsf{v}^n \text{ for } n \geq 0\}$$

denotes the set of infinite paths (rays) in $\mathbb{V}$. It may be identified with $\mathbb{N}^\infty$ via

$$(\mathsf{v}^n)_{n\geq 0} \longleftrightarrow (\mathsf{v}_n)_{n\geq 0}$$

where $\mathsf{v}^n = \mathsf{v}_1...\mathsf{v}_n$ for each $n \geq 1$. This will be done hereafter wherever useful.

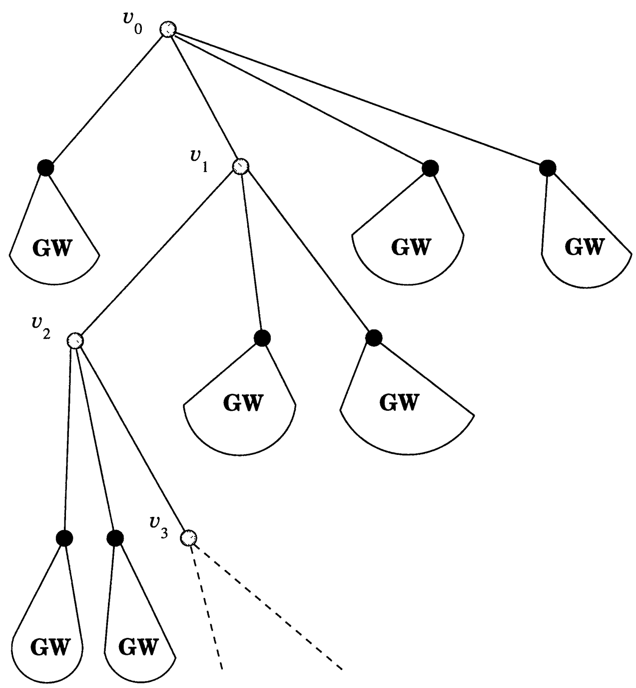


**Fig. 5.1** A size biased Galton-Watson tree with distinguished path $(\upsilon_0, \upsilon_1, ...)$

The next step is to turn $(\widehat{\boldsymbol{GW}}, \boldsymbol{V})$ into a random element in $\mathbb{T} \times \partial\mathbb{V}$ by endowing the latter space with a suitable $\sigma$-field. The subsequent lemma, which is proved in a similar manner as Lemma 4.2, provides us with a metric on the enlarged Ulam-Harris tree $\overline{\mathbb{V}} := \mathbb{V} \cup \partial\mathbb{V}$. Notice that the ordering introduced in Definition 4.6 as well as the minimum relation $\wedge$ extend to $\overline{\mathbb{V}}$ in an obvious manner.

**Lemma 5.4.** *Defining* $\rho : \overline{\mathbb{V}} \times \overline{\mathbb{V}} \to [0,1]$ *by*

$$\rho(\mathsf{v},\mathsf{w}) \;=\; e^{-|\mathsf{v}\wedge\mathsf{w}|}$$

*with* $e^{-\infty} := 0$*, the pair* $(\overline{\mathbb{V}},\rho)$ *forms a compact metric space with topological boundary* $\partial\mathbb{V}$ *and countable, dense and open subset* $\mathbb{V}$*.*

*Proof.* Problem 5.10. □

In analogy to the definition of $\mathscr{B}(\mathbb{T})$ we now define

$$\mathbb{B}(\mathsf{v},\varepsilon) = \{\mathsf{w} \in \overline{\mathbb{V}} : \rho(\mathsf{v},\mathsf{w}) < \varepsilon\} \tag{5.4}$$

for $\mathsf{v} \in \overline{\mathbb{V}}$, $\varepsilon > 0$ and then

$$\mathscr{B}(\overline{\mathbb{V}}) \;=\; \sigma\big(\{\mathbb{B}(\mathsf{v},\varepsilon) : \mathsf{v} \in \overline{\mathbb{V}},\, \varepsilon > 0\}\big).$$

By the separability of $(\overline{\mathbb{V}},\rho)$, this is again the Borel $\sigma$-field with respect to $\rho$ with countable generator $\{\mathbb{B}(\mathsf{v},\varepsilon) : \mathsf{v} \in \mathbb{V},\, \varepsilon > 0\}$ [☞ Problem 5.11]. The good news for our purposes comes next.

**Lemma 5.5.** *The mapping* $(\widehat{\boldsymbol{GW}},\boldsymbol{V}) : \Omega \to \mathbb{T} \times \overline{\mathbb{V}}$ *is* $\mathfrak{A}$-$\mathscr{B}(\mathbb{T})\otimes\mathscr{B}(\overline{\mathbb{V}})$*-measurable.*

*Proof.* It suffices to verify [☞ [5, Remark after Thm. 22.2]] that

(i) $\widehat{\boldsymbol{GW}}$ is $\mathfrak{A}$-$\mathscr{B}(\mathbb{T})$-measurable and
(ii) $\boldsymbol{V}$ is $\mathfrak{A}$-$\mathscr{B}(\overline{\mathbb{V}})$-measurable.

As for (i), we proceed as in Lemma 4.11(a). Given $\tau \in \mathbb{T}$ and $n \in \mathbb{N}$, we obtain

$$\begin{aligned}\widehat{\boldsymbol{GW}}^{-1}([\tau]_n) \;&=\; \Big\{\omega \in \Omega : \widehat{\boldsymbol{GW}}_{|n}(\omega) = \tau_{|n}\Big\}\\ &\in\; \sigma\Big((\widehat{X}_k,U_k)_{0\le k<n},\,(X_{\mathsf{v}})_{|\mathsf{v}|<n}\Big) \;\subset\; \mathfrak{A}\end{aligned}$$

and thus the asserted measurability.

Turning to (ii), let $\mathsf{u} = \mathsf{u}_1\mathsf{u}_2... \in \mathbb{N}^\infty$, $\mathsf{u}^0 := \varnothing$, $\mathsf{u}^n := \mathsf{u}_1...\mathsf{u}_n$ for $n \ge 1$ and $0 < \varepsilon < 1$, w.l.o.g. $\varepsilon = e^{-m}$ for some $m \ge 1$. Then

$$\boldsymbol{V}^{-1}(\mathbb{B}(\mathsf{u},\varepsilon)) \;=\; \{\omega \in \Omega : \upsilon_m(\omega) = \mathsf{u}^m\} \;\in\; \sigma\Big((\upsilon_k,U_k)_{0\le k<m}\Big) \;\subset\; \mathfrak{A}$$

gives the desired conclusion. □

We now put

$$\widehat{\mathsf{GW}}_* \;:=\; \mathbb{P}\Big((\widehat{\boldsymbol{GW}},\boldsymbol{V}) \in \cdot\Big) \quad \text{and} \quad \widehat{\mathsf{GW}} \;:=\; \mathbb{P}\Big(\widehat{\boldsymbol{GW}} \in \cdot\Big)$$

and note the following counterpart of Prop. 4.13.

**Proposition 5.6.** *Let $(\widehat{\boldsymbol{GW}},\boldsymbol{V})$ be a size-biased spinal GWT with distribution $\widehat{\mathsf{GW}}_*$. Then the following assertion holds true for any $n\in\mathbb{N}$: If $I_n \subset \mathbb{N}^n$ and $\mathsf{v} \in I_n$ are such that $\mathbb{P}(\widehat{\boldsymbol{GW}} \in I_n, \upsilon_n = \mathsf{v}) > 0$, then the mappings $(\Theta_{\mathsf{v}}\widehat{\boldsymbol{GW}}^{\mathsf{v}}, \Theta_{\mathsf{v}}\boldsymbol{V})$ and $\Theta_{\mathsf{u}}\widehat{\boldsymbol{GW}}^{\mathsf{u}}$, $\mathsf{u}\in \mathsf{i}_n\backslash\{\mathsf{v}\}$ are measurable and conditionally independent given $(\widehat{\boldsymbol{GW}},\upsilon_n) = (I_n,\mathsf{v})$ with distribution $\widehat{\mathsf{GW}}_*$ and $\widehat{\mathsf{GW}}$, respectively.*

*Proof.* Problem 5.12. □

The relation between $\mathsf{GW}$ and $\widehat{\mathsf{GW}}$ forms the key for the techniques developed hereafter and is described by the next result. For $\tau\in\mathbb{T}$, $n\in\mathbb{N}_0$ and $\mathsf{u}\in\tau_n$, we define the set

$$[\tau;\mathsf{u}]_n \;=\; \Big\{(\tau',(\mathsf{v}^k)_{k\ge 0}) \in \mathbb{T}\times\partial\mathbb{V} : \tau' \in [\tau]_n \text{ and } \mathsf{v}^n = \mathsf{u}\Big\}.$$

which consists of all pairs $(\tau',(\mathsf{v}^k)_{k\ge 0}) \in \mathbb{T}\times\partial\mathbb{V}$ such that the tree $\tau'$ equals $\tau$ up to level $n$ and the infinite path $(\mathsf{v}^k)_{k\ge 0}$ passes through $\mathsf{u}$.

**Lemma 5.7. [Comparison lemma]** *Let $(p_n)_{n\ge 0}$ be an offspring distribution with finite positive mean $\mathsf{m}$ and put $w_n(\tau) := \mathsf{m}^{-n} z_n(\tau)$ for $\tau\in\mathbb{T}$ and $n\in\mathbb{N}_0$. Then the following assertions about $\mathsf{GW}$, $\widehat{\mathsf{GW}}$ and $\widehat{\mathsf{GW}}_*$ hold true:*

*(a)* *For all $\tau\in\mathbb{T}$ and $n\in\mathbb{N}_0$,*

$$\mathsf{GW}([\tau]_{n+1}) \;=\; p_{z_1(\tau)} \prod_{k=1}^{z_1(\tau)} \mathsf{GW}([\Theta_k\tau^k]_n). \tag{5.5}$$

*(b)* *For all $\tau\in\mathbb{T}$, $n\in\mathbb{N}_0$ and $\mathsf{u}\in\tau_n$,*

$$\widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_n) \;=\; \mathsf{m}^{-n}\,\mathsf{GW}([\tau]_n). \tag{5.6}$$

*(c)* *For all $\tau\in\mathbb{T}$ and $n\in\mathbb{N}_0$,*

$$\widehat{\mathsf{GW}}([\tau]_n) \;=\; w_n(\tau)\,\mathsf{GW}([\tau]_n). \tag{5.7}$$

*(d)* *Put $\widehat{Z}_n := z_n\circ\widehat{\boldsymbol{GW}}$ for $n\in\mathbb{N}_0$. Then $\widehat{Z}_n$ is a size-biasing of $Z_n$, that is*

$$\mathbb{P}(\widehat{Z}_n = k) \;=\; \frac{k\,\mathbb{P}(Z_n = k)}{\mathsf{m}^n} \;=\; \frac{k\,\mathbb{P}(Z_n=k)}{\mathbb{E}Z_n} \tag{5.8}$$

*for all* $k,n \in \mathbb{N}_0$.

*Proof.* (a) Let $\tau \in \mathbb{T}$, $n \in \mathbb{N}_0$ and $z_1(\tau) = l$. Since (5.5) is trivial if $l = 0$ or $p_l = 0$, let further $l \geq 1$ and $p_l > 0$. Then it follows with the help of Prop. 4.13 that

$$\begin{aligned} \mathsf{GW}([\tau]_{n+1}) &= \mathbb{P}(Z_1 = l, \Theta_k \boldsymbol{GW}^k \in [\Theta_k \tau^k]_n \text{ for } 1 \leq k \leq l) \\ &= p_l\, \mathbb{P}(\Theta_k \boldsymbol{GW}^k \in [\Theta_k \tau^k]_n \text{ for } 1 \leq k \leq l | Z_1 = l) \\ &= p_l \prod_{k=1}^{l} \mathsf{GW}([\Theta_k \tau^k]_n) \end{aligned}$$

which is the assertion.

(b) Using induction over $n$, note first that

$$\upsilon_0 = \mathsf{u} = \varnothing, \quad [\tau;\varnothing]_0 = \mathbb{T} \times \partial\mathbb{V} \quad \text{and} \quad [\tau]_0 = \mathbb{T}$$

for all $\tau \in \mathbb{T}$ imply

$$\widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_0) \;=\; \widehat{\mathsf{GW}}_*([\tau;\varnothing]_0) \;=\; 1 \;=\; \mathsf{GW}([\tau]_0)$$

for all $\tau \in \mathbb{T}$.

Now assume the assertion has been proved for some $n \in \mathbb{N}_0$, all $\tau' \in \mathbb{T}$ and $\mathsf{u}' \in \tau_n'$ (inductive hypothesis). We need the following intermediate calculation: Given $\tau \in \mathbb{T}$ with $z_1(\tau) = l \geq 1$ and $\mathsf{u} \in \tau_{n+1}$, there obviously exists a unique $j \in \{1,...,l\}$ such that $\mathsf{u} \in \tau^j$. Setting $A := \{\Theta_i \widehat{\boldsymbol{GW}}^i \in [\Theta_i \tau^i]_n \text{ for } 1 \leq i \leq l,\, i \neq j\}$, we hence infer

$$\Big\{(\widehat{\boldsymbol{GW}}, \boldsymbol{V}) \in [\tau;\mathsf{u}]_{n+1}\Big\} \;=\; \Big\{\widehat{X}_0 = l,\, U_0 = j,\, (\Theta_j \widehat{\boldsymbol{GW}}^j, \Theta_j \boldsymbol{V}) \in [\Theta_j \tau^j; \Theta_j \mathsf{u}]_n\Big\} \cap A.$$

If $p_l > 0$, which entails $\mathbb{P}(\widehat{X}_0 = l, U_0 = j) = \mathsf{m}^{-1} l p_l l^{-1} = \mathsf{m}^{-1} p_l > 0$, we obtain with the help of Prop. 5.6

$$\begin{aligned} \widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_{n+1}) &= \mathbb{P}\Big((\widehat{\boldsymbol{GW}}, \boldsymbol{V}) \in [\tau;\mathsf{u}]_{n+1}\Big) \\ &= \frac{p_l}{\mathsf{m}}\, \mathbb{P}\Big(\Big\{(\Theta_j \widehat{\boldsymbol{GW}}^j, \Theta_j \boldsymbol{V}) \in [\Theta_j \tau^j; \Theta_j \mathsf{u}]_n\Big\} \cap A \Big| \widehat{X}_0 = l, U_0 = j\Big) \\ &= \frac{p_l}{\mathsf{m}}\, \widehat{\mathsf{GW}}_*([\Theta_j \tau^j; \Theta_j \mathsf{u}]_n) \prod_{i \neq j} \mathsf{GW}([\Theta_i \tau^i]_n). \end{aligned}$$

Applying to this the inductive hypothesis and then (5.5), we obtain

$$\begin{aligned} \widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_{n+1}) &= \frac{p_l}{\mathsf{m}^{n+1}}\, \mathsf{GW}([\Theta_j \tau^j]_n) \prod_{i \neq j} \mathsf{GW}([\Theta_i \tau^i]_n) \\ &= \frac{1}{\mathsf{m}^{n+1}}\, \mathsf{GW}([\tau]_{n+1}) \end{aligned}$$

as claimed.

(c) Here the assertion follows with the help of (5.6), viz.

$$\begin{aligned}\widehat{\mathsf{GW}}([\tau]_n) &= \mathbb{P}(\widehat{\boldsymbol{GW}} \in [\tau]_n) \\ &= \sum_{\mathsf{u}\in\tau_n} \mathbb{P}(\widehat{\boldsymbol{GW}} \in [\tau]_n,\, \upsilon_n = \mathsf{u}) \\ &= \sum_{\mathsf{u}\in\tau_n} \widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_n) \\ &= \sum_{\mathsf{u}\in\tau_n} \mathsf{m}^{-n}\,\mathsf{GW}([\tau]_n) \\ &= w_n(\tau)\,\mathsf{GW}([\tau]_n)\end{aligned}$$

for all $\tau \in \mathbb{T}$ and $n \in \mathbb{N}_0$.

(d) Finally, the just shown (5.7) provides us with

$$\begin{aligned}\mathbb{P}(\widehat{Z}_n = k) &= \sum_{\tau\in\mathbb{T}_n: z_n(\tau)=k} \widehat{\mathsf{GW}}([\tau]_n) \\ &= \sum_{\tau\in\mathbb{T}_n: z_n(\tau)=k} w_n(\tau)\,\mathsf{GW}([\tau]_n) \\ &= \sum_{\tau\in\mathbb{T}_n: z_n(\tau)=k} \frac{k}{\mathsf{m}^n}\,\mathsf{GW}([\tau]_n) \\ &= \frac{k\,\mathbb{P}(Z_n = k)}{\mathsf{m}^n}\end{aligned}$$

for any $k \in \mathbb{N}$, thus (5.8). ◻

*Remark 5.8.* With the help of the parts (b) and (c) of the Comparison lemma, one can easily prove the intuitively evident result that, given $\widehat{\boldsymbol{GW}}_n = \{\mathsf{u}^1, ..., \mathsf{u}^p\}$ for some $p \in \mathbb{N}$, the conditional distribution of the $n^{th}$ vertebra $\upsilon_n$ is uniform on this set [☞ Problem 5.13].

*Remark 5.9.* Regarding the question raised at the beginning of this section, when $\widehat{\mathsf{GW}}$ is dominated by $\mathsf{GW}$, the Comparison lemma does not yet provide a complete answer, but at least shows that, for any $\tau \in \mathbb{T}$ and $n \in \mathbb{N}_0$,

$$\widehat{\mathsf{GW}}([\tau]_n) \;=\; w_n(\tau)\,\mathsf{GW}([\tau]_n) \;=\; \int_{[\tau]_n} w_n(\chi)\;\mathsf{GW}(d\chi),$$

for $w_n$ is constant on $[\tau]_n$. We content ourselves here with this statement and return to the question when reproving the Kesten-Stigum theorem in Section 5.4.

## Problems

**Problem 5.10.** Prove Lemma 5.4.

**Problem 5.11.** Let $\mathbb{B}(\mathsf{v},\varepsilon)$ be defined by (5.4). Prove that $\{\mathbb{B}(\mathsf{v},\varepsilon) : \mathsf{v} \in \mathbb{V}, \varepsilon > 0\}$ is a countable and a generator of the Borel $\sigma$-field $\mathscr{B}(\overline{\mathbb{V}})$ with respect to $\rho$.

**Problem 5.12.** Prove Prop. 5.6.

**Problem 5.13.** Let $n, p \in \mathbb{N}$ and $\mathsf{u}^1,...,\mathsf{u}^n \in \mathbb{N}^n$ be such that $\mathbb{P}(\widehat{\boldsymbol{GW}}_n = \{\mathsf{u}^1,...,\mathsf{u}^p\}) > 0$. Show that

$$\mathbb{P}\left(\upsilon_n = \mathsf{u}^k \middle| \widehat{\boldsymbol{GW}}_n = \{\mathsf{u}^1,...,\mathsf{u}^p\}\right) \;=\; \frac{1}{p}$$

for any $k = 1,...,p$, that is, $\upsilon_n$ conditioned upon $\widehat{\boldsymbol{GW}}_n = \{\mathsf{u}^1,...,\mathsf{u}^p\}$ is discrete uniform on this set.

## 5.3 Size-biased Galton-Watson trees and GWPI

In this section, we return to Galton-Watson processes with immigration (GWPI) studied in Chapter 3 and prove extensions of two results from there in the noncritical case, Cor. 3.3 by HEATHCOTE and Theorem 3.12 by SENETA, with the help of the previously developed tools. The primary reason for doing this is that, as we will show first, the *size-biased GWP* $(\widehat{Z}_n)_{n\ge 0}$, given by $\widehat{Z}_n = z_n \circ \widehat{\boldsymbol{GW}}_n$ and already encountered in the Comparison lemma 5.7, forms a noncritical GWPI and may thus be studied with the help of the afore-mentioned results.

Unlike our definition in Chapter 3, it is stipulated here for sake of convenience that a GWPI $(Y_n)_{n\ge 0}$ starts with zero individuals, i.e. $Y_0 = 0$, and is fed by an iid sequence of numbers of immigrants $\zeta_1, \zeta_2, ...$ as of generation one only. In other words, it is $(Y_{n+1})_{n\ge 0}$ that constitutes a GWPI in the sense of the original Definition 3.1. As usual, let $(p_n)_{n\ge 0}$ denote the underlying offspring distribution and $(c_n)_{n\ge 0}$ the immigration distribution, thus the distribution of the $\zeta_n$. Then

$$Y_n \;=\; \zeta_n + \sum_{k=1}^{Y_{n-1}} \xi_{n,k}, \quad n \ge 1 \tag{5.9}$$

with iid random variables $\xi_{n,k}$ having distribution $(p_n)_{n\ge 0}$ and independent of $(\zeta_n)_{n\ge 1}$. We also call $(Y_n)_{n\ge 0}$ a *GWPI with immigration sequence* $(\zeta_n)_{n\ge 1}$. It may given in alternative form as

$$Y_n \;=\; \sum_{k=1}^{n} \sum_{j=1}^{\zeta_k} Z_{n-k}(k,j), \quad n \ge 1 \tag{5.10}$$

where the $(Z_n(j,k))_{n\geq 0}$ denote iid ordinary GWP's with one ancestor and offspring distribution $(p_n)_{n\geq 0}$ which are independent of $(\zeta_n)_{n\geq 1}$. One may interpret $(Z_n(j,k))_{n\geq 0}$ as the process spawned by the $j^{th}$ immigrant in generation $k$.

### 5.3.1 Connection between GWPI and size-biased GWT's

The result we are going to show next is that in a size-biased GWT the children produced by the individuals of the spine may be viewed as immigrants. This provides a connection with GWPI that will subsequently be utilized.

**Theorem 5.14.** *Let $(Y_n)_{n\geq 0}$ be a GWPI with immigration sequence $(\zeta_n)_{n\geq 1}$ defined by $\zeta_n = \widehat{X}_n - 1$ for $n \in \mathbb{N}$. Then*

$$\mathbb{P}((Y_n)_{n\geq 0} \in \cdot) \;=\; \widehat{\mathsf{GW}}((z_n - 1)_{n\geq 0} \in \cdot) \;=\; \mathbb{P}((\widehat{Z}_n)_{n\geq 0} \in \cdot).$$

*Proof.* Obviously, $\widehat{\mathsf{GW}}(z_0 - 1 \in \cdot) = \delta_0 = \mathbb{P}(Y_0 \in \cdot)$. For $n \in \mathbb{N}_0$ and $(k_1,...,k_{n+1}) \in \mathbb{N}^{n+1}$, we obtain by making use of the independence of $\zeta_{n+1}$, $(\xi_{n+1,j})_{j\geq 1}$ and $(Y_1,...,Y_n)$ that

$$\begin{aligned}
&\mathbb{P}(Y_1 = k_1,...,Y_n = k_n, Y_{n+1} = k_{n+1}) \\
&\quad = \mathbb{P}\left(Y_1 = k_1,...,Y_n = k_n, \zeta_{n+1} + \sum_{j=1}^{k_n} \xi_{n+1,j} = k_{n+1}\right) \\
&\quad = \mathbb{P}(Y_1 = k_1,...,Y_n = k_n)\,\mathbb{P}\left(\zeta_{n+1} + \sum_{j=1}^{k_n} \xi_{n+1,j} = k_{n+1}\right) \\
&\quad = \mathbb{P}(Y_1 = k_1,...,Y_n = k_n) \sum_{l=0}^{k_{n+1}} \mathbb{P}\left(\widehat{X}_{n+1} = l+1, \sum_{j=1}^{k_n} \xi_{n+1,j} = k_{n+1} - l\right) \\
&\quad = \mathbb{P}(Y_1 = k_1,...,Y_n = k_n) \sum_{l=0}^{k_{n+1}} \frac{(l+1)p_{l+1}}{\mathsf{m}}\, p^{*k_n}_{k_{n+1}-l}.
\end{aligned} \tag{5.11}$$

On the other hand, setting $\Pi(k_1,...,k_n) := \{\tau \in \mathbb{T}_n : z_j(\tau) - 1 = k_j \text{ for } 1 \leq j \leq n\}$, we find that

$$
\begin{aligned}
&\widehat{\mathsf{GW}}(z_1-1=k_1,\ldots,z_n-1=k_n,z_{n+1}-1=k_{n+1})\\
&\quad=\sum_{\tau\in\Pi(k_1,\ldots,k_n)}\sum_{\mathsf{u}\in\tau_n}\mathbb{P}\left(\widehat{\boldsymbol{GW}}=[\tau]_n,\upsilon_n=\mathsf{u},\widehat{X}_n+\sum_{\mathsf{v}\in\tau_n\backslash\{\mathsf{u}\}}X_{\mathsf{v}}=k_{n+1}+1\right)\\
&\quad=\sum_{\tau\in\Pi(k_1,\ldots,k_n)}\sum_{\mathsf{u}\in\tau_n}\widehat{\mathsf{GW}}_*([\tau;\mathsf{u}]_n)\,\mathbb{P}\left(\widehat{X}_n+\sum_{\mathsf{v}\in\tau_n\backslash\{\mathsf{u}\}}X_{\mathsf{v}}=k_{n+1}+1\right)\\
&\quad=\left(\sum_{\tau\in\Pi(k_1,\ldots,k_n)}\widehat{\mathsf{GW}}([\tau]_n)\right)\left(\sum_{l=0}^{k_{n+1}}\frac{(l+1)p_{l+1}}{\mathsf{m}}\,p^{*k_n}_{k_{n+1}-l}\right)\\
&\quad=\widehat{\mathsf{GW}}(z_1-1=k_1,\ldots,z_n-1=k_n)\left(\sum_{l=0}^{k_{n+1}}\frac{(l+1)p_{l+1}}{\mathsf{m}}\,p^{*k_n}_{k_{n+1}-l}\right)
\end{aligned}\tag{5.12}
$$

having utilized the independence of $(\widehat{\boldsymbol{GW}},\upsilon_n)$ and $(\widehat{X}_n,(X_{\mathsf{v}})_{\mathsf{v}\in\mathbb{N}^n})$ for the second equality, and the Comparison lemma 5.7 for the third one.

By combining (5.11) and (5.12), it follows upon induction over $n$ that

$$
\mathbb{P}(Y_1=k_1,\ldots,Y_n=k_n)\;=\;\widehat{\mathsf{GW}}(z_1-1=k_1,\ldots,z_n-1=k_n)
$$

for all $n\in\mathbb{N}$ and $k_1,\ldots,k_n\in\mathbb{N}^n$ and thus the assertion. □

*Remark 5.15.* The previous result can be quite easily understood even without providing formal arguments: When removing the spine $\boldsymbol{V}$ from the size-biased GWT $\widehat{\boldsymbol{GW}}$, each generation $n\ge 1$ of the remaining population, and thus of size $z_n-1$, can be decomposed into those members that are direct descendants of the spinal vertex $\upsilon_{n-1}$, i.e. $\boldsymbol{B}_n\backslash\{\upsilon_n\}$ [☞ Section 5.2] and all other ones that are children of any individual in $\boldsymbol{A}_n=\widehat{\boldsymbol{GW}}_{n-1}\backslash\{\upsilon_{n-1}\}$. The elements of $\boldsymbol{B}_n\backslash\{\upsilon_n\}$ are interpreted as immigrants, having offspring distribution $(c_k)_{k\ge 0}$ defined by $c_k=\widehat{p}_{k+1}$, whereas all other individuals reproduce in accordance with $(p_k)_{k\ge 0}$.

The next two subsections are devoted to the already announced derivation of two results for noncritical GWPI that will be useful thereafter and are minor extensions of results already obtained in Chapter 3. More important than the latter fact, however, is that the proofs do not use gf's but rather probabilistic arguments based on the previously developed theory.

### 5.3.2 Asymptotic growth of supercritical GWPI

Let us begin with two general auxiliary lemmata the first of which is based on the well-known Borel-Cantelli lemma.

**Lemma 5.16.** *Given a sequence $(X_n)_{n\ge 0}$ of iid nonnegative random variables, the following assertions hold:*

*(a)* *If $\mathbb{E}X_0<\infty$, then* $\lim_{n\to\infty}\frac{X_n}{n}=0$ *a.s.*

*(b)* *If $\mathbb{E}X_0=\infty$, then* $\limsup_{n\to\infty}\frac{X_n}{n}=\infty$ *a.s.*

*(c)* *If $\mathbb{E}X_0<\infty$, then* $\sum_{n\ge 1}c^n e^{X_n}<\infty$ *a.s. for all* $c\in(0,1)$.

*(d)* *If $\mathbb{E}X_0=\infty$, then* $\sum_{n\ge 1}c^n e^{X_n}=\infty$ *a.s. for all* $c\in(0,1)$.

*Proof.* Using the well-known inequality $\sum_{n\ge 1}\mathbb{P}(X>n)\le\mathbb{E}X\le\sum_{n\ge 0}\mathbb{P}(X>n)$ for any nonnegative random variable $X$, we infer that

$$\sum_{n\ge 1}\mathbb{P}\left(\frac{X_n}{n}>\varepsilon\right)\ \le\ \frac{\mathbb{E}X_0}{\varepsilon}\ \le\ \sum_{n\ge 0}\mathbb{P}\left(\frac{X_n}{n}>\varepsilon\right)$$

for all $\varepsilon>0$. Hence, by the Borel-Cantelli lemma, $\mathbb{P}(n^{-1}X_n>\varepsilon\text{ i.o.})=0$ or $=1$ according as $\mathbb{E}X_0<\infty$ or $=\infty$. This yields (a) and (b).

(c) Suppose that $\mathbb{E}X_0<\infty$ and let $c\in(0,1)$ and $0<\varepsilon<-\log c$. For $\omega\in A:=\{\lim_{n\to\infty}n^{-1}X_n=0\}$ and all sufficiently large $n$, we then infer $X_n(\omega)\le n\varepsilon$ from $\mathbb{P}(A)=1$, thus

$$c^n e^{X_n(\omega)}\ \le\ e^{n(\varepsilon+\log c)}.$$

This shows the convergence of $\sum_{n\ge 1}c^n e^{X_n}$ on $A$ because $e^{\varepsilon+\log c}<1$.

(d) If $\mathbb{E}X_0=\infty$, then $\mathbb{P}(B)=1$ for $B:=\{\limsup_{n\to\infty}n^{-1}X_n=\infty\}$. As a consequence, $c^n e^{X_n(\omega)}\ge 1$ holds true for any fixed $c\in(0,1)$, $\omega\in B$ and infinitely many $n\in\mathbb{N}$, which proves the divergence of $\sum_{n\ge 1}c^n e^{X_n}$ on $B$. □

The second lemma is an extension of the martingale convergence theorem to sequences of nonnegative random variables having the submartingale property but only conditionally integrable.

**Lemma 5.17.** *Let $M=(M_n)_{n\ge 0}$ be a sequence of nonnegative random variables adapted to a filtration $(\mathscr{F}_n)_{n\ge 0}$ and satisfying the following two conditions:*

*(i)* $\mathbb{E}(M_{n+1}|\mathscr{F}_n)\ge M_n$ *a.s. for all* $n\in\mathbb{N}_0$. *[submartingale property]*

*(ii)* $\sup_{n\ge 0}\mathbb{E}(M_n|\mathscr{F}_0)<\infty$ *a.s.* *[conditional $L^1$-boundedness]*

*Then $M_n$ converges a.s. to a finite random variable $M_\infty$.*

*Proof.* Problem 5.22 □

We are ready now to prove the following variant of Theorem 3.12 by SENETA for supercritical GWPI.

**Theorem 5.18. [Seneta]** *Let $(Y_n)_{n\ge 0}$ be a supercritical GWPI with finite offspring mean $\mathsf{m}$ and immigration sequence $(\zeta_n)_{n\ge 1}$ with generic copy $\zeta$. Then the following assertions hold true:*

*(a) If $\mathbb{E}\log^+\zeta<\infty$, then $\mathsf{m}^{-n}Y_n$ converges a.s. to a finite limit $Y_\infty$.*
*(b) If $\mathbb{E}\log^+\zeta=\infty$, then*

$$\limsup_{n\to\infty}\frac{Y_n}{c^n}\;=\;\infty\quad a.s.$$

*for any $c\in\mathbb{R}_>$.*

*Proof.* (a) Putting $\mathscr{F}_0:=\sigma((\zeta_n)_{n\ge 1})$, we obtain upon using (5.9) that

$$\begin{aligned}\mathbb{E}(Y_n|\mathscr{F}_0)\;&=\;\zeta_n\;+\;\sum_{k\ge 0}\mathbb{E}\left(\mathbf{1}_{\{Y_{n-1}=k\}}\sum_{j=1}^{k}\xi_{n,j}\middle|\mathscr{F}_0\right)\\ &=\;\zeta_n\;+\;\sum_{k\ge 1}k\mathsf{m}\,\mathbb{P}(Y_{n-1}=k|\mathscr{F}_0)\\ &=\;\zeta_n\;+\;\mathsf{m}\,\mathbb{E}(Y_{n-1}|\mathscr{F}_0)\quad\text{a.s.,}\end{aligned}$$

for $\sum_{j=1}^k\xi_{n,j}$ and $(Y_{n-1},(\zeta_n)_{n\ge 1})$ are independent for each $k\in\mathbb{N}_0$. Recalling $Y_0=0$, this inductively leads to

$$\mathbb{E}\left(\frac{Y_n}{\mathsf{m}^n}\middle|\mathscr{F}_0\right)\;=\;\sum_{k=1}^{n}\frac{\zeta_k}{\mathsf{m}^k}\quad\text{a.s.}$$

for all $n\in\mathbb{N}_0$. An application of Lemma 5.16 to the sequence $(\log^+\zeta_n)_{n\ge 1}$ now shows that, if $\mathbb{E}\log^+\zeta<\infty$,

$$\sup_{n\ge 0}\mathbb{E}\left(\frac{Y_n}{\mathsf{m}^n}\middle|\mathscr{F}_0\right)\;=\;\sum_{k\ge 1}\sum_{k=1}^{n}\frac{\zeta_k}{\mathsf{m}^k}\;\le\;\sum_{k\ge 1}\frac{e^{-\log^+\zeta_k}}{\mathsf{m}^k}\;<\;\infty\quad\text{a.s.}\tag{5.13}$$

After these observations we infer the almost sure convergence of $\mathsf{m}^{-n}Y_n$ from the previous lemma if we still verify that besides its assumption (ii) [valid by (5.13)] all other assumptions there are fulfilled as well. Put $\mathscr{F}_n:=\sigma((\zeta_k)_{k\ge 1},Y_0,...,Y_n)$ for $n\in\mathbb{N}$ so that $(Y_n)_{n\ge 0}$ is adapted with respect to $(\mathscr{F}_n)_{n\ge 0}$. Furthermore,

$$\mathbb{E}\left(\frac{Y_{n+1}}{\mathsf{m}^{n+1}}\middle|\mathscr{F}_n\right)\;=\;\frac{\zeta_{n+1}}{\mathsf{m}^{n+1}}\;+\;\frac{1}{\mathsf{m}^{n+1}}\mathbb{E}\left(\sum_{k=1}^{n}\xi_{n+1,k}\middle|\mathscr{F}_n\right)\;\ge\;\frac{Y_n}{\mathsf{m}^n}\quad\text{a.s.}$$

for each $n \in \mathbb{N}_0$, because the independence of $\mathscr{F}_n$ and $(\xi_{n+1,k})_{k\geq 1}$ and the $\mathscr{F}_n$-measurability of $Y_n$ ensure that

$$\mathbb{E}\left(\sum_{k=1}^{n} \xi_{n+1,k} \middle| \mathscr{F}_n\right) \ =\ Y_n\,\mathbb{E}\xi_{n+1,1}\ =\ \mathsf{m}\,Y_n \quad \text{a.s.}$$

Having verified all assumptions of Lemma 5.17, the latter gives the desired conclusion.

(b) If $\mathbb{E}\log^+\zeta = \infty$, then Lemma 5.16 yields

$$\limsup_{n\to\infty} \frac{\log^+\zeta_n}{n}\ =\ \infty \quad \text{a.s.}$$

and therefore for each positive $c$

$$\infty\ =\ \limsup_{n\to\infty} \frac{1}{c}\exp\left(\frac{\log\zeta_n}{n}\right)\ =\ \limsup_{n\to\infty}\left(\frac{\zeta_n}{c^n}\right)^{1/n} \quad \text{a.s.}$$

As $Y_n \geq \zeta_n$ for each $n\geq 1$, we arrive at the assertion. □

*Remark 5.19.* Even when disregarding the slightly different definitions of a GWPI here and in Chapter 3 concerning its initial distribution, Theorems 3.12 and 5.18 do not match exactly. First of all, part (a) of the latter result considers the normalization $\mathsf{m}^n$, whereas Theorem 3.12 uses the Heyde-Seneta norming $k_n$ which is of the same order of magnitude only if the $\xi_{n,k}$ satisfy the (ZlogZ)-condition. Since $\mathsf{m}^{-n}k_n \to 0$ if this condition fails to hold, we conclude further in Theorem 5.18(a) that $Y_\infty$ is a.s. positive and finite under (ZlogZ) and a.s. equal to zero otherwise. If $\mathbb{E}\log^+\zeta = \infty$, then Theorem 3.12 asserts that $k_n^{-1}Y_n \to \infty$ a.s. and thus $c^{-n}Y_n \to \infty$ a.s. as well for any $0 < c < \mathsf{m}$, for $c^{-n}k_n \to \infty$ [☞ Theorem 2.1 and its proof]. This is a stronger assertion than in part (b) of the above result which, on the other hand, goes beyond Theorem 3.12 by providing information on the behavior of $c^{-n}Y_n$ also for $c \geq \mathsf{m}$.

### 5.3.3 Subcritical GWPI: Heathcote's result revisited

To take another look at Cor. 3.3 due to HEATHCOTE, we need the following generalization of the Borel-Cantelli lemma.

**Lemma 5.20.** *Let $(k_n)_{n\geq 0}$ denote a sequence of nonnegative integers, $(A_{n,k})_{n,k\geq 1}$ an array of independent events and $B_n := \bigcup_{k=1}^{k_n} A_{n,k}$ for $n \in \mathbb{N}$ (defined as $\emptyset$ if $k_n = 0$). Then the implication*

$$\mathbb{P}\left(\limsup_{n\to\infty} B_n\right) = 0 \quad \Rightarrow \quad \sum_{n\geq 1}\sum_{k=1}^{k_n} \mathbb{P}(A_{n,k}) < \infty$$

*holds true.*

*Proof (by contraposition).* If $\sum_{n\ge 1}\sum_{k=1}^{k_n}\mathbb{P}(A_{n,k})=\infty$, then

$$\begin{aligned}\mathbb{P}\Big(\liminf_{n\to\infty} B_n^c\Big) &= \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{j\ge n}\bigcap_{k=1}^{k_j} A_{j,k}^c\right)\\ &= \lim_{n\to\infty}\prod_{j\ge n}\prod_{k=1}^{k_j}\Big(1-\mathbb{P}(A_{j,k})\Big)\\ &= \lim_{n\to\infty}\exp\left(-\sum_{j\ge n}\sum_{k=1}^{k_j}\log\Big(1-\mathbb{P}(A_{j,k})\Big)\right)\\ &\le \exp\left(-\sum_{j\ge n}\sum_{k=1}^{k_j}\mathbb{P}(A_{j,k})\right) = 0\end{aligned}$$

which proves the assertion. □

**Theorem 5.21. [Heathcote]** *Let $(Y_n)_{n\ge 0}$ be a subcritical GWPI with $p_0<1$ and immigration sequence $(\zeta_n)_{n\ge 1}$ with generic copy $\zeta$. Then the following assertions hold true:*

*(a)* *If $\mathbb{E}\log^+\zeta<\infty$, then $Y_n$ converges in distribution to a nonnegative random variable $Y_\infty$.*

*(b)* *If $\mathbb{E}\log^+\zeta=\infty$, then $Y_n \xrightarrow{\mathbb{P}} \infty$, i.e.*

$$\lim_{n\to\infty}\mathbb{P}(Y_n>t) = 1$$

*for any $t\in\mathbb{R}_>$.*

By making a more precise statement in the case when $\mathbb{E}\log^+\zeta=\infty$, this result is a slight extension of Cor. 3.3.

*Proof.* It is a simple exercise [☞ Problem 5.23] to verify that (5.10) implies the distributional identity

$$Y_n \stackrel{d}{=} Y_n' := \sum_{k=1}^{n}\sum_{j=1}^{\zeta_k} Z_{k-1}(k,j) \tag{5.14}$$

for any $n\in\mathbb{N}$, where it should be recalled that the $(Z_n(k,j))_{n\ge 0}$ are iid GWP's with one ancestor and offspring distribution $(p_n)_{n\ge 0}$ and independent of the immigration sequence. The $Y_n'$, in contrast to their copies $Y_n$, have nonnegative increments $\Lambda_n := \sum_{k=1}^{\zeta_n} Z_{n-1}(n,k)$ and hence converge almost surely to the random variable

$$Y_\infty' \;:=\; \sum_{n\ge 1}\Lambda_n.$$

taking values in $\overline{\mathbb{N}}_0$. Since the $\Lambda_n$ are obviously independent and a.s. taking values in $\mathbb{N}_0$, the ordinary Borel-Cantelli lemma provides us with the implications

$$\sum_{n\ge 1}\mathbb{P}(\Lambda_n\ge 1)\begin{cases}<\infty\\ =\infty\end{cases}\quad\Rightarrow\quad \mathbb{P}(\Lambda_n\ge 1\text{ i.o.})\begin{cases}=0\\ =1\end{cases}\quad\Rightarrow\quad \mathbb{P}(Y_\infty'<\infty)\begin{cases}=1\\ =0\end{cases}$$

and thereby with

$$\mathbb{P}(Y_\infty'<\infty)\;\in\;\{0,1\}. \tag{5.15}$$

(a) Assume now $\mathbb{E}\log^+\zeta<\infty$ and let $\mathscr{F}_0=\sigma((\zeta_n)_{n\ge 1})$ as before. Since all $Z_{n-1}(n,k)$ are independent of $\mathscr{F}_0$, we infer

$$\mathbb{E}(Y_\infty'|\mathscr{F}_0)\;=\;\sum_{k\ge 1}\mathbb{E}(\Lambda_k|\mathscr{F}_0)\;=\;\sum_{k\ge 1}\mathbb{E}\left(\sum_{k=1}^{\zeta_n}Z_{n-1}(n,k)\middle|\mathscr{F}_0\right)\;=\;\sum_{n\ge 1}\frac{\zeta_n}{\mathsf{m}^{n-1}}\quad\text{a.s.}$$

Now apply Lemma 5.16 to $(\log^+\zeta_n)_{n\ge 1}$ with $c=\mathsf{m}\in(0,1)$ to infer the a.s. finiteness of the series $\sum_{n\ge 1}\mathsf{m}^n e^{\log^+\zeta_n}$ and thus

$$\sum_{n\ge 1}\mathsf{m}^{n-1}\zeta_n\;=\;\mathbb{E}(Y_\infty'|\mathscr{F}_0)\;<\;\infty\quad\text{a.s.}$$

Therefore, we arrive at the desired conclusion via (5.14), viz.

$$Y_n\;\xrightarrow{d}\;Y_\infty'\;<\;\infty\quad\text{a.s.}$$

(b) Let $\mathbb{E}\log^+\zeta=\infty$. Further assuming $\mathbb{P}(Y_\infty'=\infty)<1$ we will produce a contradiction hereafter. By (5.15),

$$\mathbb{P}(Y_\infty=\infty)\;=\;\mathbb{P}\left(\sum_{n\ge 1}\Lambda_n=\infty\right)\;=\;0,$$

which, for $D_n:=\{\Lambda_n\ge 1\}=\{\sum_{k=1}^{\zeta_n}Z_{n-1}(n,k)\ge 1\}$, entails

$$0\;=\;\mathbb{P}\left(\limsup_{n\to\infty}D_n\right)\;=\;\mathbb{P}\left(\limsup_{n\to\infty}D_n\middle|\mathscr{F}_0\right)\quad\text{a.s.}$$

Recalling the independence of all $Z_{n-1}(n,k)$ and $\mathscr{F}_0$, this means that

$$\mathbb{P}\left(\limsup_{n\to\infty} D_n \,\Big|\, \zeta_n = y_n,\, n\ge 1\right) \;=\; \mathbb{P}\left(\limsup_{n\to\infty}\left\{\sum_{k=1}^{y_n} Z_{n-1}(n,k)\ge 1\right\}\right) \;=\; 0$$

for $\mathbb{P}((\zeta_n)_{n\ge1}\in\cdot)$-almost all $\mathbf{y}=(y_n)_{k\ge1}$. For any such $\mathbf{y}$ and all $k,n\in\mathbb{N}$, put $A_{n,k}:=\{Z_{n-1}(n,k)\ge 1\}$ and $B_n:=\bigcup_{k=1}^{y_n}A_{n,k}$. Then we may invoke Lemma 5.20 (with $k_n=y_n$) to infer from $\mathbb{P}(\limsup_{n\to\infty}B_n)=0$ that

$$\begin{aligned}\sum_{n\ge1}\sum_{k=1}^{y_n}\mathbb{P}(A_{n,k}) \;&=\; \sum_{n\ge1}\sum_{k=1}^{y_n}\mathbb{P}(Z_{n-1}(n,k)\ge 1)\\ &=\; \sum_{n\ge1} y_n\,\mathbb{P}(Z_{n-1}(1,1)\ge 1) \;<\; \infty\end{aligned}$$

for $\mathbb{P}((\zeta_n)_{n\ge1}\in\cdot)$-almost all $\mathbf{y}$ and therefore

$$\sum_{n\ge1}\zeta_n\,\mathbb{P}(Z_{n-1}(1,1)\ge 1) \;<\; \infty \quad \text{a.s.}$$

By finally using the simple inequality $e^{\log^+\zeta_n}\le 1+\zeta_n$ and $\mathbb{P}(Z_{n-1}(1,1)\ge 1)\le(1-p_0)^{n-1}$, we obtain (as $0<p_0<1$)

$$\sum_{n\ge1}(1-p_0)^n e^{\log^+\zeta_n} \;<\; \infty \quad \text{a.s.}$$

and thereupon with the help of Lemma 5.16 (with $c=1-p_0$) the contradiction $\mathbb{E}\log^+\zeta<\infty$. So we have shown $\mathbb{P}(Y'_\infty=\infty)=1$, which by another appeal to (5.14) gives

$$\lim_{n\to\infty}\mathbb{P}(Y_n>t) \;=\; \lim_{n\to\infty}\mathbb{P}(Y'_n>t) \;=\; \mathbb{P}(Y'_\infty>t) \;=\; 1$$

for all $t>0$. □

## Problems

**Problem 5.22. [Proof of Lemma 5.17]** Let $(M_n,\mathscr{F}_n)_{n\ge0}$ be as in Lemma 5.17.

(a) If $H=(H_n)_{n\ge1}$ denotes a predictable (with respect to $(\mathscr{F}_n)_{n\ge0}$) sequence of bounded, nonnegative random variables and $M^{(a)}:=((M_n-a)^+)_{n\ge0}$ for $a\in\mathbb{R}$, prove that the martingale transform $H\cdot M^{(a)}=((H\cdot M^{(a)})_n)_{n\ge0}$, defined by $(H\cdot M^{(a)})_0=0$ and $(H\cdot M^{(a)})_n=\sum_{k=1}^n H_k(M_k^{(a)}-M_{k-1}^{(a)})$ for $n\ge1$, satisfies the inequality

$$\mathbb{E}((H\cdot M^{(a)})_{n+1}|\mathscr{F}_n) \;\ge\; (H\cdot M^{(a)})_n \quad \text{a.s.}$$

for any $n\in\mathbb{N}_0$.

(b) For $a,b\in\mathbb{R}$ with $a<b$, let $U_n(a,b)$ denote the number of upcrossings of the interval $[a,b]$ made by $M$ by time $n$ and put $U_\infty(a,b)=\lim_{n\to\infty}U_n(a,b)$. Prove

with the help of (a) [☞ also [36, Ch. 11]] that the *conditional upcrossing inequality*

$$(b-a)\,\mathbb{E}(U_n(a,b)|\mathscr{F}_0) \;\le\; \mathbb{E}((M_n-a)^+|\mathscr{F}_0)-\mathbb{E}((M_0-a)^+|\mathscr{F}_0) \quad (5.16)$$

holds true a.s. for all $n \in \mathbb{N}$ and $a < b$.

(c) Use (5.16) and assumption (ii) of Lemma 5.17 to show that

$$(b-a)\,\mathbb{E}(U_\infty(a,b)|\mathscr{F}_0) \;\le\; |a| \;+\; \sup_{n\ge 0}\mathbb{E}(M_n|\mathscr{F}_0) \;<\; \infty \quad \text{a.s.}$$

and then conclude the a.s. convergence of $M_n$ from this result.

**Problem 5.23.** Prove that (5.10) implies (5.14), for instance, by making use of gf's.

## 5.4 Supercritical GWP's: Another proof of the Kesten-Stigum theorem

We have gathered all necessary ingredients to proceed with a first demonstration of how the theory of GWT's and their size-biasings can be utilized to give alternative proofs of classical results for GWP's. We begin with one of the most prominent results in the supercritical case, the Kesten-Stigum theorem. As already mentioned, the general method is due to LYONS, PEMANTLE & PERES [24] and essentially rests upon the Comparison lemma 5.7, Theorems 5.14, 5.18 and the following very generally formulated lemma about the relation between two probability measures on a filtered probability space.

**Lemma 5.24.** *Let* $\mathbf{P},\mathbf{Q}$ *be two probability measures on a measurable space* $(\mathfrak{X},\mathscr{A})$ *and* $(\mathscr{A}_n)_{n\ge 0}$ *a filtration on* $\mathfrak{X}$ *such that* $\mathscr{A}=\sigma(\mathscr{A}_n, n\ge 0)$. *Let further* $\mathbf{P}_n,\mathbf{Q}_n$ *denote the restrictions of* $\mathbf{P},\mathbf{Q}$ *to* $\mathscr{A}_n$ *and suppose that* $\mathbf{P}_n \ll \mathbf{Q}_n$ *with* $X_n := \frac{d\mathbf{P}_n}{d\mathbf{Q}_n}$ *for each* $n \in \mathbb{N}_0$. *The the following assertions hold true with* $X := \limsup_{n\to\infty} X_n$*:*

*(a)* $\mathbf{P}$ *possesses the decomposition*

$$\mathbf{P}(A) \;=\; \int_A X\,d\mathbf{Q} \;+\; \mathbf{P}(A\cap\{X=\infty\}), \quad A\in\mathscr{A} \qquad (5.17)$$

*in a* $\mathbf{Q}$*-continuous and a* $\mathbf{Q}$*-singular part.*

*(b)* *The following assertions are equivalent:*
- *(b1)* $\mathbf{P} \ll \mathbf{Q}$.
- *(b2)* $X < \infty$ $\mathbf{P}$*-a.s.*
- *(b3)* $\int_{\mathfrak{X}} X\,d\mathbf{Q} = \mathbf{P}(\mathfrak{X}) = 1$.

(c) *By duality, the following assertions are equivalent as well:*
(b1) $\mathbf{P} \perp \mathbf{Q}$.
(b2) $X = \infty$ $\mathbf{P}$-*a.s.*
(b3) $\int_{\mathfrak{X}} X\, d\mathbf{Q} = 0$.

*Proof.* We restrict ourselves to the verification of (a) and leave the rather straightforward arguments for (b) and (c) to the reader [☞ Problem 5.26].

We first show that $(X_n, \mathscr{A}_n)_{n\ge 0}$, called *likelihood process of* $\mathbf{P}$ *with respect to* $\mathbf{Q}$, forms a $\mathbf{Q}$-martingale. Plainly, any $X_n$ is $\mathscr{A}_n$-measurable and also $\mathbf{Q}$-integrable, for $\int X_n\, d\mathbf{Q} = \mathbf{P}(\mathfrak{X}) = 1$. For any $n \in \mathbb{N}_0$ and $A \in \mathscr{A}_n$, we further obtain using $\mathscr{A}_n \subset \mathscr{A}_{n+1}$ that

$$\int_A X_{n+1}\, d\mathbf{Q} \;=\; \mathbf{P}_{n+1}(A) \;=\; \mathbf{P}_n(A) \;=\; \int_A X_n\, d\mathbf{Q}$$

and thus $\mathbb{E}_{\mathbf{Q}}(X_{n+1}|\mathscr{A}_n) = X_n$ $\mathbf{Q}$-a.s., which proves the martingale property.

For the proof of (5.17), we first consider the case when $\mathbf{P} \ll \mathbf{Q}$ and let $Y$ be the $\mathbf{Q}$-density of $\mathbf{P}$. For each $n \in \mathbb{N}_0$ and $A \in \mathscr{A}_n$, we then have

$$\int_A Y\, d\mathbf{Q} \;=\; \mathbf{P}(A) \;=\; \mathbf{P}_n(A) \;=\; \int_A X_n\, d\mathbf{Q}_n \;=\; \int_A X_n\, f\mathbf{Q}$$

and thus infer $X_n = \mathbb{E}_{\mathbf{Q}}(Y|\mathscr{A}_n)$ $\mathbf{Q}$-a.s. This shows that $(X_n)_{n\ge 0}$ is a ui martingale [☞ e.g. [36]] which converges a.s. and in $L^1(\mathbf{Q})$ to $Y$, hence $X = Y = \frac{d\mathbf{P}}{d\mathbf{Q}}$ $\mathbf{Q}$-a.s. and

$$\mathbf{P}(X = \infty) \;=\; \int_{\{X=\infty\}} X\, d\mathbf{Q} \;=\; 0.$$

This clearly proves (5.17).

Turning to the general case, we put $\nu := (\mathbf{P}+\mathbf{Q})/2$ and let $\nu_n$ be the restriction of $\nu$ to $\mathscr{A}_n$, that is $\nu_n = (\mathbf{P}_n + \mathbf{Q}_n)/2$ for each $n \in \mathbb{N}_0$. Then $\mathbf{P} \ll \nu$ as well as $\mathbf{Q} \ll \nu$, whence we may use the first part of the proof for $U_n := \frac{d\mathbf{P}_n}{d\nu_n}$ and $V_n := \frac{d\mathbf{Q}_n}{d\nu_n}$. Define $U := \limsup_{n\to\infty} U_n$ and $V$ accordingly. For each $n \in \mathbb{N}$ and $A \in \mathscr{A}_n$, we have

$$\int_A (U_n + V_n)\, d\nu \;=\; \mathbf{P}(A) + \mathbf{Q}(A) \;=\; 2\nu(A) \;=\; \int_A 2\, d\nu$$

implying

$$U_n + V_n = 2 \quad \nu\text{-a.s.} \tag{5.18}$$

By the same arguments as above, we find that $U_n \to U = \frac{d\mathbf{P}}{d\nu}$ and $V_n \to V = \frac{d\mathbf{Q}}{d\nu}$ $\nu$-a.s., which in combination with (5.18) further shows $\nu(U = V = 0) = 0$. Consequently, $\frac{U}{V}$ is $\nu$-a.s. well-defined, and it follows

$$\frac{U}{V} \;=\; \lim_{n\to\infty} \frac{U_n}{V_n} \;=\; \lim_{n\to\infty} \frac{d\mathbf{P}_n/d\nu_n}{d\mathbf{Q}_n/d\nu_n} \;=\; \lim_{n\to\infty} \frac{d\mathbf{P}_n}{d\mathbf{Q}_n} \;=\; \lim_{n\to\infty} X_n \;=\; X \quad \nu\text{-a.s.}$$

and particularly $\{V=0\}=\{X=\infty\}$ $\nu$-a.s. Finally, we conclude

$$\begin{aligned}\mathbf{P}(A) &= \int_A U\,d\nu\\ &= \int_{\{A\cap\{V>0\}}U\,d\nu + \int_{A\cap\{V=0\}}U\,d\nu\\ &= \int_A XV\,d\nu + \int_{A\cap\{V=0\}}U\,d\nu\\ &= \int_A X\,d\mathbf{Q} + \mathbf{P}(A\cap\{X=\infty\}),\end{aligned}$$

which is (5.17). □

Here is once again the theorem by KESTEN & STIGUM in reduced form, focussing on the main equivalent assertions.

**Theorem 5.25. [Kesten-Stigum]** *Let $(Z_n)_{n\ge 0}$ be a supercritical GWP with one ancestor, finite offspring mean $\mathsf{m}$, extinction probability $q$ and normalization $W_n=\mathsf{m}^{-n}Z_n$ for $n\in\mathbb{N}_0$ with a.s. limit $W$. Then the following assertions are equivalent:*

$$\mathbb{P}(W=0)=q, \tag{5.19}$$

$$\mathbb{E}W=1, \tag{5.20}$$

$$\mathbb{E}Z_1\log Z_1 = \sum_{n\ge 1}p_n\,n\log n < \infty. \tag{ZlogZ}$$

*Proof.* Let the $Z_n$ be given in the form $z_n\circ\boldsymbol{GW}$ for a GWT $\boldsymbol{GW}$. For $n\in\mathbb{N}_0$, we define $\mathscr{E}_n=\{\emptyset,\mathbb{T}\}\cup\{[\tau]_n:\tau\in\mathbb{T}\}$ and $\mathscr{A}_n=\sigma(\mathscr{E}_n)$. We will make use of the previous lemma with $\mathfrak{X}=\mathbb{T}$, $\mathscr{A}=\mathscr{B}(\mathbb{T})$, $\mathbf{P}=\widehat{\mathsf{GW}}$ and $\mathbf{Q}=\mathsf{GW}$. Notice that $\mathscr{A}$ is indeed equal to $\sigma(\mathscr{A}_n,n\ge 0)$ as required there, because $\mathscr{E}$ defined by (4.1) satisfies $\mathscr{E}=\bigcup_{n\ge 0}\mathscr{E}_n$ and $\mathscr{A}=\sigma(\mathscr{E})$ by (4.2).

Consider now, for $n\ge 1$,

$$\Gamma_n(B) := \int_B w_n(\tau)\,\mathsf{GW}(d\tau),\quad B\in\mathscr{B}(\mathbb{T}),$$

which defines a probability measure on $(\mathbb{T},\mathscr{B}(\mathbb{T}))$, the normalization following from

$$\Gamma_n(\mathbb{T}) = \int_{\mathbb{T}} w_n(\tau)\,\mathsf{GW}(d\tau) = \mathsf{m}^{-n}\,\mathbb{E}(w_n\circ\boldsymbol{GW}) = \mathbb{E}W_n = 1.$$

Part (c) of the Comparison lemma 5.7, with $A=[\tau]_n\in\mathscr{E}_n$, provides us with

$$\widehat{\mathsf{GW}}(A) = w_n(\tau)\,\mathsf{GW}([\tau]_n) = \int_{[\tau]_n} w_n(\chi)\,\mathsf{GW}(d\chi) = \Gamma_n(A),$$

where we have utilized that $w_n(\chi) = w_n(\tau)$ for any $\chi \in [\tau]_n$. The measures $\widehat{\mathsf{GW}}$ and $\Gamma_n$ thus coincide, first on $\mathscr{E}_n$, but then in fact on $\mathscr{A}_n = \sigma(\mathscr{E}_n)$ because $\mathscr{E}_n$ is $\cap$-stable and containing $\mathbb{T}$ [☞ [5, Thm. 5.4]]. We have thus found that

$$\widehat{\mathsf{GW}}_{|\mathscr{A}_n} \ll \mathsf{GW}_{|\mathscr{A}_n} \quad \text{and} \quad \frac{\widehat{\mathsf{GW}}_{|\mathscr{A}_n}}{\mathsf{GW}_{|\mathscr{A}_n}} = w_n$$

where we should mention the $\mathscr{A}_n$-measurability of $w_n$ which follows from

$$z_n^{-1}(\{k\}) \;=\; \{\tau \in \mathbb{T} : z_n(\tau) = k\} \;=\; \bigcup_{\tau \in \mathbb{T}_n : |\tau_n| = k} [\tau]_n \;\in\; \mathscr{A}_n$$

for any $k \in \mathbb{N}$.

For the rest of the proof let $w := \limsup_{n\to\infty} w_n$, so that $W = w \circ \boldsymbol{GW}$ $\mathbb{P}$-a.s. Having verified all assumptions of Lemma 5.24, its parts (b) and (c) provide us with the crucial dichotomy: On the one hand,

$$\int_{\mathbb{T}} w\, d\mathsf{GW} = 1 \quad \Leftrightarrow \quad \widehat{\mathsf{GW}} \ll \mathsf{GW} \quad \Leftrightarrow \quad w < \infty \;\widehat{\mathsf{GW}}\text{-a.s.} \tag{5.21}$$

and on the other hand,

$$w = 0 \;\mathsf{GW}\text{-a.s.} \quad \Leftrightarrow \quad \widehat{\mathsf{GW}} \perp \mathsf{GW} \quad \Leftrightarrow \quad w = \infty \;\widehat{\mathsf{GW}}\text{-a.s.} \tag{5.22}$$

The dichotomy will help us here because it relates the distribution of $w$ under $\mathsf{GW}$ with the one under $\widehat{\mathsf{GW}}$, for which the Theorems 5.14 and 5.18 yields further information.

"(ZlogZ)⇒(5.20)" Integrability of $Z_1 \log^+ Z_1$ obviously implies (by means of (5.2)) that

$$\mathbb{E}\log^+(\widehat{X}_1 - 1) \;=\; \mathsf{m}^{-1}\,\mathbb{E}Z_1\log^+(Z_1 - 1) \;<\; \infty,$$

for $\widehat{X}_1$ is a size-biasing of $Z_1$. Denoting by $(Y_n)_{n\ge 0}$ a GWPI on the given probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ with offspring distribution $(p_n)_{n\ge 0}$ and immigration sequence $\zeta_n = \widehat{X}_n - 1$ for $n \ge 1$, we infer from Theorem 5.18 that

$$Y_\infty \;:=\; \lim_{n\to\infty} \frac{Y_n}{\mathsf{m}^n} \;<\; \infty \quad \mathbb{P}\text{-a.s.}$$

Since, by Theorem 5.14, the distributions of $(\mathsf{m}^{-n}Y_n)_{n\ge 0}$ under $\mathbb{P}$ and of $(w_n - \mathsf{m}^{-n})_{n\ge 0}$ under $\widehat{\mathsf{GW}}$ coincide, we further infer upon using $\mathsf{m}^{-n} \to 0$ that

$$\widehat{\mathsf{GW}}(w < \infty) \;=\; \widehat{\mathsf{GW}}\left(\limsup_{n\to\infty}(w_n - \mathsf{m}^{-n}) < \infty\right) \;=\; 1,$$

which, by (5.21), finally shows

$$1 \;=\; \int_{\mathbb{T}} w\, d\mathsf{GW} \;=\; \int_{\Omega} w \circ \boldsymbol{GW}\, d\mathbb{P} \;=\; \mathbb{E}W,$$

that is (5.20).

"(5.20)⇒(5.19)" Clearly, $\mathbb{E}W = 1$ entails $\mathbb{P}(W = 0) < 1$ and thus $\mathbb{P}(W = 0) = q$ by Lemma 1.24.

"(5.19)⇒(ZlogZ)" (by contraposition) If $\mathbb{E}Z_1 \log^+ Z_1 = \infty$, we infer by similar arguments as in the part "(ZlogZ)⇒(5.20)" and by making use of Theorems 5.14 and 5.18 that

$$\widehat{\mathsf{GW}}(w = \infty) = 1$$

and therefore, by an appeal to (5.22), the desired conclusion $\mathsf{GW}(w = 0) = 1$. The details are left to the reader [☞ Problem 5.27]. □

## Problems

**Problem 5.26.** Prove the parts (b) and (c) of Lemma 5.24, of course with the help of part (a).

**Problem 5.27.** Provide the details for the proof of "(5.19)⇒(ZlogZ)" in the above proof of Theorem 5.25.

## 5.5 The limiting behavior of subcritical GWP's

If $(Z_n)_{n\geq 0}$ is a *subcritical* GWP with one ancestor and offspring mean $\mathsf{m}$, the simple estimate

$$\mathbb{P}(Z_n > 0) \;\leq\; \mathbb{E}Z_n \;=\; \mathsf{m}^n$$

for any $n \in \mathbb{N}_0$ provides first evidence on the rate of exponential decay of the survival probability $\mathbb{P}(Z_n > 0)$ to 0. As in Section 2.2, but by different methods again taken from Lyons, Pemantle & Peres [24], we will pursue the question when $\mathsf{m}^n$ is the exact rate of decay of $\mathbb{P}(Z_n > 0)$, which suggests to study the sequences

$$c_n := \mathsf{m}^{-n}\,\mathbb{P}(Z_n > 0) \quad \text{and} \quad \mathsf{m}_n^+ := \mathbb{E}(Z_n | Z_n > 0) = c_n^{-1}$$

for $n \in \mathbb{N}_0$. The already known answer, given in Theorem 2.13 by Kolmogorov, is restated and proved here as part of Theorem 5.29 after the following preparative lemma about size-biased distributions.

**Lemma 5.28.** *For each $n \in \mathbb{N}_0$, let $X_n$ be an integrable $\mathbb{N}$-valued random variable with distribution $\mathbf{P}_n := \mathbb{P}(X_n \in \cdot)$ and associated size-biasing $\widehat{X}_n$, thus $\widehat{\mathbf{P}}_n = \mathbb{P}(\widehat{X}_n \in \cdot)$. Then the following assertions hold true:*

*(a)* *If $(\widehat{\mathbf{P}}_n)_{n\geq 0}$ is tight, then* $\sup_{n\geq 0} \mathbb{E}X_n < \infty$.

*(b) If, in contrast,* $\widehat{X}_n \xrightarrow{\mathbb{P}} \infty$*, i.e.* $\lim_{n\to\infty}\mathbb{P}(\widehat{X}_n \le t) = 0$ *for all* $t \in \mathbb{R}_>$*, then* $\lim_{n\to\infty}\mathbb{E}X_n = \infty$.

*Proof.* (a) The tightness of the $\widehat{\mathbf{P}}_n$ ensures the existence of some $N \in \mathbb{N}$ such that

$$\inf_{n\ge 0}\mathbb{P}(\widehat{X}_n \in \{1,...,N\}) \;=\; \inf_{n\ge 0}\frac{1}{\mathbb{E}X_n}\sum_{k=1}^{n} k\,\mathbf{P}_n(\{k\}) \;\ge\; \frac{1}{2}$$

and therefore

$$\frac{1}{N}\sup_{n\ge 0}\mathbb{E}X_n \;\le\; \sup_{n\ge 0}\frac{\mathbb{E}X_n}{\sum_{k=1}^{N} k\,\mathbf{P}_n(\{k\})} \;\le\; 2.$$

(b) Suppose that $\liminf_{n\to\infty}\mathbb{E}X_n < \infty$ and thus w.l.o.g. (possibly after switching to a subsequence) $M := \sup_{n\ge 0}\mathbb{E}X_n < \infty$. Since then

$$\lim_{n\to\infty}\sum_{k=1}^{N} k\,\mathbf{P}_n(\{k\}) \;\le\; \lim_{n\to\infty}\frac{M}{\mathbb{E}X_n}\sum_{k=1}^{N} k\,\mathbf{P}_n(\{k\}) \;=\; M\lim_{n\to\infty}\mathbb{P}(\widehat{X}_n \le N) \;=\; 0$$

holds true for any $N \in \mathbb{N}$, we obtain

$$\lim_{n\to\infty}\mathbb{P}(X_n \le N) \;\le\; \lim_{n\to\infty}\sum_{k=1}^{N} k\,\mathbf{P}_n(\{k\}) \;=\; 0$$

and thereby

$$\sup_{n\ge 1}\mathbb{E}X_n \;\ge\; \sup_{n\ge 1}\sup_{N\ge 1} N\,\mathbb{P}(X_n > N) \;=\; \infty$$

contradicting our assumption $M < \infty$. ◻

**Theorem 5.29.** *For any GWP* $(Z_n)_{n\ge 0}$ *with finite positive offspring mean* $\mathsf{m}$ *the sequence* $c_n = \mathsf{m}^{-n}\,\mathbb{P}(Z_n > 0)$*,* $n \in \mathbb{N}_0$*, is nonincreasing and hence convergent. Moreover, in the subcritical case* $\mathsf{m} < 1$*, each of the following assertions is equivalent to* (ZlogZ)*:*

$$c \;:=\; \lim_{n\to\infty} c_n \;>\; 0, \tag{5.23}$$

$$\sup_{n\ge 0}\mu_n^+ \;<\; \infty. \tag{5.24}$$

*Proof.* As $\mu_n^+ = c_n^{-1}$, the equivalence of (5.23) and (5.24) is trivial once the monotonicity of the $c_n$ (or the $\mu_n^+$) has been verified.

For $n \in \mathbb{N}_0$, put $\mathbf{P}_n := \mathbb{P}(Z_n \in \cdot | Z_n > 0)$ and let $Z_n$ again be given in the form $z_n \circ \boldsymbol{GW}$ for a GWT $\boldsymbol{GW}$. For $\mathsf{v} \in \boldsymbol{GW}$, define $Z_n(\mathsf{v}) := z_n \circ \Theta_{\mathsf{v}} \circ \boldsymbol{GW}^{\mathsf{v}}$, which is the size of the $n^{th}$ generation of the subpopulation stemming from $\mathsf{v}$ with associated (shifted) subtree $\Theta_{\mathsf{v}} \circ \boldsymbol{GW}^{\mathsf{v}}$ (a copy of $\boldsymbol{GW}$). On $\{Z_n > 0\}$, we further define $\rho_n$ to be the individual of the first generation $\boldsymbol{GW}_1$ with minimal label having descendants in generation $n$, the number being $H_n$. Formally,

$$\rho_n := \inf\{\mathsf{v} \in \boldsymbol{GW}_1 : Z_{n-1}(\mathsf{v}) > 0\} \quad \text{and} \quad H_n := Z_{n-1}(\rho_n)\mathbf{1}_{\{Z_n>0\}}$$

for $n \in \mathbb{N}$. A simple calculation for $k,n \in \mathbb{N}$ yields

$$\begin{aligned}
\mathbb{P}(H_n = k) &= \sum_{i\ge 1} \mathbb{P}(\rho_n = i, Z_n > 0, H_n = k) \\
&= \sum_{i\ge 1} \mathbb{P}(Z_1 \ge i, Z_{n-1}(l) = 0 \text{ for } l < i, Z_{n-1}(i) = k) \\
&= \sum_{i\ge 1} \sum_{j\ge i: p_j > 0} p_j \, \mathbb{P}(Z_{n-1}(l) = 0 \text{ for } l < i, Z_{n-1}(i) = k | Z_1 = j) \\
&= \sum_{i\ge 1} \sum_{j\ge i: p_j > 0} p_j \, \mathbb{P}(Z_{n-1} = k) \, \mathbb{P}(Z_{n-1} = 0)^{i-1} \\
&= \mathbb{P}(Z_{n-1} = k) \sum_{i\ge 1} \mathbb{P}(Z_1 \ge i) \, \mathbb{P}(Z_{n-1} = 0)^{i-1},
\end{aligned}$$

where Prop. 4.13 has been used in the penultimate line. It follows that

$$\mathbb{P}(Z_n > 0) = \sum_{k\ge 1} \mathbb{P}(H_n = k) = \mathbb{P}(Z_{n-1} > 0) \sum_{i\ge 1} \mathbb{P}(Z_1 \ge i) \, \mathbb{P}(Z_{n-1} = 0)^{i-1} \tag{5.25}$$

and then upon taking ratios

$$\mathbb{P}(H_n = k | Z_n > 0) = \frac{\mathbb{P}(H_n = k)}{\mathbb{P}(Z_n > 0)} = \frac{\mathbb{P}(Z_{n-1} = k)}{\mathbb{P}(Z_{n-1} > 0)} = \mathbb{P}(Z_{n-1} = k | Z_{n-1} > 0),$$

for any $k \in \mathbb{N}$, that is $\mathbb{P}(H_n \in \cdot | Z_n > 0) = \mathbf{P}_{n-1}$ for $n \in \mathbb{N}$.

As $H_n \le Z_n$, we now obtain for $k,n \in \mathbb{N}$ that

$$\begin{aligned}
\mathbf{P}_n([k,\infty)) &= \mathbb{P}(Z_n \ge k | Z_n > 0) \\
&\ge \mathbb{P}(H_n \ge k | Z_n > 0) \\
&= \mathbb{P}(Z_{n-1} \ge k | Z_{n-1} > 0) = \mathbf{P}_{n-1}([k,\infty))
\end{aligned}$$

and thereby

$$\mathbb{E}(Z_n | Z_n > 0) = \sum_{k\ge 1} \mathbf{P}_n([k,\infty)) \ge \sum_{k\ge 1} \mathbf{P}_{n-1}([k,\infty)) = \mathbb{E}(Z_{n-1} | Z_{n-1} > 0),$$

that is the monotonicity of the sequence $(\mu_n^+)_{n\ge 0}$.

For the rest of the proof let $\mathsf{m} < 1$. Bringing the size-biasings $\widehat{Z}_n$ into play, the reader is reminded that $\widehat{Z}_n = z_n \circ \widehat{\boldsymbol{GW}}$. By part (d) of the Comparison lemma 5.7 and Problem 5.32, we have $\widehat{Z}_n \stackrel{d}{=} \widehat{\mathbf{P}}_n$, the latter as usual denoting the size-biasing of $\mathbf{P}_n$.

Hence, if $(Y_n)_{n\ge 0}$ is again a GWPI with immigration sequence $\zeta_n = \widehat{X}_n - 1$, $n \ge 1$, and offspring distribution $(p_n)_{n\ge 0}$, then also

$$\widehat{\mathbf{P}}_n \;=\; \mathbb{P}(1+Y_n \in \cdot) \tag{5.26}$$

for any $n \in \mathbb{N}_0$ by Theorem 5.14.

"(5.24)$\Rightarrow$(ZlogZ)" If $(\mu_n^+)_{n\ge 0}$ is bounded, then Lemma 5.28(b) shows that $\widehat{Z}_n$ cannot tend to $\infty$ in probability, nor can $Y_n$ by (5.26). Hence, Theorem 5.21(b) in combination with (5.2) yields

$$\mathbb{E}\log^+\zeta_1 \;=\; \mathbb{E}\log^+(\widehat{X}_1 - 1) \;=\; \mathsf{m}^{-1}\,\mathbb{E}Z_1\log^+(Z_1 - 1) \;<\; \infty \tag{5.27}$$

and thus validity of (ZlogZ).

"(ZlogZ)$\Rightarrow$(5.24)" A look at (5.27) shows that (ZlogZ) is equivalent to the integrability of $\log^+\zeta_1$. Now use part (a) of Theorem 5.21 to infer the a.s. convergence of $Y_n$ and thus the weak convergence of the $\widehat{\mathbf{P}}_n$, in particular the tightness of this sequence. By another appeal to Lemma 5.28, this finally provides us with

$$\sup_{n\ge 0}\int x\,\mathbf{P}_n \;=\; \sup_{n\ge 0}\mathbb{E}(Z_n|Z_n>0) \;<\; \infty,$$

which is the desired conclusion. □

For the next result let $\|\mathbf{P}-\mathbf{Q}\|$ denote the total variation distance of two distributions $\mathbf{P}$ and $\mathbf{Q}$ on $\mathbb{N}_0$ [☞ Appendix A.2 for a short survey of the most relevant facts about $\|\cdot\|$ and coupling]. The next result shows the convergence in total variation of $\mathbf{P}_n = \mathbb{P}(Z_n \in \cdot|Z_n > 0)$. Since, for distributions on $\mathbb{N}_0$, weak convergence and total variation convergence are actually equivalent [☞ Theorem A.8 in the Appendix], this does not improve, but only reconfirm the classical result by YAGLOM stated in Theorem 2.14. On the other hand, (5.28) gives little more than just total variation convergence of the $\mathbf{P}_n$.

**Theorem 5.30. [Yaglom]** *Given a subcritical GWP $(Z_n)_{n\ge 0}$ with one ancestor and $p_0 < 1$, the conditional distributions $\mathbf{P}_n = \mathbb{P}(Z_n \in \cdot|Z_n > 0)$ satisfy*

$$\sum_{n\ge 1}\|\mathbf{P}_n - \mathbf{P}_{n-1}\| \;<\; \infty. \tag{5.28}$$

*In particular,* $\lim_{n\to\infty}\|\mathbf{P}_n - \pi\| = 0$ *for a distribution $\pi$ on $\mathbb{N}_0$ (viz. the quasi-stationary distribution of $(Z_n)_{n\ge 0}$,* ☞ (2.25)*).*

*Proof.* We have shown in the proof of Theorem 5.29, the notation of which is naturally kept here, that the conditional distribution of the there defined random variables $H_n$ given $Z_n > 0$ equals $\mathbf{P}_{n-1}$. Therefore, $(H_n, Z_n)$ provides a coupling of $(\mathbf{P}_{n-1}, \mathbf{P}_n)$ under $\mathbb{P}(\cdot|Z_n > 0)$, and we obtain with the coupling inequality (A.5) in the Appendix that

$$\|\mathbf{P}_n - \mathbf{P}_{n-1}\| \ \le \ \mathbb{P}(H_n \neq Z_n | Z_n > 0).$$

Since the event $\{H_n \neq Z_n\} = \sum_{i \ge 1} \{\rho_n = i, H_n \neq Z_n\}$ can be decomposed as

$$\sum_{i \ge 1} \sum_{j \ge i} \left\{ Z_1 = j, \sum_{l=1}^{i-1} Z_{n-1}(l) = 0, Z_{n-1}(i) > 0, \sum_{l=i+1}^{j} Z_{n-1}(l) > 0 \right\},$$

another application of Prop. 4.13 shows

$$\begin{aligned}
\|\mathbf{P}_n - \mathbf{P}_{n-1}\| \ &\le \ \mathbb{P}(H_n \neq Z_n | Z_n > 0) \\
&= \ \frac{\mathbb{P}(Z_{n-1} > 0)}{\mathbb{P}(Z_n > 0)} \sum_{i \ge 1} \sum_{j \ge i} p_j \, \mathbb{P}(Z_{n-1} = 0)^{i-1} (1 - \mathbb{P}(Z_{n-1} = 0)^{j-i}).
\end{aligned}$$

As one can easily see with the help of gf's,

$$\frac{1}{\mathsf{m}} \ \le \ \frac{\mathbb{P}(Z_{n-1} > 0)}{\mathbb{P}(Z_n > 0)} \ \downarrow \ \frac{1}{\mathsf{m}}, \quad \text{as } n \to \infty$$

so that $\delta := \sup_{n \ge 1} \frac{\mathbb{P}(Z_{n-1} > 0)}{\mathbb{P}(Z_n > 0)} < \infty$. Defining

$$\alpha(k) \ = \ \min\{n \ge 1 : \mathbb{P}(Z_n > 0) < 1/k\}, \quad k \ge 1,$$

we obtain after a suitable rearrangement of terms that

$$\begin{aligned}
\sum_{n \ge 1} \|\mathbf{P}_n - \mathbf{P}_{n-1}\| \ &\le \ \delta \sum_{n \ge 1} \sum_{i \ge 1} \sum_{j \ge i} p_j \, \mathbb{P}(Z_{n-1} = 0)^{i-1} (1 - \mathbb{P}(Z_{n-1} = 0)^{j-i}) \\
&\le \ \delta (I_1 + I_2), \\
\text{where} \quad I_1 \ &:= \ \sum_{j \ge 1} p_j \sum_{n=1}^{\alpha(j)-1} \sum_{i=1}^{j} \mathbb{P}(Z_{n-1} = 0)^{i-1} (1 - \mathbb{P}(Z_{n-1} = 0)^{j-i}) \\
\text{and} \quad I_2 \ &:= \ \sum_{j \ge 1} p_j \sum_{n \ge \alpha(j)} \sum_{i=1}^{j} \mathbb{P}(Z_{n-1} = 0)^{i-1} (1 - \mathbb{P}(Z_{n-1} = 0)^{j-i}).
\end{aligned}$$

The expressions $I_1$ and $I_2$ will now be estimated separately. The following inequality results from the monotonicity of the $c_n$, viz.

$$\mathbb{P}(Z_{\alpha(k)-1} > 0) \ \le \ \mathsf{m}^{\alpha(k)-1-n} \, \mathbb{P}(Z_n > 0)$$

for $0 \le n < \alpha(k)$, and is utilized to give

$$
\begin{aligned}
I_1 &\le \sum_{j\ge 1} p_j \sum_{n=1}^{\alpha(j)-1} \sum_{i=1}^{j} \mathbb{P}(Z_{n-1}=0)^{i-1} \\
&\le \sum_{j\ge 1} p_j \sum_{n=1}^{\alpha(j)-1} \frac{1}{\mathbb{P}(Z_n>0)} \\
&\le \sum_{j\ge 1} p_j \sum_{n=1}^{\alpha(j)-1} \frac{\mathsf{m}^{\alpha(j)-1-n}}{\mathbb{P}(Z_{\alpha(j)-1}>0)} \\
&\le \sum_{j\ge 1} j\, p_j \sum_{n\ge 0} \mathsf{m}^n \qquad\qquad (\text{as } \mathbb{P}(Z_{\alpha(j)-1}>0)\ge 1/j) \\
&= \frac{\mathsf{m}}{1-\mathsf{m}} < \infty.
\end{aligned}
$$

For $I_2$ we proceed in a similar manner. First,

$$
\mathbb{P}(Z_n>0) \le \mathsf{m}^{n-\alpha(k)}\, \mathbb{P}(Z_{\alpha(k)}>0)
$$

for $n \ge \alpha(k)$, which in combination with $1-(1-x)^j \le jx$ for $x\in[0,1]$ and $j\in\mathbb{N}$ yields

$$
\begin{aligned}
I_2 &\le \sum_{j\ge 1} p_j \sum_{n\ge\alpha(j)} \sum_{i=1}^{j} \left(1-\mathbb{P}(Z_{n-1}=0)^{j-i}\right) \\
&\le \sum_{j\ge 1} p_j \sum_{n\ge\alpha(j)} \mathbb{P}(Z_n>0) \sum_{i=1}^{j} (j-i) \\
&\le \sum_{j\ge 1} j^2 p_j \sum_{n\ge 0} \mathsf{m}^n \mathbb{P}(Z_{\alpha(j)}>0) \\
&= \frac{\mathsf{m}}{1-\mathsf{m}} < \infty. \qquad\qquad (\text{as } \mathbb{P}(Z_{\alpha(j)}>0)\le 1/j)
\end{aligned}
$$

So we have verified that

$$
\sum_{n\ge 1} \|\mathbf{P}_n - \mathbf{P}_{n-1}\| \le \delta(I_1+I_2) < \infty
$$

which is (5.28). In particular, $(\mathbf{P}_n)_{n\ge 0}$ forms a Cauchy sequence with respect to $\|\cdot\|$ and thus converges in total variation to a distribution $\pi$. $\square$

## Problems

**Problem 5.31.** Prove the following stronger version of Lemma 5.28: The sequence $(\widehat{\mathbf{P}}_n)_{n\ge 0}$ is tight iff the sequence $(X_n)_{n\ge 0}$ is ui. [Hint: Show that tightness $(\widehat{\mathbf{P}}_n)_{n\ge 0}$ holds iff there exists a nondecreasing, unbounded function $\varphi:\mathbb{R}_\ge \to \mathbb{R}_\ge$ such that

$$\sup_{n\geq 0} \mathbb{E}\varphi(\widehat{X}_n) < \infty.$$

Then use (5.2).]

**Problem 5.32.** Let $X$ be a nonnegative integrable random variable with distribution $\mathbf{P}$ and $\mathbf{Q} := \mathbb{P}(X \in \cdot|X > 0)$. Show that $\mathbf{P}$ and $\mathbf{Q}$ have the same size-biasings, that is $\widehat{\mathbf{P}} = \widehat{\mathbf{Q}}$.

**Problem 5.33.** Let the situation and notation of Theorem 5.29 and its proof be given. Further, let $Z_1'$ be a random variable with $\mathbb{P}(Z_1' = i) = \mathsf{m}^{-1}\,\mathbb{P}(Z_1 \geq i)$ for $i \in \mathbb{N}$. Show with the help of (5.25) that

$$c_n - c_{n+1} \;=\; c_n\left(1 - \mathbb{E}(1-\mathsf{m}^n c_n)^{Z_1'-1}\right) \;\leq\; \mathsf{m}^n\,\mathbb{E}(Z_1'-1)^2$$

(which particularly shows the monotonicity of the $c_n$) and use this to conclude

$$c_n \;=\; c + O(\mathsf{m}^n), \quad n \to \infty,$$

if $(Z_n)_{n\geq 0}$ has finite offspring variance.

**Problem 5.34.** Given the situation and notation of Theorem 5.30 and its proof, prove the following assertions:

(a) If (ZlogZ) holds, then $\alpha(k) \simeq \log k/\log|\mathsf{m}|$ as $k \to \infty$.

(b) For each $p \geq 1$,

$$\sum_{n\geq 1} p_n\, n \log^p n < \infty \;\Rightarrow\; \sum_{n\geq 1} n^p\,\|\mathbf{P}_n - \mathbf{P}_{n-1}\| < \infty \;\Rightarrow\; \lim_{n\to\infty} n^p\,\|\mathbf{P}_n - \pi\| = 0$$

holds true.

(c) For each $\gamma \in (1,\mathsf{m})$,

$$\sum_{n\geq 1} n^{\gamma+1} p_n < \infty \;\Rightarrow\; \sum_{n\geq 1} \gamma^n\,\|\mathbf{P}_n - \mathbf{P}_{n-1}\| < \infty \;\Rightarrow\; \lim_{n\to\infty} \gamma^n\,\|\mathbf{P}_n - \pi\| = 0$$

holds true.

**Problem 5.35.** Given again the situation and notation of Theorem 5.30, recall that

$$p_{ij}^k \;=\; \mathbb{P}(Z_{n+k} = j|Z_n = i) \;=\; \mathbb{P}_i(Z_k = j)$$

denote the $k$-step transition probabilities of the Markov chain $(Z_n)_{n\geq 0}$. A distribution $\pi = (\pi_n)_{n\geq 0}$ is called

— $\beta$*-invariant for* $(Z_n)_{n\geq 0}$ for some $0 < \beta < 1$ if, for all $j \in \mathbb{N}_0$,

$$\beta\pi_j \;=\; \sum_{i\geq 1} \pi_i p_{ij}.$$

Notice that this entails $\pi_0 = 1$, for $p_{00} = 1$.

— *quasi-invariant for* $(Z_n)_{n\ge 0}$ if $\pi = \mathbb{P}_\pi(Z_n \in \cdot | Z_n > 0)$ for all $n \in \mathbb{N}$, i.e.

$$\pi_j \;=\; \frac{\mathbb{P}_\pi(Z_n = j)}{\mathbb{P}_\pi(Z_n > 0)} \;=\; \frac{\sum_{i\ge 1} \pi_i p_{ij}^n}{\sum_{k\ge 1}\sum_{i\ge 1} \pi_i p_{ik}^n}$$

for $j,n \in \mathbb{N}$ [☞ also (2.25)].

— *quasi-stationary for* $(Z_n)_{n\ge 0}$ if $\mathbb{P}_\lambda(Z_n \in \cdot | Z_n > 0) \xrightarrow{w} \pi$ for some initial distribution $\lambda$ on $\mathbb{N}_0$, i.e.

$$\pi_j \;=\; \lim_{n\to\infty} \mathbb{P}_\lambda(Z_n = j | Z_n > 0) \quad \text{for all } j \in \mathbb{N}_0. \tag{5.29}$$

Prove the following assertions for the limiting distribution $\pi$ in Theorem 5.30, which is obviously quasi-stationary with $\lambda = \delta_1$:

(a) $\pi$ is m-invariant.
(b) $\pi$ is quasi-invariant.
(c) $\pi$ satisfies (5.29) under each initial distribution $\lambda$ with $\lambda_0 = 0$. [Hint: Prove (5.29) for any $\lambda = \delta_k$, $k \in \mathbb{N}$, by induction over $k$.]
(d) $\pi$ is the unique quasi-invariant distribution for $(Z_n)_{n\ge 0}$.

## 5.6 And finally critical GWP's again

We close this chapter with an approach towards *critical* GWP's $(Z_n)_{n\ge 0}$ in the framework of GWT's and their size-biasings, again in essence taken from [24]. The following decomposition of $\widehat{Z}_n = z_n \circ \widehat{\boldsymbol{GW}}$ is fundamental.

Keeping the notation of the previous sections, put

$$Z_n(\mathsf{u}) \;:=\; z_n \circ \Theta_u \circ \widehat{\boldsymbol{GW}}^{\mathsf{u}}$$

for $n \in \mathbb{N}_0$ and $\mathsf{u} \in \widehat{\boldsymbol{GW}}$, and then, for $(j,n) \in \mathbb{N}^{2,\le} := \{(k,l) \in \mathbb{N}^2 : k \le l\}$,

$$\begin{aligned} S_{n,j} \;&:=\; \sum_{\rho \in \widehat{\boldsymbol{GW}}_j : \upsilon_{j-1} \preceq \rho \ne \upsilon_j} \widehat{Z}_{n-j}(\rho) \;=\; \widehat{Z}_{n-j+1}(\upsilon_{j-1}) - \widehat{Z}_{n-j}(\upsilon_j) \\ \text{and} \quad R_{n,j} \;&:=\; \sum_{\rho \in \widehat{\boldsymbol{GW}}_j : \upsilon_{j-1} \preceq \rho,\, \upsilon_j < \rho} \widehat{Z}_{n-j}(\rho). \end{aligned}$$

This means that $S_{n,j}$ counts the number of descendants of $\upsilon_{j-1}$ in $\widehat{GW}_n$, i.e. generation $n$, which are not descendants of $\upsilon_j$, while $R_{n,j}$ gives the number of descendants of $\upsilon_{j-1}$ in $\widehat{GW}_n$ whose unique ancestors in generation $j$ are positioned right of $\upsilon_j$ in the size-biased GWT $\widehat{\boldsymbol{GW}}$. Obviously, $R_{n,j} \le S_{n,j}$ for any $(j,n) \in \mathbb{N}^{2,\le}$.

The announced decomposition of $\widehat{Z}_n$ may now be stated as a telescoping sum, viz.

$$\widehat{Z}_n \;=\; 1+\widehat{Z}_n(\upsilon_0)-\widehat{Z}_n(\upsilon_n) \;=\; 1+\sum_{j=1}^{n} S_{n,j}. \tag{5.30}$$

Finally putting

$$A_{n,j} \;:=\; \{R_{n,j}=S_{n,j}\}$$

for $(j,n)\in\mathbb{N}^{2,\le}$, which describes the event that all descendants of $\upsilon_{j-1}$ in generation $n$ that are not descendants of $\upsilon_j$ have an ancestor in generation $j$ to the right of $\upsilon_j$, the subsequent lemma lists a number of relevant properties of the previously defined random variables.

**Lemma 5.36.** *Given a critical GWP $(Z_n)_{n\ge 0}$ with reproduction variance $\sigma^2 > 0$, the following assertions hold true:*

*(a) For each $n\in\mathbb{N}$, the random vectors $(R_{n,1},S_{n,1}),\dots,(R_{n,1},S_{n,1})$ and thus also the events $A_{n,1},\dots,A_{n,n}$ are independent.*
*(b) For each $(j,n)\in\mathbb{N}^{2,\le}$, $\mathbb{E}R_{n,j}=\sigma^2/2$.*
*(c) For each $(j,n)\in\mathbb{N}^{2,\le}$,*

$$\mathbb{P}(A_{n,j})=\frac{\mathbb{P}(Z_{n-j+1}>0)}{\mathbb{P}(Z_{n-j}>0)}.$$

*(d) For each $j\in\mathbb{N}$,*

$$\lim_{n\to\infty}\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}} \;=\; \frac{\sigma^2}{2}$$

*and furthermore, if $\sigma^2$ is finite,*

$$\lim_{n\to\infty}\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}^c} \;=\; 0.$$

*Proof.* (a) For $1\le j\le n$, let $(r_j,s_j)\in\mathbb{N}^{2,\le}$,

$$\begin{aligned} M_j \;&:=\; \{\rho\in\widehat{\boldsymbol{GW}}_j:\rho\succeq\upsilon_{j-1},\,\rho<\upsilon_j\}\\ \text{and}\quad N_j \;&:=\; \{\rho\in\widehat{\boldsymbol{GW}}_j:\rho\succeq\upsilon_{j-1},\,\rho>\upsilon_j\}. \end{aligned}$$

These sets divide the children $\ne\upsilon_j$ of $\upsilon_{j-1}$ into those left and right of $\upsilon_j$ in $\widehat{\boldsymbol{GW}}$.

On the event $D_n:=\{\widehat{X}_i=k_i,\,U_i=\mathsf{v}_i \text{ for } 0\le i<n\}$, we have that

$$\upsilon_j=v_0\dots\mathsf{v}_{j-1}\;\in\;\mathbb{N}^j,$$

and, given $\mathbb{P}(D_n)=\prod_{i=0}^{n-1}p_{k_i}>0$, the random variables

$$\widehat{Z}_{n-j}(\mathsf{u}),\quad 1\le j\le n,\ \mathsf{u}\in M_j\cup N_j$$

are conditionally independent and for fixed $j$ also identically distributed as $\Gamma_{n-j} := \mathbb{P}(Z_{n-j} \in \cdot)$. The last assertion can be deduced from Prop. 5.6. Since

$$\{R_{n,j} = r_j, S_{n,j} = s_j\} \;=\; \left\{\sum_{\mathsf{u}\in M_j} \widehat{Z}_{n-j}(\mathsf{u}) = s_j - r_j,\ \sum_{\mathsf{u}\in N_j} \widehat{Z}_{n-j} = r_j\right\}$$

for any $j$, it follows that

$$\begin{aligned}
&\mathbb{P}(R_{n,j} = r_j, S_{n,j} = s_j,\, 1 \le j \le n)\\
&\quad = \sum_{*} \prod_{i=0}^{n-1} p_{k_i} \prod_{j=1}^{n} \Gamma_{n-j}^{*(\mathsf{u}_{j-1}-1)}\big(\{s_j - r_j\}\big)\, \Gamma_{n-j}^{*(k_{j-1}-\mathsf{u}_{j-1})}\big(\{r_j\}\big)\\
&\quad = \prod_{j=1}^{n} \sum_{1\le \mathsf{u}_{j-1}\le k_{j-1}} p_{k_{j-1}}\, \Gamma_{n-j}^{*(\mathsf{u}_{j-1}-1)}\big(\{s_j - r_j\}\big)\, \Gamma_{n-j}^{*(k_{j-1}-\mathsf{u}_{j-1})}\big(\{r_j\}\big),
\end{aligned}$$

where $\sum_{*}$ denotes summation over all $((\mathsf{u}_0,k_0),...,(\mathsf{u}_{n-1},k_{n-1})) \in (\mathbb{N}^{2,\le})^n$ and the last line is obtained after rearrangement of this sum. This shows the asserted independence of the $(R_{n,j}, S_{n,j})$ for $j = 1,...,n$ with

$$\mathbb{P}(R_{n,j} = r, S_{n,j} = s) \;=\; \sum_{l=1}^{k} p_k\, \Gamma_{n-j}^{*(l-1)}\big(\{s-r\}\big)\, \Gamma_{n-j}^{*(k-l)}\big(\{r\}\big) \tag{5.31}$$

for $(r,s) \in \mathbb{N}^{2,\le}$ and $j = 1,...,n$.

(b) Using (5.31), we obtain

$$\begin{aligned}
\mathbb{E}R_{n,j} \;&=\; \sum_{r\ge 1} r\, \mathbb{P}(R_{n,j} = r)\\
&=\; \sum_{r\ge 1} r \sum_{k\ge 1} \sum_{l=1}^{k} p_k\, \Gamma_{n-j}^{*(k-l)}\big(\{r\}\big)\\
&=\; \sum_{k\ge 1} p_k \sum_{l=1}^{k} \sum_{r\ge 1} r\, \Gamma_{n-j}^{*(k-l)}\big(\{r\}\big)\\
&=\; \sum_{k\ge 1} p_k \sum_{l=1}^{k} (k-l)\, \underbrace{\mathbb{E}Z_{n-j}}_{=1}\\
&=\; \sum_{k\ge 1} p_k\, \frac{k(k-1)}{2} \;=\; \frac{\sigma^2}{2}.
\end{aligned}$$

(c) Here we obtain with (5.31)

$$\begin{aligned}
\mathbb{P}(A_{n,j}) &= \mathbb{P}(R_{n,j} = S_{n,j}) \\
&= \sum_{k\ge 1}\sum_{l=1}^{k} p_k \Gamma_{n-j}^{*(l-1)}\big(\{0\}\big) \\
&= \sum_{k\ge 1} p_k \sum_{l=1}^{k} \mathbb{P}(Z_{n-j}=0)^{l-1} \\
&= \frac{1}{\mathbb{P}(Z_{n-j}>0)} \sum_{k\ge 1} p_k\big(1-\mathbb{P}(Z_{n-j}=0)^k\big) \\
&= \frac{1}{\mathbb{P}(Z_{n-j}>0)} \sum_{k\ge 1} p_k\, \mathbb{P}(Z_{n-j+1}>0|Z_1=k) \\
&= \frac{\mathbb{P}(Z_{n-j+1}>0)}{\mathbb{P}(Z_{n-j}>0)}.
\end{aligned}$$

(d) Once again with the help of (5.31), we find that

$$\begin{aligned}
\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}} &= \sum_{r\ge 1} r\,\mathbb{P}(R_{n,j}=S_{n,j}=r) \\
&= \sum_{r\ge 1} r \sum_{k\ge 1}\sum_{l=1}^{k} p_k \Gamma_{n-j}^{*(l-1)}\big(\{0\}\big)\,\Gamma_{n-j}^{*(k-l)}\big(\{r\}\big) \\
&= \sum_{k\ge 1} p_k \sum_{l=1}^{k} \mathbb{P}(Z_{n-j}=0)^{l-1} \sum_{r\ge 1} r\,\Gamma_{n-j}^{*(k-l)}\big(\{r\}\big) \\
&= \sum_{k\ge 1} p_k \sum_{l=1}^{k} (k-l)\,\mathbb{P}(Z_{n-j}=0)^{l-1}.
\end{aligned}$$

which implies

$$\lim_{n\to\infty} \mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}} = \sum_{k\ge 1} p_k \sum_{l=1}^{k}(k-l) = \frac{\sigma^2}{2}$$

by an appeal to the monotone convergence theorem (using $\mathbb{P}(Z_n=0)\uparrow 1$) and the calculation in part (b). If $\sigma^2<\infty$, this further entails

$$\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}^c} = \frac{\sigma^2}{2} - \mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}} \to 0,$$

as $n\to\infty$, and the proof is complete. □

*Remark 5.37.* By similar arguments as in the previous proof, it can be shown that $\mathbb{E}S_{n,j}=\sigma^2$ for any $(j,n)\in\mathbb{N}^{2,\le}$. Since this will not be needed hereafter, we leave its proof as an exercise for the reader [☞ Problem 5.44].

The above proof has shown that $\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}}$, $\mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}^c}$ and $\mathbb{P}(A_{n,j})$ depend on $(j,n)$ only through its difference $n-j$. Therefore, we can define

$$\alpha_{n-j} := \mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}}, \quad \beta_{n-j} := \mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}^c} \quad\text{and}\quad \gamma_{n-j} := \mathbb{P}(A_{n,j}),$$

for which Lemma 5.36 provides us with

$$\lim_{n\to\infty}\alpha_n = \frac{\sigma^2}{2}, \quad \lim_{n\to\infty}\gamma_n = 1 \quad \text{and} \quad \lim_{n\to\infty}\beta_n = 0 \tag{5.32}$$

where the last statement further requires $\sigma^2 < \infty$.

We are now able to give probabilistic proof of (2.31) and (2.32) in Theorem 2.24, first derived by KOLMOGOROV [18] under $\mathbb{E}Z_1^3 < \infty$ and later in full generality by KESTEN, NEY & SPITZER [16].

**Theorem 5.38. [Kolmogorov]** *Let $(Z_n)_{n\ge 0}$ be a critical GWP with finite and positive reproduction variance $\sigma^2$. Then*

$$\lim_{n\to\infty} n\,\mathbb{P}(Z_n > 0) \;=\; \lim_{n\to\infty}\mathbb{E}\left(\frac{Z_n}{n}\middle| Z_n > 0\right)^{-1} \;=\; \frac{2}{\sigma^2}$$

*with the usual convention that $\frac{1}{\infty} := 0$.*

*Proof.* For $n \in \mathbb{N}$, we define

$$R_n \;=\; 1 + \sum_{j=1}^{n} R_{n,j}.$$

Since $R_{n,j}$ counts the number of individuals in $\widehat{\boldsymbol{GW}}_n$ which are greater than $\upsilon_n$ and stemming from $\upsilon_{j-1}$, but not from $\upsilon_j$, we see that $R_n$ just accounts for the total number of individuals greater or equal to $\upsilon_n$.

For a better understanding of the subsequent arguments that are again based on a coupling, we first give the following informations: It is straightforward to infer from Lemma 5.36(b) that $n^{-1}\,\mathbb{E}R_n$ has the same limit, viz. $\sigma^2/2$, as asserted for $\mathbb{E}(n^{-1}Z_n|Z_n > 0)$. Therefore, we will proceed with the construction of random variables $R_n^*$, $n \in \mathbb{N}$ satisfying:

(i) $R_n^* \stackrel{d}{=} \mathbf{P}_n = \mathbb{P}(Z_n \in \cdot|Z_n > 0)$ for any $n \ge 1$ [☞ (5.37)].
(ii) If $\sigma^2 < \infty$, then $n^{-1}\,\mathbb{E}|R_n - R_n^*| \to 0$ as $n \to \infty$ [☞ (5.32)].

However, this construction is by no means obvious and requires at first the insight that $\widehat{Z}_n$ conditioned upon the event

$$A_n \;:=\; \{\upsilon_n \text{ is leftmost individual in } \widehat{\boldsymbol{GW}}_n\} \;=\; \{\upsilon_n = \min \widehat{\boldsymbol{GW}}_n\}$$

has conditional distribution $\mathbf{P}_n$ as well [☞ (5.34)]. As $R_{n,j} \le S_{n,j}$ for $(j,n) \in \mathbb{N}^{2,\le}$, we have

$$A_n \;=\; \{R_n = \widehat{Z}_n\} \;=\; \bigcap_{j=1}^{n}\{R_{n,j} = S_{n,j}\} \;=\; \bigcap_{j=1}^{n} A_{n,j} \tag{5.33}$$

and thus infer with the help of Lemma 5.36 that

$$\mathbb{P}(A_n) \;=\; \prod_{j=1}^{n}\mathbb{P}(A_{n,j}) \;=\; \prod_{j=1}^{n}\frac{\mathbb{P}(Z_{n-j+1}>0)}{\mathbb{P}(Z_{n-j}>0)} \;=\; \frac{\mathbb{P}(Z_n>0)}{\mathbb{P}(Z_0>0)} \;=\; \mathbb{P}(Z_n>0).$$

By combining this with the calculation

$$\begin{aligned}\mathbb{P}(\{\widehat{Z}_n=k\}\cap A_n) \;&=\; \mathbb{P}(\widehat{Z}_n=k,\, \upsilon_n=\min\widehat{\boldsymbol{GW}}_n)\\ &=\; \sum_{\tau\in\mathbb{T}_n:z_n(\tau)=k}\widehat{\mathsf{GW}}_*\big([\tau;\min\tau_n]_n\big)\\ &=\; \sum_{\tau\in\mathbb{T}_n:z_n(\tau)=k}\widehat{\mathsf{GW}}([\tau]_n)\\ &=\; \mathbb{P}(Z_n=k),\end{aligned}$$

where the Comparsion lemma 5.7 has been utilized in the penultimate line, we arrive at the announced distributional identity

$$\mathbb{P}(\{\widehat{Z}_n\in\cdot\}\cap A_n) \;=\; \mathbb{P}(Z_n\in\cdot|Z_n>0) \;=\; \mathbf{P}_n. \tag{5.34}$$

Now we can turn to the construction of the $R_n^*$ with distribution $\mathbf{P}_n$. For $n\ge 1$, let $R'_{n,1},\dots,R'_{n,n}$ be independent random variables on $(\Omega,\mathfrak{A},\mathbb{P})$ which are independent of $(R_{n,1},S_{n,1}),\dots,(R_{n,n},S_{n,n})$ and such that

$$\mathbb{P}(R'_{n,j}\in\cdot) \;=\; \mathbb{P}(R_{n,j}\in\cdot|A_{n,j}) \;=\; \mathbb{P}(S_{n,j}\in\cdot|A_{n,j})$$

for $1\le j\le n$. Here $\mathbb{P}(A_{n,j})=\gamma_{n-j}>0$ should be recalled. Then defining

$$R_{n,j}^* \;=\; R_{n,j}\mathbf{1}_{A_{n,j}}+R'_{n,j}\mathbf{1}_{A_{n,j}^c},\quad 1\le j\le n \tag{5.35}$$

and further

$$R_n^* \;=\; 1+\sum_{j=1}^{n}R_{n,j}^*, \tag{5.36}$$

the random variables $R_{n,1}^*,\dots,R_{n,n}^*$ are also independent by Lemma 5.36, and we with the help of the stated independence assumptions that

$$\begin{aligned}\mathbb{P}(R_{n,j}^*=k) \;&=\; \mathbb{P}(\{R_{n,j}=k\}\cap A_{n,j})+\mathbb{P}(\{R'_{n,j}=k\}\cap A_{n,j}^c)\\ &=\; \mathbb{P}(\{S_{n,j}=k\}\cap A_{n,j})+\mathbb{P}(R'_{n,j}=k)\,\mathbb{P}(A_{n,j}^c)\\ &=\; \mathbb{P}(S_{n,j}=k|A_{n,j})\big(\mathbb{P}(A_{n,j})+\mathbb{P}(A_{n,j}^c)\big)\\ &=\; \mathbb{P}(S_{n,j}=k|A_{n,j})\end{aligned}$$

for $k\in\mathbb{N}$ and $1\le j\le n$ and thereupon, by (5.30) and (5.33), that

$$\begin{aligned}
\mathbb{P}(R_n^*=k) &= \sum_{k_1+\ldots+k_n=k-1} \mathbb{P}(R_{n,j}^*=k_j \text{ for } 1\le j\le n) \\
&= \sum_{k_1+\ldots+k_n=k-1} \prod_{j=1}^n \mathbb{P}(R_{n,j}^*=k_j) \\
&= \sum_{k_1+\ldots+k_n=k-1} \prod_{j=1}^n \mathbb{P}(S_{n,j}^*=k_j|A_{n,j}) \\
&= \frac{1}{\mathbb{P}(A_n)} \sum_{k_1+\ldots+k_n=k-1} \mathbb{P}(\{S_{n,j}^*=k_j \text{ for } 1\le j\le n\}\cap A_{n,j}) \\
&= \frac{1}{\mathbb{P}(A_n)} \mathbb{P}(\{\widehat{Z}_n=k\}\cap A_n) \\
&= \mathbb{P}(\widehat{Z}_n=k|A_n).
\end{aligned}$$

In view of (5.34), this yields as announced in (i) above that

$$\mathbb{P}(R_n^*\in\cdot) \;=\; \mathbf{P}_n \;=\; \mathbb{P}(Z_n\in\cdot|Z_n>0). \tag{5.37}$$

Next, we must verify (ii) and do so first under the assumption $\sigma^2<\infty$. By the definition of the $R_{n,j}^*$ and another appeal to Lemma 5.36, we then obtain

$$\begin{aligned}
\frac{1}{n}\left|\mathbb{E}(R_n-R_n^*)\right| &\le \frac{1}{n}\mathbb{E}|R_n-R_n^*| \\
&\le \frac{1}{n}\sum_{j=1}^n \mathbb{E}|R_{n,j}-R_{n,j}^*| \\
&\le \frac{1}{n}\sum_{j=1}^n \int_{A_{n,j}^c} (R_{n,j}+R_{n,j}^*)\,d\mathbb{P} \\
&= \frac{1}{n}\sum_{j=1}^n \int_{A_{n,j}^c} (R_{n,j}+R_{n,j}')\,d\mathbb{P} \\
&= \frac{1}{n}\sum_{j=1}^n \big(\beta_{n-j}+\mathbb{P}(A_{n,j}^c)\,\mathbb{E}R_{n,j}'\big) \\
&= \frac{1}{n}\sum_{j=1}^n \left(\beta_{n-j}+\frac{1-\gamma_{n-j}}{\gamma_{n-j}}\,\mathbb{E}R_{n,j}\right) \\
&= \frac{1}{n}\sum_{j=1}^n \left(\beta_{n-j}+\frac{1-\gamma_{n-j}}{\gamma_{n-j}}\,\frac{\sigma^2}{2}\right).
\end{aligned} \tag{5.38}$$

But the sum in (5.38) converges to 0, for it is the $n^{th}$ Césaro mean of a null sequence [☞ (5.32)]. Consequently,

$$\left|\frac{\mathbb{E}R_n^*}{n}-\frac{\sigma^2}{2}\right| \le \frac{|\mathbb{E}(R_n^*-R_n)|}{n} + \left|\frac{\mathbb{E}R_n}{n}-\frac{\sigma^2}{2}\right|$$
$$= o(1) + \left|\frac{1}{n}+\underbrace{\frac{1}{n}\sum_{j=1}^n \mathbb{E}R_{n,j}}_{=\sigma^2/2}-\frac{\sigma^2}{2}\right| = o(1)$$

as $n\to\infty$, which finally leads to

$$\frac{1}{n\,\mathbb{P}(Z_n>0)} = \frac{\mathbb{E}(Z_n|Z_n>0)}{n} = \frac{\mathbb{E}R_n^*}{n} = \frac{\sigma^2}{2}+o(1), \quad n\to\infty.$$

Left with the case $\sigma^2=\infty$, it follows from (5.35) and (5.36) that

$$R_n^* \ge \sum_{j=1}^n R_{n,j}\mathbf{1}_{A_{n,j}}$$

and thereby furthermore

$$\frac{1}{n\,\mathbb{P}(Z_n>0)} = \frac{\mathbb{E}R_n^*}{n} \ge \frac{1}{n}\sum_{j=1}^n \mathbb{E}R_{n,j}\mathbf{1}_{A_{n,j}} = \frac{1}{n}\sum_{j=0}^{n-1}\alpha_j \to \frac{\sigma^2}{2} = \infty,$$

as $n\to\infty$, where we have used that Césaro's theorem remains valid for sequences with limit $\infty$. □

*Remark 5.39.* Assertion (ii) in the previous proof, valid if $\sigma^2<\infty$, just states the $L^1$-convergence of $n^{-1}(R_n-R_n^*)$ to 0 and particularly implies that

$$\frac{R_n}{n}-\frac{R_n^*}{n} \stackrel{\mathbb{P}}{\to} 0, \quad \text{as } n\to\infty.$$

This fact will be utilized in the proof of Theorem 5.43, for it ensures by Slutky's theorem that convergence in distribution of $n^{-1}R_n$ and of $n^{-1}R_n^*$ are equivalent conditions, with necessarily identical limit distribution.

In order to also give an alternative proof of the last statement (2.33) in Theorem 2.24 due to YAGLOM, viz. the weak convergence of the conditional distribution of $n^{-1}Z_n$ given $Z_n>0$ to an exponential law with parameter $2/\sigma^2$, three subsequent results provide the necessary prerequisites. The first one sheds light on the distributional relation between $R_n$ and $\widehat{Z}_n$. Roughly speaking, $R_n$ is obtained from $\widehat{Z}_n$ by scaling with a factor having a uniform distribution on $(0,1)$.

**Lemma 5.40.** *Let $U$ be independent of $(\widehat{Z}_n)_{n\ge0}$ with $U \stackrel{d}{=} Unif(0,1)$. Then*

$$R_n \stackrel{d}{=} \lceil U\widehat{Z}_n\rceil$$

*for each $n \in \mathbb{N}$, where as usual $\lceil x \rceil := \min\{n \in \mathbb{Z} : n \geq x\}$ for $x \in \mathbb{R}$.*

*Proof.* We have seen in the proof of Theorem 5.38 that $R_n - 1$ counts the number of nodes in $\widehat{\boldsymbol{GW}}_n$ lying to the right of $\upsilon_n$, in particular $R_n \leq \widehat{Z}_n$. Given any $\tau \in \mathbb{T}$ with $z_n(\tau) = m$, the nodes in $\tau_n$ increasingly ordered are denoted as $\rho_n^{(1)}(\tau), ..., \rho_n^{(m)}(\tau)$. By another appeal to the Comparison lemma 5.7, we infer for $k \geq 1$ that

$$\begin{aligned}
\mathbb{P}(R_n = k) &= \sum l \geq k \mathbb{P}(\widehat{Z}_n = l, R_n = k) \\
&= \sum_{l \geq k} \sum_{\tau \in \mathbb{T}_n : z_n(\tau) = l} \widehat{\mathsf{GW}}_*([\tau; \rho_n^{(l-k+1)}(\tau)]_n) \\
&= \sum_{l \geq k} \sum_{\tau \in \mathbb{T}_n : z_n(\tau) = l} \mathsf{GW}([\tau]_n) \\
&= \sum_{l \geq k} \mathbb{P}(Z_n = l) \\
&= \sum_{l \geq k} \mathbb{P}(\widehat{Z}_n = l) \underbrace{\mathbb{P}\left(\frac{k-1}{l} < U \leq \frac{k}{l}\right)}_{=1/l} \\
&= \mathbb{P}(k - 1 < U\widehat{Z}_n \leq k) \\
&= \mathbb{P}(\lceil U\widehat{Z}_n \rceil = k).
\end{aligned}$$

This obviously proves the assertion. ◻

The second lemma, the proof of which we leave to the reader as an exercise [☞ Problem 5.45], shows compatibility of distributional convergence ands size-biasing.

**Lemma 5.41.** *Let $(X_n)_{n \geq 0}$ be a sequence of nonnegative random variables with finite positive means and associated sequence $(\widehat{X}_n)_{n \geq 0}$ of size-biasings. Suppose that $X_n \xrightarrow{d} X_0$ and that $\widehat{X}_n$ converges in distribution as well. Then $\widehat{X}_n \xrightarrow{d} \widehat{X}_0$.*

Two characterizations of the exponential distribution are the content of the third preparative result and of interest also in their own right. One of these involves once again a distributional equation which is further discussed in Problem 5.46 and 5.47. For more detailed information including extensions to the multidimensional case we refer to the articles by KOTZ & STEUTEL [19], PAKES & KHATTREE [30] and PAKES [29].

**Proposition 5.42.** *Let $X, X_1, X_2$ be iid nonnegative random variables with mean $\mu \in \mathbb{R}_>$, $\widehat{X}$ a size-biasing of $X$ and $U$ a further Unif$(0,1)$-variable independent of the afore-mentioned variables. Then the following three statements are equivalent:*

*(a)* $X \stackrel{d}{=} Exp(1/\mu)$. *(b)* $X \stackrel{d}{=} U\widehat{X}$. *(c)* $X \stackrel{d}{=} U(X_1 + X_2)$.

*Proof.* We first prove "(a)⇔(b)" and let $\varphi$ denote the LT of $X$. Since LT's determine the distribution uniquely, assertion (b) is equivalent to [use (5.2) and Fubini's theorem]

$$\begin{aligned}\varphi(t) \;&=\; \mathbb{E}e^{-tU\widehat{X}} \;=\; \int_0^1 \mathbb{E}e^{-tu\widehat{X}}\, du\\ &=\; \frac{1}{\mu}\int_0^1 \mathbb{E}Xe^{-tuX}\, du\\ &=\; \frac{1}{\mu}\mathbb{E}\left(X\int_0^1 e^{-tuX}\, du\right) \;=\; \frac{1-\varphi(t)}{\mu t}\end{aligned}$$

for all $t \in \mathbb{R}_>$ and thus

$$\varphi(t) \;=\; \frac{1/\mu}{t+1/\mu}, \quad t \in \mathbb{R}_>.$$

This gives (a) by another appeal to the uniqueness theorem for LT's.

"(a), (b)⇒(c)" may be assessed by a comparison of $\lambda\!\!\lambda$-densities [☞ 5.3 for the density of $\widehat{X}$] which immediately gives $\widehat{X} \stackrel{d}{=} X_1 + X_2 \stackrel{d}{=} \Gamma(2, 1/\mu)$.

"(c)⇒(a)" Since $X_1 + X_2$ has LT $\varphi^2$, we find the equivalence of (c) with

$$\varphi(t) \;=\; \int_0^1 \varphi(ut)^2\, du \;=\; \frac{1}{t}\int_0^t \varphi(u)^2\, du, \quad t \in \mathbb{R}_>, \tag{5.39}$$

which yields upon differentiation

$$\varphi'(t) \;=\; -\frac{1}{t^2}\int_0^t \varphi(u)^2\, du \;+\; \frac{\varphi(t)^2}{t}, \quad t \in \mathbb{R}_>. \tag{5.40}$$

A combination of both, (5.39) and (5.40), provides us with the differential equation

$$t\varphi'(t) + \varphi(t) \;=\; \varphi(t)^2, \quad t \in \mathbb{R}_>. \tag{5.41}$$

By next considering the function $\Psi := (1-\varphi)/\varphi$, positive on $\mathbb{R}_>$ and with derivative $\Psi' = -\varphi'/\varphi^2$, it follows from (5.41) that

$$\frac{\Psi'(t)}{\Psi(t)} \;=\; \frac{\varphi'(t)}{\varphi(t)^2 - \varphi(t)} \;=\; \frac{1}{t}, \quad t \in \mathbb{R}_>$$

and then upon integration over arbitrary intervals $(s,t) \subset \mathbb{R}_>$ that

$$\log\left(\frac{\Psi(t)}{\Psi(s)}\right) = \int_s^t \frac{\Psi'(u)}{\Psi(u)}\,du = \int_s^t \frac{1}{u}\,du = \log\left(\frac{t}{s}\right),$$

that is $t^{-1}\Psi(t) = s^{-1}\Psi(s)$ for all $t > s \in \mathbb{R}_>$. Consequently, $\Psi(t) = at$ on $\mathbb{R}_>$ for some $a \in \mathbb{R}_>$ or, equivalently,

$$\varphi(t) = \frac{1}{1+\Psi(t)} = \frac{1}{1+at} = \frac{1/a}{t+1/a}, \quad t \in \mathbb{R}_>,$$

which means that $X$ must have an exponential law with parameter $1/a$. But $a$ then is the expectation of $X$ and thus $a = \mu$. ☐

Now we are ready to give a probabilistic proof of the convergence result (2.33) due to YAGLOM [37].

**Theorem 5.43. [Yaglom]** *For a critical GWP $(Z_n)_{n\ge 0}$ with offspring variance $\sigma^2 \in \mathbb{R}_>$, the conditional law of $Z_n$ given $Z_n > 0$ converges weakly to an exponential law with parameter $2/\sigma^2$, that is*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Z_n}{t} \le t \,\middle|\, Z_n > 0\right) = 1 - e^{-2t/\sigma^2}$$

*for any $t \in \mathbb{R}_>$.*

*Proof.* We keep the notation from Theorem 5.38 and its proof. By (5.37), we must show $n^{-1}R_n^* \xrightarrow{d} \mathrm{Exp}(2/\sigma^2)$. Since $\sigma^2$ is finite, we know that $\mathbb{E}R_n^*/n \to \sigma^2/2$ [☞ proof of Theorem 5.38], $\mathbb{E}\widehat{Z}_n = \mathbb{E}Z_n^2/\mathbb{E}Z_n = \mathbb{V}\mathrm{ar}\,Z_n + 1 = n\sigma^2 + 1$ [☞ Prop. 1.4],

$$\sup_{n\ge 1}\frac{\mathbb{E}R_n}{n} = \frac{\sigma^2}{2} + 1 < \infty, \quad \kappa := \sup_{n\ge 1}\frac{\mathbb{E}R_n^*}{n} < \infty \quad \text{and} \quad \sup_{n\ge 1}\frac{\mathbb{E}\widehat{Z}_n}{n} = \sigma^2 + 1 < \infty,$$

which easily implies tightness of the distribution families

$$\left\{\mathbb{P}\left(\frac{R_n}{n} \in \cdot\right) : n \ge 1\right\}, \left\{\mathbb{P}\left(\frac{R_n^*}{n} \in \cdot\right) : n \ge 1\right\} \text{ and } \left\{\mathbb{P}\left(\frac{\widehat{Z}_n^*}{n} \in \cdot\right) : n \ge 1\right\}.$$

The Helly-Bray selection theorem in combination with $n^{-1}(R_n - R_n^*) \xrightarrow{\mathbb{P}} 0$ [☞ Remark 5.39] ensures the existence of random variables $S, T$, w.l.o.g. defined on the same probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, such that

$$\frac{R_{n_k}}{n_k} \xrightarrow{d} S, \quad \frac{R_{n_k}^*}{n_k} \xrightarrow{d} S \quad \text{and} \quad \frac{\widehat{Z}_{n_k}^*}{n_k} \xrightarrow{d} T \tag{5.42}$$

for a suitable subsequence $(n_k)_{k\ge 1}$. Moreover, using (5.37), we find that

$$\begin{aligned}\mathbb{P}(\widehat{Z}_n/n\le t) &= \frac{\mathbb{E}Z_n\mathbf{1}_{\{Z_n/n\le t\}}}{\mathbb{E}Z_n}\\ &= \frac{\mathbb{E}(Z_n\mathbf{1}_{\{Z_n/n\le t\}}|Z_n>0)}{\mathbb{E}(Z_n|Z_n>0)}\\ &= \frac{\mathbb{E}(\mathbf{1}_{\{R_n^*/n\le t\}}R_n^*/n)}{\mathbb{E}R_n^*/n}\end{aligned}$$

for $t\in\mathbb{R}_>$ and $n\in\mathbb{N}$, so that $\widehat{Z}_n/n$ is a size-biasing of $R_n^*/n$, in particular

$$\mathbb{E}(\mathbf{1}_{\{R_n^*/n>t\}}R_n^*/n) \le \kappa\,\mathbb{P}(\widehat{Z}_n/n>t)$$

for $t\in\mathbb{R}_>$ and $n\in\mathbb{N}$. Now the tightness of $\{\mathbb{P}(\widehat{Z}_n/n\in\cdot):n\ge 1\}$ entails the uniform integrability of the sequence $(R_n^*/n)_{n\ge 1}$ and thus

$$\mathbb{E}S = \lim_{k\to\infty}\frac{\mathbb{E}R_{n_k}^*}{n_k} = \frac{\sigma^2}{2} \in \mathbb{R}_>. \tag{5.43}$$

Next, an application of Lemma 5.41 (with $X_0=S$, $X_k=R_{n_k}^*/n_k$ and $\widehat{X}_k=\widehat{Z}_{n_k}/n_k$ for $k\ge 1$) yields $T\stackrel{d}{=}\widehat{S}$ for any size-biasing $\widehat{S}$ of $S$ and therefore

$$\frac{\widehat{Z}_{n_k}}{n_k} \stackrel{d}{\to} \widehat{S},\quad k\to\infty.$$

Let $U$ be a $\mathit{Unif}(0,1)$-variable independent of $\widehat{S}$ and $(\widehat{Z}_n)_{n\ge 0}$. Then

$$\frac{U\widehat{Z}_{n_k}}{n_k} \stackrel{d}{\to} U\widehat{S},\quad k\to\infty.$$

Since

$$0 \le \Delta_n := \frac{\lceil U\widehat{Z}_n\rceil}{n}-\frac{U\widehat{Z}_n}{n} \le \frac{1}{n} \to 0,\quad n\to\infty$$

we infer with the help of Lemma 5.40 that

$$\frac{R_{n_k}}{n_k} \stackrel{d}{=} \frac{\lceil U\widehat{Z}_{n_k}\rceil}{n_k} = \frac{U\widehat{Z}_{n_k}}{n_k}+\Delta_{n_k} \stackrel{d}{\to} U\widehat{S},\quad k\to\infty.$$

By combining this with the first statement in (5.42), we see that $S\stackrel{d}{=}U\widehat{S}$ which in turn shows $S\stackrel{d}{=}\mathit{Exp}(2/\sigma^2)$ by invoking Prop. 5.42 and recalling $\mathbb{E}S=\sigma^2/2$ from above [☞ (5.43)].

The proof is finally completed by noting that the previous arguments obviously apply to *any* subsequence $(n_k)_{k\ge 1}$ such that (5.42) holds. In other words, any distributionally convergent $n_k^{-1}R_{n_k}^*$ has limit $Exp(2/\sigma^2)$ whence $n^{-1}R_n^*$ itself converges to this distribution, as required. $\square$

## Problems

**Problem 5.44.** Prove that $\mathbb{E}S_{n,j} = \sigma^2$ for any $(j,n) \in \mathbb{N}^{2,\leq}$ as claimed in Remark 5.37.

**Problem 5.45.** Prove Lemma 5.41. [Hint: Argue that it suffices to show vague convergence of $\mathbb{P}(\widehat{X}_n \in \cdot)$ to $\mathbb{P}(\widehat{X}_0 \in \cdot)$ and use Problem 5.31.]

**Problem 5.46.** Consider the situation of Prop. 5.42 and let $\mathbf{T}_2 = \bigcup_{n\geq 0}\{1,2\}^n$ be the infinite binary tree. Then let $\{U_\mathsf{v} : \mathsf{v} \in \mathbf{T}_2\}$ be family of iid random variables with uniform distribution on $[0,1]$. For $\mathsf{v} = \mathsf{v}_1...\mathsf{v}_n \in \mathbf{T}_2$, put

$$T(\mathsf{v}) \;=\; (T_1(\mathsf{v}), T_2(\mathsf{v})) \;:=\; (U_\mathsf{v}, U_{\mathsf{v})})$$

and then

$$L(\mathsf{v}) \;:=\; T_{\mathsf{v}_1}(\varnothing)T_{\mathsf{v}_2}(\mathsf{v}_1)\cdot\ldots\cdot T_{\mathsf{v}_n}(\mathsf{v}_1...\mathsf{v}_{n-1}) \;=\; U_\varnothing U_{\mathsf{v}_1}\cdot\ldots\cdot U_{\mathsf{v}_1...\mathsf{v}_{n-1}}.$$

This may be interpreted as follows: To each edge $(\mathsf{v}, \mathsf{v}i)$ in $\mathbf{T}_2$ the weight $T_i(\mathsf{v})$ is assigned, which upon multiplication of edge weights leads to a total weight $L(\mathsf{v})$ for the unique path connecting the root with $\mathsf{v}$.

Finally, let $\{X_\mathsf{v} : \mathsf{v} \in \mathbf{T}_2\}$ be a second family of iid random variables with common distribution $F$, finite and positive mean $\mu$, and independent of $\{U_\mathsf{v} : \mathsf{v} \in \mathbf{T}_2\}$. Prove the following assertions:

(a) If $F = Exp(1/\mu)$, then

$$X_\varnothing \stackrel{d}{=} \sum_{\mathsf{v}\in\mathbf{T}:|v|=n} L(\mathsf{v})X_\mathsf{v}$$

for all $n \in \mathbb{N}$.

(b) For general $F$ with expectation $\mu$ and finite variance

$$\sum_{\mathsf{v}\in\mathbf{T}:|v|=n} L(\mathsf{v})X_\mathsf{v} \stackrel{d}{\to} Exp(1/\mu) \quad \text{as } n\to\infty.$$

[Hint: Consider $\Delta_n := \sum_{\mathsf{v}\in\mathbf{T}:|v|=n} L(\mathsf{v})(X_\mathsf{v} - X_\mathsf{v}')$ for a family $\{X_\mathsf{v}' : \mathsf{v} \in \mathbf{T}_2\}$ of iid $Exp(1/\mu)$-variables which is independent of all $X_\mathsf{v}$ and $L(\mathsf{v})$, $\mathsf{v} \in \mathbf{T}_2$. Prove $\Delta_n \stackrel{\mathbb{P}}{\to} 0$ by computing $\mathbb{V}\mathrm{ar}\,\Delta_n$ and use (a).]

(c) If $F = \delta_1$, then

$$W_n \;:=\; \sum_{\mathsf{v}\in\mathbf{T}:|v|=n} L(\mathsf{v})X_\mathsf{v} \;=\; \sum_{\mathsf{v}\in\mathbf{T}:|v|=n} L(\mathsf{v}), \quad n\geq 0$$

constitutes a martingales which converges a.s. to some $W \stackrel{d}{=} Exp(1/\mu)$.

**Problem 5.47.** Still in the situation of Prop. 5.42, let $U_1, U_2$ be *two* independent $Unif(0,1)$-variables independent of $X, X_1, X_2$. Prove the equivalence of

(a) $X \stackrel{d}{=} \Gamma(2, 1/\mu)$ (b) $X \stackrel{d}{=} U_1 X_1 + U_2 X_2$.

## References


1. Alsmeyer, G.: Renewal, recurrence and regeneration (in preparation). Springer-Verlag, New York (2010+)
2. Asmussen, S., Hering, H.: Branching processes, *Progress in Probability and Statistics*, vol. 3. Birkhäuser Boston Inc., Boston, MA (1983)
3. Athreya, K.B.: On the maximum sequence in a critical branching process. Ann. Probab. **16**(2), 502–507 (1988)
4. Athreya, K.B., Ney, P.E.: Branching processes. Springer-Verlag, New York (1972). Die Grundlehren der mathematischen Wissenschaften, Band 196
5. Bauer, H.: Measure and integration theory, *de Gruyter Studies in Mathematics*, vol. 26. Walter de Gruyter & Co., Berlin (2001). Translated from the German by Robert B. Burckel
6. Breiman, L.: Probability. Addison-Wesley Publishing Company, Reading, Mass. (1968)
7. Chow, Y.S., Teicher, H.: Probability theory. Independence, interchangeability, martingales, $3^{rd}$ edn. Springer Texts in Statistics. Springer-Verlag, New York (1997)
8. Dwass, M.: The total progeny in a branching process and a related random walk. J. Appl. Probability **6**, 682–686 (1969)
9. Feller, W.: An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons Inc., New York (1971)
10. Foster, J.H.: Branching processes involving immigration. ProQuest LLC, Ann Arbor, MI (1969). Thesis (Ph.D.) –The University of Wisconsin - Madison
11. Foster, J.H., Williamson, J.A.: Limit theorems for the Galton-Watson process with time-dependent immigration. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **20**, 227–235 (1971)
12. Hall, P., Heyde, C.C.: Martingale limit theory and its application. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1980). Probability and Mathematical Statistics
13. Harris, T.E.: The theory of branching processes. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin (1963)
14. Heathcote, C.R.: A branching process allowing immigration. J. Roy. Statist. Soc. Ser. B **27**, 138–143 (1965). Corrigendum and comments in B **28**, 213–217 (1966)
15. Heyde, C.C.: Extension of a result of Seneta for the super-critical Galton-Watson process. Ann. Math. Statist. **41**, 739–742 (1970)
16. Kesten, H., Ney, P., Spitzer, F.: The Galton-Watson process with mean one and finite variance. Theory Probab. Appl. **11**, 513–540 (1966)
17. Kesten, H., Stigum, B.P.: A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Statist. **37**, 1211–1223 (1966)
18. Kolmogorov, A.: Zur Lösung einer biologischen Aufgabe. Mitt. Forsch.-Inst. Math. Mech. Kujbyschew-Univ. Tomsk **2**, 1–6 (1938)
19. Kotz, S., Steutel, F.W.: Note on a characterization of exponential distributions. Statist. Probab. Lett. **6**(3), 201–203 (1988)
20. Levinson, N.: Limiting theorems for Galton-Watson branching process. Illinois J. Math. **3**, 554–565 (1959)
21. Lotka, A.J.: The extinction of families – I. J. Wash. Acad. Sci. **21**, 377–380 (1931)
22. Lotka, A.J.: The extinction of families – II. J. Wash. Acad. Sci. **21**, 453–459 (1931)
23. Lotka, A.J.: Théorie analytique des associations biologiques, deuxieme partie, vol. 780. Hermann, Paris (1939)

24. Lyons, R., Pemantle, R., Peres, Y.: Conceptual proofs of $L\log L$ criteria for mean behavior of branching processes. Ann. Probab. **23**(3), 1125–1138 (1995)
25. Pakes, A.G.: On the critical Galton-Watson process with immigration. J. Austral. Math. Soc. **12**, 476–482 (1971)
26. Pakes, A.G.: Some limit theorems for the total progeny of a branching process. Advances in Appl. Probability **3**, 176–192 (1971)
27. Pakes, A.G.: Further results on the critical Galton-Watson process with immigration. J. Austral. Math. Soc. **13**, 277–290 (1972)
28. Pakes, A.G.: Remarks on the maxima of a martingale sequence with application to the simple critical branching process. J. Appl. Probab. **24**(3), 768–772 (1987)
29. Pakes, A.G.: Length biasing and laws equivalent to the log-normal. J. Math. Anal. Appl. **197**(3), 825–854 (1996)
30. Pakes, A.G., Khattree, R.: Length-biasing, characterizations of laws and the moment problem. Austral. J. Statist. **34**(2), 307–322 (1992)
31. Seneta, E.: Functional equations and the Galton-Watson process. Advances in Appl. Probability **1**, 1–42 (1969)
32. Seneta, E.: An explicit-limit theorem for the critical Galton-Watson process with immigration. J. Roy. Statist. Soc. Ser. B **32**, 149–152 (1970)
33. Seneta, E.: On the supercritical Galton-Watson process with immigration. Math. Biosci. **7**, 9–14 (1970)
34. Stromberg, K.R.: Introduction to classical real analysis. Wadsworth International, Belmont, Calif. (1981). Wadsworth International Mathematics Series
35. Weiner, H.: Moments of the maximum in a critical branching process. J. Appl. Probab. **21**(4), 920–923 (1984)
36. Williams, D.: Probability with martingales. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge (1991)
37. Yaglom, A.M.: Certain limit theorems of the theory of branching random processes. Doklady Akad. Nauk SSSR (N.S.) **56**, 795–798 (1947)