Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER Stochastic Processes and their Applications 115 (2005) 1658–1676
stochastic processes and their applications
www.elsevier.com/locate/spa

Functional limit theorems for strongly subcritical branching processes in random environment ☆

V.I. Afanasyev[a], J. Geiger[b,*], G. Kersting[c], V.A. Vatutin[a]

[a]*Department of Discrete Mathematics, Steklov Institute, 8 Gubkin Street, 119 991 Moscow, Russia*
[b]*Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, D-67653 Kaiserslautern, Germany*
[c]*Fachbereich Mathematik, Universität Frankfurt, Fach 187,D-60054 Frankfurt am Main, Germany*

Received 14 September 2004; received in revised form 27 April 2005; accepted 27 April 2005
Available online 31 May 2005

Abstract

For a strongly subcritical branching process $(Z_n)_{n\geqslant 0}$ in random environment the non-extinction probability at generation n decays at the same exponential rate as the expected generation size and given non-extinction at n the conditional distribution of Z_n has a weak limit. Here we prove conditional functional limit theorems for the generation size process $(Z_k)_{0\leqslant k\leqslant n}$ as well as for the random environment. We show that given the population survives up to generation n the environmental sequence still evolves in an i.i.d. fashion and that the conditioned generation size process converges in distribution to a positive recurrent Markov chain.
© 2005 Elsevier B.V. All rights reserved.

MSC: primary 60J80; secondary 60G50; 60F17

Keywords: Branching process; Random environment; Random walk; Change of measure; Positive recurrent Markov chain; Functional limit theorem

☆This paper is part of a project supported by the German Research Foundation (DFG) and the Russian Foundation of Basic Research (Grant 436 Rus 113/683 and in part Grant 02-01-00266). Research of fourth author supported in part by RFBR-NWO Grant 047.016.013.

*Corresponding author. Tel.: +49 631 2053619; fax: +49 631 2052748.

E-mail addresses: viafan@mail.ru (V.I. Afanasyev), jgeiger@mathematik.uni-kl.de (J. Geiger), kersting@math.uni-frankfurt.de (G. Kersting), vatutin@mi.ras.ru (V.A. Vatutin).

0304-4149/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2005.05.001

1. Introduction and main results

For a branching process in random environment it is assumed that the offspring distribution of the individuals varies in a random fashion, independently from one generation to the other. Conditioned on the environment individuals reproduce independently of each other. Let Q_n be the random offspring distribution of an individual at generation $n-1$ and let Z_n denote the number of individuals at generation n. Then Z_n is the sum of Z_{n-1} independent random variables, each of which has distribution Q_n. To give a formal definition let Δ be the space of probability measures on $\mathbb{N}_0:=\{0,1,\dots\}$ which equipped with the metric of total variation is a Polish space. Let Q be a random variable taking values in Δ. Then, an infinite sequence $\Pi=(Q_1,Q_2,\dots)$ of i.i.d. copies of Q is said to form a *random environment*. A sequence of $\mathbb{N}_0$-valued random variables $Z_0,Z_1,\dots$ is called a *branching process in the random environment* Π, if Z_0 is independent of Π and given Π the process $Z=(Z_0,Z_1,\dots)$ is a Markov chain with

$$\mathscr{L}(Z_n \mid Z_{n-1}=z,\, \Pi=(q_1,q_2,\dots)) = q_n^{*z} \tag{1.1}$$

for every $n\in\mathbb{N}$, $z\in\mathbb{N}_0$ and $q_1,q_2,\dots\in\Delta$, where q^{*z} is the z-fold convolution of the measure q. The corresponding probability measure on the underlying probability space will be denoted by $\mathbf{P}$. Note that the transition probabilities P_{xy} of the Markov chain $(Z_n)_{n\geqslant 0}$ are

$$P_{xy} = \mathbf{E}[Q^{*x}(\{y\})], \quad x,y\in\mathbb{N}_0. \tag{1.2}$$

In the following we assume that the process starts with a single founding ancestor, $Z_0=1$ a.s., and that $\mathbf{P}\{Q=\delta_0\}=0$, where δ_x denotes unit point mass at x. Note, however, that in general Z is not the superposition of Z_0 independent copies of the process started at $Z_0=1$. The second assumption is no loss of generality since if $\mathbf{P}\{Q\neq\delta_0\}=:\alpha<1$, then $\mathscr{L}(Z_n)=\alpha^n\mathscr{L}(Z_n \mid Q_k\neq\delta_0, 1\leqslant k\leqslant n)+(1-\alpha^n)\delta_0$.

It turns out that the asymptotic behavior of the generation size process Z is determined in the main by the associated random walk $S=(S_n)_{n\geqslant 0}$. This random walk has initial state $S_0=0$ and increments $X_n=S_n-S_{n-1}$, $n\geqslant 1$ defined as

$$X_n := \log m(Q_n),$$

where

$$m(q) := \sum_{y=0}^{\infty} y\,q(\{y\})$$

is the mean of the offspring distribution $q\in\Delta$. In view of (1.1) and the assumption $Z_0=1$ a.s. the conditional expectation of Z_n given the environment Π can be expressed by means of S as

$$\mathbf{E}[Z_n \mid \Pi] = \prod_{k=1}^{n} m(Q_k) \;=\; \exp(S_n) \quad \mathbf{P}\text{-a.s.} \tag{1.3}$$

Averaging over the environment gives

$$\mathbf{E}[Z_n] = (\mathbf{E}[m(Q)])^n. \tag{1.4}$$

If the random walk S drifts to $-\infty$, then the branching process is said to be *subcritical*. It is customary to assume that $X^+ = \log^+ m(Q)$ has finite mean. Then subcriticality corresponds to $\mathbf{E}[\log m(Q)] < 0$. For such processes the conditional non-extinction probability at n decays at an exponential rate for almost every environment. This fact is an immediate consequence of the strong law of large numbers and the first moment estimate

$$\begin{aligned}\mathbf{P}\{Z_n > 0 \mid \Pi\} &= \min_{0 \leqslant k \leqslant n} \mathbf{P}\{Z_k > 0 \mid \Pi\} \\ &\leqslant \min_{0 \leqslant k \leqslant n} \mathbf{E}[Z_k \mid \Pi] = \exp\left(\min_{0 \leqslant k \leqslant n} S_k\right) \quad \mathbf{P}\text{-a.s.}\end{aligned} \tag{1.5}$$

As was first observed by Afanasyev [1] and later independently by Dekking [6] the asymptotic behavior of subcritical branching processes in random environment essentially depends on the sign of $\mathbf{E}[m(Q) \log m(Q)]$. Accordingly there are three different cases, namely the *weakly* subcritical, the *intermediate* subcritical and the *strongly* subcritical case (see, e.g., [10] for the detailed classification).

The present article is part of a series of publications (having started with paper [3] on the critical case) in which we try to develop the characteristic properties of the different cases. For a comparative discussion we refer the reader to [5].

Here we study the strongly subcritical case:

Assumption A1.

$$\mathbf{E}[m(Q) \log m(Q)] < 0.$$

By Jensen's inequality, A1 implies $\mathbf{E}[m(Q)] \log \mathbf{E}[m(Q)] < 0$, which means that

$$\mathbf{E}[m(Q)] < 1. \tag{1.6}$$

Again using Jensen's inequality we see that (1.6) entails

$$\mathbf{E}[\log m(Q)] < 0. \tag{1.7}$$

Our second assumption is an integrability condition on Q.

Assumption A2.

$$\mathbf{E}[Z_1 \log^+ Z_1] < \infty.$$

Here are some instances, where Assumptions A1 and A2 are satisfied.

Examples.

1. The classical Galton–Watson branching process is a special case of branching processes in random environment with $\mathbf{P}\{Q = q\} = 1$ for some $q \in \Delta$. If the Galton–Watson process is subcritical, i.e., if $m(q) < 1$, then Assumption A1 holds. For subcritical Galton–Watson processes A2 is well-known to be a necessary and

sufficient condition for the mean generation size $\mathbf{E}Z_n = m(q)^n$ to give the right decay rate of the survival probability $\mathbf{P}\{Z_n > 0\}$ (see, e.g., [4, Corollary 2 in Section 1.11]).

2. A2 holds if the random offspring distribution Q has bounded support. In particular, the results to follow hold for any strongly subcritical binary branching process (where individuals have either two children or none).
3. Let $\eta(Q)$ be the standardized second factorial moment of Q,
$$\eta(Q) := (m(Q))^{-2} \sum_{y=0}^{\infty} y(y-1)\, Q(\{y\}).$$
We claim that Assumption A1 and the integrability condition
$$\mathbf{E}[m(Q) \log^+ \eta(Q)] < \infty \tag{1.8}$$
imply A2. Indeed, observe that Jensen's inequality implies
$$\sum_{y=1}^{\infty} \log y \, \frac{yQ(\{y\})}{m(Q)} \leqslant \log(1 + m(Q)\eta(Q)) \leqslant 1 + \log^+ m(Q) + \log^+ \eta(Q) \quad \mathbf{P}\text{-a.s.}$$
Multiplying either side by $m(Q)$ and taking expectations gives
$$\mathbf{E}[Z_1 \log^+ Z_1] \leqslant \mathbf{E}[m(Q)] + \mathbf{E}[m(Q) \log^+ m(Q)] + \mathbf{E}[m(Q) \log^+ \eta(Q)] < \infty.$$
4. If Q is a Poisson distribution with random mean, then $\eta(Q) = 1$ a.s., while if Q is a random geometric distribution on $\mathbb{N}_0$, then $\eta(Q) = 2$ a.s. Hence, in these cases condition (1.8) is redundant (recall (1.6)) and we merely require the random walk S to satisfy A1.

In many aspects the longtime behavior of strongly subcritical branching processes in random environment resembles the asymptotic behavior of classical subcritical Galton–Watson branching processes. One such similarity is the fact that the first moment estimate $\mathbf{P}\{Z_n > 0\} \leqslant \mathbf{E}[Z_n]$ gives the right decay of the non-extinction probability at generation n up to a constant. (We note that all limit theorems in this paper are under the law $\mathbf{P}$ which is what is called the annealed approach. Notation $a_n \sim b_n$ is used to indicate that the two sequences are asymptotically equivalent, i.e., $a_n/b_n \to 1$ as $n \to \infty$.)

Theorem 1.1. *Assume* A1 *and* A2. *Then, as* $n \to \infty$,
$$\mathbf{P}\{Z_n > 0\} \sim \theta \mathbf{E}[Z_n] \tag{1.9}$$
for some $0 < \theta \leqslant 1$.

This result is due to Guivarc'h and Liu (Theorem 1.2 (a) in [11]). It was originally proved by D'Souza and Hambly [7] under an extra moment assumption.

In a subcritical Galton–Watson branching process the nth generation size has a conditional weak limit law given non-extinction at n. The next result, which is Theorem 1.1 in [10], shows that the same holds true for a strongly subcritical branching process in random environment.

Theorem 1.2. *Assume* A1 *and* A2. *Then there is a probability measure r with weights* r_z, $z \in \mathbb{N}$ *so that*

$$\lim_{n\to\infty} \mathbf{P}\{Z_n = z \mid Z_n > 0\} = r_z, \quad z \in \mathbb{N}. \tag{1.10}$$

Note that Fatou's lemma implies $m(r) \leqslant \theta^{-1}$. Hence, by Theorem 1.1, we have

$$m(r) < \infty. \tag{1.11}$$

Below we will see that the two quantities $m(r)$ and θ^{-1} agree (Corollary 2.3). We remark that the proof of Theorem 1.2 (and the results to follow) depends only on the asymptotics (1.9) but does not use the integrability condition A2 explicitly.

We now come to the main results of this paper. Clearly, conditioning the population on non-extinction at n also has an effect on the random environment Π. Note that there are various ways how the unlikely event, that the population survives until some late generation n, might occur. E.g., the population might be lucky to find an extraordinary productive environment in which chances for survival are high. However, it might also be that the population evolves in a typical environment, still by good luck it manages to avoid extinction. Theorem 1.3 below shows that given non-extinction at n the environment still evolves in an i.i.d. fashion. The new (random) offspring law is more productive but still subcritical. Hence, the situation is something in between the two scenarios described above.

To give the precise statement we introduce a measure $\widehat{\mathbf{P}}$ on the σ-field generated by $Z_1, Z_2, \ldots; Q_1, Q_2, \ldots$ which essentially describes the asymptotic behavior of (Z, Π) conditioned on $Z_n > 0$. The measure is obtained from $\mathbf{P}$ by the following transformation: For every non-negative measurable ψ on $\mathbb{N}_0^k \times \Delta^k$, $k \geqslant 1$ let

$$\widehat{\mathbf{E}}[\psi(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)] := \frac{\mathbf{E}[Z_k \psi(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)]}{(\mathbf{E}[m(Q)])^k}. \tag{1.12}$$

To see that (under suitable regularity conditions on the underlying probability space) relation (1.12) defines a probability measure on $\sigma(Z_1, Z_2, \ldots; Q_1, Q_2, \ldots)$ observe that the following consistency condition holds: If functions ψ_k and ψ_{k+1} satisfy

$$\psi_{k+1}(z_1, \ldots, z_{k+1}; q_1, \ldots, q_{k+1}) = \psi_k(z_1, \ldots, z_k; q_1, \ldots, q_k)$$

for all $z_i \in \mathbb{N}_0$ and $q_i \in \Delta$, $1 \leqslant i \leqslant k+1$, then

$$\begin{aligned}
&\widehat{\mathbf{E}}[\psi_{k+1}(Z_1, \ldots, Z_{k+1}; Q_1, \ldots, Q_{k+1})] \\
&\quad = \frac{\mathbf{E}[Z_{k+1} \psi_k(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)]}{(\mathbf{E}[m(Q)])^{k+1}} \\
&\quad = \frac{\mathbf{E}[\psi_k(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)\mathbf{E}[Z_{k+1} \mid Z_1, \ldots, Z_k; \Pi]]}{(\mathbf{E}[m(Q)])^{k+1}} \\
&\quad = \frac{\mathbf{E}[\psi_k(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)\, m(Q_{k+1})\, Z_k]}{(\mathbf{E}[m(Q)])^{k+1}} \\
&\quad = \widehat{\mathbf{E}}[\psi_k(Z_1, \ldots, Z_k; Q_1, \ldots, Q_k)],
\end{aligned}$$

where for the last two equalities we have used relation (1.1) and the independence of Q_{k+1} and $(Z_1, \dots, Z_k; Q_1, \dots, Q_k)$.

For the distribution of the environmental sequence under the new measure $\widehat{\mathbf{P}}$ note that (1.3) gives

$$\widehat{\mathbf{E}}[\psi(Q_1, \dots, Q_k)] = \frac{\mathbf{E}[\psi(Q_1, \dots, Q_k)\mathbf{E}[Z_k \mid \Pi]]}{(\mathbf{E}[m(Q)])^k} = \frac{\mathbf{E}[\exp(S_k)\psi(Q_1, \dots, Q_k)]}{(\mathbf{E}[m(Q)])^k} \tag{1.13}$$

for every $k \in \mathbb{N}$ and non-negative measurable ψ on Δ^k. Relation (1.13) and the fact that the density $\exp(S_k) = \prod_{j=1}^k m(Q_j)$ has product structure show that under $\widehat{\mathbf{P}}$ the random measures $Q_j, j \geqslant 1$ are still i.i.d. Their common law is the *size-biased* distribution given by

$$\widehat{\mathbf{E}}[\psi(Q)] = \frac{\mathbf{E}[m(Q)\psi(Q)]}{\mathbf{E}[m(Q)]}. \tag{1.14}$$

Note that the measure $\widehat{\mathbf{P}}$ favors offspring distributions with large mean: A reproduction law $q \in \Delta$ is $m(q)/\mathbf{E}[m(Q)]$ times as likely as under the measure $\mathbf{P}$. Using twice Jensen's inequality we see that

$$\mathbf{E}[m(Q)]\mathbf{E}[\log m(Q)] \leqslant \mathbf{E}[m(Q)] \log \mathbf{E}[m(Q)] \leqslant \mathbf{E}[m(Q) \log m(Q)].$$

Hence, by Assumption A1, we have

$$\mathbf{E}[X] = \mathbf{E}[\log m(Q)] \leqslant \widehat{\mathbf{E}}[\log m(Q)] = \widehat{\mathbf{E}}[X] < 0, \tag{1.15}$$

i.e., under $\widehat{\mathbf{P}}$ the drift of the random walk is increased but remains negative.

Theorem 1.3. *Assume* A1 *and* A2. *Let* $i_{n,j}$, $n \in \mathbb{N}$, $1 \leqslant j \leqslant k$ *be non-negative integers with* $1 \leqslant i_{n,1} < i_{n,2} < \cdots < i_{n,k} \leqslant n$ *and* $n - i_{n,k} \to \infty$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} \mathbf{P}\{Q_{i_{n,1}} \in B_1, \dots, Q_{i_{n,k}} \in B_k \mid Z_n > 0\} = \prod_{j=1}^{k} \widehat{\mathbf{P}}\{Q \in B_j\} \tag{1.16}$$

for every $k \in \mathbb{N}$ *and Borel sets* $B_1, \dots, B_k \subset \Delta$.

Moreover, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left\{ \sup_{0 \leqslant t \leqslant 1} \left| \frac{1}{n} S_{\lfloor nt \rfloor} - t\,\widehat{\mathbf{E}}[X] \right| \geqslant \varepsilon \mid Z_n > 0 \right\} = 0.$$

Remarks. Let us explain in an informal manner the intuition behind Theorem 1.3 and the significance of Assumption A1. The change in the distribution of the environment Π when conditioning on the event $\{Z_n > 0\}$ is captured by the formula

$$\mathbf{P}\{\Pi \in d\pi \mid Z_n > 0\} = c_n \mathbf{P}\{Z_n > 0 \mid \Pi = \pi\} \mathbf{P}\{\Pi \in d\pi\} \tag{1.17}$$

for every $\pi = (q_1, q_2, \ldots) \in \Delta^{\mathbb{N}}$ with normalizing constant $c_n = (\mathbf{P}\{Z_n > 0\})^{-1}$. In view of (1.5) it is plausible that one can rewrite (1.17) as

$$\mathbf{P}\{\Pi \in d\pi \mid Z_n > 0\} \approx c'_n \exp\left(\min_{0 \leqslant k \leqslant n} s_k\right) \mathbf{P}\{\Pi \in d\pi\}, \tag{1.18}$$

where $s_k := \sum_{j=1}^{k} \log m(q_j)$ and $c'_n = (\mathbf{E}[\exp(\min_{k \leqslant n} S_k)])^{-1}$. We claim that under A1 approximation (1.18) simplifies to

$$\mathbf{P}\{\Pi \in d\pi \mid Z_n > 0\} \approx c''_n \exp(s_n) \mathbf{P}\{\Pi \in d\pi\}, \tag{1.19}$$

where $c''_n = (\mathbf{E}[m(Q)])^{-n}$. Indeed, note that in order to pass from (1.18) to (1.19) one needs L_1-convergence of the ratio of the two densities w.r.t. to the new measure $\mathscr{L}(\Pi \mid Z_n > 0)$. However, w.r.t. to the measure defined on the right-hand side of (1.19) the process $(S_k)_{0 \leqslant k \leqslant n}$ performs a random walk with drift $\widehat{\mathbf{E}}[X]$, which is negative under A1 (see the discussion following (1.13)). In this situation the difference $S_n - \min_{0 \leqslant k \leqslant n} S_k$ is asymptotically independent of any initial piece of the random walk and has a weak limit as $n \to \infty$.

The behavior of the conditioned environment in the strongly subcritical case is in sharp contrast to the other subcritical cases and the critical case, where conditioning the population on survival at some late generation leads to dependence among the states of the environmental sequence (see [3,5]).

Our next theorem describes the dynamics of the generation size process $(Z_k)_{0 \leqslant k \leqslant n}$ given non-extinction at n. Note that conditioning on the event $\{Z_n > 0\}$ not only has an effect on the environment, but also affects the individuals' reproduction within the new environment.

To prepare for the result we first identify the distribution of the generation size process Z under the measure $\widehat{\mathbf{P}}$ as the law of a certain Markov chain. Note that (1.12) implies (set $z_0 := 1$)

$$\begin{aligned}\widehat{\mathbf{P}}\{Z_1 = z_1, \ldots, Z_k = z_k\} &= \frac{z_k \mathbf{P}\{Z_1 = z_1, \ldots, Z_k = z_k\}}{(\mathbf{E}[m(Q)])^k} \\ &= \frac{z_1 \cdots z_k}{z_0 \cdots z_{k-1}} \frac{P_{z_0 z_1} \cdots P_{z_{k-1} z_k}}{(\mathbf{E}[m(Q)])^k} = \prod_{j=1}^{k} \widehat{P}_{z_{j-1} z_j}\end{aligned} \tag{1.20}$$

for every $z_j \in \mathbb{N}$, $1 \leqslant j \leqslant k$, where

$$\widehat{P}_{xy} := \frac{y P_{xy}}{x \mathbf{E}[m(Q)]}, \quad x, y \in \mathbb{N}. \tag{1.21}$$

By linearity of expectation, the $\widehat{P}_{xy}$ sum to 1 for every $x \in \mathbb{N}$ (recall (1.2)). Relation (1.20) shows that

$$\widehat{\mathbf{P}}\{Z_1 = z_1, \ldots, Z_k = z_k \mid Z_k = z_k\} = \mathbf{P}\{Z_1 = z_1, \ldots, Z_k = z_k \mid Z_k = z_k\} \tag{1.22}$$

for every $z_1, \ldots, z_k \in \mathbb{N}$. Hence, the two Markov chains with transition matrices P and $\widehat{P}$, respectively, have the same distribution if initial and final states are fixed. The unconditional distributions, however, are notably different since for every x the measure $(\widehat{P}_{xy})$ is the size-biasing of (P_{xy}). In particular, the support of $(\widehat{P}_{xy})$ is $\mathbb{N}$ rather than $\mathbb{N}_0$.

For later reference we state the following formula for the k-step transition probabilities of the $\widehat{P}$-chain:

$$\widehat{\mathbf{P}}\{Z_{j+k} = y \mid Z_j = x\} = \widehat{P}^k_{xy} = \frac{yP^k_{xy}}{x(\mathbf{E}[m(Q)])^k} \tag{1.23}$$

for every $x, y \in \mathbb{N}$ and $j, k \geqslant 0$. Here $\widehat{P}^k, P^k$ denote the kth power of the transition matrices $\widehat{P}$ and P.

Remark. For classical Galton–Watson processes the Markov chain with transition matrix $\widehat{P}$ is called the Q-process of the branching process (see [4, Section 1.14]).

Having introduced the limiting object we can now state the functional limit theorem for the conditioned generation size process. (We use $\mathscr{L}(X)$ and $\mathscr{L}(X \mid A)$ for the distribution or conditional distribution of the random variable X given the event A and write $d_{\mathrm{TV}}[\mu, \nu]$ for the total variation distance between probability measures μ and ν.)

Theorem 1.4. *Assume* A1 *and* A2. *Let k_n, $n \geqslant 1$ be a sequence of non-negative integers with $k_n \leqslant n$ and $n - k_n \to \infty$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} d_{\mathrm{TV}}[\mathscr{L}(Z_1, \ldots, Z_{k_n} \mid Z_n > 0), \mathscr{L}(\widehat{Z}_1, \ldots, \widehat{Z}_{k_n})] = 0 \tag{1.24}$$

and

$$\lim_{n\to\infty} d_{\mathrm{TV}}[\mathscr{L}(Z_n, \ldots, Z_{n-k_n} \mid Z_n > 0), \mathscr{L}(\widetilde{Z}_0, \ldots, \widetilde{Z}_{k_n})] = 0. \tag{1.25}$$

Here, the process $(\widehat{Z}_j)_{j\geqslant 0}$ is a Markov chain with transition matrix $\widehat{P}$ started at $\widehat{Z}_0 = 1$. The chain converges towards its stationary distribution $\hat{r}$ with weights

$$\hat{r}_x = \frac{x r_x}{m(r)}, \quad x \in \mathbb{N}. \tag{1.26}$$

The process $(\widetilde{Z}_j)_{j\geqslant 0}$ is a Markov chain with the time-reversed transition matrix $\widetilde{P}$,

$$\widetilde{P}_{xy} = \frac{\hat{r}_y}{\hat{r}_x} \widehat{P}_{yx}, \quad x, y \in \operatorname{supp} \hat{r} \tag{1.27}$$

and initial distribution r. This chain converges in distribution to $\hat{r}$, too.

Remarks.

1. The result displays a unique feature of strongly subcritical branching processes in random environment (among the subcritical cases): The population conditioned on non-extinction at n stays small throughout the time interval from 0 to n. Moreover, every once in a while there are certain regeneration epochs. Those are times when all individuals stem from the same individual of the previous generation. Note that this does not necessarily mean that the population has declined to a single individual (e.g., if branching is binary then $\mathbf{P}\{Z_n \text{ is even}\} = 1$ for all $n \geqslant 1$).

2. The processes $(\widehat{Z}_j)_{j\geqslant 0}$ and $(\widetilde{Z}_j)_{j\geqslant 0}$ could be defined on some different probability space. Therefore we denote the corresponding probabilities and expectations by $\mathbb{P}$ and $\mathbb{E}$. Note that for every $j\geqslant 0$ the random variable $\widehat{Z}_j$ has the size-biased distribution

$$\mathbb{P}\{\widehat{Z}_j = z\} = \frac{z\mathbf{P}\{Z_j = z\}}{\mathbf{E}[Z_j]}, \quad z \in \mathbb{N}.$$

This change of measure is quite intuitive: Since the conditioned environment is still subcritical the event that two or more individuals at generation j have a descendant at n is asymptotically negligible, so that (compare Lemma 2.4 below)

$$\mathbf{P}\{Z_n > 0 \mid Z_j = z\} \sim z\mathbf{P}\{Z_{n-j} > 0\} \quad \text{as } n \to \infty.$$

Consequently,

$$\mathbf{P}\{Z_j = z \mid Z_n > 0\} \sim z\mathbf{P}\{Z_j = z\} \frac{\mathbf{P}\{Z_{n-j} > 0\}}{\mathbf{P}\{Z_n > 0\}} \sim \mathbb{P}\{\widehat{Z}_j = z\}.$$

3. As a consequence of Theorem 1.4, conditioned on $Z_n > 0$ the random variables $Z_{\lfloor nt_1 \rfloor}, \ldots, Z_{\lfloor nt_k \rfloor}$ are asymptotically i.i.d. for distinct $t_j \in (0,1)$, $1 \leqslant j \leqslant k$. This fact had been established in [2] for the case where Q a.s. has a linear fractional generating function (which means that the offspring law $Q(\cdot \cap \mathbb{N})/Q(\mathbb{N})$ is geometric with random mean). In this special case also results on the reduced tree spanned by the individuals of generation n and the root have been obtained (see [9]).

2. Proofs

To prepare for the proofs of Theorems 1.3 and 1.4 we first establish the asserted properties of the Markov chain with transition matrix $\widehat{P}$. The fact that the chain converges towards its equilibrium distribution will be an immediate consequence.

Proposition 2.1. *Assume* A1 *and* A2.

(i) *The probability measure* $\hat{r}$ *from* (1.26) *is an invariant distribution for* $\widehat{P}$,

$$\sum_{y \in \mathbb{N}} \hat{r}_y \widehat{P}_{yx} = \hat{r}_x, \quad x \in \mathbb{N}. \tag{2.1}$$

(ii) *The chain has a single recurrent class* $\widehat{R} = \operatorname{supp} \hat{r}$. *The class* $\widehat{R}$ *is positive recurrent and aperiodic.*

(iii) *For whatever initial state the chain eventually hits* $\widehat{R}$,

$$\lim_{k\to\infty} \sum_{y \in \hat{R}} \widehat{P}^k_{xy} = 1, \quad x \in \mathbb{N}.$$

Proof. By (1.11), the measure $\hat{r}$ has total mass 1. To prove the invariance of the measure $\hat{r}$ for $\widehat{P}$ note that in view of (1.21) and (1.26), condition (2.1) is equivalent to

$$\sum_{y \in \mathbb{N}} r_y P_{yx} = \mathbf{E}[m(Q)]\, r_x, \quad x \in \mathbb{N}. \tag{2.2}$$

From Theorems 1.1 and 1.2 we obtain

$$\begin{aligned} r_y P_{yx} &= \lim_{n\to\infty} \frac{\mathbf{P}\{Z_n = y\}\mathbf{P}\{Z_{n+1} = x \mid Z_n = y\}}{\mathbf{P}\{Z_n > 0\}} \\ &= \mathbf{E}[m(Q)] \lim_{n\to\infty} \mathbf{P}\{Z_n = y, Z_{n+1} = x \mid Z_{n+1} > 0\} \end{aligned} \tag{2.3}$$

for every $x, y \in \mathbb{N}$. Again using Theorems 1.1 and 1.2 we see that (2.3) implies

$$\begin{aligned} \mathbf{E}[m(Q)]\, r_x &= \lim_{n\to\infty} \left(\sum_{y=1}^{z} \mathbf{E}[m(Q)]\, \mathbf{P}\{Z_n = y, Z_{n+1} = x \mid Z_{n+1} > 0\} \right. \\ &\qquad \left. + \mathbf{P}\{Z_{n+1} > 0 \mid Z_n > 0\}\, \mathbf{P}\{Z_n > z, Z_{n+1} = x \mid Z_{n+1} > 0\} \right) \\ &= \sum_{y=1}^{z} r_y P_{yx} + \lim_{n\to\infty} \mathbf{P}\{Z_n > z, Z_{n+1} = x \mid Z_n > 0\} \end{aligned} \tag{2.4}$$

for every $x, z \in \mathbb{N}$. Hence,

$$0 \leqslant \mathbf{E}[m(Q)]\, r_x - \sum_{y=1}^{z} r_y P_{yx} \leqslant \limsup_{n\to\infty} \mathbf{P}\{Z_n > z \mid Z_n > 0\} = \sum_{y=z+1}^{\infty} r_y. \tag{2.5}$$

Letting $z \to \infty$ in (2.5) gives (2.2).

To prove part (ii) we first show that there are states which can be reached from any other state of the chain in a single step. By the assumed subcriticality, there exists $z \in \mathbb{N}$ with

$$\mathbf{P}\{Q(\{0\}) > 0, Q(\{z\}) > 0\} > 0, \tag{2.6}$$

i.e., in the original branching process individuals of the same generation may have both 0 or z children with positive probability. For such z (recall (1.2) and (1.21))

$$\widehat{P}_{xz} = \frac{z\mathbf{E}[Q^{*x}(\{z\})]}{x\mathbf{E}[m(Q)]} \geqslant \frac{z\mathbf{E}[Q(\{z\})(Q(\{0\}))^{x-1}]}{x\mathbf{E}[m(Q)]} > 0 \tag{2.7}$$

for every $x \in \mathbb{N}$. The second assertion of the proposition now follows from standard results from Markov chain theory: Since any invariant probability distribution is supported by positive recurrent states (see, e.g., the criterion in Section XV.7 of [8]), part (i) of the proposition shows that the chain has at least one such class. In view of (2.7) there can be at most one recurrent class. Clearly, this class $\widehat{R}$, say, contains all z which satisfy (2.7). Since $\widehat{P}_{zz} > 0$ for such z, the class is aperiodic. The fact that $\widehat{R} = \operatorname{supp} \hat{r}$ again follows from part (i), because the equilibrium weight $\hat{r}_x$ is the reciprocal of the expected return time to x (see, e.g., [8, Theorem 1 in Section XV.7]).

We finally prove (iii). In view of (2.7) it will be sufficient to show that the chain cannot escape to ∞ with positive probability,

$$\widehat{\mathbf{P}}\{Z_n \to \infty \mid Z_0 = x\} = 0, \quad x \in \mathbb{N}. \tag{2.8}$$

This in turn will follow from the stochastic monotonicity of the chain which we will establish first. We claim that

$$\widehat{P}_{xy} = \widehat{\mathbf{E}}[\widehat{Q} * Q^{*(x-1)}(\{y\})], \quad x, y \in \mathbb{N}, \tag{2.9}$$

where the random measure $\widehat{Q}$ is obtained from Q by size-biasing,

$$\widehat{Q}(\{y\}) = \frac{yQ(\{y\})}{m(Q)}, \quad y \in \mathbb{N}.$$

Indeed, in view of (1.2), (1.14) and (1.21) we have

$$\begin{aligned}
\widehat{P}_{xy} &= \frac{y}{x}\,\widehat{\mathbf{E}}\left[\frac{Q^{*x}(\{y\})}{m(Q)}\right] \\
&= \frac{1}{x}\,\widehat{\mathbf{E}}\left[(m(Q))^{-1} \sum_{y_1+\cdots+y_x=y} (y_1 + \cdots + y_x) Q(\{y_1\}) \cdots Q(\{y_x\})\right] \\
&= \widehat{\mathbf{E}}\left[(m(Q))^{-1} \sum_{y_1+\cdots+y_x=y} y_1 Q(\{y_1\}) \cdots Q(\{y_x\})\right] \\
&= \widehat{\mathbf{E}}\left[\sum_{y_1+\cdots+y_x=y} \widehat{Q}(\{y_1\})\, Q(\{y_2\}) \cdots Q(\{y_x\})\right] \\
&= \widehat{\mathbf{E}}[\widehat{Q} * Q^{*(x-1)}(\{y\})].
\end{aligned}$$

Identity (2.9) shows that $\widehat{P}$ is monotone, i.e. $(\widehat{P}_{xy})_{y\in\mathbb{N}}$ is stochastically increasing with x:

$$\begin{aligned}
\sum_{y=1}^{z} \widehat{P}_{xy} &= \widehat{\mathbf{E}}[\widehat{Q} * Q^{*(x-1)}(\{1, \ldots, z\})] \\
&\geqslant \widehat{\mathbf{E}}[\widehat{Q} * Q^{*(x'-1)}(\{1, \ldots, z\})] = \sum_{y=1}^{z} \widehat{P}_{x'y}
\end{aligned}$$

for every $z \in \mathbb{N}$ and $x \leqslant x'$. A standard coupling argument shows that we can construct versions of the chain started at $x \leqslant x'$, respectively, so that with probability 1 the process started at x is always below the one started at x'.

Now suppose that $\widehat{R} = \operatorname{supp} \hat{r}$ is unbounded so that for every $x \in \mathbb{N}$, there exists $x' \in \widehat{R}$ with $x' \geqslant x$. Using monotonicity of $\widehat{P}$ and the fact that $\widehat{R}$ is a recurrent class we obtain

$$\widehat{\mathbf{P}}\{Z_n \to \infty \mid Z_0 = x\} \leqslant \widehat{\mathbf{P}}\{Z_n \to \infty \mid Z_0 = x'\} = 0.$$

It is easy to see that $\widehat{R} = \operatorname{supp} \hat{r}$ can only be bounded if $P_{1y} = 0$ for all $y \geqslant 2$ (and then $\widehat{R} = \{1\}$). In this case the chains with transition matrices P and $\widehat{P}$ have decreasing paths and (2.8) is trivially true.

To deduce (iii) from (2.8) note that with probability 1 each transient state is visited only finitely often. Therefore, $\{Z_n \notin \widehat{R} \text{ for all } n\} = \{Z_n \to \infty\}$ a.s. and thus

$$\lim_{k\to\infty} \sum_{y\in\hat{R}} \widehat{P}^k_{xy} = \widehat{\mathbf{P}}\{Z_n \in \widehat{R} \text{ for some } n \,|\, Z_0 = x\} = 1. \qquad \square$$

Remark. The chain with transition matrix $\widehat{P}$ can have transient states. In fact, it might well be that the event that the time of the first exit from the set of transient states is later than n has positive probability for all n. E.g., if

$$\mathbf{P}\{Q = \delta_1\} = \alpha \quad \text{and} \quad \mathbf{P}\{Q = p\delta_0 + (1-p)\delta_2\} = 1-\alpha$$

for some $0<\alpha<1$ and $\frac{1}{2}<p<1$, then

$$\mathbf{P}\{Z_k = 1 \text{ for all } 1 \leqslant k \leqslant n \,|\, Z_n > 0\} > 0$$

even though $\operatorname{supp} \hat{r} = 2\mathbb{N}$. The chain with the time-reversed transition matrix $\widetilde{P}$, however, is always an irreducible recurrent Markov chain with state space $\operatorname{supp} \hat{r}$.

Corollary 2.2. *Assume* A1 *and* A2. *For whatever initial state the chain converges towards its equilibrium distribution,*

$$\lim_{k\to\infty} \widehat{P}^k_{xy} = \hat{r}_y, \quad x, y \in \mathbb{N}. \tag{2.10}$$

Proof. When restricted to $\widehat{R} = \operatorname{supp} \hat{r}$ the chain is positive recurrent, aperiodic and irreducible. Hence, for $x \in \widehat{R}$ the claim follows from the standard convergence theorem for Markov chains (see, e.g., [8, Theorem 1 in Section XV.7]). To extend the result to general x use part (iii) of Proposition 2.1. $\square$

An immediate consequence of the weak convergence result (2.10) is uniform integrability of the Z_n conditioned on non-extinction at n.

Corollary 2.3. *Assume* A1 *and* A2. *Then*

$$\lim_{z\to\infty} \limsup_{n\to\infty} \mathbf{E}[Z_n I\{Z_n > z\} \,|\, Z_n > 0] = 0 \tag{2.11}$$

and

$$m(r) = \theta^{-1}. \tag{2.12}$$

Proof. Using first (1.12) and then Theorem 1.1 and Corollary 2.2, we get

$$\begin{aligned}\mathbf{E}[Z_n I\{Z_n > z\} \,|\, Z_n > 0] &= \frac{(\mathbf{E}[m(Q)])^n}{\mathbf{P}\{Z_n > 0\}} \widehat{\mathbf{P}}\{Z_n > z\} \\ &\to \theta^{-1} \sum_{y=z+1}^{\infty} \hat{r}_y \quad \text{as } n \to \infty.\end{aligned}$$

Letting $z \to \infty$ gives (2.11). For (2.12) recall that weak convergence and uniform integrability imply convergence of the means. Hence, Theorem 1.2 and (2.11) give

$$\theta^{-1} = \lim_{n\to\infty} \frac{\mathbf{E}[Z_n]}{\mathbf{P}\{Z_n > 0\}} = \lim_{n\to\infty} \mathbf{E}[Z_n \,|\, Z_n > 0] = m(r). \qquad \square$$

We will establish one more preliminary result before we prove Theorems 1.3 and 1.4. Let $e_{k,n}$ be the conditional extinction probability at n given Π when $Z_k = 1$,

$$e_{k,n} = e_{k,n}(\Pi) := \mathbf{P}\{Z_n = 0 \mid Z_k = 1, \Pi\}, \quad 0 \leqslant k \leqslant n. \tag{2.13}$$

In view of (1.1) we have

$$\mathbf{P}\{Z_n > 0 \mid Z_k, \Pi\} = 1 - e_{k,n}^{Z_k} \quad \mathbf{P}\text{-a.s.} \tag{2.14}$$

The following lemma states that for a binomially distributed random variable Y_n with random parameters Z_{k_n} and $1 - e_{k_n,n}$ the quantities $\mathbf{E}Y_n$ and $\mathbf{P}\{Y_n \geqslant 1\}$ are asymptotically equivalent. Note that Y_n is the number of individuals at generation k_n which have a descendant at n.

Lemma 2.4. *Assume* A1 *and* A2. *Let k_n, $n \geqslant 1$ be a sequence of non-negative integers with $k_n \leqslant n$ and $n - k_n \to \infty$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} \frac{\mathbf{E}\,|\,Z_{k_n}(1 - e_{k_n,n}) - (1 - e_{k_n,n}^{Z_{k_n}})\,|}{(\mathbf{E}[m(Q)])^n} = 0. \tag{2.15}$$

Proof. The inequality $1 - x^j \leqslant j(1 - x)$ for $0 \leqslant x \leqslant 1$, $j \in \mathbb{N}_0$ (with the usual convention $0^0 = 1$) implies

$$0 \leqslant 1 - e_{k_n,n}^{Z_{k_n}} \leqslant Z_{k_n}(1 - e_{k_n,n}) \quad \mathbf{P}\text{-a.s.} \tag{2.16}$$

Also, note that, by independence and stationarity of the Q_j under $\mathbf{P}$ and relations (1.4) and (1.9), we have

$$\lim_{n\to\infty} \frac{\mathbf{E}[Z_{k_n}(1 - e_{k_n,n})]}{(\mathbf{E}[m(Q)])^n} = \lim_{n\to\infty} \frac{\mathbf{E}[Z_{k_n}]\,\mathbf{P}\{Z_{n-k_n} > 0\}}{(\mathbf{E}[m(Q)])^{k_n + n - k_n}} = \theta. \tag{2.17}$$

Hence, to prove (2.15) it suffices to show

$$\liminf_{n\to\infty} \frac{\mathbf{E}[1 - e_{k_n,n}^{Z_{k_n}}]}{(\mathbf{E}[m(Q)])^n} \geqslant \theta. \tag{2.18}$$

To establish (2.18) use $1 - x^j \geqslant j(1 - x)x^j$ for $0 \leqslant x \leqslant 1$, $j \in \mathbb{N}_0$ and independence of the Q_j to deduce

$$\begin{aligned} \mathbf{E}[1 - e_{k_n,n}^{Z_{k_n}}] &\geqslant \mathbf{E}[Z_{k_n}(1 - e_{k_n,n})\, e_{k_n,n}^{Z_{k_n}}] \\ &\geqslant (1 - \varepsilon)^z\, \mathbf{E}[Z_{k_n}(1 - e_{k_n,n});\, Z_{k_n} \leqslant z,\, e_{k_n,n} \geqslant 1 - \varepsilon] \\ &= (1 - \varepsilon)^z\, \mathbf{E}[Z_{k_n};\, Z_{k_n} \leqslant z]\mathbf{E}[1 - e_{k_n,n};\, 1 - e_{k_n,n} \leqslant \varepsilon] \end{aligned} \tag{2.19}$$

for every $\varepsilon > 0$ and $z \in \mathbb{N}_0$. For the first expectation on the right-hand side of (2.19) note that, by (1.12),

$$\frac{\mathbf{E}[Z_{k_n};\, Z_{k_n} \leqslant z]}{(\mathbf{E}[m(Q)])^{k_n}} = 1 - \widehat{\mathbf{P}}\{Z_{k_n} > z\}. \tag{2.20}$$

For the other expectation observe that the first moment inequality (1.5) and relations (1.13) and (2.13) give

$$\frac{\mathbf{E}[1-e_{k_n,n};\,1-e_{k_n,n}>\varepsilon]}{(\mathbf{E}[m(Q)])^{n-k_n}} \leqslant \frac{\mathbf{E}[\exp(S_n-S_{k_n});\,\exp(S_n-S_{k_n})>\varepsilon]}{(\mathbf{E}[m(Q)])^{n-k_n}}$$
$$= \widehat{\mathbf{P}}\{S_{n-k_n}>\log\varepsilon\}. \tag{2.21}$$

The probability on the right-hand side of (2.21) tends to 0 as $n\to\infty$ by the law of large numbers (recall (1.15)). Hence, an application of Theorem 1.1 yields

$$\lim_{n\to\infty}\frac{\mathbf{E}[1-e_{k_n,n};\,1-e_{k_n,n}\leqslant\varepsilon]}{(\mathbf{E}[m(Q)])^{n-k_n}}=\theta. \tag{2.22}$$

Combining (2.20) and (2.22) with (2.19) we obtain

$$\liminf_{n\to\infty}\frac{\mathbf{E}[1-e_{k_n,n}^{Z_{k_n}}]}{(\mathbf{E}[m(Q)])^{n}}\geqslant\theta(1-\varepsilon)^z\left(1-\limsup_{n\to\infty}\widehat{\mathbf{P}}\{Z_{k_n}>z\}\right) \tag{2.23}$$

for every $\varepsilon>0$ and $z\in\mathbb{N}_0$. The weak convergence result (2.10) shows that the random variables Z_{k_n}, $n\geqslant 1$ are tight w.r.t. $\widehat{\mathbf{P}}$. Hence,

$$\lim_{z\to\infty}\limsup_{n\to\infty}\widehat{\mathbf{P}}\{Z_{k_n}>z\}=0.$$

Letting first $\varepsilon\to 0$ and then $z\to\infty$ in (2.23) gives (2.18). $\square$

Proof of Theorem 1.3. Let $k\in\mathbb{N}$ and let $B_1,\ldots,B_k$ be Borel subsets of Δ. Using first (2.14) and then Lemma 2.4, Theorem 1.1 and the independence of the Q_j, we obtain

$$\begin{aligned}
&\mathbf{P}\{Q_{i_{n,1}}\in B_1,\ldots,Q_{i_{n,k}}\in B_k\mid Z_n>0\}\\
&\quad=(\mathbf{P}\{Z_n>0\})^{-1}\mathbf{E}[\mathbf{P}\{Q_{i_{n,1}}\in B_1,\ldots,Q_{i_{n,k}}\in B_k,\ Z_n>0\mid Z_{i_{n,k}},\Pi\}]\\
&\quad=(\mathbf{P}\{Z_n>0\})^{-1}\mathbf{E}[1-e_{i_{n,k},n}^{Z_{i_{n,k}}};\ Q_{i_{n,j}}\in B_j\text{ for all }1\leqslant j\leqslant k]\\
&\quad=\frac{\mathbf{E}[1-e_{i_{n,k},n}]}{\mathbf{P}\{Z_n>0\}}\,\mathbf{E}[Z_{i_{n,k}};\ Q_{i_{n,j}}\in B_j\text{ for all }1\leqslant j\leqslant k]+\mathrm{o}(1).
\end{aligned} \tag{2.24}$$

By shift-invariance of the measure $\mathbf{P}$ and again using Theorem 1.1 we deduce from (2.24) that

$$\begin{aligned}
&\mathbf{P}\{Q_{i_{n,1}}\in B_1,\ldots,Q_{i_{n,k}}\in B_k\mid Z_n>0\}\\
&\quad=\frac{\mathbf{E}[Z_{i_{n,k}};\ Q_{i_{n,j}}\in B_j\text{ for all }1\leqslant j\leqslant k]}{(\mathbf{E}[m(Q)])^{i_{n,k}}}+\mathrm{o}(1).
\end{aligned} \tag{2.25}$$

We now show that the first term on the right-hand side of (2.25) equals $\prod_{j=1}^{k} \widehat{\mathbf{P}}\{Q \in B_j\}$. Observe that for every $1 \leqslant i_1 < \cdots < i_k$ and $k \geqslant 2$ we have

$$\begin{aligned}
&\mathbf{E}[Z_{i_k};\ Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k] \\
&\quad = \mathbf{E}[\mathbf{E}[Z_{i_k} I\{Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k\} \mid Q_1, \ldots, Q_{i_{k-1}}, Z_{i_{k-1}}]] \\
&\quad = \mathbf{E}[\mathbf{E}[Z_{i_k} I\{Q_{i_k} \in B_k\} \mid Z_{i_{k-1}}];\ Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k-1].
\end{aligned} \tag{2.26}$$

For the conditional expectation on the right-hand side of (2.26) note that the shift-invariance of $\mathbf{P}$ and relations (1.3) and (1.13) imply

$$\begin{aligned}
\mathbf{E}[Z_{i_k} I\{Q_{i_k} \in B_k\} \mid Z_{i_{k-1}}] &= \mathbf{E}[\mathbf{E}[Z_{i_k} I\{Q_{i_k} \in B_k\} \mid Q_{i_{k-1}+1}, \ldots, Q_{i_k}, Z_{i_{k-1}}] \mid Z_{i_{k-1}}] \\
&= \mathbf{E}[Z_{i_{k-1}} \exp(S_{i_k} - S_{i_{k-1}}) I\{Q_{i_k} \in B_k\} \mid Z_{i_{k-1}}] \\
&= \mathbf{E}[\exp(S_{i_k - i_{k-1}});\ Q_{i_k - i_{k-1}} \in B_k]\, Z_{i_{k-1}} \\
&= (\mathbf{E}[m(Q)])^{i_k - i_{k-1}}\, \widehat{\mathbf{P}}\{Q \in B_k\}\, Z_{i_{k-1}} \quad \mathbf{P}\text{-a.s.}
\end{aligned} \tag{2.27}$$

Plugging (2.27) into (2.26) gives

$$\begin{aligned}
&\mathbf{E}[Z_{i_k};\ Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k] \\
&\quad = (\mathbf{E}[m(Q)])^{i_k - i_{k-1}}\, \widehat{\mathbf{P}}\{Q \in B_k\}\, \mathbf{E}[Z_{i_{k-1}};\ Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k-1]
\end{aligned} \tag{2.28}$$

for every $k \geqslant 2$. For $k = 1$ relation (1.12) implies

$$\mathbf{E}[Z_{i_1};\ Q_{i_1} \in B_1] = (\mathbf{E}[m(Q)])^{i_1} \widehat{\mathbf{P}}\{Q \in B_1\}.$$

Iterating equation (2.28) we now deduce

$$\mathbf{E}[Z_{i_k};\ Q_{i_j} \in B_j \text{ for all } 1 \leqslant j \leqslant k] = (\mathbf{E}[m(Q)])^{i_k} \prod_{j=1}^{k} \widehat{\mathbf{P}}\{Q \in B_j\} \tag{2.29}$$

for every $k \in \mathbb{N}$. Combining (2.29) with (2.25) establishes the first assertion of Theorem 1.3.

For the second part of the theorem fix $\varepsilon > 0$ and let

$$A_{\varepsilon,n} := \left\{ \sup_{0 \leqslant t \leqslant 1} \left| \frac{1}{n} S_{\lfloor nt \rfloor} - t\, \widehat{\mathbf{E}}[X] \right| \geqslant \varepsilon \right\}.$$

Using first inequality (1.5) and then relation (1.13) and Theorem 1.1 we obtain

$$\begin{aligned}
\mathbf{P}\{A_{\varepsilon,n} \mid Z_n > 0\} &= (\mathbf{P}\{Z_n > 0\})^{-1} \mathbf{E}[\mathbf{E}[I\{A_{\varepsilon,n}\} I\{Z_n > 0\} \mid \Pi]] \\
&\leqslant (\mathbf{P}\{Z_n > 0\})^{-1} \mathbf{E}[\exp(S_n);\ A_{\varepsilon,n}] \\
&= \frac{(\mathbf{E}[m(Q)])^n}{\mathbf{P}\{Z_n > 0\}}\, \widehat{\mathbf{P}}\{A_{\varepsilon,n}\} \\
&= \theta^{-1} \widehat{\mathbf{P}}\{A_{\varepsilon,n}\}(1 + \mathrm{o}(1)).
\end{aligned} \tag{2.30}$$

Now let

$$N_\varepsilon := \sup\left\{ k \geqslant 1 :\ |S_k - k\widehat{\mathbf{E}}[X]| \geqslant \frac{\varepsilon}{2} k \right\}.$$

Clearly, $|S_k - k\widehat{\mathbf{E}}[X]| < (\varepsilon/2)n$ for $N_\varepsilon < k \leqslant n$. Hence, for n large ($n \geqslant 2\varepsilon^{-1}|\widehat{\mathbf{E}}[X]|$) the triangle inequality yields

$$A_{\varepsilon,n} \subset \bigcup_{k=1}^{n} \left\{ |S_k - k\widehat{\mathbf{E}}[X]| \geqslant \frac{\varepsilon}{2} n \right\} \subset \bigcup_{k=1}^{N_\varepsilon} \left\{ |S_k - k\widehat{\mathbf{E}}[X]| \geqslant \frac{\varepsilon}{2} n \right\}.$$

Since $\widehat{\mathbf{P}}\{N_\varepsilon < \infty\} = 1$ by means of the strong law of large numbers, we get

$$\limsup_{n\to\infty} \mathbf{P}\{A_{\varepsilon,n} \,|\, Z_n > 0\} \leqslant \theta^{-1} \limsup_{n\to\infty} \widehat{\mathbf{P}}\left\{ \max_{1\leqslant k \leqslant N_\varepsilon} |S_k - k\widehat{\mathbf{E}}[X]| \geqslant \frac{\varepsilon}{2} n \right\} = 0.$$

This completes the proof of Theorem 1.3. $\square$

Proof of Theorem 1.4. Recall that the total variation distance between probability measures μ and ν on a discrete space S is

$$d_{\mathrm{TV}}[\mu, \nu] = \frac{1}{2} \sum_{x\in S} |\mu(x) - \nu(x)|. \tag{2.31}$$

We first prove assertion (1.24). For every $k \leqslant n$ and $z_1, \ldots, z_k \in \mathbb{N}$ the law of total probability and relation (1.22) imply

$$\begin{aligned}
&\mathbf{P}\{Z_1 = z_1, \ldots, Z_k = z_k \,|\, Z_n > 0\} \\
&\quad = \sum_{y\in\mathbb{N}} \widehat{\mathbf{P}}\{Z_1 = z_1, \ldots, Z_k = z_k \,|\, Z_n = y\} \, \mathbf{P}\{Z_n = y \,|\, Z_n > 0\} \\
&\quad = \widehat{\mathbf{P}}\{Z_1 = z_1, \ldots, Z_k = z_k\} \sum_{y\in\mathbb{N}} \frac{\widehat{P}^{n-k}_{z_k y}}{\widehat{P}^{n}_{1y}} \mathbf{P}\{Z_n = y \,|\, Z_n > 0\} \\
&\quad = \mathbb{P}\{\widehat{Z}_1 = z_1, \ldots, \widehat{Z}_k = z_k\} \, h(k, n, z_k),
\end{aligned} \tag{2.32}$$

where

$$h(k, n, z) = \sum_{y\in\mathbb{N}} \frac{\widehat{P}^{n-k}_{zy}}{\widehat{P}^{n}_{1y}} \mathbf{P}\{Z_n = y \,|\, Z_n > 0\}, \quad z \in \mathbb{N}.$$

Putting together (2.31) and (2.32) gives

$$\begin{aligned}
&d_{\mathrm{TV}}[\mathscr{L}(Z_1, \ldots, Z_{k_n} \,|\, Z_n > 0), \mathscr{L}(\widehat{Z}_1, \ldots, \widehat{Z}_{k_n})] \\
&\quad = \frac{1}{2} \sum_{z_1, \ldots, z_{k_n} \in \mathbb{N}} \mathbb{P}\{\widehat{Z}_1 = z_1, \ldots, \widehat{Z}_{k_n} = z_{k_n}\} \, |1 - h(k_n, n, z_{k_n})| \\
&\quad = \tfrac{1}{2} \mathbb{E}\, |1 - h(k_n, n, \widehat{Z}_{k_n})|.
\end{aligned} \tag{2.33}$$

Now observe that, by Theorem 1.1 and Corollary 2.2,

$$\lim_{n\to\infty} \frac{\widehat{P}^{n-k_n}_{zy}}{\widehat{P}^{n}_{1y}} \mathbf{P}\{Z_n = y \,|\, Z_n > 0\} = r_y, \quad y, z \in \mathbb{N}. \tag{2.34}$$

Moreover, relation (1.23) and Theorem 1.1 imply

$$\frac{\widehat{P}_{zy}^{n-k_n}}{\widehat{P}_{1y}^{n}}\,\mathbf{P}\{Z_n=y\,|\,Z_n>0\}=\frac{P_{1y}^{n}\,\widehat{P}_{zy}^{n-k_n}}{\widehat{P}_{1y}^{n}\,\mathbf{P}\{Z_n>0\}}=\frac{(\mathbf{E}[m(Q)])^{n}}{\mathbf{P}\{Z_n>0\}}\,\frac{\widehat{P}_{zy}^{n-k_n}}{y}\leqslant c\widehat{P}_{zy}^{n-k_n} \tag{2.35}$$

for some $c<\infty$. Using again Corollary 2.2 we see that

$$\lim_{x\to\infty}\limsup_{n\to\infty}\sum_{y=x+1}^{\infty}\frac{\widehat{P}_{zy}^{n-k_n}}{\widehat{P}_{1y}^{n}}\,\mathbf{P}\{Z_n=y\,|\,Z_n>0\}\leqslant c\lim_{x\to\infty}\left(1-\sum_{y=1}^{x}\hat{r}_y\right)=0 \tag{2.36}$$

for every $z\in\mathbb{N}$. Relations (2.34) and (2.36) show that we may interchange summation and limiting procedures to obtain

$$\lim_{n\to\infty}h(k_n,n,z)=\sum_{y\in\mathbb{N}}\lim_{n\to\infty}\frac{\widehat{P}_{zy}^{n-k_n}}{\widehat{P}_{1y}^{n}}\,\mathbf{P}\{Z_n=y\,|\,Z_n>0\}=\sum_{y\in\mathbb{N}}r_y=1 \tag{2.37}$$

for every $z\in\mathbb{N}$. Since h is a non-negative function bounded by c (by (2.35)) and the family $\mathscr{L}(\widehat{Z}_{k_n})$, $n\geqslant 1$ is tight (by Corollary 2.2), we can use relation (2.37) to conclude

$$\lim_{n\to\infty}\mathbb{E}\,|\,1-h(k_n,n,\widehat{Z}_{k_n})\,|=0. \tag{2.38}$$

Assertion (1.24) follows from (2.33) and (2.38).

The second assertion is proved in much the same way as (1.24). Let $k\leqslant n$ and $z_0,\dots,z_k\in\operatorname{supp}\hat{r}$. Recalling the definitions of $\widehat{P},\widetilde{P}$ and $\hat{r}$ from (1.21), (1.26) and (1.27) we obtain

$$\begin{aligned}\mathbf{P}\{Z_n=z_0,\dots,Z_{n-k}=z_k\,|\,Z_n>0\}&=\frac{P_{1z_k}^{n-k}}{\mathbf{P}\{Z_n>0\}}\prod_{j=1}^{k}P_{z_jz_{j-1}}\\&=\frac{(\mathbf{E}[m(Q)])^{n}}{m(r)\,\mathbf{P}\{Z_n>0\}}\,\frac{\widehat{P}_{1z_k}^{n-k}}{\hat{r}_{z_k}}\,r_{z_0}\prod_{j=1}^{k}\widetilde{P}_{z_{j-1}z_j}\\&=\mathbb{P}(\widetilde{Z}_0=z_0,\dots,\widetilde{Z}_k=z_k)\,\bar{h}(k,n,z_k),\end{aligned} \tag{2.39}$$

where

$$\bar{h}(k,n,z)=\frac{(\mathbf{E}[m(Q)])^{n}}{m(r)\,\mathbf{P}\{Z_n>0\}}\,\frac{\widehat{P}_{1z}^{n-k}}{\hat{r}_z},\quad z\in\operatorname{supp}\hat{r}. \tag{2.40}$$

Hence,

$$\begin{aligned}
&d_{\mathrm{TV}}[\mathscr{L}(Z_n,\ldots,Z_{n-k_n}\,|\,Z_n>0),\mathscr{L}(\widetilde{Z}_0,\ldots,\widetilde{Z}_{k_n})]\\
&\quad=\frac{1}{2}\sum_{z_0,\ldots,z_{k_n}\in\operatorname{supp}\hat{r}}\mathbb{P}\{\widetilde{Z}_0=z_0,\ldots,\widetilde{Z}_{k_n}=z_{k_n}\}\,|\,1-\bar{h}(k_n,n,z_{k_n})|\\
&\qquad+\tfrac{1}{2}\mathbf{P}\{Z_j\notin\operatorname{supp}\hat{r}\text{ for some }n-k_n\leqslant j\leqslant n\,|\,Z_n>0\}\\
&\quad=\tfrac{1}{2}\mathbb{E}\,|\,1-\bar{h}(k_n,n,\widetilde{Z}_{k_n})\,|+\tfrac{1}{2}\mathbf{P}\{Z_{n-k_n}\notin\operatorname{supp}\hat{r}\,|\,Z_n>0\},
\end{aligned}\tag{2.41}$$

where for the last equality we have used the fact that if $Z_j\in\operatorname{supp}\hat{r}=\widehat{R}$, then $Z_{j+1}\in\widehat{R}\cup\{0\}$.

Clearly, to prove (1.25) we may assume $k_n\to\infty$ with no loss of generality. Then the first part of Theorem 1.4 and Corollary 2.2 imply

$$\lim_{n\to\infty}\mathbf{P}\{Z_{n-k_n}\notin\operatorname{supp}\hat{r}\,|\,Z_n>0\}=0.\tag{2.42}$$

For the other term on the right-hand side of (2.41) note that, by Theorem 1.1 and Corollaries 2.2 and 2.3,

$$\lim_{n\to\infty}\bar{h}(k_n,n,z)=1\tag{2.43}$$

for every $z\in\operatorname{supp}\hat{r}$. Hence, by the triangle inequality,

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{E}\,|\,1-\bar{h}(k_n,n,\widetilde{Z}_{k_n})|\\
&\quad\leqslant\limsup_{n\to\infty}\mathbb{P}\{\widetilde{Z}_{k_n}>z\}+\limsup_{n\to\infty}\mathbb{E}[\bar{h}(k_n,n,\widetilde{Z}_{k_n});\,\widetilde{Z}_{k_n}>z].
\end{aligned}\tag{2.44}$$

By Corollary 2.2, the first term on the right-hand side of (2.44) tends to 0 as $z\to\infty$. For the second term observe that, by (1.26), (1.27) and Proposition 2.1(i),

$$\mathbb{P}\{\widetilde{Z}_{k_n}=y\}=\sum_{x\in\mathbb{N}}r_x\widetilde{P}^{k_n}_{xy}\leqslant m(r)\sum_{x\in\mathbb{N}}\hat{r}_x\widetilde{P}^{k_n}_{xy}=m(r)\hat{r}_y$$

for every $y\in\operatorname{supp}\hat{r}$. Consequently (recall (2.40), Theorem 1.1 and Corollary 2.2 and set $\bar{h}(k,n,y):=0$, if $y\notin\operatorname{supp}\hat{r}$),

$$\begin{aligned}
\limsup_{n\to\infty}\mathbb{E}[\bar{h}(k_n,n,\widetilde{Z}_{k_n});\,\widetilde{Z}_{k_n}>z]&=\limsup_{n\to\infty}\sum_{y=z+1}^{\infty}\bar{h}(k_n,n,y)\,\mathbb{P}\{\widetilde{Z}_{k_n}=y\}\\
&\leqslant\limsup_{n\to\infty}\frac{(\mathbf{E}[m(Q)])^n}{\mathbf{P}\{Z_n>0\}}\sum_{y=z+1}^{\infty}\widehat{P}^{n-k_n}_{1y}\\
&=\theta^{-1}\left(1-\sum_{y=1}^{z}\hat{r}_y\right)\to 0\quad\text{as }z\to\infty.
\end{aligned}\tag{2.45}$$

Letting $z\to\infty$ in (2.44) we obtain

$$\lim_{n\to\infty}\mathbb{E}\,|\,1-\bar{h}(k_n,n,\widetilde{Z}_{k_n})|=0.\tag{2.46}$$

Putting together (2.41), (2.42) and (2.46) proves (1.25).

The asserted properties of the transition matrix $\widehat{P}$ have already been established in the proofs of Proposition 2.1 and Corollary 2.2. The convergence of the $\widetilde{P}$-chain is immediate from (1.27) and (2.10). $\square$

References

[1] V.I. Afanasyev, Limit theorems for a conditional random walk and some applications, Diss. Cand. Sci., Moscow, MSU, 1980.

[2] V.I. Afanasyev, Limit theorems for intermediately subcritical and strongly subcritical branching processes in a random environment, Discrete Math. Appl. 11 (2001) 105–132.

[3] V.I. Afanasyev, J. Geiger, G. Kersting, V.A. Vatutin, Criticality for branching processes in random environment, Ann. Probab. 33 (2005) 645–673.

[4] K.B. Athreya, P. Ney, Branching Processes, Springer, Berlin, 1972.

[5] M. Birkner, J. Geiger, G. Kersting, Branching processes in random environment—a view on critical and subcritical cases, in: J.-D. Deuschel, A. Greven (Eds.), Interacting Stochastic Systems, Springer, Berlin, 2005, pp. 269–291.

[6] F.M. Dekking, On the survival probability of a branching process in a finite state i.i.d. environment, Stochastic Processes Appl. 27 (1988) 151–157.

[7] J.S. D'Souza, B.M. Hambly, On the survival probability of a branching process in a random environment, Adv. Appl. Probab. 29 (1997) 38–55.

[8] W. Feller, An Introduction to Probability Theory and its Applications. I, Wiley, New York, 1968.

[9] K. Fleischmann, V.A. Vatutin, Reduced subcritical Galton–Watson processes in a random environment, Adv. Appl. Probab. 31 (1999) 88–111.

[10] J. Geiger, G. Kersting, V.A. Vatutin, Limit theorems for subcritical branching processes in random environment, Ann. Inst. H. Poincaré Probab. Statist. 39 (2003) 593–620.

[11] Y. Guivarc'h, Q. Liu, Propriétés asymptotiques des processus de branchement en environnement aléatoire, C. R. Acad. Sci. Paris Sér. I Math. 332 (4) (2001) 339–344.