# Reinforced Galton–Watson processes III: Empirical offspring distributions

Jean Bertoin<sup>\*</sup> Bastien Mallein<sup>†</sup>

#### Abstract

Reinforced Galton–Watson processes describe the dynamics of a population where reproduction events are reinforced, in the sense that offspring numbers of forebears can be repeated randomly by descendants. More specifically, the evolution depends on the empirical offspring distribution of each individual along its ancestral lineage. We are interested here in asymptotic properties of the empirical distributions observed in the population, such as concentration, evanescence and persistence. For this, we incorporate tools from the theory of large deviations to our preceding analysis [4, 5].

**Keywords:** Reinforced Galton–Watson process, empirical offspring distributions, large deviations.

AMS subject classifications: 60J80; 60J85

# 1 Introduction

Given a locally finite rooted tree, which we think of as encoding the genealogical structure of a population, we look at out-degree sequences along branches from the root. For any vertex v different from the root, this defines the *empirical offspring distribution* at v by

$$\boldsymbol{\mu}_{v} \coloneqq \frac{1}{|v|} \sum_{i=0}^{|v|-1} \boldsymbol{\delta}_{d(v_{i})}, \tag{1.1}$$

where  $|v| \ge 1$  denotes the generation of v,  $d(v_i)$  the number of children begotten by the forebear  $v_i$  of v at generation i < |v|, and, as usual,  $\delta$  designates a Dirac measure.

The classical Galton–Watson tree is an elementary population model in which individuals reproduce independently and with a fixed reproduction law. An easy combination of the well-known many-to-one formula for Galton–Watson trees and Sanov's theorem entails that for any small enough neighborhood G of some probability measure  $\boldsymbol{\varrho}$  on  $\mathbb{N} = \{1, 2, \ldots\}$ , the mean number of vertices v

<sup>\*</sup>Institute of Mathematics, University of Zurich, Switzerland, jean.bertoin@math.uzh.ch. <sup>†</sup>Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, UPS, F-31062

Toulouse Cedex 9, France, bastien.mallein@math.univ-toulouse.fr.

at generation n with  $\mu_v \in G$  goes to 0 or to  $\infty$  exponentially fast as  $n \to \infty$ , depending on whether the entropy of  $\rho$  relative to  $\nu$ ,

$$H(\boldsymbol{\varrho}|\boldsymbol{\nu}) \coloneqq \sum \varrho(k) \log\left(\frac{\varrho(k)}{\nu(k)}\right), \qquad (1.2)$$

is larger or smaller than

$$\langle \boldsymbol{\varrho}, \ln \rangle \coloneqq \sum \varrho(k) \ln k.$$
 (1.3)

Here and throughout the rest of this text, the notation  $\sum$  is used for summation over k, and we denote by ln the restriction of the natural logarithm log to  $\mathbb{N}$ .

In the first case,  $\boldsymbol{\varrho}$  is almost-surely evanescent for the empirical offspring distributions of the Galton–Watson tree, in the sense that with probability one, we can find a neighborhood G of  $\boldsymbol{\varrho}$  such that the number of vertices v with  $\boldsymbol{\mu}_v \in G$  is finite. In the second case,  $\boldsymbol{\varrho}$  is persistent with positive probability, as a.s. on the event of survival of the tree, for any neighbourhood G of  $\boldsymbol{\varrho}$ , there will be infinitely many vertices v such that  $\boldsymbol{\mu}_v \in G$ . Precise stronger statements will be given and proven in Section 2.3; they should be viewed as benchmarks for the present study.

Reinforced Galton–Watson processes were introduced in [4] by incorporating random repetitions of reproduction events in the evolution of usual Galton– Watson processes. This depends on a memory parameter  $q \in (0, 1)$ , which accounts for the frequency of repetitions. For every  $n \geq 1$ , conditionally given the offspring numbers of individuals up to generation n-1, individuals at generation n reproduce independently one from the others and as follows. With probability q, the number of children of v is sampled from the empirical offspring distribution  $\mu_v$ , and with complementary probability 1-q, it is rather sampled from the reproduction law  $\nu$ . In words, for every individual, with the obvious exception of the ancestor, we pick a forebear uniformly at random on its ancestral lineage. Then, either with probability q, this individual begets the same number of children as the selected forebear, or with complementary probability 1-q, the number of its children is an independent sample from the reproduction law  $\nu$ .

The growth of the averaged population size of reinforced Galton–Watson processes has been determined in [4] using generating function techniques, and then the survival and a.s. growth rate have been further studied in [5] introducing martingale techniques. Here, we incorporate tools from large deviations theory to our analysis and investigate asymptotic properties of empirical offspring distributions of individuals at large generations. Our main result, Theorem 4.1, states an exponential concentration around a specific reproduction law, and provides a sufficient condition for a given reproduction law to be evanescent or persistent; see Definition 2.1 below for a precise definition.

The rest of the present work is organized as follow. Section 2 is devoted to a few preliminaries on trees, basic background from large deviations theory, and the presentation of benchmark results about empirical offspring distributions in classical Galton-Watson processes. Section 3 deals with versions of Sanov's Large Deviations Principle for reinforced sampling; it mainly stems from earlier contributions by Budhiraja and Waterbury [12] and ourselves [4]. Our main results are presented and proven in Section 4. Last, in Section 5, we also point at a sufficient condition for survival of reinforced Galton–Watson processes,

which generalizes than that obtained previously in [5] and follows directly from our analysis.

# 2 Preliminaries

In this section, we first introduce some definition and properties related to the study of genealogical trees. In particular, we define the notions of <u>evanescence</u> and <u>persistence</u> for empirical offspring distributions. We then recall in Section 2.2 the Large Deviations Principles on finite alphabets, before applying it to the Galton-Watson tree in Section 2.3.

#### 2.1 Some definition and a many-to-one formula

We fix a probability measure  $\boldsymbol{\nu}$  on  $\mathbb{Z}_+ = \{0, 1, \ldots\}$  with support

$$S = \{k \in \mathbb{Z}_+ : \nu(k) > 0\}.$$

We assume throughout this work that S is finite<sup>1</sup>,  $\#S < \infty$ ; the uninteresting case when  $S \cap \mathbb{N}$  is a singleton is further ruled out. Recall that the restriction of the logarithmic function to S is denoted here by  $\ln^2$ ; it will play an important role for this study. This requires a few obvious conventions when  $0 \in S$ , namely

$$\ln 0 = -\infty , \ e^{-\infty} = 0 , \ 0 \ln 0 = 0.$$
(2.1)

We write  $\mathcal{P}_S \subset \mathbb{R}^S$  for the simplex of probability measures  $\boldsymbol{\varrho}$  on S; the support of such  $\boldsymbol{\varrho}$  may of course be a strict subset of S. We use the notation

$$\langle \boldsymbol{\varrho}, \lambda \rangle = \sum \varrho(k) \lambda(k), \qquad \lambda \in [-\infty, \infty)^S,$$

where, by an earlier convention, the sum runs over  $k \in S$  (in particular, observe that  $\langle \boldsymbol{\varrho}, \ln \rangle = -\infty$  iff  $\varrho(0) > 0$ ).

Next, consider any rooted tree, say T, such that out-degrees of vertices of T always belong to S. It is convenient to use implicitly the Ulam-Harris-Neveu framework<sup>3</sup>, which enables us to view T as a subset of the Ulam tree  $\bigcup_{n\geq 0} \mathbb{N}^n$ . We stress that T is fixed (deterministic) in this section; however the same notation T will rather refer to a random tree from Section 2.3 onwards. Recall that empirical offspring distributions  $\boldsymbol{\mu}_v$  of vertices of T have been introduced in (1.1), so that  $\boldsymbol{\mu}_v \in \mathcal{P}_S$ . We now give a key definition for the present work.

**Definition 2.1.** We say that  $\boldsymbol{\varrho} \in \mathcal{P}_S$  is:

(i) evanescent if there exists some neighborhood G of  $\rho$  in  $\mathcal{P}_S$  such that

$$\#\{v \in T : \boldsymbol{\mu}_v \in G\} < \infty,$$

<sup>&</sup>lt;sup>1</sup>This restriction comes from our previous work [4, 5], in which we observed that a reinforced Galton-Watson process with a reproduction law of unbounded support would grow at a super-exponential rate, thus would require very different methods to study.

<sup>&</sup>lt;sup>2</sup>We found convenient to write  $\ln k$  for the logarithm of a positive integer k and  $\log x$  when the argument is a positive real number x, even though of course the first function is merely the restriction of the second to  $\mathbb{Z}_+$ . That way, we treat  $\ln$  as a vector in  $\mathbb{R}^S$ .

<sup>&</sup>lt;sup>3</sup>In short, this amounts to enumerating progenies in T and agreeing that the name of a child is obtained by aggregating its rank in the progeny at the end of the name of its parent; see e.g. [20].

(ii) weakly persistent if  $\rho$  is not evanescent, that is, if

$$\#\{v \in T : \boldsymbol{\mu}_v \in G\} = \infty$$

for any neighborhood G of  $\boldsymbol{\varrho}$  in  $\mathcal{P}_S$ ,

(iii) strongly persistent if there exists an infinite line of descent  $(v_n)_{n\geq 1}$  in T with

$$\lim_{n\to\infty}\boldsymbol{\mu}_{v_n}=\boldsymbol{\varrho}$$

- Remark 2.2. 1. The definition is only relevant for infinite trees, since by a compactness argument, T is finite if and only if all probability laws on S are evanescent.
  - 2. Plainly, a law  $\boldsymbol{\varrho}$  with  $\boldsymbol{\varrho}(0) > 0$  cannot arise as an empirical offspring distribution, and thus is necessarily evanescent.

Example 2.3. We determine the persistent and evanescent laws for the tree T constructed as follows: starting with a binary tree, we graft to each node of that tree an infinite line. In this tree, each individual has either 1 or 3 children, depending on whether they belong to the original binary tree or were later added. This tree can be written with Ulam-Harris notation as:

$$T = \left\{ u \in \bigcup_{n \in \mathbb{Z}_+} \{1, 2, 3\}^n : \text{ if } u(i) = 3 \text{ then } u(j) = 1 \text{ for all } i+1 \le j \le |u| \right\}.$$

The strongly persistent measures are  $\delta_1$  and  $\delta_3$ , whereas  $t\delta_1 + (1-t)\delta_3$  is a weakly persistent measure for all  $t \in (0,1)$ . There are no evanescent law in  $\mathcal{P}_{\{1,3\}}$  in that example.



Figure 1: A representation of the tree T constructed in Example 2.3 encoded with its Ulam–Harris notation. The vertices of the original binary tree are represented as circles.

In the forthcoming Section 2.3, we will analyze the case when T is random and encodes a Galton–Watson process with reproduction law  $\nu$ . We study the case of a multitype Galton–Watson tree in Section 2.4, generalizing the example above. The reinforced version, which is the main purpose of this work, will then be considered in Section 4.

We end this section by presenting a version of the well-known many-to-one formula to compute the number of vertices at a fixed generation with a given empirical offspring distribution in T. Imagine that we distinguish a (possibly finite) line of descent  $(U_n)$ , randomly and recursively as follows. The root of T is of course the first element  $U_0$  of the distinguished line. Then, for each  $n \ge 0$ , as long as the distinguished vertex  $U_n$  is not a leaf of T,  $U_{n+1}$  is chosen uniformly at random among the offspring of  $U_n$ . If  $U_n$  is a leaf of T, then the distinguished line ends.

We shall refer to  $(U_n)$  as a harmonic line<sup>4</sup> in T. For each generation k for which the vertex  $U_k$  is well-defined, we write  $\hat{d}_k = d(U_k)$  for the number of children on the harmonic line at generation k. If the line has stopped by generation k, we write  $\hat{d}_k = 0$  by convention. For all  $n \in \mathbb{N}$ , we denote by

$$\widehat{\boldsymbol{\mu}}_n = rac{1}{n} \sum_{j=0}^{n-1} \boldsymbol{\delta}_{\widehat{d}_j},$$

for the empirical offspring distribution at generation n of the harmonic line. Note that  $\hat{\mu}_n = \mu_{U_n}$  provided the line survives up to generation n. We also remark that  $\hat{\mu}(0) > 0$  if and only if the line stopped before generation n.

**Lemma 2.4** (Many-to-one formula). Take any  $(x_0, \ldots, x_{n-1}) \in (S \setminus \{0\})^n$  for some  $n \ge 1$  and set  $\mu = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{x_j}$ . Using the notation (1.3), we have

$$\begin{aligned} &\#\{v \in T : |v| = n \text{ and } d_{v_j} = x_j \text{ for all } j < n\} \\ &= \mathbb{P}\left(\widehat{d}_0 = x_0, \dots, \widehat{d}_{n-1} = x_{n-1}\right) \times \exp(n\langle \boldsymbol{\mu}, \ln \rangle), \end{aligned}$$

In particular, for any probability measure  $\rho \in \mathcal{P}_S$  such that the measure  $n\rho$  is integer-valued, there is the identity

$$#\{v \in T : |v| = n \text{ and } \boldsymbol{\mu}_v = \boldsymbol{\varrho}\} = \mathbb{P}\left(\boldsymbol{\mu}_{U_n} = \boldsymbol{\varrho}\right) \times \exp(n\langle \boldsymbol{\varrho}, \ln\rangle).$$

*Remark* 2.5. We stress that in the case  $\rho(0) > 0$ ,  $\rho$  cannot be an empirical offspring distribution, and that the right-hand side in the statement equals 0 by our conventions.

*Proof.* It is immediate from the definition of  $U_n$  that for all  $v \in T$  with |v| = n, we have

$$\mathbb{P}(U_n = v) = \mathbb{P}(U_1 = v_1, \dots, U_n = v_n) = \prod_{j=0}^{n-1} \frac{1}{d(v_j)}$$

<sup>&</sup>lt;sup>4</sup>To avoid a possible misunderstanding, we stress that  $U_n$  shall not be thought of as vertex picked uniformly at random amongst vertices at generation n, but rather results by selecting a child uniformly at random in a progeny, generation after generation. Observe that for the tree in Example 2.3,  $U_2$  belongs to the binary subtree of T with probability 4/9, despite these vertices representing 4/7th of the second generation. Indeed, at each step, the harmonic line has probability 2/3 to stay in the binary subtree.

As a result, for any  $n \ge 1$  and  $(x_0, \ldots, x_{n-1}) \in S^n$ , there is the identity

$$\mathbb{P}(d(U_0) = x_0, \dots, d(U_{n-1}) = x_{n-1})$$
  
=  $\sum_{|v|=n} \mathbb{P}(U_n = v) \mathbf{1}_{\{d(v_0) = x_0, \dots, d(v_{n-1}) = x_{n-1}\}}$   
=  $\frac{1}{\prod_{j=0}^{n-1} x_j} \#\{v \in T : |v| = n \text{ and } d(v_i) = x_i \text{ for all } i = 0, \dots, n-1\}.$ 

The first claim follows readily rewriting  $\prod_{j=0}^{n-1} x_j = \exp(n\langle \boldsymbol{\mu}, \ln \rangle)$ , the second one by summing over all  $(x_0, \ldots, x_{n-1})$  such that  $\sum_{j=0}^{n-1} \boldsymbol{\delta}_{x_j} = n\boldsymbol{\varrho}$ .

#### 2.2 Basic background from large deviations theory

We briefly recall in this section a few key facts from large deviations theory for finite alphabets that will be needed later on, referring e.g. to [13, Chapters 2 and 3] and [11] for details and references.

We denote the logarithmic moment generating function of  $\nu$  by

$$\Lambda_0(\lambda) \coloneqq \log\left(\sum \nu(k) \mathrm{e}^{\lambda(k)}\right), \qquad \lambda \in \mathbb{R}^S;$$

the index 0 in  $\Lambda_0$  is meant to indicate later on that we are dealing with the memory parameter q = 0. Its Fenchel-Legendre transform is defined as

$$\Lambda^*_0(oldsymbol{\mu})\coloneqq \sup_{\lambda\in\mathbb{R}^S}\left(\langleoldsymbol{\mu},\lambda
angle-\Lambda_0(\lambda)
ight),\qquadoldsymbol{\mu}\in\mathbb{R}^S.$$

Recall also from (1.2) that  $H(\boldsymbol{\varrho}|\boldsymbol{\nu})$  stands for the entropy of  $\boldsymbol{\varrho}$  relative to  $\boldsymbol{\nu}$ , and that the variational formula allows for the identification

$$H(\boldsymbol{\varrho}|\boldsymbol{\nu}) = \Lambda_0^*(\boldsymbol{\varrho}). \tag{2.2}$$

The relative entropy of a measure is linked to large deviations theory by the fundamental theorem of Sanov [13, Theorem 2.1.10]. Given some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  and a sequence  $(\xi_i)_{i>1}$  of random variables in S, we write

$$L_n \coloneqq \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i}, \qquad n \ge 1$$

for the sequence of empirical measures they induce. Recall that  $(L_n)_{n\geq 1}$  satisfies the Large Deviation Principle (LDP) with rate function  $I : \mathcal{P}_S \to [0, \infty]$  if the latter is a lower semi-continuous function, and if we have for every closed  $F \subset \mathcal{P}_S$ ,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(L_n \in F) \le -\inf_{\boldsymbol{\varrho} \in F} I(\boldsymbol{\varrho}),$$

and for every open  $G \subset \mathcal{P}_S$ ,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(L_n \in G) \ge -\inf_{\boldsymbol{\varrho} \in G} I(\boldsymbol{\varrho})$$

In this setting, Sanov's Theorem states that if  $\mathbb{P}_0$  is a probability measure under which the sequence  $(\xi_i)$  is i.i.d. with law  $\boldsymbol{\nu}$ , then the sequence of empirical measures  $(L_n)_{n\geq 1}$  satisfies the LDP under  $\mathbb{P}_0$  with rate function

$$I_0 = H(\cdot|\boldsymbol{\nu}) = \Lambda_0^*. \tag{2.3}$$

The Gibbs Conditioning Principle (see [13, Section 3.3]) is an important consequence of Sanov's Theorem; it can be stated as follows. Recall that  $(\xi_i)$ is i.i.d. with law  $\boldsymbol{\nu}$  under  $\mathbb{P}_0$ , and consider any closed convex subspace  $\Gamma$  of  $\mathcal{P}_S$ with a non-empty interior. For any  $n \geq 1$ , under the conditional probability measure  $\mathbb{P}_0(\cdot \mid L_n \in \Gamma)$ , the sequence  $(\xi_i)_{1 \leq i \leq n}$  is exchangeable, and if  $\boldsymbol{\nu}_{\Gamma,n}$ denotes its one-dimensional marginal distribution,

$$\nu_{\Gamma,n}(k) \coloneqq \mathbb{P}_0(\xi_1 = k \mid L_n \in \Gamma), \qquad k \in S,$$

then

$$\lim_{n\to\infty}\boldsymbol{\nu}_{\Gamma,n} = \arg\min_{\Gamma} H(\cdot|\boldsymbol{\nu}),$$

where the right-hand side stands for the unique<sup>5</sup>  $\gamma_0 \in \Gamma$  such that

$$H(\boldsymbol{\gamma}_0|\boldsymbol{\nu}) = \min_{\boldsymbol{\Gamma}} H(\cdot|\boldsymbol{\nu})$$

A related result in this setting concerns the most likely trajectory taken by the process of empirical measures conditionally on the rare event  $\{L_n \in \Gamma\}$ . For any sequence of positive integers  $(k(n))_{n\geq 1}$  with

$$\lim_{n \to \infty} k(n) = \infty \quad \text{and} \quad \limsup_{n \to \infty} k(n)/n \leq 1,$$

there is the convergence in probability

$$\lim_{n \to \infty} \mathbb{P}_0(\operatorname{dist}_{\mathcal{P}_S}(L_{k(n)}, \gamma_0) < \varepsilon \mid L_n \in \Gamma) = 1, \quad \text{for all } \varepsilon > 0, \quad (2.4)$$

where dist<sub> $\mathcal{P}_S$ </sub> stands, say, for the maximal distance on  $\mathcal{P}_S$ . In other words, the most likely trajectory for  $L_n$  to end up in  $\Gamma$  consists in staying at all time within a ball of radius  $\varepsilon$  centred at  $\gamma_0$ . We also refer to Mogulskii's Theorem [13, Section 5.1] for a much deeper sample path LDP for random walks.

### 2.3 The benchmark case of Galton–Watson trees

This section focuses on the case when T is a Galton–Watson tree with reproduction law  $\nu$ , i.e. without reinforcement. We write  $\mathbb{P}_0$  for the distribution of T, where the index 0 is again meant to stress that the memory parameter is q = 0. We write also

$$m_{\pmb{\nu}}\coloneqq \sum k\nu(k)$$

for the mean reproduction number, and  $\overline{\boldsymbol{\nu}}_0$  for the size-biased reproduction law, that is

$$\overline{\nu}_0(k) = k\nu(k)/m_{\nu}, \qquad k \in S$$

The following result will serve for us as a benchmark in the study of the reinforced setting; it analyzes the empirical offspring distribution at a large generation n of T. Although it is well-known that a typical individual at the nth generation of a Galton-Watson process has an empirical offspring distribution close to the size-biased reproduction law  $\overline{\nu}_0$  (see e.g. [3]), estimations of the number of individuals with a given empirical offspring distribution appear to be less well-known.

<sup>&</sup>lt;sup>5</sup>Using the compactness of  $\Gamma$ , and the convexity of  $\Gamma$  and strict convexity of  $H(\cdot|\nu)$ , the existence of this unique minimizer is guaranteed.

**Theorem 2.6.** Recall the notation (1.3). The following assertions hold for usual Galton–Watson trees:

(i) Concentration around the size-biased reproduction law: For every neighborhood G of  $\overline{\nu}_0$ , there exists  $\varepsilon = \varepsilon(G) > 0$  such that for all  $n \ge 1$ ,

$$\mathbb{E}_{0}(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \notin G\})) \leq e^{-\varepsilon n} \mathbb{E}_{0}(\#\{v \in T : |v| = n\})$$

(ii) **Evanescent laws:** Any  $\boldsymbol{\varrho} \in \mathcal{P}_S$  with

$$\langle \boldsymbol{\varrho}, \ln \rangle < \Lambda_0^*(\boldsymbol{\varrho})$$

is evanescent,  $\mathbb{P}_0$ -a.s.

(iii) Strongly persistent laws: For any  $\rho \in \mathcal{P}_S$  with

 $\langle \boldsymbol{\varrho}, \ln \rangle > H(\boldsymbol{\varrho} | \boldsymbol{\nu}),$ 

 $\varrho$  is strongly persistent,  $\mathbb{P}_0$ -a.s. conditionally on non-extinction.

The quantities  $\Lambda_0^*(\boldsymbol{\varrho})$  and  $H(\boldsymbol{\varrho}|\boldsymbol{\nu})$  in (ii) and (iii) are identical by the variational formula (2.2); we used on purpose these two different naming conventions to make comparison with the reinforced case in the forthcoming Theorem 4.1 more transparent. We also observe that if  $\varrho(0) > 0$ , then  $\langle \boldsymbol{\varrho}, \ln \rangle = -\infty$  and the condition in (ii) is automatically fulfilled (we already pointed out that such a probability measure is necessarily evanescent).

- Remark 2.7. 1. We know that  $\mathbb{E}_0(\#\{v \in T : |v| = n\}) = m_{\nu}^n$ . Further, in the critical or subcritical case  $m_{\nu} \leq 1$ , T is finite and any law  $\boldsymbol{\varrho} \in \mathcal{P}_S$ is obviously evanescent,  $\mathbb{P}_0$ -a.s. This does not contradict this claim of concentration around the size-biased reproduction law in expectation.
  - 2. Theorem 2.6 essentially shows that any distribution is either evanescent of strongly persistent for a Galton-Watson tree (with edge cases when  $\langle \boldsymbol{\varrho}, \ln \rangle = H(\boldsymbol{\varrho}|\boldsymbol{\nu})$  to be separately treated). We consider in Section 2.4 an example of a random tree with weakly –but not strongly– persistent measures.
  - 3. Given the concentration of the empirical offspring distribution of the population at time n around  $\overline{\nu}_0$ , it is worth noting that immediate algebra yield the identity

$$\log m_{\nu} - H(\boldsymbol{\varrho}|\boldsymbol{\overline{\nu}}) = \langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho}|\boldsymbol{\nu}). \tag{2.5}$$

This gives an intuitive reason for the importance of comparing the quantities (1.2) with (1.3) as mentioned in the introduction to estimate the presence of individuals with an empirical offspring distribution around  $\rho$ .

Theorem 2.6 can be seen as a disguised version of [8, Theorem A] that describe the distribution of individuals in a multidimensional branching random walk. To see the connection, let us remark that

$$v \in T \mapsto (|v|\mu_v(j), j \in S)$$

is a branching random walk on  $\mathbb{Z}^S$ . In this branching random walk, an individuals at position  $x \in \mathbb{Z}^S$  creates j children at position  $x + e_j$  with probability  $\nu(j)$ , where  $e_j(i) = \mathbf{1}_{\{i=j\}}$ . Consequently, the set of non-empty vertices at the *n*th generation of this branching random walk can be well-approached by nC, where  $C = \{ \boldsymbol{x} \in \mathbb{R}^S : \Lambda^*(\boldsymbol{x}) < m_{\boldsymbol{\nu}} \}$ . More precise estimates, such as the ones obtained in [10, Theorem 5] for non-lattice branching random walks, could be adapted to the current setting as well.

The proof of Theorem 2.6 will easily follow from the well-known spinal decomposition which we now briefly present, tailored for our purposes. Let  $a: S \to \mathbb{R}_+$  be a nonnegative function of the out-degrees such that

$$\sum_{k\in S} ka(k)\nu(k) = 1.$$
(2.6)

The value of a(0) being irrelevant for (2.6), we always assume, without loss of generality, that a(0) = 0. We then denote by  $\boldsymbol{\nu}^a$  the probability measure on S given by

$$\nu^a(k) = ka(k)\nu(k).$$

The best known and most important example is when  $a(k) = 1/m_{\nu}$  for all  $k \ge 1$ ; then  $\nu^a$  is the size-biased version  $\overline{\nu}_0$  of the reproduction law.

Recall that for every vertex  $v \in T$ , d(v) denotes its out-degree, |v| its generation, and  $v_j$  its forebear at generation  $j \leq |v|$ . It is seen from (2.6) and the branching property that the process

$$W_n^a \coloneqq \sum_{|v|=n} \prod_{j=0}^{n-1} a(d(v_j)), \qquad n \ge 1,$$

is a  $\mathbb{P}_0$ -martingale with expectation  $\mathbb{E}_0(W_n^a) = 1$ . In the case when  $a(k) \equiv 1/m_{\nu}$ ,  $W_n^a$  is the fundamental martingale for Galton–Watson processes appearing e.g. in [2, Theorem 1 on page 9]. This enables us to define a probability measure  $\mathbb{P}_0^a$  describing the joint law of a random tree T together with an infinite random line of descent  $(V_n^a)_{n\geq 0}$ , usually referred to as a spine, as follows. For any  $n\geq 1$ , any subtree  $t_n$  of the Ulam tree with height n and any v vertex of  $t_n$  at height n, if we write  $T_n$  for the subtree obtained by pruning T at generation n, then

$$\mathbb{P}_{0}^{a}(T_{n} = t_{n}, V_{n}^{a} = v) = \mathbb{P}_{0}(T_{n} = t_{n}) \prod_{j=0}^{n-1} a(d_{t_{n}}(v_{j})), \qquad (2.7)$$

where  $d_{t_n}(v_i)$  stands for the out-degree in  $t_n$  of the forebear  $v_i$  of v at generation i < n.

Remark 2.8. In the case  $\nu(0) = 0$  when the probability of an empty progeny equals zero, the harmonic line  $(U_n)$  of Section 2.1 is infinite and can also be viewed as a spine  $(V_n^a)$  for the function  $a(k) \equiv k^{-1} / \sum j^{-1} \nu(j)$ . The martingale is then trivial,  $W_n^a \equiv 1$ .

We can now state a necessary and sufficient condition for the uniform integrability of the martingale  $(W_n^a)$ , extending the Kesten-Stigum  $L \log L$  criterion [18] for the case  $a \equiv 1/m_{\nu}$ . This result can be seen as a version of Biggins' theorem [7], we adapt here a proof by Lyons [19] of this fact.

**Lemma 2.9.** The following assertions hold:

- (i) Under  $\mathbb{P}^a_0$ , the sequence  $(d(V^a_n))_{n\geq 0}$  of out-degrees along the spine is i.i.d. with law  $\boldsymbol{\nu}^a$ .
- (ii) If

$$\sum \nu^a(k) \log a(k) < 0,$$

then  $(W_n^a)_{n\geq 1}$  is a uniformly integrable martingale under  $\mathbb{P}_0$ . Otherwise  $\lim_{n\to\infty} W_n^a = 0$  a.s.

*Proof.* The first assertion is immediate from (2.7). For the second, note from the first that under  $\mathbb{P}_0^a$ , the process

$$\log\left(\prod_{j=0}^{n-1} a(d(V_j^a))\right) = \sum_{j=0}^{n-1} \log a(d(V_j^a)), \qquad n \ge 1,$$

is a random walk with drift

$$\mathbb{E}^a_0\left(\log d(V^a_n)\right) = \sum \nu^a(k) \log a(k)$$

If  $\sum \nu^a(k) \log a(k) < 0$ , then  $\log \left(\prod_{j=0}^{n-1} a(d(V_j^a))\right)$  is a random walk with negative drift. It follows from the law of large numbers that

$$\sum_{n=1}^{\infty}\prod_{j=0}^{n-1}a(d(V_{j}^{a}))<\infty,\qquad\mathbb{P}_{0}^{a}\text{-a.s.}$$

Since out-degrees are bounded, we deduce from the spinal decomposition (see [9, Section 12]) that  $\limsup_{n\to\infty} W_n^a < \infty$ ,  $\mathbb{P}_0^a$ -a.s. By Durrett's criterion (see [14, Theorem 4.3.5] or [9, Theorem 3]), this entails in turns the uniform integrability under  $\mathbb{P}_0$ .

Reciprocally, if  $\sum \nu^a(k) \log a(k) \ge 0$ , the drift of  $\log \left( \prod_{j=0}^{n-1} a(d(V_j^a)) \right)$  is well-defined and non-negative. Thus

$$\limsup_{n \to \infty} W_n^a \ge \limsup_{n \to \infty} \prod_{j=0}^{n-1} a(d(V_j^a)) = \infty \quad \mathbb{P}_0^a \text{-a.s.}$$

Applying again Durrett's criterion, we deduce that  $W_n^a \to 0$  a.s. under  $\mathbb{P}_0$ .  $\Box$ 

We now have all the ingredients needed to establish Theorem 2.6.

*Proof of Theorem 2.6.* (i) The average population size of a Galton–Watson process at generation  $n \ge 0$  is

$$\mathbb{E}_0(\#\{v \in T : |v| = n\}) = m_{\boldsymbol{\nu}}^n.$$

Take  $a(k) \equiv 1/m_{\nu}$  and write  $\overline{\mathbb{P}}_0 = \mathbb{P}_0^a$ . Since then

$$\prod_{j=0}^{n-1} a(d_{t_n}(v_j)) = m_{\nu}^{-n}$$

for any vertex  $v_n$  at generation n in any tree  $t_n$  with height n, we deduce from (2.7) the identity

$$\frac{\mathbb{E}_0(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \notin G\})}{\mathbb{E}_0(\#\{v \in T : |v| = n\})} = \overline{\mathbb{P}}_0(L_n \in F),$$

where  $F = \mathcal{P}_S \backslash G$  and

$$L_n \coloneqq \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{d(V_{i-1}^a)}.$$

As we know from Lemma 2.9(i) that under  $\overline{\mathbb{P}}_0$ ,  $L_n$  is the empirical measure of i.i.d. variables distributed according to the size-biased reproduction law  $\overline{\nu}_0$ , we deduce from Sanov's large deviations upper-bound that

$$\limsup_{n \to \infty} \frac{1}{n} \log \overline{\mathbb{P}}_0(L_n \in F) \le -\inf_{\boldsymbol{\varrho} \in F} H(\boldsymbol{\varrho} | \overline{\boldsymbol{\nu}}_0).$$

The entropy relative to  $\overline{\nu}_0$ ,  $\rho \mapsto H(\rho | \overline{\nu}_0)$ , is a continuous map on  $\mathcal{P}_S$  which is strictly positive except at  $\overline{\nu}_0$ . Since F avoids a neighborhood of  $\overline{\nu}_0$ , the right-hand side above is strictly negative.

(ii) Let  $\rho$  be as in the statement. By continuity of the entropy relative to  $\nu$ , we can find a closed neighborhood F of  $\rho$  such that

$$\sum \varrho'(k) \log k < H(\varrho'|\nu), \qquad \text{for all } \varrho' \in F.$$

On the other hand, we have seen in (i) that

$$\mathbb{E}_0(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in F\}) = m_{\boldsymbol{\nu}}^n \times \overline{\mathbb{P}}_0(L_n \in F);$$

and then again by Sanov's large deviations upper-bound,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_0(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in F\}) \le \log(m_{\boldsymbol{\nu}}) - \inf_{\boldsymbol{\varrho}' \in F} H(\boldsymbol{\varrho}'|\overline{\boldsymbol{\nu}}_0).$$

It now suffices to write

$$H(\boldsymbol{\varrho}'|\boldsymbol{\overline{\nu}}_0) = \sum \varrho'(k) \log \left(\frac{\varrho'(k)m_{\boldsymbol{\nu}}}{k\nu(k)}\right) = H(\boldsymbol{\varrho}'|\boldsymbol{\nu}) + \log(m_{\boldsymbol{\nu}}) - \sum \varrho'(k) \ln k,$$

see (2.5), to get

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_0(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in F\}) < 0$$

Summing over generations yields

$$\mathbb{E}_0(\#\{v\in T:\boldsymbol{\mu}_v\in F\})<\infty,$$

and we conclude that  $\boldsymbol{\varrho}$  is evanescent,  $\mathbb{P}_0$ -a.s.

(iii) Let  $\boldsymbol{\varrho}$  be as in the statement; in particular  $\varrho(0) = 0$ . We set  $a(k) := \varrho(k)/(k\nu(k))$  for  $k \in S$  –which satisfies (2.6)– so that  $\boldsymbol{\varrho} = \boldsymbol{\nu}^a$ . By Lemma 2.9(i) and the law of large numbers for the sequence  $(d(V_n^a))_{n\geq 0}$  of out-degrees along

the spine, we see that  $\rho$  is strongly persistent  $\mathbb{P}^a_0$ -a.s. On the other hand, we have by the variational formula that

$$\sum \nu^{a}(k) \ln k = \langle \boldsymbol{\varrho}, \ln \rangle > H(\boldsymbol{\varrho} | \boldsymbol{\nu}) = \sup_{\lambda \in \mathbb{R}^{S}} \left( \langle \boldsymbol{\varrho}, \lambda \rangle - \Lambda_{0}(\lambda) \right).$$

Applying this inequality to  $\lambda(k) = \log(ka(k))$ , so that  $\Lambda_0(\lambda) = 0$ , we have

$$\sum \nu^{a}(k) \ln k > \sum \nu^{a}(k) \log(ka(k)),$$

so the assumption of Lemma 2.9(ii) holds.

Note from (2.7) that for every  $n \geq 1$ , the distribution of  $T_n$  under  $\mathbb{P}_0^a$  is absolutely continuous with respect to that under  $\mathbb{P}_0$  with density  $W_n^a$ , so the uniform integrability of the martingale  $W_n^a$  enables us to deduce that the distribution of whole tree T under  $\mathbb{P}_0^a$  is again absolutely continuous with respect to that under  $\mathbb{P}_0$ , with density given by the terminal value  $W_\infty^a$ . Hence  $\boldsymbol{\varrho}$  is strongly persistent with a strictly positive probability under  $\mathbb{P}_0$ , and the stronger claim conditionally on non-extinction follows from a standard argument involving the branching property.

We now conclude this section with a refinement of Theorem 2.6 which will be useful in the next section. This result allows to estimate the almost sure growth rate of the number of individuals with an empirical distribution in a neighbourhood of  $\rho$ .

**Lemma 2.10.** Let  $\boldsymbol{\varrho} \in \mathcal{P}_S$  and G an open neighborhood of  $\boldsymbol{\varrho}$ .

(i) If  $m_{\nu} > 1$  and  $\langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho}|\boldsymbol{\nu}) > 0$ , then we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \# \{ v \in T : |v| = n \& \boldsymbol{\mu}_v \in G \} \ge \langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho} | \boldsymbol{\nu}),$$

 $\mathbb{P}_0$ -a.s. on the survival event (i.e. when T is infinite).

(ii) If  $m_{\nu} < 1$ , then we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_0(\exists v\in T:|v|=n\,\&\,\boldsymbol{\mu}_v\in G)\geq\langle\boldsymbol{\varrho},\mathrm{ln}\rangle-H(\boldsymbol{\varrho}|\boldsymbol{\nu}).$$

*Proof.* (i) We prove this result using an analogue reasoning to [6]. Let B be an open (convex) ball centered at  $\boldsymbol{\varrho}$  with radius r > 0 sufficiently small so that the ball of same center and radius 2r is contained in G. We recall that, by Lemma 2.9 and Sanov's large deviations theorem, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}_0(\#\{v\in T: |v|=n\,\&\,\boldsymbol{\mu}_v\in\overline{G}\})\geq \langle\boldsymbol{\varrho},\ln\rangle-H(\boldsymbol{\varrho}|\boldsymbol{\nu}).$$

Using that this quantity is positive by our assumptions, we construct a supercritical Galton-Watson tree embedded in T as follows. For all  $\varepsilon > 0$  small enough, we choose K such that

$$\frac{1}{K}\log \mathbb{E}_0(\#\{v \in T : |v| = K \& \boldsymbol{\mu}_v \in B\}) \ge \langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho}|\boldsymbol{\nu}) - \varepsilon > 0 \quad (2.8)$$

The first generation of that Galton-Watson tree  $T^{(K)}$  is given by  $\{|v| = K : \mu_v \in B\}$ , the second generation by the descendant of those with an empirical

offspring distribution between times K and 2K that belong to G, and so on. By (2.8), this Galton-Watson tree is supercritical, with

$$\liminf_{p \to \infty} \frac{1}{pK} \# \{ v \in T^{(K)} : |v| = p \} \ge \langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho} | \boldsymbol{\nu}) - \varepsilon,$$

a.s. on the survival of  $T^{(K)}$ . Moreover, by convexity of B, all individuals  $v \in T^{(K)}$  verify that  $\mu_v \in B$ . Additionally, for all p > 0 large enough, we have  $\mu_{v_p} \in G$ . We conclude that

$$\mathbb{P}_0\left(\liminf_{n\to\infty}\frac{1}{n}\log\#\{v\in T: |v|=n\,\&\,\boldsymbol{\mu}_v\in G\}\geq \langle\boldsymbol{\varrho},\ln\rangle-H(\boldsymbol{\varrho}|\boldsymbol{\nu})\right)>0.$$

We then use the 0-1 law for Galton-Watson processes, since our event of interest is hereditary, to conclude.

(ii) Without loss of generality again, we assume here that G is an open ball centered at  $\rho$ . The second moment method yields

$$\mathbb{P}_{0}(\exists v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G) \geq \frac{\mathbb{E}_{0} \left( \#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G\} \right)^{2}}{\mathbb{E}_{0} \left( \#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G\}^{2} \right)}.$$

By Lemma 2.9, we recall that

$$\begin{split} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}_0 \left( \# \{ v \in T : |v| = n \& \boldsymbol{\mu}_v \in G \} \right) \\ = \log m_{\boldsymbol{\nu}} + \liminf_{n \to \infty} \frac{1}{n} \log \overline{\mathbb{P}}_0(L_n \in G) \\ \ge - \inf_{\boldsymbol{\mu} \in G} (H(\boldsymbol{\mu} | \boldsymbol{\nu}) - \langle \boldsymbol{\mu}, \ln \rangle), \end{split}$$

with the same notation and computations as in the proof of Theorem 2.6(i). On the other hand, we have

$$\mathbb{E}_{0} \left( \#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G\}^{2} \right) \\= m_{\boldsymbol{\nu}}^{n} \overline{\mathbb{E}}_{0} \left( \#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G\} \mathbf{1}_{\{L_{n} \in G\}} \right) \\= m_{\boldsymbol{\nu}}^{n} \sum_{k=0}^{n-1} \overline{\mathbb{E}}_{0} \left( \#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \in G, |v \wedge U_{n}| = k\} \mathbf{1}_{\{L_{n} \in G\}} \right) \\+ m_{\boldsymbol{\nu}}^{n} \mathbb{P}_{0} (L_{n} \in G),$$

where  $|v \wedge U_n|$  represents the time at which v splits from the harmonic line  $U_n$ . We write  $M = \sup S$  for the largest possible number of children of an individual, and remark that for all  $k \leq n$ , we have

$$\overline{\mathbb{E}}_0\left(\#\{v\in T: |v|=n\,\&\,\boldsymbol{\mu}_v\in G, |v\wedge U_n|=k\}\mathbf{1}_{\{L_n\in G\}}\right) \le Mm_{\boldsymbol{\nu}}^{n-k}\overline{\mathbb{P}}_0(L_n\in G).$$

Therefore

$$\mathbb{E}_0\left(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in G\}^2\right) \le (n+1)Mm_{\boldsymbol{\nu}}^n \mathbb{P}_0(L_n \in G).$$

As a result, another call to Sanov's large deviations theorem yields

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_0 \left( \# \{ v \in T : |v| = n \& \boldsymbol{\mu}_v \in G \}^2 \right) \le -\inf_{\boldsymbol{\mu} \in \overline{G}} (H(\boldsymbol{\mu}|\boldsymbol{\nu}) - \langle \boldsymbol{\mu}, \ln \rangle).$$

We conclude by continuity of  $\boldsymbol{\mu} \mapsto H(\boldsymbol{\mu}|\boldsymbol{\nu}) - \langle \boldsymbol{\mu}, \ln \rangle$  that

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_0(\exists v \in T : |v| = n \& \boldsymbol{\mu}_v \in G) \ge -\inf_{\boldsymbol{\mu} \in G} H(\boldsymbol{\mu}|\boldsymbol{\nu}) - \langle \boldsymbol{\mu}, \ln \rangle$$
$$\ge \langle \boldsymbol{\varrho}, \ln \rangle - H(\boldsymbol{\varrho}|\boldsymbol{\nu}). \qquad \Box$$

#### 2.4 Weak persistence in a two-type Galton-Watson tree

The purpose of this section simply to exhibit natural examples of random trees in which there exist weakly but not strongly persistent laws for the empirical offspring distributions. We stress that we do not aim to obtain either a general result or optimal conditions.

Let  $\nu$ ,  $\nu'$  be two probability distributions on  $\mathbb{Z}_+$  with finite support and such that

$$m_{\nu} = \sum k\nu(k) > 1 > m_{\nu'} = \sum k\nu'(k).$$
 (2.9)

So a  $\nu$ -Galton-Watson tree is infinite with positive probability, while a  $\nu'$ -Galton-Watson tree is a.s. finite. We consider a two-type Galton-Watson process, where each individual has a type in  $\{1, 2\}$ , and given its type, reproduces independently from the rest of the population. We suppose that individuals of type 1 reproduce by creating a random number of children of type 1 with law  $\nu$ , as well as one child of type 2. Particles of type 2 only create children of type 2, making a random number with law  $\nu'$ .

Remarking that the total number of children of each individual of type 1 is given by k + 1 for some k such that  $\nu(k) > 0$ , we write

$$\overline{S} = \{k \ge 2 : \nu(k-1) > 0\} \cup \{k \ge 1 : \nu'(k) > 0\},\$$

with S the support of  $\nu$ . We also define the operator  $\tau$  associating to a probability measure  $\boldsymbol{\varrho}$  its pushforward by  $k \mapsto k-1$ . We prove in this section the following result.

**Proposition 2.11.** With the two-types tree T defined above, a law  $\boldsymbol{\varrho} \in \mathcal{P}_{\overline{S}}$  is

- (i) strongly persistent if  $\langle \tau \boldsymbol{\varrho}, \ln \rangle > H(\tau \boldsymbol{\varrho} | \boldsymbol{\nu});$
- (ii) weakly persistent if there exists  $s \in (0,1)$  and  $\mu, \mu'$  with

$$\boldsymbol{\varrho} = s\boldsymbol{\mu} + (1-s)\boldsymbol{\mu}',$$

such that the following two inequalities hold:

$$\langle \tau \boldsymbol{\mu}, \ln \rangle > H(\tau \boldsymbol{\mu} | \boldsymbol{\nu})$$

and

$$s\langle \tau \boldsymbol{\mu}, \ln \rangle + (1-s)\langle \boldsymbol{\mu}', \ln \rangle > sH(\tau \boldsymbol{\mu}|\boldsymbol{\nu}) + (1-s)H(\boldsymbol{\mu}'|\boldsymbol{\nu}').$$

Using this proposition, one can construct a variety of random trees with weakly –but not strongly– persistent offspring distributions. For example, if there exists  $k \in \mathbb{N}$  such that  $\nu'(k) > 0$  but  $\nu(k-1) = 0$ , then any individual v such that  $\mu_v(k) > 0$  must be of type 2. However, by (2.9), almost surely there is no infinite line of individuals of type 2. Consequently, any law  $\boldsymbol{\varrho}$  such that  $\varrho(k) > 0$  cannot be strongly persistent.

Proof of Proposition 2.11. (i) This is an immediate consequence of Theorem 2.6, using that the subtree  $T^{(1)}$  of T, consisting of only individuals of type 1 is a Galton-Watson tree with reproduction law  $\boldsymbol{\nu}$ , and the empirical offspring distribution of  $v \in T^{(1)}$  verifies  $\boldsymbol{\mu}_v^{(1)} = \tau \boldsymbol{\mu}_v$ , with  $\boldsymbol{\mu}_v^{(1)}$  representing the empirical offspring distribution of v as an element of  $T^{(1)}$ .

(ii) Let  $\boldsymbol{\varrho}, \boldsymbol{\mu}, \boldsymbol{\mu}'$  and s as in the statement. By continuity of  $H(\cdot|\boldsymbol{\nu})$  and  $H(\cdot|\boldsymbol{\nu}')$ , we observe that there exist open neighborhoods O, G, G' of  $\boldsymbol{\varrho}, \boldsymbol{\mu}, \boldsymbol{\mu}'$  respectively, such that for all  $(\boldsymbol{\xi}, \boldsymbol{\xi}') \in G \times G', s\boldsymbol{\xi} + (1-s)\boldsymbol{\xi}' \in O$ . We fix  $\varepsilon > 0$  such that

$$\langle \tau \boldsymbol{\mu}, \ln \rangle - H(\tau \boldsymbol{\mu} | \boldsymbol{\nu}) - \varepsilon > 0$$
  
and  $s \langle \tau \boldsymbol{\mu}, \ln \rangle + (1-s) \langle \boldsymbol{\mu}', \ln \rangle - (sH(\tau \boldsymbol{\mu} | \boldsymbol{\nu}) + (1-s)H(\boldsymbol{\mu}' | \boldsymbol{\nu}')) - 2\varepsilon > 0.$ 

By Lemma 2.10(i), we observe that almost surely on the survival event, for all n large enough, there will be at least

$$\exp\left(sn\left(\langle \tau\boldsymbol{\mu},\ln\rangle-H(\tau\boldsymbol{\mu}|\boldsymbol{\nu})-\varepsilon\right)\right)$$

individuals v of type 1 at generation  $\lfloor ns \rfloor$  with empirical offspring distribution  $\mu_v \in G$ . In addition, by Lemma 2.10(ii), each of these individuals will give birth to one individual of type 2, that has probability at least

$$\exp\left((1-s)n\left(\langle \boldsymbol{\mu}', \ln \rangle - H(\boldsymbol{\mu}'|\boldsymbol{\nu}') - \varepsilon\right)\right)$$

to start a Galton-Watson tree of reproduction law  $\nu'$  creating at least one descendant at generation  $n - \lfloor ns \rfloor - 1$  with an empirical offspring distribution in G'. Then, using Borel-Cantelli lemma, we conclude that almost surely on the survival event, for all n large enough, there will be an individual |v| = n in Twith an empirical offspring distribution in O, completing the proof.

# 3 Reinforced sampling

We return here to the setting of Section 2.2, where the notation  $\mathbb{P}_0$  was used for a probability measure under which the  $\xi_i$  are i.i.d. with law  $\nu$ . The dynamics of reinforced Galton–Watson processes that have been depicted in the Introduction incite us to define another probability measure denoted by  $\mathbb{P}_q$ , where  $q \in (0, 1)$ is the memory parameter, under which the sequence  $(\xi_i)$  rather describes balls added successively to an urn in a Pólya type process.

Namely, we view S as a set of colors and imagine an urn containing colored balls. In some probability space  $(\Omega, \mathcal{A}, \mathbb{P}_q)$ , we pick a ball uniformly at random in the urn, observe its color and return it to the urn together with a new ball. The color of the new ball is the same as that just sampled with probability q, and with complementary probability 1 - q, its color is rather chosen randomly according to  $\boldsymbol{\nu}$ . We then iterate independently, supposing for definitiveness that at time n = 1, we add to an initially empty urn a single ball with random color  $\xi_1$  distributed according to  $\boldsymbol{\nu}$ . We write  $\xi_n$  for the color of the *n*-th ball added to the urn. It should be plain that under  $\mathbb{P}_q$ , each variable  $\xi_i$  has again the law  $\boldsymbol{\nu}$ ; however the sequence with memory  $(\xi_i)_{i\geq 1}$  is not i.i.d., nor Markovian, nor even stationary.

Just as in the previous section, we denote by  $(L_n)_{n\geq 1}$  the sequence of empirical measures induced by the sequence of colors  $(\xi_i)_{i\geq 1}$ . Perhaps the most

basic result in the reinforced setting is that the Law of Large Numbers remains valid, namely,  $L_n$  converges to  $\boldsymbol{\nu}$  as  $n \to \infty$ ,  $\mathbb{P}_q$ -a.s. This is indeed a special case of a much more general result for random urn schemes, see e.g. [1] or [17]. The main purpose of this section is to investigate the effect of reinforcement on Sanov's Theorem and some of its consequences. On our way, we briefly study a function that has a crucial role in the analysis.

#### 3.1 Reinforced Sanov's theorem

Budhiraja and Waterbury [12] have established the LDP for a more general family of reinforced chains on a finite alphabet, in which the rate function is given in the form of an optimization problem. In the present setting, their main result can be stated as follows.

**Theorem 3.1** (After Budhiraja and Waterbury [12]). The sequence of empirical measures  $(L_n)$  satisfies the LDP under  $\mathbb{P}_q$  with rate function  $I_q : \mathcal{P}_S \to \mathbb{R}_+$  given by

$$I_q(\boldsymbol{\varrho}) = \inf_{\boldsymbol{\eta} \in \mathcal{U}(\boldsymbol{\varrho})} \int_0^1 H(\boldsymbol{\eta}_s | q \boldsymbol{\psi}_s + (1-q)\boldsymbol{\nu}) \mathrm{d}s,$$

where for all  $t \in (0, 1]$ 

$$\boldsymbol{\psi}_t = \frac{1}{t} \int_0^t \boldsymbol{\eta}_s \mathrm{d}s, \qquad (3.1)$$

and  $\mathcal{U}(\boldsymbol{\varrho})$  denotes the collection of controls  $\boldsymbol{\eta} = (\boldsymbol{\eta}_s)_{s \in [0,1]}$  in  $\mathcal{P}_S$  such that  $\boldsymbol{\psi}_1 = \int_0^1 \boldsymbol{\eta}_s \mathrm{d}s = \boldsymbol{\varrho}.$ 

*Proof.* Specializing [12, Theorem 2] gives the LDP for  $(L_n)$  with rate function

$$I_q(\boldsymbol{\varrho}) \coloneqq \inf_{\boldsymbol{\eta} \in \mathcal{V}(\boldsymbol{\varrho})} \int_0^\infty e^{-s} H(\boldsymbol{\eta}_s | q \boldsymbol{\phi}_s + (1-q)\boldsymbol{\nu}) ds,$$

where

$$\boldsymbol{\phi}_t = \boldsymbol{\varrho} - \int_0^t \boldsymbol{\eta}_s \mathrm{d}s + \int_0^t \boldsymbol{\phi}_s \mathrm{d}s, \qquad (3.2)$$

and  $\mathcal{V}(\boldsymbol{\varrho})$  denotes the collection of controls  $\boldsymbol{\eta} = (\boldsymbol{\eta}_s)_{s\geq 0}$  in  $\mathcal{P}_S$  such that  $\boldsymbol{\phi}_t \in \mathcal{P}_S$  for all  $t \geq 0$ . We then perform the simple change of variables  $s \mapsto 1 - e^{-s}$  and get

$$\begin{split} &\int_0^\infty e^{-s} H(\boldsymbol{\eta}_s | q \boldsymbol{\phi}_s + (1-q) \boldsymbol{\nu}) ds \\ &= \int_0^1 H(\boldsymbol{\eta}_{-\log(1-s)} | q \boldsymbol{\phi}_{-\log(1-s)} + (1-q) \boldsymbol{\nu}) ds \end{split}$$

On the other hand, we remark that if  $\eta \in \mathcal{V}(\varrho)$ , then the solution to (3.2) is

$$\boldsymbol{\phi}_{-\log(1-t)} = \frac{1}{1-t} \left( \boldsymbol{\varrho} - \int_0^t \boldsymbol{\eta}_{-\log(1-s)} \mathrm{d}s \right) = \frac{1}{1-t} \int_t^1 \boldsymbol{\eta}_{-\log(1-s)} \mathrm{d}s,$$

where for the second equality, we used that

$$\lim_{t\to\infty} \mathrm{e}^{-t} \boldsymbol{\phi}_t = 0 = \boldsymbol{\varrho} - \int_0^\infty e^{-s} \boldsymbol{\eta}_s \mathrm{d}s.$$

Then, writing  $\overline{\eta}_t = \eta_{-\log t}$  and  $\psi_t = \phi_{-\log t}$ , we obtain the alternative formula for  $I_q$  of the statement, remarking that  $\mathcal{P}(S)$  is convex, therefore for any possible control,  $\frac{1}{t} \int_0^t \overline{\eta}_s ds \in \mathcal{P}(S)$ .

Solving explicitly the optimization problem in Theorem 3.1 to get an explicit expression for the rate function  $I_q$  does not seem easy, and the purpose of the next section is to offer an alternative characterization which will be much simpler to analyze. Before this, we point at the following upper-bound which may be quite sharp, at least in simple cases, as Example 3.4 below suggests.

**Corollary 3.2.** For every  $\rho \in \mathcal{P}_S$ , there is the upper-bound

$$I_q(\boldsymbol{\varrho}) \le H(\boldsymbol{\varrho}|q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}). \tag{3.3}$$

Moreover, if  $\boldsymbol{\varrho}$  is equivalent to  $\boldsymbol{\nu}$ , that is,  $\varrho(k) > 0$  for all  $k \in S$ , and if  $\boldsymbol{\varrho} \neq \boldsymbol{\nu}$ , then the inequality (3.3) is strict.

*Proof.* The first claim is immediate from Theorem 3.1. Indeed, for the constant control  $\eta_s \equiv \varrho$ , we have  $\psi_t \equiv \varrho$ , and (3.3) follows.

Suppose now that  $\boldsymbol{\varrho} \neq \boldsymbol{\nu}$  and that  $\boldsymbol{\varrho}(k) > 0$  for all  $k \in S$ . Set

$$\boldsymbol{\varrho}_x \coloneqq \boldsymbol{\varrho} + x(\boldsymbol{\varrho} - \boldsymbol{\nu}) \quad \text{for } x \in \mathbb{R},$$

so  $\boldsymbol{\varrho}_x \in \mathcal{P}_S$  provided that |x| is small enough. The function

$$h(x,y) \coloneqq H(\boldsymbol{\varrho}_x \mid q\boldsymbol{\varrho}_y + (1-q)\boldsymbol{\nu})$$

is  $\mathcal{C}^{\infty}$  on some neighborhood of (0,0), and we have

$$h(x,y) = h(0,0) + xg_1 + yg_2 + O(x^2 + y^2),$$
(3.4)

where  $(g_1, g_2) = \nabla h(0, 0)$ .

The key observation is that

$$g_2 < 0.$$
 (3.5)

Indeed, for y > 0 sufficiently small, we can express  $q \varrho_y + (1-q)\nu$  as a convex combination of  $q \varrho + (1-q)\nu$  and  $\varrho$ , namely

$$q\boldsymbol{\varrho}_{y} + (1-q)\boldsymbol{\nu} = \left(1 - \frac{qy}{1-q}\right)(q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}) + \frac{qy}{1-q}\boldsymbol{\varrho}.$$

By convexity of the function  $-\log$ , this yields

$$h(0,y) \le \left(1 - \frac{qy}{1-q}\right)h(0,0) + \frac{qy}{1-q}H(\boldsymbol{\varrho} \mid \boldsymbol{\varrho}) = \left(1 - \frac{qy}{1-q}\right)h(0,0),$$

and then

$$\partial_y h(0,0) \le -\frac{q}{1-q}h(0,0).$$

We have h(0,0) > 0 as  $\boldsymbol{\varrho} \neq q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}$ , and (3.5) holds.

We now return to the setting of Theorem 3.1 and consider the control  $\eta$  given by

$$\boldsymbol{\eta}_t = \begin{cases} \boldsymbol{\varrho}_{\varepsilon} & \text{if } 0 \leq t \leq 1/2, \\ \boldsymbol{\varrho}_{-\varepsilon} & \text{if } 1/2 < t \leq 1, \end{cases}$$

for some sufficiently small  $\varepsilon > 0$ . We get

$$\boldsymbol{\psi}_t = \begin{cases} \boldsymbol{\varrho}_{\varepsilon} & \text{if } 0 \leq t \leq 1/2, \\ \boldsymbol{\varrho}_{\varepsilon(1/t-1)} & \text{if } 1/2 < t \leq 1. \end{cases}$$

By the Taylor expansion (3.4), we deduce

$$\begin{split} &\int_{0}^{1} H(\boldsymbol{\eta}_{s} | q \boldsymbol{\psi}_{s} + (1-q) \boldsymbol{\nu}) \mathrm{d}s \\ &= \frac{1}{2} h(\varepsilon, \varepsilon) + \int_{1/2}^{1} h(-\varepsilon, \varepsilon (1/t-1)) \mathrm{d}t \\ &= \frac{1}{2} h(0,0) + \frac{1}{2} (g_{1} + g_{2}) \varepsilon + \frac{1}{2} h(0,0) - \frac{1}{2} g_{1} \varepsilon + g_{2} \varepsilon \int_{1/2}^{1} (1/t-1) \mathrm{d}t + O(\varepsilon^{2}) \\ &= h(0,0) + g_{2} \varepsilon \log 2 + O(\varepsilon^{2}). \end{split}$$

Since we know from (3.5) that  $g_2 < 0$ , we can choose  $\varepsilon > 0$  sufficiently small so that the integral in the right-hand side is strictly less than h(0,0).

#### 3.2 Fenchel-Legendre identification of the rate function

Using results of [4], we provide an alternative representation of the rate function  $I_q$  as the Fenchel-Lengendre transform of a convex function  $\Lambda_q : \mathbb{R}^S \to \mathbb{R}_+$ , similar to (2.3). In order to deal later on with probability measures with support strictly included in S, it is convenient to allow some coordinates of  $\Lambda_q$  to be equal to  $-\infty$ ; recall the convention (2.1). For every  $\lambda \in [-\infty, \infty)^S$ , we set

$$\Lambda_q(\lambda) \coloneqq \log q - \log \left( \int_0^\infty \prod (1 - t \mathrm{e}^{\lambda(k)})_+^{\nu(k)(1-q)/q} \,\mathrm{d}t \right), \tag{3.6}$$

where  $x_+$  denotes the positive part of a real number x and the product in the integral runs over  $k \in S$  (or, equivalently, over the k's with  $\lambda_k > -\infty$ ). Needless to say, the right-hand side is interpreted as  $-\infty$  when  $\lambda \equiv -\infty$ .

Theorem 3.3. We have:

- (i) For every  $\lambda \in [-\infty, \infty)^S$ , there is the limit  $\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_q \left( \exp(\langle L_n, n\lambda \rangle) \right) = \Lambda_q(\lambda).$
- (ii) As a consequence, the rate function  $I_q$  in Theorem 3.1 can be identified as the Fenchel-Legendre transform  $\Lambda_q^*$  of  $\Lambda_q$ ,

$$I_q(\boldsymbol{\varrho}) = \Lambda_q^*(\boldsymbol{\varrho}) \coloneqq \sup_{\lambda \in \mathbb{R}^S} \left( \langle \boldsymbol{\varrho}, \lambda \rangle - \Lambda_q(\lambda) \right), \qquad \boldsymbol{\varrho} \in \mathcal{P}_S.$$

*Proof.* (i) Essentially, this is a weaker (logarithmic) version of the main result of [4]; see also the earlier articles [15, 16] in the case #S = 2. Actually, the calculations in [4] are mostly done for the special case  $\lambda_k = \ln k$ , but the method works just as well for arbitrary  $\lambda$ . For the reader's convenience, we briefly present below the guiding line from [4], and skip explicit calculations to avoid duplication.

To start with, we set  $a_k := e^{\lambda_k}$  for  $k \in S$ , and point at the identities

$$\exp(\langle L_n, n\lambda \rangle) = \prod_{i=1}^n a_{\xi_i} = \prod a_k^{\#\{i \le n: \xi_i = k\}},$$

where the product on the right-hand side runs for  $k \in S$ . Next, introduce an independent standard Yule process  $Y = (Y(t))_{t \ge 0}$  and define for all  $t \ge 0$  and  $k \in S$ ,

$$Y_k(t) \coloneqq \#\{i \le Y(t) : \xi_i = k\}.$$

We argued in [4, Section 2] that  $\mathbf{Y}(t) = (Y_k(t))_{k \in S}$  is a multitype Yule process, with space of types S. Since Y(t) has the geometric distribution with parameter  $e^{-t}$ , this yields

$$e^{-t}\sum_{n=0}^{\infty} (1 - e^{-t})^n \mathbb{E}_q \left( \exp(\langle L_n, n\lambda \rangle) \right) = \mathbb{E}_q \left( \prod a_k^{Y_k(t)} \right);$$
(3.7)

see [4, Lemma 2.1].

The probability generating function of  $\mathbf{Y}(t)$  that appears in the right-hand side of (3.7) has been determined in [4, Sections 3 and 4] by solving a system of non-linear differential equations; see notably Proposition 4.3 and Corollary 4.4 there. In particular, we can derive the asymptotic behavior of the right-hand side of (3.7) as  $t \to \infty$ .

All what is needed now is to deduce the asymptotic behavior of the moment generating function  $\mathbb{E}_q(\exp(\langle L_n, n\lambda \rangle))$  as  $n \to \infty$  from that of (3.7) as  $t \to \infty$ . This is achieved by the singularity analysis performed in [4, Section 5], under the condition that the maximum of  $\lambda$  is reached at a single location. One gets

$$\lim_{n \to \infty} \exp(-n\Lambda_q(\lambda)) \mathbb{E}_q \left( \exp(\langle L_n, n\lambda \rangle) \right) = c$$

for some explicit constant c > 0; see [4, Theorem 1] for the special case  $\lambda_k = \ln k$ .

It only remains to consider the situation where the maximum of  $\lambda$  is reached at two or more location, say  $j_1, \ldots, j_\ell \in S$  with  $\ell \geq 2$ . But this case immediately reduces to the preceding by considering a quotient space. Specifically set  $\widetilde{S} =$  $S \setminus \{j_2, \ldots, j_\ell\}$ , and then  $\widetilde{\nu}(i) = \nu(i)$  for  $i \in \widetilde{S} \setminus \{j_1\}$  and  $\widetilde{\nu}(j_1) = \nu(j_1) + \cdots + \nu(j_\ell)$ . Then it should be plain that, in the obvious notation, the distribution of variable  $\langle L_n, n\lambda \rangle$  under  $\mathbb{P}_q$  is the same as that of  $\langle \widetilde{L}_n, n\widetilde{\lambda} \rangle$  under  $\widetilde{\mathbb{P}}_q$ , where  $\widetilde{\lambda}$  is the restriction of  $\lambda$  to  $\widetilde{S}$ . We conclude that (i) still holds in this situation.

(ii) This follows from (i) by an application of the Gärtner-Ellis Theorem [13, Theorem 2.3.6] and the uniqueness of the rate function, see e.g. [13, Lemma 4.1.4] or [11, Theorem 1.15].  $\Box$ 

The Fenchel-Legendre transform  $\Lambda_q^*$  can be evaluated numerically, but unfortunately, it seems that there is no both general and explicit expression for it (like the variational formula (2.2) in the case without memory q = 0), because the function  $\Lambda_q$  in (3.6) also cannot be computed explicitly in the first place. There are nonetheless situations where calculations are available, for instance whenever  $(1/q - 1)\nu$  is integer valued<sup>6</sup>. The example below illustrates the simplest case of all.

<sup>&</sup>lt;sup>6</sup>Then  $t \to \prod (1 - te^{\lambda(k)})^{\nu(k)(1-q)/q}$  is a polynom function in the variable t, which can thus be integrated explicitly.

*Example* 3.4. Suppose that  $S = \{1, 2\}$ , that the reproduction law is given by  $\nu(j) = 1/2$  for j = 1, 2. Suppose further that q = 1/3, so that (1 - q)/q = 2. Then  $\Lambda_q(x, y)$  is a symmetric function and for any  $x \leq y$ , we find:

$$\Lambda_{1/3}(x,y) = \log(2) + y - \log(3 - e^{x-y}).$$

Thus

$$\nabla \Lambda_{1/3}(x,y) = \left(\frac{e^{x-y}}{3 - e^{x-y}}, 1 - \frac{e^{x-y}}{3 - e^{x-y}}\right),$$

and we further get for any 0 ,

$$\Lambda_{1/3}^*(p,1-p) = \Lambda_{1/3}^*(1-p,p) = p \log\left(\frac{3p}{p+1}\right) - \log(2) + \log\left(\frac{3}{p+1}\right).$$

On the other hand, we have for  $\boldsymbol{\varrho} = (p, 1-p)$  that

$$H(\boldsymbol{\varrho}|q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}) = p\log\left(\frac{3p}{p+1}\right) + (1-p)\log\left(\frac{3(1-p)}{2-p}\right),$$

and the upper-bound (3.3) is remarkably sharp in this case; see Figure 2.



Figure 2: Graphs of  $H(\boldsymbol{\varrho}|q\boldsymbol{\varrho}+(1-q)\boldsymbol{\nu})$  (blue) and of  $\Lambda_q^*(\boldsymbol{\varrho})$  (red) for  $\boldsymbol{\varrho} = (p, 1-p)$  with  $p \in [0, 1/2]$  in Example 3.4. The function  $H(\cdot|q \cdot +(1-q)\boldsymbol{\nu})$  appears as a very good approximation to  $\Lambda_q^*$ .

Remark 3.5. For every  $\lambda \in [-\infty, \infty)^S$  with  $\overline{\lambda} \coloneqq \sup_S \lambda > -\infty$ , the product in (3.6) can be bounded from above by  $\mathbf{1}_{\{t \leq \exp(-\overline{\lambda})\}}$ . It follows that

$$\Lambda_q(\lambda) \ge \log q + \lambda$$

and then, one has for every  $\boldsymbol{\varrho} \in \mathcal{P}_S$ ,

$$\Lambda_q^*(\boldsymbol{\varrho}) \le -\log q.$$

Observe that this agrees with (3.3), since obviously  $H(\boldsymbol{\varrho}|q\boldsymbol{\varrho}+(1-q)\boldsymbol{\nu}) \leq -\log q$ .

It follows from Theorem 3.3 that  $\Lambda_q$  is a convex function on  $\mathbb{R}^S$ , a fact that however does not seem easy to verify directly. At the opposite, the next first two claims are readily checked directly from the definition (3.6), without appealing to Theorem 3.3.

Lemma 3.6. The following assertions hold:

- (i) For any  $\lambda \in [-\infty, \infty)^S$  and any  $j \in S$  with  $\lambda(j) > -\infty$ ,  $\Lambda_q$  has a strictly positive partial derivative in the j-direction at  $\lambda$ ,  $\partial_j \Lambda_q(\lambda) > 0$ . For  $\lambda$  with  $\lambda(j) = -\infty$ , we further set  $\partial_j \Lambda_q(\lambda) = 0$ .
- (ii) The gradient of  $\Lambda_q$ ,  $\nabla \Lambda_q := (\partial_j \Lambda_q)_{j \in S}$ , induces a continuous function on  $\{\lambda \in [-\infty, \infty)^S : \lambda_k > -\infty \text{ for some } k \in S\}$  with values in  $\mathcal{P}_S$ . In particular, the function  $\Lambda_q$  is  $\mathcal{C}^1$  on  $\mathbb{R}^S$ .
- (iii) For any  $\lambda \in [-\infty,\infty)^S$  with  $\lambda \not\equiv -\infty$  and  $\boldsymbol{\varrho} = \nabla \Lambda_q(\lambda)$ , one has the identity

$$\Lambda_q(\lambda) = \langle oldsymbol{arrho},\lambda
angle - \Lambda_q^*(oldsymbol{arrho}).$$

For any  $\boldsymbol{\varrho}' \in \mathcal{P}_s$  with  $\boldsymbol{\varrho}' \neq \nabla \Lambda_q(\lambda)$ , one has also

$$\Lambda_q(\lambda) > \langle \boldsymbol{\varrho}', \lambda \rangle - \Lambda_q^*(\boldsymbol{\varrho}')$$

Proof. (i-ii) Indeed, the Leibniz integral rule gives

$$\partial_j \Lambda_q(\lambda) = \nu(j)(1/q-1) \frac{\int_0^\infty \frac{t e^{\lambda(j)}}{1 - t e^{\lambda(j)}} \prod (1 - t e^{\lambda(k)})_+^{\nu(k)(1-q)/q} dt}{\int_0^\infty \prod (1 - t e^{\lambda(k)})_+^{\nu(k)(1-q)/q} dt} > 0$$

An integration by parts shows that for any  $\lambda \in [-\infty, \infty)^S$  with  $\lambda \not\equiv -\infty$ , we have

$$\int_{0}^{\infty} \sum_{j \in S} \nu(j) (1/q - 1) \frac{t e^{\lambda(j)}}{1 - t e^{\lambda(j)}} \prod (1 - t e^{\lambda(k)})_{+}^{\nu(k)(1-q)/q} dt$$
$$= \int_{0}^{\infty} \prod (1 - t e^{\lambda(k)})_{+}^{\nu(k)(1-q)/q} dt,$$

which translates into

$$\sum_{j \in S} \partial_j \Lambda_q(\lambda) = 1.$$

All the other claims are straightforward.

(iii) Theorem 3.3(i) enables us to apply [13, Lemma 2.3.9].

We deduce some basic properties of the rate function  $\Lambda_q^*$ .

Corollary 3.7. The following assertions hold:

- (i) The function  $\Lambda_q^*$  is strictly convex on  $\mathcal{P}_S$ .
- (ii) For every  $\boldsymbol{\varrho} \in \mathcal{P}_S$ , there exists  $\lambda_{\boldsymbol{\varrho}} \in [-\infty,\infty)^S$  with  $\lambda_{\boldsymbol{\varrho}}(k) > -\infty$  iff  $\boldsymbol{\varrho}(k) > 0$ , such that

$$\Lambda_q^*(\boldsymbol{\varrho}) = \langle \boldsymbol{\varrho}, \lambda_{\boldsymbol{\varrho}} \rangle - \Lambda_q(\lambda_{\boldsymbol{\varrho}}).$$

Moreover we have

$$\boldsymbol{\varrho} = \nabla \Lambda_q(\lambda_{\boldsymbol{\varrho}}).$$

*Proof.* (i) Indeed, the Fenchel-Legendre transform of a  $C^1$  function is always strictly convex; see e.g. [21, Theorem 11.13].

(ii) Consider any sequence  $(\lambda_n)$  in  $\mathbb{R}^S$  such that  $\langle \boldsymbol{\varrho}, \lambda_n \rangle - \Lambda_q(\lambda_n)$  converges to  $\Lambda_q^*(\boldsymbol{\varrho})$  as  $n \to \infty$ . Observe that  $\Lambda_q(\lambda + c) = c + \Lambda_q(\lambda)$  for any  $\lambda \in \mathbb{R}^S$  and any  $c \in \mathbb{R}$ , so there is no loss of generality in assuming that  $\overline{\lambda}_n := \max_S \lambda_n = 0$ for every *n*. Then,  $\lambda_n$  converges along some subsequence to, say,  $\lambda_{\boldsymbol{\varrho}} \in [-\infty, 0]^S$ with  $\overline{\lambda}_{\boldsymbol{\varrho}} = 0$ , and this yields the identity

$$\Lambda_q^*(\boldsymbol{\varrho}) = \langle \boldsymbol{\varrho}, \lambda_{\boldsymbol{\varrho}} \rangle - \Lambda_q(\lambda_{\boldsymbol{\varrho}}).$$

Obviously  $\lambda_{\boldsymbol{\varrho}}(k) = -\infty$  can only occur when  $\boldsymbol{\varrho}(k) = 0$ . In the converse direction, suppose that  $\lambda_{\boldsymbol{\varrho}}(j) > -\infty$  for some  $j \in S$  with  $\boldsymbol{\varrho}(j) = 0$ . Let  $\lambda' \in [-\infty, \infty)^S$ defined by  $\lambda'(k) = \lambda_{\boldsymbol{\varrho}}(k)$  if  $k \neq j$  and  $\lambda'(j) = -\infty$ . Then  $\langle \boldsymbol{\varrho}, \lambda_{\boldsymbol{\varrho}} \rangle = \langle \boldsymbol{\varrho}, \lambda' \rangle$ and  $\Lambda_q(\lambda_{\boldsymbol{\varrho}}) > \Lambda_q(\lambda')$ , which is absurd. Last, the claim that the gradient of  $\Lambda_q$ evaluated at  $\lambda_{\boldsymbol{\varrho}}$  equals  $\boldsymbol{\varrho}$  is plain, as  $\lambda_{\boldsymbol{\varrho}}$  is a critical point of  $\langle \boldsymbol{\varrho}, \cdot \rangle - \Lambda_q(\cdot)$ .  $\Box$ 

The strict convexity of the rate function  $\Lambda_q^*$  readily yields the convergence in probability of the empirical distribution conditionally on a large class of rare events, which is a reinforced version of (2.4); see also e.g [11, Theorem 1.4].

**Corollary 3.8.** Let  $\Gamma$  be a closed convex subspace of  $\mathcal{P}_S$  with a non-empty interior. Then there is the convergence in probability

$$\lim_{q \to \infty} \mathbb{P}_q(\operatorname{dist}_{\mathcal{P}_S}(L_n, \gamma_q) < \varepsilon \mid L_n \in \Gamma) = 1, \quad \text{for all } \varepsilon > 0,$$

where  $\gamma_q = \arg \min_{\Gamma} \Lambda_q^*$  is the unique location on  $\Gamma$  at which  $\Lambda_q^*$  attains its minimum.

Remark 3.9. As a consequence of this result and Corollary 3.2, we observe that contrarily to the usual Sanov's theorem, the path followed by the process of empirical measures, conditioned on being close to  $\boldsymbol{\varrho}$  at a large time n, is not nearly constant. Heuristically, this is due to the fact that it is less costly to influence empirical measures at the start of the process than thereafter. Informally, a thrifty strategy is thus to drive L during a relatively short time  $\lfloor \delta n \rfloor$  towards a state from which the probability of reaching a neighborhood of  $\boldsymbol{\varrho}$  at time n is high.

## 4 Reinforced Galton–Watson processes

We now state the main results of this paper, which should be viewed as a reinforced version of Theorem 2.6. Recall the notation (1.3), as well as that in Theorems 3.1 and 3.3, and also Lemma 3.6. Last, note that for q = 0, the probability measure  $\nabla \Lambda_0(\ln)$  coincides with the size-biased version of the reproduction law  $\overline{\nu}_0$ .

**Theorem 4.1.** The following claims hold for reinforced Galton–Watson trees with memory parameter 0 < q < 1:

(i) Exponential concentration: Define

$$\overline{\boldsymbol{\nu}}_q \coloneqq \nabla \Lambda_q(\ln).$$

Then  $\overline{\nu}_q \in \mathcal{P}_S$  and for every neighborhood G of  $\overline{\nu}_q$ , there exists  $\varepsilon = \varepsilon(G) > 0$  such that for all  $n \ge 1$ ,

 $\mathbb{E}_{q}(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_{v} \notin G\})) \leq e^{-\varepsilon n} \mathbb{E}_{q}(\#\{v \in T : |v| = n\}).$ 

(ii) **Evanescent laws:** Any  $\boldsymbol{\varrho} \in \mathcal{P}_S$  with

$$\langle \boldsymbol{\varrho}, \ln \rangle < \Lambda_a^*(\boldsymbol{\varrho})$$

is evanescent,  $\mathbb{P}_q$ -a.s.

(iii) Strongly persistent laws: For any  $\rho \in \mathcal{P}_S$  with

$$\langle \boldsymbol{\varrho}, \ln \rangle > H(\boldsymbol{\varrho} | q \boldsymbol{\varrho} + (1-q) \boldsymbol{\nu}),$$

the probability under  $\mathbb{P}_q$  that  $\varrho$  is strongly persistent is strictly positive.

Example 4.2. As in Example 3.4, consider the reproduction law  $\boldsymbol{\nu} = \frac{1}{2}\boldsymbol{\delta}_1 + \frac{1}{2}\boldsymbol{\delta}_2$ . By Theorem 2.6(i), the empirical offspring distributions of vertices at generation  $n \gg 1$  concentrated around the size-biased law  $\boldsymbol{\overline{\nu}}_0 = \frac{1}{3}\boldsymbol{\delta}_1 + \frac{2}{3}\boldsymbol{\delta}_2$  in the Galton–Watson case, whereas for the reinforced version with memory parameter q = 1/3, Theorem 4.1(i) shows that they rather concentrate around  $\frac{1}{5}\boldsymbol{\delta}_1 + \frac{4}{5}\boldsymbol{\delta}_2$ .

*Proof of (i).* The proof relies on previous estimates obtained in [4]. We shall use these estimates, together with the many-to-one lemma, to compute the average number of individuals at time n with a given empirical distribution.

We know from [4, Theorem 1.1] that

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_q(\#\{v \in T : |v| = n\}) = \Lambda_q(\ln).$$

$$(4.1)$$

Next, the many-to-one formula of Lemma 2.4 entails that for any  $n \ge 1$  and any measurable  $B \subset \mathcal{P}_S$ , there is the identity

$$\mathbb{E}_q(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in B\}) = \mathbb{E}_q\left(\exp(n\langle \boldsymbol{\mu}_{U_n}, \ln\rangle) \mathbf{1}_{\boldsymbol{\mu}_{U_n} \in B}\right).$$

The construction of the out-degree sequence  $(d(U_n))$  along a harmonic line under  $\mathbb{P}_q$  mirrors that of a reinforced sequence  $(\xi_n)$  until the first appearance of the value 0 (which never occurs a.s. if and only if  $\nu(0) = 0$ , i.e. if and only if there are no empty offsprings). More precisely the former has the same distribution as the latter stopped at the first index n at which  $\xi_n = 0$ . Using that  $\exp(n\langle L_n, \ln \rangle) = 0$  if 0 appears in the sequence, in terms of the empirical measure, this yields the equality

$$\mathbb{E}_q(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in B\}) = \mathbb{E}_q(\exp(n\langle L_n, \ln\rangle) \mathbf{1}_{L_n \in B}).$$

On the other hand, we know from Lemma 3.6(iii) that

$$\Lambda_q(\ln) = \langle \overline{\boldsymbol{\nu}}_q, \ln \rangle - \Lambda_q^*(\overline{\boldsymbol{\nu}}_q)$$

and that

$$\Lambda_q(\ln) > \langle \boldsymbol{\varrho}, \ln \rangle - \Lambda_q^*(\boldsymbol{\varrho}) \qquad \text{for all } \boldsymbol{\varrho} \in \mathcal{P}_S, \ \boldsymbol{\varrho} \neq \overline{\boldsymbol{\nu}}_q.$$

By the Laplace principle, we deduce from Theorem 3.1 and Theorem 3.3(ii) that for any closed  $F \in \mathcal{P}_S$  avoiding some neighborhood of  $\overline{\nu}_q$ ,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_q \left( \exp(n \langle L_n, \ln \rangle \mathbf{1}_{L_n \in F}) \le \max_{\boldsymbol{\varrho} \in F} \left( \langle \boldsymbol{\varrho}, \ln \rangle - \Lambda_q^*(\boldsymbol{\varrho}) \right) < \Lambda_q(\ln).$$

Recalling (4.1), this established the exponential concentration.

The argument for (ii) is similar, and actually a bit simpler.

*Proof of (ii).* Assume  $\langle \boldsymbol{\varrho}, \ln \rangle < \Lambda_q^*(\boldsymbol{\varrho})$ ; by the lower semi-continuity of  $\Lambda_q^*$ , we can find a closed neighborhood F of  $\boldsymbol{\varrho}$  such that

$$\sup_{\boldsymbol{\varrho}'\in F} \left(\langle \boldsymbol{\varrho}', \ln\rangle - \Lambda_q^*(\boldsymbol{\varrho}')\right) = -2c < 0.$$

Then, just as in (i), we have for all n sufficiently large

$$\mathbb{E}_q(\#\{v \in T : |v| = n \& \boldsymbol{\mu}_v \in F\}) \le e^{-cn},$$

and a summation over all generations enables us to conclude that

$$\mathbb{E}_q(\#\{v \in T : \boldsymbol{\mu}_v \in F\}) < \infty.$$

In particular,  $\boldsymbol{\varrho}$  is evanescent  $\mathbb{P}_q$ -a.s.

The proof of (iii) relies on the next technical estimate, which can be seen as an extension of results obtained in [5, Section 5].

**Lemma 4.3.** Consider a function  $a: S \rightarrow [0, 1/q)$  such that

$$\sum \frac{\nu(j)}{1 - qa(j)} = 1/(1 - q),$$

and then define  $\pi_a \in \mathcal{P}_S$  by

$$\pi_a(k) \coloneqq \frac{(1-q)a(k)}{1-qa(k)}\nu(k), \qquad k \in S.$$

If

$$\sum \pi_a(k) \log\left(\frac{a(k)}{k}\right) < 0,$$

then the probability under  $\mathbb{P}_q$  that  $\pi_a$  is strongly persistent is strictly positive.

Taking Lemma 4.3 for granted, we can easily establish the last part of Theorem 4.1.

Proof of (iii). Let  $\boldsymbol{\varrho} \in \mathcal{P}_S$ ; define

$$a(k) \coloneqq \frac{\varrho(k)}{q\varrho(k) + (1-q)\nu(k)}, \qquad k \in S.$$

 $\mathbf{So}$ 

$$H(\boldsymbol{\varrho}|q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}) = \sum \varrho(k)\log a(k)$$

and we have also  $\rho = \pi_a$ . Therefore the condition of the statement

$$\langle \boldsymbol{\varrho}, \ln \rangle > H(\boldsymbol{\varrho} | q \boldsymbol{\varrho} + (1-q) \boldsymbol{\nu})$$

reads

$$\sum \pi_a(k) \log\left(\frac{a(k)}{k}\right) < 0$$

This enables us to invoke Lemma 4.3.

r

The proof of Lemma 4.3 below may look indirect, however the guiding line is close to that of Theorem 2.6(iii) in the case q = 0 without reinforcement. In short, under a new probability measure  $\mathbb{P}_q^a$ , we will construct a spine such that the empirical offspring distribution along the spine approaches  $\pi_a$  as the generation goes to infinity. Then, we will have to check that the distribution of the tree under  $\mathbb{P}_q^a$  is absolutely continuous with respect to the initial law  $\mathbb{P}_q$ . This is the role of the inequality in the statement. This argument is similar to that used in [5, Section 5]. To minimize the overlap with our previous work, we shall just sketch here the proof of Lemma 4.3, focussing on changes needed in key calculations. Of course, reader wishing to follow details may want to have [5] at hand as well.

Sketch of the proof of Lemma 4.3. Let a be as in the statement. We first introduce

$$n_a(\boldsymbol{\varrho}) \coloneqq \sum_{k \in S} a(k)(q \varrho(k) + (1-q)\nu(k)), \quad \boldsymbol{\varrho} \in \mathcal{P}_S.$$

This quantity should be interpreted as the conditional mean value of a(d(v)) for the reinforced Galton–Watson process, given that v is a vertex with empirical offspring distribution  $\mu_v = \rho$ . We further define, for any vertex v of the tree aside the root,

$$\Phi^{a}(v) \coloneqq \prod_{j=0}^{|v|-1} \frac{a(d(v_{j}))/d(v_{j})}{m_{a}(\boldsymbol{\mu}_{v_{j}})},$$
(4.2)

where, as previously,  $v_j$  denotes the forebear of v at generation j, and we agree for definitiveness that for j = 0,  $m_a(\boldsymbol{\mu}_{\varnothing}) = \sum a(k)\nu(k)$ . It is immediate that the process

$$M_n^a = \sum_{|v|=n} \Phi^a(v), \qquad n \ge 1$$

$$(4.3)$$

is a nonnegative martingale with unit mean under  $\mathbb{P}_q$ .

Much as for (2.7), this martingale serves to define a probability measure  $\mathbb{P}_q^a$  for the joint law of a random tree T together with a spine  $(V_n^a)_{n\geq 0}$ . Specifically, for any  $n \geq 1$ , any subtree  $t_n$  of the Ulam tree with height n and any v vertex of  $t_n$  at height n, if we write  $T_n$  for the subtree obtained by pruning T at generation n, then we set

$$\mathbb{P}_q^a \left( T_n = t_n, V_n^a = v \right) \coloneqq \mathbb{P}_q \left( T_n = t_n \right) \Phi^a(v). \tag{4.4}$$

The next key step of the proof lies in the observation that the evolution of the degrees  $(d(V_n^a))$  along the spine can be described as a generalized Pólya urn, whose scheme however differs from that in Section 3. See [5, Section 6.2].

Specifically, think again of S as a set of colors; we also add a special color denoted by  $\star$ . We define an urn process with balls having colors in  $S \cup \{\star\}$ as follows. Let the urn contain two balls at the initial time n = 0, one with a random color sampled according to  $\nu$  and one with color  $\star$ . Imagine that a ball with color  $k \in S$  has activity qa(k), meaning that the probability that it is picked at some random drawing from the urn is proportional to qa(k), whereas a ball with color  $\star$  has activity  $(1-q) \sum a(j)\nu(j)$ . At each step, a ball is drawn at random in the urn with probability proportional to its activity, its color is observed and the ball is then replaced in the urn. If the color of the sampled ball is  $j \in S$ , then we add to the urn one ball with color j and one ball with color  $\star$  and a second ball with random color in S, such that probability of adding the color k is proportional to  $a(k)\nu(k)$ . Then the sequence of the successive colors of the non- $\star$  balls added to the urn step after step, has the same law as the sequence  $(d(V_n^n))$  of out-degrees under  $\mathbb{P}_q^a$ .

The upshot of the description above is that it enables us to determine the asymptotic behavior of  $(d(V_n^a))$  using classical results on generalized Pólya urns due to Athreya and Karlin [1], see also Janson [17]. This requires the computation of the principal spectral elements of the mean replacement matrix of the urn re-weighted by activities. Namely, introduce  $A = (A_{i,j})_{i,j \in S \cup \{\star\}}$  defined for  $i, j \in S$  by

$$A_{i,j} = qa(i)\boldsymbol{\delta}_i(j), \quad A_{i,\star} = qa(i), \quad A_{\star,j} = (1-q)a(j)\nu(j),$$

and finally,

$$A_{\star,\star} = (1-q) \sum a(k)\nu(k).$$

One readily finds from our assumption (see [5, Lemma 5.5] for similar calculations) that the leading eigenvalue of A is 1 and the left-eigenvector can then be chosen to coincide with  $\pi_a$  on S.

The first consequence of the above spectral analysis for the Pólya urn is that the empirical measure of out-degrees along the spine converges  $\mathbb{P}_q^a$ -a.s. to  $\pi_a$ as the generation goes to infinity. On the other hand, one immediately verifies that  $m_a(\pi_a) = 1$ . Thus

$$\lim_{n \to \infty} \frac{1}{n} \log \Phi^a(V_n^a) = \sum \pi_a(k) \log \left(\frac{a(k)}{k}\right) < 0, \qquad \mathbb{P}_q^a\text{-a.s.},$$

and therefore the series  $\sum_{n=1}^{\infty} \Phi^a(V_n^a)$  converges  $\mathbb{P}_q^a$ -a.s. Using the same argument as in the proof of Lemma 2.9(ii), we conclude that the martingale  $M_n^a$  is uniformly integrable under  $\mathbb{P}_q$ , and it follows that  $\pi_a$  is strongly persistent with a strictly positive probability under  $\mathbb{P}_q$ .

As an immediate application, we point out that any law on S with no atom at 0 can be made strongly persistent provided that the reinforcement is sufficiently strong.

**Corollary 4.4.** Let any  $\boldsymbol{\varrho} \in \mathcal{P}_S$  with  $\varrho(0) = 0$  and  $\varrho(1) < 1$ . We may choose the memory parameter sufficiently large such that

$$\langle \boldsymbol{\varrho}, \ln \rangle > (1-q)H(\boldsymbol{\varrho}|\boldsymbol{\nu}),$$

and then the probability under  $\mathbb{P}_q$  that  $\boldsymbol{\varrho}$  is strongly persistent is strictly positive. In particular, if

$$rac{\langle \boldsymbol{\varrho}, \ln 
angle}{H(\boldsymbol{\varrho} | \boldsymbol{\nu})} \in (1 - q, 1),$$

then  $\boldsymbol{\varrho}$  is evanescent  $\mathbb{P}_0$ -a.s. and strongly persistent with strictly positive  $\mathbb{P}_q$ -probability.

*Proof.* Indeed  $\langle \boldsymbol{\varrho}, \ln \rangle > 0$  and  $H(\boldsymbol{\varrho}|\boldsymbol{\nu}) < \infty$ , so we may choose  $q \in (0, 1)$  as in the statement. It follows from the convexity of the relative entropy that

$$(1-q)H(\boldsymbol{\varrho}|\boldsymbol{\nu}) \geq H(\boldsymbol{\varrho}|q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}),$$

and we conclude the proof of the first claim with an appeal to Theorem 4.1(iii). The second assertion follows, by Theorem 2.6(ii) and the variational formula (2.2).

The (contraposition of the) final result of this section result provides an elementary necessary condition for a law to be strongly persistent with positive probability.

**Proposition 4.5.** For any  $\boldsymbol{\varrho} \in \mathcal{P}_S$  with

$$qm_{\boldsymbol{\varrho}} + (1-q)m_{\boldsymbol{\nu}} < 1,$$

the probability under  $\mathbb{P}_q$  that  $\varrho$  being strongly persistent is zero.

*Proof.* Indeed there is then a neighborhood G of  $\rho$  such that

$$\sup_{\mathbf{q}' \in G} \left( q m_{\mathbf{q}'} + (1 - q) m_{\mathbf{\nu}} \right) < 1.$$
(4.5)

Under  $\mathbb{P}_q$ , consider any vertex v in T with  $\mu_v \in G$ , and prune its descent as soon as the empirical offspring distribution exits G. By subcriticality, (4.5) entails that the pruned descent of v is finite,  $\mathbb{P}_q$ -a.s.

Remark 4.6. Comparing Theorem 4.1(iii) and Proposition 4.5 shows that

$$\langle \boldsymbol{\varrho}, \ln \rangle \leq H(\boldsymbol{\varrho}, q\boldsymbol{\varrho} + (1-q)\boldsymbol{\nu}) \text{ as soon as } qm_{\boldsymbol{\varrho}} + (1-q)m_{\boldsymbol{\nu}} = m_{q\boldsymbol{\varrho}+(1-q)\boldsymbol{\nu}} < 1.$$

Actually, it follows immediately from Gibbs inequality that

 $m_{q\boldsymbol{\rho}+(1-q)\boldsymbol{\nu}} \leq 1 \Longrightarrow \langle \boldsymbol{\varrho}, \ln \rangle \leq H(\boldsymbol{\varrho}, q\boldsymbol{\varrho}+(1-q)\boldsymbol{\nu}).$ 

## 5 A sufficient condition for survival

We now conclude this work by pointing out that Lemma 4.3 also yields the following sufficient condition for the survival of reinforced Galton-Watson processes. Introduce the space of functions on S

$$\mathcal{A} := \left\{ a : S \to [0, 1/q) : \sum \frac{\nu(j)}{1 - qa(j)} = \frac{1}{1 - q} \right\},$$

and the function  $J: \mathcal{A} \to (-\infty, \infty]$  given by

$$J(a) = \sum \nu(k) \frac{(1-q)a(k)}{1-qa(k)} \log\left(\frac{a(k)}{k}\right).$$

We also recall that the principal branch  $W_0$  of the Lambert W function is defined for every  $x \ge -1/e$  as the unique solution  $y \ge -1$  to the equation  $ye^y = x$ . **Corollary 5.1.** The reinforced Galton–Watson process survives for ever with a strictly positive  $\mathbb{P}_q$ -probability whenever

$$\min_{A} J < 0.$$

More precisely, J always reaches its minimum at a unique location  $\underline{a} = \arg \min J$ , which is determined as the unique  $a \in \mathcal{A}$  such that there exists some C > 0 with

$$a(j) = -q^{-1}W_0(-Cj), \qquad j \in S.$$

This is improves upon [5, Theorem 1.3], in which the sufficient condition for survival is given by the stronger condition J(b) < 0, where  $b \in \mathcal{A}$  is proportional to the identity, i.e. b(k) = ck for a suitable c > 0. In the converse direction, we believe that the reinforced Galton–Watson process should become eventually extinct a.s. when  $\min_{\mathcal{A}} J = J(\underline{a}) > 0$ .

*Proof.* First, recall the notation of Lemma 4.3 and note that

$$J(a) = \sum \pi_a(k) \log\left(\frac{a(k)}{k}\right)$$

So if we assume that  $J(\underline{a}) < 0$ , then the requirement of Lemma 4.3 holds for  $\pi = \pi_{\underline{a}}$ , and the latter is strongly persistent with a strictly positive  $\mathbb{P}_{q}$ probability. A fortiori the reinforced Galton–Watson process survives with a strictly positive probability.

The proof of the second claim uses the method of Lagrange multipliers. To start with, note that if  $0 \in S$  and  $a \in \mathcal{A}$  with a(0) > 0, then  $J(a) = \infty$ . An argument of compactness and continuity shows that there exists some  $\underline{a} \in \mathcal{A}$ with  $\underline{a}(0) = 0$  whenever  $0 \in S$ , such that  $J(\underline{a}) = \inf J = \min J$ . Next, the function  $x \mapsto x \log x$  has a finite derivative at any x > 0, whereas its rightderivative at x = 0 is  $-\infty$ . It follows that if  $a \in \mathcal{A}$  has a(k) = 0 for some  $k \neq 0$ , then a is not a local minimum of J; hence  $\underline{a}(k) > 0$  for all  $k \in S \setminus \{0\}$ .

The function J is  $\mathcal{C}^1$  on  $\{a: S \setminus \{0\} \to (0, 1/q)\}$ , with a partial derivative in the k-th coordinate given by

$$\partial_k J(a) = \frac{(1-q)\nu(k)}{1-qa(k)} \left(1 + \frac{1}{1-qa(k)}\log\left(\frac{a(k)}{k}\right)\right)$$

So, if we introduce the constraint function

$$g(a) = \sum \frac{\nu(j)}{1 - qa(j)},$$

which has  $\partial_k g(a) = q\nu(k)/(1-qa(k))^2$ , then  $\mathcal{A} = \{a : g(a) = 1/(1-q)\}$  and the criticality of  $\underline{a}$  forces the vectors  $\partial_k J(\underline{a})$  and  $\partial_k g(\underline{a})$  to be proportional. This entails the existence of some constant  $c \in \mathbb{R}$  such that

$$\log(\underline{a}(k)/k) + 1 - q\underline{a}(k) = c \quad \text{for all } k \in S \setminus \{0\}.$$

We rewrite the equation above as

$$-qa(k)e^{-q\underline{a}(k)} = -kC,$$

with  $C = e^c > 0$ . Thus  $\underline{a}(k) = -q^{-1}W(-Ck)$ , with W being one of the two branches of Lambert's W function. But since  $W_{-1}(-Ck) \leq -1$ , only  $\underline{a}(j) = -q^{-1}W_0(-Cj)$  fits the requirement qa(k) < 1. Since  $W_0$  is monotone increasing, the constraint  $g(\underline{a}) = 1/(1-q)$  determines C, and we conclude that J has indeed a unique minimum at the location given in the statement.  $\Box$ 

**Funding.** The first author was partially supported by the project SNSF 200020 212115: Random processes with genealogy dependent dynamics. The second author was partially supported by the MITI interdisciplinary program 80PRIME GEx-MBB and the ANR MBAP-P (ANR-24-CE40-1833) project.

### References

- Krishna B. Athreya and Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. Ann. Math. Statist., 39:1801–1817, 1968.
- [2] Krishna B. Athreya and Peter E. Ney. <u>Branching processes</u>. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].
- [3] Romain Azaïs and Benoît Henry. Maximum likelihood estimation for spinal-structured trees. <u>Journal of Applied Probability</u>, 62(1):234–268, 2025.
- [4] Jean Bertoin and Bastien Mallein. Reinforced Galton–Watson processes I: Malthusian exponents. <u>Random Structures & Algorithms</u>, 65(2):387–410, 2024.
- [5] Jean Bertoin and Bastien Mallein. Reinforced Galton-Watson processes II: Large time behaviors. Ann. Appl. Probab., (to appear), 2025+.
- [6] J. D. Biggins. Chernoff's theorem in the branching random walk. <u>J. Appl.</u> <u>Probab.</u>, 14:630–636, 1977.
- [7] J. D. Biggins. Martingale convergence in the branching random walk. <u>J.</u> Appl. Probab., 14:25–37, 1977.
- [8] J. D. Biggins. The asymptotic shape of the branching random walk. <u>Adv.</u> Appl. Probab., 10:62–84, 1978.
- [9] J. D. Biggins and A. E. Kyprianou. Measure change in multitype branching. Adv. Appl. Probab., 36(2):544–581, 2004.
- [10] J. D. Biggins and A. Rahimzadeh Sani. Convergence results on multitype, multivariate branching random walks. <u>Adv. Appl. Probab.</u>, 37(3):681–705, 2005.
- [11] Amarjit Budhiraja and Paul Dupuis. <u>Analysis and approximation of rare events</u>. Representations and weak convergence methods, volume 94 of Probab. Theory Stoch. Model. New York, NY: Springer, 2019.
- [12] Amarjit Budhiraja and Adam Waterbury. Empirical measure large deviations for reinforced chains on finite spaces. <u>Systems & Control Letters</u>, 169:105379, 2022.
- [13] Amir Dembo and Ofer Zeitouni. <u>Large deviations techniques and applications.</u>, volume 38 of <u>Appl. Math. (N. Y.)</u>. New York, NY: Springer, 2nd ed. edition, 1998.

- [14] Rick Durrett. Probability. Theory and examples, volume 49 of Camb. Ser. Stat. Probab. Math. Cambridge: Cambridge University Press, 5th edition edition, 2019.
- [15] Simone Franchini. Large deviations for generalized Polya urns with arbitrary urn function. <u>Stochastic Processes Appl.</u>, 127(10):3372–3411, 2017.
- [16] Simone Franchini and Riccardo Balzan. Large-deviation theory of increasing returns. Phys. Rev. E, 107:064142, Jun 2023.
- [17] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. <u>Stochastic Processes Appl.</u>, 110(2):171–245, 2004.
- [18] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Stat., 37:1211–1223, 1966.
- [19] Russell Lyons. A simple path to Biggins' martingale convergence for branching random walk. In <u>Classical and modern branching processes</u>. <u>Proceedings of the IMA workshop, Minneapolis, MN, USA, June 13–17, 1994</u>, pages 217–221. New York, NY: Springer, 1997.
- [20] J. Neveu. Arbres et processus de Galton-Watson. (Trees and Galton-Watson processes). <u>Ann. Inst. Henri Poincaré, Probab. Stat.</u>, 22:199–207, 1986.
- [21] R. Tyrrell Rockafellar and Roger J.-B. Wets. <u>Variational analysis</u>, volume 317 of Grundlehren Math. Wiss. Berlin: Springer, 1998.