

# Large deviations for i.i.d. replications of the total progeny of a Galton–Watson process

Claudio Macci\*, Barbara Pacchiarotti

*Dipartimento di Matematica, Università di Roma Tor Vergata,  
Via della Ricerca Scientifica, I-00133 Rome, Italy*

[macci@mat.uniroma2.it](mailto:macci@mat.uniroma2.it) (C. Macci), [pacchiar@mat.uniroma2.it](mailto:pacchiar@mat.uniroma2.it) (B. Pacchiarotti)

Received: 27 September 2016, Revised: 15 December 2016, Accepted: 17 December 2016,  
Published online: 11 January 2017

**Abstract** The Galton–Watson process is the simplest example of a branching process. The relationship between the offspring distribution, and, when the extinction occurs almost surely, the distribution of the total progeny is well known. In this paper, we illustrate the relationship between these two distributions when we consider the large deviation rate function (provided by Cramér’s theorem) for empirical means of i.i.d. random variables. We also consider the case with a random initial population. In the final part, we present large deviation results for sequences of estimators of the offspring mean based on i.i.d. replications of total progeny.

**Keywords** Cramér’s theorem, initial random population, estimators of offspring mean  
**2010 MSC** [60F10](#), [60J80](#), [62F10](#), [62F12](#)

## 1 Introduction

There is a vast literature on branching processes. Here we cite the monographs [[1](#), [3](#), [12](#)]; moreover, we also cite the monographs [[18](#)] for the multitype case, [[10](#)], which focuses on statistical inference, and [[13](#)] and [[15](#)] for applications in biology.

The simplest example of a branching process is the Galton–Watson process. We consider the case of a population that has a unique individual at the beginning and all the individuals (of all generations) live for a unitary time; moreover, at the end of their

---

\*Corresponding author.

lifetimes, every individual of the population (of every generation) produces a random number of new individuals acting independently of all the rest, according to a specific fixed distribution. So, if we consider a sequence of random variables  $\{V_n : n \geq 0\}$  such that  $V_n$  is the population size at time  $n$  (for all  $n \geq 0$ ), we have  $V_0 = 1$  and

$$V_n := \sum_{k=1}^{V_{n-1}} X_{n,k} \quad (\text{for } n \geq 1),$$

where  $\{X_{n,i} : n, i \geq 1\}$  is a family of nonnegative integer-valued i.i.d. random variables. In other words,  $X_{n,1}, \dots, X_{n,V_{n-1}}$  represent the offspring generated at time  $n$  by each of  $V_{n-1}$  individuals that live at time  $n - 1$ . We recall some other preliminaries on the Galton–Watson process in Section 2, where, in particular, we consider a slightly different notation to allow the case with a random initial population (instead of the case with a unitary initial population cited before).

In this paper, we present large deviation results. The theory of large deviations is a collection of techniques that gives an asymptotic estimate of small probabilities in an exponential scale (see, e.g., [6] as a reference). We recall some preliminaries in Section 2. The literature on large deviations for branching processes is large. Here we essentially recall some references with results concerning the Galton–Watson process.

In several references, the large-time behavior for the supercritical case is studied, namely the case where the offspring mean  $\mu$  is strictly larger than one (in such a case, the extinction probability is strictly less than one). Here we recall [2] (see also [4] for the multitype case), [5], where the main object is the study of the tails of  $W := \lim_{n \rightarrow \infty} V_n / \mu^n$ , [19] with a careful analysis based on harmonic moments of  $\{V_n : n \geq 0\}$ , [20] (and [21]) with some conditional large deviation results based on some local limit theorems, [8] where the central role of some “lower deviation probabilities” is highlighted for the study of the asymptotic behavior of the Lotka–Nagaev estimator  $V_{n+1} / V_n$  of  $\mu$ .

Other references study the most likely paths to extinction at some time  $n_0$  when the initial population  $k$  is large. The idea is to consider the representation of a branching process with initial population equal to  $k$  as a sum of  $k$  i.i.d. replications of the process with a unitary initial population; in this case, Cramér’s theorem for empirical means of i.i.d. random variables (on  $R^{n_0}$ ) plays a crucial role. A most likely path to extinction in [16] (see also [17]) is a trajectory that minimizes the rate function among the paths that reach the level 0 at time  $n_0$ . A generalization of this concept for the most likely paths to reach a level  $b \geq 0$  can be found in [11].

In this paper, we are interested in a different direction. Namely, we are interested in the empirical means of i.i.d. replications of the total progeny of a Galton–Watson process. The total progenies of branching processes are studied in several references: here we cite the old references [7, 14, 22] for a Galton–Watson process, and [9] (see Section 2.2) among the references concerning different branching processes. The total progeny of a Galton–Watson process is an almost surely finite random variable when the extinction occurs almost surely, and therefore the supercritical case will not be considered. Some relationships between the offspring distribution and the total progeny distribution of a Galton–Watson process are well known (see (3) for the probability mass functions and (4) for the probability generating functions).

A new relationship is provided by Proposition 1, where we illustrate how the rate function for the empirical means of total progenies can be expressed in terms of the analogous rate function for the empirical means of a single progeny. This is a quite natural problem to investigate large deviations, and, as we can expect, (4) has an important role in the proof; in fact, the large deviation rate function for empirical means of i.i.d. random variables (provided by Cramér’s theorem recalled below; see Theorem 1) is given by the Legendre transform of the logarithm of the (common) moment generating function of the random variables. Moreover, the relationship provided by Proposition 1 can have interest in information theory because the involved rate functions can be expressed in terms of suitable relative entropies (or Kullback–Leibler divergences); see, for example, [23] for a discussion on the rate function expressions in terms of the relative entropy.

Another result presented in this paper is Proposition 2, that is a version of Proposition 1, where the initial population  $V_0$  is a random variable with a suitable distribution. Finally, in Propositions 3 and 4, we prove large deviation results for some estimators of the offspring mean  $\mu$  in terms of i.i.d. replications of the total progeny and of the initial population (we are considering the case where the initial population  $V_0$  is a random variable as in Proposition 2).

We conclude with the outline of the paper. We start with some preliminaries in Section 2. In Section 3, we prove the results concerning the large deviation rate functions related to Cramér’s theorem. Finally, in Section 4, we prove the large deviation results for the estimators of the offspring mean  $\mu$ .

## 2 Preliminaries

We start with some preliminaries on the Galton–Watson process. In the second part, we recall some preliminaries on large deviations.

### 2.1 Preliminaries on Galton–Watson process

Here we introduce a slightly different notation, and, moreover, we recall some preliminaries in order to define the total progeny of a Galton–Watson process.

We start with some notation concerning the offspring distribution (note that  $\mu_f$  defined further coincides with  $\mu$  in the Introduction):

- the probability mass function  $p_h := P(X_{n,i} = h)$  (for all integer  $h \geq 0$ );
- the probability generating function  $f(s) := \sum_{h \geq 0} s^h p_h$ ;
- the mean value  $\mu_f := \sum_{h \geq 0} h p_h$  (and we have  $\mu_f = f'(1)$ ).

Moreover, we introduce the analogous items for the initial population:

- the probability mass function  $\{q_r : r \geq 0\}$  (see (1));
- the probability generating function  $g(s) := \sum_{r \geq 0} s^r q_r$ ;
- the mean value  $\mu_g := \sum_{r \geq 0} r q_r$  (and we have  $\mu_g = g'(1)$ ).

So, from now on, we consider the following slightly different notation:

$$\{V_n^{f,g} : n \geq 0\}$$

(in place of  $\{V_n : n \geq 0\}$  presented before). More precisely:

- the probability generating function of  $V_0^{f,g}$  is  $g$  (so  $V_0^{f,g}$  does not depend on  $f$ ), and therefore

$$q_r := P(V_0^{f,g} = r) \quad (\text{for all integer } r \geq 0); \quad (1)$$

- for a family of i.i.d. random variables  $\{X_{n,i} : n, i \geq 1\}$  with probability generating function  $f$ , we have

$$V_n^{f,g} := \sum_{i=1}^{V_{n-1}^{f,g}} X_{n,i} \quad (\text{for all } n \geq 1).$$

**Remark 1.** Note that  $\{V_n^{f,g} : n \geq 0\}$  here corresponds to  $\{V_n : n \geq 0\}$  presented in the Introduction if  $q_1 = 1$  or, equivalently, if  $g = \text{id}$  (i.e.  $g(s) = s$  for all  $s$ ).

If we consider the extinction probability

$$p_{\text{ext}}^{f,g} := P(\{V_n^{f,g} = 0 \text{ for some } n \geq 0\}),$$

then it is known that we have

$$p_{\text{ext}}^{f,\text{id}} = \min\{s \in [0, 1] : f(s) = s\};$$

moreover, if  $p_0 > 0$ , then we have  $p_{\text{ext}}^{f,\text{id}} = 1$  if  $\mu_f \leq 1$  and  $p_{\text{ext}}^{f,\text{id}} \in (0, 1)$  if  $\mu_f > 1$ . More generally, we have

$$p_{\text{ext}}^{f,g} := q_0 + \sum_{n \geq 1} (p_{\text{ext}}^{f,\text{id}})^n q_n = g(p_{\text{ext}}^{f,\text{id}}),$$

and, if  $q_0 < 1$  (we obviously have  $p_{\text{ext}}^{f,g} = 1$  if  $q_0 = 1$ ), then we have the following cases:

$$\begin{aligned} p_{\text{ext}}^{f,g} &= g(0) = q_0 && \text{if } p_0 = 0; \\ p_{\text{ext}}^{f,g} &= g(1) = 1 && \text{if } p_0 > 0 \text{ and } \mu_f \leq 1; \\ p_{\text{ext}}^{f,g} &\in (q_0, 1) && \text{if } p_0 > 0 \text{ and } \mu_f > 1. \end{aligned}$$

Then, if  $p_0 > 0$  and  $\mu_f \leq 1$ , then the random variable  $Y^{f,g}$  defined by

$$Y^{f,g} := \sum_{i=0}^{\tau-1} V_i^{f,g}, \quad \text{where } \tau := \inf\{n \geq 0 : V_n^{f,g} = 0\},$$

is almost surely finite and provides the total progeny of  $\{V_n^{f,g} : n \geq 0\}$ . In view of what follows, we consider the probability generating function

$$\mathcal{G}_{f,g}(s) := \sum_{k \geq 0} s^k \pi_k^{f,g},$$

where  $\{\pi_k^{f,g} : k \geq 0\}$  is the probability mass function of the random variable  $Y^{f,g}$ . Moreover, we have the mean value

$$\nu^{f,g} := \sum_{k \geq 0} k \pi_k^{f,g}, \quad \text{and we have } \nu^{f,g} = \frac{\mu_g}{1 - \mu_f}; \quad (2)$$

in particular,  $v^{f,g} = \frac{\mu_g}{1-\mu_f}$  even if  $\mu_f = 1$ , namely

$$v^{f,g} = \begin{cases} \infty & \text{if } \mu_g > 0 \text{ (and } \mu_f = 1) \\ 0 & \text{if } \mu_g = 0 \text{ (and } \mu_f = 1). \end{cases}$$

Finally, we recall some well-known connections between total progeny and offspring distributions (see e.g. [7]): for the probability mass functions, we have

$$\pi_k^{f,\text{id}} = \frac{1}{k} \cdot p_{k-1}^{*k}, \quad (3)$$

where  $\{p_h^{*n} : h \geq 0\}$  is the  $n$ th power of convolution of  $\{p_h : h \geq 0\}$ ; for the probability generating functions, we have

$$\mathcal{G}_{f,\text{id}}(s) = s f(\mathcal{G}_{f,\text{id}}(s)). \quad (4)$$

## 2.2 Preliminaries on large deviations

We start with the concept of large deviation principle (LDP). A sequence of random variables  $\{W_n : n \geq 1\}$  taking values in a topological space  $\mathcal{W}$  satisfies the LDP with rate function  $I : \mathcal{W} \rightarrow [0, \infty]$  if  $I$  is a lower semicontinuous function,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in O) \geq - \inf_{w \in O} I(w) \quad \text{for all open sets } O,$$

and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in C) \leq - \inf_{w \in C} I(w) \quad \text{for all closed sets } C.$$

We also recall that a rate function  $I$  is said to be good if all its level sets  $\{\{w \in \mathcal{W} : I(w) \leq \eta\} : \eta \geq 0\}$  are compact.

**Remark 2.** If  $P(W_n \in S) = 1$  for some closed set  $S$  (at least eventually with respect to  $n$ ), then  $I(w) = \infty$  for  $w \notin S$ ; this can be checked by taking the lower bound for the open set  $O = S^c$ .

In particular, we refer to Cramér’s theorem on  $\mathbb{R}^d$  (see e.g. Theorems 2.2.3 and 2.2.30 in [6] for the cases  $d = 1$  and  $d \geq 2$ ), and we recall its statement. We remark that, in this paper, we consider the cases  $d = 1$  (in such a case, the rate function need not to be a good rate function) and  $d = 2$ . Moreover, we use the symbol  $\langle \cdot, \cdot \rangle$  for the inner product in  $\mathbb{R}^d$ .

**Theorem 1** (Cramér’s theorem). *Let  $\{W_n : n \geq 1\}$  be a sequence of i.i.d.  $\mathbb{R}^d$ -valued random variables, and let  $\{\bar{W}_n : n \geq 1\}$  be the sequence of empirical means defined by  $\bar{W}_n := \frac{1}{n} \sum_{k=1}^n W_k$  (for all  $n \geq 1$ ).*

(i) *If  $d = 1$ , then  $\{\bar{W}_n : n \geq 1\}$  satisfies the LDP with rate function  $I$  defined by*

$$I(w) := \sup_{\theta \in \mathbb{R}} \{\theta w - \log \mathbb{E}[e^{\theta W_1}]\}.$$

(ii) *If  $d \geq 2$  and the origin of  $\mathbb{R}^d$  belongs to the interior of the set  $\{\theta \in \mathbb{R}^d : \log \mathbb{E}[e^{\langle \theta, W_1 \rangle}] < \infty\}$ , then  $\{\bar{W}_n : n \geq 1\}$  satisfies the LDP with good rate function  $I$  defined by*

$$I(w) := \sup_{\theta \in \mathbb{R}^d} \{\langle \theta, w \rangle - \log \mathbb{E}[e^{\langle \theta, W_1 \rangle}]\}.$$

### 3 Applications of Cramér's theorem

The aim of this section is to prove Propositions 1 and 2. In view of this, we recall Lemmas 1 and 2, which give two immediate applications of Cramér's theorem (Theorem 1) with  $d = 1$ ; in Lemma 2, we consider the case with a unitary initial population almost surely (thus, as stated Remark 1, the case with  $q_1 = 1$  or, equivalently,  $g = \text{id}$ ).

**Lemma 1** (Cramér's theorem for offspring distribution). *Let  $\{X_n : n \geq 1\}$  be i.i.d. random variables with probability generating function  $f$ . Let  $\{\bar{X}_n : n \geq 1\}$  be the sequence of empirical means defined by  $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$  (for all  $n \geq 1$ ). Then  $\{\bar{X}_n : n \geq 1\}$  satisfies the LDP with rate function  $I_f$  defined by  $I_f(x) := \sup_{\alpha \in \mathbb{R}} \{\alpha x - \log f(e^\alpha)\}$ .*

**Lemma 2** (Cramér's theorem for total progeny distribution with  $g = \text{id}$ ). *Assume that  $p_0 > 0$  and  $\mu_f \leq 1$ . Let  $\{Y_n : n \geq 1\}$  be i.i.d. random variables with probability generating function  $\mathcal{G}_{f,\text{id}}$ . Let  $\{\bar{Y}_n : n \geq 1\}$  be the sequence of empirical means defined by  $\bar{Y}_n := \frac{1}{n} \sum_{k=1}^n Y_k$  (for all  $n \geq 1$ ). Then  $\{\bar{Y}_n : n \geq 1\}$  satisfies the LDP with rate function  $I_{\mathcal{G}_{f,\text{id}}}$  defined by  $I_{\mathcal{G}_{f,\text{id}}}(y) := \sup_{\beta \in \mathbb{R}} \{\beta y - \log \mathcal{G}_{f,\text{id}}(e^\beta)\}$ .*

Now we can prove our main results. We start with Proposition 1, which provides an expression for  $I_{\mathcal{G}_{f,\text{id}}}$  in terms of  $I_f$ .

**Proposition 1.** *Let  $I_f$  and  $I_{\mathcal{G}_{f,\text{id}}}$  be the rate functions in Lemmas 1 and 2. Then we have  $I_{\mathcal{G}_{f,\text{id}}}(y) = y I_f(\frac{y-1}{y})$  for all  $y \geq 1$ .*

**Proof.** We remark that

$$I_f(x) := \sup_{\alpha \in \mathcal{D}(f)} \{\alpha x - \log f(e^\alpha)\},$$

where  $\mathcal{D}(f) := \{\alpha \in \mathbb{R} : f(e^\alpha) < \infty\}$ , and

$$I_{\mathcal{G}_{f,\text{id}}}(x) := \sup_{\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})} \{\beta y - \log \mathcal{G}_{f,\text{id}}(e^\beta)\},$$

where  $\mathcal{D}(\mathcal{G}_{f,\text{id}}) := \{\beta \in \mathbb{R} : \mathcal{G}_{f,\text{id}}(e^\beta) < \infty\}$ , by Lemmas 1 and 2, respectively.

Moreover, the function  $\alpha : \mathcal{D}(\mathcal{G}_{f,\text{id}}) \rightarrow \mathcal{D}(f)$  defined by

$$\alpha(\beta) := \log \mathcal{G}_{f,\text{id}}(e^\beta)$$

is a bijection. This can be checked noting that  $\alpha(\beta) \in \mathcal{D}(f)$  (for all  $\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})$ ) because  $f(e^{\alpha(\beta)}) = f(\mathcal{G}_{f,\text{id}}(e^\beta)) = \frac{\mathcal{G}_{f,\text{id}}(e^\beta)}{e^\beta} < \infty$  (here we take into account (4)); moreover, its inverse  $\beta : \mathcal{D}(f) \rightarrow \mathcal{D}(\mathcal{G}_{f,\text{id}})$  is defined by

$$\beta(\alpha) := \log \mathcal{G}_{f,\text{id}}^{-1}(e^\alpha)$$

(where  $\mathcal{G}_{f,\text{id}}^{-1}$  is the inverse of  $\mathcal{G}_{f,\text{id}}$ ), and  $\beta(\alpha) \in \mathcal{D}(\mathcal{G}_{f,\text{id}})$  (for all  $\alpha \in \mathcal{D}(f)$ ) because  $\mathcal{G}_{f,\text{id}}(e^{\beta(\alpha)}) = e^\alpha < \infty$ .

Thus, we can set  $\alpha = \log \mathcal{G}_{f,\text{id}}(e^\beta)$  (for  $\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})$ ) in the expression of  $I_f(x)$ , and we get

$$I_f(x) = \sup_{\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})} \left\{ \log \mathcal{G}_{f,\text{id}}(e^\beta)x - \log f(\mathcal{G}_{f,\text{id}}(e^\beta)) \right\}.$$

Then (we take into account (4) in the second equality below)

$$\begin{aligned} I_f(x) &= \sup_{\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})} \left\{ \log \mathcal{G}_{f,\text{id}}(e^\beta)x - \log(e^{-\beta} e^\beta f(\mathcal{G}_{f,\text{id}}(e^\beta))) \right\} \\ &= \sup_{\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})} \left\{ \log \mathcal{G}_{f,\text{id}}(e^\beta)x + \beta - \log \mathcal{G}_{f,\text{id}}(e^\beta) \right\} \\ &= \sup_{\beta \in \mathcal{D}(\mathcal{G}_{f,\text{id}})} \left\{ \beta - (1-x) \log \mathcal{G}_{f,\text{id}}(e^\beta) \right\}, \end{aligned}$$

and, for  $x \in [0, 1)$ , we get

$$I_f(x) = (1-x) I_{\mathcal{G}_{f,\text{id}}}\left(\frac{1}{1-x}\right).$$

We conclude by taking  $x = \frac{y-1}{y}$  for  $y \geq 1$  (thus,  $x \in [0, 1)$ ), and we obtain the desired equality with some easy computations.  $\square$

Now we present Proposition 2, which concerns the LDP for the empirical means of i.i.d. bivariate random variables  $\{(Y_n, Z_n) : n \geq 1\}$  distributed as  $(Y^{f,g}, V_0^{f,g})$ . In particular, we obtain an expression for the rate function  $I_{\mathcal{G}_{f,g,g}}$  in terms of  $I_f$  in Lemma 1 and  $I_g$  defined by

$$I_g(z) := \sup_{\gamma \in \mathbb{R}} \left\{ \gamma z - \log g(e^\gamma) \right\}. \quad (5)$$

**Proposition 2.** *Let  $\{(Y_n, Z_n) : n \geq 1\}$  be i.i.d. random variables distributed as  $(Y^{f,g}, V_0^{f,g})$ . Assume that  $\mathbb{E}[e^{\beta Y^{f,g} + \gamma V_0^{f,g}}]$  is finite in a neighborhood of  $(\beta, \gamma) = (0, 0)$ . Let  $\{(\bar{Y}_n, \bar{Z}_n) : n \geq 1\}$  be the sequence of empirical means defined by  $(\bar{Y}_n, \bar{Z}_n) := (\frac{1}{n} \sum_{k=1}^n Y_k, \frac{1}{n} \sum_{k=1}^n Z_k)$  (for all  $n \geq 1$ ). Then  $\{(\bar{Y}_n, \bar{Z}_n) : n \geq 1\}$  satisfies the LDP with good rate function  $I_{\mathcal{G}_{f,g,g}}$  defined by*

$$I_{\mathcal{G}_{f,g,g}}(y, z) = \begin{cases} y I_f\left(\frac{y-z}{y}\right) + I_g(z) & \text{if } y \geq z > 0, \\ I_g(0) & \text{if } y = z = 0, \\ \infty & \text{otherwise.} \end{cases}$$

**Remark 3.** We are assuming (implicitly) that  $p_0 > 0$  and  $\mu_f \leq 1$ ; in fact, since we require that  $\mathbb{E}[e^{\beta Y^{f,g} + \gamma V_0^{f,g}}]$  is finite in a neighborhood of  $(\beta, \gamma) = (0, 0)$ , we are assuming that  $\mu_f < 1$  and  $\mu_g < \infty$ .

**Proof.** The LDP is a consequence of Cramér's theorem (Theorem 1) with  $d = 2$ , and the rate function  $I_{\mathcal{G}_{f,g,g}}$  is defined by

$$I_{\mathcal{G}_{f,g,g}}(y, z) := \sup_{\beta, \gamma \in \mathbb{R}} \left\{ \beta y + \gamma z - \log \mathbb{E}[e^{\beta Y^{f,g} + \gamma V_0^{f,g}}] \right\}.$$

Throughout the proof, we restrict our attention on the pairs  $(y, z)$  such that  $y \geq z \geq 0$ . In fact, almost surely, we have  $Y^{f,g} \geq V_0^{f,g} \geq 0$ , and therefore  $\bar{Y}_n \geq \bar{Z}_n \geq 0$ ; thus, by Remark 2 we have  $I_{\mathcal{G}_{f,g}}(y, z) = \infty$  if condition  $y \geq z \geq 0$  fails.

We remark that  $\mathbb{E}[s^{Y^{f,g}} | V_0^{f,g}] = (\mathcal{G}_{f,\text{id}}(s))^{V_0^{f,g}}$ , and therefore

$$\mathbb{E}[e^{\beta Y^{f,g} + \gamma V_0^{f,g}}] = \mathbb{E}[e^{\gamma V_0^{f,g}} (\mathcal{G}_{f,\text{id}}(e^\beta))^{V_0^{f,g}}] = g(e^\gamma \mathcal{G}_{f,\text{id}}(e^\beta));$$

thus,

$$I_{\mathcal{G}_{f,g}}(y, z) = \sup_{\beta, \gamma \in \mathbb{R}} \{ \beta y + \gamma z - \log g(e^\gamma \mathcal{G}_{f,\text{id}}(e^\beta)) \}.$$

Furthermore, the function

$$(\beta, \gamma) \mapsto (\beta, \gamma + \log \mathcal{G}_{f,\text{id}}(e^\beta))$$

is a bijection defined on  $\mathcal{D}(\mathcal{G}_{f,\text{id}}) \times \mathbb{R}$ , where

$$\mathcal{D}(\mathcal{G}_{f,\text{id}}) := \{ \beta \in \mathbb{R} : \mathcal{G}_{f,\text{id}}(e^\beta) < \infty \}$$

as in the proof of Proposition 1; then, for  $\delta := \gamma + \log \mathcal{G}_{f,\text{id}}(e^\beta)$ , we obtain

$$I_{\mathcal{G}_{f,g}}(y, z) = \sup_{\beta, \delta \in \mathbb{R}} \{ \beta y + (\delta - \log \mathcal{G}_{f,\text{id}}(e^\beta))z - \log g(e^\delta) \}.$$

Thus, we have (note that the last equality holds by Proposition 1)

$$\begin{aligned} I_{\mathcal{G}_{f,g}}(y, z) &\leq \sup_{\beta \in \mathbb{R}} \{ \beta y + z \log \mathcal{G}_{f,\text{id}}(e^\beta) \} + \sup_{\delta \in \mathbb{R}} \{ \delta z - \log g(e^\delta) \} \\ &= \begin{cases} z I_{\mathcal{G}_{f,\text{id}}}(y/z) + I_g(z) & \text{if } y \geq z > 0, \\ I_g(0) & \text{if } y = z = 0, \\ \infty & \text{otherwise.} \end{cases} \\ &= \begin{cases} y I_f(\frac{y-z}{y}) + I_g(z) & \text{if } y \geq z > 0, \\ I_g(0) & \text{if } y = z = 0, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

We conclude by showing the inverse inequality

$$I_{\mathcal{G}_{f,g}}(y, z) \geq \sup_{\beta \in \mathbb{R}} \{ \beta y + z \log \mathcal{G}_{f,\text{id}}(e^\beta) \} + \sup_{\delta \in \mathbb{R}} \{ \delta z - \log g(e^\delta) \}. \quad (6)$$

To this end, we take two sequences  $\{\beta_n : n \geq 1\}$  and  $\{\delta_n : n \geq 1\}$  such that

$$\lim_{n \rightarrow \infty} \beta_n y - z \log \mathcal{G}_{f,\text{id}}(e^{\beta_n}) = \sup_{\beta \in \mathbb{R}} \{ \beta y + z \log \mathcal{G}_{f,\text{id}}(e^\beta) \}$$

and

$$\lim_{n \rightarrow \infty} \delta_n z - \log g(e^{\delta_n}) = \sup_{\delta \in \mathbb{R}} \{ \delta z - \log g(e^\delta) \}.$$

Then we have

$$I_{\mathcal{G}_{f,g}}(y, z) \geq \beta_n y + (\delta_n - \log \mathcal{G}_{f,\text{id}}(e^{\beta_n}))z - \log g(e^{\delta_n}),$$

and we get (6) letting  $n$  go to infinity.  $\square$

#### 4 Large deviations for estimators of $\mu_f$

In this section, we prove two LDPs for two sequences of estimators of the offspring mean  $\mu_f$ . Namely, if  $\{(\bar{Y}_n, \bar{Z}_n) : n \geq 1\}$  is the sequence in Proposition 2 (see also the precise assumptions in Remark 3; in particular, we have  $\mu_f < 1$ ), then we consider:

1.  $\{\frac{\bar{Y}_n - \bar{Z}_n}{\bar{Y}_n} : n \geq 1\}$ ;
2.  $\{\frac{\bar{Y}_n - \mu_g}{\bar{Y}_n} : n \geq 1\}$ .

Obviously, these estimators are well defined if the denominators  $\bar{Y}_n$  are different from zero; then, in order to have well-defined estimators, we always assume that  $q_0 = 0$  (where  $q_0$  is as in (1)), and, noting that, in general,  $I_g(0) = -\log q_0$ , we have

$$I_g(0) = \infty.$$

Moreover, both sequences converge to  $\frac{v^{f,g} - \mu_g}{v^{f,g}} = \mu_f$  as  $n \rightarrow \infty$  (see  $v^{f,g}$  in (2)), and they coincide when the initial population is deterministic (equal to  $\mu_g$  almost surely).

The LDPs of these two sequences are proved in Propositions 3 and 4. Moreover, Corollary 1 and Remark 4 concern the comparison between the convergence of the first sequence  $\{\frac{\bar{Y}_n - \bar{Z}_n}{\bar{Y}_n} : n \geq 1\}$  and its analogue when the initial population is deterministic (equal to the mean). Propositions 3 and 4 are proved by combining the contraction principle (see e.g. Theorem 4.2.1 in [6]) and Proposition 2 (note that the rate function  $I_{\mathcal{G}_{f,g,g}}$  in Proposition 2 is good, as it is required to apply the contraction principle). We remark that, in the proofs of Propositions 3 and 4, we take into account that  $I_{\mathcal{G}_{f,g,g}}(0, 0) = \infty$  by Proposition 2 and  $I_g(0) = \infty$ . At the end of this section, we present some remarks on the comparison between the rate functions in Propositions 3 and 4 (Remarks 5 and 6).

We start with the LDP of the first sequence of estimators.

**Proposition 3.** *Assume the same hypotheses of Proposition 2 and  $q_0 = 0$ . Let  $\{(Y_n, Z_n) : n \geq 1\}$  be i.i.d. random variables distributed as  $(Y^{f,g}, V_0^{f,g})$ . Let  $\{(\bar{Y}_n, \bar{Z}_n) : n \geq 1\}$  be the sequence of empirical means defined by  $(\bar{Y}_n, \bar{Z}_n) := (\frac{1}{n} \sum_{k=1}^n Y_k, \frac{1}{n} \sum_{k=1}^n Z_k)$  (for all  $n \geq 1$ ). Then  $\{\frac{\bar{Y}_n - \bar{Z}_n}{\bar{Y}_n} : n \geq 1\}$  satisfies the LDP with good rate function  $J_{\mathcal{G}_{f,g,g}}$  defined by*

$$J_{\mathcal{G}_{f,g,g}}(x) := \begin{cases} -\log g(e^{-\frac{I_f(x)}{1-x}}) & \text{if } x \in [0, 1), \\ \infty & \text{otherwise.} \end{cases}$$

**Proof.** By Proposition 2 and the contraction principle we have the LDP of  $\{\frac{\bar{Y}_n - \bar{Z}_n}{\bar{Y}_n} : n \geq 1\}$  with good rate function  $J_{\mathcal{G}_{f,g,g}}$  defined by

$$J_{\mathcal{G}_{f,g,g}}(x) := \inf \left\{ I_{\mathcal{G}_{f,g,g}}(y, z) : y \geq z > 0, \frac{y-z}{y} = x \right\}.$$

The case  $x \notin [0, 1)$  is trivial because we have the infimum over the empty set. For  $x \in [0, 1)$ , we rewrite this expression as follows (where we take into account the expression of the rate function  $I_{\mathcal{G}_{f,g,g}}$  in Proposition 2):

$$\begin{aligned}
J_{\mathcal{G}_{f,g,g}}(x) &= \inf \left\{ I_{\mathcal{G}_{f,g,g}} \left( \frac{z}{1-x}, z \right) : z > 0 \right\} \\
&= \inf \left\{ \frac{z}{1-x} I_f \left( \frac{\frac{z}{1-x} - z}{\frac{z}{1-x}} \right) + I_g(z) : z > 0 \right\} \\
&= \inf \left\{ \frac{z}{1-x} I_f(x) + I_g(z) : z > 0 \right\} \\
&= - \sup \left\{ -z \frac{I_f(x)}{1-x} - I_g(z) : z > 0 \right\};
\end{aligned}$$

thus, since  $I_g(z) = \infty$  for  $z \leq 0$ , we obtain  $J_{\mathcal{G}_{f,g,g}}(x) = -\log g(e^{-\frac{I_f(x)}{1-x}})$  by taking into account the definition of  $I_g$  in (5) and the well-known properties of Legendre transforms (see e.g. Lemma 4.5.8 in [6]; see also Lemma 2.2.5(a) and Exercise 2.2.22 in [6] for the convexity and the lower semicontinuity of  $\gamma \mapsto \log g(e^\gamma)$ ).  $\square$

We have an immediate consequence of this proposition that concerns the case with a deterministic initial population equal to  $\mu_g$  (almost surely). Namely, if we consider the probability generating function  $g_\diamond$  defined by  $g_\diamond(s) := s^{\mu_g}$  (for all  $s$ ), then we mean the case  $g = g_\diamond$ , and therefore:

- $V_0^{f,g_\diamond} = \mu_g$  almost surely; thus,  $Z_n = \mu_g$  and  $\bar{Z}_n = \mu_g$  almost surely (for all  $n \geq 1$ );
- $\{Y_n^{f,g_\diamond} : n \geq 1\}$  are i.i.d. random variables distributed as  $Y^{f,g_\diamond}$ , that is,

$$Y^{f,g_\diamond} := \mu_g + \sum_{i=1}^{\tau} V_i^{f,g_\diamond}, \quad \text{where } \tau := \inf\{n \geq 0 : V_n^{f,g_\diamond} = 0\};$$

- the rate function  $J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}$  is

$$J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}(x) = \begin{cases} \mu_g \cdot \frac{I_f(x)}{1-x} & \text{if } x \in [0, 1), \\ \infty & \text{otherwise,} \end{cases} \quad (7)$$

by Proposition 3.

**Corollary 1** (Comparison between  $J_{\mathcal{G}_{f,g,g}}$  in Proposition 3 and  $J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}$ ). *We have  $J_{\mathcal{G}_{f,g,g}}(x) \leq J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}(x)$  for all  $x \in \mathbb{R}$ . Moreover the inequality turns into an equality if and only if we have one of the following cases:*

- $x \notin [0, 1)$  and  $J_{\mathcal{G}_{f,g,g}}(x) = J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}(x) = \infty$ ;
- $x = \mu_f$  and  $J_{\mathcal{G}_{f,g,g}}(x) = J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}(x) = 0$ ;
- $V_0^{f,g}$  is deterministic, equal to  $\mu_g$ , and  $J_{\mathcal{G}_{f,g,g}}(x) = J_{\mathcal{G}_{f,g_\diamond,g_\diamond}}(x)$  for all  $x \in \mathbb{R}$ .

**Proof.** The case  $x \notin [0, 1)$  is trivial. On the contrary, if  $x \in [0, 1)$ , then by Jensen's inequality we have

$$-\log g(e^{-\frac{I_f(x)}{1-x}}) = -\log \mathbb{E}[e^{-\frac{I_f(x)}{1-x} \cdot V_0^{f,g}}] \leq \mu_g \cdot \frac{I_f(x)}{1-x};$$

moreover, the cases where the inequality turns into an equality follow from the well-known properties of Jensen’s inequality.  $\square$

**Remark 4** (Comparison between convergence of estimators of  $\mu_f$ ). Assume that  $\mu_f > 0$  and the initial population is not deterministic. Then there exists  $\eta > 0$  such that

$$0 < J_{\mathcal{G}_{f,g},g}(x) < J_{\mathcal{G}_{f,g_0},g_0}(x) \quad \text{for } x \in (\mu_f - \eta, \mu_f + \eta) \setminus \{\mu_f\}. \quad (8)$$

Thus, we can say that  $\{\frac{\bar{Y}_n^{f,g_0} - \mu_g}{\bar{Y}_n^{f,g_0}} : n \geq 1\}$  converges to  $\mu_f$  (as  $n \rightarrow \infty$ ) faster than  $\{\frac{\bar{Y}_n^{f,g} - \bar{Z}_n}{\bar{Y}_n^{f,g}} : n \geq 1\}$ ; in fact, we can find  $\varepsilon > 0$  such that

$$\lim_{n \rightarrow \infty} \frac{P(|\frac{\bar{Y}_n^{f,g_0} - \mu_g}{\bar{Y}_n^{f,g_0}} - \mu_f| \geq \varepsilon)}{P(|\frac{\bar{Y}_n^{f,g} - \bar{Z}_n}{\bar{Y}_n^{f,g}} - \mu_f| \geq \varepsilon)} = 0.$$

We can repeat the same argument to say that  $\{\frac{\bar{Y}_n^{f,g_0} - \mu_g}{\bar{Y}_n^{f,g_0}} : n \geq 1\}$  converges to  $\mu_f$  (as  $n \rightarrow \infty$ ) faster than  $\{\bar{X}_n : n \geq 1\}$  in Lemma 1. In fact, we have  $V_0^{f,g_0} = \mu_g$  almost surely,  $\mu_g$  is an integer, and, since  $\mu_g > 0$  because  $q_0 = 0$ , we have  $\mu_g \geq 1$ ; then we have

$$J_{\mathcal{G}_{f,g_0},g_0}(x) = \mu_g \cdot \frac{I_f(x)}{1-x} > I_f(x) > 0 \quad \text{for all } x \in (0, 1) \setminus \{\mu_f\}$$

(we can also consider the case  $x = 0$  if  $\mu_g > 1$ ).

Now we present the LDP for the second sequence of estimators.

**Proposition 4.** Assume the same hypotheses of Proposition 2 and  $q_0 = 0$ . Let  $\{Y_n : n \geq 1\}$  be i.i.d. random variables distributed as  $Y^{f,g}$ . Let  $\{\bar{Y}_n : n \geq 1\}$  be the sequence of empirical means defined by  $\bar{Y}_n := \frac{1}{n} \sum_{k=1}^n Y_k$  (for all  $n \geq 1$ ). Then  $\{\frac{\bar{Y}_n - \mu_g}{\bar{Y}_n} : n \geq 1\}$  satisfies the LDP with good rate function  $J_{\mu_g}$  defined by

$$J_{\mu_g}(x) := \begin{cases} \inf\{\frac{\mu_g}{1-x} I_f(\frac{\mu_g - z}{1-x}) + I_g(z) : z > 0\} & \text{if } x < 1, \\ \infty & \text{if } x \geq 1. \end{cases}$$

**Proof.** By Proposition 2 and the contraction principle we have the LDP of  $\{\frac{\bar{Y}_n - \mu_g}{\bar{Y}_n} : n \geq 1\}$  with good rate function  $J_{\mu_g}$  defined by

$$J_{\mu_g}(x) := \inf\left\{ I_{\mathcal{G}_{f,g},g}(y, z) : y \geq z > 0, \frac{y - \mu_g}{y} = x \right\}.$$

The case  $x \geq 1$  is trivial because we have the infimum over the empty set (we recall that  $\mu_g > 0$  because  $q_0 = 0$ ). For  $x < 1$ , we have

$$J_{\mu_g}(x) = \inf\left\{ I_{\mathcal{G}_{f,g},g}\left(\frac{\mu_g}{1-x}, z\right) : z > 0 \right\},$$

and we obtain the desired formula by taking into account the expression of the rate function  $I_{\mathcal{G}_{f,g},g}$  in Proposition 2.  $\square$

**Remark 5** (We can have  $J_{\mu_g}(x) < \infty$  for some  $x < 0$ ). We know that, for  $J_{\mathcal{G}_{f,g},g}$  in Proposition 3, we have  $J_{\mathcal{G}_{f,g},g}(x) = \infty$  for  $x \notin [0, 1)$ . On the contrary, as we see, we could have  $J_{\mu_g}(x) < \infty$  for some  $x < 0$ . In order to explain this fact, we denote the minimum value  $r$  such that  $q_r > 0$  by  $r_{\min}$ ; then we have  $\mu_g \geq r_{\min}$ ; moreover, we have  $\mu_g > r_{\min}$  if  $q_{r_{\min}} < 1$ . In conclusion, we can say that if  $\mu_g > r_{\min}$ , then the range of negative values  $x$  such that  $J_{\mu_g}(x) < \infty$  is

$$x \geq 1 - \frac{\mu_g}{r_{\min}}; \quad (9)$$

in fact, for  $x < 1$ , both  $I_f\left(\frac{\frac{\mu_g}{1-x} - z}{\frac{\mu_g}{1-x}}\right)$  and  $I_g(z)$  are finite for  $z \in [r_{\min}, \frac{\mu_g}{1-x}]$ , and therefore we can say that  $J_{\mu_g}(x) < \infty$  if  $r_{\min} \leq \frac{\mu_g}{1-x}$  or, equivalently, if (9) holds.

**Remark 6** (Estimators of  $\mu_f$  when  $\mu_f = 0$ ). If  $\mu_f = 0$ , that is,  $f(s) = 1$  for all  $s$  or, equivalently,  $p_0 = 1$ , then the rate function in Proposition 3 is

$$J_{\mathcal{G}_{f,g},g}(x) = \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Then it is easy to check that  $J_{\mathcal{G}_{f,g},g}$  coincides with  $I_f$ , and therefore  $J_{\mathcal{G}_{f,g},g}$  coincides with  $J_{\mathcal{G}_{f,g \circ g \circ g \circ g}}$  in (7) (note that, in particular, we cannot have the strict inequalities in (8) in Remark 4 stated for the case  $\mu_f > 0$ ). Finally, if  $\mu_f = 0$  (and as usual  $q_0 = 0$  or, equivalently,  $\mu_g > 0$ ), then we have  $z = \frac{\mu_g}{1-x}$  in the variational formula of the rate function in Proposition 4, and therefore

$$J_{\mu_g}(x) = \begin{cases} I_g\left(\frac{\mu_g}{1-x}\right) & \text{if } 1 - \frac{\mu_g}{r_{\min}} \leq x < 1, \\ \infty & \text{otherwise.} \end{cases} \quad (10)$$

Note the rate function in (10) can also be derived by combining the contraction principle and the rate function  $I_g$  for the empirical means  $\{\bar{Z}_n : n \geq 1\}$ ; in fact, we have  $\{\frac{\bar{Y}_n - \mu_g}{\bar{Y}_n} : n \geq 1\} = \{\frac{\bar{Z}_n - \mu_g}{\bar{Z}_n} : n \geq 1\}$ , and the rate function  $I_g$  is good by the hypotheses of Proposition 4 (see Proposition 2 and Remark 3). Finally, we also note that inequality (9) appears in the rate function expression (10).

## Acknowledgments

The authors thank a referee for suggesting shorter proofs of Propositions 1 and 2. The support of GNAMPA (INDAM) is acknowledged.

## References

- [1] Asmussen, S., Hering, H.: Branching Processes. Birkhäuser, Boston (1983). [MR0701538](#). doi:[10.1007/978-1-4615-8155-0](#)
- [2] Athreya, K.B.: Large deviation rates for branching processes. I. Single type case. Ann. Appl. Probab. **4**, 779–790 (1994). [MR1284985](#). doi:[10.1214/aop/1177004971](#)
- [3] Athreya, K.B., Ney, P.E.: Branching Processes. Springer, New York, Heidelberg (1972). [MR0373040](#)
- [4] Athreya, K.B., Vidyashankar, A.N.: Large deviation rates for branching processes. II. The multitype case. Ann. Appl. Probab. **5**, 566–576 (1995). [MR1336883](#). doi:[10.1214/aop/1177004778](#)

- [5] Biggins, J.D., Bingham, N.H.: Large deviations in the supercritical branching process. *Adv. Appl. Probab.* **25**, 757–772 (1993). [MR1241927](#). doi:[10.2307/1427790](#)
- [6] Dembo, A., Zeitouni, O.: *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York (1998). [MR1619036](#). doi:[10.1007/978-1-4612-5320-4](#)
- [7] Dwass, M.: The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6**, 682–686 (1969). [MR0253433](#). doi:[10.1017/S0021900200026711](#)
- [8] Fleischmann, K., Wachtel, V.: Lower deviation probabilities for supercritical Galton–Watson processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **43**, 233–255 (2007). [MR2303121](#). doi:[10.1016/j.anihpb.2006.03.001](#)
- [9] González, M., Molina, M.: On the partial and total progeny of a bisexual Galton–Watson branching process. *Appl. Stoch. Models Data Anal.* **13**, 225–232 (1998). [MR1628630](#). doi:[10.1002/\(SICI\)1099-0747\(199709/12\)13:3/4<225::AID-ASM316>3.0.CO;2-9](#)
- [10] Guttorp, P.: *Statistical Inference for Branching Processes*. John Wiley and Sons, New York (1991). [MR1254434](#)
- [11] Hamza, K., Klebaner, F.C.: How did we get here? *J. Appl. Probab. A* **51**, 63–72 (2014). [MR3317350](#). doi:[10.1239/jap/1417528467](#)
- [12] Harris, T.E.: *The Theory of Branching Processes*. Springer, Berlin (1963). [MR0163361](#)
- [13] Jagers, P.: *Branching Processes with Biological Applications*. John Wiley and Sons, London, New York, Sydney (1975). [MR0488341](#)
- [14] Kennedy, D.P.: The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12**, 800–806 (1975). [MR0386042](#). doi:[10.1017/S0021900200048750](#)
- [15] Kimmel, M., Axelrod, D.E.: *Branching Processes in Biology*, 2nd edn. Springer, New York (2015). [MR3310028](#). doi:[10.1007/978-1-4939-1559-0](#)
- [16] Klebaner, F.C., Liptser, R.: Likely path to extinction in simple branching models with large initial population. *J. Appl. Math. Stoch. Anal.* **23**, 60376 (2006). [MR2221000](#). doi:[10.1155/JAMSA/2006/60376](#)
- [17] Klebaner, F.C., Liptser, R.: Large deviations analysis of extinction in branching models. *Math. Popul. Stud.* **15**, 55–69 (2008). [MR2380468](#). doi:[10.1080/08898480701792477](#)
- [18] Mode, C.J.: *Multitype Branching Processes. Theory and Applications. Modern Analytic and Computational Methods in Science and Mathematics*, vol. 34. American Elsevier Publishing Co., Inc., New York (1971). [MR0279901](#)
- [19] Ney, P.E., Vidyashankar, A.N.: Harmonic moments and large deviation rates for supercritical branching processes. *Ann. Appl. Probab.* **13**, 475–489 (2003). [MR1970272](#). doi:[10.1214/aoap/1050689589](#)
- [20] Ney, P.E., Vidyashankar, A.N.: Local limit theory and large deviations for supercritical branching processes. *Ann. Appl. Probab.* **14**, 1135–1166 (2004). [MR2071418](#). doi:[10.1214/105051604000000242](#)
- [21] Ney, P.E., Vidyashankar, A.N.: Corrections and acknowledgment for: “Local limit theory and large deviations for supercritical branching processes”. *Ann. Appl. Probab.* **16**, 2272–2272 (2006). [MR2288722](#). doi:[10.1214/105051606000000574](#)
- [22] Pakes, A.G.: Some limit theorems for the total progeny of a branching process. *Adv. Appl. Probab.* **3**, 176–192 (1971). [MR0283892](#). doi:[10.1017/S0001867800037629](#)
- [23] Varadhan, S.R.S.: Large deviations and entropy. In: Greven, A., Keller, G., Warnecke, G. (eds.): *Entropy*, pp. 199–214. Princeton University Press, Princeton (2003). [MR2035822](#)