

## ON LARGEST OFFSPRING IN A CRITICAL BRANCHING PROCESS WITH FINITE VARIANCE

JEAN BERTOIN,\* *Universität Zürich*

### Abstract

Continuing the work in Bertoin (2011) we study the distribution of the maximal number  $X_k^*$  of offspring amongst all individuals in a critical Galton–Watson process started with  $k$  ancestors, treating the case when the reproduction law has a regularly varying tail  $\bar{F}$  with index  $-\alpha$  for  $\alpha > 2$  (and, hence, finite variance). We show that  $X_k^*$  suitably normalized converges in distribution to a Fréchet law with shape parameter  $\alpha/2$ ; this contrasts sharply with the case  $1 < \alpha < 2$  when the variance is infinite. More generally, we obtain a weak limit theorem for the offspring sequence ranked in decreasing order, in terms of atoms of a certain doubly stochastic Poisson measure.

*Keywords:* Branching process; maximal offspring; extreme value theory; Cox process

2010 Mathematics Subject Classification: Primary 60F05; 60J80

### 1. Introduction and main results

Let  $\mathbf{p} = (p(n))_{n \in \mathbb{N}}$  be a probability measure on the set of nonnegative integers; we view  $\mathbf{p}$  as the law of a random variable  $X$  which gives the number of children of a typical individual. We always assume criticality, i.e.

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} np(n) = 1,$$

and implicitly exclude the degenerate case  $\mathbf{p} = \delta_1$ . We consider a Galton–Watson process having reproduction law  $\mathbf{p}$ . When the latter starts from  $k \geq 1$  ancestors, we write  $T_k$  for the size of the total population, and then, for  $i = 1, \dots, T_k$ , we denote by  $X_i$  the number of children of the  $i$ th individual, where individuals in the branching process are enumerated by some arbitrary procedure (e.g. breadth-first search). This note is concerned with weak limit theorems as  $k \rightarrow \infty$  for the maximal offspring

$$X_k^* = \max\{X_i : 1 \leq i \leq T_k\},$$

and completes results of a previous work [2] which we now briefly recall. Plainly,  $X_k^*$  can be viewed as the maximum of  $k$  independent and identically distributed (i.i.d.) copies of  $X_1^*$ , and one should naturally expect extreme value distributions to appear in the limit.

Let  $\bar{F}(n) = \sum_{i=n+1}^{\infty} p(i)$  denote the tail distribution function of  $\mathbf{p}$ , and suppose that

$$\bar{F}(n) = n^{-\alpha} \ell(n), \tag{1}$$

where  $\ell$  is a slowly varying function and  $\alpha \geq 1$ . The main result in [2] is that, for  $1 < \alpha < 2$ ,  $k^{-1} X_k^*$  converges in law as  $k \rightarrow \infty$  to a Fréchet distribution with shape parameter 1.

Received 25 September 2012.

\* Postal address: Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.  
Email address: jean.bertoin@math.uzh.ch

It is rather surprising that this weak limit theorem depends on  $\alpha$  only through the scale parameter of the Fréchet law; in particular, the normalization of  $X_k^*$  is independent of  $\alpha$  and the slowly varying function  $\ell$  plays no role at all. The assumption that  $1 < \alpha < 2$  implies that the variance

$$\sigma^2 = \sum_{n=0}^{\infty} n^2 p(n) - 1$$

of the reproduction law is infinite, and it is therefore natural to wonder whether a similar phenomenon also occurs in the case  $\alpha > 2$  for which the variance  $\sigma^2$  is finite. The approach used in [2] for  $1 < \alpha < 2$  relies on the connection popularized by Harris between branching processes and left-continuous random walks, stable limit theorems for the latter, and the structure of jumps of stable Lévy processes. It does not extend to the case  $\alpha > 2$  since Brownian motion then arises in the limit and the jump structure becomes degenerate.

In order to state the first result of the present contribution, it is convenient to introduce an asymptotic inverse of the tail distribution  $\bar{F}$ , that is, a function  $\varphi : (0, 1) \rightarrow \mathbb{R}_+$  such that

$$\bar{F}(\varphi(\varepsilon)) \sim \varepsilon \quad \text{as } \varepsilon \rightarrow 0+.$$

It is well known that such a function exists when (1) holds; furthermore,  $\varphi$  then varies regularly at  $0+$  with index  $-1/\alpha$  and can be chosen to be nonincreasing.

**Theorem 1.** *Assume that (1) holds with  $\alpha > 2$ . Then, for every  $x > 0$ , we have*

$$\lim_{k \rightarrow \infty} \mathbb{P}(X_k^* < x\varphi(k^{-2})) = \exp\left(-\frac{\sqrt{2}}{\sigma} x^{-\alpha/2}\right).$$

Theorem 1 contrasts sharply with the results of [2]; this points at the following transition for critical branching processes with a reproduction law fulfilling (1) and started from  $k \gg 1$  ancestors. For  $1 < \alpha < 2$ , the maximal offspring is always of order  $k$ , while, for  $\alpha > 2$ , it is only  $k^{2/\alpha}$ . More precisely, the maximal offspring for two such branching processes may have very different asymptotic behaviors even when the two reproduction laws have the same finite variance, whereas the asymptotic behaviors are always essentially the same when both variances are infinite.

It is well known and easy to prove that when the reproduction law  $\mathbf{p}$  has finite variance  $\sigma^2$ , the Laplace transform of  $T_k$ , the total population generated by  $k$  ancestors, fulfills

$$\lim_{k \rightarrow \infty} \mathbb{E}(\exp(-ak^{-2}T_k)) = \exp\left(-\frac{\sqrt{2a}}{\sigma}\right), \quad a \geq 0,$$

or, equivalently, that  $k^{-2}T_k$  converges weakly towards some stable( $\frac{1}{2}$ ) variable  $\tau$ . Comparing with Theorem 1, this suggests that it should be interesting to also take into account the role of  $T_k$  in the study of maximal offspring.

In this direction, introduce a sequence  $(\eta_i)_{i \in \mathbb{N}}$  of i.i.d. variables with law  $\mathbf{p}$ , which we further suppose to be independent of the branching process, and, hence, of  $T_k$ . So  $(X_i)_{i \in \mathbb{N}}$  and  $(\eta_i)_{i \in \mathbb{N}}$  have the same law, but plainly the partial sequences  $(X_i)_{i \leq T_k}$  and  $(\eta_i)_{i \leq T_k}$  do not have the same distribution since  $(X_i)_{i \in \mathbb{N}}$  and  $T_k$  are not independent (and more precisely there is the identity  $T_k = X_1 + \dots + X_{T_k} + 1$ ). Nonetheless, it is easy to check that

$$\eta_k^* = \max\{\eta_i : 1 \leq i \leq T_k\}$$

has the same asymptotical behavior as  $X_k^*$  as  $k \rightarrow \infty$ , namely,

$$\lim_{k \rightarrow \infty} \mathbb{P}(\eta_k^* < x\varphi(k^{-2})) = \exp\left(-\frac{\sqrt{2}}{\sigma}x^{-\alpha/2}\right).$$

This reflects the fact that when the reproduction law fulfills (1) with  $\alpha > 2$ , and the number of ancestors is large, the most prolific individual in a branching process has a negligible impact on the whole process. It is interesting to stress that this phenomenon ceases for  $1 < \alpha < 2$ . More precisely, in the case  $1 < \alpha < 2$ , even though the maximal offspring  $X_k^*$  is order  $k$  and thus much smaller than the total population  $T_k$  which has order  $k^\alpha$ ,  $k^{-1}\eta_k^*$  and  $k^{-1}X_k^*$  do not have the same asymptotic distribution; see Comment 2 of [2, Section 3].

Our second result extends Theorem 1 by considering more generally not just the largest offspring, but also the second, third, ... largest, jointly with the total population size. Specifically, we write

$$X_{k,1}^* = X_k^* \geq X_{k,2}^* \geq X_{k,3}^* \geq \dots$$

for the ordered sequence of offspring numbers  $\{X_i : 1 \leq i \leq T_k\}$ .

**Theorem 2.** *Assume that (1) holds with  $\alpha > 2$ . Then*

$$\left(k^{-2}T_k; \frac{1}{\varphi(k^{-2})}X_{k,1}^*, \frac{1}{\varphi(k^{-2})}X_{k,2}^*, \dots, \frac{1}{\varphi(k^{-2})}X_{k,T_k}^*\right)$$

converges in the sense of finite-dimensional distributions as  $k \rightarrow \infty$  towards  $(\tau; \xi_1, \xi_2, \dots)$ , where  $\tau$  is a positive stable( $\frac{1}{2}$ ) variable with law

$$\frac{\sigma}{\sqrt{2\pi}t^3} \exp\left(-\frac{\sigma^2}{2t}\right) dt, \quad t > 0,$$

and  $\xi_1 > \xi_2 > \dots$  denotes the sequence of the atoms ranked in the decreasing order of a random measure  $M$  on  $(0, \infty)$  such that, conditionally on  $\tau = t$ ,  $M$  is Poisson with intensity  $\alpha tx^{-\alpha-1} dx$ .

In particular, we see that the sequence of offspring numbers, properly normalized and ranked in decreasing order, converges weakly towards the ranked sequence of the atoms of a doubly stochastic Poisson measure  $M$  (also known as a Cox process).

The rest of this work is organized as follows. Proofs of the two theorems are given in the next section, relying on estimates for Laplace transforms of linear functionals of offspring variables. We then gather in Section 3 a number of comments and observations, including some pointers to related literature.

## 2. Proofs

Our approach for establishing Theorems 1 and 2 relies on standard techniques for estimating Laplace transforms of linear functionals of the empirical measure of offspring. One could also prove Theorem 2 by first rephrasing it in terms of random walks and then using more sophisticated tools of weak convergence of càdlàg semimartingales in Skorokhod's space; however, this would require knowing *a priori* the correct orders of magnitudes and also identifying *a priori* dominant contributions and their relations. Therefore we prefer to develop arguments which are both more elementary and natural.

We also point out that Theorem 1 can be recovered from Theorem 2 and the following elementary calculation. Taking Theorem 2 for granted, we have, for every  $x > 0$ ,

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathbb{P}(X_{k,1}^* \leq \varphi(k^{-2})x) &= \mathbb{P}(\xi_1 \leq x) \\ &= \mathbb{P}(M((x, \infty)) = 0) \\ &= \mathbb{E}(\exp(-x^{-\alpha} \tau)) \\ &= \exp\left(-\frac{\sqrt{2}}{\sigma} x^{-\alpha/2}\right). \end{aligned}$$

However, it would probably not have been easy to guess *a priori* the correct renormalization of the largest offspring without knowing Theorem 1 beforehand.

Our starting point is based on the following elementary feature. Consider a function  $f : \mathbb{N} \rightarrow [0, \infty]$ ; we set

$$H(f) = \sum_{i=1}^{T_1} f(X_i)$$

and aim at evaluating the Laplace transform

$$\mathcal{L}(f) = \mathbb{E}(\exp(-H(f))).$$

In this direction, we observe from the branching property that  $\mathcal{L}(f)$  solves the equation

$$s = \sum_{n=0}^{\infty} p(n)e^{-f(n)} s^n, \quad s \in [0, 1]. \tag{2}$$

Since the entire function

$$g_f : s \mapsto \sum_{n=0}^{\infty} p(n)e^{-f(n)} s^n$$

has derivative  $g'_f \leq 1$  on  $[0, 1]$  (recall that  $\mathbf{p}$  is critical) and  $0 < g_f(0) < g_f(1) < 1$ , (2) has a unique solution and thus determines  $\mathcal{L}(f)$ .

**2.1. Proof of Theorem 1**

The heart of the proof of Theorem 1 lies in the following lemma. At this stage, we do not require (1) to hold.

**Lemma 1.** *Let  $\mathbf{p}$  be a critical reproduction law with finite variance  $\sigma^2$ . Suppose that its tail distribution has  $\bar{F}(x) > 0$  for all  $x > 0$  and that*

$$\lim_{y \rightarrow \infty} \limsup_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = 0.$$

Then

$$\mathbb{P}(X_1^* > x) \sim \sqrt{\frac{2\bar{F}(x)}{\sigma^2}} \quad \text{as } x \rightarrow \infty.$$

**Remark.** Note that  $X_1^*$  may have infinite expectation; this occurs for instance when  $\bar{F}(x) \sim (x \ln x)^{-2}$  as  $x \rightarrow \infty$ .

*Proof of Lemma 1.* Let  $x > 0$ , and define  $f(n) = 0$  for  $n \leq x$  and  $f(n) = \infty$  for  $n > x$ . So

$$H(f) = \begin{cases} 0 & \text{if } X_1^* \leq x, \\ \infty & \text{if } X_1^* > x, \end{cases}$$

and if we set  $\mathcal{L}(f) := \varrho(x)$  then  $\varrho(x) = \mathbb{P}(X_1^* \leq x)$  is the distribution function of  $X_1^*$ . In particular, (2) shows that  $\varrho(x)$  is the unique solution to the equation

$$\varrho(x) = \sum_{n \leq x} p(n)\varrho(x)^n, \quad \varrho(x) \in [0, 1]. \tag{3}$$

Now introduce the generating function  $g$  of  $\mathbf{p}$ :

$$g(s) = \sum_{n=0}^{\infty} p(n)s^n, \quad s \in [0, 1].$$

Since  $\mathbf{p}$  is critical with variance  $\sigma^2$ , we have

$$g(s) = s + \frac{1}{2}\sigma^2(1-s)^2 + o((1-s)^2), \quad s \rightarrow 1-.$$

Combining with (3), we obtain the estimate as  $x \rightarrow \infty$ :

$$\frac{1}{2}\sigma^2\bar{\varrho}(x)^2 + o(\bar{\varrho}(x)^2) = \sum_{n > x} p(n)\varrho(x)^n = \bar{F}(x) \sum_{n=1}^{\infty} \frac{p(x+n)}{\bar{F}(x)}\varrho(x)^{n+x}. \tag{4}$$

Here  $\bar{\varrho} = 1 - \varrho$  is the tail distribution of  $X_1^*$ . We now claim that

$$\sum_{n=1}^{\infty} \frac{p(x+n)}{\bar{F}(x)}\varrho(x)^{n+x} \sim 1 \quad \text{as } x \rightarrow \infty,$$

which will complete the proof of the statement, thanks to (4).

Indeed, on the one hand, since  $(p(x+n)/\bar{F}(x), n \geq 1)$  is a probability measure and  $\varrho(x) \leq 1$ , there is the obvious upper bound

$$\sum_{n=1}^{\infty} \frac{p(x+n)}{\bar{F}(x)}\varrho(x)^{n+x} \leq 1. \tag{5}$$

On the other hand, for every fixed  $y > 0$ , we have

$$\sum_{n=1}^{yx} \frac{p(x+n)}{\bar{F}(x)}\varrho(x)^{n+x} \geq \frac{\bar{F}(x) - \bar{F}(x(1+y))}{\bar{F}(x)}\varrho(x)^{x(1+y)}. \tag{6}$$

We now see from (4) and (5) that

$$\bar{\varrho}(x) \leq \sqrt{\frac{2\bar{F}(x)}{\sigma^2}} + o(\bar{\varrho}(x)).$$

Furthermore,  $\bar{F}(x) = o(x^{-2})$  because  $\mathbf{p}$  has a finite variance. Hence,  $\bar{\varrho}(x) = o(1/x)$ , which in turn entails that

$$\lim_{x \rightarrow \infty} \varrho(x)^{(1+y)x} = 1.$$

Plugging this into (6), we get

$$\liminf_{x \rightarrow \infty} \sum_{n=1}^{\infty} \frac{p(x+n)}{\bar{F}(x)} \varrho(x)^{n+x} \geq \liminf_{x \rightarrow \infty} \frac{\bar{F}(x) - \bar{F}(x(1+y))}{\bar{F}(x)},$$

and since  $y$  can be chosen arbitrarily large, we conclude from the assumption of the lemma that

$$\liminf_{x \rightarrow \infty} \sum_{n=1}^{\infty} \frac{p(x+n)}{\bar{F}(x)} \varrho^{n+x}(x) \geq 1,$$

which completes the proof.

If we now assume that (1) holds with  $\alpha > 2$  then the conditions of Lemma 1 are fulfilled. Theorem 1 then follows from the fact that, because different ancestors produce i.i.d. branching processes,  $X_k^*$  can be viewed as the maximum of  $k$  i.i.d. copies of  $X_1^*$ , the estimate of Lemma 1, and the classical result of Gnedenko (see, for instance, Proposition 1.11 of [10]).

**2.2. Proof of Theorem 2**

We consider a continuous function  $f: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $f \equiv 0$  on some neighborhood of 0, and  $a \geq 0$ . Recall that  $\varphi$  is an asymptotic inverse of  $\bar{F}$ . For every  $k \geq 1$ , we write

$$f_{k,a}(x) = ak^{-2} + f\left(\frac{x}{\varphi(k^{-2})}\right).$$

The key step consists in estimating

$$\bar{\mathcal{L}}(f_{k,a}) = 1 - \mathcal{L}(f_{k,a}) = 1 - \mathbb{E}\left(\exp\left(-ak^{-2}T_1 - \sum_{i=1}^{T_1} f\left(\frac{X_i}{\varphi(k^{-2})}\right)\right)\right)$$

as  $k \rightarrow \infty$ .

**Lemma 2.** *Assume that (1) holds with  $\alpha > 2$ . In the notation above, we have*

$$\lim_{k \rightarrow \infty} k \bar{\mathcal{L}}(f_{k,a}) = \frac{\sqrt{2}}{\sigma} \left( a + \alpha \int_0^\infty (1 - e^{-f(x)}) x^{-\alpha-1} dx \right)^{1/2} \quad \text{as } k \rightarrow \infty.$$

*Proof.* For the sake of simplicity, we assume that  $f \equiv 0$  on  $[0, 1]$ , the general case requiring only slightly heavier notation. The calculations are closely related to those in the proof of Lemma 1, and we will therefore sometimes provide fewer details. We first recall from (2) that there is the identity

$$\mathcal{L}(f_{k,a}) = \sum_{n=0}^{\infty} p(n) \exp\left(-ak^{-2} - f\left(\frac{n}{\varphi(k^{-2})}\right)\right) \mathcal{L}(f_{k,a})^n,$$

and then, using the expansion of the generating function  $g$ , we get

$$\frac{1}{2} \sigma^2 \bar{\mathcal{L}}(f_{k,a})^2 + o(\bar{\mathcal{L}}(f_{k,a})^2) = \sum_{n=0}^{\infty} p(n) \left( 1 - \exp\left(-ak^{-2} - f\left(\frac{n}{\varphi(k^{-2})}\right)\right) \right) \mathcal{L}(f_{k,a})^n. \quad (7)$$

We first use the facts that  $f \equiv 0$  on  $[0, 1]$  and  $\mathcal{L}(f_{k,a}) \leq 1$ , and get the upper bounds

$$\frac{1}{2}\sigma^2 \bar{\mathcal{L}}(f_{k,a})^2 + o(\bar{\mathcal{L}}(f_{k,a})^2) \leq \sum_{n \leq \varphi(k^{-2})} p(n)(1 - e^{-ak^{-2}}) + \sum_{n > \varphi(k^{-2})} p(n) \leq ak^{-2} + \bar{F}(\varphi(k^{-2})).$$

Since  $\bar{F}(\varphi(k^{-2})) \sim k^{-2}$ , we conclude that  $\bar{\mathcal{L}}(f_{k,a}) = O(1/k)$ . Recall that  $\varphi$  is regularly varying at  $0+$  with index  $-1/\alpha > -\frac{1}{2}$ , in particular,  $\varphi(k^{-2}) = o(k)$ , and, therefore,

$$\lim_{k \rightarrow \infty} \mathcal{L}(f_{k,a})^{\varphi(k^{-2})} = 1. \tag{8}$$

We need to estimate the right-hand side of (7) as  $k \rightarrow \infty$ ; in this direction it is convenient to decompose the sum depending on whether  $n \leq \varphi(k^{-2})$  or  $n > \varphi(k^{-2})$ , as the summand has different asymptotic behaviors on these two regions. So first, as  $f \equiv 0$  on  $[0, 1]$ , we deduce from (8) that

$$\sum_{n \leq \varphi(k^{-2})} p(n) \left( 1 - \exp\left(-ak^{-2} - f\left(\frac{n}{\varphi(k^{-2})}\right)\right) \right) \mathcal{L}(f_{k,a})^n \sim \sum_{n \leq \varphi(k^{-2})} p(n)(1 - e^{-ak^{-2}}) \sim ak^{-2}.$$

Second, assumption (1) ensures that, for every bounded and continuous function  $h : [1, \infty) \rightarrow \mathbb{R}$ , we have

$$\sum_{n > \varphi(k^{-2})} p(n) h\left(\frac{n}{\varphi(k^{-2})}\right) \sim \bar{F}(\varphi(k^{-2}))\alpha \int_1^\infty h(x)x^{-\alpha-1} dx.$$

Since  $\mathcal{L}(f_{k,a}) \leq 1$  and  $\bar{F}(\varphi(k^{-2})) \sim k^{-2}$ , we then readily deduce from (8) that

$$\sum_{n > \varphi(k^{-2})} p(n) \left( 1 - \exp\left(-ak^{-2} - f\left(\frac{n}{\varphi(k^{-2})}\right)\right) \right) \mathcal{L}(f_{k,a})^n \sim k^{-2}\alpha \int_0^\infty (1 - e^{-f(x)})x^{-\alpha-1} dx.$$

These two estimates combined with (7) complete the proof of the lemma.

We can now tackle the proof of Theorem 2. Because the branching processes generated by distinct ancestors are independent, we have

$$\mathbb{E}\left(\exp\left(-ak^{-2}T_k - \sum_{n=1}^{T_k} f\left(\frac{X_n}{\varphi(k^{-2})}\right)\right)\right) = \mathcal{L}(f_{k,a})^k,$$

and it follows from Lemma 2 that

$$\lim_{k \rightarrow \infty} \mathbb{E}\left(\exp\left(-ak^{-2}T_k - \sum_{n=1}^{T_k} f\left(\frac{X_n}{\varphi(k^{-2})}\right)\right)\right) = \exp\left(-\frac{\sqrt{2}}{\sigma}\left(a + \alpha \int_0^\infty (1 - e^{-f(x)})x^{-\alpha-1} dx\right)^{1/2}\right). \tag{9}$$

We next identify the right-hand side of (9) as the joint Laplace transform of  $(\tau, \langle M, f \rangle)$ , where  $\tau$  is a random variable and  $M$  a random point measure on  $(0, \infty]$  whose joint law is specified in Theorem 2. Indeed, we have then, first,

$$\mathbb{E}(\exp(-a\tau)) = \exp\left(-\sqrt{\frac{2a}{\sigma^2}}\right),$$

second,

$$\mathbb{E}(e^{-\langle M, f \rangle} \mid \tau = t) = \exp\left(-t\alpha \int_0^\infty (1 - e^{-f(x)})x^{-\alpha-1} dx\right),$$

and, finally,

$$\mathbb{E}(\exp(-a\tau - \langle M, f \rangle)) = \exp\left(-\frac{\sqrt{2}}{\sigma} \left(a + \alpha \int_0^\infty (1 - e^{-f(x)})x^{-\alpha-1} dx\right)^{1/2}\right).$$

Next introduce, for each  $k \geq 1$ , the empirical measure of rescaled offspring when the branching process has  $k$  ancestors:

$$M'_k = \sum_{i=1}^{T_k} \delta_{X_i/\varphi(k^{-2})}.$$

We ignore the atom of  $M'_k$  at 0 to get a random point measure  $M_k$  on  $(0, \infty]$ , and rephrase (9) as

$$\lim_{k \rightarrow \infty} \mathbb{E}(\exp(-ak^{-2}T_k - \langle M_k, f \rangle)) = \mathbb{E}(\exp(-a\tau - \langle M, f \rangle))$$

for all  $a \geq 0$  and continuous functions  $f: (0, \infty] \rightarrow \mathbb{R}_+$  which have compact support. According to Theorem 14.16 of [5], this ensures that the pair  $(k^{-2}T_k, M_k)$  converges weakly as  $k \rightarrow \infty$  towards  $(\tau, M)$ , where the space of Radon measures on  $(0, \infty]$  is endowed with the vague topology. This entails Theorem 2.

### 3. Miscellaneous remarks

1. The literature surveyed by Yanev [12], [13] contains a variety of results about extremes for individuals in branching processes. In particular, Rahinov and Yanev [9] have characterized the asymptotic behavior as  $n \rightarrow \infty$  of the maximal offspring at the  $n$ th generation conditionally on the event that the extinction has not yet occurred; previously, Pakes [7] considered asymptotics of the largest score up to and including the  $n$ th generation, again conditionally on the event that extinction does not occur before the  $n$ th generation. In a different but related direction, namely the distribution of the maximum of the branching process itself, we refer the reader to [1], [3], [6], and [11].

2. If one replaces assumption (1) in Theorem 1 by the condition that the tail distribution of the reproduction law decays exponentially, say  $\bar{F}(x) \sim ae^{-bx}$  for some  $a, b > 0$ , then the same argument shows that  $X_k^* - 2b^{-1} \ln k$  now converges weakly as  $k \rightarrow \infty$  to a Gumbel distribution. In the same way, the analogue of Theorem 2 can be stated as follows. As  $k \rightarrow \infty$ ,

$$(k^{-2}T_k; X_{k,1}^* - 2b^{-1} \ln k, X_{k,2}^* - 2b^{-1} \ln k, \dots, X_{k,T_k}^* - 2b^{-1} \ln k)$$

converges in the sense of finite-dimensional distributions towards  $(\tau; x_1, x_2, \dots)$ , where  $\tau$  has the law

$$\frac{\sigma}{\sqrt{2\pi t^3}} \exp\left(-\frac{\sigma^2}{2t}\right) dt, \quad t > 0,$$



and, conditionally on  $\tau = t$ ,  $x_1 > x_2 > \dots$  denotes the sequence of atoms ranked in decreasing order of a Poisson measure on  $(-\infty, \infty)$  with intensity  $tabe^{-bx} dx$ .

3. We point out that (2) has a natural interpretation in terms of branching processes. In this direction, we may view  $\mathbf{p}_f = (p_f(n) = p(n)e^{-f(n)})_{n \in \mathbb{N}}$  as a probability measure on  $\mathbb{N} \cup \{\infty\}$  by setting  $p_f(\infty) = 1 - \sum_{n=0}^{\infty} p(n)e^{-f(n)}$ . Then

$$\sum_{n=0}^{\infty} p(n)e^{-f(n)}s^n, \quad 0 \leq s < 1,$$

should be thought of as the generating function of  $\mathbf{p}_f$ , and if we consider a Galton–Watson branching process with values in  $\mathbb{N} \cup \{\infty\}$  and reproduction law  $\mathbf{p}_f$  (of course, this process is absorbed at  $\infty$  as soon as an individual has an infinite offspring), then we recognize (2) as the equation for the probability of extinction.

In particular, the well-known formula of Dwass [4] for the distribution of the total population size yields the semiexplicit formula

$$\mathcal{L}(f) = \sum_{n=1}^{\infty} \frac{1}{n} \mathbf{p}_f^{*n}(n-1),$$

where  $\mathbf{p}_f^{*j} = \mathbf{p}_f * \dots * \mathbf{p}_f$  denotes the  $j$ th convolution power of  $\mathbf{p}_f$ . Unfortunately, this expression seems too complicated to be of some use in practice.

4. It is natural to view  $H(f)$  as a Hamiltonian,  $\mathcal{L}(f)$  as a partition function, and introduce the Gibbs measure

$$\mathbb{Q} = \frac{e^{-H(f)}}{\mathcal{L}(f)} \mathbb{P},$$

where  $\mathbb{P}$  denotes the law of the Galton–Watson process with reproduction law  $\mathbf{p}$  and started from a single ancestor. Using (2) and, for instance, Otter’s formula (see, e.g. Formula (304) of Pitman [8]), it is easily seen that  $\mathbb{Q}$  is again the law of a Galton–Watson process with reproduction law given by

$$q(n) = \mathcal{L}(f)^{n-1} e^{-f(n)} p(n), \quad n \in \mathbb{N}.$$

In particular, it is now elementary to compute the relative entropy  $D(\mathbb{Q} \parallel \mathbb{P})$  of  $\mathbb{Q}$  with respect to  $\mathbb{P}$ . Indeed, excluding implicitly the trivial case  $f \equiv 0$ , Wald’s lemma yields

$$\mathbb{E}_{\mathbb{Q}}(H(f)) = \langle \mathbf{q}, f \rangle \mathbb{E}_{\mathbb{Q}}(T_1),$$

with  $\langle \mathbf{q}, f \rangle = \sum_{n=0}^{\infty} f(n)q(n)$ , and it is a standard fact for subcritical branching processes that

$$\frac{1}{\mathbb{E}_{\mathbb{Q}}(T_1)} = 1 - \mathbb{E}_{\mathbb{Q}}(X_1) = 1 - \sum_{n=0}^{\infty} nq(n).$$

5. It would be interesting to obtain an analogue of Theorem 2 in the case  $1 < \alpha < 2$ . As a first modest step, one should probably try to understand the impact of the maximal offspring  $X_k^*$  on the total population  $T_k$ .

### References

[1] ATHREYA, K. B. (1988). On the maximum sequence in a critical branching process. *Ann. Prob.* **16**, 502–507.  
 [2] BERTOIN, J. (2011). On the maximal offspring in a critical branching process with infinite variance. *J. Appl. Prob.* **48**, 576–582.  
 [3] BOROVKOV, K. A. AND VATUTIN, V. A. (1996). On distribution tails and expectations of maxima in critical branching processes. *J. Appl. Prob.* **33**, 614–622.

- [4] DWASS, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Prob.* **6**, 682–686.
- [5] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
- [6] LINDVALL, T. (1976). On the maximum of a branching process. *Scand. J. Statist.* **3**, 209–214.
- [7] PAKES, A. G. (1998). Extreme order statistics on Galton-Watson trees. *Metrika* **47**, 95–117.
- [8] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. (Lecture Notes Math. **1875**), Springer, Berlin.
- [9] RAHIMOV, I. AND YANEV, G. P. (1999). On maximum family size in branching processes. *J. Appl. Prob.* **36**, 632–643.
- [10] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- [11] VATUTIN, V. A., WACHTEL, V. AND FLEISCHMANN, K. (2008). Critical Galton-Watson branching processes: the maximum of the total number of particles within a large window. *Theory Prob. Appl.* **52**, 470–492.
- [12] YANEV, G. P. (2007). Revisiting offspring maxima in branching processes. *Pliska Stud. Math. Bulgar.* **18**, 401–426.
- [13] YANEV, G. P. (2008). A review of offspring extremes in branching processes. In *Records and Branching Processes*, eds M. Ahsanullah and G. P. Yanev, Nova Science, pp. 127–145.