*J. Appl. Prob. Spec. Vol.* **51A,** 63–72 (2014)


# CELEBRATING 50 YEARS OF THE APPLIED PROBABILITY TRUST

Edited by
S. ASMUSSEN, P. JAGERS, I. MOLCHANOV and L. C. G. ROGERS

Part 3. Biological applications

## HOW DID WE GET HERE?

KAIS HAMZA, *Monash University*
School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.
Email address: kais.hamza@monash.edu

FIMA C. KLEBANER, *Monash University*
School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.
Email address: fima.klebaner@monash.edu

APPLIED PROBABILITY TRUST
DECEMBER 2014

https://doi.org/10.1239/jap/1417528467 Published online by Cambridge University Press

# HOW DID WE GET HERE?

BY KAIS HAMZA AND FIMA C. KLEBANER

**Abstract**

Looking at a large branching population we determine along which path the population that started at 1 at time 0 ended up in $B$ at time $N$. The result describes the density process, that is, population numbers divided by the initial number $K$ (where $K$ is assumed to be large). The model considered is that of a Galton–Watson process. It is found that in some cases population paths exhibit the strange feature that population numbers go down and then increase. This phenomenon requires further investigation. The technique uses large deviations, and the rate function based on Cramer's theorem is given. It also involves analysis of existence of solutions of a certain algebraic equation.

*Keywords:* Branching process; large deviations; most likely path

2010 Mathematics Subject Classification: Primary 60J80
Secondary 60F10

## 1. Introduction

Consider a branching population started with a large number $K$ at time 0, and let $Z_n$ denote its size at time $n = 0, 1, 2, \ldots$. Since $K$ is large, the branching property implies that the law of large numbers holds. Consequently, the density process at time $n$, $Z_n/K$, is approximated by a deterministic function. Suppose that we observe $Z_N/K = B$ at time $N$. Then there is a function ending in $B$ such that its small neighbourhood contains the most likely path of the process $Z_n/K$, $0 \leq n \leq N$. We find such paths. While many paths exhibit monotonicity, in some cases monotonicity is lost, leading to a strange phenomenon in which the population numbers drop significantly below $B$ before they increase to $B$.

The method of obtaining the most likely path uses the theory of large deviations, and the main result of this paper is stated in Theorem 1. Every path that is not in the vicinity of the mean behaviour predicted by the law of large numbers represents a large deviation in the stochastic process. Large deviations theory asymptotically gives the probability that the path of the process is in a neighbourhood of a function $u$ on $[0, N]$. While we continue to think of $u$ as a function on $[0, N]$, it is worth noting that, because of the finiteness of the time index set $[0, N]$, $u$ is in fact an $(N + 1)$-dimensional vector. Denoting by $\rho(u, w)$ a metric on the functions $u$ and $w$, the probability that the random path $W$ is in a $\delta$-neighbourhood of $u$ is given approximately by

$$\mathbb{P}\{\rho(W, u) \leq \delta\} \approx \mathrm{e}^{-KI(u)},$$

where $I(u)$ (the rate function) depends only on $u$, and a precise meaning of the approximation, denoted be '$\approx$', is given by the limit relation

$$\lim_{\delta\to 0}\lim_{K\to\infty}\frac{1}{K}\ln\mathbb{P}\{\rho(W, u) \leq \delta\} = -I(u).$$

If we consider paths $u$ that originate at $A$ and end in $B$ at time $N$, the path that minimizes the rate function $I(u)$, say $u^*$, will be called *the most likely path* from $A$ to $B$. Since we are concerned with the density process $Z_n/Z_0$, the initial state $A$ is in fact 1.

Actually, the following more general statement holds. The path $u^*$ maximizes the probability amongst a much larger set of paths $F$ that start at $A$ and end in $B$, and not just for those paths being in a $\delta$-neighbourhood:

$$\lim_{K\to\infty} \frac{1}{K} \ln \mathbb{P}\{W \in F\} = -\inf_{u\in F} I(u) = -I(u^*).$$

While the general theory of large deviations is well known, explicit minimization results are seldom obtained. In our applications we find such paths explicitly, and this allows us to study their interesting features. The branching property allows us to obtain the large deviations rate function for the Galton–Watson process by using the elementary Cramér theorem on large deviations for random vectors.

Large deviations for branching processes have been considered in the literature; see, e.g. [1, 3, 8, 10]. However, with the exception of [8], these papers are concerned with the behaviour of the sequence $Z_n/Z_{n-1}$ as $n$ approaches $\infty$, whereas we focus on large deviations for density processes on a finite interval and indexed by the initial size $K$.

## 2. Results

Recall the definition of the Galton–Watson process, which serves as a basic stochastic model for growth in discrete time; see, e.g. [2, 5]. Let $Z_0$ be a positive integer, and let

$$Z_{n+1} = \sum_{i=1}^{Z_n} \eta_{i,n},$$

where the random variables $\eta_{i,n}$ are independent and identically distributed (i.i.d.). By convention, $\sum_{i=1}^{0} = 0$.

Let $(p_l)_{l\geq 0}$ denote the probability sequence of the offspring distribution $\mathbb{P}\{\eta_{i,n} = \ell\} = p_\ell$, and let $f$ be its probability generating function (PGF)

$$f(s) = \mathbb{E}[s^{\eta_{i,n}}] = \sum_{\ell=0}^{\infty} p_\ell s^\ell.$$

Let $R$ denote the radius of convergence of $f$; assume throughout the paper that $R > 1$. This ensures the existence of exponential moments $\mathbb{E}[\mathrm{e}^{t\eta_{i,n}}]$ for some positive $t$, a necessary requirement for Cramér's large deviations result to hold (see Theorem 3 below).

Often the initial number of particles is taken to be 1. To distinguish such processes, we denote them by $X$, i.e. $X$ denotes a Galton–Watson process with $X_0 = 1$. In this case, the well-known basic result on Galton–Watson processes states that the distribution of $X_n$ has PGF $f_n$, the $n$th iterate of $f$, $f_n(s) = f(f_{n-1}(s))$, $f_0(s) = s$.

An immediate corollary is that $\mathbb{P}\{X_n = 0\} = f_n(0)$, and since $f_n'(s) = \prod_{i=0}^{n-1} f'(f_i(s))$, it follows that when the mean exists it is given by

$$\mathbb{E}[X_n] = f_n'(1) = (f'(1))^n = m^n,$$

where $m = f'(1)$ is the offspring mean.

**Theorem 1.** *Let $Z_n$ be a Galton–Watson process started with a large number $Z_0 = K$ of particles. Suppose that the equation*

$$B = y\frac{f_N'(y)}{f_N(y)} \tag{1}$$

*has at least one nonnegative solution, $y_N$. Then the most likely path of the process $Z_n/K$, $0 \le n \le N$, to go from 1 at time 0 to $B$ at time $N$ is given by*

$$u_n^* = \frac{f_{N-n}(y_N)}{f_N(y_N)}\frac{f_N'(y_N)}{f_{N-n}'(y_N)}, \qquad n = 0, 1, \dots, N. \tag{2}$$

Recall that the best (in the mean-squared error sense) predictor of a random variable is its mean. It turns out that the most likely path is the mean of the process, but under a new measure $\widetilde{\mathbb{P}}$ (tilted distribution) which makes the mean of $Z_N$ equal to $B$. In it, $\Lambda_B$ is known as the *Esscher transform* of $X_N$.

**Theorem 2.** *Let $y_N$ solve (1), and let $\widetilde{\mathbb{P}}$ be defined by*

$$\mathrm{d}\widetilde{\mathbb{P}} = \Lambda_B\,\mathrm{d}\mathbb{P}, \qquad \Lambda_B = \frac{y_N^{X_N}}{\mathbb{E}[y_N^{X_N}]}.$$

*Then, for all $0 \le n \le N$,*

$$u_n^* = \widetilde{\mathbb{E}}[X_n].$$

The special case $B = 0$ gives the most likely path to extinction. This result was obtained in [9] and is given here for completeness. Note that, in the case $p_0 = 0$, $yf_N'(y)/f_N(y)$ is extended by continuity to 1 at $y = 0$.

**Corollary 1.** *If $p_0p_1 > 0$, the most likely path to extinction (2) can be written as*

$$u_n^* = \frac{\mathbb{P}\{X_{N-n} = 0\}}{\mathbb{P}\{X_N = 0\}}\frac{\mathbb{P}\{X_N = 1\}}{\mathbb{P}\{X_{N-n} = 1\}},$$

*or, more simply, $u_n^* = \mathbb{E}[X_n \mid X_N = 0]$.*

*Proof.* If $B = 0$ then $y = 0$ is a solution of (1); furthermore, $f_n(0) = \mathbb{P}\{X_n = 0\}$ and $f_n'(0) = \mathbb{P}\{X_n = 1\}$. The second representation is a direct consequence of Theorem 2, where in this case

$$\Lambda_B = \frac{1}{\mathbb{P}\{X_N = 0\}}\mathbf{1}_{\{X_N = 0\}}.$$

Owing to the branching property, large deviations for Galton–Watson processes follow directly from Cramér's theorem for random vectors. Let $Z$ be a Galton–Watson process started with $Z_0 = K$ particles; then

$$Z_n = \sum_{i=1}^{K} X_n^{(i)},$$

where the $X^{(i)}$ are independent copies of the Galton–Watson process started with a single particle $X_0^{(i)} = 1$. As observed above, these processes on $[0, N]$ are vectors and Cramér's theorem applies.

Recall Cramér's theorem in $\mathbb{R}^N$ (see, e.g. [4]). It states that the rate function is given by the transform of the logarithm of the moment generating function.

**Theorem 3.** (Cramér.) *Let $(X^i)_{i=1}^{\infty}$ be i.i.d. random vectors in $\mathbb{R}^N$ with log-moment generating function $G$. Then, for any $u = (u_1, \dots, u_N)$ in $\mathbb{R}^N$,*

$$\lim_{\delta \to 0} \lim_{K \to \infty} \frac{1}{K} \ln \mathbb{P}\{\rho(W, u) \le \delta\} = -G^*(u),$$

*where $W$ is the average random vector $(1/K)\sum_{i=1}^{K} X^i$ and $G^*$ is the Legendre–Fenchel transform of the log-moment generating function of the vector $X^i = (X_1^i, \dots, X_N^i)$,*

$$G^*(u) := \sup_{\lambda_1, \dots, \lambda_N} \left\{ \sum_{n=1}^{N} \lambda_n u_n - G(\lambda_1, \dots, \lambda_N) \right\}.$$

The next result explicitly gives the Legendre–Fenchel transform of the Galton–Watson process $X$.

**Theorem 4.** *The Legendre–Fenchel transform of the Galton–Watson process $X = (X_0, X_1, \dots, X_N)$, with $X_0 = 1$, is given by*

$$G^*(u) = \sum_{n=0}^{N-1} u_n g^*\left(\frac{u_{n+1}}{u_n}\right), \tag{3}$$

*where*

$$g^*(v) = \sup_t \{tv - g(t)\}$$

*is the Legendre–Fenchel transform of $g$, the log-moment generating function of the offspring distribution.*

**Remark 1.** The rate function was derived in [7] for density-dependent branching processes and in [6], using different techniques, for Markov chains.

## 3. Example: the linear-fractional branching process

In this section we detail the case of the linear-fractional branching process. We exhibit a variety of situations and behaviours of the most likely path. The attraction of this example lies in the fact that there are available explicit analytic expressions for the iterates of the probability generating function, $f$. Recall (see, e.g. [5, Chapter I.7.1]) that, for the linear-fractional branching process, $p_k = cp^{k-1}$, $k \ge 1$, and $p_0 = (1 - c - p)/(1 - p)$, $0 < p < 1$, $0 < c < 1$, and $c + p \le 1$. Furthermore, as mentioned, an expression for the iterates of $f$, the generating function, is available, thus allowing the exact computation of $u_n^*$ and even $y_N$.

**Example 1.** (*Nondecreasing linear-fractional branching process.*) If $c = 1 - p$ then $p_0 = 0$, $m = 1/(1 - p) > 1$, the process is supercritical, and $f_n(s) = s/[m^n - (m^n - 1)s]$. The equation $B = y\, f_N'(y)/f_N(y)$ has the unique nonzero solution

$$y_N = \frac{(B - 1)m^N}{B(m^N - 1)}.$$

For $B > 1$, this solution is positive and the most likely path is given by

$$u_n^* = \frac{m^N - B}{m^N - 1} + \frac{(B - 1)}{m^N - 1} m^n.$$

The graphs in Figure 1 depict typical behaviour for the most likely path in two particular cases.

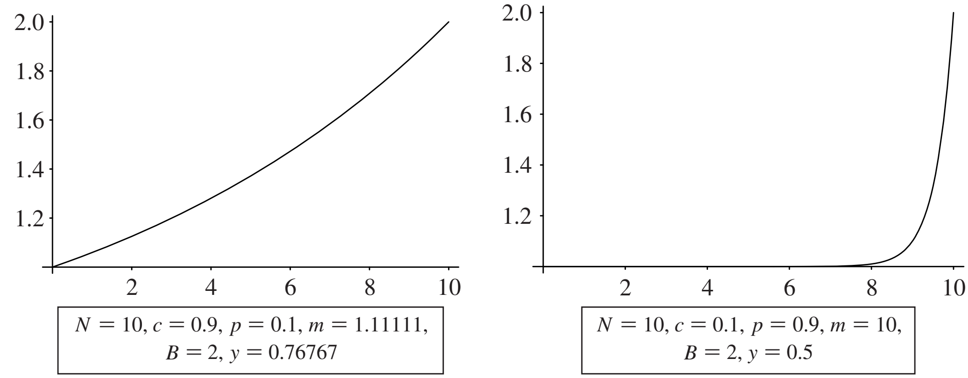



FIGURE 1: From 1 to 2 in ten steps for two nondecreasing linear-fractional branching processes.

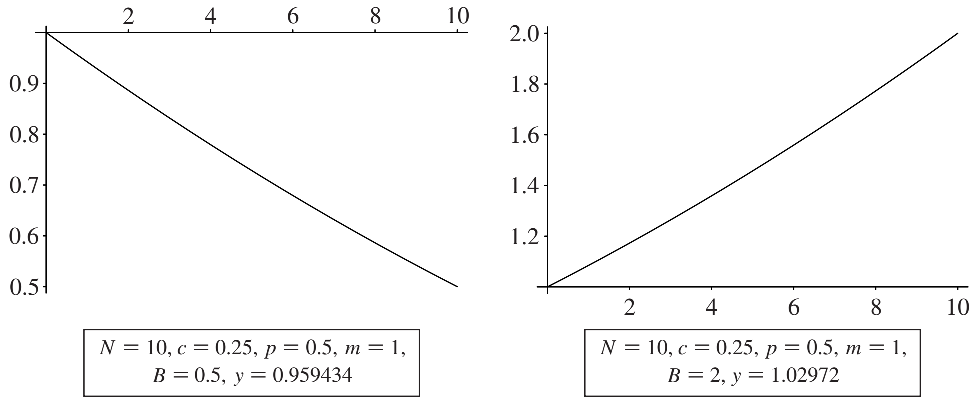



FIGURE 2: From 1 to 0.5 and 2 in ten steps for a critical linear-fractional branching process.

While in some cases the most likely path is, as may be expected, close to a straight line, in other cases, and possibly unexpectedly, it can display distinctly different behaviour. In this case, the process is nondecreasing and if started from 1, $Z_n/K$ cannot reach any level below 1. Furthermore, for $B < 1$,

$$B = \frac{y f_N'(y)}{f_N(y)}$$

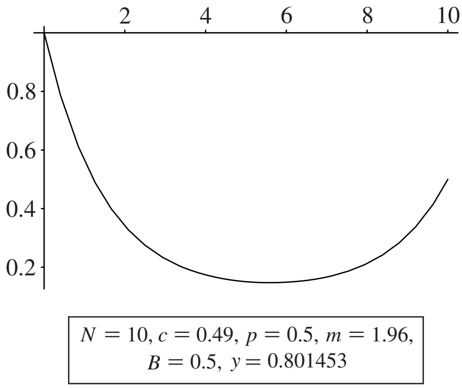
has a negative solution.

**Example 2.** (*Critical linear-fractional branching process.*) If $c = (1-p)^2$ then $m = 1$, the process is critical, and

$$f_n(s) = \frac{np - (np + p - 1)s}{1 - p + np - nps}.$$

In this case, $B = y\, f_N'(y)/f_N(y)$ has a unique solution $y\, f_N'(y)/f_N(y)$ increases from 0 to $+\infty$), and the most likely path can be computed numerically (see Figure 2).

**Example 3.** (*General linear-fractional branching process.*) Possibly the most interesting situation is obtained for a supercritical, nonmonotone linear-fractional branching process ($p_0 \neq 0$). For careful choices of $c$ and $p$, the most likely path can display nonmonotone behaviour (see Figure 3).



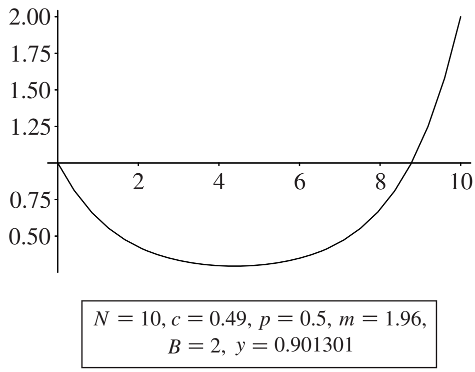




FIGURE 3: From 1 to 0.5 and 2 in ten steps for a supercritical linear-fractional branching process.

## 4. Proofs

### 4.1. Proof of Theorem 1

We break the proof of Theorem 1 into several lemmas. We obtain $u_n^*$ by minimizing the function $G^*$ given in Theorem 3 with two points fixed: $u_0 = 1$, $u_N = B$,

$$G^*(u_1, \dots, u_{N-1}) = \sum_{n=0}^{N-1} u_i g^*\left(\frac{u_{n+1}}{u_n}\right).$$

As a convex function, any critical point of $G^*$ (i.e. a sequence $(u_1, \dots, u_{N-1})$) is a global minimum. To identify such critical points, we need the properties of $g^*$ and its derivative given in Lemma 3 below. It turns out that this analysis is not trivial.

For $x, y \ge 0$, write $x \prec y$ to mean $x \le y$ if $y$ is finite, and $x < \infty$ if $y$ is infinite.

While we apply the following lemmas to the offspring random variables $\eta_{i,n}$, the statements are general and refer to a generic random variable, $\xi$, with generating function $f$, and are given here for completeness.

Let $R$ be the radius of convergence of the generating function $f$, define

$$\omega = \lim_{z \uparrow R} f(z),$$

and let $g$ be the corresponding log-moment generating function, $g(t) := \ln f(\mathrm{e}^t)$. Then $g$ is convex and increasing, and since the offspring distribution has exponential moments, $R > 1$. It also follows that, unless $p_0 + p_1 = 1$, $\lim_{z \uparrow R} f(z)/z > 1$ (and is possibly infinite). Furthermore, $g'(t) = \mathrm{e}^t f'(\mathrm{e}^t)/f(\mathrm{e}^t)$ is defined on $(-\infty, \ln R)$ with values in $(0, \theta)$, where

$$\theta = \lim_{z \uparrow R} \frac{z f'(z)}{f(z)}.$$

There are four cases to consider depending on whether $R$ and $\theta$ are finite or infinite. Lemmas 1 and 2 below describe what happens when $\theta < +\infty$. Note that $\omega = +\infty$ whenever $R = +\infty$.

**Lemma 1.** *If $R = +\infty$ then $\xi \prec \theta$ almost surely (a.s.), i.e. if $R = +\infty$ then either $\theta = +\infty$, or $\theta < +\infty$ and $\xi \le \theta$.*

*Proof.* Suppose that $R = +\infty$ and $\theta < +\infty$; consider the function $g_\theta(t) := \theta t - g(t)$. Now $g_\theta'(t)$ decreases from $\theta = g_\theta'(-\infty)$ to $0 = g_\theta'(+\infty)$, so $g_\theta'(t) \ge 0$ for all $t$; hence, $g_\theta(t)$

https://doi.org/10.1239/jap/1417528467 Published online by Cambridge University Press

is strictly increasing on $\mathbb{R}$. Write

$$\theta t - g(t) = -\ln(\mathrm{e}^{-\theta t}\mathbb{E}[\mathrm{e}^{t\xi}]) = -\ln(\mathbb{E}[\mathrm{e}^{t(\xi-\theta)},\ \xi \le \theta] + \mathbb{E}[\mathrm{e}^{t(\xi-\theta)},\ \xi > \theta]).$$

If $\mathbb{P}\{\xi > \theta\} > 0$ then, by monotone convergence, the right-hand side term tends to $-\infty$ as $t \to +\infty$, which contradicts the earlier observation that $g_\theta(t)$ is strictly increasing on $\mathbb{R}$. It follows immediately that $\xi \le \theta$ a.s.

Note that if $K := \operatorname{ess\,sup} \xi < \infty$ then $\theta = K$. Thus, stronger than Lemma 1, if $R = +\infty$ then $\theta < +\infty$ if and only if $K := \operatorname{ess\,sup} \xi < \infty$, and then $\theta = K$.

**Lemma 2.** *If $R < +\infty$ and $\theta < +\infty$, then $\omega < +\infty$.*

*Proof.* Suppose that $R < +\infty$ and $\theta < +\infty$. Consider again the function $g_\theta(t)$. Since $g_\theta'(t) \ge 0$ as in the proof of Lemma 1, $g_\theta(t)$ is strictly increasing on $(-\infty, \ln R)$ with a supremum at $\ln R$ equal to $\theta \ln R - \ln \omega$, which automatically requires that $\omega$ be finite.

When $R < +\infty$ and $\theta = +\infty$, $\omega$ can either be finite or infinite.

**Lemma 3.** *For unbounded $\xi$, $g^*(v) = \sup_{t<\ln R}\{vt - g(t)\}$ is finite for $v \ge 0$ (or for $v > 0$ if $p_0 = 0$) and*

$$g^*(v) = \begin{cases} -\ln p_0 & \text{for } v = 0,\\ v(g')^{-1}(v) - g((g')^{-1}(v)) & \text{for } 0 < v < \theta,\\ v\ln R - \ln\omega & \text{for } v \ge \theta \text{ and } \theta < +\infty. \end{cases}$$

*However, if $R = +\infty$ and $\theta < +\infty$ so that $\xi$ is bounded by $\theta$, then $g^*$ is finite on $[0, \theta]$ (or on $(0, \theta]$ if $p_0 = 0$) and*

$$g^*(v) = \begin{cases} -\ln p_0 & \text{for } v = 0,\\ v(g')^{-1}(v) - g((g')^{-1}(v)) & \text{for } 0 < v < \theta,\\ -\ln\mathbb{P}(\xi = \theta) & \text{for } v = \theta. \end{cases}$$

*Furthermore, in all cases,*

$$(g^*)'(v) = \begin{cases} (g')^{-1}(v) & \text{for } 0 < v < \theta,\\ \ln R & \text{for } v > \theta,\ R < +\infty,\ \text{and } \theta < +\infty. \end{cases}$$

Next we discuss the existence of a critical point of $G^*$ and proceed to identify it, when it exists. To this end, define $\gamma = (g^*)'$, and recall that all paths considered here satisfy $u_0 = 1$ and $u_N = B$.

**Lemma 4.** *For each $n = 1, \dots, N-1$, $G^*$, as a function of $u_n$ alone, has exactly one critical point. It satisfies the equation*

$$g\left(\gamma\left(\frac{u_{n+1}}{u_n}\right)\right) = \gamma\left(\frac{u_n}{u_{n-1}}\right). \tag{4}$$

*Proof.* We have

$$\begin{aligned} \frac{\partial}{\partial u_n} G^*(u_1, \dots, u_{N-1}) &= \frac{\partial}{\partial u_n}\left(u_n g^*\left(\frac{u_{n+1}}{u_n}\right) + u_{n-1} g^*\left(\frac{u_n}{u_{n-1}}\right)\right)\\ &= g^*\left(\frac{u_{n+1}}{u_n}\right) - \frac{u_{n+1}}{u_n}(g^*)'\left(\frac{u_{n+1}}{u_n}\right) + (g^*)'\left(\frac{u_n}{u_{n-1}}\right),\\ &= -g\left(\gamma\left(\frac{u_{n+1}}{u_n}\right)\right) + \gamma\left(\frac{u_n}{u_{n-1}}\right) \quad \text{(using Lemma 3).} \end{aligned}$$

https://doi.org/10.1239/jap/1417528467 Published online by Cambridge University Press

When $\theta = +\infty$,

$$\frac{\partial}{\partial u_n} G^*(u_1, \dots, u_{N-1}) = -g\left(\gamma\left(\frac{u_{n+1}}{u_n}\right)\right) + \gamma\left(\frac{u_n}{u_{n-1}}\right) = -\ln(\mathrm{e}^{-\gamma(u_n/u_{n-1})} f(\mathrm{e}^{\gamma(u_{n+1}/u_n)})).$$

The right-hand side tends to $-\infty$ as $u_n \to 0$, and tends to $\ln(R/p_0)$ (or $+\infty$) as $u_n \to +\infty$. In particular, there exists a unique $u_n$ such that $\partial G^*(u_1, \dots, u_{N-1})/\partial u_n = 0$.

When $\theta < +\infty$, then $-g(\gamma(u_{n+1}/u_n)) + \gamma(u_n/u_{n-1})$ approaches a quantity that is less than (or possibly equal to) $-\ln(f(R)/R)$ which itself is negative (since $f(R)/R > 1$). On the other hand, it approaches $\ln(R/p_0)$ (or $+\infty$) as $u_n$ approaches $+\infty$. In particular, there exists a unique $u_n$ such that $\partial G^*(u_1, \dots, u_{N-1})/\partial u_n = 0$.

The existence of a unique critical point for $G^*$, as a function of $u_n$ (alone), is now guaranteed.

It is important to note at this point that the existence of a 'global' critical point $(u_1^*, \dots, u_{N-1}^*)$ is not yet settled: the one-dimensional critical points need not be realised as a single global critical point.

**Proposition 1.** *If $G^*$ has a critical point $(u_1^*, \dots, u_{N-1}^*)$ then it must satisfy (2) and $y_N$ must satisfy (1). Conversely, if (1) has a solution then $G^*$ has a critical point of the form (2).*

*Proof.* Assume now that a (single) global critical point exists, and introduce the function $h = \gamma^{-1} \circ g \circ \gamma$. Note that, for $v \in (0, +\infty)$, $\gamma(v) \in (-\infty, \ln R)$, $g[\gamma(v)] \in (0, \ln \omega)$, and $\gamma^{-1}(g[\gamma(v)])$ is well defined. Lemma 3 gives a precise relationship between $\gamma = (g^*)'$, its inverse, and $g'$. Observe that (4) yields a nonlinear recursion for $v_n = u_n^*/u_{n-1}^*$: $v_n = h(v_{n+1})$. Write $h_n$ for the $n$th iterate of $h$ and $g_n$ for the $n$th iterate of $g$. Then, for any $n = 1, \dots, N-1$,

$$v_n = h_{N-n}(v_N) = (\gamma^{-1} \circ g_{N-n} \circ \gamma)(v_N).$$

Let $y_N = \mathrm{e}^{\gamma(v_N)}$. Then

$$h_{N-n}(v_N) = \gamma^{-1}(\ln[f_{N-n}(y_N)]) = f_{N-n}(y_N) \frac{f'(f_{N-n}(y_N))}{f_{N-n+1}(y_N)}$$

and

$$\begin{aligned} u_n^* &= \prod_{i=1}^{n} v_i \\ &= \prod_{i=1}^{n} h_{N-i}(v_N) \\ &= \prod_{i=1}^{n} f_{N-i}(y_N) \frac{f'(f_{N-i}(y_N))}{f_{N-i+1}(y_N)} \\ &= \frac{f_{N-n}(y_N)}{f_N(y_N)} \prod_{i=1}^{n} f'(f_{N-i}(y_N)) \\ &= \frac{f_{N-n}(y_N)}{f_N(y_N)} \frac{f_N'(y_N)}{f_{N-n}'(y_N)}. \end{aligned}$$

In particular,

$$B = u_N^* = y_N \frac{f_N'(y_N)}{f_N(y_N)}.$$

Therefore, it follows that if a global critical point exists, it must be of the form stated in Theorem 1, and (1) has at least one solution.

https://doi.org/10.1239/jap/1417528467 Published online by Cambridge University Press

Conversely, if (1) has a solution then the above shows that

$$u_n^* = \frac{f_{N-n}(y_N)}{f_N(y_N)} \frac{f_N'(y_N)}{f_{N-n}'(y_N)}$$

minimises $G^*$. The proof is complete.

### 4.2. Proof Theorem 2

From the definition of $y_N$ in (1), $B = y_N f_N'(y_N)/f_N(y_N) = \mathbb{E}[X_N y_N^{X_N}]/\mathbb{E}[y_N^{X_N}]$. Consider the change of measure $\mathrm{d}\widetilde{\mathbb{P}} = \Lambda_B \,\mathrm{d}\mathbb{P}$ with Radon–Nikodym derivative $\Lambda_B$ as stated in the theorem. The PGF of $Z_n$ under the new measure is

$$\tilde{f}_{X_n}(s) := \widetilde{\mathbb{E}}[s^{X_n}] = \mathbb{E}[\Lambda_B s^{X_n}] = \frac{\mathbb{E}[s^{X_n} y_N^{X_N}]}{f_N(y_N)}.$$

Appealing to the branching property under $\mathbb{P}$,

$$X_N = \sum_{i=1}^{X_n} X_{N-n}^{(i)} \quad \text{for } n < N, \tag{5}$$

where the branching processes $X^{(i)}$ are independent copies started at 1 with PGF $f$ and are also independent of $X_n$. It follows that

$$\tilde{f}_{X_n}(s) = \frac{\mathbb{E}[\mathbb{E}(s^{X_n} y_N^{\sum_{i=1}^{X_n} X_{N-n}^{(i)}}) \mid X_n]}{f_N(y_N)} = \frac{f_n(s f_{N-n}(y_N))}{f_N(y_N)}.$$

It is now easy to see that the mean of the process under the new measure is $u_n^*$:

$$\widetilde{\mathbb{E}}[X_n] = \tilde{f}_{X_n}'(1) \;=\; \frac{f_{N-n}(y)}{f_N(y)} f_n'(f_{N-n}(y)) = u_n^*,$$

where, by the chain rule, $f_n'(f_{N-n}(y)) = f_N'(y)/f_{N-n}'(y)$.

### 4.3. Proof of Theorem 4

The result is proved in [9] under more stringent conditions; it is included here for completeness. First note that the log-moment generating function of a Galton–Watson vector satisfies $\psi_1(\lambda) = g(\lambda)$, and, for $N > 1$,

$$\psi_N(\lambda_1, \ldots, \lambda_N) = g(\lambda_1 + \psi_{N-1}(\lambda_2, \ldots, \lambda_N)). \tag{6}$$

Equation (6) is derived by applying the branching property (5) and routine conditioning on $X_1$. Then (3) is proved by induction.

To emphasise the dependence on $N$, write $\psi_N$ for $G$ and $\psi_N^*$ for $G^*$. Then

$$\begin{aligned}
\psi_N^*(u_1, u_2, \ldots, u_N) &= \sup_{\lambda_1, \lambda_2, \ldots, \lambda_N} \left\{ \sum_{j=1}^N \lambda_j u_j - \psi_N(\lambda_1, \ldots, \lambda_N) \right\} \\
&= \sup_{\lambda_1, \lambda_2, \ldots, \lambda_N} \left\{ \sum_{j=1}^N \lambda_j u_j - g(\lambda_1 + \psi_{N-1}(\lambda_2, \ldots, \lambda_N)) \right\} \\
&= \sup_{\lambda_2, \ldots, \lambda_N} \left\{ \sum_{j=2}^N \lambda_j u_j + \sup_{\lambda_1}\{\lambda_1 u_1 - g(\lambda_1 + \psi_{N-1}(\lambda_2, \ldots, \lambda_N))\} \right\}.
\end{aligned}$$

Changing variables in the inner supremum ($\lambda = \lambda_1 + \psi_{N-1}(\lambda_2, \dots, \lambda_N)$) yields

$$\begin{aligned}
\psi_N^*(u_1, u_2, \dots, u_N) &= \sup_{\lambda_2,\dots,\lambda_N} \left\{ \sum_{j=2}^{N} \lambda_j u_j + \sup_{\lambda} \{\lambda u_1 - \psi_{N-1}(\lambda_2, \dots, \lambda_N) u_1 - g(\lambda)\} \right\} \\
&= \sup_{\lambda_2,\dots,\lambda_N} \left\{ \sum_{j=2}^{N} \lambda_j u_j - \psi_{N-1}(\lambda_2, \dots, \lambda_N) u_1 + g^*(u_1) \right\} \\
&= u_0 g^*\left(\frac{u_1}{u_0}\right) + u_1 \psi_{N-1}^*\left(\frac{u_2}{u_1}, \dots, \frac{u_N}{u_1}\right),
\end{aligned}$$

using $u_0 = 1$. Proceeding by recursion, it follows that

$$\psi_N^*(u_1, \dots, u_N) = \sum_{j=0}^{N} u_j g^*\left(\frac{u_{j+1}}{u_j}\right).$$

## References


[1] Athreya, K. B. (1994). Large deviation rates for branching processes I. Single type case. *Ann. Appl. Prob.* **4,** 779–790.

[2] Athreya, K. B. and Ney, P. E. (2004). *Branching Processes.* Springer, New York.

[3] Biggins, J. D. and Bingham, N. H. (1993). Large deviations in the supercritical branching process. *Adv. Appl. Prob.* **25,** 757–772.

[4] Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett Publishers, Boston, MA.

[5] Harris, T. E. (1963). *The Theory of Branching Processes.* Springer, Berlin.

[6] Kifer, Y. (1990). A discrete-time version of the Wentzell-Freidlin theory. *Ann. Prob.* **18**, 1676–1692.

[7] Klebaner, F. C. and Zeitouni, O. (1994). The exit problem for a class of density-dependent branching systems. *Ann. Appl. Prob.* **4,** 1188–1205.

[8] Klebaner, F. C. and Liptser, R. (2006). Likely path to extinction in simple branching models with large initial population. *J. Appl. Math. Stoch. Anal.*, 23pp.

[9] Klebaner, F. C. and Liptser, R. (2009). Large deviations analysis of extinction in branching models. *Math. Pop. Studies* **15,** 55–69.

[10] Ney, P. E. and Vidyashankar, A. N. (2004). Local limit theory and large deviations for supercritical branching processes. *Ann. Appl. Prob.* **14,** 1135–1166.



KAIS HAMZA, *Monash University*

School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.
Email address: kais.hamza@monash.edu

FIMA C. KLEBANER, *Monash University*

School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.
Email address: fima.klebaner@monash.edu