The Time to Extinction of Branching Processes and Log-convexity: I

By

M. C. Bhattacharjee

Indian Institute of Management Calcutta

and

Florida State University

October, 1986

FSU Technical Report No. M-749

Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

AMS subject classifications. Primary 60J80, Secondary 60K10, 60E99.
The Time to Extinction of Branching Processes
and Log-convexity: I

M. C. BHATTACHARJEE

Indian Institute of Management Calcutta and Florida State University

ABSTRACT

We show that the time to extinction in critical or subcritical Galton-Watson and Markov branching processes has the anti-aging property of log-convex density, and therefore has the decreasing failure rate (DFR) property of reliability theory. Apart from providing new insights into the structure of such extinction time distributions which cannot generally be expressed in a closed form; a consequence of our result is that one can invoke sharp reliability bounds to provide very simple bounds on the tail and other characteristics of the extinction time distribution. The limit distribution of the residual time to extinction in the subcritical case also follows as a direct consequence. A sequel to this paper will further consider the critical case and other ramifications of the log-convexity of the extinction time distribution.

Key Words: Galton-Watson and Markov branching processes, Time to extinction, Reliability Theory, Log-convexity and DFR property.

AMS subject classifications. Primary 60J80, Secondary 60K10, 60E99.
1. Introduction

The time to extinction \( T \) of a critical or subcritical branching process, whose distribution cannot be computed in a closed form except in very special cases, has received the attention of several authors. Since for a large class of such processes, \( \text{var}(T) \) grows exponentially faster than \( T \) as the mean progeny size \( m \to 1 \); previous efforts concentrated either on asymptotic results or on finding appropriate bounds on the distribution \( T \) and some of its parameters such as the mean and the percentiles. Not much is known about the underlying probabilistic structure of the distribution of the time to extinction. In this paper we consider this question for Galton-Watson and Markov branching processes.

For Galton-Watson processes; among the earliest results not restricted to specific offspring distributions, the first is due to Kolmogoroff (see, Athreya and Ney (1972)) in the critical case \( (m = 1) \) which asymptotically estimates \( P(T > n) \) when the offspring distribution has a finite variance. Heathcote and Seneta (1966) provided bounds for \( ET \) and \( \mu = m^n/P(T > n) \) in the subcritical case \( (m < 1) \) with finite offspring variance. Pollak (1971) provided additional and sometimes improved estimates for \( ET \) and \( \mu \); but the scope of his results is restricted to offspring p.g.f.s. satisfying certain inequalities involving three derivatives of the p.g.f. Other results include those by Seneta (1967) and Pollak (1969) for the special case of Poisson progeny — useful in genetic applications. Agresti (1974) finds bounds for \( P(T \leq n) \), \( ET \) and the percentiles in the subcritical case: results which apply to all Galton-Watson processes with finite offspring variance and remain good for \( m \approx 1 \), of interest in many applications. The underlying method of attack here, following an approach due to Seneta (1967), has been to sandwich the offspring p.g.f. by suitably chosen ‘fractional linear generating functions’; \( (f. l. g. f. s.) \), a feature preserved under repeated self-compositions of p.g.f.s. Comparatively less is known in the

We prove that for any subcritical or critical Galton-Watson or Markov branching process, the time to extinction has a *decreasing failure rate* (DFR) distribution. Indeed it satisfies the stronger property that the probability of extinction at the $n$th generation (the probability density of the time to extinction in the Markov branching case) is log-convex in $n$ (log-convex). For supercritical processes, the result remains true for an appropriately modified r. v. Such a finding has several pay-offs. First, it sheds light on the underlying probabilistic structure of the distribution of the time to extinction. Second, it shows how a large variety of DFR distributions can arise in ‘natural’ settings. Finally it yields new dividends in the form of sharp bounds for the tail, the moments and other features of the distribution of $T$, which we obtain by exploiting the log-convexity.

In Sections 2 and 3 we study the time to extinction in the discrete case. Section 4 considers the case of continuous time Markov branching processes. In a subsequent paper, we shall consider further application of these ideas. Our methods are relatively simple and in essence consist of superimposing some relevent ideas from reliability theory on the framework of branching processes.

2. Reliability Characteristics of the Time to Extinction: Galton-Watson Processes

Formally, let $\{Z_n; n = 0, 1, 2, \ldots\}$ be a Galton-Watson process defined on a probability space $(\Omega, \mathcal{F}, P)$ with an offspring distribution $\{p_j; j = 0, 1, 2, \ldots\}$, $p_j \geq 0$, $\sum_j p_j = 1$ with p. g. f.

\[
(2-1) \quad f(s) = \sum_{j=0}^{\infty} p_j s^j, \quad 0 \leq s \leq 1.
\]

Assume, as is usual, that we start with a single ancestor ($Z_0 = 1$ w. p. 1) so that $Z_1$ is the ‘offspring’ variable; and further that $p_0 + p_1 < 1$, $p_j \neq 1$, any $j$,
to avoid inessential trivialities. The conditioning $Z_0 = 1$ will remain suppressed throughout. Notationally we follow Athreya and Ney (1972), hereafter referred to as AN.

Thus $f(s) = E(s^{Z_1})$ and if $f_n(s) =: E(s^{Z_n})$ is the p. g. f. of the $n$th generation; then by conditioning on $Z_1$ it is standard that $f_n(s), n \geq 1 (f_1 \equiv f)$ are the iterates of the offspring p. g. f. $f(s), 0 \leq s \leq 1$:

$$f_n(s) = f(f_{n-1}(s)) = f_{n-1}(f(s)).$$

Let $m = EZ_1$ and $\sigma^2 = \text{var} Z_1$ when the latter exists and let

$$T = \inf\{n : Z_n = 0\}$$

be the time to extinction. If the process is subcritical ($m < 1$) or critical ($m = 1$), then $T < \infty$ w. p. 1. The classical results on the distribution of $T$ (see AN) are:

a) subcritical case ($m < 1$)

$$(2-2) \quad (i) \quad P(T > n) \sim m^n Q(0),$$

where $Q(s)$ on $[0, 1)$ is the unique solution of

$$Q(f(s)) = mQ(s)$$

satisfying $Q(1) = 0, Q'(s) \to 1$ as $s \to 1$.

(ii) $\theta < -\log m \implies E(e^{\theta T}) < \infty$

so that $T$ has moments of all orders.

b) critical case ($m = 1$)

$$(2-3) \quad \sigma^2 \equiv \text{var} Z_1 < \infty \implies P(T > n) \sim 2/(n\sigma^2).$$
Seneta (1967) demonstrated that

\[ m = 1 \implies ET \leq \int_0^1 \frac{1-s}{f(s)-s} ds, \]

so that ET can be finite even in the critical case. This bound remains valid for the subcritical case \( m < 1 \).

Setting \( u_n =: f_n(0) \) which \( \uparrow \) with \( n \); we see the sequence \( u_n \) satisfies the recurrence relation

\[ u_n = f(u_{n-1}), \quad n \geq 1 \]

with \( u_0 =: 0 \). Thus the time \( T \) to extinction has distribution

\[ P(T > n) = P(Z_n > 0) = 1 - u_n, \tag{2-4} \]

and hence the mass function,

\[ q_n =: P(T = n) = u_n - u_{n-1} = \Delta u_{n-1}, \]

for \( n \geq 1 \). To motivate our main result (Theorem 2.3), it is instructive to look first at the failure theoretic properties of \( T \) which rules out the possibility of any 'aging' behaviour. The failure rate function \( r_n \) of the discrete r. v. \( T \) is defined by the conditional probabilities

\[ r_n =: P(T = n | T \geq n) = \frac{u_n - u_{n-1}}{1-u_{n-1}}, \quad n \geq 1. \tag{2-5} \]

**Lemma 2.1.** As \( n \to \infty \),

\[ m < 1 \implies r_n \to 1 - m \]

\[ m = 1 \text{ and } \sigma^2 < \infty \implies r_n \to 0. \]

This implies \( T \) cannot be IFR; viz., if \( m < 1 \), then \( r_1 = u_1 = f(0) = p_0 \geq 1 - m = \lim r_n \), since \( m = \sum_{j=1}^\infty j p_j \geq \sum_{j=1}^\infty p_j = 1 - p_0 \); while in the critical case
the lemma implies that we can find \( n \) and \( N > n \) such that \( r_n > \epsilon > r_N \) for all sufficiently small \( \epsilon \).

We postpone the proof of Lemma 2.1 which is of independent interest and has other consequences (Corollary 3.1) to Section 3. For the present purpose of investigating the 'reliability characteristics' of \( T \), we next note that it is possible to strengthen the conclusion that \( T \) cannot be IFR-aging to

**Lemma 2.2.** The distribution of \( T \) cannot be NBUE-aging in the strict sense.

*Proof. i) subcritical case \( (m < 1) \). From convexity of the p. g. f. \( f(s) \) we have (see e.g., AN, p. 39),

\[
(1 - s)(1 - p_0)^n \leq 1 - f_n(s) \leq m^n, \quad n \geq 1
\]

which, as \( s \to 0 \), yields

\[
(1 - p_0)^n \leq P(T > n) \leq m^n, \quad n \geq 0
\]

so that in particular

\[
(2-6) \quad p_0^{-1}(1 - p_0)^n \leq \sum_{k=n}^{\infty} P(T > k) \leq (1 - m)^{-1}m^n, \quad n \geq 0.
\]

At \( n = 0 \), this yields

\[
(2-7) \quad p_0^{-1} \leq ET \leq (1 - m)^{-1}, \quad \text{if} \quad m < 1.
\]

Hence,

\[
\sum_{k=1}^{\infty} P(T > k) = ET - 1 \geq ET \cdot P(T > 1),
\]

since \( P(T > 1) = 1 - P(Z_1 = 0) = 1 - p_0 \leq 1 - (ET)^{-1} \) by (2-7). Thus

\[
E(T - 1|T > 1) \not\leq ET
\]

6
contradicting the strict NBUE property.

ii) Critical case \((m = 1, \sigma^2 < \infty)\). If \(m = 1\) and \(ET = \infty\), \(T\) obviously cannot be NBUE (which requires a finite mean) and is trivially NWUE. If \(m = 1\) and \(ET < \infty\), then under \(\sigma^2 < \infty\) we show that the NBUE property

\[
\sum_{k=n}^{\infty} P(T > k) \leq (ET) P(T > n)
\]

cannot hold for all \(n\).

From the standard result (2.3) for critical Galton-Watson processes, we have

\[
(2-8) \quad n(ET) P(T > n) \rightarrow 2ET/\sigma^2 < 2N/\sigma^2,
\]

choosing \(N\) large enough so that \(ET < N\), possible since \(ET\) is finite. Note

\[
n \sum_{k=n}^{\infty} P(T > k) > n \sum_{k=n}^{n+N} P(T > k) \geq n \sum_{k=n}^{n+N} \frac{k}{n+N} P(T > k)
\]

\[
(2-9) \quad = \frac{n}{n+N} \sum_{k=n}^{n+N} kP(T > k) \rightarrow 2N/\sigma^2
\]

as \(n \rightarrow \infty\). Combining (2-8) and (2-9) together shows that for all sufficiently large \(n\),

\[
\sum_{k=n}^{\infty} P(T > k) > ET \cdot P(T > n)
\]

so that \(T\) cannot be NBUE. \(<>\)

We now prove our main result.

**Theorem 2.3.** The time to extinction \(T\) in a subcritical or critical Galton-Watson Process has the anti-aging property:

\[
P(T = n) \text{ is log--convex in } n.
\]
Hence $T$ is DFR.

Proof: We have

$$u_n - u_{n-1} = f(u_{n-1}) - f(u_{n-2}) = (u_{n-1} - u_{n-2}) f'(\delta_{n-2}),$$

for some $\delta_{n-2}$ satisfying $u_{n-2} < \delta_{n-2} < u_{n-1}$. Thus,

$$q_n \equiv P(T = n) = \Delta u_{n-1} = \Delta u_{n-2} \cdot f'(\delta_{n-2})$$

$$= q_{n-1} f'(\delta_{n-2}), \quad n \geq 2,$$

A fortiori, by a repeated application:

$$q_{n+1} = q_n f'(\delta_{n-1}) = q_{n-1} f'(\delta_{n-2}) f'(\delta_{n-1}), \quad n \geq 1$$

where $u_{n-1} < \delta_{n-1} < u_n$. This implies

$$q_{n+1} q_{n-1} - q_n^2 = q_{n-1} f'(\delta_{n-2}) (f'(\delta_{n-1}) - f'(\delta_{n-2})) \geq 0$$

for all $n \geq 2$, since $f(s)$ is $\uparrow$ on $[0,1]$, and $u_n \uparrow$ in $n$ implies $\delta_n \uparrow$ in $n$. Thus $P(T = n)$ is log-convex, i.e., $T$ has the discrete version of the 'log-convex' density property.

Since log-convexity is preserved under tail summation (Keilson (1979)), $P(T > n)$ is also log-convex, i.e., $T$ has the DFR property. $<>$

What can we say if extinction is not certain? Since in the spirit of (2-10), for any Galton-Watson process with a probability of extinction $q \in (0,1]$,

$$u_{n+1} - u_n = (u_n - u_{n-1}) f'(\delta_{n-1}), \quad n \geq 1$$

with $u_{n-1} < \delta_{n-1} < u_n$; we see $\delta_n \uparrow q$ since $u_n \uparrow q$ as $n \uparrow \infty$. This implies that for each $n$, $f'(\delta_n) \leq \gamma =: f'(q) \leq 1$. Thus, $u_{n+1} - u_n \leq u_n - u_{n-1}$; i.e., $u_n$ is
concave for all Galton-Watson processes, suggesting the possible log-convexity of $P(Z_n > 0) = 1 - u_n$ in general.

In the supercritical case, the anti-aging character of the time to extinction in the sense of Theorem 2.4 in fact remains true with an appropriate modification. This can be achieved either by shifting the mass at infinity to a mass at zero, or by conditioning on extinction. Let $q \in (0,1)$ be the probability of extinction given $EZ_1 = m > 1$ and let

$$Y = T1_{T < \infty} = T \quad \text{if} \quad T < \infty$$

$$0 \quad \text{if} \quad T = \infty.$$ 

COROLLARY 2.4. In the supercritical case with $q \neq 0$, both $Y$ and $T|T < \infty$ have log-convex probabilities and thus have the DFR property.

**Proof.** Conditioned on extinction, we get

$$P(T > n|T < \infty) = \left\{P(T < \infty) - P(T \leq n)\right\}/P(T < \infty)$$

$$= 1 - \frac{(u_n/q)}{1 - \frac{u_n}{q}}$$

so that

$$P(T = n|T < \infty) = \frac{(u_n - u_{n-1})}{q}, \quad n \geq 1$$

similarly for the r.v. $X$ in (2-12), $P(Y = 0) = 1 - q$, $P(Y = n) = qP(T > n|T < \infty) = (u_n - u_{n-1})$, $n \geq 1$. The proposition now follows from the log-convexity of $(u_n - u_{n-1})$ in Theorem 2.4. 

**Remark.** For the supercritical process $Z_n$ with improper time to extinction $T$, let $B = \{Z_n = 0 \text{ some } n \geq 1\}$. It is easily seen that $T|T < \infty$ is equivalent
in distribution to the time to extinction $T^*$ of the induced subcritical process $Z^*_n$ defined by the p. g. f.

$$f^*(s) = q^{-1} f(qs), \quad 0 \leq s \leq 1$$

the induced subcritical process $Z^*_n$ being equivalent to the original process $Z_n$ relativized to the set of extinction $B$.

3. Applications of the Anti-aging Property of $T$ in the Discrete Case

The log-convexity of $P(T = n)$ and $P(T > n)$ can be exploited to develop new results including bounds on the tail, parameters such as the mean time to extinction, percentiles, movements and other associated characteristics of the distribution of $T$.

3.1. The Failure Rate of the Time to Extinction

We begin by looking at the asymptotic behavior of $P(Z_n = 0 | Z_{n-1} > 0) = P(T = n | T > n - 1)$, the failure rate of $T$ defined in (2-5).

Proof of Lemma 2.1. The probability of extinction of the immediately next generation, given the process was alive in the $(n-1)$st. generation, is

$$r_n = P(T = n | T \geq n) = \frac{u_n - u_{n-1}}{1 - u_{n-1}}, \quad n \geq 1.$$

This claim follows by noting that, in the critical case ($m = 1$) with finite offspring variance $\sigma^2$, Kolmogoroff's asymptotic estimate (2.3) implies

$$1 - r_n = \frac{n P(T > n)}{nP(T > n - 1)} \xrightarrow{2/\sigma^2} 1;$$

while in the subcritical case ($m < 1$),

$$1 - r_n = P(T > n | T \geq n) = \frac{1 - u_n}{1 - u_{n-1}} \rightarrow m$$
by (2.2).

A more direct proof in the subcritical case is as follows.

Letting $P_n(i,j)$ denote the $n$-step transition probabilities, we have

$$1 - r_n = \frac{1 - u_n}{1 - u_{n-1}} = \frac{\sum_{j \neq 0} P_n(1,j)}{\sum_{j \neq 0} P_{n-1}(1,j)} = \frac{P_n(1,1)}{P_{n-1}(1,1)} \frac{\sum_{j \neq 0} \{P_n(1,j)/P_n(1,1)\}}{\sum_{j \neq 0} \{P_{n-1}(1,j)/P_{n-1}(1,1)\}} \to m$$

(3-1) as $n \to \infty$, since by the 'monotone ratio theorem' (AN, Lemma 2 and Theorem 2, p. 12-13), in the subcritical case,

$$P_n(1,1)/P_{n-1}(1,1) \to \gamma =: f'(q) \equiv f'(1) = m < 1$$

while,

$$P_n(1,j)/P_n(1,1) \to \pi_j, \quad j \geq 1$$

the stationary measure associated with the subcritical process under $p_1 \neq 0$, and $\sum_j \pi_j$ converges for $m < 1$.  

As an immediate consequence of Lemma 2.1, we get

**Corollary 3.1.** In the subcritical case, the residual time to extinction converges in distribution to a geometric distribution with mean $(1 - m)^{-1}$.

**Proof.** The residual time to extinction $T - n | T > n$ has tail

$$P(T - n > k | T > n) = \frac{P(T > n + k)}{P(t > n)} = \frac{1 - u_{n+k}}{1 - u_n}$$

$$= \prod_{j=0}^{k-1} \left( \frac{1 - u_{n+j+1}}{1 - u_{n+j}} \right) \to m^k, \quad k \geq 0$$

as $n \to \infty$, by repeatedly applying (3-1). Thus

$$P(T - n = k | T > n) \to (1 - m) m^{k-1}; \quad k = 1, 2, \ldots$$
as \( n \to \infty \).

\[ \xi_p = \inf \{ j : P(T > j) \leq 1 - p \} \]

be the \( p \)th percentile of the time to extinction, \( 0 < p < 1 \) and let

\[ \log \theta(p) =: \xi_p^{-1} \log(1 - p). \]

**Theorem 3.2.** (i) \( P(T > n) \leq (\geq) \theta^n(p) \), if \( n \leq (\geq) \xi_p \).

(ii) In the subcritical case,

\[ P(T > n) \leq m^n, \]
\[ P(T = n) \leq p_0 m^{n-1}. \]

**Proof.** (i) \( P^{1/n}(T > n) \) increases in \( n \), since \( T \) is DFR. Hence

\[ n < \xi_p \implies \left\{ P(T > n) \right\}^{1/n} \leq \left\{ P(T > \xi_p) \right\}^{1/\xi_p} = (1 - p)^{-1} \theta(p). \]

For \( n > \xi_p \), the inequality is reserved.

(ii) \[ P(T > n) = 1 - u_n = \prod_{j=0}^{n-1} (1 - r_j) \leq m^n, \quad n \geq 0, \]

since the failure rate \( r_n \) is decreasing and \( r_n \to 1 - m \) by Lemma 2.1. To bound \( P(T = n) \equiv q_n \), note that since \( q_n \) is log-convex (Theorem 2.4), using (2-11) we get

\[ \frac{q_{n+1}}{q_n} = f(t(\xi_{n-1})) \uparrow f(t(1)) = m. \]
Thus

\[ (T = n) \equiv q_n = q_1 \prod_{j=1}^{n-1} (q_{j+1}/q_j) \leq p_0^{n-1}, \quad n \geq 1 \]

since \( q_1 = p_0 \).

Let \( \eta_T \) be the coefficient of variation of the time to extinction.

**Theorem 3.3.** For any Galton-Watson process which is either subcritical, or critical with \( ET < \infty \), we have

(i) \( \xi_p \geq \{\log(1 - p)\}/\{\log(1 - \frac{1-p}{ET})\}, \quad 0 < p < 1 \)

(ii) \( \text{Var } T \geq ET(ET - 1) \)

(iii) \( \eta_T \geq 1 - (ET)^{-1} \)

(iv) \( ET^k \geq k!(ET)(ET - 1)^{k-1}, \quad k = 1, 2, \ldots \)

(v) the p. g. f. of \( T \) satisfies

\[ E(s^T) \geq s/(s + (1-s)ET), \quad 0 \leq s \leq 1 \]

The bounds in (ii)-(v) are sharp in the subcritical case.

(vi) \( T \) is infinitely divisible.

The next Lemma, instrumental in proving (ii)-(v), is a variant of a result in Reliability Theory (see e. g. Barlow and Proschan (1975), p. 112). For any r. v. \( X \) defined on the (non-negative/positive) integers with \( EX < \infty \), let \( X^* \) be geometric on the same domain with \( EX^* = EX \) and let \( \tilde{X} \) denote the induced r. v. with distribution

\[ P(\tilde{X} = n) = P(X > n)/EX, \quad n \geq 0. \]

Thus, corresponding to the extinction time \( T \), \( T^* \) is a geometric r. v. on the positive integers with

\[ P(T^* = n) = \alpha(1 - \alpha)^{n-1}, \quad n \geq 1 \]

such that \( ET^* = ET \); then \( \alpha = (ET)^{-1} \).
**Lemma 3.4.** Suppose $ET < \infty$. Then for any non-negative function $h$ on the integers,

$$Eh(\tilde{T}) \geq Eh(\tilde{T}^*) .$$

**Proof.** Since $P(T > n)$ is log-convex while $P(T^* > n) = (1 - (ET)^{-1})^n$ is log-linear, the former sequence crosses the latter from below, i.e., there exists an $n_0$ such that $c_n := P(T > n) - P(T^* > n)$ satisfies $c_n \leq 0$ if $n \leq n_0$ while $c_n > 0$ if $n > n_0$. Since $ET = ET^*$, we have $\sum_n c_n = 0$; so that

$$\sum_{n=0}^{\infty} h(n) \{P(T > n) - P(T^* > n)\}$$

$$= \left( \sum_{n \leq n_0} + \sum_{n > n_0} \right) c_n \{h(n) - h(n_0)\} \geq 0 ;$$

since the summand, being product of two factors of the same sign, is always non-negative. Hence

$$E\{h(\tilde{T}) - h(\tilde{T}^*)\}$$

$$= (ET)^{-1} \sum_{n=0}^{\infty} h(n) \{P(T > n) - P(T^* > n)\} \geq 0. \quad <>$$

**Proof of Theorem 3.3.** (i) Use Theorem 3.2 to get

$$ET > \sum_{n \geq \xi_p} P(T > n) \geq \{1 - \theta(p)\}^{-1} \theta(\xi_p)(p) \cdot \frac{1 - p}{1 - \theta(p)}$$

i.e.,

$$1 - (1 - p)^{\xi_p} \equiv 1 - \theta(p) \geq (1 - p)/ET$$

which yields the desired inequality.
(ii) For any integer valued r. v. $X$, using (3-2) check that

$$EX \cdot E\tilde{X} = \sum_{n=1}^{\infty} nP(X > n)$$

$$= \frac{1}{2} \sum_{n=2}^{\infty} n(n-1)P(X = n)$$

$$= \frac{1}{2} EX(X - 1).$$

(3-3)

Hence, for the time to extinction $T$, using Lemma 3.3 and recalling $ET^* = ET$, we have

$$ET(T - 1) = 2ET \cdot E\tilde{T} \geq 2ET \cdot E\tilde{T}^* = ET^*(T^* - 1).$$

This is equivalent to both $ET^2 \geq ET^{2*}$ and var $T \geq$ var $T^*$, again in virtue of the fact that $ET^* \geq ET$ by construction. Since $T^*$ is geometric on $\{1, 2, \cdots\}$ with parameter $\alpha = 1/ET$,

$$\text{var } T^* = \frac{(1 - \alpha)}{\alpha^2} = ET(ET - 1).$$

This proves (ii). Note that (ii) implies $ET^2 \geq 2(ET)^2 - ET$ and hence,

$$\eta^2_T = \frac{\text{var } T}{(ET)^2} \geq \frac{\text{var } T^*}{(ET)^2} = \eta^2_{T^*} = 1 - \alpha \equiv 1 - \frac{1}{ET}.$$ 

In particular, for the subcritical case, using the bound (2-7), we get

$$\eta^2_T \geq 1 - p_0$$

(3-4)

(iv) follows by arguments similar to (ii) which can be used to prove that there exist $b_k$ such that

$$ET^{(k)} \geq b_k^{-1}(ET)(ET^*^{(k-1)}) = ET^{* (k)}$$

$$= k \frac{(1 - \alpha)^{k-1}}{\alpha^k} = k!(ET)(ET - 1)^{k-1}, \quad k \geq 1.$$ 

(3-5)
The asymmetry in this bound is due to the fact that the support of $T$ is the set of positive integers and not \( \{0,1,2,\cdots\} \). Considering the r. v. $T_o := (T - 1)$, we can get the more symmetric result

\begin{equation}
ET_o^{(k)} \geq k!(ET_o)^k \quad k = 1, 2, \cdots.
\end{equation}

(v) If $\psi(s) := E(s^T), 0 \leq s \leq 1$ denotes the p. g. f. of $T$, choosing $1 - h(n) = s^n$ in Lemma 3.3,

\[
\frac{1 - \psi(s)}{1 - s} = \sum_{n=0}^{\infty} s^n P(T > n) \leq \sum_{n=0}^{\infty} s^n P(T^* > n) = \frac{1 - E(s^{T^*})}{1 - s}
\]

i.e.

\[
\psi(s) \geq E(s^{T^*}) = s\alpha / \{1 - s(1 - \alpha)\} = s / \{s + (1 - s)ET\}
\]

since $ET = \alpha^{-1}$.

The bounds in (ii)-(v) are sharp in the subcritical case. This can be seen by taking $f(s) = p + (1 - p)s$, the Bernoulli offspring distribution for which the extinction time is geometric with $ET = p^{-1}$.

(vi) Recalling the log-convexity of $q_n \equiv P(T = n), n \geq 1$; namely

\[
q_n^2 \leq q_{n+1}q_{n-1}, \quad n \geq 2;
\]

we further note that

\[
q_1 = P(T = 1) = p_0 > 0,
\]

and

\[
q_2 = P(T = 2) = E(p_0^{Z_1}|Z_1 > 0) > 0,
\]

since $f(p_0) = p_0 + (1 - p_0)E(p_0^{Z_1}|Z_1 > 0)$ and $f(p_0) > p_0 \in (0, 1)$. By a sufficient condition of Warde and Katti (1971), we conclude the infinite divisibility of $T$. <
Using known bounds for $ET$ in our Theorem 3.3 provides even simpler and more readily computable bounds on the various parameters of the distribution of $T$. The bound in (3-4) is one such. Similarly, the p. g. f. of the time to extinction is bounded by

$$E(s^T) \geq s(1-m)/(1-sm)$$

if $m < 1$,

$$s/(s + (1-s) \int_0^1 \frac{1-z}{f(z) - z} dz)$$

if $m = m_1$,

using (2-7) and Seneta's (1967) upperbound for $ET$ in the critical case.

4. The Markov Branching Case

For a Markov branching process $\{Z(t); t \geq 0\}$ with an offspring p. g. f. $f(s)$ as in (2-1), the time to extinction

$$(4-1) \quad T = \inf \{t > 0 : Z(t) = 0\},$$

is honest whenever $f'(1) \equiv m \leq 1$. Narayan (1981) derived $E(T|T < \infty)$ for any Markov branching process and used this to show that

$$ET \leq \frac{1}{2a} \left( \frac{1}{p_0} + \frac{1}{1-m} \right)$$

in the subcritical case and is achieved by the Bernoulli offspring distribution ($a^{-1} = \text{mean life of particles}$). If $F(s,t)$ is the p. g. f. of $Z(t)$, then in the subcritical case we also have the following asymptotic estimate (Ansmussen and Hering (1983)) of the tail of $T$:

$$(4-2) \quad P(T > t) = 1 - F(0,t) \sim e^{\lambda t} L(e^{\lambda t})$$

as $t \to \infty$, for some $L$ slowly varying at 0; where $\lambda = a(m-1) < 0$.

Besides these results, not much else appears to be known about the distribution of the time to extinction of Markov branching processes. We show that
the anti-aging structure of the time to extinction in Theorem 2.3 remains true
for Markov branching processes.

**Theorem 4.1.** For a subcritical or critical Markov branching process with a
single ancestor, the time to extinction $T$ has a log-convex density and hence is
DFR.

*Proof.* The offspring distribution has a p.g.f. $f(s)$ with mean $m = f'(1) \leq 1$. Given $Z(0) = 1$ w. p. 1 let

$$(4-3) \quad P_{ij}(t) = P\{Z(t) = j | Z(0) = i\}$$

be the transition probability function of the Markov branching process \{Z(t); t \geq 0\} starting with $Z(0) = 1$. We omit the conditioning $Z(0) = 1$ as being understood throughout. Let

$$(4-4) \quad F(s, t) = E s^{Z(t)} = \sum_{j=0}^{\infty} P_{1j}(t) s^j, \quad 0 \leq s \leq 1$$

be the p.g.f. of $Z(t)$. Then $T = \inf\{t > 0 : Z(t) = 0\}$ has tail

$$P(T > t) = 1 - P(Z(t) = 0) = 1 - F(0, t),$$

and hence a density

$$g(t) = F'(0, t) = P_{11}(t),$$

where $F'(s, t) =: \frac{\partial}{\partial s} F(s, t)$. Now, $F(s, t)$ satisfies the well-known functional equation

$$(4-5) \quad F(s, t + u) = F(F(s, t), u)$$

viz.,

$$F(s, t + u) = E\{E(s^{Z(t+u)}|Z(u))\} = E(\{F(s, t)\}^{Z(u)}): \text{ by the Markov property}$$

$$= F(F(s, t), u).$$

18
Hence differentiating (4-5) with respect to $s$, we get

$$F'(s, t + u) = F'(F(s, t), u)F'(s, t).$$

Setting $s = 0$, we get

$$P_{11}(t + u) = F(0, t + u) = F'(F(0, t), u)F'(0, t) = F'(P(T \leq t), u)P_{11}(t).$$

This implies

$$(4-6) \quad \frac{P_{11}(t + u)}{P_{11}(t)} = F'(P(T \leq t), u) \uparrow \text{ in } t, \quad u > 0,$$

since, $F'(s, u) \uparrow$ in $s$ for each $u$ and $P(T \leq t) \uparrow$ in $t$. Thus $T$ has log-convex density, while log-convex density $\implies$ DFR property. \hspace{1cm} <>

Note that in contrast with Heathcote and Seneta (1966), Pollak (1971) and Agresti (1974), we did not have to make an assumption of finite offspring-variance ($f'(1) < \infty$) in either Theorem 2.3 or Theorem 4.1. Consequently all direct implications of the anti-aging property of the distribution of $T$ will also be free of this assumption.

It is also instructive to contrast our results on the time to extinction for Galton-Watson and Markov branching processes to some result of Assaf, Shaked and Shantikumar (1985) who investigate a different but a somewhat similar problem of first passage times in a finite state continuous time Markov process which can be uniformized such that the embedded Markov chain has a $TP_r$ (totally positive of order $r$) transition matrix. They show that certain first passage times for such processes with $TP_r$ transition matrix has a $PF_r$ (Poyla frequency function of order $r$) density. Note that our result of $T$ having the anti-aging property of log-convex density – cannot be derived as an application of the results of Assaf, Shaked and Shantikumar (ibid.); also we do not make any
assumption on the offspring distribution which would imply a total positivity property of the transition matrix of the Markov chain determined by a Markov branching process.

All reliability theoretic consequences of log-convex density and hence the implied DFR property (see e.g., Barlow and Proschan (1965), (1975)) are immediately applicable to the distribution of the time to extinction, when $m \leq 1$. We thus have for example:

\[(i) \quad P(T > t) \leq e^{t/ET}, \quad \text{if} \quad t < ET, \]
\[
(ET)e^{-1}/t, \quad \text{if} \quad t > ET.
\]

In the critical case, the bound is non-trivial only if $ET < \infty$

\[(ii) \quad ET^r \geq (\leq) r!(ET)^r, \quad r \geq (\leq) 1.
\]

In particular, the coefficient of variation $\eta_T$ satisfies

\[\eta_T \geq 1\]

\[(iii) \quad \int_0^\infty e^{-\theta t}dP(T \leq t) \geq (1 + \theta \cdot ET)^{-1}, \quad \theta > 0\]

\[(iv) \quad \int_0^\infty e^{-\theta t} \phi(t) P(T > t)dt \geq \int_0^\infty e^{-(\theta + \frac{\phi^2}{t})t} \phi(t)dt,\]

for all non-decreasing $\phi$ on $(0, \infty)$.

These results are parallel to Theorem 3.2 in the discrete case. The bounds are sharp, being achieved by the Bernoulli offspring distribution $f(s) = p + (1 - p)s$.

Using the reliability characteristics of the distribution of $T$, in the spirit of Barlow and Proschan (Theorem 4.6, p. 30, 1965), we also get the following bounds for the percentiles of the time to extinction.
COROLLARY 4.2. For critical (with $ET < \infty$) or subcritical Markov branching processes,

(i) $\xi_p \leq \{-\frac{\ln(1-p)}{1-p}\}ET \quad 0 < p < 1$, 

If $p \leq 1 - e^{-1}$, this can be tightened to the sharp bound

\begin{equation}
(4-7) \quad \xi_p \leq \{-\ln(1 - p)\} \cdot ET.
\end{equation}

In particular, the median time to extinction satisfies

\begin{equation}
(4-8) \quad \xi_{1/2} \leq \ln 2 \int_0^1 \frac{1 - s}{f(s) - s} ds \leq (\ln 2)(1 - m)^{-1}.
\end{equation}

(ii) For all sufficiently large $p$,

$\xi_p \leq (\mu e^{-1})/(1 - p)$.

Proof. Using the fact that $P^{1/t}(T > t)$ is $\uparrow$ as a consequence of the DFR property implied by the log-convex density of $T$,

\[ ET \geq \int_{\xi_p}^{\infty} P(T > t) dt \geq \int_{\xi_p}^{\infty} \left\{ P(T > \xi_p) \right\}^{t/\xi_p} dt = \frac{(1 - p)\xi_p}{\ln(1 - p)} \]

from which the claim follows. Note the DFR property implies $T$ is absolutely continuous on $(0, \infty)$ and further that our time to extinction has no jump at zero; thus $P(T > \xi_p) = 1 - p$. If $0 < p \leq 1 - e^{-1}$, then $P(T > \xi_p) = 1 - p \geq e^{-1} \geq P(T \geq \mu) \implies \xi_p \leq \mu$. This implies

\[ 1 - p = P(T > \xi_p) \leq e^{-\xi_p/ET}, \]

which yields the required tightening of the former bound. This bound is sharp, being achieved by the Bernoulli offspring distribution which results in the exponential time to extinction.
The bound (4-1) obviously applies with \( p = \frac{1}{2} \). Replacing \( ET \) by Seneta's bound and noting \( f(s) \leq 1 - m(1 - s) \) in the subcritical case, by the convexity of the p. g. f. \( f(s) \), we obtain (4-8).

Finally if \( p \in (0,1) \) is sufficiently large, then \( \xi_p > \mu \); a-posteriori,

\[
1 - p = P(T > \xi_p) \leq (ET)e^{-1}/\xi_p.
\]

As another application of the anti-aging property of the time to extinction in the Markov branching case, we have the following counterpart of Corollary 3.1.

**Corollary 4.3.** In subcritical Markov branching processes with mean particle life \( a^{-1} \), the residual time to extinction converges in distribution to an exponential with mean \( \{a(1 - m)\}^{-1} \).

**Proof.** As \( t \to \infty \),

\[
P(T - t > x | T > t) = \frac{1 - F(0, t + x)}{1 - F(0, t)}
\]

\[
\to \lim_{t \to \infty} \frac{F'(0, t + x)}{F'(0, t)} \equiv \lim_{t \to \infty} \frac{P_{11}(t + x)}{P_{11}(t)}.
\]

But, by (4.6),

\[
\frac{P_{11}(t + x)}{P_{11}(t)} = F'(P(T \leq t), x) = \sum_{j=1}^{\infty} jP_{1j}(x)\{P(T \leq t)\}^{j-1}
\]

\[
\to \sum_{j=1}^{\infty} jP_{1j}(x), \quad \text{as } t \to \infty
\]

\[
= E\overline{Z}(x) = e^{-a(1 - m)x},
\]

the last step by the well known result on the expected size of the process at time \( x \). Due to the log-convexity of \( P_{11}(x) \), the limit is in fact monotone:

\[
P(T - t > x | T > t) \uparrow e^{-a(1 - m)x}.
\]
of the density of $T$ is stronger than what is actually necessary to conclude the limiting exponentiality of the residual extinction time. Here is an alternative proof.

**From the asymptotic estimate (4-2) of the tail of $T$, it follows that $T$ is ‘age-smooth’** (Bhattacharjee (1985)); i.e., $P(T > \ln x)$ is $\lambda$-varying with $\lambda = -a(1 - m) < 0$. By Theorem 3.6 in Bhattacharjee (*ibid.*), it follows that $T$ belongs to the residual lifetime domain of attraction of the exponential; i.e.,

$$P(T > t + \frac{x}{a(1 - m)}|T > t) \rightarrow e^{-x}$$

as $t \rightarrow \infty$. This follows by using the representation (4-2) in computing the left hand side above. Note, as a consequence, we must then have

$$e(t) =: E(T - t|T > t),$$

the mean residual time to extinction, also as a possible norming factor. Thus

$$P(T > t + xe(t)|T > t) \rightarrow e^{-x},$$

by the results of Balkema and deHaan (1974), and that

$$e(t) \sim \{a(1 - m)\}^{-1}$$

(Theorem 2.2, Bhattacharjee (1985)).

Since $P_{11}(t + x) \sim P_{11}(t)$ when $m = 1$, Corollary 4.3 in its present form and its direct-analog in discrete time breaks down in the critical case. We will consider the asymptotic behavior in this case and other ramifications of the distribution of extinction time in non-supercritical processes in a sequel to this article.
Moshe Shaked and George Shantikumar (1986), after being appraised by the author in a private communication of our main results (Theorems 2.3 and 4.1), have recently given an alternative proof using likelihood ratio ordering.

Acknowledgement. The author thanks Eugene Seneta for his careful reading and comments on an earlier version of the manuscript which covered the results in Sections 2 and 3.
REFERENCES


