*Bernoulli* **6**(2), 2000, 323–338

# Change of measures for Markov chains and the LlogL theorem for branching processes

KRISHNA B. ATHREYA

*Departments of Mathematics and Statistics, Iowa State University, Ames IA 50011, USA.*
*E-mail: athreya@math.iastate.edu*

Let $P(. , .)$ be a probability transition function on a measurable space $(M, \mathbf{M})$. Let $V(.)$ be a strictly positive eigenfunction of $P$ with eigenvalue $\rho > 0$. Let

$$\tilde{P}(x, \mathrm{d}y) \equiv \frac{V(y)P(x, \mathrm{d}y)}{\rho V(x)}.$$

Then $\tilde{P}(. , .)$ is also a transition function. Let $P_x$ and $\tilde{P}_x$ denote respectively the probability distribution of a Markov chain $\{X_j\}_0^\infty$ with $X_0 = x$ and transition functions $P$ and $\tilde{P}$. Conditions for $\tilde{P}_x$ to be dominated by $P_x$ or to be singular with respect to $P_x$ are given in terms of the martingale sequence $W_n \equiv V(X_n)/\rho^n$ and its limit. This is applied to establish an LlogL theorem for supercritical branching processes with an arbitrary type space.

*Keywords:* change of measures; Markov chains; martingales; measure-valued branching processes

## 1. Introduction

Recently Lyons *et al*. (1995) (see also Kurtz *et al*. 1997; Lyons 1997) used a result from measure theory to give a probabilistic proof of the LlogL theorem of Kesten and Stigum (1966) for branching processes in single- and multiple cases. In this paper their techniques are extended to a Markov chain context and then used to prove an LlogL theorem for measure-valued branching processes on a general type space.

## 2. Markov chains

Let $(M, \mathbf{M})$ be a measurable space and $P(. , .)$ be a transition probability function on it. Thus, for each $x$ in $M$, $P(x, .)$ is a probability measure on $\mathbf{M}$ and for each $A$ in $\mathbf{M}$, $P(. , A)$ is an $\mathbf{M}$-measurable function on $M$. Let $v(.)$ be a strictly positive function on $(M, \mathbf{M})$ such that, for some $\rho > 0$,

$$\int v(y)P(x, \mathrm{d}y) = \rho v(x) \qquad \text{for all } x \text{ in } M \tag{1}$$

and

$$\tilde{P}(x, A) \equiv \left(\int_A v(y)P(x, \mathrm{d}y)\right)(\rho v(x))^{-1}. \tag{2}$$

Then $\tilde{P}$ is also a transition function. We exclude the special case when $v(x) \equiv 1$ since in this case $\rho = 1$ and $\tilde{P} = P$.

We now introduce some notation and definitions. Let $\Omega \equiv M^\infty$, the space of all $M$-valued functions on $\{0, 1, 2, \ldots\}$. Let $X_n(\omega) \equiv \omega(n)$, the coordinate projection for $n = 0, 1, 2, \ldots$. Write $F_n \equiv \sigma(X_0, X_1, \ldots, X_n)$, the σ-algebra generated by $X_0, X_1, \ldots, X_n$, $B \equiv \sigma(X_0, X_1, \ldots, X_n, \ldots)$, $W_n \equiv v(X_n)/\rho^n v(X_0)$ and $\pi_n(\omega) \equiv (X_0, X_1, \ldots, X_n)$. Let $P_x$ be the probability measure on $(\Omega, B)$ that with probability one makes $\{X_j\}_0^\infty$ a Markov chain with $X_0 = x$, and transition function $P$, and let $P_{x,n}$ be the restriction of $P_x$ to $F_n$, and $\tilde{P}_x$, $\tilde{P}_{x,n}$ the corresponding quantities with transition function $\tilde{P}$.

Using the obvious notation, we see that

$$\begin{aligned}
\tilde{P}_{x,n}(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \cdots \times \mathrm{d}x_n) &= \tilde{P}(x, \mathrm{d}x_1)\tilde{P}(x_1, \mathrm{d}s_2) \ldots \tilde{P}(x_{n-1}, \mathrm{d}x_n) \\
&= \frac{v(x_1)P(x, \mathrm{d}x_1)}{\rho v(x)} \frac{v(x_2)P(x_1, \mathrm{d}x_2)}{\rho v(x_1)} \cdots \frac{v(X_n)P(x_{n-1}, \mathrm{d}x_n)}{\rho v(x_{n-1})} \\
&= v(x_n)\frac{P(x, \mathrm{d}x_1)P(x_1, \mathrm{d}x_2) \ldots P(x_{n-1}, \mathrm{d}x_n)}{\rho^n v(x)} \\
&= \frac{v(x_n)}{\rho^n v(x)} P_{x,n}(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \cdots \times \mathrm{d}x_n),
\end{aligned}$$

leading to the following proposition.

**Proposition 1.** *For each $n \geqslant 1$, $\tilde{P}_{x,n}$ is dominated by $P_{x,n}$ with the Radon–Nikodym derivative $W_n$.*

Next, using (1) and the Markov property we see that under $P_x$

$$\mathrm{E}(W_{n+1}|F_n) = \int \frac{v(y)P(X_n, \mathrm{d}y)}{\rho^{n+1}v(X_0)} = \frac{\rho v(X_n)}{\rho^{n+1}v(X_0)} = \frac{v(X_n)}{\rho^n v(X_0)} = W_n.$$

Also under $\tilde{P}_x$

$$\begin{aligned}
\tilde{\mathrm{E}}_x(W_{n+1}^{-1}|F_n) &= \tilde{\mathrm{E}}_x\left(\rho^{n+1}\frac{v(X_0)}{v(X_{n+1})}\middle| F_n\right) \\
&= \rho^{n+1}v(X_0)\int \frac{1}{v(y)}\tilde{P}(X_n, \mathrm{d}y) \\
&= \rho^{n+1}v(X_0)\int \frac{v(y)P(X_n, \mathrm{d}y)}{v(y)\rho v(X_n)} \\
&= \frac{\rho^{n+1}v(X_0)}{v(X_n)}\int P(X_n, \mathrm{d}y) \\
&= W_n^{-1}.
\end{aligned}$$

So we have the following proposition.

**Proposition 2.** *Under $P_x$, $\{W_n, F_n\}_0^\infty$ is a non-negative martingale and under $\tilde{P}_x$, $\{W_n^{-1}, F_n\}_0^\infty$ is a non-negative martingale.*

***Remark 1.*** The kernel $\tilde{P}$ defined in (2) is known in the literature as the *tilted kernel* and is a standard tool especially in the study of large deviations. Also, as pointed out by a referee, if we define the space-time Markov chain $Y_n \equiv (X_n, n)$ and set $h(x, n) \equiv \rho^{-n} v(x)$ then $h(\cdot)$ is a harmonic function and hence $W_n \equiv h(Y_n)$ is a martingale. For more information on this see Rogers and Williams (1994).

By the martingale convergence theorem the sequence $W_n$ converges with probability one under $P_x$. Let

$$W(\omega) \equiv \overline{\lim}_n W_n(\omega). \tag{3}$$

Thus $W(\omega)$ is actually the limit of $W_n(\omega)$ on a set of probability one under $P_x$. For any $A \in F_k,\ k < \infty$,

$$\tilde{P}_x(A) = \tilde{P}_{x,k}(A) = \tilde{P}_{x,n}(A), \qquad \text{for } n \geqslant k,$$

$$= \int_A W_n \, \mathrm{d}P_{x,n} = \int_A W_n \, \mathrm{d}P_x.$$

Now fix $k$ and let $n \to \infty$. By Fatou's lemma we have

$$\tilde{P}_x(A) \geqslant \int_A W \, \mathrm{d}P_x. \tag{4}$$

This being true for $A \in F_k$ for any $k$, (4) holds for all $A \in B$. The question as to when equality holds in (4) is answered by the following theorem.

**Theorem 1.** *For all $A \in B$*

$$\tilde{P}_x(A \cap (\mathrm{W} < \infty)) = \int_A W \, \mathrm{d}P_x,$$

*and hence*

$$\tilde{P}_x(A) = \int_A W \, \mathrm{d}P_x + \tilde{P}_x(A \cap (W = \infty)).$$

This theorem is a special case of a more general result in measure theory (Durrett 1996).

**Theorem 2.** *Let $(\Omega, B)$ be a measurable space and $\{F_n\}_0^\infty$ a filtration such that $B = \sigma(\cup_0^\infty F_n)$. Let $\mu$ and $\tilde{\mu}$ be two probability measures such that for each $n$ the restrictions $\mu_n$ and $\tilde{\mu}_n$ of $\mu$ and $\tilde{\mu}$ to $F_n$ respectively are such that $\tilde{\mu}_n$ is dominated by $\mu_n$ with derivative $W_n$. Let $W = \overline{\lim}\, W_n$. Then*

(a) *$\{W_n, F_n\}_0^\infty$ is a martingale under $\mu$ and so $W = \lim_n W_n$ with probability one with respect to $\mu$;*

(b) *for any* $A \in B$,

$$\tilde{\mu}(A) = \int_A W\,\mathrm{d}\mu + \tilde{\mu}(A \cap (\mathrm{W} = \infty));$$

(c) *if* $\tilde{\mu}_{\mathrm{a}}(A) \equiv \int_A W\,\mathrm{d}\mu$ *and* $\tilde{\mu}_{\mathrm{s}}(A) = \tilde{\mu}(A \cap (W = \infty))$, *then* $\tilde{\mu} = \tilde{\mu}_{\mathrm{a}} + \tilde{\mu}_{\mathrm{s}}$ *is the unique Lebesgue–Radon–Nikodym decomposition of* $\tilde{\mu}$ *with respect to* $\mu$.

**Corollary 1.**

(a) $\tilde{\mu}$ *is dominated by* $\mu$ *if and only if* $\int_\Omega W\,\mathrm{d}\mu = 1$ *if and only if* $\tilde{\mu}(W = \infty) = 0$.
(b) $\tilde{\mu}$ *is singular with respect to* $\mu$ *if and only if* $\mu(W = 0) = 1$ *if and only if* $\tilde{\mu}(W = \infty) = 1$.

Thus equality holds in (4) for all $A \in B$ if and only if $\tilde{P}_x$ is dominated by $P_x$ if and only if $\tilde{P}_x(W = \infty) = 0$. Although the proof of Theorem 2 is available in the literature (Durrett 1996, p. 242), a simple proof is given below to make this paper self-contained.

***Proof of Theorem 2.*** (a) For all $A \in \mathscr{F}_n$, $\int_A W_{n+1}\,\mathrm{d}\mu = \tilde{\mu}_{n+1}(A) = \tilde{\mu}_n(A) = \int_A W_n\,\mathrm{d}\mu$ and so under $\mu$, $\mathrm{E}(W_{n+1}|\mathscr{F}_n) = W_n$ with probability one.

(b) Let $M_{k,n}(\omega) \equiv \sup_{k \leqslant j \leqslant n} W_j(\omega)$. Then, for each $k$, $\{M_{k,n}(\omega)\}_{n=k}^{\infty}$ is a non-decreasing sequence whose limit $M_k(\omega)$ is $\sup_{k \leqslant j} W_j(\omega)$. Next, $\{M_k(\omega)\}_{k=1}^{\infty}$ is a non-increasing sequence whose limit is $W(\omega) = \overline{\lim}_n W_n(\omega)$. Now fix $k_0$ and $N < \infty$. Let $A \in \mathscr{F}_{k_0}$. Then for $n \geqslant k \geqslant k_0$, $B_{k,n} \equiv A \cap (M_{k,n} \leqslant N) \in \mathscr{F}_n$ and so

$$\tilde{\mu}(B_{k,n}) = \int_{B_{k,n}} W_n\,\mathrm{d}\mu = \int W_n(\omega) I_{B_{k,n}}(\omega)\,\mathrm{d}\mu. \tag{5a}$$

As $n \to \infty$, $I_{B_{k,n}}(\omega) \to I_{B_k}(\omega)$ for all $\omega$, where $B_k = A \cap (M_k \leqslant N)$. Also under $\mu$, $W_n(\omega) \to W(\omega)$ with probability one. So, by the bounded convergence theorem (applied to both sides of (5a)), we obtain

$$\tilde{\mu}(B_k) = \int W(\omega) I_{B_k}(\omega)\,\mathrm{d}\mu.$$

Now let $N \to \infty$. By the monotone convergence theorem applied to both sides,

$$\tilde{\mu}(A \cap (M_k < \infty)) = \int_A W(\omega) I_{(M_k < \infty)}(\omega)\,\mathrm{d}\mu.$$

Next, as $k \to \infty$, $I_{(M_k < \infty)}(\omega)$ increases to $I_{(W < \infty)}(\omega)$. Another application of the monotone convergence theorem yields

$$\tilde{\mu}(A \cap (W < \infty)) = \int_A W(\omega) I_{(W < \infty)}(\omega)\,\mathrm{d}\mu = \int_A W\,\mathrm{d}\mu \tag{5b}$$

since $\mu(W < \infty) = 1$. Since (5b) is true for every $A \in \mathscr{F}_{k_0}$ and $k_0 < \infty$, it is true for $A \in \cup_0^\infty \mathscr{F}_k$ and hence for all $A \in B$. Finally, for any $A \in B$,

$$\tilde{\mu}(A) = \tilde{\mu}(A \cap (W < \infty)) + \tilde{\mu}(A \cap (W = \infty)),$$

so (b) follows.

(c) Clearly, $\tilde{\mu}_a$ in (c) is absolutely continuous with respect to $\mu$ and $\tilde{\mu}_s$ is singular with respect to $\mu$ since $\tilde{\mu}_s(W < \infty) = 0$ and $\mu(W = \infty) = 0$. The uniqueness follows since both $\mu$ and $\tilde{\mu}$ are finite measures. □

Next, we apply Corollary 1 to prove the LlogL theorem for Galton–Watson processes with arbitrary type space.

## 3. An application to branching processes

Let $(S, \mathbf{S})$ be a measurable space. Let $M \equiv \{\mu : \mu(\cdot) = \sum_{i=1}^{n} \delta_{x_i}(\cdot)$ for some $n < \infty$, $x_1, x_2, \ldots, x_n \in S\}$ where $\delta_x(\cdot)$ is the delta measure at $x$, that is, $\delta_x(A) = 1$ if $x \in A$ and 0 if $x \notin A$. Let $\mathbf{M}$ be the σ-algebra generated by sets of the form $\{\mu : \mu(A) = k\}$, where $A \in \mathbf{S}$ and $k \in \{0, 1, 2 \ldots\}$. By a point process on $(S, \mathbf{S})$ we mean a random mapping $\xi$ from some probability space $(\Omega, B, P)$ to $(M, \mathbf{M})$. It is clear that $M$ is closed under addition. Let, for each $x$ in $S$, $P^x(\cdot)$ denote a probability measure on $(M, \mathbf{M})$.

Given the family of probability measures $\{P^x : x \in S\}$, one can generate an $M$-valued Markov chain $\{Z_n\}_0^\infty$ as follows. Starting with $Z_0 = \sum_{i=1}^{z_0} \delta_{x_{0i}}$, let $\xi^{x_{0i}}$, $i = 1, 2, \ldots, z_0$, be independent point processes (that is, $M$-valued random variables) such that $\xi^{x_{0i}}$ has distribution $P^{x_{0i}}(\cdot)$. If we think of $Z_0$ as the zeroth generation, then the first generation $Z_1$ is given by

$$Z_1 = \sum_{i=1}^{z_0} \xi^{x_{0i}}.$$

If $Z_1(S) = z_1$, then we can rewrite $Z_1$ as

$$Z_1 = \sum_{j=1}^{z_1} \delta_{x_{1j}} \tag{6}$$

and $\{x_{1j} : j = 1, 2, \ldots, z_1\}$ are the types of the first-generation individuals. Similarly, given $Z_n = \sum_{i=1}^{z_n} \delta_{x_{ni}}$ where $Z_n = Z_n(S)$, and $Z_j : j \leq n$, generate independent point processes $\xi^{x_{ni}}$, $i = 1, 2, \ldots, Z_n$, such that $\xi^{x_{ni}}$ has distribution $P^{x_{ni}}(\cdot)$. Then set

$$Z_{n+1} \equiv \sum_{i=1}^{z_n} \xi^{x_{ni}} = \sum_{j=1}^{z_{n+1}} \delta_{x_{n+1,j}}, \tag{7}$$

where $z_{n+1} = Z_{n+1}(S)$

***Definition 1.*** *The Markov chain $\{Z_n\}_0^\infty$ is called a measure-valued Galton–Watson branching process with type space $S$, initial population $Z_0$ and offspring distribution family $P^x(\cdot)$; $x \in S$.*

When $S$ is a singleton this reduces to the simple Galton–Watson branching process. When $S$ is a finite set of size $k$, this becomes the multitype Galton–Watson branching process; see Athreya and Ney (1972) for definition and properties. Many continuous-time processes, such as the single- and multitype Bellman–Harris processes, branching Markov processes and branching random walks, can be cast as measure-valued branching processes in the above sense when considered at discrete time points $t = n\Delta$, $n = 0, 1, 2, \ldots$. For example, the single-type Bellman–Harris process may be viewed as a measure-valued branching process with $S = [0, \infty]$ and **S** the Borel σ-algebra of $S$, for each $x$, $P^x(\cdot)$ is the probability distribution of the vector $\xi^x$ of ages at time $\Delta$ in a Bellman–Harris process initiated by one particle of age $x$ at time 0.

Let $m(x, A) = \mathrm{E}\xi^x(A)$ be the *mean kernel*. Let $\rho > 1$ and $v : S \to (0, \infty)$ be an **S**-measurable eigenfunction of the mean kernel $m$ with eigenvalue $\rho$. That is,

$$\int_S v(y) m(x, \mathrm{d}y) = \rho v(x). \tag{8a}$$

Let $V : M \to (0, \infty)$ be defined by

$$V(\mu) \equiv \int v \,\mathrm{d}\mu \equiv \sum_1^n v(x_i) \tag{8b}$$

if $\mu = \sum_1^n \delta_{x_i}$.

Then from (7) we see that

$$\begin{aligned} \mathrm{E}(V(Z_{n+1}) | Z_0, Z_1, \ldots, Z_n) &= \mathrm{E}(V(Z_{n+1}) | Z_n) \\ &= \mathrm{E}\left( \sum_{i=1}^{z_n} V(\xi^{x_{ni}}) | Z_n \right) \\ &= \mathrm{E}\left( \rho \sum_{i=1}^{z_n} v(x_{ni}) \right) = \rho V(Z_n) \end{aligned}$$

by virtue of (8).

Thus $V$ is an eigenfunction for the Markov chain $\{Z_n\}_0^\infty$ with eigenvalue $\rho$. Let $P(.\,, .)$ denote the transition function of $\{Z_n\}_0^\infty$.

For any initial value $z$ in $M$ let $P_z$ and $\tilde{P}_z$ be the distribution of the Markov chain with initial condition $z$ and transition function $P$ and $\tilde{P}$, where

$$\tilde{P}(z, \mathrm{d}\mu) = \frac{V(\mu) P(z, \mathrm{d}\mu)}{\rho V(z)} \tag{9}$$

as in Section 2.

The results of Section 2 on the absolute continuity or singularity of $P_z$ and $\tilde{P}_z$ will now be used to establish a condition for the non-triviality of the limit random variable $W$ of the martingale

$$W_n = \frac{V(Z_n)}{\rho^n} \tag{10}$$

under $P_{Z_0}$ for the Galton–Watson branching process $\{Z_n\}$.

It follows from Corollary 1 that, for $Z_0 \neq 0$,

$$P_{Z_0}(W = 0) = 1 \text{ if and only if } \tilde{P}_{Z_0}(W = \infty) = 1 \tag{11}$$

and

$$\mathrm{E}_{Z_0} W = V(Z_0) \text{ if and only if } \tilde{P}_{Z_0}(W = \infty) = 0. \tag{12}$$

When $S$ is a singleton Lyons *et al.* (1995) showed that, under $\tilde{P}_{Z_0}$, the Markov chain $\{Z_n\}_0^\infty$ is a branching process with an immigration component and used a simple criterion for the two cases $\tilde{P}_{Z_0}(W = \infty) = 1$ and $\tilde{P}_{Z_0}(W = \infty) = 0$. It turns out this is a *dichotomy*, that is, $\tilde{P}_{Z_0}(W = \infty)$ is either 1 or 0, and that the former prevails if and only if the LlogL condition of Kesten and Stigum (1966) holds, that is, if and only if $\mathrm{E} Z_1 \log Z_1 < \infty$, where $Z_0 = 1$.

Our goal now is to show that $\tilde{P}_{Z_0}$ can still be interpreted as the distribution of a measure-valued branching process with an immigration component and to seek sufficient conditions for $P_{Z_0}(W = \infty)$ to be one and also for it to be zero. In a number of special cases this becomes a dichotomy.

Here is a probabilistic description of the $\tilde{P}$ Markov chain. For any non-negative measurable function $f$ and a measure $\mu$ on $(S, \mathbf{S})$ let

$$(f, \mu) \equiv \int f \, \mathrm{d}\mu,$$

and for any $(M, \mathbf{M})$ random variable $\xi$ its moment generating functional

$$M_\xi(f) = \mathrm{E}(\mathrm{e}^{-(f,\xi)}).$$

It is known that $M_\xi(.)$ determines the probability distribution of $\xi$.

Let $\{Z_n\}_0^\infty$ be a Markov chain with values in $(M, \mathbf{M})$ and transition function $\tilde{P}$ defined in (9), that is,

$$\tilde{P}(m, \mathrm{d}m') = \frac{V(m')P(m, \mathrm{d}m')}{\rho V(m)}$$

where $V(\cdot)$ is as in (8a); $v$ is a non-negative function on $(S, \mathbf{S})$ such that, for any $x$ in $S$, $\mathrm{E} V(\xi^x) = \rho v(x)$, $\xi^x$ being a point process with distribution $P^x$; and, for $m = \sum_1^n \delta_{x_i}$, $P(m, \mathrm{d}m') = P(\sum_1^n \xi^{x_i} \in \mathrm{d}m')$ where $\xi^{x_i}$, $i = 1, 2 \ldots, n$, are independent point processes with $\xi^{x_i}$ having distribution $P^{x_i}$.

Thus, under $\tilde{P}$, the moment generating functional of $Z_1$ given $Z_0$ is

$$\begin{aligned} M_{Z_1|Z_0}(f) &= \tilde{\mathrm{E}}(\mathrm{e}^{-(f,Z_1)}|Z_0) \\ &= \mathrm{E}\left(\frac{\mathrm{e}^{-(f,Z_1)}V(Z_1)}{\rho V(Z_0)}\middle| Z_0\right), \end{aligned}$$

where $\tilde{\mathrm{E}}$ denotes expectation under $\tilde{P}$ and E denotes expectation under $P$. But under $P$, if $Z_0 = \sum_1^n \delta_{x_i}$, then $Z_1$ may be written as

$$Z_1 = \sum_1^n \xi^{x_i}$$

where $\{\xi^{x_i}, i = 1, 2 \ldots\}$ are being independent, $\xi^{x_i}$ having distribution $P^{x_i}$. So

$$M_{Z_1|Z_0}(f) = \mathrm{E}\left(\frac{\exp\{-(f, \sum_1^n \xi^{x_i})\} V(\sum_1^n \xi^{x_i})}{\rho(\sum_1^n v(x_i))}\right).$$

Since $V(\sum_1^n \xi^{x_i}) = \sum_1^n V(\xi^{x_i})$,

$$\begin{aligned} M_{Z_1|Z_0}(f) &= \sum_{j=1}^n \frac{v(x_j)}{(\sum_1^n v(x_i))} \mathrm{E}\left(\frac{\mathrm{e}^{-(f,\xi^{x_j})} V(\xi^{x_j})}{\rho v(x_j)} \prod_{i\neq j} \mathrm{e}^{-(f,\xi^{x_i})}\right) \\ &= \sum_{j=1}^n \frac{v(x_j)}{(\sum_1^n v(x_i))} \mathrm{E}\left(\frac{\mathrm{e}^{-(f,\xi^{x_j})} V(\xi^{x_j})}{\rho v(x_j)}\right) \prod_{i\neq j} \mathrm{E}(\mathrm{e}^{-(f,\xi^{x_i})}) \text{ (by independence)} \\ &= \sum_{j=1}^n \frac{v(x_j)}{(\sum_1^n v(x_i))} \mathrm{E}(\mathrm{e}^{-(f,\tilde{\xi}^{x_j})}) \prod_{i\neq j} \mathrm{E}(\mathrm{e}^{-(f,\xi^{x_i})}), \end{aligned}$$

where $\tilde{\xi}^x$ is an $M$-valued random variable with probability distribution

$$P(\tilde{\xi}^x \in \mathrm{d}m) = \frac{V(m)P(\xi^x \in \mathrm{d}m)}{\rho v(x)}. \tag{13}$$

This shows that the Markov chain $\{Z_n\}_0^\infty$ with transition function $\tilde{P}$ evolves in the manner now described. Given $Z_n = (x_{n1}, x_{n2}, \ldots x_{nz_n})$, $Z_{n+1}$ is generated as follows:

(i) First pick the individual $x_{nj}$ with probability $v(x_{nj})/\sum_1^{z_n} v(x_{ni})$ and choose its offspring point process $\tilde{\xi}^{x_{nj}}$ according to the $V(\cdot)$-biased probability distribution $\tilde{P}^x(\mathrm{d}m) = V(m)P^x(\mathrm{d}m)/\rho v(x)$.

(ii) For all the other individuals choose the offspring point process $\xi^{x_{ni}}$ according to the original probability distribution $P^{x_{ni}}(\mathrm{d}m)$.

(iii) Choose $\tilde{\xi}^{x_{nj}}$ and $\xi^{x_{ni}} i \neq j$ all independently.

(iv) Set $Z_{n+1} = \tilde{\xi}^{x_{nj^*}} + \sum_{i\neq j^*} \xi^{x_{ni}}$, (14)

where $j^*$ is the index of the individual chosen according to (i).

The above construction is similar to that of Lyons *et al*. (1995). (The measure corresponding to $\tilde{P}_{Z_0,n}$ is a sort of average of the one introduced by Lyons *et al*. (1995) that keeps track of the 'spine' $\{x_{nj^*}, n = 1, 2, \ldots\}$.) For some Galton–Watson processes their more elaborate construction is not necessary.

The idea of using a $V(\cdot)$-biased distribution is similar to 'size biasing' in population genetics literature and also occurs in the work of Waymire and Williams (1996).

Thus

$$\frac{V(Z_{n+1})}{\rho^{n+1}} = \sum_{i\neq j^*} \frac{V(\xi^{x_{ni}})}{\rho^{n+1}} + \frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}}. \tag{15}$$

The condition for $P(W=0)=1$ is $\tilde{P}(W=\infty)=1$. So if $\overline{\lim}(V(\tilde{\xi}^{x_{nj^*}}))/(\rho^{n+1})=\infty$ with probability one then, under $\tilde{P}$, $\overline{\lim}\, W_{n+1} \geqslant \overline{\lim}(V(\tilde{\xi}^{x_{nj^*}}))/(\rho^{n+1})=\infty$ with probability one and hence $P_{z_0}(W=0)$ would be one.

A sufficient condition for $P_{Z_0}(W=0)=1$ is that, for $\tilde{\xi}^x$ as in (13),

$$\inf_x P(V(\tilde{\xi}^x)>t) \equiv \underline{h}(t), \qquad t>0, \tag{16a}$$

satisfies

$$\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u=\infty. \tag{16b}$$

This is so because, for all $K>0$,

$$\tilde{P}\left(\frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}} \geqslant K|F_n\right) \geqslant \underline{h}(K\rho^{n+1})$$

and (16b) implies $\Sigma\underline{h}(K\rho^{n+1})=\infty$ yielding, by the conditional Borel–Cantelli lemma (Durrett 1996, p. 240), the conclusion that

$$\overline{\lim}\,\frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}} \geqslant K \qquad \text{with probability one.} \tag{17}$$

This being true for every $K=1, 2, \ldots,$ $\overline{\lim}(V(\tilde{\xi}^{x_{nj^*}}))/\rho^{n+1})=\infty$ with probability one.

Next we look for a sufficient condition for $\mathrm{E}_{Z_0}(W)=1$. This is equivalent to $\tilde{P}_{Z_0}(W=\infty)=0$. Consider the condition that, for $\tilde{\xi}^x$ as in (13),

$$\overline{h}(t) \equiv \sup_x P(V(\tilde{\xi}^x)>t) \tag{18a}$$

satisfies

$$\int_1^\infty \overline{h}(\mathrm{e}^u)\,\mathrm{d}u<\infty. \tag{18b}$$

It follows from (15) that

$$\begin{aligned}
\tilde{\mathrm{E}}\left(\frac{V(Z_{n+1})}{\rho^{n+1}}\middle| Z_n, \tilde{\xi}^{x_{nj^*}}\right) &= \sum_{i\neq j^*}\frac{\rho V(x_{ni})}{\rho^{n+1}}+\frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}} \\
&\leqslant \sum_i \frac{V(x_{ni})}{\rho^n}+\frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}} \qquad (\text{since } V(\cdot)\geqslant 0) \\
&= \frac{V(Z_n)}{\rho^n}+\frac{V(\tilde{\xi}^{x^{nj^*}})}{\rho^{n+1}} \qquad (19a).
\end{aligned}$$

Iterating the above yields,

$$\tilde{\mathrm{E}}\left(\frac{V(Z_{n+1})}{\rho^{n+1}}\middle| Z_{n-1}, \tilde{\xi}^{x_{n-1,j^*}}, \tilde{\xi}^{x_{nj^*}}\right) = E\left(\frac{V(Z_n)}{\rho^n}\middle| Z_{n-1}, \tilde{\xi}^{x_{n-1,j^*}}, \tilde{\xi}^{x_{nj^*}}\right) + \frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}}$$

$$\leqslant \frac{V(Z_{n-1})}{\rho^{n-1}} + \frac{V(\tilde{\xi}^{x_{n-1,j^*}})}{\rho^n} + \frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{n+1}},$$

and hence

$$\tilde{\mathrm{E}}\left(\frac{V(Z_{n+1})}{\rho^{n+1}}\middle| Z_0, \tilde{\xi}^{x_{0j^*}}, \tilde{\xi}^{x_{ij^*}}, \ldots, \tilde{\xi}^{x_{nj^*}}\right) \leqslant V(Z_0) + \sum_{r=0}^{n} \frac{V(\tilde{\xi}^{x_{rj^*}})}{\rho^{r+1}}$$

$$\leqslant V(Z_0) + \sum_{r=0}^{\infty} \frac{V(\tilde{\xi}^{x_{nj^*}})}{\rho^{r+1}} \equiv W^*, \text{ say.} \tag{19b}$$

Next,

$$\hat{P}\left(\frac{V(\tilde{\xi}^{x_{rj^*}})}{\rho^r} \geqslant \delta^r\right) = \tilde{\mathrm{E}}\left(P\left(\frac{V(\tilde{\xi}^{x_{rj^*}})}{\rho^r} \geqslant \delta^r \middle| x_{rj^*}\right)\right)$$

$$\leqslant \overline{h}((\rho\delta)^r)$$

where $\overline{h}$ is as in (18a). By (18b), $\sum_r \overline{h}((\rho\delta)^r) < \infty$ if $0 < \delta < 1$ is chosen such that $\rho\delta > 1$. By Borel–Cantelli this implies that, with probability one under $\tilde{P}$, $V(\tilde{\xi}^{x_{nj^*}})/\rho^r \leqslant \delta^r$ for all but a finite number of $r$, and hence that $W^* < \infty$ with probability one under $\tilde{P}$ (since $0 < \delta < 1$).

Next, from Proposition 2, under $\tilde{P}$, the sequence $\{W_n^{-1} : n = 0, 1, 2 \ldots\}$ is a non-negative martingale and hence $\lim W_n = W \leqslant \infty$ exists with probability one under $\tilde{P}$. Let $\tilde{G}_n$ be the σ-algebra generated by $Z_0$ and $\tilde{\xi}^{x_{rj^*}}$ $r = 0, 1, 2, \ldots n$ and $\tilde{G} = \sigma(\bigcup_0^\infty \tilde{G}_n)$. Then, by Fatou,

$$\tilde{\mathrm{E}}(W|\tilde{G}) \leqslant \underline{\lim}\, \tilde{\mathrm{E}}(W_n|G).$$

But $\tilde{\mathrm{E}}(W_n|G) \leqslant \tilde{\mathrm{E}}(\tilde{\mathrm{E}}(W_n|G_n)|G) \leqslant \tilde{\mathrm{E}}(W^*|G) = W^*$, since $W^*$ is $G$-measurable. Thus

$$\tilde{\mathrm{E}}(W|\tilde{G}) < \infty \text{ with probability one under } \tilde{P}_{z_0}$$

and hence

$$\tilde{P}_{Z_0}(W < \infty) = 1 \quad \text{or} \quad \tilde{P}_{Z_0}(W = \infty) = 0.$$

So under (18b) we conclude that

$$\mathrm{E}_{Z_0} W = 1 \text{ under } P_Z.$$

Summarizing the above discussion we have the following:

**Theorem 3.** *Let* $\{Z_n\}_0^\infty$ *be a measure-valued branching process with type space* $(S, \mathbf{S})$ *and offspring distribution family* $\{P^x : x \in S\}$ *as in Definition 1. Let* $\rho > 1$, $v : S \to (0, \infty)$ *and* $V : M \to (0, \infty)$ *satisfy (8a) and (8b). Let* $W_n = V(Z_n)/\rho^n$. *Let* $\underline{h}(t) \equiv \inf_x P(V(\tilde{\xi}^x) > t)$ *and*

$\overline{h}(t) \equiv \sup_x P(V(\tilde{\xi}^x) > t)$, *where* $\tilde{\xi}^x$ *has distribution defined by (13). Then for any non-zero non-trivial* $Z_0$,

(*i*) $$\lim_n W_n = W \text{ exits with probability one under } P_{Z_0};$$

(*ii*) $$P_{Z_0}(W = 0) = 1 \text{ if} \int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u = \infty;$$

(*iii*) $$\mathrm{E}_{Z_0} W = V(Z_0) \text{ if} \int_1^\infty \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty.$$

***Remark 2.*** In many cases the two conditions $\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u = \infty$ and $\int_1^\infty \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$ become a dichotomy. That is, $\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$ implies $\int_1^\infty \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$.

***Remark 3.*** There are other versions of the LlogL theorem for the general state space case. Asmussen and Herring (1983) give a version with some compactness type conditions on the mean kernel. Kesten (1989) has a version in the countably infinite type case. The present author has not attempted to deduce these previously known results from Theorem 3 above. It does appear that in terms of hypotheses Theorem 3 above is perhaps more transparent and simpler to verify than those in the quoted works.

# 4. Examples

## 4.1. Multitype Galton–Watson process

Let $S = \{1, 2, \ldots, k\}$. An individual located at site $i$ will be referred to as of type $i$. Let $\boldsymbol{\xi}^i$ denote the random vector of offspring of a type $i$ individual. Let $m_{ij} = \mathrm{E}(\xi^i_j)$, where $\xi^i_j$ is the $j$th coordinate of $\boldsymbol{\xi}^i$. Assume there is no extinction, that is, $P(\boldsymbol{\xi}^i = \mathbf{0}) = 0$ for all $i$ where $\mathbf{0}$ is the vector of zeros. Assume simple irreducibility, that is, $0 < m_{ij} < \infty$ for all $i$, $j$.

Let $1 < \rho < \infty$ be the Perron–Froebenius maximal eigenvalue of $M = ((m_{ij}))$ with corresponding left and right eigenvectors $\mathbf{u}$ and $\mathbf{v}$ respectively normalized so that $\mathbf{u} \cdot \mathbf{1} = 1$ and $\mathbf{u} \cdot \mathbf{v} = 1$ where $\mathbf{1}$ is the vector of ones and $\cdot$ refers to dot product.

Let $\tilde{\boldsymbol{\xi}}^i$ be the random vector with $\mathbf{v}$-biased distribution

$$P(\tilde{\boldsymbol{\xi}}^i = \mathbf{j}) = \frac{\mathbf{j} \cdot \mathbf{v} P(\tilde{\boldsymbol{\xi}}^i = \mathbf{j})}{\rho v_i}.$$

Let $h_i(t) = P(\mathbf{v} \cdot \tilde{\boldsymbol{\xi}}^i > t)$ for $t > 0$.

We first consider sufficiency. Clearly $\overline{h}(t) \equiv \sup_i h_i(t) \leqslant \sum_{i=1}^k h_i(t)$.

Thus $\int_1^\infty h_i(\mathrm{e}^u)\,\mathrm{d}u < \infty$ for all $i$ implies $\int_1^\infty \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$. But

$$\int_1^\infty h_i(\mathrm{e}^u)\,\mathrm{d}u = \int_1^\infty P(\mathbf{v} \cdot \tilde{\boldsymbol{\xi}}^i > \mathrm{e}^u)\,\mathrm{d}u = \int_1^\infty \left( \sum_{\mathbf{j}} \frac{\mathbf{v} \cdot \mathbf{j}}{\rho v_i} P(\boldsymbol{\xi}^i = \mathbf{j}) I(\mathbf{v} \cdot \mathbf{j} > \mathrm{e}^u) \right) \mathrm{d}u, \tag{20}$$

where $I(t > \mathrm{e}^u) = 1$ if $t > \mathrm{e}^u$ and 0 if $t \geqslant \mathrm{e}^u$. The above integral equals

$$\sum_{\mathbf{j}} \frac{\mathbf{v}\cdot\mathbf{j}}{\rho v_i} P(\boldsymbol{\xi}^i = \mathbf{j}) \int_1^\infty I(\mathbf{v}\cdot\mathbf{j} > \mathrm{e}^u)\,\mathrm{d}u.$$

Since for $t > \mathrm{e}$, $\int_1^\infty I(t > \mathrm{e}^u))\,\mathrm{d}u = \log t$, it follows that

$$\int_1^\infty h_i(\mathrm{e}^u)\,\mathrm{d}u < \infty \text{ if and only if } \mathrm{E}(\mathbf{v}\boldsymbol{\xi}^i)\log(\mathbf{v}\cdot\boldsymbol{\xi}^i) < \infty. \tag{21}$$

Thus, Theorem 3(iii) yields the sufficiency part of the Kesten–Stigum theorem (see Kesten and Stigum 1966) under the assumption $0 < m_{ij} < \infty$ for all $i$, $j$.

Turning now to the necessary part, consider the chain $\{Z_{2n}: n = 0, 1, 2 \ldots\}$ which is also a Galton–Watson branching process. Let

$$h_{i2}(t) = P(V(\tilde{Z}_2) > t \mid Z_0 = e_i),$$

$$h_i(t) = P(V(\tilde{Z}_1) > t \mid Z_1 = e_i).$$

Once again assuming simple irreducibility, that is, $m_{ij} > 0$ for all $i$, $j$, it can be seen that for every $i$, $j$, there exist $C_{ij} > 0$ such that

$$h_{i2}(t) \geqslant C_{ij} h_j(t).$$

Now suppose

$$\mathrm{E}(\mathbf{v}\cdot\boldsymbol{\xi}^{(j)})\log(\mathbf{v}\cdot\boldsymbol{\xi}^{(j)}) = \infty \qquad \text{for some } j = j_0. \tag{22}$$

Then

$$\underline{h}(t) = \inf_i h_{i2}(t) \geqslant C h_{j_0}(t),$$

where $C = \inf_i C_{ij_0}$ and $\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u \geqslant C\int_1^\infty h_{j_0}(\mathrm{e}^u)\,\mathrm{d}u$. But by (21) this last integral is $\infty$ under (22). Now by Theorem 3(ii) it follows that $W = 0$ with probability one and the necessary part of the Kesten–Stigum theorem holds (see Kesten and Stigum 1966).

The above arguments can be extended to the general irreducible non-singular case when there exists an $r$ such that $M^r$ has all strictly positive entries by considering the Galton–Watson process along the sequence $nr$, $n = 0, 1, 2, \ldots$.

## 4.2. Single-type Bellman–Harris process

Let $\{p_j\}^\infty$ be a probability distribution and $G(\cdot)$ be a non-lattice probability distribution on $(0, \infty)$. Let $S = [0, \infty)$ and $\mathbf{S} = B[0, \infty)$, the Borel σ-algebra. For each $x > 0$, let $\{\xi_t^x\}$ be the point process corresponding to the ages of all the individuals present at time $t$ in a Bellman–Harris process initiated by one particle of age $x$ at time 0 and with offspring distribution $\{p_j\}$ and lifetime distribution $G$. Then, for any $\Delta > 0$, the sequence $Z_n = \xi_{n\Delta}^x$, $n = 0, 1, 2, 3, \ldots,$ is a measure-valued branching process of the type treated in Section 3 with type space $S$ and offspring family $\{P^x(\cdot): x \in S\}$ given by

$$P^x(\cdot) = P(\xi_\Delta^x \varepsilon \cdot)$$

Let $\alpha > 0$ be the Malthusian parameter defined by

$$m\int_{[0,\infty)} \mathrm{e}^{-\alpha u}\,\mathrm{d}G(u) = 1, \tag{23a}$$

where $1 < m = \Sigma j p_j < \infty$. For all $x \geqslant 0$ such that $1 - G(x) > 0$, let

$$\begin{aligned} v(x) &\equiv \left(\int_{[x,\infty)} \mathrm{e}^{-\alpha u}\,\mathrm{d}G(u)\right)\mathrm{e}^{\alpha x}(1 - G(x))^{-1} \\ &= \mathrm{E}(\mathrm{e}^{-\alpha L_x}), \end{aligned} \tag{23b}$$

where $L_x$ denotes the time of death of an ancestor whose age is $x$ so that

$$P(L_x > t) = \frac{1 - G(x+t)}{(1 - G(x))} \qquad \text{for } t \geqslant 0.$$

If $T = \sup\{x\colon 1 - G(x) > 0\}$ then the effective type space is $S = [0, T]$. We set $v(T) = 1$ since $L_T = 0$ with probability one. It can be shown that (see Athreya and Ney 1972)

$$\mathrm{E}V(\xi_t^x) = \mathrm{e}^{\alpha t} v(x).$$

Consider an ancestor of age $x$ who dies at time $L_x$ and produces $N$ offspring. Let $\{\xi_t^{0,i}\colon t \geqslant 0\}$, $i = 1, 2, \ldots,$ be independent and identically distributed copies of the process $\{\xi_t^0\colon t \geqslant 0\}$ and independent of $L_x$ and $N$. Then the process $\{\xi_t^x\colon t \geqslant 0\}$ for this ancestor may be written as:

$$\xi_t^x = \begin{cases} x + t, & L_x > t, \\ \sum_{i=1}^{N} \xi_{t-L_x}^{0,i}, & L_x \leqslant t. \end{cases}$$

Let $\Delta = 1$ and $h_x(t) = P(V(\tilde{\xi}_1^x) > t)$ for $t \geqslant 0$. Then from the definition of $\tilde{\xi}^x$ as in (13) we obtain

$$h_x(t) = \frac{\mathrm{E}(V(\xi_1^x)\colon V(\xi_1^x) > t)}{\mathrm{e}^{\alpha} v(x)} \tag{24}$$

Since $v(x) = \mathrm{E}(\mathrm{e}^{-\alpha L_x})$ for $x < T$ and 1 for $x = T$, $v(\cdot)$ is always in [0,1]. Thus,

$$V(\xi_1^x) = \begin{cases} v(x+1), & L_x > 1, \\ \sum_1^{N} V(\xi_{1-L_x}^{0,i}), & L_x \leqslant 1. \end{cases} \tag{25}$$

Since there is no extinction, $\xi_t^{0,i}([0, \infty))$ is non-decreasing in $t$ and, $v(\cdot)$ being less than or equal to 1, we obtain

$$\sum_1^N V(\xi_{1-L_x}^{0,i}) \leqslant \sum_1^N \xi_1^{0,i} = \sum_1^N Z_i = Y, \text{ say.} \tag{26}$$

By the conditional independence of $N$, $L_x$ and $\sum_1^N Z_i$ we have, for $t > 1$,

$$h_x(t) \leqslant \mathrm{E}(Y : Y > t)\frac{P(L_x \leqslant 1)}{\mathrm{e}^{\alpha}v(x)}. \tag{27}$$

Since $\mathrm{e}^{\alpha}v(x) = \mathrm{E}(\mathrm{e}^{\alpha(1-L_x)}) \geqslant P(L_x \leqslant 1)$, we obtain

$$\overline{h}(t) = \sup_x h_x(t) \leqslant \mathrm{E}(Y : Y > t) \equiv K_1(t), \text{ say.} \tag{28}$$

So $\int_1^{\infty} \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$ if $\int_1^{\infty} K_1(\mathrm{e}^u)\,\mathrm{d}u < \infty$. But

$$\begin{aligned}\int_1^{\infty} K_1(\mathrm{e}^u)\,\mathrm{d}u &= \int_1^{\infty} \mathrm{E}(YI(Y > \mathrm{e}^u))\,\mathrm{d}u \\ &= \mathrm{E}\left(\int_1^{\infty} YI(Y > \mathrm{e}^u)\,\mathrm{d}u\right) \\ &= \mathrm{E}(Y \log Y : Y > \mathrm{e}) \\ &\leqslant \mathrm{E}Y(\log Y).\end{aligned}$$

From the definition of $Y$ in (26) and the independence of $N$ and $\{Z_i\}$ it follows that

$$\begin{aligned}\mathrm{E}(Y \log Y) &= \mathrm{E}((N \log N)\bar{Z}) + \mathrm{E}(N\bar{Z}\log\bar{Z}), \qquad \text{where } \bar{Z} = \frac{1}{N}\sum_1^N Z_i, \\ &= \mathrm{E}(\mathrm{E}((N \log N)\bar{Z}|N)) + \mathrm{E}(N\bar{Z}\log\bar{Z}).\end{aligned}$$

But

$$\mathrm{E}((N \log N)\bar{Z}|N) = (N \log N)\mathrm{E}(Z_1) \tag{29}$$

and by the convexity of the function $x \log x$, for $x > 0$,

$$\bar{Z}\log\bar{Z} \leqslant \frac{1}{N}\sum_1^N Z_i \log Z_i,$$

so that

$$\begin{aligned}\mathrm{E}(N\bar{Z}\log\bar{Z}) &\leqslant \mathrm{E}\left(\sum_1^N Z_1 \log Z_i\right) \\ &= \mathrm{E}(Z_1 \log Z_i)(\mathrm{E}N).\end{aligned}$$

It is known (see Athreya and Ney 1972) that $\mathrm{E}N \log N = \Sigma j(\log j)p_j < \infty$ implies $\mathrm{E}Z_1 \log Z_1 < \infty$ and hence $\mathrm{E}Y \log Y < \infty$. Thus $\Sigma j(\log j)p_j < \infty$ implies $\int_1^{\infty} \overline{h}(\mathrm{e}^u)\,\mathrm{d}u < \infty$.

Now consider the measure-valued branching process $\{\xi_n^0,\ n = 1, 2, \ldots\}$ and the associated martingale sequence $\{W_n = \mathrm{e}^{-\alpha n}V(\xi_n^0)\}_0^{\infty}$. By Theorem 3(iii), we see that $\Sigma j(\log j)p_j < \infty$ implies $W_n$ has a non-trivial limit. This is the 'if' part of Kesten–Stigum theorem for the Bellman–Harris process.

For the only if part we make the assumption that

$$\delta = \inf_x P(L_x \leqslant 1) > 0. \tag{30}$$

Then

$$v(x) = \mathrm{E}(\mathrm{e}^{-\alpha L_x}) \geqslant \mathrm{e}^{-\alpha} P(L_x \leqslant 1) \geqslant \mathrm{e}^{-\alpha}\delta = c, \text{ say.}$$

So

$$V(\xi_1^x) \geqslant c\left(\sum_1^N Z_i\right) I(L_x \leqslant 1)$$

and hence $h_x(\cdot)$ defined in (24) satisfies

$$\begin{aligned} h_x(t) &\geqslant c\mathrm{E}(Y : cY > t)\frac{P(L_x \leqslant 1)}{\mathrm{e}^{\alpha} v(x)} \\ &\geqslant c\mathrm{E}(y : cY > t)\delta. \end{aligned}$$

Thus

$$\underline{h}(t) = \inf_x h_x(t) \geqslant c\delta \mathrm{E}(Y : Y > t/c)$$

and

$$\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u = \infty \quad \text{if} \quad \int_1^\infty \mathrm{E}(Y : Y > t/c)\,\mathrm{d}t = \infty,$$

that is, if $\mathrm{E}Y(\log Y) = \infty$.

It can be seen from (27) and (28) that $\mathrm{E}N \log N = \Sigma j(\log j) p_j = \infty$ implies $\mathrm{E}Y(\log Y) = \infty$. Thus we conclude that $\Sigma j(\log j) p_j = \infty$ and $\delta \equiv \inf_x P(L_x \leqslant 1) > 0$ imply $\int_1^\infty \underline{h}(\mathrm{e}^u)\,\mathrm{d}u = \infty$ and hence that $S_n \to 0$ with probability one. The same argument works if there is a $t_0 > 0$ such that $\inf_x P(L_x \leqslant t_0) > 0$.

It is possible to drop this last condition with a slightly more involved argument to show

$$\Sigma P(V(\tilde{\xi}_n^{x_{nj^*}}) > K\mathrm{e}^{\alpha n} | \mathscr{F}_n) = \infty$$

and hence (17). This argument looks at the empirical distribution of $\{x_{nj}\}$ at time $n$ and establishes that, for some $0 < a < T$, the proportion of $x_{nj} \leqslant a$ is bounded below by a positive quantity.

The argument for the single-type case above can be extended to the multitype Bellman–Harris case; see Athreya and Rama Murthy (1977) for a statement of the LlogL theorem in this case.

## Acknowledgement

The author would like to thank the two referees for their thorough examination of the paper and their many suggestions for improvement.

# References


Asmussen, S. and Herring, H. (1983) *Branching Processes*. Boston: Birkhäuser.

Athreya, K.B. and Ney, P. (1972) *Branching Processes*. New York: Springer-Verlag.

Athreya, K.B. and Rama Murthy, K. (1977) The convergence of the state distribution in multidimensional Crump–Mode–Jagers processes. *J. Indian Math. Soc.*, **41**, 27–57.

Durrett, R. (1996) *Probability: Theory and Examples*, 2nd edition. Belmont, CA: Duxbury Press.

Kesten, H. (1989) Supercritical branching processes with countably many types and the sizes of random cantor sets. In T.N. Anderson, K.B. Athreya and D. Iglehart (eds), *Probability, Statistics and Mathematics Papers in Honor of Samuel Karlin*, pp. 108–121. New York: Academic Press.

Kesten, H. and Stigum, B.P. (1966) A limit theorem for multidimensional Galton–Watson process. *Ann. Math. Statist.*, **37**, 1211–1223.

Kurtz, T.G., Lyons, R., Pemantle, R. and Peres, Y. (1997) A conceptual proof of the Kesten–Stigum theorem for multitype branching processes. In K.B. Athreya and P. Jagers (eds), *Classical and Modern Branching Processes*, IMA Vol. Math. Appl. 84, pp. 181–186. New York: Springer-Verlag.

Lyons, R. (1997) A simple path to Biggins Martingale convergence for branching random walk. In K.B. Athreya and P. Jagers (eds), *Classical and Modern Branching Processes*, IMA Vol. Math. Appl. 84, pp. 217–222. New York: Springer-Verlag.

Lyons, R., Pemantle, R. and Peres, Y. (1995) Conceptual proofs of LlogL criteria for mean behaviour of branching processes, *Ann. Probab.*, **23**, 1125–1138.

Rogers, L.G.C. and Williams, D. (1994) *Diffusions, Markov Processes and Martingales*, 2nd edn, Vol. 1. Chichester: Wiley.

Waymire, E.C. and Williams, S.C. (1996) A cascade decomposition theory with applications to Markov and exchangeable cascades, *Trans. Amer. Math. Soc.*, **348**, 585–632.


Received May 1997 and revised February 1999