Stochastic Processes and their Applications 8 (1978) 101–116.

THE GENEALOGY OF CRITICAL BRANCHING PROCESSES*

R. DURRETT

Department of Mathematics, University of California, Los Angeles, CA 90024, USA

Received 12 July 1977
Revised 31 May 1978

In this paper we will obtain results concerning the distribution of generations and the degree of relationship of the individuals in a critical branching process $\{Z(t), t \geq 0\}$ and we will apply these results to obtain a "central limit theorem" for critical branching random walks.

AMS 1970 Subj. Class.: Primary 60J80; Secondary 60F05
critical branching process,
generation numbers, degree of relationship,
branching random walk,
central limit theorem

1. Introduction

In this paper we will solve two problems concerning the genealogy of critical branching processes. We will obtain a limit theorem for the generation number of a randomly chosen individual and we will prove some results which describe the relationship in a randomly chosen finite set of particles.

To obtain the result for the generation numbers we will study the distributions of two related random variables: δ_n the time of the first death of a member of generation n and σ_n the time of death of the last member of generation n. In Section 2 we show that if $Z(t), t \geq 0$ is a critical age-dependent branching process in which the particles have lifetimes with a distribution G which has mean $\mu < \infty$, then $(\delta_n/n \mid 0 \leq \delta_n < \infty)$ and $(\sigma_n/n \mid 0 \leq \sigma_n < \infty)$ converge in probability to μ. Using this result we can show that if U_t is the generation number of a particle picked at random from those above at time t $(U_t/t \mid Z(t) > 0) \to 1/\mu$ as $t \to \infty$.

The proofs of the limit results for $(\delta_n/n \mid 0 \leq \delta_n < \infty)$ and $(\sigma_n/n \mid 0 \leq \sigma_n < \infty)$ are based on a result which is an extension of a theorem proved by Hammersley (1974) who studied the limiting behavior of δ_n/n for a supercritical branching process. The first part of the proof which we give is exactly the same as Hammersley's proof, but after formula (2.6) the argument is different and is simpler. This should not be surprising since the conclusions are different in the supercritical case—Hammersley

* This research was supported by NSF Grant No. MCS 72-04591.

(1974), Samuels (1971), and Biggins (1976) have shown that as $n \to \infty$

$$\delta_n/n \to \gamma, \quad n/U_n \to \mu', \quad \text{and} \quad \sigma_n/n \to \Gamma \tag{1}$$

where the constants $\gamma < \mu' < \mu < \Gamma$ are defined in terms of the Laplace transform $\int e^{-\theta t}\, dG(t)$ (see [12, p. 664]; [17, p. 657]; [3, pp. 458, 455]). If $G(t) = 1 - e^{-t}$ and the offspring distribution has mean $m > 1$ then $\mu' = 1/m$ and $\gamma < \Gamma$ are the two roots of $x\, e^{1-x} = 1/m$. From this we see that the results we obtained in the critical case are just the ones which would be obtained by extrapolating from the supercritical case.

When we consider results about the degree of relationship this is no longer the case. If we take two individuals alive at time t they will be in generations U_t^1 and U_t^2 and have a last common ancestor who died at time $D_t \leq t$ and who was a member of generation U_t^{12}. In the supercritical case Buhler (1972) has shown the following result (see Theorem 4.2, p. 471).

Theorem. *For $1 \leq i < j \leq k$ let $R_t^{ij} = (U_t^i - U_t^{ij}) + (U_t^j - U_t^{ij})$ be the degrees of relationship of k individuals chosen randomly at time t in a Markov branching process. If the offspring distribution has finite variance then on $\{Z_t > 0$ for all $t \geq 0\}$ the random variables*

$$(R_t^{ij} - 2mt)(2mt)^{-1/2}, \quad 1 \leq i < j \leq k \tag{2}$$

converge weakly to R^{ij} as $t \to \infty$. The random variables R^{ij} have a joint normal distribution with $\mathbf{E}R^{ij} = 0$ and

$$\operatorname{cov}(R^{ij}, R^{kl}) = \begin{cases} 1 & \text{if} \quad \{i, j\} = \{k, l\}, \\ \frac{1}{2} & \text{if} \quad |\{i, j\} \cap \{k, l\}| = 1, \\ 0 & \text{if} \quad \{i, j\} \cap \{k, l\} = \emptyset. \end{cases}$$

The result implies in particular that R^{ij}/t converges in probability to $2m$. Now from the result of Samuels (1971) mentioned above U_t^1/t and U_t^2/t converge to m in probability so we have that $(U_t^1 + U_t^2 - 2R_t^{12})/t$ converges in probability to 0. This result suggests that as $t \to \infty$, D_t/t converges to 0 in probability. An even stronger statement is true. It is easy to show (see formula (1) of Section 3) that if $s_t \to \infty$, then

$$\lim_{t \to \infty} \mathbf{P}\{D_t > s_t\} = 0. \tag{3}$$

Eq. (3) shows that individuals in a supercritical branching process are in general very distant relatives. In the critical case the situation is much different. In Section 3 we show that for Markov branching processes

$$\lim_{t \to \infty} \mathbf{P}\{D_t > rt\} = (1 - r)\left(1 + 2 \sum_{k=1}^{\infty} \frac{r^k}{k+2}\right). \tag{4}$$

Differentiating formula (4) with respect to r gives that the limit distribution has a density

$$h(r)=\tfrac{1}{3}+2\sum_{k=1}^{\infty} r^k\left(\frac{k+1}{(k+2)(k+3)}\right) \tag{5}$$

which is monotone increasing and $\sim -2\log(1-r)$ as $r\to 1$. This result shows that two individuals chosen at random from a critical branching process are fairly closely related.

We can also describe the relationships between more than two particles. To do this we consider $N_t(s)$ the number of individuals alive at time t which have offspring alive at time t. On $\{Z(t)=0\}$, $N_t(s)\equiv 0$. On $\{Z(t)>0\}$, $N_t(s)\geq 1$ and is monotone increasing. In Section 3 we compute that if $Z(t)$ is a Markov branching process, $0\leq r_1<r_2<1$, and $1\leq j\leq k$

$$\lim_{t\to\infty} \mathbf{P}(N_t(r_1t)=k \mid Z(t)>0)=r_1^{k-1}(1-r_1), \tag{6}$$

$$\lim_{t\to\infty} \mathbf{P}(N_t(r_2t)=k \mid N_t(r_1t)=j)=\binom{k-1}{k-j}\left(\frac{1-r_2}{1-r_1}\right)^j\left(\frac{r_2-r_1}{1-r_1}\right)^{k-j} \tag{7}$$

Since $\{N_t(rt),\ 0\leq r<1\}$ is a Markov process for each t it is easy to use (6), (7) and the monotonicity of the paths to show that for all $\varepsilon>0$, $\{N_t(rt),\ 0\leq r\leq 1-\varepsilon\}$ converges weakly as a sequence of random elements of $D[0, 1-\varepsilon]$ to a limit process with transition probabilities given by (6) and (7).

The limit result for $\{N_t(rt),\ 0\leq t<1\}$ gives a fairly complete description of the relationship in a critical branching process. In Section 4 we use this result to determine what occurs in a critical Markov branching process when we associate with each particle a position in $\mathbf{R}^d$ according to the following rules:

(i) A particle does not move during its lifetime.

(ii) When a particle located at x dies and has j offspring the positions $X_{1j},\dots,X_{jj}$ of its offspring have a joint distribution given by

$$\mathbf{P}(X_{1j}\leq x_1,\dots,X_{jj}\leq x_j\}=\psi_j(x_1-x,\dots,x_j-x). \tag{8}$$

Let ψ_{ij} be the distributon of X_{ij}, let p_j be the probability of having j offspring and let

$$\Psi(x)=\sum_{j=1}^{\infty} p_j\sum_{i=1}^{j}\psi_{ij}(x) \tag{9}$$

be the expected number of the offspring which will be born at a position $\leq x$ when the parent was located at 0.

These processes, called branching random walks, were introduced by Ney who studied them in the supercritical (age-dependent) case in (1965) and obtained a limit theorem for the number of particles in $\nu t+xt^{1/2}+(-\infty,0]^d$ where $\nu=\int y\Psi(dy)$.

Recently Kaplan and Asmussen (1976) have shown that if the displacements of the offspring from the parent are independent and if the distribution ψ_{ij} is independent of i and j then there is a limit theorem for the number of particles in $\nu t + xt^{1/2} + [0, h]^d$.

In Section 4 we will study the corresponding problems for critical Markov branching random walks under the assumption that

(a) the displacements of the offspring from the parent are independent and identically distributed, or

(b) $p_j > 0$ for only one $j > 0$.

Our main result is a limit theorem for the family history of a randomly chosen particle. This result states that if for each $t \geq 0$ we pick a particle at random from those alive at time t and let $Y_t(s)$ be the position of its ancestor which was alive at time s then as $t \to \infty$, $(t^{-1/2}(Y_t(rt) - \nu rt), 0 \leq r \leq 1)$ converges weakly to a Brownian motion with mean 0 and covariance $\Sigma'_{ij} = \int y_i y_j \Psi(\mathrm{d}y) - \nu_i \nu_j$.

Combining this conclusion with the results of Section 3 suggests that if we let $\eta_s \subseteq \mathbf{R}^d$ be the position of the parties alive at time s and define

$$\eta_s^t = \{x : x \in \eta_s \text{ and } x \text{ has offspring alive at time } t\},$$

then as $t \to \infty$ $(t^{-1/2}(\eta_{rt}^t - \nu rt), 0 \leq r \leq 1 \mid \eta_t \neq \emptyset)$ converges to a limit process which can be constructed from $N(s)$ (the limit of $N_t(s)$) by the following rules:

(i) One particle is present at time 0 and a new particle is born at each time of discontinuity of $N_t(s)$.

(ii) No particle dies before time $r = 1$.

(iii) All living particles move according to independent Brownian motions with mean 0 and covariance Σ'.

(iv) A new particle is born at the location of a particle chosen at random from those alive at the time of its birth.

The reason for interest in a "functional limit theorem" of this type is that it should imply that if we let $\eta_t(S)$ be the number of points in $\eta_t \cap S$ then the distribution of

$$(t^{-1}\eta_t(\nu t + xt^{1/2} + (-\infty, 0]^d) \mid \eta_t \neq 0) \tag{10}$$

converges to a limit which has the same distribution as the corresponding quantity for the conditional branching Brownian motion described above. This limit theorem, while plausible, is difficult to make rigorous. To do this we would have to construct a space of multi-valued functions in which the processes can converge, prove convergence of finite dimensional distributions, check tightness, and show that the functional we have defined above is continuous. The details are too lengthy to include in this paper but we may consider them in a future publication.

The situation becomes even more complicated when we consider limit theorems for $\eta_t(\nu t + xt^{1/2} + [0, h]^d)$. We conjecture that if A is a bounded open set with $|\partial A| = 0$ and we let $A_t = \nu t + xt^{1/2} + A$, then there are monotone functions H_1 and $H_{d,A}$, $d \geq 3$

so that for $x>0$:

$$\text{if } d=1 \quad \lim_{t\to\infty} (t)\mathbf{P}\{\eta_t(A_t)/|A|t^{1/2}>x\}=H_1(x),$$

$$\text{if } d=2 \quad \lim_{t\to\infty} (t\log t)\mathbf{P}\{\eta_t(A_t)/|A|\log t>x\}=4\,\mathrm{e}^{-x/8\pi},$$

$$\text{if } d\geqslant 3 \quad \lim_{t\to\infty} (t^{d/2})\mathbf{P}\{\eta_t(A_t)/|A|>x\}=H_{d,A}(x).$$

The first two results were proved for branching Brownian motions by Fleischman (1977) who gave the Laplace transform and some of the moments of $H_1(x)$. (It is not an exponential or gamma distribution.) The second result has been generalized to branching random walks (see Durrett (1977), Section 8). It seems likely that the first result also holds for branching random walks but I have not tried to show this. A more interesting unsolved problem is to show that the limits in $d=1,2$ hold for $x=0$. In the case $d=2$ this would say that $t\log t\ \mathbf{P}\{\eta_t(A)>0\}\to 4$.

The third statement is a new conjecture. This guess is supported by the observation that from the local central limit theorem (for random walks) $t^{d/2}\mathbf{E}\eta_t(A)$ is bounded away from ∞ and the fact that from Theorems 6.2 and 7.1 in Durrett (1977) $t^{d/2}\mathbf{P}\{\eta_t(A)>0\}$ is bounded away from 0.

2. The distribution of generations

In this section we will study the distribution of the branching process among generations and in particular obtain a limit law for the generation number of a particle chosen at random from those alive at time t. To do this we will begin by studying the distribution of the random time interval in which particles of generation n may be present. The left end point of this interval is β_n the time of the first birth of a member of generation n, the right end point is σ_n the time of death of the last member of generation n.

To study the limiting behaviour of β_n it is convenient to introduce a slightly smaller sequence of random variables. Let δ_{n-1} be the time a death first occurs to an individual of generation $n-1$. It is easy to obtain a recurrence relation for the distibution of the δ_n, $n\geq 1$. Let $K_n(t)=\mathbf{P}\{\delta_n-\delta_0\leq t\}$ (we set $\delta_n=\infty$ if the nth generation is empty). By conditioning on the number of offspring in the first generation we get

$$1-K_n(t)=\sum_{j=0}^{\infty} p_j(1-(K_{n-1}*G(t))^j$$

or

$$K_n(t)=1-h(1-(K_{n-1}*G(t)))$$

where h is the probability generating function of the offspring distribution $\{p_j, j\geq 0\}$.

Letting $Q(z)=1-h(1-z)$ we can write the above as

$$K_n(t)=Q(K_{n-1}*G(t)). \tag{1}$$

There is a similar recurrence relation which gives the distribution of σ_n. Let $L_n(t)=\mathbf{P}\{\sigma_n-\sigma_0\leq t\}$ (we set $\sigma_n=-\infty$ if the nth generation is empty). Now for the last death to occur before time t all the deaths have to occur before t so

$$L_n(t)=\sum_{j=0}^{\infty} p_j(L_{n-1}*G(t))^j=h(L_{n-1}*G(t)).$$

If we let $L_n^-(t)=\mathbf{P}\{\sigma_0-\sigma_n\leq t\}$ and $G^-(t)=\mathbf{P}\{-\sigma_0\leq t\}$, then

$$L_n^-(t)=1-h(L_{n-1}*G(-t-))=Q(L_{n-1}^-*G^-(t)). \tag{2}$$

Since K_n and L_n^- are defined by the same recurrence relation we will follow the approach of Hammersley (1974) and focus on the function Q to derive our limit theorems. To demonstrate which properties of Q are needed for our conclusions to hold we have formulated the next result in greater generality than is required by our applications.

Theorem 1. *Let Q be a concave nondecreasing function defined on $0\leq z\leq 1$ which has $Q(z)<z$ for all $z>0$, $Q(0)=0$, and $Q'(0)=1$. Suppose the sequence of distributions F_n is given by $F_{n+1}=Q(F_n*G)$ where $F_0(x)=0$ or 1 according as $x<0$ or $x\geq 0$ and where G is a proper distribution with mean μ. If we let $f_n=F_n(\infty)$, then $f_n\downarrow 0$ and*

$$f_n^{-1}F_n(nx)\to 1 \quad \text{for } x>\mu. \tag{3}$$

If (i) $\int_{-\infty}^{0} e^{\theta t}\,dG(t)<\infty$ *for some* $\theta<0$, *or*

(ii) $\int_{-\infty}^{0} t^2\,dG(t)<\infty$, $Q''(0)=-2\lambda<-\infty$ *and* $(Q(z)-(z-\lambda z^2))/z^2$ *decreases to* 0 *as* $z\downarrow 0$, *then*

$$f_n^{-1}F_n(nx)\to 0 \quad \text{for } x<\mu. \tag{4}$$

Let T_n be a random variable with distribution F_n. If (ii) *holds and* $\int_0^\infty t^2\,dG(t)<\infty$, *then* $\mathbf{E}(|T_n/n-\mu|\,|T_n<\infty)\to 0$.

Proof. We will first check that $f_n\downarrow 0$. From the recurrence relation

$$f_{n+1}=F_{n+1}(\infty)=Q(F_n*G(\infty))=Q(F_n(\infty)=Q(f_n)\geq f_n$$

so f_n is decreasing. Since $\lim_{n\to\infty} f_n$ is a fixed point of Q it follows from our assumptions that $f_n\downarrow 0$.

Let $C_n(t)=f_n^{-1}F_n(t)$. The distributions F_n have $F_{n+1}=Q(F_n*G)$ so the distributions C_n satisfy

$$C_{n+1}=f_{n+1}^{-1}Q(f_nC_n*G).$$

Let $Q_n(z)=f_{n+1}^{-1}Q(f_nz)$. Q_n is a concave function with $Q_n(0)=0$ and $Q_n(1)=1$ so $Q_n(z)\geq z$ for all n. Define a sequence of distributions by $G_{n+1}=G_n*G$, $G_0=C_0$.

If $C_n \geq G_n$, then

$$C_{n+1} = Q_n(C_n * G) \geq C_n * G \geq G_n * G = G_{n+1} \tag{5}$$

so it follows by induction that $C_n \geq G_n$ for all n. Since the mean of G is finite it follows from the weak large numbers that $G_n(nx) \to 1$ for $x > \mu$ and so $C_n(nx) \to 1$ for $x > \mu$.

To prove $C_n(nx) \to 0$ for $x < \mu$ we need to obtain an upper bound for the sequence C_n. To do this we observe $Q_n'(0) = f_n/f_{n+1}$ and Q_n is concave so $Q_n(z) \leq z f_n/f_{n+1}$ and

$$C_{n+1} = Q_n(C_n * G) \leq \frac{f_n}{f_{n+1}}(C_n * G).$$

Since $C_0 = G_0$ it follows by induction that

$$C_n \leq \prod_{m=1}^{n} \frac{f_{m-1}}{f_m} G_n = G_n/f_n \quad (f_0 = 1).$$

Again it follows from the weak law of large numbers that $G_n(nx) \to 0$ for $x < \mu$ but f_n also $\to 0$ so we cannot conclude that $C_n(nx) \to 0$ until we can show that G_n goes to 0 at a faster rate than f_n.

The conditions we need to do this depend upon the information we have about the rate of convergence of f_n to 0. From the definitions of Q and f_n it follows that $f_{n+1}/f_n = Q(f_n)/f_n \to 1$. To conclude that $C_n(nx) \to 0$ without additional information about the sequence f_n we have to show $G_n(nx) \to 0$ exponentially rapidly. This is guaranteed if $\varphi(\theta) = \int_{-\infty}^{\infty} e^{\theta t}\, dG(t) < \infty$ for some $\theta < 0$. To see this observe that

$$e^{\theta n x} G_n(nx) \leq \varphi(\theta)^n \quad \text{for all } \theta \leq 0,$$

so

$$G_n(nx) \leq \Big(\inf_{\theta \leq 0} e^{-\theta x} \varphi(\theta)\Big)^n.$$

Now $\lim_{\theta \uparrow 0} (e^{-\theta x}\varphi(\theta) - 1)/\theta = -(\mu - x)$ so if $x < \mu$, $\inf_{\theta \leq 0} e^{-\theta x}\varphi(\theta) < 1$. From this it is easy to conclude that

$$f_n^{-1} F_n(nx) = C_n(nx) \leq G_n(nx)/f_n \to 0$$

which completes the proof of (i).

To obtain results for the more general distributions allowed in (ii) we need more precise information about the sequence f_n. It follows from a result of Keten, Ney, and Spitzer (1966) that if $Q(z) = 1 - h(1-z)$ where h is a probability generating function with $h''(1) = 2\lambda \in (0, \infty)$, then $f_n \sim (n\lambda)^{-1}$ as $n \to \infty$. Only a few properties of probability generating functions are utilized in the proof of this result. From the proof given in Athreya, Ney (1972, pp. 20–21) it follows that the result may be reformulated as follows (see eq. (7) on p. 21 in [1]).

Lemma 2. *Suppose Q is a concave nondecreasing function with $Q(0) = 1$, $Q'(0) = 1$, and $Q''(0) = -2\lambda$, $\lambda \in (0, \infty)$. Let $f_0 = 1$ and for $n \geq 1$ let $f_n = Q(f_{n-1})$. If $(Q(z) - (z - \lambda z^2))/z^2$ decreases to zero as $z \downarrow 0$, then $f_n \sim (n\lambda)^{-1}$ as $n \to \infty$.*

To conclude that $C_n(nx)\to 0$ for $x<\mu$ now we have to show $nG_n(nx)\to 0$. To do this we will use a result of Baum and Katz (Theorem 4 in [2]): If $X_1, X_2, \ldots$ are independent and identically distributed and $S_n = X_1 + \cdots + X_n$ then the following two statements are equivalent for each $\alpha \geq 0$:

(a) $n^\alpha \mathbf{P}\{|S_n - n\mu| > n\varepsilon\} \to 0$ for all $\varepsilon > 0$, and

(b) $n^{\alpha+1}\mathbf{P}\{|X_1| > n\} \to 0$ and $\mathbf{E}[X_1 - \mu; |X_1| \leq n] \to 0$.

To use this theorem we have to truncate the distribution G. Let $X_1^M, X_2^M, \ldots$ be independent random variables with distribution

$$G^M(x) = \begin{Bmatrix} G(x) & x < M \\ 1 & x \geq M \end{Bmatrix}.$$

Let $S_n^M = X_1^M + \cdots + X_n^M$. Applying the result above we see that if $\int_{-\infty}^0 t^2 \, \mathrm{d}G(t) < \infty$ then $n^2\mathbf{P}\{|X_1^M| > n\} \to 0$ and so

$$n\mathbf{P}\{|S_n^M - n\mu_M| > n\varepsilon\} \to 0 \quad \text{for all } \varepsilon > 0$$

where $\mu_M = \int t \, \mathrm{d}G^M(t)$. Now

$$nG_n(nx) \leq n\mathbf{P}\{S_n^M \leq nx\} = n\mathbf{P}\{S_n^M - n\mu_M \leq n(x - \mu_M)\}$$

so if $x<\mu$ and M is chosen so that $x - \mu_M < 0$, then it follows that $f_n^{-1}F_n(nx) \leq f_n^{-1}G(nx) \to 0$ which completes the proof of (ii).

At this point we have shown that if T_n is a random variable with distribution F_n and (i) or (ii) holds, then conditionally on $T_n < \infty$, T_n/n converges in probability to μ. To complete the proof of Theorem 1 we have to show that if (ii) holds, then $\mathbf{E}(|T_n/n - \mu| \,|\, T_n < \infty) \to 0$.

The first step is easy and does not require assumption (ii). If $\varepsilon > 0$, then from (5)

$$\int_{-\infty}^{(\mu-\varepsilon)n} x/n \, \mathrm{d}G_n(x) \leq \int_{-\infty}^{(\mu-\varepsilon)n} x/n \, \mathrm{d}C_n(x) \leq (\mu-\varepsilon)C_n((\mu-\varepsilon)n).$$

We have already shown that the right hand side converges to 0. To prove that the left hand side has the same limit we observe that if S_n is the sum of n independent random variables with distribution G

$$\int_{-\infty}^{(\mu-\varepsilon)n} \mu - (x/n) \, \mathrm{d}G_n(x) \leq \mathbf{E}|S_n/n - \mu|$$

so

$$\int_{-\infty}^{(\mu-\varepsilon)n} x/n \, \mathrm{d}G_n(x) \geq \mu\mathbf{P}\{S_n/n \leq (\mu-\varepsilon)\} - \mathbf{E}|S_n/n - \mu|$$

and the right hand side converges to 0 by the mean ergodic theorem (see Breiman, p. 117).

To estimate the other tail we use the same approach. If $\varepsilon > 0$ from (6)

$$(\mu+\varepsilon)[1 - C_n((\mu+\varepsilon)n)] \leq \int_{(\mu+\varepsilon)n}^{\infty} x/n \, \mathrm{d}C_n(x) \leq f_n^{-1}\int_{(\mu+\varepsilon)n}^{\infty} x/n \, \mathrm{d}G_n(x)$$

and the left hand side converges to 0. To see that the right side also goes to 0 we use the following equality:

$$f_n^{-1}\int_{(\mu+\varepsilon)n}^{\infty} x/n \, \mathrm{d}G_n(x) =$$
$$= (nf_n)^{-1} n\mathbf{E}[S_n/n-\mu;\, S_n>(\mu+\varepsilon)n] + (nf_n)^{-1} n\mu \mathbf{P}\{S_n>(\mu+\varepsilon)n\}.$$

To estimate the first piece we observe that under our assumption $(nf_n)^{-1}\to\lambda<\infty$ and $\mathbf{E}(S_n/n-\mu)^2\to 0$ (see Chung, p. 103), so

$$n\mathbf{E}[S_n/n-\mu;\, S_n>(\mu+\varepsilon)n]\leq \varepsilon^{-1}\mathbf{E}[(S_n/n-\mu)^2;\, S_n>(\mu+\varepsilon)n]\to 0.$$

To complete the computation we observe that from the theorem of Baum and Katz it follows that $n\mathbf{P}\{S_n>(\mu+\varepsilon)n\}\to 0$ and so

$$\int_{(\mu+\varepsilon)n}^{\infty} x/n \, \mathrm{d}C_n(x)\to 0.$$

Combining this with the fact that the integral over $(-\infty, (\mu-\varepsilon)n)$ converges to 0 shows $\mathbf{E}(|T_n/n-\mu|\,|T_n<\infty)\to 0$ and comples the proof of Theorem 1.

Applying Theorem 1 to the sequences δ_n and σ_n defined earlier gives

Theorem 3. *For any lifetime distribution G with a finite mean μ $(\delta_n/n \mid \delta_n<\infty)$ converges to μ in probability. If G has a finite second moment, then $(\sigma_n/n \mid \sigma_n>-\infty)$ also converges to μ in probability.*

This result shows that if G has a finite second moment and the nth generation is not empty then most of the particles are born and die at times close to $n\mu$. This suggests that if U_t is the generation number of an individual picked at random from those alive at time t then $(U_t/t \mid Z(t)>0)$ converges in probability to $1/\mu$. To prove this we have to examine the effect on the nth generation of conditioning on $Z(t)>0$. The key to the analysis is the observation that from Theorem 3 if $\varepsilon>0$, then as $t\to\infty$ $(\delta_{t/(\mu-\varepsilon)}/t \mid \delta_{t/(\mu-\varepsilon)}<\infty)$ converges to $\mu/(\mu-\varepsilon)$ in probability so

$$\mathbf{P}(Z(t)>0 \mid \delta_{t/(\mu-\varepsilon)}<\infty)\to 1. \tag{7}$$

With this formula it is easy to compute that $(U_t/t \mid Z(t)>0)$ converges in probability to $1/\mu$. In fact we can prove a stronger statement:

Theorem 4. *If $t\to\infty$, then*

$$\mathbf{P}(\infty>\sigma_{\rho t/\mu}>t \mid Z(t)>0)\to 0 \quad \text{for } \rho<1$$

and

$$\mathbf{P}(\delta_{\rho t/\mu}<t \mid Z(t)>0)\to 0 \quad \text{for } \rho>1.$$

Proof. Let $\rho<1$ and $\varepsilon>0$.

$$\mathbf{P}(\sigma_{\rho t/\mu}>t \mid Z(t)>0)\le \mathbf{P}(\delta_{t/\mu-\varepsilon}=\infty \mid Z(t)>0) + \mathbf{P}(\sigma_{\rho t/\mu}>t \mid \delta_{t/\mu-\varepsilon}<\infty, Z(t)>0)\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty \mid Z(t)>0). \tag{8}$$

Now the second expression on the right hand side is

$$\le \mathbf{P}(\sigma_{\rho t/\mu}>t \mid \delta_{t/\mu-\varepsilon}<\infty, Z(t)>0)$$

and $\rho<1<\mu/\mu-\varepsilon$ so the above is

$$\le \frac{\mathbf{P}(\sigma_{\rho t/\mu}>t, \delta_{\rho t/\mu}<\infty)}{\mathbf{P}(\delta_{\rho t/\mu}<\infty)}\,\frac{\mathbf{P}(\delta_{\rho t/\mu}<\infty)}{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty)}\,\frac{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty)}{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty, Z(t)>0)}.$$

Now from Theorem 3, Lemma 2, and (7) we have

$$\mathbf{P}(\sigma_{\rho t/\mu}>t \mid \delta_{\rho t/\mu}<\infty)\to 0,$$

$$\mathbf{P}(\delta_{\rho t/\mu}<\infty)/\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty)\to \mu/\rho(\mu-\varepsilon)$$

and

$$\mathbf{P}(Z(t)>0 \mid \delta_{t/\mu-\varepsilon}<\infty)\to 1,$$

so the second term in the right hand side of (8) converges to 0. To estimate the other term we observe that

$$\mathbf{P}(\delta_{t/\mu-\varepsilon}=\infty \mid Z(t)>0)=1-\left(\frac{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty, Z(t)>0)}{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty)}\cdot\frac{\mathbf{P}(\delta_{t/\mu-\varepsilon}<\infty)}{\mathbf{P}(Z(t)>0)}\right)$$

and from (7) and Lemma 2 we have that the right-hand side converges to $1-(\mu-\varepsilon)/\mu=\varepsilon/\mu$. This shows that

$$\limsup_{t\to\infty} \mathbf{P}(\sigma_{\rho t/\mu}>t \mid Z(t)>0)\le \varepsilon/\mu \quad \text{for all } \varepsilon>0$$

which proves the first statement.

To prove the second statement we observe that if $\rho>1$

$$\mathbf{P}(\delta_{\rho t/\mu}<t \mid Z(t)>0)=\frac{\mathbf{P}(\delta_{\rho t/\mu}<t, Z(t)>0)}{\mathbf{P}(Z(t)>0)} \le \frac{\mathbf{P}(\delta_{\rho t/\mu}<t)}{\mathbf{P}(\delta_{\rho t/\mu}<\infty)}\cdot\frac{\mathbf{P}(\delta_{\rho t/\mu}<\infty)}{\mathbf{P}(\delta_{\rho t/\mu}<\infty, Z(t)>0)}$$

and the last expression converges to 0 by Theorem 3 and (7).

3. The degree of relationship

If $Z(t)>0$ and we pick two individuals at random (with replacement) from those alive at time t they will have a last common ancestor who died at a time $D_t\le t$. In this section we will obtain a limit law for $(D_t/t \mid Z(t)>0)$ and derive a result which describes precisely how all the individuals alive at time t are related.

The starting point for our analysis is the following formula. Let $f_t = \mathbf{P}\{Z(t)>0\}$, then

$$\mathbf{P}(D(t)>s \mid Z(t)>0) =$$
$$= f_t^{-1} \sum_{n=1}^{\infty} \mathbf{P}\{Z(s)=n\} \sum_{k=1}^{n} \binom{n}{k} f_{t-s}^k (1-f_{t-s})^{n-k} \mathbf{P}_{t,s,k} \tag{1}$$

where $\mathbf{P}_{t,s,k}$ is the probability two individuals chosen at random at time t have the same ancestor at time s conditioned on the event that exactly k individuals alive at time s have offspring at time t. The probability $\mathbf{P}_{t,s,k}$ may be computed as follows: If we let $X_1^u, X_2^u, \ldots$ be independent and identically distributed random variables with the same distribution as $(Z(u) \mid Z(u)>0)$ and if we let $S_k^u = X_1^u + \cdots + X_k^u$, then $\mathbf{P}_{t,s,k} = k\mathbf{E}(X_1^{t-s}/S_k^{t-s})^2$.

Now from [2, p. 113] as $u \to \infty$, $X_i^u/\lambda u$ converges to X_i^* an exponentially distributed random variable with mean 1. Since $0 \leq X_1^u/S_k^u \leq 1$ it follows that if $t-s \to \infty$

$$\mathbf{P}_{t,s,k} \to k\mathbf{E}(X_1^*/S_k^*)^2$$

where S_k^* is the sum of X_1^* and $k-1$ independent random variables with the same distribution. To compute $\mathbf{E}(X_1^*/S_k^*)^2$ we observe that if $u<v$ and $k \geq 2$

$$\mathbf{P}(X_1^* = u \mid S_k^* = v) = \mathrm{e}^{-u}\left(\frac{(v-u)^{k-2}}{k-2!}\,\mathrm{e}^{-(v-u)}\right)\Big/\left(\frac{v^{k-1}}{k-1!}\,\mathrm{e}^{-v}\right)$$
$$= (k-1)(v-u)^{k-2}/v^{k-1}.$$

Letting $x = u/v$ gives $\mathbf{P}(X_1^*/S_k^* = x \mid S_k^* = v) = (k-1)(1-x)^{k-2}$ and $\mathbf{P}(X_1^*/S_k^* \leq y) = 1-(1-y)^{k-1}$ for $0 \leq y \leq 1$. From this it follows that if $k \geq 2$

$$\mathbf{E}(X_1^*/S_k^*)^2 = \int_0^1 2y(1-y)^{k-1}\,\mathrm{d}y = 2/k(k+1)$$

and $\lim_{t-s\to\infty} \mathbf{P}_{t,s,k} = 2/k+1$. Observe that the formula is valid for $k=1$.

Having obtained this information it is easy to compute the limit of $\mathbf{P}\{D_t > s\}$ when $t \to \infty$ and $s/t \to r < 1$. To do this we observe that from [1, pp. 19–20], and a simple computation we have

$$f_t^{-1} f_s \sim t/s \to 1/r, \tag{3}$$

$$f_s^{-1}\mathbf{P}\{Z_s/\lambda s \in \mathrm{d}x\} \Rightarrow \mathrm{e}^{-x}\,\mathrm{d}x \tag{4}$$

and

$$\binom{\lambda sx}{k} f_{t-s}^k (1-f_{t-s})^{\lambda sx-k} \sim (k!)^{-1}(\lambda sx f_{t-s})^k (1-f_{t-s})^{\lambda sx} \tag{5}$$

so

$$\lim_{t\to\infty} \mathbf{P}\{D_t > s\} = r^{-1}\int_0^\infty dx\, e^{-x}\sum_{k=1}^\infty \frac{(xr/1-r)^k}{k!}e^{-xr/1-r}\frac{2}{k+1}$$
$$= 2r^{-1}\sum_{k=1}^\infty \left(\frac{r}{1-r}\right)^k\int_0^\infty dx\, e^{-x/1-r}\frac{x^k}{k+1!}.$$

The integral in the sum above is $(1-r)^{k+1}/k+1$ so the limit theorem can be written as

$$\lim_{t\to\infty} \mathbf{P}\{D_t > rt\} = (1-r)\left(2\sum_{k=1}^\infty \frac{r^{k-1}}{k+1}\right). \tag{6}$$

It is comforting to note that the limit is a distribution which is concentrated on [0, 1].

Differentiating the formula above with respect to r gives that the limit distribution has a density

$$h(r) = \tfrac{1}{3} + 2\sum_{k=1}^\infty r^k\left(\frac{k+1}{(k+2)(k+3)}\right) \tag{7}$$

which is monotone increasing and $\sim -2\log(1-r)$ as $r\to 1$.

The results given above describe the degree of relationship of two individuals chosen at random at time t. To describe the relationships in larger sets of particles we will consider $N_t(s)$ the number of individuals alive at time s which have descendants alive at time t. $N_t(s)$ is a nondecreasing function of s and *if* $k\geq 1$

$$\mathbf{P}(N_t(s)=k \mid Z(t)>0) = f_t^{-1}\sum_{j=k}^\infty \mathbf{P}\{Z(s)=j\}\binom{j}{k} f_{t-s}^k(1-f_{t-s})^{j-k}.$$

From (3), (4), and (5) it follows that if $t\to\infty$, then

$$\lim_{t\to\infty} \mathbf{P}(N_t(rt)=k \mid Z(t)>0) =$$
$$= r^{-1}\int_0^\infty dx\, e^{-x}\frac{(xr/1-r)^k}{k!}e^{-xr/1-r} = r^{k-1}(1-r).$$

This shows that as $t\to\infty$, $N_t(rt)-1$ converges weakly to a geometric distribution with success parameter $(1-r)$.

The next step in determining the limiting behavior of $\{N_t(rt), 0\leq r\leq 1\}$ is to compute that the finite dimensional distributions converge. Now the distribution of $(Z(s+u), 0\leq u\leq t-s \mid N_t(s)=k)$ is the same as that of the sum of k independent processes with the same distribution as $(Z(u), 0\leq u\leq t-s \mid Z(t-s)>0)$ so using the one dimensional convergence result shows that if $t\to\infty$ and $0\leq r_1<r_2<1$, then

$$\lim_{t\to\infty} \mathbf{P}(N_t(r_2t)=k \mid N_t(r_1t)=j) = \binom{k-1}{k-j}\left(\frac{1-r_2}{1-r_1}\right)^j\left(\frac{r_2-r_1}{1-r_1}\right)^{k-j} \quad \text{for } k\geq j. \tag{9}$$

The right-hand side is the probability of observing the jth occurrence of an event on the kth trial in a sequence of independent events with probability $(1-r_2)/(1-r_1)$ (see Feller, Vol. I, p. 165).

Since $\{N_t(rt), 0 \leq r \leq 1\}$ is a Markov process for each t combining the results of (8) and (9) shows that the finite dimensional distributions converge. Since each path of $N_t(rt)$ is monotone increasing and for all $r<1$ the collection $\{N_t(rt), t \geq 0\}$ of one dimensional distributions is tight it follows that as $t \to \infty$, $\{N_t(rt), 0 \leq r \leq 1-\varepsilon\}$ converges weakly as a sequence of random elements of $D[0, 1-\varepsilon]$ to a limit process $N_\infty(r)$ with transition probabilities given by (8) and (9).

The paths of the limit process are monotone increasing and have $N_\infty(r) \uparrow \infty$ as $r \uparrow 1$. From formula (9) it follows that if $r_1 < 1$, $j \geq 1$, $k \geq j+2$, then

$$\lim_{r_2 \to r_1} (r_2 - r_1)^{-1} \mathbf{P}(N_\infty(r_2) = k \mid N_\infty(r_1) = j) = 0$$

so $N_\infty(r)$, $0 \leq r < 1$ almost surely has no jumps of size ≥ 2. This result is the one we should expect—it is unlikely that two survivors will be born at the same time.

From the limit theorem for $\{N_t(rt), 0 \leq r \leq 1\}$ we can obtain results about the relationships in a set of $k \geq 3$ particles chosen at random (with replacement) from those alive at time t. To state these results we need to introduce some notation: for $1 \leq i < j \leq k$ let D_t^{ij} be the time of death of the last common ancestor of the particles which were chosen at the ith and jth selections. To obtain limit laws for $t^{-1} D_t^{ij}$ from those we have derived for $N_t(s)$ we observe that from the lack of memory property of the exponential distribution it follows that when $N_t(s)$ increases the new offspring is born to an individual chosen at random.

From this limit law for $N_t(rt)$ it is (theoretically) possible to calculate the limit distribution of $(D_t^{ij}, 1 \leq i < j \leq k)$ and obtain results for critical branching processes which correspond to the results obtained by Buhler in the supercritical case. We have done this for $k=3$. In this case either $D_t^{12} = D_t^{13}$, $D_t^{23} = D_t^{12}$, or $D_t^{13} = D_t^{23}$ so there are only two random variables to study $D_t' = D_t^{12} \wedge D_t^{13} \wedge D_t^{23}$ and $D_t'' = D_t^{12} \vee D_t^{13} \vee D_t^{23}$. We can compute the limit of D_t'/t in the same way we computed the limit of D_t/t.

The first step is to observe that

$$\mathbf{P}(D_t' > s \mid Z(t) > 0) =$$

$$= f_t^{-1} \sum_{n=1}^{\infty} \mathbf{P}\{Z(s) = n\} \sum_{k=1}^{n} \binom{n}{k} f_{t-s}^k (1 - f_{t-s})^{n-k} \mathbf{P}'_{t,s,k} \tag{10}$$

where $\mathbf{P}'_{t,s,k}$ is the probability that three individuals chosen from those alive at time t have the same ancestor at time s. If X_1^u and S_1^u are the random variables defined after formula (1) then

$$\mathbf{P}'_{t,s,k} = k\mathbf{E}(X_1^{t-s}/S_k^{t-s})^3.$$

From this it follows that if $t - s \to \infty$

$$\mathbf{P}'_{t,s,k} \to k\mathbf{E}(X_1^*/S_k^*)^3 \tag{11}$$

where X_1^* is an exponentially distributed random variable with mean 1 and S_k^* is the sum of X_1^* and $k-1$ independent random variables with the same distribution. From this it follows that

$$\mathbf{E}(X_1^*/S_k^*)^3 = \int_0^1 3y^2(1-y)^{k-1}\,dy = \frac{3\cdot 2}{k(k+1)(k+2)}.$$

Using (3), (4), (5), (10), and (11) gives that

$$\lim_{t\to\infty} \mathbf{P}(D'_t > rt \mid Z(t)>0) = \sum_{k=1}^{\infty} r^{k-1} \frac{3!}{(k+1)(k+2)} \int dx\, e^{-x/1-r} \frac{(x/1-r)^k}{k!}$$

$$= (1-r) \sum_{k=1}^{\infty} r^{k-1} \frac{3!}{(k+1)(k+2)}.$$

By proceeding in a similar way we can compute the limit of $(D''_t/t \mid Z(t)>0)$. To do this we observe that $\mathbf{P}(D''_t < rt \mid Z(t)>0)$ is equal to the right hand side of (10) with $\mathbf{P}'_{t,s,k}$ replaced by $\mathbf{P}''_{t,s,k}$ the probability that three individuals chosen at random from those alive at time t have three different ancestors at time s. If X_1^u and S_1^u are the random variables defined after formula (1), then

$$\mathbf{P}''_{t,s,k} = \mathbf{E}\Bigg[\sum_{\substack{h,i,j\\ \text{distinct}}} X_h^{t-s} X_i^{t-s} X_j^{t-s} / (S_k^{t-s})^3\Bigg].$$

Again as $u\to\infty$, $X_1^u/\lambda u$ converges in distibution to X_1^* an exponentially distributed random variable with mean 1 so whenever $t-s\to\infty$

$$\mathbf{P}''_{t,s,k} \to \iiiint \frac{uvw}{(u+v+w+x)^3} e^{-(u+v+w+x)} \frac{x^{k-4}}{k-4!}\,dx\,dw\,dv\,du.$$

If we let $\mathbf{P}''_{\infty,k}$ denote the right hand side of (13) then

$$\lim_{t\to\infty} \mathbf{P}(D''_t < rt \mid Z(t)>0) = (1-r) \sum_{k=1}^{\infty} r^{k-1} \mathbf{P}''_{\infty,k}.$$

By a similar, more tedious, computation we could compute the limit of $\mathbf{P}(D'_t < r_1 t < r_2 t < D''_t \mid Z(t)>0)$ but we would probably not obtain a useful formula.

4. Branching random walks

In this section we will consider the limiting behavior of critical Markov branching random walks. We will begin by considering the family history of a particle which is picked at random from those alive at time t. For each time $s \leq t$ this particle has a unique ancestor alive at time s. Let $Y_t(s)$ be the position of this ancestor and $V_t(s)$ be its generation number. Now since the particles have exponential lifetimes the survival of a "family name" from s until time t is independent of the history of the process before time s. From this we see that the distribution of $(V_t(s) \mid \eta_t \neq \emptyset)$ is the same as the distribution of $(U_s \mid \eta_t \neq \emptyset)$. With this fact it is easy to compute the limit of $V_t(rt)/t$. To do this we let $\varepsilon > 0$ and observe that

$$\begin{aligned} &\mathbf{P}(U_{rt} > (r+\varepsilon)t \mid \eta_t \neq \emptyset) \leq \\ &\quad \leq \mathbf{P}(U_{rt} > (r+\varepsilon)t \mid \eta_{rt} \neq \emptyset)\frac{\mathbf{P}(\eta_{rt} \neq \emptyset)}{\mathbf{P}(\eta_t \neq \emptyset)}. \end{aligned} \tag{1}$$

From Theorem 2.4 and Lemma 2.2 it follows that the right hand side of (1) converges to 0 and $\mathbf{P}(U_{rt} > (r+\varepsilon)t \mid \eta_t \neq \emptyset) \to 0$. A similar argument shows $\mathbf{P}(U_{rt} < (r-\varepsilon)t \mid \eta_t \neq \emptyset)$ so $V_t(rt)/t$ converges to r in probability.

From this we see that if $t \to \infty$ and $s/t \to r$ then $V_t(s)/t$ converges to r in probability. Since the limit of the one dimensional distributions are degenerate it follows that for any $0 \leq r_1 < r_2 \cdots < r_n \leq 1$, $(V_t(r_1 t)/t, \ldots, V_t(r_n t)/t)$ converges in probability to $(r_1, \ldots, r_n)$. Since $V_t(s)$ is a monotone function of s and function $\varphi(r) = r$ is continuous it follows that $V_t(rt)/t$ converges to φ in probability (as a sequence of random element of D).

At this point we would like to obtain a limit law for $\{Y_t(rt), 0 \leq r \leq 1\}$. To do this we need to know the distribution of a particle alive at time rt given that he has a descendant alive at time t. This distribution can be computed in the two cases mentioned in the introduction:

(a) the displacements of the offspring from their parents are independent and have a common distribution which does not depend on the number of offspring produced or

(b) $p_j > 0$ for only one $j > 0$.

Under either of these assumptions the steps taken by our particle and its ancestors are independent random variables with distribution Ψ. From this it follows that $Y_t(rt) = S_{V_t(rt)}$ where S_n, $n \geq 1$ is a random walk which is independent of $V_t(rt)$ and takes steps with distribution Ψ.

Let $\nu = \int y \Psi(\mathrm{d}y)$. It follows from the central limit theorem that for all $t \geq 0$ $(S_{[nt]} - nt\nu)/n^{1/2}$ converges to a d-dimensional Normal distribution with covariance $\Sigma'_{ij} = \int y_i y_j \Psi(\mathrm{d}y) - \nu_i \nu_j$. Now $Y_t(rt) \overset{\mathrm{d}}{=} S_{V_t(rt)}$ and we have shown that $V_t(rt)/t$ converges in probability to r so it follows from formulas (17.7)–(17.9) in Billingsley (1968) that

$$\{Y_t(rt), 0 \leq r \leq 1)\} \quad \text{converges weakly to} \quad \{W(r), 0 \leq r \leq 1\}, \tag{2}$$

a multidimensional Brownian motion with mean zero and covariance matrix Σ'.

References

[1] K. Athreya and P. Ney, Branching Processes (Springer-Verlag, Berlin, 1972).

[2] L. Baum and M. Katz, Convergence rates in the laws of large numbers, Trans. Amer. Math. Soc. 120 (1965) 108–123.

[3] J.D. Biggins, The first and last birth problem for a multi-type age-dependent branching process, Adv. Appl. Probability 8 (1976) 446–459.
[4] P. Billingsley, Weak Convergence of Probability Measures, (John Wiley, New York, 1968).
[5] L. Brieman, Probability (Addison Wesley, Reading, MA, 1968).
[6] W. Buhler, Generations and degree of relationship in supercritical Markov branching processes, Z. fur Wahr. und verw. Geb. 18 (1971) 141–152.
[7] W. Buhler, The distribution of generations and other aspects of the family structure of branching processes, Proc. Sixth Berkeley Symp. Vol. III (1972) 463–480.
[8] K.L. Chung, A Course in Probability Theory (John Wiley, New York, 1968; 2nd ed. 1974).
[9] R. Durrett, An infinite particle system with additive interactions, Advances in Appl. Probability, to appear.
[10] W. Feller, An Introduction to Probability Theory and Its Application, Vol. I (John Wiley, New York, 1950; 1970).
[11] J. Fleischman, Limiting distributions for branching random fields, to appear in Trans. Amer. Math. Soc. (1977).
[12] J. M. Hammersley, Postulates for subadditive processes, Ann. Probability 3 (1974) 790–801.
[13] N. Kaplan and S. Asmussen, Branching random walks, II, Stoch. Processes Appl. 4 (1976) 15–31.
[14] H. Kesten, P. Ney and F. Spitzer, The Galton Watson process with mean one and finite variance, Theor. Probability Appl. 11 (1966) 513–540.
[15] J.F.C. Kingman, The first birth problem for an age dependent branching process, Ann. Probability 2 (1975) 652–680.
[16] P. Ney, The convergence of a random distribution function associated with a branching process, J. Math. Analysis Appl. 12 (1965) 316–327.
[17] M. Samuels, Distribution of the branching process population among generations, J. Appl. Probability 8 (1971) 655–667.
[18] S. Sawyer, Branching diffusion processes in population genetics, Adv. Appl. Probability 8 (1976) 659–689.