SOME MATHEMATICAL MODELS FOR BRANCHING PROCESSES

T. E. HARRIS
THE RAND CORPORATION

1. Introduction

The stochastic processes variously called branching, birth, or multiplicative processes have been considered by writers in many different fields during the past eighty odd years. We shall not try to characterize such processes mathematically, although certain related mathematical properties will appear in all the processes we study. Physically speaking we may say that they represent the evolution of aggregates or systems whose components can reproduce, be transformed, and die, the transitions being governed by probability laws. The examples which have been most frequently considered in applications are the propagation of human and animal species and genes, nuclear chain reactions, and electronic cascade phenomena. The first, and probably best known mathematical model, which we shall consider in section 2, arose in connection with the problem of "the extinction of family surnames," and was treated by Galton and Watson [1] as far back as 1874.

As we should expect, the mathematical models which are simple enough to make possible a thorough analytic treatment of the subject are often radical oversimplifications of reality. Nevertheless, certain practical applications of the theory have been possible.

For a good historical account of the subject, including many references to applications, as well as interesting original work, we refer the reader to papers by M. S. Bartlett [2] and David G. Kendall [3]. Their bibliographies, together with that at the end of this paper, give (not completely exhaustive) references to much of what has been written in the field. It is unfortunate that some work done during the war, and classified, is still not available.

The present paper considers a number of stochastic processes which have been used as models for branching phenomena. We shall be particularly concerned with limiting theorems and limiting distributions giving the behavior of the systems studied after long periods of time. One pattern recurs often enough to make the following statement plausible, although a general mathematical formulation has not been given. It is strongly suggested by results of Everett and Ulam [4] and various results of Harris [5] and Bellman and Harris [6].

Consider a family of objects. Each object is described at a given instant of time by a vector quantity x, where x may describe the age, energy, position in space, or a combination of these or other traits. The quantity x for a given object may vary with time in a deterministic or a random fashion. In addition, there is a law for the probability that an object of "type" x, existing at time t, will produce (or be transformed into) a given aggregate of objects at time $t' > t$; for example, we may prescribe, for the disjoint sets $X_1, X_2, \ldots$, of x-values and the integers $k_1, k_2, \ldots$, the probability that starting with an object of type x at t, there will be k_1 objects at t' whose x-coordinates belong to X_1, k_2 whose coordinates belong to X_2, etc. Thus we might prescribe the probability that an object of age x be transformed into k_1 objects of age $\leqq x_1$ and k_2 of age $> x_1$.

Now let $N(t)$ be the number of objects at time t, and let $P_t(X)$ be the "distribution" of the population at t. [$P_t(X)$ is not a probability distribution; it is a random quantity giving the number of objects in the set X at t.] Assume that the system dealt with is one which will grow in size without limit as $t \to \infty$. Then, under various conditions, it will be true that

(a) $N(t)/E[N(t)]$ converges with probability 1 as $t \to \infty$ to a *random variable*, $E[N(t)]$ being the expected value of $N(t)$.

(b) $P_t(X)/N(t)$ converges with probability 1, in some sense, to a *constant* distribution $Q(X)$; (that is, the same for all realizations of the system).

It appears to be a matter of considerable interest to determine broad conditions under which (a) and (b) are true. They will be demonstrated for some of the systems considered in this paper.

In addition to limiting theorems we shall consider various results especially applicable to the classical model of Galton and Watson, and its multidimensional generalization. In particular we shall describe briefly some work of the Russian school not yet available in English.

2. The simple iterative scheme

In this section we consider the original Galton-Watson model. It has been used by many writers, and many of its properties have been discovered and rediscovered several times. In spite of its simplicity it is of considerable importance, partly because there are intrinsically interesting mathematical problems connected with it, many of them still unsolved; partly because many results connected with it can be wholly or partially generalized to more complicated models.

In this scheme we consider an initial object (ancestor) forming the zero generation. This object has probabilities p_r, $r = 0, 1, 2, \ldots$, of producing r objects, which will constitute the first generation. Each object in the first generation has the same probabilities as the ancestor of producing a given number of "children," independently of what is produced by any other object in its generation or preceding ones. Formally we can define the sequence of random variables z_n, $n = 0, 1, \ldots$, where z_n is the number of objects in the n-th generation, by

$$P(z_0 = 1) = 1,$$
$$P(z_1 = r) = p_r, \qquad r = 0, 1, \ldots,$$

and the requirement that if $z_n = j$, then z_{n+1} is the sum of j independent random variables each having the same distribution as z_1. (If $z_n = 0$, $z_{n+1} = 0$.)

We define the generating function of z_1 by

$$f(s) = \sum_{r=0}^{\infty} p_r s^r.$$

Throughout section 2 we shall assume, unless the contrary is stated, that $\sum r^2 p_r < \infty$. This insures the existence of second moments for all the random variables with which we shall be concerned. We shall also exclude the trivial cases (1) $p_0 + p_1 = 1$, and (2) $f(s) = s^k$ for some integer k.

The following facts are then well known.

(a) The generating function of z_n is $f_n(s)$, defined by

$$f_0(s) = s,$$
$$f_{n+1}(s) = f[f_n(s)], \qquad n = 0, 1, \ldots .$$

(b) Let

$$E z_1 = f'(1) = m,$$
$$\text{variance } (z_1) = f''(1) + m - m^2 = \sigma^2 .$$

Then

$$E z_n = m^n, \tag{2.1}$$
$$\text{var } (z_n) = \sigma^2 m^n (m^n - 1) / (m^2 - m), \qquad m \neq 1;$$
$$\text{var } (z_n) = n\sigma^2, \qquad m = 1.$$

(c) If $m \leqq 1$ the probability is 1 that $z_n = 0$ for some n. If $m > 1$, let a be the unique root in the half open interval [0, 1) of the equation

$$a = f(a).$$

Then with probability a, $z_n = 0$ for some n and with probability $1 - a$, $z_n \to \infty$.

(d) If $m > 1$ the random variable

$$w_n = \frac{z_n}{m^n}$$

converges in distribution to a random variable w whose moment generating function

$$\phi(s) = E e^{sw}$$

satisfies

$$\phi(ms) = f[\phi(s)], \qquad Re(s) \leqq 0. \tag{2.2}$$

As we shall see, many properties of the distribution of w can be deduced from (2.2).

The result (a) is originally due to Galton and Watson and has been rediscovered a number of times; (b) and (c) have likewise been found several times; (d) appears to be due first to Hawkins and Ulam [7] and was obtained independently by Yaglom [8].

We next consider convergence of the actual sample sequences w_n. For this purpose we note the important relations

$$E(z_{n+p} | z_n) = E(z_{n+p} | z_n, z_{n-1}, \ldots, z_0) \tag{2.3}$$
$$= m^p z_n, \qquad p = 0, 1, \ldots;$$
$$E(z_{n+p} z_n) = m^p E z_n^2, \qquad p = 0, 1, \ldots .$$

From (2.3), and the definition of $w_n = z_n/m^n$, we have

$$(2.4)\qquad \begin{aligned} E(w_{n+p}|w_n) &= E(w_{n+p}|w_n, w_{n-1}, \ldots, w_0) \\ &= w_n, \qquad p = 0, 1, \ldots; \\ E(w_{n+p}w_n) &= Ew_n^2, \qquad p = 0, 1, \ldots . \end{aligned}$$

The relations (2.3) and (2.4), or something analogous, hold in all the models we shall consider.

We have already mentioned that if $m \leqq 1$ the sequence z_n, and hence w_n, converges to 0 with probability 1. If $m > 1$ we have

THEOREM 1. *If $m > 1$, w_n converges to a random variable w with probability* 1.

The proof follows from (2.4), according to which

$$(2.5)\qquad E(w_{n+p} - w_n)^2 = Ew_{n+p}^2 - Ew_n^2 .$$

From (1),

$$(2.6)\qquad Ew_n^2 = 1 + \frac{\sigma^2}{m^2 - m} - \frac{\sigma^2}{m^n(m^2 - m)}$$

whence

$$(2.7)\qquad E(w_{n+p} - w_n)^2 = O(m^{-n}), \qquad m > 1 .$$

From (2.7) it follows that w_n converges in mean square to a random variable w and that

$$(2.8)\qquad \sum_{n=1}^{\infty} E(w - w_n)^2 = \sum_{n=1}^{\infty} O(m^{-n}) < \infty .$$

From (2.8) it then follows that w_n converges with probability 1 to w.

It should be noted that by virtue of (2.4) the random variables w_n form what Doob has called a *martingale*. Moreover, since $w_n \geqq 0$ we have $E|w_n| = Ew_n = 1$. Since the quantities $E|w_n|$ are uniformly bounded it follows from a theorem of Doob [9, p. 460] that the w_n converge with probability 1. Moreover, this argument does not require the existence of second moments, which we have assumed. However, the argument depending on second moments appears easier to generalize to more elaborate models. It also gives a bound for the rate of convergence.

The functional equation (2.2), sometimes called Koenigs' equation, sometimes Schröder's equation [10], [11], [12], after 19th century mathematicians who studied it, can be used to find the behavior of $\phi(s)$ on the imaginary s-axis, the negative real s-axis, and, if it exists, on the positive real s-axis. Then, using various kinds of Tauberian theorems, properties of the distribution of w can be inferred. Some details can be found in [5], but a great deal remains to be done in this direction.

We now consider some limiting theorems of a different sort. The fact that $z_n \to 0$ when $m \leqq 1$ makes the limiting situation look uninteresting. However, Yaglom [8] noticed that we get nontrivial limiting distributions in this case if we consider the conditional distribution of z_n, given that $z_n \neq 0$.

THEOREM 2 (Yaglom). *Let $g_n(s)$ be the conditional generating function of z_n, given $z_n \neq 0$,*

$$g_n(s) = \sum_{r=1}^{\infty} s^r P(z_n = r \mid z_n \neq 0) = \frac{f_n(s) - f_n(0)}{1 - f_n(0)} .$$

Then if $m < 1$,

$$\lim_{n \to \infty} g_n(s) = g(s)$$

where $g(s)$ *satisfies the functional equation*

$$g[f(s)] = m g(s) + 1 - m, \qquad |s| \leqq 1,$$

$$g(1) = 1, \quad g'(1-) = K,$$

$$K = \lim_{n \to \infty} \frac{1 - f_n(0)}{m^n}.$$

The proof is carried out using the classical work of Koenigs.

The limiting distributions considered in theorems 1 and 2 can assume a great variety of forms for any value of m, depending on the exact form of $f(s)$. It is therefore noteworthy that when $m = 1$ there is a universal limiting distribution, as seen in theorem 3, first proved for the special case $f(s) = e^{s-1}$ by Fisher [13].

THEOREM 3 (Yaglom). *Assume* $m = 1$ *and* $\sum r^3 p_r < \infty$. *Then*

$$\lim_{n \to \infty} P\{2 z_n / [n f''(1)] \leqq u \mid z_n \neq 0\} = 0$$

$$\text{if } u < 0 \text{ and } 1 - e^{-u} \text{ if } u \geqq 0.$$

If $m = 1$, $E(z_n | z_n \neq 0) \sim n f''(1)/2$, $n \to \infty$. The proof of theorem 3 is carried out using a theorem of Fatou [14] on iteration of functions in the neighborhood of a fixed point with derivative 1.

Another type of limiting distribution is of some interest in the case $m = 1$. We may consider the distribution of z_n, given $z_{n+p} \neq 0$, where p is a positive integer. The generating function of this distribution approaches a limit as $n \to \infty$, the limit being $s f_n'(s)$. In fact we can define in this way a conditional probability measure on the subspace (of zero measure) of sequences (z_n) which never vanish. A precise statement of the limiting result is clumsy, but it may be given informally as

THEOREM 4. *Assume* $m = 1, \sum r^3 p_r < \infty$. *Suppose that* n *and* $n' - n$ *are both large. If extinction has not occurred after* n' *generations then* $z_n/(1 + n\sigma^2)$ *has approximately the probability law whose density is* $4ue^{-2u}du$, $u > 0$.

The proof is by means of the theorem of Fatou, using the relation

$$f_n'(s) = \prod_{j=0}^{n-1} f'[f_j(s)].$$

Besides z_n, another random variable of interest is

$$Z = 1 + z_1 + \dots,$$

where Z is the total progeny produced in all generations. We have seen that if $m \leqq 1$, Z is finite with probability 1, and we can consider its probability distribution.

Let $Q(s)$ be the generating function for Z,

$$Q(s) = \sum_{r=1}^{\infty} s^r P(Z = r) = \sum_{r=1}^{\infty} Q_r s^r.$$

Hawkins and Ulam [7] and Otter [15] have shown that $Q(s)$ satisfies the functional equation

$$Q(s) = \sum_{r=1}^{\infty} Q_r s^r = sf[Q(s)] = s\sum_{r=0}^{\infty} p_r [Q(s)]^r. \tag{2.9}$$

Otter has investigated the equation (2.9) and has obtained an asymptotic expression for the coefficients Q_r.

Further results and details can be found in Otter's paper. Otter bases his work on a probability measure defined not on the space of sequences (z_n) but on the space of "trees." For example, one distinguishes the progeny of the second son of the fourth son, etc.

We shall only mention another group of problems about which little is known, those concerned with finding the distribution of upper or lower bounds for z_n as n ranges over various sets of values. The distribution for the number of generations to extinction, in case $m < 1$, has been discussed in [5]. Another distribution of interest, about which nothing seems to be known, is that of $\sup_{1\leqq n<\infty} z_n$, if $m \leqq 1$, and of $\sup_{1\leqq n<\infty} z_n/m^n$ if $m > 1$.

3. The multidimensional iterative scheme

We consider in this section the generalization of the model of section 2 to the multidimensional case. Specifically, we consider a sequence of vector random variables $\mathbf{z}_n = (z_n^1, \ldots, z_n^k)$ where z_n^i represents the number of objects of the i-th type in the n-th generation. (We shall use bold face lower case letters for vectors, bold face upper case letters for matrices.) The types $1, \ldots, k$ may be thought of as representing energy levels in the case of nuclear particles, or age groups in the case of biological organisms, etc. We assume that an object of type i existing in the n-th generation has a probability $p^i(r_1, \ldots, r_k)$ of producing in the next generation r_1 objects of type $1, \ldots, r_k$ objects of type k, independently of past history or of what is produced by other objects. The probabilities $p^i(r_1, \ldots, r_k)$, together with specification of the initial aggregate $\mathbf{z}_0$ determine the probability law for the sequence $(\mathbf{z}_n)$.

Much of the theory of these processes has been developed by Everett and Ulam [4], Sevast'yanov [16], and Sevast'yanov and Kolmogorov [17]. We shall summarize some of their work and give some further results. I wish to thank Drs. Everett and Ulam for permission to quote some of their results which have not yet appeared in the journals.

Define $f^i(\mathbf{s})$ and $f_n^i(\mathbf{s})$, $i = 1, \ldots, k$; $n = 1, 2, \ldots$, by

$$f^i(\mathbf{s}) = f_1^i(\mathbf{s}) = \sum_{r_1, \ldots, r_k \geqq 0} p^i(r_1, \ldots, r_k)\, s_1^{r_1} \ldots s_k^{r_k},$$

$$f_n^i(\mathbf{s}) = \sum_{r_1, \ldots, r_k \geqq 0} P(z_n^1 = r_1, \ldots, z_n^k = r_k)\, s_1^{r_1} \ldots s_k^{r_k}.$$

We then have the relations

$$f_{n+1}^i(\mathbf{s}) = f^i[f_n^1(\mathbf{s}), \ldots, f_n^k(\mathbf{s})], \qquad i = 1, \ldots, k; \quad n = 1, 2, \ldots. \tag{3.1}$$

Now define the first moment matrix

$$M = (m_{ij}),$$

$$m_{ij} = \left.\frac{\partial f^i}{\partial s_j}\right|_{s_1 = \ldots = s_k = 1},$$

where m_{ij} is the expected number of objects of type j produced in a single generation by a single object of type i. We exclude the trivial case where all the m_{ij} vanish.

Differentiation of (3.1) at $s_1 = \ldots = s_k = 1$ gives

$$Ez_n = z_0 M^n. \tag{3.2}$$

We shall consider the second moment matrix later. We shall assume that the second moments $E(z_1^i z_1^j)$, $i, j = 1, \ldots, k$, are finite.

Since all elements of M are nonnegative we can use the well known fact that M has a positive characteristic root λ which is at least as large in absolute value as any other characteristic root and which corresponds to a characteristic vector all of whose elements are nonnegative. If all the m_{ij} are positive, λ is simple and larger in absolute value than any other characteristic root, and every component of the corresponding characteristic vector is positive. We shall reserve the letter λ throughout section 3 for the largest positive characteristic root of M.

By analogy with the classification of states in Markoff chains we can introduce the notion of a *closed group* of types [17], a set of types whose progeny all belong to the set. A closed group is *indecomposable* if it does not contain two disjoint closed subgroups. If the types $1, \ldots, k$ form an indecomposable group, we shall speak of an *indecomposable system*. A closed group is called a *final group* if, with probability 1, the progeny in the next generation of an object in the group is exactly one object in the group, and if the group contains no proper closed subgroup with this property.

A process such that for any z_0 complete extinction is bound to occur is called *degenerate*.

THEOREM 5 (Sevast'yanov). *In order that a process be degenerate it is necessary and sufficient that (a)* $\lambda \leqq 1$ *and (b) there are no final groups.*

We shall say that a process is *completed* if only objects belonging to final groups remain. Suppose that there are K final groups, $K \leqq k$, and let $q^i(r_1, \ldots, r_K)$ be the probability that if the initial object is of type i, the process will be completed, with r_1 objects in final group $1, \ldots, r_K$ in final group K, remaining. For simplicity (and in connection with theorem 6 *only*) we suppose that an object which dies is transformed into a particular type which represents a death state. This type then forms a final group.

Let $H_1, \ldots, H_K$ be sets of integers, H_r being the numbers corresponding to the types in the r-th final group. Let

$$\psi^i(s) = \sum q^i(r_1, \ldots, r_K) s_1^{r_1} \ldots s_K^{r_K}.$$

THEOREM 6 (Sevast'yanov and Kolmogorov). *The functions* $\psi^i(s)$, *for* $|s| \leqq 1$, *are uniquely determined by the equations*

$$\psi^i(s) = f^i[\psi^1(s), \ldots, \psi^k(s)], \qquad i \notin H_1 + \ldots + H_K; \tag{3.3}$$
$$\psi^i(s) = s_r, \qquad i \in H_r, \qquad r = 1, \ldots, K.$$

The quantities $\psi^i(1) \leqq 1$ are the probabilities of "completion," as previously defined, if the initial ancestor is of type i, and can be obtained by solving (3.3) with $s = 1$.

If there are no final groups, then we may consider the probability a^i that extinction will occur, if the initial object was of type i (dropping henceforth the convention that the "death state" is one of the types). The quantities a^i are then determined by the equations

$$a^i = f^i(a^1, \ldots, a^k), \qquad i = 1, \ldots, k. \tag{3.4}$$

If we make the further assumption that for each $i, j = 1, \ldots, k$ there is a positive probability that an object of type i will have among its progeny in some future generation an object of type j, then $a^i = 1$ for a single i implies $a^i = 1$ for all i, and $a^i = 1$ if and only if $\lambda \leqq 1$.

Theorems 5 and 6, in slightly less general form, were proved by Everett and Ulam [4].

We shall say that a system is *positively regular* if λ is simple and larger in magnitude than any other characteristic root and if for every $i, j = 1, \ldots, k$ there is a positive probability that an object of type i will have in some generation of its progeny an object of type j. If every element of M is positive, the positively regular case is assured.

In the positively regular case we have, from matrix theory,

$$M = \lambda M_1 + N$$

where $M_1M_1 = M_1$, $M_1N = 0$, $NM_1 = 0$, and every element of M_1 is positive. The matrix M_1 has rank 1 and in fact has the form

$$M_1 = (\mu^i\nu^j)$$

where $\mu^1, \ldots, \mu^k$ are the components of the right eigenvector of λ and $\nu^1, \ldots, \nu^k$ are those of the left eigenvector. The components μ^i and ν^i are all positive, and with the proper normalization we then have

$$\sum_j m_{ij}\mu^j = \lambda\mu^i, \qquad \sum_i \nu^i m_{ij} = \lambda\nu^j, \qquad \sum_i \mu\,\nu^i = 1.$$

Moreover, there is an a_1, $0 < a_1 < 1$, such that

$$\frac{\|N^n\|}{\lambda^n} = O(a_1^n), \qquad n \to \infty, \tag{3.5}$$

where $\|N^n\|$ is the sum of the absolute values of the elements of N^n.

During the remainder of this section we consider only the positively regular case with $\lambda > 1$.

THEOREM 7a. *Suppose that* $\lambda > 1$ *and the system is positively regular. Then the* *random variable*

$$\frac{1}{\lambda_n}(z_n^1 + \ldots + z_n^k) = S_n$$

converges with probability 1 *to a random variable* S.

THEOREM 7b (Everett and Ulam). *Suppose* $\lambda > 1$ *and the system is positively regular. If the system does not become extinct the ratios* $z_n^1 \colon z_n^2 \colon \ldots \colon z_n^k$ *approach with probability* 1 *the ratios* $v^1 \colon v^2 \colon \ldots \colon v^k$ *of the components of the left eigenvector of* λ.

Theorem 7b, which appeared in a declassified Los Alamos report [4] written in 1948, is part of extensive results to be published later. As stated by the authors, it applied to the space of "trees" or "graphs"; compare the remark on Otter's work in section 2. As we shall see, 7b can be used to prove 7a. However, we shall also outline a simple method different from that of [4] which proves the two together, and gives an error term.

Analogously with (3.3), section 2, we have

$$E(z_{n+p} \mid z_n) = z_n M^p . \tag{3.6}$$

If we divide both sides of (3.6) by λ^{n+p} and multiply (scalar product) both sides on the right by the column vector μ', the right eigenvector of λ, we get

$$E\left(\frac{z_{n+p}\mu'}{\lambda^{n+p}} \,\middle|\, z_n\right) = \frac{z_n M^p \mu'}{\lambda^{n+p}} = \frac{z_n \mu'}{\lambda^n} . \tag{3.7}$$

Let ξ_n denote the scalar random variable defined by

$$\xi_n = \frac{z_n \mu'}{\lambda^n} .$$

Then (7) gives

$$E(\xi_{n+p} \mid \xi_n) = \xi_n . \tag{3.8}$$

Use of the theorem of Doob referred to in section 2 proves convergence with probability 1 of the sequence ξ_n. Knowing from theorem 7b that the direction of z_n (if it does not eventually vanish) approaches a limit, the fact which we have just observed, that the scalar product $z_n \mu'/\lambda^n$ converges, proves that each component of the vector z_n/λ^n converges, and theorem 7a follows.

As an alternative proof of theorems 7a and 7b we consider the sequence $w_n = z_n/\lambda^n$. From (3.6)

$$E(w_{n+p} \mid w_n) = \frac{w_n M^p}{\lambda^p} . \tag{3.9}$$

Relation (3.9) looks very much like its one dimensional analogue, (2.4). In fact, if M has an inverse, the resemblance can be made more striking. However, this line of argument, based on Doob's theorem, has not been carried out. Instead we argue on the moments as follows. From (3.9) and remarks made below it will appear that

$$E[(w_{n+p}^i - w_n^i)(w_{n+p}^j - w_n^j)] = O(a_2^n) , \qquad i, j = 1, \ldots, k ; \quad 0 < a_2 < 1 . \tag{3.10}$$

In particular,

$$(w_{n+p}^i - w_n^i)^2 = O(a_2^n) , \qquad i = 1, \ldots, k . \tag{3.11}$$

Thus for each i, w_n^i converges with probability 1 to a random variable w^i and we have theorem 7a. We shall see below that the second moment matrix of the w^i, (Ew^iw^j) has rank 1. Thus the w^i are perfectly correlated. Also, it can be shown that for each i, $w^i = 0$ if and only if extinction occurs, with probability 1. Hence, if $w^i \neq 0$ the ratio w^i/w^j is the same as the ratio Ew^i/Ew^j, with probability 1. Since, as we shall see, $Ew^i/Ew^j = \nu^i/\nu^j$, theorem 7b follows.

To obtain (3.10) we first examine the second moment matrix of z_n. If $u = (u^i)$ and $v = (v^i)$ are vectors we denote the matrix (u^iv^j) by $u'v$. If u and v are random variables then $E(u'v) = [E(u^iv^j)]$, while the *variance* of u is

$$E(u'u) - Eu'Eu = [E(u^iu^j) - Eu^iEu^j] .$$

We shall use M' to denote the transpose of M.

Let e_r be the k-component vector with 1 in the r-th place and zeros elsewhere. Define

$$B_r = E(z_1'z_1 \mid z_0 = e_r) = \Big(E(z_1^iz_1^j \mid z_0 = e_r)\Big);$$

$$E(z_1^iz_1^j \mid z_0 = e_r) = \frac{\partial^2 f^r}{\partial s_i \partial s_j}\bigg|_{s_1 = \cdots = s_k = 1} \qquad \text{if } i \neq j ;$$

$$\frac{\partial^2 f^r}{\partial s_i^2}\bigg|_{s_1 = \cdots = s_k = 1} = E[(z_1^i)^2 - z_1^i \mid z_0 = e_r] ;$$

$$V_r = \text{variance of } z_1, \text{ given } z_0 = e_r ,$$

$$= B_r - E(z_1' \mid z_0 = e_r)\, E(z_1 \mid z_0 = e_r) = B_r - M'e_r'e_rM ;$$

$$C_n = E(z_n'z_n) ; \qquad q_n = E z_n = (q_n^i) .$$

From elementary considerations we get

$$C_{n+1} = M'C_nM + \sum_{i=1}^{k} V_i q_n^i , \qquad n = 0, 1, \ldots ; \tag{3.12}$$

$$C_0 = z_0'z_0 ;$$

whence

$$\frac{C_n}{\lambda^{2n}} = \frac{M'^n}{\lambda^n} C_0 \frac{M^n}{\lambda^n} + \frac{1}{\lambda^2} \sum_{j=1}^{n} \frac{M'^{n-j}}{\lambda^{n-j}} \left(\sum_{i=1}^{k} V_i \frac{q_{j-1}^i}{\lambda^{2j-2}} \right) \frac{M^{n-j}}{\lambda^{n-j}} . \tag{3.13}$$

From (3.13) we derive

$$\frac{C_n}{\lambda^{2n}} = C + O(a_3^n) , \qquad 0 < a_3 < 1 ; \tag{3.14}$$

$$C = M_1' \Big\{ C_0 + \frac{1}{\lambda^2} \sum_{i=1}^{k} \delta^i V_i \Big\} M_1 ,$$

(recalling that $M_1 = \lim_{n \to \infty} M^n/\lambda^n$), where δ^i is the i-th component of the vector $z_0(I - M/\lambda^2)^{-1}$, I being the identity matrix.

From (3.14) and the relation of M_1 to M we see that C has rank 1 and satisfies

$$CM^n = \lambda^n C = M'^n C , \qquad n = 0, 1, \ldots . \tag{3.15}$$

Now consider $w_n = z_n/\lambda^n$. If n and p are nonnegative integers,

$$(3.16)\quad E[(w'_{n+p} - w'_n)(w_{n+p} - w_n)] = \frac{C_{n+p}}{\lambda^{2n+2p}} + \frac{C_n}{\lambda^{2n}} - E(w'_n w_{n+p}) - E(w'_{n+p} w_n).$$

But from (3.15), (3.14), and (3.6),

$$(3.17)\quad E(w'_n w_{n+p}) = \frac{E(z'_n z_{n+p})}{\lambda^{2n+p}} = \frac{E(z'_n z_n) M^p}{\lambda^{2n}\lambda^p} =$$

$$[C + O(a_3^n)] \frac{M^p}{\lambda^p} = C + O(a_3^n), \qquad 0 < a_3 < 1;$$

similarly $E(w'_{n+p} w_n) = C + O(a_3^n)$. From (3.17), (3.16), and (3.14), we have (3.10), q.e.d.

Regarding the distribution of w, we have

$$(3.18)\quad Ew = z_0 M_1 = \sum_i z_0^i \mu^i (\nu^1, \nu^2, \ldots, \nu^k),$$

$$Ew'Ew = M'_1 z'_0 z_0 M_1 = M'_1 C_0 M_1$$

and, from (3.14) and (3.18)

$$(3.19)\quad \operatorname{var}(w) = \frac{1}{\lambda^2} M'_1 \Big(\sum_{i=1}^{k} \delta^i V_i \Big) M_1 .$$

We recall that V_i is the variance matrix of z_1, given $z_0 = e_i$, and δ^i is the i-th component of the vector $z_0(I - M/\lambda^2)^{-1}$.

The w^i, being perfectly correlated, and the random variable S of theorem 7a, all have the same distribution except for constant factors.

Let $\phi^i(s)$ be the moment generating function of S, if there was initially one object of type i. It is then not difficult to show that the functions $\phi^i(s)$ satisfy

$$(3.20)\quad \phi^i(\lambda s) = f^i[\phi^1(s), \ldots, \phi^k(s)], \quad Re(s) \leqq 0, \quad i = 1, \ldots, k.$$

The functions $\phi^i(s)$ are uniquely determined by (3.20) and the requirements

$$\phi^i(0) = 1, \qquad \frac{d\phi^i}{ds}(0-) = \mu^i(\nu^1 + \ldots + \nu^k)$$
$$= E(S \mid z_0 = e_i),$$

where e_i has 1 in the i-th place and zeros elsewhere.

4. Continuous time parameter, Markov case

Feller [18] was apparently the first to discuss branching or birth processes where a continuous time parameter is involved, and since then there has been an extensive literature. For references we refer again to Kendall [3]. The point of departure for these treatments has usually been the specification of functions $b(n, t)$ and $d(n, t)$ where $b(n, t)\,dt$ is the probability of a birth and $d(n, t)\,dt$ the probability of a death between t and $t + dt$ if the size of the population at t is n. When these functions are specified, differential equations can be obtained for the probability of a given number of objects at t. Most treatments have assumed the birth and death rates to be independent of the age of the objects, although allowing them to depend on absolute time. We shall discuss the question of age dependence in section 4.

The model which we now consider is determined as follows. Consider an object existing at time t. Assume that there is a probability.

$$b_r \Delta t + o(\Delta t),$$

that the object is transformed into r objects, $r = 0, 2, 3, \ldots$ between t and $t + \Delta t$, $\Delta t > 0$, where

$$b = b_0 + b_2 + b_3 + \ldots < \infty,$$

and a probability $1 - b\Delta t + o(\Delta t)$ of not being transformed. We assume that the transformation probabilities are independent of the age of the object and the number of other objects existing. Then if there is initially a single object at $t = 0$, and if $p_r(t)$ is the probability that there are r objects at t we have the equations

$$\frac{dp_r(t)}{dt} = (r+1) b_0 p_{r+1}(t) - rbp_r(t) + (r-1) b_2 p_{r-1}(t) + \ldots + b_r p_1(t), \qquad r = 0, 1, \ldots \tag{4.1}$$

with the initial conditions

$$p_r(0) = 0, \qquad r \neq 1; \quad p_1(0) = 1. \tag{4.2}$$

We assume, henceforth, that the b_r are independent of time.

Various special cases of equations (4.1) have been studies both directly and by means of the generation function

$$F(s, t) = \sum_{r=0}^{\infty} p_r(t) s^r \tag{4.3}$$

which satisfies the equation

$$\frac{\partial F(s, t)}{\partial t} = \xi(s) \frac{\partial F(s, t)}{\partial s} \tag{4.4}$$

where

$$\xi(s) = b_0 - bs + b_2 s^2 + \ldots. \tag{4.5}$$

It is well known that if we define

$$f_n(s) = F(s, nh), \qquad n = 0, 1, \ldots, h > 0,$$

the functions $f_n(s)$ are the successive iterates of the function $f(s) = f_1(s) = F(s, h)$. Thus every scheme of the sort determined by equations (4.1) has a simple iterative scheme imbedded in it.

The converse of this statement is not true. It is clear that if $f(s)$ is an arbitrary generating function there is not in general a scheme defined by a set of equations (4.1), and a value t_0 of t such that

$$F(s, t_0) = f(s). \tag{4.6}$$

This is obvious, for example, if $f(s)$ is a polynomial of degree $\geqq 2$. We are thus led to ask under what circumstances it is possible, when a generating function $f(s)$

is given, to find a family of generating functions $F(s, t)$ obtained from equations of type (4.1), with $F(s, t_0) = f(s)$ for some t_0. To be precise we shall say that a probability generating function $f(s)$ belongs to *class C* (written $f \in C$) if there exists a family $F(s, t)$,

$$F(s, t) = \sum_{r=0}^{\infty} p_r(t)\, s^r$$

such that

$$\text{(4.7)} \quad \text{(a)} \ p_r(t) \geqq 0, \sum_{r=0}^{\infty} p_r(t) = 1;$$

$$\text{(b)} \ F[F(s, t_1), t_2] = F(s, t_1 + t_2), \qquad t_1 \geqq 0 \qquad t_2 \geqq 0, \qquad |s| \leqq 1 \cdot$$

$$\text{(c)} \ F(s, 1) = f(s) = \sum_{r=0}^{\infty} p_r s^r;$$

$$\text{(d)} \ F(s, t) = s + t\xi(s) + o(t), \qquad t \to 0, \qquad |s| \leqq 1;$$

where $\xi(s)$ is some function defined for $|s| \leqq 1$ and $o(t)/t \to 0$ uniformly in s. Some kind of regularity condition is necessary and (d) is convenient, although weaker looking assumptions could be substituted. From (d) we see that $\frac{\partial F}{\partial t}(s, t)\,|_{t=0} = \xi(s)$, and $\xi(s)$ is a power series convergent in the unit circle.

Using classical work on the iteration of functions and the general theory of Markov processes, we can now determine whether a given $f \in C$, provided $f(0) = p_0 = 0$. The literature on iteration goes back to Abel and is vast. I should like to thank Professor David Hawkins who first called this field to my attention.

We note that if $p_0 = p_1 = 0$, $f(s)$ can never belong to class C; in what follows we set aside this case as well as the trivial case $p_1 = 1$.

THEOREM 8a. *If* $p_0 = 0$, $0 < p_1 < 1$, *a necessary and sufficient condition that* $f(s)$ *belong to class C is that each of the quantities* b_r, $r = 2, 3, \ldots$, *should be non-negative, where the* b_r *are defined by the recurrence relations*

$$\text{(4.8)} \qquad b_r = \frac{1}{p_1 - p_1^r} \sum_{j=1}^{r-1} b_j [\beta_{rj} - (r - j + 1)\, p_{r-j+1}], \qquad r = 2, 3, \ldots;$$

$$b_1 = \log p_1 .$$

The exact value of $b_1 = -b$ is unimportant so long as it is negative; β_{rj} is the coefficient of s^r in $[f(s)]^j$.

The criterion of theorem 8a is not very satisfactory since it is often difficult to apply, and since it does not give any obvious relationship between membership in class C and the general analytic properties of $f(s)$. We can, however, give

THEOREM 8b. *If* $p_0 = 0$, $0 < p_1 < 1$, *and* $f(s)$ *is entire,* $f(s)$ *does not belong to class C.*

We show first that the criterion of theorem 8a is sufficient. From the classical work of Koenigs [10] we know that if $f(s)$ is any function analytic in a neighborhood of $s = 0$, with $f(0) = 0$ and $0 < f'(0) < 1$, there is a family

$$F(s, t) = \sum_{r=1}^{\infty} p_r(t)\, s^r$$

satisfying (4.7—b, c and d) in a neighborhood of $s = 0$. The function $\xi(s)$ is regular near $s = 0$ and satisfies

$$\xi[F(s, t)] = \xi(s)\frac{\partial F(s, t)}{\partial s}, \tag{4.9}$$

$$\frac{\partial F(s, t)}{\partial t} = \xi(s)\frac{\partial F(s, t)}{\partial s}, \tag{4.10}$$

$$\xi'(0) = \log p_1, \qquad \xi'(1-) = \log m, \tag{4.11}$$

which are evident from (4.7—b and d). Putting $t = 1$ in (4.9) gives

$$\xi[f(s)] = \xi(s)f'(s), \qquad |s| \leqq 1. \tag{4.12}$$

From (4.12) it is clear that $\xi(0) = 0$; (4.11) and (4.12) then determine uniquely the coefficients in the power series for $\xi(s)$, which are just the numbers b_r of (4.8) with $b_1 = \log p_1$. If the b_r, $r \geqq 2$, are nonnegative then $\xi(s)$ must be analytic in the interior of the unit circle since (4.12) shows that $\xi(s)$ is analytic for every real s between 0 and 1; moreover, from (4.12) we have $\xi(1) = 0$ so that

$$\sum_{j=1}^{\infty} b_j = 0. \tag{4.13}$$

From (4.10) we see that the $p_r(t)$ satisfy the equations

$$\frac{dp_r(t)}{dt} = \sum_{j=0}^{r-1} (r - j)\, b_{j+1}p_{r-j}(t), \qquad r = 1, 2, \ldots. \tag{4.14}$$

But now it is easily seen that equations (4.14), which are in the standard form of the differential equations for discontinuous Markov processes, have nonnegative solutions uniquely determined by the initial conditions

$$p_r(0) = 0, \qquad r \neq 1; \quad p_1(0) = 1,$$

which come from (4.7—d). The general theory insures that $\sum_{r=1}^{\infty} p_r(t) \leqq 1$, but the fact that $F(1, 1) = f(1) = 1$ shows that $\sum_{r=1}^{\infty} p_r(t) = 1$ for all $t \geqq 0$. Thus the nonnegativeness of the b_r is sufficient. The necessity is obvious from (4.14).

We shall obtain theorem 8b as a by product of theorem 9.

We observe that if $f \in C$, with $f(0) = 0$, $\sum r^2 p_r < \infty$ implies $\sum r^2 b_r < \infty$; that is, $\xi''(1-) < \infty$. This can be deduced from (4.12). Moreover, $\xi''(1-) < \infty$ implies $\sum r^2 p_r < \infty$; this seems to be difficult to obtain from (4.12) directly but can be demonstrated by actual construction of $F(s, t)$ as the solution of (4.10).

Suppose then that $f \in C$, $f(0) = 0$, $\sum r^2 p_r < \infty$. From section 2 we know that the random variables z_n/m^n converge with probability 1 to a random variable w whose moment generating function $\phi(s)$ satisfies $\phi(ms) = f[\phi(s)]$. Alternatively we can say that the random variable $z(t)$ whose probabilities are defined by (4.14) is such that for any $h > 0$ the sequence $z(nh)/Ez(nh)$, $n = 1, 2, \ldots$, converges with probability 1 to w.[1] Since it is often the probabilities b_j which are known initially rather than the p_r, it is convenient to determine $\phi(s)$ in terms of $\xi(s)$.

THEOREM 9. *Suppose* $f \in C$, $f(0) = 0$, $\sum r^2 p_r < \infty$. *The inverse function of* $\phi(s)$ *is given by*

$$\phi^{-1}(u) = -(1-u)\exp\left\{\int_1^u \left[\frac{\xi'(1)}{\xi(y)} + \frac{1}{1-y}\right] dy\right\}, \qquad 0 < u \leqq 1. \tag{4.15}$$

(We recall that $\xi'(1) = \log f'(1) = \log m$.) There are several ways of getting (4.15). We start from (4.12), which implies

$$\begin{aligned} \xi[f_n(s)] &= \xi(s) \prod_{j=0}^{n-1} f'[f_j(s)] \\ &= \xi(s) f_n'(s). \end{aligned} \tag{4.16}$$

Since the moment generating function of z_n/m^n is $f_n(e^{s/m^n})$ we have

$$\lim_{n\to\infty} f_n(e^{s/m^n}) = \phi(s), \qquad Re(s) \leqq 0, \tag{4.17}$$

$$\begin{aligned} \lim_{n\to\infty} \frac{d}{ds}[f_n(e^{s/m^n})] &= \lim_{n\to\infty} \frac{e^{s/m^n} f_n'(e^{s/m^n})}{m^n} \\ &= \phi'(s), \qquad Re(s) \leqq 0. \end{aligned} \tag{4.18}$$

We can justify (4.18) from the fact that $f_n(e^{s/m^n})$, $n = 1, 2, \ldots$, are moment generating functions for the random variables z_n/m^n, whose second moments are uniformly bounded. Now replace s by e^{s/m^n} in (4.16), $Re(s) \leqq 0$, and let $n \to \infty$. The left side of (4.16) approaches $\xi[\phi(s)]$ while the right side approaches $s\xi'(1)\phi'(s)$. We thus have the differential equation

$$\xi[\phi(s)] = \xi'(1)\, s\phi'(s), \qquad -\infty < s \leqq 0, \tag{4.19}$$

which, together with the condition

$$\left.\frac{d}{du}\phi^{-1}(u)\right|_{u=1-} = 1$$

gives (4.15).

Theorem 9 holds even if $f(0) \neq 0$ provided $m = f'(1) > 1$ [that is, even if $b_0 \neq 0$ provided $\xi'(1) > 0$]. As an example we consider the case

$$\xi(s) = \mu - (\mu + \lambda)s + \lambda s^2, \qquad 0 \leqq \mu < \lambda,$$

obtaining

$$\phi(s) = \frac{\mu}{\lambda} + \frac{\lambda - \mu}{\lambda} \cdot \frac{1}{1 - \dfrac{s\lambda}{\lambda - \mu}}.$$

Thus the random variable w is zero with probability μ/λ and otherwise has an exponential distribution. This is a special case of D. G. Kendall's "generalized birth and death process" [19].

[1] From the theory of Doob [9] it follows that if t_n is any sequence approaching ∞, $z(t_n)/Ez(t_n) \to w$ with probability 1.

Suppose now that $f(s)$ belongs to class C with $f(0) = 0$, and that $f(s)$ is entire. Then it is easily seen that $\phi(s)$ is entire, and $\phi^{-1}(u)/(1 - u)$ is analytic in a neighborhood of $u = 1$; using (4.15) this means that $\xi(s)$ is analytic in a neighborhood of $s = 1$. Since $f(s) > s$ for $s > 1$, and $f'(s) > 0$, this means that (4.12) can be used to extend $\xi(s)$ analytically to the whole positive axis. Since $1/\xi(y) = O(1/y^2)$, $y \to \infty$, we have

$$\int_2^\infty \frac{dy}{\xi(y)} < \infty . \tag{4.20}$$

From (4.20) and (4.15) we see that $\phi^{-1}(u)$ approaches a finite limit L as $u \to \infty$; that is, $\phi(L) = \infty$. But this contradicts the assumption that ϕ is entire. Thus theorem 8b is proved.

5. Age dependent processes

In the branching processes arising in biology the probability that an object existing at some time be transformed in a given time interval is not independent of the age of the object; in other words, the age specific birth and death rates are not constant. This means that the random variable $z(t)$, the number of objects at t, is not a Markov process. There are then several possibilities. We may accept the non-Markov character of $z(t)$ and work with it as well as we can; or we may choose to describe the state of the system at t by a *function* $z(t, x)$, the number of objects whose age is less than x, thus restoring the Markov character; or we may approximate by introducing a finite number of types, corresponding to age groups, and use a model of the multidimensional sort discussed in section 3 or the multidimensional extensions of the model of section 4.

We shall consider the following model. An object (of age 0) existing at $t = 0$ has a cumulative life length distribution $G(t)$. At the end of its life the object is transformed into r objects with probability q_r, $r = 0, 1, \ldots$, each having the same life length distribution $G(t)$, and so on. If the transformation is always binary we have the case of bacteriological fission, with which we shall be primarily concerned. We shall summarize some results obtained by Bellman and Harris [6] on the distribution of $z(t)$ and then consider the $z(t, x)$ process. Certain results about $z(t, x)$ for a related but more complicated model have been given recently by Kendall [3], who has also studied our variable $z(t)$ for the case where $G(t)$ is a convolution of exponential distributions [20].

Let

$$p_r(t) = P[z(t) = r], \qquad F(s, t) = \sum_{r=0}^{\infty} p_r(t) s^r .$$

If the initial object is transformed into r objects at time $y < t$, the generating function for the number of objects at t is $[F(s, t - y)]^r$. Thus we see that $F(s, t)$ satisfies

$$F(s, t) = \int_0^t h[F(s, t-y)]\, dG(y) + s[1 - G(t)] \tag{5.1}$$

where we have put

$$h(s) = \sum_{r=0}^{\infty} q_r s^r .$$

Equation (5.1) determines $F(s, t)$ when $G(t)$ and $h(s)$ are given. Arguments similar to those used for the model of section 2 show that there is a positive probability that $z(t)$ never vanishes (and therefore goes to ∞) if and only if $\sum rq_r > 1$.

If $G(t)$ is a step function with a single step, equation (5.1) gives the iterative scheme of section 2; if $G(t) = 1 - e^{-ct}$, where c is constant we have the Markov case of section 4 and in fact equation (5.1) can be reduced to a partial differential equation of the type of (4.4) in this case.

In the remainder of this section we shall be exclusively concerned with the case $\sum rq_r > 1$. For simplicity of exposition we restrict ourselves to the binary case $h(s) = s^2$. We also assume that $G(t)$ has a density function of bounded total variation,

$$G(t) = \int_0^t g(y)\,dy, \qquad \int_0^\infty |\,dg(y)\,| < \infty. \tag{5.2}$$

The case where $G(t)$ does not have a density is discussed in [21], a brief account of which appears in [6].

Equation (5.1) then takes the form

$$F(s, t) = \int_0^t F^2(s, t-y)\, g(y)\,dy + s\,[1 - G(t)]. \tag{5.3}$$

Similarly, defining

$$F_2(s_1, s_2; t_1, t_2) = \sum_{r_1, r_2} P[z(t_1) = r_1,\ z(t_2) = r_2]\, s_1^{r_1} s_2^{r_2}$$

we have, for $t_1 \leqq t_2$,

$$\begin{aligned} F_2(s_1, s_2; t_1, t_2) &= \int_0^{t_1} F_2^2(s_1, s_2; t_1 - y, t_2 - y)\, g(y)\,dy \\ &\quad + s_1 \int_{t_1}^{t_2} F^2(s_2, t_2 - y)\, g(y)\,dy + s_1 s_2\,[1 - G(t_2)]. \end{aligned} \tag{5.4}$$

Define

$$m_1(t) = Ez(t), \qquad m_2(t, h) = E[z(t)\, z(t+h)], \qquad h \geqq 0.$$

Then $m_1(t)$ satisfies the renewal equation

$$m_1(t) = 2\int_0^t m_1(t-y)\, g(y)\,dy + 1 - G(t)$$

and $m_2(t, h)$ satisfies a similar equation. Under the hypotheses (5.2) we have

$$m_1(t) = n_1 e^{\beta t}\,[1 + O(e^{-\epsilon_1 t})], \qquad \epsilon_1 > 0, \tag{5.5}$$

$$m_2(t, h) = \frac{2 n_1^2 I_2 e^{2\beta t + \beta h}\,[1 + O(e^{-\epsilon_2 t})]}{1 - 2 I_2}, \qquad \epsilon_2 > 0, \tag{5.6}$$

where

$$n_1 = \frac{1}{4\beta \int_0^\infty e^{-\beta y} y\, g(y)\,dy}, \qquad I_2 = \int_0^\infty e^{-2\beta y} g(y)\,dy \tag{5.7}$$

and β is the positive number satisfying

$$\frac{1}{2}=\int_0^{\infty} e^{-\beta y} g(y)\, dy . \tag{5.8}$$

It should be noted that the $O(e^{-\epsilon_2 t})$ in (5.6) is independent of h. The derivation of (5.5) and (5.6), using well known methods, is given in [21].

The importance of (5.6) is evident if we define

$$w(t) = \frac{z(t)}{n_1 e^{\beta t}} .$$

Formula (5.6) gives

$$E[z(t+h)\, z(t)] = e^{ah} E[z(t)]^2 [1 + O(e^{-\epsilon_2 t})] . \tag{5.9}$$

In the case of the Markov processes discussed in section 4, the $O(e^{-\epsilon_2 t})$ on the right side of (5.9) is replaced by 0. However, (5.9) is sufficient for our purpose, for using it shows that

$$E[w(t+h) - w(t)]^2 = O(e^{-\epsilon_3 t}) , \qquad t \to \infty , \qquad \epsilon_3 > 0 . \tag{5.10}$$

From (5.10) we have

THEOREM 10. *Under the assumptions*

$$h(s) = s^2 , \qquad G(t) = \int_0^t g(y)\, dy , \qquad \int_0^{\infty} |dg(y)| < \infty ,$$

the random variable $z(t)/(n_1 e^{\beta t})$ *converges to a random variable* w *with probability* 1 *in the sense that for each* $h > 0$ *the sequence* $z(nh)/(n_1 e^{\beta nh})$, $n = 1, 2, \ldots$, *converges with probability* 1 *to* w.

Rather than the sequence nh we could pick any sequence t_n such that $\sum_n O(e^{-\epsilon_3 t_n}) < \infty$. Presumably, $w(t)$ converges to w with probability 1 in the usual sense also.

Defining $\phi(s) = Ee^{sw}$, $Re(s) \leqq 0$, it can be shown that $\phi(s)$ satisfies

$$\phi(s) = \int^{\infty} \phi^2(se^{-\beta y})\, g(y)\, dy , \qquad Re(s) \leqq 0 . \tag{5.11}$$

From (5.11) can be obtained bounds for the magnitude of $\phi(it)$ and $\phi'(it)$, t real, as $t \to \pm \infty$, whence we get

THEOREM 11. *The distribution of* w *is absolutely continuous.*

Details are in [21].

We now consider the process $z(t, x)$, where

(5.12) $z(t, x)$ = number of objects in existence at time t of age $\leqq x$

and we introduce

(5.13) $M(t, y, x)$ = expected number of objects of age $\leqq x$ at t if there was one object of age y at time 0 .

The function $z(t, x)$ has been considered often in deterministic population studies. It is known that under certain conditions the age structure of a population (under deterministic assumptions) converges to a limiting value (see, for example, Leslie [22] and Lotka [23]). We shall give a probabilistic analogue of this result for our model which is also an analogue of the "ratio theorem" of Everett and Ulam (theorem 7b) and is likewise connected with a result of Doob [24] in renewal theory.

Let us first consider, heuristically, some properties of $M(t, y, x)$. Suppose the age structure of the population at some time t_1 is given by the function $z(t_1, x)$. The expected number of objects of age $\leqq x$ at time $t_1 + t_2$ is then

$$\int_0^{\infty} M(t_2, y, x)\, d_y z(t_1, y)\,. \tag{5.14}$$

Let $z(t)$ be the vector quantity representing $z(t, x)$, considered as a function of x. We may then define the operator $\boldsymbol{M}^{\tau}$ by the requirement that

$$\boldsymbol{M}^{\tau} H(x) = \int_0^{\infty} M(\tau, y, x)\, dH(y) \tag{5.15}$$

for any function $H(x)$ of bounded variation on $(0, \infty)$. The operators $\boldsymbol{M}^{\tau}$ have the property that

$$\boldsymbol{M}^{\tau_1}\boldsymbol{M}^{\tau_2} = \boldsymbol{M}^{\tau_1+\tau_2}\,.$$

We can now write the symbolic equation

$$E[\boldsymbol{z}(t+h) \mid \boldsymbol{z}(\tau),\ \tau \leqq t] = \boldsymbol{M}^{h}\boldsymbol{z}(t)\,, \tag{5.16}$$

which is the analogue of (2.3) and (3.6).

The operators $\boldsymbol{M}^{\tau}$ might be called "positive"; that is, they leave the cone of increasing functions of bounded variation invariant. Thus one would expect some of the classical theory on matrices with positive elements to carry over; in particular, the existence of a $\lambda > 0$ and an $A(x)$, an increasing function of x, such that for any $H = H(x)$ which is increasing and of bounded variation,

$$\lim_{t\to\infty} \frac{\boldsymbol{M}^{t}H}{\lambda^{t}} = c(H)\, A \tag{5.17}$$

where $c(H)$ is a linear functional of H; we have not made precise in what sense the limit in (5.17) should exist.

Although a general theory for positive operators has been given by Krein and Rutman [25] and Bohnenblust and Karlin [26], it does not appear to be readily applicable to the present case. However, we can get the results we need by using the fact that $M(t, y, x)$ satisfies renewal type equations.

We may define the age distribution of the population at t by the ratio $z(t, x)/z(t)$. We already know from theorem 10 the behavior of $z(t)$ for large t. We complete this with

THEOREM 12. *Define*

$$A(x) = \frac{\int_0^{x} e^{-\beta u}[1 - G(u)]\, du}{\int_0^{\infty} e^{-\beta u}[1 - G(u)]\, du}\,, \tag{5.18}$$

(5.19) $$D(t) = \sup_{0<x<\infty} \left| A(x) - \frac{z(t, x)}{z(t)} \right| .$$

Under the assumptions of theorem 10, $D(t) \to 0$ *with probability* 1 (*in the sense of theorem* 10) *as* $t \to \infty$.

The function $A(x)$ is the analogue of the "stable age distributions" considered in deterministic population theory.

Although for simplicity we are assuming an initial object of age 0 for theorems 10, 11 and 12, the modifications for an initial object of arbitrary age are obvious.

It is not hard to see that in order to prove theorem 12 it is sufficient to show that for *each* $x > 0$,

(5.20) $$\left| A(x) - \frac{z(t, x)}{z(t)} \right| \to 0, \qquad t \to \infty .$$

We do this by defining

(5.21) $$w(t, x) = \frac{z(t, x)}{n_1 e^{\beta t}}$$

and showing that $w(t, x) \to A(x)w$. The methods are similar to those of [21] and we only sketch the proof.

Let $F(y, x, s, t)$ be the generating function for the number of objects of age $\leqq x$ at t if initially there was one object of age y. Then F satisfies

(5.22) $$F(y, x, s, t) = sJ(x - y - t)[1 - G(y, t)] + \int_0^t F^2(0, x, s, t-u) \frac{g(y+u)\,du}{1 - G(y)},$$

where $J(t)$ is the Heaviside function: $J(t) = 0$ for $t < 0$, $J(t) = 1$ for $t \geqq 0$; $G(y, t)$ is the lifelength distribution for an object of age y,

$$G(y, t) = \frac{G(t+y) - G(y)}{1 - G(y)}$$

with the convention that $G(y, t)$ and $g(y + u)/[1 - G(y)]$ are to be taken as 0 if $G(y) = 1$.

Differentiation of (5.22) with respect to s at $s = 1$ gives

(5.23) $$M(t, y, x) = J(x - y - t)[1 - G(y, t)] + 2\int_0^t M(t-u, 0, x) \frac{g(y+u)}{1 - G(y)}\,du .$$

The assumptions we have made enable us to deduce from (5.23), in the manner of [21], that (putting $y = 0$)

(5.24) $$M(t, 0, x) = \frac{A(x)\,e^{\beta t}}{4\beta \int_0^\infty t e^{-\beta t} g(t)\,dt} + O(e^{(\beta-\epsilon)t}), \qquad t \to \infty;\ \epsilon > 0 .$$

We have used ϵ to represent a positive number, not necessarily the same each time.

Thus we see that the *expected* distribution of ages settles down to $A(x)$. To see that the *actual* distribution of ages does so we have to consider the joint distribution of $z(t_1, x_1)$ and $z(t_2, x_2)$.

Define $F(y, x_1, s_1, x_2, s_2, t_1, t_2)$ for $t_1 < t_2$ as the joint generating function for $z(t_1, x_1)$ and $z(t_2, x_2)$, given that the initial object had age y. This function satisfies an integral equation similar to (5.22), which we forbear to write down because of its length; differentiation of this equation gives a renewal type equation for the expected value of the product $z(t, x_1)z(t + \tau, x_2)$ given that the initial object was of age y. The methods used to obtain (5.24) then show that for any $\tau \geqq 0$

$$(5.25)\quad E[z(t, x_1)\, z(t+\tau, x_2)] = \frac{2n_1^2 I_2 A(x_1)\, A(x_2)\, e^{2\beta t+\beta\tau}[1+O(e^{-\epsilon t})]}{1-2I_2},$$

where n_1 and I_2 are defined by (5.7). Equation (5.25) is applicable when the initial object was of age 0. The term $O(e^{-\epsilon t})$ is independent of τ. From (5.25) we can show that for any $x_1 \geqq 0$, $x_2 \geqq 0$, $\tau \geqq 0$

$$(5.26)\quad E[A(x_2)\, w(t, x_1) - A(x_1)\, w(t+\tau, x_2)]^2 = O(e^{-\epsilon t}), \qquad \epsilon > 0,\ t \to \infty,$$

where $w(t, x)$ is defined by (5.21). From (5.26) we see that $w(t, x)$ converges in mean square to $A(x)w$, where w is the random variable of theorem 10. Theorem 12 follows from this fact.

6. Mutations

A multidimensional analogue for the model of section 5 can be constructed, in which k types of objects are considered. The asymptotic behavior of the moments can be discussed using systems of renewal type equations. We shall not pursue this topic further, but consider a special model for bacteriological mutation.

When bacteria are attacked by a bacterial virus, certain of the bacteria are sometimes resistant to the virus and can transmit this power of resistance to their descendants. *A priori*, two hypotheses would appear to be possible: (a) the resistant bacteria arose as mutations before the virus was added; (b) there is a small probability that any bacterium survives an attack of a virus; bacteria which survive an attack transmit immunity to their descendants. In case (a) the bacteria which survive the original onslaught will occur in "clones" of various sizes, each being the descendants of a mutant. In case (b) the survivors are randomly distributed throughout the medium.

The problem of distinguishing between (a) and (b) was attacked, using statistical methods, by Luria and Delbrück [27]. As part of the problem it is necessary to consider the distribution of the number of bacteria of the mutant form (that is, mutants or their descendants) at a given time if the hypothesis (a) is true. The model chosen by Luria and Delbrück was as follows. The main bacterial culture is assumed to grow deterministically, the number at time t being

$$N(t) = Ne^t$$

where N is the initial number. The probability that a mutation arises in the time interval $(t, t + dt)$ is taken as $\rho Ne^t dt + o(dt)$. The descendants of a mutant increase deterministically, the number of descendants at time τ after the mutant arises being e^τ.

Let $\zeta(t)$ be the number of the mutant form at time t. Under their hypotheses, Luria and Delbrück give the formulas

$$E\zeta(t) = \rho t N e^t$$
$$\text{variance}\,[\zeta(t)] = \rho N e^t (e^t - 1)\,.$$

The distribution of mutations has also been considered by Coulson and Lea, as mentioned in [28], who apparently assume deterministic growth for the main population while the mutants multiply in an age independent probabilistic fashion. They then determined the generating function for the number of the mutant form at a given time.

We shall consider the application of age dependent theory to the mutation problem. We retain the assumption that the parent culture grows deterministically. We have seen in section 5 that this is approximately true (under the hypotheses of that section) for cultures with a large number of individuals. Following Luria and Delbrück we take the size of the colony at t to be Ne^t, with a probability $\rho Ne^t dt + o(dt)$ of a mutation between t and $t + dt$. Since the number of mutants is relatively small, we take a probabilistic model for their growth, assuming that $F(s, t)$ is the generating function for the number of descendants of a mutant existing at a time t after creation of the mutant; $F(s, t)$ may be the function of section 5, but could be taken as some other function.

Let $H(N, s, t)$ be the generating function for the number of the mutant form at t, if the mother colony had size N at $t = 0$, no mutants then being present. The usual type of reasoning shows that

$$(6.1)\quad H(N,\ s,\ )t = e^{-\rho N(e^t-1)} + \int_0^t \rho N e^y e^{-\rho N(e^y-1)} F(s,\ t-y)\, H(N e^y,\ s,\ t-y)\, dy\,.$$

We can solve (6.1) easily if we think of our stochastic process as depending on the two "time" parameters N and t. Considered as a function of N, the process is infinitely divisible: for each $N_1 \geqq 0$, $N_2 \geqq 0$,

$$(6.2)\qquad H(N_1, s, t)\, H(N_2, s, t) = H(N_1 + N_2, s, t)\,.$$

From (6.2) we can write

$$(6.3)\qquad H(N, s, t) = \exp\,[NL(s, t)]$$

where $NL(e^s, t)$ is the cumulant generating function for the number of mutations at time t. To determine $L(s, t)$ we substitute (6.3) into (6.1). Considering the quotient

$$\frac{H(N,\ s,\ t) - e^{-\rho N(e^t-1)}}{N}$$

as $N \to 0$ shows that

$$(6.4)\qquad L(s,\ t) = -\rho(e^t - 1) + \rho\int_0^t e^y F(s,\ t-y)\, dy\,.$$

Letting $\zeta(t)$ be the number of the mutant form at time t, we have from (6.3) and (6.4)

$$(6.5)\qquad E[\zeta(t)] = \rho N \int_0^t e^y m(t-y)\, dy\,,$$

where $m(t) = \left.\frac{\partial F(s, t)}{\partial s}\right|_{s=1}$. If $m(t) \sim n_1 e^t$, it follows from (6.5) that

$$E[\zeta(t)] \sim \rho N n_1 t e^t. \tag{6.6}$$

Similar expressions for higher moments can be found.

It should be remembered that (6.6) remains valid only until the number of mutant individuals becomes nonnegligible compared with the nonmutant variety. Clearly it cannot hold for very large t since ultimately the expression on the right side of (6.6) becomes larger than Ne^t.

REFERENCES

[1] H. W. WATSON and F. GALTON, "On the probability of the extinction of families," *Jour. Anthrop. Inst.*, Vol. 4 (1874), pp. 138–144.

[2] M. S. BARTLETT, "Some evolutionary stochastic processes," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 11 (1949), pp. 211–229.

[3] D. G. KENDALL, "Stochastic processes and population growth," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 11 (1949), pp. 230–264.

[4] C. J. EVERETT and S. ULAM, "Multiplicative systems in several variables," I, II, III, AECD-2164 (Los Alamos declassified document 534), AECD-2165 (Los Alamos declassified document 533), and LA-707; 1948.

[5] T. E. HARRIS, "Branching processes," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 474–494.

[6] R. BELLMAN and T. E. HARRIS, "On the theory of age-dependent stochastic branching processes," *Proc. Nat. Acad. Sci.*, Vol. 34 (1948), pp. 601–604.

[7] D. HAWKINS and S. ULAM, "Theory of multiplicative processes," I, Los Alamos declassified document 265, 1944.

[8] A. M. YAGLOM, "Certain limit theorems in the theory of branching random processes," Russian, *Doklady Akad. Nauk. U.S.S.R.* (N.S.), Vol. 56 (1947), pp. 795–798.

[9] J. L. DOOB, "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp. 455–486.

[10] M. G. KOENIGS, "Recherches sur les intégrales de certaines equations fonctionnelles," *Annales Sci. de l'École Normale Sup. de Paris, Suppl.*, Vol. 1, Ser. 3 (1884), pp. S3–S41.

[11] ERNST SCHRÖDER, "Ueber iterirte functionen," *Math. Annalen*, Vol. 3 (1871), pp. 296–322.

[12] J. HADAMARD, "Two works on iteration and related questions," *Bull. Amer. Math. Soc.*, Vol. 50 (1944), pp. 67–75.

[13] R. A. FISHER, *The Genetical Theory of Natural Selection*, Oxford University Press, Oxford, 1930.

[14] P. FATOU, "Sur les equations fonctionnelles," *Bull. Soc. Math. de France*, Vol. 47 (1919), pp. 191–204.

[15] R. OTTER, "The multiplicative process," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 206–224.

[16] B. A. SEVAST'YANOV, "On the theory of branching random processes," Russian, *Doklady Akad. Nauk. U.S.S.R.*, Vol. 59 (1948), pp. 1407–1410.

[17] A. N. KOLMOGOROV and B. A. SEVAST'YANOV, "The calculation of final probabilities for branching random processes," *Doklady Akad. Nauk. U.S.S.R.* (N.S.), Vol. 56 (1947), pp. 783–786.

[18] W. FELLER, "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein," *Acta Biotheoretica*, Vol. 5 (1939), pp. 11–40.

[19] D. G. KENDALL, "The generalized birth-and-death process," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 1–15.

[20] ———, "On the role of variable generation time in the development of a stochastic birth process," *Biometrika*, Vol. 35 (1948), pp. 316–330.

[21] R. BELLMAN and T. HARRIS, "On the theory of age-dependent stochastic branching processes," to appear.

[22] P. H. LESLIE, "Some further notes on the use of matrices in population mathematics," *Biometrika*, Vol. 35 (1948), pp. 213–245.

[23] A. J. LOTKA, "Théorie analytique des associations biologiques," *Actualités Sci.*, No. 780, Hermann, Paris (1930), pp. 1–149.

[24] J. L. DOOB, "Renewal theory from the point of view of the theory of probability," *Trans. Amer. Math. Soc.*, Vol. 63 (1948), pp. 422–438.

[25] M. G. KREIN and M. A. RUTMAN, "Linear operators leaving invariant a cone in a Banach space," *Uspehi Matem. Nauk.* (N.S.), Vol. 3, No. 1 (23) (1948), pp. 3–95. See *Math. Reviews*, Vol. 10 (1949), pp. 256–257.

[26] F. BOHNENBLUST and S. KARLIN, to appear.

[27] S. E. LURIA and M. DELBRÜCK, "Mutations of bacteria from virus sensitivity to virus resistance," *Genetics*, Vol. 28 (1943), pp. 491–511.

[28] C. A. COULSON and D. E. LEA, see "Discussion on the papers," *Jour. Roy. Stat. Soc.*, Ser. B, Vol. 11 (1949), pp. 269–270.