Stochastic Processes and Population Growth

By David G. Kendall
(Magdalen College, Oxford)

(Received January 1st, 1949)

Contents

1. *Introduction*

(i) *Preliminary remarks.*—The classical theory of population growth treats the size of the population as a continuous variable, supposes this to be distributed continuously throughout the whole range of ages, and presumes the modifications produced in the course of time by the phenomena of birth, death and ageing to proceed in a deterministic manner. For its mathematical apparatus the theory is therefore able to adapt the differential and integral equations employed to describe similar phenomena in physical science. In practice the data are always grouped in intervals of age and epoch, and there is therefore great advantage in having also available the matrix technique developed by P. H. Leslie and others, which handles the problem in this discrete (but still deterministic) form throughout.

It is clear that a more refined analysis must take into account the role of chance effects in the development of the population, and starting with the work of W. Feller in 1939 a good deal of attention has recently been paid to this question. A large number of stochastic models have been constructed which describe, on various hypotheses, the chance fluctuations in the total population size, and it will be seen from the detailed account given in section 2 of the present paper that a fair approximation to the real situation can be obtained when the organism envisaged is of an elementary nature.

The application of the same technique to the description of human populations meets with two major difficulties. In the first place, although it has usually been found convenient to ignore numerical differences between the two sexes and to discuss only the growth of the female component of the population, the male component being supposed to adjust its numbers accordingly, attention has recently been drawn to the inadequacy of this approximation in some important problems.* Even in the deterministic theory it is difficult to improve on this procedure, but for a stochastic model the difficulties are still greater. It should also be noted that while in many problems it may be permissible to assume an equality in the expected numbers of the two sexes, it is unsatisfactory in a stochastic model to have to ignore chance departures from such a relationship. A few rather general comments on this problem will be given in section 2 (ix), but no serious attempt will be made here to solve it.

The second difficulty is associated with the age-distribution of the population, and the controlling influence which it exerts on the rate of growth by way (for example) of an age-specific birth rate. A solution to the problem of describing stochastic fluctuations in the age-distribution will be presented here in section 3, but it must be mentioned that a good deal of further work remains to be done. The methods used await a rigorous formulation, while the application of the solution to any practical case will lead to an integral equation† of a type rather less familiar than that which occurs in the deterministic theory.

It must, of course, be emphasized that in the large majority of population problems, especially on the human scale, the numbers of individuals are so large (the role played by chance fluctuations being correspondingly so small) that a description of the deterministic type provides virtually all that is required. Nevertheless in a growing number of particular problems a stochastic treatment has proved necessary, and this is one justification for the attempt made here to construct a sketch of a complete theory of population growth from the standpoint of the theory of probability. There is also a secondary value in investigations of this kind, for at a time when general stochastic processes are receiving such widespread attention there is much to be gained from the detailed study of a variety of particular examples.

(ii) *Sketch of the deterministic theory.*—The literature concerning population mathematics from the deterministic standpoint is now very extensive; a good general review will be found in a recent article by A. J. Lotka (1945) (see also the papers by E. C. Rhodes in the 1940 *Journal*), while for convenience of reference later it will be useful to recall here the simplest ideas and formulae of the subject.

In all forms of presentation, whether continuous or discrete, the fundamental assumption in the construction of "deterministic" models of population growth is as follows: *the future development of the population can be exactly predicted once its state at some initial epoch is completely specified.* This specification of the "state" of the population may have to be very detailed; for example, even if the development of the total population size alone is being considered, it will be necessary to specify the initial age-distribution if the birth and death rates are age-specific. The most elementary models, however, achieve their simplification by ignoring the age-structure of the population, and they involve the assumption that at any instant of time all its members have identical potentialities. If to this be added the further assumption that the members alive at a given epoch generate sub-populations in independence of one another, then it will follow that the total population size $N(t)$ must satisfy a differential equation of the form

$$\frac{dN}{dt} = \nu N \qquad (1)$$

where ν is the intrinsic rate of growth *per capita* per unit of time. The solution (when ν is constant) is, of course,

$$N(t) = N(0)\, e^{\nu t}, \qquad (2)$$

the law of Malthus, while if the intrinsic rate of growth is subject to a secular variation the solution is

$$N(t) = N(0) \exp \{ \int_0^t \nu(\tau)\, d\tau \} \qquad (3)$$

which assumes the previous form when the time-coordinate is suitably transformed.

If now the assumption of independence be dropped and replaced by the hypothesis that ν falls off linearly with an increase in the total population size (because of overcrowding, for

* See for example, P. H. Karmel (1947).
† Equation (111).

example), then with $\nu = \nu_0(1 - N/\alpha)$ (where ν_0 and α are constants) the differential equation takes on the form

$$\frac{dN}{dt} = \nu_0 N(1 - N/\alpha) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

with the solution

$$N(t) = \frac{\alpha N(0) e^{\nu_0 t}}{\alpha + N(0)[e^{\nu_0 t} - 1]}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

the "logistic" law of Verhulst, Pearl, and Reed. For small t this is not much different from the exponential law (2), but now as t tends to infinity the population size (if $\nu_0 > 0$) approaches the upper limiting value α. Laws of growth of the form (5) have been used with success to describe the growth-phenomena displayed in many sets of circumstances, both by bacteria and by human populations.

In the more detailed study of human populations consideration of the age-structure is both necessary and usual. The following account will illustrate briefly the type of analysis then employed; as explained in section 1 (i), it refers to the female component of the population only. Let $\lambda(x)\, dt$ denote the expected number of female children born to a female of age x in the infinitesimal time interval dt, and let $\mu\, dt$ be the probability that a female, alive at the epoch t, will die during the subsequent time-interval of length dt. (It is assumed here for simplicity that the death rate μ is not age-specific; this is admissible as a first approximation, and shortens the formulae considerably.) Then if $n(x, t)dx$ is the expected number of females at time t in the age-group $(x, x + dx)$, and if $b(t)dt$ is the expected number of female births occurring in the time-interval $(t, t + dt)$, it will follow that

$$n(x, t) = e^{-\mu x} b(t - x) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and

$$b(t) = \int_0^\infty \lambda(y)\, n(y, t)\, dy, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

so that

$$b(t) = g(t) + \int_0^t \lambda(y) e^{-\mu y} b(t - y)\, dy \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where

$$g(t) = e^{-\mu t} \int_0^\infty \lambda(t + Y)\, n(Y, 0)\, dY \quad . \quad . \quad . \quad . \quad . \quad (9)$$

the last expression being completely determined if the total population size and age-structure are known at the initial epoch $t = 0$.

The future (expected) history of the population can thus be fully described once the birth-function $b(t)$ is known, and this in turn is determined by the integral equation (8). This is of a type which has been extensively studied, and reference may be made to a paper by W. Feller (1941), which contains a very thorough investigation from the present point of view. There is a solution in terms of Laplace transforms which can in principle always be used when the functions $\lambda(y)$ and $g(t)$ are given in a definite analytical form, and this is, of course, fundamental to the theoretical discussion of (8), but for practical purposes one of three other methods is more suitable. The first of these starts with $g(t)$ as an approximation to $b(t)$, and obtains successively closer approximations by continued substitution in the integral equation. The second assumes for the solution an expansion of the form

$$b(t) = \Sigma\, b_r e^{\omega_r t},$$

and includes a procedure for computing the constants, while the third approximates to the integral in (8) by a finite sum. This last procedure is closely related to the work of P. H. Leslie (1945), who has developed the theory of the model with "discrete time" *ab initio*, and his method has the advantage that the age-distribution of the population at the times $t = 0, 1, 2, \ldots$, can be calculated successively by continually pre-multiplying a column-vector representing the initial age-distribution by a certain matrix. As no use will be made here of the representation in terms of "discrete time" the interested reader is referred to the papers of Leslie, and to the similar work of H. Bernardelli and E. G. Lewis mentioned in the list of references.

It is convenient to conclude this section with a discussion of the specially simple case when *both* of λ and μ are independent of the age of the individual. Although the assumption of a birth rate which is not age-specific would be most unrealistic, it corresponds to the simplest possible mathematical situation, and is a natural step to the consideration of more complicated models.

The first point to notice is that the form of the solution changes sharply as one crosses the locus $x = t$ (when $t = 0$ is the initial epoch). Above this line no consideration of the integral equation is required, and the solution is

$$n(x, t) = e^{-\mu t}\, n(x - t, 0) \qquad (x > t) \quad . \quad . \quad . \quad . \quad (10)$$

In the interval $(0 < x < t)$, however, the state of affairs is quite different, as one is then concerned with individuals born after the initial epoch. If the initial total population size is

$$N(0) = \int_0^\infty n(y, 0)\, dy,$$

the general solution can be found from that obtaining when $N(0) = 1$ by multiplying the right-hand side of each equation by $N(0)$.

When there is only a single "ancestor"

$$g(t) = \lambda\, e^{-\mu t},$$

and the integral equation takes the form

$$e^{\mu t}\, b(t) = \lambda + \lambda \int_0^t e^{\mu\tau}\, b(\tau)\, d\tau,$$

the (unique) solution being

$$b(t) = \lambda\, e^{(\lambda - \mu)t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

so that

$$n(x, t) = \lambda\, e^{-\lambda x + (\lambda - \mu)t} \qquad (0 < x < t) \quad . \quad . \quad . \quad . \quad (12)$$

Thus the population will grow or decrease exponentially, or will remain stationary, according as $\lambda > \mu$, $\lambda < \mu$ or $\lambda = \mu$, but in each case the relative age-structure in the *generated* population is the same, being the section of the distribution

$$\lambda\, e^{-\lambda x}\, dx \qquad (0 < x < \infty)$$

cut off by the interval $0 < x < t$.

(iii) *General discussion of stochastic models.*—It is evident that the deterministic theory of population growth outlined in the preceding section is not really adequate for the description of all the phenomena one might wish to study, for it fails to take into account the role of chance fluctuations in the development of the process. What is really needed is a stochastic theory, and the early recognition of this fact is well shown by a remarkable investigation which appeared in 1874.

In the previous year A. de Candolle, writing about the work of Francis Galton and others on the extinction of the families of men of note, remarked, "je n'ai pas rencontré la réflexion bien importante qu'ils auraient dû faire de l'extinction *inévitable* des noms de famille. Évidemment tous les noms doivent s'éteindre." Galton, however, had already been thinking about this matter, and in the *Educational Times* of 1873 he proposed the following problem: "A large nation, of whom . . . the adult males . . . each bear separate surnames, colonize a district. . . . Find . . . what proportion of the surnames will have become extinct after r generations." The only answer received being obviously false, Galton referred the question to his friend the Rev. H. W. Watson,* and their joint findings were published in 1874. Watson's calculations provide all that is necessary for the solution, and he recognized the inevitability of ultimate extinction in a population for which the average birth and death rates exactly balance, though he seems wrongly to have supposed that his conclusion would apply whatever the expected net rate of growth.

More than fifty years later† the same problem was considered by the late A. K. Erlang, and after his death it was proposed to the readers of *Matematisk Tidsskrift*; this time the complete solution was given by J. F. Steffensen (1930, 1933), at first in ignorance of the earlier work. He showed‡ that if

$$f(x) \equiv a_0 + a_1x + a_2x^2 + \ldots$$

is the function generating the probabilities $a_0, a_1, a_2, \ldots$, that a newborn male will eventually have 0, 1, 2, . . . sons, then there is always a positive chance that the male line of descent will ultimately become extinct, and (as Erlang had conjectured) it is given by the smallest root of the equation

$$f(x) = x$$

in the interval $a_0 \leqslant x \leqslant 1$. Ultimate extinction will be certain if and only if the expected number of male children,

$$a_1 + 2a_2 + 3a_3 + \ldots,$$

is less than *or equal* to unity.

The Watson-Steffensen formulae have since been applied by A. J. Lotka (1931) to the data for white males in the population of the United States (1920), the value found for the ultimate chance of extinction of a male line being 0·88.

Mention should be made at this point of the place occupied by the problem of the extinction of surnames in relation to the more general questions studied by R. A. Fisher (1930)§ and others concerning the extinction of rare genetic characters, and in this connection it is interesting to note that a surname is inherited as if it were associated with a gene totally sex-linked in Y.

All these stochastic models were formulated in terms of "discrete time," and it was apparently not until W. Feller's paper in 1939 that a systematic investigation was made of the possibility of treating population growth as a *temporally continuous* stochastic process.|| The cardinal assumption is now that the growth of a population can be represented by the development of a Markoff process; this is to say, that the state of the population at time t can be described by the value of a random variable $\boldsymbol{X}(t)$ with the following property:

$$\text{Distr}\,\{\boldsymbol{X}(t) \mid \boldsymbol{X}(t_0)\} \equiv \text{Distr}\,\{\boldsymbol{X}(t) \mid \boldsymbol{X}(\tau) \text{ for all } \tau \leqslant t_0\}, \text{ whenever } t_0 < t.$$

The nature of the variable $\boldsymbol{X}(t)$ differs from model to model. In the simplest birth-and-death process, to be discussed in section 2 (i), it is a non-negative integer measuring the total population

* Henry William Watson (1827–1903), Sc.D., F.R.S., was 2nd Wrangler in 1850. He was at one time Fellow of Trinity College, Cambridge, but resigned on his marriage, and his original mathematical work was all done in the country parish of Berkswell, which living he held from 1865 until shortly before his death in 1903. He is chiefly remembered as one of the founders of the Alpine Club in 1857, and as the author of books on mathematical physics. One of these, on the kinetic theory of gases, appeared in 1876, two or three years after the events just described. There is an account of his collaboration with Galton in Karl Pearson's *Life* of the latter, which also contains an interesting series of letters to de Candolle.

† More recently some very general questions of this sort have engaged the attention of the Russian school of mathematicians. A systematic review of their work is difficult because of the language obstacle, but some notes have been included in the list of references (A. N. Kolmogoroff, N. A. Dmitriev, A. M. Yaglom, and B. A. Savost'yanov; 1947 and 1948).

‡ See also the discussion of the problem in Professor Bartlett's contribution to this symposium; he refers to a similar but independent solution by I. J. Good.

§ An earlier treatment of the problem of the survival of individual genes will be found in Fisher's paper of 1922, while a recent review (with some further developments) has been given by J. B. S. Haldane (1949). A closely related problem arising in electronics has been discussed by P. M. Woodward (1947).

|| After writing this I came across a most interesting paper (1914) by the late A. G. McKendrick (1876–1943), in which temporally continuous processes are introduced and studied in relation to a group of biological problems. McKendrick's work will be referred to again in section 2 (v) of the present paper; his other relevant publications (often in journals not usually seen by mathematicians) are listed in the obituary notice by W. F. Harvey in the *Edinburgh Medical Journal* (**50**, 500–6).

size, while in the multiple-phase process to be described in section 2 (viii) it is a vector variable whose several components enumerate the individuals in the several phases. The most complicated situation occurs, however, in section 3 when $\boldsymbol{X}(t)$ is a "portmanteau" variable describing not only the size of the total population, but also the exact age of each individual in it. This type of generalization is inevitable if a more realistic description of the actual process is sought and at the same time the "Markoff" property is to be retained, for it is then essential that a statement of the realized "value" of $\boldsymbol{X}(t)$ shall specify completely everything about the state of the population which is relevant to the prediction of its future development.

In all the models to be discussed in this paper, however, the variable $\boldsymbol{X}(t)$ will play what is essentially an enumerative role. Thus, in the most complicated case, its "value" at time t will be the *function* $N(x, t)$ which enumerates completely the actual population in the sense that*

$$\int_{x_1}^{x_2} dN(x, t)$$

is the actual number of individuals which at time t occupy the age-group (x_1, x_2). This is in contrast to one of the models discussed by Feller in his 1939 paper in which the total population size (age-structure being ignored) is treated as a *continuous* variable (instead of as an integer changing only by discrete jumps). Feller's reasons for adopting this mode of description are expressed with reference to a continuously changing "total life-energy." I do not know how far this is an appropriate conception, but the mathematical techniques employed by Feller in the development of the idea are rather different from those with which this paper is concerned, and it is not convenient to discuss them further here.

2. *Stochastic Fluctuations in the Total Population Size*

(i) *The simple birth-and-death process.*—The first example of a population process of the present type was apparently that introduced by W. H. Furry in 1937. The physical application which he had in mind will be mentioned in section 2 (iv); for the moment his model will be considered with its biological interpretation. From this point of view the system under discussion is a population of organisms multiplying in accordance with the following rules:

(*a*) the sub-populations generated by two co-existing individuals develop in complete independence of one another;

(*b*) an individual existing at time t has a chance

$$\lambda\, dt + o(dt)$$

of multiplying by binary fission during the following time-interval of length dt;

(*c*) the "birth rate" λ is the same for all individuals in the population at all times t.

The rule (*b*) will usually be interpreted in the sense that at each birth just one new member is added to the population, but of course mathematically (and because the age-structure of the population is being ignored) it is not possible to distinguish between this and an alternative interpretation in which a parent always dies at birth and is replaced by two new members.

Let N_0 be the number of individuals at the initial time $t = 0$, and suppose for the moment that $N_0 = 1$; let $p_n(t)$ be the probability that the population size $N(t)$ has the value n at time t. Then

$$\frac{d}{dt}\, p_n(t) = (n-1)\,\lambda\, p_{n-1}(t) - n\lambda\, p_n(t) \qquad (n > 1),$$

and

$$\frac{d}{dt}\, p_1(t) = \qquad -\lambda\, p_1(t), \qquad \ldots \ldots \quad (13)$$

and so the probability-generating function

$$\varphi(z, t) = \sum_{(n)} z^n\, p_n(t) \quad \ldots \ldots \quad (14)$$

* Here and elsewhere in this paper the Stieltjes integration is to be understood as being performed with regard to the *age* variable, x, in $dN(x, t)$.

must satisfy the partial differential equation

$$\frac{\partial \varphi}{\partial t} = \lambda\, z(z-1) \frac{\partial \varphi}{\partial z}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

the most general solution to which is of the form

$$\varphi(z, t) = \Phi\{(1 - 1/z)\, e^{\lambda t}\}.$$

But

$$\varphi(z, 0) = \Phi\{1 - 1/z\} = z,$$

and so

$$\Phi(Z) = 1/(1 - Z),$$

and

$$\varphi(z, t) = ze^{-\lambda t}\,[1 - z(1 - e^{-\lambda t})]^{-1} \quad . \quad . \quad . \quad . \quad (16)$$

Expansion in powers of z now gives the probability distribution for the population size at any time t; it is

$$p_n(t) = e^{-\lambda t}\,(1 - e^{-\lambda t})^{n-1} \qquad (n \geqslant 1), \quad . \quad . \quad . \quad . \quad (17)$$

a distribution of geometric form, the common ratio of which approaches unity as t tends to infinity.

From this simple example one can begin to appreciate a number of points which prove to be of more general importance. In the first place, because of the assumption (a), the population size at time t, when $N_0 > 1$, is the sum of N_0 independent variables for each of which the probability distribution is given by the law (17), and so the distribution of $N(t)$ in general can readily be obtained by raising the function given by (16) to the power N_0, and then expanding in powers of z as before. The resulting distribution is obviously of negative-binomial type.

Again, when $N_0 = 1$, the mean $\bar{N}$ of $N(t)$ can readily be calculated to be

$$\bar{N}(t) = \frac{\partial}{\partial z}\,\varphi(1, t) = e^{\lambda t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and similarly for the variance of $N(t)$ one finds

$$\operatorname{Var}\{N(t)\} = \frac{\partial^2}{\partial z^2}\,\varphi(1, t) - \bar{N}(\bar{N} - 1) = e^{\lambda t}\,(e^{\lambda t} - 1) \quad . \quad . \quad . \quad (19)$$

When $N_0 > 1$, the right-hand side of each equation is to be multiplied by N_0.

It will be noticed that the mean growth of the process, given by (18), follows the same exponential law as that which appeared in (2) as the solution to the simplest deterministic equation (with λ now taking the place of ν). All the other stochastic models to be examined, with one exception, similarly mimic in their mean behaviour the corresponding deterministic model; this is the justification of the assertion sometimes made that the deterministic theory is simply an account of the expectation behaviour of the random variables which occur in the stochastic formulation. That this is *not* generally true was first pointed out by Feller (1939), and the example to which he was then referring (of great importance for its own sake) will be mentioned here in section 2 (vii).

In the deterministic theory it made no difference whether the intrinsic rate of growth ν was purely reproductive in origin, or was really a balance, $\lambda - \mu$, between a birth rate λ and a death rate μ. In the stochastic theory this is no longer true, and the birth-and-death process is quite distinct from the pure birth process just described.

The birth-and-death process was introduced by Feller (1939), who obtained the formulae (23) and (24) below. It was reconsidered in connexion with a physical application by N. Arley (1943), and the equations giving its probability-structure were first solved by C. Palm in an unpublished letter to Arley. (See Arley and Borchsenius (1945), for an account of Palm's work.) The difference from the Furry process is that there is now a mortality effect acting in the following way:

(*d*) an individual existing at time t has a chance

$$\mu dt + o(dt)$$

of dying in the following time-interval of length dt;

(*e*) the "death rate" μ is the same for all individuals at all times t.

The equation to be satisfied by the generating function is now

$$\frac{\partial \varphi}{\partial t} = (\lambda z - \mu)(z - 1)\frac{\partial \varphi}{\partial z} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (20)$$

and in accordance with a remark made in one of the preceding paragraphs it will be enough to state the solution when $N_0 = 1$, since the solution for a general number of ancestors can be obtained from this. Palm's formulae can be written

$$p_0(t) = \xi_t \text{ and } p_n(t) = [1 - p_0(t)]\,(1 - \eta_t)\,\eta_t^{\,n-1} \qquad (n \geqslant 1) \quad . \quad . \qquad (21)$$

where

$$\frac{\xi_t}{\mu} = \frac{\eta_t}{\lambda} = \frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (22)$$

so that the distribution (when $N_0 = 1$) is now a geometric series with a modified zero term, the mean population size being

$$\bar{N}(t) = e^{(\lambda-\mu)t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (23)$$

while

$$\text{Var}\,\{N(t)\} = \frac{\lambda + \mu}{\lambda - \mu}\, e^{(\lambda-\mu)t}\,\{e^{(\lambda-\mu)t} - 1\} \quad . \quad . \quad . \quad . \quad . \qquad (24)$$

When $\lambda = \mu$, so that the net expected rate of growth is zero and the mean population size is stationary, the solution to (20) assumes a somewhat different form which can most readily be obtained by letting $\lambda - \mu$ approach zero in the preceding formulae. It then appears that

$$p_0(t) = \xi_t = \eta_t = \lambda t/(1 + \lambda t) \quad . \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

while

$$\bar{N} = 1 \quad \text{and} \quad \text{Var}\,\{N(t)\} = 2\lambda t \quad . \quad . \quad . \quad . \quad . \quad . \qquad (26)$$

(when $N_0 = 1$), and of course these results can also be obtained by a direct solution of the original equation.

A most important observation is that for *all* N_0,

$$\lim_{t\to\infty} p_0(t) = 1 \qquad \text{when } \lambda \leqslant \mu$$

and

$$\lim_{t\to\infty} p_0(t) = (\mu/\lambda)^{N_0} \qquad \text{when } \lambda > \mu \quad . \quad . \quad . \quad . \quad . \qquad (27)$$

This limit can be interpreted as the probability of the extinction of the population in a finite time, so that one can say that there will be "almost certain" extinction whenever $\lambda \leqslant \mu$, and in particular when $\lambda = \mu$ and the mean size of the population is constant. The above result, which is true whatever the initial number of individuals may be, brings out very clearly the inadequacy of the deterministic description of the growth of a population, and it throws fresh light on the questions first discussed by de Candolle, Galton and Watson in 1873. It was first established in this form by M. S. Bartlett in 1946, although a similar result for his "life-energy" process was given by Feller in 1939.

Now that a death rate μ has been introduced it is convenient to examine more carefully the model which has been set up, and to enquire whether it could possibly be used to describe in broad outline the behaviour of human populations. It will be recalled that, in this context, the condition (*b*) is to be interpreted as meaning:

(*b′*) a female existing at time t has a chance

$$\lambda dt + o(dt)$$

of producing a single female child during the following time-interval of length dt; and of course the discussion here relates to the fortunes of the female component of the population only. The chance of death and the chance of reproduction are supposed to operate independently, the chance of reproduction and then subsequent death (as also the chance of multiple births) being assumed to be of the order $(dt)^2$, and so irrelevant to the formation of the differential-difference equations corresponding to (13).

One can usefully calculate a number of distributions connected with the life-history of a single individual; for example, the total lifetime T will be distributed according to the law

$$\mu\, e^{-\mu T} dT \qquad (0 < T < \infty) \quad . \quad . \quad . \quad . \quad . \quad (28)$$

which is not at all bad as a first approximation. On the other hand the generation time τ (which in this model could be either the time from birth to first child, or the time from one child to the next) has the distribution

$$\lambda\, e^{-\lambda\tau}\, d\tau \qquad (0 < \tau < \infty) \quad . \quad . \quad . \quad . \quad . \quad (29)$$

when the mortality effect is not in operation, and this is a bad approximation even for bacteria. A more complicated process giving an improved distribution for the generation time will be described in section 2 (viii). Finally one can calculate the distribution of the total number r of first-generation progeny produced by an individual during "its" lifetime; this is the geometric series

$$\left(1 - \frac{\lambda}{\lambda + \mu}\right)\left(\frac{\lambda}{\lambda + \mu}\right)^r \qquad (r = 0, 1, 2 \ldots) \quad . \quad . \quad . \quad . \quad (30)$$

Material for an interesting numerical comparison will be found in Lotka's empirical study of the distribution of r for white *males* in the United States of 1920 (Lotka, 1931). He found that for $r = 1, 2, 3, \ldots, 10$ a good approximation was given to the observed figures by the simple formula

$$P_r = 0{\cdot}4099\,(0{\cdot}5586)^r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

for the probability that a new-born male will eventually have just r sons. If this distribution is cut off after $r = 10$, the value for P_0 required by the law of total probability is $0{\cdot}4828$, while if the series be continued to infinity the appropriate value is $0{\cdot}4813$. The empirical value for P_0 (the probability that a new-born male will have no sons) is $0{\cdot}4981$. Thus it appears that for this human population the distribution of r is approximately geometric in form, with a modified zero term, and it is also clear that an *un*modified geometric series like (30) would at least reproduce very crudely the order of magnitude of the observed figures.

To illustrate this, the following table compares Lotka's observed distribution with the *complete* geometric series,

$$P_r = (1 - \tfrac{1}{2})\,(\tfrac{1}{2})^r \qquad (r = 0, 1, 2, \ldots) \quad . \quad . \quad . \quad . \quad . \quad (32)$$

corresponding to a birth-rate/death-rate ratio of $\lambda/\mu = 1$.

Value of r	P_r, *observed*	P_r, *from* (32)
0	0·4981	0·5000
1	2103	2500
2	1270	1250
3	0730	0625
4	0418	0313
5	0241	0156
6	0132	0078
7	0069	0039
8	0035	0020
9	0015	0010
10	0005	0005

(ii) *Birth-and-death processes with time-dependent rates.*—It was pointed out in section 1 (ii) that the deterministic model represented by equation (1) has essentially the same type of law of growth even when the intrinsic growth rate depends on the time, and it is at once clear from equation (15) that the same is true of the Furry process; any variation of λ with t can be eliminated formally by a simple change of time scale. For the birth-and-death process, however, the situation is more complicated; there are now two quantities, λ and μ, in which secular variations may take place, and it is not possible to reduce the equation (20) to a standard form simply by changing the time-scale except in the rather special case when the ratio λ/μ is constant.

The complete solution to (20) when λ and μ are general functions of t has, however, been obtained (Kendall, 1948*a*), and it is found that (for $N_0 = 1$) the distribution of population size still has the modified geometric form (21). The quantities ξ_t and η_t are to be determined from

$$\xi_t = 1 - \frac{e^{-\varrho}}{W} \quad \text{and} \quad \eta_t = 1 - \frac{1}{W} \quad . \quad . \quad . \quad . \quad . \quad (33)$$

where

$$\rho(t) = \int_0^t \{\mu(\tau) - \lambda(\tau)\}\, d\tau \quad . \quad . \quad . \quad . \quad . \quad (34)$$

and

$$W = e^{-\varrho}\{1 + \int_0^t e^{\varrho(\tau)}\, \mu(\tau)\, d\tau\} \quad . \quad . \quad . \quad . \quad . \quad (35)$$

The mean growth of the process is given by

$$\bar{N}(t) = e^{-\varrho(t)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

and

$$\text{Var}\,\{N(t)\} = e^{-2\varrho} \int_0^t e^{\varrho(\tau)}\, \{\lambda(\tau) + \mu(\tau)\}\, d\tau \quad . \quad . \quad . \quad . \quad (37)$$

(iii) *Problems of extinction, the cumulative process, periodic processes.*—The formulae of the last section make possible a more general treatment of the extinction problem*; from the formula $p_0(t) = \xi_t^{N_0}$ it will in fact readily be found that *a necessary and sufficient condition for the certain extinction of the population after the lapse of a sufficiently long period of time is that the integral*

$$\int_0^\infty e^{\varrho(\tau)}\, \mu(\tau)\, d\tau \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

should be divergent. If, however, the integral is convergent and has the value J, there is still a non-zero probability that the population will be extinguished, and it is given by

$$\{J/(1 + J)\}^{N_0}\, . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

This quantity vanishes only when μ is identically zero and the population, of course, then develops according to a pure birth process of the Furry type. It is convenient to call a process *transient* when $J = \infty$.

For a transient process it is useful to consider the random variable T which is its "age" at the moment of extinction; the distribution of T is

$$p_0'(T)\, dT \qquad (0 < T < \infty) \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Thus for a simple balanced process ($\lambda = \mu =$ constant) descended from a single ancestor ($N_0 = 1$) the T-distribution is

$$\frac{\lambda\, dT}{(1 + \lambda T)^2} \qquad (0 < T < \infty).$$

The median of the T-distribution is determined by the condition $p_0(T_m) = \frac{1}{2}$, or (when $N_0 = 1$) by

$$\int_0^{T_m} e^{\varrho(\tau)}\, \mu(\tau) d\tau = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

* For a detailed account of the topics discussed in this section, see my paper 1948*a*.

Suppose now that a population consists initially of N_0 members, and that its growth is governed by a process of transient type. A large time elapses, and it is then observed that, by a fluke, the population has so far escaped extinction. It is then extremely probable that all its surviving members are descended from just one of the original ancestors. For the associated conditional probability is

$$\frac{N_0 (1 - \xi_t) \xi_t^{N_0 - 1}}{1 - \xi_t^{N_0}} = \frac{N_0 \xi_t^{N_0 - 1}}{1 + \xi_t + \xi_t^2 + \ldots + \xi_t^{N_0 - 1}},$$

which tends to 1 as t tends to infinity. In the problem of de Candolle and Galton this is, of course, the probability that all members of the population now bear the same surname. (In the surname problem the process describes the growth of the male instead of the female component of the population.)

In some applications one is less interested in the random variable $N(t)$ than in the associated variable $M(t)$ defined as follows: $M(0) = N_0$, while for $t > 0$, the function $M(t)$ shares all the *positive* jumps of $N(t)$. This may be called the associated *cumulative* process, and it is useful in the description of a bacterial colony such that while individuals lose their reproductive powers (this is the equivalent of the μ-effect) they are never lost from the total population count. To discuss the joint distribution of $M(t)$ and $N(t)$ one must introduce the generating function

$$\psi(z, w, t) = \sum_{(n)} \sum_{(m)} p_{n,m}(t) z^n w^m \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where $p_{n,m}(t)$ is the probability that $N(t) = n$ and $M(t) = m$. It will be found that ψ satisfies the equation

$$\frac{\partial \psi}{\partial t} = \{\lambda w z^2 - (\lambda + \mu) z + \mu\} \frac{\partial \psi}{\partial z} \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

the associated boundary condition (when $N_0 = 1$) being

$$\psi(z, w, 0) = zw.$$

It is possible to find the means, variances and covariance of $M(t)$ and $N(t)$ in the general case by transforming (43) into an equation for the cumulant-generating function

$$K(u, v, t) = \log \psi(e^u, e^v, t),$$

and expanding in powers of u and v. The complete solution, i.e., the determination of the $p_{n,m}(t)$, has only been carried out when λ and μ are constants, and then it is found that for a transient process ($\lambda \leqslant \mu$), when $N_0 = 1$, the asymptotic marginal frequency distribution of $M(t)$, as t tends to infinity, is

$$\frac{\lambda + \mu}{2\lambda} \frac{(2M)!}{2^{2M} (M!)^2} \frac{x^M}{2M - 1} \qquad (M = 1, 2, 3, \ldots) \quad . \quad . \quad (44)$$

where

$$x = \frac{4 \lambda \mu}{(\lambda + \mu)^2}.$$

An interesting application of the formulae of this and the last section occurs when the functions λ and μ are each *periodic* with period ω and have equal average values. It will then be found that, when $t = k\omega$ (for each positive integer k),

$$\overline{N} = N_0 \text{ and } \operatorname{Var}\{N(t)\} = kN_0 \int_0^\omega e^{\varrho(\tau)} \{\lambda(\tau) + \mu(\tau)\} d\tau \quad . \quad . \quad . \quad (45)$$

Thus, although the *expected* value of $N(t)$ repeats itself regularly, the periodicity will in practice be obscured in the course of time by the steady increase (with increasing t) of the magnitude of the random fluctuations as measured by Var $\{N(t)\}$. It is easy to see that such processes are necessarily of transient type.

(iv) *Applications to physics.*—The Furry process discussed at the beginning of section 2 (i) was introduced in connection with a physical application, and this was also the main subject of Arley's investigation. It is therefore appropriate at this point to depart a little from the main subject of this paper and to describe briefly the physical problem concerned.

When a single "particle" of the "soft" component of cosmic radiation falls on a material slab (say of lead) its effect is multiplied to form a cascade shower, the development of which parallels in many ways the growth of a population. The postulated mechanism is as follows: a photon travelling an infinitesimal distance dt in the material has a probability $\pi\, dt$ of being absorbed and emitting a positive and negative electron-pair. Similarly an electron of either sort, on travelling a distance dt, has a chance $\varepsilon\, dt$ of giving up some of its energy and emitting a photon. In practice the electrons alone are observed, so that the shower of electrons has the appearance of growing continually by the production of twin progeny.

Several stochastic models have been proposed to represent the phenomenon, but in each case the mechanism has been simplified to a considerable extent in order to obtain a more manageable mathematical description. One of the most realistic schemes is that of Arley; he discusses a system of particles of two species, which he calls m-particles and n-particles, multiplying according to the following rules:

(α) particles of each kind have a probability λdt of being absorbed in a time interval dt and at the same time giving birth to two particles of the other kind;

(β) particles of each kind have a probability μdt of being absorbed in a time interval dt without giving birth to any other particles.

The process is thus almost the same as the birth-and-death process, except that on reproduction a particle changes its species and at the same time gives birth to a particle of the same (new) species. In order to represent the progressive energy-degradation of the particles, Arley further postulates that $\mu = \mu_1 t$ while λ is a constant, λ_0. It is easy to write down the equation similar to (20) for the generating function

$$\chi(z, w, t) = \sum_{(m)} \sum_{(n)} p_{m,n}(t)\, z^m w^n,$$

where $p_{m,n}(t)$ is the probability that at time t there are m of the m-particles and n of the n-particles, but this has never been solved. Arley did, however, calculate the means, variances and co-variance of m and n at any time t. He also constructed an approximate solution, and checked its accuracy by a step-by-step integration.

More recently Bartlett has examined a similar mechanism in which, after absorption by the λ-effect, two new particles are produced as follows:

both of the first kind,	probability p^2;
both of the second kind,	probability q^2;
one of each kind,	probability $2pq$.

Here, of course, $p + q = 1$. He has solved completely the distribution problem for this model, including the extension to the case when λ and μ are general functions of the time t. A more detailed account of this work will be found in Professor Bartlett's own contribution to the present symposium.

(v) *Immigration, the negative binomial distribution, and the logarithmic series.*—The remainder of this paper will be concerned solely with the biological interpretation of the equations. One interesting variant* of the simple birth-and-death process is obtained if one adds the assumption:

(f) in an infinitesimal time interval dt there is a chance κdt that a single member will be added to the population by immigration from the outside world.

The characteristic feature of the κ-effect is that it acts at an expected rate which is independent of the size to which the population has grown; it is this fact which makes it distinct from the λ-effect. There is nothing to be gained by introducing an effect to represent emigration, for this would simply result in an increased value of the constant μ. The generating function for the (λ, μ, κ) process satisfies the equation

$$\frac{\partial \varphi}{\partial t} = (\lambda z - \mu)(z - 1)\frac{\partial \varphi}{\partial z} + \kappa(z - 1)\varphi \quad . \quad . \quad . \quad . \quad (46)$$

* Reference may be made to my paper 1948*b*.

while the associated boundary condition is

$$\varphi(z, 0) = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

if initially the population is zero. Solution along the usual lines leads to the result

$$\varphi = \left(\frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)t} - \mu}\right)^{\kappa/\lambda} \left\{1 - z\,\frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right\}^{-\kappa/\lambda} \quad . \quad . \quad . \quad . \quad (48)$$

when $\lambda \neq \mu$, and to

$$\varphi = (1 + \lambda t)^{-\kappa/\lambda} \left\{1 - \frac{\lambda t z}{1 + \lambda t}\right\}^{-\kappa/\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

when λ and μ are equal.

The mean number of individuals is

$$\frac{\kappa}{\lambda - \mu}\{e^{(\lambda-\mu)t} - 1\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

when $\lambda \neq \mu$, and is

$$\kappa t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

when λ and μ are equal.

It is of some interest to consider the limiting form of (48) when $\lambda < \mu$ and the time t tends to infinity. The limiting generating function is

$$\varphi = \left(1 - \frac{\lambda}{\mu}\right)^{\kappa/\lambda} \left\{1 - \frac{\lambda z}{\mu}\right\}^{-\kappa/\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

and so the mean population size for large t is

$$\frac{\kappa}{\mu - \lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

and these formulae relate to the stable distribution of population size which immigration can just maintain against the excess of μ over λ.

It is clear from (48) that the distribution will still be negative binomial in form for every finite value of t when $\mu = 0$, the process then being one of immigration and reproduction only. This derivation of the negative binomial distribution is in fact well known; it appears to have been given first by the late A. G. McKendrick (1914), and was later re-discovered by G. Pólya and F. Eggenberger (for references see W. Feller (1943) and O. Lundberg (1940)).

On the other hand, when $\lambda = 0$, so that the process is one of immigration and mortality (or emigration) only, the distribution assumes the Poisson form. It is not quite easy to see this from (48), and it is much simpler to write $\lambda = 0$ in (46) and to solve afresh to obtain

$$\varphi = \exp\left\{\frac{\kappa}{\mu}(1 - e^{-\mu t})(z - 1)\right\},$$

from which the result is obvious.

An interesting application of this conclusion concerns the work of A. Milne (1943) and R. A. Fisher (1941–2) on the distribution of the numbers of ticks (*Ixodes ricinus* L.) carried by individual sheep. The observed distribution is of the negative binomial type, and departures from a Poisson distribution were also observed when the random variable concerned was the number of ticks collected on a standard blanket trailed over a given length of ground in a specified manner. If one knew nothing of the biology of the problem, the negative binomial infestation of sheep might be explained in either of two distinct ways. It could be the result of a birth-death-and-immigration process, or it could be the result of sampling from a heterogeneous super-population of Poisson distributions (the relevant theory is that of M. Greenwood and G. U. Yule (1920)). Milne gives the second explanation, pointing out that the distribution of ticks on the ground will be essentially "patchy" in character, and his view is confirmed by the blanket counts. For if the parasites were scattered uniformly and independently on the ground the blanket results should give a realization of the (λ, μ, κ) process with $\lambda = 0$, κ equal to the rate at which ticks are collected, and μ to the rate at which they are swept off the blanket again by the vegetation. A Poisson distribution for the blanket counts would thus have been expected, and this was not observed.

This dual character of the negative binomial distribution has been discussed with great clarity by Feller in the paper just mentioned*; it is particularly important because in so many applications both types of mechanism are at first sight conceivable.

Just as in R. A. Fisher's original derivation (Fisher, Corbet and Williams (1943)) the logarithmic series distribution

$$\frac{x^n}{ny} \qquad (n = 1, 2, 3, \ldots) \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

where $x = 1 - e^{-y}$, can be obtained as a limiting form of the negative binomial distribution for the (λ, μ, κ) process by first considering the conditional distribution of population size *n when it is known that* $n > 0$, and then letting the immigration constant κ tend to zero. It follows from (52) that if $\lambda < \mu$, if κ is exceedingly small, and if a colony of positive size is known to exist, then the number of its members will follow approximately the distribution (54) with

$$x = \lambda/\mu.$$

Thus the dual character of the negative binomial distribution is shared by the logarithmic series; this may be relevant to some of the applications of the latter.

(vi) *Two-point boundary problems.*—The problems usually solved in connection with stochastic processes concern the prediction of the future course of the process when as a boundary condition its state at some initial instant is given. But it is sometimes of interest to consider also two-point boundary problems of the following kind:

given that $N(t_1) = n_1$ *and* $N(t_2) = n_2$, *where* $t_1 < t_2$, *what can be said about the probable course of the process when* $t_1 < t < t_2$?

Some theorems dealing with questions of this sort have been obtained by P. Lévy (1943) for the additive process which represents the Brownian movement; in this section a simple problem of the same type will be discussed in connection with the birth-and-death process.

Let $[\varphi(z; t, t')]^{N'}$, where $t' < t$, be the function generating the probabilities $p_n(t)$ for the birth-and-death process at epoch t when the initial condition is $N(t') = N'$; this function can easily be computed with the aid of formulae which have already been given. Let the boundary conditions be $N(0) = N_0$ and $N(T) = 0$, and let the probability distribution of $m \equiv N(t)$ (given this information) be $\{q_m(t)\}$, where $0 < t < T$. It is then possible to obtain a simple expression for the generating function

$$\psi(w, t) = \sum_{(m)} w^m q_m(t)$$

in terms of the function φ.

If the probability that $N(t_2) = n_2$ when it is known only that $N(t_1) = n_1$ (where $t_1 < t_2$) be denoted by

$$P\begin{pmatrix} n_2, & t_2 \\ n_1, & t_1 \end{pmatrix},$$

then

$$\sum_{(n_2)} z^{n_2} P\begin{pmatrix} n_2, & t_2 \\ n_1, & t_1 \end{pmatrix} = [\varphi(z;\ t_2, t_1)]^{n_1},$$

and

$$P\begin{pmatrix} 0, & t_2 \\ n_1, & t_1 \end{pmatrix} = [\varphi(0;\ t_2, t_1)]^{n_1}.$$

Also

$$q_m(t)\, P\begin{pmatrix} 0, & T \\ N_0, & 0 \end{pmatrix} = P\begin{pmatrix} m, & t \\ N_0, & 0 \end{pmatrix} P\begin{pmatrix} 0, & T \\ m, & t \end{pmatrix},$$

and so

$$\psi(w, t) = \left[\frac{\varphi\{w\varphi(0;\ T, t);\ t, 0\}}{\varphi(0;\ T, 0)}\right]^{N_0} \quad . \quad . \quad . \quad . \quad . \quad (55)$$

* See also J. O. Irwin (1941).

The simplest special case occurs when λ and μ are *equal* constants, for then

$$\varphi(z;\ t_2, t_1) = \frac{\lambda\,\Delta t + z(1 - \lambda\,\Delta t)}{1 + \lambda\,\Delta t}\left\{1 - \frac{z\,\lambda\,\Delta t}{1 + \lambda\Delta t}\right\}^{-1},$$

where $\Delta t = t_2 - t_1$, and so ψ is the $N_0{}^{\text{th}}$ power of

$$\frac{1 + \lambda T}{\lambda T(1 + \lambda t)}\left\{\lambda t + w(1 - \lambda t)\frac{\lambda(T - t)}{1 + \lambda(T - t)}\right\}\left\{1 - w\frac{\lambda t\,\lambda(T - t)}{(1 + \lambda t)(1 + \lambda\,T - t)}\right\}^{-1} \quad . \quad (56)$$

On writing $w = 0$, one thus obtains the following result: *if* $N(0) = N_0$ *and* $N(T) = 0$, *then the conditional probability that* $N(t) = 0$ *is*

$$\left\{\frac{1 + 1/(\lambda T)}{1 + 1/(\lambda t)}\right\}^{N_0}, \quad \textit{when} \quad 0 < t < T \quad . \quad . \quad . \quad . \quad . \quad (57)$$

(vii) *Some properties of the "logistic" processes.*—Feller, in his 1939 paper, also examined the birth-and-death processes associated with the logistic law of growth. One general process of this kind differs from that discussed in section 2 (i) only in that the birth and death rates are now linearly dependent on the instantaneous population size, so that

$$\lambda \equiv \alpha[N_2 - N(t)] \quad \text{and} \quad \mu \equiv \beta[N(t) - N_1] \qquad (N_1 < N_2), \quad . \quad . \quad (58)$$

say, where α, β, N_1 and N_2 are absolute constants. For such a process the independent development of the sub-populations stemming from two individuals alive at a given epoch no longer obtains, and accordingly the effect of changing the initial size of the population can no longer be allowed for by raising the generating function to a suitable power.

If N_1 and N_2 are integers, it is to be understood that $N(0)$ lies in the closed interval (N_1, N_2), and then it is intuitively "clear" that $N(t)$ will in its ultimate behaviour display stationary fluctuations of a stochastic character in the region between these two limits, the effect of the initial conditions progressively becoming less important. Thus a high value of $N(t)$, near N_2, implies a low birth rate and a high death rate, and so on the whole one would then expect $N(t)$ to decrease. On the other hand, a low value near N_1 will mean a low death rate and high birth rate, and on balance a tendency for $N(t)$ to increase. The two boundaries can, of course, never be crossed because the birth rate vanishes when $N(t) = N_2$ and the death rate vanishes when $N(t) = N_1$. It would be very interesting to investigate the structure of these fluctuations with the aid of the techniques developed for the analysis of time series; one would begin by trying to determine the auto-correlations (which would presumably always be positive), but I do not propose to pursue this matter here.

It is easy to write down the partial differential equation which must be satisfied by the probability-generating function for the logistic process; it is

$$\frac{\partial\varphi}{\partial t} = \beta(1 - z)\frac{\partial}{\partial z}\left[z\frac{\partial\varphi}{\partial z} - N_1\,\varphi\right] - \alpha(1 - z)z\frac{\partial}{\partial z}\left[N_2\varphi - z\frac{\partial\varphi}{\partial z}\right] \quad . \quad . \quad (59)$$

and though this is linear it is of the second order, so that no simple general solution is available. If one sets $\beta = 0$ one obtains the "logistic" generalization of the Furry process, the differential-difference equations for which are of the general form

$$\frac{d}{dt}p_n(t) = \pi_{n-1}\,p_{n-1}(t) - \pi_n\,p_n(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

Feller (page 21 of his 1939 paper) has given the general solution to the equations (60), and this can always be used in the investigation of purely reproductive processes of the Furry type. The unmodified equation (59), however, is still unsolved, and one is here confronted with a difficulty which appears whenever a term in the corresponding deterministic equation is quadratic in the unknown functions enumerating the population, for this always means that second-order partial derivatives will occur in the equation for the probability-generating function. The same trouble arises, for example, when one tries to set up stochastic equations describing the "struggle for life" (in the sense of Volterra) between two competing species, or between prey and predator.

Feller's own investigation of the logistic processes led to the interesting result that they do *not* show the simple relation which might have been expected between the development of the deterministic process and the *mean* development of the stochastic process, and he gives a detailed examination of a number of instances of this effect, which is another consequence of the non-linear character of the equations.

I propose to show here how the equation (59) can be used to give a formal indication of what might be expected of the probable behaviour of the stochastic process when t tends to infinity; it is not, of course, suggested that the following argument dispenses with the need for a rigorous discussion, if one can be found.

Suppose that $\varphi(z, t)$ has a limit $\Phi(z)$ as t approaches infinity; then this function $\Phi(z)$ will generate probabilities which will represent the ultimate "stable" distribution of the values of $N(t)$. Since in the associated deterministic model $N(t)$ tends to a unique limiting value which is, in fact,

$$\frac{\alpha N_2 + \beta N_1}{\alpha + \beta} \qquad (61)$$

when t tends to infinity, it might at first be expected that in the stochastic model $N(t)$ would also tend to a limit $N(\infty)$, which would, however, be a random variable fluctuating from one realization of the process to the next in accordance with the probabilities generated by $\Phi(z)$. But there is nothing in the structure of the model to suggest that $N(t)$, having approached a limiting value $N(\infty)$, would be specially likely to stay there, and it appears to be a better guess to suppose that as t tends to infinity $N(t)$ becomes independent of its initial value, and fluctuates between the bounding values N_1 and N_2 in such a way that P_n (the coefficient of z^n in $\Phi(z)$) measures the fraction of time spent in the "state" $N(t) = n$.

Continuing this formal argument, one would expect $\Phi(z)$ to satisfy the equation

$$\beta \frac{d}{dz}\left[z\frac{d\Phi}{dz} - N_1\,\Phi\right] = \alpha z \frac{d}{dz}\left[N_2\,\Phi - z\frac{d\Phi}{dz}\right] \qquad (62)$$

(obtained by setting the left-hand side of (59) equal to zero), and (62) is an equation which it is quite easy to solve. The usual technique for solution in series leads to a pair of solutions only one of which is relevant here, and this gives

$$P_{N_1+m} = \frac{C}{N_1+m}\binom{N_2-N_1}{m}\alpha^m\,\beta^{(N_2-N_1)-m} \qquad (63)$$

where the constant C is to be adjusted to make

$$\sum_{n=N_1}^{N_2} P_n = 1.$$

One limiting form is worthy of special notice; this occurs when

$$N_2 \to \infty, \quad \alpha \to 0, \quad \alpha N_2 = \lambda_0,$$
$$N_1 = 1, \quad \beta = \mu_1,$$

so that

$$\lambda \equiv \lambda_0 \quad \text{and} \quad \mu \equiv \mu_1[N(t) - 1] \qquad (64)$$

The population size $N(t)$ can then range through all the values given by the positive integers, and

$$P_n = [e^{\lambda_0/\mu_1} - 1]^{-1}\frac{(\lambda_0/\mu_1)^n}{n!} \qquad (n = 1, 2, 3, \ldots) \qquad (65)$$

so that the ultimate distribution is a modification of that of Poisson.

The mean value of this distribution is

$$\sum_{n=1}^{\infty} n\,P_n = \frac{x}{1-e^{-x}} \qquad (x = \lambda_0/\mu_1),$$

while from (61) the deterministic limiting population size is

$$\lim \frac{\alpha N_2 + \beta N_1}{\alpha + \beta} = 1 + x.$$

It will be noticed that

$$\frac{x}{1 - e^{-x}} < 1 + x,$$

so that these results confirm the effect discovered by Feller.

(viii) *More realistic treatment of the generation time.*—It was noted in section 2 (i) that if the birth-and-death processes considered here are to provide an adequate description of the mode of growth of populations of bacterial organisms, a more satisfactory representation of the distribution of generation time will have to be found.

Such a modification is given by what I have called the multiple-phase process; the idea underlying its construction is briefly as follows:* Actual measurements (for example, on *B. aerogenes*) by C. D. Kelly and Otto Rahn (1932) have shown that in practice the generation time τ has a peaked distribution with a marked non-zero mode (for the organism just mentioned, at 30°C., the modal value is about half-an-hour, while the coefficient of variation for the distribution is about 25 per cent.). Without attempting to "fit" such a curve in detail it at least appears that a good qualitative representation of the kind of τ-variation actually found would be obtained if the τ-distribution were assumed to be of the χ^2-form, modified to allow of a suitable scale factor. On the other hand, it is not difficult to modify the simple birth-and-death process to give this required form to the theoretical τ-distribution; according to Kelly and Rahn there is no mortality during the initial stages of the growth of a bacterial colony, if the conditions are sufficiently favourable, and so in the model about to be described the death rate will be set equal to zero.

Suppose that the organism can exist in any one of k phases, and that reproduction (by binary fission) takes places only in the k^{th} phase. An individual starts in the first phase, and after a time τ_1 distributed according to the law

$$e^{-k\lambda\tau_1}\, k\,\lambda\, d\tau_1 \qquad (0 < \tau_1 < \infty)$$

it jumps into the second phase. This process is repeated, the lifetimes in the several phases being independent, and finally the particle terminates its residence in the k^{th} phase by dividing to form two individuals each in the first phase.

If at any time $(n_1, n_2, \ldots, n_k)$ are the numbers of individuals in the several phases, and if

$$N(t) = n_1 + n_2 + \ldots + n_k,$$

then this last variable represents the total size of a population multiplying according to a birth process for which the generation time

$$\tau = \tau_1 + \tau_2 + \ldots + \tau_k$$

is distributed as

$$\frac{1}{2k\lambda}\chi^2_{2k}.$$

When $k = 1$, there is no difference between this and the Furry process; on the other hand, if k is allowed to approach infinity the process assumes a deterministic form in which there is an exact doubling of the whole population at regular intervals of time $1/\lambda$. A rough analysis of the data of Kelly and Rahn for *B. aerogenes* has shown a value of k of the order of 20 to be appropriate to that organism.

The equations governing the development of the process are a little complicated, and only the results will be described. For large t the mean population size (if initially there is one individual, in the first phase) is asymptotic to

$$\frac{2^{1/k}}{2\,\alpha_k} e^{\alpha_k \lambda t} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

* The details will be found in my paper 1948*c*. Some of the results concerning the expectation behaviour of the multiple-phase process are related to those of A. W. Brown (1940).

where

$$\alpha_k = k(2^{1/k} - 1).$$

The constant α_k tends to log 2 as k tends to infinity, while when $k = 20$ it has the value $0{\cdot}705$ (as against the limiting value $0{\cdot}693$).

The size of the random fluctuations about this mean value are best described by the coefficient of variation of the population size, which is given (with fair accuracy for all k) by

$$\log 2 \sqrt{\frac{2}{k}} \quad \text{(for large } t\text{)} \qquad (67)$$

Of course with more general initial conditions these formulae require modification.

It is also of interest to examine what happens when t is fixed and the parameter k is very large. The calculations here involve certain formal operations which need justification, but they indicate that for large k the population size is asymptotically normally distributed about the mean value $2^{\lambda t}$ with the variance

$$\frac{2\,(\log 2)^2}{k}\, 2^{\lambda t}\,(2^{\lambda t} - 1).$$

Thus the formula

$$\text{C. of V. } \{N(t)\} \simeq \log 2 \left(\frac{2}{k}\right)^{\frac{1}{2}} (1 - 2^{-\lambda t})^{\frac{1}{2}}$$

appears to hold whenever *one* of k and t is sufficiently large.

Now log 2 . $\sqrt{2}$ is equal to $0{\cdot}98$, and so the fact that the coefficient of variation of the generation time τ is equal to $1/k^{\frac{1}{2}}$ is at least a curious coincidence. There appears to be more to it than this, however, for further formal calculations suggest that the approximate relationship

$$\text{C. of V. } \{N(t)\} \simeq \text{C. of V. } (\tau), \quad \text{for large } t \qquad (68)$$

(when $N(0) = 1$) holds for *every* stochastic birth process in which the distribution of generation time is sufficiently closely concentrated about its mean value.

(ix) *The problem of the two sexes.*—In one of the discussions at the recent Oxford Conference of the Society several speakers pointed out that the possibility of variations in the relative numbers of the two sexes has been too long neglected in population mathematics. The present section is intended to touch on some of the characteristic difficulties of this problem, and to suggest some crude approximations to a solution. A more profound analysis of the question (from the deterministic standpoint) will be found in the papers of P. H. Karmel, to which reference has already been made.

First it is convenient to consider how one could modify the simplest deterministic model, represented in the mathematics by equation (1). If $M(t)$ and $F(t)$ are the numbers of males and females respectively, the most natural way to generalize (1) is to write

$$\begin{aligned} dM/dt &= -\mu M + \tfrac{1}{2}\Lambda(M, F), \\ dF/dt &= -\mu F + \tfrac{1}{2}\Lambda(M, F), \end{aligned} \qquad (69)$$

where $\Lambda(M, F)$ is symmetric in M and F and represents the contribution from the birth rate. (It is assumed* that the death rate is the same for the two sexes, and that each birth is equally likely to add a new male or a new female to the population.) Subtraction gives

$$\frac{d}{dt}(M - F) = -\mu(M - F)$$

and so

$$M(t) - F(t) = [M(0) - F(0)]\, e^{-\mu t} \qquad (70)$$

i.e. any initial preponderance of one sex over the other disappears in the course of time.

If one keeps to models of this type, further integration of the equations must be preceded by a more detailed assumption about the function Λ. To represent random mating one might set it proportional to MF; this however implies a total number of births per unit time varying as the

* These assumptions can, of course, only be admitted as preliminary approximations.

square of the total population size (when the sex-ratio happens to be constant), and as might be expected the solution takes on an unstable character. To illustrate this, let $M_0 = F_0$, so that $M = F$ for all t, in accordance with (70); M and F then both satisfy an equation of the form

$$dM/dt = -\mu M(1 - M/\alpha) \quad . \quad . \quad . \quad . \quad . \quad (71)$$

where α is a constant. This can be solved as in section 1 (ii), and it will be found that the nature of the solution is entirely different for different initial values of the population size. Thus M tends to zero when t tends to infinity if $M_0 < \alpha$, while M is constant for all t if $M_0 = \alpha$. If $M_0 > \alpha$, M tends to infinity as t approaches the *finite* value

$$T = -\frac{1}{\mu}\log\left(1 - \frac{\alpha}{M_0}\right).$$

Similar but more complicated phenomena are displayed by the solution in the more general case when $M_0 \neq F_0$.

These difficulties are avoided if Λ is linear in the total population size, and in particular if

$$\Lambda = 2\lambda\sqrt{(MF)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

The constant λ is then the birth rate per head per unit of time when M and F happen to be equal. The equations are most easily solved in this case by writing $M = R^2$ and $F = S^2$, so that

$$dR/dt = -\tfrac{1}{2}\mu R + \tfrac{1}{2}\lambda S,$$
$$dS/dt = -\tfrac{1}{2}\mu S + \tfrac{1}{2}\lambda R.$$

Addition and integration now show that M and F tend jointly to infinity when $\lambda > \mu$, and to zero when $\lambda < \mu$, as t tends to infinity. When $\lambda = \mu$, $R + S$ is constant and M and F tend to the same non-zero limit.

A somewhat simpler but less realistic model is obtained if

$$\Lambda = \lambda(M + F), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

so that the birth rate depends on the arithmetic instead of on the geometric mean of M and F. The equations are now

$$dM/dt = -\mu M + \tfrac{1}{2}\lambda(M + F),$$
$$dF/dt = -\mu F + \tfrac{1}{2}\lambda(M + F),$$

and the solution behaves qualitatively in the same way as before, the critical relation between the constants being still $\lambda = \mu$.

Perhaps the most realistic model is that corresponding to the choice

$$\Lambda = 2\lambda \min(M, F) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

The equations are easily integrable because the algebraic sign of $(M - F)$ is the same for all t, in virtue of (70), and so if there is initially an excess of females,

$$dM/dt = -\mu M + \lambda M,$$
$$dF/dt = -\mu F + \lambda M,$$

and

$$M = M_0 e^{(\lambda-\mu)t},$$
$$F = M - (M_0 - F_0)e^{-\mu t}.$$

The qualitative behaviour of the solution is the same as for the two preceding models.

Interesting new features arise if one discriminates between married and unmarried persons. Let M, F and N denote at time t the numbers of unmarried males, unmarried females, and married couples, respectively. Then the natural generalization of the preceding model is governed by the equations

$$dM/dt = -\mu M + \lambda N + \mu N - \mathrm{K}(M, F),$$
$$dF/dt = -\mu F + \lambda N + \mu N - \mathrm{K}(M, F),$$
$$dN/dt = -2\mu N + \mathrm{K}(M, F),$$

where μ is the death rate per head per unit of time, λ is the birth rate per married person per unit of time, and $K(M, F)$ is the marriage rate per unit of time. (The age-distribution is, of course, being neglected throughout.) Subtraction of the first and second equations shows that

$$M - F = (M_0 - F_0)\, e^{-\mu t},$$

so that as before any initial excess of males or females disappears in the course of time. On the other hand, if the first and third equations are considered together, it is easy to complete the description of the solution when

$$K(M, F) = 2\nu \min (M, F) \quad . \quad . \quad . \quad . \quad . \quad (75)$$

Suppose, for example, that there is initially an excess of females; then

$$\begin{aligned} dM/dt &= -(\mu + 2\nu)\, M + (\lambda + \mu)N, \\ dN/dt &= \quad 2\nu\, M \quad - \quad 2\mu\, N, \end{aligned}$$

so that both M and N are of the form

$$Ae^{p_1 t} + Be^{p_2 t},$$

where p_1 and p_2 are the roots of the equation

$$p^2 + (3\mu + 2\nu)p + 2\mu^2 + 2\nu(\mu - \lambda) = 0.$$

The roots are always real and distinct, and one of them is always negative, so that it is the greater root which determines the character of the solution. In general it can be seen that the population size tends to infinity, approaches a finite limit, or tends to zero according as $\lambda > \lambda_1$, $\lambda = \lambda_1$, or $\lambda < \lambda_1$ where

$$\lambda_1 = \mu\,(1 + \mu/\nu) \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

One can now review briefly the prospects of being able to express these modes of population growth in stochastic form; it is evident that the problem would be a very difficult one. Assumptions such as (72), (74) and (75) seem impossible to discuss with the generating function technique, and one is driven back to that associated with (71), and to (73). The first of these must be dismissed because of the unstable character of the deterministic solution (in any case, the partial differential equation for the generating function would prove to be of the second order, and so would defy solution by the usual methods). The second, and its equivalent

$$K(M, F) = \nu\,(M + F) \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

in the final formulation of the problem, while leading to equations for which a solution could certainly be found, still has the disadvantage that it implies a non-vanishing overall marriage rate even when all the unmarried persons are of one sex.

With these somewhat unhelpful comments the topic will now be left. If it is thought worth while investigating the stochastic model incorporating the assumption (77), reference should be made to Professor Bartlett's discussion of his form of the Arley "cosmic ray" process (referred to here in section 2 (iv)), for the two problems are very similar.

(x) *Estimation of the birth rate for the purely reproductive process.*—For the most part this paper is concerned with the *construction* of models which may be useful in describing phenomena in the real world, and in general no attention will here be paid to the quite different problems of *estimation* which will arise as soon as a specific application is being considered. These matters are very important, but their investigation properly belongs to a later stage, which may, of course (if the models proposed prove unacceptable to biologists), never arrive. However, the temptation to anticipate is irresistible, and I shall work out one problem of this kind which may give some idea of the type of result which is to be expected.

Consider the simple birth process introduced by Furry, and suppose that the development of a single realization of it is observed over a period of time of length T, the initial number of individuals being N_0. Let the observations be made at the epochs

$$t = 0, \tau, 2\tau, \ldots, k\tau = T,$$

the observed population numbers being respectively

$$N_0, N_1, N_2, \ldots, N_k.$$

It is required to estimate λ, and when the maximum likelihood method is employed for this purpose it must be remembered that the standard theorems about minimum asymptotic variance, etc., apply when a large number of independent replicates of the data are the basis of the estimation. This means, not the observation of a single population for a very long period of time, but the parallel observation of a large number of independent populations, each for the same period of time. Accordingly, in the following formulae, the number of replicates R is to be understood to be very large.

Now

$$\text{Prob}\,(N_1, N_2, \ldots, N_k \mid N_0) = \prod_{i=0}^{k-1} \text{Prob}\,(N_{i+1} \mid N_i),$$

and

$$\text{Prob}\,(N_{i+1} \mid N_i) = C(N_i, N_{i+1})\, e^{-N_i\lambda\tau}\,(1 - e^{-\lambda\tau})^{N_{i+1}-N_i},$$

so that the relevant part of the logarithm of the likelihood is

$$L = (N_k - N_0)\log(1 - e^{-\lambda\tau}) - \lambda\tau \sum_{i=0}^{k-1} N_i,$$

differentiation with regard to λ giving

$$\frac{\partial L}{\partial \lambda} = \frac{(N_k - N_0)\,\tau}{e^{\lambda\tau} - 1} - \tau \sum_{i=0}^{k-1} N_i,$$

and

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{(N_k - N_0)\,\tau^2 e^{\lambda\tau}}{(e^{\lambda\tau} - 1)^2}.$$

Thus the maximum likelihood estimate of λ is l, where*

$$e^{l\tau} = \frac{N_1 + N_2 + \ldots + N_k}{N_0 + N_1 + \ldots + N_{k-1}}, \qquad (78)$$

and the asymptotic variance of l, for large R, is

$$\frac{\lambda^2}{N_0 R (e^{\lambda T} - 1)} \left\{\frac{\sinh \frac{1}{2}\lambda T/k}{\frac{1}{2}\lambda T/k}\right\}^2. \qquad (79)$$

It would be interesting to know whether the right-hand side of (78) is an unbiased estimate of the transformed parameter $e^{\lambda\tau}$.

Now the crude procedure for estimating λ would be to put $\lambda = l_1$ where

$$e^{l_1 T} = N_k/N_0;$$

this is equivalent to setting $k = 1$ in the above formulae, and so the asymptotic sampling variance would then be

$$\frac{\lambda^2}{N_0 R(e^{\lambda T} - 1)} \left\{\frac{\sinh \frac{1}{2}\lambda T}{\frac{1}{2}\lambda T}\right\}^2.$$

* When combining the information from the R replicates, the contributions to the numerator and denominator of (78) must be totalled before division.

On the other hand, the whole of the information will be available when $k = \infty$, for then the circumstances of every transition in the population size can be examined. On considering the asymptotic forms (for large k and fixed T) of the formulae (78) and (79) it will be found that the equation of estimation is then

$$l = \frac{N(T) - N(0)}{\int_0^T N(u)\,du}, \qquad (80)$$

the sampling variance of the estimate (for a large number of replicates) being

$$\frac{\lambda^2}{N_0 R(e^{\lambda T} - 1)}. \qquad (81)$$

If one associates this with 100 per cent. efficiency, then

$$100\left\{\frac{\frac{1}{2}\lambda\tau}{\sinh \frac{1}{2}\lambda\tau}\right\}^2 \qquad (82)$$

gives the relative efficiency obtained when observing at discrete intervals of time τ, instead of continuously.

In illustration of this result a few numerical values will be given. The table shows, for a small range of values of $\lambda\tau$, both the relative efficiency of intermittent observation, and the expected factor by which the population will increase between one observation and the next. It will be seen that very little information is lost unless this factor is as large as 4 or 5.

$\lambda\tau$	$e^{\lambda\tau}$	*Efficiency (per cent.).*
0·5	1·65	98
1·0	2·72	92
1·5	4·48	83
2·0	7·39	72
2·5	12·2	61
3·0	20·1	50

3. *Stochastic Fluctuations in the Age-distribution*

(i) *Generalities. The moment-generating functional.*—As was explained in the introduction, the "state" of the process at time t will now be described by the *function* $N(x, t)$; this specifies completely the age-distribution of the population, in the sense that

$$\int_{x_1}^{x_2} dN(x, t)$$

is the actual number of individuals in the age-group (x_1, x_2). The initial age-distribution $N(x, 0)$ will be supposed given, and the structure of the process will have been determined when, for any epoch t and for any finite set of non-overlapping intervals

$$(x_r, y_r) \qquad (r = 1, 2, \ldots, k),$$

the joint probability distribution of the random variables

$$\int_{x_r}^{y_r} dN(x, t) \qquad (r = 1, 2, \ldots, k) \qquad (83)$$

can be written down.

The model whose behaviour is to be studied in this way can be defined by the following assumptions:

(*a*) The sub-populations generated by two co-existing individuals develop in complete independence of one another.

(*b*) An individual of age x existing at the epoch t has a chance

$$\lambda(x)\,dt + o(dt)$$

of producing a new individual of age zero during the subsequent time-interval of length dt; it follows that if

$$\Lambda = \int_0^\infty \lambda(x)\,d\,N(x, t), \quad = \Sigma\,\lambda(x),$$

where the sum is extended over all members of the population alive at time t, then the chance of there being exactly r births in the interval $(t, t + dt)$ is P_r, where

$$P_0 = 1 - \Lambda\,dt + o(dt),$$

$$P_1 = \Lambda\,dt + o(dt),$$

and

$$P_r = o(dt) \quad \text{when} \quad r \geqslant 2.$$

Thus r is asymptotically a Poisson variable, its mean and variance being equal (to the first order in dt).

(*c*) The birth rate $\lambda(x)$ varies in any manner with the age (x) of the parent, but is independent of the epoch (t).

(*d*) An individual of age x existing at the epoch t has a chance

$$\mu(x)\,dt + o(dt)$$

of dying during the subsequent time interval of length dt.

(*e*) The death rate $\mu(x)$ varies in any manner with the age (x), but is independent of the epoch (t).

The actual problems of distribution are, as might be expected, extremely difficult, and it is much easier to study the fluctuations possible in the age-distribution by the equivalent of a method of moments. Such a course was followed in the only previous treatment of this problem, by M. S. Bartlett (1947), who approximated to the continuous model just described by one employing discrete intervals of age and epoch.

In the discrete model, the function $N(x, t)$ is replaced by the vector-variable $\mathbf{N}(t)$ whose components

$$n_1(t), n_2(t), \ldots, n_m(t)$$

enumerate the numbers of individuals in each of the several age-groups, and the simplest description of the process is then in terms of the moment-generating function

$$E\left\{\exp\left(\sum_{i=1}^{m} \theta_i\, n_i(t)\right)\right\}, \qquad \ldots \ldots \quad (84)$$

from which the joint probability distribution of the variables can be obtained by the usual inversion formula. More briefly one can write (84) in the form

$$E\left\{\exp\Big(\Theta \cdot \mathbf{N}(t)\Big)\right\}$$

where Θ is an arbitrary *vector* whose components are $\theta_1, \theta_2, \ldots, \theta_m$. This at once suggests that the continuous model can most conveniently be described by means of the moment-generating *functional*

$$M[\theta(x);\ t] = E\left\{\exp\left(\int_0^\infty \theta(x)\,dN(x, t)\right)\right\} \qquad \ldots \ldots \quad (85)$$

where $\theta(x)$ is now an arbitrary *function*. The expression (85) is called a functional because its argument is not a number θ nor a vector Θ but a function $\theta(x)$.

It is easy to see, in a formal way, that the specification of (85) completely determines the required probability structure of the process. Thus, the joint moment-generating function for the variables (83) can be obtained by substituting for $\theta(x)$ in (85) the "value" defined by

$$\begin{aligned} \theta(x) &\equiv \theta_r, \text{ in } x_r \leqslant x < y_r \ (r = 1, 2, \ldots, k), \\ &\equiv 0, \qquad \text{elsewhere,} \end{aligned}$$

and the joint probability distribution can then be found with the aid of the usual inversion theorem.

On the other hand, there exists an expansion of the cumulant-generating functional,

$$K[\theta(x);\ t] = \log M[\theta(x);\ t], \qquad (86)$$

which determines the *cumulant-functions* α, β and γ, defined* as follows:

$$\begin{aligned} E[dN(x, t)] &= \alpha(x, t)\, dx + o(dx), \\ \text{Var}\, [dN(x, t)] &= \beta(x, t)\, dx + o(dx), \\ \text{Cov}\, [dN(x, t), dN(y, t)] &= \gamma(x, y, t)\, dx\, dy + o(dx\, dy) \end{aligned} \qquad (87)$$

To demonstrate this, write $u\theta_0(x)$ for $\theta(x)$, so that $K[u\theta_0(x);\ t]$ is the cumulant-generating function (when regarded as a function of u alone) for the random variable

$$\int_0^\infty \theta_0(x)\, dN(x, t).$$

Thus the coefficient of u in K is

$$E[\int_0^\infty \theta_0(x)\, dN(x, t)] = \int_0^\infty \theta_0(x)\, \alpha(x, t)\, dx,$$

while that of $\frac{1}{2}u^2$ is

$$\begin{aligned} \text{Var}\, [\int_0^\infty \theta_0(x)\, dN(x, t)] = \int_0^\infty [\theta_0(x)]^2\, \beta(x, t)\, dx + \\ + \int_0^\infty \int_0^\infty \theta_0(x)\, \theta_0(y)\, \gamma(x, y, t)\, dx\, dy, \end{aligned}$$

i.e. the expansion will be of the form

$$\begin{aligned} K[\theta(x);\ t] = \int_0^\infty \theta(x)\, \alpha(x, t)\, dx + \tfrac{1}{2}\int_0^\infty [\theta(x)]^2\, \beta(x, t)\, dx + \\ + \tfrac{1}{2}\int_0^\infty \int_0^\infty \theta(x)\, \theta(y)\, \gamma(x, y, t)\, dx\, dy + \ldots \end{aligned} \qquad (88)$$

on writing again $u\theta_0 \equiv \theta$.

My first method of procedure (following Professor Bartlett's discussion of the discrete problem) was to study the development in time of the cumulant-generating functional, and then by expansion to obtain differential and integral equations for the cumulant-functions α, β and γ. In this way one obtains†

$$K[\theta(x);\ t + dt] = K[\theta^*(x);\ t] + o(dt) \qquad (89)$$

where

$$\theta^*(x) = \theta(x) + dt[\theta'(x) - \mu(x)\{1 - e^{-\theta(x)}\} + \lambda(x)\{e^{\theta(0)} - 1\}],$$

which can be replaced by

$$\begin{aligned} \theta(x) + dt[\theta'(x) - \mu(x)\, \theta(x) + \lambda(x)\, \theta(0) + \\ + \tfrac{1}{2}\, \mu(x)\{\theta(x)\}^2 + \tfrac{1}{2}\, \lambda(x)\{\theta(0)\}^2] \end{aligned} \qquad (90)$$

* See also the remarks at the beginning of section 3 (ii).

† Details of this calculation will be found in the Appendix.

if only the linear and quadratic terms in θ are required. (Here $\theta'(x)$ denotes the first derivative of $\theta(x)$.) Expansion then gives

$$\frac{\partial\alpha}{\partial t} + \frac{\partial\alpha}{\partial x} = -\mu\alpha,$$

$$\frac{\partial\beta}{\partial t} + \frac{\partial\beta}{\partial x} = \mu\alpha - 2\mu\beta,$$

and

$$\frac{\partial\gamma}{\partial t} + \frac{\partial\gamma}{\partial x} + \frac{\partial\gamma}{\partial y} = -[\mu(x) + \mu(y)]\gamma, \quad . \quad . \quad . \quad . \quad . \quad (91)$$

partial differential equations which have the general solutions

$$\alpha(x, t) = e^{-m(x)} A(t - x),$$

$$\beta(x, t) = \alpha(x, t) + e^{-2m(x)} B(t - x),$$

and

$$\gamma(x, y, t) = e^{-m(x)-m(y)} C(t - x, t - y), \quad . \quad . \quad . \quad . \quad . \quad (92)$$

where

$$m(v) = \int^{v} \mu(u)\, du.$$

The remaining equations given by the expansion method are the boundary conditions

$$\alpha(0, t) = \int_0^\infty \lambda(x)\, \alpha(x, t)\, dx = \beta(0, t),$$

and

$$\lambda(x)\, \beta(x, t) + \tfrac{1}{2}\int_0^\infty \gamma(x, u, t)\, \lambda(u)\, du + \tfrac{1}{2}\int_0^\infty \gamma(u, x, t)\, \lambda(u)\, du = \tfrac{1}{2}[\gamma(x, 0, t) + \gamma(0, x, t)] \quad . \quad . \quad (93)$$

To these must be added, of course, the symmetry condition

$$\gamma(x, y, t) = \gamma(y, x, t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (94)$$

It was by pursuing this method of attack that many of the following results were first obtained, but I propose in this paper to use a different line of argument which is much simpler and is at the same time perhaps more easy to justify. It is free, moreover, from a grave disadvantage of the cumulant-generating-functional method, associated with the fact that the solutions turn out to have $x = t$ as a locus of discontinuity; no doubt due allowance could be made for this in effecting the necessary integrations by parts, but I am sure the difficulty is best avoided.

Professor Bartlett has pointed out to me that the idea of the moment- or cumulant-generating functional is not new; it was introduced simultaneously in 1947 by S. Bochner (in a fundamental study of stochastic processes of general type) and by Le Cam (in an analysis of the relationship of rainfall to river-flows).

(ii) *Integral equations for the linear and quadratic cumulant-functions.*—In these problems it is acceptable as a first approximation to assume that the death rate $\mu(x)$ is a constant, μ, independent of the age x of the individual; as this assumption produces a great simplification of the formulae while leaving apparent the generality of the method, I shall adopt it throughout. On the other hand, the dependence of the birth rate $\lambda(x)$ on the age x of parent is very important and cannot be ignored without departing very much from reality; accordingly it will be retained in the present section, but in section 3 (iii) I shall develop the consequences of the (quite unreal) assumption, λ = constant, because it represents the simplest (and so a very instructive) special case.

It is convenient to begin by stating several lemmas concerning conditional distributions. If the suffices 1 and 2 denote two successive steps of the averaging process, then

$$E_{12}(x) = E_1[E_2(x)],$$

$$\mathrm{Var}_{12}(x) = E_1[\mathrm{Var}_2(x)] + \mathrm{Var}_1[E_2(x)],$$

and

$$\mathrm{Cov}_{12}(x, y) = E_1 [\mathrm{Cov}_2 (x, y)] + \mathrm{Cov}_1 [E_2 (x), E_2 (y)] \quad . \quad . \quad . \quad (95)$$

The proofs are immediate.

Next it will be useful to introduce the cumulant-functions α, β and γ in a slightly different way, which is rather more satisfactory than their previous definition by means of the equations (87). Consider first the random variable $dN(x, t)$ enumerating the individuals in the asymptotically small age-group $(x, x + dx)$, where $x < t$. It is then assumed that

$$\begin{aligned} dN(x, t) &= 0 \text{ with probability } 1 - \alpha(x, t)\, dx + o(dx), \\ &= 1 \quad ,, \qquad ,, \qquad \alpha(x, t)\, dx + o(dx), \\ &\geqslant 2 \quad ,, \qquad ,, \qquad o(dx). \end{aligned}$$

Evidently one can say that $dN(x, t)$ is to the first order a Poisson variable with the (asymptotically small) mean value $\alpha(x, t)\, dx$, its variance $\beta(x, t)\, dx$ being obviously (to this order) equal to its mean.

The joint distribution of $dN(x, t)$ and $dN(y, t)$ must now be considered, when x and y are both less than t; the leading terms in this are taken to be

	$dN(x, t) = 0$	$dN(x, t) = 1$
$dN(y, t) = 0$	$1 - \alpha(x, t)\, dx - \alpha(y, t)\, dy$ $+ \Gamma(x, y, t)\, dx\, dy$	$\alpha(x, t)\, dx$ $- \Gamma(x, y, t)\, dx\, dy$
$dN(y, t) = 1$	$\alpha(y, t)\, dy$ $- \Gamma(x, y, t)\, dx\, dy$	$\Gamma(x, y, t)\, dx\, dy$

where $\Gamma(x, y, t) = \gamma(x, y, t) + \alpha(x, t)\, \alpha(y, t)$. This gives (to the first order) the correct marginal distributions, while the introduction of the terms in γ allows for the lack of independence between the two variables, the leading term in their covariance being $\gamma(x, y, t)\, dx\, dy$, as previously stated.

It will be supposed that initially there is just one member of the population, of age X, so that

$$N(x, 0) = 0 \quad (x < X) \quad \text{and} \quad N(x, 0) = 1 \quad (x > X) \quad . \quad . \quad . \quad (96)$$

The fortunes of this single ancestral individual are best described by introducing a special variable $z(t)$ which is equal to 1 if the ancestor is still alive at the epoch t, and to zero otherwise. The chance that $z = 1$ is thus $e^{-\mu t}$. More complicated initial conditions can, of course, be represented by a superposition of solutions of this type.

Now consider the variable $dN(x, t)$, where $x < t$, relative to the conditions at the epoch $t - x$ (supposed given, for the moment). Conditionally it is a binomial variable with

$$\text{“}n\text{”} = dN(0, t - x) \quad \text{and} \quad \text{“}p\text{”} = e^{-\mu x}.$$

Here $dN(0, t - x)$ is effectively the number of births in a time interval of length dx located at the epoch $t - x$, and so it has the (conditional) mean value

$$\left\{ \lambda(X + t - x)\, z(t - x) + \int_0^{t-x} \lambda(y)\, dN(y, t - x) \right\} dx;$$

also the mean and variance of $dN(x, t)$ have been seen to be equal (to the first order) when $x < t$, so that

$$\alpha(x, t) = \beta(x, t) \quad \text{if} \quad x < t \quad . \quad . \quad . \quad . \quad . \quad (97)$$

Accordingly, both these functions are determined for $x < t$ by the integral equation

$$\alpha(x, t) = \lambda(X + t - x)\, e^{-\mu t} + e^{-\mu x} \int_0^{t-x} \lambda(y)\, \alpha(y, t - x)\, dy,$$
$$= e^{-\mu x}\, \varphi(t - x), \text{ say,}$$

and thus the function $\varphi(v)$ (which is the expected number of births per unit time at the epoch v) satisfies

$$\varphi(v) = \lambda(X + v)\, e^{-\mu v} + \int_0^v \lambda(y)\, e^{-\mu y}\, \varphi(v - y)\, dy \quad . \quad . \quad . \quad (98)$$

This is an integral equation of the type discussed by Feller (it is, of course, identical with (8) of the deterministic theory), and it follows from Feller's Theorem 2 that there exists just one non-negative solution which is bounded in every finite interval. In principle, therefore, the cumulant-functions α and β can always be uniquely determined when $x < t$.

The "Poisson" character of $dN(x, t)$ (for $x < t$) obviously continues to hold when there is a perfectly general initial distribution $dN(x, 0)$. But if $x > t$, $dN(x, t)$ is a *binomial* variable with

$$\text{“}n\text{”} = dN(x - t, 0) \quad \text{and} \quad \text{“}p\text{”} = e^{-\mu t},$$

so that then

$$\alpha(x, t)\, dx = e^{-\mu t}\, dN(x - t, 0)$$

and

$$\beta(x, t)\, dx = e^{-\mu t}\, (1 - e^{-\mu t})\, dN(x - t, 0) \quad . \quad . \quad . \quad . \quad (99)$$

Returning to the initial condition (96), it is next necessary to evaluate the *conditional* means

$$\alpha_1(x, t)\, dx = E\{dN(x, t) \mid z(t) = 1\}$$

and

$$\alpha_0(x, t)\, dx = E\{dN(x, t) \mid z(t) = 0\} \quad . \quad . \quad . \quad . \quad . \quad (100)$$

for $x < t$. Only one need be found, because

$$\alpha(x, t) = e^{-\mu t}\, \alpha_1(x, t) + (1 - e^{-\mu t})\, \alpha_0(x, t), \quad . \quad . \quad . \quad . \quad (101)$$

and it is simplest to determine α_1. An argument similar to that just given shows that

$$\alpha_1(x, t) = \lambda(X + t - x)\, e^{-\mu x} + e^{-\mu x} \int_0^{t-x} \lambda(y)\, \alpha_1(y, t - x)\, dy,$$
$$= e^{-\mu x}\, \varphi_1(t - x), \qquad \text{say,}$$

where

$$\varphi_1(v) = \lambda(X + v) + \int_0^v \lambda(y)\, e^{-\mu y}\, \varphi_1(v - y)\, dy \, . \quad . \quad . \quad . \quad (102)$$

Once again Feller's theorem applies and so there is just one solution of the type required.

From the nature of the problem, the cumulant-function γ satisfies the identities

$$\gamma(x, y, t) = \gamma(y, x, t),$$

and

$$\gamma(x, y, t) = 0 \quad \text{when} \quad x > t \quad \text{and} \quad y > t; \quad . \quad . \quad . \quad . \quad (103)$$

it need therefore only be determined when $x < t < y$ and when $x < y < t$. If $x < t < y$, it is enough to have evaluated

$$\kappa(x, t)\, dx = \operatorname{Cov}\{z(t), dN(x, t)\} \qquad (x < t) \quad . \quad . \quad . \quad (104)$$

when the initial condition is given by (96), and it is easy to express this in terms of functions which have already been found, because

$$\kappa(x, t) = e^{-\mu t}\, (1 - e^{-\mu t})\, (\alpha_1 - \alpha_0) \quad . \quad . \quad . \quad . \quad (105)$$
$$= e^{-\mu t}\, (\alpha_1 - \alpha)$$

in virtue of (101). Alternatively one can note that κ satisfies the equation

$$\kappa(x, t) = \lambda(X + t - x)\, e^{-\mu t} (e^{-\mu x} - e^{-\mu t}) + e^{-2\mu x} \int_0^{t-x} \lambda(y)\, \kappa(y, t - x)\, dy,$$

so that

$$\chi(v) = \lambda(X + v)(1 - e^{-\mu v}) + \int_0^{v} \lambda(y)\, e^{-\mu y} \chi(v - y)\, dy \qquad (106)$$

where*

$$\kappa(x, t) = e^{-\mu x - \mu t} \chi(t - x);$$

Feller's theorem applies as before, and so there exists a solution κ which is never negative.

Finally one must set up an equation to determine γ when $x < y < t$; this is slightly more difficult. The last identity in (95) implies that

$$\gamma(x, y, t)\, dx\, dy = e^{-2\mu x} \operatorname{Cov} \{dN(0, t - x), dN(y - x, t - x)\} \qquad (107)$$

and then a continuation of the previous type of argument† shows that

$$\begin{aligned}\gamma(x, y, t) = {} & \lambda(X + t - x)\, e^{-2\mu x} \kappa(y - x, t - x) + \\ & + \lambda(y - x)\, e^{-2\mu x} \beta(y - x, t - x) + \\ & + e^{-2\mu x} \int_0^{t-x} \lambda(u)\, \gamma(y - x, u, t - x)\, du \qquad (108)\end{aligned}$$

The last equation is more suitably written in the longer form

$$\begin{aligned}\gamma(x, y, t) = {} & \lambda(X + t - x)\, e^{-2\mu x} \kappa(y - x, t - x) + \\ & + \lambda(y - x)\, e^{-2\mu x} \beta(y - x, t - x) + \\ & + e^{-2\mu x} \int_{y-x}^{t-x} \lambda(u)\, \gamma(y - x, u, t - x)\, du + \\ & + e^{-2\mu x} \int_0^{y-x} \lambda(u)\, \gamma(u, y - x, t - x)\, du \qquad (109)\end{aligned}$$

because all the functions $\gamma(\xi, \eta, \zeta)$ then have arguments satisfying the inequalities $\xi < \eta < \zeta$. It is clear that γ must be of the form

$$\gamma(x, y, t) = e^{-2\mu t} \psi(y - x, t - y) \qquad (110)$$

and so the reduced function ψ is to be found from

$$\begin{aligned}\psi(U, V) = {} & e^{2\mu(U+V)} \lambda(X + U + V)\, \kappa(U, U + V) + \\ & + e^{2\mu(U+V)} \lambda(U)\, \beta(U, U + V) + \\ & + \int_0^{V} \lambda(U + \tau)\, \psi(\tau, V - \tau)\, d\tau + \\ & + \int_0^{U} \lambda(U - \tau)\, \psi(\tau, V)\, d\tau \qquad (111)\end{aligned}$$

This appears to be a new type of integral equation, an analytical investigation of which will be necessary before the present formulae can be applied to any practical case. The simplest situation is, of course, that when λ is constant; the equation then has an elementary solution

* Of course $\chi = \varphi_1 - \varphi$.

† In working through these arguments it will be found very helpful to consider the life-history of a population element in relation to a plane diagram in which t and x are Cartesian coordinates.

which will be given in the next section, but it is not in itself of practical relevance. It is worth noting that if one carries out the Laplace transformation

$$\int_0^\infty e^{-sV}\,\psi(U, V)\,dV = \Psi(U, s)$$

the equation (111) becomes

$$\Psi(U, s) = C(U, s) + \int_0^\infty \lambda(U + \tau)\,e^{-s\tau}\,\Psi(\tau, s)\,d\tau + \int_0^U \lambda(U - \tau)\,\Psi(\tau, s)\,d\tau,$$

where $C(U, s)$ does not involve ψ; this provides a link between (111) and the more familiar types of integral equation, but it will not be further investigated here.

(iii) *Solution of the integral equations when the birth and death rates are not age-specific.*—When $\lambda(x)$ has the constant value λ whatever the age of parent, the integral equations of the preceding section can all be solved explicitly by an elementary method. In the following account most of the details will be omitted because the method cannot be employed when λ is a general function of x; it is worth noting, however, that the way in which the solutions are obtained makes it quite clear that they are unique.

First, much as in section 1 (ii), one finds that

$$\alpha(x, t) = \beta(x, t) = \lambda e^{(\lambda-\mu)t-\lambda x} \text{ when } x < t, \qquad (112)$$

and that

$$\kappa(x, t) = \frac{\lambda\mu}{\lambda - \mu}\,e^{-\mu x-\mu t}\,\{e^{(\lambda-\mu)(t-x)} - 1\} \text{ when } x < t; \qquad (113)$$

it will be noticed that κ is never negative.

The solution of the γ-equation is rather more complicated, and an outline of the method will be given. The preceding results show that $\psi(U, V)$, the function defined by (110), must satisfy the equation

$$\psi(U, V) = \frac{\lambda^2\mu}{\lambda - \mu}\,(e^{\lambda V} - e^{\mu V}) + \lambda^2\,e^{(\lambda+\mu)V+\mu U} + \lambda\int_0^V \psi(\tau, V - \tau)\,d\tau + \lambda\int_0^U \psi(\tau, V)\,d\tau.$$

Thus

$$\frac{\partial\psi}{\partial U} = \lambda\psi + \lambda^2\mu\,e^{(\lambda+\mu)V+\mu U},$$

and so

$$\psi(U, V) = F(V)\,e^{\lambda U} - \frac{\lambda^2\mu}{\lambda - \mu}\,e^{(\lambda+\mu)V+\mu U}.$$

Substitution now gives an equation for $F(V)$ which is easily solved, and one obtains finally

$$\gamma(x, y, t) = \frac{\lambda^2}{\lambda - \mu}\,e^{\lambda(t-y)-2\mu t}\{\lambda e^{\lambda(t-x)} - \mu e^{\mu(t-x)} + \mu e^{\lambda(y-x)}\,[e^{\lambda(t-y)} - e^{\mu(t-y)}]\},$$

when $x < y < t$. It will be noticed that this is non-negative. The formula for γ can be re-written as

$$\lambda^2\left(\frac{\lambda + \mu}{\lambda - \mu}\right)e^{2(\lambda-\mu)t-\lambda(x+y)} - \frac{\lambda^2\mu}{\lambda - \mu}\,e^{(\lambda-\mu)t}\,\{e^{-\lambda y-\mu x} + e^{-\mu y-\lambda x}\} \qquad (114)$$

and then, since the right-hand side is symmetric in x and y, it follows that this result must be true whenever x and y are both less than t. (It could not have been foreseen that the same analytic form would obtain for $x < y < t$ and $y < x < t$.)

The equations have been solved separately for the case $\lambda = \mu$, and the results obtained agree with the limiting forms of the above expressions when the difference $\lambda - \mu$ tends to zero; they are

$$\alpha(x, t) = \beta(x, t) = \lambda e^{-\lambda x} \qquad (x < t),$$

$$\kappa(x, t) = \lambda^2(t - x)\,e^{-\lambda x-\lambda t} \qquad (x < t),$$

and

$$\gamma(x, y, t) = \lambda^2\{1 + \lambda(2t - x - y)\}\,e^{-\lambda x-\lambda y} \qquad (x, y < t) \qquad (115)$$

These formulae for the case $\lambda = \mu$ will be used to indicate how a more general form of initial condition can be introduced. If the initial age-distribution $N(x, 0)$ is written $N_0(x)$, where

$$\int_0^\infty dN_0(x) = N(\infty, 0) = N_0,$$

so that N_0 is the initial *total* population size, one easily finds that

$$\alpha(x, t)\, dx = e^{-\lambda t}\, dN_0(x - t) \qquad (x > t),$$

and

$$\beta(x, t)\, dx = e^{-\lambda t}(1 - e^{-\lambda t})\, dN_0(x - t) \qquad (x > t),$$

while

$$\gamma(x, y, t)\, dy = \lambda^2(t - x)\, e^{-\lambda x - \lambda t}\, dN_0(y - t) \quad . \quad . \quad . \quad (116)$$

when $x < t < y$. When x and y are both less than t the previous results hold as before, the only modification required being the multiplication of the right-hand side of each formula by N_0.

A specially interesting example is that in which $dN_0(x)$ is approximately given by $N_0 \lambda e^{-\lambda x}\, dx$, so that the initial distribution is an approximate realization of the "stable" age-distribution. One then finds (per unit of initial population) that

$$\alpha(x, t) = \lambda e^{-\lambda x} \qquad \text{(for all } x \text{ and } t\text{)},$$

$$\begin{aligned} \beta(x, t) &= \lambda e^{-\lambda x} && (x < t), \\ &= \lambda e^{-\lambda x}(1 - e^{-\lambda t}) && (x > t), \end{aligned}$$

and

$$\begin{aligned} \gamma(x, y, t) &= \lambda^2\{1 + \lambda(2t - x - y)\}\, e^{-\lambda x - \lambda y} && (x < y < t), \\ &= \lambda^3(t - x)\, e^{-\lambda x - \lambda y} && (x < t < y), \\ &= 0 && (t < x < y), \quad . \quad . \quad (117) \end{aligned}$$

while, of course, $\gamma(y, x, t) \equiv \gamma(x, y, t)$.

I now return to the more general situation when λ and μ need not be equal. The relations*

$$E\{\int_{x_1}^{x_2} dN\} = \int_{x_1}^{x_2} \alpha(x, t)\, dx,$$

$$\text{Var}\{\int_{x_1}^{x_2} dN\} = \int_{x_1}^{x_2} \beta(x, t)\, dx + \int_{x_1}^{x_2}\int_{x_1}^{x_2} \gamma(x, y, t)\, dx\, dy,$$

and

$$\text{Cov}\{\int_{x_1}^{x_2} dN, \int_{x_3}^{x_4} dN\} = \int_{x_1}^{x_2}\int_{x_3}^{x_4} \gamma(x, y, t)\, dx\, dy \qquad (x_1 < x_2 < x_3 < x_4) \quad . \quad (118)$$

can be used to investigate the linear and quadratic moments associated with a grouped population. Thus, when there is just a single ancestor,

$$E\{\int_{x_1}^{x_2} dN\} = e^{(\lambda - \mu)t}(e^{-\lambda x_1} - e^{-\lambda x_2}) \qquad (x_1 < x_2 < t)$$

and

$$\begin{aligned} \text{Var}\{\int_{x_1}^{x_2} dN\} = {} & e^{(\lambda - \mu)t}(e^{-\lambda x_1} - e^{-\lambda x_2}) + \\ & + \left(\frac{\lambda + \mu}{\lambda - \mu}\right) e^{2(\lambda - \mu)t}(e^{-\lambda x_1} - e^{-\lambda x_2})^2 - \\ & - \left(\frac{2\lambda}{\lambda - \mu}\right) e^{(\lambda - \mu)t}(e^{-\lambda x_1} - e^{-\lambda x_2})(e^{-\mu x_1} - e^{-\mu x_2}) \\ & \qquad (x_1 < x_2 < t), \quad (119) \end{aligned}$$

* dN will now be written instead of $dN(x, t)$, for typographical convenience.

so that (when there are N_0 ancestors) the coefficient of variation,

$$\text{C. of V.}\ \{\int_{x_1}^{x_2} dN\},$$

approaches (for large t) the limiting value

$$\sqrt{\left(\frac{\lambda+\mu}{\lambda-\mu}\cdot\frac{1}{N_0}\right)} \quad \text{when} \quad \lambda > \mu \quad . \quad . \quad . \quad . \quad (120)$$

When $\lambda = \mu$ it is asymptotic to

$$\sqrt{\left(\frac{2\lambda t}{N_0}\right)}$$

and when $\lambda < \mu$ it tends to infinity like $e^{\frac{1}{2}(\mu-\lambda)t}$.

Similar calculations show that the covariance of the two random variables

$$\int_{x_1}^{x_2} dN \quad \text{and} \quad \int_{y_1}^{y_2} dN \qquad (x_1 < x_2 < y_1 < y_2 < t)$$

is given by

$$\left(\frac{\lambda+\mu}{\lambda-\mu}\right) e^{2(\lambda-\mu)t}\,(e^{-\lambda x_1} - e^{-\lambda x_2})\,(e^{-\lambda y_1} - e^{-\lambda y_2}) -$$
$$- \left(\frac{\lambda}{\lambda-\mu}\right) e^{(\lambda-\mu)t}\{(e^{-\lambda x_1} - e^{-\lambda x_2})\,(e^{-\mu y_1} - e^{-\mu y_2}) + (e^{-\lambda y_1} - e^{-\lambda y_2})(e^{-\mu x_1} - e^{-\mu x_2})\} \qquad (121)$$

if $N_0 = 1$. Thus, for any value of N_0, the correlation coefficient ρ of the two variables tends to 1 as t tends to infinity when $\lambda \geqslant \mu$, while it approaches the limiting value

$$\left(\frac{2\lambda\mu}{\mu-\lambda}\right) \sqrt{(\delta x \,.\, \delta y)}\, e^{-\frac{1}{2}\mu(x+y)} \quad \text{Cosh}\left\{\frac{(\mu-\lambda)(y-x)}{2}\right\}. \quad . \quad . \quad . \quad (122)$$

if $\lambda < \mu$ and if the intervals of grouping

$$\delta x = x_2 - x_1 \quad \text{and} \quad \delta y = y_2 - y_1$$

are small.

(iv) *Problems of distribution.*—The preceding calculations only enable one to discuss stochastic fluctuations in the age-distribution in so far as these are measurable by their linear and quadratic moments, and it is desirable that a complete theory should give a more detailed account of what can happen by supplying the joint probability distribution of the numbers in any given set of age groups. A general attack on this problem does not at the moment seem likely to succeed, but some progress can be made if one continues with the simple (but unrealistic) assumption that the birth and death rates λ and μ are constants. In what follows it will in addition be supposed that $\lambda = \mu$, so that the population is in a state which would be judged stationary if only the expectation values of the various quantities were considered. This is at once the most interesting case, and also that for which the formulae are most simple. The eventual result of the calculation will be the probability distribution of the random variable

$$\int_{x}^{x+a} dN(y, t) \qquad (x < x + a < t) \quad . \quad . \quad . \quad . \quad (123)$$

when at the initial time $t = 0$ there is just one ancestral individual; this is the modified geometric-series distribution generated by the function $\psi(z, t)$ of equation (124).

Let $p_m(t)$ be the probability that the variable (123) assumes the value m at time t; then the generating function

$$\begin{aligned}\psi(z, t) &= \sum_{(m)} p_m(t)\, z^m \\ &= E\,(z^m) \\ &= E\,.\,[1 - e^{-\lambda x} + z e^{-\lambda x}]^{N(a,\, t-x)},\end{aligned}$$

when account is taken of the variable $N(a, t - x)$.

Next, consider a simple birth-and-death process for which $\lambda = \mu$, developing from a single ancestor at the initial epoch $\tau = 0$, and let $P_{r,s}(\tau)$ be the probability that at the subsequent epoch τ the total population is $r + s$ while the total number of *progeny* is r (so that $s = 1$ if the ancestor is still alive, and otherwise $s = 0$). Let the generating function for this bivariate distribution be

$$\chi(z, w, \tau) = \sum_{(r)} \sum_{(s)} P_{r,s}(\tau)\, z^r w^s.$$

Then evidently

$$\psi(z, t) = E\,.\,\{\chi[1 - e^{-\lambda x} + ze^{-\lambda x}, 1, a]\}^{N(\infty,\, t-x-a)}$$

when account is duly taken of the probability distribution of the variable $N(\infty, t - x - a)$. But this last variable is the *total* population size in a birth-and-death process, with $\lambda = \mu$, at epoch $t - x - a$, when initially there is just one member, and so from the results of section 2 (i) it follows that

$$\psi(z, t) = \frac{\lambda(t - x - a) + [1 - \lambda(t - x - a)]\,\chi[1 - e^{-\lambda x} + ze^{-\lambda x}, 1, a]}{[1 + \lambda(t - x - a)] - \lambda(t - x - a)\,\chi[1 - e^{-\lambda x} + ze^{-\lambda x}, 1, a]}.$$

There remains the problem of determining the function χ; now obviously one can write

$$\chi(z, w, \tau) = e^{-\lambda\tau}\, w\,\chi_1(z, \tau) + (1 - e^{-\lambda\tau})\,\chi_0(z, \tau),$$

and so

$$\chi(z, 1, \tau) = e^{-\lambda\tau}\chi_1 + (1 - e^{-\lambda\tau})\,\chi_0.$$

The function χ_1 can easily be found from the results of section 2 (v), for it is the generating function for a (λ, μ, κ) process in which $\lambda = \mu = \kappa$; thus

$$\chi_1(z, \tau) = 1/(1 + \lambda\tau - \lambda\tau z).$$

Again, $\chi(z, z, \tau)$ is the generating function for an ordinary birth-and-death process when the ancestor and its progeny are not discriminated; thus

$$\frac{\lambda\tau + z(1 - \lambda\tau)}{1 + \lambda\tau - z\lambda\tau} = e^{-\lambda\tau}\, z\,\chi_1 + (1 - e^{-\lambda\tau})\,\chi_0,$$

and so

$$\chi(z, w, \tau) = \frac{\lambda\tau + z(1 - \lambda\tau) + (w - z)\, e^{-\lambda\tau}}{1 + \lambda\tau - \lambda\tau z}.$$

Hence finally

$$\psi(z, t) = \frac{1 + (1 - z)\, e^{-\lambda x}\,\{\lambda(t - x - a)(1 - e^{-\lambda a}) + (e^{-\lambda a} - 1 + \lambda a)\}}{1 + (1 - z)\, e^{-\lambda x}\,\{\lambda(t - x - a)\,(1 - e^{-\lambda a}) + \lambda a\}} \quad . \qquad (124)$$

It can now be verified without very much trouble that the confluent forms for $\lambda = \mu$ of the formulae (119) agree with the mean and variance of the above distribution.

It will be seen that (for a single initial ancestor) the random variable

$$\int_x^{x+a} dN(y, t) \qquad (x < x + a < t)$$

has a probability distribution of geometric-series form, with a modified zero term. The probability that there are *no* individuals in the age-group $(x, x + a)$ at time t is

$$\frac{e^{\lambda x} + \lambda(t - x - a)(1 - e^{-\lambda a}) + (e^{-\lambda a} - 1 + \lambda a)}{e^{\lambda x} + \lambda(t - x - a)(1 - e^{-\lambda a}) + \lambda a} \quad . \quad . \quad . \qquad (125)$$

This approaches certainty,

(1) if t is fixed and a tends to zero,

or

(2) if a is fixed and t tends to infinity.

4. *Concluding Remarks*

In this paper I have tried to develop the theory of the birth-and-death processes in a way that will interest both the biologist and the mathematician; obviously very much remains to be done, and it seems fitting to conclude by mentioning a few outstanding problems on which further work is required:

(*a*) The line of argument used in section 2 (vii) may also lead to interesting results when applied to what may be called "competition" processes—the stochastic equivalents of the "struggle for life" situations envisaged by Volterra. It will not usually be the case that the ultimate behaviour of the process is independent of the initial conditions, and so a systematic attack on parabolic equations like (59) will probably be needed.

(*b*) There are many estimation problems similar to that dealt with here in section 2 (x) which should reward further investigation.

(*c*) The need for a more general solution of the integral equation (111) has already been stressed.

(*d*) The use made here of the moment-generating functional suggests a number of interesting problems which could profitably be considered in relation to the work of Bochner on general stochastic processes.

(*e*) Two possible applications may be mentioned. The first concerns experiments designed to throw light on the origin of bacterial resistance to such agents as penicillin and streptomycin; the fundamental idea on which the experiments are based was introduced by S. E. Luria and M. Delbrück (1943) when studying the transformation of bacteria from virus sensitivity to virus resistance, and their method involves the consideration of what is essentially a (λ, μ, κ) process with λ constant, μ zero, and κ a known function of the time.

(*f*) On quite another scale it is possible that the methods developed here may find an application to the work of Thomas Park (1948) on competition between experimental populations of flour beetles (*Tribolium confusum* Duval and *Tribolium castaneum* Herbst). He has maintained carefully controlled experiments over a period of four years, and a number of interesting phenomena of extinction were observed. The theoretical investigation suggested at (*a*) above would have an obvious bearing on the interpretation of the work of Park and his school.

In conclusion I should like to express my deep gratitude to Professor M. S. Bartlett and to the many other friends whose advice and encouragement have been a continual help to me during the course of this work.

Appendix.—*The Temporal Development of the Cumulant-generating Functional.*

The cumulant-generating functional has been defined as

$$K[\theta(x);t] = \log E\{\exp(\int_0^\infty \theta(x)\, d\,N(x,t))\}$$

$$= \log E\{\exp \Sigma_t\, \theta(x)\}$$

(where the summation is over all members of the population alive at the epoch t), and its development in time is governed by the "differential equation" stated at (89) in the main text of this paper. It seems desirable to give here a few more details of the rather complicated argument which leads to this result.

Consider the conditional expectation of

$$\exp \int_0^\infty \theta(x)\, d\,N(x, t+dt) = \exp\, \Sigma_{t+dt}\, \theta(x),$$

when the age-distribution $N(x,t)$ at the epoch t is given. It is equal to

$$\Pi_t\{e^{\theta(x+dt)}(1-\mu(x)dt) + \mu(x)dt\}\, \{e^{\theta(o)}\Sigma_t\lambda(x)dt + 1 - \Sigma_t\, \lambda(x)dt\},$$

which to the first order in dt is the exponential of

$$\Sigma_t \{\theta(x) + \theta'(x)dt - \mu(x)\, dt\, (1 - e^{-\theta(x)}) + \lambda(x)\, dt\, (e^{\theta(o)} - 1)\}.$$

The conclusion (89) now follows on averaging over the conditions obtaining at the epoch t.

Professor Bartlett has pointed out to me that the introduction of the derivative $\theta'(x)$ is quite unnecessary, and that the first two terms inside the bracket in the last formula could be replaced by $\theta(x + dt)$. This is also a desirable modification, because in practice $\theta(x)$ would usually be identified with a discontinuous function.

References

(a) *General.*—The two most important references are:

Feller, W. (1939), "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung," *Acta Biotheoretica*, **5**, 11–40. (Abstract in *Math. Reviews*, **1**, 22).

Bartlett, M. S. (1947), *Stochastic Processes.* (Notes of a course given at the University of North Carolina in the Fall Quarter, 1946.) (There is a copy in the Society's library.)

The remaining references to the literature have been roughly grouped according to subject.

(b) *Deterministic treatment of population problems.*—General reviews of this field will be found in the articles by Rhodes (1940) and Lotka (1945).

Bernardelli, H. (1941), "Population waves," *J. Burma Research Soc.*, **31**, 1–18.

Feller, W. (1941), "On the integral equation of renewal theory," *Ann. Math. Stat.*, **12**, 243–67.

Karmel, P. H. (1947), "The relations between male and female reproduction rates," *Population Studies*, **1**, 249–74. (See also two further papers by Karmel, on the same topic, in volumes 1 and 2 of the same journal.)

Leslie, P. H. (1945), "On the use of matrices in certain population mathematics," *Biometrika*, **33**, 183–212.

Lewis, E. G. (1942), "On the generation and growth of a population," *Sankhyā*, **6**, 93–6.

Lotka, A. J. (1945), "Population analysis as a chapter in the mathematical theory of evolution"; a contribution to *Essays on Growth and Form presented to D'Arcy Wentworth Thompson*, ed. by W. E. le Gros Clark and P. B. Medawar, Oxford.

—— (1948), "Application of recurrent series in renewal theory," *Ann. Math. Stat.*, **19**, 190–206.

Rhodes, E. C. (1940), "Population mathematics," *J. Roy. Stat. Soc.*, **103**, 61–89, 218–45, and 362–87.

(c) *The extinction of surnames, and related topics.*

de Candolle, A. (1873), *Histoire des Sciences et des Savants depuis deux siècles.* Geneva. (See especially pp. 388–9.)

Watson, H. W., and Galton, F. (1873), Problem 4001, in *Mathematical Questions, with their Solutions, from the 'Educational Times,'* **19**, 103–5.

—— (1874), "On the probability of the extinction of families," *J. Anthrop. Inst.*, **4**, 138–44.

Galton, F. (1889), *Natural Inheritance.* London. (The 1874 paper, slightly modified, forms Appendix F to this work.)

Pearson, K. (1914–30), *The Life, Letters, and Labours of Francis Galton.* Cambridge.

Erlang, A. K. (1929), Problem 15 in *Matematisk Tidsskrift*, B, p. 36.

Steffensen, J. F. (1930), "Om Sandsynligheden for at Afkommet uddør," *Matematisk Tidsskrift*, B, pp. 19–23.

—— (1933), "Deux problèmes du calcul des probabilités," *Ann. Inst.H. Poincaré*, **3**, 319–44.

Lotka, A. J. (1931), "The extinction of families," *J. Washington Acad. of Sciences*, **21**, 377–80 and 453–9.

Fisher, R. A. (1922), "On the dominance ratio," *Proc. Roy. Soc. Edin.*, **42**, 321–41.

—— (1930), *The Genetical Theory of Natural Selection.* Oxford.

Haldane, J. B. S. (1949), "Some statistical problems arising in genetics," *J. Roy. Stat. Soc.*, B, **11**, 1–9,

Woodward, P. M. (1947), "A statistical theory of cascade multiplication," *Proc. Cambridge Phil. Soc.*, **44**, 404–12.

Reference must also be made to the paper by Yaglom included under heading (e), and to three papers by T. E. Harris, of which abstracts alone have been published (in *Ann. Math. Stat.*, **18**, 611; **19**, 116; and **19**, 433).*

* The following have now appeared: T. E. Harris, "Branching processes," *Ann. Math. Stat.*, **19** (1948), 474–94; and R. Bellman and T. E. Harris, "On the theory of age-dependent stochastic branching processes," *Proc. National Acad. Sci.*, **34**, (1948), 601–4. The second of these papers is concerned with the effect of assuming a general distribution for what I have called the "generation time."

(d) *General references for section* 2.

Arley, N. (1943), *On the Theory of Stochastic Processes, and their Application to the Theory of Cosmic Radiation.* Copenhagen.
——, and Borchsenius, V. (1945), "On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics," *Acta Math.*, **76**, 261–322 (especially pp. 298–9).
Brown, A. W. (1940), "A note on the use of a Pearson Type III function in renewal theory," *Ann. Math. Stat.*, **11**, 448–53.
Feller, W. (1943), "On a general class of 'contagious' distributions," *ibid.*, **14**, 389–400.
Fisher, R. A. (1941–2), "The negative binomial distribution," *Ann. Eugen.*, **11**, 182–7.
——, Corbet, A. S., and Williams, C. B. (1943), "The relation between the number of species and the number of individuals in a random sample of an animal population," *J. Animal Ecol.*, **12**, 42–58.
Furry, W. H. (1937), "On fluctuation phenomena in the passage of high energy electrons through lead," *Phys. Rev.*, **52**, 569–81.
Greenwood, M., and Yule, G. U. (1920), "An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents," *J. Roy. Stat. Soc.*, **83**, 255–79.
Irwin, J. O. (1941), Comments on a paper by E. G. Chambers and G. U. Yule in *J. Roy. Stat. Soc. Suppl.*, **7**, 101–7.
Kelly, C. D., and Rahn, O. (1932), "The growth rate of individual bacterial cells," *J. Bacteriology*, **23**, 147–53.
Kendall, D. G. (1948*a*), "On the generalized 'birth-and-death' process," *Ann. Math. Stat.*, **19**, 1–15.
—— (1948*b*), "On some modes of population growth leading to R. A. Fisher's logarithmic series distribution," *Biometrika*, **35**, 6–15.
—— (1948*c*), "On the role of variable generation time in the development of a stochastic birth process," *ibid.*, **35**, 316–30.
Lévy, P. (1943), "Un théorème d'invariance projective relatif au mouvement Brownien," *Comm. Math. Helvetici*, **16**, 242–8.
Lundberg, O. (1940), *On Random Processes and their Application to Sickness and Accident Statistics.* Uppsala.
McKendrick, A. G. (1914), "Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of a progressive nature," *Proc. London Math. Soc.* (2), **13**, 401–16.
Milne, A. (1943), "The comparison of sheep-tick populations (*Ixodes ricinus* L.)," *Ann. Appl. Biol.*, **30**, 240–50.

(e) *The work of the Russian school on "branching" stochastic processes.*

Kolmogoroff, A. N., and Dmitriev, N. A. (1947), "Branching stochastic processes," *C.R. (Doklady) Acad. Sci. U.R.S.S.* (N.S.), **56**, 5–8. (English.)
——, and Savost'yanov, B. A. (1947), "The calculation of final probabilities for branching random processes," *Doklady Akad. Nauk S.S.S.R.* (N.S.), **56**, 783–6. (Russian: abstract in *Math. Reviews*, **9**, 149.)
Savost'yanov, B. A. (1948), "On the theory of branching random processes," *ibid.*, **59**, 1407–10. (Russian: abstract in *Math. Reviews*, **9**, 451.)
Yaglom, A. M. (1947), "Certain limit theorems of the theory of branching random processes," *ibid.*, **56**, 795–8. (Russian: abstract in *Math. Reviews*, **9**, 149.)

This series of papers is chiefly concerned with what might be called "the problem of k sexes, all female"; that is to say, it deals with a population of objects of k types $T_1, T_2, \ldots, T_k$, such that in unit time an object of type T_i has a chance $p_i(n_1, n_2, \ldots, n_k)$ of being transformed into n_1 objects of type T_1, n_2 objects of type $T_2, \ldots$, and n_k objects of type T_k. This abstract scheme covers a large variety of particular problems in biology and physics. The authors derive a number of limit theorems about the ultimate behaviour of systems of this sort, and as an example Kolmogoroff and Dmitriev consider the simple birth-and-death process discussed by Feller and Arley, and derive Palm's formulae (21) and (22) by the generating function method. The paper by Yaglom gives a very detailed discussion of the Galton-Watson "surname" problem, and includes a number of remarkable results about the ultimate distribution.

(f) *General references for sections* 3 *and* 4.

Bochner, S. (1947), "Stochastic processes," *Ann. of Maths.*, **48**, 1014–61.
Le Cam (1947), "Un instrument d'étude des fonctions aléatoires: la fonctionnelle caractéristique," *Comptes Rendus*, **224**, 710–1.
Luria, S. E., and Delbrück, M. (1943), "Mutations of bacteria from virus sensitivity to virus resistance," *Genetics*, **28**, 491–511.
Park, Thomas (1948), "Experimental studies of interspecies competition, I," *Ecol. Monographs*, **18**, 265–308.

DISCUSSION ON SYMPOSIUM ON STOCHASTIC PROCESSES

The CHAIRMAN invited the reader of each paper to comment on the other papers.

Mr. J. E. MOYAL: I should like to express my appreciation of Professor Bartlett's and Mr. Kendall's papers, both, in my opinion, fundamental contributions to the subject of the theory of stochastic processes. I should also like to thank the Society for giving me this opportunity of expressing my views on the applications of this theory to physical problems.

Kendall in his paper introduces the function $N(x, t)$, specifying the age distribution of a population at time t. He then treats the problem by the method of moment-generating functionals, which amounts roughly to considering the increments $\delta N(x, t)$ as an infinite set of correlated variates attached to each t, the whole set forming a Markoff process in t. I would suggest an equivalent approach to the problem, which is to consider $N(x, t)$ as a stochastic process in the two independent variables: the time t, and the age x. This formulation is, perhaps, improved by treating N as a random function of *points* on the t-axis, and of *intervals* or sets of intervals on the x-axis (the last is in analogy with the treatment of additive processes; the reason why one wants to express N as a function of intervals is connected with the difficulty of defining a population *density* in age, because the probability of $\delta N \neq 0$ tends to 0 when $\delta x \to 0$). The process thus defined would be Markovian in t, but would not in general, as is clearly shown in Kendall's discussion, form either an additive or a Markoff process in x. Let us then write in Kendall's notation,

$$N(\Delta x, t) = \int_{x_1}^{x_2} dN(x, t) \qquad (\Delta x = x_2 - x_1)$$

We may say, generalizing the simple definition of a random function, that such a process will be completely specified by the joint distribution

$$F_k(N_1, N_2, \ldots, N_k; \ \Delta x_1, t_1, \ldots, \Delta x_k, t_k)$$

of the population numbers $N_1, \ldots, N_k$ in any k age intervals $\Delta x_1, \ldots, \Delta x_k$, at the times $t_1, \ldots, t_k$. Thinking of the process in these terms is a mere matter of terminology. The more practical point I wish to stress is that one could extend to random functions of several variables the technique invented to derive the differential equations satisfied by the generating functions of single-dimensional processes; the formulations I suggest would make this extension applicable to the problem of age-distribution in population theory.

The problem of cosmic ray showers discussed by Bartlett and Kendall is essentially of the same nature. In Kendall's notation it is a process defined by *three* correlated functions: $N_+(E, \xi)$ $N_-(E, \xi)$ and $N_p(E, \xi)$, specifying respectively the *energy* distributions of positive electrons, negative electrons and photons at a distance ξ from the point of inception of the shower: this distance plays thus the role of the time in the age-distribution problem, while the energy E plays the role of the age. This is really a special case of the more general theory of quantum kinetics. In §8.6 and §8.7 of my paper I discussed this theory, limiting myself to the case of a *discrete* energy spectrum. I believe an approach of the type suggested above will be more appropriate in order to discuss the case where the energy spectrum is *continuous*, or partly continuous and partly discrete.

Professor M. S. BARTLETT: When I first approached Mr. Moyal and Mr. Kendall and suggested that we might co-operate in presenting this symposium they willingly assented. I would like to take this opportunity of thanking them for their magnificent response in their two comprehensive papers.

On these papers I have only two trivial comments. In §3.6 Mr. Moyal has defined the Markoff process merely in terms of the distribution at any one time t being dependent only on the known state of the system at the last available (and known) time and not on still previous times. In order to be quite general one should stipulate that the *simultaneous* distribution at a number of instants depends only on the state at the last available previous time.*

The other point is in Mr. Kendall's paper in connection with his discussion of the moment-generating functional. When discussing the use of the arbitrary function $\theta(x)$ of the variable x

* After the meeting I realized that this comment is unnecessary, as it may easily be deduced by elementary probability logic that the two definitions are equivalent. Such simple logical arguments are often useful; for example, it may also be deduced that corresponding to any Markoff process evolving forwards in time there is an equivalent Markoff process considered in the reverse direction, and also a "linear nearest neighbour system" (in which only the nearest known states before and after time t affect the probability at time t), and conversely. It should, perhaps, be added that the explicit probabilities in the "reversed" process do not bear a very simple relation to those for the "forward" process unless the Markoff process is either stationary or deterministic.

he has introduced in equation (90) the derivative $\theta'(x)$. This does not matter much in a formal treatment, but it seems better to avoid using such a derivative, using instead the differential $d\theta(x)$.

Mr. Kendall has done much interesting research on the origins of some of the distributions which arise. In particular he has referred to a paper in 1914 by McKendrick who discussed some stochastic processes defined for continuous time and derived the negative binomial distribution. As I have noted in my paper, a distribution may arise in various ways. In the case of the negative binomial there was the heterogeneity assumption considered by Greenwood and Yule in 1920, and later in physics the same distribution was obtained again by a stochastic process argument; but the latter, as Kendall has pointed out, had already been given by McKendrick in 1914—truly a remarkable achievement.

Mr. DAVID G. KENDALL: I greatly enjoyed reading Mr. Moyal's paper, and found particularly interesting the account which he gives of the pure-mathematical approach to stochastic-process theory. In my own work on population processes I have adopted quite deliberately the point of view of the applied mathematician, because the whole object of these preliminary skirmishings has been a rapid exploration of the territory. Sooner or later, of course, consolidation will be necessary, and it is good to see that the heavy artillery of Hilbert space is already fully deployed!

There are two detailed points arising out of Professor Bartlett's paper which I should like to mention here, and the first concerns his very ingenious application of the theory of the Brownian motion to the construction of an elementary "intuitive" approach to the Kolmogoroff test of "goodness of fit"; this is especially welcome at the present time, because the publication last year of Smirnoff's tables (in the *Annals of Mathematical Statistics*) put the test on a sound practical basis, and important applications are to be expected. (An example of what can be done in this way will be found in the Introduction to Wold's *Table of Random Normal Deviates*, recently issued in the *Tracts for Computers* series.) It will be remembered that there is a closely related test of "goodness of fit" which employs the statistic

$$\omega^2 = \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \, dF(x);$$

tests of this type were first introduced by Cramér and by von Mises, but in its final form the theory is due to Smirnoff. Now it seems to me that one might profitably develop Professor Bartlett's line of argument in connection with this test also; possibly one would have to consider the distribution of the random variable,

$$\int_0^T \{x(t)\}^2 \, dt,$$

where $y = x(t)$ denotes a *conditioned* Brownian motion, the conditions being $x(0) = x(T) = 0$. Distributions of such "Wiener functionals" have been given by Cameron, Martin, Erdös and Kac,* and Lévy's work on two-point boundary problems would also be relevant.

My second point concerns Professor Bartlett's equation (28), which excited me very much when he first showed it to me because it suggested the word *group*, which has such a high emotional content for contemporary mathematicians. Actually it is not in general a group but a semi-group which is relevant, and it is not difficult to see that if

$$\varphi\,(z, t) \equiv \sum_{(n)} z^n \, p_n(t)$$

is the generating function associated with the size n of a population stemming from a single individual at $t = 0$, then φ satisfies the functional equation

$$\varphi(z, t + \tau) = \varphi\{\varphi(z, t), \tau\}.$$

Thus if the family of transformations $\{T_t\}$ is defined by

$$T_t \, . \, z \equiv \varphi(z, t),$$

it will be seen that $T_{t+\tau} = T_\tau T_t$, and that $T_0 = 1$, so that the T_t form a one-parameter semi-group with parameter t. In the special case of the simple birth-and-death process it happens that the

*See the paper by M. Kac, *Trans. American Math. Soc.*, **65** (1949), 1–13, for details and further references.

extension to a complete group can be made by formally adjoining the transformations T_{-t} associated with negative values of the parameter, and then one has $T_{-t}\, T_t = 1$.

The functional equation is, of course, just another form of the Chapman-Kolmogoroff equation, and so one may expect a semi-group to be associated with every stochastic process of the Markoff type. This is so, and some of the details have been worked out in Hille's recent book on *Functional Analysis and Semi-Groups*. In the present instance it may be noted that the technique we have used for determining φ amounts to writing down the infinitesimal transformation

$$T_\varepsilon \, . \, z = z + \varepsilon(\lambda z - \mu)(z - 1),$$

and from it generating the finite equations of the group in the usual way. Also it can now be seen that Bartlett's equation (28) is associated with a famous result in the classical theory of continuous groups. This asserts that a one-parameter continuous group can always be reduced, by suitably transforming the variable and the parameter, to the canonical form

$$T_u \, . \, \zeta \equiv \zeta + u$$

— the group of additions.

The CHAIRMAN: It is my pleasant duty to propose from the chair a vote of thanks to the three speakers, and in so doing to emphasize how great an honour I consider it to preside on what is, I think, an epoch-making occasion. These three papers contribute greatly to the unification of our subject.

Perhaps I may be allowed to indulge in a few personal reminiscences. In 1931, being very young and bold, I read a paper to the Study Group of this Society entitled "Is the Universe Statistical?" I felt then, perhaps rather dimly, that many of the ideas being developed in modern mathematical physics were the same statistical ideas that we were using in other fields, mainly biological; but to develop the connection properly was then quite beyond my capacity.

A year or two before the war I remember having a discussion with Professor Bartlett on the same subject, and I thought he might give us a contribution on it to be added to our "Recent Advances" series of articles, but we were a little discouraged by a most distinguished authority and nothing came of the idea at the time.

At the next stage in the story we ought, I think, to remember the name of our former but still very active Assistant Secretary, Miss Thorburn, for it was she who, in 1940, noticed a quiet young man working in the reading-room at the Society and realized his ability. He had just come over from Paris and was enabled to go to Cambridge shortly afterwards. He had some very original ideas on the statistical basis of modern physics, and these attracted the attention of Harold Jeffries and later of Dirac. That young man is with us to-day in the person of Mr. Moyal, and in the paper he has presented we have the full fruits of his efforts.

A great deal of the technical detail it contains I hardly understand, but I have no doubt whatever of the importance of Mr. Moyal's paper in the clarification and unification of ideas.

I much enjoyed, too, the clarity of Mr. D. G. Kendall's exposition, and I foresee that it will have much influence in applications to my own subject.

Professor Bartlett's paper opens endless vistas of new research in the applied field. There are only two points of detail on which I should like to comment. On p. 242 of Mr. Kendall's paper he shows two ways in which the negative binomial distribution can arise. There are at least two others. Speaking in terms of accidents, it may arise not only from unequal accident proneness, but also, in a population of people equally prone to start with, from the probability of an accident depending on the number previously suffered. It may also arise in the case of multiple accidents, as I pointed out in 1941 in the discussion on Chambers' and Yule's paper to which Mr. Kendall referred.

It is worth while pointing out that equation (20) on p. 217 was given by the late H. E. Soper in an important appendix to a paper by W. R. Thompson in the *Bulletin of Entomological Research* for March, 1931. The discussion is elementary in character and has Soper's invariable clearness of exposition.

May I remind you once more of a saying of Quetelet's which I last quoted at a meeting held by this Society in 1934: "*L'urne que nous interrogeons c'est la Nature.*" To what extent is it true that we may reverse this and say: "*La Nature que nous interrogeons c'est une urne* "? The papers to which we have listened may help to provide the answer.

The vote of thanks was put to the meeting and carried unanimously.

Dr. NIELS ARLEY: I should like to express my thanks for the invitation to take part in this meeting. There are many interesting points on which I might comment, but I will confine myself to only a few on the paper by Mr. Moyal.

First, a question of history: on p. 156 of the paper it is stated that the method of generating functions was first introduced into the field of statistical mechanics by Professor Bartlett in a paper in 1937. This statement is not quite correct, as the whole book on statistical mechanics by Fowler, like the earlier papers by Darwin and Fowler from 1922, is entirely based on that method.

Next I should like to mention that the formalism of quantum mechanics also can be conveniently developed by means of characteristic functions, as I shall show in a paper to be published elsewhere.

Finally, I should like to ask Mr. Moyal to give more detailed information on whether and under what conditions the phase-space density $F(p_i, q_i, t)$ will (as $t \to \infty$) converge towards the canonical distribution of Gibbs, exp $[(\psi - H)/kT]$. I think the only satisfactory theoretical justification for using the canonical distribution for the description of equilibrium states would be to formulate the starting-point of statistical mechanics so that F tends towards Gibbs' function when $t \to \infty$, corresponding to the experimental finding that the physical system described by F does actually tend towards an equilibrium state, all physical properties of which experience shows can be described satisfactorily by means of Gibbs' function.

After these few comments I should like to indicate how the exact distribution of the general "birth-and-death" process can be deduced in a quite elementary way. I am sure that stochastic processes will have an ever-increasing field of application in the future. Their only drawback is that except in the simplest cases it is either extremely complicated or even impossible to solve the equations exactly or even approximately. I therefore think there is a certain interest in deducing the solution by as elementary a method as possible. I give the following outline of a paper which will appear in the next number of *Skandinavisk Aktuarietidsskrift.**

If $\lambda(t)dt$ and $\gamma(t)dt$ denote the asymptotic probabilities of "birth" and "death," respectively, in the time interval dt at t, the probabilities $P(n,t)$ of $\mathbf{n} = n$ at t will satisfy the infinite system of simultaneous differential equations

$$\frac{d}{dt} P(n,t) = (n+1)\gamma(t)P(n+1,t) + (n-1)\lambda(t)P(n-1,t)$$

$$- n(\lambda(t) + \gamma(t))\, P(n,t) \qquad (n \geq 1) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\frac{d}{dt} P(0,t) = \gamma(t)\, P(1,t), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

with the initial conditions

$$P(1,0) = 1, \;\; P(n,0) = 0 \; (n \neq 1). \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

For $\gamma = 0$ these equations can be solved directly and the $P(n,t)$ are then found to be the terms in a geometric series. It is therefore a natural hypothesis in the general case, $\gamma \neq 0$, to assume that one may write

$$P(n,t) = \varphi(t)\, \psi^{n-1}(t) \qquad (n \geq 1)$$

$$P(0,t) = 1 - \sum_{n=1}^{\infty} P(n,t). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Inserting (4) into (1) and (2) one finds that $\varphi(t)$ and $\psi(t)$ satisfy two differential equations which are in fact *independent* of n, thus proving the correctness of the hypothesis (4). Next, the analytical expressions for φ and ψ are most conveniently obtained from the moments of the first and second orders, which from (4) are found to be

$$\mu_1(t) = \bar{\mathbf{n}} = \frac{\varphi(t)}{(1 - \psi(t))^2},$$

$$\mu_2(t) = \overline{\mathbf{n}^2} = \varphi(t)\frac{1 + \psi(t)}{(1 - \psi(t))^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Due to (1), μ_1 and μ_2 satisfy the equations

$$\frac{d}{dt}\mu_1 = (\lambda - \gamma)\mu_1,$$

$$\frac{d}{dt}\mu_2 = 2(\lambda - \gamma)\mu_2 + (\lambda + \gamma)\mu_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

* This has now appeared *Skand. Akt. Tidsskr.*, 1949, p. 21.

which can easily be solved. Inserting the results into (5) and solving for φ and ψ we obtain, of course, exactly the same expressions for $P(n,t)$ as D. G. Kendall has previously obtained by means of the much more complicated calculations of the method of generating functions.

Professor C. A. COULSON (King's College, London): One conclusion has already come out of this discussion, and it is that in one form or another stochastic processes play an almost central part in nearly every branch of physics. I believe this is of considerable importance at the present time, and that we have here an example in the theoretical field similar in many respects to what is often found in the experimental field. I mean that progress in experimental physics has often followed the development of some new apparatus—the existence of the apparatus has both elucidated old problems and stimulated the attack on new ones. Examples spring so easily to mind that it is almost superfluous to recall the effects on the development of experimental physics of such new tools as the mass spectrograph, the electron microscope, coincidence counters in cosmic ray work or the photographic emulsion technique for the study of nuclear fission. I mention all this, however, because it does seem to me that the recent understanding of this central theoretical technique of random events and processes may do for theoretical physics something of the same kind as the experimental techniques have done for experimental physics. I remember Professor Born, of Edinburgh, saying once, after a colloquium on matrix methods, that when he was developing the theory of crystal vibrations, the whole subject would have marched forward vastly more surely and rapidly if he had had available the resources of present-day matrix calculus. I think that we may reasonably hope to see considerable stimulus given to some of the other branches of theoretical physics as a result of the sort of analysis that has been expounded this afternoon. For that reason, if for no other, the whole discussion seems most timely.

After these general comments what I have now to say may seem rather trivial; but among the many applications of the birth-and-death processes discussed by Kendall there is one in which I have taken some interest, and which may be worth describing.

It has been found that bacteria are able to undergo mutations in which the new bacterium appears as a closely related but nevertheless quite distinct organism, propagating itself by division in much the same way as the normal non-mutated variety. For example, *B. coli* normally grow in a medium which does not need to contain the substance biotin. But a mutated form exists which is unable to grow without the presence of biotin. Mutations from one type to the other occur, particularly in the presence of radiation, or (sometimes) suitable chemicals, so that it becomes a matter of some importance to determine the mutation rate. This general situation is quite common; everyone is aware of the serious effects which result from the developed resistance of various forms of coccus to penicillin; but resistance is also developed to certain other chemicals, and to radiation; there are also changes in bacteria which affect fermentation reactions. In some respects, however, the best-known example of this situation occurs in the experiments of Luria and Delbruck referred to by Kendall, in which bacteria acquire an immunity against certain bacteriophages which are able to destroy the normal form, but not the mutant. Again it is important to be able to tell the mutation rate. Such rates appear to be very small, of the order of 10^{-6}, so that out of every 10^6 divisions of an organism there is one mutation. This means that the number of mutants in a culture is not generally very large, and a complete quantitative discussion of the process, of the kind outlined by Kendall, is highly necessary.

Some three years ago the late Dr. D. E. Lea, of Cambridge, and I became interested in this problem; it is, of course, a restricted version of the λ, μ, κ process dealt with in more general terms by Kendall. In many respects our analysis is similar to his.

We imagine that a culture is started containing one normal bacterium; and that this develops by division in the usual way, so that the total number n of bacteria present at time t is $e^{\lambda t}$. From time to time one of these mutates to the new form, and the new form itself grows by division, similarly to the original type. The problem is to discuss the probability distribution of the number of mutants after time t.

We can see at once that such a probability distribution will be very skew. For if a mutation occurs early on in the growth, it will have "fathered" a very large number of progeny; but if it has occurred quite near the end of the time of observation, there will not have been sufficient time for the new species to multiply appreciably. As a result the probability $p_n(t)$ that at time t there are just n mutants is such that there is a small, but non-zero, value of p_n even for very large n, and a much larger value for small n. This extreme skewness—as I shall show in a moment—presents interesting problems in the important need to estimate the mutation rate as accurately as possible from the observed distribution $p_n(t)$.

In principle the problem of determining $p_n(t)$ has been solved by Kendall. The answer is that p_n is the coefficient of x^n in the expansion of a certain generating function (his equations 48 and 49). Now that is all very well if you are a mathematician. But if you want to find out the probability of, say, there being 30 to 40 mutants, you have to expand the generating function in powers of x up to x^{40}. The matter is only put in this form to show how utterly impossible it is—which shows that there is a world of difference between a solution suitable for a mathematician and a bacteriologist. Lea and I found that we could make an approximation in the generating function (equivalent roughly to replacing the total number N of bacteria in certain expressions by N–1) and a new generating function resulted which permitted reasonably simple recurrence relations to be found between successive p_n. As a result we were able to prepare tables, now in course of publication, of p_n for values of n up to 64.

But to return to the question of estimating the mutation rate; Lea was able to show that although p_n had an extremely skew distribution, it was possible by a suitable combination of quantities to obtain a variate whose distribution was almost exactly normal. This is the quantity

$$\left(\frac{a}{n/m - \log m + b} - c\right)$$

where a, b, c are constants, n is the number of mutants at a time when the mean number of mutations that have occurred is m. m is simply related to the quantities introduced by Kendall. The existence of such an effectively normal variate could not easily be shown theoretically, but it does enable the ordinary experimental scientist to use the sort of technique with which he is familiar in order to estimate the mutation rate.

We have investigated several distinct ways of estimating this rate—which is really the goal of most of the experiments. There is, naturally enough, a variety of ways in which this could be done. For example, we could use—

(i) The chance of there being no mutations at all; this has frequently been done and was a reasonably good method if the mean number of mutations does not exceed about 7; but it rapidly worsens when there are more than this number.

(ii) The median of a set of parallel cultures.

(iii) The mean value of the normal variate, determining the mutation rate in such a way that the mean value is zero.

(iv) The method of maximum likelihood, in the form developed by R. A. Fisher. This last method is the best of all four, because it uses all the information available from the experiments that have been made, whereas the other three methods, in varying degrees, fail to make the best use of the observed material.

I mention all this partly because of the concern that Lea and I had in the matter; partly because the problem itself possesses an intrinsic interest in many applications, and partly also because it seems to me to be important to realize that what is an effective and final solution for the mathematician is, in itself, only part of the way towards a complete solution of the problem which would satisfy the man or woman actually doing the experiments. Such a person is trying his or her best to estimate certain parameters, the success or reliability of which will then be used to decide whether the interpretation originally proposed for some phenomenon is or is not a valid one.

For these reasons I should like to plead that the mathematicians who develop these theories should also make themselves responsible for seeing how they work out in actual practice.

Professor M. Greenwood: My only excuse, apart from the normal talkativeness of elderly folk, for accepting the invitation to speak is that I may say how profoundly interested I am in what is now happening.

The idea which fills me with hope and which has much impressed me arises out of a remark by Professor Bartlett in his paper in regard to experimental epidemiology. I personally had a long experience of experimental epidemiology when working with the late Professor Topley. The difficulty I have always had about it is that experimental epidemiology is an extremely expensive and slow process. Probably in the course of our sixteen or eighteen years' collaboration Topley and I have sacrificed on the altar of science 200,000 mice. One starts with a small number of infected and healthy mice and builds up the population by admitting healthy immigrants. It takes some years to discover how to do this without casual infection spoiling an experiment. A single experiment continued for years uses up thousands of mice, but is just one experiment to estimate stochastic effects. I have in mind a long experiment with a particular virus in which the population was in equilibrium for a long time. The population only fluctuated in what might be regarded as a purely random way. Then, three or four days after a heat wave one or two mice died, and, after a pause, something began to happen. Mortality increased faster and faster and the herd, which consisted in equilibrium of about 400 mice, was reduced to 30 or 40.

The temptation to believe that, somehow, the heat wave was the "cause" of the disaster was irresistible; but we remembered that this was just one experiment, and an experiment we could not repeat. Even if the belief were justified, *how* did the extrinsic factor operate? One can imagine a number of stochastic possibilities, but has not that power of continuous mathematical thought commanding a good mathematical technique which would attain to an arithmetically useful solution. The researches of Professor Bartlett and the other speakers should encourage the experimental epidemiologists of the future because they contain a promise that, in the time to come, experimenters will be provided with an arithmetical technique which will enable them to solve problems as yet insoluble. Already, for instance, in the Medical Research Council's report, *Experimental Epidemiology* (Special Report Series, No. 209, S.O., 1936), and in a series of paper in the *Journal of Hygiene* and elsewhere, there are plenty of arithmetical data upon which a computor can try out new methods of research.

It has been a pleasure to be present at what, I think, is an epoch-making occasion, the revelation of a new organon.

Dr. I. J. Good: I would like first to make a few remarks concerning branching processes with *discrete time*. The Rev. H. W. Watson's theorem* of the iterated function can be extended in two different directions. Both extensions are trivial and presumably known. The first is when the fertility of the individuals depends on time, i.e. on the ordinal number of the generation; the second is when there are several types of individuals, all "female."

(i) Let the probability generating functions for the numbers of children of an individual in the successive generations be $f_0(x), f_1(x), f_2(x), \ldots$ Then the probability generating function for the number of descendants in the r^{th} generation of one individual in the zero$^{\text{th}}$ generation is (with the usual independence assumptions)

$$f_r(f_{r-1}(\ldots f_0(x)\ldots)).$$

(ii) For simplicity consider the case of two types of individual. Let the probability generating functions for the numbers of children of an individual of the two types be $F(x,y)$ and $G(x,y)$, where, for example, the coefficient of x^3y^7 in $F(x,y)$ is the probability that an individual of the first type will have three children of the first type and seven of the second type. Then the probability generating functions for the numbers of descendants in the r^{th} generation are $F_r(x,y)$ and $G_r(x,y)$, where $F_1(x,y) = F(x,y)$, $F_r(x,y) = F(F_{r-1}(x,y), G_{r-1}(x,y)) = F_{r-1}(F(x,y), G(x,y))$, etc.

The ideas of (i) and (ii) can, of course, be combined into a single generalization of Watson's theorem. (See equation (3) below.)

Now consider the following problem. What is the probability generating function for the *total* number of descendants, in the first r generations, of one individual (under H. W. Watson's simple assumptions)? Let the probability generating function for the number of children of one individual be $f(x)$. Then the required probability generating function is

$$g_r(x) = f(xf(xf\ldots xf(x)\ldots)), \qquad (1)$$

where $g(x) = g_1(x) = f(x)$, $g_{s+1}(x) = g(xg_s(x))$ $\quad (s = 1, 2, 3, \ldots)$.

This result is proved, among others, in my forthcoming note in the *Proc. Cam. Phil. Soc.*, to which Professor Bartlett has already referred. From this result it is easy to deduce the first and higher moments. A corollary is that if the expected number of children of an individual is $k < 1$, then the probability generating function $y = y(x)$ for the total number of descendants of an individual by the time the line becomes extinct satisfies the functional equation

$$y = f(xy) \qquad (2)$$

The expected number of descendants is $k/(1-k)$ and the variance is $\sigma^2/(1-k)^3$, where σ^2 is the variance of the number of children of an individual, assuming this variance exists.

It is interesting to observe that (1) can be deduced from the generalization of H. W. Watson's theorem to the case of two sexes, both female. (A direct proof is given in my other manuscript.) Call the two types of individual "type I" and "type II." In this model the individuals of type I are the "real" individuals, while those of type II are introduced as a mathematical trick. (They can be regarded as ghosts.) Every individual always has exactly one child of type II. Therefore, with the previous notation,

$$F(x,y) = y.f(x),\ G(x,y) = y.$$

The total number of individuals in the r^{th} generation in the new model is equal to the number of individuals of type I in the whole tree, i.e. it is equal to the total number of individuals in the first r generations (including the zero$^{\text{th}}$ generation) in the original problem. Therefore the required

* See Professor Bartlett's paper, equation (20).

probability generating function is $F_r(x,x)$, or rather $x^{-1}F_r(x,x)$ if we wish to exclude the original individual. The result (1) now follows by observing that

$$G_{n+1}(x,y) = G(F_n(x,y),\ G_n(x,y)) = G_n(x,y) = \ldots = y,$$

and

$$\begin{aligned} F_{n+1}(x,y) &= F(F_n(x,y),\ G_n(x,y)) \\ &= G_n(x,y).f(F_n(x,y)) = y.f(F_n(x,y)). \end{aligned}$$

Secondly, I will deal with *transition from discrete to continuous time*. For purposes of calculation, especially on an electronic digital computer, it is necessary to use models with discrete time, even when the time is continuous in the original problem (if time ever is continuous). In theoretical work the reverse process can be used.

Suppose that there are several types of individuals, and that for an individual of type x existing at time t (*i.e.* generation number t) the probability generating function for the numbers of children is

$$f_{x,t}(z_1,\ z_2,\ z_3,\ \ldots)\ (x = 1, 2, 3, \ldots\ ;\ t = 0, 1, 2, \ldots).$$

Let $F_{x,t}(z_1, z_2, \ldots)$ be the probability generating function for the numbers of descendants in the t^{th} generation of an individual of type x which exists at time 0. (For example, $F_{x,0} = z_x$, $F_{x,1} = f_{x,1}$.). Then the general generalization of H. W. Watson's theorem, already referred to implicitly, is that

$$F_{x,t+1}(z_1,\ z_2,\ \ldots) = f_{x,t}\{F_{1,t}(z_1,\ z_2,\ \ldots),\ F_{2,t}(z_1,\ z_2,\ \ldots),\ \ldots\}. \quad . \quad . \quad (3)$$

Let $u_x = F_{x,t}(z_1, z_2, \ldots)$ and subtract u_x from both sides of equation (3). Then if the time interval is small we have approximately,

$$\frac{du_x}{dt} = f_{x,t}(u_1,\ u_2,\ \ldots) - u_x \qquad (x = 1, 2, 3, \ldots) \quad . \quad . \quad . \quad (4)$$

Equations (4), which generalize equation (29) of Professor Bartlett's paper, are exact for any problem with continuous time which can be arbitrarily well approximated by a problem with discrete time and small time intervals.

If the different types of individuals are fundamentally of the same type but of different ages, the age of an individual being denoted by x, then the ordinary differential equations (4) will usually go over into a partial differential equation.* (If there are two fundamental types then there will be two partial differential equations, and so on.) As an example, consider the birth and death process, allowing for variability of fertility both with age and with time. Then $f_{x,t}$ is approximately of the form

$$f_{x,t}(z_1,\ z_2,\ \ldots) = \mu(x,t) + \{1 - \lambda(x,t) - \mu(x,t)\}z_{x+1} + \lambda(x,t)z_1 z_{x+1},$$

if the unit of time is very small. If the unit for measuring age is also very small, then u_{x+1} is approximately $u_x + \frac{\partial u_x}{\partial x}$, and we obtain from (4) the apparently new type † of partial differential equation:

$$\frac{\partial u}{\partial t} = \mu + \lambda u_1 u - (\lambda + \mu)u + (1 - \lambda - \mu + \lambda u_1)\frac{\partial u}{\partial x}, \quad . \quad . \quad . \quad (5)$$

where λ and μ are known functions of x and t, $u = u_x = F_{x,t}(z_1, z_2, \ldots)$ and $u_1 = F_{1,t}(z_1, z_2, \ldots)$. If λ and μ are independent of x, then so also is u, and (5) reduces to

$$\frac{du}{dt} = (\lambda u - \mu)\,(u - 1) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This is the auxiliary equation for Mr. Kendall's equation (20). (It is of Bernoulli's form with dependent variable u–1.)

Similarly the cosmic ray problem, as formulated for example in Professor Bartlett's paper, section III(4), can be reduced to the solution of the two ordinary differential equations,

$$\frac{du}{dt} = \lambda_1\,(v^2 - u) + \mu_1\,(1 - u),$$

$$\frac{dv}{dt} = \lambda_2\,(uv - v) + \mu_2\,(1 - v),$$

* *Cf.* the reference to M. S. Bartlett (1947) in Mr. Kendall's paper, section 3 (i).
† The dependent variable occurs in two forms, u and u_1.

where λ_1, etc., are functions of t. This is a new method for obtaining the auxiliary equations for equation (32) of Professor Bartlett's paper.

Mr. B. J. PRENDIVILLE: My comments relate in general to discrete-valued Markoff processes for continuous time in which there are only a finite number of possible states or values of the system (these will, for convenience, be termed *finite* processes), and in particular to those finite processes where the transitions occurring are from a given state to one of its immediate neighbours in some scheme of classification of states.

Besides the one-variable processes which obviously fall into this particular category there are other processes apparently with more than one variable which by some re-arrangement can also be considered. For example, we have the atomic collision problem in J. E. Moyal's paper, in which a "state," for the special case of just four energy-levels, is a particular set of values of $(n_r, n_s; n_u, n_v)$. Since the total number of particles in this system is constant this state can be relabelled to read $(n_r, n_r + d_1; n - n_r, n - n_r + d_2)$, where $n_r + n_s + n_u + n_v = 2n + d_1 + d_2$ and $n_s - n_r = d_1$, $n_v - n_u = d_2$. This reduces the problem to the consideration of the random behaviour of just one variable, n_r. A similar reduction is available in the one-term radiation problem and in other processes of a like nature.

Two methods are in general use for tackling Markoff processes for continuous time—the generating function technique resulting in a partial differential equation for the probability generating function; and the algebraic approach in which a vector $\bar{p}$ representing the probabilities of the various states appears in an equation of the form $d\bar{p}/dt = A\bar{p}$, where A is the matrix of the transition probabilities. In many examples the generating function technique has been used to provide a simple direct solution, but in other cases the solution by this procedure cannot be completed without a knowledge of the eigenvalues of the matrix A, and it may be noted that for all finite processes a knowledge of these eigenvalues determines whether or not an equilibrium solution exists.

For those finite processes in which transitions take place only between neighbouring states I have shown that apart from one zero eigenvalue the others are all real and negative. This implies the existence of an equilibrium solution independent of the time and the initial conditions, and therefore justifies the assumption tc that effect in some of the problems considered by D. G. Kendall and J. E. Moyal.

Similar results can be obtained for processes with more general transitions, and for those with more than one variable which are not included in the above theory, provided always that the number of possible states is finite. This extension includes the conservative processes discussed by Professor Bartlett, and also the alternative specification of the system, mentioned in the footnote to that section of his paper which is needed when births are considered. The Collision and Radiation problems for a finite number of energy-levels are also included. In this wider class of processes there is at least one zero eigenvalue (the possibility of a multiple zero root, although it appears unlikely in most problems, cannot in general be disregarded); from some results of A. Brauer it is possible to show that the other eigenvalues all have a negative real part. If the zero eigenvalue can be shown to be a simple root the process tends to a limiting equilibrium distribution independent of the initial conditions, but if it is a multiple root the equilibrium distribution is not necessarily unique, in the sense that there will be some dependence on the initial conditions. I am indebted to D. G. Kendall for his referring me to this work of Brauer and for his useful comments on it.

As an example, it has been possible to solve by both approaches a birth and death problem of non-transient type similar to D. G. Kendall's "logistic" in that the population size is confined between two fixed levels. Instead of assuming the birth and death rates to be $\alpha(N_2 - N)$, $\beta(N_1 - N)$, $(N_1 < N < N_2)$, we may take them as $\alpha\left(\frac{N_2}{N} - 1\right)$, $\beta\left(1 - \frac{N_1}{N}\right)$ respectively. This reduces the problem to linear form and so makes it easier to handle. Until more is known of the estimation problems associated with these processes it is hardly possible to say which assumption is more reasonable, but in the limit as $N_2 \to \infty$ D. G. Kendall's form probably gives a more realistic model and this alternative is mentioned here primarily to illustrate the general principles. It is hoped to publish a more detailed account of these results in due course.

Mr. P. ARMITAGE: I have recently become interested in the application of Markoff processes to bacterial mutation (which is one of the problems mentioned by Mr. Kendall at the end of his paper, and about which Professor Coulson has been speaking this evening). Some of my results have probably been anticipated by Professor Coulson, but there are, I believe, one or two differences between our approaches.

I should like first, if I may, to refer to the description of ordinary bacterial growth by means of Markoff processes. Mr. Kendall has pointed out that the simple discontinuous birth process introduced by Furry in 1937 gives a distribution of generation times which is of negative exponential form, and he has proposed an ingenious method of keeping the Markoff hypothesis and yet obtaining a distribution of generation times which approximates closely to that observed in practice. What has worried me about this latter approach is that the assumption is still retained that the distribution of generation time (or lifetime) of an individual organism is independent of the ancestral history of the organism, whereas, according to bacteriologists with whom I have discussed this point, the lifetime of an individual is negatively correlated with the lifetime of its parent. That is to say, if any organism takes rather longer than usual to divide, its two offspring will tend to divide after rather shorter lifetimes than usual. The effect of this is, of course, to make the whole process more nearly deterministic than Mr. Kendall's model would allow, and it may not be worth while introducing a stochastic model at all. It would be useful to have some sort of experimental evidence as to the variance of the population size at time t. I should be very interested to hear what Mr. Kendall thinks about this point.

Mr. Kendall has mentioned on p. 262 the paper by Luria and Delbrück on bacterial mutation. I think it would be useful to describe briefly their model (I have not retained their original nomenclature). They assume that two strains of organism X and Y each grow deterministically and exponentially at a rate $(a + g)$ from initial sizes x_0 and y_0. Each X organism has, during an interval dt, a constant chance gdt of mutating to a Y organism.

Under the assumption that g is small in comparison with a, Luria and Delbrück obtain by a fairly simple argument the mean and variance of the size of population Y at time t. By a generalization of their method, it can be shown that the r^{th} cumulant of the distribution $(r > 1)$ is

$$\kappa_r = \frac{x_0 g}{(r-1)a}\{e^{rat} - e^{at}\}.$$

When g is not small in comparison with a the method breaks down, and we have to use the standard methods for Markoff processes. The interesting thing is that this is a "mixed" process in the sense that the population sizes x and y vary both continuously (by exponential growth) and discontinuously (by mutation). (I have followed Feller in using the term "mixed" in this connection. It is a different use from that of Professor Bartlett.)

The difference equation for the probability density function $f(y,t)$ may now be written down. (There is no need to introduce x, since $x + y$ grows deterministically.) It differs from those found in purely discontinuous processes in that the expression for $\frac{\partial f(y,t)}{\partial t}$ has a term in $y\,\frac{\partial f(y,t)}{\partial y}$.

We can define the probability generating function $\varphi(z,t) = \int f(y,t) z^y dy$, and, following Professor Bartlett, write down the differential equation satisfied by φ. This now includes a term $z \log z\,\frac{\partial \varphi}{\partial z}$. I have not solved the equation explicitly, but as usual the early cumulants may be obtained easily. As we should expect, they agree with the formula above to order g/a.

I have mentioned this model because if we are willing to assume deterministic growth when no mutation takes place, then when there is mutation we get these mixed processes.

The method can, of course, be extended to cover the case when the growth rates are unequal, and when there is back-mutation, *i.e.* mutations occur not only from X to Y, but also from Y to X. In these cases the process must be defined in terms of both variables, x and y. The most important outstanding problem is, I think, that of finding the best method of estimating the two growth rates and the two mutation rates, and I look forward to reading the contributions of Professor Coulson and Mr. Kendall to this problem of estimation.

Mr. D. V. Lindley: I should like to thank Mr. Moyal for his extremely interesting paper, the mathematical sections of which I found most stimulating. I am still not clear, however, what a random function is, and doubt whether any completely satisfactory definition is available. So long as this is true we have to be careful in discussing certain probabilities. For example, one definition (Kolmogoroff, 1933) of a random function requires a multivariate distribution to be given at any n points of time. But Kolmogoroff in giving this definition points out that this is not final because it is not, in general, possible to discuss the probability of the random function being less than some constant in a given time interval. This is just the probability that Professor

Bartlett uses in his paper. Later work does not clear up this point except in certain special cases, and therefore as far as I can see the general problem is still unsolved. If no definition is available, then we cannot be clear either about the notions we are discussing or whether the operations we performed are, in fact, valid. On p. 212 Professor Bartlett had said the "mathematical complications . . . can be overcome by suitable definitions." What were these definitions?

[At this stage Professor Bartlett and Mr. Moyal drew Mr. Lindley's attention to the discussion of this point in Mr. Moyal's paper, the relevant part of which Mr. Moyal then summarized. Mr. Lindley subsequently continued:]

Since Mr. Moyal's remarks during my contribution to the discussion merely repeat what he had already said in the paper I feel we are no further forward. The essence of the difficulty is this: Mr. Moyal says in §3.3, "definitions (of continuous-time processes) . . . have not led so far to any useful applications." But is not Professor Bartlett's paper very useful, and would he not, if asked what the expressions he uses and operations he performs mean, have to use such definitions? These difficulties are stated, far better than I can state them, in a summary by Cramer (*Ann. math. Stat.*, **18**, 165–193 (1947)).

Professor G. A. BARNARD: I must admit I share Mr. Lindley's doubts about the foundations of the general theory of continuous stochastic processes, but I do not think such doubts should be taken too seriously. It may well turn out that a general theory of arbitrary stochastic processes is no more possible than a general theory of entirely arbitrary functions, while none the less the applications of particular types of stochastic process, discussed in the papers we have heard, are entirely valid.

I wonder whether an approach to the theory of random variables which I have found useful might be extended to the case of stochastic processes. Whereas Kolmogoroff, in his classical work, starts off with a space on which a non-negative measure is defined, and then defines a random variable as a function defined on such a space, I prefer to define a random variable simply as a function $X(t)$, defined over some set, the range of t. I then introduce a group of mappings of this set into itself. Where the range of t is a finite number n of objects, the group of mappings will be the symmetric group S_n of all permutations of the objects among themselves; in some continuous cases the group will be the group of measure-preserving transformations. If $\mu(t) = t'$ is such a mapping, I then say that the function $X(t)$ is statistically equivalent to the function $X(\mu(t))$. The statistical properties of a function are then those, roughly speaking, which it shares with all the other functions with which it is statistically equivalent.

From this point of view we would be led to define a stochastic process as a function of two variables, $X(s,t)$. We then introduce a group of transformations μ of the range of s into itself, leaving the range of t unaltered, and consider $X(s,t)$ statistically equivalent to $X(\mu(s), t)$. Considered from the point of view of the statistical properties so defined, the function of two variables would then be a stochastic process with t as the parameter. It is easy to deal with discrete stochastic processes in this way, but I do not yet see how to define functions $X(s,t)$ which will correspond to non-trivial continuous processes.

Dr. M. G. KENDALL: I share some of Mr. Lindley's difficulties about the continuous random process. In most economic work associated with time series it is doubtful whether any process is continuous if the interval is small enough, although in the Stock Market the interval has to be a matter of hours before discontinuity is important. There are, however, cases in which a continuous record is available, *e.g.* in thermometric or barometric graphs; but even here it may be found that for very small time intervals there are essential discontinuities.

One possible approach is to assume that all natural phenomena are ultimately discontinuous, but that a mathematical account can nevertheless be given of them in terms of the "continuous stochastic process." Another is to observe that although there may be continuity in the position of particles for example, there may be discontinuities in their velocities or accelerations. Neither of these entirely removes my difficulty in picturing a continuous random process in the ordinary sense, and I feel that there are certainly difficulties of a nomenclature and possibly difficulties of logic to be cleared up here.

Professor TINTNER: Dr. Kendall has mentioned the problem of applications to economics. Stochastic difference equations have been used in this connection, but never stochastic differential equations. I believe that these latter are more appropriate in dealing with these problems.

My paper on the "simple" theory of business fluctuations, published in *Econometrica*,* 1942, may give a starting point. If one starts from a general Walrasian equilibrium in an economy with

* Tintner, G., "A 'simple' theory of business fluctuations," *Econometrica*, **10** (1942), 317.

n commodities and services, it is necessary to make simplifying assumptions in order to have a manageable mathematical problem. Let us assume that demand and supply functions are linear; further, that they depend upon expected rather than upon prevailing prices. Assume further that expected prices are linear functions of the existing prices and the time derivatives of these prices. Let p be the set of prices, a vector with n components. Denote by dp/dt the vector of the time derivatives of the prices. Let A and B be two matrices with n columns and rows. Then we have a system of linear differential equations with constant coefficients:

$$A.p + B.(dp/dt) = 0.$$

The nature of the solutions depends upon the roots of the determinantal equation:

$$| A + \lambda B | = 0.$$

It is interesting to note that both the exponential trend and also cyclical fluctuations may arise from the system. A tentative verification was presented in 1944 with the use of American data.*

To make the system stochastic, we may introduce a random vector E with n components, which appears on the right-hand side of the above system. But actually we also have to introduce errors of observations in the prices, which we denote by e. Let us assume that we observe, not the theoretical prices p, but the prices affected with errors of observation, $P = p + e$. These are not observed continuously, but only on a set of points in time $t_1, t_2 \ldots t_n$. Having this information, what conclusions can be drawn about the structure of the system? This is evidently a very difficult estimation problem, especially since the samples in question may not be large.

Mr. QUENOUILLE: I wonder, in view of the discussion having moved on to the applications of a stochastic theory, whether Professor Bartlett can answer two questions on practical aspects. In a statistical appendix to a paper by Professor Blackman published about ten years ago the author considered the distribution of plant density, but Professor Blackman on that occasion referred to the Poisson distribution of plants, and using this distribution he derived a relation for plant density. This was proportional to, or, in fact, for a unit quadrat, equal to minus the logarithm of the number of empty quadrats, *i.e.* $- \log q$. The same proportionality held whether one introduced the Poisson theory or not. Professor Blackman gave diagrams for the frequencies of densities of plants, which showed in certain cases that the Poisson distribution did not hold. These frequencies seem to agree with some of the theories put forward at this meeting; the negative binomial distribution appears to be a reasonable fit, and it would seem possible to build a somewhat wider theory to incorporate an index of dispersion based on the negative binomial distribution. Has Professor Bartlett in any way come into contact with that problem?

In the second place the negative binomial seems to have cropped up quite a number of times in the course of the papers, as has the Poisson distribution and the binomial. It seems that a stage is being reached at which there are a number of models and a number of explanations, but not sufficient data. The same seems true in regard to the use of stochastic processes in economic data in which it is possible to explain economic observations by two or three possible models.

I now take as an example some data provided by Dr. M. G. Kendall on sheep populations:

Observed and Theoretical Serial Correlations of Kendall's Sheep-Population Series

Serial correlation	1	2	3	4	5	6	7	8
Observed	0·60	− 0·15	− 0·60	− 0·54	− 0·14	0·14	0·20	0·12
Kendall's theoretical	(0·60)	(− 0·15)	− 0·62	− 0·54	− 0·09	0·33	0·42	0·19
Moving average theoretical	0·62	− 0·13	− 0·58	− 0·52	− 0·20	0·04	0·12	0·09

The first two rows gave the observed serial correlations and theoretical values calculated from a second order autoregressive scheme. The third row gives the theoretical serial correlations for a simple moving average of the type $u_n + 2u_{n-1} + 2u_{n-2} + u_{n-3}$ when the effect of trend elimination has been taken into account. Evidently, either method presents a reasonable fit.

There are any number of ways of satisfying any particular stochastic process, and it seems, in some ways, a little dangerous to adhere to one. I say this in order to emphasize what I believe to be a bad omission, not on the part of the authors of the three papers, but in the general development of the subject, namely, the omission of tests to determine whether any particular set of stochastic observations could be reasonably represented by any particular theory. In many fields of application there do seem to be, at the moment, more theories than observations.

* Tintner, G., "The 'simple' theory of business fluctuations: a tentative verification," *Review of Economic Statistics*," **26** (1944), 148.

The following written contribution was received after the meeting:

Dr. K. J. Le Couteur: There are many physical experiments, such as nuclear disintegration, in which a single primary particle gives rise to several secondary particles by a process which is itself unobservable. A possible, much simplified, mathematical model is a population in which each member has probabilities $\lambda(t)dt$ of giving birth $(\mu - \gamma)\,(t)dt$ of death and $\gamma(t)dt$ of emigration in the time interval t to $t + dt$. Only the emigrants are supposed observable, and one has to calculate the total emigration out of a population generated from a single ancestor. The behaviour of the population itself is governed by Kendall's (1948*a*) theory with death-rate μ; in the physical problem it is always ultimately extinguished.

If $p(n,r,t)$ denotes the probability that at time t the population contains n particles and that a total of r have emigrated, the differential equation for the generating function,

$$\varphi(z,s,t) = \Sigma p(n,r,t) z^n s^r, \qquad (1)$$

is

$$\frac{\partial\varphi}{\partial t} = \{\lambda z^2 - (\lambda + \mu)z + \mu - \gamma + \gamma s\}\frac{\partial\varphi}{\partial z} \qquad (2)$$

with a boundary condition $\varphi = z$ at $t = 0$.

For constant s this is a Riccati equation. The solution is of the form

$$\varphi(z,s,t) = e^{\zeta}\,\frac{\xi + (1 - \xi - \eta)z}{1 - \eta z} \qquad (3)$$

where ξ, η. ζ are functions of s and t. The conditions on ξ, η, ζ are found to be ($\zeta' = \frac{\partial\zeta}{\partial t}$, etc.),

$$\zeta'(1-\eta)^2 = \gamma(s-1)\,(1-\xi)\,(1-\eta)$$
$$-\zeta'\eta(1-\xi-\eta) + (\eta\xi' - \xi\eta') + \eta' = \lambda(1-\xi)\,(1-\eta).$$
$$\zeta'\xi + \xi' = (\mu - \gamma + \gamma s)\,(1-\xi)\,(1-\eta). \qquad (4)$$

With $s = 1$ and $\zeta' = 0$, these equations become identical with Kendall's (1948*a*). The further substitution

$$U = 1 - \xi \qquad V = 1 - \eta$$

leads to

$$\zeta' = \gamma(s-1)\frac{U}{V} \qquad (5a)$$

$$V' = -\gamma(s-1) + (\mu - \lambda + 2\gamma(s-1))V - (\mu - \gamma + \gamma s)V^2 \qquad (5b)$$

$$\left(\frac{U}{V}\right)' = \frac{U}{V}\left\{\lambda - \mu - 2\gamma(s-1) + 2\gamma\frac{s-1}{V}\right\} - \gamma(s-1)\left(\frac{U}{V}\right)^2. \qquad (5c)$$

Equations 5 can be solved completely if λ,μ,γ are independent of t, though in this simple case it is easier to proceed directly. In the general case, if a solution of the Riccati equation (5b) is known, the remaining equations can be integrated by quadratures.

Define

$$\delta(s,t) = \mu - \lambda + 2(s-1)\gamma - 2(s-1)\gamma/\nu, \qquad (6)$$

then the solution of (5c) is

$$\frac{V}{U} = e^{P}\{1 + \int_0^t e^{-P(\tau)}(s-1)\gamma d\tau\} \qquad (7)$$

with

$$P(s,t) = \int_0^t \delta(s,\tau)d\tau \qquad (8)$$

Then (5a) gives

$$\zeta(t) = \int_0 \gamma(s-1)\left(\frac{U}{V}\right)d\tau = \int_0^t\left\{-\delta - \left(\frac{U}{V}\right)'\Big/\left(\frac{U}{V}\right)\right\}d\tau$$

so that

$$e^{\zeta(s,t)} = \frac{V}{U}e^{-P} = 1 + \int_0^t e^{-P(\tau)}(s-1)\gamma d\tau. \qquad (9)$$

According to the definition (3), $e^{\zeta(s,t)}$ is the p.g.f. for the emigration up to time t. In the physical problem mentioned above, only the total emigration after an infinite time is observable. This determines only $\zeta(s,\infty)$, and so yields only very limited information about the basic probabilities λ,μ,γ.

It is easy to give explicit expressions for the first and second factorial moments of the number of emigrants. With the notation $P(1,t) = \rho(t)$, they are

$$\mu_1(T) = \int_0^T e^{-\varrho(t)}\gamma(t)dt, \qquad (10)$$

$$\mu_2(T) = -\int_0^T dt\, e^{-\varrho(t)}2\gamma\left(\frac{\partial P}{\partial s}\right)_{s=1} = \int_0^T dt\, e^{-\varrho(t)}2\gamma \int_0^t 2\gamma\frac{1-V}{V}\, d\tau. \qquad (11)$$

The quantities which appear in (10) and (11) must be evaluated with $s = 1$ and so may be taken directly from Kendall's (1948a) paper.* Thus μ_1 and μ_2 become

$$\mu_1(T) = \int_0^T \bar{n}_t\gamma(t)dt, \qquad (12$$

which is obvious, and

$$\mu_2(T) = \int_0^T dt\; 2\gamma e^{-\varrho(t)} \int_0^t 2\gamma(W-1)d\tau. \qquad (13)$$

The function $\delta(s, t)$ which determines ζ through (8) and (9) can be calculated directly from an equation equivalent to (5*b*),

$$2\delta' = (\lambda - \mu)^2 - 4\lambda\gamma\,(s - 1) - \delta^2 \qquad (14)$$

with initial conditions $\delta = \mu - \lambda$ at $t = 0$.

In the special case of constant λ,μ,γ the solution, due to J. E. Moyal, is

$$\varphi(s,t=\infty) = \frac{\mu+\lambda}{2\lambda} - \left\{\left(\frac{\mu+\lambda}{2\lambda}\right)^2 - \frac{\mu-\gamma+\gamma s}{\lambda}\right\}^{\frac{1}{2}}.$$

The CHAIRMAN then invited the authors to reply to the discussion.

Mr. J. E. MOYAL: In view of the lateness of the hour I will defer my comments. I should like, however, to thank the various speakers for their interesting and illuminating discussion.

Mr. MOYAL subsequently replied as follows:

Professor Bartlett, in a footnote to his remarks, states that the "forward" and "reverse" transition probabilities for a Markoff process bear no simple relationship to each other, unless the process is either stationary or deterministic. I would substitute the word "unitary" for deterministic, meaning by this term that the temporal transformation generated by the transition probabilities is unitary. Let us write $p(x, t \mid x_0, t_0)$ for the "forward" transition probability, $q(x_0, t_0 \mid x, t)$ for the "reverse" $(t > t_0)$. The condition for the transformation to be unitary is easily seen to be that the two should be equal:

$$q(x_0, t_0 \mid x, t) = p(x, t \mid x_0, t_0).$$

p must then fulfil the two relations

$$\int dx_0\, p(x, t \mid x_0, t_0)\, p(x', t \mid x_0, t_0)\, dx_0 = \delta(x - x')$$

$$\int dx_0\, p(x, t \mid x_0, t_0)\, p(x, t \mid x'_0, t_0)\, dx = \delta(x_0 - x'_0).$$

For a discrete and finite process it is well known that these conditions imply determinism, but it is not clear that they do so in general: i.e. for processes with denumerably or continuously infinite sets of possible values of the variates. What is clear from the above integral relations is that there can be no "diffusion," no mixing of "probability fluid" elements. If the system is *not* deterministic, then at least if the probabilities starting from two distinct points x_0, x'_0 at t_0 spread to intervals (or sets of intervals) Δx, $\Delta x'$ at t, the latter must *not overlap*. I may mention that I raised this question of the meaning of a unitary Markoff process in my paper on "Quantum mechanics as a statistical theory" (reference (2) in my present paper).

I was much impressed by Dr. Arley's ingenious method of solving the "birth-and-death" process. To answer his query regarding the general conditions under which the phase-space

* Kendall, D. G. (1948*a*), *Ann. Math. Stat.*, **19**, 1–15.

distributions of statistical mechanics tend to equilibrium is a "tall order." What he asks for substantially are the general conditions of validity of the H-theorem in statistical mechanics, and this, in spite of its respectable age, is still an open and controversial question; to answer it adequately would involve writing another whole paper. I shall limit myself therefore to a few, not very precise, comments. Briefly, I believe the H-theorem valid for the distribution of any given mechanical system *only* if one introduces some source of "randomness." Mathematically, this may be expressed as follows: if a completely isolated system satisfies Liouville's theorem, $df/dt = 0$, one can easily verify that the entropy of the system, defined by $S = -k\int f \log f \, d\tau$, is constant ($dS/dt = 0$), and hence there is *no* tendency to equilibrium. Completely isolated systems do not, however, exist in nature; there is always a certain degree of uncontrolled interaction between the system and the outside world, which introduces a more-or-less random disturbance of the system. It is this random disturbance which gives rise to the tendency to equilibrium. Mathematically, we may say that the distribution function $f(p_i, q_i, t)$ will tend to an equilibrium value independent of t only when

$$\frac{df}{dt} = Sf,$$

where S is some "statistical operator" operating on f, which represents some source of randomness acting on the system. This may arise in several ways:

(*a*) f may be the distribution of a part of a larger system. Boltzmann's famous "proof" of the H-theorem refers in reality to the distribution and entropy of a single molecule in a gas. The "random" element is introduced by the interaction of this molecule with the rest of the gas. When Boltzmann extends his "proof" to the whole gas, he merely adds the entropies of the individual molecules, which is equivalent to neglecting their dependence in the expression of their joint distribution function, and considering the interactions which give rise to this dependence as the "random" disturbance.

One has a similar situation for a system in contact with a "heat-bath," which must lead to the canonical ensemble as the equilibrium distribution for the system, just as the interactions lead to the Maxwell-Boltzmann distribution for the individual molecules of a classical gas.

(*b*) The H-theorem will be satisfied by the distribution functions of a system of Brownian particles, where, of course, the "random" impulses are provided by the thermal motion of the molecules, and the "statistical operator" S is a "diffusion" operator of the form $\Sigma \nabla_u^2 f$.

(*c*) Several authors arrive at a general H-theorem for a distribution, say F, which is obtained from the phase-space distribution f of the system by averaging f over small but finite elements of the phase-space (*cf.* Tolman, ref. (55) in my paper). Here again the random element is introduced through the ignored motions inside these elements.

(*d*) There is an equivalent procedure in the quantum theory of statistical assemblies, which consists in averaging over the phases of the wave-functions (*cf.* Tolman, *loc. cit.*), leading to what is called a mixed or impure state. This has the effect of turning the process into a Markoff one, which should lead to an H-theorem for the distributions under fairly wide conditions. This question is discussed by Mr. Prendiville in his contribution.

With regard to Mr. Lindley's remarks, I find it difficult to add anything to what I have already said in the body of my paper, where I specifically dealt with the points he raises in §3.1, §3.2, and more particularly in §3.3.

Professor Barnard's remarks on the definition of a stochastic process sounded most intriguing, but I must confess that I was unable to grasp his meaning fully. I should be glad to see a more detailed exposition of his ideas.

I was much interested in Professor M. G. Kendall's remarks regarding the continuity of stochastic processes. There is certainly no difficulty in building up mathematical models of processes which are continuous—or in fact differentiable to any order and even analytic. There is also no difficulty in building models which are not continuous, or which are continuous but not differentiable, etc. I feel that there is an element of choice, or rather of convenience, in which model we choose to fit the observations. In fact, I discussed this very point in my paper (§7.1) in connection with the theory of the Brownian motion. I pointed out that the choice there boils down to a question of *scale*: one might regard the velocities of the Brownian particle as differentiable on the molecular scale, but it is actually more convenient mathematically to regard them as non-differentiable at the (much coarser) scale at which we measure these velocities.

Mr. Quenouille's remarks on the necessity of tests in order to determine the most suitable stochastic process model to fit observations do not as a rule apply to physical problems, because in these problems one has usually independent reasons for putting forward a particular theoretical model, and then proceeds to make observations in order to test this particular theory. The problem of estimation is usually quite a separate one from that of solving the equations of the theory, because it depends on the experimental conditions. One may, in fact, have quite different sets of experiments to test the same theory, each of which may raise different estimation problems.

Professor M. S. BARTLETT: I would also like to defer my reply apart from expressing my gratitude to the various speakers for their comments. There is, however, one note of explanation I wish to make in view of Dr. Arley's reference to the use of the moment-generating function in statistical mechanics. Mr. Moyal's suggestion that its introduction is due to me is, of course, exaggerated. For a variable taking only integral values it is equivalent to the probability-generating function, which was already well known in statistical mechanics under the name of the partition function. Nevertheless, the physicists had not, at the time I wrote the note to which Mr. Moyal has referred, always realized this equivalence. In particular, Fowler's own discussion of fluctuations in his treatise on statistical mechanics is unnecessarily cumbersome, and it was, in fact, Fowler himself who suggested the publication of my note.

Professor BARTLETT subsequently replied as follows:

With regard to the further point raised by Dr. Arley of the use of the characteristic function in quantum mechanics, I imagine that this has by now been considered independently by several writers. An elementary discussion was, for example, included in the unpublished pre-war manuscript of mine to which the Chairman referred, any random variable whose expectation provides a characteristic function being represented by a unitary operator. The more interesting problem recently discussed by Moyal and others is the form of the operator corresponding to the joint characteristic function of two non-commuting variables.

On re-reading my manuscript, I found one of the concluding paragraphs so relevant to some of the statistical mechanics developments in Mr. Moyal's present paper* that I may, perhaps, be allowed to quote it:

". . . the quantum-statistical mechanics of assemblies in equilibrium has by now established itself in a very strong position. Perhaps the chief theoretical difficulty outstanding is the absence of a proper conception of equilibrium in the present technique. The concept of equilibrium must involve the notion of duration, but attempts to demonstrate that the equilibrium state is a limiting stable state independent of initial conditions are rather unsatisfactory, and do not fall in with the simple formalism of the equilibrium theory. The success of the method as an abstract technique independent of particular theories of transition and interaction mechanisms is very striking, but I believe that a complete theory must involve the latter, though they must first be transformed into an abstract technique capable of fitting in with the corresponding partition function technique of equilibrium states."

I was especially glad to hear the Chairman's tribute to Mr. Moyal, who has done so much, not least by his own work, to impress on us in this country the importance of the theory of stochastic processes.

I still think, like Mr. Moyal, that Mr. Lindley's queries have been already answered quite generally in Moyal's paper. Mr. Lindley should, however, also read, when it appears, the unpublished work by Professor Pitt which Mr. Moyal has mentioned (due to appear in *Proc. Camb. Phil. Soc.*). I will merely add a comment on the particular application in my own paper, to which Mr. Lindley has referred, concerned with the probability of an additive process $X(t)$ being contained within some boundary for all t in an interval. A special definition sufficient to cover this type of problem was introduced by Khintchine some years ago, the probability being defined as the upper bound of the probability of the boundary being exceeded at least once at an arbitrary set of points $t_1, \ldots, t_n$ (*cf.* H. Cramer, "Deux conférences sur la théorie des probabilités," *Skand. Aktuarietidskr.* (1941), p. 34, especially p. 64, or C.-O. Segerdahl, section 5, reference (10) of my paper, the latter discussing the equivalence of this definition to the simpler one arising from the choice of *equidistant* points t suggested briefly in my own paper).

While thanking Mr. D. G. Kendall for his two remarks on my paper, I have at present no direct comments to add. In connection, however, with equation (27), I will note further that it is possible to obtain the limiting distribution of $Y = Xe^{-kt}$, where X is the untransformed variable and k is the rate of increase of its mean (*cf.* the discussion for discrete time by T. E. Harris in his *Ann. Math. Stat.*, **19** (1948), p. 474, paper on "Branching Processes"). The moment-generating function for Y ultimately satisfies the equation (for one initial individual):

$$k\theta \frac{\partial M}{\partial \theta} = h(\theta),$$

with solution

$$M = \xi^{-1}\left(\frac{1}{k}\log\theta + \text{constant}\right),$$

* *Cf.* also his written reply to Dr. Arley's query on the conditions under which the distributions of statistical mechanics tend to equilibrium.

where the constant is determined from the constant unit mean for Y. Thus in the elementary birth-and-death process with birth and death rates λ and μ respectively ($\lambda > \mu$), we obtain

$$M = \frac{\lambda - \mu(1 + \theta)}{\lambda(1 - \theta) - \mu},$$

representing the chance of extinction as a discrete component μ/λ at the origin, and a continuous component in the form of an exponential distribution.

Mr. Prendiville's finite processes include, as he notes, my conservative processes. My footnote on the inclusion of "births" thus becomes rather redundant, for other modifications such as "marriages" can also be included if we are prepared to interpret the state of the system in a wide enough sense. The system must then, however, be considered as one super-individual, the independence of the individuals in the conservative (or conservative plus births) process being lost. Thus, while the formalism of the generating function method would still apply, there seems nothing to be gained from it in the general case, and the equivalent equation $dp/dt = Ap$ mentioned by Prendiville may be considered directly.

With regard to Professor Tintner's problem, there is no formal difficulty in solving the first stochastic system he proposes (*cf.* equation (14) of the last section of my North Carolina lecture notes, where, however, only stationary schemes were being considered). But, as he is probably aware, the introduction of further observational errors in all the variables is a very unpleasant complication to the theoretical problem of establishing structure.

I agree with Mr. Quenouille's suggestion that stochastic process theory should assist in the interpretation of ecological distributions. For a negative binomial distribution with probability-generating function $(1 + \sigma - \sigma z)^{-}$, the mean density is proportional to log-percentage absence if σ remains constant. The implications of this requirement in terms of any particular mechanism giving rise to the negative binomial remains to be studied.

Finally, it is relevant to my discussion of a rather simplified model of the cosmic ray cascade shower to call attention to an exact treatment of the second-order moment fluctuations, the energies of the particles being allowed to vary over a continuous range. This has been recently achieved by Bhabha and Alladi Ramakrishnan using a limiting procedure, and Ramakrishnan has further derived the same equations by developing more direct methods, this representing an independent development of the direct methods used by D. G. Kendall in his discussion of a population's age-structure, and recalling Moyal's suggestion at the Symposium for dealing with the cosmic ray problem.

Mr. David G. Kendall: While asking for permission to defer my reply, I should like to express my gratitude for the lively and stimulating discussion, and to say how pleasant it has been to watch the Fellows of the Society coming down firmly on the side of mathematical rigour.

Mr. Kendall subsequently replied as follows: Professor Bartlett and Mr. Moyal have already dealt with most of the topics raised in the discussion, and in adding a few comments to theirs I will try not unduly to lengthen an already long story.

I am most grateful to Dr. Arley for having come all the way from Denmark for the meeting, and I was much impressed by his simple and elegant method for solving the equations to the birth-and-death process when the rates vary with the time. I am sure Dr. le Couteur, whose communication is concerned with a generalization of the same problem, will want to consider whether Dr. Arley's device can be employed to simplify his analysis in a similar way.

I much admire Mr. Prendiville's method for constructing a birth-and-death process of bounded non-transient type for which the partial differential equation is linear and so easily solvable. His discussion of the existence and uniqueness of a limiting distribution helps to make much clearer some of the operations in my paper. Some questions of this sort are discussed very thoroughly by Fréchet in the second volume of his famous book on probability; I mention this as a reference which may assist Mr. Prendiville in the promised extension of his work.

Several speakers have referred to the age-distribution problem. Mr. Moyal's formulation (a good deal more general than anything I have considered) proposes the determination of the joint distribution of the numbers in several age-groups at several instants of time. As far as the restricted problem is concerned, in which one only considers the distributional properties of the age-structure at a single instant, the remarks made by Mr. Good in the second section of his contribution seem to me very important. Incidentally, I think there is a slight error in section 1 of his note: I believe the second formula should read

$$f_0(f_1(\ldots f_r(x)\ldots)).$$

As Professor Coulson remarks, the process investigated by the late Dr. Lea and himself is of the (λ, μ, κ) type, actually with $\mu = 0$, but their formulae cannot be deduced from my (48) and (49) because these last equations only apply when κ is constant, and in the problem referred to,

$$\kappa = \varepsilon\, e^{\lambda t},$$

where ε is the mutation rate per cell division, and the normal form is supposed to grow deterministically. Mr. Armitage's remarks also concern this topic. In assuming the stochastic independence of generation time in parent and progeny I was more or less following the conclusions of Kelly and Rahn (the only reference in this field known to me). They state that the rate of fission appears not in general to be an inherited character, and that the two cells resulting from one division may have widely differing generation times. They do, however, mention the negative correlation referred to by Mr. Armitage, in the sense that the division of a cell may be delayed for some time and then two divisions may follow rapidly. The impression I gained was that though the cell wall had not divided, the contents had already separated and had commenced to function as two units; it therefore seemed adequate to consider this as no more than an occasional masking of the normal behaviour. This was, of course, only an ill-informed guess on my part, and if there is any later experimental data bearing upon this point I should be glad to know about it. If it is correct to treat bacterial growth as an effectively deterministic process, then it should be possible to demonstrate this either macroscopically, as Mr. Armitage suggests, by examining the observed fluctuations in population size, or microscopically, by extending the detailed observations of the Kelly and Rahn type. A decision on this matter would seem to be an important preliminary to the determination of mutation rates.

Dr. Irwin, speaking from the Chair, has recalled some of the pioneer workers in this field, whose results have often been re-discovered and described in the new terminology. I have been very much struck, at all stages of the investigation, by the number of such surprises for a while concealed in the literature. I am sure there must be many more references to early work which I have omitted in ignorance; I regret none more than that to a paper of Mr. Yule* containing a derivation of what has since been called the Furry process. Professor Bartlett has mentioned Alladi Ramakrishnan's independent introduction of the cumulant-function technique. Of the other new work which has appeared since my paper was written the most important appears to be that of Richard Otter,† who has linked up the problems of the growth and extinction of populations with the combinatorial theory of "trees." This was being discussed‡ in the pages of the *Educational Times* only a few years after Galton there proposed his "surname" problem, and we may wonder equally at the versatility of the contributors to that remarkable periodical and at the near-century which has had to elapse for these two topics to become reunited.

The records of the Symposium would be incomplete without a closing tribute to the late Professor Major Greenwood. I am sure my fellow-speakers must have shared my great pleasure in his presence at the meeting and in his contribution to the discussion. His death represents a personal loss even to those of us who are newcomers to the Society; the departure of a hoped-for friend, the locking of the door to a storehouse of matchless counsel.

* *Phil. Trans. Roy. Soc.* (B), **213** (1925), 21–87.

† *Ann. Math. Stat.*, **20** (1949), 206–24.

‡ *Educational Times*, **30** (1878), 81–5. See also (for example) Cayley's 1875 *Report* to the British Association for the Advancement of Science.