DIFFUSION PROCESSES IN GENETICS

WILLIAM FELLER
PRINCETON UNIVERSITY

1. Introduction

The theory of evolution provides examples of stochastic processes which have not yet been treated systematically. We find here many open problems whose solution promises new insights into the general theory.

There exists a huge literature on the mathematical theory of evolution and statistical genetics, but existing methods and results are due almost entirely to R. A. Fisher and Sewall Wright.[1] They have attacked individual problems with great ingenuity and an admirable resourcefulness, and had in some instances to discover for themselves isolated facts of the general theory of stochastic processes. However, as is only natural with such pioneer work, it is not easy to penetrate to the mathematical core of the arguments or to discover the explicit and implicit assumptions underlying the theory. In the following an attempt is made to formulate the basic mathematical problems and to discuss their connection with other stochastic processes. Such a systematic approach leads automatically to more general formulations which may be useful at least for a better understanding of the underlying assumptions.

We are concerned with mathematical models of population growth. Relatively small populations require discrete models, but for large populations it is possible to apply a continuous approximation, and this leads to processes of the diffusion type.

By way of introduction we start with the simple *branching process* which became popular in connection with its application to nuclear chain reactions, but which had been previously used by Galton in a discussion of the survival of family names, and by R. A. Fisher in his treatment of the survival of mutant genes. This branching process describes the simplest possible populations: the individuals are of like kind, and there is absolutely no interaction among them so that they are statistically independent. Thus, contrary to a widespread belief, the branching process does not represent an isolated type of stochastic process and is remarkable mainly because of its simplicity. For example, in the case of a growing population the process in its later stages converges to a diffusion process regulated by the Fokker-Planck equation describing the simplest growth (compare appendix II, and end of section 5).

Serious difficulties arise if one wishes to construct population models with in-

This research was done under an Office of Naval Research contract at Cornell University for developing probability theory.

[1] See references [5], [12], and [13]. It is difficult to give useful references to original papers, since these are mostly highly technical and inaccessible to nonspecialists.

teractions among the individuals.[2] The situation grows worse if the population consists of different types of individuals. Explicit examples of such populations with various types of interactions between individuals are provided by various breeding systems. Their treatment leads to Markov chains which are in general finite. [4, chaps. 15, 16].

The populations with which we would like to deal in genetics are far too complicated for a mathematical treatment, and it proved necessary to introduce artificial simplifications. Thus, almost the entire theory is devoted to the case where the population consists of only two types of individuals, called a- and A-genes. Each individual may have a certain number of descendants of either kind. We have here a bivariate branching process with the added difficulty that the individuals are not statistically independent. A mathematical theory of such bivariate branching processes is extremely desirable, but even a beginning is hard to get. For example, there are three possibilities of extinction (only a-genes, only A-genes, or both). If the elementary transition probabilities of the process are known, the extinction probabilities can be calculated from an infinite system of linear equations, but these in general do not have unique solutions.

In fact, the bivariate branching process leads to such difficulties that apparently not one single truly bivariate case has been treated in the literature. In the theory of evolution this difficulty is overcome by the assumption of *a constant population size:* if there are j a-genes, the number of A-genes is $2N - j$, and thus it suffices to study the univariate population of a-genes. Presumably this is a justifiable approximation in specific applications, but it is of interest to note that essential features of the whole mathematical theory depend on this assumption. Dropping it will lead us to an entirely new theoretical model (the practical applicability of which is not discussed).

In section 3 we discuss a discrete model of a population of fixed size of a- and A-genes. The probability relations which connect the numbers of a-genes in two succeeding generations reflect random mating and take into account mutation pressure and selective advantages in a way which seems to correspond to the basic assumptions of the accepted theories.[3]

In section 10 the assumption of constant population size is dropped and a truly bivariate model is constructed which takes into account selective advantages in a more flexible way.

It is known[4] that an essential part of Wright's theory is mathematically equiva-

[2] A special problem arises in connection with differentiation according to age. If there are m possible age groups then the population may be classified into m classes and the process becomes an m-variate process of the Markov type to which the remarks apply. This is an artificial device and often it is preferable to insist in discussing only the total population size. However, in this case the past history has after effects, and the process is univariate but not of the Markov type. In practice, this is a compelling reason to neglect effects of age differences, and we shall adhere to this tacit convention throughout the paper.

[3] However, our formalization leads to a 'sampling variance' which is somewhat larger than the usually accepted value.

[4] Apparently [14] this was first observed by Kolmogorov who had previously treated a particular problem of genetics as a diffusion problem. However, as early as 1922 R. A. Fisher was led to an equation of the heat conduction type, and this was before Kolmogorov's famous paper [7] on stochastic processes. For Fisher's method and references to the original papers compare [5, p. 88]. A variation of Wright's arguments which leads directly to the diffusion equation was given by Malécot [11].

lent to assuming a certain diffusion equation for the *gene frequency* (that is, the proportion of a-genes). This equation is discussed in section 7 and obtained in section 8 from the model of section 2 by a limiting process which is actually implicitly implied in Wright's theory. Now this diffusion equation is of a peculiar type and it should be realized that the limiting process in question is but one in a family of possible processes (depending on the ratios of mutation rates to population size).

In general (section 9) we obtain in the limit *ordinary* diffusion equations with drift, and thus *normal distributions* for the deviation of the gene frequency from the equilibrium point. Presumably this corresponds to actual phenomena in nature. However this may be, it is of mathematical interest that the singular diffusion equation of Wright's theory appears as a degenerate case of ordinary diffusion.

This diffusion equation (as well as others occurring in population theory) are of a *singular type* and lead to new types of boundary conditions and mathematical problems which have not yet been investigated.* Of significance is the fact that (even in the case of the diffusion equation of Wright's theory) a *stable solution* $u(x)$ *can exist but nevertheless the actual solution may approach, not* $u(x)$, *but a certain fraction* $qu(x)$. The possibility of such an occurrence seems to have been completely overlooked in the literature (section 7).

The same limiting process which leads from the model of section 3 to the diffusion equation of Wright's theory can be applied to our new bivariate model and leads to a diffusion equation in two dimensions. It is very different from the univariate diffusion equation, but the latter is contained in it (section 10). The solutions of this new equation are not discussed, but it is to be observed that *in no truly bivariate case does the gene frequency satisfy a diffusion equation* (sections 6 and 10). In fact, if the population size is not constant, then the gene frequency is not a random variable of a Markov process. Thus, conceptually at least, the assumption of a constant population size plays a larger role than would appear on the surface.

2. The branching process and Markov chains

The classical branching process is too well known to require treatment here,[5] and we shall not use its special properties. However, we start from it to tie it in with our notations, and to discuss its relation with the diffusion equation of section 5.

In each succeeding generation the population of the classical branching process consists of like individuals, each of which may give birth to 0, 1, 2, . . . direct descendants (that is, elements of the succeeding generation). It is assumed that the corresponding probabilities $p_0, p_1, p_2, \ldots \left(\text{with } \sum p_k = 1\right)$ are the same for all individuals and, furthermore, that the individuals of any generations are statistically independent of each other. Denote the number of individuals in the n-th generation by $Z^{(n)}$. This is a random variable of a Markov chain, and the whole theory hinges on the *transition probabilities* p_{jk} (the conditional probability that $Z^{(n+1)} = k$ given that $Z^{(n)} = j$).

* *Added in proof.* A systematic theory, including the new boundary condition, is to appear in the *Annals of Math.*

[5] Compare, for example, [4, chapter 11].

An explicit expression for p_{jk} is cumbersome, but (as for many Markov chains) the use of generating functions makes the required calculations possible. Putting

$$f(x) = \sum_{k=0}^{\infty} p_k x^k , \tag{2.1}$$

the classical theory depends on the easily verified remark that (due to the assumption of statistical independence)

$$p_{jk} = \text{coefficient of } x^k \text{ in } f^j(x) . \tag{2.2}$$

From this it follows that if the original population size is r, then the generating function of the population size $Z^{(n)}$ in the n-th generation is given by $f_n^r(x)$ where

$$f_1(x) = f(x) , \qquad f_2(x) = f[f(x)] , \ldots , f_{n+1}(x) = f_n[f(x)] . \tag{2.3}$$

The main point of interest to us is the fact that the definition of the process permits us to calculate the transition probabilities p_{jk}; these, together with the initial population size, completely determine all probability relations of the process. In particular, the *n-step transition probabilities* $p_{jk}^{(n)}$ (giving the probability that if the population size $Z^{(m)}$ is at any time j, it will n generations later amount to k), can be calculated recursively from $p_{jk}^{(1)} = p_{jk}$,

$$p_{jk}^{(n+1)} = \sum_{\nu} p_{j\nu}^{(n)} p_{\nu k} . \tag{2.4}$$

As a further illustration consider the *extinction probability* ϵ_r that the population size will even be reduced to zero if the initial population size is r. This case can occur only if $p_0 > 0$. In the terminology of Markov chains the population size 0 is an 'absorbing state' (or 'trap'), which once it is reached becomes a permanent state. It follows then easily [4, p. 334] that the extinction probabilities ϵ_r are a solution of the infinite system of equations

$$\epsilon_r = \sum_{\nu=1}^{\infty} p_{r\nu}\epsilon_\nu + p_{r0} . \tag{2.5}$$

The fact that $p_{r\nu}$ is the ν-th coefficient in the power series for $f^r(x)$ suggests investigating the possibility of a solution of the form $\epsilon_r = \lambda^r$. With it the right hand member reduces to $f^r(\lambda)$, and hence (2.5) requires that $\lambda^r = f^r(\lambda)$, or

$$\lambda = f(\lambda) . \tag{2.6}$$

Conversely, if λ is a solution of (2.6), then $\epsilon_r = \lambda^r$ is a solution of (2.5). Now $\lambda = 1$ is always a root of (2.6), but if $f'(1) > 1$ there exists another root $\lambda < 1$. In this case, therefore, the solution of the system (2.5) is not unique. It follows from the general theory (or the recursion formula for the probability of extinction within n generations) that we are interested in the smallest solution of (2.5), and this determines our extinction probabilities uniquely. We have thus derived the well known basic theorem concerning branching processes from the general theory of Markov chains.[6]

[6] As a further example of the applicability of the general theory to the branching process let us note that all states except 0 are transient so that $p_{ik}^{(n)} \to 0$ as $n \to \infty$ ($k \neq 0$). A deeper theorem of

Leaving the branching process we arrive at more general population models by assuming appropriate transition probabilities p_{jk}. The simplest generalization consists in permitting the fertility to depend on the actual population size; in our notation this means that the probability distribution $\{p_k\}$ changes from generation to generation and depends on the instantaneous population size.

In theory the methods apply also to *bivariate populations*, but to describe the actual state of the population we then require two integers. Accordingly the transition probabilities will now be of the form $p(j_1, j_2; k_1, k_2)$: this is the (conditional) probability that a generation will consist of k_1 individuals of the first kind and k_2 individuals of the second kind, given that the corresponding numbers for the preceding generation are j_1 and j_2. Everything becomes rather involved. For example, the three types of extinction probabilities can be determined from infinite systems of linear equations analogous to (2.5), but the solutions are not unique and are hard to get even in the simplest cases. As was mentioned in the introduction, apparently no truly bivariate model has ever been treated, and this should be a challenge to mathematicians.

3. Discrete model with constant population size

With respect to a particular pair of genes, say a and A, each individual belongs to one of the three genotypes (a, a), (a, A), or (A, A), and we should actually consider a trivariate population of N individuals. However, an accepted standard simplification consists in introducing the genes themselves as elements of the population so that we deal with a population of $2N$ elements which are either a or A. Here N is a fixed number.

For a first orientation let us begin with the simple *random mating* case disregarding mutation pressures and selective forces. Mathematically speaking, the assumption of random mating amounts to saying that the $2N$ genes of any generation are formed in $2N$ independent trials: if the parent population consists of j a-genes and $2N - j$ A-genes, then each trial results in a or A with probabilities

$$p_j = \frac{j}{2N}, \qquad q_j = 1 - \frac{j}{2N}, \tag{3.1}$$

respectively. In other words, we have now a Markov chain with transition probabilities given by the binomial distribution

$$p_{jk} = \binom{2N}{k} p_j^k q_j^{2N-k}. \tag{3.2}$$

These, together with the initial population size, determine the entire process. We have now the two 'absorbing states' 0 and $2N$ which, once reached, are perpetuated: in biological language these occurrences are called *fixation*. The probability ϵ_r of a fixation at 0 (extinction of a-genes) if the initial number of a-genes is r, is again given by the system of equations (2.5), which is now finite. Its solution is easily found to be $\epsilon_r = 1 - r/2N$. The speed with which this fixation is reached can be

Doeblin [1] states that there exists a sequence $\lambda_n \to \infty$ such that $\lambda_n^{-1} p_{ik}^{(n)}$ converges to a limit $a_k > 0$ which is independent of j. It is easily verified that in our case the sequence λ_n can be chosen as $\lambda_n = f''^n(1)$. It follows that *the probability distribution of the normed variable* Z_n/λ_n *converges.* • This interesting fact is contained in stronger results of T. E. Harris [6].

judged from the largest non-trivial characteristic value of the matrix (p_{jk}). This turns out to be $1 - 1/2N$. This means that if a_n is the probability of no fixation within n generations, a_{n+1}/a_n is approximately $1 - 1/2N$. This is an agreement with estimates of the ratio of decay obtained by quite different methods by Fisher and Wright.[7] It also agrees with a result to be obtained from an approximating diffusion equation (compare section 8).

We now modify the model so as to take into account mutations. We assume that at the formation of each new generation each gene has the possibility to mutate, that is, to change into a gene of the other kind. More precisely we assume that the mutation

$$a \to A \text{ occurs with probability } \alpha_1 \tag{3.3}$$

$$A \to a \text{ occurs with probability } \alpha_2 .$$

Again we assume that the $2N$ genes are statistically independent. This means that if a particular generation consists of j a-genes and $2N - j$ A-genes, the $2N$ genes of the next generation are formed by $2N$ independent trials each of which results in a with probability

$$p_j = \frac{j}{2N}(1 - \alpha_1) + \left(1 - \frac{j}{2N}\right)\alpha_2 \tag{3.4}$$

and in A with probability

$$q_j = \frac{j}{2N}\alpha_1 + \left(1 - \frac{j}{2N}\right)(1 - \alpha_2) . \tag{3.5}$$

With this new definition of p_j and q_j the transition probabilities p_{jk} are still given by (3.2).

We still have a simple Markov chain, but the process exhibits an entirely new character. Thus, if $\alpha_1 > 0$, $\alpha_2 > 0$ there is no possibility of fixations. Instead, the n-step transition probabilities $p_{jk}^{(n)}$ [cf. (2.4)] approach a *steady state probability distribution* $\{x_k\}$ which is given by the solution of the system of equations [obtained from (2.4) letting $n \to \infty$]

$$x_k = \sum_{\nu=1}^{2N} x_\nu p_{\nu k} . \tag{3.6}$$

Whatever the initial composition of the population, the probability that the n-th generation is composed of k a-genes and $2N - k$ A-genes tends to x_k. The distribution $\{x_k\}$ is called the *steady state gene frequency distribution* and is of greatest interest. An explicit formula for it is not known, and most of the work consists in deriving suitable approximations.

Selective forces in nature convey advantages to one kind of gene and disadvantages to the other. Qualitatively they act in the same direction as mutations and it is therefore possible, in a first approximation, to account for them in the same way. At any rate, the usual models of genetic theories amount to assuming

[7] Compare, for example, [5, p. 87] or [12, p. 16]. The Markov chain (3.2) was considered by Malécot [9]. He tried to determine the first characteristic value approximately from the consideration of an integral equation resembling (3.2). His approximation is $1 - 1/(2N + 1)$. Formula (8.1) gives the characteristic values of the more general chain of the text, and reduces to the present case if $\alpha_1 = \alpha_2 = 0$. That $1 - 1/2N$ is a characteristic value is easily verified by noting that $x_k = k^2 - k$ is a characteristic vector.

that the coefficients a_i in (3.4) and (3.5) represent the combined influence of mutation pressure and selective forces. Accordingly, we shall introduce no further improvement in our scheme.

It must be emphasized, however, that with a constant population size any decrease say of a-genes implies an equal increase of A-genes. If the population size changes then the same ratio of the numbers of a- and A-genes may represent different compositions of the population and, strictly speaking, *selective pressure can not be a function of the gene ratios*. Accordingly, any theory derived from the present model (or an equivalent to it) *depends essentially on the assumption of a constant population* size and the number of a-genes must not be identified with the a-gene frequency in a variable population. Only a bivariate model permits to treat the latter, and such models lead to new types of equations (section 10).

Let the random variable $Z^{(n)}$ represent the number of a-genes in the n-th generation, and $\phi_n = Z^{(n)}/2N$ the corresponding *gene frequency*. At a time when $Z^{(n)} = j$ the *expected change of the gene frequency in the next generation is*

$$E(\Delta\phi_n) = a_2 - (a_1 + a_2)\frac{j}{2N} \tag{3.7}$$

and the corresponding variance

$$\text{var}(\Delta\phi_n) = \frac{1}{2N} p_j q_j . \tag{3.8}$$

This quantity corresponds to the '*sampling variance*' of S. Wright. His point of departure is the expression (3.7), but the sampling variance is usually identified with the expression obtained from (3.8) by putting $a_1 = a_2 = 0$. The larger variance (3.8) takes into account that chance fluctuations are due not only to the random mating process, but also to mutations and selective forces.

4. The general diffusion equation

Let $Z(t)$ be a random variable (such as gene frequency, population size, etc.) depending on the time t which we now agree to treat as a continuous parameter. The stochastic process described by $Z(t)$ is said to be of the Markov type if, roughly speaking, future changes depend on the present state, but not on the past history which led to this present state. All processes which we shall consider are of the Markov type.[8]

The variable $Z(t)$ may change in jumps: such is the case, for example, if $Z(t)$ represents the number of incoming telephone calls, or the size of a population as long as we count individuals and do not introduce the usual continuous approximation. With discontinuous processes the probability of a change during a small time interval $(t, t + h)$ is small (of the order of magnitude h), but *if* a change occurs, it is of finite magnitude. Opposed to such processes are processes of the diffusion type where $Z(t)$ changes continually. In this case there is certainty that during any time interval, however small, $Z(t)$ undergoes some change, but for small time intervals this change is practically sure to be small. More precisely, the probability that $|Z(t + h) - Z(t)| > \epsilon$ is of smaller order of magnitude than h.[9]

[8] Compare footnote 1.

[9] For this definition compare [2] where the equation is derived under more general conditions than stated in the text. The whole theory, of course, is due to Kolmogorov [7]. It should be noted

As in the case of finite chains, the whole process is determined by the initial value of $Z(t)$ at time 0 and by the *transition probability density*. By this we mean the function $p(t; z_1, x)$ which gives the (conditional) probability density that $Z(t_0 + t) = x$ given that $Z(t_0) = z$ at some fixed time t_0. This is the continuous analogue of the transition probabilities $p_{jk}^{(n)}$ of section 2. In the discrete case all the $p_{jk}^{(n)}$ could be expressed in terms of the one step transition probabilities p_{jk}. It is the most remarkable fact of diffusion theory that $p(t; z, x)$ can be calculated for all t if only the *infinitesimal mean displacement* $a(x)$ *and the infinitesimal variance* $2b(x)$ are known. These quantities are defined in analogy with (3.7) and (3.8) as follows.

Suppose it is known that at some fixed time t_0 we have $Z(t_0) = z$. The expected value of $Z(t)$ at a later time $t_0 + h$ is then

$$E[Z(t_0 + h)] = \int x p(h; z, x)\, dx \tag{4.1}$$

and the expected value of the increase (the so called mean displacement) is

$$E[\Delta Z(t_0)] = \int (x - z)\, p(h; z, x)\, dx . \tag{4.2}$$

Similarly

$$\text{var}\,[\Delta Z(t_0)] = \int (x - z)^2 p(h; z, x)\, dx . \tag{4.3}$$

These quantities are functions of h and z, and for small h they must be of the order of magnitude h. It is only a mild regularity assumption that as $h \to 0$

$$\frac{1}{h} E[\Delta Z(t_0)] \to a(z) \tag{4.4}$$

$$\frac{1}{h} \text{var}\,[\Delta Z(t_0)] \to 2b(z) .$$

These are the infinitesimal mean displacement and variance.

It was shown by Kolmogorov [7] *that the probability density* $u(t, x)$ of $Z(t)$ *satisfies the diffusion equation* (or 'Fokker-Planck equation')

$$u_t(t, x) = \{b(x)\, u(t, x)\}_{xx} - \{a(x)\, u(t, x)\}_x \tag{4.5}$$

(where subscripts indicate partial derivatives). Given the coefficients $a(x)$ and $b(x)$ and the initial distribution $u(0, x)$ of $Z(0)$, the solution $u(t, x)$ of (4.5) is in general[10] uniquely determined. In particular, if initially $Z(0) = z$, then the distribution $u(t, x)$ of $Z(t)$ coincides with $p(t; z, x)$ so that, for every fixed z, the transition

that there exist other types of Markov processes. Furthermore, it is in the text tacitly assumed that the transition probability depends only on the length of the time interval, but not on its position; in the general case the coefficients of the diffusion equation depend on x and t.

[10] We are here slurring over a delicate point of greatest interest from the point of view of differential equations. In physical diffusion theory the solution depends not only on the initial distribution, but also on boundary conditions. By contrast, our equations are of a *singular type* with the coefficient $b(x)$ vanishing on the boundaries. This leads to a new phenomenon, namely *natural boundaries* where no conditions need, or can, be imposed. In some cases there exists only one solution, in others there are many solutions, but only one corresponding to our problem. The theory of such singular equations is practically *terra incognita* and presents many interesting problems. It will be seen that the theory of evolution leads to boundary conditions of an altogether new type (compare section 7).

probability $p(t; z, x)$ is a solution of (4.5). In other words, the entire process is determined by the two coefficients $a(x)$ and $b(x)$.

Population growth is, strictly speaking, a discontinuous process, but for *large populations* the continuous model becomes a reasonable approximation, provided an appropriate time scale is used. In fact, it seems that this approximation comes closer to realities than many other simplifying assumptions. Most models used in evolution theory assume continuous growth, although this assumption may be more or less hidden. It follows then that in the univariate case such a model is mathematically equivalent to a particular choice of the coefficients $a(x)$ and $b(x)$.[11] In particular, as has been noted in the introduction, an essential part of the results of S. Wright can be derived from a certain diffusion equation. We shall discuss this aspect in section 7, but a few preparations seem desirable.

5. Exponential population growth and the branching process

The simplest illustration of the possibilities of the method is provided by a model which is the continuous analogue of the classical branching process and is actually an approximation to it when the population is large, that is, in its later phases.

We wish to describe a continuous population growth in which the individuals multiply and are statistically independent of each other with a fertility function independent of the total population size. In this case the infinitestimal mean displacement and variance are necessarily proportional to the instantaneous population size, and the diffusion equation (4.5) takes on the form

$$u_t(t, x) = \beta\{xu(t, x)\}_{xx} - \alpha\{xu(t, x)\}_x, \qquad 0 < x < \infty, \tag{5.1}$$

where α and β are constants. Their numerical values depend on the choice of units, and using appropriate scales on the t- and x-axes one can always achieve that $\beta = 1$, $\alpha = 1$, -1, or 0. The coefficient α measures the *drift* and is positive or negative according as the population average increases or decreases.

The equation (5.1) contains implicitly all probability relations governing our process and we proceed to verify that this model leads to sensible results and to compare it to the branching process.

The following explicit solution of (5.1) is given for the sake of completeness only. If the population size at time $t = 0$ is ξ, and $\alpha \neq 0$, then[12]

$$u(t, x) = \frac{\alpha}{\beta(e^{\alpha t} - 1)} \left(\frac{\xi e^{\alpha t}}{x}\right)^{1/2} I_1\left(\frac{2\alpha(\xi x e^{\alpha t})^{1/2}}{\beta(e^{\alpha t} - 1)}\right) e^{-\alpha(\xi e^{\alpha t} + x)/\beta(e^{\alpha t} - 1)} \tag{5.2}$$

$$= \frac{\xi \alpha^2 e^{\alpha t}}{\beta^2(e^{\alpha t} - 1)^2} e^{-\alpha(\xi e^{\alpha t} + x)/\beta(e^{\alpha t} - 1)} \sum_{\nu=0}^{\infty} \frac{1}{\nu!(\nu + 1)!} \left\{\frac{\alpha(\xi x e^{\alpha t})^{1/2}}{\beta(e^{\alpha t} - 1)}\right\}^{2\nu}.$$

[11] This is, of course, true only if one disregards after effects due to variations in age etc., which would make the process non-Markovian.

[12] Compare [3] where more general models are treated, but no proofs are given. Equation (5.1) is an example of a singular diffusion equation with a *unique* solution, where no boundary conditions can be imposed. (Compare footnote 10.) More curious is the equation (4.5) with $b(x) = \beta x$, $a(x) = \alpha x + \nu$ [which for $\nu = 0$ reduces to (5.1)]. Here uniqueness and various types of boundary effects depend on whether $\nu < 0$, $0 < \nu < 1$, or $\nu > 1$ (unpublished results).

The *expected population size*

$$M(t) = E[Z(t)] = \int_0^\infty xu(t, x)\,dx \tag{5.3}$$

can be obtained from (5.2), but it is simpler to note that easy manipulations on (5.1) show that $M'(t) = \alpha M(t)$, and hence

$$M(t) = \xi e^{\alpha t}. \tag{5.4}$$

Thus the population average increases exponentially, as should be expected since the rate of increase is proportional to the instantaneous population size. In a similar way one sees that

$$\text{var}\,[Z(t)] = 2\frac{\beta}{\alpha} e^{\alpha t}(e^{\alpha t} - 1). \tag{5.5}$$

Finally, the probability that the population dies out before time t is

$$\delta(t) = 1 - \int_0^\infty u(t, x)\,dx = e^{-\alpha\xi e^{\alpha t}/\beta(e^{\alpha t}-1)} \tag{5.6}$$

and hence we find for the probability of *ultimate extinction*

$$\lim_{t\to\infty} \delta(t) = \begin{cases} e^{-\alpha\xi/\beta} & \text{if} \quad \alpha > 0, \\ 1 & \text{if} \quad \alpha \leqq 0. \end{cases} \tag{5.7}$$

These results agree closely with the corresponding results for the simple *branching process* of section 2. There too the extinction probability is of the form γ^ξ. The generating function of the population size $Z^{(n)}$ in the n-th generation is $f^\xi(x)$, where ξ is the initial population size. As is well known (compare, for example, [4] or [6]),

$$E(Z^{(n)}) = \xi\mu^n \tag{5.8}$$

and

$$\text{var}\,(Z^{(n)}) = \xi\frac{\lambda}{\mu(\mu-1)}\mu^n(\mu^n - 1) \tag{5.9}$$

where $\mu = f'(1)$ is the expectation of the number of direct descendants and $\lambda = f''(1) + f'(1) - f'^2(1)$ the variance.

These formulas are not only of the same form as (5.4) and (5.5), but the latter follow from (5.8) and (5.9) as a limiting case for large populations.

To see this note that if in the branching process the population size is at any time x, then the size of the next generation has mean μx and variance λx. This means that the quantities $\mu - 1$ and λ have the significance of α and β in the continuous process.

Consider now a branching process in which the population is very large, and let us agree to measure the population in units of N individuals. In the new units each individual has weight $1/N$ and therefore the quantities $\mu - 1$ and λ are of the order of magnitude $1/N$. Accordingly we write

$$\mu - 1 = \frac{\alpha}{N}, \qquad \lambda = \frac{\beta}{N}. \tag{5.10}$$

We consider the process after a number of generations which is of the order of

magnitude N. This amounts to introducing a time scale in which one generation corresponds to a time interval $\Delta t = 1/N$, and an interval of length t contains some $n = tN$ generations. With these units (5.8) becomes

$$\xi\left(1+\frac{a}{N}\right)^{tN} \sim \xi e^{at} \tag{5.11}$$

and (5.9)

$$\xi \frac{\beta}{a\left(1+\frac{a}{N}\right)}\left(1+\frac{a}{N}\right)^{tN}\left\{\left(1+\frac{a}{N}\right)^{tN}-1\right\} \sim \xi \frac{\beta}{a} e^{at}(e^{at}-1). \tag{5.12}$$

In other words, *as $N \to \infty$ the formulas for the branching process go over into those for the continuous process.* That the same is true for the probability distributions will be shown in appendix II.

6. Gene frequency in the bivariate case

As we shall see in the next section, the theory of S. Wright assumes a diffusion equation for the gene frequency, that is, the proportion of a-genes in the population. With a population of constant size this frequency is proportional to the number of a-genes. This is not so in general. In any realistic bivariate model the numbers of a- and A-genes will be negatively correlated, but if the population size is not constant, the correlation coefficient will not be -1. In this case the *gene frequency cannot depend on a diffusion equation, and indeed, is not subject to a Markovian process.* This point will be helpful for a general understanding of the theory and may be illustrated by means of a model which is the extreme counterpart to the case of a constant population size: when the numbers of a- and A-genes are uncorrelated.

Consider two statistically independent populations each subject to the process described in the preceding section. Let the corresponding sizes be $X(t)$ and $Y(t)$, and the corresponding densities $u(t, x)$ and $v(t, y)$ respectively. The bivariate density of the pair $[X(t), Y(t)]$ has then the density $w(t, x, y) = u(t, x)v(t, y)$. Since $u(t, x)$ and $v(t, y)$ are solutions of equations of the form (5.1), it is easily seen that $w(t, x, y)$ satisfies the equation,

$$w_t(t, x, y) = \beta_1 \{xw(t, x, y)\}_{xx} + \beta_2 \{yw(t, x, y)\}_{yy} - a_1 \{xw(t, x, y)\}_x - a_2 \{yw(t, x, y)\}_y, \tag{6.1}$$

which would have been derived directly from the general theory for the bivariate case.

From $w(t, x, y)$ it is in theory possible to calculate the probability distributions of other random variables connected with the combined population. For example, the total population size is $S(t) = X(t) + Y(t)$, and the quantity corresponding to the gene frequency is $Z(t) = X(t)/S(t)$. Their probability densities are

$$\int_0^x w(t, \xi, x-\xi)\, d\xi \tag{6.2}$$

and

$$\frac{1}{x^2}\int_0^\infty \xi w\left(t, \xi, \frac{1-x}{x}\xi\right) d\xi, \qquad 0 < x < 1, \tag{6.3}$$

respectively. It is readily seen that neither is a solution of an equation of a diffusion

type. The reason is that neither $S(t)$ nor $Z(t)$ are Markovian. In fact, it is intuitively clear that a prolonged series of observations on $S(t)$ or $Z(t)$ will permit better predictions as to future developments than would the sole knowledge of the present state; this, however, is but another way of saying that the past history has an after effect.

7. The diffusion equation of Wright's theory

As has been mentioned before, the theory of S. Wright is essentially equivalent to the assumption[13] that the gene frequency satisfies a diffusion equation of the form

$$(7.1)\quad u_t(t, x) = \{\beta x(1 - x)\, u(t, x)\}_{xx} - \{[\gamma_2 - (\gamma_1 + \gamma_2)\, x]\, u(t, x)\}_x .$$

The theory hinges on the assumption that the gene frequency satisfies some diffusion equation, and on the particular choice of the coefficients $b(x)$ and $a(x)$. We have seen that in a bivariate population the gene frequency does not satisfy a diffusion equation unless *the population size is constant.* This, then, is theoretically an essential assumption of the theory. Presumably, however, it is in practice satisfied to a sufficient degree.

As for the coefficients, we shall derive them (section 8) by a passage to the limit from the Markov chain of section 2. This method reveals the assumptions underlying the theory and shows also what happens under other assumptions concerning the ratio of mutation pressure to population size.

First, however, let us describe a curious phenomenon concerning the *steady state.* If, as $t \to \infty$, the solution $u(t, x)$ of (4.5) tends to a limit $u(x)$, the latter must be a solution of the ordinary differential equation[14]

$$(7.2)\qquad \{b(x)\, u(x)\}_{xx} - \{a(x)\, u(x)\}_x = 0 .$$

Various forms of solutions were discussed by S. Wright.

If our diffusion were of the regular type, we would be assured that whenever (7.2) admits of a solution, the limiting relation $u(t, x) \to u(x)$ takes place, so that $u(x)$ describes the later stages of the process. Unfortunately this is *not* so for our problems, and mathematical problems of a new kind confront us.

Consider (7.1) in the simple case $\gamma_1 = \gamma_2 = 0$ (absence of mutation and selection pressures). We know that the discrete model in this case implies certainty of ultimate fixation, and it would be unfortunate for the continuous model if the results did not agree. Actually in our case $u(t, x) \to 0$, as desired. However, (7.2) has the solution $u(x) = C\{x(1 - x)\}^{-1}$, which is in reality not related to our problem.[15]

[13] It should be understood that even for a specific biological population the coefficients in (7.1) vary from place to place, that is, are functions of two additional geographic parameters. Geographic *migrations* constitute a new diffusion process so that Wright's theory actually envisages a compound diffusion process in a three dimensional phase space. In practice this process is simplified by discretizing it partially: this is done by means of the notion of population *isolates*, for which the coefficients are fixed. Even in this case, however, additional terms are introduced to account for migration effects, so that the theory is much more complicated than presented in the text.

[14] That the integration constant occurring on the right side of (7.2) vanishes follows from the fact that the left side represents the flow of probability mass at the origin.

[15] It is true that this is not a probability density, but this fact was erroneously attributed to the limited applicability of the continuous model near the boundaries, and it was assumed that $u(x)$ gives a reasonable approximation at least in the central part.

Even if the solution of (7.2) is a probability density it is not legitimate to conclude that $u(t, x) \to u(x)$. In general, there exists a fraction μ such that

$$u(t, x) \to \mu u(t, x), \qquad 0 < \mu < 1. \tag{7.3}$$

The difference $1 - \mu$ gives the probability of 'pure' populations in the steady state. At the beginning of the process probability mass flows out to the boundaries $x = 0$ and $x = 1$, but constant mutation lets part of this mass flow back into the interval where it is once more subject to the diffusion process. The net effect is that the masses concentrated at $x = 0$ and $x = 1$ increase steadily to a saturation point which is maintained in the steady state.

The actual determination of μ presents a difficult problem and what we have just described is a stochastic process of a type not yet studied and which deserves attention.

8. Passage to the limit

We proceed to show how the coefficients in (7.1) can be obtained by a limiting process from the discrete model of section 3.

We know that the probability distribution $\{x_k\}$ of the population size $Z^{(n)}$ in this model approaches a steady state distribution given by (3.6). The speed with which $p_{jk}^{(n)} \to x_k$ depends on the characteristic values $\lambda_0, \lambda_1, \ldots, \lambda_{2N}$ of the matrix (p_{jk}). It will be shown in appendix I that

$$\lambda_r = (1 - a_1 - a_2)^r \binom{2N}{r} \frac{r!}{(2N)^r}, \qquad r = 0, 1, \ldots, 2N. \tag{8.1}$$

[It is remarkable that λ_r depends only on the sum $a_1 + a_2$, but a similar statement is true also for the diffusion equation (7.1)]. Roughly speaking, $\lambda_0 = 1$ determines the steady state distribution $\{x_0\}$, and the next largest value determines the speed of the convergence. In fact, it is known from matrix theory that $p_{jk}^{(n)}$ can be written in the form

$$p_{jk}^{(n)} = x_k + a_1\lambda_1^n + a_2\lambda_2^n + \ldots + a_{2N}\lambda_{2N}^n, \tag{8.2}$$

where the coefficients a_ν depend on j and k, but not on n. Since $\lambda_1 > \lambda_2 > \ldots > \lambda_{2N}$ it follows that $p_{jk}^{(n)} - x_k$ decreases roughly as

$$\lambda_1^n = (1 - a_1 - a_2)^n \left(1 - \frac{1}{2N}\right)^n. \tag{8.3}$$

Here the first factor represents the influence of the mutation and selection pressures, the second the 'sampling variance.' Wright's theory considers mainly the case where the a_i are of the order of magnitude of N^{-1}, and we put

$$a_1 = \frac{\beta_1}{N}, \qquad a_2 = \frac{\beta_2}{N}. \tag{8.4}$$

We consider the β_i constant and study the asymptotic behavior of our model as $N \to \infty$. (Other possibilities are described in section 9.)

To observe a tendency to equilibrium we require a number of generations which is of the order of magnitude N. We, therefore, choose the time scale so that the

time required for one generation becomes $\Delta t = 1/N$: a time interval of duration t then corresponds to some $t/\Delta t = tN$ generations. For large populations we treat both time and the gene frequency as continuous variables and put

$$t = \frac{n}{N}, \qquad x = \frac{k}{2N}. \tag{8.5}$$

In the new notations the mean (3.7) and variance (3.8) of the change in population size correspond to the quantities occurring in equation (4.4). Hence,

$$a(x) = \lim_{N\to\infty} \frac{1}{\Delta t} E(\Delta\phi_n) = \beta_2 - (\beta_1 + \beta_2)\, x \tag{8.6}$$

and similarly

$$2\,b(x) = \lim_{N\to\infty} \frac{1}{\Delta t} \operatorname{var}(\Delta\phi_n) = \tfrac{1}{2} x(1-x). \tag{8.7}$$

We get thus the coefficients in (7.1) with $\gamma = \frac{1}{4}$ (this numerical value may be changed at will by a proper choice of scales).

It is noteworthy that[16] the solutions of (7.1) can be written in the form

$$u(t, x) = \sum e^{-\nu_r t} a_r f_r(x), \tag{8.8}$$

analogous to (8.2). Here

$$\nu_r = r\{\beta_1 + \beta_2 + \tfrac{1}{4}(r-1)\}. \tag{8.9}$$

It should be proved that our passage to the limit actually leads from (8.2) to (8.8). This has not yet been done, but at least it is easily seen that

$$\lim_{N\to\infty} \lambda_r^n = e^{-\nu_r t}, \qquad t = n\Delta t = \frac{n}{N}. \tag{8.10}$$

9. Other possibilities

The described passage to the limit which led to Wright's diffusion equation (7.1) is different from the familiar similar process in physical diffusion theory where the ratio $\Delta x/\Delta t$ tends to infinity rather than to a constant. It rests entirely on the assumption (8.4). We shall now show that *any modification of this assumption leads to a nonsingular diffusion equation of the familiar type (to normal distributions)*.

A continuous approximation to the discrete model of section 3 is sensible only if α_1 and α_2 are small. Accordingly we shall put

$$\alpha_1 = \beta_1\epsilon, \qquad \alpha_2 = \beta_2\epsilon. \tag{9.1}$$

We shall keep β_1 and β_2 fixed and perform the passage to the limit $\epsilon \to 0$, $N \to \infty$. In the preceding section we have solved the problem for the special case $\epsilon = 1/N$. Since the units of measurement are arbitrary, there remain essentially only the cases $N\epsilon \to 0$ and $N\epsilon \to \infty$.

A glance at (8.1) shows that if $N\epsilon \to 0$ the influence of the coefficients α_i becomes, in the limit, negligible; the process is asymptotically equivalent to the case of a pure random mating without mutation and selection. We assume therefore that

$$N\epsilon \to \infty. \tag{9.2}$$

[16] Compare the thesis by S. Goldberg (not yet published) giving explicit solutions of (7.1).

In this case the influence of the mutation pressure is predominant. Now this pressure is directed towards the *equilibrium point*

$$k = \frac{\alpha_2}{\alpha_1 + \alpha_2} = \frac{\beta_2}{\beta_1 + \beta_2}. \tag{9.3}$$

It is to be expected that once this equilibrium point is reached, large deviation from it will not occur. As a matter of fact, the gene frequency now tends in probability to the equilibrium point, and the fluctuation is concerned with the deviations from the equilibrium point measured in a scale in which they remain finite. In other words, we shall replace the population size $Z^{(n)}$ by the variable

$$Y^{(n)} = \frac{1}{\delta}\left\{\frac{1}{2N} Z^{(n)} - \frac{\beta_2}{\beta_1 + \beta_2}\right\}, \tag{9.4}$$

where $\delta > 0$ is a scale unit still to be determined.

Since the speed of the convergence towards the steady state depends essentially on α_1 and α_2, the number of generations required for continuous changes will be of the order of magnitude $1/\epsilon$. Accordingly, we introduce a time scale in which each generation requires time

$$\Delta t = \epsilon. \tag{9.5}$$

If in any generation $Y^{(n)} = x$ then the expectation of the change $Y^{(n+1)} - Y^{(n)} = Y^{(n+1)} - x$ turns out to be

$$-x(\alpha_1 + \alpha_2) = -x(\beta_1 + \beta_2)\,\Delta t. \tag{9.6}$$

In accordance with the definition (4.4) we get therefore for the linear coefficient

$$a(x) = -x(\beta_1 + \beta_2). \tag{9.7}$$

Similarly

$$\text{var}(\Delta Y^{(n)}) = \frac{1}{2N\delta^2} p_j q_j \sim \frac{1}{2N\delta^2} \frac{\beta_1\beta_2}{(\beta_1 + \beta_2)^2}. \tag{9.8}$$

Hence we get $b(x) = 1$ provided we choose δ so that the last expression becomes $2\Delta t$. Accordingly we put

$$\delta^2 = \frac{1}{4N\epsilon} \frac{\beta_1\beta_2}{(\beta_1 + \beta_2)^2}. \tag{9.9}$$

In view of (9.2) this quantity tends to 0. We thus are led to a process $Y(t)$ with a probability density satisfying *the diffusion equation*

$$u_t(t, x) = u_{xx}(t, x) + (\beta_1 + \beta_2)\{xu(t, x)\}_x. \tag{9.10}$$

Here the range of x is the *entire axis*. For an initial value $Y(0) = \eta$ (corresponding to a gene frequency $\beta_2/(\beta_1 + \beta_2) + \delta\eta$), the solution of (9.10) is

$$u(t, x) = \left\{\frac{\beta_1 + \beta_2}{2\pi(1 - e^{-2t(\beta_1+\beta_2)})}\right\}^{1/2} \exp\left\{-\frac{(\beta_1 + \beta_2)(x + \eta e^{-(\beta_1+\beta_2)t})}{2(1 - e^{-2(\beta_1+\beta_2)t})}\right\}. \tag{9.11}$$

Equation (9.10) is an ordinary diffusion equation with a *drift toward the origin*, and (9.11) is a normal probability density with mean and variance tending 0 and $1/(\beta_1 + \beta_2)$, respectively.

10. A bivariate model

The purpose of this section is to generalize the discrete model of section 3 by dropping the assumption that the population size is constant. This permits a more flexible treatment of selection forces. Many schemes can be devised, but for illustrative purposes it is preferable to describe a particular model rather than to go into vague generalities.

The composition of the population in the n-th generation is described by a pair of integers (j_1, j_2) indicating the numbers of a- and A-genes, respectively. These numbers are random variables to be denoted by $X_1^{(n)}$ and $X_2^{(n)}$. We wish to describe a natural scheme leading to explicit expressions for the transition probabilities $p(j_1, j_2; k_1, k_2)$ (that is, the probability that $X_1^{(n+1)} = k_1$ and $X_2^{(n+1)} = k_2$ given that $X_1^{(n)} = j_1$, $X_2^{(n)} = j_2$).

In the model of section 3 the new generation is formed in $2N$ independent trials. We want to take into account the possibility that, say, an a-gene dies without contributing to the succeeding generation and without thus indirectly increasing the (absolute) number of A-genes. Instead of keeping the population size rigidly constant we can adjust the scheme so that only the expected population size remains fixed. For that purpose we accept the following model.

If in a generation the totals of a- and A-genes are j_1 and j_2, then *the succeeding generation is formed in* $N = \sigma_1 j_1 + \sigma_2 j_2$ *independent multiple trials as follows. Each trial consists in first selecting an a- or an A-gene, with corresponding probabilities* $\sigma_1 j_1/N$ *and* $\sigma_2 j_2/N$. *If the result is an a-gene it has probability* ν_1 *to die without further influencing the next generation; if it survives, it has probability* α_1 *to mutate into an A-gene.* These act in a similar way with corresponding probabilities ν_2 and α_2.

In other words, each of the $N = \sigma_1 j_1 + \sigma_2 j_2$ trials has three possible outcomes. It results in an a-gene with probability

$$p_1 = \frac{\sigma_1 j_1}{N}(1 - \nu_1)(1 - \alpha_1) + \frac{\sigma_2 j_2}{N}(1 - \nu_2)\alpha_2; \tag{10.1}$$

the outcome is A with probability

$$p_2 = \frac{\sigma_1 j_1}{N}(1 - \nu_1)\alpha_1 + \frac{\sigma_2 j_2}{N}(1 - \nu_2)\alpha_2; \tag{10.2}$$

finally, the outcome is 'no contribution to the new generation' with probability $1 - p_1 - p_2$. Hence the transition probability is given by the trinomial distribution

$$p(j_1, j_2; k_1, k_2) = \frac{N!}{k_1! k_2! (N - k_1 - k_2)!} p_1^{k_1} p_2^{k_2} (1 - p_1 - p_2)^{N - k_1 - k_2}. \tag{10.3}$$

The essential difference between this scheme and that of section 3 is that in the present case the population can increase indefinitely or die out. No special results have been worked out for this new scheme, and we shall be satisfied with deriving the diffusion equation to which it leads in the same way as the model of section 3 leads to Wright's equation (7.1).

If in the n-th generation $X_1^{(n)} = j_1$, $X_2^{(n)} = j_2$, then we have for the increments $\Delta X_i = X_i^{(n+1)} - X_i^{(n)}$

$$\begin{aligned} E(\Delta X_i) &= Np_i - j_i, \\ \text{var}\,(X_i) &= Np_i(1 - p_i). \end{aligned} \tag{10.4}$$

New to the present case is only the covariance between the two increments,

$$\text{cov}\,(X_1, X_2) = -Np_1p_2 \tag{10.5}$$

(or a correlation coefficient $-\{p_1p_2/(1-p_1)(1-p_2)\}^{1/2}$ instead of -1 as in section 3).

Since we are in the bivariate case, the diffusion equation will be of the form

$$u_t(t, x, y) = \sum_{i,j=1}^{2} \{ b_{ij}(x_1, x_2)\, u(t, x_1, x_2) \}_{x_i x_j} - \sum_{i=1}^{2} \{ a_i(x_1, x_2)\, u(t, x_1, x_2) \}_{x_i} . \tag{10.6}$$

A special case is equation (6.1), where $b_{12} = 0$ because of the assumed independence of $X(t)$ and $Y(t)$. The coefficients in (10.6) (the infinitesimal mean displacements and variances) are now to be obtained by a limiting process from (10.4) and (10.5).

Again we denote the time between two succeeding generations by $\Delta t = \epsilon$ and measure the population in units of ϵ^{-1}. In section 8 we had $\epsilon = 1/2N$, but now N is not a constant. As there we put

$$a_i = \beta_i \Delta t\,, \qquad i = 1, 2\,. \tag{10.7}$$

Fluctuations in population size will be slow only if the survival chances $1 - \nu_i$ are close to the inverses of σ_i, and we put in analogy with (10.7)

$$1 - \nu_i = \frac{1 + \mu_i \Delta t}{\sigma_i}\,, \qquad i = 1, 2\,. \tag{10.8}$$

In the new units the expected change of the number of a-genes during a time interval of length Δt is

$$\begin{aligned} \epsilon E\,(\Delta X_1) &= \epsilon\,(Np_1 - j_1) \\ &= \Delta t\,\{-x_1\,(\beta_1 + \mu_1 - \beta_1\mu_1\epsilon) + x_2\beta_2\mu_2\epsilon\} \end{aligned} \tag{10.9}$$

and hence

$$a_i\,(x_1, x_2) = -x_i\,(\beta_i + \mu_i)\,, \qquad i = 1, 2\,. \tag{10.10}$$

A similar calculation for the variances leads to

$$b_{ii}(x_1, x_2) = \tfrac{1}{2} x_i \left(1 - \frac{x_i}{\sigma_1 x_1 + \sigma_2 x_2}\right), \tag{10.11}$$

and to

$$b_{12}(x_1, x_2) = \tfrac{1}{2} \frac{x_1 x_2}{\sigma_1 x_1 + \sigma_2 x_2}\,. \tag{10.12}$$

We have thus a bivariate diffusion equation in which the coefficients are not linear functions. If $\sigma_1 = \sigma_2 = 1$ the expressions (10.10) and (10.11) reduce to those in Wright's theory except that in the latter $x_1 + x_2$ is replaced by 1. Even if we put $\sigma_1 = \sigma_2 = 1$, our model is more general than the diffusion equation (7.1). In fact, with $\sigma_1 = \sigma_2 = 1$ the determinant $b_{11}b_{22} - b_{12}^2$ vanishes, and the introduction of new variables $\xi = x_1 + x_2$, $\eta = x_1 - x_2$ will reduce the second order terms in

(10.6) to a single second derivative. However, both variables still appear in the linear terms. If $\sigma_1 = \sigma_2 = 1$, $\mu_1 = \mu_2$, then (10.6) reduces to the equation (7.1) of Wright's theory.

APPENDIX I. THE CHARACTERISTIC VALUES OF THE MATRIX (3.2)

We want to verify formula (8.1) for λ_r and, incidentally, calculate the characteristic column vectors of the general matrix (p_{jk}) defined by (3.2) with p_j, q_j given in (3.4) and (3.5). For this purpose we have to show that for $r = 0, 1, \ldots, 2N$ the system of linear equations

$$\sum_{k=0}^{2N} p_{jk}x_k = \lambda_r x_j \tag{11.1}$$

admits of a nontrivial solution $\{x_0, \ldots, x_{2N}\}$.

Put for abbreviation

$$k_{(\nu)} = k\,(k-1)\ldots(k-\nu+1) \tag{11.2}$$

and note that

$$\sum_{k=0}^{2N} p_{jk}k_{(\nu)} = \frac{d^\nu}{dx^\nu}(q_j + p_j x)^{2N}\Big|_{x=1} = (2N)_{(\nu)}p_j^\nu . \tag{11.3}$$

We now prove that it is always possible to determine constants $a_0, a_1, \ldots, a_r$ (not all of them zero) so that

$$x_k = a_r k_{(r)} + a_{r-1}k_{(r-1)} + \ldots + a_1 k + a_0 \tag{11.4}$$

is a solution of (11.1). This means that *for the characteristic vector belonging to* λ_r *the component* x_k *is a polynomial in* k *of degree at most* r.

The assertion to be proved is that the a_i can be chosen so as to satisfy the $2N + 1$ equations

$$\sum_{\nu=0}^{r} a_\nu (2N)_{(\nu)}p_j^\nu = \lambda_r \sum_{\nu=0}^{r} a_\nu j_{(\nu)} . \tag{11.5}$$

Now both p_j^ν and $j_{(\nu)}$ are polynomials of degree ν in j, and it is possible to write

$$p_j^\nu = \sum_{s=0}^{\nu} c_{s,\nu}\, j_{(s)} , \tag{11.6}$$

where the $c_{s,\nu}$ are independent of j. Equating the coefficients in (11.5) we get

$$\lambda_r a_t = \sum_{\nu=t}^{r} a_\nu (2N)_{(\nu)}\, c_{t,\nu} , \qquad t = 0, 1, \ldots, r . \tag{11.7}$$

But

$$c_{\nu,\nu} = (1 - a_1 - a_2)^\nu N^{-\nu} \tag{11.8}$$

so that $(2N)_\nu c_{\nu,\nu} = \lambda_\nu$, and hence (11.7) is satisfied for $\nu = r$ and arbitrary a_r. Put $a_r = 1$. Then (11.7) permits us to calculate in succession $a_{r-1}, a_{r-2}, \ldots$ and

this process breaks down only if for some $\nu < r$ one has $\lambda_\nu = \lambda_r$. In this case one can put $a_r = 0$, and thus we get $2N + 1$ independent characteristic vectors.

APPENDIX II. THE PASSAGE FROM THE BRANCHING PROCESS TO THE DIFFUSION EQUATION (5.1)

The results of section 5 show that a direct passage from the simple branching process of section 2 to the equation (5.1) must be possible. We effect this passage to the limit formally: it is not difficult to justify the steps, since the necessary regularity properties of the generating functions $f_n(x)$ were established by Harris [6].

We consider the case of large populations, that is, the later stages of a branching process in which the mean number of descendants $\mu = f'(1) > 1$. We introduce units for measuring time and population size, such that during 'small' time intervals the fluctuations will be 'small.' This requires that in the new units the quantity

$$\mu = 1 + \epsilon \tag{12.1}$$

must be small. Accordingly we introduce new units in such a way that an individual in the old counting and the time of one generation correspond to

$$\Delta x = \Delta t = \epsilon . \tag{12.2}$$

Since the population size will no longer be an integer, we pass from the generating function $f(x)$ to the characteristic function $f(e^{iz})$ or, in the new units, $f(e^{iz\epsilon})$.

Put

$$f''(1) = \lambda , \tag{12.3}$$

so that the variance of the number of direct descendants becomes

$$\sigma^2 = \lambda + \mu - \mu^2 . \tag{12.4}$$

Then, as $\epsilon \to 0$

$$\begin{aligned} 1 - f(e^{iz\epsilon}) &= (1 - e^{iz\epsilon})\,\mu + \tfrac{1}{2}(1 - e^{iz\epsilon})^2 \lambda + O(\epsilon^3) \\ &= -iz\mu\epsilon - \tfrac{1}{2}z^2(\lambda + \mu)\,\epsilon^2 + O(\epsilon^3) \end{aligned} \tag{12.5}$$

and hence

$$\log f(e^{iz\epsilon}) = i\epsilon z \left\{ z + \epsilon\left(z + \frac{i}{2} z^2\sigma^2 \right) + O(\epsilon^2) \right\} . \tag{12.6}$$

Now put

$$\phi(t, z) = f_n(e^{iz\epsilon}) , \tag{12.7}$$

where $t = n\epsilon$. Then

$$\begin{aligned} \phi(t+\epsilon, z) &= f_n[f(e^{iz\epsilon})] \\ &= \phi\left(t, \frac{1}{i\epsilon} \log f(e^{iz\epsilon}) \right) \\ &= \phi\left[t,\ z + \epsilon\left(z + \frac{i}{2} z^2\sigma^2 \right) + O(\epsilon^2) \right] . \end{aligned} \tag{12.8}$$

In the limit we are led to the differential equation

$$\phi_t(t, z) = \left(z + \frac{i}{2} z^2\sigma^2 \right) \phi_z(t, z) . \tag{12.9}$$

It follows that $\phi(t, z)$ is the characteristic function of a density $u(t, x)$ which satisfies the diffusion equation

$$u_t(t, x) = \frac{\sigma^2}{2}\{xu(t, x)\}_{xx} - \{xu(t, x)\}_x \tag{12.10}$$

and the boundary condition $u(t, 0) = 0$. Now (12.10) is a special case of (5.1), and the boundary condition follows from the fact that in the branching process the probability mass flowing out into the origin tends to zero.

Harris' remark that in the branching process the distribution of $Z^{(n)}\mu - n$ converges holds also for our differential equation.

REFERENCES

[1] W. DOEBLIN, "Sur deux problèmes de M. Kolmogoroff concernant les chaines dénombrables," *Bull. Soc. Math. France*, Vol. 66 (1939), pp. 1–11.

[2] W. FELLER, "Zur Theorie der stochastischen Prozesse," *Math. Annalen*, Vol. 113 (1936), pp. 113–160.

[3] ———, "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung," *Acta Biotheoretica*, Vol. 5 (1939), pp. 11–40.

[4] ———, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1950.

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

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

[7] A. KOLMOGOROV, "Über die analytischen Methoden der Wahrscheinlichkeitsrechnung," *Math. Annalen*, Vol. 94 (1931), pp. 415–458.

[8] ———, "Deviations from Hardy's formula in partial isolation," *C. R. Acad. Sci. URSS*, Vol. 3, No. 8 (1935), pp. 129–132.

[9] G. MALÉCOT, "Sur un problème de probabilités en chaine que pose la génétique," *C. R. Acad. Sci. Paris*, Vol. 219 (1944), pp. 379–381.

[10] ———, *Les Mathématiques de l'Hérédité*, Masson et Cie, Paris, 1948.

[11] ———, "Les processus stochastiques de la génétique," to be published.

[12] S. WRIGHT, "Statistical genetics in relation to evolution," *Actualités Scientifiques et Industrielles*, No. 802, Hermann et Cie, Paris, 1939.

[13] ———, "Statistical genetics and evolution. (Gibbs Lecture)," *Bull. Amer. Math. Soc.*, Vol. 48 (1942), pp. 223–246.

[14] ———, "The differential equation of the distribution of gene frequencies," *Proc. Nat. Acad. Sci.*, Vol. 31 (1945), pp. 382–389.