

165/H

# THE MATHEMATICS OF BIOLOGY

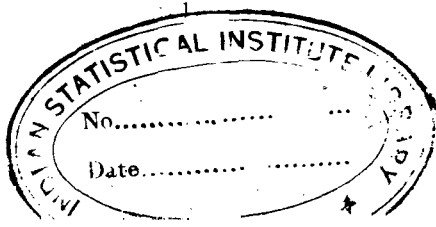
By Professor J. B. S. HALDANE, F.R.S.

*A Lecture delivered before a Joint Meeting of the Society with the Mathematical and Physical Society, on Tuesday, the 6th of November, 1951, the PRESIDENT being in the Chair.*

MUCH of the mathematics which biologists use falls into one of two categories. Some of it is concerned with physical or chemical processes occurring in the instruments we use or the living systems which we study. For example, Swann and Mitchison are now working on the design of the polarizing microscope, which has, till lately, been neglected as a tool for investigating cell structure. About 25 years ago Briggs and I produced a differential equation concerning enzyme action, which we solved in the trivial case of the steady state. Chance was able to study the first few milliseconds of an enzyme action, and much to my surprise, found that the equation described them reasonably well. But he had to use a machine of the analogical type to compute its solution. The other kind of mathematics which we use on a large scale are those of statistics. The distributions of measurements in a thousand men and a thousand castings are probably not very unlike, though the human measurements doubtless have the larger coefficient of variation. Many of the statistical methods used in industry are based on methods invented by Pearson and others for use on biological data.

I am going to deal with problems which have no exact analogies (so far as I know) in physics, chemistry, or engineering. Let us consider the growth of a population. Clearly the simplest case is when the probabilities of dying or of producing offspring depend on the age of the individual only, and not on the epoch or the population density for example. We can approximate to such cases in a laboratory with mice or insects kept under standard conditions.

Let  $t$  be the epoch (e.g. 1951.853),  $x$  the age of an individual (e.g. 4 days or 40 years). Let  $P(t)$  be the number of females in the population at time  $t$ . We neglect the males, because *ex hypothesi* there are enough to fertilize the females. Let  $R$  be the mean number of female progeny produced by a female in a lifetime. We count every female born or hatched as a potential mother. The actual number  $r$  of which  $R$  is the mean may thus be 0 in 99% of cases, and about 100 in the other 1%. Let  $\phi = \ln R$ . Then if  $R > 1$ ,  $\phi > 0$ , the population will ultimately



increase exponentially until our initial assumptions cease to hold. If we are dealing with insects or plants with one generation per year, clearly

$$P(t) = P_0 R^t = P_0 e^{\phi t}.$$

Next let us suppose that, if we consider a group of females,  $f(x)\delta x$  is the probability that one of their daughters will be born between maternal ages  $x$  and  $x+\delta x$ . Then if  $B(t)\delta t$  is the number born between times  $t$  and  $t+\delta t$ ,

$$B(t) = R \int_0^\infty B(t-x) f(x) dx \quad (1)$$

For at time  $t$  there are  $B(t-x)\delta x$  mothers aged  $x$  to  $x+\delta x$ . Also if  $l(x)$  is the probability of living to age  $x$  or over

$$P(t) = \int_0^\infty B(t-x) l(x) dx \quad (2)$$

Now  $f(x)$  is a distribution function, so  $\int_0^\infty f(x) dx = 1$ . To solve (1) put  $B(t) = A e^{\lambda t}$ . Then

$$\int_0^\infty e^{-\lambda x} f(x) dx = R^{-1}$$

But the left-hand side is the moment-generating function of the maternal age distribution (or if  $\lambda = -iz$ , the characteristic function). Taking logarithms we have

$$-\phi \pm 2m\pi i = K(-\lambda) = -\kappa_1 \lambda + \frac{\kappa_2}{2!} \lambda^2 - \frac{\kappa_3}{3!} \lambda^3 + \quad (3)$$

$\kappa_r$  is the  $r$ -th cumulant of the distribution, that is to say  $\kappa_1$  is its mean,  $\kappa_2$  its variance,  $\kappa_3$  its skewness,  $\kappa_4$  its kurtosis, and so on.

This equation, which is due to Lotka, can easily be solved numerically in three cases. One is when the distribution of maternal ages is nearly normal. The human distribution, though skewed, is near enough to make the corrections relatively small. The second is when it is nearly represented by a curve rising steeply from zero at puberty, and falling exponentially. This applies to some insects. The third is when it is concentrated in a number of short breeding seasons. If the number is finite this gives an algebraic equation for  $e\lambda$ , but its real root is most easily obtained from (3) if breeding can go on for 10 or 20 years. (3) has only one real root, representing an exponential increase or decrease, and in general an infinity of complex roots representing oscillations round it, so that

$$B(\lambda) = \sum_{m=0}^{\infty} A_m e^{\lambda m^2}$$

$$\lambda_0 = \frac{\phi}{\kappa_1} \left[ 1 + \frac{\kappa_2 \phi}{2! \kappa_1^2} + \frac{3\kappa_2^2 - \kappa_1 \kappa_3}{3! \kappa_1^4} \phi^2 + \right. \\ \left. + \frac{15\kappa_2^3 - 10\kappa_1 \kappa_2 \kappa_3 + \kappa_1^2 \kappa_4}{4! \kappa_1^6} \phi^3 + \dots \right].$$

$$\begin{aligned} \lambda_m = \lambda_0 - \frac{2m^2\pi^2}{\kappa_1^2} & \left[ \kappa_2 + \frac{(3\kappa_2^2 - \kappa_1\kappa_3)}{\kappa_1^2} \phi \right. \\ & \left. + \frac{5\kappa_2^3 - 10\kappa_1\kappa_2\kappa_3 + \kappa_1^2\kappa_4}{6\kappa_1^4} (3\phi^2 - 2m^2\pi^2) + \dots \right] \\ \pm \frac{2m\pi i}{\kappa_1} & \left[ 1 + \frac{\kappa_2 \phi}{\kappa_1^2} + \frac{(3\kappa_2^2 - \kappa_1\kappa_3) (3\phi^2 - 4m^2\pi^2) + \dots}{6\kappa_1^4} \right] \quad (4) \end{aligned}$$

So the period of the slowest and least damped oscillation is about a generation. Clearly  $P(t)$  is given by a similar series.

Now let us see what happens if, in the trivial case of annual generations,  $R$  depends on the population density. Let  $y_n$  be the logarithm of the density in year  $n$ , and let  $R_n$  be a function of  $y_n$ . Then

$$y_{n+1} = y_n + \phi_n = y_n + \phi(y_n)$$

$\phi(y_n)$  passes through a positive maximum value as  $y_n$  increases. That is to say the population increases when it lies between certain limits. We assume that  $\phi'(y_n)$  is always negative.

At high densities the population diminishes owing to competition for space, or more rarely for food, and especially to disease due to overcrowding. It also falls off at low densities, owing to the difficulty of finding mates, to the lack of various forms of mutual aid, and so on.

Thus  $\phi = 0$  for two values of  $y_n$ , giving two equilibria, of which the lower is certainly unstable, the upper may or may not be stable. If we measure density so that  $y = 0$  at the upper equilibrium, then it is easily shown that it will be approached directly after small displacements in either direction if

$0 > \phi'(0) > -1$ . If  $-1 > \phi'(0) > -2$  it will be approached by a series of oscillations. If  $\phi'(0) < -2$ , small oscillations round it will be unstable. With a graph of the form shown the density will never exceed the upper equilibrium density in two consecutive years, but may fall below it for a number of consecutive years. If we put  $y_n + \phi(y_n) = \Psi(y_n)$ , then there will be periodic oscillations of period  $k$  years if  $\Psi^k(y_n) - y_n = 0$ , where  $\Psi^k$  denotes the  $k$ -fold iteration of the function  $\Psi$ . It is easy to find a criterion for the stability of such periodic oscillations. We see that too sharp a regulation of population density will lead to instability, and this is true in general. However, if generations overlap, not only  $R$  but  $f(x)$  will be a function of the density, and the analysis would be extremely complicated.

We see then that even the one-species case, which corresponds to the physical case of a body moving in a defined field, can be very difficult. The case of two or three interacting species has been dealt with extensively by Volterra, and his results are, of course, fundamental for the theory of integral equations. But they are, in fact, an oversimplification of the biological facts almost as gross as my equations (1) and (2). Nicholson and Bailey (1) gave what seems to me a more

realistic account of one type of two-species relationship, that between an insect host and an insect parasite. The host has adequate food. The parasite lays eggs in its larvae, which always kill them, and not more than one parasite emerges from a host. Both species have an annual generation. Let  $x$  be the density of adult hosts,  $y$  that of parasites (adult females only). Let  $R$  be the rate of increase of hosts in the absence of parasites, and  $a$  the mean area searched by a host female in her lifetime. In a case investigated by Varley  $a$  was about 0.25 square metre. Now, if there are  $y$  parasites per square kilometre, a given host larva has a probability  $1 - ay$  of escaping any given one of them, and  $(1 - ay)^y$  or nearly  $e^{-ay}$  of escaping them all. Thus  $e^{-ay}$  of the larvae escape parasitism, and  $1 - e^{-ay}$  of those which mature hatch out parasites. So if  $x', y'$  are the frequencies next year

$$\left. \begin{aligned} x' &= Rxe^{-ay} \\ y' &= Rx(1 - e^{-ay}) \end{aligned} \right\} \quad (5)$$

Note that if the parasites were co-operative and moderately intelligent their areas would be chosen not to overlap, so the whole ground would be covered, and we should have  $x' = y'' = 0$ ; so both species would become extinct. If they were somewhat more intelligent than our own species they would adopt a system of "egg control" calculated to give a stable and probably maximal value of  $y$ . The system is in equilibrium if  $x' = x = X, y' = y = Y$ , whence

$$\left. \begin{aligned} X &= \frac{\ln R}{a(R - 1)} \\ Y &= \frac{\ln R}{a} \end{aligned} \right\} \quad (6)$$

Now suppose that, as a result of natural selection (say for a better sense of smell) the parasite manages to increase  $a$ , the area effectively searched, then both  $X$  and  $Y$  will diminish. Similarly, if the host becomes more efficient at laying eggs or hiding from other enemies, so that  $R$  is increased, then its adult, though not its larval, numbers at equilibrium will be diminished, while those of the parasite will be slightly increased. Natural selection does not always lead to survival!

The equilibrium is unstable. For if we take  $a$  as our unit of area, and put  $x = X + \xi, y = Y + \eta$ , we have, when  $\xi$  and  $\eta$  are small

$$\begin{aligned} \xi' &= \xi - \frac{\ln R}{R - 1} \eta \\ \eta' &= (R - 1) \xi + \eta \end{aligned}$$

approximately. So small oscillations are of the form  $\xi = A_1 e^{-\lambda_1 t} +$

$$A_2 e^{-\lambda_2 t} \text{ where } \lambda_1 \text{ and } \lambda_2 \text{ are the roots of } \begin{vmatrix} 1 - \lambda & -\frac{\ln R}{R-1} \\ R-1 & 1 - \lambda \end{vmatrix} = 0$$

or  $\lambda = 1 \pm \sqrt{-\ln R}$ . That is to say in successive years

$$x = X + A(1 + \ln R)^{1/2} \sin(B + \theta t)$$

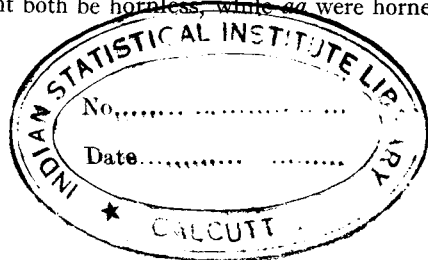
where  $\cos 2\theta = \frac{1 - \ln R}{1 + \ln R}$ , and a similar equation for  $y$ . Thus  $x$  and

$y$  have undamped oscillations of a period longer than 4 years, and less than 8 years if  $R > e$ . In fact  $R$  may exceed 10. The maxima of  $y$  lag about a quarter of a period behind those of  $x$ .

That is to say the regulation of population size has a considerable lag. In actual fact the logarithms of the densities of insect populations which are believed to be regulated in this way show somewhat irregular oscillations which are more or less sinusoidal with periods which are often of the order of 6 years. They are not, of course, undamped, presumably because when  $x$  becomes very large, other limiting factors come into play. (Graphs were shown by the lecturer).

This is a two-species case at the level of simplicity corresponding to  $P(t) = R^t$  in the one-species case. A good deal of calculations have been done on interaction between three or more species, but they seem to be rather far removed from adequate numerical data. Graphs based on actual numbers of pupae picked up were shown demonstrating that the density of an insect species may be at least 10,000 times greater in one year than in another. But no one except Varley (2) has made a serious attempt to classify all the causes of death of an insect in the course of its life cycle, together with at least some of the causes which affect its fertility. The study of fluctuations in natural animal populations is almost confined to Canadian fur-bearers and German insect pests of pine-woods. When the subject has been thoroughly studied I have no doubt that the theories here put forward will appear completely trivial; but they may indicate the kind of mathematical problems requiring solution.

I have devoted much more time to the study of changes within a population all of whose members can breed together. I shall confine myself this evening to the differences defined by a single pair of what are called allelomorphic genes  $A$  and  $a$ . There are three sorts of individuals,  $AA$ ,  $Aa$  and  $aa$ .  $AA$  and  $aa$ , which are called homozygotes, give an  $A$  or  $a$  gene as the case may be to each child.  $Aa$ , which are called heterozygotes, are equally likely to contribute  $A$  or  $a$ . So we can write down the expectations from all possible matings, e.g.  $AA \times AA \rightarrow AA$ ,  $Aa \times aa \rightarrow \frac{1}{2} Aa + \frac{1}{2} aa$ ,  $Aa \times Aa \rightarrow \frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa$ . Here  $AA$  might be white,  $Aa$  roan, and  $aa$  fully coloured cattle, or  $AA$  and  $Aa$  might both be hornless, while  $aa$  were horned.



Suppose we can classify a large population into  $m$  classes, then we have  $m$  frequencies,  $x_n, y_n, z_n, \dots$  in the  $n$ -th generation. We ask, given the mating system, the selective values of the different classes, the amount of immigration, mutation (that is to say change from  $A$  to  $\bar{a}$  or conversely), and so on, what will be the values of  $x_{n+1}, y_{n+1}, z_{n+1}$ , etc. We may be able to write down  $m$  equations ( $m-1$  independent)

$$\begin{aligned} x_{n+1} &= f_1(x_n, y_n, z_n, \dots) \\ y_{n+1} &= f_2(x_n, y_n, z_n, \dots) \\ z_{n+1} &= f_3(x_n, y_n, z_n, \dots) \\ &\dots \dots \dots \end{aligned} \quad (7)$$

If they are all linear, we can think of  $[x_n, y_n, z_n, \dots]$  as a vector converted into another vector by a matrix, and our problem, if we want to predict what will happen after  $n$  generations, is to iterate this matrix. If they are not linear we have a much harder problem unless the changes from one generation to another are so small that we can treat (7) as a set of simultaneous differential equations. These may be soluble by standard methods. I shall only discuss non-linear recurrence relations involving one or two variables here. When generations overlap, each of our equations becomes an integral equation of the general type of (1). A set of such equations is not intractable provided they are linear. When they are non-linear, approximate solutions can sometimes be found by fairly straightforward methods, but the full treatment (Norton, (3)) of even the simplest is very difficult.

If, however, the population is not so large that the number in each group can be regarded as infinite, the problem becomes a stochastic one. A point in an  $(m-1)$  dimensional simplex in one generation is represented not by a point, but by a cloud, in the next generation. The problems arising in this case have occasionally been solved fairly completely, but I cannot deal with them here. A further group of problems arises if the composition of the population differs in different parts of an area, and migration between them is possible. Local races adapted to their environments may or may not be able to persist. Here we have at best partial differential equations like those involved in the theory of heat flow or diffusion, but distressingly non-linear where one would wish the opposite. Again a few have been solved. Most have not even been stated.

To come back to actual problems, suppose we have  $AA, Aa$ , and  $aa$  individual plants with frequencies  $x_n, y_n$  and  $z_n$  in the  $n$ -th generation, and arrange that they should be self-fertilized, then

$$\left. \begin{aligned} x_{n+1} &= x_n + \frac{1}{4}y_n \\ y_{n+1} &= \frac{1}{2}y_n \\ z_{n+1} &= \frac{1}{4}y_n + z_n \end{aligned} \right\} \quad (8)$$

Clearly  $x_n = x_0 + \frac{1}{2}(1 - 2^{-n})y_0$ , and so on, but we can regard this trivial result in two ways which are illuminating.  $[x_n, y_n, z_n]$  is derived from  $[x_0, y_0, z_0]$  by the  $n$ -fold iteration of the matrix

$$\begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix} \quad \text{And we have solved a trivial stochastic problem.}$$

We have effectively divided our original population up into populations of 1, each plant being so far from any other that it cannot cross with it. We have found what will be the probable "composition" of a population of one after  $n$  generations.

The next problem can be regarded either as the problem of what will almost certainly happen in a very large population in which brother-sister mating is made compulsory after a certain time, or as the problem of what will probably happen to mice confined in a cage so small that in each generation there is just room for one breeding pair. In this case if we classify together various types of mating, with frequencies:—

$$w_n \begin{cases} AA \times AA, \\ aa \times aa \end{cases} x_n \begin{cases} AA \times Aa, \\ aa \times Aa \end{cases} y_n \begin{cases} Aa \times Aa, \\ Aa \times Aa \end{cases} z_n \begin{cases} AA \times aa, \\ aa \times aa \end{cases}$$

then  $[w_n, x_n, y_n, z_n]$  is converted into  $[w_{n+1}, x_{n+1}, y_{n+1}, z_{n+1}]$  by the matrix

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{8} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 1 \\ 0 & 0 & \frac{1}{8} & 0 \end{pmatrix}$$

This has three latent roots  $\lambda_1, \lambda_2,$  and  $\lambda_3$ , of which the largest is  $\lambda_1 = .8090$ , so  $w_n = 1 - B_1\lambda_1^n - B_2\lambda_2^n - B_3\lambda_3^n$ , and so on, where  $B_1, B_2, B_3$  depend on the initial conditions. In fact, however, if we consider the possible mating types when each parent can contribute any of four allelomorphic genes, and we do not lump together  $AA \times Aa$  with  $aa \times Aa$ , and so forth, we have to deal with a  $55 \times 55$  matrix, which has 11 latent roots. The corresponding problem for a population of 4 in which incest is avoided by marriage with a double first cousin presents us with a  $222,111 \times 222,111$  matrix! I do not know how many latent roots it has, but Wright (4, 5) by a remarkably ingenious and powerful method, managed to short-circuit the problem and find the numerically largest of the latent roots, which is what matters in the long run. He also solved the problem of what happens with random mating in a cage (or on an island) where there is just room for  $N$  breeders. The leading term in the solution of the corresponding integral equation can also be calculated. But if anyone wishes to solve these problems by matrix methods I shall be glad to show him or her how to construct the matrices if he or she will pay for the necessary paper and ink. Other similar problems involve matrices with algebraical coefficients. A few have been solved.

The simplest non-linear case arises as follows. Suppose  $AA$  and  $Aa$  are equally viable, but that the survival before maturity of  $aa$  individuals is  $1 - k$  ( $1 \gg k \gg -\infty$ ) of that of the others. So a population of  $x AA : y Aa : z aa$  at birth is represented by  $x AA : y Aa : (1-k)z aa$  at maturity. Their fertilities are supposed equal, and mating is at random. This can be shown to be very nearly true in man for many characters. This does not mean that we are promiscuous, but that, for example, the probability that your wife will belong to a given blood group is roughly the frequency of that group in England, and does not depend on the group to which you belong yourself. Random mating can be shown to be equivalent to a pooling of genes in each generation. So if the ratio of  $A$  to  $a$  is  $u_n$ , the  $n$ -th generation consists of

$$u_n^2 AA : 2u_n Aa : 1 aa$$

and the breeders, after selection, of

$$u_n^2 AA : 2u_n Aa : (1 - k) aa$$

It follows at once that

$$u_{n+1} = \frac{u_n(u_n + 1)}{u_n + 1 - k} \quad (9)$$

This equation is immediately soluble in three cases. If  $k = -\infty$ ,  $u_{n+1} = 0$ . If  $k = 1$ ,  $u_{n+1} = u_n + 1$ , or  $u_n = u_0 + n$ . And if  $k$  is very small, we

can write it as a differential equation  $\frac{du_n}{dn} = \frac{ku_n}{u_n + 1}$ , whence

$$kn = u_n - u_0 + \ln \left( \frac{u_n}{u_0} \right) \quad (10)$$

In other cases we have to iterate. The problem can be solved in two ways. Consider the equation

$$x_{n+1} = \phi(x_n) \quad (11)$$

We will transform  $x_n$ , if necessary, so that  $\phi(0) = 0$ , that is to say, 0 is an equilibrium, stable, unstable or mixed. There are four types of solution in the neighbourhood of such a point. For simplicity I shall suppose that  $\phi''(0) \neq 0$ . If  $\phi'(0) = 0$ , then

$$n + P = \ln \left[ -\ln x_n - \sum_{r=1}^{\infty} a_r x_n^r \right] \div \ln 2 \quad (12)$$

If  $0 < |\phi'(0)| < 1$ , then

$$n + P = a_0 \ln x_n + \sum_{r=1}^{\infty} a_r x_n^r \quad (13)$$

If  $|\phi'(0)| = 1$ , then

$$n + P = a_{-1} x_n^{-1} + a_0 \ln x_n + \sum_{r=1}^{\infty} a_r x_n^r \quad (14)$$

If  $|\phi'(0)| > 1$ , then (13) holds again.



Each of these equations can be written in several different ways,  $x_n$  being given in terms of  $n$ . The values of the coefficients  $a_r$  can be found by substitution, and equating coefficients of appropriate powers of  $x_n$ . In the first two cases the equilibrium is stable, in the last unstable. In the third (so long as we confine ourselves to a real variable), it is usually stable on one side and unstable on the other, that is to say effectively stable or unstable if  $x_n$  cannot change sign.  $P$  is an arbitrary periodic of unit period, that is to say a constant if  $n$  is an integer. (12) and (13) have a radius of convergence, (14) is often an asymptotic expansion. These expressions are to be found in the literature, but I know of nowhere where they are collected, and no systematic discussion of them. The only really awkward case is when  $|\phi'(0)| = 1$  and  $\phi'(0)$  is complex but not a root of unity. In this case  $x_n$  takes the bit between its teeth.

Solutions of equation (9) are, near  $u_n = 0$ , if  $k < 0$

$$(n + P)\ln(1 - k) = \ln u_n - u_n + k \left[ \frac{u_n^2}{2(2 - k)} - \frac{(3 - 5k + k^2)u_n^3}{3(2 - k)(3 - 3k + k^2)} + \dots \right] \tag{15}$$

and when  $u_n$  is large and  $k$  positive

$$k(n + P) = u_n + (1 - k) \left[ (1 + \frac{1}{2}k)\ln u_n - \frac{1}{2}k\ln(u_n + 1) + \frac{k^2}{3(u_n + 1)^2} \left\{ 1 + \frac{5(1 - k)}{4(u_n + 1)} + \frac{(25 - 40k + 23k^2)}{80(u_n + 1)^2} + \dots \right\} \right] \tag{16}$$

To find an expression which holds over the whole range of  $u_n$  we expand  $n$  in ascending powers of  $k$ . If  $x_n = \frac{u_n + 1}{u_n}$ ,

$$\Delta x_n \equiv \frac{k(x_n - 1)^2}{x_n}$$

Consider the very general expression

$$\Delta x_n = k \Psi(x_n) \tag{17}$$

$$\text{Then } \frac{dx_n}{dn} = \Delta x_n - \frac{1}{2}\Delta^2 x_n + \frac{1}{3}\Delta^3 x_n \dots$$

$$\begin{aligned} \text{So } n + P &= \int (\Delta x_n - \frac{1}{2}\Delta^2 x_n + \frac{1}{3}\Delta^3 x_n + \dots)^{-1} dx_n \\ &= k^{-1} \int [\Psi(x)]^{-1} [1 + \frac{1}{2}k\Psi'(x) - \frac{1}{12}k^2\{\Psi'(x)\}^2 + \Psi(x)\Psi''(x) + \dots] dx \end{aligned} \tag{18}$$

I have worked out the first six terms of this series, which was, in fact, derived by a more rigorous method, but still needs much investigation. The solution of (9) by means of (18) takes the form

$$n + P = k^{-1} [u_n + (1 - k) \ln(u_n + 1)] + \ln(1 + u_n^{-1}) / \ln(1 - k) - (7u_n + 3)k/12(u_n + 1)^2 + \dots \quad (19)$$

The first two terms are a very good fit. When  $k$  lies between  $\frac{1}{2}$  and  $-1$  the error is under 2%.

But if we have even two parameters  $x_n$  and  $y_n$  in each generation, the corresponding problem is much harder and the solutions obtained are only valid asymptotically.

I have spent much of my life in trying to develop a mathematical theory of natural selection. Equation (9) is a quite artificially simple case, though occasionally a good approximation to the facts. Its solution takes us about as near to an adequate mathematical theory of evolution

as the solution of  $\frac{d^2x}{dt^2} = g$  takes us to an adequate stellar mechanics.

Nevertheless, it may suggest the lines on which such a theory would have to be developed. We have, in fact, got a little further; and perhaps we shall not get much further yet till we have better observations on natural and experimental populations. This will be a slow job for a very simple reason. It is usually very easy to measure a length, angle, weight, or time, with an error of less than 1%. To estimate a frequency of less than  $\frac{1}{2}$  with this precision entails the counting of at least 40,000 individuals if we are to have 95% confidence that our frequency is correct within 1% of its deduced value.

There is, unfortunately, no satisfactory book on the subject. The best is probably Malécot's *Les Mathématiques de l'Hérédité*, which deals with a number of elegantly soluble problems, but is apt to ignore those which so far defy complete solution. I have only skimmed the subject, saying nothing about statistical methods, or about what happens when  $k$  of equation (9) changes in time or space. In the former case we can use Bessel functions when  $k$  changes slowly enough. In the latter we should, I think, need a computing machine unless it changes very abruptly. But I hope that I have shown you that biology raises its own mathematical problems. I think that for a long time the central theme may be sampling theory, because not only do we observe samples of populations, but any population is a sample of the possible progeny of its parents. But if I am right, iteration theory will be almost equally important, and it is perhaps here that mathematicians could be of the greatest help to us. So I end with a challenge. Horns are a recessive character in cattle. If, in a large random mating population, horned bulls are not bred from, but there is no selection in cows, then if  $x_n$  and  $y_n$  are the frequencies of the gene for horns in females and males of the  $n$ -th generation

$$\left. \begin{aligned} x_{n+1} &= \frac{1}{2}(x_n + y_n) \\ y_{n+1} &= \frac{x_n + y_n - 2x_n y_n}{2(1 - x_n y_n)} \end{aligned} \right\} \quad (20)$$

It is not very hard to show that in the long run

$$n + P = 2x_n^{-1} - \log_e x_n - \frac{3}{4}x_n - \frac{1}{3}x_n^2 - \frac{19}{288}x_n^3 + O(x_n^4).$$

but where is the second arbitrary constant, which arises because both  $x_0$  and  $y_0$  can have any values between 0 and 1? Its coefficient tends to zero quicker than any power of  $x_n$ , but that is as far as I can get. Any takers?

## REFERENCES

1. NICHOLSON, A. J. and BAILEY, V. H., 1935, *Proc. Zool. Soc.* 551.
2. VARLEY, G. C., 1947, *Journ. An. Ecol.*, 16, 139.
3. NORTON, H. I. J., 1928, *Proc. Lond. Math. Soc.*, 28, 1.
4. WRIGHT, S., 1921, *Genetics*, 6, 111.
5. WRIGHT, S. 1931, *Genetics*, 16, 97.

