A Biologist's Mathematics

David R. Causton

M.Sc., Ph.D., D.I.C.

Lecturer, Department of Botany & Microbiology,
University College of Wales, Aberystwyth.

Edward Arnold

I

Why Mathematics in Biology?

Probably the most significant event that occurred during the rise of man to pre-eminence, from being merely 'another animal', was the development and use of language. At first, language was only spoken, but it did enable relatively large quantities of information to be communicated from one individual to another. More than that, language also provided a means for controlling and monitoring thought itself. The important point is that language provides concepts which can be manipulated independently of the objects to which they refer, and later it becomes possible to think logically without reference to any particular objects at all. Hence, during the history of man, abstract thought became feasible, and from this beginning arose philosophy.

If we had to select an area of application of language that has been outstandingly successful, we should undoubtedly choose the expression of human emotion. The evidence for this is clear when one considers the achievements of oratory and literature. However, when it comes to conveying scientific information, ordinary language is less successful. This is because it is almost impossible to convey precise meaning, since most words in a language have more than one meaning, even if these meanings differ only marginally. Again, because of these various shades of meaning, a particular word means a slightly different thing to different people. This is fine for the literary use of a language where part of the onus for interpretation lies with the reader or listener, but it is not so good for scientific use, where data and hypotheses must be presented with complete unambiguity.

A long time, perhaps many thousands of years, after the rise of language, a new kind of medium for transferring information from one individual to another was evolving. This was at the time when the first of the ancient civilisations – Egypt and Sumer – were flourishing. These civilizations, besides being the architects and builders of large buildings an other monuments, also developed a calendar. Such achievements required precision: precision in measurement, precision in the subsequent manipul: tion of measurement, and precision in the transmission of such data to other people. All this could be achieved through another kind of language – the language of mathematics.

The mathematical language is just the opposite to ordinary language in that its elements are precise and unambiguous, or at least should be. Every quantity and symbol used can be accurately defined in terms of earlier quantities and symbols already defined. Thus, mathematics is built up on precision and fact, whereas ordinary languages are, to some extent, based on the variability and imprecision of human feelings and emotions.

THE DEVELOPMENT OF A SCIENCE

The study of any science, viewed historically, consists of two main phases. First it is studied almost exclusively from a qualitative point of view; but after an initial period, quantitative methods come to be used increasingly. One of the main reasons for this type of development, from the qualitative to the quantitative, is that a science begins as an observational study, progresses to an experimental, and finally to a theoretical study. At first, phenomena are observed as they occur in nature; later scientific work consists of performing experiments, drawing inferences from the results, and then trying to formulate general laws. By their nature, some sciences have to make the jump from observation to theory without the intervening benefits of experimentation. Astronomy is an obvious example here, and it is remarkable that observational and theoretical astronomy have proceeded side by side for several millennia.

At the present time, physics is the science which is pre-eminent in the use of mathematics. Many physical phenomena are rather less complex than are those of other sciences, and the subject has progressed through the stages of observation and experiment, and has emerged as a theoretical science. This is not to say that observation and experiment are not still carried out in physics; the main point is that physics has reached the stage in which there exists a substantial body of theory, mathematical in nature, which has its origins in observation and experiment. In physics at the present time, the experimental and theoretical sides of the subject are of equal importance.

The phenomena of chemistry are often more complex than those of physics, and this subject has not progressed quite as far as physics on the

theoretical side. Although there have been spectacular advances in theoretical chemistry in recent years, chemistry as a whole is still, at present, somewhat more of an experimental science than is physics.

In biology, the situation is very different. Firstly, biological phenomena are highly complex; secondly, there is almost unlimited scope for pure observation of biological phenomena. Hence it is mainly only in the present century that biology has become an experimental science, whereas experimental work in the physical sciences has been undertaken for several hundred years. As a result, it is only now that 'theoretical biology' is tentatively emerging.

From the above remarks, it would seem that mathematical theory is the ultimate aim in a scientific discipline. This is true; not for its own sake, but because in the last resort the phenomena of nature can only be explained in the precise terms of mathematics. Consider this example from physics.

Observation: a stick immersed in water at an angle to the surface (other than a right-angle) so that part of the stick is out of the water and part submerged; the stick appears bent at the surface of the water.

Experiment: a vessel of water is set up on the laboratory bench, and rays of light are traced through the water for various angles of the incident beam; it is found that at an air–water surface (assume that the vessel is made of very thin glass) the ratio of the sine of the angle of the light beam on the air side of the surface to the sine of the angle of the beam on the water side of the surface is constant, and this constant ratio is called the refractive index.

Theoretical deductions: this experimental result can be used in conjunction with facts gleaned from experiments on other phenomena of light, such as reflection, diffraction, interference, to establish knowledge on the nature of light. For instance, it has been found that the velocity of light in a dense medium is less than in a sparse one. This latter experimental finding coupled with the result of the refraction experiment can be analysed mathematically to show that light travels in a wave form.

For this particular example, there is an obvious relationship between observation, experiment, and theoretical deduction. Such examples can be multiplied many fold. In chemistry, we observe a particular reaction, and we experiment to find out the exact conditions under which the reaction occurs. When we then enquire why this particular reaction occurs and not some other, it is necessary to look to the concepts of physical and theoretical chemistry, both of which are founded on mathematics.

Whether or not all biological phenomena can be explained by the physical sciences, or that ultimately it is found that the property of life is 'something extra', it is already quite evident that the manifestations of 'life' can be explained in terms of the physical sciences, particularly chemistry. Since the physical sciences are based on mathematics, so also, indirectly, are the biological sciences.

In summary, experimental results are usually in a quantitative form, even in biology, and therefore sound theoretical deductions can normally only be made by mathematical analysis. This is why, ultimately, mathematics is indispensable to any science; and so any scientist, whatever his or her speciality, should have an adequate knowledge of mathematics.

BIOLOGY, MATHEMATICS, AND STATISTICS

The mathematical model

From the penultimate paragraph of the previous section, one might infer that the utility of mathematics to the biologist is indirect, arising only after experimental results have been interpreted by the concepts of physical science. This, however, is not so. Mathematics is applied directly to the results of biological observation and experiment in a similar manner to the physical sciences but, because of the complexity of the phenomena, its application is much more difficult.

In the present state of biological knowledge, it is impossible to apply a rigorous mathematical analysis to a biological system, such as may be applied, for example, to an electric circuit. What is done, however, is to construct a **mathematical model** of the phenomenon in which we are interested. Certain **assumptions** about the system have first to be made, and put into mathematical form. These assumptions are based on current knowledge obtained from previous observations and experiments. Next, appropriate **mathematical methods** are applied to the assumptions to achieve an end result which **simulates** the system under study. The simulated result can then be compared with what actually happens. If agreement between the theoretical result and the observed happening is good, then we gain further insight into the process under study; and moreover, we can use the model for predictive purposes. In any science, an ultimate aim is **prediction**. For instance, in an electrical circuit we can predict how the current will change for a given change of resistance, using the mathematical description of the circuit. In a biological system that has been 'described' mathematically by means of a model, predictions of what will happen under certain changes of conditions can also be made. If the results of using the model do not agree with actuality, then one or more of our basic assumptions must be wrong (assuming the absence of mathematical errors!), and so, in a negative sense, our knowledge is still increased. An example of the construction of a very simple mathematical model is given in Chapter 6.

Statistics in biology

There is yet another complication to the would-be user of quantitative methods in biology, and that is **variability**. The phenomenon of variability is not confined to biology, but arises whenever experimental work is undertaken. Even in the physical sciences, repetitions of a single experiment will give slightly different results, e.g. measurement of the refractive index of a substance, or the location of an end point in volumetric analysis. This kind of variability, which is called **experimental variability** or **experimental error**, arises solely because a human being attempts to measure something; the something does not change, but the reactions of the human being during the conduct of the experiment do change.

Experimental variability also occurs in biology, but here it is considerably augmented by the variability inherent in biological material. If we measure the refractive index of a block of glass very carefully, we are safe in asserting that our result *is* the refractive index of this kind of glass, under the conditions of the experiment. On the other hand, if we measure the increase in height of a sunflower plant over one day, we certainly cannot say that this is the growth rate of sunflower plants in general, even under the same experimental conditions. The same plant may have a noticeably different growth rate at an earlier or later stage in its growth; and even if we take two plants which germinated from the same source of seed at the same time, they will almost certainly show different growth rates at any instant, aside from experimental error. So to be able to make any sort of general statement about the growth rate of sunflower plants of a given age and under defined conditions, we have to measure several plants and take an average. This immediately raises the question as to how reliable our result is, and this cannot be answered without employing a branch of mathematics known as **statistics**.

Even when no experiment is involved and one is only trying to summarize observations usefully and build a model from them, a simple mathematical approach may not be very satisfactory because of the variability of biological material. A good illustration is afforded by Example 10.3 on page 186. Read the general description of the situation through, note that a mathematical expression is used to describe the situation, and then carefully read the questions asked, each one of which obviously requires a single numerical answer. Now, without worrying about how the answers were obtained, read the last sentence of each of the three sections, and note that each answer is a precise figure. Bearing in mind that a 'cohort' in this context is a natural stand of similar-aged plants, it is quite obvious that these precise answers are only statements of likely results around which actual results will deviate to a greater or lesser extent. One immediately asks, 'How much deviation can be expected?'. The **deterministic** mathematical model that has been erected to describe the situation in this example cannot answer such a question. If, however, the same model had been set up, but with an added feature – a **probability** structure – then we should be in a position to answer questions like the above. The mathematical model would now be a **stochastic** model; it is much more realistic, and more complex.

Both the mathematics of stochastic models, and of statistical methods for the analysis of experiments, are based upon the same theoretical subject – probability and statistics. It is a branch of applied mathematics in the broad sense, not in the narrow sense that the term 'applied mathematics' is often used to denote applications to physics. Therefore, probability theory and statistical science are based on mathematics, and a good knowledge of the subject is necessary for their study. In this book, we shall not deal with probability and statistics. Our concern will be with such topics in mathematics that are of a general nature, topics that have direct biological relevance, and also those that form a basis for the study of statistical science about which the modern biologist needs to know.

EXAMPLE 10.3

In a cohort of foxglove (***Digitalis purpurea***) plants, the number, y, surviving at time t (measured in months from the emergence of the seedlings) was found to conform to the equation

$$y = 100e^{-0.2310t}.$$

(a) What was the original number of seedlings in the cohort?
(b) What was the half-life of the cohort, i.e. the time when half the original number of individuals had died?
(c) The foxglove is a biennial, germinating in spring and flowering in the summer of the following year. Assuming that 15 months are required for vegetative growth, how many individuals of the cohort would be likely to survive to flower?
(a) Since time is measured from seedling emergence, we put $t = 0$; then, from the equation, $y = 100$. So the original number of seedlings was 100.
(b) To calculate the time at which half the number of seedlings have died, we substitute the number surviving into the equation for y and solve for t. In this case, half the original number surviving gives $y = 50$, and so

$$50 = 100e^{-0.2310t}$$

or

$$0.5 = e^{-0.2310t}$$

Taking the reciprocal of both sides gives

$$e^{0.2310t} = 2$$

Take natural logarithms of both sides

$$0.2310t = \log_e 2$$

Now

$$\log_e 2 = 0.6931,$$

and so

$$t = \frac{0.6931}{0.2310} \simeq 3$$

Therefore 50% of the plants die during the first three months from seedling emergence.
(c) As we are told that 15 months are required for vegetative growth, we merely substitute $t = 15$ into the equation and solve for y.

$$y = 100e^{-0.2310 \times 15}$$

i.e.

$$y = 100e^{-3.465}$$

or

$$y = 100/e^{3.465}$$

Now $e^{3.465} = 31.98$, and so

$$y \simeq 3.127$$

Hence, 3 plants are likely to survive to flower.

It should be noted that an exponential function, such as the one shown in the above example, would be fitted to data in its linear form, equation 10.16. In the foxglove cohort, observations on the number of individuals surviving would be made from time to time, and then a graph would be plotted of $\log_e$ (number of individuals) against time. The points should lie roughly on a straight line, and such a line may then be fitted in the way mentioned in Chapter 3 (page 51) and values of the constants $\log_e a$ (and hence a) and b obtained.