BASIC MATHEMATICS
FOR THE
BIOLOGICAL AND SOCIAL SCIENCES

by

F. H. C. MARRIOTT
Department of Biomathematics, University of Oxford

PERGAMON PRESS
Oxford · London · Edinburgh · New York
Toronto · Sydney · Paris · Braunschweig
1966

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1
Pergamon Press (Scotland) Ltd., 2 & 3 Teviot Place, Edinburgh 1
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523
Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1
Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
Pergamon Press S.A.R.L., 24 rue des Écoles, Paris 5[e]
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

First edition 1970

Library of Congress Catalog Card No. 73–99863

Printed in Great Britain by A. Wheaton & Co., Exeter

08 006664 X (flexicover)
08 006663 1 (hard cover)

To my wife
but for whom
this book would have been written long ago

PREFACE

THE biological and social sciences are becoming more and more mathematical. Statistical methods in agriculture, diffusion in cells, the study of biological control mechanisms, mathematical models in economics, the application of information theory to neurophysiology, population models—these are all examples of recently developed methods that demand considerable mathematical skill.

Traditionally, students in these subjects do not receive much mathematical training. Textbooks of applied mathematics are rather strongly biased towards applications in mechanics, engineering and some branches of physics, while theoretical books are more concerned with the logical foundations of the subject than with its value as a tool.

This book is intended to fill the gap. It starts, after some preliminaries, with the introduction of the infinitesimal calculus, and goes on to deal with scalar and vector quantities, complex numbers and the simplest types of differential equation. Statistical methods are not discussed; there are several good texts available, and it would not be possible to do justice to the subject in a single chapter. The examples—apart from those that are simply exercises—are taken from biology, economics and related subjects, or from probability theory and physics. They differ considerably in difficulty, and those marked * are probably better left to a second reading.

ACKNOWLEDGMENTS

I AM grateful to Mrs. K. Earnshaw for typing a difficult manuscript, to my wife for help with the figures, and to Dr. J. F. Bithell for reading the proofs.

I must particularly acknowledge the help of Mr. W. J. Langford, whose careful and patient criticism has enabled me to improve the book in many respects and to remove many mistakes.

CHAPTER 1

BASIC ALGEBRA

1.1. Symbols and notation

Most of the symbols introduced in this book will be discussed when they are first needed. Some, however, are so important and save so much writing that they are worth explaining at the beginning.

Factorials

The product $1.2.3 \ldots (n-1).n$, where n is a whole number, is called *factorial* n, and written $n!$ It represents the number of ways of arranging n different objects—the number of *permutations* of n objects. There are n ways of picking out the first; when that has been chosen, there are $n-1$ left, and so $n-1$ ways of picking out the second, and so on. For example, two things can be arranged in two ways, three in six ways, four in twenty-four ways. Factorial n increases very rapidly as n increases; $5! = 120$, $6! = 720$, and $10! = 3628800$.

Notice that $n! = n.(n-1)!$ when $n \geqslant 2$. This is true also if $n = 1$, if conventionally $0!$ is taken equal to 1.

The sigma notation

An abbreviation for the sum of a series of terms of the same sort is often needed. The Greek capital sigma, Σ, is used to represent "the sum of terms like", and the highest and lowest values are indicated above and below the Σ sign.

The geometrical series $a + ar + ar^2, \ldots, ar^n$ may be written

$$a \sum_{i=0}^{n} r^i.$$

The index i in this expression is just a dummy, taking in succession the values 0,1,2 to n, and the Σ indicates that all these values are added together.

Functions

When two variable numbers, represented by x and y, are related by an equation that gives the value of y for each value of x in some range, y is said to be a *function* of x. For example:

$$\begin{aligned} y &= x^2, & -\infty < x < \infty, \\ y &= \log_{10}x, & 0 < x < \infty, \\ y &= \sqrt{(1 - x^2)}, & -1 \leqslant x \leqslant 1. \end{aligned}$$

In each case, the value of y is defined for each value of x within the range specified. For $y = x^2$, this range extends over all positive and negative values of x. The value of $\log_{10}x$ is defined only for positive values of x. This is also true of the square root, and so $\sqrt{(1 - x^2)}$ is defined only when $x^2 \leqslant 1$.

For each of these functions, the value of x gives the value of y uniquely—at least, if the positive value of the square root is taken. They are therefore called *single-valued functions* of x. If the last function had been defined by $y^2 = 1 - x^2$, so that $y = \pm\sqrt{(1 - x^2)}$, each value of x would have given two values of y. They are also *continuous*; that is, there is no sudden jump in the value of y when x changes by a small amount. It is possible to define discontinuous functions, for example

$$\left\{\begin{array}{ll} y = -1 & -\infty < x < 0 \\ y = 1 & 0 \leqslant x < \infty \end{array}\right\}.$$

Functions of this sort are not of much interest in practical applications, though they are important in mathematical theory.

It is convenient to have a general notation for "y is a function of x" in this sense. Writing $y = f(x)$, or any similar expression ($y = g(x)$, $y = U(x)$, $y = \phi(x)$, etc.) means that for each value of x, in some range, the function $f(x)$ will define one or more values of y. It will generally be assumed that each value of x defines just one value of y, so that $f(x)$ is a single-valued function of x. The notation can be generalized, writing, for example, $z = f(x,y)$ to mean that the value of z is given when x and y are both known; and writing $f(x,y) = 0$ when x and y are related, but not necessarily by an explicit function $y = f(x)$ or $x = g(y)$.

The modulus sign $|x|$ is used to indicate the positive value of x; $|x| = x$ if $x \geqslant 0$ and $|x| = -x$ if $x \leqslant 0$. The notation will be given an extended meaning in connexion with vectors and complex numbers.

1.2. The binomial theorem

It is easy to see, by direct multiplication, that

$$(1 + x)^2 = 1 + 2x + x^2,$$
$$(1 + x)^3 = 1 + 3x + 3x^2 + x^3,$$
$$(1 + x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4.$$

The binomial theorem is simply a method of writing down a general expression for $(1 + x)^n$ in powers of x, without having to multiply it out every time.

Consider now the product of n factors. The form of the expansion is clearer in the more general case:

$$(a_1 + x)(a_2 + x)(a_3 + x) \ldots (a_n + x).$$

When this expression is multiplied out, the coefficient of x^r consists of terms involving the product of $n - r$ of the a's; for example, there is a term $a_1a_2a_3, \ldots, a_{n-r}x^r$, when the x^r is derived from the last r factors, and the constant from the first $n - r$. In fact, the coefficient of x^r is the sum of *all possible* products of $n - r$ chosen from the n a's. Thus if $n = 5$, the coefficient of x^3 in

$(a_1 + x)(a_2 + x)(a_3 + x)(a_4 + x)(a_5 + x)$ is $a_1a_2 + a_1a_3 + a_1a_4 + a_1a_5 + a_2a_3 + a_2a_4 + a_2a_5 + a_3a_4 + a_3a_5 + a_4a_5$.

The number of products, each of $n - r$ a's, in the coefficient of x^r is the *number of ways of choosing $n - r$ items from n*. Since each group of $n - r$ chosen corresponds to r left out, this is equally the *number of ways of choosing r items from n*. Now consider the number of ways of choosing *r in a particular order*. There are n ways of choosing the first, $n - 1$ ways of choosing the second from the remaining $n - 1$, $n - 2$ ways of choosing the third, and so on, until the rth is chosen from the remaining $n - r + 1$ items. The total number is $n(n - 1) \ldots (n - r + 1)$. This includes all possible *arrangements* of the r items, for example a_1a_2 as well as a_2a_1. To get the number of ways of choosing r from n regardless of order, it is necessary to divide by the number of ways of arranging r items, that is $r(r - 1)(r - 2) \ldots 1 = r!$ Hence the number of ways of choosing r items from n is

$$\frac{n(n-1)\ldots(n-r+1)}{r!}.$$

This expression is so important that it is useful to introduce a special notation, and write

$$\binom{n}{r} = \frac{n(n-1)\ldots(n-r+1)}{r!}.$$

Notice that this can also be written

$$\frac{n!}{r!(n-r)!}.$$

This brings out the symmetry of the expression; clearly if r and $n - r$ are interchanged, the value of $\binom{n}{r}$ is unaltered, so that $\binom{n}{r} = \binom{n}{n-r}$. The first definition of $\binom{n}{r}$ is more general, however. It gives $\binom{n}{r}$ for *any* non-zero value of n, provided r is a positive integer (or zero), while $n!/\{r!(n-r)!\}$ has a meaning only when n is a positive integer. The importance of this will be seen in Chapter 7.

Now the coefficient of x^r in $(a_1 + x)(a_2 + x) \dots (a_n + x)$ is the sum of $\binom{n}{r}$ terms like $a_1a_2, \dots, a_{n-r}$. For example, if $n = 5$, $r = 3$ the number of terms is $\dfrac{5!}{3!\,2!} = 10$, the ten terms listed above. If now $a_1 = a_2 = \dots = a_n = 1$, each term is 1, and the coefficient of x^r is $\binom{n}{r}$, so we may write

$$(1 + x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r. \tag{1}$$

This is the *binomial theorem*, and the coefficients $\binom{n}{r}$ are called *binomial coefficients*. A rather more general form is

$$(x + y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}. \tag{2}$$

Notice that if $r > n$, $n(n-1)\dots(n-r+1) = 0$. If n were not a positive integer, this expression would never become zero, and the expansion would continue indefinitely. It will be shown in Chapter 7 that this infinite series is, in some circumstances, a valid representation of $(1 + x)^n$.

An application to probability theory

An interesting application of the binomial distribution in probability theory gives the probability of r events in n independent trials when the probability of an event in a single trial is p. This probability may be written

$$P(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r} \tag{3}$$

where $q = 1 - p$ is the probability of no event in a single trial. The result is easily proved. The probability that the first r trials are successful and the following $n - r$ unsuccessful is $p^r q^{n-r}$, and the same is true for any *specified* arrangement of r successes and $n - r$ failures. The number of possible ways of choosing r trials out of the n as successes is

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

so the total probability $P(r)$ has the value given.

It is, of course, certain that r takes one of the values $0,1 \ldots n$, so that

$$\sum_{r=0}^{n} P(r) = 1.$$

But

$$\sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} = (p + q)^n$$

(using (2)), and this is unity as $p + q = 1$, confirming the result.

For example, suppose the probability that a child is a boy or girl is $\frac{1}{2}$, and that this probability is unrelated to the sex of previous children or to any genetical characteristics of the parents. Then the probabilities of 0,1,2,3,4 boys in a family of 4 are given by the terms of the binomial expansion of $(\frac{1}{2} + \frac{1}{2})^4$, that is $\frac{1}{16}$, $\frac{4}{16}$, $\frac{6}{16}$, $\frac{4}{16}$, $\frac{1}{16}$. They add up, of course, to 1.

1.3. Partial fractions

The method of combining fractions by putting them over a common denominator is familiar, for example

$$\frac{2}{x+3} - \frac{3}{2x+5} = \frac{2(2x+5) - 3(x+3)}{(x+3)(2x+5)} = \frac{x+1}{2x^2+11x+15}.$$

It is often useful to break up an expression like that on the right, with a polynomial denominator, into a sum or difference of simpler fractions like those on the left. This is always possible if the denominator can be split up into linear factors. To do this example in reverse, assume

$$\frac{x+1}{(x+3)(2x+5)} = \frac{A}{x+3} + \frac{B}{2x+5}.$$

Now this is not an equation in x; A and B must be chosen to make it an algebraic identity. In other words, the coefficient of x and the constant term in $A(2x+5) + B(x+3)$ must both be 1, since this expression must be precisely $x + 1$. This gives two simultaneous equations for A and B,

$$2A + B = 1,$$
$$5A + 3B = 1$$

and the solution of these equations gives $A = 2$, $B = -3$.

In this case, the constants A and B can be evaluated more simply by putting in particular values of x to make the multipliers of B and A, in turn, zero. Since $A(2x+5) + B(x+3) = x + 1$ is an identity, true for all values of x, it must be true for $x = -3$. Putting this value in the equation gives at once $A(-1) = -2$, or $A = 2$. Similarly, putting $x = -\frac{5}{2}$ gives $B(\frac{1}{2}) = -\frac{3}{2}$, $B = -3$.

This method is quite generally applicable, but there are two slight complications. In the first place, if the numerator involves higher powers of x than the denominator—or as high—they must be removed first by division. For example:

$$\frac{x^3 + x^2 - x + 1}{x^2 - 1} = \frac{x(x^2-1) + (x^2-1) + 2}{x^2 - 1}$$

$$= x + 1 + \frac{2}{(x+1)(x-1)}.$$

Now assume the last fraction is equal to $A/(x+1) + B/(x-1)$. Then $A + B = 0$, $B - A = 2$, so $A = -1$, $B = 1$ and the fraction finally becomes $x + 1 + 1/(x-1) - 1/(x+1)$.

Secondly, it may not be possible to factorize the denominator into linear factors. The expression can then still be expressed in partial fractions, but the numerators will not necessarily be constants. For example, if

$$\frac{2}{(x-1)(x^2+1)} \text{ is written as } \frac{A}{x-1} + \frac{B}{x^2+1},$$

A and B must satisfy the equations $A = 0$, $B = 0$, $A - B = 2$, which is impossible. But if it is written

$$\frac{A}{x-1} + \frac{Bx+C}{x^2+1},$$

then $A + B = 0$, $C - B = 0$, $A - C = 2$ so that $A = 1$, $B = -1$, $C = -1$, and

$$\frac{2}{(x-1)(x^2+1)} = \frac{1}{x-1} - \frac{x+1}{x^2+1}.$$

When a quadratic factor is present, the corresponding numerator must be assumed to be a linear function.

If there is a repeated factor in the denominator, say $(x + a)^3$, the result can be expressed in partial fractions in the form

$$\frac{A}{x+a} + \frac{B}{(x+a)^2} + \frac{C}{(x+a)^3}.$$

This gives the right number of equations for A, B and C. For example:

$$\frac{1}{(x-1)(x+1)^2} = \frac{A}{x-1} + \frac{B}{x+1} + \frac{C}{(x+1)^2}.$$

Equating coefficients:
$$A + B = 0,$$
$$2A + C = 0,$$
$$A - B - C = 1.$$

Hence $A = \frac{1}{4}$, $B = -\frac{1}{4}$, $C = -\frac{1}{2}$, and the expansion becomes:

$$\frac{1}{(x-1)(x+1)^2} = \frac{1}{4}\left\{\frac{1}{x-1} - \frac{1}{x+1} - \frac{2}{(x+1)^2}\right\}.$$

Hence any rational function—that is, a ratio of two polynomials—can be resolved into a sum of terms of the form:

(a) powers of x and a constant term,
(b) simple fractions like $A/(x + a)$,
(c) fractions of the form $B/(x + a)^r$,
(d) fractions of the form $(Cx + D)/(x^2 + bx + c)$,
(e) fractions of the form $(Ex + F)/(x^2 + bx + c)^r$,

provided only that the denominator can be expressed as a product of linear terms, quadratic terms and linear and quadratic terms raised to a power. This is always possible in theory, but may be complicated in practice.

Examples

1. Write out in full $\sum_{r=0}^{5} \frac{x^r}{r!}$.

2. Sketch the graph of $y = |x - 5|$.

3. Expand $(2x + 3)^3$ by the binomial theorem, and check the result by direct multiplication.

4. Write down the first four terms of the binomial expansion of $(1 - x)^{-1}$. [Note that $n!$ is not defined when n is negative, but it is easy to use the binomial coefficients in the form $n(n - 1) \ldots (n - r + 1)/r!$.] Confirm that these terms give a good approximation to $1/(1 - x)$ when x is small.

In examples 5–10 resolve the expression into partial fractions:

5. $\frac{1}{x^2 - 4}$.

6. $\frac{x^5 + 2x^2 - 4x - 6}{x^2 - 4}$.

7. $\frac{x}{x^4 + 1}$.

8. $\frac{x^2 + 6x + 3}{(x + 1)^2}$.

9. $\frac{1}{(x - a)(x - b)}$.

10. $\frac{x^3 + x^2 + 2x + 3}{(x^2 + 1)^2}$.

11. Show that $(x + y + z)^n$

$$= \sum_{p=0}^{n} \sum_{q=0}^{n-p} \frac{n!}{p!q!(n-p-q)!} x^p y^q z^{n-p-q}.$$

[This is the extension of the binomial theorem to three components. It is easily proved by the method used to prove the binomial theorem.]

12. According to genetical theory, if each parent carries both dominant and recessive genes of a characteristic, one-quarter of the offspring should be of the recessive type. Calculate the probability that 0,1,2,3,4 and 5 of a family of 5 are of the recessive type.

13. A sum of money is invested at 2% compound interest for 10 years. Show, from the binomial theorem, that it is increased by a factor 1·2190, and check the answer using logarithms. (Note that when x is small, the first few terms in the expansion of $(1 + x)^n$ give a very good approximation.)

14. Show that if the largest term in the expansion of $(x + y)^n$ is the term in $x^r y^{n-r}$, then $r/(n - r + 1) \leqslant x/y \leqslant (r + 1)/(n - r)$. Show that $r = nx/(x + y)$ satisfies these inequalities.

[Note that this is not necessarily a whole number, so it gives only an approximation to the value of r giving the largest term. The true value is the integer on one side or the other of this number. There may be two equal largest terms, so one of the inequalities—but not both—may be an equality.]

15. There are N fish in a lake. Of these, n_1 are caught, marked and released. A little later, n_2 more are caught, and of these r are found to be marked. Show that the most probable value of r is approximately n_1n_2/N, so that the value of N may be estimated as n_1n_2/r.

[This capture–recapture technique is widely used to estimate natural populations. It depends rather critically on the assumptions (i) that the marked animals intermingle completely with the rest of the population, and (ii) that they are neither easier nor harder to catch.]

CHAPTER 2

GRAPHICAL METHODS

2.1. Introduction

This chapter falls into two sections. The first deals with the graphical presentation of data, whether in the form of diagrams in the text of a paper or slides to accompany spoken presentation. The second part is concerned with the graphical representation of mathematical relationships. This is the subject known as analytical geometry, but here only the most obvious features will be discussed. The aim is to be able to answer the question "If I plot this curve, what will it look like?"

The two parts are, of course, closely linked. On the one hand, a set of experimental results may suggest that two quantities are related, and it may be useful to fit some purely empirical curve to represent the relationship, or perhaps to predict one from the other. Alternatively, theoretical considerations may suggest a mathematical relationship, and the question is whether experimental results agree with this relationship and what values of any constant parameters in the equation give the best fit.

At the end of the chapter, there is some discussion of three-dimensional coordinates. These are important in fields as widely separated as biochemistry, microscopy and forestry; in all these subjects it is valuable to visualize solid objects, and what will happen if sections are cut through them in various ways. A more theoretical aspect is that, just as a functional relationship between two variables can be represented by a curve in two dimensions, so a relationship between three variables corresponds to a surface in space. It is helpful to be able to visualize the features of surfaces in three (or even more!) dimensions.

2.2. The graphical presentation of data

A diagram can often convey the most important features of a set of data more quickly and more effectively than tables of numerical values and desciiptions of their properties. Figure 2.1 is an example. It is easy to see that in the lower curve there is a relationship between the variables, that it is roughly linear for

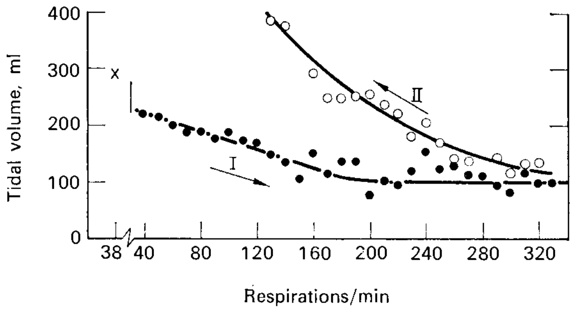

FIG. 2.1. [*J. Physiol.* **190**, 247, fig. 4, lower section.] See text.

most of the range but seems to flatten for the largest values, and that there is considerable scatter about this curve. The second curve is similar in general form, but is steeper and rises to higher values. The figure shows this much more clearly than a table of values would do, and much more clearly than the description.

A diagram is designed to convey a message effectively without either confusing or misleading. Many diagrams lose much of their point because they are too small, too crowded with symbols, or have scales badly chosen. It is notoriously easy to produce quite different impressions by plotting data in different ways, and graphs can often be used, deliberately or accidentally, to suggest quite false conclusions.

A few points should be observed in presenting any results graphically:

(i) Use the scales intelligently. Most types of graph paper have scales divided into tenths. The larger divisions can conveniently represent 2, 5, or 10 or even 4 units, but division by 3 or 7 is inconvenient and should be avoided if possible. The results should not all lie in one part of the graph. It is not always necessary to extend the scales to zero, and there is usually no objection to "cutting off the bottom" of a graph (but this can be misleading in that it suggests bigger fluctuations in the figures than have actually occurred—see Huff (1954)). If two sets of data are intended, or are likely, to be compared, it is essential that the scales should be the same.

(ii) Do not try to demonstrate too many features in the same diagram. It is all too easy to pepper a graph with symbols, using different types to represent different classes of data, and fitting different curves to each set. The result is usually to obscure the point the graph is intended to illustrate. Often it is better to show mean values rather than all the points—a graph is an illustration, not a statistical proof. Simplicity is particularly important in graphs shown as lantern slides, which are usually only visible for a very short time.

(iii) Any symbols should be large enough to be seen clearly. Most graphs are drawn to be reproduced, and the processes, either of printing or making a slide for projection, tend to make things less clear. Further, the actual size is usually changed. A diagram that looks perfectly clear on the drawing board may be unintelligible when printed, scaled down to quarter-size, or viewed from the back of a lecture theatre when projected slightly out of focus. So use large, bold, symbols—and if they crowd into each other, there are probably too many of them.

(iv) If anyone is likely to study your results, to do statistical or other calculations on them, give them in a table as well as graphically. Many editors do not like this, but anyone who has tried to read values from a published diagram, usually with the squares of the graph paper suppressed and the scales barely indicated, will appreciate the point.

(v) It is often difficult to decide whether to fit a theoretical curve to data points, to join them up, to draw a rough curve through them by eye, or to leave them alone. If you do fit a curve, indicate how it was done. In particular, if some theoretical relationship is suggested, it is important to make it clear whether the

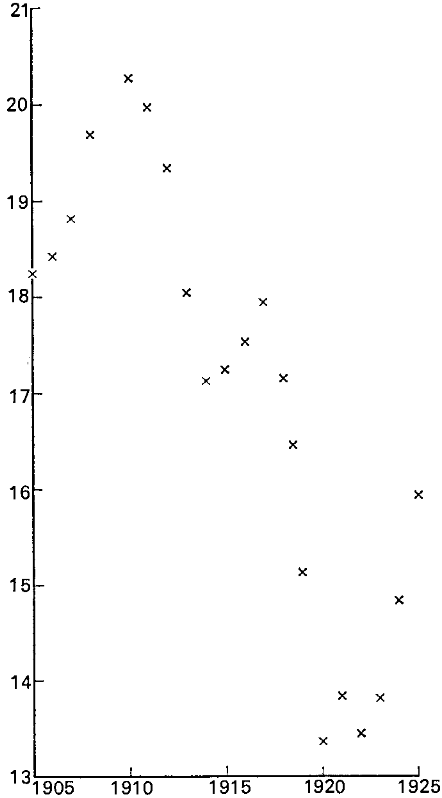

FIG. 2.2. The sheep population of England and Wales (millions) between 1905 and 1925. Replot these figures, (a) extending the ordinate scale to include zero, and (b) doubling the abscissa scale and halving the ordinate scale. Note the differences in the appearance of the graphs, and note how much easier it would be to replot from a table of numerical values.

line drawn on the graph represents that relationship or is merely an empirical curve drawn through the points.

Figures 2.2 and 2.3 illustrate some of these rules.

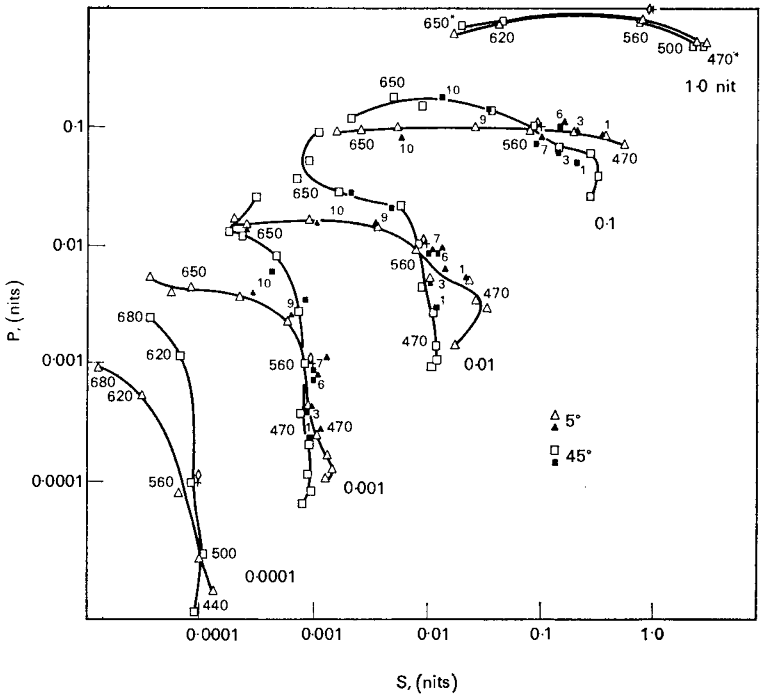

FIG. 2.3. [*Vision Res.* 7, 625, fig. 3.] An over-elaborate graph. The context would perhaps make it more comprehensible, but there are far too many curves and symbols shown on the same figure.

2.3. Special types of graph

Changes of scale

When x and y represent corresponding measurements of some sort, it is natural to plot these measurements one against the other, taking y as the ordinate and x as the abscissa. There is no reason,