Spatial analysis in archaeology

IAN HODDER AND
CLIVE ORTON

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON · NEW YORK · MELBOURNE

Published by the Syndics of the Cambridge University Press
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
Bentley House, 200 Euston Road, London NW1 2DB
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1976

Text set in 10/12 pt Monotype Baskerville, printed by letterpress, and bound in Great Britain at
The Pitman Press, Bath

Library of Congress Cataloguing in Publication Data

Hodder, Ian.

Spatial analysis in archaeology.

(New studies in archaeology; 1)

Includes bibliographical references and index.

1. Archaeology–Methodology. 2. Archaeology–Statistical methods. I. Orton, Clive, 1948–
II. Title. III. Series.
CC75.H57 930'.1'0285 75-44582
ISBN 0 521 21080 1

PREFACE

This book consists largely of the results obtained from research in Cambridge by Ian Hodder from 1971–4. Here much encouragement was received from David Clarke at Peterhouse and from a number of friends doing research at Cambridge in the same period. Great stimulus was obtained through contact with Dr A. Cliff, Department of Geography, Cambridge, without whose time and attention the work would have been impossible. He also wrote some of the computer programmes used in the study. Ian Robertson helped in the solving of some statistical and computational problems, and Ron Martin advised on the model described on page 183. Anything of merit learned as an undergraduate must derive from the staff of the Institute of Archaeology in London in the period from 1968 to 1971, in particular Professor J. Evans and Professor F. R. Hodson.

Clive Orton, until recently in the statistical department of the Ministry for Agriculture, Fisheries and Food, and the Southwark Archaeological Excavation Committee, was encouraged to apply his statistical attention to the problems of archaeology by Professor F. R. Hodson and Professor D. G. Kendall, and this book forms part of the result.

We should like to thank Dr M. Rowlands (middle Bronze Age palstave data) and Dr M. Fulford (Romano-British fine pottery data) for allowing us to use unpublished data, and Dr R. Reece for help with his Roman coin data. Dr J. Alexander drew the work of R. H. Atkin to our attention, Dr R. Coleman provided some useful comments on edge-effects in nearest-neighbour analysis, and A. E. Brown very kindly allowed us to try out our ideas on his extra-mural students. Mrs P. M. E. Altham read and commented on parts of the manuscript. Mrs Jackson helped with some of the typing. Sincerest thanks go to Françoise Hivernel for her time and patience both in typing and in many other ways.

January 1975

I.H.
C.R.O.

1
Introduction

1.1 *General introduction*

The main aim of this work is to suggest to archaeologists that there is a potential for more detailed and systematic study of spatial patterning in archaeological data. The distribution map lies behind some of the most central themes in archaeology such as trade, diffusion and culture. The map of archaeological data also has a value for chronology. For example, Clark (1957) has suggested that the degree of overlap between distributions of different cultural assemblages can give an indication of their contemporaneity (cf. Willey and Phillips 1958, 32). 'The distribution map is one of the main instruments of archaeological research and exposition, but because it is a commonplace of books and papers, do not let us forget what it is trying to do – to accomplish and to demonstrate the totality of information about some archaeological fact, to study the total evidence in space regarding one aspect of the material remains of the past' (Daniel 1962, 80). 'For the past thirty or forty years archaeological distribution maps have been one of the main weapons in the armoury of the prehistorian' (Clark 1957, 153). However, the development of spatial studies in archaeology has been slow. Early prehistorians were mainly concerned with establishing chronological sequences and they did not always concern themselves with the geographical extent of the cultures they examined. 'It was because of this that archaeological mapping made little headway until well into the twentieth century. . . . It was not until 1912 that Crawford first used distribution maps to argue questions of cultural history' (Clark *ibid.*). But it is only in the last few years that systematic methods for the examination of archaeological maps have begun to be used. Because of this past neglect, most of the methods which have been used recently in archaeology and which are employed in the study presented here have been introduced and adapted from other disciplines, in particular geography and plant ecology. 'The distribution of artifacts in space is only now, through the application of locational analysis, undergoing systematic study. All the notions of random and regular spacing, of central place theory and settlement hierarchy, and of correlations among distributions, have yet to be assimilated to prehistoric archaeology' (Renfrew 1973*b*, 250). This work is an attempt to develop such systematic study, although the field is very wide and one cannot hope to cover all aspects nor to solve all the associated problems.

An appraisal of the role of spatial studies in archaeology is felt necessary for three reasons. The first is that previous work in this field has been limited in its aims and methods which were often uncritical and did not aid detailed interpretation. Second, subjective assessments of distributions can be dangerous, and third, some methods are needed to handle the large amounts of distributional information that are now becoming available.

As an illustration of the first point and as an example of the early approach to distribution maps in archaeology, Fox's study of *The personality of Britain* (first edition 1932, revised 1943) may be considered. Fox's aim in this study was as follows. 'I shall endeavour to express the character of Britain in prehistoric and early historic ages, and to indicate the effect of the environment she afforded on the distribution and fates of her inhabitants and her invaders' (Fox 1943, 10). 'The most convenient line of approach is to find out, by the study of distribution maps, where in this island early Man actually lived and laboured' (*ibid.*, 11). 'The essay is concerned to establish principles and not to present the prehistory of Britain' (*ibid.*, 14).

The method employed to achieve this aim was the visual interpretation of a large collection of distribution maps of different periods. General similarities between the distributions were sought. A major difference, for example, was noted between distributions with a western (for example megalithic monuments) and an eastern (for example Beakers) bias. Such differences then had to be explained and interpreted. 'The first question that arises is how these different distributions are to be explained. Must they be considered in isolation, or are there underlying and constant factors to be taken into account in framing any rational explanation of them? In these pages the existence of such dominating factors will be made manifest' (*ibid.*, 14). It was found that 'geographical position and form suffice to explain, in large measure, the two chief variations in distribution' (*ibid.*, 15). Thus, southeast England is well situated for contact with and influences from neighbouring parts of Europe, while the western British Isles absorbed influences moving along the Atlantic routes. The physiography of Britain also played a part, with the western and northern highland zone being markedly different from the lowland zone. 'It is easy to understand why the major physical factors should exert so powerful an influence on distributions. Lowland country, with its insignificant hills and easy contours, is more easily overrun by invaders than highland. The difficulties which mountainous country presents to an invader are well known; moreover, the highlander lives a harder life and is less easily conquered, still less easily displaced, than the lowlander' (*ibid.*, 33). Differences in climate and the impact of these on man's economy were also seen to affect distributions. For example, 'the distribution of the "damp" oak woodland and Man's dislike for it explains many curious features in the prehistoric maps' (*ibid.*, 58).

In modern times, the aims of Fox's study seem limited. In pursuit of his desire to establish general principles, 'a given distributional situation may be, as here, expressed in the simplest terms available, stripped of those complexities which make the pattern of human life and activity so interesting, and which it is the business of the prehistorian and historian to elucidate' (*ibid.*, 14). 'It is true that some of the massed maps show, by a variety of symbols, the diversity of material which goes to build a culture pattern; but usually it is the resultant pattern only which is relevant to my purpose' (*ibid.*, 14). In addition, the visual methods employed often seem uncritical. There is no detailed examination of the degree of correlation between maps nor of whether a distribution indicates patterns of site destruction or fieldwork intensity, an invasion, trade or social contact. An invasion hypothesis is usually preferred although there is little discussion as to why this should be so.

The inadequacy of early studies of distributions of artifacts can also be seen in the common aim to establish prehistoric trade routes. An example is the work of Sprockhoff (1930) discussed by Stjernquist (1966, 8–9). Sprockhoff's network of trade routes for the Bronze Age is shown in fig. 1.1. Two methods were combined to produce this map. One was to map out finds of imported goods and finds from hoards. These hoards were not discussed in detail but were assumed to be trade hoards. This method was combined with one in which the author used verifiable stretches of mediaeval trade routes. An assumption was made that the network of trade routes had not changed much during the period from the Bronze Age to the Middle Ages. De Navarro (1925) also constructed 'trans-continental trade routes'. In his case these were derived from the distribution of amber finds. In these and other studies

> none of the scholars engaged on the problems concerning trade routes has tried to analyse the distribution maps in greater detail. The trade routes have been drawn with the intention of indicating the principal direction of the flow of goods. There is no detailed analysis of the economic and topographical conditions, although the topographical aspects have been taken into account in studies of limited areas. The conclusions drawn, however, presuppose the opinion that places where finds of imported objects have been made mark a trade route. There is an underlying hazy conception of what this means and from the literature one can see that there is a confusion in thought between the views on the goods as evidence of a trade route and the goods as evidence of a market [Stjernquist 1966, 14].

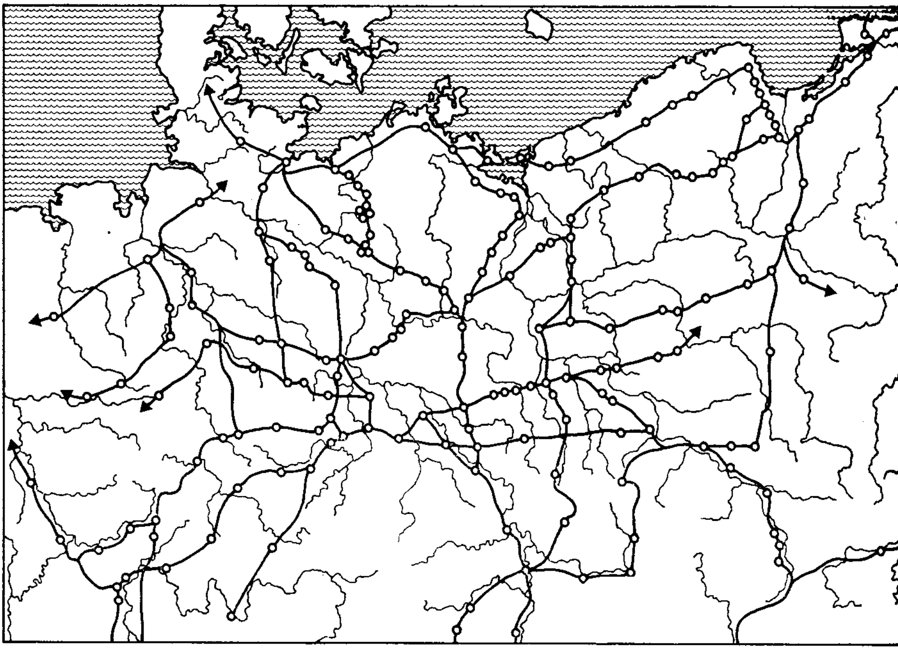

Fig. 1.1. Trade routes for the Bronze Age (after Sprockhoff). Source: Stjernquist 1966.

A second reason for developing spatial studies in archaeology results from the subjectivity involved in map interpretation. 'It has been shown that the ability of the map-user to discriminate and evaluate the information contained in the map is not free from subjective elements and that the more information contained in a map the more ambiguity and uncertainty there is likely to be as regards the interpretation to be put upon it' (Harvey 1969, 377). It is possible, however, to measure some aspects of map information and to develop more rigorous methods of map interpretation. 'Until the recent introduction of a statistical definition of spatial uniformity based on nearest neighbour analysis, . . . it was difficult rigorously to measure dot-patterns. The more traditional "eye-ball" methods are not really satisfactory' (Garner 1967, 310).

The subjectivity of map interpretation may not be immediately apparent and it is perhaps worth providing some examples (see also p. 31). In fig. 1.2 to 1.5 points have been allocated at random to a bounded area (point co-ordinates obtained from a random numbers table). By allowing a certain flexibility in the approach to these distributions we can identify structure even though the pattern is random. For example, if the points are considered as sites, in fig. 1.2 circles can be drawn around certain of the site points (cf. the approach followed by Stanford 1972). Some pairing and clustering might be

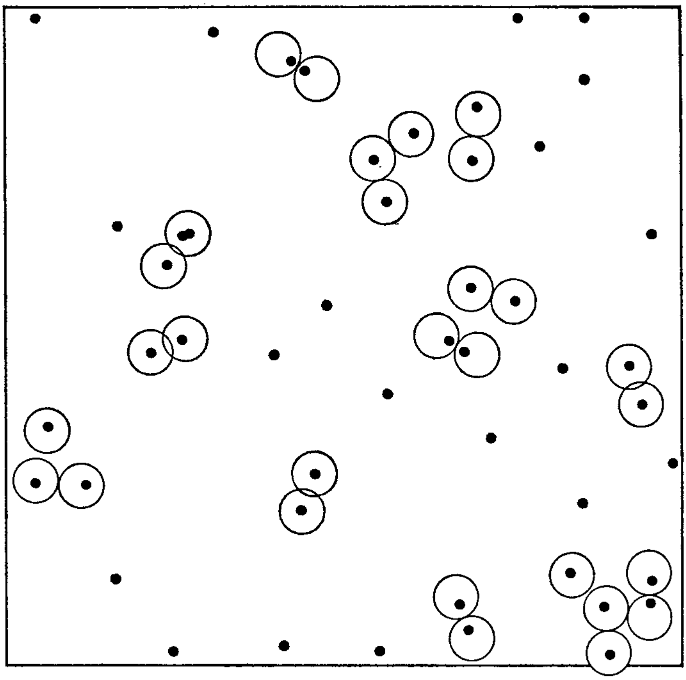

Fig. 1.2. Points placed at random in a bounded area. Circles suggest spheres of influence around 'sites'.

suggested, with an otherwise fairly even scatter of single sites. Depending on the context, a number of hypotheses could be put forward to explain this 'structured' pattern. For example, the pairings and clusters might be seen to indicate those sites which have moved location, the circles reflecting the area of land used up by these shifting agriculturalists (cf. Clarke 1972, 25). Alternatively, the regularly spaced single sites might be seen as major service centres with peripheral clustering of minor sites in areas of least competition from the main centres.

In fig. 1.3 we can test a hypothesis that the points in part of the study area are regularly spaced by placing over them a network of hexagons whose orientation and size we are allowed to alter at will (cf. Clarke 1968, 508–9). We notice that one site occurs in most hexagons and therefore conclude that the hypothesis of regular spacing is correct. Since it is the hexagon shape which has been used we might even invoke aspects of central place theory to explain the distribution. An 'advantage' of archaeological data is its incomplete nature. In cases where the model does not fit we introduce this factor. For example, empty cells might be said to predict where further sites will be found or to suggest where they have been destroyed. Hexagon cells containing

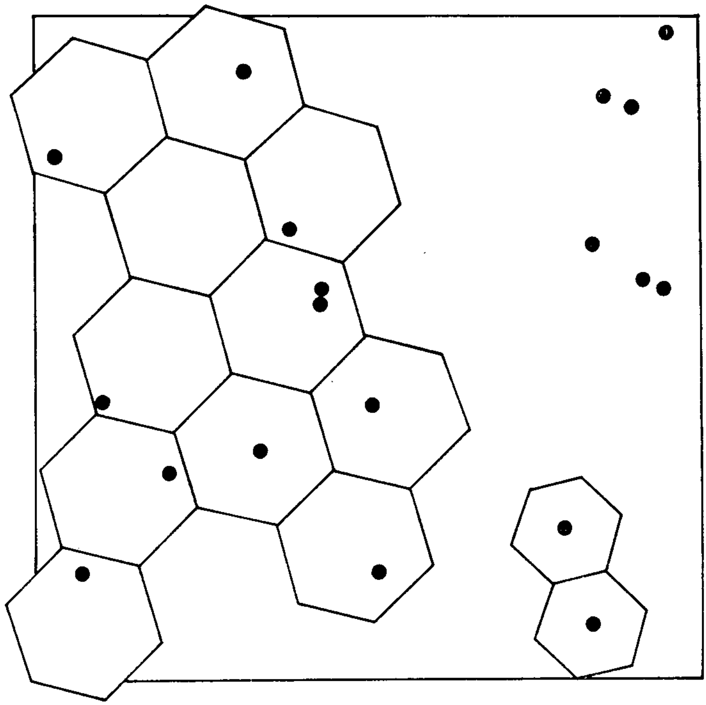

Fig. 1.3. Hexagons placed over a random distribution to suggest regular spacing.

two rather than one site could be explained by suggesting that the sites are not precisely contemporaneous.

If a hypothesis of clustering of sites is preferred, contours can be drawn around the same pattern of sites as in fig. 1.4. This time (fig. 1.5) agglomeration is suggested, not regularity.

These examples serve to underline the dangers of a non-rigorous approach to map analysis and interpretation when, as with archaeological data, little is known of the spatial process which produced the pattern.

Developments in the spatial analysis of archaeological data are needed for a third reason. Recently, large bodies of distributional information have been collected which are difficult to examine without some advance in the available techniques. For example, for one phase of the northwest German early Bronze Age, Bergmann (1970) collected over 112 distribution maps of different artifact types. On each of these, different symbols were used to indicate the context of discovery. Handling the similiarities and variations between all this information is not easy, and a reappraisal (p. 211) identified additional patterning to that noticed in the visual sorting. In a study of 542 middle Bronze Age palstaves in southern England (Rowlands and Hodder, unpublished),

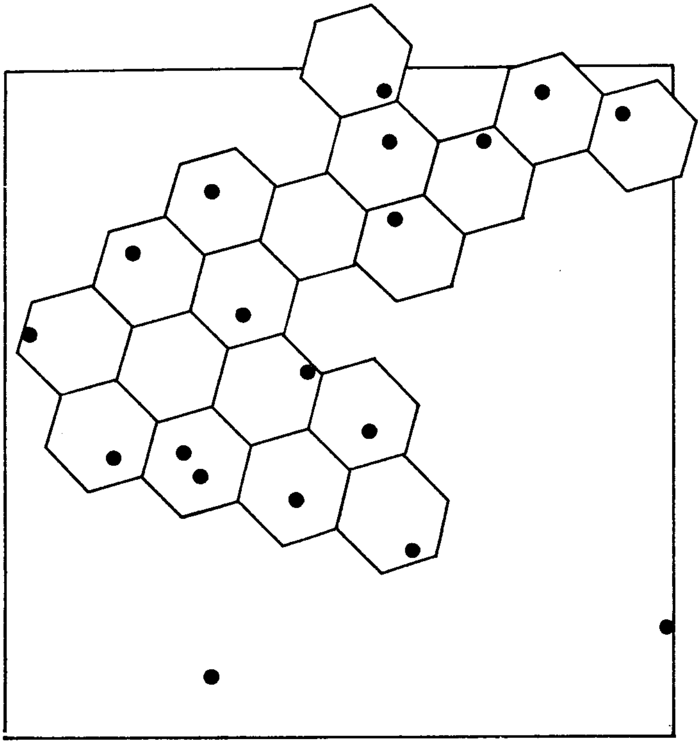

Fig. 1.4. Hexagons placed over a random distribution to suggest regular spacing.

the spatial structure of the many palstave types in each of four typological groupings of the material was to be examined for two divisions of the palstaves. To produce the large number of necessary maps would have been time consuming, difficult to present in a published form, and extremely difficult to interpret. By using some of the techniques to be discussed in this work, the problem became manageable.

In response to the need for a development in spatial studies, recent archaeological work has been advancing in two main ways. The first development is an attempt to describe and analyse distributions in a more rigorous way in order to obtain greater precision and reliability. Examples of this approach are the work of Whallon (1973; 1974) and Dacey (1973). These and other examples will be discussed in the following chapters. With more characteristics of the distributions defined there is a better basis for interpretation. In general a quantitative and/or statistical approach is involved. Because of the difficulties in using statistical tests on archaeological data (which will be discussed below), the rigour which is achieved in this way may be more apparent than real. However, some rigour is sometimes found in attempts to use

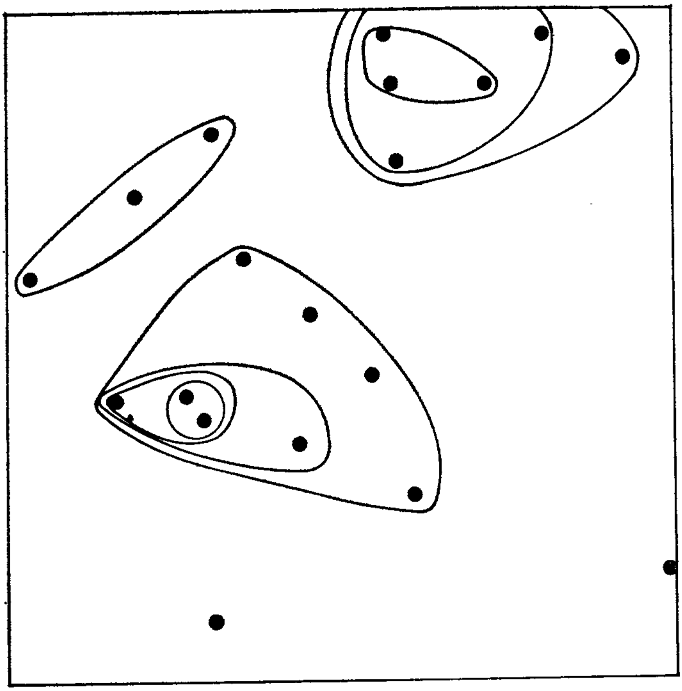

Fig. 1.5. An impression of clustering in a random pattern.

clearly defined and repeatable analytical procedures. A second recent change is a greater emphasis on the process leading to the form of an archaeological pattern. This was an area of interest for Fox, but there was no real attempt to distinguish invasion from other mechanisms. With better techniques we can hope to differentiate different processes such as, for example, different types of diffusion. Anomalies in spatial trends might be related to varying social conditions etc. Indeed, along these lines Renfrew (1969; 1972*b*) has made some generalisations about trade from regression curves, and Hogg (1971) has modelled the process of dispersal of Iron Age coins.

One difficulty which will become apparent during the course of this study is that of inferring process from form. One spatial pattern may be produced by a variety of different spatial processes (Harvey 1968; King 1962) and it must be part of the task to determine the possible range of the variety. Ways of differentiating between alternative hypotheses about the same spatial form will be examined (p. 88) but often one must look to non-spatial evidence to corroborate or disprove theories about spatial processes. As an example of the difficulty the distribution of sites can be considered. It will be shown that the form of many locational patterns is comparable to a Poisson distribution. This suggests that the settlements are located as a random and independent process. Much locational theory suggests quite the opposite.

> It is known that the probability of a store locating in any area is conditional upon a number of factors, not the least important of which is the relative location of other stores. Therefore, a simple model such as the Poisson law hardly is suggested by theory, and while it may serve as a convenient first approximation of the location pattern, it reduces to simplicity a situation which already has been acknowledged as a complex one. Besides, it is almost certain that other probability models could be found which would fit the observed facts equally well, and unless there is theory to guide us in our choice, one model may appear no better than the others [King 1969, 43].

The interpretation of random patterns will be discussed in section 4.1.

One approach that will continually be used to model spatial patterns is to use random or stochastic processes. This method has been widely found to be useful in modelling human behaviour and it has already begun to be used in archaeology (Isaac 1972, 178; Thomas 1972). Its use for examining spatial patterns should perhaps be explained. We expect non-random spatial patterns because we know that individual behaviour is not random but is constrained and determined by, for example, kinship factors in the exchange of goods and physical factors in the location of sites. However, it will be found that non-random behaviour is often not apparent in the spatial patterns. Many of the observed archaeological patterns have a form which is similar to patterns produced by a random process. If the form of the pattern is similar to the end result of a random process, this does not necessarily mean that the process which produced the observed pattern was random. It is possible, however, that, given a 'satellite view', aggregate human behaviour is often best simulated by a random process, or by very simple models incorporating a strong random element. This view has been put forward by Curry (1964; 1967) and developed by, for example, Cliff and Ord (1973; 1974). According to Curry, 'every decision may be optimal from a particular point of view and yet the resulting actions as a whole may appear as random. Lack of information, social ties, and so on will change an economic optimising solution but not the randomness formulation' (1964, 138). It is possible to consider behaviour as rational when all the constraints of a decision are known, and this is the level at which social anthropologists are able to work in studying human interaction (for example Barnes 1972). However, especially in a dynamic framework, there is such a large number of decisions being taken, rarely coincident in time and being separately motivated under differing circumstances and degrees of information, that comprehension of rationality on a wide scale is impossible. Thus, 'men, motivated by various ideas, act so that from the point of view of the locational structure as a whole their actions appear as random' (Curry 1964, 145–6).

It is perhaps helpful to compare this notion of random aggregate behaviour with that of entropy in Information Theory (Harvey 1969, 462; Wilson 1970). If a system contains n elements and it behaves in such a way that if the value of one element in the system is known all the other values can be predicted, then such a system is highly organised. In a similar system the values of $n - 1$ elements might be known, but the value of the nth element still cannot be predicted. Such a system is disorganised and is in a state of high entropy. In certain situations there may be a variety of choices available for any action so that the aggregate pattern of actions shows little order and provides little information about the actions. This is comparable to the notion of randomness discussed above. Information 'may be regarded as the measure of the amount of organisation (as opposed to randomness), in the system' (Klir and Valach 1967, 58). 'The results of an unrestrained random process can be defined as showing zero order. Order is achieved by placing constraints on the freedom of choice of action. This variety of available choice may be called entropy and is the complement of the degree of ordering' (Curry 1964, 144).

1.2 *Statistical introduction*

It would not be possible in a book of this length to give even an introduction to basic statistical theory, or to the use of the more common statistical techniques in general, without overwhelming the rest of the text. There is of course no need to do so, as there are many textbooks on the subject. For a straightforward introduction to the theoretical aspects we recommend Lindgren (first edition 1960 or second edition 1968, page references will refer to the first edition), while for a more practical approach Davies (third edition 1961) or Davies and Goldsmith (1972) is recommended. Where possible, these works will be quoted as source material for particular aspects of statistical theory or technique.

1.2.1 *Notation*

A certain amount of mathematical notation is needed in the presentation of techniques. Most of the symbols should be familiar – a few of the less common are given below; this section can of course be skipped by those familiar with them.

(i) *Subscripts*: are lower case letters, or numbers, slightly below a symbol, and indicating that that symbol represents a particular item out of a set. For example, in section 4.3 the symbol 'p' is used to denote the predicted proportion of the distance between two major towns at which a minor town is situated. There are several (n) roads, each with its own particular value of p, and these are denoted by p_1 for the first road, p_2 for the second and so on up to p_n for the nth. The expression p_i is commonly used for an unspecified ith proportion. Double or multiple subscripts can be used – for example in section 4.3 p_{1i} means the ith proportion if hypothesis 1 is true, and p_{2i} means the ith proportion if hypothesis 2 is true.

(ii) Σ *and* Π (sigma *and* pi): it is often necessary to express a multiple sum, for example in section 3.1 the average nearest-neighbour distance is calculated by summing r_1, r_2 up to r_n and dividing by n. This can be written as $\frac{1}{n}(r_1 + r_2 + \ldots + r_n)$ but it is simpler and shorter to write it as $\frac{1}{n}\sum_{i=1}^{n} r_i$; the 'range' of the subscript i (from 1 to n) means that one starts summing at r_1, and finishes at r_n. Where the range is obvious from the context, the expression can safely be shortened to $\frac{1}{n}\Sigma r$, and further to $\bar{r}$ ('r-bar').

Multiple products (e.g. $r_1 \times r_2 \times \ldots \times r_n$) are also used. The standard mathematical shorthand for this is $\prod_{i=1}^{n} r_i$, or just Πr.

(iii) *Factorials*: a special case of the multiple product is the factorial. Written as $n!$, it is the product $1 \times 2 \times 3 \times \ldots \times n$. An important piece of notation related to the factorial is the *binomial coefficient*, written as $\binom{n}{r}$, and standing for $\frac{n!}{r! \times (n-r)!}$. It will be used extensively in chapter 6. Two important properties of the binomial coefficient are that $\binom{n}{r} = \binom{n}{n-r}$ and $\binom{n}{0} = 1$.

(iv) *Estimates*: an estimate of an unknown parameter is often denoted by a '^' or 'hat'. For example, in section 3.2 the parameter λ (lambda) has to be estimated from the site density, and this estimate is written as $\hat{\lambda}$.

1.2.2. *Distribution functions*

The important concepts of random variable and distribution function are discussed by Lindgren (pp. 31–90). ('Distribution' in this sense is *not* the same as the 'distribution' of a distribution map.) Suppose a variable x (for example, a squared nearest-neighbour distance as in section 3.2, p. 47) can take as its value any one of a set of real numbers, and that the probability (P) of x being less than or equal to a certain value, say ω (omega) is written as $P(x \leqslant \omega)$.

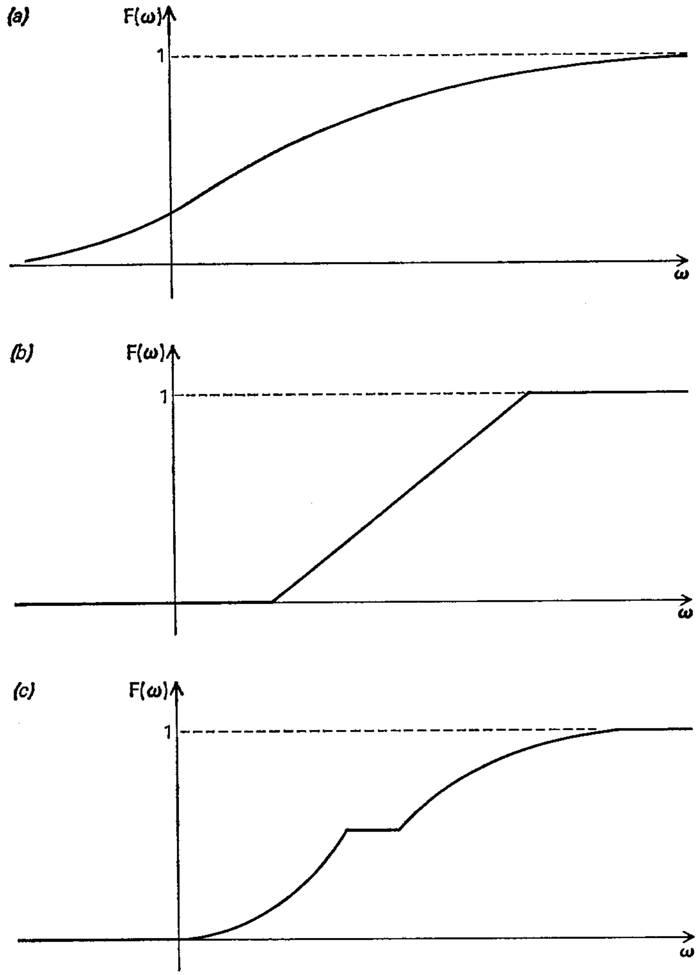

Fig. 1.6 (*a–c*). Examples of various sorts of cumulative distribution functions. For explanation of symbols see text.

Then the function $F(\omega) = P(x \leqslant \omega)$ is a cumulative distribution function (cdf). A function of this sort will clearly increase (or stay the same) as the value of ω increases, and it will never be less than 0 or greater than 1 (because there cannot be a probability less than 0 or greater than 1). A few simple examples are shown in fig. 1.6. A related function is the density function $f(\omega)$, which is in fact the derivative of the cdf, i.e. from the example just mentioned $f(\omega) = \frac{d}{d\omega} F(\omega) = \frac{d}{d\omega} P(x \leqslant \omega)$. It has the property that if the graph of $f(x)$ against x is drawn, the area to the left of the value $x = \omega$ and beneath the curve is equal to $F(\omega)$ (the shaded area of fig. 1.7). Also the probability of

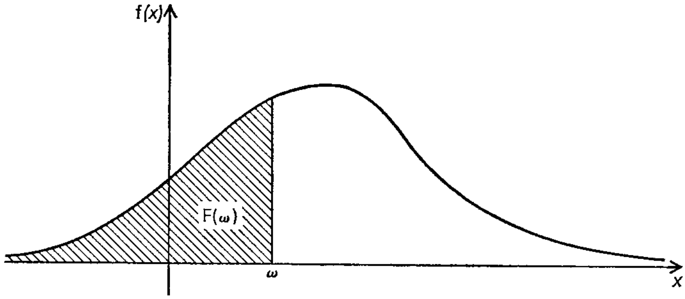

Fig. 1.7. Example of density function showing relationship between it and the associated cumulative density function. For explanation of symbols see text.

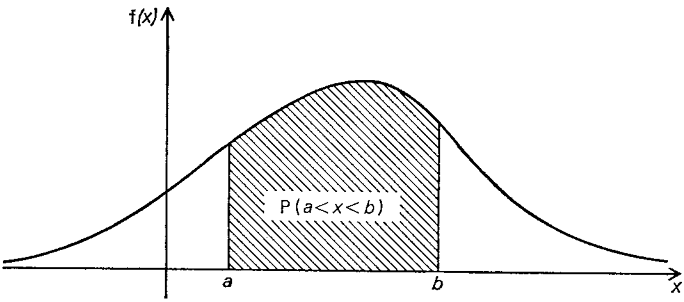

Fig. 1.8. Example of a density function, showing probability interpretation of the area beneath the curve.

x falling between two values a and b, $P(a < x < b)$, equals the area beneath the curve between the values a and b (fig. 1.8). The value of ω for which $F(\omega) = \alpha\%$ is known as the alpha-percentile of the distribution.

Also in section 3.2 (p. 47), *conditional* cdfs and density functions are used. They are distinguished by a short vertical line between the variable, in this case ω, and the condition (c). For example, $f(\omega | 0 \leqslant \omega \leqslant c)$ is the density function of ω on the condition that it must not be less than 0 nor greater than c.

Sometimes the variable can only take one of a finite (strictly speaking, countable) number of different values (for example, the number of sites in a grid-square can only be a whole number not exceeding the total number of sites). Such variables are called *discrete*, and their distributions are *discrete distributions*. They are characterised by the probabilities associated with the possible values the variable could take, which are written as $p_k = P(x = x_k)$ where x_k is one of the possible values x_0, x_1, x_2, etc.

The distributions most commonly used in this book are:

(i) The Normal distribution, $f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ (Lindgren, p. 88).

(ii) The (negative) exponential distribution, $f(x) = \lambda \exp(-\lambda x)$, $x > 0$, (Lindgren, p. 82) which are both continuous distributions, and

(iii) The Poisson distribution, $p_k = \exp(-m) \times m^k/m!$ (Lindgren, p. 76).

(iv) The negative binomial distribution, $p_k = \frac{(r+k-1)!}{k!(r-1)!} p^r q^k$, where $q = 1 - p$, (Lindgren, p. 142) which are both discrete distributions.

Of these four distributions, two – the negative exponential and the Poisson – each have one parameter (λ and m respectively) which determine their shape, while two – the Normal and the negative binomial – have two parameters each (μ (mu), σ (sigma), and p, r respectively).

If one is considering two (or more) variables, it may be that the conditional distribution of one, given the value of the other, is independent of that value (Lindgren, p. 103). In other words, the value of the second variable may tell us nothing about the value of the first. In such circumstances, the variables are said to be *independent*. Many statistical tests are based on the assumption that the variables concerned are independent, and will not otherwise be valid.

If two variables are not independent, one may be interested in the degree of relationship or 'coherence' between them. The most common measure of coherence is the correlation coefficient, usually denoted by ρ (rho). It can take any value from -1 to $+1$: if it is zero the variables are said to be 'uncorrelated', while a value of $+1$ indicates perfect positive correlation and -1 perfect negative correlation. Uncorrelated variables need not necessarily be independent, but independent variables are always uncorrelated (see section 5.1).

Suppose we have a random sample $\mathbf{x} = x_1, x_2, \ldots, x_n$ from a distribution with density function $f(x)$, which is determined by a single parameter θ (theta), and can therefore be written as $f(x, \theta)$. The density function will take a separate value for each of the x_i – $f(x_1, \theta)$, etc. Multiplying all these together gives us a new function $L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$ which is known as the *likelihood function* (Lindgren, p. 222). This can be used to estimate the parameter θ: *the maximum likelihood estimate* of θ being that value which maximises the likelihood function $L(\theta)$. However, we shall be more interested in comparing likelihood functions for different values of θ (see next section).

1.2.3 *Statistical tests*

Suppose we have a hypothesis (often called a *null* hypothesis if it is a particularly simple one) about a certain situation, which we wish to test by reference to a certain set of data. For example, in section 3.1 we use the hypothesis that a point-pattern is random in order to study a distribution pattern of sites. In such a situation there are two sorts of mistake one can make – (i) falsely rejecting a true hypothesis (called a type I error) and (ii) falsely accepting an untrue hypothesis (called a type II error). In devising a statistical test we have to agree on an acceptable chance of making a type I error – it might be a very small chance if the implications of rejecting the hypothesis are far-reaching, or it might be larger if we are using the hypothesis as an 'Aunt Sally' and are not committed to it. This chance is called the *significance level* of the test – commonly used values are 5 % and 1 %, but any value can be used. The test will then be to reject the hypothesis if some statistic based on the data (known as a *test statistic*) exceeds a *critical value* which depends on the test being used and on the significance level chosen. If, having performed the test, we say that 'the result is significant at the 5 % level' we mean that, if the null hypothesis were true, the probability of obtaining a set of data no more favourable to the hypothesis than our actual set of data, is 5 % or less.

As a simple example, suppose we take a sample of n observations from a Normal distribution with known variance 1 but unknown mean, and that our null hypothesis is that the mean $\theta = 0$. We choose 5 % as our significance level, which (via tables of the Normal distribution) tells us to reject the null hypothesis if the mean of our observations is outside the range $-1.96/\sqrt{n}$ to $+1.96/\sqrt{n}$ (the critical values). We calculate the mean (our test statistic), find that its value lies outside this range and say that it differs significantly from zero, at the 5 % level.

A rather different situation occurs if we have two hypotheses and wish to choose between them. This can be done (as in section 4.3) by calculating the likelihood function (or its greatest value) under each hypothesis and dividing one by the other to form a *likelihood ratio* (LR). There will be a critical value of this ratio, below which one hypothesis will be accepted and above which the other (see p. 82).

The topic of hypothesis testing is covered much more fully by Lindgren (pp. 232–67).

A particular case is the situation when the null hypothesis specifies the form of the parent distribution of which our data is a sample – for example, that our data are a random sample from a Poisson distribution. In such cases a *goodness-of-fit* test is often used. The well-known χ^2 (chi-squared) test (Cochran, 1952) is one such, but there are many others. Such tests are usually employed when the possible alternative to the hypothesis is not well defined. Lindgren (p. 296) points out a disconcerting feature of goodness-of-fit tests – that if the sample is large enough then the null hypothesis would almost certainly be rejected, since the true state of the situation, while perhaps very close to our null hypothesis, is probably not *exactly* as specified in the null hypothesis. The question to be asked may then be 'is the difference practically significant?' rather than 'is it statistically significant?'.