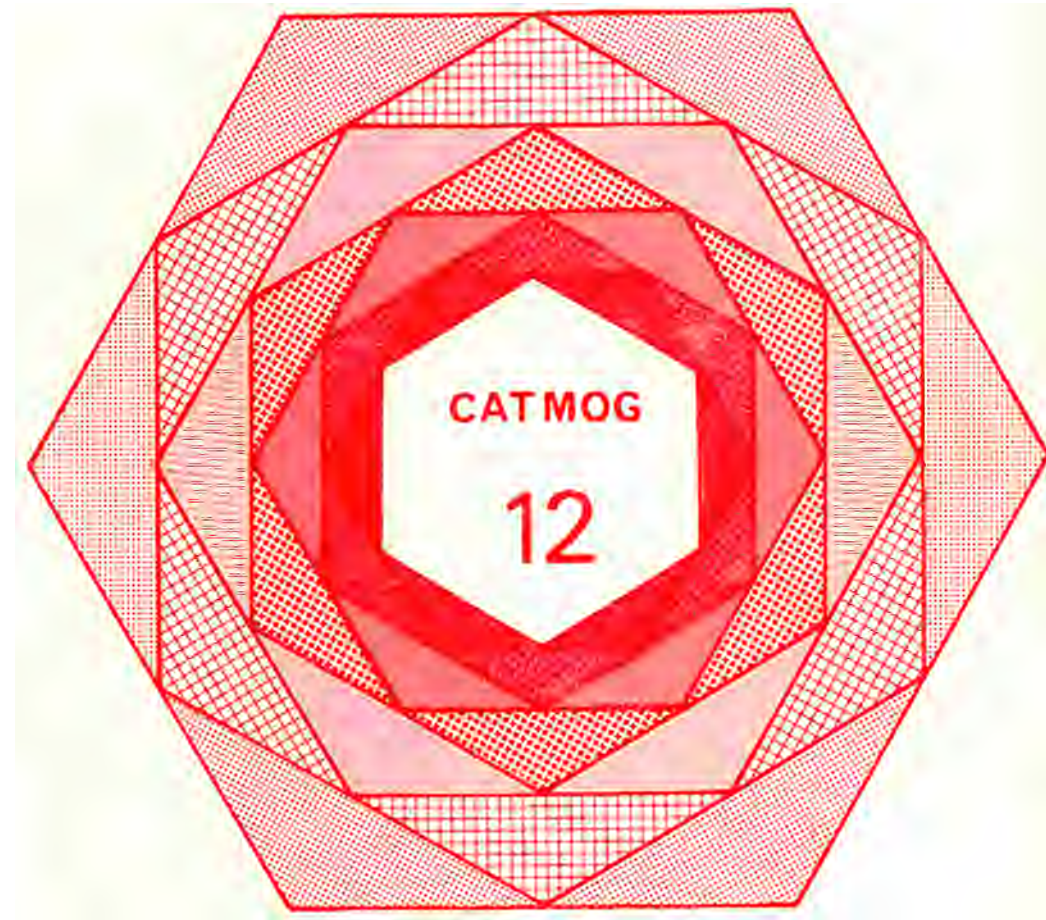


# AN INTRODUCTION TO QUADRAT ANALYSIS

R.W.Thomas



ISSN 0306-6142

ISBN 0 902246 66 6

© 1977 R.W. Thomas

CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of class-work. Each book is written by an author currently working with the technique or concept he describes.

AN INTRODUCTION TO QUADRAT ANALYSIS

by

R. W. Thomas  
(University of Manchester)

CONTENTS

1. Introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography. N.J. Thrift

Other titles are in preparation, and any suggestions as to new titles should be sent to the Editor, Dr P.J. Taylor, Dept of Geography, University of Newcastle upon Tyne. Prospective authors are invited to submit an outline of their proposed text for consideration.

*This series, Concepts and Techniques in Modern Geography is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers. For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London, S.W.7.*

*The series is published by Geo Abstracts Ltd., University of East Anglia, Norwich, NR4 7TJ, to whom all other enquiries should be addressed.*

	PaCe_
I <u>INTRODUCTION</u>	
(i) Background	3
(ii) Data collection methods	5
(iii) The methodology of quadrat analysis	6
(iv) Other properties of point patterns	8
II <u>INDEPENDENCE IN SPACE</u>	
(i) The multiplication axiom for independent events	9
(ii) Binomial coefficients	10
(iii) The binomial distribution	10
(iv) The Poisson distribution as a limit of the binomial	12
III <u>GOODNESS-OF-FIT TESTS</u>	
(i) The Chi-square test	14
(ii) The Kolmogorov-Smirnov D Statistic	15
(iii) The variance/mean ratio	17
IV <u>DEPENDENCE IN SPACE</u>	
(i) Uniform and clustered patterns	18
(ii) The negative binomial distribution	20
(iii) Moments and maximum likelihood estimation of k	21
(iv) Other distributions	24
V <u>GEOGRAPHICAL APPLICATIONS AND THEIR PROBLEMS</u>	25
(i) Karst depressions and the problem of model specification	26
(ii) The scale problem	27

VI	<u>ALTERNATIVE APPROACHES</u>	
	(i) Multinomial coefficients and state descriptions	30
	(ii) The entropy maximising distribution	32
	(iii) Redundancy	34
VII	<u>CONCLUSIONS</u>	36
	<u>BIBLIOGRAPHY</u>	37
	<u>APPENDIX</u>	41

Acknowledgement

Thanks are due to Peter Lloyd for permission to reproduce material from the N.W. Industrial Data Bank; to Fiona Hill for typing the original manuscript, and to Clive Thomas for preparing the diagrams.

1. INTRODUCTION

(i) Background

Quadrat analysis embraces a variety of mathematical and statistical techniques which are designed to measure properties of point patterns. These techniques are of inherent interest to geographers because they provide answers to fundamental questions about the relationships between points in space. However, the first applications of the quadrat method appear in the literature of plant ecology, beginning with a paper by Gleason (1920). In plant ecology quadrat methods are used to analyse spatial properties of plant communities, but it is only recently that geographers have taken a serious interest in these techniques. Geographical point patterns which have been subjected to quadrat analysis include the distribution of shops in urban areas (Rogers, 1965, 1969 c), the distribution of karst depressions in a limestone region (McConnell and Horn, 1972), and the adoption of agricultural innovations by rural populations (Harvey, 1966).

Fig.1(i) shows the distribution of vehicle factories in the Merseyside conurbation (1966), which is a representative example of a geographical point pattern suitable for quadrat analysis. The geographer looking at this map might ask himself the following questions. 'Does the pattern of points suggest that the location of one factory influences the location of other factories?' In other words, 'are factory locations in some way dependent on one another?' Alternatively, he may ask the simpler, and even more fundamental, question, 'is there evidence in the pattern which indicates that the points are totally unrelated?' In other words, 'is the pattern just a random distribution of points?' Subsequently, we shall see that the idea of a random point pattern requires careful definition, but Fig.1(ii) gives a visual impression of a random pattern.

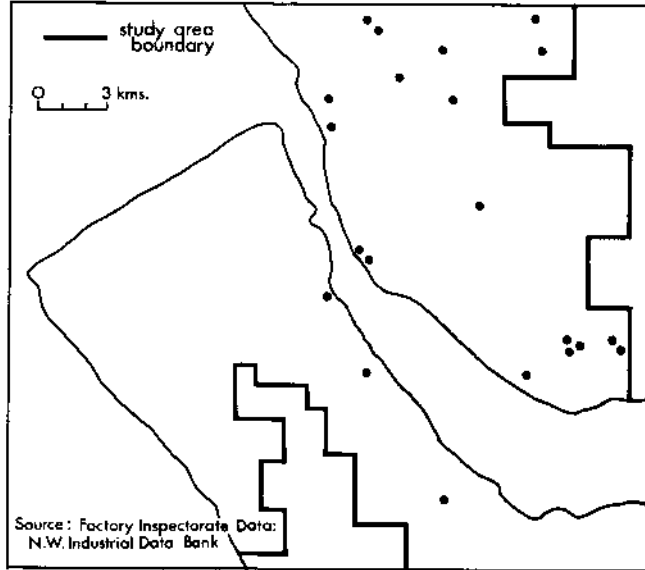
Quadrat methods attempt to answer such questions by adapting some of the basic mathematical ideas in probability theory to analyse the frequency distribution of a point pattern. By frequency we mean the manner in which density of points varies over the study area. Fig.2(i) illustrates a hypothetical pattern of 21 points which we shall use as a simple numerical example. We can measure the frequency distribution of this pattern by laying a square grid over the pattern (Fig.2(ii)) and counting the number of points falling within each cell. We will define the individual observation as  $I_n$ , which is the number of points located in the  $i$ th cell. For any pattern there will be  $n$  cells and  $r$  points. In this example  $n = 36$  and  $r = 21$ . We obtain the frequency distribution of the point pattern by counting the number of cells containing exactly  $m$  points for all values of  $m$  between 0 and  $r$ . We will refer to these counts by the term  $n_m$ . In our example  $n_{m=0} = 20$  indicating that the pattern contains 20 empty cells. The set of values  $n_m$  is called the frequency array and it is this array that provides the basic data for quadrat analysis.

It will prove helpful if we note two obvious characteristics of the frequency array describing a particular point pattern. To be a valid representation of the pattern the frequency array must obey the following two constraints:

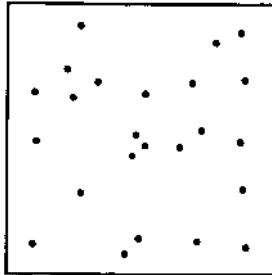
$$\sum_{m=0}^r n_m = n \quad (1)$$

that is, the sum over the number of cells containing  $m$  points must equal the total number of cells, and

i Merseyside, 1966: Vehicle plants (S.I.C. Order 8)



ii A random pattern



Points located by random co-ordinates.

Fig. 1 : Some example patterns

$$\sum_{m=0}^r n_{m \cdot m} = r \quad (2)$$

that is, multiplying the number of cells containing  $m$  points by  $m$ , and summing over  $m$ , must give the total number of points comprising the pattern. The calculation of both these formulas is shown in Fig.2(ii).

(ii) Data collection methods

The design of a quadrat experiment begins with defining the limits of the study region. Any definition chosen depends on the individual nature of the objects that form the points, although the following guidelines provide a general background to this problem. First, every part of the study area should be a possible location for a point. Thus, if we are studying the frequency distribution of karst depressions, the study area should be exclusively a limestone region. The inclusion of other rock types would lead to an erroneous frequency array because the number of empty cells would be overestimated. Second, it should be remembered that the size of the study area determines the level of resolution of the problem. For instance, any discussion of our factory pattern is immediately restricted to the intra-urban scale by our study area definition. The remaining problems of study area definition are specific to certain types of quadrat experiments and are treated at the relevant points in the text.

Having defined the study area we then construct the frequency array. This is achieved by adopting the procedure of either quadrat censusing, or quadrat sampling. We have already described quadrat censusing which involved deciding on a cell size and then laying a contiguous grid of these cells over the study area. Fig.2(ii) illustrates the results of censusing the hypothetical pattern with two different cell sizes. The problem of selecting an appropriate cell size is quite complex and is treated separately in sections V (ii) and VI (iii). Quadrat sampling depends on selecting a single cell of a predetermined size and then randomly placing the cell over the study area  $n$  times. With quadrat sampling the frequency distribution is obtained by counting the number of points lying within the quadrat on each of the  $n$  random placings and then aggregating these  $n$  observations to form the frequency array,  $n_m$ . Thus quadrat censusing provides a frequency description of the entire population of points located within the study area, while quadrat sampling yields an estimate of the frequency distribution obtained from a random spatial sample. In the case of quadrat censusing the average density of points depends on  $n$  and  $r$  such that the average number of points per cell ( $\lambda$ ) is given by

$$\lambda = r/n \quad (3)$$

whereas from quadrat sampling our estimate of the average density ( $\bar{m}$ ) is given by

$$\bar{m} = \sum_{i=1}^n m_i/n \quad (4)$$

where  $m_i$  is the number of points found in the  $i^{\text{th}}$  quadrat sample. The variance of the observations around  $\bar{m}$  and  $\lambda$  is given by

$$s_m^2 = \left( \sum_{i=1}^n m_i^2 - \frac{\left( \sum_{i=1}^n m_i \right)^2}{n} \right) / (n-1) \quad (5)$$

Usually the choice between adopting censusing or sampling depends on the phenomenon being studied. For instance, the study of a community of plants in a local habitat would most easily be achieved by sampling, whereas for the distribution of shop types in an urban area censusing is more appropriate because here the total population of shops could easily be identified. However, the majority of the statistical theory of quadrat analysis is based on the premise that the data has been collected by sampling. Consequently, the nature of the mathematical model to be tested may well determine the data collection procedure.

(iii) The methodology of quadrat analysis

We begin by observing that dividing the individual elements of the observed frequency array ( $n_m$ ) by  $n$  (the total number of cells) gives us a probability distribution which represents the likelihood of finding exactly  $x$  points in a single quadrat selected at random from the census. Therefore, if we require the probability that a randomly selected quadrat contains exactly  $x$  points, then the probability that  $x=m$  is given by

$$P_{\{x=m\}} = P_m = n_m/n \quad m = 0, 1, \dots, r \quad (6)$$

where  $m$  can take on values between 0 and  $r$ . For the frequency array given in Fig.2(ii)a the probability that a randomly selected quadrat contains 0 points is given by

$$P_{\{x=0\}} = P_0 = 20/36 = .556$$

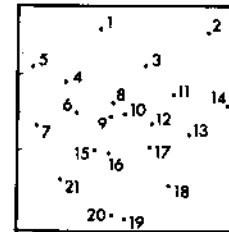
The distribution of  $P(x=m)$  possesses the property common to all probability distributions which is that the sum of the individual probabilities forming the distribution must be 1 (unity), such that

$$\sum_{m=0}^r P_m = 1 \quad (7)$$

The central idea in the quadrat method is that we construct theories, in the form of probability distributions, to give predicted probabilities which may be compared with each of the individual probabilities in the observed frequency distribution. The theoretical probability distribution is obtained by making sensible assumptions about the process governing the evolution of the point pattern. From those assumptions we deduce the probability distribution that will give the appropriate prediction of the frequency distribution of the pattern. Finally, we compare the predicted probabilities with the observed probabilities obtained by sampling or censusing the pattern. If there exists a close correspondence between the predicted and the observed probabilities then we accept the assumptions incorporated in the theoretical distribution as being the most likely explanation for the process governing the evolution of the pattern. Conversely, if the theory bears little relation to the observed distribution then we reject the assumptions as being an untenable process explanation.

An example will help clarify this reasoning. The simplest assumption we can make about any point pattern is that its distribution is controlled by a random process. That is, the location of any point within the pattern occurred independently of all other points. Given this assumption it is quite easy to deduce a random probability distribution capable of predicting a

i A HYPOTHETICAL POINT PATTERN



$r = \text{no. of points} = 21$

ii FREQUENCY

a. Small scale  
 $n = \text{no. of cells} = 36$

0	0	1	0	0	1
1	1	0	1	0	0
0	1	3	0	1	1
1	0	2	2	1	0
0	1	0	0	1	0
0	0	2	0	0	0

$m$	$n_m$	$n_m m$
0	20	0
1	12	12
2	3	6
3	1	3
$\sum_{m=0}^r n_m$	$n = 36$	
$\sum_{m=0}^r n_m m$	$r = 21$	

$$\lambda = 21/36 = .583$$

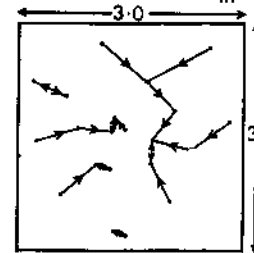
b. Large scale  $n = 9$

2	2	1
2	7	3
1	2	1

$m$	$n_m$	$n_m m$
0	0	0
1	3	3
2	4	8
3	1	3
7	1	7
$\sum_{m=0}^r n_m$	9	
$\sum_{m=0}^r n_m m$	21	

$$\lambda = 21/9 = 2.333$$

iii NEAREST NEIGHBOUR DISTANCES



$d_i = \text{Distance between } i^{\text{th}} \text{ point and its nearest neighbour.}$

$$\bar{d}_0 = \frac{\sum_{i=1}^r d_i}{r} = \text{Observed mean nearest neighbour distance}$$

$$= .93/21 = .443$$

iv CONTIGUITY

a. An alternative census

0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	1	1	2
0	0	0	1	1	2
0	0	0	1	1	2
0	0	0	1	1	3

b. Binary map

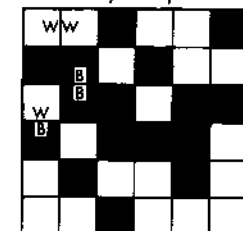


Fig. 2 : Approaches to the analysis of a hypothetical point pattern

random frequency array. We can then compare the predicted random probabilities with the observed probabilities in order to test the validity of our assumption of randomness. Alternatively, we can construct probability models which make assumptions about the location of a point being dependent on the location of the other points forming the pattern, and clearly the precise nature of these assumptions will depend on the particular phenomena we are studying.

Any probability distribution designed to give a prediction for the observed frequency array must satisfy, or at least closely approximate, the following condition

$$\sum_{m=0}^r P_m \cdot m = r/n = \lambda \quad (8)$$

This condition states that, when the predicted probabilities ( $P_m$ ) are multiplied through by successive values of  $m$ , the sum of these values must equal the observed mean number of points per cell. This condition is equivalent to equation (2) divided through by  $n$  and makes good sense because the theory should predict the same number of points in the observed distribution.

(iv) Other properties of point patterns

Quadrat analysis is one of a number of different approaches that geographers have taken to the analysis of point patterns. Consequently, in order to appreciate the role of quadrat analysis in geography it is necessary to have some knowledge of the alternative procedures. Here we present a brief sketch of the two major alternatives; nearest neighbour analysis and contiguity analysis.

Nearest neighbour analysis focuses on the distance between each point in the pattern and its nearest neighbour (see Fig.2(iii)), and of particular interest is the value of the observed mean nearest neighbour distance

$$\bar{d}_0 = \sum_{i=1}^r d_i / r \quad (9)$$

The impetus for geographical interest in nearest neighbour analysis was provided by Clark and Evans (1954) who derived a sampling theory for the observed mean nearest neighbour distance. Their work provides a significance test for the difference between  $\bar{d}_0$  and the expected mean nearest neighbour distance ( $\bar{d}_e$ ) in a random pattern of points.

It must be remembered that quadrat methods ignore the spatial arrangement of the pattern. For example, the highly ordered point pattern depicted in Fig.2(iv)a produces the same frequency array as the pattern in Fig.2(ii)a. Therefore a considerable body of statistical and geographical research has been devoted to developing techniques which measure the arrangement of a pattern. Such techniques are generally referred to as contiguity analysis. For example, we can transform the hypothetical census in Fig.2(ii)a into a binary map (Fig.2(iv)b) by defining empty cells as white, and cells containing at least one point as black. The simplest form of contiguity analysis is concerned with tests based on the sampling theory of the distribution of contacts between cells. These contacts are defined as white-white, black-white

and black-black. The tests are made between the observed distribution of contacts and the expected distribution in a random pattern. Contiguity tests have been developed for regular grids (Cliff, 1968) and irregular spatial units (Geary, 1954), and more recently many of the central assumptions of contiguity analysis have been incorporated in the general treatment of the problems of spatial autocorrelation (Cliff and Ord, 1973). Indeed, all the approaches to point patterns described here may be subsumed under the general heading of stochastic process theory in geography, and Hepple (1974) has provided a comprehensive survey of this field.

In writing this monograph it has been assumed that the reader is familiar with basic descriptive statistics and the more elementary notions of significance testing. The additional probability theory required to understand quadrat methods is presented in sections II, IV and VI. The more complicated theoretical sections have been starred and may be omitted on first reading. Section II of the monograph begins by describing the reasoning that leads to the establishment of random probability distributions. Section III describes the statistical procedures we may adopt to test for the goodness-of-fit between a theoretical and the observed distribution, while Section IV considers the more difficult problem of constructing probability distributions which model situations where the locations of points are dependent on one another. The remaining sections examine some of the problems arising from the application of quadrat methods in geography and also offer some alternative approaches to these problems.

## II. INDEPENDENCE IN SPACE

(i) The multiplication axiom for independent events

Before we can begin to construct probability models for the random frequency distribution we need to define exactly what is meant by an independent event in probability theory.

Many problems in probability theory may be envisaged as an experiment with a finite number of mutually exclusive outcomes ( $n$ ). If we term each of these outcomes an event, then the probability of the occurrence of one of these events in a single trial on the experiment is given by the symbol  $p(E)$ . For instance, if our experiment was the tossing of an unbiased coin, then the experiment has two outcomes, heads and tails. The probability for each of these equally likely outcomes in a single trial is given by

$$p(E_H) = p(E_T) = 1/n = \frac{1}{2} \quad (10)$$

If in a single trial the specified event does not occur, we say that the complementary event ( $\bar{E}$ ) has happened.  $\bar{E}$  includes all the events that comprise the experiment other than the specified event, and its probability is given by

$$p(\bar{E}) = 1 - p(E_i) \quad (11)$$

where  $E_i$  is the specified event. The reader may wish to check that the definition of the probability of a complementary event satisfies the condition given in equation (7).

Two events are said to be independent if the occurrence of one event in an experiment does not influence the occurrence of the other. The idea of independence is defined in the multiplication axiom which states that, if  $E_1$

and  $E_2$  are two independent events the probability that they both happen is the product of their individual probabilities'. Symbolically we write this definition as

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2) \quad (12)$$

where  $\cap$  is the set theory symbol for intersection, and the term  $p(E_1 \cap E_2)$  is the probability of  $E_2$  following  $E_1$  in two trials on an experiment. For example, in our unbiased coin tossing experiment the multiplication axiom gives the probability of Heads being followed by tails in two tosses of the coin as

$$p(E_H \cap E_T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

More generally, for experiments with  $n$  outcomes, the multiplication axiom may be written as

$$p(E_1 \cap E_2 \cap \dots \cap E_n) = \prod_{i=1}^n p(E_i) \quad (13)$$

where  $\prod_{i=1}^n$  is the symbol for the product of all the terms from  $i$  to  $n$ .

(ii) Binomial Coefficients

Binomial coefficients are combinational numbers which define the total number of different sequences in which an event can occur exactly  $m$  times in  $r$  trials on an experiment. For example, it is easy to see that there are 6 sequences in which 2 heads can occur in 4 tosses of a coin, these are

HHTT  
HTTH  
HTHT  
TTHH  
THTT  
THTH

In this case the answer was obtained by experimentation, however, the general solution for the number of sequences in which an event occurs  $m$  times in  $r$  trials is given by the binomial coefficient. Formally this is expressed as

$$\binom{r}{m} = \frac{r!}{m! (r-m)!} = \frac{r(r-1) \dots (r-m+1)}{m!} \quad (14)$$

where  $r! = 1 \times 2 \times 3 \times \dots \times r$

Enumeration of the binomial coefficient for this problem gives the result we obtained by experimentation

$$\binom{4}{2} = \frac{1 \times 2 \times 3 \times 4}{(1 \times 2) \times (1 \times 2)} = 6$$

(iii) The binomial distribution

We can now combine the ideas incorporated in the multiplication axiom and binomial coefficients to obtain the distribution that predicts the random frequency array. We require an answer to this question, if  $r$  points are placed independently, and one at a time into a grid composed of  $n$  equal sized cells, what is the probability that a single specified cell contains exactly  $m$  points at the conclusion of the experiment? Because the cells are equal sized the probability that one point falls in the specified cell in a single trial is given by

$$p(E) = 1/n = p \quad (15)$$

Conversely, the probability that the point does not land in the specified cell, such that the complementary event occurs, is given by

$$p(\bar{E}) = 1 - (1/n) = q \quad (16)$$

We also know that there are  $\binom{r}{m}$  different sequences, or outcomes of the experiment, which result in  $m$  points being found in the specified cell. If we take a single one of these sequences we can use the multiplication axiom to obtain the probability of its occurrence. The event  $E$  occurs  $m$  times in the sequence, each time with a probability  $p$ . Because the points are placed in the grid independently we obtain a probability of  $p^m$  for  $m$  points being placed in the specified cell. However, because in total  $r$  points are placed in the grid, for a specified cell to contain  $m$  points the complementary event (points not landing in the desired cell) must have occurred  $r-m$  times, each time with a probability  $q$ . Therefore, the probability of a single sequence occurring with  $m$  points being placed in the specified cell is given by

$$p^m q^{r-m} \quad (17)$$

However, because there are exactly  $\binom{r}{m}$  sequences which answer our question, each occurring with the probability (17), then, for the specified cell to contain exactly  $x$  points the probability that  $x = m$  is given by the binomial distribution

$$P_{(x=m)} = P_m = \binom{r}{m} p^m q^{r-m} \quad m = 0, 1, \dots, r \quad (18)$$

If we evaluate this expression for the hypothetical pattern in Fig.2(ii)a we obtain the following results:

$$r = 21, n = 36, \text{ therefore, } p = 1/36 = .027, \text{ and } q = 1 - .027 = .973$$

$$P_0 = \binom{21}{0} \times .027^0 \times .973^{21} = .553$$

(Note that  $0! = 1$ , and  $.027^0 = 1$ )

$$P_1 = \binom{21}{1} \times .027^1 \times .973^{20} = .332 \quad (\text{Note that } 1! = 1)$$

Further evaluation of the binomial distribution for values of  $m$  between 2 and  $r$  leads to a discrete probability distribution which describes the probability of finding  $m$  points in the specified cell when the pattern of points evolved under a random, or independent, process. Table 1 lists these values of  $P_m$  for values of  $m$  between 0 and 4. Because we have assumed that the  $n$  cells are identical, if we multiply these probabilities by  $n$ , we obtain the predicted number of cells containing  $m$  points for the whole pattern. This final calculation provides a random frequency array which can be compared with the observed frequency array ( $n_m$ ). Again, these calculations are listed in Table 1.

Subsequently we shall find that two useful summary measures of the predicted frequency array ( $P_m, n$ ) are its mean and variance. The mean, or first moment ( $Ex(m)$ ), of the binomial distribution depends on  $p$ , the probability of placing a single point in a specified cell, and  $r$ , the total number of points, such that expected, or mean value of  $m$  is given by

$$Ex(m) = rp = r/n = \lambda \quad (19)$$

Table 1. Probabilities and frequencies for the hypothetical data set

m	Binomial		Poisson		Observed	
	$P_m$	$P_{m \cdot n}$	$P_m$	$P_{m \cdot n}$	$n_m/n$	$n_m$
0	.553	19.9	.556	20.1	.556	20
1	.332	12.0	.327	11.7	.333	12
2	.095	3.4	.095	3.4	.083	3
3	.017	0.6	.019	0.7	.028	1
4	.002	0.1	.003	0.1	.000	0

The variance of the binomial distribution measures the spread of the values of  $m$  around the mean ( $\lambda$ ). The variance is defined as the difference between the second moment (the expected value of  $m^2$ , denoted by  $Ex(m^2)$ ) and the square of the first moment (the mean), and for the binomial distribution the variance is given by

$$\sigma^2 = Ex(m^2) - [Ex(m)]^2 = rpq \quad (20)$$

(iv) The Poisson distribution as a limit of the binomial

In most cases where we wish to apply the binomial in point pattern analysis  $n$  and  $r$  usually have large values, consequently  $p$  tends to be small. When  $n$  and  $r$  are large it can be shown that the Poisson distribution gives very close approximations to the results obtained from the binomial. In fact, the Poisson is a limiting case of the binomial distribution, and the proof is as follows: (this proof may be omitted on the first reading)

Let  $r$  and  $n$  tend to infinity such that the ration  $r/n$  tends to a constant value of  $\lambda$ , then put  $p = \lambda/r$  (from 19). The binomial distribution may now be written as

$$P_m = \binom{r}{m} \left(\frac{\lambda}{r}\right)^m \left(1 - \frac{\lambda}{r}\right)^{r-m} \quad (\text{from 18}) \quad (21)$$

$$= \frac{r(r-1)\dots(r-m+1)}{m!} \cdot \left(\frac{\lambda}{r}\right)^m \cdot \left(1 - \frac{\lambda}{r}\right)^{r-m} \quad (22)$$

(from 14)

The right hand side of this expression can be rearranged to form

$$P_m = \left(1 - \frac{\lambda}{r}\right)^r \cdot \left(\frac{\lambda}{m!}\right)^m \cdot \left(1 - \frac{\lambda}{r}\right) \cdot \left(1 - \frac{2\lambda}{r}\right) \dots \left(1 - \frac{(m-1)\lambda}{r}\right) \cdot \left(1 - \frac{\lambda}{r}\right)^{-m} \quad (23)$$

Because we are letting  $r$  tend to infinity all the terms in this expression, with the exception of the first two, tend to values of 1. This gives

$$P_m = \left(1 - \frac{\lambda}{r}\right)^r \cdot \frac{\lambda^m}{m!} \quad (24)$$

Again, because  $r$  is large, the first term in this expression is approximated by  $e^{-\lambda}$ , where  $e = 2.7183$  (the base of Napierian logarithms), giving \*

$$P_m = \frac{e^{-\lambda} \lambda^m}{m!} \quad m=0,1 \dots \quad (25)$$

which is the formula for the Poisson distribution. The mean, or expected value of  $m$  for this distribution is  $\lambda$ , and a characteristic of the Poisson is that the variance is also given by  $\lambda$ .

Because formula (25) is more easily evaluated than (18), the Poisson is usually chosen in preference to the binomial as the predictor of the random frequency array. For our hypothetical pattern, where  $\lambda = 21/36 = .583$  the calculation of the Poisson probabilities for the first three values of  $m$  are as follows:

$$P_0 = \frac{2.7183^{-.583} \times .583^0}{0!} = 2.7183^{-.583} = .556$$

(Note that a table for values of  $e^{-\lambda}$  is given in the appendix)

$$P_1 = \frac{2.7183^{-.583} \times .583^1}{1!} = 2.7183^{-.583} \times .583 = .333$$

$$P_2 = \frac{2.7183^{-.583} \times .583^2}{1 \times 2} = .083$$

The remaining large values of  $P_m$  and  $P_{m \cdot n}$  for this problem are listed in Table 1. Usually, for any probability distribution, it is not necessary to calculate probabilities for values of  $m$  where  $P_{m \cdot n}$  is less than 0.1. The closeness of the Poisson's approximation for the binomial is illustrated by the small differences between the respective results listed in Table 1. However, as a general rule, the Poisson will only give an acceptable approximation if all the following conditions are fulfilled:

$$\begin{aligned} n &\geq 50 & (26) \\ p &< 0.1 & (27) \\ \lambda &< 5.0 & (28) \end{aligned}$$

The deductions leading to the establishment of the random probability model presented here are based on results in Feller (1957) and Gray (1967, p.52). However, in the literature of quadrat analysis alternative derivations of the Poisson and the binomial models may be found in Rogers (1974, p.3) and Greig-Smith (1964, p.12). They begin by assuming that the study area is capable of being sub-divided into an infinite number of quadrats, that is space is continuous, their deductions then lead to the establishment of the Poisson as the pure random model, with the binomial being a limiting case of the Poisson capable of predicting minor departures from randomness. Essentially, our derivation is appropriate for quadrat censuring where the study area is divided into a discrete number of cells, while the alternative 'continuous space' derivation is suitable for quadrat sampling. However, in practical applications of these models the differences between the two predictions are usually too small to influence the interpretation of results.



### III. GOODNESS-OF-FIT TESTS

As the name suggests, goodness-of-fit tests are statistical procedures for deciding whether a model prediction gives a close enough representation of an observed problem for the assumptions of the model to be accepted as giving an adequate explanation for the processes controlling the problem. We have already mentioned the existence of statistical procedures for testing observed nearest neighbour distances for randomness, and in this section we will be concerned with the more commonly used goodness-of-fit procedures in quadrat analysis. The example used is the appropriateness of the Poisson model. However, with the exception of the variance/mean ratio, the tests may be applied to all the probability distributions that occur in quadrat analysis.

#### (ii) The Chi-square test

This is the most commonly used goodness-of-fit statistic which tests the degree of correspondence between two frequency distributions grouped into identical classes. The formula for the test statistic is,

$$\chi^2 \text{ (w - 1 d.f.)} = \sum_{j=1}^w \frac{(O_j - E_j)^2}{E_j} \quad (29)$$

where  $O_j$  is the observed frequency in the  $j^{\text{th}}$  class,  $E_j$  is the expected, or predicted frequency in the  $j^{\text{th}}$  class, and  $w$  is the total number of classes. The test is made with  $w - 1$  degrees of freedom (d.f.). Inspection of the formula will reveal that  $\chi^2$  tends to zero as the correspondence between the observed and expected frequencies becomes perfect, and tends to be large as the fit becomes poorer. In quadrat analysis the expected frequency distribution is the prediction from the probability model ( $P_m \cdot n$ ) and the observed frequency distribution is the data obtained from the quadrat census or sample ( $n_m$ ). Thus we may rewrite (29) as:

$$\chi^2 \text{ (w - 1 d.f.)} = \sum_{m=0}^w \frac{(n_m - (P_m \cdot n))^2}{P_m \cdot n} \quad (30)$$

One of the problems with  $\chi^2$  is that its value is sensitive to small frequencies, which tend to inflate the value of the statistic. Clearly an expected frequency of zero will send the value of  $\chi^2$  to infinity. Consequently, for tests using the  $\chi^2$  distribution to be meaningful some convention is needed to place a lower limit on the number of observations in a frequency class. Statisticians disagree on the precise form of this convention, although the current opinion is that in quadrat analysis it is possible to reduce the lower limit to 1 without serious loss of accuracy (see Rogers (1974, p.68)). In order to test the null hypothesis that there is no significant difference between the frequency array shown in Fig.2(ii)a and the predicted Poisson probabilities, the data in Table 1 was arranged for the  $\chi^2$  test. Following the grouping criterion it is necessary to merge the frequencies for  $m = 2, 3, 4$  into a single class of  $m \geq 2$ . The results of the subsequent  $\chi^2$  test are shown in Table 2. The test is made at the  $\alpha = .05$  significance level which means that the probability of making Type I errors (wrongly rejecting the null hypothesis ( $\alpha$ )) is also .05, because by definition  $p = \alpha$ . In this instance we accept the null hypothesis and conclude that the observed pattern has a random frequency distribution.

Table 2:  $\chi^2$  test

$m$	$n_m$	$P_m \cdot n$ (Poisson)	$\frac{(n_m - (P_m \cdot n))^2}{P_m \cdot n}$
0	20	20.1	.001
1	12	11.7	.008
$\geq 2$	4	4.2	.009

Calculated  $\chi^2$  (2 d.f.) = .018

$H_0: f(n_m) = f(P_m \cdot n) \quad p = .05 \quad \alpha = .05$

Critical  $\chi^2$  (2 d.f.) = 5.991

Decision: Accept  $H_0$

Many statisticians, for example Mead (1974), are sceptical about the value of the  $\chi^2$  test in the quadrat analysis, and this doubt concerns the power of the test in such applications. The power of a test is defined as  $1 - \beta$ , where  $\beta$  is the probability of making Type II errors (wrongly accepting the null hypothesis). Generally the  $\chi^2$  test is regarded as being low powered, and consequently application of the test will often fail to detect a genuine difference between the observed and expected frequency distributions.

A second problem that arises with  $\chi^2$  also concerns the probability of Type II errors. For most statistical tests the null hypothesis is the converse of the actual idea that is being tested such that the research worker is hoping to reject the null hypothesis and so confirm his original idea. Tests are designed in this way because statisticians are conservative and will only accept an idea when there is a low probability ( $\alpha$ ) of the difference being due to sampling error. However, this is not the usual case in quadrat analysis. Normally the research worker wishes to demonstrate a close correspondence between the observed and expected distributions by accepting the null hypothesis. It follows that if the research worker is to maintain a conservative statistical approach he must select a significance level which minimises the probability of Type II errors. Therefore the severity of the test described by Table 2 would be increased if the significance level were also increased, say to  $p = .20$ , because raising the significance level to higher probabilities reduces the value of  $\beta$ . For a full discussion of the Chi-square test in quadrat analysis the reader is referred to Rogers (1974, Ch.5).

#### (ii) The Kolmogorov-Smirnov D statistic

A simple, powerful, and often used alternative to the  $\chi^2$  test is the Kolmogorov-Smirnov D statistic. The statistic measures goodness-of-fit by testing the maximum deviation between the predicted cumulative frequency distribution and the observed cumulative frequency distribution for a significant difference. Assuming that each observation ( $m_i$ ) is part of a random quadrat sample of size  $n$ , then the magnitude of the deviation is dependent solely on  $n$ , such that the statistic D may be defined as

$$D(n \text{ d.f.}) = \frac{\max |Cu.P_m.n - Cu.n_m|}{n} \quad (31)$$

Where  $Cu.P_m.n$  is the predicted cumulative frequency distribution and  $Cu.n_m$  is the observed cumulative frequency distribution. The statistic is used to test the null hypothesis

$$H_0 : D = 0.0 \quad (32)$$

against the alternative

$$H_1 : D > 0.0 \quad (33)$$

at a predefined significance level. Tables for critical values of  $D$  at a variety of significance levels may be found in Siegel (1956) and Lindgren (1975). The test has been applied to the observed and Poisson data listed in Table 1 and the results at the  $p = .20$  significance level are listed in Table 3. The test produces the same conclusion as Chi-square, that is the observed distribution is adequately described by the Poisson model.

Table 3. The Kolmogorov-Smirnov D statistic

m	(1) Cu.P <sub>m</sub> .n (Poisson)	(2) Cu.n <sub>m</sub>	(1) - (2)
0	20.1	20	.1
1	31.8	32	.2
2	35.2	35	.2
3	35.9	36	.1
4	36.0	36	0

$$\text{Calculated } D_{(36 \text{ d.f.})} = .2/36 = .006$$

$$H_0 : D = 0.0 \quad p = .20 \quad \alpha = .20$$

$$\text{Critical } D_{(36 \text{ d.f.})} = .179$$

Decision : Accept  $H_0$

The  $D$  statistic possesses the advantage over  $\chi^2$  that the full range of values of  $m$  are considered in the evaluation of the test statistic, although a corresponding disadvantage is the loss of detail associated with the  $D$  statistic because only the maximum deviation contributes to the final value of  $D$ . Again the problem arises with the  $D$  statistic that the researcher is usually hoping to accept the null hypotheses, which makes the minimization of Type II errors of paramount importance in the selection of a significance level. Fig.3 summarises the results of some experimental work by Lindgren (1975) on the relationship between Type I and Type II errors for the  $D$  statistic when the sample size ( $n$ ) is 10. It can be seen that the probability  $\beta$  falls as  $\alpha$  increases, and for this particular experiment the rate of decline in the value of  $\beta$  tails off when  $\alpha$  is approximately equal to .20. Naturally, the overall magnitude of  $\beta$  for any  $\alpha$  would be reduced if the sample size were increased, and it is generally accepted that the selection of a significance level of  $p = .20$  produces a sensible compromise in the attempt to keep both  $\alpha$  and  $\beta$  tolerably low.

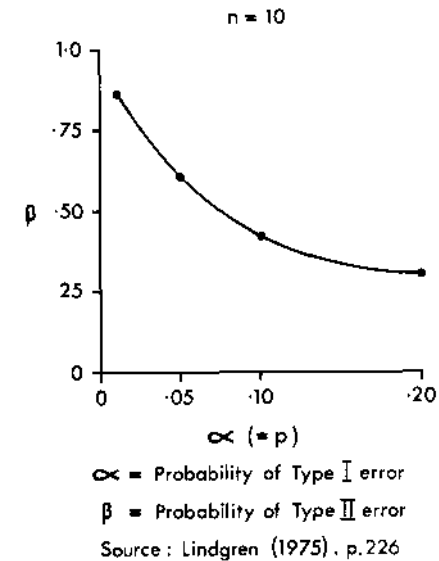


Fig.3: The relationship between Type I and Type II errors for the  $D$  statistic

(iii) The Variance/mean ratio

For an observed frequency array the variance/mean ratio is given by

$$s^2_m / \bar{m} = (5)/(4) \quad (\text{see p.6}) \quad (34)$$

The ratio defined by (34) is used in the construction of a test statistic for assessing whether the observed frequency array was generated by a random process. The test is founded on the property of the Poisson distribution that its mean equals its variance. Consequently, if the observed frequency array is random the observed variance/mean ratio should tend to a value of 1, which gives the null hypothesis

$$H_0 : s^2_m / \bar{m} = 1.0 \quad (35)$$

When the variance/mean ratio is derived from a random quadrat sample the difference between the observed ratio and unity may be tested for significance with a two-tailed  $t$  test with  $n-1$  degrees of freedom. The standard error of this difference is given by  $\sqrt{2/(n-1)}$  such that the calculated value of  $t$  is

$$\text{calc. } t_{(n-1 \text{ d.f.})} = \frac{s^2_m / \bar{m} - 1.0}{\sqrt{2/(n-1)}} \quad (36)$$

Applying this test to our example data produces the results listed in Table 4. Again we conclude that the observed frequency array was generated by a random process. If, in another application of the test, the null hypothesis were rejected  $H_1$  would be accepted and we would conclude that the point pattern

Table 4. The variance/mean ratio

$$\bar{m} = .583 \quad s^2_m = .593 \quad s^2_m/\bar{m} = 1.016$$

$$\text{Calculated } t_{(35 \text{ d.f.})} = \frac{1.016 - 1.0}{\sqrt{2/35}} = .067$$

$$H_0 : s^2_m/\bar{m} = 1.0 \quad p = .05 \quad \alpha = .05$$

$$H_1 : s^2_m/\bar{m} = 1.0$$

$$\text{Critical } t_{(35 \text{ d.f.})} = 2.03$$

Decision: Accept  $H_0$ , reject  $H_1$

is tending to be either, a clustered distribution, or, uniformly distributed in space. These secondary characteristics of the variance/mean ratio are discussed fully in the following section.

None of the tests described here are sufficient measures of goodness-of-fit in their own right, and it is common practice to apply a number of tests before deciding if a model gives a sufficiently good prediction of the observed frequency distribution. Furthermore, the statistical theory for all these tests is based on the premise that the frequency data were collected by random quadrat sampling. If the data are derived by censusing then the tests are not really appropriate, although in practice they are often applied to census data as a guide to interpretation. The three tests we have described were selected because they are the most widely applied. However, numerous other goodness-of-fit tests for quadrat analysis have been constructed and the reader is referred to Greig-Smith (1964) and Mead (1974) for interesting discussions of their performance.

#### IV. DEPENDENCE IN SPACE

##### (i) Uniform and clustered patterns

So far we have limited our discussion to the way in which randomness, or independence manifests itself in space. However, of far greater intrinsic interest to the geographer is discovering how non-randomness arises in point patterns by attempting to model the processes that give rise to these patterns.

Figure 4 illustrates some extreme examples of non-random patterns. Patterns (a) and (b) are examples of uniform patterns; (a) depicts a triangular distribution of points, while in (b) the points are located on the corners of a square lattice. Uniform patterns are usually regarded as being diagnostic of a competitive process such that points compete for space in the plane. Thus if we imagine a uniform pattern evolving through time then the presence of a mint in plane will have the effect of lowering the probability of subsequent points being located in its immediate vicinity. A typical example of a uniform point pattern is the distribution of settlements in fairly evenly populated regions. Here there is competition between towns for market areas and consequently the settlements repel each other to create a fairly uniform distribution.

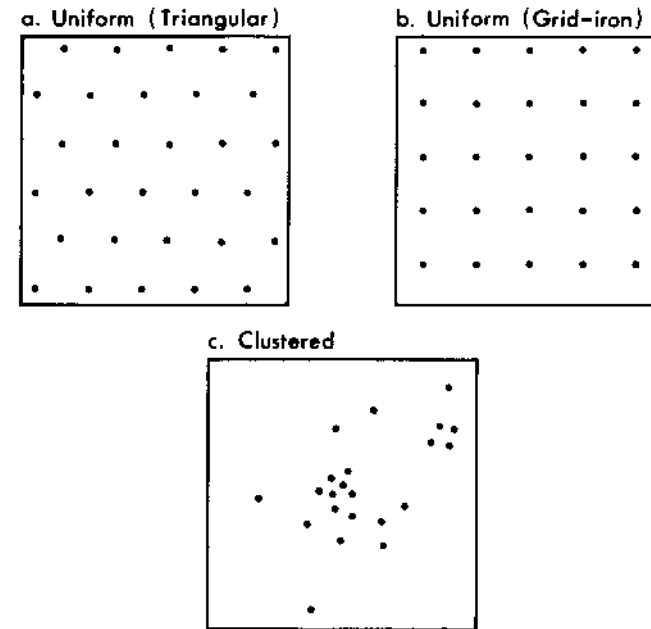


Fig. 4: Some ordered point patterns

If we take a quadrat census of a completely uniform pattern such that there are equal numbers of points in each cell, then the mean of this census will be identical to each observation ( $m_i$ ), and because all  $m_i$  are equal the variance will be zero. Consequently, the value of the variance/mean ratio will also be zero. Accordingly, values of the variance/mean ratio between one and zero are indicative of uniformity, which becomes more pronounced as the ratio tends to zero.

Clustered point patterns (Fig. 4c) are thought to be the result of contagious processes where, as the pattern evolves, the location of a point in a cell increases the probability of subsequent points being located in that cell. Phenomena that diffuse through time and space are usually governed by contagious processes. For example, Hagerstrand's (1967) renowned work on the adoption of agricultural innovations by Swedish farmers demonstrated that the decision by a farmer to adopt an innovation was the result of verbal contact with a farmer already making use of the innovation. Given that social contacts in rural areas tend to be made over short distances then, if we study the distribution of adopters among a rural population of farmers, the initial patterns of adoption are likely to be highly clustered (see Harvey, 1966). If we envisage an infinite plane where an infinite number of points tend to cluster in the same location then the variance of the resulting frequency array will tend to infinity together with the value of the variance/mean ratio.

(ii) The negative binomial distribution

In geographical applications of quadrat analysis clustered distributions have been found to be far more prevalent than uniform distributions. For this reason, the majority of this section will be concerned with probability distributions which model the clustering of points. In particular we shall be most concerned with the properties of the negative binomial because it is with this distribution that geographers have had the greatest success in fitting observed frequency arrays.

The negative binomial is derived in the following manner. Assume that points are assigned to an infinite grid independently of time. However, as distinct from the assumptions of the binomial, we now assume that the probability of a point being placed in a specified cell increases linearly with the number of points already placed in that cell. Consequently, at any one point in time, the probabilities of cells receiving a point are not equal, but are directly related to the existing distribution. Naturally this process will generate a clustered distribution, and it can be proved (see Rogers, 1974, p.16) that probability distribution embodying these assumptions, such that the probability of a specified cell containing exactly  $x$  points is, given by

$$P_{(X=m)} = P_m = \binom{k+m-1}{m} \cdot \left(\frac{p}{1+p}\right)^m \cdot \left(\frac{1}{1+p}\right)^k \quad m=0,1,\dots \quad (37)$$

$$\text{where } \lambda = kp \quad (38)$$

$$\sigma^2 = kp(1+p) \quad (39)$$

$$\text{consequently, } p = \lambda/k \quad (40)$$

It can be seen that the probabilities predicted by equation (37) are dependent solely on the values given to the two parameters  $\lambda$  and  $k$ . We have already defined  $\lambda$  as the point density. Not surprisingly,  $k$  measures the degree of clustering associated with the contagious process, and its exact value is specified within the limits zero to infinity. As  $k$  tends to infinity the clustering disappears and the negative binomial tends to the Poisson distribution, and as  $k$  approaches zero the distribution converges on an exceptionally clustered logarithmic distribution (see Bliss and Fisher, 1953). Inspection of the variance/mean ratio for this distribution ((39)/(38)) shows that its value must always be greater than one, irrespective of the values of  $\lambda$  and  $k$ . When  $k$  is not a positive integer, which is usually the case, the probabilities for the negative binomial are obtained by solving the following density function which is an approximation for (37).

$$P_{(m=0)} = 1/(1+p)^k \quad (41)$$

$$P_m = \frac{(k+m-1)R}{m} \cdot P_{(m-1)} \quad m = 1,2,\dots \quad (42)$$

$$\text{where } R = p/(1+p) \quad (43)$$

The distribution described by equation (37) may be deduced from a number of different premises concerning the processes giving rise to the clustering (see Dacey, 1968). For quadrat analysis the two most important processes which give rise to the negative binomial are termed generalized and compound processes. The formal mathematical description of these two processes is quite complex, however, the basic distinction between the two is quite easily

understood. The preceding derivation of the negative binomial is based on the assumption that the distribution is a result of a generalized process. Here the clustering is the result of some basic affinity between the points being studied. Compound processes are the result of some basic inhomogeneity in the population of points. For instance, if in our innovation adoption example the density of the farm population varied significantly over the study area, we could observe clustering in the distribution of adopters not because of short distance social contacts between farmers, but because there were high densities of farmers in lowland areas and low densities in upland areas. The clustering would be a result of the parameter  $\lambda$  varying over the study area, not the result of a 'genuine' contagious process. Because these two sets of assumptions lead to the same predicted frequency distribution the design of a quadrat sampling experiment must make clear whether the generalised or compound model is appropriate. The generalised distribution is the more precisely defined model in geographical terms, but for its assumptions to hold the research worker must be confident that both the points and the study area are fairly homogeneous in nature.

(iii) Moments and maximum likelihood estimation of  $k$

Ideally, when we test the validity of the assumptions of a probability model as an explanation for the processes controlling a point pattern, we should possess sufficient knowledge of these processes to specify the values of the model parameters from a priori reasoning. For example, if we wished to fit the negative binomial to the agricultural innovation problem we should know the values of  $\lambda$  and  $k$  that are appropriate in the test situation. However, this is rarely the case, and usually research workers are forced to adopt statistical estimation procedures as a substitute for theoretical reasoning. The replacement of deductive reasoning with inductive statistical procedures at a crucial stage in the analysis is the major weakness in the logic of the quadrat method. The situation is made more complicated by the availability of a number of different estimation procedures, and we need to distinguish carefully between their respective properties. The two most commonly used methods are moments and maximum likelihood estimation.

Statistical estimation procedures are designed such that some predefined property of the data is preserved in the model prediction. When we fitted the Poisson distribution to the hypothetical data set (Table 1) we used formula (3) to obtain an estimate for the Poisson's single parameter  $\lambda$ . By adopting this procedure the mean ( $\lambda$ ), or first moment, of the Poisson prediction is made equal to the mean ( $\bar{m}$ ) of our data set. Naturally enough such procedures, which ensure that one or more of the model's moments is equal to its observed value, are known as moments estimation. For the negative binomial these procedures are complicated by the fact that two parameters have to be estimated. Briefly, the moments estimation is designed to find values of  $\lambda$  and  $k$  such that the first and second moments of the negative binomial prediction are equal to the mean and variance of the data set. The following equations give the required estimates (denoted by  $\hat{\lambda}$  over the parameter)

$$\hat{\lambda} = \bar{m} \quad (44)$$

$$\hat{k} = \bar{m}^2 / (s_m^2 - \bar{m}) \quad (45)$$

These equations demonstrate that the logical outcome of the moments procedure is that both  $\lambda$  and  $k$  are estimated using the observed mean and variance. Table 5(i) shows the results of fitting the moments estimate of the negative binomial to the frequency distribution of farmers who adopted T.B. control between 1900-24. It may be noted that mean and variance of the predicted negative binomial are equal to the observed mean and variance.

Table 5: Fitting the negative binomial to data for the adoption of T.B. control in S. Sweden (1900-24). Source: Harvey (1966) after Hågerstrand (1967)

m	(i)	(ii)	(iii)
	Negative Binomial (moments estimation)	Negative Binomial (Maximum likelihood estimation)	Observed frequency
	(P <sub>m</sub> ·n)	(P <sub>m</sub> ·n)	(n <sub>m</sub> )
0	62.4	62.9	63
1	15.7	15.1	15
2	6.3	6.2	6
3	2.9	2.9	3
4	1.4	1.4	1
5	-	-	1
6	-	-	1
≥5	1.3	1.5	
$\hat{\lambda}$	.567	.567	$\bar{m}$ .567
$\hat{\sigma}^2$	1.282	1.338	$s_m^2$ 1.282
$\hat{k}$	.449	.416	-
calc. $\chi^2$ (5 d.f.)	.546	.292	-
crit. $\chi^2$ (5 d.f.)	11.1	11.1	-
p = .05			

The fundamental principle in maximum likelihood estimation is that we obtain estimates of the model parameters such that the observed frequency distribution is predicted as closely as is possible by the corresponding model probabilities. Mathematically this procedure may be defined as finding that particular value of the model parameter  $y$  which maximises the value of the following likelihood function:

$$L(y_1, y_2, \dots, y_h) = \prod_{m=0}^w (P_m)^{n_m} \quad (46)$$

where  $w$  is the largest observed value of  $m$  and  $P_m$  are the predicted probabilities for the observed frequency distribution  $n_m$ . Table 6 illustrates the arithmetical properties of the likelihood function. Under the row  $n_m$  is given the observed frequency array for a pattern consisting of 4 points and 6 cells. Assume we want to fit some hypothetical probability distribution to this data set and that the distribution has one parameter  $y$ . We wish to find the maximum likelihood estimate for  $y$ . When  $y$  is estimated to be some specific value  $y_1$ , the model probabilities  $\{P_m(y_1)\}$  exactly predict the

Table 6: Evaluating the likelihood function

m	0	1	2	$L(y_1)$
$n_m$	3	2	1	
$P_m(y_1)$	.500	.333	.167	.002314
$P_m(y_2)$	.490	.343	.167	.002311
$P_m(y_3)$	.510	.323	.167	.002311

observed frequency array and the value of the likelihood function is,  $L(y_1) = (.500^{3.0}) \times (.333^{2.0}) \times (.167^{1.0}) = .002314$ . When  $y$  is estimated to be of a value either  $y_2$  or  $y_3$  the model probabilities slightly under-predict and over-predict the largest frequency ( $n_0$ ) respectively. For both these cases the value of the likelihood function is less than we obtained with  $y_1$ . This example demonstrates that the value of the likelihood function is maximised when the parameter estimate provides a perfect fit between the model probabilities and the data. Therefore, the value of  $y_1$  is the maximum likelihood estimate.

In practice probability models are usually unable to give a perfect prediction for the observed frequencies irrespective of the parameter values. In such cases the maximum likelihood estimate of the parameter value will give a less than perfect prediction. However, it is always true that maximum likelihood estimation will give the best prediction possible, and for this reason the method is preferred to moments estimation.

The main problem in using maximum likelihood estimation is that finding the parameter value which maximises (46) is usually a complex procedure both algebraically and arithmetically. The solution is simple only for the Poisson and binomial distributions because in these cases the only parameter to be estimated is the mean ( $\lambda$ ), and it turns out that the maximum likelihood estimate of the mean is identical to the moments estimate. However, if other model parameters need to be estimated from the data the maximum likelihood procedures are awkward. For instance, in order to obtain the maximum likelihood estimate of the negative binomial  $k$  parameter it is necessary to carry out a complex iterative procedure which is too involved to incorporate in this monograph. The reader who wishes to follow up this topic is referred to a paper by Bliss and Fisher (1953) which gives a clear account of maximum likelihood estimation for the negative binomial  $k$  parameter.

Table 5 (ii) shows the results of fitting the negative binomial to the innovation adoption data by maximum likelihood methods, and the results illustrate well the different principles involved in the two estimation procedures. The maximum likelihood solution improves the goodness-of-fit, which is indicated by the lower  $\chi^2$  value. It is noticeable that the greatest improvement in fit has been obtained for the two largest frequency classes and this occurs because the likelihood function is especially sensitive to large values of  $n_m$ . Indeed, maximum likelihood estimation may crudely be described as

'modal' estimation. It may also be noted that the maximum likelihood estimate does not preserve the value of the observed variance in the model prediction.

In the present example both estimation procedures give good fits and the respective  $\chi^2$  tests indicate that we should accept the null hypothesis and conclude that the frequency distribution of innovation adopters is the result of a contagious process. However, in many applications of the negative binomial the two estimation procedures yield widely differing results, and in such instances it is wise to select the maximum likelihood estimate of k.

(iv) Other distributions

Many other probability distributions exist for describing certain types of dependency between points, and most of these distributions can be derived as mixtures of the three basic models we have already discussed: the binomial, the Poisson, and the negative binomial. One interesting example is the Neyman A distribution (Neyman, 1939) which may be derived as a compound model resulting from the mixture of two Poisson processes (see Rogers, 1974). This model assumes that if clusters of points are laid down randomly in space such that the average number of points per cluster also follows a Poisson distribution, then the probability of finding exactly x points in a specified cell is given by

$$P_{\{x=m\}} - P_{(m)} = \frac{e^{-va} a^m}{m} \cdot \sum_{i=0}^{\infty} \frac{i^m}{i!} (ve^{-a})^i \quad m = 0, 1, \dots \quad (47)$$

$(0^0 \equiv 1)$

with a mean and variance defined respectively by

$$\lambda = va \quad (48)$$

$$\sigma^2 = va(a + 1) \quad (49)$$

The moments estimates for a and v are given by

$$\hat{a} = \{s^2_m - \bar{m}\} / \bar{m} \quad (50)$$

$$\hat{v} = \bar{m}^2 / \{s^2_m - \bar{m}\} \quad (51)$$

Neyman derived this distribution specifically to model the distribution of insect larvae crawling away from recently hatched egg clusters. He assumed that the egg clusters were distributed randomly in space and also that mean number of eggs per cluster also followed a random distribution. Here the parameter 'a' measures the mean number of clusters per unit area, and v the mean number of eggs per cluster. Clearly this distribution has a highly specialised derivation and, to date, although geographical applications of the distribution have been suggested (see Harvey, 1966), the model has not provided adequate enough predictions for its assumptions to be considered as a suitable process explanation for these problems.

One of the few examples of a probability distribution capable of describing uniformity in point patterns, with variance/mean ratios less than one has been developed by Dacey (1964). This model was derived to model the distribution of large towns (population greater than 2,500) in Iowa. In this example the cells of the census are Iowa's counties which were originally delimited, fortuitously it seems, as a square grid. Dacey assumed that two different processes should account for the frequency distribution of Iowa towns. County seats, with populations greater than 2,500, will tend to be distributed uniformly one to a cell, while all other urban places will be randomly distributed and follow the Poisson model. Consequently, large county seats are distributed with a density p which is, by definition, restricted to values between 0 and 1, while all other urban places are distributed with a density y. The overall average number of towns per county is denoted as usual by  $\lambda$ .

Dacey has shown that the distribution which satisfies these assumptions, such that the probability that a specified county contains x towns

from both populations, is given by

$$P_{\{x=m\}} = P_m = \frac{(1-p)y^m e^{-y}}{m!} + \frac{pmy^{(m-1)} e^{-y}}{m!} \quad m=0, 1, \dots \quad (52)$$

with a mean and variance defined by

$$\lambda = y + p \quad (53)$$

$$\sigma^2 = \lambda - p^2 \quad (54)$$

If y and p are unknown their values may be given by the following moments estimates

$$\hat{p} = \sqrt{\bar{m} - s^2_m} \quad 0 \leq \hat{p} \leq 1 \quad (55)$$

$$\hat{y} = \bar{m} - \hat{p} \quad (56)$$

Notice that as p approaches 0, indicating that all county seats have populations less than 2,500, the first term in (52) tends to the Poisson distribution with a mean equal to y, while the second term tends to 0 and vanishes. Alternatively, as p tends to 1 the model's variance/mean ratio, (54)/(53), becomes increasingly smaller to indicate that the predicted frequency array is representing an increasingly more uniform pattern. This model has been found to fit the frequency distribution of Iowa towns for all U.S. census periods between 1840-1950. However, Dacey's distribution is similar to the Newman A in that its assumptions restrict geographical applications to its original derivation. This is also true of some other contagious probability models which have made brief appearances in the geographical literature on quadrat analysis such as the Thomas double Poisson and Polya-Aeppli distributions (see Olsson, 1967). \*

V. GEOGRAPHICAL APPLICATIONS AND THEIR PROBLEMS

The existing geographical applications of quadrat analysis have tended to deal with two types of point pattern. The first involves the modelling of contagious processes in human populations, while the second type may be classified as attempts to explain and describe structural features of the landscape in homogeneous regions. An example of the first problem is Harvey's (1966) attempt to fit the negative binomial and other contagious probability models to the distribution of agricultural innovations adopters in S. Sweden, work which we described in the previous section. Similarly, Reynolds (1974) has fitted the negative binomial to the distribution of voters for winning mayoral candidates in Indianapolis. Reynolds suggests that social contacts within a voter's immediate urban environment will tend to influence his voting decision in favour of the dominant political party in that area, which results in a clustered distribution of party voters. However, structural applications are the more prevalent in geography, and these include the analysis of karst depressions in limestone regions (LaValle 1967; McConnell and Horn, 1972), factory and shop distributions in urban areas (Thomas and Reeve, 1976; Rogers, 1965 and 1969c; Sibley, 1972) and the distribution of houses in Puerto Rico (Dacey, 1968). Here we will review some of these diverse applications while paying particular attention to the two methodological problems which influence any application of quadrat analysis: model specification, and the influence of quadrat size on the results - known as the scale problem.

(i) Karst depressions and the problem of model specification

Conflicting evidence relating to the processes that control the frequency distribution of karst depressions in limestone regions reported in the works of Lavalley and McConnell and Horn provides an interesting example of how the failure to specify the parameters of a probability model on a priori grounds may lead to quite different interpretations of the same results.

Karst depressions are of two major types: dolines and collapse sinks. Dolines tend to be small features, usually ranging in depth between 3 and 10 metres, and are formed above the water table both by solution along zones of weakness in the rock and by the ponding of surface run-off. Collapse sinks are the result of cavern roof collapse and their distribution is dependent on the subterranean drainage system. McConnell and Horn proposed that the distributions of both types of depression would be controlled by a random process such that a double Poisson model would account for the frequency distribution of all depressions. We have already seen that the operation of two random processes at different densities leads to a clustered distribution, and in this instance a double Poisson model developed by Schilling (1947) was selected as being the most appropriate. Alternatively, Lavalley suggested that the following process was the more likely. Independently of time, the random occurrence of a karst depression leads to an increased probability of subsequent depressions being formed in that cell because, around the original depression, local erosional processes would be accelerated by increased diversion of run-off into the subterranean drainage system. This second description leads to the selection of the negative binomial as the appropriate probability model for the frequency distribution of karst depressions.

Table 7 illustrates the results of fitting the negative binomial, by maximum likelihood estimation, and Schilling's double Poisson, by moments estimation, to McConnell and Horn's data for the distribution of karst depressions in the Mitchell Plain, Indiana. The Kolmogorov-Smirnov D statistic shows that the negative binomial gives an adequate description at the  $p = .15$  significance level, while the double Poisson gives a slightly better fit at the  $p = .20$  significance level. This situation, where two theories derived from different assumptions are found to fit the same data set, is termed complementarity. The problem arises because in both instances the model parameters are estimated from the data. Both models make assumptions about the location of points in time and naturally the model parameters are deduced from these assumptions; indeed, Feller (1943) has concluded that it is impossible to distinguish between two contagious distributions on the single criterion of the observed frequency array. Clearly, further information on the evolution of the observed pattern is required before any confident conclusions can be drawn from this analysis.

However, given these reservations, it is possible to marshal both statistical and geomorphological evidence in support of the double Poisson as the more plausible explanation of karst depression formation. McConnell and Horn argue that the presence of a depression in a cell is unlikely to increase the probability of further depressions being found in that cell. They suggest that the diversion of run-off in the cell increases the efficiency of the surface drainage system and serves to enlarge the original depression rather than cause the formation of new depressions. Moreover, the acceptance of the double Poisson is made with a lower probability of Type II errors than the negative binomial. Finally, it may be noted that the products of the estimated parameters of the double Poisson,  $b_1y_1 = .27$  and  $b_2y_2 = 1.87$  are the probable number of collapse sinks and dolines respectively to be found in any one cell.

Table 7: The distribution of karst depressions in S. Indiana

m	Observed Frequency	Negative Binomial	Double Poisson	Geometric
0	112	102.3	107	105.9
1	53	70.4	64	69.0
2	35	46.5	37	45.0
3	43	30.3	29	29.3
4	22	19.6	26	19.1
5	23	12.6	19	12.5
6	6	8.1	11	8.1
7	5	5.2	6	5.3
8	3	3.3	2	3.4
9	1	2.1	1	2.2
10	1	1.3	-	1.4
Parameter Estimates		$\hat{k} = 1.090$ $\hat{\lambda} = 1.87$	$\hat{b} = .5481$ $\hat{b}_1 = .4519$ $\hat{y}_1 = .4839$ $\hat{y}_2 = 3.5510$	$\hat{\lambda} = 1.87$
D (calc)		.0641	.0263	.0523
D (crit, $p = .20$ )		.0614	.0614	.0614
D (crit, $p = .15$ )		.0654	.0654	.0654
Redundancy	.0342	-	-	-

(ii) The scale problem

Any application of quadrat methods is affected by the scale problem because the selection of quadrat size is always an arbitrary procedure, and it is often true that the particular scale of analysis selected may influence the subsequent interpretation of results. In particular, if the hypothesis of randomness in the observed pattern is to be accepted, it must be shown that either the Poisson or the binomial fits the observed frequency array at a variety of different scales. If either of these models do not fit at any one scale, then the hypothesis of randomness must be rejected for all scales and an alternative model of dependence sought. Similarly, when fitting models of dependence the research worker must show that the model's parameter values do not vary significantly with changes in quadrat size, otherwise scale is influencing the interpretation of results in some unknown manner. Dacey (1968) after

successfully fitting the negative binomial to house distributions in Puerto Rico was able to demonstrate that the model's parameters were fairly stable over three different scales.

It has been argued that the scale problem is intrinsically useful to geographers because it encourages them to study the influence of space on the processes controlling a pattern. However, to date, the results of such research have not been very illuminating. Rogers (1974), in a study of shop distributions in Ljubljana, Yugoslavia, was able to demonstrate that different quadrat sizes were appropriate for different shop type distributions. For any distribution the optimal quadrat size was defined as that which maximizes the power of the  $\chi^2$  test for a given level of significance'. (p.108). However, although Rogers identified optimal scales he was unable to interpret their geographical significance.

One of the most well-known, and ingenious treatments of scale effects is the approach taken by Greig-Smith (1952). His scheme is to test for randomness at a variety of scales within a square quadrat census where the number of cells on each axis is some power of 2. The test is again based on the property of the Poisson distribution that its mean equals its variance, and is designed as a hierarchical analysis of variance. Figure 5 illustrates an example of a 4 x 4 census that could be used for the test. The data are the number of points in each cell (mi). The census is partitioned into a set of nested blocks by dividing the whole census in half, each half is divided in half, etc., keeping the halves as near square as possible, until the individual cells form the final blocks. In the 4 x 4 case 4 divisions are required (see fig.5). Let  $X_j(i)$  be the total number of points located in the  $i$ th block of size  $j$ , where  $j$  is the number of cells in that block. We may now define the variation (or sums of squares) between blocks of size  $j$  nested within blocks of size  $2j$  as

$$SS_j = \sum_{i=1}^{n/j} \left( X_j(i) \right)^2 / j - \sum_{i=1}^{n/2j} \left( X_{2j}(i) \right)^2 / 2j \quad (57)$$

The total sum of squares for all  $n$  cells is

$$TSS = SS_1 + SS_2 + SS_4 + \dots + SS_{n/2} \quad (58)$$

and the degrees of freedom for sums of squares between blocks of size  $j$  nested within blocks of size  $2j$  is given by,

$$\text{degrees of freedom } SS_j = n/2j \quad (59)$$

with the corresponding variance estimates ( $MS_j$ ) equal to

$$MS_j = 2jSS_j/n \quad (60)$$

The evaluation of all these formulae for an example problem is shown in Figure 5. The ratios  $MS_2/MS_1, MS_4/MS_1, \dots, MS_{n/2}/MS_1$  are variance ratios, each of which may be tested for significant departures from randomness using the F distribution. The test is made at the desired significance level with  $n/2$  degrees of freedom for the denominator ( $MS_1$ ) and  $n/2j$  degrees of freedom for the numerator.

The analysis of variance for the hypothetical data set illustrated in Figure 5 indicates that at the  $p = .05$  significance level the pattern is accepted as being random for all scales, with the variation between quarters within halves displaying the greatest tendency towards non-randomness. If, for any particular scale, the null hypothesis is rejected, then this is

i. DATA

0	0	2	0	0	1	2	1	1	3	8	8
0	1	0	1	2	5	3	2	7	5		
0	3	2	0								
2	2	1	2								
j=1				j=2				j=4		j=8	

ii. CALCULATING THE SUMS OF SQUARES (Equation (57))

$$SS_8 = \left( \frac{8^2}{8} + \frac{8^2}{8} \right) - \frac{16^2}{16} = 16 - 16 = 0$$

$$SS_4 = \left( \frac{1^2}{4} + \frac{3^2}{4} + \frac{7^2}{4} + \frac{5^2}{4} \right) - 16 = 21 - 16 = 5$$

$$SS_2 = \left( \frac{9^2}{2} + 2 \left( \frac{1^2}{2} \right) + 3 \left( \frac{2^2}{2} \right) + \frac{3^2}{2} + \frac{5^2}{2} \right) - 21 = 24 - 21 = 3$$

$$SS_1 = \left( 7(0^2) + 3(1^2) + 5(2^2) + 3^2 \right) - 24 = 32 - 24 = 8$$

$$TSS = \frac{16}{16}$$

iii. VARIANCE TABLE

Source	Sum of squares (SS <sub>j</sub> )	Degrees of freedom (n/2j)	Variance estimate (ms <sub>j</sub> )	Calc F (ms <sub>j</sub> /ms <sub>1</sub> )	Critical F, p=.05	Decision
Between halves (j=8)	0	1	0.0	0.0	5.318	Accept
Between quarters (j=4) within halves	5	2	2.5	2.5	4.450	Accept
Between eighths (j=2) within quarters	3	4	0.75	0.75	3.838	Accept
Between cells (j=1) within eighths	8	8	1.0	-	-	-

iv. VARIANCE ESTIMATE PLOTTED AGAINST BLOCK SIZE (j)

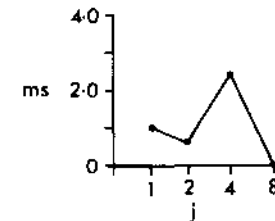


Fig. 5: Hierarchical analysis of variance for scale effects



evidence for clustering at that scale, and Greig-Smith has suggested that the size of quadrat at that scale will be related to the 'mean area of clumping' in the pattern. The form of the test described here does not measure tendencies towards uniformity in the pattern. However, alternative significance tests for this property do exist (see Mead, 1974).

The main statistical problem with the F test is that once the hypothesis of randomness has been rejected for a single scale the frequency distribution can no longer be assumed to follow the Poisson distribution. Consequently the tests at all other scales become invalid. When this is the case a useful qualitative interpretation of scale effects can be gained from the graph of variance estimates plotted against scales (see fig. 5 (iv)). Nevertheless, the sustained interest of statisticians in the Greig-Smith procedure is an indication of its importance, and recently both Mead (1974) and Zahl (1974) have suggested improvements to the basic method.

VI ALTERNATIVE APPROACHES

(1) Multinomial coefficients and state descriptions

We have seen that the major problem in quadrat analysis is that an observed point pattern contains insufficient information to test all the assumptions of the probability models. This difficulty prompted the author to develop an alternative approach to quadrat analysis based solely on the information contained in the quadrat census (Thomas and Reeve, 1976). The method is based on the assumptions of a probability distribution known as Bose-Einstein statistics, which provides an alternative definition of equal likelihood, or randomness to those embodied in the binomial and Poisson distributions.

It will help our exposition if we first redefine some of the terminology of quadrat analysis. Remembering that a point pattern consists of an arrangement of  $r$  points among  $n$  cells, we define  $n$  and  $r$  as macro-state descriptions of the point pattern. A meso-state description of the pattern is any frequency array whose individual elements ( $n_m$ ) satisfy the macro-state description such that

$$\sum_{m=0}^r n_m = n \tag{1}$$

and

$$\sum_{m=0}^r n_m \cdot m = r \tag{2}$$

Clearly there will be a number of meso-states which satisfy these conditions, and two of the meso-states associated with  $n = 9$  and  $r = 5$  are shown in Figure 6. Lastly we can define a micro-state description as any arrangement of the  $r$  points among the  $n$  cells. Obviously there will be a large number of micro-states associated with any macro-state description, and a smaller number associated with any one meso-state description. Indeed, it can easily be proved (see Gray, 1967, p.97) that the total number of micro-states associated with any macro-state description is given by the binomial coefficient

$$\binom{n+r-1}{r} \tag{61}$$

i **MACRO-STATE** :  $n=9, r=5$ ; Total no of micro-states =

$$\binom{n+r-1}{r} = \binom{13}{5} = 1287$$

ii **MESO-STATES** (two examples)

(a)		(b)	
m	$n_m$	m	$n_m$
0	6	0	5
1	2	1	3
2	0	2	1
3	1		

$$\sum_{m=0}^r n_m = 9$$

$$\sum_{m=0}^r n_m m = 5$$

iii **MICRO-STATES**

Two of the

$$W = \frac{9!}{6! 2! 0! 1!} = 252$$

micro-states associated with (a)

Two of the

$$W = \frac{9!}{5! 3! 1!} = 504$$

micro-states associated with (b)

0	1	0	0	0	0	0	0	0	2	0	0
0	0	3	0	0	0	0	0	1	0	0	0
0	1	0	1	1	3	1	1	2	1	1	1

Fig. 6: State descriptions and their properties

Further, the total number of micro-states associated with a meso-state description is given by the multinomial coefficient

$$W = \frac{n!}{r \prod_{m=0}^r n_m!} \quad (62)$$

Examples of all these definitions and properties are given in Figure 6.

(ii) The entropy maximising distribution

We can now derive a probability distribution for predicting the most likely frequency distribution (meso-state). We wish to find that particular meso-state description that can create the greatest number of micro-state descriptions, because this description will be the most likely in the absence of controls on the pattern. Mathematically, this requires us to obtain the meso-state description which maximises the value of equation (62) subject to the conditions (1) and (2). Such a meso-state will be the most likely because it can arise in the greatest number of ways. Moreover, the solution to this problem will be an entropy-maximising solution because  $\ln(W)$  is one of the definitions of entropy or uncertainty. It is useful to re-state this problem in terms of the probability distribution which, when multiplied through by  $n$ , will predict the most likely meso-state. Here we wish to find that probability distribution whose individual probabilities ( $P_m$ ) maximise the value of Shannon's measure of the entropy of a probability distribution (Shannon and Weaver, 1949) given by

$$H_r + 1 = \sum_{m=0}^r P_m \log(1/P_m) \quad (63)$$

subject to the conditions

$$\sum_{m=0}^r P_m = 1 \quad (7)$$

and

$$\sum_{m=0}^r P_m \cdot m = r/n \quad (8)$$

Because  $H$  is linearly related to  $\ln(W)$  the aims of these two problems are identical.

We now demonstrate that the Bose-Einstein probability distribution gives the solution to these problems. We know that the total number of micro-states associated with  $n$  and  $r$  is given by (61). It can also be proved (see Feller, 1957) that the total number of times  $m$  points are found in a specified cell of each of the  $\binom{n+r-1}{r}$  micro-states is given by the binomial coefficient

$$\binom{n+r-m-2}{r-m} \quad (64)$$

This value holds for any cell because the problem is symmetrical.

Consequently, if we wish to find the probability that a specified cell contains exactly  $x$  points when all micro-states associated with  $n$  and  $r$  are assumed to be equally likely, then the probability  $x = m$  is given by

$$P_{(x=m)} = P_m = \binom{n+r-m-2}{r-m} / \binom{n+r-1}{r} \quad m = 0, 1, \dots, r \quad (65)$$

This is probability distribution known as Bose-Einstein statistics. It follows by conjecture that, because this distribution assumes all micro-states associated with the macro-state description to be equally likely, then, if we multiply the successive probabilities in (65) through by  $n$ , we will obtain the meso-state description of the pattern that maximises the value of (62) subject to (1) and (2). Similarly, the probabilities themselves will maximise the value of (63) subject to (7) and (8). Simulated verification of these conjectures is given in Thomas and Reeve (1976).

If we let  $r$  and  $n$  tend to infinity such that the average number of points per cell ( $\lambda$ ) becomes constant, then it can be shown that the geometric distribution given by

$$P_m = \lambda^m / (1 + \lambda)^{m+1} \quad m = 0, 1, \dots \quad (66)$$

gives a close approximation to the probabilities obtained from equation (65). Indeed, the approximation holds for quite small values of  $n$  and  $r$ . It is interesting to note that the geometric distribution is the discrete form of the negative exponential distribution (where  $m$  is a continuous variable) which is the entropy function in Wilson's (1970) family of spatial interaction models.

By abandoning the assumption that points are placed in cells independently of one another through time, we have deduced a probability distribution that gives a quite different prediction for the most likely frequency array than the binomial or Poisson. The difference between the binomial and the Bose-Einstein distribution arises because the binomial model assumes the points are distinguishable from one another at the micro-state level. For example, although an interchange of a pair of points in one of the micro-states in Figure 6 leaves the form of the micro-state unaltered, under the binomial definition of equal likelihood each possible interchange of a pair of points creates another permissible, distinguishable distribution of points. In fact, under the assumption of the binomial there are  $n^r$  equally likely distributions which is always much larger than the  $\binom{n+r-1}{r}$  equally likely micro-states assumed in the Bose-Einstein model.

The choice of a random probability model for a particular problem depends on the amount of information that is available to the research worker. If the observed point pattern is the only information then the Bose-Einstein model is appropriate, but if the evolution of the pattern can be traced over time the Poisson or binomial will be appropriate.

For the karst depression data listed in Table 7 the Bose-Einstein assumptions are appropriate and the geometric distribution has been found to fit the observed frequency array at the  $p = .20$  level of the Kolmogorov-Smirnov  $D$  statistic. This result implies the observed frequency distribution is the most likely we could expect given the information we have available. Moreover, because their assumptions involve time, neither the negative binomial nor double Poisson can be substantiated as plausible explanations for the processes controlling the pattern on the present evidence.

(iii) Redundancy

The entropy-maximising property of the Bose-Einstein model enables us to calculate Shannon's indices of relative entropy and redundancy for any observed frequency distribution. Relative entropy is the ratio between the observed entropy of the frequency array and its maximum possible entropy. Using equation (63) we can define this quantity as.

$$\text{Relative Entropy} = \frac{\sum_{m=0}^r n_m/n \log (n/n_m)}{\sum_{m=0}^r P_m \log (1/P_m)} \quad (67)$$

where  $(n_m/n)$ 's are the observed probabilities and  $P_m$  are the probabilities generated by either equation (65) or (66). Relative entropy takes on values between zero and one and measures the extent to which the observed pattern is free to vary. The calculation of the relative entropy can be simplified by using the expression,  $\log(1 + \lambda) + \lambda \log \{(1 + \lambda)/\lambda\}$  to calculate the maximum entropy which is the denominator of (67). This formula for the maximum entropy is based on results presented in Webber (1976). Redundancy is the complement of relative entropy and is defined as

$$\text{Redundancy} = 1 - (67) \quad (68)$$

Redundancy measures the extent to which the pattern is controlled by some unknown processes which we will term rather clumsily as 'unhypothesised information'. Again the index takes on values between zero and one. A redundancy of zero indicates a perfect correspondence between the observed frequency array and the Bose-Einstein prediction. A redundancy of one occurs when the pattern is totally controlled, and this upper limit is achieved by totally uniform and, under certain conditions, totally clustered patterns. Such patterns are totally controlled by unhypothesised information and may easily be reproduced by simple mechanical rules.

Unlike other indices of point pattern dispersion, redundancy does not discriminate between clustered and uniform patterns, because in their extreme cases both tend to a redundancy of one. However, a theoretical property of the entropy-maximising geometric distribution helps remedy this deficiency. When the negative binomial  $k$  parameter takes on a value of one the negative binomial is identical to the geometric distribution. Therefore, in cases where the Bose-Einstein definition of equal likelihood is appropriate, we can define all observed patterns with estimated  $k$  parameters (from equation (45)) of less than one as tending to be clustered, and all other patterns as tending to be uniform. Incidentally, the fact that the negative binomial is equivalent to the geometric distribution when  $k$  equals one, vividly illustrates the degree of difference in the prediction of the most likely frequency array between the geometric and the Poisson models, remembering that the Poisson is equivalent to a negative binomial with  $k$  equal to infinity.

Redundancy possesses two distinct advantages over other indices of point pattern dispersion. Firstly, it is based on a simpler definition of equal likelihood and therefore has a more general range of applications. Secondly, values of redundancy are not influenced by the value of the point density  $(\lambda)$ .

Table 8: Merseyside Factory Location: 1966

Standard Industrial Classification (1958)	Redundancy	$\hat{k}$ (max. likelihood)	$s_m^2/\bar{m}$	$\lambda$	$r$	$n$
Ship-building(7)	.137	.076	2.359	.097	40	412
Vehicles (8)	.061	.082	1.524	.051	21	412
Clothing and Footwear (12)	.497	.044	39.208	.430	177	412

Other indices, such as the variance/mean ratio and the negative binomial  $k$  parameter are dependent on  $\lambda$ , which makes them difficult to use when one wishes to compare properties of different point patterns.

Table 8 is designed to illustrate the merits of the redundancy index in a comparative situation. The results are based on Factory Inspectorate Records of industrial plant location in Greater Merseyside, 1966 (see fig.1(i)). Point patterns of factories for three industrial classes have been analysed on a 412 kilometre square census covering the built-up area of the conurbation. Here the assumptions of the Bose-Einstein model are appropriate because no additional information is available on the evolution of the patterns. The subsequent redundancy values are easy to sustain from geographical reasoning. Vehicle factories are large independent concerns which have few factors influencing their location at the intra-urban scale. Consequently, the low redundancy and the close fit of the frequency array with the geometric distribution (not illustrated) are hardly surprising. The one obvious constraint on ship-building firms is a coastal location. However, at the one square kilometre scale a large number of cells in Greater Merseyside have coastal locations, suggesting that the coastal locations appear to exert a minor degree of control at this scale. Again the redundancy of .137, indicating a slight degree of clustering, sustains this interpretation. Conversely, clothing firms are small scale and highly inter-dependent with historically centralised urban locations, and these locational characteristics are well illustrated by the high degree of control indicated by the .497 redundancy. The interpretation of the variance/mean ratios and the negative binomial  $k$  parameters for these patterns do not appear to be so intuitively reasonable. Here all three patterns are interpreted as having clustered distributions, with conflicting estimates for the degree of clustering exhibited by vehicles and shipbuilding occurring between the  $k$  parameter and the variance/mean ratio. These last results illustrate well the difficulties of using density dependent indices in a comparative analysis.

Finally, because redundancy is invariant with point density, we can use the index to measure directly the influence of cell size on the organisation of the pattern. If we analyse a pattern successively at a predetermined number of block sizes ( $j$ ) then we can define the index,

$$G = \max.R_j - \min.R_j \quad (69)$$

where  $R_j$  is the redundancy for the  $j$ th block size.  $G$  then measures the minimum influence of quadrat size on the pattern within the selected values of  $j$ . We define  $G$  as the minimum influence because cell size effects could well be contained within the quantity  $\min.R_j$ .

**Table 9:** Redundancy as a measure of scale effects

S.I.C. (1958)	j = 1 n = 412	2 206	4 103	16 26	G
Ship-building	.137	.105	.200	.343	.238
Vehicles	.061	.036	.028	.045	.033
Clothing and Footwear	.497	.484	.484	.463	.034

Table 9 illustrates the application of the G index to the Merseyside factory location data at four different block sizes. We see that only ship-building seems to be influenced by cell size, with a G index equal to .238. Here the redundancy of the frequency array becomes progressively larger as the block size is increased and we can interpret this result using Greig-Smith's idea of the 'mean area of clumping'. Because relatively fewer cells are coastal as the block size increases, the frequency array becomes more clustered as proportionately more points appear in single cells. Consequently, the scale with the highest redundancy will be the optimum block size for the identification of major clusters of ship-building firms.

## VII. CONCLUSIONS

Point patterns are one of the most elementary ways of representing a geographical variable. Yet we have seen that by making quite simple assumptions about relationships between the points we can deduce quite sophisticated probability models for predicting observed frequency distributions. Perhaps one of the most intriguing aspects of quadrat analysis is the interplay between mathematical, statistical and geographical reasoning associated with the interpretation of results.

The mathematics of quadrat analysis is concerned with the deduction of logically consistent probability models from a given set of assumptions. Two diametrically opposite problems are embodied in the process of deduction. First it is possible to arrive at the same model from different assumptions. This is the case with compound and generalised versions of contagious probability models. Conversely, we may obtain quite different models for essentially the same idea, by making apparently minor modifications to the original assumptions. We found this to be the case with the binomial and Bose-Einstein descriptions of equal likelihood. These theoretical problems require considerable geographical understanding of a particular pattern before sensible choices can be made between different assumptions and models. The most vital judgements are qualitative and not quantitative.

The statistical problems associated with the quadrat method involve more subjective judgements. The selection of powerful goodness-of-fit tests and efficient estimating procedures may still lead to erroneous conclusions if insufficient information is available to test all the assumptions of the model. One of the more recent innovations in the geographical applications of quadrat analysis has been the testing of bivariate probability models. Here two different patterns are modelled simultaneously and Rogers and Martin (1971)

have tried to fit a number of these models to point patterns of shops and urban populations. However, here again the problem of model specification becomes even more intractable because the number of parameters to be estimated increases with the number of variables. When designing quadrat analysis experiments one is usually faced with a choice between complex theories which are difficult to test, and simple indices which are useful for comparative studies but which have limited explanatory power. Nevertheless, the reciprocal relationship between theory and data which characterises the quadrat method often provides useful insights into many geographical problems.

## BIBLIOGRAPHY

### A. Theoretical and Statistical

- Anscombe, F.J. (1950), Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37, 358-82.
- Bliss, C.I. & Fisher, R.A. (1953), Fitting the negative binomial to biological data, and note on the efficient fitting of the negative binomial. *Biometrics* .9 , 176-200.
- Bliss, C.I. & Owen, A.R.G. (1958), Negative binomial distributions with a common k. *Biometrika*, 45, 37-58.
- Clark, P.J. & Evans, F.C. (1954), Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, 35, 445-453.
- Dacey, M.F. (1964), Modified Poisson probability law for point patterns more regular than random. *Annals of the Association of American Geographers*, 54, 559-565.
- Dacey, M.F. (1966), A county seat model for the areal pattern of an urban system. *Geographical Review*, 56, 527-45.
- Feller, W. (1943), On a general class of contagious distributions. *Annals of Mathematical Statistics*, 14, 389-400.
- Feller, W. (1957), *An Introduction to Probability Theory and its Applications*. Part I (New York: Wiley)
- Fisher, R.A. (1941), The negative binomial distribution. *Annals of Eugenics*, 11, 182-7.
- Gray, J.R. (1967), *Probability*. (Edinburgh and London: Oliver and Boyd)
- Greig-Smith, P. (1952), The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany (New Series)*, 16, 293-312.
- Greig-Smith, P. (1964), *Quantitative Plant Ecology*, 2nd Edition. (London: Butterworths)
- Gurland, J. (1958), A generalized class of contagious distributions. *Biometrics*, 14, 229-49.
- Holgate, P. (1965), Some new tests for randomness. *Journal of Ecology*, 53, 261-66.
- Katti, S.K. & Gurland, J. (1962), Efficiency of certain methods of estimation for the negative binomial and Neyman A distributions. *Biometrika*, 49, 215-26.

- Kershaw, K.A. (1957), The use of cover and frequency in the detection of pattern in plant communities. *Ecology*, 38, 291-99.
- Matern, B. (1960), Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden Fran Statens Skogsforskningsinstitut, Band 49*, 1-144.
- McConnell, H. (1966), Quadrat methods in map analysis. *Discussion Paper No. 3, Department of Geography, University of Iowa*.
- Mead, R. (1974), A test for spatial pattern at several scales using data from a grid of contiguous quadrats. *Biometrics*, 30, 295-307.
- Moore, P.G. (1953), A test for non-randomness in plant populations. *Annals of Botany (New Series)*, 17, 57-62.
- Morisita, M. (1959), Measuring the dispersion of individuals and analysis of the distributional patterns. *Memoirs of the Faculty of Science Kyushu University, Series E, 2*, 215-235.
- Neyman, J. (1939), On a new class of contagious distribution, applicable in entomology and bacteriology. *Annals of Mathematical Statistics*, 10, 35-37.
- Ord, J.K. (1970), The negative binomial and quadrat sampling. in: *Random Counts in Scientific Work: Vol. 2, Random counts in biomedical and social sciences*, ed G.P. Patil, (University of Pennsylvania Press), 15-63.
- Quenouille, M.H. (1949), A relation between logarithmic, Poisson and negative binomial series. *Biometrics*, 5, 162-64.
- Rogers, A. (1969a), Quadrat analysis of urban dispersion: 1. Theoretical techniques. *Environment and Planning*, 1, 47-80.
- Rogers, A. & Gomar, N. (1969b), Statistical inference in quadrat analysis. *Geographical Analysis*, 1, 370-84.
- Rogers, A. & Martin, J. (1971), Quadrat analysis of urban dispersion: 3. Bivariate models. *Environment and Planning*, 3, 433-50.
- Rogers, A. & Raquillet, R. (1972), Quadrat analysis of urban dispersion: 4. Spatial Sampling. *Environment and Planning*, 4, 331-45.
- Rogers, A. (1974), *Statistical analysis of spatial dispersion*. (London: Pion)
- Schilling, W. (1947), A frequency distribution represented as the sum of two Poisson distributions. *Journal of the American Statistical Association*, 42, 407-24.
- Skellam, J.G. (1952), Studies in statistical ecology. 1. Spatial patterns. *Biometrika*, 39, 346-62.
- Thomas, M. (1949), A generalization of Poisson's limit for use in ecology. *Biometrika*, 36, 18-25.
- Thomas, R.W. & Reeve, D.E. (1976), The role of Bose-Einstein statistics in point pattern analysis. *Geographical Analysis*, 8, 113-36.
- Thompson, H.R. (1955), Spatial point processes, with applications to ecology. *Biometrika*, 42, 102-15.

- Thompson, H.R. (1958), The statistical study of plant distributions using a grid of quadrats. *Australian Journal of Botany*, 6, 322-42.
- Zahl, S. (1974), Applications of the S-method to the analysis of spatial pattern. *Biometrics*, 30, 513-24.
- B. Applications
- Dacey, M.F. (1968), An empirical study of the areal distribution of houses in Puerto Rico. *Transactions of the Institute of British Geographers*, 45, 51-70.
- Getis, A. (1964), Temporal analysis of land-use patterns with the use of nearest neighbour and quadrat methods. *Annals of the Association of American Geographers*, 54, 391-399.
- Gleason, H.A. (1920), Some applications of the quadrat method, *Bulletin, Torrey Botanical Club*, 47, 21-33.
- Harvey, D.W. (1966), Geographical processes and the analysis of point patterns. *Transactions of the Institute of British Geographers* 40, 81-95.
- Harvey, D.W. (1967), Some methodological problems in the use of the Neyman Type A and negative binomial probability distributions for the analysis of spatial point patterns. *Transactions of the Institute of British Geographers*, 42, 81-95.
- LaValle, P.D. (1967), Geographical processes and the analysis of karst depressions within limestone regions. (Abs) *Annals of the Association of American Geographers*, 57, 794.
- McConnell, H. & Horn, J.M. (1972), Probabilities of surface karst. in: *Spatial analysis in geomorphology*, ed R.J. Chorley, (London: Methuen), 111-34.
- Olsson, G. (1968), Complementary models: a study of colonization maps. *Geografiska Annaler*, 50B, 115-32.
- Reynolds, D.R. (1974), Spatial contagion in political influence processes. in: *Locational approaches to power and conflict*, ed K.R. Cox, D.R. Reynolds and S. Rokkan, (New York: Wiley), 233-74.
- Rogers, A. (1965), A stochastic analysis of the spatial clustering of retail establishments. *Journal of the American Statistical Association*, 60, 1094-1103.
- Rogers, A. (1969c), Quadrat analysis of urban dispersion: 2. Case studies of urban retail systems. *Environment and Planning*, 1, 155-71.
- Sibley, D. (1972), Strategy and tactics in the selection of shop locations. *Area*, 4, 151-56.
- C. General reading and related topics
- Cliff, A.D. (1968), The neighbourhood effect in the diffusion of innovations. *Transactions of the Institute of British Geographers*, 44, 75-84.
- Cliff, A.D. & Ord, J.K. (1973), *Spatial Autocorrelation*, (London: Pion).
- Geary, R.C. (1954), The contiguity ratio and statistical mapping. *Incorporated Statistician*, 5, 115-41.

Hagerstrand, T. (1967), *Innovation diffusion as a spatial process*. (Translated by A. Pred), (University of Chicago Press)

Harvey, D.W. (1968), Pattern, process and the scale problem in geographical research. *Transactions of the Institute of British Geographers*, 45, 71-78.

Hepple, L.W. (1974), The impact of stochastic process theory upon spatial analysis in human geography. in: *progress in Geography*, ed C. Board, R.J. Chorley, P. Haggett and D.R. Stoddart, 6, 89-142.

King, L.J. (1969), *Statistical analysis in geography*, (Englewood Cliffs: Prentice-Hall).

Lindgren, B.W. (1975), *Basic ideas of statistics*, (New York: Macmillan).

Olsson, G. (1967), Central place systems, spatial interaction and stochastic processes. *Papers, Regional Science Association*, 18, 13-46.

Siegel, S. (1956), *Nonparametric statistics*, (New York: McGraw-Hill).

Shannon, C.E. & Weaver, W. (1949), *The mathematical theory of communication*, (Urbana: University of Illinois Press).

Webber, M.J. (1976), Elementary entropy maximising probability distributions: analysis and interpretation. *Economic Geography*, 52, 218-27.

Wilson, A.G. (1968), Notes on some concepts in social physics. *Papers, Regional Science Association*, 22, 159-93.

Wilson, A.G. (1970), *Entropy in urban and regional modelling*. (London: Pion).

APPENDIX

$\lambda$	VALUES OF $e^{-\lambda}$ ( $0 < \lambda < 1$ )									
	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
.1	.9048	.8958	.8869	.8781	.8694	.8607	.8521	.8437	.8353	.8270
.2	.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
.3	.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
.4	.6703	.6637	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
.5	.6065	.6005	.5945	.5886	.5827	.5769	.5712	.5655	.5599	.5543
.6	.5488	.5434	.5379	.5326	.5273	.5220	.5169	.5117	.5066	.5016
.7	.4966	.4916	.4868	.4819	.4771	.4724	.4677	.4630	.4584	.4538
.8	.4493	.4449	.4404	.4360	.4317	.4274	.4232	.4190	.4148	.4107
.9	.4066	.4025	.3985	.3946	.3906	.3867	.3829	.3791	.3753	.3716

$\lambda$	$e^{-\lambda}$									
	1	2	3	4	5	6	7	8	9	10
0.0	.36788	.13534	.04979	.01832	.00674	.00248	.00091	.00034	.00012	.00005

( $\lambda = 1, 2, \dots, 10$ )

NOTE: To obtain values of  $e^{-\lambda}$  for other values of  $\lambda$ , use the law of exponents.

EG.  $e^{-2.41} = (e^{-2.06})(e^{-0.41}) = (.13534)(.6637) = .08983$