Some Statistical Aspects of Spatial Distribution Models for Plants and Trees

PETER DIGGLE

Department of Operational Efficiency
Swedish University of Agricultural Sciences
S-770 73 Garpenberg, Sweden

Abstract

ODC 524.31:562.45/46

The paper is an account of some statistical analyses carried out in conjunction with a silvicultural research project in the department of Operational Efficiency, College of Forestry, Garpenberg, Sweden (Eriksson, L & Eriksson, O in preparation 1981).

In Section 2, a statistic due to Moran (1950) is used to detect spatial interaction amongst counts of the numbers of young trees in sample plots laid out in a systematic grid arrangement.

Section 3 discusses the construction of bivariate distributions for the numbers of wild and planted trees in a sample plot.

Section 4 considers the relationship between the distributions of the number of trees in a plot and their total basal area.

Section 5 is a review of statistical methods for use in connection with preliminary surveys of large areas of forest. In particular, Section 5 discusses tests of spatial randomness for a single species, tests of independence between two species, and estimation of the number of trees per unit area.

Preface

The origin of this report dates back to 1978. At this time Dr Peter Diggle was employed as a guest-researcher at the College. The aim of Dr Diggle's work at the College was to introduce advanced mathematical and statistical models in forestry research.

Dr Diggle was specially employed to develop some biological models for a planning system for silvicultural treatments and for various cutting systems. To tackle this problem a research-team has been working to develop an advanced system for planning the above mentioned activities. Since Dr Diggle left Sweden as a guest-researcher he has been working as a consultant to the project. And we see in this report the results of his effort in this field.

College of Forestry
Garpenberg, March 1981
Sten Nilsson

Contents

1 Introduction

1.1 General introduction

Research workers in the department of Operational Efficiency have been working for several years to develop a planning system for silvicultural treatments and for various cutting options. The goal of this system is to provide a long-term silvicultural strategy for a forest company.

Today in Sweden, the practical consequences of, for example, regeneration or pre-commercial thinning operations are evaluated by such simple criteria as number of living plants one year after planting, number of stems after pre-commercial thinning, etc. However, the first income from a stand comes some 50 to 80 years after regeneration. It is therefore necessary to analyse the effects of a silvicultural treatment on the complete life-cycle of a stand. No such complete description is currently available.

To tackle this problem, a computerised planning system has been developed. (Eriksson, L & Eriksson, O in preparation 1981.) The two principle components of this system are
(i) a biological model
(ii) an economic model.

The biological model simulates the decelopment of a stand from regeneration to final felling. Parameters in the model can be varied to accommodate various policies for regeneration, pre-commercial cleaning, thinning, etc. Principally because of the long time-scale involved, no data are available concerning the complete development of a stand. Several different kinds of basic biological research data have therefore been used in the model-building process. These data-sets relate to investigations carried out in different departments of the university, and each such investigation has been directed towards answering relatively specific questions concerned with particular aspects of stand development.

The need to link these aspects together in the biological model has generated a number of loosely related statistical problems. This report presents a number of "sub-models" which are now used within the full biological simulation model, and it should therefore be read in conjunction with Eriksson, L & Eriksson, O (in preparation). However, we hope that some of the statistical problems discussed in the report will prove of wider interest in the general context of forestry research.

1.2 Sources of data

We make extensive use of data collected in the PTAX investigations (Hultén & Jansson 1972). These data include information on the survival or death of young planted trees and on the incidence of naturally regenerated trees in circular plots of radius 2 m, laid out systematically in a number of stands distributed throughout Sweden. The number of plots per stand

Figure 1. Typical example of sampling design for PTAX data.

varies considerably, but is typically of the order of 40 per stand, excluding plots which are unsuitable for planting. The spacing between adjacent plots also varies according to the overall area of the stand, but is typically of the order of 30 m. Figure 1 shows a typical example of the layout of sampling plots within a stand.

We also use typestand data, in particular the typestands for cleaning described in Gustafsson (1974). Each typestand consists of a 25 m by 20 m rectangle within which the locations and species of each stem are recorded, together with additional information including the cleaning grade of each stem, from 1 to 7; where:

grade   1 indicates trees to be removed to get a stand of 2800 stems/ha
grade 1–2 indicates trees to be removed to get a stand of 2400 stems/ha

grade 1–3 indicates trees to be removed to get a stand of 2000 stems/ha
grade 1–4 indicates trees to be removed to get a stand of 1600 stems/ha
grade 1–5 indicates trees to be removed to get a stand of 1200 stems/ha
grade 1–6 indicates trees to be removed to get a stand of   800 stems/ha
grade   7 indicates all remaining trees

1.3 Plan of report

In Section 2 we define the concept of spatial interaction for a set of observations recorded at fixed spatial locations, and show how a test for spatial interaction can be applied to the PTAX data in an exploratory way, to identify stands with interesting spatial structure.

In Section 3 we describe a number of probability distributions which arise in the analysis of quadrat count data. We motivate the various distributions by considering both the mode of regeneration (natural or planted) and the possible existence of environmental heterogeneity. We then discuss an extension to the bivariate case, based on a class of bivarate distributions introduced by Plackett (1965). The PTAX data again provide an illustrative application, and we note the difficulties which arise in making a combined interpretation of the results from a large number of stands.

Section 4 investigates a specific problem concerning variation in yield within a stand. In young stands, this is conventionally measured by the variance-to-mean ratio of the number of individual stems within a sample plot. In older stands, the accepted measure is the coefficient of variation of the total basal area within a sample plot. We explore the relationship between these two quantities, both by theoretical arguments and empirically, using typestand data.

Section 5 discusses the general topic of assessing spatial patterns in forest stands, using practicable, distance-based sampling procedures. Topics covered includes tests of spatial randomness for a single species, tests for association between species and robust estimation of the mean number of stems per unit area.

Section 6 offers a general discussion of the results obtained. We acknowledge the limitations of our formal analyses, and note some additional problems which, through lack of time, have not yet been tackled. Throughout the report, we shall assume that the reader is familiar with the basic concepts and terminology of mathematical statistics, including probability distributions, probability density functions, expectation, the likelihood function, parameter estimation and hypothesis testing. For detailed discussion of these, we refer the reader to one of the many intermediate texts which are available, such as Lindgren (1968) or Hogg & Craig (1970).

2 Spatial interaction

2.1 Introduction

When measurements are taken at a number of spatial locations it often happens that observed values from adjacent locations are relatively more similar, or dissimilar, than could resonably be expected to have occured by chance. Recognition of this empirical phenomenon leads to the following definition of *spatial interaction*.

Consider a finite set of random variables Y_i: $i = 1, \ldots, n$ associated with locations $\underline{x}_i$: $i = 1, \ldots, n$ in some region of the plane. The $\underline{x}_i$ are termed *sites* and the Y_i *site-variables*. The simplest possible structure for the site-variables is that they should be mutually independent and identically distributed. This defines the state of *complete spatial randomness* (henceforth CSR) for the system $\{(Y_i, \underline{x}_i): i = 1, \ldots, n\}$; any form of dependency amongst the site-variables defines a state of *spatial interaction*.

For CSR it is necessary, but not strictly sufficient, that the following hypothesis should hold:

H: the joint distribution of the Y_i is invariant under random permutation of the labels $i = 1, \ldots, n$.

In seeking to establish the existence or non-existence of spatial interaction in a given set of data, it is the hypothesis H that we shall test. The implication of H is that the sites convey no relevant information about the joint distribution of the site-variables.

It is important to recognise two qualitatively different types of departure from CSR; the site variables may be mutually independent but not identically distributed, or identically distributed but dependent. Techniques for the investigation of CSR are motivated by the latter phenomenon and, for example, we would not expect our tests to be particularly appropriate for the investigation of smooth trends in spatial data. If smooth trends are suspected, one way to identify them is by a trend surface analysis (Watson 1971). This is a form of regression analysis in which the basic assumption is that the site-variables are mutually independent, with common variance σ^2 and expectations

$$E[Y_i] = f(\underline{x}_i),$$

where $f(\cdot)$ is in some prescribed parametric class, typically a low order polynomial. Regression methods can also be used when relevant explanatory variables are available at the various sites. In almost all applications, it is known *a priori* that H is untrue. In these circumstances, rejection of H is of no intrinsic interest but is a minimal pre-requisite for any serious attempt to interpret the observed spatial distribution. In particular, we note in Section 2.2 below that the standard test for spatial interaction takes the form of a correlation coefficient between pairs of observations from neighbouring sites.

2.2 A test for spatial interaction

The form of the hypothesis H suggests that we should carry out a test of CSR by comparing the value u_1 of a statistic u calculated from the data with the values u_i: $i = 2, \ldots, m$ of the same statistic, but calculated under each of $m - 1$ independent random permutations of the Y_i amongst the $\underline{x}_i$. If $u_{(i)}$ denote the ordered u_i,

$$u_{(1)} > u_{(2)} > \ldots > u_{(m)}$$

it is then exactly true under H that

$$P\{u_1 = u_{(i)}\} = m^{-1}: i = 1, \ldots, m,$$

and the rank of u_1 provides an exact test of H. In practice, $m = 100$ is adequate for testing at conventional levels of significance (Marriott 1979) and conveniently allows attained significance levels to be quoted in percentages. For further discussion of this idea of a "Monte Carlo Test", see Barnard (1963) and Hope (1968). Besag & Diggle (1977) describe several applications to spatial data.

It follows from our earlier discussion that if a test of H against an alternative of spatial interaction is required, the statistic u should reflect dependence amongst the site-variables. The standard statistic, due to Moran (1950), therefore takes the form of a correlation coefficient between pairs of observations from neighbouring sites, in the following precise sense.

Given observations y_i from sites $\underline{x}_i$: $i = 1, \ldots, n$, let

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$$

and write

$$z_i = y_i - \bar{y}: i = 1, \ldots, n.$$

Define a connection matrix D to have off-diagonal elements

$$\delta_{ij} = \begin{cases} 1: \text{if sites } \underline{x}_i \text{ and } \underline{x}_j \text{ are } \textit{connected} \\ 0: \text{otherwise} \end{cases}$$

and diagonal elements

$$\delta_{ij} = 0$$

Further define

$$2A = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij},$$

$$2D = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \delta_{ij} \right) \left[\left\{ \left(\sum_{j=1}^{n} \delta_{ij} \right) - 1 \right\} \right]$$

and

$$b_2 = \left(n^{-1} \sum_{i=1}^{n} z_i^4 \right) \Big/ \left(n^{-1} \sum_{i=1}^{n} z_i^2 \right)^2.$$

Moran's statistic can be written as

$$u = (n/2A) \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} \, z_i \, z_j \, / \sum_{i=1}^{n} z_i^2$$

and under H

$$E(u) = -(n-1)^{-1}$$

$$E(u^2) = \left\{ \begin{matrix} n[4A(n^2-3n+3)-8(A+D)n+12A^2] \\ -b_2[4A(n^2-n)-16(A+D)n+24A^2] \end{matrix} \right\} \, / \, \{4A^2(n-1)\,(n-2)\,(n-3)\}$$

These results, due to Cliff & Ord (1973, Ch. 2), can be used to provide an asymptotically valid test of H based on a normal approximation to the distribution of u under H, without recourse to Monte Carlo randomisation. However, the Monte Carlo approach provides an exact test for small n and a useful check on the asymptotics for larger n.

The discussion so far has begged the question of how connections between sites are to be established. In general terms, it would seem reasonable to connect pairs of sites whose separation distance d_{ij} is relatively small. Distance need not be the only criterion, and Cliff & Ord (1973) discuss further possibilities. Figure 2 shows a system in which connections are established via the *Dirichlet tessellation* of the sites. This partitions the region under examination into polygonal *cells* C_i associated with the x_i, each C_i consisting of all points closer to x_i than to any other x_j. The connections in Figure 2 have been established between all pairs of sites whose corresponding cells share a common boundary. Some mathematical properties of the Dirichlet tessellation are given by Rogers (1964). Green & Sibson (1978) provide a very efficient algorithm for the automatic determination of the tessellation, which can cope with n of the order of several thousand.
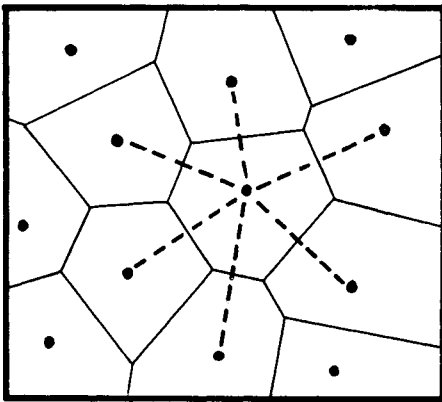

Figure 2. The Dirichlet tessellation of a set of 12 sites in a rectangular region. Connections are established between sites whose Dirichlet cells share a common boundary, and are shown, as dashed lines, for one of sites.

2.3 Application to PTAX data

We recall that in the PTAX investigations, a number of circular plots of radius 2 m are laid out in a square lattice arrangement within each of a number of stands, the spacing between plots being of the order of 30 m. It is therefore reasonable to identify the plot centres as point sites, x_i. Amongst the many possible site-variables are

Y_i = number of wild head-plants recorded at inventory

and

R_i = number of surviving head-plants recorded at inventory, from n_i initially planted.

The discussion below relates to 41 stands in central Sweden, planted in 1972 and subsequently recorded in 1975. The stands are irregular in shape, and include areas of land occupied by roads, water etc. and therefore unsuitable for planting. The number of plots without such impediment is typically of the order of 40 per stand.

For the site-variables Y_i, it is not unreasonable to test H using the observed values $z_i = y_i - \bar{y}$. On the other hand, a more reasonable assumption for the R_i is that their marginal distributions are binomial

$R_i \sim \text{Bin}(n_i; p_i)$

Strictly, the hypothesis H is not relevant in this situation: even if the R_i are mutually independent and there is no spatial variation in the environment so that all the p_i are equal, variations amongst the n_i may nevertheless induce a form of spatial interaction amongst the R_i. If the spatial distribution of the n_i is of interest it should be tested in its own right. We assume that the variable of primary interest is the proportion of survivors amongst planted trees and therefore base a test on standardised observations,

$z_i{}^* = (r_i - n_i \, \hat{p}) / \sqrt{n_i \, \hat{p}(1-\hat{p})}$

where

$$\hat{p} = \sum_{i=1}^{n} r_i / \sum_{i=1}^{n} n_i$$

is the observed proportion of survivors in the n plots. Under the assumption $p_i = p$ for all i, each $z_i{}^*$ is a realisation of a random variable with approximately zero mean and unit variance so that H is, in a sense, almost satisfied.

Connections are established between nearest neighbours in vertical, horizontal and diagonal directions. Specifically, if d_{ij} denotes the distance between sites $\underline{x}_i$ and $\underline{x}_j$, in units of the lattice spacing, we define

$$\delta_{ij} = \begin{cases} 1: & 0 < d_{ij} \le \sqrt{2} \\ 0: & \text{otherwise} \end{cases}$$

2.4 Results

Despite the attempt to produce a relatively homogeneous selection of 41 stands, the results vary widely. Figure 3 gives a simple graphical summary, based on the fact that under H, the rank of u_1 is uniformly distributed on the integers 1 to 100 inclusive. The figure shows an excess of low ranks for the statistics calculated from the planted survivors, indicative of positive spatial interaction – mortality levels in near-neighbouring plots are relatively similar. In contrast, the overall results for the wild head-plants appear to be compatible with H. Formal support for these conclusions is provided by the application of the Kolmogorov-Smirnov goodness-of-fit test (see, for
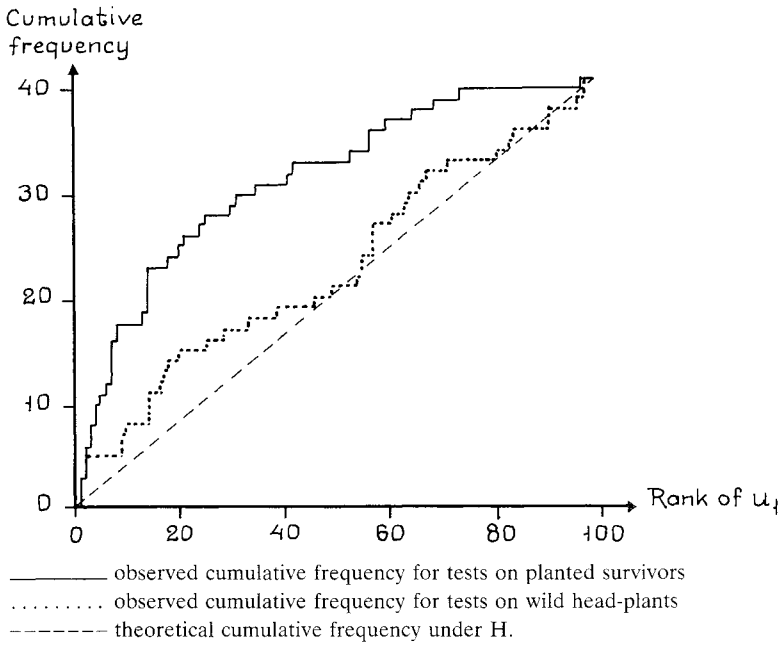
Figure 3. Tests for spatial interaction in 41 stands.

——————— observed cumulative frequency for tests on planted survivors
.... observed cumulative frequency for tests on wild head-plants
—————— theoretical cumulative frequency under H.

example, Lindgren 1968, Ch. 6). In the case of the planted survivors, the maximum absolute deviation between the observed and theoretical cumulative frequencies in Figure 3 far exceeds the 1 % critical value, whilst for the wild head-plants the corresponding maximum absolute deviation is less than the 20 % critical value.

A more detailed record of the results is given in Table 1. One disappointing feature is that there is no discernible relationship between the lattice spacing and the observed value of u_1 (Figure 4). We therefore suggest that when spatial interaction is indicated, its probable source is the more or less smooth pattern of spatial variation in some relevant, but as yet unidentified, environmental variables, rather than any direct stochastic interaction between site-variables from neighbouring plots.

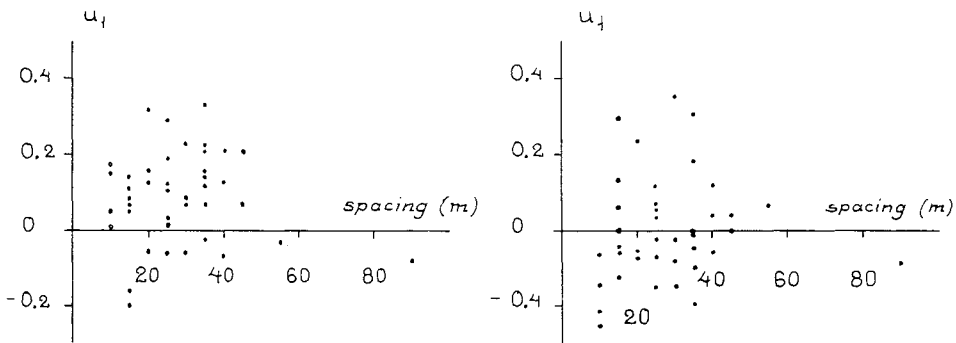(a) Planted survivors   (b) Wild head-plants

Figure 4. Values of u_1 plotted against lattice spacing.

Table 1. Tests for spatial interaction in 41 stands.

Stand number	Number of plots	Lattice spacing (metres)	Test of H applied to planted survivors		wild head-plants	
			u_1	rank*	u_1	rank*
147	34	40	− 0.071	60	− 0.047	64
149	34	35	0.332	1	− 0.007	14
150	30	35	0.206	2	0.314	2
151	33	35	0.123	7	0.003	28
152	48	35	− 0.023	53	0.185	2
154	33	30	0.082	20	− 0.143	83
159	45	40	0.201	3	0.040	25
161	50	55	− 0.026	41	0.066	10
162	44	45	0.206	2	0.042	18
165	37	10	0.050	25	− 0.137	96
217	34	15	− 0.161	97	− 0.050	57
218	45	20	0.319	1	0.237	1
219	47	25	0.012	31	0.054	20
220	22	25	− 0.061	65	− 0.154	84
221	44	30	0.067	24	0.361	1
222	25	35	− 0.068	57	− 0.192	91
223	35	35	0.148	4	− 0.040	57
224	36	20	0.129	5	− 0.053	54
225	36	20	0.158	6	− 0.068	57
226	18	25	0.121	14	0.125	9
227	21	30	− 0.058	42	− 0.083	61
228	27	35	0.225	2	− 0.189	98
229	38	45	0.074	21	− 0.002	38
230	34	15	0.074	14	0.071	14
236	27	15	0.082	13	0.140	9
237	21	10	0.011	35	− 0.213	91
238	20	10	0.147	14	− 0.250	98
241	43	20	− 0.053	57	− 0.050	55
255	42	25	0.289	1	0.071	16
256	25	30	0.227	77	− 0.023	46
257	42	10	0.169	3	− 0.056	66
258	12	15	− 0.197	74	0.000	33
260	38	15	0.140	7	− 0.118	81
262	40	15	0.110	8	0.304	1
266	28	25	0.106	14	0.041	17
269	44	25	0.020	30	− 0.069	63
270	40	15	0.047	18	− 0.039	55
277	42	25	0.188	4	− 0.018	49
297	38	40	0.130	8	0.122	14
298	44	90	− 0.079	71	− 0.077	71
300	27	35	0.136	7	− 0.091	67

* Rank of u_1 amongst u_i: $i = 1, \ldots, 100$. Low rank suggests positive spatial interaction, high rank negative spatial interaction.

To provide a specific illustration we examine in detail the planted trees in stand 223, for which $u_1 = 0.148$ and the Monte Carlo test gave a one-sided attained significance level of 4 %, only mildly indicative of positive spatial interaction. Figure 5 shows the crude proportions of survivors in the 35 plots with o indicating less than 0.05, 1 indicating 0.05 to 0.15, and so on
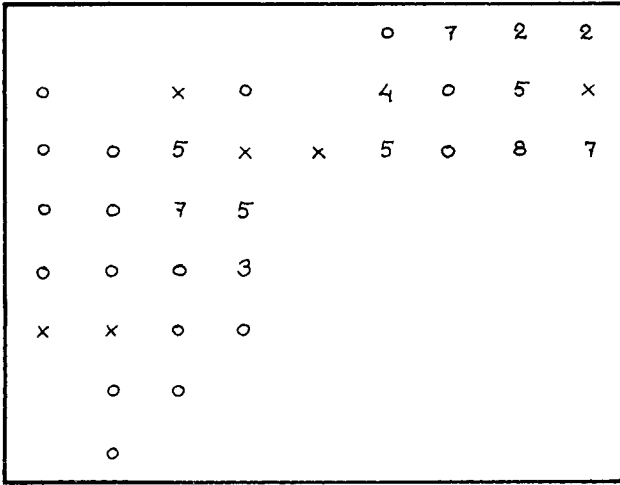
13

Figure 5. Proportions of survivors for each plot in stand 223, in tenths, cross indicating at least 95 % survival.

until x represents at least 0.95. There appears to be high mortality on the left hand side of the stand, and although this *could* be due to direct stochastic interaction between neighbouring plots it seems more plausible to interpret it as evidence of smooth spatial variation in some unidentified explanatory variable or variables.

We remark that the interpretation of this type of analysis will often be ambiguous. For, given a fixed number of site-variables and a single realisation of each, the pattern of observations will usually be compatible either with a structure of dependent, but identically distributed site-variables, or with one of independent, but non-identically distributed site-variables. The ambiguity can be resolved either by identifying relevant explanatory variables or by replication under controlled conditions, or by non-statistical arguments.

3 Quadrat count distributions

3.1 Introduction

Quadrat count analysis is a long-established tool for the quantitative description of biological populations. The basic sampling unit, or *quadrat*, is a small region of prescribed size and shape located in the population of interest according to some systematic or random sampling design. Within each of a number of such quadrats, the numbers of individuals of one or several types are recorded, leading to data of the form

$(x_i, y_i, \ldots): i=1, \ldots, m,$

where the x_i, y_i etc. denote the numbers of individuals of the specified types found in the ith of m quadrats.

The simplest form of statistical analysis for such data consists of fitting a discrete distribution to the counts of individuals of a specified type. More complex forms of analysis involve an attempt to describe patterns of dependence either between counts of different types of individual within a quadrat, or between counts of a specified type of individual in spatially adjacent quadrats. It is the former type of generalisation which we discuss here.

We remark that a quadrat was originally a one metre square, used by the Uppsala school of plant ecologists as the basic sampling unit in fieldwork (Du Rietz 1929, 1930). The term is now used to denote any small sample area, not necessarily square.

3.2 Univariate quadrat count distributions

Formally, any discrete distribution defined on the non-negative integers can be fitted to quadrat count data. However, it is desirable that the actual distributions used in a particular application should be compatible with a plausible underlying chance mechanism. With this in mind we select four classes of distribution for detailed study. For a wider discussion of discrete distributions, see Patil (1965) or Patil & Joshi (1968).

3.2.1 *The Poisson distribution*

Suppose that seeds are distributed completely at random in the plane, in the sense that

a) the average number of seeds per unit area does not vary over the plane and
b) seeds are located independently of one another.

These two postulates constitute an informal definition of a planar Poisson point process, for which the number of points (i.e. seeds) in an arbitrary

region of area A follows a Poisson distribution with mean λA, for some $\lambda > 0$; if, furthermore, each seed survives to produce a plant with probability p, independently of all other seeds, the number of plants in A again follows a Poisson distribution, with mean $p\lambda A$.

These remarks suggest that the Poisson distribution provides a natural working hypothesis for quadrat counts in naturally regenerated populations. The Poisson probability distribution is

$$p_x = e^{-\mu} \mu^x/x!: x = 0, 1, \ldots \tag{3.2.1}$$

where the single parameter $\mu > 0$ is the mean of the distribution. The variance of the distribution is also μ; this equality of mean and variance is a useful characteristic property, and is used in Section 5.3 as the basis of a test of the hypothesis that counts follow a Poisson distribution.

3.2.2 *The negative binomial distribution*

Suppose, rather informally, that (3.2.1) applies locally, but that the mean number of plants per quadrat varies over the plane because of environmental heterogeneity. If the source of this environmental variation is both known and observable, the constant μ in (3.2.1) can be replaced by a regression-like equation,

$$\mu_i = \mu(z_i, t_i, \ldots): i = 1, \ldots, m,$$

where the z_i, t_i, etc., are the values of explanatory variables observed in the ith of m quadrats. In the absence of such information, we can instead describe variation in μ as stochastic; we attach a distribution to μ and suppose that independent realisations from this distribution determine the actual values of μ_i for the m quadrats. If μ is a continuous random variable with probability density function $f(\cdot)$, then (3.2.1) is replaced by a mixed Poisson distribution.

In particular, if μ has a gamma distribution with

$$f(\mu) = (k/\lambda)^k \mu^{k-1} e^{k\mu/\lambda}/\Gamma(k): \mu \geq 0,$$

then

$$p_x = \frac{\Gamma(k + x)}{\Gamma(k) \Gamma(x + 1)} \left(\frac{\lambda}{\lambda+k}\right)^x \left(\frac{k}{\lambda+k}\right)^k: x = 0, 1, \ldots \tag{3.2.2}$$

which defines the negative binomial distribution. Here, $\lambda > 0$ is the mean of the distribution and $k > 0$ determines the variance via the equation

$$\mathrm{Var}(X) = \lambda + \lambda^2/k.$$

Thus, k^{-1} can be regarded as a measure of heterogeneity. Notice that $\mathrm{Var}(X) \to \lambda$ as $k \to \infty$, and in this limiting case (3.2.2) reverts to (3.2.1) with $\mu = \lambda$. Also, $\Gamma(\cdot)$ denotes the gamma function, which is extensively tabulated (e.g. Abramowitz & Segun 1965, Ch. 6).

The negative binomial distribution therefore provides a second possible model for quadrat counts in a naturally regenerated population and has often been used for this purpose. Evans (1953) gives a number of ecological examples. Eneroth (1945), Tirén (1949) and others describe applications in Swedish forestry research. However, it should be appreciated that the derivation given here makes no explicit reference to any underlying spatial

process and it is in fact an open question as to whether *any* realistic spatial process incorporating random heterogeneity can lead to a negative binomial quadrat count distribution. See, for example, Matérn (1971) and especially the discussion thereof.

3.2.3 *The binomial distribution*

Suppose that n plants are placed in a given region and that each plant survives with probability p, independently of all other plants. Then the number of surviving plants follows a binomial distribution, defined by

$$p_x = \binom{n}{x} p^x (1-p)^{n-x}: x = 0, 1, \ldots, n. \tag{3.2.3}$$

This suggest that the binomial distribution is a plausible model for quadrat counts in an artifically regenerated population for which initial planting records are available.

3.2.4 *The beta-binomial distribution*

As in (3.2.2) we can argue that (3.2.3) might apply locally, but that the probability p will vary over the plane because of a heterogeneous environment. It might then be reasonable to postulate stochastic variation in p, leading to a mixed binomial distribution. In particular, if p has a beta distribution with

$$f(p) = \{B(a, b)\}^{-1} p^{a-1} (1-p)^{b-1}: 0 \leq p \leq 1,$$

we obtain the beta-binomial distribution,

$$p_x = \binom{n}{x} B(a+x, b+n-x) / B(a, b): x = 0, 1, \ldots, n \tag{3.2.4}$$

where $B(\cdot)$ defines the beta function,

$$B(a, b) = \lceil(a)\lceil(b) / \lceil(a+b).$$

Following Griffiths (1973), a convenient re-parameterisation of (3.2.4) is to

$$\Theta = (a+b)^{-1}; \pi = a\Theta; \varrho = 1 - \pi$$

in which case the mean of the distribution (3.2.4) is $n\pi$. Also, (3.2.3) is obtained as a special case with $\Theta = 0$; thus, Θ is a measure of heterogeneity.

General comments made in (3.2.2) about the device of postulating stochastic variation in the parameter of a distribution, and in particular the warnings about the uncertain relationship of (3.2.2) to any underlying spatial process, remain applicable to the distribution (3.2.4). All that can be said concerning the choice of the gamma and beta distributions in deriving (3.2.2) and (3.2.4) respectively is that each provides a flexible, two-parameter class of distributions which can assume many different shapes. In the case of the gamma distribution, $f(\cdot)$ is a monotone decreasing function for $k \leq 1$, is unimodal and positively skewed for $k > 1$, and becomes symmetric in the limit $k \rightarrow \infty$. For the beta distribution, $f(\cdot)$ may be monotone increasing ($a > 1$, $b \leq 1$ or $a = 1$, $b < 1$), monotone decreasing ($a \leq 1$, $b > 1$ or $a < 1$, $b = 1$), uniform ($a = b = 1$), unimodal ($a > 1$, $b > 1$) or U-shaped ($a < 1$, $b < 1$).

3.3 Bivariate distributions

In developing a flexible class of bivariate distributions which could be used as models for bivariate quadrat counts, we should like to proceed via plausible mechanistic assumptions concerning the way in which dependence between two types of individual might arise. This appears to be difficult, and we are aware of no successful investigations along these lines. We therefore adopt a purely statistical approach, and ask whether we can provide a class of bivariate distributions which incorporates two univariate distributions of the type discussed in Section 3.2, together with one or more additional parameters describing the nature of the dependence between the two sets of quadrat counts.

Let random variables X and Y denote the counts of two types of individual in a particular quadrat and suppose that the *marginal distributions* of X and Y have been specified as

$$p_x = P\{X=x\}: x=0, 1, \ldots$$

and

$$q_y = P\{Y=y\}: y=0, 1, \ldots$$

Write the *marginal distribution functions* of X and Y as

$$F(x) = \sum_{t=0}^{x} p_t = P\ \{X \leq x\}: x=0, 1, \ldots$$

and

$$G(y) = \sum_{t=0}^{y} q_t = P\ \{Y \leq y\}: y=0, 1, \ldots$$

and define a *bivariate distribution function* for X and Y by

$$H(x, y) = P\ \{X \leq x, Y \leq y\}: x=0, 1, \ldots; y=0, 1, \ldots$$

We require a specification of $H(\cdot)$ which is compatible with prescribed forms for $F(\cdot)$ and $G(\cdot)$, i.e. we require $H(\cdot)$ to satisfy

$$H(x, \infty) = F(x); \ H(\infty, y) = G(y).$$

A solution to this problem has been provided by Plackett (1965). Plackett's class of bivariate distributions is defined by

$$H(x,y) = \begin{cases} \left[S(x, y) - \{S^2(x, y) - 4\psi(\psi-1)\ F(x)\ G(y)\}^{\frac{1}{2}}\right]/\{2(\psi-1)\}: 0<\psi\neq1 \\ F(x)\ G(y): \psi=1 \end{cases} \quad (3.3.1)$$

where

$$S(x,y) = 1 + \{F(x) + G(y)\}\ (\psi-1).$$

The joint probability distribution $r_{xy} = P\ \{X=x, Y=y\}$ can be recovered from (3.3.2) as

$$r_{xy} = H(x, y) - H(x-1, y) - H(x, y-1) + H(x-1, y-1).$$

In evaluating r_{xy} we must remember that for a non-negative random variable X, $F(x)=0$ whenever $x<0$, and similarly for Y and $G(y)$. Despite its complicated form Plackett's family has two attractive features:

(i) it holds for *any* prescribed $F(\cdot)$ and $G(\cdot)$

(ii) it can embrace perfect positive association ($\psi \to \infty$), independence ($\psi=1$) and pefect negative association ($\psi=0$).

A simpler-looking class of distributions is obtained by writing $\alpha=\psi-1$, expanding (3.3.1) as a power series in α and discarding terms in α^2 and higher powers. This gives

$$H(x, y)=F(x)\, G(y)\, [1+\alpha\, \{1-F(x)\}\, \{1-G(y)\}] \tag{3.3.2}$$

The price paid for the simplicity of this formula is a loss of generality. Equation (3.3.2) defines a valid bivariate distribution function $H(\cdot)$ only for $|\alpha|<1$, which in turn restricts the strength of the association between X and Y. For further discussion, see Mardia (1970).

3.4 Likelihood analysis

We have now assembled a fairly wide range of distributions for the description of bivariate quadrat counts. For naturally regenerated populations, the negative binomial distribution (3.2.2) provides a two-parameter class of marginal distributions which includes as a limiting case the one-parameter Poisson. For artificially regenerated populations, the beta-binomial distribution (3.2.4) fills a similar role, with the one-parameter binomial as a special case. In either event, the parameters of the distribution have a natural descriptive interpretation in terms of average values (λ for the negative binomial, π for the beta-binomial) and variability or heterogeneity (k^{-1} for the negative binomial, Θ for the beta-binomial). For any combination of marginal distributions, Plackett's distribution family (3.3.1) contributes a fifth parameter to describe the dependence between the two types of individuals, in a manner which again admits of a fairly natural descriptive interpretation. We now consider the problem of model selection and parameter estimation within this framework.

Quite generally, consider a set of observations $\{x_i: i=1, \ldots, m\}$ arising as an independent random sample from a probability distribution $\{p_x: x=0, 1, \ldots\}$, which incorporates a vector of parameters $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_p)$. The log-likelihood function for $\underset{\sim}{\Theta}$ is defined as

$$L(\underset{\sim}{\Theta})=\sum_{i=1}^{n}\log(p_{x_i})$$

and the value of $\underset{\sim}{\Theta}$ which maximises $L(\cdot)$ is the maximum likelihood estimate of $\underset{\sim}{\Theta}$, written as $\hat{\underset{\sim}{\Theta}}$.

A useful general result about maximum likelihood estimation is the following. Write $\mathrm{Var}(\hat{\underset{\sim}{\Theta}})$ for the $p \times p$ matrix with $(i, j)^{th}$ element

$$v_{ij}=\mathrm{Cov}(\hat{\Theta}_i, \hat{\Theta}_j)$$

Then, asymptotically and under mild regularity conditions,

$$\mathrm{Var}(\hat{\underset{\sim}{\Theta}})=\left\{ E\left[-\frac{\partial^2\, L(\underset{\sim}{\Theta})}{\partial \underset{\sim}{\Theta}^2} \right] \right\}^{-1} \tag{3.4.1}$$

where the term in curly brackets is to be interpreted as a $p \times p$ matrix with $(i, j)^{th}$ element.

$$E\left[-\frac{\partial^2 L(\underline{\Theta})}{\partial \Theta_i \, \partial \Theta_j}\right]$$

and the expectation is with respect to the assumed distribution of the x_i: $i=1, \ldots, m$. In practice, if the expectation raises technical difficulties, approximate large sample standard errors for $\hat{\Theta}_i$ are obtained by omitting the expectation in (3.4.1) and replacing the unknown $\underline{\Theta}$ by $\hat{\underline{\Theta}}$.

A second useful result concerns tests of hypotheses. The hypothesis that the assumed distribution $\{p_x: x=0, 1, \ldots\}$ is appropriate can be written symbolically as

$H_1: \underline{\Theta} = (\Theta_1, \ldots, \Theta_p)$,

for unspecified values of Θ_i. Suppose that, within H_1, we wish to test the hypothesis that some of the p parameters are redundant, namely,

$H_0: \underline{\Theta} = (\Theta_1, \ldots, \Theta_q, 0, \ldots 0)$

for some $q < p$. If we then write L_i for the maximised log-likelihood under the hypothesis H_i, we have that *if H_0 is true* the quantity $2(L_1 - L_0)$ is distributed approximately as chi-squared on $p-q$ degreees of freedom; significantly large values cast doubt on the validity of H_0 within the wider hypothesis H_1. These ideas extend immediately to the case of bivariate data $(x_i, y_i): i=1, \ldots, m$ and an assumed distribution $\{r_{xy}: x, y=0, 1, \ldots\}$. For a detailed account at a fairly advanced level, see Silvey (1975).

The application of likelihood analysis to the problem of fitting a bivariate quadrat count distribution is illustrated by Figure 6. Note that for each marginal distribution we apply a test of H_0 against an alternative H_1 and adopt the appropriate one-parameter or two-parameter marginal distribution according as the result of the test is, respectively, non-significant or significant. These tests determine the prescription for $F(\cdot)$ and $G(\cdot)$ in (3.3.1), within which we proceed to test the hypothesis, $H_0: \psi=0$, of independence. The reference to "marginal parameters" in Figure 6 refers to the fact that the maximum likelihood estimates of, for example, μ and p in a bivariate Poisson-binomial distribution with data $\{(x_i, y_i): i=1, \ldots, m\}$ are not in general the same as the maximum likelihood estimates from two univariate analyses with respective data $\{x_i: i=1, \ldots, m\}$ and $\{y_i: i=1, \ldots, m\}$. It should be noted that in the application described in Section 3.5, the actual differences were rather small. For reference, we give here the estimation formulae for the various types of univariate distribution under consideration.

3.4.1 *Poisson*

The maximum likelihood estimator is

$$\hat{\mu} = \sum_{i=1}^{m} x_i/m$$

with estimated variance
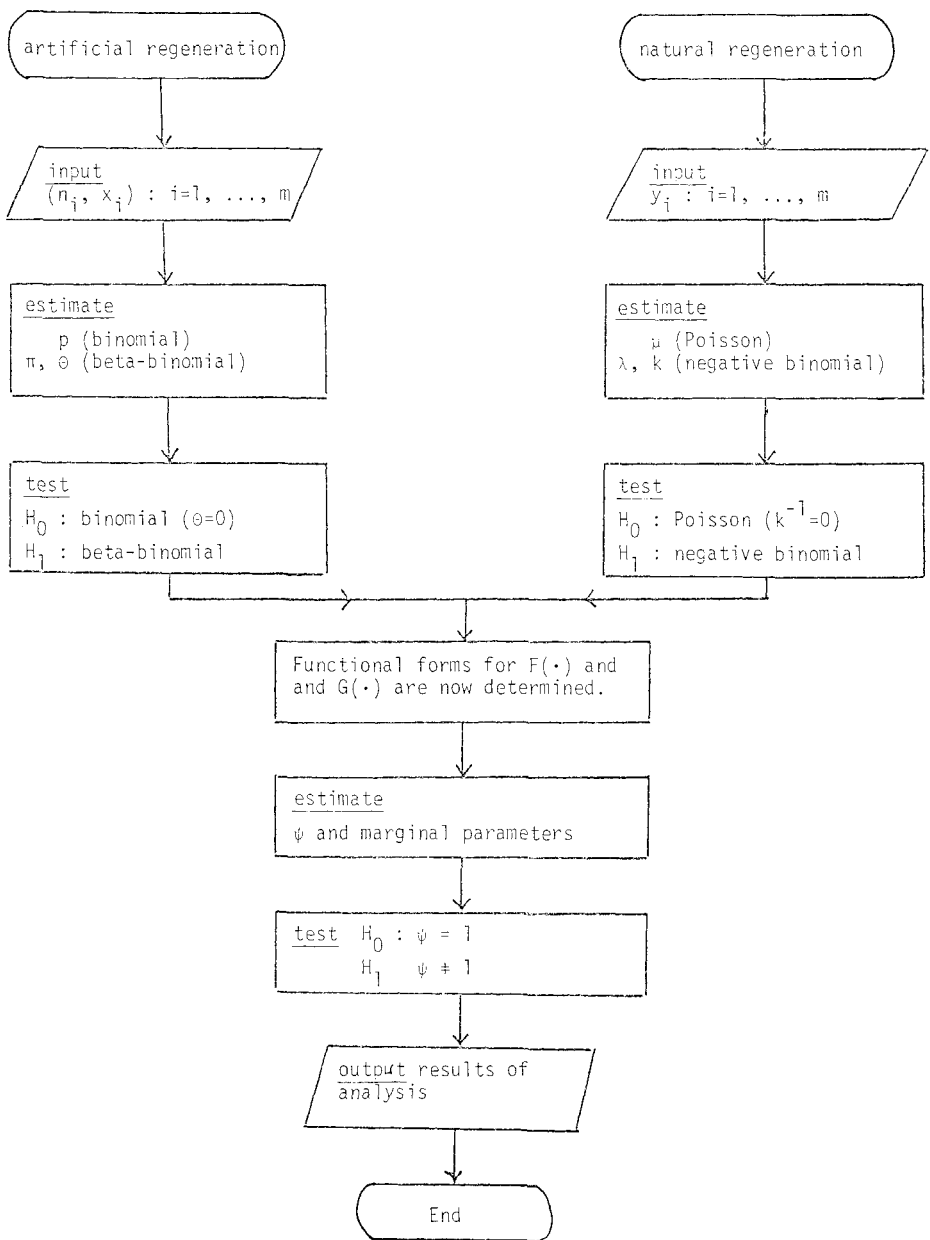
$$\text{Var}(\hat{\mu}) \simeq \hat{\mu}/m$$

artificial regeneration

natural regeneration

input
$(\overline{n_i}, x_i) : i=1, \ldots, m$

input
$\overline{y_i} : i=1, \ldots, m$

estimate
  p (binomial)
π, Θ (beta-binomial)

estimate
  μ (Poisson)
λ, k (negative binomial)

test
H_0 : binomial ($\Theta=0$)
H_1 : beta-binomial

test
H_0 : Poisson ($k^{-1}=0$)
H_1 : negative binomial

Functional forms for $F(\cdot)$ and
and $G(\cdot)$ are now determined.

estimate
ψ and marginal parameters

test H_0 : $\psi = 1$
     H_1 $\psi \neq 1$

output results of
analysis

End

Figure 6. Flow-chart for likelihood analysis.

3.4.2 *Negative binomial*

The log-likelihood is maximised numerically, using the simplex algorithm of Nelder & Mead (1965). Maximum likelihood estimates $\hat{\lambda}$ and $\hat{k}$ are uncorrelated, with estimated variances

$$\text{Var}(\hat{\lambda}) \simeq \hat{\lambda}\,(\hat{\lambda}+\hat{k})\,/\,(m\hat{k})$$

and

$$\text{Var}(\hat{k}) \simeq \left\{ m\,\text{E}\left[\sum_{j=1}^{x}(\hat{k}+x-j)^2 \right] - m\,\hat{\lambda}\,/\,\hat{k}(\hat{\lambda}+\hat{k}) \right\}^{-1}$$

where the expectation is with respect to x whose distribution is (3.2.2) with parameters $\hat{\lambda}$ and $\hat{k}$.

3.4.3 *Binomial*

The maximum likelihood estimator is

$$\hat{p} = \sum_{i=1}^{m} x_i \,/\, \sum_{i=1}^{m} n_i$$

with estimated variance

$$\text{Var}(\hat{p}) \simeq \hat{p}\,(1-\hat{p})\,/\,\sum_{i=1}^{m} n_i$$

3.4.4 *Beta-binomial*

Numerical maximisation of the log-likelihood is again effected by the Nelder-Mead algorithm. To evaluate the variance-covariance matrix of (π, Θ) write

$$d_{11}\,(n,\,x) = \sum_{j=0}^{x-1}(\hat{\pi}+j\hat{\Theta})^{-2} + \sum_{j=0}^{n-x-1}(\hat{\varrho}+j\hat{\Theta})^{-2}$$

$$d_{12}\,(n,\,x) = \sum_{j=0}^{x-1}j(\hat{\pi}+j\hat{\Theta})^{-2} - \sum_{j=0}^{n-x-1}j(\hat{\varrho}+j\hat{\Theta})^{-2}$$

and

$$d_{22}\,(n,\,x) = \sum_{j=0}^{x-1}j^2(\hat{\pi}+j\hat{\Theta})^{-2} + \sum_{j=0}^{n-x-1}j^2(\hat{\varrho}+j\hat{\Theta})^{-2} + \sum_{j=1}^{n-1}j^2(1+j\hat{\Theta})^{-2}$$

where we recall that $\varrho=1-\pi$. Also define

$$e_{rs} = \sum_{i=1}^{m}\text{E}\,[d_{rs}(n_i,\,X)]$$

where the expectation is with respect to X whose distribution is (3.2.4) with parameters

$$a = \hat{\pi}\,\hat{\Theta}^{-1};\; b = (1-\hat{\pi})\,\hat{\Theta}^{-1}$$

Then,

$$\text{Var} \, (\hat{\pi}, \hat{\Theta}) \simeq (e_{22} \, e_{11} - e_{12})^{-1} \begin{bmatrix} e_{22} & -e_{12} \\ -e_{12} & e_{11} \end{bmatrix}$$

3.5 Application to PTAX data

We consider the analysis of counts in circular sample plots within a stand in the PTAX investigations. We consider three such variables,

X=Number of planted survivors
Y=Number of wild softwood plants
Z=Number of wild hardwood plants

This generates three possible bivariate analyses, all of which utilize Plackett's class of bivariate distributions with appropriately specified marginal distributions: binomial or beta-binomial for X, Poisson or negative binomial for Y and for Z.

Evidently, a complete trivariate analysis would be preferable, but we know of no suitable family of trivariate distributions. Another point which should be emphasised at the outset is that these analyses ignore the spatial arrangement of the sample plots. A new technical difficulty with the PTAX-data is that in the case of the wild hardwood plants the count Z is recorded exactly only for values less than 10. We therefore group values of 10 or more, and modify the assumed marginal distribution accordingly: thus, p_0, p_1, ..., p_9 are specified by the Poisson or negative binomial formulae (3.2.1) and (3.2.2) respectively, but the upper tail probabilities are replaced by a single probability,

$$p^* = P \, \{Z \geq 10\} = 1 - (p_0 + p_1 + \ldots + p_9)$$

3.6 Results

Table 2 shows the results of this analysis for a single stand, number 1202 which has 46 plots recorded as being suitable for planting. Note that
(i)   for planted survivors the binomial distribution is accepted against the beta-binomial
(ii) for both wild softwood and wild hardwood the Poisson distribution is rejected overwhelmingly in favour of the negative binomial.
(iii) independence is accepted between planted survivors and wild softwood, but rejected in favour of dependence in the other two cases. The estimated values of ψ are greater than 1, indicating positive dependence.

Result (iii) is somewhat atypical in that for most stands the analysis suggested dependence only between wild softwood and wild hardwood.

This type of analysis has been applied to 185 stands, with somewhat mixed results. The proportion of planted survivors in a stand varies between 0.11 and 0.98, with a mean and standard deviation of 0.69 and 0.19 respectively. The binomial is rejected at the 5 % level of significance on 65 out of 185 occasions. The average number of wild softwood plants per plot varies from 0.02 to 5.18 with a mean and standard deviation of 1.41 and 1.09, and the Poisson is rejected in favour of the negative binomial on 140 out of 185 occasions. For wild hardwoods, the corresponding figures are 0.21 to 12.50, 4.50 and 2.83, 182 out of 185.

Table 2. Bivariate quadrat count analysis of stand 1202.

a) *Marginal distributions:* L_0 and L_1 indicate maximised log-likelihoods for appropriate one parameter and two parameter distributions respectively.

Variable	L_0	L_1	$2(L_1-L_0)$	p-value	Conclusion
X	−29.08	−28.89	0.38	0.5221	binomial
Y	−488.63	−40.12	897.02	0.0000	negative binomial
Z	−132.98	−74.72	116.52	0.0000	negative binomial

b) *Bivariate distributions:* L_0 and L_1 indicate maximised log-likelihoods for independence ($\psi=1$) and general bivariate distribution, using marginal distributions implied by table 3a.

Variables	L_0	L_1	$2(L_1-L_0)$	p-value	Conclusion	$\hat{\psi}$
X, Y	−69.20	−69.17	0.06	0.3938	independent	1.21
X, Z	−103.80	−100.67	6.26	0.0120	dependent	6.97
Y, Z	−114.84	−112.34	5.00	0.0249	dependent	3.60

Parameter estimates for marginal distributions

X: binomial $\hat{p}=0.73$
Y: negative binomial $\hat{\lambda}=1.03$, $\hat{k}=0.10$
Z: negative binomial $\hat{\lambda}=2.64$, $\hat{k}=0.21$

The distribution of p-values in testing for dependence between planted survivors and wild softwood appears fairly uniform; for example 23 out of 185 p-values are less than 0.10, which is close to expectation in the absence of any genuine dependence. A similar picture emerges with regard to planted survivors and wild hardwood, with 23 p-values less than 0.10. However, for wild softwood and wild hardwood 59 p-values are less than 0.10 and 135 stands give $\hat{\psi}>1$ so that, overall, there appears to be some genuine positive dependence in this case.

Attempts have been made to relate the estimated values of the various parameters to a number of explanatory variables at stand level. We find significant, but small, positive correlations between the proportion of planted survivors in a stand and the average number of wild hardwood plants per plot, and between the average numbers of wild softwood and wild hardwood plants per plot. In conclusion, the PTAX data show enormous variation between the various stands, with no clear indications in the recorded data of why this should be so. We claim only that the analysis indicates in broad terms the type of marginal distributions and dependence structures which might reasonably be used in a simulation model.

4 A relationship between two measures of variation in yield within a stand

4.1 Introduction

One objective of a silvicultural programme is to achieve a uniform growth, in the sense that the "yield per unit area" in different parts of the stand should be roughly constant. In young stands, the variation in yield per unit area over the stand is measured by first counting the numbers of stems in sample plots of a standard size, distributed randomly through a stand, and then calculating the sample variance-to-mean ratio of these stem-counts. In older stands, a similar sampling procedure is used, but total basal area is now a more useful measurement than stem-count and the variation in yield per unit area is therefore expressed by the sample coefficient of variation of the total basal area per plot.

In order to establish a theoretical relationship between these two quantities we suppose that the number of stems in a sample plot of a specified area is a random variable N with mean μ_n and variance σ_n^2; we write $v_n = \sigma_n^2/\mu_n$. We suppose also that basal area for a single stem is a random variable X with mean μ_x and variance σ_x^2; we write $c_x = \sigma_x/\mu_x$. Then, the total basal area per plot is a random variable Y which is related to N and X by

$$Y = \sum_{i=1}^{N} X_i \qquad (4.1.1)$$

The X_i are assumed to be *independent* random variables, all with the same distribution as X. We use μ_y, σ_y^2 and $c_y = \sigma_y/\mu_y$ to denote the mean, variance and coefficient of variation of Y.

We shall use the notation of conditional expectation, and require the following results concerning two random variables, X and Y:

$$E[Y] = E_x[E_y(Y \mid X)] \qquad (4.1.2)$$

$$Var(Y) = Var_x[E_y(Y \mid X)] + E_x[Var_y(Y \mid X)] \qquad (4.1.3)$$

Note that the subscript on the expectation symbol indicates the source of the random variation (X or Y) with respect to which the expectation is taken. Further discussion can be found in most intermediate texts in mathematical statistics, for example Hogg & Craig (1970).

4.2 Relationships for plots of a single size

From (4.1.1) we deduce that $E(Y \mid N) = N E(X_i) = N\mu_x$, and $Var(Y \mid N) = N Var(X_i) = N\sigma_x^2$. Now, using (4.1.2) and (4.1.3) we further deduce that

$$\mu_y = \mu_n \mu_x \qquad \text{and} \qquad \sigma_y^2 = \sigma_n^2 \mu_x^2 + \sigma_x^2 \mu_n,$$

whence $\qquad c_y^2 = \mu_n^{-1}(v_n + c_x^2) \qquad (4.2.1)$

These formulae are given in the appendix to Matérn (1972).

Note that (4.2.1) assumes only that individual basal areas are independent. A comparison between (4.2.1) and the corresponding empirical relationship, calculated by sampling the typestand data, will be presented in section (4.4) in order to provide an indirect check on the validity of this assumption.

4.3 Relationships for plots of different sizes

To correspond to current silvicultural practice we now extend the discussion by assuming that the total basal area measurements in older stands refer to larger plots than those used to obtain stem-counts in younger stands. We therefore introduce a random variable N_0, the stem-count in a plot of unit area, with mean, variance and variance-to-mean ratio μ_0, σ_0^2 and $v_0 = \sigma_0^2/\mu_0$. If now N refers to stem-count in a plot of area A>1, and counts in non-overlapping sub-plots are independent, it follows that

$$\mu_n = A\mu_0; \quad \sigma_n^2 = A\sigma_0^2; \quad v_n = v_0,$$

and these relationships may be substituted into (4.2.1) to give

$$c_y^2 = (A\mu_0)^{-1} \ (v_0 + c_x^2) \tag{4.3.1}$$

As before, this theoretical result will be checked empirically by reference to the typestand data. However, we remark that the assumption of independent counts in non-overlapping sub-plots implies quite severe restrictions on the nature of the underlying spatial point process of stem locations. In particular, in Section 3 we have already discussed the idea that the local environment at a point within a stand may exhibit random spatial variation over the stand, and this provided a heuristic justification for using the negative binomial and beta-binomial distributions to describe stem-count data. If this spatial variation is "smooth", in the sense that adjacent plots enjoy similar environmental conditions and hence similar mean number of stems per unit area, the assumption of independent stem-counts in non-overlapping sub-plots is violated. By the same token, if such spatial variation affects vigour of growth, stem basal areas within a plot will be positively correlated and this will inflate the variance of the total basal area per plot. Conversely, competitive interactions between stems within a plot may reduce the variance of the total basal area per plot.

4.4 Application to typestand data

Each of ten cleaning stands were sampled randomly by 64 plots of radius 0.5 m, by 16 plots of radius 1.0 m and by 4 plots of radius 2.0 m, and the sample analogues of the various quantities in the relationship (4.2.1) were calculated, after applying each of the seven possible cleaning regimes as described in Section 1.2. This gives a total of 210 pairs of observed and predicted values of c_y which can be plotted in various ways.

First, Figure 7 shows a scatterplot of all 210 pairs, which shows no systematic departure from equality of observed and predicted values. We might expect that any dependence between stem basal areas would be
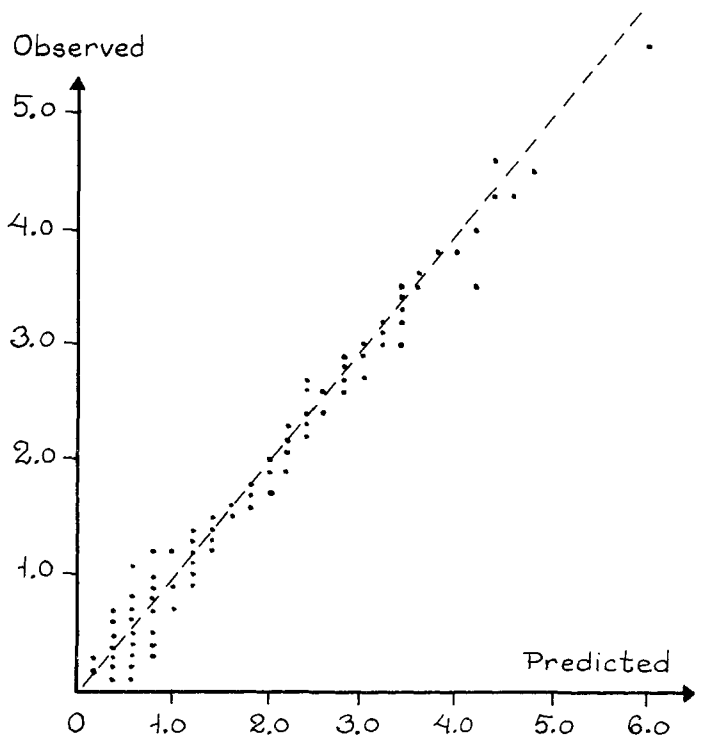
Figure 7. Observed and predicated values of coefficient of variation of total basal area per plots (10 stands, 7 cleaning regimes, 3 plot radii).

accentuated as the cleaning regime proceeds from 0 to 6, the higher numbered regimes being progressively more selective. It is therefore reassuring that when the 30 pairs for each cleaning regime are plotted separately, the approximate equality of observed and predicted values is maintained. Figure 8 shows the scatterplots corresponding to cleaning regimes 0, 3 and 6.

Separate scatterplots for the three different plot radii are shown in Figure 9. These show approximate equality between observed and predicted values, although in all three cases there is a slight preponderance of observations below the bisector. As noted earlier, two contrary phenomena operate to modify the predicted relationship (4.2.1) – smooth environmental variation and competetive interactions induce positive and negative correlation, respectively, between basal areas of neighbouring stems. The implication of our results is that in the typestand data, the net effect of these two phenomena is small.

Finally, we note that (4.2.1) and (4.3.1) together imply that c_y should be inversely proportional to plot radius. The observed values of c_y for plots of radius 0.5, 1.0 and 2.0 meters, averaged over ten stands and seven cleaning regimes, are 2.79 (0.10), 1.36 (0.07) and 0.48 (0.03), respectively. Figures in parentheses are standard errors. The first two of these average values are in approximately the predicted ratio of two to one, but the average value for the largest plot size is significantly smaller than predicted. Table 3 presents observed values of c_y separately for each of the seven cleaning
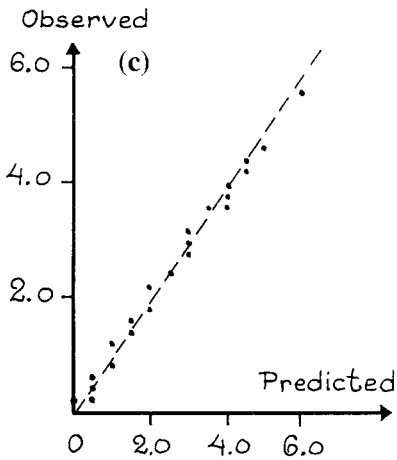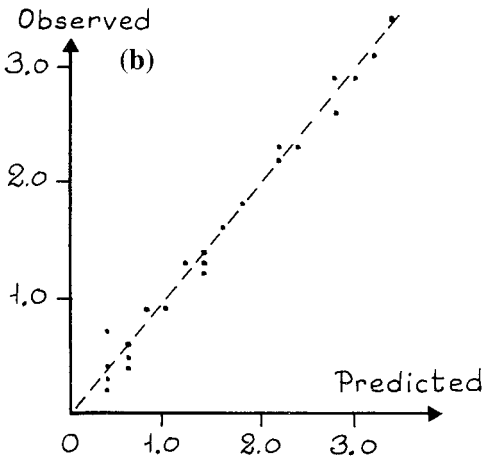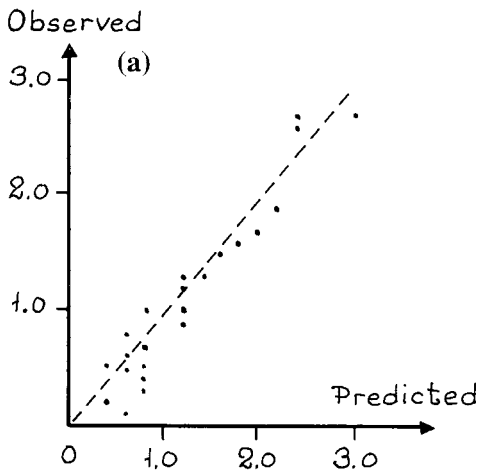
Figure 8. Observed and predicted values of coefficient of variation of total basal area per plot (10 stands, 3 plot radii).
(a) cleaning regime 0 (no cleaning).
(b) cleaning regime 3 (cleaning to 2000 stems/hectare).
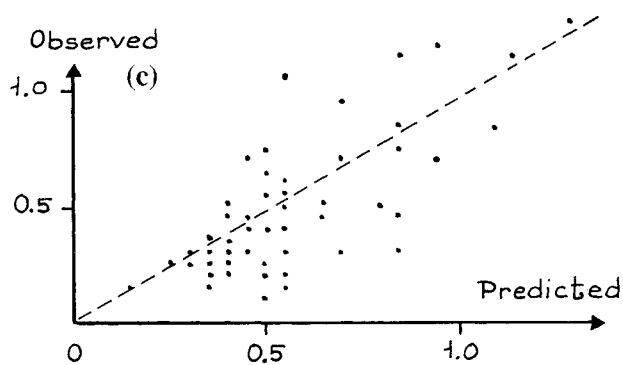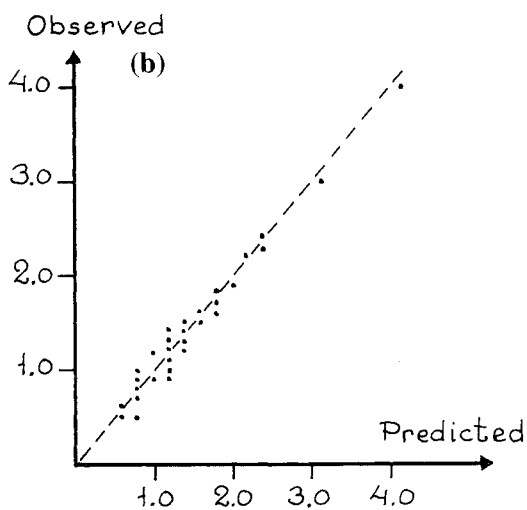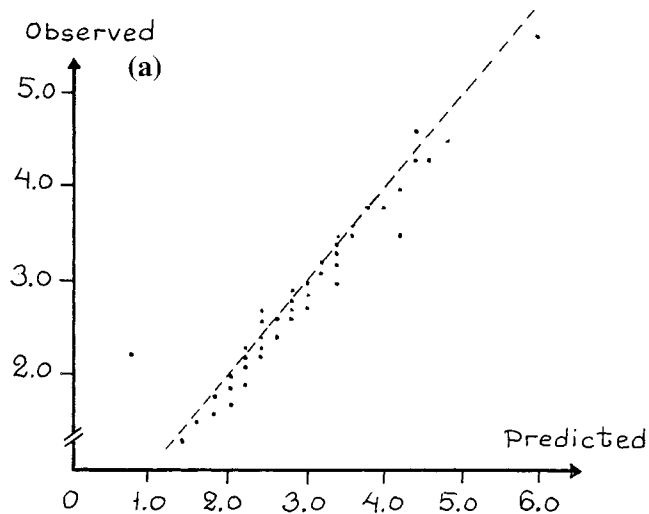(c) cleaning regime 6 (cleaning to 800 stems/hectare).

Figure 9.
Observed and predicted values
of coefficient of variation of total
basal area per plot (10 stands,
7 cleaning regimes).
(a) plot radius 0.5 m.
(b) plot radius 1.0 m.
(c) plot radius 2.0 m.

Table 3. Observed coefficients of variation of total basal area per plot, averaged over ten typestands. (Standard errors in parentheses.)

Plot radius	Cleaning regime						
	0	1	2	3	4	5	6
0.5 m	1.86 (0.18)	2.27 (0.14)	2.43 (0.13)	2.71 (0.12)	3.00 (0.14)	3.34 (0.18)	3.94 (0.25)
1.0 m	0.92 (0.12)	1.10 (0.12)	1.13 (0.13)	1.22 (0.11)	1.43 (0.11)	1.56 (0.14)	2.17 (0.26)
2.0 m	0.42 (0.06)	0.40 (0.06)	0.39 (0.07)	0.43 (0.07)	0.51 (0.12)	0.58 (0.13)	0.64 (0.11)

regimes, and shows that the average values of c_y for the largest plots drift progressively below the values implicit in the relationship (4.3.1) as the cleaning regime changes from 0 to 6. This is not unreasonable, since one aim of the cleaning operation is to produce an even spatial distribution of good quality stems. Two plausible consequences of this are that the variance of the number of stems per plot does not increase in proportion to its area and the basal areas of near-neighbouring stems exhibit negative spatial autocorrelation. Both these factors would contribute to a reduced coefficient of variation of total basal area per plot.

We should perhaps emphasise that the area covered by each typestand is rather small. In practice, problems caused by gross environmental variation within a stand may be more acute, with consequent distortion of the predicted relationship (4.2.1).

5 Statistical methods for sparsely sampled point patterns

5.1 Introduction

The locations of trees in a stand constitute a *spatial point pattern*, by which we mean a set of point locations, henceforth *events*, in some essentially planar region. When the events are of two or more distinguishable types, for example trees of different species, we shall call the resulting superposition of the several component patterns a *multivariate spatial point pattern*.

A parameter of obvious relevance in the assessment of stand quality, at least in the early stages of stand development, is the *intensity*, or mean number of stems per unit area. Equally relevant is the spatial distribution of stems through the stand, the ideal being an even, or "regular" spatial distribution (see Section 5.2 below). In a mixed stand, it is also important to understand the nature of any associations which may exist between different species. In accord with historical development we first discuss at some length the concept of pattern and the use of quadrat counts or distance measurements in testing the hypothesis of a completely *random* spatial point pattern. We then describe some methods of intensity estimation which are intended to be robust against departures from a completely random pattern, and conclude with some comments on the problem of testing associations in a multivariate pattern.
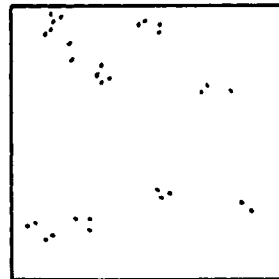
Figure 10 shows three fictitious stands which contain the same number of stems but display markedly different *patterns*. The first is a *regular* pattern, the second *aggregated* and the third completely *random*, in a sense which will be made precise in Section 5.2. We see that the description "regular" implies a more or less even distribution of stems over the stand, whilst "aggregated" implies the opposite effect.

One approach to the analysis of pattern is to envisage a continuous range of variation, from regularity through randomness to aggregation. This one-
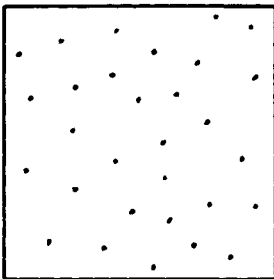
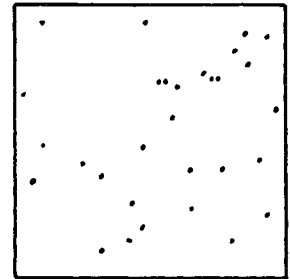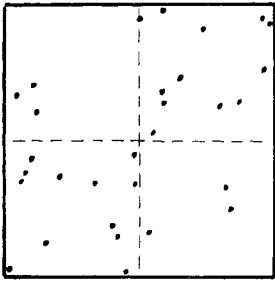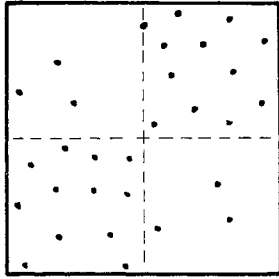(a) a regular pattern   (b) an aggregated pattern   (c) a random pattern

Figure 10. Three fictitious stands, each containing 30 stems in the unit square.

(a) heterogeneous environment (high intensity in top-right and bottom-left-corners).

(b) heterogeneous environment as above, plus local competitive interactions.

Figure 11. Two further fictitious stands, each containing 30 stems in the unit square.

dimensional interpretation is an oversimplification. For example, aggregated patterns can arise either through clustering of individual stems or as a consequence of environmental heterogeneity; elements of aggregation and regularity can be combined when local competitive interactions are superimposed on a heterogeneous environment. Figure 11 gives some more fictitious stands which display these features. We conclude that a single parameter cannot in general provide an adequate description of pattern, and modern statistical techniques recognise this. Thus, for example, Mead (1974) discusses the detection of different "scales of pattern", whilst Ripley (1977) and Diggle (1978, 1979) are concerned with fitting explicit stochastic models.

In field-work, a complete map of the stand will not be available and these modern techniques are inappropriate. Instead, data are extracted from the underlying pattern by one of two fundamentally different types of sampling technique, usually referred to as *quadrat sampling* and *distance sampling*. Quadrat sampling consists of recording how many events are in each of a number of randomly located plots in the stand, whilst distance sampling consists of recording the distances from each of a number of randomly located points to neighbouring events, defined in various ways: a simple example would be to record the distance from each point to the nearest event. In this context an "index of pattern" will provide a useful, albeit limited, description of pattern which might contribute towards the evaluation of an individual stand.

Sections 2, 3 and 4 of this report were concerned with particular aspects of the analysis of quadrat count data obtained from stands covering a relatively small area. The emphasis in the present section is much more on the analysis of data obtained from a preliminary survey of a much larger area.

In Section 5.2 we give a precise definition of a completely random pattern as a realisation of a homogeneous, planar Poisson process, and derive som useful distributional results for this process. In the remainder of Section 5 we use the term *complete spatial randomness,* henceforth CSR, to mean that the pattern in question forms a partial realisation of a homogeneous, planar Poisson process. Note that CSR was used in a different, but analogous, sense in Section 2.

Our subsequent discussion of tests of CSR therefore carries the implicit assumption that any statistic which is used to test CSR can also be used as an index of pattern. We anticipate that a good index will be one which gives a powerful test against both regular and aggregated alternatives to the Poisson process; whilst a powerful test will not necessarily give good discrimination between different *degrees* of regularity or aggregation it is certainly the case that a weak test cannot do so. Unless explicitly stated otherwise, we shall always arrange that significantly small values of a test statistic indicate regularity, and significantly large values aggregation, in the underlying pattern.

5.2 The Poisson process

As noted above, the Poisson process is used as an idealised standard of complete spatial randomness, relative to which pattern may be assessed. Informally, the Poisson process consists of events distributed independently and uniformly over the planar region occupied by the stand. Thus, there is no spatial variation in *intensity,* defined as mean number of events per unit area, nor are there any interactions between events. For the pattern shown as Figure 10c, the events were generated as an independent random sample from the uniform distribution over the unit square.

More formally, the Poisson process is defined by a single parameter λ, the intensity of the process, such that if $p_n(A)$ denotes the probability distribution of the number $N(A)$ of events in an area A, then for small A,

(i) $p_n(A) = \begin{cases} 1 - \lambda A + o(A) & : \quad n = 0 \\ \lambda A + o(A) & : \quad n = 1 \\ o(A) & : \quad n > 1, \end{cases}$

where $o(\cdot)$ means "of smaller order of magnitude than", and
(ii) for any two disjoint areas A and B, $N(A)$ and $N(B)$ are statistically independent.

Then, $N(A)$ follows a Poisson distribution,

$$p_n(A) = e^{-\lambda A}(\lambda A)^n/n! : n = 0, 1, 2, \ldots, \tag{5.2.1}$$

the *quadrat count distribution* for the Poisson process, and (ii) ensures that counts in disjoint areas are independent. The derivation of (5.2.1) is given for a Poisson process in one, temporal dimension by Cox & Lewis (1966, Ch. 2). A heuristic justification of the result is obtained by considering a large number n_o of events in a large area A_o. Each event lies within A with small probability A/A_o, independently of all other events. The number of events in A therefore follows a binomial distribution, to which the Poisson distribution provides a good approximation for large n_o and small A/A_o. The result (5.2.1) follows, with $\lambda = n_o/A_o$.

Various *distance distributions* for the Poisson process can be derived by simple geometrical arguments, using (5.2.1) and property (ii). For example, let X denote the distance from a randomly located point to the nearest event, then

$$P\{X>x\}=P\{N(\pi x^2)=0\}=\exp(-\lambda\pi x^2),$$

so the PDF of X is

$$f(x)=2\lambda\pi x\exp(-\lambda\pi x^2): x\geq 0, \tag{5.2.2}$$

a result which has reappeard many times in the ecological literature since its original derivation by Hertz in 1909 (Holgate, 1972). The distribution (5.2.2) also applies to the distance from a randomly selected *event* to the nearest other event. Suppose, however, that E is the nearest event to a randomly located point P, and F the nearest other event to E. Clearly, the distance X≡PE follows the distribution (5.2.2) but Y≡EF does not. Figure
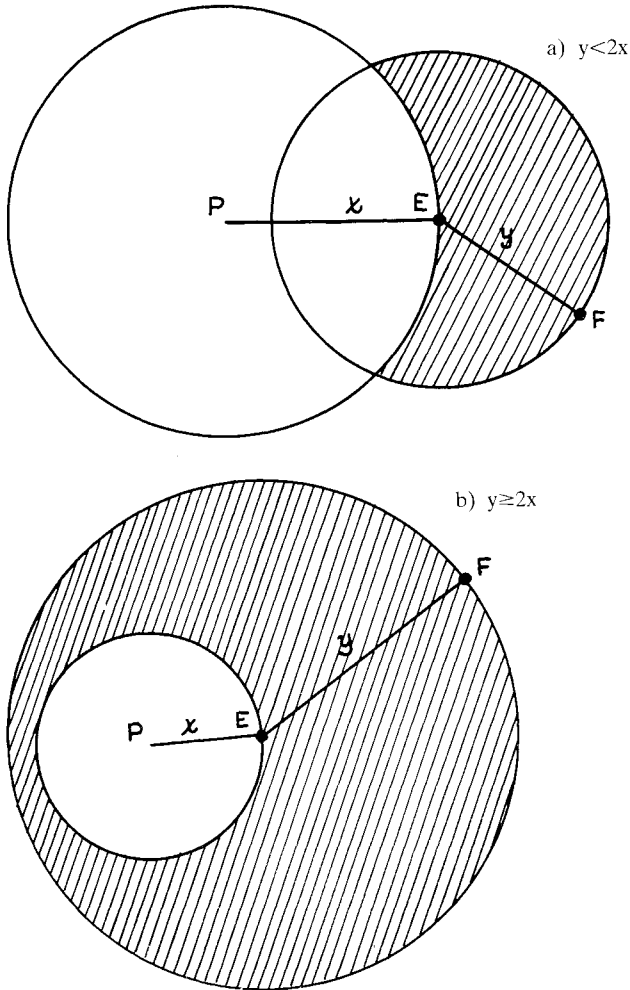

Figure 12. The nearest event to a randomly located point.
P is a randomly located point, E the nearest event to P and F the nearest event to E.
Distance PE=x, distance EF=y. Shaded area is region searched for event F.

12 shows that the conditional distribution of Y, given X=x, is determined by

$$P \{Y>y \mid X=x\}=\exp \{-\lambda A(x,y)\},$$

where A(x,y) is the shaded area in Figure 12,

$$A(x,y)=\begin{cases} y^2\pi-(y^2\Phi+x^2\Theta-xy \sin\Phi) & : & y<2x \\ (y^2-x^2)\pi & : & y\geq 2x, \end{cases}$$

cos $\Phi=y/2x$ and $\Theta=\pi-2\Phi$. The conditional PDF of Y, given X=x is

$$g(y|x)=\frac{\partial A(x,y)}{\partial y} \lambda\exp\{-\lambda(A(x,y))\} \tag{5.2.3}$$

and the joint PDF of X and Y is

$$h(x,y)=f(x)g(y|x). \tag{5.2.4}$$

This distribution, and its use for assessing pattern, is discussed by Cox & Lewis (1976). Similar arguments can be used to derive various other distance distributions, some of which are discussed in Section 5.4.

5.3 The quadrat sampling method

The quadrat sampling method consists of locating m quadrats of a specified size and shape at random within a stand, and observing the numbers n_i: i=1, ..., m of events within each quadrat. Under CSR the n_i are an independent random sample from the Poisson distribution (5.2.1) which is characterised by the equality of its mean and variance, irrespective of the value of λ. Thus, a natural test statistic is the sample variance-to-mean ratio, or *index of dispersion*,

$$I=(m-1)^{-1}\sum_{i=1}^{m} (n_i-\bar{n})^2/\bar{n}$$

Under CSR, $(m-1)I$ follows a chi-squared distribution with $m-1$ degrees of freedom and the expectation of I is 1. The index of dispersion was first used in this context by Fisher, Thornton & Mackenzie (1922); subsequent ecological applications include those reported by Blackman (1935) and by Ghent (1963). See also Sections 3 and 4 of this report.

With regard to implementation, the power of a test of CSR against any specified alternative obviously increases with m, but also depends in an unpredictable way on the size and shape of the individual quadrats. Some recent results in this area are given by Perry & Mead (1979). Economic considerations will presumably imply some upper limit on m and on the total quadrat area. There will also be some practical constraint on the individual quadrat size. In forestry work, circular quadrats are presumably the easiest to use, and have the advantage that any "edge-effects" are thereby minimised. A number of authors, including Ghent (1963) have suggested that a systematic bias may be introduced by a tendency to count events which in fact lie just outside the quadrats, in the mistaken belief that an empty quadrat contains no information. Finally, the total quadrat area should be no more than a small fraction of the total area occupied by the stand. For further discussion of this last point, see Section 5.5.

The index of dispersion appears to have no serious rivals as a test statistic based on quadrat sampling. For example, Cormack (1979) notes that both Lloyd's (1967) "index of crowding", $\{\sum n_i(n_i-1)\}/m\bar{n}$, and Morisita's (1959) "index of aggregation", $\{\sum n_i(n_i-1)\}/\{\bar{n}(m\bar{n}-1)\}$, need to be converted to $(m-1)I$ for purposes of testing CSR.
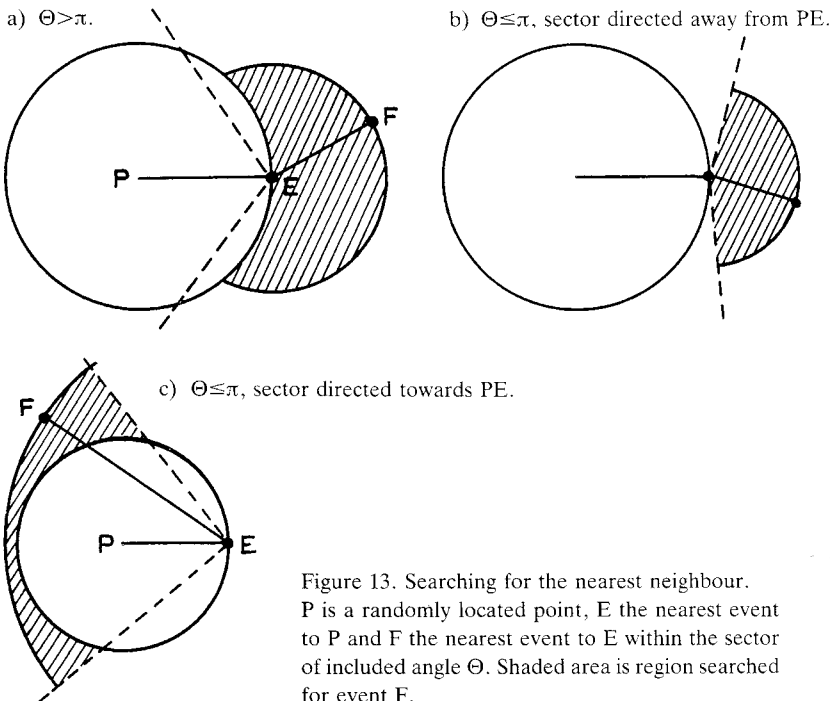
5.4 Distance sampling methods

A number of distance sampling methods were developed in American forestry (e.g. Cottam & Curtis 1949) as a solution to the practical problems raised by quadrat sampling in mature forests. Statistical aspects of these sampling procedures have since been investigated by many authors, and the large number of distance-based test statistics now available contrasts sharply with the situation for quadrat sampling.

Persson (1964) gives a detailed review of early developments in this field. Another useful reference is Holgate (1972) whilst some more recent papers are cited explicity in the remainder of this section.

5.4.1 *Further results for the Poisson process*

A transformation to $U=\pi X^2$ in the result (5.2.2) gives the PDF $f(u)=\frac{1}{2}e^{-u/2}$, which defines a chi-squared distribution on two degrees of freedom. More generally, we have the following:

Let $X_{k,\Theta}$ denote any distance obtained by searching for the kth nearest event in a Poisson process, within a sector of included angle Θ, and let

a) $\Theta>\pi$.

b) $\Theta\leq\pi$, sector directed away from PE.

c) $\Theta\leq\pi$, sector directed towards PE.

Figure 13. Searching for the nearest neighbour. P is a randomly located point, E the nearest event to P and F the nearest event to E within the sector of included angle Θ. Shaded area is region searched for event F.

$U_k = \Theta \lambda X_{k,\Theta}^2$. Then U_k is distributed as chi-squared on $2k$ degrees of freedom.

Note in particular that this applies equally to distances measured from a randomly located point or a randomly selected event. It can also apply to distances measured from the nearest event to a randomly located point, but only if $\Theta \leq \pi$ since otherwise the area searched cannot be a sector. Figure 13 shows why the orientation of the sector is also important in this case.

In constructing distance-based test statistics we shall need to combine distances measured from a number of sample points or events. In doing this, we shall use property (ii) of the Poisson process, whereby distances obtained by searching disjoint areas are statistically independent. We shall further assume, with reservations to be noted in Section 5.5, that distances derived from different sample points or events are also statistically independent.

5.4.2 *Tests based on univariate sampling procedures*

Let $x_1, \ldots, x_m$ be the distances from each of m randomly located points to the nearest event. Under CSR the squared distances $U_i = x_i^2$ are an independent random sample from an exponential distribution, which has coefficient of variation unity. A possible test statistic is therefore the sample squared coefficient of variation,

$$c = s_u^2 / \bar{u}^2, \tag{5.4.2.1}$$

where $\bar{u} = m^{-1} \sum u_i$ and $s_u^2 = (m-1)^{-1} \{ \sum u_i^2 - (\sum u_i)^2 / m \}$.

Diggle (1973) shows that the large sample mean and standard deviation of c under CSR are 1 and $2/\sqrt{m}$ respectively. Eberhardt (1967) proposed a similar statistic, but based on the distances x_i themselves. Eberhardt's statistic is

$$e = m (\sum x_i^2) / (\sum x_i)^2, \tag{5.4.2.2}$$

whose large sample mean under CSR is $4/\pi \approx 1.27$. Tables for testing CSR are given by Hines & Hines (1979).

Both the statistics c and e can be adapted easily to sampling procedures which involve some restriction on the direction of search for nearest neighbours. Two such procedures which deserve mention in view of their possible practical advantages are illustrated in Figure 14. In Cottam, Curtis & Hale's (1953) "point-centred quarter", four observations are taken at each sample point; this is obviously attractive if the effort required to locate sample points in the field is non-trivial. Catana's (1963) "wandering quarter" has a similar advantage in that a single starting point generates a sequence of observations as the field-worker traverses the stand. Note, in either scheme, the fixed orientation of the sectors within which nearest neighbours are sought.

5.4.3 *Tests based on bivariate sampling procedures*

A number of tests exploit the intuitive idea that point-to-event distances will typically be smaller than event-to-event distances in regular patterns,
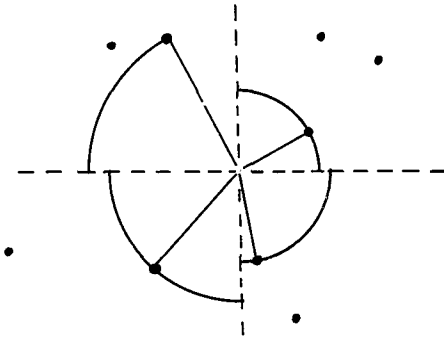
(a) point-centred quarter.
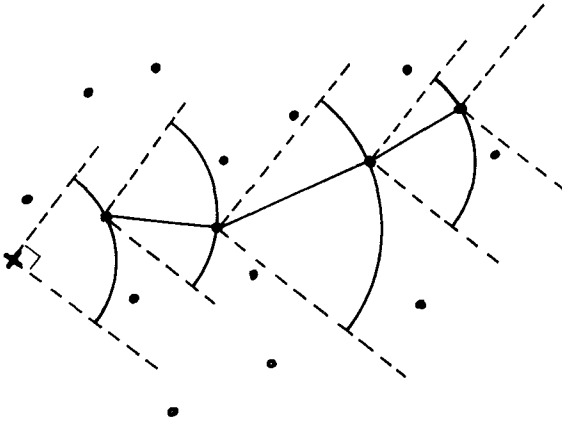
(b) wandering quarter.

Figure 14. Two distance sampling procedures.
In each case, x denotes a sample point and solid lines indicate recorded distances.
Dashed lines de lineate areas of search.

and larger in aggregated patterns. In particular, let $x_1, \ldots, x_m$ be distances from each of m sample points to the nearest event within a sector of included angle Θ, and $y_1, \ldots, y_m$ the corresponding distances from each of m events to the nearest event within a sector of included angle Φ. Under CSR $\Theta\lambda\Sigma x_i^2$ and $\Phi\lambda\Sigma x_i^2$ are independently and identically distributed as chi-squared on 2m degrees of freedom, and the ratio $\Theta\Sigma x_i^2/\Phi\Sigma y_i^2$ is distributed as F on 2m and 2m degrees of freedom.

This approach to testing CSR was first suggested by Hopkins (1954), who assumed that the distances y_i would be measured from randomly selected events, with $\Theta=\Phi=2\pi$, giving test statistic

$$a=\Sigma x_i^2/\Sigma y_i^2. \tag{5.4.3.1}$$

The random selection of an event requires complete enumeration of the stand, which is presumably impractical. Byth & Ripley (1980) suggest a complete enumeration within a number of quadrats laid out systematically over the stand, giving n distances $y_1, \ldots y_n$ from which a random sample of size m can then be drawn. They recommend m quadrats of such a size that they will contain a total of about 5 m events. There would seem to be considerable difficulties in implementing this procedure in the field.
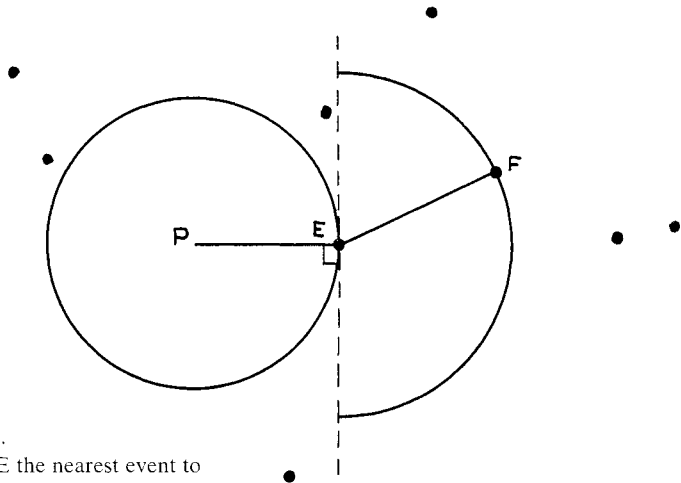
Figure 15. T-square sampling.
P is randomly located point, E the nearest event to P and F the nearest event to P within the half-plane delimited by the perpendicular to PE through E. Distance PE=x, distance EF=y.

Besag & Gleaves (1973) propose a "T-square" sampling procedure which is operationally simpler but leads to tests with reduced power, for equal m, in comparison with Hopkins' test. The sampling scheme is illustrated in Figure 15. We see that $\Theta=2\pi$, $\Phi=\pi$ and the test statistic is

$$t=2\Sigma x_i^2/\Sigma y_i^2. \tag{5.4.3.2}$$

Note in particular the orientation of the half-plane of search for the y_i-distances. A variant of t is

$$t^*=m^{-1}\Sigma\{x_i^2/(x_i^2+\tfrac{1}{2}y_i^2)\}, \tag{5.4.3.3}$$

whose sampling distribution under CSR is approximately Normal, with mean $\tfrac{1}{2}$ and variance $(12m)^{-1}$. Hines & Hines (1979) introduce a third statistic, which is Eberhardt's statistic (5.4.2.2), but calculated from the x_i and the $y_i/\sqrt{2}$, treated as a single sample of 2m observations.

Holgate (1965) uses distances (x_{1i}, x_{2i}) from each of m sample points to the nearest and second nearest events. Two possible test statistics are

$$h=\Sigma x_{1i}^2/(\Sigma x_{2i}^2-\Sigma x_{1i}^2) \tag{5.4.3.4}$$

and

$$h^*=\Sigma(x_{1i}^2/x_{2i}^2), \tag{5.4.3.5}$$

whose sampling distributions under CSR are the same as for the corresponding T-square statistics t and t^*.

Cox & Lewis (1976) tackle the complicated distribution which arises when the x_i are distances from sample points to nearest events and the y_i are distances from those events to *their* nearest events. They use only the $m_o\leq m$ pairs (x_i, y_i) for which $y_i<2x_i$ and consider the sequence of transformations $\Theta=2\sin^{-1}(y_i/2x_i)$, $w_i=\{2\pi-(\pi+\Theta_i)\cos\Theta_i+\sin\Theta_i\}^{-1}$ and $r_i=4(1-\pi w_i)/3$. This leads to a test statistic

$$cl=m_o^{-1}\Sigma r_i \tag{5.4.3.6}$$

39

whose distribution under CSR is approximately Normal, with mean $\frac{1}{2}$ and variance $(12m_o)^{-1}$. Cormack (1977) shows that pairs (x_i, y_i) with $y_i \geq 2x_i$ are uninformative.

A number of power comparisons of these statistics, based largely on results from simulations of various regular and aggregated alternatives, have been published. Diggle, Besag & Gleaves (1976) compare the statistics (5.4.3.1) to (5.4.3.5). They conclude that Hopkins' a is the most powerful, followed by t and t*, whilst Holgate's h and h* are relatively weak, especially so against regular alternatives. These findings are confirmed by Hines & Hines (1979) and Byth & Ripley (1980). Hines & Hines also suggest that Cox & Lewis' cl has very similar power characteristics to t*, despite its discarding some of the data, and that Eberhardt's statistic can give a very powerful test when used in conjunction with the T-square sampling procedure.

5.4.4 *Tests for heterogeneity*

All the tests described so far in this section are sensitive to changes in the "small-scale" pattern presented by the events in question. They are less effective in detecting departures from CSR in which CSR applies locally, but with possibly different intensities in different parts of the stand; this is particularly so of the bivariate procedures described in Section 5.4.3.

As a specific example, we first show how the point-centred quarter sampling procedure can be used in this context. Let x_{ij}: $j = 1,2,3,4$ be the distances obtained from the ith sample point. Then, $\pi \lambda_i (x_{i1}^2 + x_{i2}^2 + x_{i3}^2 + x_{i4}^2)/2$ is distributed as chi-squared on 8 degrees of freedom and for m sample points the hypothesis of CSR can be expressed as $\lambda_1 = \lambda_2 = \ldots = \lambda_m$. The likelihood ratio test of CSR against the alternative of unrestricted λ_i was given, in a different context, by Bartlett (1937). The test statistic is

$$b = 4\{m \log(\Sigma u_i/m) - \Sigma \log u_i\},$$

where $u = x_{i1}^2 + x_{i2}^2 + x_{i3}^2 + x_{i4}^2$, and the approximate sampling distribution of b under CSR is chi-squared on $m-1$ degrees of freedom. A significantly large value suggests rejection of CSR. The distributional approximation is improved if b is multiplied by a correction factor $c_m = 24m/(25m + 1)$. The same test can be applied to other distance sampling procedures, although the correction factor varies according to the particular sampling procedure used. In particular, Diggle (1977a) recommends a version based on T-square sampling, and intended to be used in conjunction with t*. The b-statistic uses $u = x^2 + \frac{1}{2}y^2$, and the correction factor is $12m/(13m+1)$. In the present context, the sampling distribution of t* is the same, whether or not the λ_i are equal. Thus, t* can be used to detect regular or aggregated patterns, but if t* does not reject CSR, b can be used to detect heterogeneity, giving a four-way classification of patterns as regular, random, heterogeneous or aggregated.

The b-statistic has been constructed for the specific purpose of detecting spatial variation in intensity, of the type illustrated by Figure 11a, whilst the tests described in Section 5.4.3 are concerned more with measuring "local" effects of aggregation, in the sense illustrated by Figure 10b, or regularity,

as in Figure 10a. The univariate procedures discussed in Section 5.4.2 can be thought of as general purpose procedures for which the alternative to CSR is left unspecified. Thus the choice of a test statistic may, and indeed should, be influenced by the range of alternatives to CSR which are under consideration. Further tests could be devised to test other specific classes of alternative. See, for example, Brown & Rothery's (1978) discussion of tests designed to detect local regularity combined with long-range aggregation.

5.5 Robust intensity estimation

The usual approach to intensity estimation is to derive an estimator which has attractive statistical properties for the special case of CSR, and then to investigate the extent to which its performance deteriorates under various types of departure from CSR. Quite generally, we use λ to denote the mean number of events per unit area for a spatial point pattern, consistent with its earlier usage as the intensity parameter of a homogeneous planar Poisson process.

Using quadrat sampling, with m quadrats each of area A, the natural estimator for λ is

$$\hat{\lambda}=\sum_{i=1}^{m} n_i / (mA)$$

For a realisation of CSR, $\hat{\lambda}$ is unbiased, with variance $\lambda/(mA)$. More generally, $\hat{\lambda}$ is *always* unbiased if the quadrats are randomly located, but its variance depends in an unpredictable way on the size and shape of the quadrats, as well as on the total quadrat area.

As with tests of CSR, practical considerations have led to the development of various distance-based estimators as alternatives to $\hat{\lambda}$. Using a distance-based approach it becomes more natural to estimate the inverse parameter $\gamma=\lambda^{-1}$, the mean area per event.

Under CSR, the maximum likelihood estimator for γ based on distances x_i from each of m randomly located points to the nearest neighbouring event is

$$\hat{\gamma}=\pi\sum_{i=1}^{m} x_i^2/m, \qquad (5.5.1)$$

which is unbiased for γ with variance γ^2/m. In contrast to $\hat{\lambda}$, unbiasedness is not guaranteed under any form of departure from CSR; in particular, in aggregated patterns the sample points tend to fall in empty spaces between clusters of events, leading to large values of x_i and positive bias in $\hat{\gamma}$.

This undesirable situation can be alleviated if estimators based on point-to-event and event-to-event distances are combined, essentially for the reason given at the beginning of Section 5.4.3. Diggle (1975, 1977b) investigates two estimators based on observations (x_i, y_i): $i=1, \ldots, m$ obtained by T-square sampling. The estimators are

$$\hat{\gamma}=\pi(\Sigma x_i^2+\tfrac{1}{2}\Sigma y_i^2) / (2 m)$$

and

$$\gamma_T^*=(\pi / m) \sqrt{(\tfrac{1}{2}\Sigma x_i^2 \, \Sigma y_i^2)}.$$

41

Of these, $\hat{\gamma}_T$ is the maximum likelihood estimator under CSR, but simulation studies suggest that γ_T^* enjoys better robustness properties. Byth (1980) suggests that a further improvement in robustness can be achieved by using

$$\tilde{\gamma}_T = 2\sqrt{2} \ (\Sigma x_i \ \Sigma y_i) \ / \ m^2.$$

The intuitive explanation given for this is that the use of squared distances in γ_T^* allows an occasional very large measurement to exert a disproportionate influence on the estimate of γ.

In a series of papers reviewed by Warren & Batcheler (1979), C L Batcheler develops an estimator whose starting point is a variant of (5.5.1) in which distances greater than some pre-determined threshold are grouped. An emprically determined correction factor is then applied to this estimator (or, strictly, its inverse regarded as an estimator for λ) in order to increase its robustness. Warren & Batcheler report that this estimator has worked well in a variety of applications, but Byth (1980) obtains relatively poor results from a simulation study.

Cox (1976) also discusses the use of a correction factor to improve robustness. His estimator takes as its starting point the bivariate sampling procedure used for the Cox & Lewis test of CSR. Preliminary results for this estimator again appear promising, but comparisons have not yet been made with other estimators.

Patil, Burnham & Kovner (1979) devise a distance-based estimator which is unbiased for any stationary point process which does not produce coincident points. The price paid for this form of robustness is an apparently large increase in variance and, again, no comparisons have been made with other estimators.

5.6 Spatial association

When two species are present in a stand, an analysis of the spatial pattern presented by each component species can be supplemented by tests of the hypothesis that the two patterns are generated by independent spatial point processes. When k species are present, the application of $\frac{1}{2}k(k-1)$ such tests applied to each pair of species is loosely analogous to the calculation of a correlation matrix for conventional multivariate data. As with tests of CSR, rejection or acceptance of the hypothesis under test is of limited interest *per se*, but rather should be seen as a means towards the end of describing, at least in a qualitative sense, the nature of any spatial associations which are detected.

Given the obvious practical relevance of this problem, the literature on it is surprisingly sparse. One possible explanation is that to be useful, a test for spatial association must be non-parametric in the sense that it should not make restrictive assumptions about the nature of the component patterns. In fact, it is not difficult to produce such non-parametric tests, but the power characteristics of these tests in a spatial contest have not yet been investigated in any depth. The following comments are based largely on results in Diggle & Cox (1981), and on the earlier work of Goodall (1965).

We consider a bivariate pattern, i.e. one formed by two species, and wish to test the hypothesis that this pattern is generated by two independent,

stationary, but otherwise unspecified point processes (by "stationary", we mean roughly that the intensity does not vary systematically over the area in question). Any of the following results can form the basis of a test of this hypothesis:

(i)   counts, N_1 and N_2 say, of the numbers of trees of each species in an arbitrarily located quadrat are independent random variables.

(ii) distances, X_1 and X_2 say, from an arbitrary point to the nearest tree of species 1 and species 2 respectively are independent random variables.

(iii) if E_1 and E_2 are the locations of the nearest tree of species 1 and 2 respectively to an arbitrary point P, the directed angle E_1PE_2 is uniformly distributed on the interval 0 to 360 degrees.

(iv) If X_1 and Y_1 are the distances from an arbitrary point to the nearest tree of species 1, and from an arbitrary tree of species 2 to the nearest tree of species 1, then X_1 and Y_1 are identically distributed random variables (and similarly for X_2 and Y_2).

Diggle & Cox (1981) discuss tests based on results (ii) to (iv) inclusive; their recommendation is to test for spatial association using a rank correlation coefficient calculated from observed values of X_1 and X_2.

If a quadrat-based sampling method is adopted, a natural test statistic would be the chi-squared based on the two-way frequency table of counts. We are aware of no published comparisons of distance-based and quadrat-based methods in this context.

5.7 Sampling in the field

We have not attempted to compare quadrat-based and distance-based tests and estimators, since the choice between these two sampling methods may ultimately depend more on practical considerations than on theoretical arguments, and such practical considerations will vary between applications. With regard to estimation, the guarantee of an unbiased estimator of intensity is a persuasive argument in favour of quadrat sampling, but the more robust distance-based estimators work well for moderate departures from CSR, and may give better results per man-hour in the field. Similarly with tests of CSR, the only valid comparison between quadrat-based and distance-based procedures would be one made on the basis of equal effort in the field.

All of the distance-based procedures require a number of sample points to be located within the stand. Conceptually the simplest scheme, but one which may be difficult in practice, is to locate sample points independently at random according to a uniform distribution over some designated region A. An alternative procedure would be to take points at fixed intervals along one or more straight lines drawn through the stand, or even to define a regular grid of points. Whichever procedure is adopted, it is vital that the field-worker is given precise instructions. Otherwise, there may be a tendency to avoid "difficult" spots and to choose sample points in relatively open areas of the stand, thus biasing the results. Also, the following points should be noted.

Firstly, distances measured from points too near the edge of the stand will tend to be larger than those measured from points well within the

stand. This represents a further source of bias which can be eliminated if the sampling region A is chosen to lie well within the region of interest. On the other hand, if A is made too small, the analysis may not be relevant to the stand as a whole. As a rough guide, A should be chosen so that no observed distance is likely to exceed the minimum distance between the edge of A and the edge of the stand.

Secondly, the number of sample points m should be small in relation to the number of events n in the sampling region A. Diggle, Besag & Gleaves (1976) recommend that the "sampling ratio" m/n should not exceed 0.1. Byth & Ripley (1980) suggest that for random sampling a value of $m/n \leq 0.05$ is preferable, but that a larger sampling ratio is acceptable for m points in a regular grid.

A further advantage of a regular grid is that it ensures a "representative" sample of points and allows for a retrospective assessment of any "smoothly varying" spatial phenomena. A possible danger is that the sampling grid *may* coincide with some underlying regularity generated by any management of the stand at an earlier stage of growth. Is this likely to occur in practice?

For quadrat-based methods, it is again important that the area under investigation should not be over-sampled. However, we know of no empirically determined guidelines for an acceptable sampling intensity, such as are cited above for distance-based methods.

6 Concluding remarks

Some of the analyses described in this report have proved more successful than others, and a recurrent theme has been the difficulty of drawing authoritative conclusions from observational data. In experimental work, it is often possible to formulate a precise statistical model for the data, and to justify the various assumptions by reference to the experimental procedure which, *inter alia*, attempts to control the effects of potential extraneous sources of variation. With observational studies, no such control is possible, it becomes less easy to justify assumptions, and any structure in the processes of primary interest may be obscured by apparently unstructured variation arising through uncontrolled factors, for example the microenvironment from which a tree draws nourishment.

The investigation in Section 2 is essentially exploratory in nature. It is valuable only if it points the investigator towards scientifically relevant hypotheses concerning the nature of spatial variation in young forest stands.

Section 3 is motivated by the need to provide a collection of quadrat count distributions which can then be used in simulations of the biological model of stand development. In this respect, it is disappointing that no satisfactory class of trivariate distributions has been identified, although the empirical evidence from an analysis of PTAX data suggests that univariate and bivariate distributions may be adequate in the present context.

Section 4 also attempts to answer a specific requirement for the biological simulation model. Theoretical relationships between two methods of assessing the random variation in plot-yields are derived, and shown to correspond reasonably well with empirical relationships observed for type-stand data. It would be interesting to investigate further the possible explanations for such discrepancies as are revealed between theory and data.

Section 5 is primarily a review of the extensive literature on statistical methods for sparsely sampled spatial point patterns. This remains an area of considerable research activity. In particular, there is a need both for further theoretical study of multivariate patterns and for practical investigation of field-sampling techniques to assess their usefulness (or otherwise!) to the forester.

7 References

Abramowitz, M & Segun, I A, 1965: Handbook of Mathematical Functions. New York, Dover.

Barnard, G A, 1963: Contribution to the Discussion of Professor Bartlett's paper. J R Statist Soc. B 25, 294.

Bartlett, M S, 1937: Properties of sufficiency and statistical tests. Proc. R. Soc A 160, 268–82.

Besag, J & Diggle, P J, 1977: Simple Monte Carlo tests for spatial pattern. Appl. Statist. 26, 327–33.

Besag, J E & Gleaves, J T, 1973: On the detection of spatial pattern in plant communities. Bull. Int. Statist. Inst. 45 (1) 153–8.

Blackman, G E, 1935: A study by statistical methods of the distribution of species in grassland associations. Ann. Bot. 49, 749–77.

Brown, D & Rothery, P, 1978: Randomness and local regularity of points in a plane. Biometrika 65, 115–22.

Byth, K, 1980: On robust distance-based intensity estimators. Unpublished typescript.

Byth, K & Ripley, B D, 1980: On sampling spatial patterns by distance methods. Biometrics 36, 279–84.

Catana, A J (1963). The wandering quarter method of estimating population density. Ecology 44, 349–60.

Cliff, A D & Ord, J K, 1973: Spatial autocorrelation. London: Pion.

Cormack, R M, 1977: The invariance of Cox & Lewis' statistic for the analysis of spatial patterns. Biometrika 64, 143–4.

– 1979: Spatial aspects of competition between individuals. In Spatial & Temporal Analysis in Ecology (ed R M Cormack, J K Ord), 151–211. Fairland: International Cooperative Publishing House.

Cottam, G & Curtis, J T, 1949: A method for making rapid surveys of woodlands, by means of pairs of randomly selected trees. Ecology 30, 101–4.

Cottam, G, Curtis, J T & Hale, B W, 1953: Some sampling characteristics of a population of randomly dispersed individuals. Ecology 34, 741–57.

Cox, D R & Lewis, P A W, 1966: The Statistical Analysis of Series of Events. London: Methuen.

Cox, T F, 1976: The robust estimation of the density of a forest stand using a new conditioned distance method. Biometrika 63, 493–500.

Cox, T F & Lewis, T, 1976: A conditioned distance ratio method for analysing spatial patterns. Biometrika 63, 483–92.

Diggle, P J, 1975: Robust density estimation using distance methods. Biometrika 62, 39–48.

– 1977a: The detection of random heterogeneity in plant populations. Biometrics 33, 390–4.

– 1977b: A note on robust density estimation for spatial point patterns. Biometrika 64, 91–5.

– 1978: On parameter estimation for spatial point processes. J R Statist Soc. B 40, 178–81.

– 1979: On parameter estimation and goodness-of-fit testing for spatial point patterns. Biometrics 35, 87–101.

Diggle, P J, Besag, J E & Gleaves, J T, 1976: Statistical analysis of spatial point patterns by means of distance methods. Biometrics 32, 659–67.

Diggle, P J & Cox, T F (1981). On sparse sampling methods and statistical analysis of multivariate spatial point patterns. Proc 43rd Session, Int Statist. Inst (to appear).

Du Rietz, G E, 1929: The fundamental units of vegetation. Proceedings of the International Congress of Plant Science. Ithaca, 1, 123–7.

– 1930: Classification and nomenclature of vegetation. Svensk Botanisk Tidskrift, 24, 489–503.

Eberhardt, L L, 1967: Some developments in distance sampling. Biometrics 23, 207–16.

Eneroth, O, 1945: Om frömängden vid fläcksådd samt om sambandet mellan plantantal per ha och slutenhet vid självsådd. Norrl. Skogsvårdsförb. Tidskrift, 161–222.

Evans, D A, 1953: Experimental evidence concerning contagious distributions in ecology. Biometrika, 40, 186–211.

Fisher, R A, Thornton, H G & Mackenzie, W A, 1922: The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples. Ann. Appl. Bot. 9, 325–59.

Ghent, A W, 1963: Studies of regeneration of forest stands devastated by spruce budworm. For. Sci. 9, 295–310.

Goodall, D W, 1965: Plot-less tests of interspecific association. J Ecol. 53, 197–210.

Green, P J & Sibson, R, 1978: Computing Dirichlet tessellations in the plane. Comp J 21, 168–73.

Griffiths, D A, 1973: Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics, 29, 637–48.

Gustafsson, R, 1974: Typestands for cleaning. Dept

Operational Efficiency, Royal College of Forestry, Research Notes No 70.

Hines, W G S & Hines, R J O, 1979: The Eberhardt statistic and the detection of nonrandomness of spatial point distributions. Biometrika 66, 73–80.

Hogg, R V & Craig, A T, 1970: Introduction to Mathematical Statistics (3rd ed). London: Collier-Macmillan.

Holgate, P, 1965: Some new tests of randomness. J Ecol. 53, 261–6.

– 1972: The use of distance methods for the analysis of spatial distributions of points. In Stochastic Point Processes (ed P A W Lewis), 122–35. New York: Wiley.

Hope, A C A, 1968: A simplified Monte Carlo significance test procedure. J R Statist. Soc. B 30, 582–98.

Hopkins, B, 1954: A new method of determining the type of distribution of plant individuals. Ann. Bot. 18, 213–26.

Hultén, H & Jansson, K-Å, 1972: Fältinstruktion, planttaxering rutin 1, PTAX R1. Intern rapport, Skogshögskolan, Institutionen för skogsföryngring.

Lindgren, B W, 1968: Statistical Theory. London: Collier-Macmillan.

Lloyd, M, 1967: Mean crowding. J Animal Ecol. 36, 1–30.

Mardia, K V, 1970. Families of Bivariate Distributions. London: Griffin.

Marriott, F H C, 1979: Monte Carlo tests: how many simulations? Appl. Statist. 28, 75–7.

Matérn, B, 1971: Doubly stochastic Poisson processes in the plane. In Statistical Ecology Vol. 1 (ed G P Patil, E C Pielou & W E Waters), 195–213. Pennsylvania State Univ. Press.

– 1972: The precision of basal area estimates. For Sci, 18, 123–5.

Mead, R, 1974: A test for spatial pattern at several scales using data from a grid of contiguous quadrats. Biometrics 30, 295–307.

Moran, P A P, 1950: Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Morisita, M, 1957: A new method for the estimation of density by the spacing method, applicable to non-randomly distributed populations (in Japanese). Physiol and Ecol. 7, 134–44.

Nelder, J A & Mead, R, 1965: A simplex method for function minimisation. Comp. J. 7, 308–13.

Patil, G P, 1965: Classical and Contagious Discrete Distributions. New York: Pergamon Press.

Patil, G P & Joshi, S W, 1968: A Dictionary and Bibliography of Discrete Distributions. Edinburgh: Oliver & Boyd.

Patil S A, Burnham, K P & Kovner, J L, 1979: Nonparametric estimation of plant density by the distance method. Biometrics 35, 597–604.

Perry, J N & Mead, R, 1979: On the power of the index of dispersion test to detect spatial pattern. Biometrics 35, 613–22.

Plackett, R L, 1965. A class of bivariate distributions. J Amer. Statist. Ass. 60, 516–22.

Ripley, B D, 1977: Modelling spatial patterns (with Discussion). J R Statist. Soc. B 39, 172–212.

Rogers, C A, 1964. Packing & Covering. London: Cambridge University Press.

Silvey, S D, 1975: Statistical Inference. London: Chapman & Hall.

Tirén, L, 1949: Om den naturliga föryngringen på obrända hyggen i norrländsk granskog. Meddel. från Statens Skogsforskningsinst., Band 38:9, (210 pp).

Warren, W A & Batcheler, C L, 1979: The density of spatial patterns: robust estimation through distance methods. In Spatial & Temporal Analysis in Ecology, ed R M Cormack & J K Ord, 247–70. Fairland: International Co-operative Publiching House.

Watson, G S, 1971: Trend-surface analysis. J Int Ass. Math Geol. 3, 215–26.