Spatial Interaction and the Statistical Analysis of Lattice Systems

By JULIAN BESAG

University of Liverpool

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, March 13th, 1974, Professor J. DURBIN in the Chair]

SUMMARY

The formulation of conditional probability models for finite systems of spatially interacting random variables is examined. A simple alternative proof of the Hammersley–Clifford theorem is presented and the theorem is then used to construct specific spatial schemes on and off the lattice. Particular emphasis is placed upon practical applications of the models in plant ecology when the variates are binary or Gaussian. Some aspects of infinite lattice Gaussian processes are discussed. Methods of statistical analysis for lattice schemes are proposed, including a very flexible coding technique. The methods are illustrated by two numerical examples. It is maintained throughout that the conditional probability approach to the specification and analysis of spatial interaction is more attractive than the alternative joint probability approach.

Keywords: MARKOV FIELDS; SPATIAL INTERACTION; AUTO-MODELS; NEAREST-NEIGHBOUR SCHEMES; STATISTICAL ANALYSIS OF LATTICE SCHEMES; CODING TECHNIQUES; SIMULTANEOUS BILATERAL AUTOREGRESSIONS; CONDITIONAL PROBABILITY MODELS

1. INTRODUCTION

IN this paper, we examine some stochastic models which may be used to describe certain types of spatial processes. Potential applications of the models occur in plant ecology and the paper concludes with two detailed numerical examples in this area. At a formal level, we shall largely be concerned with a rather arbitrary system, consisting of a finite set of sites, each site having associated with it a univariate random variable. In most ecological applications, the sites will represent points or regions in the Euclidean plane and will often be subject to a rigid lattice structure. For example, Cochran (1936) discusses the incidence of spotted wilt over a rectangular array of tomato plants. The disease is transmitted by insects and, after an initial period of time, we should clearly expect to observe clusters of infected plants. The formulation of spatial stochastic models will be considered in Sections 2–5 of the paper. Once having set up a model to describe a particular situation, we should then hope to be able to estimate any unknown parameters and to test the goodness-of-fit of the model on the basis of observation. We shall discuss the statistical analysis of lattice schemes in Sections 6 and 7.

We begin by making some general comments on the types of spatial systems which we shall, and shall not, be discussing. Firstly, we shall not be concerned here with any random distribution which may be associated with the locations of the sites themselves. Indeed, when setting up models in practice, we shall require quite specific information on the relative positions of sites, in order to assess the likely interdependence between the associated random variables. Secondly, although, as in

Cochran's example above, the system may, in reality, have developed continuously through time, we shall always assume that observation on it is only available at an isolated instant; hence, we shall not be concerned here with the setting up of spatial–temporal schemes. This has the important consequence that our models will not be mechanistic and must be seen as merely attempts at describing the "here and now" of a wider process. In many practical situations, this is a reasonable standpoint, since we can only observe the variables at a single point in time (for example, the yields of fruit trees in an orchard) but, in other cases, a spatial–temporal approach may be more appropriate. In fact, the states of the tomato plants, in Cochran's example, were observed at three separate points in time and it is probably most profitable to use a classical temporal autoregression to analyse the system. A similar comment applies to the hop plants data of Freeman (1953). Ideally, even when dealing with a process at a single instant of time, we should first set up an intuitively plausible spatial–temporal model and then derive the resulting instantaneous spatial structure. This can sometimes be done if we are prepared to assume stationarity in both time and space (see Bartlett, 1971a) but, unfortunately, such an assumption is unlikely to be realistic in our context. However, when this approach is justifiable, it is of course helpful to check that our spatial models are consistent with it; for simple examples, see Besag (1972a). Otherwise, regarding the transient spatial structure of a spatial–temporal process, this is almost always intractable and hence there exists a need to set up and examine purely spatial schemes without recourse to temporal considerations.

The following examples are intended as typical illustrations of the spatial situations we shall have in mind. They are classified according to the nature of

(a) the system of sites (regular or irregular),
(b) the individual sites (points or regions) and
(c) the associated random variables (discrete or continuous).

1.1. A *regular lattice of point sites with discrete variables* commonly occurs under experimental conditions in plant ecology. Examples include the pattern of infection in an array of plants (Cochran, 1936, as described above; Freeman, 1953, on the incidence of nettlehead virus in hop plants) and the presence or absence of mature plants seeded on a lattice and subsequently under severe competition for survival (data kindly supplied by Dr E. D. Ford, Institute of Tree Biology, Edinburgh, relates to dwarf French marigolds on a triangular lattice of side 2 cm). Often, as above, the data are binary.

1.2. A *regular lattice of point sites with continuous variables* commonly occurs in agricultural experiments, where individual plant yields are measured (Mead, 1966, 1967, 1968, on competition models; Batchelor and Reed, 1918, on fruit trees). It is often reasonable to assume that the variates have a multivariate normal distribution.

1.3. A *regular lattice of regions with discrete variables* arises in sampling an irregularly distributed population when a rectangular grid is placed over an area and counts are made of the number of individuals in each quadrat (Professor P. Greig-Smith on *Carex arenaria*, in Bartlett, 1971b; Gleaves, 1973, on *Plantago lanceolata*; Clarke, 1946, and Feller, 1957, p. 150, on flying-bomb hits in South London during World War II; Matui, 1968, on the locations of farms and villages in an area of Japan). In plant ecology, the quadrats are often so small that few contain more than a single plant and it is then reasonable to reduce the data to a binary (presence/absence) form.

1.4. A *regular lattice of regions with continuous variables* typically occurs in field trials where aggregate yields are measured (Mercer and Hall, 1911, on wheat plots). Multivariate normality is often a reasonable assumption.

1.5. *Irregular point sites with discrete variables* arise in sampling natural plant populations. Examples include the presence or absence of infection in individuals and the variety of plant at each site in a multi-species community.

1.6. *Irregular point sites with continuous variables* again occur in sampling natural plant populations (Brown, 1965, on tree diameters in pine forests; Mead, 1971, on competition models).

1.7. *Irregular regions with discrete or continuous variables* have applications particularly in a geographical context, with regions defined by administrative boundaries (O'Sullivan, 1969, and Ord, 1974, on aspects of the economy of Eire).

It has previously been stated that in the practical construction of spatial models, we shall require precise information concerning the relative positions of the various sites. Where the sites are regions, rather than points, the data are by definition, aggregate data and the assumption of single, uniquely defined locations for each of the associated variables is clearly open to criticism. For example, quadrat counts (Section 1.3) are usually used to examine spatial pattern rather than spatial interaction. Further comments will appear in Section 5.

Combinations of the above situations may occur. For example, in competition experiments where yields are measured, "missing observations" may be due to intense competition and should then be specifically accounted for by the introduction of mixed distributions. We shall not contemplate such situations here.

2. Conditional Probability Approach to Spatial Processes

There appear to be two main approaches to the specification of spatial stochastic processes. These stem from the non-equivalent definitions of a "nearest-neighbour" system, originally due to Whittle (1963) and Bartlett (1955 Section 2.2, 1967, 1968), respectively. Suppose, for definiteness, that we temporarily restrict attention to a rectangular lattice with sites labelled by integer pairs (i,j) and with an associated set of random variables $\{X_{i,j}\}$. For the moment, we ignore any problems concerning the finiteness or otherwise of the lattice. Then Whittle's basic definition requires that the *joint* probability distribution of the variates should be of the product form

$$\prod_{i,j} Q_{i,j}(x_{i,j};\, x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}), \tag{2.1}$$

where $x_{i,j}$ is a value of the random variable, $X_{i,j}$. On the other hand, Bartlett's definition requires that the *conditional* probability distribution of $X_{i,j}$, given all other site values, should depend only upon the values at the four nearest sites to (i,j), namely $x_{i-1,j}$, $x_{i+1,j}$, $x_{i,j-1}$ and $x_{i,j+1}$. Whilst the conditional probability formulation may be said to have rather more intuitive appeal, this is marred by a number of disadvantages.

Firstly, there is no obvious method of deducing the joint probability structure associated with a conditional probability model. Secondly, the conditional probability structure itself is subject to some unobvious and highly restrictive consistency conditions. When these are enforced, it can be shown (Brook, 1964) that the conditional probability formulation is degenerate with respect to (2.1). Thirdly, it has been remarked by Whittle (1963) that the natural specification of an equilibrium process in statistical mechanics is in terms of the joint distribution rather than the conditional distribution of the variables.

These problems were partially investigated in a previous paper (Besag, 1972a). The constraints on the conditional probability structure were identified for homogeneous systems and found to be so severe that they actually generated particular spatial models, given the nature of the variables. Had these models failed to retain any practical appeal, then there would have been little further scope for discussion. However, this is not the case. For example, with binary variables, the conditional probability formulation *necessarily* generates just that basic model (the Ising model of ferromagnetism) which has been at the centre of so much work in statistical mechanics. Thus, although this model may *classically* be formulated in terms of joint probabilities, it is generated in a natural way through basic *conditional* probability assumptions. This fact may also be related to the problem of degeneracy. There is surely no indignity in studying a subclass of schemes provided that subclass is of interest in its own right. However, we go further. Suppose we consider wider classes of conditional probability models in which the conditional distribution of $X_{i,j}$ is allowed to depend upon the values at more remote sites. We can build up a hierarchy of models, more and more general,which eventually will include the scheme (2.1) and any particular generalization of it. That is, we extend the concept of first-, second- and higher-order Markov chains in one dimension to the realm of spatial processes. There is then no longer any degeneracy associated with the conditional probability models. This is the approach taken in the present paper. It has been made possible by the advent of the celebrated Hammersley–Clifford theorem which, sadly, has remained unpublished by its authors.

Finally, in this section, we examine the problems and implications of deriving the joint probability structure associated with the site variables, given their individual conditional distributions. We no longer restrict attention to "nearest-neighbour" models nor even to lattice schemes but instead consider a fairly arbitary system of sites. Suppose then that we are concerned with a *finite* collection of random variables, $X_1, \ldots, X_n$, which are associated with sites labelled $1, \ldots, n$, respectively. For each site, $P(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, the conditional distribution of X_i, given all other site values, is specified and we require the joint probability distribution of the variables. Our terminology will be appropriate to discrete variables but the arguments equally extend to the continuous case.

We make the following important assumption: if $x_1, \ldots, x_n$ can individually occur at the sites $1, \ldots, n$, respectively, then they can occur together. Formally, if $P(x_i) > 0$ for each i, then $P(x_1, \ldots, x_n) > 0$. This is called the *positivity* condition by Hammersley and Clifford (1971) and will be assumed throughout the present paper. It is usually satisfied in practice. We define the sample space Ω to be the set of all possible realizations $\mathbf{x} = (x_1, \ldots, x_n)$ of the system. That is, $\Omega = \{\mathbf{x}: P(\mathbf{x}) > 0\}$. It then follows that for any two given realizations $\mathbf{x}$ and $\mathbf{y} \in \Omega$,

$$\frac{P(\mathbf{x})}{P(\mathbf{y})} = \prod_{i=1}^{n} \frac{P(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{P(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}. \tag{2.2}$$

The proof of this result resembles that of equation (6) in Besag (1972a). Clearly, we may write

$$P(\mathbf{x}) = P(x_n | x_1, \ldots, x_{n-1}) P(x_1, \ldots, x_{n-1}); \tag{2.3}$$

however, $P(x_1, \ldots, x_{n-1})$ cannot be factorized in a useful way since, for example, $P(x_{n-1} | x_1, \ldots, x_{n-2})$ is not easily obtained from the given conditional distributions.

Nevertheless, we can introduce y_n, write

$$P(\mathbf{x}) = \frac{P(x_n | x_1, \ldots, x_{n-1})}{P(y_n | x_1, \ldots, x_{n-1})} P(x_1, \ldots, x_{n-1}, y_n)$$

and now operate on x_{n-1} in $P(x_1, \ldots, x_{n-1}, y_n)$. This yields

$$P(x_1, \ldots, x_{n-1}, y_n) = \frac{P(x_{n-1} | x_1, \ldots, x_{n-2}, y_n)}{P(y_{n-1} | x_1, \ldots, x_{n-2}, y_n)} P(x_1, \ldots, x_{n-2}, y_{n-1}, y_n),$$

after the similar introduction of y_{n-1}. Continuing the reduction process, we eventually arrive at equation (2.2) which clearly determines the joint probability structure of the system in terms of the given conditional probability distributions. We require the positivity condition merely to ensure that each term in the denominator of (2.2) is non-zero.

Equation (2.2) highlights the two fundamental difficulties concerning the specification of a system through its conditional probability structure. Firstly, the labelling of individual sites in the system being arbitrary implies that many factorizations of $P(\mathbf{x})/P(\mathbf{y})$ are possible. All of these must, of course, be equivalent and this, in turn, implies the existence of severe restrictions on the available functional forms of the conditional probability distributions in order to achieve a mathematically consistent joint probability structure. This problem has been investigated by Lévy (1948), Brook (1964), Spitzer (1971), Hammersley and Clifford (1971) and Besag (1972a) and we discuss it in detail in the next section. Secondly, whilst expressions for the relative probabilities of two realizations may be fairly straightforward, those for absolute probabilities, in general, involve an extremely awkward normalizing function with the consequence that direct approaches to statistical inference through the likelihood function are rarely possible. We shall have to negotiate this problem in Section 6 of the paper.

3. Markov Fields and the Hammersley–Clifford Theorem

In this section, we examine the constraints on the functional form of the conditional probability distribution available at each of the sites. We restate a theorem of Hammersley and Clifford (1971) and give a simple alternative proof. This theorem, which has received considerable attention recently, is essential to the construction of valid spatial schemes through the conditional probability approach. We begin by describing the problem more precisely. Our definitions will closely follow those of Hammersley and Clifford.

The first definition determines the *set of neighbours* for each site. Thus, site $j\,(\neq i)$ is said to be a neighbour of site i if and only if the functional form of $P(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is dependent upon the variable x_j. As the simplest example, suppose that $X_1, \ldots, X_n$ is a Markov chain. Then it is easily shown that site $i\,(2 \leqslant i \leqslant n-1)$ has neighbours $i-1$ and $i+1$ whilst the sites 1 and n have the single neighbours 2 and $n-1$, respectively. For a more interesting spatial example, suppose the sites form a finite rectangular lattice and are now conveniently labelled by integer pairs (i,j). Then, if $P(x_{i,j} |$ all other site values) depends only upon $x_{i,j}$, $x_{i-1,j}$, $x_{i+1,j}$, $x_{i,j-1}$ and $x_{i,j+1}$ for each internal site (i,j), we have a so-called "nearest-neighbour" lattice scheme. In such a case, each internal site (i,j) has four neighbours, namely $(i-1,j)$, $(i+1,j)$, $(i,j-1)$ and $(i,j+1)$. (There is a slight inconsistency in the usage of the word "neighbour" here: this will be resolved in later sections by introducing the

term "first-order" scheme.) Any system of n sites, each with specified neighbours, clearly generates a class of valid stochastic schemes. We call any member of this class a *Markov field.* Our aim is to be able to identify the class in any given situation.

Any set of sites which either consists of a single site or else in which every site is a neighbour of every other site in the set is called a *clique.* Thus, in the "nearest-neighbour" situation described above, there are cliques of the form $\{(i,j)\}$, $\{(i-1,j),(i,j)\}$ and $\{(i,j-1),(i,j)\}$ over the entire lattice, possibly with adjustments at the boundary. The definition of a clique is crucial to the construction of valid Markov fields.

We now make two assumptions, again following Hammersley and Clifford. Firstly, we suppose that there are only a finite number of values available at each site, although we shall relax this condition later in the section. Secondly, we assume that the value zero is available at each site. If this is originally untrue, it can always be subsequently brought about by re-indexing the values taken at the offending sites, a procedure which will be illustrated in Section 4.3. This second assumption, which is therefore made for purely technical reasons, ensures that, under the positivity condition, an entire realization of zeros is possible. That is, $P(\mathbf{0})>0$ and we may legitimately define

$$Q(\mathbf{x}) \equiv \ln\{P(\mathbf{x})/P(\mathbf{0})\} \tag{3.1}$$

for any $\mathbf{x}\in\Omega$. Lastly given any $\mathbf{x}\in\Omega$, we write $\mathbf{x}_i$ for the realization

$$(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n).$$

The problem to which Hammersley and Clifford addressed themselves may now be stated as follows: *given the neighbours of each site, what is the most general form which $Q(\mathbf{x})$ may take in order to give a valid probability structure to the system?* Since

$$\begin{aligned}\exp\{Q(\mathbf{x})-Q(\mathbf{x}_i)\} &= P(\mathbf{x})/P(\mathbf{x}_i)\\ &= P(x_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)/P(0 \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),\end{aligned} \tag{3.2}$$

the solution to this problem immediately gives the most general form which may be taken by the conditional probability distribution at each site.

In dealing with the rather general situation described above, the Hammersley–Clifford theorem superseded the comparatively pedestrian results which had been obtained for "nearest-neighbour" systems on the k-dimensional finite cubic lattice (Spitzer, 1971; Besag, 1972a). However, the original method of proof is circuitous and requires the development of an operational calculus (the "blackening algebra"). A simple alternative statement and proof of the theorem rest upon the observation that for any probability distribution $P(\mathbf{x})$, subject to the above conditions, there exists an expansion of $Q(\mathbf{x})$, unique on Ω and of the form

$$\begin{aligned}Q(\mathbf{x}) = \sum_{1\leqslant i\leqslant n} x_i G_i(x_i) + \sum\sum_{1\leqslant i<j\leqslant n} x_i x_j G_{i,j}(x_i, x_j) + \sum\sum\sum_{1\leqslant i<j<k\leqslant n} x_i x_j x_k G_{i,j,k}(x_i, x_j, x_k) + \ldots\\ + x_1 x_2 \ldots x_n G_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n).\end{aligned} \tag{3.3}$$

For example, we have

$$x_i G_i(x_i) \equiv Q(0, \ldots, 0, x_i, 0, \ldots, 0) - Q(\mathbf{0}),$$

with analogous difference formulae for the higher order G-functions. With the above notation, Hammersley and Clifford's result may be stated in the following manner:

for any $1 \leqslant i < j < \ldots < s \leqslant n$, *the function* $G_{i,j,\ldots,s}$ *in* (3.3) *may be non-null if and only if the sites* $i, j, \ldots, s$ *form a clique. Subject to this restriction, the G-functions may be chosen arbitrarily.* Thus, given the neighbours of each site, we can immediately write down the most general form for $Q(\mathbf{x})$ and hence for the conditional distributions. We shall see examples of this later on.

Proof of theorem. It follows from equation (3.2) that, for any $\mathbf{x} \in \Omega$, $Q(\mathbf{x}) - Q(\mathbf{x}_i)$ can only depend upon x_i itself and the values at sites which are neighbours of site i. Without loss of generality, we shall only consider site 1 in detail. We then have, from equation (3.3),

$$Q(\mathbf{x}) - Q(\mathbf{x}_1) = x_1 \Big\{ G_1(x_1) + \sum_{2 \leqslant j \leqslant n} x_j G_{1,j}(x_1, x_j) + \sum\sum_{2 \leqslant j < k \leqslant n} x_j x_k G_{1,j,k}(x_1, x_j, x_k) + \ldots + x_2 x_3 \ldots x_n G_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \Big\}.$$

Now suppose site l ($\neq 1$) is *not* a neighbour of site 1. Then $Q(\mathbf{x}) - Q(\mathbf{x}_1)$ must be independent of x_l for all $\mathbf{x} \in \Omega$. Putting $x_i = 0$ for $i \neq 1$ or l, we immediately see that $G_{1,l}(x_1, x_l) = 0$ on Ω. Similarly, by other suitable choices of $\mathbf{x}$, it is easily seen successively that all 3-, 4-, ..., n-variable G-functions involving both x_1 and x_l must be null. The analogous result holds for any pair of sites which are not neighbours of each other and hence, in general, $G_{i,j,\ldots,s}$ can only be non-null if the sites $i, j, \ldots, s$ form a clique.

On the other hand, any set of G-functions gives rise to a valid probability distribution $P(\mathbf{x})$ which satisfies the positivity condition. Also since $Q(\mathbf{x}) - Q(\mathbf{x}_i)$ depends only upon x_l if there is a non-null G-function involving both x_i and x_l, it follows that the same is true of $P(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. This completes the proof.

We now consider some simple extensions of the theorem. Suppose firstly that the variates can take a denumerably infinite set of values. Then the theorem still holds if, in the second part, we impose the added restriction that the G-functions be chosen such that $\sum \exp Q(\mathbf{x})$ is finite, where the summation is over all $\mathbf{x} \in \Omega$. Similarly, if the variates each have absolutely continuous distributions and we interpret $P(\mathbf{x})$ and allied quantities as probability densities, the theorem holds provided we ensure that $\exp Q(\mathbf{x})$ is integrable over all $\mathbf{x}$. These additional requirements must not be taken lightly, as we shall see by examples in Section 4. Finally, we may consider the case of multivariate rather than univariate site variables. In particular, suppose that the random vector at site i has ν_i components. Then we may replace that site by ν_i notional sites, each of which is associated with a single component of the random vector. An appropriate system of neighbours may then be constructed and the univariate theorem be applied in the usual way. We shall not consider the multivariate situation any further in the present paper.

As a straightforward corollary to the theorem, it may easily be established that for any given Markov field

$$P(X_i = x_i,\ X_j = x_j,\ \ldots,\ X_s = x_s \,|\, \text{all other site values})$$

depends only upon $x_i, x_j, \ldots, x_s$ and the values at sites neighbouring sites $i, j, \ldots, s$. In the Hammersley–Clifford terminology, the *local* and *global* Markovian properties are equivalent.

In practice, we shall usually find that the sites occur in a finite region of Euclidean space and that they often fall naturally into two sets: those which are internal to the

system and those which form its boundary (or boundaries). In constructing a Markov field, it is quite likely that we are able to make reasonable assumptions concerning the conditional distribution associated with each of the internal sites but that problems arise at the boundary of the system. Such problems may usually be by-passed by considering the joint distribution of the internal site variables conditional upon fixed (observed) boundary values. We need then only specify the neighbours and associated conditional probability structure for each of the internal sites in order to define uniquely the above joint distribution. This is a particularly useful approach for lattice systems.

The positivity condition remains as yet unconquered and it would be of considerable theoretical interest to learn the effect of its relaxation. On the other hand, it is probably fair to say that the result would be of little practical significance in the analysis of spatial interaction with given site locations.

Finally, we note that, for discrete variables, a further proof of the Hammersley–Clifford theorem has been given by Grimmett (1973). This is apparently based upon the Möbius inversion theorem (Rota, 1964). Other references on the specification of Markov fields include Averintsev (1970), Preston (1973) and Sherman (1973).

4. Some Spatial Schemes Associated with the Exponential Family

In the next two sections, we become more specific in our discussion of spatial schemes. The present section deals with a particular subclass of Markov fields and with some of the models which are generated by it, whilst Sections 5.1, 5.2 and 5.3 are more concerned with practical aspects of conditional probability models. In Section 5.4, the simultaneous autoregressive approach (Mead, 1971; Ord, 1974) to finite spatial systems is discussed, again from the conditional probability viewpoint. Finally, in Section 5.5, stationary auto-normal models on the infinite regular lattice are defined and compared with the stationary simultaneous autoregressions of Whittle (1954).

In the remainder of this paper, we shall use the function $p_i(.)$ to denote the conditional probability distribution (or density function) of X_i given all other site values. Thus $p_i(.)$ is a function of x_i and of the values at sites neighbouring site i. Wherever possible, the arguments of $p_i(.)$ will be omitted.

4.1. *Auto-models*

Given n sites, labelled $1, \ldots, n$, and the set of neighbours for each, we have seen in Section 3 how the Hammersley–Clifford theorem generates the class of valid probability distributions associated with the site variables $X_1, \ldots, X_n$. Within this general framework, we shall in Section 4.2 consider particular schemes for which $Q(\mathbf{x})$ is well defined and has the representation

$$Q(\mathbf{x}) = \sum_{1 \leqslant i \leqslant n} x_i G_i(x_i) + \sum\sum_{1 \leqslant i < j \leqslant n} \beta_{i,j} x_i x_j, \tag{4.1}$$

where $\beta_{i,j} = 0$ unless sites i and j are neighbours of each other. Such schemes will be termed *auto-models*.

In order to motivate this definition, it is convenient to consider the wider formulation below. Suppose we make the following assumptions.

Assumption 1. The probability structure of the system is dependent only upon contributions from cliques containing no more than two sites. That is, when well defined, the expansion (3.3) becomes

$$Q(\mathbf{x}) = \sum_{1\leqslant i\leqslant n} x_i G_i(x_i) + \sum\sum_{1\leqslant i<j\leqslant n} x_i x_j G_{i,j}(x_i, x_j), \tag{4.2}$$

where $G_{i,j}(.)\equiv 0$ unless sites i and j are neighbours.

Assumption 2. The conditional probability distribution associated with each of the sites belongs to the *exponential family* of distributions (Kendall and Stuart, 1961, p. 12). That is, for each i,

$$\ln p_i(x_i; \ldots) = A_i(.)\, B_i(x_i) + C_i(x_i) + D_i(.), \tag{4.3}$$

where the functions B_i and C_i are of specified form and A_i and D_i are functions of the values at sites neighbouring site i. A valid choice of A_i determines the type of dependence upon neighbouring site values and D_i is then the appropriate normalizing function.

It is shown in Section 4.3 that as a direct consequence of Assumptions 1 and 2, A_i must satisfy

$$A_i(.)\equiv \alpha_i + \sum_{j=1}^{n} \beta_{i,j} B_j(x_j), \tag{4.4}$$

where $\beta_{j,i}\equiv\beta_{i,j}$ and $\beta_{i,j} = 0$ unless sites i and j are neighbours of each other. Hence, it follows, when appropriate, that $G_{i,j}$ in equation (4.2) has the form

$$G_{i,j}(x_i, x_j)\equiv \beta_{i,j} H_i(x_i) H_j(x_j), \tag{4.5}$$

where $x_i H_i(x_i) = B_i(x_i) - B_i(0)$. Thus we generate the class of auto-models by making the additional requirement that, for each i, the function B_i is linear in x_i.

Superficially, auto-models might appear to form quite a useful subclass of Markov fields. Assumption 1 is not only satisfied for any rectangular lattice "nearest-neighbour" scheme but can also be taken as a fairly natural starting point in much wider lattice and non-lattice situations. Further, the linearity of B_i is satisfied by the most common members of the exponential family. However, the assumptions are, in fact, so restrictive, as seen through equation (4.4), that they often produce models which, in the end result, are devoid of any intuitive appeal at all. In Section 4.2, a range of auto-models has been included and hopefully illustrates both ends of the spectrum. Practical applications of two of the models will be discussed in later sections.

It is clear that, in terms of equation (4.1), auto-models have conditional probability structure satisfying

$$p_i(x_i; \ldots)/p_i(0; \ldots) = \exp\left[x_i\left\{G_i(x_i) + \sum_{j=1}^{n} \beta_{i,j} x_j\right\}\right], \tag{4.6}$$

where again $\beta_{j,i}\equiv\beta_{i,j}$ and $\beta_{i,j} = 0$ unless sites i and j are neighbours of each other. The models can further be classified according to the form which $p_i(.)$ takes and this leads to the introduction of terms such as *auto-normal*, *auto-logistic* and *auto-binomial* to describe specific spatial schemes.

In the subsequent discussion, it will be assumed, unless otherwise stated, that any parameters $\beta_{i,j}$ are at least subject to the conditions following equation (4.6). Ranges of summation will be omitted wherever possible and these should then be apparent by comparison with equation (4.1) or (4.6).

4.2. *Some Specific Auto-models*

4.2.1. *Binary schemes*

For any finite system of binary (zero–one) variables, the only occasions upon which a given non-null G-function can contribute to $Q(\mathbf{x})$ in the expansion (3.3) are those upon which each of its arguments is unity. We may therefore replace all non-null G-functions by single arbitrary parameters, without any loss of generality, and this leads to the multivariate logistic models of Cox (1972). One would hope in practice that only a fairly limited number of non-zero parameters need to be included. In particular, if the only non-zero parameters are those associated with cliques consisting of single sites and of pairs of sites, we have an auto-logistic model for which we may write

$$Q(\mathbf{x}) = \sum \alpha_i x_i + \sum\sum \beta_{i,j} x_i x_j. \tag{4.7}$$

It follows that

$$p_i(.) = \exp\{x_i(\alpha_i + \sum \beta_{i,j} x_j)\}/\{1 + \exp(\alpha_i + \sum \beta_{i,j} x_j)\}, \tag{4.8}$$

analogous to a classical logistic model (Cox, 1970, Chapter 1), except that here the explanatory variables are themselves observations on the process.

4.2.2. *Gaussian schemes*

In many practical situations, especially those arising in plant ecology, it is reasonable to assume that the joint distribution of the site variables (plant yields), possibly after suitable transformation, is multivariate normal. It is evident that any such scheme is an *auto-normal* scheme. In particular, we shall consider schemes for which

$$p_i(.) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\sigma^{-2}\{x_i - \mu_i - \sum \beta_{i,j}(x_j - \mu_j)\}^2]. \tag{4.9}$$

Using the factorization (2.2) or otherwise, this leads to the joint density function,

$$P(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{1}{2}n} |\mathbf{B}|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\sigma^{-2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{B}(\mathbf{x}-\boldsymbol{\mu})\}, \tag{4.10}$$

where $\boldsymbol{\mu}$ is the $n \times 1$ vector of arbitrary finite means, μ_i, and $\mathbf{B}$ is the $n \times n$ matrix whose diagonal elements are unity and whose off-diagonal (i,j) element is $-\beta_{i,j}$. Clearly $\mathbf{B}$ is symmetric but of course we also require $\mathbf{B}$ to be positive definite in order for the formulation to be valid.

At this point, it is perhaps worth indicating the distinction between the process (4.9) defined above, for which

$$E(X_i | \text{ all other site values}) = \mu_i + \sum \beta_{i,j}(x_j - \mu_j), \tag{4.11}$$

and the process defined by the set of n *simultaneous* autoregressive equations, typically

$$X_i = \mu_i + \sum \beta_{i,j}(X_j - \mu_j) + \varepsilon_i, \tag{4.12}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Gaussian variates, each with zero mean and variance σ^2. In contrast to equation (4.10), the latter process has joint probability density function,

$$P(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{1}{2}n} |\mathbf{B}| \exp\{-\tfrac{1}{2}\sigma^{-2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \mathbf{B}(\mathbf{x}-\boldsymbol{\mu})\}, \tag{4.13}$$

where $\mathbf{B}$ is defined as before. Also, it is no longer necessary that $\beta_{j,i} \equiv \beta_{i,j}$, only that $\mathbf{B}$ should be non-singular. Further aspects of simultaneous autoregressive schemes will be discussed in Sections 5.4 and 5.5.

4.2.3. *Auto-binomial schemes*

Suppose that X_i has a conditional binomial distribution with associated "sample size" m_i and "probability of success" θ_i which is dependent upon the neighbouring site values. Then $H_i(x_i) \equiv 1$ and, under Assumption 1, the odds of "success" to "failure" must satisfy

$$\ln\{\theta_i/(1-\theta_i)\} = \alpha_i + \sum \beta_{i,j} x_j.$$

When $m_i = 1$ for all i, we again have the auto-logistic model.

4.2.4. *Auto-Poisson schemes*

Suppose that X_i has a conditional Poisson distribution with mean μ_i dependent upon the neighbouring site values. Again $H_i(x_i) \equiv 1$ and, under Assumption 1, μ_i is subject to the form

$$\mu_i = \exp(\alpha_i + \sum \beta_{i,j} x_j).$$

Further, since the range of X_i is infinite, we must ensure that $\exp Q(\mathbf{x})$ is summable over $\mathbf{x}$. We show below that this requires the further restriction $\beta_{i,j} \leqslant 0$ for all i and j.

We have

$$Q(\mathbf{x}) = \sum\{\alpha_i x_i - \ln(x!)\} + \sum\sum \beta_{i,j} x_i x_j.$$

Clearly $\exp Q(\mathbf{x})$ must be summable when each $\beta_{i,j} = 0$ so the same holds when each $\beta_{i,j} \leqslant 0$. To show the necessity of the condition, we consider the distribution of the pair of variates (X_1, X_2) given that all other site values are equal to zero. The odds of the realization (x_1, x_2) to the realization $(0, 0)$ are then

$$\exp Q(x_1, x_2, 0, \ldots, 0) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \beta_{1,2} x_1 x_2)/(x_1! x_2!),$$

for non-negative integers x_1 and x_2. We certainly require that the sum of this quantity over all x_1 and x_2 converges and this is only true when $\beta_{1,2} \leqslant 0$. Similarly, we require $\beta_{i,j} \leqslant 0$ for all i and j. This restriction is severe and necessarily implies a "competitive" rather than "co-operative" interaction between auto-Poisson variates.

4.2.5. *Auto-exponential schemes*

Suppose that X_i has a conditional negative exponential distribution with mean μ_i dependent upon the values at sites neighbouring site i. Once more $H_i(x_i) \equiv 1$ and, under Assumption 1, μ_i must take the form $(\alpha_i + \sum \beta_{i,j} x_j)^{-1}$. The scheme is valid provided $\alpha_i > 0$ and $\beta_{i,j} \geqslant 0$ but the conditional probability structure appears to lack any form of intuitive appeal. Analogous statements hold for all gamma-type distributions.

4.3. *Proof of Equation* (4.4)

In order to establish the result (4.4) under Conditions 1 and 2, we begin by assuming that $\ln p_i(0; \ldots)$ is well behaved, relaxing this condition later. For convenience, we shall write A_i and D_i of equation (4.3) as functions of

$$(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$$

although in reality they depend only upon the values at sites neighbouring site i. Since $\ln p_i(0; \ldots)$ is well behaved, $Q(\mathbf{x})$ is well defined (under the positivity condition) and has the representation (4.2) according to Assumption 1. Equations (4.2) and

(4.3) may now be related through equation (3.2). Putting $x_j = 0$ for all $j \neq i$, we obtain, for each i,

$$x_i G_i(x_i) = A_i(\mathbf{0})\{B_i(x_i) - B_i(0)\} + C_i(x_i) - C_i(0). \tag{4.14}$$

Now suppose sites 1 and 2 are neighbours of each other. Putting $x_j = 0$ for $j \geqslant 3$ and again using equation (3.2) to link (4.2) and (4.3), we obtain, for $i = 1$,

$$x_1 G_1(x_1) + x_1 x_2 G_{1,2}(x_1, x_2) = A_1(0, x_2, 0, \ldots, 0)\{B_1(x_1) - B_1(0)\} + C_1(x_1) - C_1(0)$$

and, for $i = 2$,

$$x_2 G_2(x_2) + x_1 x_2 G_{1,2}(x_1, x_2) = A_2(x_1, 0, \ldots, 0)\{B_2(x_2) - B_2(0)\} + C_2(x_2) - C_2(0).$$

Combining these two equations with (4.14), we deduce that

$$x_1 x_2 G_{1,2}(x_1, x_2) = \beta_{1,2}\{B_1(x_1) - B_1(0)\}\{B_2(x_2) - B_2(0)\},$$

where $\beta_{1,2}$ is a constant. More generally, if sites i and j are neighbours and $i < j$,

$$x_i x_j G_{i,j}(x_i, x_j) = \beta_{i,j}\{B_i(x_i) - B_i(0)\}\{B_j(x_j) - B_j(0)\}. \tag{4.15}$$

The result (4.4) is easily deduced from (4.14) and (4.15).

The condition that $\ln p_i(0; \ldots)$ is well behaved is not satisfied by all members of the exponential family. However, in cases where $\ln p_i(0; \ldots)$ degenerates as, for example, with most gamma distributions, we may use a simple transformation on the X_i's to affirm that (4.4) still holds. Suppose, without loss of generality, that $0 < p_i(1; \ldots) < \infty$ and in that case let $Y_i = \ln X_i$ at each site. Then the conditional probability structure of the process $\{Y_i\}$ also lies within the exponential family of distributions but there is no degeneracy associated with the value $Y_i = 0$. The previous arguments may then be applied to show that A_i still satisfies equation (4.4).

5. Some Two-dimensional Spatial Schemes and Their Applications

5.1. *Finite Lattice Schemes*

In practice, the construction of conditional probability models on a finite regular lattice is simplified by the existence of a fairly natural hierarchy in the choice of neighbours for each site. For simplicity, and because it occurs most frequently in practice, we shall primarily discuss the rectangular lattice with sites defined by integer pairs (i,j) over some finite region. Where the notation becomes a little unwieldy, the reader may find it helpful to sketch and label the sites appropriately. The simplest model which allows for local stochastic interaction between the variates $X_{i,j}$ is then the *first-order* Markov scheme (or "nearest-neighbour" model) in which each interior site (i,j) is deemed to have four neighbours, namely $(i-1,j)$, $(i+1,j)$, $(i,j-1)$ and $(i,j+1)$. If, as suggested in Section 3, we now interpret $Q(\mathbf{x})$ as being concerned with the distribution of the internal site variables conditional upon given boundary values, the representation (3.3) in the Hammersley–Clifford theorem can be written

$$\begin{aligned} Q(\mathbf{x}) = \sum x_{i,j}\phi_{i,j}(x_{i,j}) + \sum x_{i,j}x_{i+1,j}\psi_{1,i,j}(x_{i,j}, x_{i+1,j}) \\ + \sum x_{i,j}x_{i,j+1}\psi_{2,i,j}(x_{i,j}, x_{i,j+1}), \end{aligned} \tag{5.1}$$

where $\{\phi_{i,j}\}$, $\{\psi_{1,i,j}\}$ and $\{\psi_{2,i,j}\}$ are arbitrary sets of functions, subject to the summability of $Q(\mathbf{x})$, and the ranges of summation in (5.1) are such that each clique,

involving at least one site internal to the system, contributes a single term to the representation. Writing (x, t, t', u, u') for the partial realization

$$(x_{i,j}, x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}),$$

the conditional probability structure at the site (i,j) is given by

$$p_{i,j}(x;\, t,t',u,u') = \exp\{f_{i,j}(x;\, t,t',u,u')\}/\sum \exp\{f_{i,j}(z;\, t,t',u,u')\}, \qquad (5.2)$$

where

$$f_{i,j}(.) = x\{\phi_{i,j}(x) + t\psi_{1,i-1,j}(t,x) + t'\psi_{1,i,j}(x,t') + u\psi_{2,i,j-1}(u,x) + u'\psi_{2,i,j}(x,u')\}$$

and the summation, or integration in the case of continuous variates, extends over all values z, possible at (i,j). In any given practical situation, the ϕ-, ψ_1- and ψ_2-functions can then be chosen to give an appropriate distributional form for $p_{i,j}(.)$. For the scheme to be spatially homogeneous, these functions must be independent of position (i,j) on the lattice. We then have the special case discussed by Besag (1972a). If, further, $\psi_1 = \psi_2$, the scheme is said to be isotropic.

The idea of a first-order scheme may easily be extended to produce higher-order schemes. Thus a *second-order* scheme allows (i,j) to have the additional neighbours $(i-1,j-1)$, $(i+1,j+1)$, $(i-1,j+1)$ and $(i+1,j-1)$, whilst a *third-order* scheme further includes the sites $(i-2,j)$, $(i+2,j)$, $(i,j-2)$ and $(i,j+2)$. To obtain $Q(\mathbf{x})$, we merely add a contributory term for each clique which involves at least one site internal to the system. For example, a homogeneous second-order scheme has

$$\begin{aligned} Q(\mathbf{x}) = {} & \sum x_{i,j}\phi(.) + \sum x_{i,j}x_{i+1,j}\psi_1(.) + \sum x_{i,j}x_{i,j+1}\psi_2(.) \\ & + \sum x_{i,j}x_{i+1,j+1}\psi_3(.) + \sum x_{i,j}x_{i+1,j-1}\psi_4(.) \\ & + \sum x_{i,j}x_{i+1,j}x_{i,j+1}\xi_1(.) + \sum x_{i,j}x_{i+1,j}x_{i+1,j+1}\xi_2(.) \\ & + \sum x_{i,j}x_{i+1,j}x_{i+1,j-1}\xi_3(.) + \sum x_{i,j}x_{i,j+1}x_{i+1,j+1}\xi_4(.) \\ & + \sum x_{i,j}x_{i+1,j}x_{i,j+1}x_{i+1,j+1}\delta(.), \end{aligned}$$

where the arguments of each function are its individual multipliers; thus

$$\delta(.) \equiv \delta(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1})$$

and so on. In specific examples, the apparent complexity of the expressions may be very much reduced. However, it is felt that, unless the variables are Gaussian, third- and higher-order schemes will almost always be too unwieldy to be of much practical use.

First- and second-order schemes may easily be constructed for other lattice systems in two or more dimensions. Amongst these, the plane triangle lattice is of particular interest, firstly because it frequently occurs in practice and secondly because a first-order scheme on a triangular lattice, for which each internal site has six neighbours, is likely to be more realistic than the corresponding scheme on a rectangular lattice.

5.2. *Specific Finite Lattice Schemes*

5.2.1. *Binary data*

It is clear from equation (5.1) that the homogeneous first-order scheme for zero-one variables on a rectangular lattice is given by

$$Q(\mathbf{x}) = \alpha \sum x_{i,j} + \beta_1 \sum x_{i,j}x_{i+1,j} + \beta_2 \sum x_{i,j}x_{i,j+1},$$

where α, β_1 and β_2 are arbitrary parameters. This leads to the conditional probability structure

$$p_{i,j}(x;\, t, t', u, u') = \frac{\exp\left[x\{\alpha+\beta_1(t+t')+\beta_2(u+u')\}\right]}{1+\exp\{\alpha+\beta_1(t+t')+\beta_2(u+u')\}}, \tag{5.3}$$

in the notation of Section 5.1. The scheme is necessarily auto-logistic.

For the second-order scheme, there are cliques of sizes three and four and there is no longer any need for the scheme to be auto-logistic. Thus, if we additionally write (v, v', w, w') for the partial realization $(x_{i-1,j-1}, x_{i+1,j+1}, x_{i-1,j+1}, x_{i+1,j-1})$, we find that $p_{i,j}(.)$ is now given by an expression similar to (5.3) but with the terms in curly brackets $\{\}$ replaced by

$$\begin{aligned}\alpha+\beta_1(t+t')+\beta_2(u+u')+\gamma_1(v+v')+\gamma_2(w+w') \\ +\xi_1(tu+u'w+w't')+\xi_2(tv+v'u'+ut')+\xi_3(tw+w'u+u't') \\ +\xi_4(tu'+uv+v't')+\eta(tuv+t'u'v'+tu'w+t'uw').\end{aligned} \tag{5.4}$$

The scheme is only auto-logistic if the ξ- and η-parameters are all zero.

Incidentally, this is a convenient point at which to mention the first-order binary scheme on a triangular lattice, for this can be thought of as a scheme on a rectangular lattice in which (i,j) has the *six* neighbours $(i-1,j)$, $(i+1,j)$, $(i,j-1)$, $(i,j+1)$, $(i-1,j-1)$ and $(i+1,j+1)$. The homogeneous first-order scheme is thus obtained from (5.4) by putting $\gamma_2 = \xi_1 = \xi_3 = \eta = 0$. The scheme is auto-logistic only if, in addition, $\xi_2 = \xi_4 = 0$.

Regarding applications of the rectangular lattice models, we shall, in Section 7.1, analyse Gleaves's *Plantago lanceolata* data using the first- and second-order isotropic auto-logistic schemes. However, none of the sets of data, cited in Section 1, appears to provide a convincing demonstration of low-order auto-logistic behaviour. It is hoped that more "appropriate" sets of data will become available in the future. A number of remarks are made in this context. Firstly, in order to carry out a detailed statistical analysis of spatial interaction, rather than merely test for independence or estimate the parameters of a model, it is usually the case that fairly extensive data are required. For example, spatial models have been fitted to Greig-Smith's data by Bartlett (1971b), using the spectral approximation technique of Bartlett and Besag (1969), and by Besag (1972c), using the coding technique of Section 6. The respective models are similar, though not equivalent, and each appears to give a fairly satisfactory fit. However, the last statement should be viewed with some scepticism since the goodness-of-fit tests available for such a small system (24×24) are very weak. This will be illustrated by the more detailed analysis of Gleaves's data.

Secondly, it is stressed that the lower-order homogeneous schemes, under discussion here, have been specifically designed with *local* stochastic interaction in mind; in particular, it is unreasonable to apply them in situations where there is evidence of gross heterogeneity over the lattice. For example, the hop plant data of Freeman (1953) display a fairly clear dichotomy between the upper and lower halves of the lattice, the former being relatively disease free (Bartlett, 1974). Thirdly, the use of lattice schemes on Greig-Smith's and Gleaves's data is, of course, an artifice: as remarked in Section 1, these examples are really concerned with spatial pattern rather than spatial interaction. Furthermore, as is well known, the size of quadrat used when collecting such data can profoundly influence the results of the subsequent

statistical analysis. Incidentally, from a numerical viewpoint, it is most efficient to arrange the quadrat size so that 0's and 1's occur with approximately equal frequency. An alternative procedure might be to adopt some sort of nested analysis (Greig-Smith, 1964).

The criticisms above are not intended to paint a particularly gloomy picture but merely to point out some limitations of the models. It is maintained that auto-logistic analyses can be useful in practice; the models, having once been established, are easy to interpret and, even when rejected, can aid an understanding of the data and of the underlying spatial situation.

5.2.2. *Gaussian variables*

It has already been stated in Section 2.2 that auto-normal schemes are of relevance to many ecological situations. For a finite rectangular lattice system, two homogeneous schemes are of particular practical interest. They are the first-order scheme for which $X_{i,j}$, given all other site values, is normally distributed with mean

$$\alpha+\beta_1(x_{i-1,j}+x_{i+1,j})+\beta_2(x_{i,j-1}+x_{i,j+1}) \tag{5.5}$$

and constant variance σ^2 and the second-order scheme for which $X_{i,j}$, given all other site values, is normally distributed with mean

$$\begin{aligned}&\alpha+\beta_1(x_{i-1,j}+x_{i+1,j})+\beta_2(x_{i,j-1}+x_{i,j+1})\\&+\gamma_1(x_{i-1,j-1}+x_{i+1,j+1})+\gamma_2(x_{i-1,j+1}+x_{i+1,j-1})\end{aligned} \tag{5.6}$$

and constant variance, σ^2. Such schemes can, for example, be used for the analysis of crop yields in uniformity trials when, perhaps through local fluctuations in soil fertility or the influence of competition, it is no longer reasonable to assume statistical independence. This is illustrated in Section 7, using the classical wheat plots data of Mercer and Hall (1911).

In more general experimental situations, it is possible to set up inhomogeneous auto-normal schemes to account for stochastic interaction between the variables. For example, one can replace α in the expressions (5.5) and (5,6) by $\alpha_{i,j}$, allowing this to depend deterministically upon the treatment combination at (i,j), in the usual way. Such schemes can still be analysed by the coding methods which will be discussed in Section 6. It is suggested that there is a need for further research here, particularly into the use of specially constructed experimental designs which take advantage of both the model and the coding analysis.

At this point, it is perhaps worth while anticipating the results of Section 5.4 in order to re-emphasize the distinction between the present approach and that based upon *simultaneous* autoregressive schemes. Removing means for simplicity, suppose we consider the scheme defined by the equations,

$$X_{i,j}=\beta_1 X_{i-1,j}+\beta_1' X_{i+1,j}+\beta_2 X_{i,j-1}+\beta_2' X_{i,j+1}+\varepsilon_{i,j} \tag{5.7}$$

over some finite region, with appropriate adjustments at the boundary of the system, where the $\varepsilon_{i,j}$'s are independent Gaussian error variates with common variance. The analogous scheme on a finite triangular lattice has been examined by Mead (1967). It might well be assumed that, at least when $\beta_1'=\beta_1$ and $\beta_2'=\beta_2$, the conditional expectation structure of the process (5.7) would tally with the expression (5.5), putting $\alpha=0$. However, this is not at all the case: in fact, the process (5.7) has

conditional expectation structure defined by

$$\begin{aligned}(1+\beta_1^2+\beta_1'^2+\beta_2^2+\beta_2'^2)\,E(X_{i,j}|\text{ all other site values})\\ =(\beta_1+\beta_1')(x_{i-1,j}+x_{i+1,j})+(\beta_2+\beta_2')(x_{i,j-1}+x_{i,j+1})\\ -(\beta_1\beta_2'+\beta_1'\beta_2)(x_{i-1,j-1}+x_{i+1,j+1})-(\beta_1\beta_2+\beta_1'\beta_2')(x_{i-1,j+1}+x_{i+1,j-1})\\ -\beta_1\beta_1'(x_{i-2,j}+x_{i+2,j})-\beta_2\beta_2'(x_{i,j-2}+x_{i,j+2}),\end{aligned} \tag{5.8}$$

consistent with a special case (since there are only four independent β-parameters rather than six) in the class of *third-order* auto-normal schemes. The peculiar conditional expectation structure arises because of the bilateral nature of the auto-regression; that is, in contrast with the unilateral time series situation, $\varepsilon_{i,j}$ is *not* independent of the remaining right-hand-side variables in (5.7). Some previous comments concerning the conditional probability structure of simultaneously defined schemes have been made by Bartlett (1971b), Besag (1972a) and Moran (1973a, b).

5.3. *Non-lattice Systems*

We now turn to the construction of models for which there are a finite number of irregularly distributed, but co-planar, sites. As stated in Section 1, we shall only be concerned here with the distribution of the site variables X_i $(i = 1, \ldots, n)$, given the knowledge of their respective locations, and not with an investigation of the spatial pattern associated with the sites themselves. The first problem is in the choice of neighbours for each site. If the sites comprise a finite system of closed irregular regions in the form of a mosaic, such as counties or states in a country, it will usually be natural to include as neighbours of a given site i, those sites to which it is adjacent. In addition, it may be felt necessary to include more remote sites whose influence is, nevertheless, felt to be of direct consequence to the site i variable.

Alternatively, if the sites constitute a finite set of irregularly distributed points in the plane, a rather more arbitrary criterion of neighbourhood must be adopted. However, the situation can be reduced to the preceding one if we can find an intuitively plausible method of defining appropriate territories for each site. One possibility is to construct the Voronyi polygons (or Dirichlet cells) for the system. The polygon of site i is defined by the union of those points in the plane which lie nearer to site i than to any other site. This formulation clearly produces a unique set of non-overlapping convex territories, often capable of a crude physical or biological interpretation. It appears to have been first used in practice by Brown (1965) in a study of local stand density in pine forests. Brown interpreted the polygon of any particular tree as defining the "area potentially available" to it. If, in general, two sites are deemed to be neighbours only when their polygons are adjacent, it is evident that each internal site must have at least three neighbours and that cliques of more than four sites cannot occur. With this definition, Brown's pine trees each have approximately six neighbours, as might commonly be expected in situations where competitive influences tend to produce a naturally or artificially imposed regularity on the pattern of sites. A slight, but artificial, reduction in complexity occurs if we further stipulate that in order for two sites to be neighbours, the line joining them must pass through their common polygon side. Cliques can then contain no more than three members.

Mead (1971) and Ord (1974) have each used the Voronyi polygons of a system to set up and examine simultaneous autoregressive schemes such as (4.12).

Whatever the eventual choice of the neighbourhood criterion, we may derive the most general form for the available conditional probability structure in any particular situation by applying the Hammersley–Clifford theorem. Some specific schemes have been given in Section 4.2. In particular, we discuss the use of the auto-normal scheme (4.9). The first task is to reduce the dimensionality of the parameter space by relating the μ_i's and $\beta_{i,j}$'s in some intuitively reasonable way. In the case of point sites, suppose the Voronyi polygons are constructed and that $d_{i,j}$ represents the distance between neighbouring sites i and j whilst $l_{i,j}$ represents the length of their common polygon side. It is then often feasible to relate each μ_i and non-zero $\beta_{i,j}$ to the corresponding $d_{i,j}$ and $l_{i,j}$. The symmetry property of the $\beta_{i,j}$'s arises naturally. Specific suggestions for use in the scheme (4.12) have been made by Mead (1971) in the context of plant competition models. Analogous suggestions are made by Ord (1974) in a geographical context. These suggestions could equally be implemented in the case of conditional probability models.

5.4. *Simultaneous Autoregressive Schemes*

At various stages, reference has been made to the simultaneous autoregressive schemes (4.12) and (5.7). We now determine their associated conditional probability structure since this is a facet of the models which has occasionally been misunderstood in the past. In fact, it is convenient to widen the formulation somewhat by considering schemes of the form

$$\sum b_{i,j} X_j = Z_i \tag{5.9}$$

for $i = 1, \ldots, n$, or, in matrix notation, $\mathbf{BX} = \mathbf{Z}$, where $\mathbf{B}$ is an $n \times n$ non-singular matrix and $\mathbf{Z}$ is a vector of independent continuous random variables. In practice, the matrix $\mathbf{B}$ will often be fairly sparse. We neither demand that the Z_i's are identically distributed nor that they are Gaussian. Let $f_i(.)$ denote the density function of Z_i. Then $X_1, \ldots, X_n$ have joint density,

$$P(\mathbf{x}) = \|\mathbf{B}\| f_1(\mathbf{b}_1^{\mathrm{T}}\mathbf{x}) f_2(\mathbf{b}_2^{\mathrm{T}}\mathbf{x}) \ldots f_n(\mathbf{b}_n^{\mathrm{T}}\mathbf{x}),$$

where $\mathbf{b}_i^{\mathrm{T}}$ denotes the ith row of $\mathbf{B}$. The conditional probability structure at site i is then immediately obtainable from equation (3.2) or an analogue thereof. In particular, the result (5.8) is easily deduced.

More generally, suppose we say that site $j \neq k$ is *acquainted* with site k if and only if, for some i, $b_{i,j} \neq 0$ and $b_{i,k} \neq 0$; that is, if and only if at least one of the equations (5.9) depends upon both X_j and X_k. Then it is easily seen that the conditional distribution of X_k can at most depend upon the values at sites acquainted with site k. That is, the neighbours of any site are included in its acquaintances.

In a given practical situation, the sets of acquaintances and neighbours of a site may well be identical but this is not necessarily so. Suppose, for example, we consider the process

$$X_{i,j} = \beta_1 X_{i-1,j} + \beta_2 X_{i,j-1} + \beta_3 X_{i-1,j-1} + \varepsilon_{i,j}, \tag{5.10}$$

defined, for convenience, over a $p \times q$ finite rectangular torus lattice, where the $\varepsilon_{i,j}$'s are independent Gaussian variables with zero means and common variances. Then (i,j) has acquaintances $(i-1,j)$, $(i+1,j)$, $(i,j-1)$, $(i,j+1)$, $(i-1,j+1)$, $(i+1,j-1)$, $(i-1,j+2)$ and $(i+1,j-2)$ provided β_1, β_2 and β_3 are non-zero. In general, these sites will also constitute the set of neighbours of (i,j). However, suppose $\beta_3 = \beta_1\beta_2$; then the sites $(i-1,j+1)$ and $(i+1,j-1)$ are no longer neighbours of (i,j). In fact,

we shall find in Section 6 that this result provides a useful approach to problems of statistical inference, for it enables unilateral approximations to first-order auto-normal schemes to be constructed.

5.5. *Stationary Auto-normal Processes on an Infinite Lattice*

We define a stationary Gaussian process $\{X_{i,j}: i,j = 0, \pm 1, \ldots\}$ to be a finite-order auto-normal process on the infinite rectangular lattice if it has autocovariance generating function (a.c.g.f.) equal to

$$K(1-\sum\sum b_{k,l} z_1^k z_2^l)^{-1}, \tag{5.11}$$

where (i) only a finite number of the real coefficients $b_{k,l}$ are non-zero, (ii) $b_{0,0} = 0$, (iii) $b_{-k,-l} \equiv b_{k,l}$ and (iv) $\sum\sum b_{k,l} z_1^k z_2^l < 1$ whenever $|z_1| = |z_2| = 1$. K is a constant and is related to the variance of $X_{i,j}$. The ranges of summation are to be taken as $-\infty$ to $+\infty$ unless otherwise stated.

The existence of such processes was demonstrated by Rosanov (1967) and in certain special cases by Moran (1973a, b). Moran (1973b) included a simplified account of some of Rosanov's paper and we shall use this below to discuss the structure of the schemes. Firstly, however, we reintroduce the concept of neighbourhood. That is, for a stationary Gaussian process with a.c.g.f. of the form (5.11), we define the site $(i-k, j-l)$ to be a neighbour of (i,j) if and only if $b_{k,l} \neq 0$. We now show that this accords with our finite system definition.

It follows from equation (5.11) that, provided $|r|+|s|>0$,

$$\rho_{r,s} = \sum\sum b_{k,l} \rho_{r-k,s-l}, \tag{5.12}$$

where $\rho_{r,s}$ denotes the autocorrelation of lags r and s in i and j, respectively. Now let $\{\varepsilon_{i,j}: i,j = 0, \pm 1, \ldots\}$ be a doubly infinite set of variates defined by

$$X_{i,j} = \alpha + \sum\sum b_{k,l} X_{i-k,j-l} + \varepsilon_{i,j}, \tag{5.13}$$

where $\alpha = (1-\sum\sum b_{k,l})\mu$ and $\mu = E(X_{i,j})$. Then the $\varepsilon_{i,j}$'s are stationary Gaussian variables with zero means and common variances σ^2, say. Also the equations (5.12) imply that $\varepsilon_{i,j}$ and $X_{i',j'}$ are uncorrelated provided $|i-i'|+|j-j'|>0$. This result together with (5.13), implies the following: given the values at any finite set of sites which includes the neighbours of (i,j), $X_{i,j}$ has conditional mean $\alpha + \sum\sum b_{k,l} x_{i-k,j-l}$ and conditional variance σ^2 independent of the actual surrounding values.

Thus, we have confirmed that the present criterion of neighbourhood is consistent with that for finite systems and that the properties of stationary, infinite lattice, auto-normal schemes are in accordance with those of the homogeneous, finite lattice schemes. In particular, we may define first-, second- and higher-order schemes analogous to those appearing in Section 5.2.2.

Finally, we make some remarks concerning the infinite lattice schemes proposed by Whittle (1954). Removing means for simplicity, Whittle considered simultaneously defined stationary processes in the class

$$\sum\sum a_{k,l} X_{i-k,j-l} = Z_{i,j}, \tag{5.14}$$

where $\{Z_{i,j}: i,j = 0, \pm 1, \ldots\}$ is a doubly infinite set of independent Gaussian variates, each with zero mean and variance v. The scheme (5.14) has a.c.g.f.

$$v(\sum\sum a_{k,l} z_1^k z_2^l)^{-1} (\sum\sum a_{k,l} z_1^{-k} z_2^{-l})^{-1}. \tag{5.15}$$

If the number of non-zero coefficients $a_{k,l}$ is finite, we shall refer to (5.14) as being a "finite-order" Whittle scheme. It is clear from (5.15) that any such scheme has a finite-order auto-normal representation. The converse is in general untrue: for example, even the first-order auto-normal scheme does not have a finite-order simultaneous autoregressive representation unless β_1 or $\beta_2 = 0$. One is therefore led to pose the following question: when using finite-order schemes in the statistical analysis of spatial data, are there *a priori* reasons for restricting attention to the particular finite-order schemes generated by (5.14) or should the wider range of auto-normal models be considered? Note that when the number of sites is finite, there is, for Gaussian variates, a complete, but somewhat artificial, correspondence between the classes of simultaneous and conditional probability models.

A further point which is relevant, whether the number of sites is finite or infinite, is illustrated by the following example. The most general bilateral scheme used by Whittle in examining the wheat plot data of Mercer and Hall (1911) was the infinite lattice analogue of (5.7), namely

$$X_{i,j} = \beta_1 X_{i-1,j} + \beta_1' X_{i+1,j} + \beta_2 X_{i,j-1} + \beta_2' X_{i,j+1} + Z_{i,j}. \tag{5.16}$$

Firstly, this process again has the rather peculiar conditional expectation structure (5.8), but secondly, as noted by Whittle, there is an ambiguity in the identity of parameters. That is, if we interchange β_1 and β_1' and also β_2 and β_2', we obtain a process with the identical probability structure. For the scheme (4.12), the same holds true if we interchange $\mathbf{B}$ and $\mathbf{B}^{\mathrm{T}}$. This seems rather unsatisfactory. In the time series situation, the problem does not arise if one invokes the usual assumption that past influences future, not vice versa. With spatial schemes, the problem can be overcome if we are content to examine merely the conditional probability structure of the process, given by equation (5.8) in the present context. It is suggested that such considerations again support the use of the conditional probability approach to spatial systems. A further comment appears in Section 7 of the paper.

We note that first- and second-order stationary auto-normal schemes on the infinite lattice were first proposed by Lévy (1948) but that, as remarked by Moran (1973a), existence was assumed without formal justification. Moran himself concentrates almost exclusively on the first-order scheme.

6. Statistical Analysis of Lattice Systems

In this section, we propose some methods of parameter estimation and some goodness-of-fit tests applicable to spatial Markov schemes defined over a rectangular lattice. The methods may be extended to other regular lattice systems, notably the triangular lattice, and, in part, to some non-lattice situations. In practice, it would appear that, amongst lattice schemes, it is the ones of first and second order which are of most interest and it is these upon which we shall concentrate. It has already been established in Section 2 that, generally speaking, a direct approach to statistical inference through maximum likelihood is intractable because of the extremely awkward nature of the normalizing function. We therefore seek alternative techniques. The exceptional case occurs when the variates have an auto-normal structure, for which the normalizing function may often be evaluated numerically without too much effort, even in some non-lattice situations. Each of the methods will be illustrated in Section 7 of the paper.

6.1. *Coding Methods on the Rectangular Lattice*

We assume, in the notation of Section 4, that the conditional distributions, $p_{i,j}(.)$, are of a given functional form but collectively contain a number of unknown parameters whose values are to be estimated on the basis of a single realization, $\mathbf{x}$, of the system. Coding methods of parameter estimation were introduced by Besag (1972c), in the context of binary data, but they are equally available in more general situations.

In order to fit a first-order scheme, we begin by labelling the interior sites of the lattice, alternately × and ., as shown in Fig. 1. It is then immediately clear that,

. × . × . × . × . ×
× . × . × . × . × .
. × . × . × . × . ×
× . × . × . × . × .

FIG. 1. Coding pattern for a first-order scheme.

according to the first-order Markov assumption, the variables associated with the × sites, given the observed values at all other sites, are mutually independent. This results in the simple conditional likelihood,

$$\prod p_{i,j}(x_{i,j};\, x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}),$$

for the × site values, the product being taken over all × sites. Conditional maximum-likelihood estimates of the unknown parameters can then be obtained in the usual way. Alternative estimates may be obtained by maximizing the likelihood function for the . site values conditional upon the remainder (or, that is, using a unit shift in the coding pattern). The two procedures are likely to be highly dependent but, nevertheless, it is reasonable, in practice, to carry out both and then combine the results appropriately.

In order to estimate the parameters of a second-order scheme, we may code the internal sites as shown in Fig. 2. Again considering the joint distribution of the × site variables given the . site values, we may obtain conditional maximum-likelihood

. × . × . × . × . ×
. . . .
. × . × . × . × . ×
. . . .

FIG. 2. Coding pattern for a second-order scheme.

estimates of the parameters. By performing shifts of the entire coding framework over the lattice, four sets of estimates are available and these may then be combined appropriately.

Using the coding methods, we may easily construct likelihood-ratio tests to examine the goodness of fit of particular schemes. Here, we stress three points. Firstly, it is highly desirable that the wider class of schemes against which we test is one which has intuitive spatial appeal, otherwise the test is likely to be weak. This is, of course, an obvious comment but one which, in the limited statistical work on spatial analysis, has sometimes been neglected. Secondly, the two maximized likelihoods we obtain must be strictly comparable. For example, if the fit of a scheme of

first order is being examined against one of second order, the resulting likelihood-ratio test will only be valid if both the schemes have been fitted to the same set of data—that is, using the Fig. 2 coding in each case. Thirdly, there will be more than one test available (under shifts in coding) and these should be considered collectively. Whilst precise combination of the results may not be possible, they can usually be amalgamated in some conservative way. These points will be illustrated in Section 7.

The efficiency of coding techniques can to a limited extent be investigated following the methods of Ord (1974). Also of relevance are the papers by Ogawara (1951), Williams (1952) and Hannan (1955a, b) and some comments by Plackett (1960, p. 121), all on coding methods for Markov chains. The coding techniques will not, in general, be fully efficient but their great advantage lies in their simplicity and flexibility. Some results will be reported elsewhere but further investigation of the techniques is still required.

6.2. *Unilateral Approximations on the Rectangular Lattice*

An alternative estimation procedure for homogeneous first-order spatial schemes involves the construction of a simpler process which has approximately the required probability structure but which is much easier to handle. The approach is similar (equivalent for stationary auto-normal schemes) to that of Bartlett and Besag (1969). We begin by defining the set of *predecessors* of any site (i,j) in the positive quadrant to consist of those sites (k,l) on the lattice which satisfy either (i) $l<j$ or (ii) $l=j$ and $k<i$. We may then generate a *unilateral* stochastic process $\{X_{i,j}: i>0, j>0\}$ in the positive quadrant by specifying the distribution of each variable $X_{i,j}$ conditional upon the values at sites which are predecessors of (i,j). In practice, we shall allow the distribution of $X_{i,j}$ to depend only on a limited number of predecessor values. Such a process is a natural extension of a classical one-dimensional finite auto-regressive time series into two dimensions and is well defined if sufficient initial values are given. Special cases of such schemes have been discussed by Bartlett and Besag (1969), Bartlett (1971b) and Besag (1972b). By a judicious choice of the unilateral scheme, we may obtain a reasonable approximation to a given first-order spatial scheme. The more predecessor values we allow $X_{i,j}$ to depend upon, the better the approximation can be made. The great advantage of a unilateral scheme is that its likelihood function is easily written down and parameter estimation may be effected by straightforward maximum likelihood.

As the simplest general illustration, we consider unilateral processes of the form

$$P(x_{i,j}|\text{ all predecessors}) = q(x_{i,j};\, x_{i-1,j}, x_{i,j-1}).$$

The joint probability distribution of the variables $X_{i,j}$ $(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n)$ is given by

$$\prod_{i=1}^{m}\prod_{j=1}^{n} q(x_{i,j};\, x_{i-1,j}, x_{i,j-1})$$

and, hence, for any interior site (i,j) we have, in the notation of Section 4, the bilateral structure

$$\frac{p_{i,j}(x;\ldots)}{p_{i,j}(x^*;\ldots)} = \frac{q(x;\,t,u)q(t';\,x,w')q(u';\,w,x)}{q(x^*;\,t,u)q(t';\,x^*,w')q(u';\,w,x^*)}.$$

That is, the conditional distribution, $P(x_{i,j}|$ all other site values), depends not only upon $x_{i-1,j}$, $x_{i+1,j}$, $x_{i,j-1}$ and $x_{i,j+1}$ but also upon $x_{i-1,j+1}$ and $x_{i+1,j-1}$. Nevertheless,

the primary dependence is upon the former set of values and, by a suitable choice of $q(.)$, we may use the unilateral process as an approximation to a given homogeneous first-order spatial scheme. For a better approximation, we may consider unilateral processes of the form,

$$P(x_{i,j} | \text{ all predecessors}) = q(x_{i,j};\ x_{i-1,j}, x_{i,j-1}, x_{i+1,j-1})$$

and so on. The method will be illustrated for an auto-normal scheme in Section 7.

6.3. *Maximum-likelihood Estimation for Auto-normal Schemes*

We begin by considering the estimation of the parameters in an auto-normal scheme of the form (4.9) but subject to the restriction $\boldsymbol{\mu} = \mathbf{0}$. We assume that the dimensionality of the parameter space is reduced through $\mathbf{B}$ having a particular structure and that σ^2 is both unknown and independent of the $\beta_{i,j}$'s. For a given realization $\mathbf{x}$, the corresponding likelihood function is then equal to

$$(2\pi\sigma^2)^{-\frac{1}{2}n} |\mathbf{B}|^{\frac{1}{2}} \exp(-\tfrac{1}{2}\sigma^{-2}\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}). \tag{6.1}$$

It follows that the maximum-likelihood estimate of σ^2 will be given by

$$\hat{\sigma}^2 = n^{-1}\mathbf{x}\hat{\mathbf{B}}\mathbf{x}, \tag{6.2}$$

once $\hat{\mathbf{B}}$, the maximum-likelihood estimate of $\mathbf{B}$, has been found. Substituting (6.2) into (6.1), we find that $\hat{\mathbf{B}}$ may be obtained by minimizing

$$-n^{-1}\ln|\mathbf{B}| + \ln(\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}). \tag{6.3}$$

The problem of implementing maximum-likelihood estimation therefore rests upon the evaluation of the determinant, $|\mathbf{B}|$. We now examine how this relates to existing research into simultaneous autoregressions.

Suppose then that we temporarily abandon the auto-normal model above and decide instead to fit a simultaneous scheme of the form (4.12), again subject to $\boldsymbol{\mu} = \mathbf{0}$ and with $\mathbf{B}$ having the same structure as in (6.1). Provided (6.1) is valid so is the present, but different, scheme. The likelihood function now becomes

$$(2\pi\sigma^2)^{-\frac{1}{2}n} |\mathbf{B}| \exp(-\tfrac{1}{2}\sigma^{-2}\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{x}) \tag{6.4}$$

and the new estimate of $\mathbf{B}$ must be found by minimizing

$$-2n^{-1}\ln|\mathbf{B}| + \ln(\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{x}). \tag{6.5}$$

Again the only real difficulty centres upon the evaluation of the determinant $|\mathbf{B}|$, a point which we may, in a sense, now turn to advantage. Suppose that we wish to fit the auto-normal scheme associated with (6.1) to a given set of data. Then it follows that we may use existing approaches to fitting simultaneous autoregressive schemes provided that these can cope with the likelihood function (6.4). Indeed with minor modifications, we may use any existing computer programs. It is probably fair to say that thus far the simultaneous and conditional probability schools have tended to suggest the same structure for $\mathbf{B}$ in a given problem. This, together with the previous remarks, implies that it would be relatively straightforward to conduct a useful comparative investigation of the two approaches for some given sets of data.

As regards minimizing (6.5), computational progress has been made by Mead (1967) on small (triangular) lattices and by Ord (1974) in non-lattice situations where the number of sites is fairly limited (about 40 or less) and there are only one or two

unknown parameters determining **B**. The reader is referred to their papers for further details. As regards large lattices for which we may sometimes view the data as being a partial realization of a stationary Gaussian infinite lattice process, we may use the semi-analytical result of Whittle (1954) which is summarized below.

Whittle showed, for the simultaneous autoregression (5.14), that, given a partial realization of the process over n sites, the term $n^{-1}\ln|\mathbf{B}|$ in (6.5) can be approximated by the coefficient of $z_1^0 z_2^0$ in the power series expansion of

$$\ln(\sum\sum a_{k,l} z_1^k z_2^l).$$

There is no complication if the variates have equal, but non-zero, means. Thus, in order to fit a particular auto-normal scheme of the form (5.11) from a partial realization of the process over n sites, we need to minimize (6.3), where $n^{-1}\ln|\mathbf{B}|$ is the absolute term in the power series expansion of

$$\ln(1-\sum\sum b_{k,l} z_1^k z_2^l)$$

and where, neglecting boundary effects,

$$\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x} = C_{0,0} - \sum\sum b_{k,l} C_{k,l}$$

and $C_{k,l}$ denotes the empirical autocovariance of lags k and l in i and j, respectively (cf. Whittle, 1954).

For example, with the first-order scheme, analogous to (5.5), we minimize

$$-\Lambda(\boldsymbol{\beta}) + \ln(C_{0,0} - 2\beta_1 C_{1,0} - 2\beta_2 C_{0,1}),$$

where $\Lambda(\boldsymbol{\beta})$ is the absolute term in the power series expansion of

$$\ln\{1-\beta_1(z_1+z_1^{-1})-\beta_2(z_2+z_2^{-1})\}.$$

With the second-order scheme, analogous to (5.6), we minimize

$$-\Lambda(\boldsymbol{\beta},\boldsymbol{\gamma}) + \ln(C_{0,0} - 2\beta_1 C_{1,0} - 2\beta_2 C_{0,1} - 2\gamma_1 C_{1,1} - 2\gamma_2 C_{1,-1}),$$

where $\Lambda(\boldsymbol{\beta},\boldsymbol{\gamma})$ is the absolute term in the power series expansion of

$$\ln\{1-\beta_1(z_1+z_1^{-1})-\beta_2(z_2+z_2^{-1})-\gamma_1(z_1 z_2+z_1^{-1}z_2^{-1})-\gamma_2(z_1 z_2^{-1}+z_1^{-1}z_2)\}.$$

The absolute terms can easily be evaluated for given parameter values by appropriate numerical Fourier inversion. The expression (6.3) may be minimized by, for example, the Newton–Raphson technique. Convergence, in the limited work thus far, has been extremely rapid. A numerical example is included in Section 7.

Finally, we note an analogy between the fitting of stationary auto-normal schemes in the analysis of spatial data and the fitting of autoregressive schemes in classical time-series analysis. That is, considering a particular scheme in the class (5.11), suppose that the corresponding autocorrelations are denoted by $\rho_{k,l}$. Then the effect of large-sample maximum-likelihood estimation is to ensure perfect agreement between $\rho_{k,l}$ and the corresponding sample autocorrelation $r_{k,l}$ whenever $b_{k,l} \neq 0$ in the original formulation. Thus, for a first-order scheme, the fit ensures that $\rho_{1,0} = r_{1,0}$ and $\rho_{0,1} = r_{0,1}$. For the second-order scheme, we additionally fit $\rho_{1,1} = r_{1,1}$ and $\rho_{1,-1} = r_{1,-1}$. In general, there is no such interpretation for simultaneously defined autoregressions. This may suggest that the auto-normal schemes are, in fact, a more natural extension of classical temporal autoregressions to spatial situations.

7. NUMERICAL EXAMPLES

7.1. *Auto-logistic Analysis of* Plantago lanceolata *Data*

Observations on *Plantago lanceolata* were made over an apparently homogeneous area of lead–zinc tailings in defunct mine workings, Treloggan, Flintshire. The sampling frame consisted of a transect 10×940 with grid size 2 cm $\times$ 2 cm. Counts were made of the number of seedlings and number of adults in each of the 9,400 quadrats. As the grid size is rather small, the analysis below is based upon the pooled data and only presence/absence of plants in each quadrat is considered. The latter simplification results in little loss of information. The data were kindly collected by Dr J. T. Gleaves of the Department of Botany, University of Liverpool.

We let $x_{i,j} = 0/1$ denote absence/presence of *Plantago lanceolata* in the (i,j)th quadrat. There is no reason to expect asymmetry in the system and we shall therefore only consider auto-logistic schemes of the form

(a) *isotropic first-order scheme* for which

$$P(x_{i,j} | \text{ all other values}) = \frac{\exp\{(\alpha+\beta y_{i,j})\, x_{i,j}\}}{1+\exp(\alpha+\beta y_{i,j})},$$

where $y_{i,j} = x_{i-1,j}+x_{i+1,j}+x_{i,j-1}+x_{i,j+1}$, and

(b) *isotropic second-order scheme* for which

$$P(x_{i,j} | \text{ all other values}) = \frac{\exp\{(\alpha+\beta y_{i,j}+\gamma z_{i,j})\, x_{i,j}\}}{1+\exp(\alpha+\beta y_{i,j}+\gamma z_{i,j})},$$

where, in addition, $z_{i,j} = x_{i-1,j-1}+x_{i+1,j+1}+x_{i-1,j+1}+x_{i+1,j-1}$.

Note that a full isotropic second-order scheme would involve two further parameters, corresponding to cliques of triples and quadruples, respectively. A further comment appears later.

Parameter estimates can be obtained for schemes (a) and (b) using the coding techniques described in Section 6.1. For scheme (a), estimates for α and β, under Fig. 1 codings, are given in Table 1. The respective observed and expected frequencies

TABLE 1

Auto-logistic analysis of Plantago lanceolata *data: parameter estimates for scheme (a) under Fig. 1 codings*

	$\hat{\alpha}$	$\hat{\beta}$
First analysis	−2·254	0·724
Second analysis	−2·141	0·748
Approx. S.E.	0·07	0·04
Mean estimate	−2·198	0·736

appear in Tables 2 and 3, and these may be used to conduct simple chi-squared goodness-of-fit tests for the scheme. The resulting statistics, each on 3 degrees of freedom (d.fr.), are 4·84 and 2·90, respectively, suggesting a satisfactory fit. However, interpreting these tests as likelihood-ratio tests, we see that the wider hypothesis,

against which we are examining scheme (a), is itself rather specialized (and unattractive) since it still assumes independence between the columns within the body of the tables. Thus, the above type of test is not recommended although its use may be unavoidable where there is a shortage of data.

TABLE 2

Auto-logistic analysis of Plantago lanceolata *data: observed and expected frequencies for first analysis in Table* 1

$x_{i,j}$	$y_{i,j} = 0$	1	2	3	4	*Total*
0	1798 (1786·5)	847 (856·6)	329 (338·7)	101 (98·9)	25 (19·3)	3100
1	176 (187·5)	195 (185·4)	161 (151·3)	89 (91·1)	31 (36·7)	652
Total	1974	1042	490	190	56	3752

TABLE 3

Auto-logistic analysis of Plantago lanceolata *data: observed and expected frequencies for second analysis in Table* 1

$x_{i,j}$	$y_{i,j} = 0$	1	2	3	4	*Total*
0	1826 (1820·1)	838 (852·2)	293 (281·9)	83 (86·2)	13 (12·5)	3053
1	208 (213·9)	226 (211·8)	137 (148·1)	99 (95·8)	29 (29·5)	699
Total	2034	1064	430	182	42	3752

For scheme (b), the parameter estimates, under Fig. 2 codings, are given in Table 4. Scheme (a) may also be fitted under these codings and this is done in Table 5. Hence, we may examine the goodness of fit of scheme (a) within the class (b) in each of the four cases, using the usual likelihood-ratio test. The resulting statistics, each on 1 d.fr., are 49·9, 60·6, 49·4 and 48·6! It is now clear that scheme (a) is hopelessly inadequate in describing the system and that we must be wary of non-significant results in the type of test previously described. Incidentally, it is of interest to note the correspondence between Tables 1 and 5.

A typical set of observed and expected frequencies under scheme (b) is recorded in Table 6 and may be used to produce a simple chi-squared test for scheme (b) itself. The results corresponding to each of the four analyses are given in Table 7 and, combining these conservatively, leads to the rejection of scheme (b) at the 5 per cent level. Thus, despite the preceding remarks, it is still quite possible for the weaker form of test to be useful in practice.

It might be of interest to fit the full second-order or higher-order schemes to the data but, once fitted, a large number of cells would be found empty or nearly empty in the contingency tables, resulting in the invalidity of the usual distributional assumptions concerning goodness-of-fit tests. Furthermore, the objections (see Section 5.2.1) to Markov lattice models for quadrat schemes suggest that a more detailed analysis on the present lines is unlikely to be particularly helpful.

TABLE 4

Auto-logistic analysis of Plantago lanceolata *data: parameter estimates for scheme (b) under Fig. 2 codings*

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
First analysis	−2·534	0·518	0·497
Second analysis	−2·459	0·529	0·528
Third analysis	−2·369	0·543	0·456
Fourth analysis	−2·456	0·515	0·487
Approx. S.E.	0·10	0·07	0·07
Mean estimate	−2·455	0·526	0·492

TABLE 5

Auto-logistic analysis of Plantago lanceolata *data: parameter estimates for scheme (a) under Fig. 2 codings*

	$\hat{\alpha}$	$\hat{\beta}$
First analysis	−2·294	0·736
Second analysis	−2·149	0·744
Third analysis	−2·133	0·753
Fourth analysis	−2·215	0·712
Approx. S.E.	0·09	0·06
Mean estimate	−2·198	0·736

TABLE 6

Auto-logistic analysis of Plantago lanceolata *data: observed and expected frequencies for first analysis in Table* 4

$z_{i,j}$	$x_{i,j}$	$y_{i,j}=0$	1	2	3	4
0	0	628 (617·1)	207 (220·6)	61 (60·5)	6 (6·5)	1 (1·8)
	1	38 (48·9)	43 (29·4)	13 (13·5)	3 (2·5)	2 (1·2)
1	0	223 (218·5)	135 (141·1)	61 (56·3)	16 (14·8)	3 (3·4)
	1	24 (28·5)	37 (30·9)	16 (20·7)	8 (9·2)	4 (3·6)
2	0	49 (54·4)	60 (54·4)	36 (37·4)	18 (14·4)	4 (3·3)
	1	17 (11·6)	14 (19·6)	24 (22·6)	11 (14·6)	5 (5·7)
3	0	8 (8·9)	7 (6·9)	17 (19·6)	7 (9·4)	4 (1·8)
	1	4 (3·1)	4 (4·1)	22 (19·4)	18 (15·6)	3 (5·2)
4	0	0 (0·0)	0 (0·0)	4 (3·8)	1 (0·5)	0 (0·5)
	1	0 (0·0)	0 (0·0)	6 (6·2)	1 (1·5)	3 (2·5)

TABLE 7

Auto-logistic analysis of Plantago lanceolata *data: goodness-of-fit tests for scheme* (*b*)

Analysis	*Test statistic*	*D.fr.* (*approx.*)	*Significance*
First	28·3	20	almost 10%
Second	19·6	21	—
Third	38·7	21	almost 1%
Fourth	38·8	22	2%

7.2. *Auto-normal Analysis of Mercer and Hall Wheat Plots*

Mercer and Hall (1911) present the results of a uniformity trial concerning 500 wheat plots, each 11 ft × 10·82 ft, arranged as a 20 × 25 rectangular array. Two measurements, grain yield and straw yield, were made on each plot. Whittle (1954) analysed the grain yields, fitting various stationary normal autoregressions, as briefly described in Section 6.3 of the present paper. We shall analyse the same set of data but on the basis of the homogeneous first- and second-order schemes, (5.5) and (5.6).

7.2.1. *Coding methods*

In Tables 8 and 9 we record the parameter estimates for the schemes (5.5) and (5.6), respectively, using the coding techniques. The various analyses, within each table, refer to shifts in coding pattern, as previously described. Scheme (5.5) is also fitted under the Fig. 2 codings (Table 10) in order to test for the significance of the parameters γ_1 and γ_2. A typical analysis of variance is given in Table 11. Over the four coding shifts, the respective F ratios for the combined effect of γ_1 and γ_2 are 0·9

TABLE 8

Auto-normal analysis of wheat plots data: parameter estimates for scheme (5.5) *under Fig.* 1 *codings*

	$\hat{\beta}_1$	$\hat{\beta}_2$
First analysis	0·332	0·128
Second analysis	0·354	0·166
Approx. S.E.	0·03	0·03
Mean estimate	0·343	0·147

(2 and 103 d.fr.), 0·06 (2 and 94 d.fr.), 1.1 (2 and 103 d.fr.) and 1·2 (2 and 94 d.fr.). Each of these statistics suggests that the first-order scheme provides an adequate description of the system (incidentally, 0·06 is a very reasonable observation from the $F_{2,94}$ distribution). However, for further comments concerning the model and the parameter estimates, see Section 7.2.4.

TABLE 9

Auto-normal analysis of wheat plots data: parameter estimates for scheme (5.6) *under Fig.* 2 *codings*

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
First analysis	0·344	0·043	0·079	−0·062
Second analysis	0·318	0·085	0·016	0·011
Third analysis	0·407	0·243	−0·067	−0·034
Fourth analysis	0·361	0·236	−0·092	−0·041
Approx. S.E.	0·05	0·06	0·07	0·06
Mean estimate	0·358	0·152	−0·016	−0·032

TABLE 10

Auto-normal analysis of wheat plots data: parameter estimates for scheme (5.5) *under Fig.* 2 *codings*

	$\hat{\beta}_1$	$\hat{\beta}_2$
First analysis	0·348	0·052
Second analysis	0·321	0·104
Third analysis	0·393	0·199
Fourth analysis	0·340	0·168
Approx. S.E.	0·05	0·05
Mean estimate	0·350	0·131

TABLE 11

Auto-normal analysis of wheat plots data: first analysis of variance under Fig. 2 *codings*

Effect	*Sum of squares*	*D.fr.*	*Mean squares*	*F ratio*
β_1, β_2	9·63	2	4·81	45·5
γ_1, γ_2	0·19	2	0·10	0·9
Residual	10·89	103	0·106	
Total	20·71	107		

7.2.2. *Unilateral approximations to the first-order scheme*

If we treat the data as a partial realization from the stationary infinite lattice version of (5.5), we have an a.c.g.f. proportional to

$$\{1-\beta_1(z_1+z_1^{-1})-\beta_2(z_2+z_2^{-1})\}^{-1}. \tag{7.1}$$

As a first unilateral approximation, we may use the stationary autoregression,

$$X_{i,j} = b_1 X_{i-1,j} + b_2 X_{i,j-1} + Z_{i,j}, \tag{7.2}$$

where $\{Z_{i,j}: i,j = 0, \pm 1, \ldots\}$ is a doubly infinite set of independent Gaussian variates, each with zero mean and equal variance. The scheme (7.1) has a.c.g.f. proportional to

$$\{1+b_1^2+b_2^2-b_1(z_1+z_1^{-1})-b_2(z_2+z_2^{-1})+b_1b_2(z_1z_2^{-1}+z_1^{-1}z_2)\}^{-1},$$

which is clearly a first approximation to (7.1). Fitting the scheme (7.2) results in parameter estimates $\hat{b}_1 = 0{\cdot}488$ and $\hat{b}_2 = 0{\cdot}202$ and an estimated a.c.g.f. proportional to

$$\{1-0{\cdot}382(z_1+z_1^{-1})-0{\cdot}158(z_2+z_2^{-1})+0{\cdot}077(z_1z_2^{-1}+z_1^{-1}z_2)\}^{-1}.$$

The values 0·382 and 0·158 may therefore be interpreted as crude estimates of β_1 and β_2 in (7.1). However, these estimates are somewhat arbitrarily formed and little confidence should be placed in them. For example, there is no real reason why $\hat{b}_1$ and $\hat{b}_2$ themselves should not be used to estimate β_1 and β_2; the fact is that β_1 and β_2 are really too large for the first approximation to be of much use. Thus, we consider the second unilateral approximation.

The distribution of $X_{i,j}$, conditional upon all other values, in the scheme (7.2) involves an unwanted dependency upon $x_{i+1,j-1}$ and $x_{i-1,j+1}$. This may be removed by modifying the scheme to

$$X_{i,j} = b_1X_{i-1,j}+b_2X_{i,j-1}+b_1b_2X_{i+1,j-1}+Z_{i,j},$$

with a.c.g.f. proportional to

$$\{1+b_1^2+b_2^2+b_1^2b_2^2-b_1(1-b_2^2)(z_1+z_1^{-1})-b_2(z_2+z_2^{-1})+b_1^2b_2(z_1^2z_2^{-1}+z_1^{-2}z_2)\}^{-1}.$$

Fitting this scheme gives parameter estimates $\hat{b}_1 = 0{\cdot}483$ and $\hat{b}_2 = 0{\cdot}150$ and an estimated a.c.g.f. proportional to

$$\{1-0{\cdot}374(z_1+z_1^{-1})-0{\cdot}119(z_2+z_2^{-1})+0{\cdot}028(z_1^2z_2^{-1}+z_1^{-2}z_2)\}^{-1}.$$

The values 0·374 and 0·119 may therefore be interpreted as better estimates of β_1 and β_2 in (5.5). By continuing the procedure, next introducing $X_{i+2,j-1}$ and so on, a sequence of unilateral approximations to the scheme (5.5) may be generated (although there is probably little point in going further than the third one). These unilateral schemes also have the advantage that their analytical correlation structure is available in a region of the plane (Besag, 1972b). However, in more general situations, the approximation technique is likely to be rather cumbersome and is not particularly recommended.

Whittle (1954) also fits the unilateral scheme (7.2) to the Mercer and Hall data, but in its own right rather than as an approximation. Whittle notes that the fit is "surprisingly" good in comparison with some of his bilateral autoregressions. This might now be explained by interpreting the scheme as an approximation to an auto-normal scheme.

7.2.3. *Adaptation of Whittle's method*

Under the assumption of stationarity and neglecting edge effects, we may fit the schemes (5.5) and (5.6) as outlined in Section 6.3. Parameter estimation can be carried out iteratively using the Newton–Raphson technique, with the normalizing function, Λ, being evaluated by numerical integration at each stage. The results are given in

Table 12. The likelihood-ratio test for the scheme (5.5) within the class (5.6) gives a chi-squared statistic 2·69 on 2 d.fr. and again (see Section 7.2.1) the first-order auto-normal scheme appears satisfactory.

TABLE 12

Auto-normal analysis of wheat plots data: parameter estimates using Whittle's method for stationary schemes

Scheme	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
First order (5.5)	0·368	0·107	—	—
Second order (5.6)	0·381	0·160	−0·015	−0·056

7.2.4. *More realistic models*

Concerning the Mercer and Hall data, it was pointed out by Whittle (1954) that the simple simultaneous autoregression (5.16) does not really reflect the observed correlation structure. This is also true for the first- and second-order (stationary) auto-normal schemes. The disparity between the observed and fitted correlograms can easily be seen from Tables 13, 14 and 15. The entries in Table 13 have been

TABLE 13

Observed autocorrelations for the wheat plots data

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$l=-3$	0·1880	0·1602	0·1509	0·1276	0·1352
$l=-2$	0·1510	0·0234	0·0020	−0·0137	−0·1039
$l=-1$	0·2923	0·1853	0·1349	0·0788	0·0878
$l=0$	1·0000	0·5252	0·4055	0·3639	0·3561
$l=1$	0·2923	0·2354	0·1799	0·1205	0·1399
$l=2$	0·1510	0·1285	0·0999	0·0749	0·0859
$l=3$	0·1880	0·1935	0·2483	0·2415	0·2284

TABLE 14

Fitted autocorrelations for the first-order scheme in Table 12

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$l=-3$	0·0463	0·0441	0·0385	0·0316	0·0248
$l=-2$	0·1102	0·1007	0·0822	0·0628	0·0459
$l=-1$	0·2923	0·2364	0·1691	0·1151	0·0766
$l=0$	1·0000	0·5252	0·2913	0·1690	0·1015
$l=1$	0·2923	0·2364	0·1691	0·1151	0·0766
$l=2$	0·1102	0·1007	0·0822	0·0628	0·0459
$l=3$	0·0463	0·0441	0·0385	0·0316	0·0248

TABLE 15

Fitted autocorrelations for the second-order scheme in Table 12

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$l=-3$	0·0333	0·0254	0·0184	0·0129	0·0088
$l=-2$	0·0950	0·0676	0·0462	0·0306	0·0200
$l=-1$	0·2923	0·1853	0·1153	0·0710	0·0434
$l=0$	1·0000	0·5252	0·2818	0·1542	0·0859
$l=1$	0·2923	0·2354	0·1659	0·1097	0·0701
$l=2$	0·0950	0·0926	0·0771	0·0587	0·0421
$l=3$	0·0333	0·0354	0·0328	0·0276	0·0218

copied from Whittle's paper. Whittle gives a number of possible explanations of the observed correlogram behaviour. Amongst these is the fact that the data are integrated observations of growth over plots rather than point observations. As with quadrat counts, this renders a simple Markov assumption somewhat dubious. A further suggestion (Patankar, 1954) is that the process is non-stationary, but we leave this for the moment and treat the observed correlogram at face value. The question then is: how can we reproduce a fitted correlation structure which tallies with Table 13, without the associated scheme becoming too artificial? The answer may well lie in the use of a third-order auto-normal scheme, for then $r_{2,0}$ and $r_{0,2}$ can be fitted exactly and it is here that the trouble really seems to be. Note that, with the inclusion of third-order terms, the second-order terms may now have a significant role to play. The results of fitting the third-order scheme will be reported in due course.

In the *Biometrika* paper immediately following Whittle's, Patankar (1954) also examined some spatial aspects of the Mercer and Hall data but, notably, only after removing a significant linear trend running from West to East. Thus, in Table 16, we give a typical auto-normal analysis of variance, constructed by the coding method

TABLE 16

Auto-normal analysis of wheat plots with the inclusion of a linear trend term: first analysis of variance under Fig. 2 *codings*

Effect	*Sum of squares*	*D.fr.*	*Mean squares*	*F ratio*
Trend (τ)	2·03	1	2·03	19·0
β (τ removed)	7·61	2	3·81	35·6
γ (τ, β removed)	0·19	2	0·09	0·8
Residual	10·88	102	0·107	
Total	20·71	107		

but including trend removal. The estimates of the parameters are only slightly changed, as might be expected in such a situation, and the overall conclusions concerning influence of diagonally nearest plots remains unchanged.

Thirdly, we note a disconcerting property of some of the coding fits. That is, on a number of occasions, they individually give parameter estimates which are inconsistent with a stationary auto-normal scheme. For example, with the first-order

scheme (5.5), the second analysis in Table 8 produces estimates of β_1 and β_2 whose sum exceeds 0·5. Similar inconsistencies occur even after the removal of trend. Whilst it is so happens that, for each model fitted, the *mean* estimates are feasible, the individual values again suggest that the models are not entirely appropriate.

Summarizing, it cannot be claimed that the present auto-normal schemes have been successful in reflecting the overall probabilistic structure of the wheat plots process. Further analysis is required, although it is felt to be perhaps more important to examine a range of examples rather than to concentrate in too much detail upon a single set of classical data.

8. Concluding Remarks

In the preceding sections, an attempt has been made to establish that a conditional probability approach to spatial processes is not only feasible but is also desirable. It has been suggested, firstly, that the conditional probability approach has greater intuitive appeal to the practising statistician than the alternative joint probability approach; secondly, that the existence of the Hammersley–Clifford theorem has almost entirely removed any consistency problems and, further, can easily be used as a tool for the construction of conditional probability models in many situations; thirdly, that the basic lattice models under the conditional probability approach yield naturally to a very simple parameter estimation procedure (the coding technique) and, at least for binary and Gaussian variates, to straightforward goodness-of-fit tests. For Gaussian variates, maximum likelihood appears equally available for both simultaneous and conditional probability models of similar complexity. As regards the joint probability approach, it is not clear to the present author how, outside the Gaussian situation, the models are to be used in practice. How, for example, would Gleaves's binary data be analysed?

On the other hand, the two examples discussed in Section 7 of the paper are far from convincing in demonstrating that simple conditional probability schemes provide satisfactory models for spatial processes. It is felt to be pertinent that, in each case, the data were derived from regions of the plane rather than point sites. There is clearly a need for more practical analyses to be undertaken. Some alternative suggestions on the specification of lattice models for aggregated data would also be of great interest.

Acknowledgements

I am indebted to those authors who have allowed me to quote their work prior to publication and also to the participants in the Liverpool "Spatial Pattern and Interaction" meeting in November 1973 for valuable discussions, especially Tim Gleaves who collected the *Plantago lanceolata* data and carried out the associated preliminary computations. Finally, I should like to express my thanks to Professor M. S. Bartlett for his continued interest and encouragement.

References

Averintsev, M. B. (1970). On a method of describing complete parameter fields. *Problemy Peredaci Informatsii*, **6**, 100–109.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge: University Press.

—— (1967). Inference and stochastic processes. *J. R. Statist. Soc.* A, **130**, 457–477.

—— (1968). A further note on nearest neighbour models. *J. R. Statist. Soc.* A, **131**, 579–580.

—— (1971a). Physical nearest-neighbour models and non-linear time series. *J. Appl. Prob.*, **8**, 222–232.

BARTLETT, M. S. (1971b). Two-dimensional nearest-neighbour systems and their ecological applications. *Statistical Ecology*, Vol. 1, pp. 179–194. Pennsylvania State University Press.

—— (1974). The statistical analysis of spatial pattern. Presented at the Third Conference on Stochastic Processes and their Applications, Sheffield, August 1973.

BARTLETT, M. S. and BESAG, J. E. (1969). Correlation properties of some nearest-neighbour models. *Bull. Int. Statist. Inst.*, **43**, Book 2, 191–193.

BATCHELOR, L. D. and REED, H. S. (1918). Relation of the variability of yields of fruit trees to the accuracy of field trials. *J. Agric. Res.*, **12**, 245–283.

BESAG, J. E. (1972a). Nearest-neighbour systems and the auto-logistic model for binary data. *J. R. Statist. Soc.* B, **34**, 75–83.

—— (1972b). On the correlation structure of some two-dimensional stationary processes. *Biometrika*, **59**, 43–48.

—— (1972c). On the statistical analysis of nearest-neighbour systems. Proceedings of the European Meeting of Statisticians, Budapest, August 1972 (to appear).

BROOK, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, **51**, 481–483.

BROWN, G. S. (1965). Point density in stems per acre. *N.Z. For. Serv. Res.* Note **38**, 1–11.

CLARKE, R. D. (1946). An application of the Poisson distribution. *J. Inst. Actuar.*, **72**, 481.

COCHRAN, W. G. (1936). The statistical analysis of the distribution of field counts of diseased plants. *J. R. Statist. Soc. Supplt*, **3**, 49–67.

COX, D. R. (1970). *Analysis of Binary Data*. London: Methuen.

—— (1972). The analysis of multivariate binary data. *Appl. Statist.*, **21**, 113–120.

FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*, Vol. 1. New York: Wiley.

FREEMAN, G. H. (1953). Spread of disease in a rectangular plantation with vacancies. *Biometrika*, **40**, 287–305.

GLEAVES, J. T. (1973). *Unpublished Ph.D. Thesis*, University of Liverpool.

GREIG-SMITH, P. (1964). *Quantitative Plant Ecology*, 2nd ed. London: Butterworth.

GRIMMETT, G. R. (1973). A theorem about random fields. *Bull. London Math. Soc.*, **5**, 81–84.

HAMMERSLEY, J. M. and CLIFFORD, P. (1971). Markov fields on finite graphs and lattices (unpublished).

HANNAN, E. J. (1955a). Exact tests for serial correlation. *Biometrika*, **42**, 133–142.

—— (1955b). An exact test for correlation between time series. *Biometrika*, **42**, 316–326.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, Vol. 2. London: Griffin.

LÉVY, P. (1948). Chaînes doubles de Markoff et fonctions aléatoires de deux variables. *C. r. Acad. Sci., Paris*, **226**, 53–55.

MATUI, I. (1968). Statistical study of the distribution of scattered villages in two regions of the Tonami plain, Toyami prefecture. In *Spatial Analysis*. Berry, B. J. and Marble, D. F., Eds. New Jersey: Prentice-Hall.

MEAD, R. (1966). A relationship between individual plant spacing and yield. *Ann. Bot.*, **30**, 301–309.

—— (1967). A mathematical model for the estimation of interplant competition. *Biometrics*, **23**, 189–205.

—— (1968). Measurement of competition between individual plants in a population. *J. Ecol.*, **56**, 35–45.

—— (1971). Models for interplant competition in irregularly distributed populations. *Statistical Ecology*, Vol. 2, pp. 13–32. Pennsylvania State University Press.

MERCER, W. B. and HALL, A. D. (1911). The experimental error of field trials. *J. Agric. Sci.*, **4**, 107–132.

MORAN, P. A. P. (1973a). A Gaussian Markovian process on a square lattice. *J. Appl. Prob.*, **10**, 54–62.

—— (1973b). Necessary conditions for Markovian processes on a lattice. *J. Appl. Prob.*, **10**, 605–612.

OGAWARA, M. (1951). A note on the test of serial correlation coefficients. *Ann. Math. Statist.*, **22**, 115–118.

ORD, J. K. (1974). Estimation methods for models of spatial interaction. *J. Amer. Statist. Ass.* (to appear).

O'SULLIVAN, P. M. (1969). *Transportation Networks and the Irish Economy*. L.S.E. Geographical Papers No. 4. London: Weidenfeld and Nicholson.

PATANKAR, V. N. (1954). The goodness of fit of frequency distributions obtained from stochastic processes. *Biometrika*, **41**, 450–462.
PLACKETT, R. L. (1960). *Principles of Regression Analysis*. Oxford: Clarendon Press.
PRESTON, C. J. (1973). Generalised Gibbs states and Markov random fields. *Adv. Appl. Prob.*, **5**, 242–261.
ROSANOV, YU. A. (1967). On the Gaussian homogeneous fields with given conditional distributions. *Theor. Probability Appl.*, **12**, 381–391.
ROTA, G. C. (1964). On the foundations of combinatorial theory. *Z'tscher. Wahrsch'theorie & verw. Geb.*, **2**, 340–368.
SHERMAN, S. (1973). Markov random fields and Gibbs random fields. *Israel J. Math.*, **14**, 92–103.
SPITZER, F. (1971). Markov random fields and Gibbs ensembles. *Amer. Math. Monthly*, **78**, 142–154.
WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434–449.
—— (1963). Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, **40**, 974–994.
WILLIAMS, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika*, **39**, 151–167.

DISCUSSION OF MR BESAG'S PAPER

Professor D. R. COX (Imperial College): The paper is original, lucid and comprehensive. The topic is important and notoriously difficult. It is a pleasure to congratulate Mr Besag.

Statistical subjects can be characterized qualitatively by their statistical analysis to stochastic model ratio. This is rather low in the present subject and therefore the emphasis in the present paper on models for the analysis of data is very welcome. Nevertheless understanding of the conditional models may be helped by relating them to temporal–spatial models, and in particular to their stationary distributions. It would be interesting to know what general connections can be established between Mr Besag's auto-models and stationary distributions of simple temporal–spatial processes.

Mr Besag remarks on the possible advantages of a triangular rather than a rectangular lattice. Has this been tried numerically on examples of aggregated responses, where a comparison with a square lattice could be made? The physical meaning of a first-order scheme is more appealing for a triangular lattice than for a square one. There is possibly a qualitative connection with results on the location of points in sampling for a mean (Dalenius *et al.*, 1961).

Experimental design aspects are mentioned briefly in the paper. The link here is, I think, with the method of Papadakis (Bartlett, 1938). In this the treatment effects are estimated after adjustment by analysis of covariance on the residuals on neighbouring plots. The one-dimensional version of this has been related by Atkinson (1969) to an autoregressive process and Mr Besag's discussion probably provides a framework for the two-dimensional theory.

The sections of the paper on the coding method are of particularly general interest as illustrating one more technique for simplifying complicated likelihoods. There are many outstanding questions; a qualitative explanation of the high efficiency in the one-dimensional case (Hannan, 1958) might throw light on the two-dimensional behaviour.

I propose a cordial vote of thanks to Mr Besag for his excellent paper.

Dr A. G. HAWKES (University of Durham): Like many people who have proposed or seconded votes of thanks at meetings of this Society I am distinguished by the fact that I know little about the subject of the paper. Of course I have seen papers previously on the analysis of distributions on lattices. Having received the distinct impression that they contained rather nasty, messy mathematics and unpleasant computation, since I had no desperate need to understand them, I put them on one side and, apart from a brief look, tended not to work through them carefully. Therefore I am extremely grateful to Mr Besag for presenting this paper with his elegant general treatment of distributions on lattices—or, indeed, for any multivariate distribution at all—and his interesting general results.

In addition, he gives a simple and flexible class of automodels, a simple method of analysis and some nice practical examples.

In consequence of this he makes the methods which he proposes readily available for use by practising statisticians, and I am sure as a result of this these methods will become used quite widely.

The examples he proposes show that the models are not likely to be satisfactory for all spatial problems. Hopefully, in time we will accumulate a fund of examples in which they seem to be adequate. Most of the examples mentioned by Mr Besag are concerned with plant ecology. I thought it might be interesting to return to the place from where most of the terminology comes, and consider problems in sociology in which the dots would represent houses, and in which we are concerned with social interactions of neighbours and cliques of neighbours. One kind of problem in this connection which would be interesting, although it is not dealt with at all in this paper, would be to identify what the cliques were in that sort of situation. There would not be the same sort of homogeneous type system as that described by Mr Besag; the cliques would differ over the whole area, and it would be interesting to try to determine what they were.

There are problems too for the academic researcher: what is the efficiency of the coding procedure? Apart from the Gaussian model, what other models are there which can be tackled by the complete likelihood method without too much nasty computation? Apart from the auto-models, what other simple kinds of models can be found which might be generally useful? In particular, it might be interesting to know whether Poisson models or exponential models could be found which exhibited some kind of co-operation rather than competition which Mr Besag mentioned.

The paper is extremely comprehensive and does not leave me much to say about it. I have two small comments: first, I would have welcomed some suggestion on how the results of the hypothesis tests on the different codings should be combined, apart from a rather vague statement that they should be combined in some suitable conservative manner. My inclination would be to take the average of the exact significance levels, or P values if you like, as a reasonable sort of measure.

Secondly, on the question of the restrictions on the parameter values which there have to be in order to make the schemes valid, in a number of places Mr Besag gives us general results. I am thinking in particular of the auto-normal case; they are not, in fact, absolutely trivial to work out in practice for special cases. For the simple special cases such as first-, second- and third-order auto-normal I would have welcomed it if he could have written these out more explicitly and prominently for the people who will be using them, in order to see exactly what is going on.

With the coding methods there is no reason why estimate values should be obtained which are admissible—because the method of estimation does not impose those restraints. However, with the likelihood approach there should be estimates obtained which are admissible always. I think Mr Besag did not emphasize that point sufficiently strongly. There is a short paragraph at the end of Section 7 of the paper, in which Mr Besag appears to say, almost as an afterthought, "Oh, by the way, half my estimates are inadmissible". I felt that should have been emphasized a little more when he went through the discussion of the examples.

These are minor quibbles only, and they do not detract at all from the excellent paper which I am sure will be valuable in stimulating a lot of work, both practical and theoretical, in the future. I am very pleased to second the vote of thanks.

The vote of thanks was passed by acclamation.

Dr PETER CLIFFORD (University of Bristol): I am happy to have this opportunity of adding my congratulations to the author on this stimulating paper. My invitation to join this discussion arrived only this morning and I take it from this that my comments should be brief. The last occasion on which I joined a discussion of this learned society was, I believe, in 1964. As time has gone by, I have looked at those few cryptic sentences of mine and the realization has dawned on me that my comments were worthless. The author of

that paper was kind enough not to expose my ignorance and I hope Mr Besag will be so minded. In an effort to improve my contribution I have studied recent published discussions of the Royal Statistical Society. I was particularly impressed that Fellows have been moved to speak in rhyme. With such thoughts in mind I would like to propose a minor change in notation. In the translated works of Dobrushin we frequently see Markov stream for Markov chain. How much more pleasant it would be to see Markov meadow for Markov field. Thus the poets in our audience could stroll by Markov streams through Markov meadows.

My specific comments concern firstly the use of coding. Since all tests for auto-models are conditional on the boundary, it is a question of what else we choose to condition on. For first-order schemes the author bases his inference on the behaviour of half the sites with half considered to be constant. Let us call this 50 per cent utilization. In an effort to increase the utilization I would like to suggest the following scheme which conditions on every *third* diagonal.

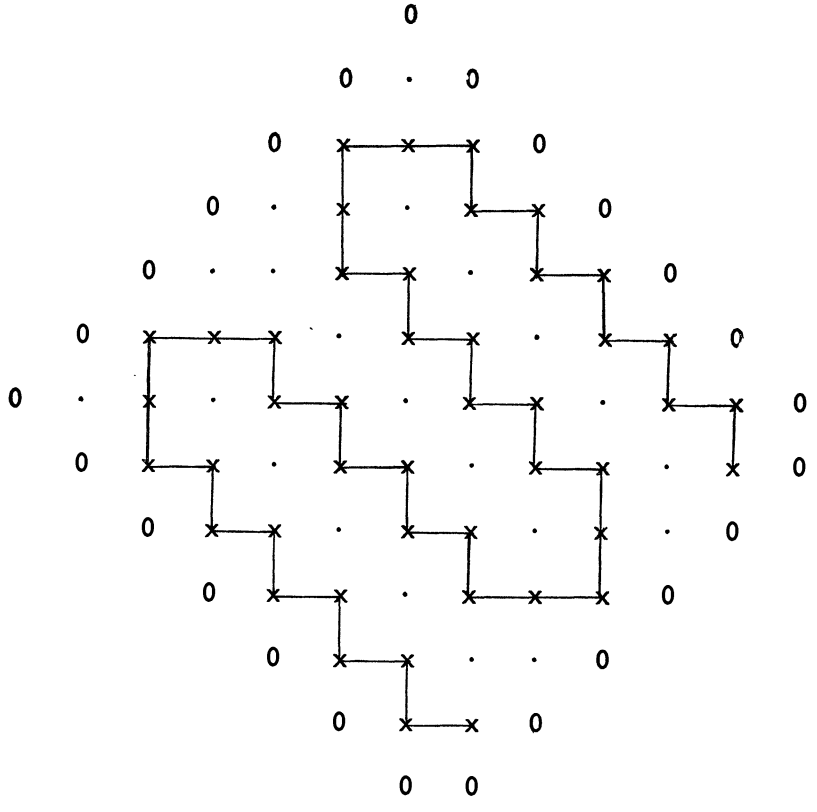

FIG. 1. Coding scheme for first order Markov field, exposing underlying conditional Markov chain.

Observe that between these diagonals there is, conditional on the diagonals, a Markov chain (or stream) whose likelihood can be written down. The test for the presence of a first-order effect then reduces to a test for a first-order effect in a chain, this, of course,

corresponds to 67 per cent utilization. For the triangular and second-order rectangular lattice the utilization is improved from 25 to 50 per cent. Greater economies will be achieved in higher dimensions. Thus one can produce tests with a greater degree of utilization by tapping one's knowledge of Markov chains. Again there will be the problem of combining several *dependent* test statistics in the most efficient manner—there are more than four different ways of coding for conditionally independent sites in the manner of Fig. 2 in the paper (by slipping any pair of columns down by one unit) and there are many routes that the Markov stream can follow as it meanders through the meadow.

My second point is a general one concerning the applicability of auto-models to data. In most situations the data represent a time cross-section of a spatial–temporal process. The aim should be to estimate the relative values of the parameters of this process from the data. In this way one would hope to be able to predict the development and say something about the past of the process. Auto-models do arise from spatial processes but to my knowledge only as equilibrium distributions for either a time reversible local birth–death process or a process in which particles move with Spitzer's speed–change interaction. Since the spread of plants is not likely to be in equilibrium nor to be a purely local phenomenon and the plants do not move around then even if we could estimate the parameters we would not feel confident in using them to predict the future. My personal feeling is that there is a need for more research into the dynamics of spatial processes as, for example, Clifford and Sudbury (1973).

My final comment concerns the paper by Hammersley and myself. Whatever the historical reasons for not publishing in 1971 the paper has clearly been superseded by the work of others and notably by the excellent exposition we have heard today.

Professor P. WHITTLE (University of Cambridge): Mr Besag has given us a substantial, interesting and most useful paper. It collects together a great amount of material and clarifies a number of points admirably. For example, the Hammersley–Clifford theorem now appears very much more accessible.

On the question of joint versus conditional specifications, there is probably no point in being very partisan, as there are plainly arguments on both sides. I must confess, however, that I still find a conditional specification natural only in a context where the conditionings occur in a natural ("temporal") sequence, so that the sequence of conditioning sigma-fields is monotone, and there are no unobvious restrictions on the conditional probabilities, as there are in abundance in the spatial case. One feels, however, that all these questions will find their proper resolution when the process is embedded in a spatio-temporal process, as Professor Bartlett has done in the paper quoted (1971a) and as I did myself in a special case in my paper on topographic correlation (Whittle, 1962). Properties of the equilibrium spatial process follow nicely in some cases from those of the spatio-temporal process. For example, Spitzer has shown that a process which is both spatially and temporally first-order Markov (with appropriate definition), and is moreover reversible, has an equilibrium spatial distribution which is second-order Markov.

The "awkward normalizing factor" which Mr Besag mentions at the end of Section 2 is of course important in its own right, and none other than the partition function in statistical mechanical contexts.

Mr Besag indicates some hope that conditional descriptions will resolve the questions of non-identifiability mentioned at the end of Section 5. In fact, though, as long as the model is Gaussian these non-identifiabilities are intrinsic and unavoidable, however one may have chosen to specify the Gaussian process.

Finally, I might mention that both Mr Besag and I might have achieved a better fit with our models had we docked the central spike off the correlogram, which presumably reflects uncorrelated noise superimposed on the variable of the spatial model. In one of the cases I examined in Whittle (1954) the correlogram followed the predicted form (a Bessel function) very well except just at the origin.

Dr KEITH ORD (University of Warwick): First of all, I should like to join with earlier contributors in thanking Mr Besag for his stimulating paper, which, in the best tradition of this Society, gave us both sound theory and sound practice.

The first point on which I should like to comment is the question of joint as against conditional specifications.

Let us consider a temporally stationary model in discrete time for n sites of the form

$$\underset{(n\times n)(n\times 1)}{\mathbf{A}\mathbf{X}_t} + \underset{(n\times m)(m\times 1)}{\mathbf{B}\mathbf{Z}_t} = \underset{(n\times 1)}{\boldsymbol{\varepsilon}_t}.$$

The notation is chosen deliberately to make a comparison with simultaneous equation models in econometrics, and $\mathbf{A}, \mathbf{B}$ represent matrices of parameters, $\mathbf{Z}_t$ a matrix of m exogeneous or lagged endogeneous variables and $\boldsymbol{\varepsilon}_t$ the error vector. If the model is specialized to

(i) $\mathbf{A} = \mathbf{I}$, $\mathbf{Z}_t$ lagged endogeneous only, then the resulting spatial model for $\mathbf{X}_t$ is of the conditional form (cf. Bartlett, 1971a);

(ii) $\mathbf{A} \neq \mathbf{I}$, then simultaneous, rather than lagged, dependence is allowed and the joint model follows;

(iii) $\mathbf{Z}_t = \boldsymbol{\varepsilon}_t$, then a simultaneous moving average model results.

The purpose of these examples is to illustrate that the choice between joint and conditional models is, at least in one important case, a choice between theories of simultaneous and lagged dependence. To date, econometrics has operated entirely in the joint model framework, but the other side of the fence should be explored, at least.

My other point relates to the evaluation of *ML* estimators discussed in Section 6.3. Since my earlier work in this area (Ord, 1974), it has been found possible to evaluate the determinant $|\mathbf{B}|$ analytically for the square lattice for first-order models of form (5.5), through use of the eigenvalues of the connection matrices. While Mr Besag's numerical approach appears to have worked well, I wonder whether we do not need more analytic results in this direction if realistic mixed regressive, auto-regressive, models are to be made operational?

Mr R. MEAD (University of Reading): First I must say how much I have learned from conversations with Mr Besag and tonight from this paper. I feel that he has constructed a most impressive general theory of a class of spatial interaction models, exposing the structure of the conditional probability approach and its relationship with the joint probability approach. As an essentially applied statistician I am always conscious of the need for that generality of approach which is the hallmark of the mathematical statistician. Yet, at the same time, I distrust generalization because of its Procrustean tendency to force different problems into a single format. Thus my comments tonight are concerned with doubts as to the practical use of the methods of this paper.

When discussing the application of the spatial interaction models to problems of spatial patterns the author comments (Section 5.2.1): "An alternative procedure might be to adopt some sort of nested analysis (Greig-Smith)." I believe that the approach of this paper and the Greig-Smith approach are in no sense alternatives since they are attempting to answer different questions. The auto-logistic models are concerned with small-scale patterns on the scale of one or two quadrat widths. Greig-Smith's methods are concerned with detecting the several scales of pattern up to a scale of 32 or 64 quadrat widths. It is a mistake to assume that there is one form of pattern which will be detected more or less efficiently by different methods. The method of pattern detection must be related to the questions the ecologist wants to answer. Unfortunately there is no mention in Section 7.1 of the objectives in collecting the data on *Plantago lanceolata* so that it is not possible to judge whether the auto-logistic models are appropriate here. However, my experience with ecologists is that they usually assume that spatial distributions contain patterns at possibly several differing scales, which they wish to detect. So the answer to the author's question in Section 8 as to how Gleaves's data could be analysed using joint probability

models is perhaps that neither form of spatial interaction model is appropriate for analysing spatial patterns.

This brings me on to my other comments which concern the conditional–joint comparison. In my experience the data available for looking at inter-plant competition usually consist of small regular lattice arrays of plants (typically 30–100 plants) for a number of different treatments often with gaps or markedly atypical plants, so that the usable array is much smaller still. The object in measuring competition is to compare the degree of competitions in different situations. To make these comparisons we need information on the precision of our estimates. The paper includes, in various tables quoted in Section 7, approximate standard errors of the parameter estimates. Is anything known about the small-sample properties of these standard errors? There appears to be no method yet of examining the sampling distributions of the parameter estimates of the auto-models. Using the joint-probability models however it is always possible to obtain sampling distributions, albeit by simulation methods.

So, in spite of the very convincing arguments advanced in the paper that conditional models are preferable, I am still not convinced of their uniform superiority over the joint-probability models. I agree that conditional models appear to have every advantage save this one of sampling distribution; but I want to use these models to provide answers for practical problems and so unless Mr Besag has something more up his sleeve I shall stick with the joint-probability models.

The following contributions were received in writing:

Dr J. M. HAMMERSLEY (University of Oxford): I am much impressed by Mr Besag's very interesting and far-ranging paper on spatial interaction. The variety and flexibility of his treatment will greatly contribute to our understanding of these very difficult problems of co-operative phenomena.

In his paper he expresses regret that Dr Clifford and I decided not to publish our joint paper; so perhaps I might explain, by way of an historical note, our reasons for this decision. Essentially our theorem states that the probabilities associated with a Markov field must satisfy certain algebraic identities. In proving this result, we assumed a *positivity condition*, namely that no probability should be zero. Mr Besag says that "the positivity condition remains unconquered and it would be of considerable theoretical interest to learn the effect of its relaxation. On the other hand, it is probably fair to say that the result would be of little practical significance in the analysis of spatial interaction with given site locations." However, with this we do not agree: in many of the most important practical applications to statistical mechanics, the physical system is subject to constraints which prevent the system from assuming certain *forbidden states*. Nevertheless, a forbidden state can be subsumed into the general theory if we interpret it as a state having zero probability. So it seemed to us not only aesthetically desirable but also practically important to amend our proof in order to make the theorem independent of any positivity condition. Moreover, at first sight the necessary amendments appeared to be quite straightforward. For example, we felt that we ought to be able to include zero probabilities by means of some limiting argument. After all, algebraic identities are very simple and robust mathematical objects, which are unlikely to fall to pieces when taken to a limit. Indeed I remember from my undergraduate days a stock device of considerable use in the Mathematical Tripos; it was known as the Principle of the Irrelevance of Algebraic Inequalities; and it asserted that, if a polynomial in x was identically zero in some open interval $a<x<b$, then this polynomial must be identically zero for all x. And this principle could easily be extended to polynomials in several variables $x_1, x_2, \ldots, x_n$. Now we could certainly express our theorem in terms of polynomials (in the probabilities $p_1, p_2, \ldots, p_n$) being identically zero for all $p_i>0$; and hence these polynomials ought to remain identically zero for $p_i \geqslant 0$. In short, the positivity inequalities $p_i>0$ looked irrelevant. So much for hand-waving mathematics. On the other hand, wriggle as we

might, we were unable to convert this reasoning into a watertight argument. The very good reason for our failure was the unexpected discovery by a graduate student, Mr John Moussouris, of a counter-example! His discovery, and his subsequent development and discussion of what happens in the absence of the positivity condition, afforded the substance of his M.Sc. Thesis at Oxford. This thesis (which is being published in full in the *Journal of Statistical Physics* under the title "Gibbs and Markov random systems with constraints") supersedes the Hammersley–Clifford paper (hence our decision to leave our paper unpublished) as well as other published accounts (such as those by Averintsev, Spitzer, Grimmett, etc. quoted in Mr Besag's references). There are still some open questions (which will be found at the end of Moussouris's paper†); but, by and large, it can now be said that the positivity condition is "conquered".

Professor M. S. BARTLETT (University of Oxford): I regret that my absence in Australia has prevented my coming to this meeting, but it is indeed a pleasure to send my congratulations to the author for this very comprehensive and valuable paper on nearest-neighbour lattice models, containing both exposition of basic theory and of statistical techniques of analysing spatial interaction. I find his discussion of auto-normal and auto-logistic model-fitting especially interesting and useful.

With the auto-normal model his ingenious coding methods are not perhaps so necessary in view of the possibility of maximum-likelihood estimation, adapted either from Whittle's Jacobian technique for the simultaneous model or equivalently, as Mr Besag notes, by equating the theoretical and observed relevant nearest-neighbour correlations. With the auto-logistic model, on the other hand, the analogous second-order polynomial in the exponential also identifies the maximum-likelihood method with the equating of theoretical and observed correlations, and in view of the absence of a complete theoretical solution for the correlation (for general mean) emphasizes the current unavailability of the maximum-likelihood method in the auto-logistic case. It is clearly desirable to have further information on the efficiency of the coding method; in the meantime, Hannan's results for the one-dimensional case suggest a much better efficiency when the mean from the alternative patterns is taken.

With regard to the earlier theoretical development in the paper, and the author's basic Assumptions 1 and 2, the use of Assumption 2 has seemed to me somewhat arbitrary. The assumption does of course ensure attractive sufficiency properties for the appropriate statistics, but this in itself seems more a matter of convenience than of necessity.

Finally, with the auto-logistic model (for an infinite lattice), the author will be aware from its identity with the Ising model in physics that there is a critical point before the nearest-neighbour correlation reaches unity, and some recent results of my own (Bartlett, 1974) suggest an intriguing ambiguity of fitting as a consequence. Above the critical point the non-ergodic long-range component of the correlation appears interpretable as a non-zero mean, even in the nominally zero mean case, so that the observed mean and nearest-neighbour correlation might in some cases be fitted either by high α and low β, or by low α and high β. With these parameter values, I am assuming $-1, 1$ (not $0, 1$) as the possible x values, as in the Ising model, where this is equivalent to saying that, if magnetization is observed, it is not possible to say (unless the temperature is stated, corresponding to the value of β) whether the magnetization is a consequence of an external magnetic field ($\alpha \neq 0$, and β small) or is spontaneous ($\alpha \to 0$, but β high).

The author replied later, in writing, as follows:

I am extremely grateful to all of the discussants for their very interesting and helpful comments. This especially since the three-day week (Heath *et al.*, 1974) prevented the galleys from being available in good time.

† *Note added in proof*: Moussouris's paper has now appeared in *J. Statist. Phys.*, **10** (1974), 11–33.

Professor Cox suggests that a knowledge of the (stationary) spatial structure of simple temporal models may help an understanding of conditional (and presumably other) spatial schemes. I shall try to cite some of the available results during the course of my reply. For example, there is the class of models associated with discrete state, time-reversible Markov processes (Kendall, 1959), as mentioned both by Dr Clifford and by Professor Whittle. Such processes, more or less by definition, have stationary spatial distributions possessing obvious conditional probability structures. In Besag (1972a), I give a simple example which generates a first-order auto-logistic scheme on the rectangular lattice as stationary distribution. This model has now been simulated by Mr David Green, using Monte Carlo methods, as part of his final-year undergraduate project at Liverpool and his results provide useful information on methods of parameter estimation and goodness-of-fit tests. Preston (1973) examines the stationary distribution for the analogous binary model defined on an arbitrary finite graph and gives results which may easily be generalized. Thus, subject to positivity, *any* discrete multivariate (spatial) distribution may be generated as the stationary limit of some easily identified time-reversible process. Unfortunately, this process may have no direct physical or biological interpretation but, even in such cases, the result may still be useful for computer simulation of spatial schemes. For binary variables, other spatial–temporal processes which have been investigated include the "jumping particles" of Preston (1973) and the "spatial conflict" models of Clifford and Sudbury (1973) for which some evolutionary results are even available.

Professor Cox mentions the appeal of using hexagonal rather than square regions when fitting a first-order scheme to aggregated responses. I agree but do not yet have any numerical results. Another possibility is to investigate the influence of size of plot on the resulting analysis. Here there may be connections with Whittle (1956). Concerning Papadakis's method, it would certainly be nice to extend Dr Atkinson's work to two dimensions although I do not, at present, see this as being a simple matter!

One problem in setting up Professor Hawkes's "fund of examples" is the apparent lack of published spatial data. A collection of examples has been started by Dr Ord and we should be most grateful for additional contributions. The identification of sociological cliques, mentioned by Professor Hawkes, is an intriguing problem. Related work has, I believe, appeared in the social sciences literature concerning gravity models etc. Regarding the combination of significance levels, this could certainly do with further investigation. The simple technique which I used in Table 7 was to multiply the minimum P value (just greater than 1 per cent) by the number of tests (four), which led to the rejection of scheme (b) at the 5 per cent level. This procedure clearly gives a rather conservative test. Professor Hawkes's suggestion of taking the average P value would, I think, generate extremely conservative tests.

Professor Hawkes points out my failure to give explicit conditions on the allowable parameter values for auto-normal lattice schemes. A sufficient condition for the rectangular lattice scheme in which $X_{i,j}$ has conditional mean $\mu_{i,j}+\sum\sum b_{k,l}\,x_{i-k,j-l}$ is that

$$\sum\sum b_{k,l}\,z_1^k\,z_2^l<1$$

whenever $|z_1|=|z_2|=1$. This follows from the conditions for existence of stationary auto-normal schemes on the infinite lattice (Section 5.5). The condition is not necessary on a finite lattice but is "almost so" in most practical situations where the lattice is not too small. In some cases, exact results are available. For example, it follows from the evaluation of eigenvalues in Ord (1974) that for a first-order isotropic scheme on an $n\times n$ lattice to be valid, one requires that $|\beta|\cos\{\pi/(n+1)\}<\frac{1}{2}$ or, to all intents and purposes, $|\beta|<\frac{1}{2}$, the result for the stationary scheme. It follows from Dr Ord's discussion comments that the conditions for the anisotropic scheme are also known exactly and may therefore be compared with the asymptotic requirement, $|\beta_1|+|\beta_2|<\frac{1}{2}$. This will also tell us whether some of the individual coding estimates for the Mercer and Hall data *are* in fact inadmissible for the overall model (I only stated that they were inconsistent with a *stationary* auto-normal scheme). Assuming they are, I would not expect this to be so in

situations where one wished to accept the model. However, I certainly agree that I should have emphasized the inconsistency more strongly in the paper rather than as an "afterthought" (which it was!).

Incidentally, if one fits a third-order stationary auto-normal scheme to the wheat plots data, the overall fit is substantially improved but still leaves much to be desired. Since the first- and second-order auto-normal schemes and the simultaneous autoregressive models 1, 2, 3, 5, 6, 7 in Whittle (1954) are all special cases of the third-order auto-normal model, we may construct appropriate goodness-of-fit tests. This leads to rejection of each of the special cases at the $\frac{1}{2}$ per cent level or less, except for scheme 7 for which the P value is 7 per cent. Further comments on model-fitting appear later.

On improving the coding technique, Dr Clifford's suggestion, ingenious though it is, looks to me rather awkward to use in practice. The stream is not only inhomogeneous, its banks appear to recede as one travels along. However, I believe Dr Clifford has taken the plunge and I very much look forward to seeing his results. I agree wholeheartedly with the appeal for more research into the dynamics of spatial processes. I feel no conflict here with the aims of the present paper; rather I see the two approaches as being complementary. An examination of the evolutionary (where possible) or equilibrium spatial structure of simple but plausible spatial–temporal models can hopefully give a *qualitative* lead to the types of spatial model which may be appropriate in practice. On the other hand, in order to devise detailed spatial schemes for the quantitative statistical analysis of spatial data, it is, I believe, useful to have available a purely spatial approach which is capable of various degrees of generalization.

I certainly agree with Professor Whittle that there is no point in being particularly partisan about "joint" versus "conditional" specifications of spatial schemes. If I have sounded otherwise then I apologize. Indeed, I would think that Assumption 1 (Section 4.1) might easily be interpreted as a first step in a *joint*-probability approach based upon the representation (3.3) of $P(\mathbf{x})$. I fancy that any personal prejudice I may hold against simultaneous autoregressions stems from the difficulty I have in concentrating, at one time, upon more than a single random variable in a dependent set. As regards the nature of the conditioning sigma-fields, could I perhaps reverse Professor Whittle's view and suggest that monotonicity is only a natural concept when one is constructing a model for the *development* of a process, through cause and effect, rather than just a spatial scheme to describe its "here and now" probabilistic properties? On a related point, my comment on the identifiability of parameters (Section 5.5) may have been misleading. What I meant is that by attempting only to determine the underlying probability structure of a spatial system, rather than postulate an actual process, one avoids (*not* "overcomes") the problem of identifiability. May I stress that I think of a "spatial scheme" merely as a collection of random variables with an associated probability structure rather than as a physical model in its classical sense?

I am grateful to Professor Whittle for pointing out that each of us would have achieved a better fit to the Mercer and Hall wheat plots data had we docked the central spike off the observed correlogram; that is, had we fitted the model $X_{i,j} = Y_{i,j} + Z_{i,j}$, where $X_{i,j}$ denotes plot yield, $Z_{i,j}$ denotes uncorrelated noise and $Y_{i,j}$ is, respectively, a first-order simultaneous autoregressive process or a first-order auto-normal scheme. In either case, $Y_{i,j}$ can be thought of as reflecting variations in soil fertility and $Z_{i,j}$ as reflecting the intrinsic variability of the wheat itself, from plot to plot (cf. Whittle, 1954). Unfortunately, for either the simultaneous or conditional probability scheme, $\{X_{i,j}\}$ is no longer a finite-order process and this leads to complications in maximum-likelihood estimation and in testing goodness of fit. It is intended to discuss such problems elsewhere. However, returning to the general form of the correlogram, I doubt whether docking the central spike really satisfies Professor Whittle. His evidence (Whittle, 1956), based partially upon the results in Fairfield Smith (1938), that, in agricultural uniformity trials, covariance functions often decay as a power of distance, rather than exponentially, at least for large lags, still conflicts with both the simultaneous and conditional multinormal schemes. Indeed, Professor

Whittle (1962) has constructed a physically plausible spatial-temporal random diffusion model which, in its three-dimensional form, exhibits the suggested power-law decay. My present confusion is only compounded by the statement in Quenouille (1949), claiming that an exponential correlation function *can* be used to provide "substantial agreement with Fairfield Smith's law over a wide range of values". There is clearly scope for further discussion of such problems.

Dr Ord discusses the choice between simultaneous and conditional spatial schemes in the context of discrete time, spatial-temporal autoregressions arising in econometrics. However, suppose we consider the (zero mean) temporally stationary process represented by

$$\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t-1} = \mathbf{C}\boldsymbol{\varepsilon}_t, \tag{1}$$

where **A**, **B** and **C** are suitable matrices of constant coefficients and $\operatorname{var} \boldsymbol{\varepsilon}_t = \sigma^2 \mathbf{I}$. Writing $\boldsymbol{\Gamma} \equiv E\mathbf{X}_t \mathbf{X}_t^{\mathrm{T}}$ for the spatial covariance matrix, it follows that

$$\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathrm{T}} - \mathbf{B}\boldsymbol{\Gamma}\mathbf{B}^{\mathrm{T}} = \sigma^2 \mathbf{C}\mathbf{C}^{\mathrm{T}}. \tag{2}$$

Matrix equations of this form have attracted much attention in their own right—for a recent account see Kučera (1974)—but, so far as I know, explicit matrix solutions are not in general available. We may of course proceed in situations where the matrices in equation (2) all commute, as they will for homogeneous regular torus lattice systems and for spatially stationary infinite regular lattice systems. We then have the familiar looking result,

$$\boldsymbol{\Gamma} = \sigma^2(\mathbf{A}\mathbf{A}^{\mathrm{T}} - \mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{C}\mathbf{C}^{\mathrm{T}}.$$

This spatial structure is not in general consistent with any simple conditional or simultaneous probability scheme. However, in the special case where $\mathbf{C} = \mathbf{I}$, suppose we define $\boldsymbol{\eta}_t$ by the purely spatial relationship (analogous to equation (5.13) in the paper),

$$(\mathbf{A}\mathbf{A}^{\mathrm{T}} - \mathbf{B}\mathbf{B}^{\mathrm{T}})\mathbf{X}_t = \boldsymbol{\eta}_t.$$

Then $E(\boldsymbol{\eta}_t \mathbf{X}_t^{\mathrm{T}}) = \sigma^2 \mathbf{I}$ and, provided the system is Gaussian, the spatial conditional expectation structure follows immediately. To obtain the corresponding simultaneous formulation, we need to find a matrix **D** such that $\mathbf{D}\mathbf{D}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} - \mathbf{B}\mathbf{B}^{\mathrm{T}}$. The only simple case therefore appears to occur when $\mathbf{B} = \mathbf{0}$. But equation (1) then no longer represents a true temporal model. Dr Ord's case (ii), in which $\mathbf{B} = \mathbf{0}$ presumably, is a *definition* of a simultaneous spatial scheme, not a temporal derivation. Similarly, case (iii) appears to lack any temporal consideration.

Another class of spatial-temporal models, related to that above but arguably of more intuitive appeal, is provided by the continuous time process,

$$\Delta\mathbf{X}_t + \mathbf{B}\mathbf{X}_t \Delta t = \Delta\boldsymbol{\varepsilon}_t,$$

with the usual notation. Provided **B** and $\mathbf{B}^{\mathrm{T}}$ commute, this leads to the corresponding spatial covariance matrix $\boldsymbol{\Gamma} = \sigma^2(\mathbf{B} + \mathbf{B}^{\mathrm{T}})^{-1}$ in the temporally stationary situation. When the variables are Gaussian, it follows that the spatial conditional expectation structure is particularly simple. This accords with a result in Bartlett (1971a). Again there is, in general, no obvious correspondence with a simultaneously defined spatial scheme. Thus, I am unaware of any simple situations in which truly temporal autoregressions, either in discrete or continuous time, suggest the use of simultaneous spatial schemes rather than conditional ones. Am I missing something?

I would certainly agree with Dr Ord's second point that there is a need for more analytical results concerning the evaluation of the Jacobian in either the joint or conditional models. Dr Ord has done some pioneering work in this direction and I look forward to seeing more of his results. These should be particularly useful in examining the efficiency of special estimation techniques, such as the coding method.

Mr Mead rightly criticizes my analysis (or lack thereof) concerning the *Plantago lanceolata* transect. Dr Gleaves is primarily interested in the interaction between gene flow and spatial pattern in plant communities. As a single aspect of his studies, Dr Gleaves required a test of randomness, preferably based upon distance methods, and we therefore developed the "*T*-square" sampling technique (Besag and Gleaves, 1973), the aims of which Mr Mead has criticized elsewhere. Having used the method with apparent success, Dr Gleaves collected the quadrat data to see what else we could learn about the community from statistical analysis. As one approach, it seemed reasonable to me, bearing in mind the peculiar environment of the plants, to examine the fit of homogeneous low-order auto-logistic schemes. I agree with Mr Mead that Professor Greig-Smith's technique aims to answer different, and possibly more relevant, questions but I still think it is somewhat premature to dismiss the auto-logistic method of analysis as a tool in plant ecology. Incidentally, the fact that a scheme is formally described as "locally interactive" does not imply that the patterns it produces are local in nature (cf. the extreme case of long-range order in the Ising model).

Regarding lattices of Gaussian variables and the small-sample properties of estimators, it seems to me that a Monte Carlo examination of the associated sampling distributions is equally available with either the simultaneous or conditional probability approach. My guess is that in order to simulate the joint model (4.13), with given parameter values, Mr Mead has at some stage to invert the non-singular matrix $\mathbf{B}$. Now for the conditional scheme (4.10), the matrix $\mathbf{B}$ is symmetric positive definite and, as I am sure Mr Mead is aware, the standard method of numerically inverting such a matrix is to determine a triangular matrix $\mathbf{C}$ such that $\mathbf{C}^{\mathrm{T}}\mathbf{C} = \mathbf{B}$. The matrix that Mr Mead requires to simulate (4.10) is just $\mathbf{C}^{-1}$.

I am delighted to hear from Dr Hammersley that the positivity condition has, by and large, finally been dispensed with. I spent longer than I should care to admit, a couple of years ago, trying to overcome the condition without any form of success. The *practical* relevance of positivity finally got through to me after reading Preston (1973); at the meeting itself, I gave an example based upon voting tendencies in each constituency at a parliamentary election. I very much look forward to a detailed reading of Mr Moussouris's paper and congratulate him on his work. However, I feel sure that Dr Hammersley will agree that, in most practical data analytic situations, the basic theorem will still suffice.

Professor Bartlett, in pointing out that, for the auto-logistic lattice models, maximum-likelihood estimation is (asymptotically) equivalent to equating the relevant observed and theoretical correlations, indicates how a Monte Carlo assessment of the relative efficiency of the coding technique (or other methods), with respect to maximum likelihood, may be made, at least for totally symmetric $\{P(0) = P(1)\}$ first-order schemes. As regards analytical results on the coding technique, my own limited work suggests that for an isotropic Gaussian first-order scheme, the asymptotic efficiency of a single coding test for *independence* is unity on the square and hexagonal lattices but only 2/3 on the triangular lattice! Bearing in mind previous comments on lattice type, this last value seems somewhat annoying. However, I believe Professor P. A. P. Moran has rather more informative results on coding efficiency with $\beta \neq 0$ and I look forward to seeing these in due course.

I agree with Professor Bartlett's remarks upon the somewhat arbitrary nature of Assumption 2, which admittedly was made largely for the reasons he suggests (sufficiency and convenience). I certainly did not intend to claim that it is a "basic" assumption to the conditional probability approach. Whilst I hope the schemes illustrate several points, for example the effect of the summability condition, there is clearly a need to set up and examine other types of schemes if the conditional probability approach is to be generally viable.

Regarding Professor Bartlett's final point, which relates to the non-ergodic nature of the Ising lattice beyond the critical point, I can but tentatively suggest that the use of a conditioning set, such as with the coding technique, may circumvent the problem of inconsistent parameter estimation.

REFERENCES IN THE DISCUSSION

ATKINSON, A. C. (1969). The use of residuals as a concomitant variable. *Biometrika*, **56**, 33–41.

BARTLETT, M. S. (1938). The approximate recovery of information from field experiments with large blocks. *J. Agric. Sci.*, **28**, 418–427.

—— (1974). Physical nearest-neighbour models and non-linear time-series, III. Non-zero means and sub-critical temperatures. *J. Appl. Prob.* (in press).

BESAG, J. E. and GLEAVES, J. T. (1973). On the detection of spatial pattern in plant communities. *Bull. Int. Statist. Inst.*, **45**, Book 1, 153–158.

CLIFFORD, P. and SUDBURY, A. W. (1973). A model for spatial competition. *Biometrika*, **60**, 581–588.

DALENIUS, T., HAJEK, J. and ZUBRZYCKI, S. (1961). On plane sampling and related geometrical problems. *Proc. 4th Berkeley Symp. Math. Prob.*, **1**, 125–150.

FAIRFIELD SMITH, H. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci.*, **28**, 1–23.

KENDALL, D. G. (1959). Unitary dilations of one-parameter semi-groups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states. *Proc. London Math. Soc.* (3), **9**, 417–431.

KUČERA, V. (1974). The matrix equation $\mathbf{AX}+\mathbf{XB} = \mathbf{C}$. *SIAM J. Appl. Math.*, **26**, 15–25.

QUENOUILLE, M. H. (1949). Problems in plane sampling. *Ann. Math. Statist.*, **20**, 355–375.

WHITTLE, P. (1956). On the variation of yield variance with plot size. *Biometrika*, **43**, 337–343.

—— (1962). Topographic correlation, power-law covariance functions, and diffusion. *Biometrika*, **49**, 305–314.