Nearest-neighbour Systems and the Auto-logistic Model for Binary Data

By J. E. BESAG

University of Liverpool

[Received February 1971. Revised September 1971]

SUMMARY

Bartlett (1966) and Whittle (1963), respectively, have proposed alternative, non-equivalent definitions of nearest-neighbour systems. The former, conditional probability definition, whilst the more intuitively attractive, presents several basic problems, not least in the identification of available models. In this paper, conditional probability nearest-neighbour systems for interacting random variables on a two-dimensional rectangular lattice are examined. It is shown that, in the case of 0, 1 variables and a homogeneous system, the only possibility is a logistic-type model but in which the explanatory variables at a point are the surrounding array variables themselves. A spatial–temporal approach leading to the same model is also suggested. The final section deals with linear nearest-neighbour systems, especially for continuous variables. The results of the paper may easily be extended to three or more dimensions.

Keywords: NEAREST-NEIGHBOUR SYSTEMS; AUTO-LOGISTIC MODEL; MULTIDIMENSIONAL BINARY DATA

1. INTRODUCTION

WE shall, principally, be considering a two-dimensional rectangular lattice, each node (or site), (i,j), of which has a random variable, $X_{i,j}$, associated with it. There appear in the literature two main definitions of nearest-neighbour models which might be applied to describe the interaction between the variables $X_{i,j}$ in this situation. Whittle (1963) suggested defining nearest-neighbour models in terms of the joint probability distribution of the variables and required that this should be of the product form

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

where $x_{i,j}$ is a value of the random variable $X_{i,j}$. On the other hand, Bartlett (1966, 1967, 1968) discussed an approach through conditional probabilities, in which he supposed that

$$P\{x_{i,j} \mid \text{all other values}\} \equiv P\{x_{i,j} \mid x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}\}, \tag{2}$$

that is, dependent only upon the nearest neighbours. Whilst this approach has considerably greater intuitive appeal, it unfortunately presents a number of difficulties. Notably, there is no direct method of evaluating the joint probability distribution on the lattice and, further, the functional form of the conditional probability on the right-hand side of (2) is subject to severe consistency conditions. The latter point was first noted by Levy (1948) concerning normally distributed random variables and subsequently, and more generally, by Brook (1964) who also demonstrated that

any valid model satisfying (2) must also satisfy (1). This situation is similar (and equivalent in one dimension) to that of considering a first-order Markov chain against one of second order: the former is degenerate with respect to the latter but is, of course, of considerable interest in its own right. The object of the present paper is to give some further insight into the nature of conditional probability nearest-neighbour models.

2. Conditional Probability Nearest-neighbour Schemes

We consider a conditional probability model for variables situated at each site of a two-dimensional rectangular lattice with given boundary values. We shall be particularly concerned with binary variables; for example, in an ecological context, this might correspond to an array of plants, each of which is either infected (1) or healthy (0), or to the presence (1) or absence (0) of a plant at a site. The perimeter sites of the array may be used to provide the boundary values, although in many practical situations it would seem reasonable to append a boundary of zeros to the array, corresponding, for example, to absence of plants there. In any event, denote the set of boundary sites of the array by B, these surrounding the set of internal sites denoted by I. Then a conditional probability nearest-neighbour model is defined by

$$P\{x_{i,j} \mid \text{all other values}\} \equiv p(x_{i,j} \mid x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}) \tag{3}$$

for all $(i,j) \in I$, where $x_{i,j}$ denotes the value of the random variable $X_{i,j}$ at the (i,j) site. We assume here that the model is spatially homogeneous, that is, the function p is independent of the internal position (i,j) on the array. For simplicity of notation, we shall suppose that the perimeter of the array is rectangular, numbering the rows $i = 0(1)m+1$ and the columns $j = 0(1)n+1$ but the conclusions will hold for any shape of closed boundary.

The explicit evaluation of the joint probability distribution of the inner array variables, conditional upon given boundary values, is not entirely straightforward. Let $\mathbf{x}$ denote a realization of the entire array, including boundary values, and let $\mathbf{x}_A$ be the part of this realization in any specified region A of the array. Then we may write the joint probability distribution as

$$P\{\mathbf{x}_I \mid \mathbf{x}_B\} = p(x_{m,n} \mid x_{m-1,n}, x_{m+1,n}, x_{m,n-1}, x_{m,n+1}) P\{\mathbf{x}_{I*} \mid \mathbf{x}_B\}, \tag{4}$$

where $I*$ denotes the set of internal sites, excluding the (m,n)th. However, if we try to repeat this procedure, using $x_{m-1,n}$, say, as "pivot", we get

$$P\{\mathbf{x}_{I*} \mid \mathbf{x}_B\} = P\{x_{m-1,n} \mid \mathbf{x}_{I**\cup B}\} P\{\mathbf{x}_{I**} \mid \mathbf{x}_B\},$$

where $I**$ denotes the set of internal sites, excluding the (m,n)th and the $(m-1,n)$th. However,

$$P\{x_{m-1,n} \mid \mathbf{x}_{I**\cup B}\}$$

is not reducible to the form $p(. \mid \ldots)$ since $x_{m,n}$ is absent from the set of conditioning variables. We may overcome this difficulty as follows. With each site, (i,j), of the lattice, we also associate a non-random variable, $y_{i,j}$. If $(i,j) \in I$, we allow ourselves to choose any feasible value for $y_{i,j}$, whilst, if $(i,j) \in B$, we define $y_{i,j} = x_{i,j}$. We may

now rewrite equation (4) as

$$P\{\mathbf{x}_I|\mathbf{x}_B\} = \frac{p(x_{m,n}|x_{m-1,n}, x_{m+1,n}, x_{m,n-1}, x_{m,n+1})}{p(y_{m,n}|x_{m-1,n}, x_{m+1,n}, x_{m,n-1}, x_{m,n+1})} P\{y_{m,n}, \mathbf{x}_{I*}|\mathbf{x}_B\}$$

$$= \frac{p(x_{m,n}|x_{m-1,n}, y_{m+1,n}, x_{m,n-1}, y_{m,n+1})}{p(y_{m,n}|x_{m-1,n}, y_{m+1,n}, x_{m,n-1}, y_{m,n+1})}$$

$$\times p(x_{m-1,n}\ x_{m-2,n}, y_{m,n}, x_{m-1,n-1}, x_{m-1,n+1}) P\{y_{m,n}, \mathbf{x}_{I**}|\mathbf{x}_B\}$$

$$= \frac{p(x_{m,n}|x_{m-1,n}, y_{m+1,n}, x_{m,n-1}, y_{m,n+1})}{p(y_{m,n}|x_{m-1,n}, y_{m+1,n}, x_{m,n-1}, y_{m,n+1})}$$

$$\times \frac{p(x_{m-1,n}|x_{m-2,n}, y_{m,n}, x_{m-1,n-1}, y_{m-1,n+1})}{p(y_{m-1,n}|x_{m-2,n}, y_{m,n}, x_{m-1,n-1}, y_{m-1,n+1})}$$

$$\times P\{y_{m-1,n}, y_{m,n}, \mathbf{x}_{I**}|\mathbf{x}_B\}. \tag{5}$$

We shall assume that the values *feasible* at any internal site are unaffected by the values of the surrounding nearest neighbours, so that $p(.|\ldots)>0$, provided each of its arguments is feasible. Then equation (5) gives the probability of the realization in terms of the probability of a realization which is identical save for (possibly) the (m,n)th and $(m-1,n)$th elements. We repeat the procedure along each of the columns of the inner array in turn, until each internal site value has been used as pivot. Hence,

$$P\{\mathbf{x}_I|\mathbf{x}_B\} = \prod \left\{\frac{p(x_{i,j}|x_{i-1,j}, y_{i+1,j}, x_{i,j-1}, y_{i,j+1})}{p(y_{i,j}|x_{i-1,j}, y_{i+1,j}, x_{i,j-1}, y_{i,j+1})}\right\} P\{\mathbf{y}_I|\mathbf{x}_B\}, \tag{6}$$

where the product is taken over all $(i,j)\in I$, $\mathbf{y}_I$ denotes the set of arbitrarily chosen values associated with the inner array sites and $y_{i,j} = x_{i,j}$ for $(i,j)\in B$. Equation (6) gives the joint probability distribution of a set of inner array variables in terms of the probability of an arbitrary (fixed) array $\mathbf{y}_I$.

On the other hand, instead of commencing the procedure with the (m,n) element, we might have used the $(1,1)$ element and continued in the reverse directions, giving rise to an alternative factorization analogous to (6). Equating the two results, we have

$$\prod \left\{\frac{p(x_{i,j}|x_{i-1,j}, y_{i+1,j}, x_{i,j-1}, y_{i,j+1})}{p(y_{i,j}|x_{i-1,j}, y_{i+1,j}, x_{i,j-1}, y_{i,j+1})}\right\}$$

$$= \prod \left\{\frac{p(x_{i,j}|y_{i-1,j}, x_{i+1,j}, y_{i,j-1}, x_{i,j+1})}{p(y_{i,j}|y_{i-1,j}, x_{i+1,j}, y_{i,j-1}, x_{i,j+1})}\right\}. \tag{7}$$

The arbitrariness of the choice of $\mathbf{y}_I$ in, for example, equation (6) implies severe restrictions on the functional form of $p(.|\ldots)$ and these restrictions we investigate below.

We assume for convenience, and without loss of generality, that zero is a feasible site value and, for any set of feasible site values, we let

$$f(x|t,u,v,w) \equiv \log\{p(x|t,u,v,w)/p(0|t,u,v,w)\}.$$

Then for any given $(r,s)\in I$, we put $y_{i,j} = x_{i,j}$ in (7) unless $(i,j) = (r,s)$ or $(r+1,s)$.

Letting $y_{r,s} = y_{r+1,s} = 0$, this gives

$$f(x\,|\,t,u,v,w)+f(u\,|\,0,u',v',w')\equiv f(x\,|\,t,0,v,w)+f(u\,|\,x,u',v',w').$$

It follows that

$$f(x\,|\,t,u,v,w)\equiv g(t\,|\,0,x,0,0)+g(u\,|\,x,0,0,0)+f(x\,|\,0,0,v,w),$$

where

$$g(x\,|\,t,u,v,w)\equiv f(x\,|\,t,u,v,w)-f(x\,|\,0,0,0,0).$$

Using the analogous procedure for any given pair of internal points (r,s) and $(r,s+1)$ gives

$$g(x\,|\,t,u,v,w)\equiv g(t\,|\,0,x,0,0)+g(u\,|\,x,0,0,0)+g(v\,|\,0,0,0,x)+g(w\,|\,0,0,x,0)$$

and hence

$$g(x\,|\,0,u,0,0)\equiv g(u\,|\,x,0,0,0) \quad \text{and} \quad g(x\,|\,0,0,0,w)\equiv g(w\,|\,0,0,x,0)$$

so that

$$g(x\,|\,t,u,v,w)\equiv g(x\,|\,t,0,0,0)+g(u\,|\,x,0,0,0)+g(x\,|\,0,0,v,0)+g(w\,|\,0,0,x,0).$$

If we write

$$f(x\,|\,0,0,0,0)\equiv x\phi(x), \quad g(x\,|\,t,0,0,0)\equiv xt\psi_1(x,t) \quad \text{and} \quad g(x\,|\,0,0,v,0)\equiv xv\psi_2(x,v)$$

then

$$f(x\,|\,t,u,v,w)\equiv x\{\phi(x)+t\psi_1(x,t)+u\psi_1(u,x)+v\psi_2(x,v)+w\psi_2(w,x)\} \tag{8}$$

and

$$p(x\,|\,t,u,v,w)\equiv \exp\{f(x\,|\,t,u,v,w)\}\Big/\sum_z \exp\{f(z\,|\,t,u,v,w)\}. \tag{9}$$

If we assume, as suggested earlier, that a boundary of zeros surrounds the inner array (i.e. $\mathbf{x}_B = \mathbf{0}$), it follows from equation (6), by also putting $y_{i,j} = 0$ for all $(i,j)\in I$, that

$$P\{\mathbf{x}_I\,|\,\mathbf{x}_B = 0\} = \frac{\prod \exp\{f(x_{i,j}\,|\,x_{i-1,j},0,x_{i,j-1},0)\}}{\sum \prod \exp\{f(z_{i,j}\,|\,z_{i-1,j},0,z_{i,j-1},0)\}}, \tag{10}$$

where the summation extends over all possible inner array configurations $\mathbf{z}_I$. Hence, it is easily shown that the constraints (8) and (9) are not only necessary for a valid conditional probability nearest-neighbour model, but are also sufficient; that is, ϕ, ψ_1 and ψ_2 are arbitrary functions. Equation (10) may be seen as an explicit factorization of the form shown to exist by Brook (1964) and extends to continuous distributions provided we interpret $p(.\,|\,...)$ as a conditional density. The results may easily be adapted to three or more dimensions.

3. BINARY DATA: THE AUTO-LOGISTIC MODEL

In the case of binary data, we may rewrite (8) as

$$f(x\,|\,t,u,v,w) = x\{\alpha+\beta_1(t+u)+\beta_2(v+w)\},$$

since $x^2 \equiv x$. Thus,

$$p(x \mid t, u, v, w) = \frac{\exp\{x[\alpha + \beta_1(t+u) + \beta_2(v+w)]\}}{1 + \exp\{\alpha + \beta_1(t+u) + \beta_2(v+w)\}}, \tag{11}$$

where α, β_1 and β_2 are arbitrary real numbers. The similarity of the model specified by equation (11) to a classical logistic model (see, for example, Cox, 1970, Chapter 1) suggests that we may reasonably describe the present model as *auto*-logistic, emphasizing that here the explanatory variables are the surrounding array variables themselves.

The joint probability distribution of the inner array variables is given by equation (10) provided that $\mathbf{x}_B = \mathbf{0}$. That is,

$$P\{\mathbf{x}_I \mid \mathbf{x}_B = \mathbf{0}\} = \frac{\exp\{\sum(\alpha + \beta_1 x_{i-1,j} + \beta_2 x_{i,j-1}) x_{i,j}\}}{C(\alpha, \beta_1, \beta_2)}, \tag{12}$$

where the summation extends over all $(i,j) \in I$ and $C(\alpha, \beta_1, \beta_2)$ is a normalizing function, dependent also upon the dimensions of the array. For more general boundary conditions, the joint distribution will differ slightly from equation (12), due to end effects. However, the results (11) and (12) do not depend upon the perimeter of the array being rectangular and hold for any closed boundary. We may note, incidentally, that the statistics $\sum x_{i,j}$, $\sum x_{i,j} x_{i-1,j}$ and $\sum x_{i,j} x_{i,j-1}$ are jointly sufficient for α, β_1 and β_2, and that, if they have joint moment generating function $M(\theta, \phi_1, \phi_2)$, that is,

$$M(\theta, \phi_1, \phi_2) \equiv E[\exp\textstyle\sum\{(\theta + \phi_1 X_{i-1,j} + \phi_2 X_{i,j-1}) X_{i,j}\}],$$

then

$$M(\theta, \phi_1, \phi_2) = C(\alpha + \theta, \beta_1 + \phi_1, \beta_2 + \phi_2)/C(\alpha, \beta_1, \beta_2),$$

demonstrating the importance of the normalizing function C.

The distribution (12) corresponds to the classical Gibbs equilibrium distribution arising in the general form of the Ising model of ferromagnetism for a two-dimensional rectangular lattice (see, for example, Newell and Montroll, 1953). In that context, $\alpha + \beta_1 + \beta_2 \neq 0$ allows for an external magnetic field, whilst $\beta_1 \neq \beta_2$ allows for asymmetry in the potential interaction energy. $C(\alpha, \beta_1, \beta_2)$ is called the partition function. The result that any Ising model is a conditional probability nearest-neighbour model is, of course, easily shown directly. However, the point of interest, here, is that we have shown that the converse is also true and that, therefore, the two definitions are equivalent. This result may easily be extended to three or more dimensions. Whittle (1963) noted, in support of his own approach to nearest-neighbour models, that the natural specification of a system in statistical mechanics is in terms of a joint distribution formula rather than as a conditional probability model. However, we now find that it is exactly the class of conditional probability nearest-neighbour models which have been at the centre of interest in the study of physical lattice systems in the past. This may suggest that it is natural, in our context, to consider the conditional probability models to be the more basic starting point, rather as one would consider a first-order Markov chain to be a more basic starting point than one would a second-order chain (the analogy is, in fact, extremely close).

4. A Spatial–Temporal Approach to the Auto-logistic Model

In most practical situations, the available (binary) data will consist of a single observed array; however, this array may usually be interpreted as an observation on

a process developing through time. In this context, Bartlett (1971) has suggested a class of spatial–temporal models which have the advantage, over purely spatial ones, that time plays its classical role of providing a direction for causal relation. It is relevant to enquire whether a nearest-neighbour spatial distribution may reasonably arise from the spatial–temporal models.

Let $x_{i,j,u}$ denote the value (0 or 1) of the random variable $X_{i,j,u}$ at the (i,j) site of the array at time u and consider the Markov (in time) model

$$P\{X_{r,s,t+\Delta t} = x \mid X_{r,s,t} = x \text{ and all other values up to time } t+\Delta t, \text{ excluding } x_{r,s,t+\Delta t}\}$$
$$= [1+\Delta t \exp\{\alpha_x + \beta_{x,1}(x_{r-1,s,t}+x_{r+1,s,t}) + \beta_{x,2}(x_{r,s-1,t}+x_{r,s+1,t})\}]^{-1} + o(\Delta t) \quad (13)$$

for $x = 0$ or 1 and all r, s and t. That is, (13) gives the probability that the (r,s) element remains unchanged at 0 or unchanged at 1 in the time interval $(t, t+\Delta t]$ given all other values at or before time $t+\Delta t$. We may, for example, consider (13) as a simple model for the spread of recurrent infection in an array of plants; for 0, 1 variables a logistic model is the natural choice and (13) provides an obvious extension to include the time element. Although the explanatory variables are again array variables, the inclusion of the time element has given rise to a classical logistic model. Different sets of parameters, $(\alpha_0, \beta_{01}, \beta_{02})$ and $(\alpha_1, \beta_{11}, \beta_{12})$, relate, respectively, to the conditional probabilities of a healthy (0) plant becoming infected (1) and vice versa, giving a considerable degree of flexibility. If the right-hand side of (13) is re-written as

$$1 - \Delta t \exp\{\alpha_x + \beta_{x,1}(x_{r-1,s,t}+x_{r+1,s,t}) + \beta_{x,2}(x_{r,s-1,t}+x_{r,s+1,t})\} + o(\Delta t),$$

it may be considered as a special case of Bartlett's spatial–temporal class of models and hence it may be shown that the stationary distribution is auto-logistic with (cf. equation (11)),

$$\alpha = \alpha_0 - \alpha_1, \quad \beta_1 = \beta_{01} - \beta_{11} \quad \text{and} \quad \beta_2 = \beta_{02} - \beta_{12}.$$

The existence of a stationary distribution is ensured, in general, since the matrix of spatial–temporal conditional transition rates has non-zero elements.

There are, of course, many practical situations where the above argument is not relevant. In particular, we may be interested in the distribution on an array when no stationarity assumption can be made. For example, if infected plants do not recover then no stationary distribution exists whilst, even in a case where a stationary distribution does exist, we may be concerned with the distribution on a single array in the transient stage. In many such cases where a bilateral spatial model is required, it is still suggested that as the first approximation, the auto-logistic model is intuitively reasonable, though no formal spatial–temporal justification is given.

5. Linear Nearest-neighbour Systems

As regards problems of inference, the class of nearest-neighbour models specified by equation (10) is, in general, extremely awkward to handle, even if we make an assumption of stationarity. For example, in the case of binary systems, analytical results concerning the auto-correlation structure of the model (11) have progressed little since the work of Onsager (1944) dealing with the entirely symmetric case when $P(1) = P(0)$ and $\beta_1 = \beta_2$. On the other hand, if we have a stationary system satisfying the linear relation

$$E(X_{i,j} \mid \text{all other values}) = \xi + \eta_1(x_{i-1,j}+x_{i+1,j}) + \eta_2(x_{i,j-1}+x_{i,j+1}), \quad (14)$$

say, then this gives rise to a spectrum proportional to

$$(1-2\eta_1\cos\omega_1-2\eta_2\cos\omega_2)^{-1}, \tag{15}$$

in the usual notation, upon which goodness-of-fit tests may be based and from which the correlation structure of the process may easily be obtained numerically. It is therefore of interest to identify nearest-neighbour systems having the property (14).

Before considering this problem in greater detail we make the following remark concerning binary systems. For such a system, we have

$$E(X_{i,j}|\text{all other values})\equiv P\{X_{i,j}=1\,|\,\text{all other values}\}$$

and, since the auto-logistic model (11) does not reduce to a linear form, unless β_1 or $\beta_2=0$, the relationship (14) must be invalid for *any* binary system with β_1 and β_2 non-zero. Whilst, even in a classical regression situation, linear binary models are not usually strictly appropriate, they often provide useful approximations which are comparatively straightforward to deal with. However, in the present situation, such a linear model is not even mathematically self-consistent. Incidentally, the spectrum (15) *is* itself valid for a binary system and Bartlett and Besag (1969) have considered goodness-of-fit tests based on one-sided approximations to it, but it is stressed that this spectrum cannot arise from a binary nearest-neighbour process.

In examining linear nearest-neighbour systems we make the following restrictive assumption. We suppose that the probability distribution (or density function) of $X_{i,j}$, given all other site values, depends only upon $x_{i-1,j}+x_{i+1,j}$ and $x_{i,j-1}+x_{i,j+1}$. This may often seem intuitively reasonable as a first approximation to the real process, particularly in the case of continuous variables. Thus, for feasible site values x, t, u, v and w, we let

$$p(x|\,t,u,v,w)\equiv\exp\{A(x,t+u,v+w)\}. \tag{16}$$

Then using equation (8), we have

$$\begin{aligned}A(x,t,v)-A(0,t,v)&\equiv x\{\phi(x)+t\psi_1(x,t)+v\psi_2(x,v)\}\\&\equiv x\{\phi(x)+t\psi_1(t,x)+v\psi_2(v,x)\}\end{aligned}$$

so that

$$\begin{aligned}f(x|\,t,u,v,w)&\equiv x\{\phi(x)+(t+u)\,\psi_1(x,t+u)+(v+w)\,\psi_2(x,v+w)\}\\&\equiv x\{\phi(x)+t\psi_1(x,t)+u\psi_1(x,u)+v\psi_2(x,v)+w\psi_2(x,w)\}\end{aligned}$$

and, hence, it follows that

$$f(x|\,t,u,v,w)\equiv x\{\phi(x;\,\beta_1,\beta_2)+\beta_1(t+u)+\beta_2(v+w)\},$$

where ϕ is an arbitrary function and β_1 and β_2 are arbitrary constants. This gives the necessary and sufficient condition for a nearest-neighbour model to have a conditional distribution of the form (16).

Linear models are generally considered more appropriate for continuous variables than for discrete ones and it happens that in the former case we may derive the moment generating function corresponding to the conditional density function (16). We let

$$M(\theta)\equiv E\{\exp(X\theta)|\,t,u,v,w\},\quad \lambda_1\equiv t+u,\quad \lambda_2\equiv v+w$$

and

$$C(\lambda_1, \lambda_2) \equiv \sum_x \exp\left[x\{\phi(x;\, \beta_1, \beta_2) + \beta_1 \lambda_1 + \beta_2 \lambda_2\}\right].$$

Then

$$M(\beta_1 \theta) = C(\lambda_1 + \theta, \lambda_2)/C(\lambda_1, \lambda_2)$$

and

$$\beta_1 E(X|\, t, u, v, w) = \partial \log C(\lambda_1, \lambda_2)/\partial \lambda_1 = \beta_1(\xi + \eta_1 \lambda_1 + \eta_2 \lambda_2)$$

for a linear model. Hence,

$$\log M(\theta) = (\xi + \eta_1 \lambda_1 + \eta_2 \lambda_2)\theta + (\eta_1 \theta^2/2\beta_1)$$

with $\eta_1 \beta_2 = \eta_2 \beta_1$. Thus, for continuous array variables, in postulating a conditional probability nearest-neighbour model satisfying (14) and (16), the only possibility, subject to mild regularity conditions, is that the conditional density of $X_{i,j}$ is normal and that therefore, from equation (10), the joint distribution of the inner array variables is multivariate normal. For discrete variables, we cannot use the same technique to derive $M(\theta)$ since λ_1, for example, can only take a countable set of values, but it is conjectured that linear models in two (or more) dimensions will not occur in practice. For a discussion of quasi-linearization of binary systems, see Bartlett (1971).

Finally, the present section may to some extent be related to the "harnesses" of Hammersley (1967). However, in this context, we make the following remark. Hammersley defines a one-sided harness (in one dimension) as a process satisfying the stochastic equation

$$X_i = X_{i-1} + \epsilon_i, \tag{17}$$

where the ϵ's are independent error variables having zero mean and variance σ^2. This definition is motivated by considering a system for which one might suppose that

$$E(X_i|\, \text{all predecessors}) = X_{i-1}. \tag{18}$$

The relations (17) and (18) are, of course, consistent with each other. However, the definition of a central harness as a process satisfying

$$X_i = \tfrac{1}{2}(X_{i-1} + X_{i+1}) + \epsilon_i, \tag{19}$$

where, again, the ϵ's are independent error variables having zero mean and variance σ^2, is *not* consistent with the conditional expectation statement

$$E(X_i|\, \text{all others}) = \tfrac{1}{2}(X_{i-1} + X_{i+1}). \tag{20}$$

Whilst the harness given by (19) obeys a second-order equation, the relation (20) is more compatible with a first-order scheme.

ACKNOWLEDGEMENT

I am greatly indebted to Professor M. S. Bartlett for much helpful discussion.

REFERENCES

BARTLETT, M. S. (1966). *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.

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

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

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.

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

HAMMERSLEY, J. M. (1967). Harnesses. *Proc. 5th Berkeley Symp.*, **3**, 89–117.

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

NEWELL, G. F. and MONTROLL, E. W. (1953). On the theory of the Ising model of ferromagnetism. *Rev. Mod. Phys.*, **25**, 353–389.

ONSAGER, L. (1944). Crystal statistics, I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, **65**, 117–149.

WHITTLE, P. (1963). Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, **40**, 974–994.