
PREAMBLE

*Knowledge is of two kinds. We know a subject ourselves,
or we know where we can find information upon it.*

S. Johnson, **Boswell's Life of Johnson**

17th century: discovery by Isaac Newton of the binomial theorem-18th century: Bernoulli's publication of the first book devoted to probability theory-19th century: development of the least squares principle by Gauss and Legendre, and emergence of the subject of statistics-20th century: Kendall, Fisher, Pearson, and the emergence of multivariate statistics-and, in our own day the appearance of spatial statistics. In many of his publications, Ripley has imparted knowledge to the spatial statistics audience about Gibbsian interaction models, one of the greatest successes of this new field. This paper furnishes a history of their development, with Ripley this time imparting knowledge to the spatial statistics audience concerning where in the literature different advances can be found. The purpose of this paper is twofold, namely, (1) to document the evolution of Gibbsian interaction models, and (2) to provide examples of their use. Thus, not only have the horizons of statistical science expanded, throughout the centuries, but they also continue to expand. Martin corroborates this latter contention by noting that much of the progress chronicled by Ripley is of very recent origin.

The Editor



Gibbsian Interaction Models

Brian D. Ripley *

Department of Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK.

Overview: Gibbsian interaction models, often referred to as *Markov random fields*, have been one of the greatest successes of Spatial Statistics. They encompass conditional spatial autoregressions and a wide class of models for interacting point patterns. Originally borrowed from ideas in statistical physics, they have been applied to both regional measurements (lattice and non-lattice) and to point processes. The distinctive feature of their development in statistics has been the emphasis on statistical inference, in fitting models and estimating parameters. Over two decades a very satisfactory methodology has been developed, but this has never been documented in simple terms. This paper aims to document this progress and to provide some examples of the use of the methodology.

1. Introduction

It is common to classify the underlying forces producing a spatial pattern as either exogenous, producing *heterogeneity* or internal, producing *interaction*. The two effects tend to occur in opposite directions. For example, with point patterns (Figure 1), exogenous forces tend to produce 'patchy' patterns which are similar to those produced by clustering, whereas interaction tends to produce patterns which are more 'regular' than one would expect to happen by chance. This comment is not universal, and it is possible to produce clustered patterns by the methods described here. However, their *raison d'être* is to give us methods to describe interacting systems. These can be either systems of points such as Figure 1, in which case the interaction is reflected in their spatial positions, or systems of regions, where the interaction is reflected in their values. Geographical examples include the locations of market towns and supermarkets as point patterns (Glass & Tobler, 1971; Rogers, 1974; Ripley, 1979b) and Robert Haining's study of petrol pricing in Sheffield (Haining, 1983; Bennett & Haining, 1985).

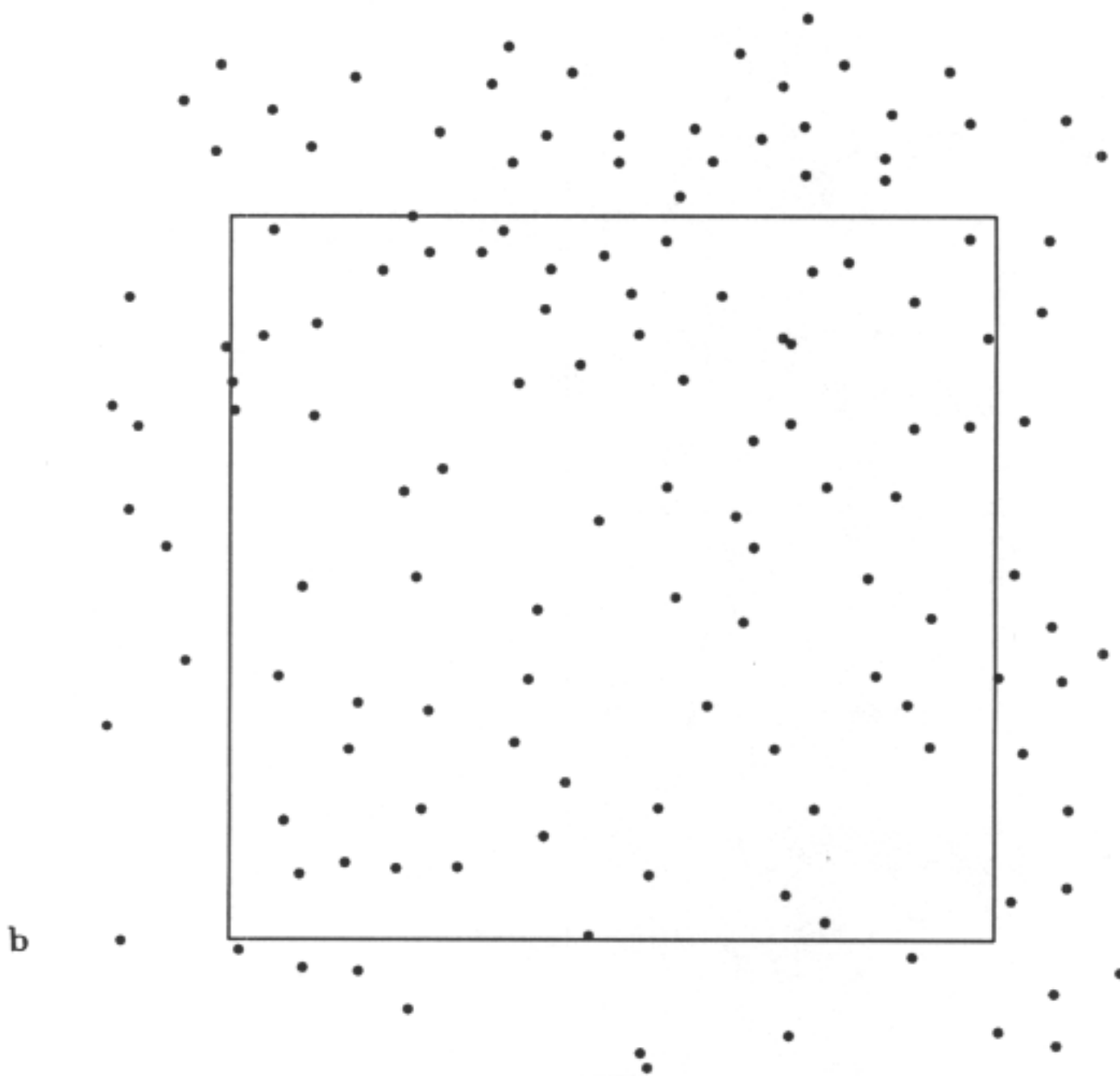
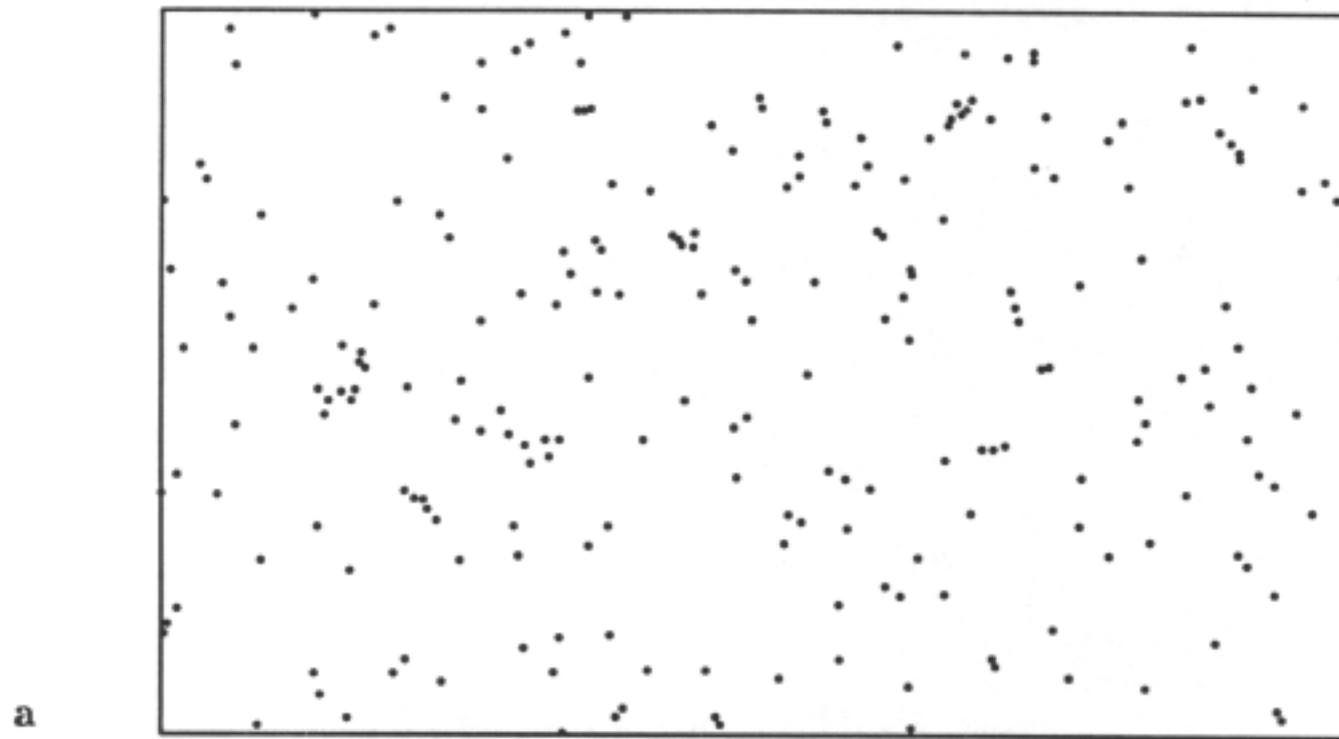
We will take for granted that studies of interaction are appropriate to at least *some* problems in geography; some of the issues here are discussed by other authors in this volume. It is worth noting that it is very difficult to consider exogenous forces and interaction simultaneously, especially for point patterns. One reason is that the interaction may be density-dependent. Suppose we studied the pattern of nests of small songbirds in a wood. Then we will expect territorial behaviour to lead to interaction and some regularity of spacing, *but* in parts of the wood where conditions are especially favourable the territories will be smaller and interaction stronger. Something similar might be expected to happen in the petrol-price competition study.

Gibbsian interaction models subsume what are often known as *Markov random fields*, in which sites or points interact if and only if they are neighbours. Of course the concept of 'neighbour' needs careful definition, but this has been extensively studied in a geographical

* After October 1, 1990, Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.

Figure 1.

Heterogeneity (a) and *interaction* (b) in point patterns. (a) shows 232 meat stores in a 6.7744.158 km ward of Tokyo (Okabe & Miti, 1984). (b) shows towns on the Spanish plain, with a 40 mile square containing 70 of the 136 points (after Glass & Tobler, 1971).



context as part of the study of spatial autocorrelation. Formally, sites form a *graph*, with an edge joining every pair of sites which is a neighbour, giving a structure such as Figure 2. For definiteness, suppose we have N sites as in Figure 2, and a random variable X_s defined at each site. Then a (joint) probability distribution for the set of random variables, $P(X_1, \dots, X_N)$, defines a Markov random field if for each site s , $P(X_s | X_t, t \neq s)$, the conditional distribution of the variable at s given the values at all other sites, depends only on the values at sites which are neighbours of s . This is a minor restriction to local interactions. The fame of Markov random fields comes from the so-called Clifford-Hammersley theorem, which states that for a Markov random field, under a *positivity* condition,

$$P(X_1, \dots, X_N) \propto \prod_{\text{cliques}} \phi(X_{i_1}, \dots, X_{i_m}) \quad (1.1)$$

Here a 'clique' is a maximally connected component of the graph, that is a set of sites all of which are neighbours of all others, and which cannot be expanded (so the term is sociologically appropriate), and ϕ is an *interaction* function depending on the variables at all sites in the clique. This result was stated by Clifford and Hammersley in the early 1970's, and a long proof was in private circulation. They never published, and it became clear that the result had simpler proofs (*e. g.* Besag, 1974; Preston, 1974, 1976). The positivity condition essentially precludes processes for which certain combinations of values of (X_1, \dots, X_N) are excluded in a way inconsistent with (1.1).

The result (1.1) is often stated in an equivalent form. Let $V = -\log_e \phi$. Then

$$P(X_1, \dots, X_N) \propto \exp\left[-\sum_{\text{cliques}} V(X_{i_1}, \dots, X_{i_m})\right]$$

In this form V is known as a *potential*, as this is the form in which it arises in statistical physics. The constant implied by the proportional sign is in most cases impossible to express in a closed form.

The importance of (1.1) is *not* the theorem, which guarantees its existence, but that it suggests a way of defining the probability distribution of interacting systems. In practice the interaction function ϕ is taken to be one for all but very small cliques. The simplest case is to consider cliques of only one site (a site being a neighbour of itself); we find

$$P(X_1, \dots, X_N) \propto \prod_{\text{sites}} \phi(X_s)$$

so the random variables are independent. The first non-trivial case is to consider only pairs of sites which are neighbours. Expression (1.1) then becomes

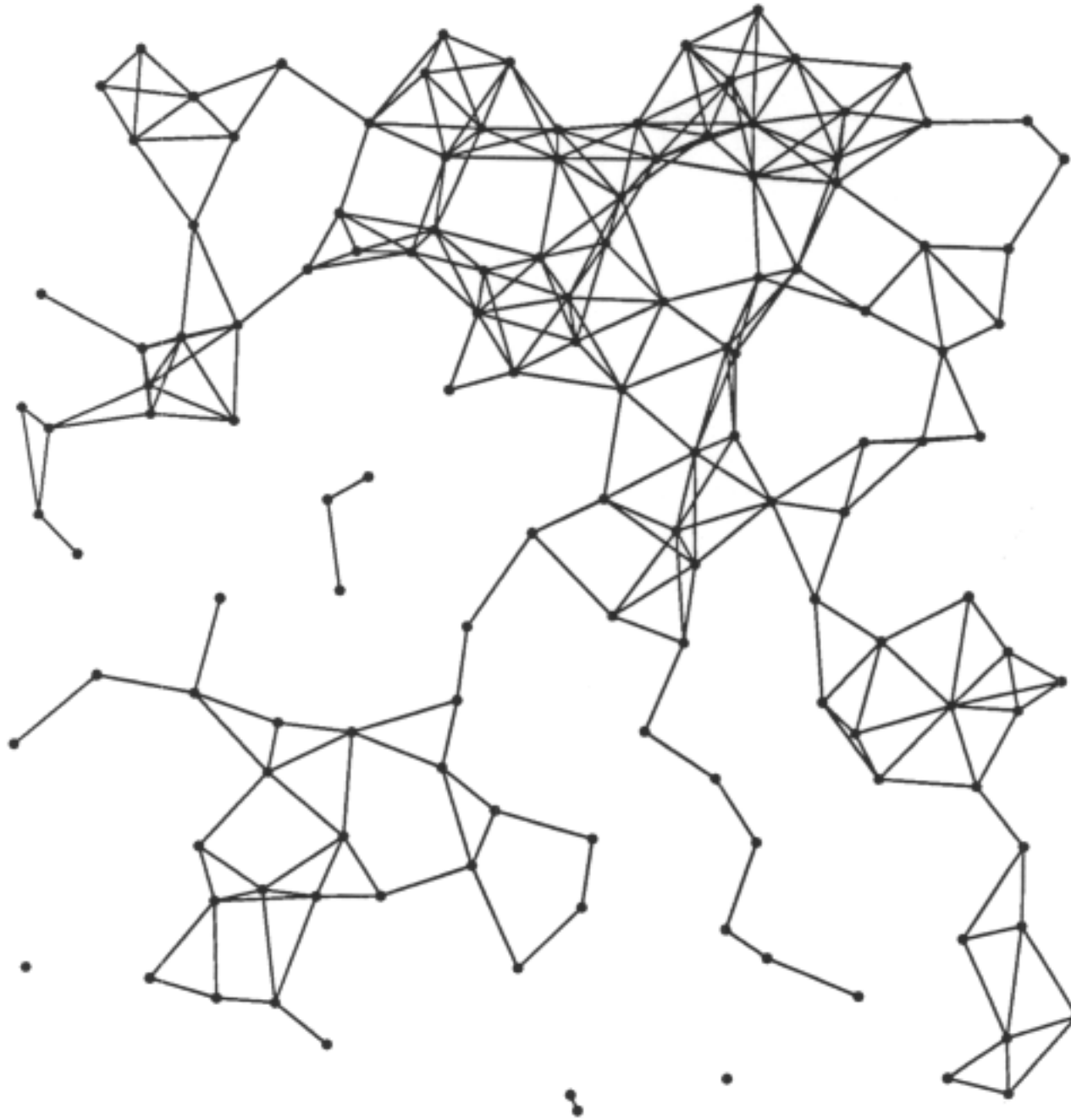
$$P(X_1, \dots, X_N) \propto \prod_{s \text{ nhbr of } t} \phi(X_s, X_t) \quad (1.2)$$

and almost all Markov random fields ever used are of this form. (Rarely, three-way interactions are considered.) It seems unnecessary to insist that sites interact only if they are neighbours, but rather we should allow a gradual diminution of interaction with distance. We can easily relax (1.2) to

$$P(X_1, \dots, X_N) \propto \prod_{s,t} \phi(X_s, X_t) \quad (1.3)$$

Figure 2.

An illustration of a graph joining the Spanish towns less than 6 miles apart.



called *pairwise interaction processes*, and these are the principal subject of this paper, as well as much recent work.

The term 'Gibbsian interaction' processes comes from statistical physics, where such models have been used for nearly a century to describe the behaviour of gases. Gravitational and electrostatic forces provide two physical examples in which interactions between objects are purely pairwise. The pairwise assumption makes sense in many other fields as well. Animals in defending territories have only pairwise fights, and this may be supposed to apply also to early human settlements. Nevertheless there are examples in which other

assumptions may be attractive. In competition studies it may be the minimum price in the town which is important, and the sizes of market towns could be governed by some notion of 'catchment area.' Thus we will define the general notion of a Gibbsian interaction process as

$$P(X_1, \dots, X_N) \propto \prod_{\text{subsets}} \phi(X_{i_1}, \dots, X_{i_m}) \quad (1.4)$$

with the presumption that the subsets of sites and the interaction function are chosen in some simple way.

An important alternative way to look at (1.4) is via the conditional distributions of the variable at each site given the values at all other sites. From (1.4) we have

$$P(X_s | X_t, t \neq s) \propto \prod \phi(X_{i_1}, \dots, X_{i_m}) \quad (1.5)$$

the product being over subsets containing s , and (1.3) becomes

$$P(X_s | X_t, t \neq s) \propto \prod_{t \text{ nhbr } s} \phi(X_s, X_t) \quad (1.6)$$

The function ϕ must be symmetric, that is, the influence of site s on site t must be the same as that of t on s .

Closely analogous ideas are used to define Gibbsian and pairwise interaction processes for point patterns. The variables X_s are replaced by the locations \mathbf{x}_s of the points, and probabilities must be re-interpreted as densities. Indeed, it is the technical details needed to make the simple formulae rigorously correct which makes the point process literature inaccessible to all but the highly mathematically trained. This is unfortunately necessary, but the treatment given in later sections is as non-mathematical as possible (and so should not be relied on as totally accurate, in that technical conditions on probability density functions have been omitted).

The distribution of market towns in a plain provides a nice example of an interacting point process. We could postulate a pairwise interaction, with strength decreasing with distance, or an interaction dependent on catchment area. Examples are shown in §4.

The Reference list at the end of the paper serves as a Bibliography to cover the papers of major methodological importance. Clearly from a biased view, I use my two monographs (Ripley, 1981, 1988) as reference sources. '*Spatial Statistics*' is meant for a general readership (but has been thought terse and mathematically tough) whereas '*Statistical Inference for Spatial Processes*' covers recent developments and their impact on statistical theory. It is intended to be mathematically tough, but is the only reference source for much of the material sketched here.

2. Regional processes

Interaction processes and Markov random fields were originally developed for lattice processes by Bartlett and Besag (then Bartlett's research assistant); Bartlett (1975) reports on those developments, which were motivated by looking at plausible space-time processes at just one time point. Much of the complication arises from the assumption of *stationary* lattice processes, that is processes defined throughout the two-dimensional lattice $\{(x, y) \mid x, y \text{ integer}\}$

in a translation-invariant way. My own belief is that processes on regular lattices are rarely useful except in connection with man-made phenomena; this includes their recent use in image analysis in which the pixels are man-made.

Besag (1975) gave an important insight that these processes do not need to be defined on a (regular) lattice, and it was in his form that they are given in our §1. If the random variables (X_1, \dots, X_n) have a normal distribution (jointly), this is defined by the vector of their means (μ_1, \dots, μ_N) and the covariance matrix $\Sigma = (\sigma_{ij})$ where $\sigma_{ij} = \text{cov}(X_i, X_j)$. Then the conditional distribution of X_i given the remaining sites is normal with mean $\mu_i + \sum_{j \neq i} b_j (X_j - \mu_j)$ and variance $\Sigma_{kk} - \mathbf{a}^T A^{-1} \mathbf{a}$, where A is the covariance matrix of the remaining sites, \mathbf{a} is the vector of covariances of site i with the remaining sites, and $\mathbf{b} = A^{-1} \mathbf{a}$. Only the mean depends on the values at the remaining sites, so we have a Markov random field if and only if $b_j = 0$ unless j is a neighbour of i . This has to be true for all i , and reduces to the assumption that the elements Σ_{ij}^{-1} are non-zero only if i and j are neighbours. Note that since only covariances need to be considered, Gaussian processes are always pairwise interaction processes.

Markov random fields are normally specified by giving the conditional distributions of X_s given the values at the remaining sites. In the Gaussian case we have found that the means are linear, so

$$E(X_s | X_t, t \neq s) = \mu_s + \sum_{t \neq s} B_{st}(X_t - \mu_t) \quad (2.1a)$$

$$\text{var}(X_s | X_t, t \neq s) = \kappa_s \quad (2.1b)$$

say, where the matrix B and numbers (κ_s) are defined by the expressions in the previous paragraph and $B_{ss} = 0$. It is important to note that we can regard (2.1) as the *definition* of the process, which in this form is called a *conditional autoregression*. Elementary calculations (Ripley, 1981, p.89) give the mean and covariance matrix of (X_s) as (μ_s) and $\Sigma = (I - B)^{-1} \text{diag}(\kappa_s)$ respectively. A covariance matrix must be positive-definite and symmetric, which imposes an awkward condition on B . In particular, we must have $B_{st}\kappa_t = B_{ts}\kappa_s$ for all distinct pairs of sites s, t . The easiest case is when the conditional variances (κ_s) are all equal, in which case the necessary and sufficient condition is that the matrix $(I - B)$ be positive-definite symmetric. I have given these conditions in some detail because they are often ignored in the literature. *Not every choice of connection matrix B in (2.1a) defines a valid conditional autoregression.*

These conditional autoregressions should not be confused with another class of spatial processes used slightly earlier, now called *simultaneous autoregressions*. These are defined by

$$X_s - \mu_s - \sum_{t \neq s} B_{st}(X_t - \mu_t) = \epsilon_s \quad (2.2)$$

where the ϵ_s are *independent* normally distributed random variables.

The Gaussian case is by far the most important, but Besag (1974) showed that there were three other interesting cases of simple Gibbsian interaction processes, the auto-logistic, auto-binomial and auto-Poisson models, with the auto-logistic being a special case of the

auto-binomial. Their specifications are, for the auto-binomial

$$X_s \sim \text{binomial}(n, p_s) \\ \log(p_s/(1-p_s)) = \alpha_s + \sum_{t \text{ nhbr } s} B_{st} X_t$$

and for the auto-Poisson

$$X_s \sim \text{Poisson}(\mu_s) \\ \log(\mu_s) = \alpha_s + \sum_{t \text{ nhbr } s} B_{st} X_t$$

where for the auto-Poisson $B_{st} \leq 0$ so only competition between sites is allowed. The auto-logistic process is the special case $n = 1$ of the auto-binomial. They correspond to the special case of

$$\phi(X_s, X_t) = B_{st} X_s X_t$$

in (1.3) or (1.6), and Besag (1974) showed that these were the only discrete distributions for which this linear form is allowable. In all these processes the crucial parameter is the matrix B governing the interactions. Knowledge of the subject is supposed to allow us to specify B and (μ_i) [or (α_i)] up to a few parameters, with the rest to be estimated from the data. For example, in studies of spatial autocorrelation (*e. g.* Cliff & Ord, 1981) it is common-place to specify a matrix W of connection weights. If this is symmetric, we might take $B = \rho W$ for small positive ρ to be estimated from the data. (In a weak sense, tests of spatial autocorrelation are tests of $\rho > 0$ vs $\rho = 0$ in this model.) In studies of plant competition, Mead (1971) based the weight function on the Voronoi polygons (also termed Dirichlet cells or Thiessen polygons) of the plants (representing 'catchment area'), and other suggestions are in Ord (1975) (although note that both these authors worked with simultaneous autoregressions).

2.1. Parameter estimation

In the early 1970's when these models were first proposed, people shied away from maximum likelihood estimation of the parameters in B and in the mean vector. The computational difficulty comes from the normalizing constant which we have conveniently ignored up to now. For a conditional autoregression the joint probability density of (X_1, \dots, X_N) is (from its specification as a multivariate normal distribution)

$$\frac{|I - B|^{1/2}}{(2\pi\kappa)^{N/2}} \exp\left[-\frac{1}{2\kappa}(\mathbf{X} - \boldsymbol{\mu})^T(I - B)(\mathbf{X} - \boldsymbol{\mu})\right]$$

so minus twice the log likelihood is given by

$$N \ln 2\pi\kappa - \ln |I - B| + (\mathbf{X} - \boldsymbol{\mu})^T(I - B)(\mathbf{X} - \boldsymbol{\mu})/\kappa \quad (2.3)$$

This is easily maximized over κ to give

$$\hat{\kappa} = N^{-1}(\mathbf{X} - \boldsymbol{\mu})^T(I - B)(\mathbf{X} - \boldsymbol{\mu})$$

so parameters in B are chosen to minimize

$$N \ln \hat{\kappa} - \ln |I - B| \quad (2.4)$$

The perceived difficulty is the determinant $|I - B|$, but for realistically sized problems it is quite straightforward to minimize (2.4) numerically. In the special case $B = \rho W$ Ord (1975) pointed out that we can exploit

$$|I - \rho W| = \prod (1 - \rho \lambda_i)$$

the product being over the eigenvalues of W . For large systems on a regular lattice there are also asymptotic approximations to $|I - B|$ dating from Whittle (1954).

However, we should ask whether maximum likelihood is desirable *per se*. The classical justifications of maximum likelihood in statistics are asymptotic, for an infinite sequence of identically distributed *independent* observations. This is not a natural asymptotic regime in spatial statistics, although Mardia & Marshall (1984) have proved that for a related type of asymptotics some classical results hold. In general, we do not even know if (2.3) is a well-behaved likelihood function. Ripley (1988, Chapter 2) demonstrates that it can fail to be concave, but that with a known mean vector and $B = \rho W$ it is unimodal. (Much of the statistical theory of likelihood functions depends on concavity.) Thus although maximum likelihood estimation is almost always possible computationally, we must not assume that it is necessarily statistically desirable. The literature on this point is often misleading or plain wrong. [For example, Upton & Fingleton (1985, p. 284) quote classical results without comments on the lack of applicability of these results, and they are by no means alone.]

An alternative, *pseudo-likelihood* estimation, was introduced by Besag (1975). The pseudo-likelihood is defined as the product of the conditional densities $P(X_s | X_t, t \neq s)$, and is treated like a likelihood. There seems no simple explanation as to why this is a sensible idea, but it has proved to work well in practice. Besag (1975) sketched a proof that the pseudo-likelihood estimator would be consistent (converge to the true value) as the problem is increased, and an elegant general proof is given by Geman & Graffigne (1987). For a conditional autoregression we have

$$\ln PL = -\frac{N}{2} \ln(2\pi\kappa) - \frac{1}{2\kappa} \sum_s [(I - B)(\mathbf{X} - \boldsymbol{\mu})]_s^2$$

so pseudo-likelihood estimation amounts to least-squares fitting to the “residuals”

$$\eta_s = X_s - \mu_s \sum_{t \neq s} B_{st}(X_t - \mu_t)$$

Be warned that (2.1) cannot be treated like an ordinary regression, and that the residuals (η_s) will themselves be spatially autocorrelated, unlike the (ϵ_s) in the simultaneous autoregression (2.2).

Pseudo-likelihood methods have been very successful, and have largely supplanted all others. Older methods such as ‘coding’ for regular lattices (Besag, 1974) should have disappeared by now!

2.2. Applications

These processes have been used less widely than their early promise suggested, and there are probably more papers on theory than applications. Part of the cause is that most attention has been to regular lattices, and as we suggested earlier, these are inevitably man-made. Another problem is defining well the set of sites to be considered. Interacting systems just do not occur in closed boxes, and it is always necessary to consider the effects of interaction with outside sites. Unless the number of sites is very large, there will be a high proportion of sites near the edge, and so it may be perilous to ignore the effect of the outside world. One reasonably successful application of spatial lattice processes has been to agricultural field trials, in which the system is closed and carefully controlled. Even there the impact has been more in emphasising the design of the trials to minimize spatial variation in soil fertility than in methods based on spatial autoregressions, except perhaps as part of rescue attempts on poorly designed experiments.

It seems that applications in image analysis will be successful. There the systems considered are vast by the standards of the 1970's, so edge effects may safely be neglected. We have been using Gibbsian interaction processes with astronomical images of 1024×656 pixels (Ripley, 1990; Molina & Ripley, 1989).

3. Point processes

Point patterns consist of n point locations within a specified domain D , specified by their Cartesian coordinates $\{\mathbf{x}_i\} = \{(x_i, y_i)\}$. They differ from classic multivariate statistics in two ways. First, the number of points n is thought of as variable, but more importantly, the coordinates of the points have very little rôle, the emphasis being on the *configuration* of the points. The domain D has to be defined carefully, since edge-effects are important. There are two rather different cases. Either D can be a natural region such as an island, or it can be thought of as a window into a much larger region, such as the examples of towns on a plain and supermarkets mentioned in §1.

Some further examples of studies of point patterns in geography are drumlins in Northern Ireland (Hill, 1973; Upton & Fingleton, 1985, pp. 68-9; Figure 7), schools in Southampton (Pinder & Witherick, 1972; Upton & Fingleton, 1985, pp. 76-7) and retail establishments in a Tokyo district (Okabe & Miki, 1984; Figure 1a).

The model for no interaction between points is the Poisson process. The number of points in D , $N(D)$, has a Poisson distribution with mean μ , and the points are independently distributed over D with density function f . Then $N(A)$, the number of points within a sub-region A of D , has a Poisson distribution with mean $\mu \int_A f$. Further, the numbers of points in non-overlapping regions are independent. This means that we can consider a domain D in isolation, as with a Poisson processes there is no interaction with points outside D . We will normally consider *homogeneous* processes in which case the density f is uniform, and $\mu = \lambda \times \text{area}(D)$ where λ , the number of points per unit area, is an important parameter known as the *intensity*.

Gibbsian interaction processes are defined by stating how much more (or less) likely each configuration of points is than under a Poisson process. For example, pairwise interaction

processes have density (with respect to a Poisson process) of

$$p(\mathbf{x}_1, \dots, \mathbf{x}_n) \propto \prod_i b(\mathbf{x}_i) \prod_{i < j} h(\mathbf{x}_i, \mathbf{x}_j) \quad (3.1)$$

The first product corresponds to a change of the underlying spatial intensity of points from $\mu \times f$ to proportional to $b \times f$, but the interaction is expressed by the second product. Since we almost always regard interaction as *stationary* throughout space, we can take $h(\mathbf{x}, \mathbf{y})$ as a function only of $d(\mathbf{x}, \mathbf{y})$, the distance between the two points. Then h can be thought of as an *interaction function*, and examples are shown in Figure 3. More general interactions are possible, but have never been found necessary. (These processes were introduced by Ripley, 1977.) We will normally consider processes for which b and f are constant, so there is no heterogeneity in the model.

The simplest special cases are the so-called Strauss processes illustrated in Figures 3(a) and 3(b). The interaction function is defined by

$$h(d) = \begin{cases} c & \text{if } d = d(\mathbf{x}, \mathbf{y}) \leq R \\ 1 & \text{otherwise.} \end{cases} \quad (3.2)$$

introduced by Kelly & Ripley (1976) following earlier (incorrect) work of Strauss (1975). The case $c = 0$ is the 'hard-core' process in which points are prohibited from being closer than distance R apart. It can be thought of as being produced by sampling from a Poisson process and throwing out all patterns which violate this condition. For $0 < c < 1$ there is a disincentive to close pairs, the density being proportional to $c^{y(R)}$ where $y(R)$ denotes the number of pairs of points closer than R . For $c = 1$ we have a Poisson process, and no interaction. For $c > 1$ the process only exists in a rather pathological way. Close pairs are encouraged, and most realizations for fixed n will have a single 'clump' of points contained within a disk of diameter R . For variable n the process cannot even be defined. A similar effect occurs whenever $h(d) > 1$ for some distance d , although this can be counter-balanced by $h(d) = 0$ for $d < R$, small.

There are a number of obvious extensions to the Strauss process. If we think of the points as having circular territories of diameter R , the interaction might depend on the area of the overlap (given by the term $[\dots]$ as a proportion of the area of the territories),

$$\ln h(d) = \begin{cases} -\theta \left[1 - \frac{2}{\pi} \left\{ \frac{d}{R} \sqrt{1 - \frac{d^2}{R^2}} + \sin^{-1} \left(\frac{d}{R} \right) \right\} \right] & \text{if } d \leq R \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

(Penttinen, 1984). Another idea is to allow more than one step in the interaction function, the 'multi-scale' process,

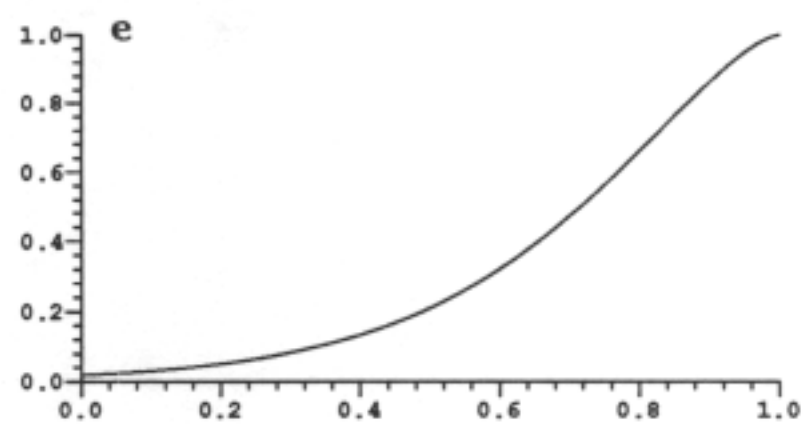
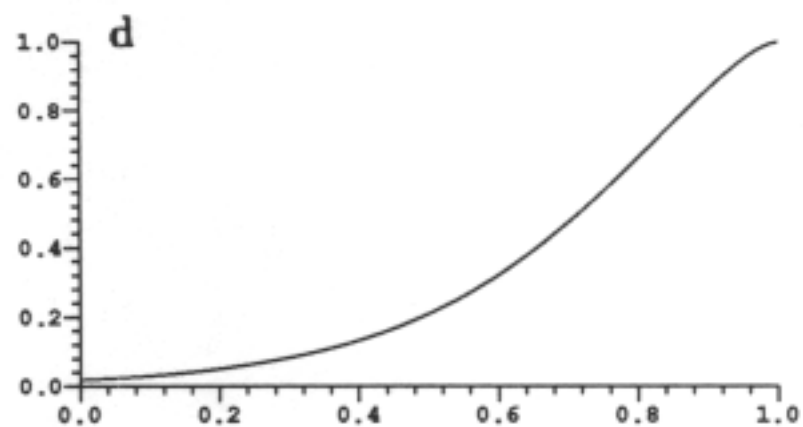
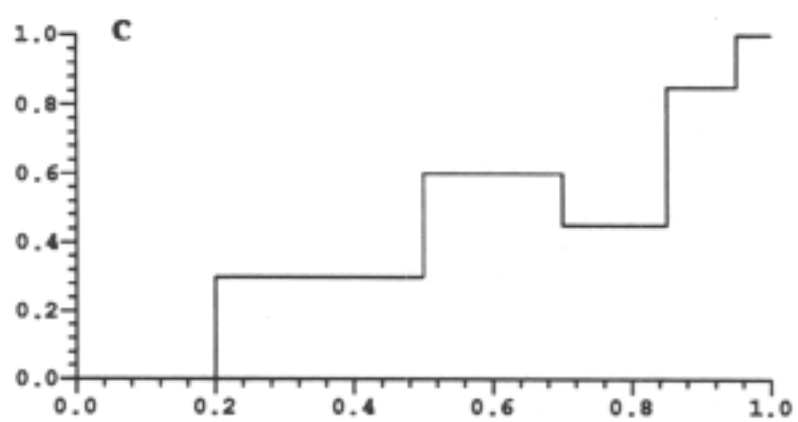
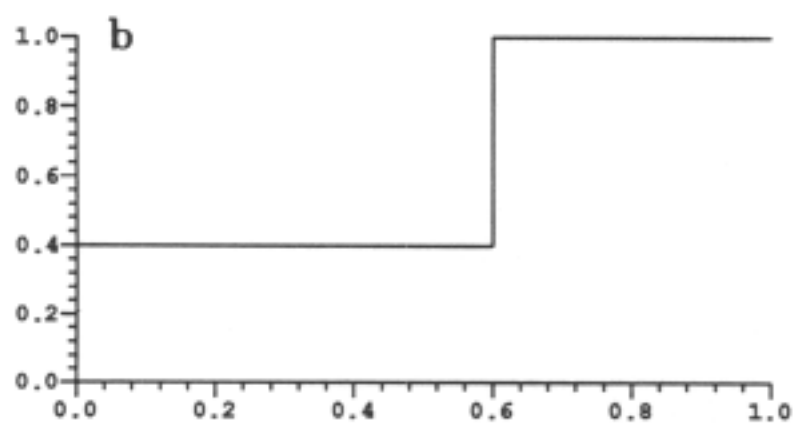
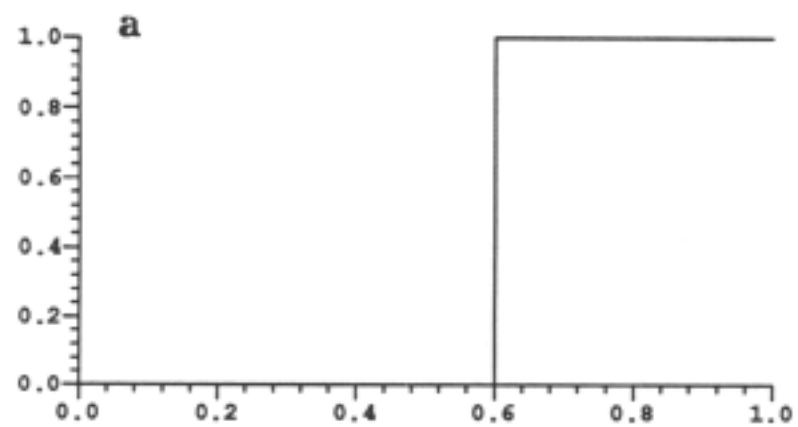
$$h(d) = \begin{cases} c_1 & \text{if } d \leq R_1 \\ c_2 & \text{if } R_1 < d \leq R_2 \\ c_3 & \text{if } R_2 < d \leq R_3 \\ \vdots & \vdots \\ 1 & \text{if } d > R_m \end{cases} \quad (3.4)$$

Pairwise interaction functions have been extensively considered in statistical physics, and other functional forms have been borrowed from that field, including the 'very-soft-core' potential (Ogata & Tanemura, 1984) illustrated in Figure 3(e),

$$h(d) = 1 - e^{-(d/\sigma)^2}$$

Figure 3.

Pairwise interaction functions h as a function of distance. (a) 'Hard-core' interaction. (b) Strauss process. (c) Multiscale process. (d) Overlap area (Penttinen, 1984) with $\theta = 4$. (e) 'Very-soft-core' (Ogata & Tanemura, 1984) with $\sigma = 0.5$.



Although some of these processes have been extensively discussed in statistical physics, the questions asked there are rather different. First, the size of the problem is very different, with of the order of 10^{23} points (molecules) so edge effects are indeed negligible. Further, the emphasis is the reverse of ours; the interaction function h is assumed known and properties of the process are required, whereas we have a sample of the process from which to learn about h .

There is an analogue for point processes of Markov random fields, the Markov point processes of Ripley & Kelly (1977).

3.1. Simulation

One very effective way to understand the definition and parameters of a statistical distribution is draw some samples from it. Indeed, in my first stochastic processes course this was given by David Kendall as a criterion for having defined a process properly, that you could simulate from it. The problem with constructive definitions such as that given earlier for the hard-core process is that they are 'in principle' only. One will eventually obtain a realization of a Poisson process with no pair of points closer than R (provided only that such a configuration can be packed into D), but the wait could be very long. Since simulation is such an effective way to understand these processes, it is important to know of reasonably efficient simulation methods.

All the known reasonably efficient simulation methods are iterative. At each stage a point is added or deleted, dependent on the configuration of the remaining points (Ripley, 1977, 1979a, 1987). The simplest case is when we prescribe the total number n of points. Then each step of the simulation consists of deleting a point and replacing it. The point to delete is chosen at random (so each of the n points is equally likely). Relabel the existing points as $\{\mathbf{x}_2, \dots, \mathbf{x}_n\}$, so the replacement point will be \mathbf{x}_1 . Then it has density proportional to

$$f_0(\mathbf{x}) = \prod_{i=2}^n h(\mathbf{x}, \mathbf{x}_i) \quad (3.5)$$

The unknown normalizing constant is once again an apparent problem, but this can be overcome by the use of *rejection sampling*, a well-known technique in simulation (Ripley, 1987). Consider the case of $h(d) \leq 1$ for all d . Then given $\{\mathbf{x}_2, \dots, \mathbf{x}_n\}$, sample \mathbf{x} from the underlying Poisson density f on the domain D , and accept it with probability $f_0(\mathbf{x})$. (The assumption on h ensures that $0 \leq f_0(\mathbf{x}) \leq 1$ and so it is indeed a probability. To accept with probability p , get your computer to generate a uniform random variable U and accept if $U < p$.) If the sample is not accepted, try again until one is.

If n is not fixed, a similar birth-and-death process is run, but additions and deletions no longer alternate. Details are given in Ripley (1977).

The theory says that this process produces samples whose distribution converges to that of the pairwise interaction process. The basic idea is due to Metropolis *et al.* (1953), the point process implementation to Ripley (1977, 1979a), the latter containing Fortran code. In practice the process has to be run for some time ($10n$ — $100n$ steps) to settle down from a reasonable starting pattern, then it can be sampled every $2n$ — $4n$ steps. For moderate n the process runs fast enough on a personal computer to be fascinating viewing. For example, on the towns data shown in Figure 1(b) simulations of the Strauss process change about 10

points per second.

3.2. Parameter estimation

Parameter estimation in pairwise-interaction point processes proved to be difficult for a decade, and an obstacle to their wider use. One can of course use trial-and-error methods, matching some aspect of the simulated patterns to the data. This was illustrated for the Spanish towns data in Ripley (1977), using my K -function to measure fit, and there are more extensive comparisons in Ripley (1988, Figure 4.2). Diggle & Gratton (1984) raised this idea to the status of a theory! However, it is both non-intuitive and rather computer-intensive.

Even the simplest cases have shown difficulties. The maximum likelihood estimate of R in a hard-core process is d_{\min} , the smallest distance between a pair of points. Since pairs closer than R cannot occur, we know $d_{\min} > R$ and hence the estimator is biased. (This is not unexpected, since the discontinuity in h at R makes this a non-regular likelihood problem.) Ripley & Silverman (1978) showed how to correct this estimator for bias. Even if R is known for the Strauss process, the maximum likelihood estimator of c is not straightforward. The density is

$$p(\mathbf{x}_1, \dots, \mathbf{x}_n) = a(b, c) b^n c^{y(R)}$$

and it is the unknown normalizing constant $a(b, c)$ which causes the trouble. To ease the notation, let us write $Y(R)$ for the random variable measuring the number of pairs of points within D closer than R . Then after some calculation we find that the maximum likelihood estimators are given by the solution in b, c to the equations

$$n = E_{b,c} N(D) \quad y(R) = E_{b,c} Y(R) \quad (3.6)$$

if the number of points is allowed to vary, otherwise b is irrelevant and c solves

$$y(R) = E_c Y(R) \quad (3.7)$$

Here $y(R)$ is the fixed number of observed close pairs, and the right-hand side is the expected number of pairs. It is intuitively obvious that this increases from zero for $c = 0$ up to the value for a Poisson process for $c = 1$. The latter depends on the shape of the domain D , but is known for several common shapes (Ripley, 1988, pp. 28-9); for small R it is approximately $n(n-1)\pi R^2/2$. It is perfectly possible for us to observe more close pairs than are the average for a Poisson process; in such a case we take $\hat{c} = 1$, although we ought to consider why we are fitting a process with interaction distance R !

Maximum likelihood estimation depends on our being able to calculate the means in (3.6). They are not known analytically, but as we shall see in §4, can be estimated by simulation. An alternative is to make approximations. When interactions are rare, we can derive the approximations (Ripley, 1988, pp. 56-7)

$$\hat{c} \approx y(R) \times \frac{2a}{n(n-1)\pi R^2} \quad (3.8)$$

for fixed n , and for variable n

$$\hat{b} \approx n/a, \quad \hat{c} \approx y(R) \times \frac{2a}{n^2\pi R^2}$$

where a denotes the area of D . These approximations correspond to fitting a straight line in c to $E_c Y(R)$, although they are derived by assuming that any point enters into just one close pair. More sophisticated approximations have been considered, notably by Ogata & Tanemura (1981, 1984), who borrowed approximations from statistical physics for 'non-ideal' gases. Details are given in Ripley (1988, pp. 59-62). For the Strauss process, they give an approximation to the log-likelihood as

$$y(R) \ln c - \frac{n^2}{2a}(c-1)\pi R^2 - \frac{5.79n^3(c-1)^3 R^4}{6a^2}$$

corresponding to

$$E_c Y(R) \approx \frac{n^2 \pi R^2}{2a} c [1 + 1.84(nR^2/a)(c-1)^2] \quad (3.9)$$

a cubic in c . (There is an error of a factor of 3 in the constants in Ripley, 1988.)

3.2.1. Pseudo-likelihood

We have seen that even in the simplest case maximum likelihood estimation causes difficulties, and just as in the regional data case, there is no reason to accord maximum likelihood methods special status from a theoretical viewpoint. This encourages us to consider pseudo-likelihood as it was so successful there. There are technical problems in the conditioning, but these can be overcome either by approximating by a lattice process (Besag, 1977; Besag, Milne & Zachary, 1982) or via the mathematically sophisticated theory of conditional intensities. The problem is that we want the probability of a point occurring at \mathbf{x} given the locations of the remaining points. This probability is zero, but we can consider

$$\frac{P(\text{point in a region } \Delta \text{ containing } \mathbf{x} \mid \text{other points})}{\text{area}(\Delta)}$$

for a small region Δ around \mathbf{x} . This is the conditional intensity $\lambda(\mathbf{x}; \mathbf{x}_1, \dots, \mathbf{x}_n)$. If \mathbf{x} is the location of one of the existing points it is omitted from the conditioning. Then the pseudo-likelihood is the product of the conditional intensity over all points in D , occupied or not, which leads to the log pseudo-likelihood as

$$\sum_1^n \ln \lambda(\mathbf{x}_i; \mathbf{x}_1, \dots, \mathbf{x}_n) - \int_D \lambda(\mathbf{x}; \mathbf{x}_1, \dots, \mathbf{x}_n) d\mathbf{x}$$

For a pairwise interaction process the conditional intensity is proportional to

$$b(\mathbf{x}) \prod_{\text{pts other than } \mathbf{x}} h(\mathbf{x}, \mathbf{x}_i)$$

so the log pseudo-likelihood is

$$\ln \left[\prod_i b(\mathbf{x}_i) \prod_{i,j} h(\mathbf{x}_i, \mathbf{x}_j) \right] - \int_D b(\mathbf{x}) \prod_i h(\mathbf{x}, \mathbf{x}_i) d\mathbf{x} \quad (3.10)$$

At first sight the first term is the likelihood (3.1) without the normalizing constant, but the second product is different in that each pair $\{i, j\}$ occurs twice. Nevertheless, pseudo-likelihood methods are very similar to maximum-likelihood ones, except that the normalizing

function is replaced by much simpler integrals. For the Strauss process the pseudo-likelihood estimators \bar{b} and \bar{c} solve

$$\begin{aligned} \bar{b} \int_D \bar{c}^{t(\mathbf{x})} d\mathbf{x} &= n \\ \bar{b} \int_D t(\mathbf{x}) \bar{c}^{t(\mathbf{x})} d\mathbf{x} &= 2y(R) \end{aligned} \tag{3.11}$$

for variable n , where $t(\mathbf{x})$ denotes the number of points of the pattern within distance R of the test point \mathbf{x} . (Conditional intensities, hence pseudo-likelihood, make no sense if n is fixed. The intensity will be zero or infinite depending on whether there are n or $n - 1$ points elsewhere.) The integrals in (3.11) can be estimated by sums over a grid of points within D , and sophisticated methods are available to find $t(\mathbf{x})$ rapidly using auxiliary data structures.

These pseudo-likelihood methods proved to be a special case of a family of moment measures derived by Takacs (1986) and Fiksel (1984). These compare the expected values of observations on the process with similar expectations conditional on a point of the pattern at \mathbf{x} . However, whereas the Takacs-Fiksel work depends on arbitrary choices of moments to compare (and is mathematically forbidding), pseudo-likelihood has some theoretical rationale.

3.2.2. Edge effects

All the parameter estimation methods mentioned up to now have been for a process defined only on the domain D . This is frequently inadequate, in that we imagine the underlying process placing the points to occur within a much larger domain D' but which is observable (or has only been recorded) within the window D . In such a case the number of points is necessarily variable. A rigorous treatment of this case is very difficult, as it would involve averaging over the positions of points outside D which might interact with those within D . Rather, we choose to correct the estimators we have seen so far for edge-effects. This is particularly easy for the Strauss process, since the only statistic which occurs is $y(R)$, the number of R -close pairs. This occurs in the second-moment K -function which is often used to describe a spatial point pattern, and so much work has been done on correcting for edge-effects, going back at least as far as Glass & Tobler (1971). Full details of the corrections proposed and of their efficacy are given in Ripley (1988, Chapter 3).

3.3. Choosing an interaction function

How will we know what shape of interaction function to pick? Almost never is there any appropriate theory to suggest a particular functional form, and it is really the shape rather than the mathematical form which we seek. Two ideas have recently been suggested. The conceptually simplest is that of Ripley (1988, p. 73). Fit a multi-scale process (3.4) to obtain a step-function (histogram-like) estimate of the interaction function h , and choose a suitable functional form to fit parametrically. This is just like plotting a histogram in univariate statistics, and choosing a family of distributions (normal, gamma, ...) from its shape.

Fortunately, estimating the parameters in the multi-scale process by pseudo-likelihood is just as easy as for the Strauss process. The equations are similar to (3.11), with c replaced by c_i and $t_i(\mathbf{x})$ denoting the number of points of the pattern whose distance from \mathbf{x} is

Brian D. Ripley

between R_{i-1} and R_i . That is, (\bar{c}_i) solve

$$\bar{b} \int_D t(\mathbf{x}) \prod_j \bar{c}_j^{t(\mathbf{x})} d\mathbf{x} = \sum_{\text{pts}} t_i(\mathbf{x}_j) = 2[y(R_i) - y(R_{i-1})]$$

for each i together with

$$\bar{b} \int_D \prod_j \bar{c}_j^{t(\mathbf{x})} d\mathbf{x} = n$$

and the ratio is a function of (c_1, \dots, c_m) which is easily estimated numerically, and so the equations are solved numerically.

Care is needed in interpreting the shape of the function, as the estimates at small distances are very variable, unlike a histogram. For a Poisson process we would have that

$$\bar{c}_i \approx [y(R_i) - y(R_{i-1})] \times \frac{2a}{n(n-1)\pi[R_i^2 - R_{i-1}^2]}$$

by an extension of (3.8), and the count of pairs $[y(R_i) - y(R_{i-1})]$ is approximately Poisson. Thus \bar{c}_i has standard deviation about the square root of the second term, and this will be large if the area of the annulus of points between R_{i-1} and R_i away from a fixed point is small. This will inevitably be the case for distances between 0 and R_1 . For example, with the Spanish towns data [Figure 1(b)] we have $n = 70$ and the area D is 40 miles square. If we take R_1 to be 1 mile, we find a standard deviation of \bar{c}_1 of about $\sqrt{[2 \cdot 40^2 / (70 \cdot 69\pi 1^2)]} \approx 0.5$ so any inference would be meaningless.

The other idea borrows from statistical physics a relationship between the interaction function h and the second-moment function K known as the Percus-Yevick formula. Whereas in physics this is used to calculate K from h , Diggle, Gates & Stibbard (1987) had the idea to reverse the process to obtain a non-parametric estimator of h ; full details are in their paper. The Percus-Yevick formula is approximate, and something of a mystery to me, but simulations have shown it to be reasonably accurate. It relates the interaction function h to the pair-correlation function g . This is defined by

$$g(t) = (2\pi t)^{-1} \frac{dK(t)}{dt}$$

where $K(t)$ is my reduced second-moment function (Ripley, 1977, 1981). Then the formula is

$$h(t) \approx g(t) / [g(t) - c(t)]$$

where $c(\cdot)$ is the solution to

$$c(t) = g(t) - 1 - \lambda \int_0^{2\pi} \int_0^\infty [g(s) - 1] c(\sqrt{t^2 + s^2 - 2ts \cos \theta}) s ds d\theta$$

Details of how $g(t)$ is estimated and how the equations are solved are given in the paper. The procedures are *not* straightforward, but in their authors' hands give good results for simulations of patterns of a few hundred points.

The cautionary note about the step-function method applies to all; one needs a very large number of points to have much success in fitting a non-parametric estimator of the shape of the interaction function h ; about fifty points is not enough, rather a few hundred are needed. Examples occur in other fields, but rarely, I suspect, in geography. Okabe & Miki (1984) do have 393 greengrocer shops in their study area (28.2km² of Tokyo) but few other cities will have so many establishments within a homogeneous area.

4. Examples

In this section we consider two examples of fitting Gibbsian interaction models to point patterns drawn from geography. All have been mentioned earlier in the text and analysed previously by various statistical methods.

4.1. Spanish towns

The data come from Glass & Tobler (1971). They show the 'cities' (or 'towns', the terms are used interchangeably) in the 40 mile square centred at latitude 2° 30' W and longitude 39° 47' N which is claimed to be "especially homogeneous in climate, physiography, transport, population density and economy." An earlier version with 69 points has been considered from various points of view in Ripley (1977, 1979b, 1988). The data shown here have been re-digitized to show towns outside the study square, and 70 came within. The external points were used for edge-correction, for example in the integrals in (3.11). Glass & Tobler suggested fitting a hard-core model with interaction distance R as 3.46 miles, this peculiar figure coming from a simple value of $\bar{\rho}$, the number of points divided by the maximum packing of discs into the same domain. This was fitted despite the presence of pairs of towns about 0.6 miles apart, and indeed there appear to be $y(R) = 30$ pairs closer than 3.5 miles, an appreciable part of the total of 2,346 pairs. Nevertheless, this is appreciably fewer pairs than we would expect for a Poisson process. Ignoring edge effects, we would expect $\lambda^2 a \pi R^2 / 2 = 57.3$ pairs. [The exact result is known from Borel (1925) as

$$\frac{\lambda^2 a^2}{2} \left[\frac{\pi R^2}{L^2} - \frac{8R^3}{3L^3} + \frac{R^4}{2L^4} \right]$$

for a square of side L , but at 53.1 this is only slightly less. (Further, to a fairly crude approximation the number of pairs will have a Poisson distribution, so the standard deviation of the number is about 7.5 and the shortfall is definitely significant. More precise and powerful tests given in Ripley (1979b) all reject the Poisson hypothesis at at least the 5% level, some at the 1% level. This suggests that a Gibbsian interaction process might be appropriate. The interaction function found by fitting a multi-scale process by pseudo-likelihood is shown in Figure 4. (Half-mile intervals were chosen, except at the smallest distances.) Remember that the values at short distances are unreliable. The Figure gives no reason to doubt that a Strauss process would be appropriate, although doubtless many other functional forms for h would also suffice.

We can fit the parameter c of the Strauss process by any of the methods discussed in §3. The exact value obtained depends on what form of edge correction (if any) is used and, of course, the method; but for $R = 3.5$ miles, values of c between 0.4 and 0.5 are suggested (Ripley, 1988, §4.6). These correspond well to the trial-and-error value of 0.5 of Ripley (1977). This is a problem in which edge correction is essential as the study region does

Figure 4.

The interaction function h fitted to the Spanish towns data by pseudo-likelihood, to suggest a parametric form.

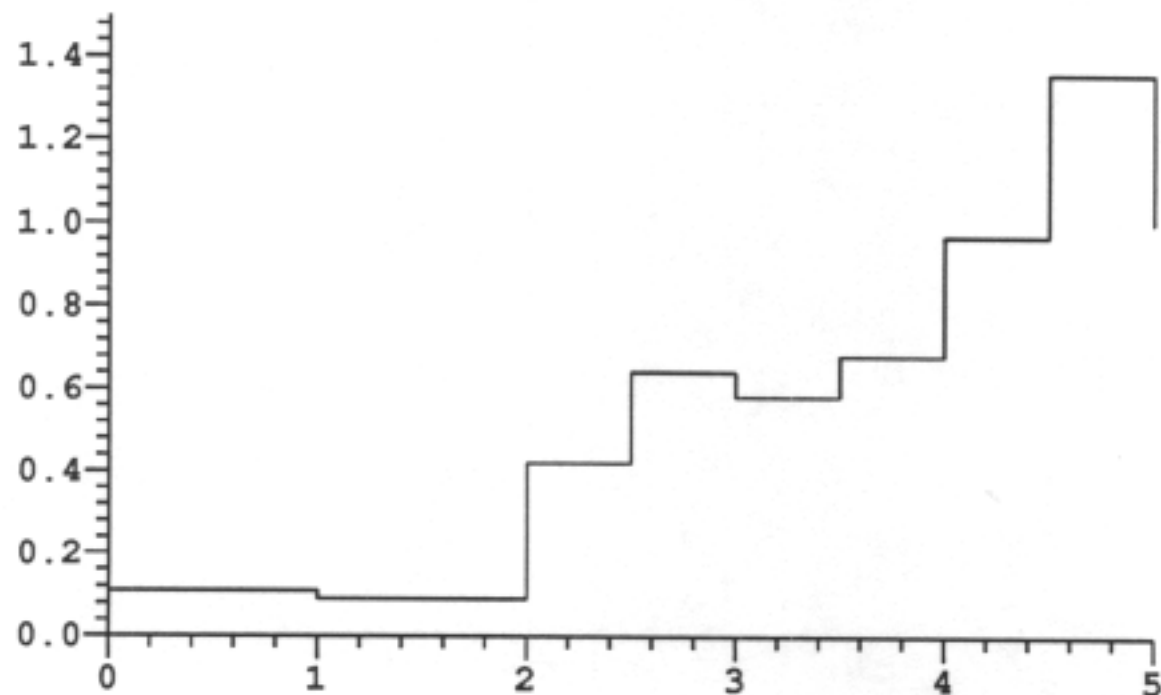
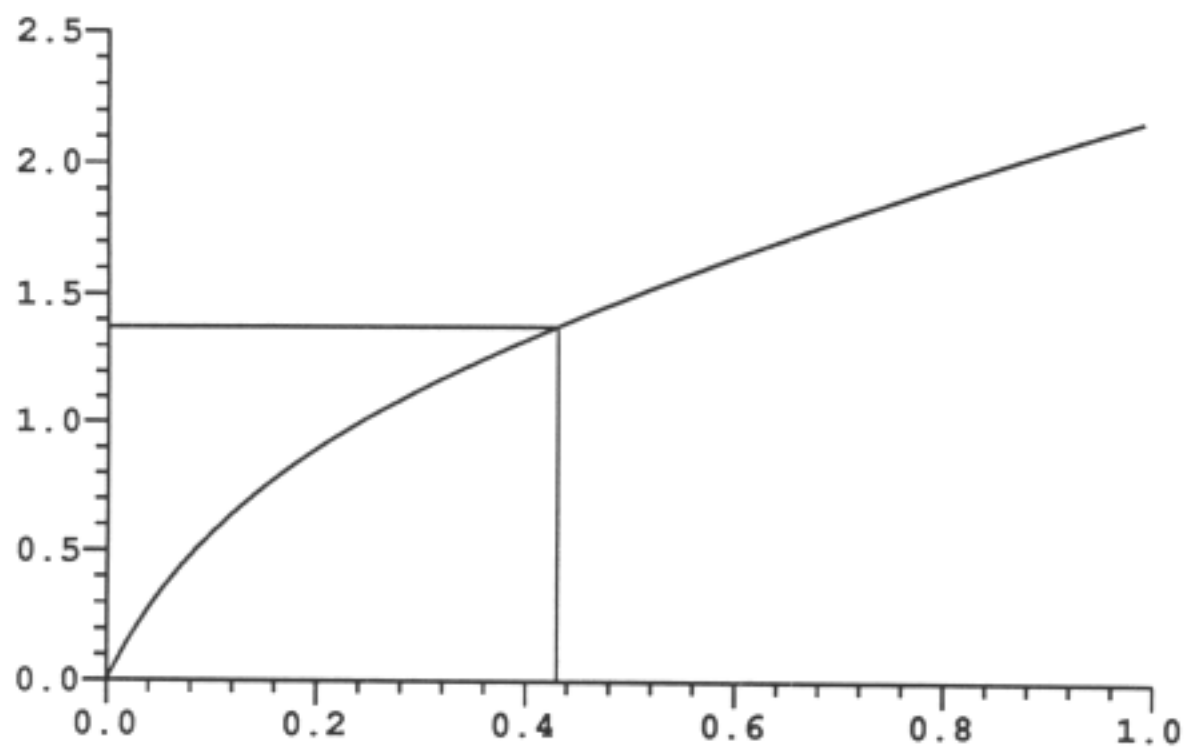


Figure 5.

Plot of the ratio of the equations (3.11) against c . The pseudo-likelihood estimator, $\hat{c} = 0.43$ is obtained from the horizontal line at $2y(R)/n = 96/70 = 1.37$. This example is for the Spanish towns with $R = 4$ miles.



not exist in isolation but is part of a very much larger plain. Fortunately the locations of adjacent points are available, but even if they were not, 'internal' edge correction methods work well and were used in my earlier studies.

To illustrate the procedures let us consider $R = 4$ miles, as this is suggested by Figure 4. For all the estimators we need $y(R) = 48$, counting as one half pairs of point with only one of them inside the square. For the pseudo-likelihood estimator the ratio of the left-hand sides of (3.11) is evaluated for a range of values of c by taking a grid of points within the square (I used 10×10) and averaging $t(\mathbf{x})c^{t(\mathbf{x})}$ and $c^{t(\mathbf{x})}$ at these points. [Remember that $t(\mathbf{x})$ is the number of towns closer than R miles to the test point \mathbf{x} .] Figure 5 shows this ratio for all values of c between zero and one. However, this range of values is not needed, as we can get a rough value of $c \approx 0.63$ from (3.8). (Note that this corresponds to fitting a straight line between the endpoints of the curve in Figure 5.) The value of the pseudo-likelihood estimator \bar{c} is 0.43.

This rough value is an approximation to the maximum likelihood estimator. We can refine it by (3.9) to be

$$c[1 + 1.29(c - 1)^2] \approx 0.62 = \frac{2ay(R)}{n^2\pi R^2}$$

with solution $c \approx 0.45$. To find the exact maximum likelihood estimator we use simulation to calculate $E_c Y(R)$ and plot this against c (Figure 6). On the other hand, the calculations for the solid line took 35 minutes on a Sun workstation (running the simulation 10,000 steps for each of twenty values of c), and still the curve is quite rough. From the observed value of $y(R) = 48$ we find $\hat{c} \approx 0.51$. As Figure 6 shows, the approximation (3.9) is good for large c , but fails for strong interactions (c small). The simulations enable us to estimate the variance of $y(R)$ and hence of \hat{c} ; this suggests a standard error of about 0.10.

Ogata & Tanemura (1984) fitted a very-soft-core model to the old dataset, using their approximation (but to higher order than is available for the Strauss process).

4.2. Drumlins in Northern Ireland

Figure 7 shows the pattern of 232 drumlins (glacial deposits) recorded by Hill (1973). Upton & Fingleton (1985) used sampling methods to analyse this pattern, and concluded "for this restricted region there is no evidence of other than a random distribution of 'plants'." More powerful methods show otherwise. For example, plots of my K-function show regularity up to 1 km, principally at distances of less than 600 m. This suggested fitting various multi-scale processes. With 5 equal steps of 200m each I obtained \bar{c}_i as 0, 0.27, 0.76, 1.30 and 1.06. This shows the interaction to be principally at shorter distances, and the most satisfactory result was obtained with an interaction function of 0.03 up to 150 m, 0.68 between 150 and 300 m, and one for larger distances.

Although there are many more points than in the previous example, the distances are smaller so the number of interacting pairs is only slightly greater [$y(300\text{m}) = 65$]. Thus the standard error of \bar{c}_2 is around 0.08.

The pattern shows clear signs of heterogeneity, especially in the North-West, so it did not seem wise to take this analysis much further.

Figure 6.

Plot of $E_c Y(R)$ against c for the Spanish towns example. The solid line is from simulations, the dashed line is equation (3.11).

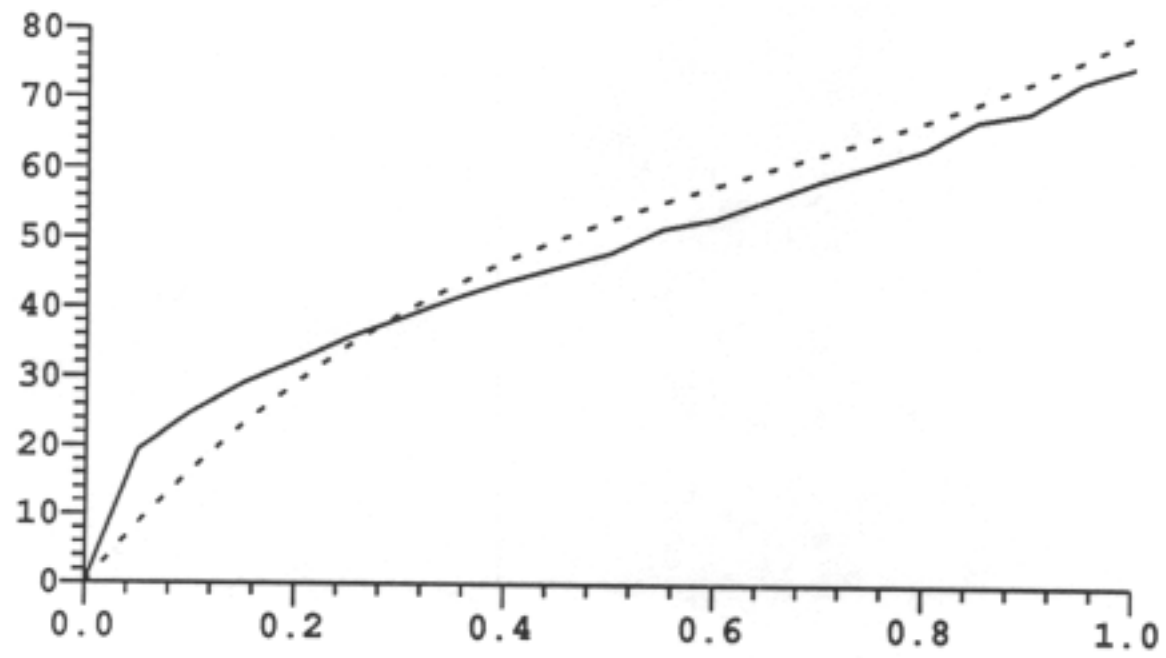
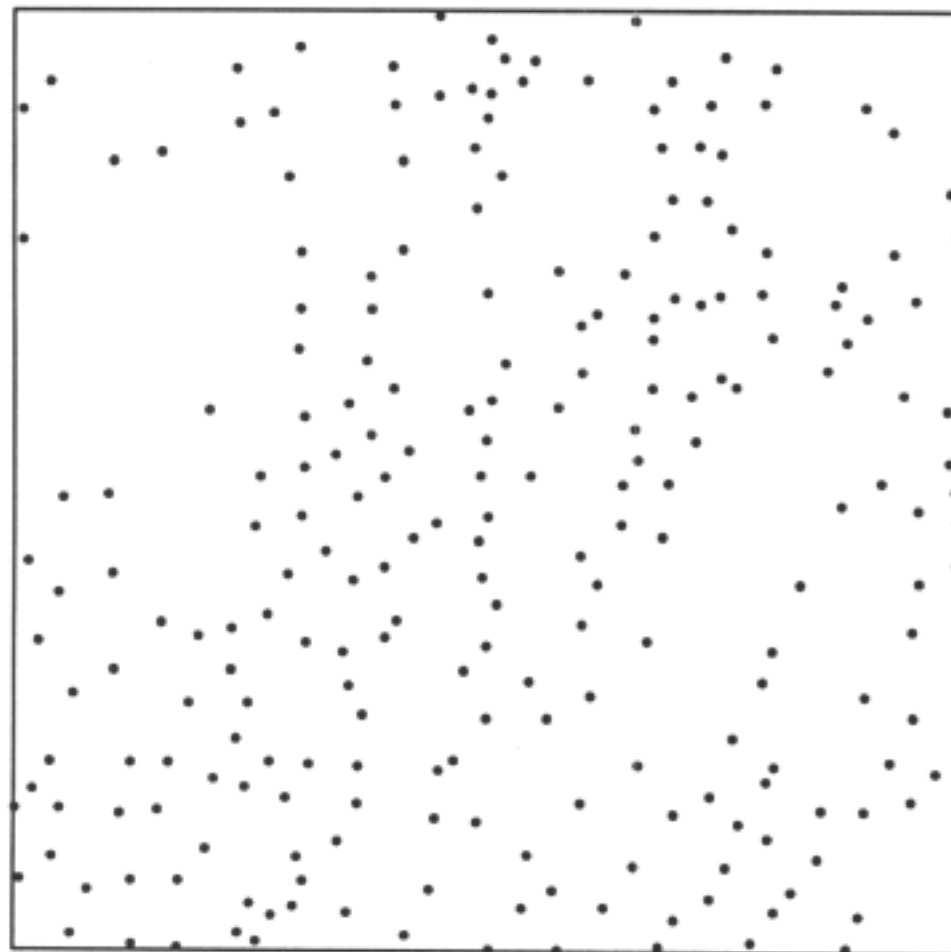


Figure 7.

Locations of 232 drumlins in an 8 km square of the Upper Ards peninsular, Co. Down. (After Hill, 1973).



4.3. Discussion

These two examples show that pairwise interaction models can be useful in geographical examples, and many more such models could have been tried. We do have to remember that parameter estimates are variable. It is difficult to quote accurate standard errors for the values quoted here, but since they are based on of the order of 50 pairs of points, and that number is approximately Poisson-distributed, we can expect standard errors of around 15% of the values quoted.

It is tempting to suggest smooth interaction functions and to interpret them. As Figure 3(d,e) shows, very different functional forms can be indistinguishable!

5. References

- Bartlett, M. S. (1971a) Two-dimensional nearest-neighbour systems and their ecological applications. In *Statistical Ecology* eds G. P. Patil, E. C. Pielou & W.E. Waters, Penn. State Univ. Press, State College, PA, I, 179-194.
- Bartlett, M. S. (1971b) Physical nearest-neighbour models and non-linear time series. *J. Appl. Prob.* **8**, 222-232.
- Bartlett, M. S. (1975) *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London.
- Bartlett, M. S. & Besag, J. (1969) Correlation properties of some nearest-neighbour systems. *Bull. Int. Statist. Inst.* **43**(2), 191-193.
- Bennett R. J. & Haining, R. J. (1985) Spatial structure and spatial interaction: Modelling approaches to the statistical analysis of geographical data. *J. Roy. Statist. Soc. A* **148**, 1-36.
- Besag, J. E. (1972a) Nearest-neighbour systems: A lemma with application to Bartlett's global solutions. *J. Appl. Prob.* **9**, 418-421.
- Besag, J. E. (1972b) Nearest-neighbour systems and the auto-logistic model for binary data. *J. Roy. Statist. Soc. B* **34**, 75-83.
- Besag, J. E. (1972c) On the correlation structure of some two-dimensional stationary processes. *Biometrika* **59**, 43-48.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. B* **36**, 192-236.
- Besag, J. (1975) Statistical analysis of non-lattice data. *The Statistician* **24**, 179-195.
- Besag, J. (1977) Some methods of statistical analysis of spatial data. *Bull. Int. Statist. Inst.* **47**(2), 77-92.
- Besag, J., Milne, R. & Zachary, S. (1982) Point process limits of lattice processes. *J. Appl. Prob.* **19**, 210-216.
- Borel, E. (1925) *Principes et Formules Classiques du Calcul des Probabilités*. Gauthier-Villars, Paris.
- Cliff, A. D. & Ord, J. K. (1981) *Spatial Processes*. Pion, London.
- Diggle, P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Diggle, P. J. & Gratton, R. J. (1984) Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. B* **46**, 193-227.

Brian D. Ripley

- Diggle, P. J., Gates, D. J. & Stibbard, A. (1987) A nonparametric estimator for pairwise-interaction point processes. *Biometrika* **74**, 763-770.
- Fiksel, T. (1984) Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektron. Inform. Kybernet.* **20**, 270-278.
- Geman, S. & Graffigne, C. (1987) Markov random fields and their applications to computer vision. *Proc. Inter. Cong. Math. 1986* ed. A.M. Gleason, Amer. Math. Soc., Providence, RI.
- Glass, L. & Tobler, W. R. (1971) Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature* **233(5314)**, 67-68.
- Haining, R. P. (1983) Modelling inter-urban price competition: An example of gasoline pricing. *J. Regional Sci.* **20**, 365-375.
- Hill, A. R. (1973) The distribution of drumlins in County Down, Ireland. *Ann. Assoc. Amer. Geogr.* **63**, 229-241.
- Kelly, F. P. & Ripley, B. D. (1976) On Strauss' model for clustering. *Biometrika* **63**, 357-360.
- Kindermann, R. & Snell, J. L. (1980) *Markov Random Fields and Their Applications* Amer. Math. Soc., Providence, RI.
- Mardia, K. V. & Marshall, R. J. (1984) Maximum likelihood estimation of models of residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Mead R. (1971) Models for interplant competition in irregularly spaced populations. In *Statistical Ecology* eds G.P. Patil, E.C. Pielou & W.E. Waters, Penn. State Univ. Press, State College, PA, **2**, 13-32
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem Phys*, **21**, 1087-1092.
- Molina, R. & Ripley, B. D. (1989) Using spatial models as priors in astronomical image analysis. *J. Appl. Statist.* **16**, 193-206.
- Ogata, Y. & Tanemura, M. (1981) Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* **33B**, 315-328.
- Ogata, Y. & Tanemura, M. (1984) Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. B* **46**, 496-518.
- Okabe, A. & Miki, F. (1984) A conditional nearest-neighbor spatial-association measure for the analysis of conditional locational interdependence. *Environment and Planning* **16**, 163-171.
- Ord, K. (1975) Estimation methods for models of spatial interaction. *J. Amer. Statist. Assoc.* **70**, 120-126.
- Penttinen, A. (1984) Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method. *Jyväskylä Studies in Computer Science, Economics and Statistics*, **7**.
- Pinder, D. A. & Witherick, M. E. (1972) The principles, practice and pitfalls of nearest-neighbour analysis. *Geography* **57**, 277-288.
- Preston, C. J. (1974) *Gibbs States on Countable Sets*. Cambridge University Press, London.

- Preston, C. J. (1976) *Random Fields*. Lecture Notes in Mathematics **534**, Springer-Verlag.
- Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. B* **39**, 172-212.
- Ripley, B. D. (1979a) Algorithm AS137. Simulating spatial patterns: Dependent samples from a multivariate density. *Appl. Statist.* **28**, 109-112.
- Ripley, B. D. (1979b) The analysis of geographical maps. In *Exploratory and Explanatory Statistical Analysis of Spatial Data*, eds C.P.A. Bartels & R.H. Ketellapper, Martinus Nijhoff, Boston, 53-72.
- Ripley, B. D. (1981) *Spatial Statistics*. Wiley, New York.
- Ripley, B. D. (1987) *Stochastic Simulation*. Wiley, New York.
- Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.
- Ripley, B. D. (1990) The use of spatial models as image priors. In *Spatial Statistics & Imaging* ed. A. Possolo, IMS Lecture Notes.
- Ripley, B. D. & Kelly, F. P. (1977) Markov point processes. *J. Lond. Math. Soc.* **15**, 188-192.
- Ripley, B. D. & Silverman, B. W. (1978) Quick tests for spatial interaction. *Biometrika* **65**, 641-642.
- Rogers, A. (1974) *Statistical Analysis of Spatial Dispersion: The Quadrat Method*. Pion, London.
- Strauss, D. G. (1975) A model for clustering. *Biometrika* **63**, 467-475.
- Takacs, R. (1986) Estimator for the pair potential of a Gibbsian point process. *Statistics* **17**, 429-433.
- Upton, G. J. G. & Fingleton, B. (1985) *Spatial Data Analysis by Example* volume I. Wiley, Chichester.
- Whittle, P. (1954) On stationary processes in the plane. *Biometrika* **41**, 434-449.

DISCUSSION

"Gibbsian interaction models"

by Brian D. Ripley

Gibbsian interaction models are usually better known as Markov random fields. These encompass many of the dependence models for regional or lattice data, and many models of point processes. The author aims to document the progress made over the past two decades in statistical inference and model fitting for these models.

About three-quarters of the paper is concerned with point processes. The Gibbsian interaction processes are defined, with particular emphasis on the Strauss processes, and the usual iterative method for simulation is given. Next there is a thorough discussion of maximum pseudolikelihood estimation. Approximate maximum likelihood is now feasible using theoretical approximations to the likelihood, or through simulation. Approximate maximum pseudolikelihood estimation is also feasible, and somewhat easier than approximate maximum likelihood. Two recent methods for choosing the interaction function are discussed. Many of these developments are very recent, and the discussion was welcome to me. Two examples are given. The first is a detailed illustration using the familiar 'Spanish towns' data, whilst the second comments on a previous analysis of some data on 'drumlins in Northern Ireland'. There are several figures that illustrate the text.

The section on regional processes is much briefer, and, apart from reviewing the usual conditional and simultaneous autoregression models, mainly consists of contrasting maximum pseudolikelihood estimation with maximum likelihood estimation. For these dependence models, both estimation methods have been used for some years (the former was introduced in 1975), although the author rightly questions whether maximum likelihood estimation is always relevant. However, his claim that 'pseudolikelihood methods have been very successful' may reflect his interest in astronomically large data sets, and non-Gaussian distributions. The results of Besag (1977) show that pseudolikelihood estimators can be very inefficient when the dependence is not small, so that their use is questionable when other estimators, that may be more efficient, are readily available. Although the author refers to a recent rigorous proof of the consistency of the pseudolikelihood estimator, he does not refer to Guyon's (1987) theorem on asymptotic Normality. There are no examples given in this section, and the only applications referred to are in agricultural field trials and astronomical image analysis.

The author's presentation is, as usual, precise and comprehensive, though the mathematics may make demands on many. Perhaps the author could have considered more the questions of how appropriate these models and methods are for geographical research, and of how they are being, or have been, used in geographical research. Whilst it may be useful to give unreferenced publications in the bibliography, it would have been helpful if this policy had been noted, and a brief note had been given with each to explain its possible relevance.

References

- Besag, J. (1977) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, **64**, 616-618.
- Guyon, X. (1987) Estimation d'un champ par pseudo-vraisemblance conditionnelle: étude asymptotique et application au cas Markovien, in *Spatial Processes and Spatial Time Series Analysis*, edited by F. Dreesbeke, pp. 15-62. Brussels: Facultés Universitaires Saint-Louis.

R. J. Martin, University of Sheffield