Markov Random Fields in Statistics

Peter Clifford

1. Introduction

For nearly a century, statisticians have been intrigued by the problems of developing a satisfactory methodology for the analysis of spatial data; see Student (1914), for an early example. It is only since the early 1970's, however, that the statistical analysis of large data sets, using flexible parametric models has become a feasible proposition.

On the practical side, progress has been made possible by the availability of relatively cheap, computerised resources for the collection and analysis of data. The study of digital images and the use of satellite data for remote sensing are prominent examples in this respect. On the methodological side, substantial progress is associated with the introduction of Markov random fields (MRFs), as a class of parametric models for spatial data (Besag 1974). Shaped by these developments, spatial statistics has emerged as perhaps the most dynamic and computer intensive of all the areas of statistical endeavour; building upon models used originally in the description of physical systems and borrowing and improving upon ideas from computational physics.

Monte Carlo methods, in particular, have played a dominant role in dealing with problems of inference. The practicalities of working with high dimensional parameter sets within a Bayesian framework, have led to the invention of refreshing and novel techniques (Geman and Geman 1984), which promise to have a profound effect on the way in which Bayesian methods are used in more general contexts and which may serve to reintegrate these methods into the main body of applied statistics.

Much of physics is concerned with providing an understanding of the spatial organisation of matter and it is not suprising that many of the ideas which have become central in the theory of spatial statistics should have their origins in physical theory. The introduction of MRFs into the theory of statistics is yet another example of the continuing transfer of knowledge from the world of theoretical physics. John Hammersley whose interests include both domains of study, was ideally placed to facilitate the process of cross-fertilisation. Others who were involved in this instance include Neyman and Besag. Neyman was responsible for bringing Hammersley and a number of other visitors, including myself, to the University of California, Berkeley in the summer of 1971. Hammersley gave an advanced course of lectures on probabilistic problems in physics, which included among other things a discussion of Spitzer's (1971) characterisation of two-state MRFs on a square lattice. This characterisation had been obtained independently by Averintsev (1970). Hammersley and I were able to generalise the results to arbitrary graphs and lattices, and to identify the central importance of the clique functions, as terms in the potential of a generalised Gibbs distribution. Hammersley returned to Oxford and sent a copy of the Berkeley paper to Besag who had already obtained partial results for rectangular lattices (Besag 1972). Besag then wrote to Hammersley with a much simpler, analytical proof of the general result, which appeared later in his very influential paper on spatial statistics (Besag 1974). Three other authors published proofs of the main theorem at about this time (Grimmett 1973; Preston 1973; Sherman 1973). A simple derivation is also possible using the factorisation theorem of Brook (1964). The basic theorem has more recently become important in non-spatial applications, most notably in the description of dependence structure for log-linear models (Ove and Strauss 1981; Darroch et al. 1980).

The Berkeley paper was never published and only a few copies were distributed. There are, however, many references to it in the literature and although the main result is stated as a named theorem in Kotz and Johnson (1983, Vol. 3, p. 570) there is, perhaps inevitably, some confusion about the exact contents. The method of proof in the unpublished paper is constructive and the operator techniques used are unusual. For these reasons it seems appropriate to take this opportunity to state the main results and to describe the methods by which they were obtained. This is done in Section 2.

The Markov property for random fields can be formulated in great generality (Preston 1974; Rozanov 1982). For statistical applications, an important step forward was the extension to point processes (Strauss 1975; Ripley and Kelly 1977). An excellent review of this topic is given by Baddeley and Møller (1989), who consider further generalisations to cover the case of marked point processes in which the neighbourhood relations for the marks are given by the graphical structure of the points.

A challenging problem is that of constructing random mosaics which are spatially Markov. A special case is the problem of subdividing two-dimensional space into regions whose boundaries are made up of line segments. In a remarkable paper, Arak (1982) showed that a time-homogeneous annihilating/birth particle system can give rise to space-time trajectories which have a two dimensional Markov property. These results were generalised by Arak and Surgailis (1989), to cover a wide class Markov polygonal fields. In Section 3 we consider how these processes might be used in the analysis of polygonal images. Some light is shed on a conjecture by Arak and Surgailis (1989) and a method of simulating the posterior distribution of a polygonal image is proposed.

2. Markov Fields on Finite Graphs

2.1. Notation

Let $G = (Z, E)$ be an undirected graph, where $Z = \{z_1, z_2, \dots, z_n\}$ is a finite set of sites and E is a set of simple edges, i.e. a set of unordered pairs of distinct sites. Two sites which form an edge are said to be *neighbours* of each other. We use capital letters $U, V, \dots, X, Y$ for subsets of Z and write $X + Y$ for the union of X and Y, and $X - Y$ for the set $\{x \in X : x \notin Y\}$. A lower-case letter stands for both an element of Z and also the associated singleton set. The set of all subsets of Z, including $\emptyset$ and Z itself is denoted by Ω. For any Y we define ∂Y, the *boundary* of Y by

$$\partial Y = \{x : (x, y) \in E, x \notin Y, y \in Y\}.$$

A set Y is said to be a *clique* if and only if

$$Y \subseteq y + \partial y, \quad \forall y \in Y;$$

in other words Y is a clique if and only if it is a singleton or if every member of Y is a neighbour of every other member of Y.

We associate with each site z_i, a finite set of colours $C_i, i = 1, 2, \dots, n$. To avoid trivial cases we will assume that the cardinality of each set is greater than one. We also assume, without loss of generality, that every set contains a colour which we can agree to call *black*. Suppose that for each z_i we select a colour from $C_i, i = 1, 2, \dots, n$. Such an assignment of colours to sites is called a *colouring* of Z. A typical colouring is denoted by χ. Let χ_Y denote the colouring obtained from χ by changing the colours on the sites in Y to black. A *partial colouring* has colours assigned on only a subset of sites. The partial colouring obtained by considering which colours have been assigned to sites in X by the colouring χ is denoted by χ^X. In particular, the colour at a site z is written as χ^z. The set of all possible colourings of Z is given by $\mathcal{C} = C_1 \times C_2 \times \cdots \times C_n$. A set Y is said to be light relative to χ if no site in Y is black under the colouring χ. We define L_χ to be the set of cliques which are light relative to χ.

Let us now consider a probability distribution on $\mathcal{C}$ with mass function P satisfying $\sum_{\chi \in \mathcal{C}} P(\chi) = 1$ and the *positivity condition* $P(\chi) > 0, \forall \chi \in \mathcal{C}$. We denote the marginal probability of the partial colouring χ^Y by $P(\chi^Y)$. This latter probability is obtained by summing P over all colourings which agree with χ on Y. We say that P is Markovian for the set X if and only if it satisfies the positivity condition and

$$P(\chi)/P(\chi^{Z-X}) = P(\chi^{X+\partial X})/P(\chi^{\partial X}), \quad \forall \chi \in \mathcal{C}.$$

We call this condition $M(X)$. If we postulate $M(z)$ for all singleton sets $z \in Z$, we say P is locally Markovian. If we postulate $M(X)$ for all $X \subseteq Z$ we say it is globally Markovian. The main theorems are as follows.

THEOREM 1. *Global and local Markov properties are equivalent.*

THEOREM 2. *P is Markovian if and only if it can be written in the form*

$$P(\chi)/P(\chi_Z) = \exp\Bigg(\sum_{Y \in L_\chi} Q(\chi^Y)\Bigg),$$

where Q is an arbitrary real-valued function of light colourings on cliques.

Furthermore, if P is Markovian then the associated function Q is given by

$$Q(\chi^Y) = \sum_{X \subseteq Y} (-1)^{|X|} \log P(\chi_{(Z-Y)+X}), \quad \forall Y \in L_\chi,$$

where $|X|$ denotes the cardinality of X.

The theorems are proved by introducing an operator algebra.

2.2. The Blackening Algebra

Let $\mathcal{R}$ be the set of all real-valued functions defined on $\mathcal{C}$. We define the *pure* blackening operator B_Y by

$$B_Y R(\chi) = R(\chi_Y), \quad R \in \mathcal{R}.$$

Since

$$B_X B_Y R(\chi) = B_X R(\chi_Y) = R(\chi_{X+Y}) = B_{X+Y} R(\chi),$$

in terms of the operators we have

$$B_X B_Y = B_Y B_X = B_{X+Y}$$

so that pure operators commute.

A *mixed* blackening operator $\alpha_1 B_{X_1} + \cdots + \alpha_m B_{X_m}$ is a finite linear combination of pure operators, where $\alpha_1, \ldots, \alpha_m$ are real-valued coefficients. For such an operator we have

$$(\alpha_1 B_{X_1} + \cdots + \alpha_m B_{X_m}) R(\chi) = \alpha_1 R(\chi_{X_1}) + \cdots + \alpha_m R(\chi_{X_m}).$$

Mixed operators multiply according to

$$\sum \alpha_i B_{X_i} \sum \beta_j B_{Yj} = \sum \alpha_i \beta_j B_{X_i+Y_j}.$$

The identity operator is denoted by $1 = B_\emptyset$ and the zero operator by 0. With the preceding definitions, the blackening operators can be seen to form a commutative algebra.

The following lemma is a simple consequence of the definitions:

Lemma 1. *If* $X \subseteq Y$ *then* $(1 - B_X)B_Y = 0$.

An operator which is equal to its square is called a projector. Every pure operator is a projector. In general, if B is a projector then so is $1 - B$. It follows that

$$B_X + B_Y - B_{X+Y} = 1 - (1 - B_X)(1 - B_Y)$$

is also a projector. In the special case $Y = Z - (X + \partial X)$, for which $X + Y = Z - \partial X$, we denote the projector by β_X, i.e.

$$\beta_X = B_X + B_{Z-(X+\partial X)} - B_{Z-\partial X} = B_X + B_{Z-(X+\partial X)}(1 - B_X).$$

We also define $B_z^* = B_{Z-(z+\partial z)}(1 - B_z)$, so that $\beta_z = B_z + B_z^*$. Finally we define the projector $\beta = \prod_{z \in Z} \beta_z$. Writing $B_Y^* = \prod_{z \in Y} B_z^*$ and $B_\emptyset^* = 1$, we have

$$\beta = \prod_{z \in Z} (B_z + B_z^*) = \sum_{Y \in \Omega} B_{Z-Y} B_Y^*. \tag{2.1}$$

Lemma 2. *If* $Y \neq \emptyset$ *and* Y *is not a light clique relative to* χ*, then* $B_Y^* R(\chi) = 0, \forall R \in \mathcal{R}$.

Proof: (i) Suppose that Y is not a clique. Then Y has two distinct elements, x, y, say, such that x is not a neighbour of y, i.e. $x \in Z - (y + \partial y)$. From Lemma 1, it follows that $B_{Z-(y+\partial y)}(1 - B_x)$, and hence B_Y^*, equals 0.
(ii) Suppose that Y is not light relative to χ, then Y contains a site z which is already black, so that $(1 - B_z)R(\chi)$, and hence $B_Y^* R(\chi)$ equals 0. ■

Let us now consider the subset of $\mathcal{R}$ which is invariant under the operator β. Denoting this subset by $I(\beta)$, we have

$$I(\beta) = \{R : \beta R = R, R \in \mathcal{R}\} = \{\beta R : R \in \mathcal{R}\}.$$

If R is arbitrary then from Lemma 2 and (2.1) we have

$$\beta R(\chi) = R(\chi_Z) + \sum_{Y \in L_\chi} B_Y^* B_{Z-Y} R(\chi). \tag{2.2}$$

Furthermore, if Y is a clique then $Y \subseteq z + \partial z$ for any $z \in Y$, so that $Z - Y \supseteq Z - (z + \partial z)$ and $B_{Z-Y} = B_{Z-(z+\partial z)} B_{Z-Y}$ by Lemma 1. It follows that for arbitrary $R \in \mathcal{R}$ we have the further simplification

$$\beta R(\chi) = R(\chi_Z) + \sum_{Y \in L_\chi} \prod_{z \in Y} (1 - B_z) R(\chi_{Z-Y}). \tag{2.3}$$

LEMMA 3. *The invariant subset $I(\beta)$ consists of those functions $R \in \mathcal{R}$ which have the representation*

$$R(\chi) = S(\chi_Z) + \sum_{X \in L_\chi} S(\chi_{Z-X}) \tag{2.4}$$

for some $S \in \mathcal{R}$.

PROOF: (i) Let R have the representation (2.4) for some $S \in \mathcal{R}$. We will apply (2.3) to show that $\beta R = R$. Since there are no light cliques in χ_Z, we have $R(\chi_Z) = S(\chi_Z)$. Notice that if $Y \in L_\chi$, then $L_{\chi_{Z-Y}}$ is just the set of all nonempty subsets of Y. It follows that when R is given by (2.4) then

$$R(\chi_{Z-Y}) = \sum_{X \subseteq Y} S(\chi_{Z-X}), \tag{2.5}$$

Furthermore,

$$\prod_{z \in Y} (1 - B_z) R(\chi_{Z-Y}) = S(\chi_{Z-Y}), \tag{2.6}$$

since if $z \in Y$ then

$$(1 - B_z) R(\chi_{Z-Y}) = R(\chi_{Z-Y}) - R(\chi_{Z-Y+z}) = \sum_{z \subseteq X \subseteq Y} S(\chi_{Z-X}).$$

From (2.3) we therefore have $\beta R = R$.
(ii) Suppose now that $R \in I(\beta)$, i.e. $R = \beta R$. Since $\prod_{z \in Y}(1 - B_z) R(\chi_{Z-Y})$ is some function of χ_{Z-Y}, say $S(\chi_{Z-Y})$, and $R(\chi_Z)$ can be taken to be $S(\chi_Z)$, it follows immediately that R can be expressed as the right-hand side of (2.4). ■

LEMMA 4. *If $X \subseteq Z$, then $I(\beta) \subseteq I(\beta_X)$.*

PROOF: Let $R \in I(\beta)$, then R will have a representation as in Lemma 3. Since β_X is linear it suffices to show that

$$\beta_X S(\chi_Z) = S(\chi_Z)$$

and

$$\beta_X S(\chi_{Z-Y}) = S(\chi_{Z-Y})$$

for all cliques Y. Writing β_X as

$$\beta_X = 1 - (1 - B_X)(1 - B_{Z-(X+\partial X)})$$

the first of the equalities follows immediately. To establish the second equality it is sufficient to show that

$$(1 - B_X)(1 - B_{Z-(X+\partial X)}) B_{Z-Y} = 0. \tag{2.7}$$

But if Y is a clique it cannot be partly in X and partly in $Z-(X+\partial X)$. Suppose that $Y \subseteq X$ then

$$Z-(X+\partial X) \subseteq Z-X \subseteq Z-Y$$

and therefore (2.7) is satisfied as a consequence of Lemma 1. Alternatively, suppose that $Y \subseteq Z-(X+\partial X)$, then $Z-Y \supseteq X+\partial X \supseteq X$, so that the equation is again satisfied by Lemma 1. ■

LEMMA 5. *The invariant set $I(\beta)$ is given by $\cap_{z\in Z} I(\beta_z)$.*

PROOF: As a special case of Lemma 4 we have $I(\beta) \subseteq I(\beta_z), \forall z \in Z$, which implies that $I(\beta) \subseteq \cap_{z\in Z} I(\beta_z)$. On the other hand, if $R \in \cap_{z\in Z} I(\beta_z)$ then $R = \beta_z R, \forall z \in Z$, and hence $R = \prod_{z\in Z} \beta_z R$, so that $R \in I(\beta)$. ■

2.3. Proofs of Theorems 1 and 2

We show firstly, that the Markov condition $M(X)$ is equivalent to

$$P(\chi)/P(\chi_X) = P(\chi_{Z-(X+\partial X)})/P(\chi_{Z-\partial X}), \quad \forall \chi \in \mathcal{C}. \tag{2.8}$$

Under condition $M(X)$ we have

$$P(\chi) = P(\chi^{X+\partial X})P(\chi^{Z-X})/P(\chi^{\partial X}), \quad \forall \chi \in \mathcal{C}. \tag{2.9}$$

Equation (2.8) then follows by making the substitutions $\chi_X, \chi_{Z-(X+\partial X)}$ and $\chi_{Z-\partial X}$, and noting that $\chi_X^{Z-X} = \chi^{Z-X}, \chi_X^{\partial X} = \chi^{\partial X}$ etc. Conversely, if (2.8) holds, then

$$P(\chi) = P(\chi_X)P(\chi_{Z-(X+\partial X)})/P(\chi_{Z-\partial X}).$$

By summation over the appropriate subsets of $\mathcal{C}$, the marginal probabilities which appear in condition $M(X)$ can now be expressed as marginal probabilities of blackened colourings, which can be simplified as in the first part of the proof. Condition $M(X)$ is then verified by substitution.

PROOF OF THEOREM 1: From (2.8), condition $M(X)$ is equivalent to

$$R(\chi) - R(\chi_X) = R(\chi_{Z-(X+\partial X)}) - R(\chi_{Z-\partial X}), \quad \forall \chi \in \mathcal{C}$$

where $R(\chi) = \log P(\chi)$. In other words,

$$\beta_X R(\chi) = R(\chi), \quad \forall \chi \in \mathcal{C}.$$

Condition $M(X)$ is therefore equivalent to $R \in I(\beta_X)$.

Theorem 1 then follows immediately since if P is locally Markovian then $R \in \cap_{z\in Z} I(\beta_z) = I(\beta) \subseteq I(\beta_X)$ by Lemma 4 and hence P is globally Markovian. ■

Proof of Theorem 2: From Lemma 3, $R \in I(\beta)$ iff

$$R(\chi) - R(\chi_Z) = \sum_{X \in L_\chi} S(\chi_{Z-X}), \quad \forall \chi \in \mathcal{C}$$

for some $S \in \mathcal{R}$. Defining $Q(\chi^X)$ to be $S(\chi_{Z-X})$, the proof of the first part of Theorem 2 is complete. For the last part, notice that $Q(\chi^Y) = S(\chi_{Z-Y})$ is given by

$$Q(\chi^Y) = \prod_{z \in Y} (1 - B_z) R(\chi_{Z-Y}),$$

as in (2.6). The result now follows since the operator $\prod_{z\in Y}(1 - B_z)$ has the expansion $\sum_{X \subseteq Y} (-1)^{|X|} B_X$. ■

3. Markov Polygonal Mosaics

The random fields described in Section 2 have proved to be useful models in the analysis of two-dimensional images (Geman and Geman 1984; Besag 1983). For image analysis, the sites of the graph, $z_1, \cdots, z_n$ correspond to pixels in a digitised picture. In the Bayesian framework χ, the unknown colouring of Z, i.e. the true scene, is treated as a realisation of a Markov random field. The observations of the pixel values $\mathcal{O} = \{O^z, z \in Z\}$ are assumed to be random corruptions of the true scene. In the simplest case, the likelihood is assumed to be proportional to

$$\exp\Bigl(\sum_{z \in Z} h(O^z \mid \chi^z)\Bigr). \tag{3.1}$$

Up to an additive constant, the logarithm of the posterior density of χ is therefore

$$\sum_{Y \in L_\chi} Q(\chi^Y) + \sum_{z \in Z} h(O^z \mid \chi^z),$$

which can expressed as

$$\sum_{Y \in L_\chi} Q^*(\chi^Y)$$

where the singleton clique functions $Q^*(\chi^z)$ have been modified by inclusion of terms from the likelihood. It follows that the family of MRFs is conjugate with likelihoods of the form (3.1).

Bayes estimates of the true scene can be made by a variety of techniques. Simulated annealing can be used to find maximum *a posteriori* estimates and the Gibbs sampler can be used to find estimates with minimum mean square error and estimates with minimum mis-classification error (Geman and Geman 1984).

When large artificial structures are present in the scene, it may be more natural to model true scenes as random mosaics which subdivide two-dimensional space into regions whose boundaries are made up of line segments. These random fields are defined on a continuous space rather than on the nodes of a graph.

3.1. Polygonal Colouring Measure

The simplest building block for polygonal fields is the Poisson line process (Kendall and Moran 1963). To describe the construction we introduce the following notation.

Let $T \subset \mathbb{R}^2$ be a convex bounded domain. Let $\mathcal{L}_T^n$ be the family of all sets of n distinct lines which intersect T and let $\mathcal{L}_T = \cup_{n=0}^{\infty} \mathcal{L}_T^n$, with $\mathcal{L}_T^0$ defined to be $\{\emptyset\}$, the family consisting of the empty set alone. We consider a Poisson line process defined on $\mathcal{L}_T$. To fix ideas we will assume that the process is homogeneous and isotropic with intensity λ, so that the number of lines crossing a disc of diameter d has a Poisson distribution with mean λd and the mean number of lines intersecting T is λd_T, where d_T is the mean diameter of T. We write μ_T for the Poisson line measure on $\mathcal{L}_T$, and we denote the conditional line measure on $\mathcal{L}_T^n$ by ν_T^n, so that

$$\mu_T(A) = \sum_{n=0}^{\infty} \frac{e^{-\lambda d_T} (\lambda d_T)^n}{n!} \nu_T^n(A \cap \mathcal{L}_T^n), \tag{3.2}$$

for events $A \subset \mathcal{L}_T$.

Suppose that C is a finite set of colours and χ maps T into C. The colouring is said to be polygonal if and only if the set of discontinuity points of χ is the union of intervals of a finite number of distinct lines, where each line contributes exactly one interval. We disregard intervals of zero length. Associated with each polygonal colouring χ there is the unique set of lines which contain the discontinuity points. We call this set $\langle\chi\rangle$. For an open set $S \subset T$ we define $\langle\chi^S\rangle$ to be the set of lines associated with discontinuities of χ on S. If S is not open, we define $\langle\chi^S\rangle$ as the limit for a sequence of diminishing open neighbourhoods of S. We denote the set of all polygonal colourings χ such that $\langle\chi\rangle = \ell$ by $\Omega_T^\ell, \ell \in \mathcal{L}_T$ and write $\Omega_T = \cup_{\ell \in \mathcal{L}_T} \Omega_T^\ell$.

The polygonal colouring measure is then defined to be

$$\gamma_T(A) = \int_{\mathcal{L}_T} |A \cap \Omega_T^\ell| \, \mu_T(d\ell), \tag{3.3}$$

where A is a measurable subset of Ω_T and $|\cdot|$ denotes cardinality. We consider distributions on Ω_T which are absolutely continuous with respect to γ_T. These can be specified by a density $f : \Omega_T \to [0, \infty)$. The associated probability measure is therefore given by

$$P_T^f(A) = \frac{\int_A f(\chi)\gamma_T(d\chi)}{\int_{\Omega_T} f(\chi)\gamma_T(d\chi)}, \tag{3.4}$$

provided that the denominator is finite.

3.2. The Uniform Density

Arak and Surgailis (1989) conjectured that it might be possible for γ_T to be finite, i.e. for $P_T^f(A)$ to be a probability measure, when f is constant. The following theorem gives a sufficient condition for this to be so.

THEOREM 3. *If* $|C| = 2$ *and* $\lambda d_T < 1$, *then* $\int_{\mathcal{L}_T} |\Omega_T^\ell| \mu_T(d\ell) < \infty$.

Before proving the theorem, we must introduce a little more notation. An extended polygonal colouring is a function $\chi^+ : \mathbb{R}^2 \to C$, whose discontinuity points are the union of intervals of lines in $\mathcal{L}_T$, but here the intervals are either semi-infinite, infinite, or of finite positive length. We write $\langle \chi^+ \rangle$ for the line set associated with χ^+. The set of all extended colourings for which $\langle \chi^+ \rangle = \ell$, is denoted by $\Theta_T^\ell, \ell \in \mathcal{L}_T$, and $\Theta_T = \cup_{\ell \in \mathcal{L}_T} \Theta_T^\ell$. We now restrict our attention to the case $|C| = 2$. Since $|\Theta_T^\ell|$ depends only on the cardinality of ℓ (say n) we will write it simply as $2c_n$; the factor 2 arising from the two possible colourings for a given discontinuity set.

LEMMA 6. *The sequence* $\{c_n\}$ *satifies the recurrence relation*

$$c_{n+1} = c_n + 4nc_{n-1} + \sum_{j=2}^{n} \frac{n!}{(n-j)!} \left(2j + \tfrac{5}{2}\right) c_{n-j}$$

with $c_0 = c_1 = 1$. *Furthermore, the power series*

$$g(u) = \sum_{n=0}^{\infty} \frac{c_n u^n}{n!}$$

has radius of convergence 1 *and is given by*

$$4 \log g(u) = -6u - u^2 + \frac{8u}{1-u} - 2\log(1-u).$$

For brevity the proof of this lemma is omitted. The proof of the theorem follows from the lemma by noting that $|\Omega_T^\ell| \le |\Theta_T^\ell|$, $\forall \ell \in \mathcal{L}_T$, so that from (3.2) and (3.3)

$$\gamma_T(\Omega_T) \le 2g(\lambda d_T)e^{-\lambda d_T}, \ \lambda d_T < 1.$$

Polygonal fields do not necessarily have a Markov property. However, Arak and Surgailis (1989) have established a sufficient condition for this to be so, namely that f is of the form $f(\chi) = e^{-F(\chi)}$, where $F : \Omega_T \to \mathbb{R} \cup \{\infty\}$ is an additive function (Rozanov 1982).

Examples of additive functions are: the total length of the intercolour boundary of χ, the number of times that the boundary between two colours turns so as to circle a particular colour and most importantly functions of the form

$$F(\chi) = \int_T k(\chi(t))\alpha(dt),$$

where α is a measure on T and $k : C \to \mathbb{R}$.

3.3. Statistical Applications

Additive functions arise naturally in statistical contexts. Thus, if χ is the true scene and observations of χ are limited to realisations of a spatial Poisson process whose intensity at point t is $\eta(\chi(t))$, $t \in T$, then the likelihood of the data is proportional to

$$\exp\Big(-\int_T \eta(\chi(t))dt + \int_T \log \eta(\chi(t))N(dt)\Big) \tag{3.5}$$

where $N(A), A \subset T$ is the counting measure of the observed point pattern. If the prior density of the true scene is proportional to $\exp\big(-F_0(\chi)\big)$, then applying Bayes Theorem, the posterior density of χ is proportional to $\exp\big(-F^*(\chi)\big)$, where

$$F^*(\chi) = F_0(\chi) + \int_T \eta(\chi(t))dt - \int_T \log \eta(\chi(t))N(dt). \tag{3.6}$$

It follows that if F_0 is additive then so is F^*. In other words, polygonal Markov fields are conjugate with Poisson sampling. It is therefore important to be able to simulate Markov polygonal fields, in particular the posterior distributions within this family.

3.4. Conditional Distributions

Polygonal Markov fields are Markov in the following sense. Let $S \subset T$ be an open set with a smooth boundary $\partial S = \bar{S} - S$ and let $\xi = (\chi^{\partial S}, \langle \chi^{\partial S} \rangle)$ then the distribution of χ^S given ξ is the same as that of χ^S given χ^{T-S}. Convex sets are of particular interest, and we will assume that S is convex from now on. Note that the information in ξ consists of the colouring on the boundary $\chi^{\partial S}$ and also the identification of those lines, intersecting ∂S, which separate different boundary colours. The conditional polygonal measure on S is

$$\gamma_S(A \mid \xi) = \int_{\mathcal{L}_S} \big|A \cap \Omega_S^{\ell}(\xi)\big| \, \mu_S(d\ell),$$

where A is a measurable subset of Ω_S and $\Omega_S^\ell(\xi)$ is the set of polygonal colourings on S which have discontinuity lines $\ell \cup \langle \chi^{\partial S} \rangle$ and which are consistent with the boundary conditions ξ. For fields which are specified by an additive function F, Arak and Surgailis (1989) have shown that the conditional distribution of χ^S is absolutely continuous with respect to $\gamma_S(\cdot \mid \xi)$, with density proportional to $\exp\bigl(-F(\chi^S)\bigr)$.

The principal advantage of having an explicit form for the conditional density, is that Monte Carlo methods, such as the Gibbs sampler can be applied to simulate the process.

3.5. Monte Carlo Simulation of Markov Polygonal Fields

The idea is to run a Markov process on the state space Ω_T, whose equilibrium will be the desired field. The set T can be taken to be a rectangle. The procedure is as follows.

Let χ be the current state of the Markov process.

(a) Select a rectangle S, at random in T.

(b) Put down a realisation of a Poisson line process with intensity ρ in S. Suppose that the lines of the process are ℓ and that they are n in number.

(c) Calculate $K(\xi, \ell)$ given by

$$[K(\xi,\ell)]^{-1} = \sum_{\omega \in \Omega_S^\ell(\xi)} e^{-F(\omega)},$$

and select a new colouring for S, from the distribution with probability mass function $K(\xi,\ell)\exp(-F(\omega))$, $\omega \in \Omega_S^\ell(\xi)$.

(d) Let

$$q = \left(\frac{\rho}{\lambda}\right)^{n_0 - n} \frac{K(\xi,\ell_0)}{K(\xi,\ell)},$$

where $\ell_0 = \langle \chi^S \rangle - \langle \chi^{\partial S} \rangle$ and n_0 is the cardinality of ℓ_0. Change the colouring on S to χ_*^S if q is greater than 1. If q is less than 1, then with probability q change the colouring to χ_*^S and with probability $1-q$ leave the colouring unchanged.

(e) Go to (a).

The algorithm is a special case of the general class of algorithms discussed by Hastings (1970). If we let

$$P(A \mid \ell) = K(\xi,\ell) \sum_{\omega \in \Omega_S^\ell(\xi)} 1_A(\omega) e^{-F(\omega)},$$

then the probability distribution for the candidate colouring χ_*^S is

$$\sum_{n=0}^{\infty} \frac{e^{-\rho d_S}(\rho d_S)^n}{n!} \int_{\mathcal{L}_S^n} P(A \mid \ell)\, \nu_S^n(d\ell),$$

which has density

$$e^{-(\rho-\lambda)d_S}(\rho/\lambda)^n e^{-F(\chi_*^S)}K(\xi,\ell),$$

with respect to $\gamma_S(\cdot \mid \xi)$. The expression in (d) is therefore the appropriate ratio of required and sampled densities. The parameter ρ can be adjusted to maximise the acceptance probability.

References

Arak, T. (1982). On Markovian random fields with a finite number of values. *4th USSR-Japan Symposium on Probability Theory and Mathematical Statistics: abstracts of communications*, Tbilisi.

Arak, T. and Surgailis, D. (1989). Markov fields with polygonal realisations. *Probability Theory and Related Fields* 80, 543–579.

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

Baddeley, A. and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, to appear.

Besag, J.E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society B* 34, 75–83.

——— (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B* 36, 192–236.

——— (1983). Discussion of paper by P. Switzer. *Bulletin of the International Statistical Institute* 47, 77–92.

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

Darroch, J.N., Lauritzen, S.L., and Speed, T.P. (1980). Markov fields and log-linear interaction for contingency tables. *Annals of Statistics* 8, 522–539.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I.E.E.E. Transactions of Pattern Analysis and Machine Intelligence* 6, 721–741.

Grimmett, G.R. (1973). A theorem about random fields. *Bulletin of the London Mathematical Society* 5, 81–84.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.

Kendall, M.G. and Moran, P.A.P. (1963). *Geometrical Probability*. Charles Griffin, London.

Kotz, S. and Johnson, N.L. (1983). *Encyclopædia of Statistical Science*. Wiley, New York.

Ove, F. and Strauss, D.J. (1981). Markov graphs. *Journal of the American Statistical Association* 86, 832–842.

Preston, C.J. (1973). Generalised Gibbs states and Markov random fields. *Advances in Applied Probability* 5, 242–261.

——— (1974). *Gibbs States on Countable Sets.* Cambridge University Press, Cambridge.

Ripley, B.D. and Kelly, F.P. (1977). Markov point processes. *Journal of the London Mathematical Society* 15, 188–192.

Rozanov, Yu.A. (1982). *Markov Random Fields.* Springer-Verlag, New York.

Sherman, S. (1973). Markov random fields and Gibbs random fields. *Israeli Journal of Mathematics* 14, 92–103.

Spitzer, F. (1971). Markov random fields and Gibbs ensembles. *American Mathematical Monthly* 78, 142–154.

Strauss, D.J. (1975). A model for clustering. *Biometrika* 62, 467–475.

Student (1914). The elimination of spurious correlation due to position in time or space. *Biometrika* 10, 179–180.

Mathematical Institute
24–29 St. Giles
Oxford OX1 3LB.