SIAM REVIEW
Vol. 28, No. 4, December 1986

004

ON A GENERAL CLASS OF MODELS FOR INTERACTION*

DAVID STRAUSS†

Abstract. This paper develops a class of probability models for configurations of interacting points in a domain. The distributions depend on a function which may be viewed as giving the potential energy of the configurations. Examples include models for interaction in a spatial region and on lattices and graphs. New models in the general class arise naturally, an example being a spatial model for points of different categories. Some general methods, including series expansions and a simulation method known from statistical mechanics, can be useful for many of these models. Several applications of this kind are considered, and some connections between the models and with the statistical mechanics are explored.

Key words. potential function, spatial distribution, lattice distribution, Markov graph, Metropolis method, stability, virial expansion

AMS(MOS) subject classifications. 60J45, 60XX

1. Introduction. This paper considers a class of probability models for interactive systems. Some examples where the models may be appropriate are:

(a) Patterns of points in Euclidean space, the points perhaps representing molecules of a gas, plants or insects in a field, or village on a map. In each case there may be attractions or repulsions between pairs of points which are close.

(b) Probabilistic colorings of a regular lattice; the colors might correspond to presence or absence of a gas molecule at a given site, or to the types of vegetation in a digitized earth satellite photograph. Again, there will typically be clustering or repulsion.

(c) In a random graph, perhaps representing a social network, an edge between two vertices might represent acquaintance between two individuals. Interactions between the edges may arise as a clustering tendency or a bias towards clique formation.

(d) Sequences of events in time, such as accidents in a large factory; the occurrence of an event at a given time may affect the likelihood of another event in a short period thereafter.

The common feature of the models considered is that the probability of a configuration depends on an "energy" function measuring the amount of interaction. Some models of this type are well known in statistical mechanics as Gibbs ensembles. In this paper, we develop a unified framework for studying interactive systems. This includes the standard physics models arising from examples (a) and (b) above, and also encompasses others, such as the interactive graph models, which do not seem to arise in physics. Because applications arise in fields such as sociology and botany, we shall be concerned with statistical questions, such as parameter estimation.

A unified framework for the models confers several benefits. It provides a convenient setting for the development of new models, an example being the spatial model for colored points considered in §6. Further, there are general methods and properties that may be applicable to any potential model. We shall see, for instance, that certain Markov graph models display a "degeneracy" closely related to the well-known instability of a spatial model. The emphasis of the paper is on general methods, and we

* Received by the editors January 16, 1985, and in revised form October 8, 1985.
† Department of Statistics, University of California at Riverside, Riverside, California 92521.

attempt to bring out some connections between the models and with statistical mechanics.

The potential models, as they will be called, have two components: a null measure μ and a potential function U. The measure determines the "structure" by specifying a neutral distribution for the number, location and coloring of the points. For example, in some applications μ specifies a fixed number of points independently and uniformly distributed in the domain. The function U may then be viewed as defining the potential energy of interaction for each configuration; in most cases U will be taken to be a sum of pairwise interactions depending on the color and interpoint distances. A model is specified by a probability density, with respect to μ, proportional to $\exp(-U)$.

Section 2 introduces formalism for potential models and indicates the connection with Markov fields and with an entropy argument. Section 3 outlines some parametric models for spatial, lattice, and graph distributions. The three remaining sections take up some general methods and apply them to various models. Section 4 concerns the possibility that the potential tends to its minimum as the system becomes large; this will be called *degeneracy*. Section 5 reviews a simulation technique known from statistical mechanics and adapts it to the Markov graph models. Section 6 considers some series expansions for the awkward normalizing constant in the probability distributions. As illustration, this is used to provide an easy derivation of some new results for a spatial model.

2. Preliminaries. In this section we define the potential models and discuss some principles from which they may be derived. To encompass the range of applications the notation needs to be rather general.

2.1. Notation and definitions. For $n \geqq 1$, let $x_1, \cdots, x_n$ be elements ("points") of a set D. This may, for example, be the set of sites of a lattice or a subset of Euclidean space. Let C be a finite set of *colors*, or *marks*, and associate with each x_i a color $y_i \in C$. For each n, the realizations, or states, are elements of $\Omega_n = \{(x_1, y_1), \cdots, (x_n, y_n)\}$, with $x_i \in D$, $y_i \in C$. We set $\Omega_0 = \varnothing$. A general state, denoted by S, is an element of the sample space $\Omega = \bigcup_{n=0}^{\infty} \{D^n \times C^n\}$.

A *potential function* U is a real symmetric function on Ω; it gives the "interactive potential energy" of configurations $S \in \Omega$. Let μ be a measure on a suitably chosen σ-field of subsets of Ω. The number of points n may be fixed or random, depending on the choice of μ. We can then define a potential model as a probability distribution P on Ω whose Radon–Nikodym derivative with respect to μ is

$$\frac{dP}{d\mu}(S) = \frac{\exp\{-U(S)\}}{Z}, \tag{2.1}$$

provided this exists. The normalizing constant, or partition function, in (2.1)

$$Z = \int \exp\{-U(S)\}\, d\mu(S) \tag{2.2}$$

will play an important role in what follows. The probability distribution of states (2.1) is sometimes called a Gibbs ensemble.

The models to be considered all have potential functions of the form

$$U(S) = \sum_{\alpha} T_\alpha(S). \tag{2.3}$$

Each α is a set of two or more integers from $\mathbb{Z}=\{1,2,\cdots\}$, and the interaction $T_\alpha(S)$ is independent of (x_i, y_i) unless $i\in\alpha$. Terms with $|\alpha|=1$ are excluded because they are not interactive, and can thus be absorbed into μ.

Let r be a metric on D, and write r_{ij} for $\|x_i-x_j\|$. We will assume homogeneity: that the T_α in (2.3) depend on x_i, x_j only through their distances r_{ij}. In many applications each T_α will involve just two arguments; then (2.3) may be written

$$U(S)=\sum_{i<j} u_{ab}(r_{ij}) \tag{2.4}$$

where $a=y_i$, $b=y_j$, and the functions u_{ab} are the *pair-potential* functions for colors a, b. The r_{ij} have argument S, but we will frequently suppress this argument for convenience. It is usually convenient to take $u(0)=\infty$, thereby making the density P vanish when the x_i are not all distinct.

2.2. The link with random fields. Suppose that D is a *finite* set, with elements $\{1,\cdots,V\}$. If S consists of n distinct elements, with $n\leqq V$, we can assign a *null* color y_0 to the remaining $V-n$. The null color is reserved exclusively for elements not in S. There then is a $(1-1)$ correspondence between realizations on Ω and *colorings* $y=(y_1,\cdots,y_V)$ of D. We consider probability distributions $P(\mathbf{y})$ on the colorings.

Write $i\sim j$ if the distribution of y_i, conditional on $\{y_k: k\neq i\}$, is dependent on y_j. The relation $\sim$ is a symmetric one. Such pairs (i, j) are *neighbors*; they define an undirected *dependence graph* $G=\{(i, j): i\sim j;\ 1\leqq i\leqq j\leqq V\}$. A probability distribution on D, together with its dependence graph, specifies a *Markov random field*. A set $A\subseteq D$ is a *clique* if it has only one element or if $(i, j)\in G$ for all $i, j\in A$. Suppose that $P(\mathbf{y})>0$ for all $\mathbf{y}$, and that there is a coloring $\mathbf{y}=\mathbf{0}$. According to the Hammersley–Clifford theorem (Besag (1974)), $Q(\mathbf{y})\equiv\ln\{P(\mathbf{y})/P(\mathbf{0})$ can be expressed uniquely as

$$Q(\mathbf{y})=\sum_{1\leqq i\leqq V} y_i\lambda_i(y_i)+\sum\sum_{i<j} y_iy_j\lambda_{ij}(y_i, y_j)+\cdots + y_1\cdots y_V\lambda_{1\cdots V}(y_1,\cdots,y_V), \tag{2.5}$$

where $\lambda_A\equiv 0$ unless A is a clique. Subject to this the λ's may be chosen arbitrarily. Now

$$U(S)=Q\{\mathbf{y}(S)\}-L\{\mathbf{y}(S)\}, \tag{2.6}$$

where L depends only on the linear terms in (2.5). Hence *the interactions* $T_\alpha(S)$ *in* (2.3) *are identically zero unless the elements of* α *form a clique in* D. Thus the dependence graph G specifies which tems appear in (2.3).

Equation (2.6) suggests that the potential function could be defined simply on the colorings of D rather than on Ω, and for finite lattices and graphs this is indeed sufficient. The full formulation of a random number of colored points at random locations seems necessary, however, if D is a continuum as, for example, in the case of spatial models.

2.3. The loglinear distributional form. In expression (2.3) for the potential function, we will generally take each $T_\alpha(S)$ to be the product of a parameter and a statistic. This results in a loglinear model for the probability distribution. Various arguments may be advanced for the loglinear form. In the case of a finite D, it arises from (2.5) with a natural parametrization. For some spatial models the loglinear form has been shown to follow from certain axioms on the probability distribution; see for example,

Kelly and Ripley (1976). Finally, the exponential dependence on U can be derived from the Maximum Entropy Principle (e.g. Jaynes (1978)). This argument, which is familiar for energy distributions in statistical mechanics, can be stated very simply and generally as follows. Let P_U be the probability of a state with potential U (for convenience we take Ω to be discrete). Take the expected potential ΣUP_U as fixed. The task is to maximize the entropy $-\Sigma P_U \ln(P_U)$ with respect to the distribution P, subject to $\Sigma P_U = 1$ and $\Sigma UP_U = \text{constant}$. A straightforward application of Lagrange multipliers then shows that $\ln(P_U)$ is a linear function of U.

3. Some potential models. In this section we briefly describe some models expressible in the general form (2.1). Various properties of the models will be discussed in subsequent sections.

3.1. Spatial models. We begin with the one color case with a fixed number of points. Take μ to be Lebesgue measure on the Borel sets of D^n, where the domain D is a bounded Borel subset of Euclidean space. The metric r is Euclidean distance. Suppose the potential function takes the pairwise additive form

$$U(S) = \sum_{1 \leqq i < j \leqq n} u(r_{ij}) \tag{3.1}$$

where $S = (x_1, \cdots, x_n)$ and $r_{ij} = |x_i - x_j|$. Then

$$P(S) = \exp\left\{ - \sum_{i<j} u(r_{ij}) \right\} \Big/ Z. \tag{3.2}$$

For various choices of u, such systems have been extensively studied in statistical mechanics, and have been applied to the statistical modelling of spatial patterns (Ogata and Tanemura (1981), (1984), Ripley (1977), (1981), Saunders et al. (1982), Strauss (1975)). We shall refer later to the "square-well potential" given by

$$u(r) = \begin{cases} \infty & \text{if } 0 \leqq r < \varepsilon, \\ -v & \text{if } \varepsilon \leqq r \leqq R, \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

and two special cases: a simple *hard core* model

$$u(r) = \begin{cases} \infty & \text{if } 0 \leqq r < \varepsilon, \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

and

$$u(r) = \begin{cases} -v & \text{if } 0 \leqq r \leqq R, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

If the number of points is to be random, we may, for example, choose μ so that n is an observation from a Poisson distribution with mean m, and μ conditional on n is uniform on D^n. In statistical mechanics a distribution defined by (2.1) with such a choice of μ is called a *grand canonical ensemble*. The distribution for fixed n is a *canonical ensemble*.

New models for colored points are simply defined for all these cases by replacing $u(\cdot)$ by a set $\{u_{ab}(\cdot)\colon a, b \in C\}$. Let there be n_a points of color a, with $\Sigma n_a = n$. Subject to this condition we may take each of $\{n_a\}$ and n to be either fixed or random by suitable choice of μ. For example, to model a data set with given $\{n_a\}$ it may be appropriate to take μ to be uniform over all configurations on D^n consistent with $\{n_a\}$

and zero elsewhere. Such models are potentially useful in biological applications; the colors might represent different species, and the interactions, perhaps corresponding to competition, will depend on the species pair in question. We will return to the colored points model in §6.

It is, of course, not necessary to take μ to be uniform over D^n. Certain choices for μ correspond to heterogeneity in the space D, such as a fertility trend in a field plot. Similar remarks apply to the lattice and graph models below.

3.2. Lattice systems. In a lattice system, the domain is taken to be a regular array of elements or *sites*. An example to be considered later is the *square lattice*, a connected subset of $\mathbb{Z}^2$, where $\mathbb{Z}$ is the set of integers. Besag (1974) discusses a number of applications. The lattice may be used to model a continuous system; an example is the so-called *lattice gas*, with two colors corresponding to presence or absence of a molecule at a site. Such representations have the advantage that the integrals arising in the spatial model are replaced by counts (of paths) in the lattice.

It will often be reasonable to take only the nearest neighbor pairs as elements of the dependence graph G. In this case we have the classical *Ising model* of statistical mechanics (Baxter (1982)). We suppose in this section that D is finite; then (2.5) applies, with $\lambda_A \equiv 0$ for $|A| > 2$. Besag (1974) discusses a number of parametric models of this form. For a realization S coloring n of the sites of D and leaving the rest empty, the potential is

$$U = \sum_{(i,j)\in G} u_{ab}(i,j) \tag{3.6}$$

where, for $(i, j) \in G$, $u_{ab}(i, j)$ is the interaction between colors a, b. Under *homogeneity* the interactions are assumed translation and rotation invariant. Then (3.6) becomes

$$U = \Sigma N_{ab} u_{ab}, \tag{3.7}$$

where N_{ab} is the number of nearest neighbor pairs with colors a, b, and u_{ab} is the corresponding nearest neighbor interaction. The simplest nontrivial case arises when there are just two colors, one of which can be regarded as null. This would be appropriate if the colors represent presence/absence. We then have $U = N\beta$, where β is an intensity parameter and N is the number of occupied nearest neighbor pairs.

The number of occupied sites n can be fixed or random. In the latter case, it is convenient to define μ so that n has a binomial distribution with parameters $(|D|, p)$ for some p in $(0,1)$ and that, conditional on n, μ is uniform over subsets of n elements of D. This formulation is again consistent with the grand canonical ensemble. The corresponding probability distribution is

$$P(S) = Z^{-1} \exp(\alpha n + \beta N) \tag{3.8}$$

where $\alpha = \ln\{p/(1-p)\}$; see, for example, Domb (1974). We shall refer to this model later.

3.3. Markov graphs. We begin by outlining some definitions and results of Frank and Strauss (1986), who also discuss sociological applications. Let I be a vertex set, with elements $i, j, \cdots$, and write I_2 for the set of unordered distinct pairs $\{i, j\}$. Elements of I_2 will be called *edges*. For simplicity we only consider here the case of an undirected finite graph H which associates one of two colors with each edge in I_2. The colors, corresponding to presence or absence of a line, are denoted by $y_{ij} = 1$ and $y_{ij} = 0$ respectively. If the y_{ij} are independent and identically distributed H is said to be a *Bernoulli graph*. H is called a *Markov graph* if y_{ij} and y_{kl}, conditional on all other

$y \in I_2$, are independent whenever $\{i, j\}$ and $\{k, l\}$ are disjoint. This means that only lines which share a common vertex are neighbors; equivalently the pair $(\{i, j\}, \{k, l\})$ is in the dependence graph G only if $\{i, j\}$ and $\{k, l\}$ intersect. Dropping some brackets, we can express the cliques in G as *triads* of form $\{ij, jk, ki\}$ and *stars of order* m of form $\{ij_1, \cdots, ij_m\}$. Under a homogeneity assumption, as in §3.2, the general potential is

$$U = \tau T + \sum_{m=2} \rho_m R_m, \tag{3.9}$$

where

$$T = \sum\sum\sum_{i<j<k} y_{ij} y_{jk} y_{ki},$$

the number of transitive triads, and $R_m = \Sigma_i \Sigma y_{ij_1} \cdots y_{ij_m}$, the number of m-stars. We absorb the term $\rho_1 R_1$ into the measure μ.

We now express the model in the general form of §2. It is convenient to take n to be the number of lines. The number of vertices is m, with $V = \binom{m}{2} \geq n$. Then D is the set of pairs $\{i, j\}$ with $1 \leqq i < j \leqq m$. A state S is an upper triangular array consisting of n ones and $V - n$ zeros. As in the lattice system, n may be taken as fixed or as binomially distributed with parameters (V, p). The neighbors of $\{i, j\} \in D$ are pairs $\{k, l\} \in D$ such that

$$\{i, j\} \cap \{k, l\} \neq \varnothing.$$

Thus as $n \to \infty$ the number of neighbors of a site grows unboundedly. This contrasts sharply with lattice systems, and has rather striking consequences as we shall see.

We shall later consider two special cases of (3.9):

$$P(S) \propto \exp(\rho R)/Z_\rho \tag{3.10}$$

where $\rho = \rho_2$, $R = R_2(S)$, and

$$P(S) \propto \exp(\tau T)/Z_\tau. \tag{3.11}$$

The first is a *clustering* model: if $\rho > 0$ the number R of interacting line pairs tends to increase, while $\rho < 0$ corresponds to repulsion between the lines. The second is a *transitivity* model.

4. Degeneracy. When n is nonrandom there are cases where the potential U will tend in probability to its minimum value as n becomes large. We shall introduce a concept of *degeneracy* to describe such behavior. Degeneracy does not necessarily mean that the model in question is unsuitable for data analysis, especially if the number of points is small. It is useful, however, to be aware of the large sample behavior. When n is random and D infinite there are cases where U does not define a model at all, in the sense that there is no finite normalizing constant for (2.1). In this section we derive some sufficient conditions for such events, and apply the conditions to the models of the previous section. It is likely that the theorem could be sharpened, but it seems in its present form to cover most cases of practical interest.

It will be convenient to distinguish the cases of random and nonrandom numbers of points.

4.1. Nonrandom number of points. Suppose that for each $n \geqq 0$ there is a null measure μ_n concentrated on $\Omega_n = D^n \times C^n$. Let P_n be the probability measure (2.1) corresponding to a potential U_n on Ω_n, and set $M_n = \inf\{U_n(S_n)\colon S_n \in \Omega_n\}$. We shall

assume that $M_n \to -\infty$ as $n \to \infty$; this will hold for all the applications considered below. For ε in the open interval $(0,1)$ write

$$A_{n,\varepsilon} = \{S_n \in \Omega_n : U_n(S_n) < (1-\varepsilon) M_n\} \tag{4.1}$$

and

$$L_{n,\varepsilon} = -\ln\{\mu_n(A_{n,\varepsilon})\}. \tag{4.2}$$

We shall say that the sequence $\{U_n\}$ is *degenerate* if, for all ε, $P_n(A_{n,\varepsilon}) \to 1$ as $n \to \infty$. The sequence is *proper* if there is an $\varepsilon > 0$ such that $P_n(A_{n,\varepsilon}) \to 0$ as $n \to \infty$.

THEOREM. (i) *The sequence* $\{U_n\}$ *is degenerate if for all* ε *in* $(0,1)$

$$\lim_{n\to\infty} \{L_{n,\varepsilon}/(-M_n)\} = 0. \tag{4.3}$$

(ii) *Given that the null mean* $E_{\mu_n}(U_n) \to -\infty$, *the sequence is proper if for some* ε *in* $(0,1)$

$$L_{n,\varepsilon} \geqq -M_n$$

for all sufficiently large n.

Proof. (i) Let δ be in $(0, \varepsilon)$. Write $\bar{A}_{n,\varepsilon}$ for $\Omega_n - A_{n,\varepsilon}$. For each n,

$$\begin{aligned} \frac{P_n(A_{n,\delta})}{P_n(\bar{A}_{n,\varepsilon})} &= \frac{\int_{A_{n,\delta}} \exp(-U_n)\, d\mu_n}{\int_{\bar{A}_{n,\varepsilon}} \exp(-U_n)\, d\mu_n} \\ &\geqq \frac{\exp\{-(1-\delta)M_n\}\exp\{-L_{n,\delta}\}}{\exp\{-(1-\varepsilon)M_n\}} \\ &= \exp\{-M_n(\varepsilon - \delta + L_{n,\delta}/M_n)\}. \end{aligned} \tag{4.4}$$

If (4.3) holds, (4.4) tends to infinity. Hence $P(\bar{A}_{n,\varepsilon}) \to 0$ as $n \to \infty$, and $\{U_n\}$ is degenerate.

(ii)

$$P_n(A_{n,\varepsilon}) = \frac{\int_{A_{n,\varepsilon}} \exp(-U_n)\, d\mu_n}{\int_{\Omega_n} (-U_n) d\mu_n} \leqq \frac{\exp\{M_n\}\exp\{-L_{n,\varepsilon}\}}{\int_{\Omega_n} \exp(-U_n)\, d\mu_n}.$$

By majorization, the denominator $\geqq \exp\{E_\mu(-U_n)\}$. It follows under the conditions of (ii) that $P_n(A_{n,\varepsilon}) \to 0$ for some ε, as required.

Degeneracy is related to the notion of instability (Ruelle (1969, §3.2)) of statistical mechanics. We shall see that in several applications $L_{n,\varepsilon}$ is bounded by a linear function of n. In this case, it follows from part (i) of the theorem that we have degeneracy if there is no $B > 0$ such that $M_n > -Bn$ for all n. This condition on M_n is just the definition of instability given by Ruelle.

4.2. Random number of points. Consider first a spatial model, D being a subset of Euclidean space. As in §3.1, we choose μ so that n has a Poisson distribution with mean m. Then a potential U on $\Omega = \bigcup_{n=0}^{\infty}(D^n \times C^n)$ defines a valid distribution if and only if

$$Z = \sum_{n=0}^{\infty} e^{-m} \frac{m^n}{n!} \int_{\Omega_n} e^{-U} d\mu \tag{4.5}$$

is finite. Let I_n be the logarithm of the integral in (4.5). It can be seen that

(i) If there is a $\delta > 0$ such that

$$n^{-1-\delta} I_n \to \infty \quad \text{as } n \to \infty, \tag{4.6}$$

then $Z = \infty$.

(ii) If there is a B such that for all n

$$n^{-1} I_n < B, \tag{4.7}$$

then $Z < \infty$.

If D is a finite set of V elements, take μ so that the distribution of n is binomial (V,p) as in §3.2. Given a potential $U(S)$ for $S \in \Omega$, we have

$$P(S) = \exp\{-U(S)\}/Z_V \tag{4.8}$$

where, corresponding to (4.5), Z_V is given by

$$Z_V = \sum_{n=0}^{V} \binom{V}{n} p^n (1-p)^{V-n} \int_{\Omega_n} e^{-U} d\mu.$$

Now suppose that we have a sequence $\{D_V\colon V = 1, 2, \cdots\}$, with corresponding potentials U_V and random variables n_V. Arguing just as in the theorem, we find that

(i) If (4.6) holds, then $\{n_V/V\} \to 1$ in probability (with respect to (4.8)) as $V \to \infty$. We shall again use the term *degenerate* for this case.

(ii) If (4.7) holds and $E_\mu\{U_V\} \to -\infty$, then for all ε in $(0,1)$

$$P\{n_V > (1-\varepsilon)V\} \to 0 \quad \text{as } V \to \infty.$$

In this case the model will again be called *proper*.

4.3. Applications. We apply these ideas to some of the models in §3.

4.3.1. A degenerate spatial model. Consider, for example, the pair potential (3.5). We discuss first the nonrandom case, with n points. If $v > 0$, the minimum energy configuration occurs when all n points are packed into a cluster with maximum separation $< R$. Denote this event by π_n. Then $\mu_n(\pi_n) > c_1 c_2^n$, for some $c_1 > 0$ and $0 < c_2 < 1$. Thus $L_{n,\varepsilon} = O(n)$, for all $\varepsilon < 1$. On the other hand $M_n = -v\binom{n}{2}$. Hence, by the theorem, the model is degenerate when $v > 0$. One way of expressing this is that, for any level of attraction v, the probability tends to 1 that an arbitrarily large proportion of the points will be packed into a tight cluster. Gates and Westcott (1982) provide calculations showing that such models, which had been fitted to empirical data sets, would almost certainly display far more clustering than was actually present in the data.

Next consider the case of random n, with μ such that n has a Poisson distribution. If $v > 0$, (4.6) holds for any $\delta < 1$. Hence in this case (3.5) does not define a model at all; in effect the system "explodes" (Kelly and Ripley (1976), Saunders et al. (1982)).

For completeness we mention the case $v < 0$, corresponding to repulsion between the points. Formally as $n \to \infty$ with D fixed the minimum energy configuration becomes "uniform." That is, the fraction of points in any region $D_1 \subset D$ tends to $\mu(D_1)/\mu(D)$. Again, $M_n = O(n^2)$ as $n \to \infty$. Hence in the limit we have a degeneracy. In most applications, however, n is small enough that $\pi R^2 n < V$, and so $M_n = 0$.

4.3.2. Lattice systems. For simplicity, we consider the two parameter model (3.8). There are V sites, of which a random number n are occupied and the remainder empty. An essential feature of lattice systems is that the number of nearest neighbors of a site

is fixed and so M_n is bounded below by $-Bn$, for some positive number B. Because of this (4.7) holds for all parameter values, and so the random n model is always proper. Similar arguments apply to other lattice models.

There is, however, the possibility of singularities in the limit as $V \to \infty$. Clearly Z in (3.8) is bounded above by $2^V \exp(BV)$, and it can be shown that the limit of $V^{-1} \ln Z$ always exists. For $\beta > \beta_c = \frac{1}{2} \sinh^{-1}(1)$ this limit as a function of α is nonanalytic at $\alpha = 0$ (Griffiths (1972)). The phenomenon is known in statistical mechanics as a *phase transition*; in the case of the lattice gas β is proportional to the inverse temperature and the gas may condense when β exceeds a *critical point* β_c (Baxter (1982, Chap. 7)). Pickard (1977) gives some distributional results for β near the critical point.

4.3.3. The Markov graph clustering model. The behavior of the ρ-model (3.10) closely resembles that of the spatial model with potential (3.5). Consider the case $\rho > 0$, with nonrandom n. Suppose that $m \sim n^\alpha$ as $n \to \infty$. We must have $\alpha \geqq \frac{1}{2}$; for ease of calculation we will assume $\alpha > 1$, though this restriction is not necessary. Since

$$A_{n,\varepsilon} \supset \{\text{all lines meet at vertex } 1\}$$

we have

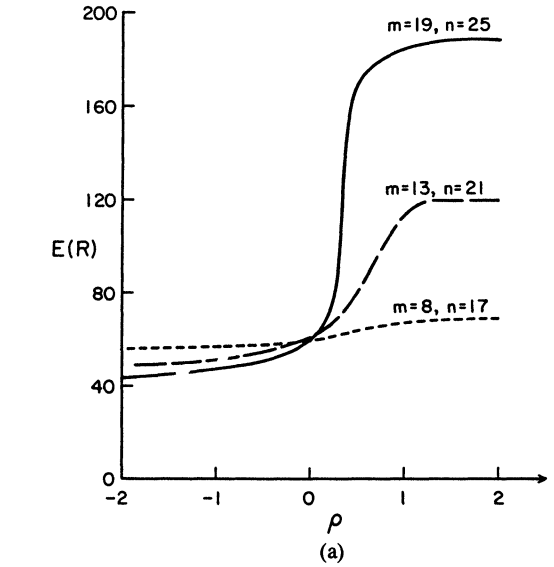
$$\mu(A_{n,\varepsilon}) > \left(\frac{2}{m}\right)^n$$

so that $L_{n,\varepsilon}$ is bounded by $An \ln n$ for some A. On the other hand $M_n = -\rho\binom{n}{2}$, so that by the theorem the model is degenerate. An interpretation is that for any value of the clustering parameter ρ the probability tends to 1 that one vertex will dominate and an arbitrarily large proportion of the lines will radiate from it. (If $\alpha < 1$ the graph will be more dense, and v vertices rather than 1 will enjoy this property, where $v \sim n/m$.) Note that the degeneracy depends on the rate of growth of m with n. For example, if $m \sim \exp(n^\alpha)$ for $\alpha > 1$ it can be shown that $n^{-2} L_{n,\varepsilon} \to \infty$ as $n \to \infty$ for all ε. By part (ii) of the theorem the model is then proper. Such a rate of growth, however, gives rise to a rather uninterestingly sparse graph.

Figure 1(a) illustrates the degeneracy. Each curve is the plot of the expectation of R against ρ. The curves were obtained by a simulation method to be described in the next section; the three pairs (m,n) were chosen so that in each case $E(R)$ has the same value (60) when $\rho = 0$. Note that as (m,n) increase, the right derivative of $E(R)$ at the origin increases. In the limit the derivative is singular at the origin.

For the random n case when $\rho > 0$ it is straightforward to verify (4.6), so that the model is degenerate. It can also be seen that whether n is fixed or random the model is proper when $\rho < 0$, which corresponds to repulsion between the lines.

4.3.4. The Markov graph transitivity model. The properties of the τ-model (3.11) are very similar to those of the ρ-model. For simplicity, suppose $n = \binom{k}{2}$, where k is an integer. When $\tau > 0$, the minimum energy configuration for n lines is a complete subgraph of k vertices. Thus $M_n = \binom{k}{3}$, the number of distinct triads. Since $M_n = O(n^{1.5})$, we again find that the model is degenerate in the nonrandom case, and also in the random case if μ is chosen as in §4.2. The nonrandom degeneracy can be interpreted as follows: if a graph has a positive transitivity tendency (however small), then as the number of lines increases the probability tends to one that an arbitrarily large fraction of the lines will coalesce into a clique. This is illustrated by Fig. 1(b), which shows the same general features as Fig. 1(a).



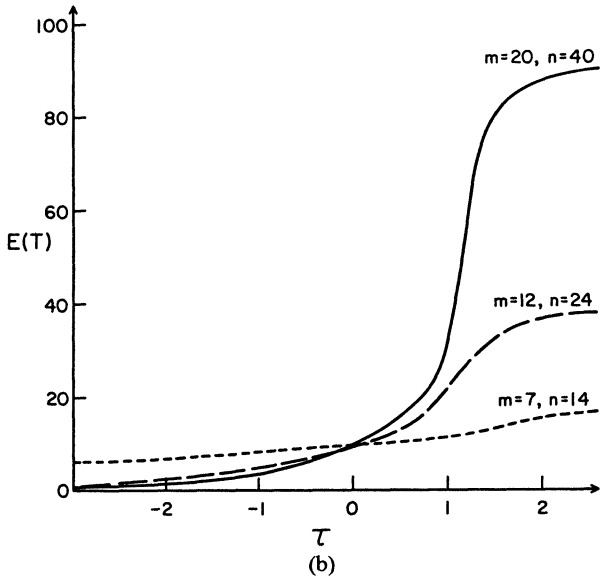


FIG. 1. *Plots of expectations against parameters for Markov graph models*; (a) $E(R)$ *against* ρ; (b) $E(T)$ *against* τ. *There are m vertices and n lines.*

5. Simulation. Simulations of the models (2.1) can be useful for a number of purposes, such as

(a) In Monte Carlo studies of the properties of estimators and test statistics.
(b) To give direct estimates of model parameters for empirical data.
(c) In Monte Carlo approximations to quantities such as Z, for comparision with analytic approximations.

Most practicable simulation methods for potential models are based on the *Metropolis method* (e.g. Hammersley and Handscomb (1964)). This generates a Markov chain on Ω whose equilibrium probabilities are of the required form (2.1). The approach has been developed in the statistical mechanics literature and has not been applied to systems such as graphs. For a review of the extensive applications to spatial and lattice schemes, see the articles in Berne (1977). In this section, we state a simplified version of the method and adapt it for the Markov graph models.

Take the state space Ω to be finite. For each $S_i \in \Omega$ define a set Ω^i of *neighbors* of S_i, with $\Omega^i \subset \Omega$ and $S_i \notin \Omega^i$. It is often convenient to set $|\Omega^i|$ equal to a constant, N say, for all i. Write P_i for $\exp\{-U(S_i)\}/Z$; note that P_i/P_j is known even if Z is not. Consider the Markov chain with transition probabilities

$$p_{ij} = \begin{cases} P_j/(NP_i) & \text{if } P_j/P_i < 1 \text{ and } j \in \Omega^i, \\ 1/N & \text{if } P_j/P_i \geqq 1 \text{ and } j \in \Omega^i, \\ c & \text{if } j = i, \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

where c is chosen so that $\sum_j p_{ij} = 1$. It is known that provided that the chain is irreducible and aperiodic it has an equilibrium distribution with probabilities P_i.

We now apply this to the Markov graph model (3.10). Suppose that there are m vertices and a fixed number n of edges (or occupied sites). We identify a realization S of the graph with its n edges. Choose any graph S_1 and generate a sequence $S_1, S_2, \cdots$ inductively as follows.

(1) At step k ($\geqq 1$), pick at random an edge $I \in S_k$ and another edge $J \notin S_k$. The graph $S' = S_k - I + J$ is a neighbor of S_k, obtained by changing one edge of S_k, and S_k, S' play the role of i, j in (5.1).
(2) Compute $\Delta R = R(S_k) - R(S')$, where R is defined in (3.10).
(3) If $\rho \Delta R \leqq 0$ set $S_{k+1} = S'$. If $\rho \Delta R > 0$, set $S_{k+1} = S'$ with probability $\exp(-\rho \Delta R)$; otherwise set $S_{k+1} = S_k$.
(4) Replace k by $k+1$ and return to (1).

The resulting Markov chain may be seen to have transition probabilities (5.1), with $N = n\{\binom{m}{2} - n\}$. Further, it is irreducible and aperiodic. It follows that it has the required equilibrium distribution (3.10).

This method was used to obtain the curves in Fig. 1(a). The procedure was

(a) Pick a value of ρ.
(b) Generate an initial configuration whose energy ρR is near the minimum.
(c) Perform a large number of steps of the simulation.
(d) Average the values of R after discarding the first few hundred (this seems sufficient to avoid the effect of the initial configuration).
(e) Repeat for other values of ρ and fit a smooth curve.

Simulation of other models in the class (3.9) is entirely analogous to the above.

The procedures may seem somewhat elaborate, but workable alternative methods are not apparent. We note that these random graph simulations are much easier to implement than, for example, Metropolis simulations for spatial models. Most of the technical problems associated with the latter (Berne (1977)) do not arise. An exception is the estimation of Z: in both cases a simple unbiased estimator for Z^{-1} is

$$D^{-n} \frac{1}{k} \sum_1^k \exp\{U(S_i)\},$$

where the S_i are realizations from the Markov chain in equilibrium, but such estimators are known to converge very slowly (Wood (1968)).

We conclude by illustrating how these simulations can be used to examine the properties of estimators. For the model (3.10) the maximum likelihood estimator of ρ satisfies

$$r = \frac{d}{d\rho} K(\hat{\rho})$$

where r is the observed value of R and $K(*)$ is its cumulant generating function when $\rho = 0$. Write the null cumulants as $\kappa_1, \kappa_2, \cdots$ as usual. If ρ is sufficiently small, we obtain a linear approximation to ρ

$$\tilde{\rho} = \frac{r - \kappa_1}{\kappa_2}. \tag{5.2}$$

The expected bias of the estimator can be shown graphically as in Fig. 2. The curve is the plot of $E(R)$ against ρ with $m = 19$ and $n = 25$, from Fig. 1(a). The oblique line is the tangent to the curve at $\rho = 0$. It can be seen that the expected proportional bias $E(\tilde{\rho} - \rho)/\rho$ is given by the ratio AB: BC. Evidently ρ is seriously biased except when ρ is small.

We note that $\tilde{\rho}$ could alternatively be used as a test statistic for the null hypothesis of randomness ($\rho = 0$).

6. Series expansions for the normalizing constant.

6.1. General considerations. It is usually necessary to know the constant Z in (2.2), at least approximately, to estimate model parameters or to compare the fit of different models. The constant can always be estimated by Monte Carlo simulation, but this may be cumbersome for practical applications. An alternative is to develop a power series for Z or $\ln Z$, the latter generally being the more convenient. The expansion may be in powers of any sufficiently small model parameter. There is an extensive literature on applications to statistical mechanics (a major reference being the volumes edited by Domb and Green (1974)), but the methods developed there have yet to be explored in other fields.

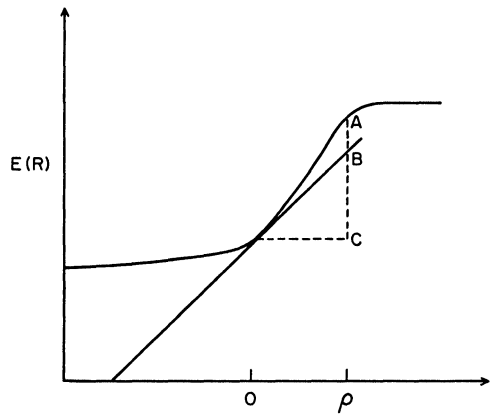


FIG. 2. *Bias of an estimator for the ρ-model. For each ρ the expected proportional bias in AB : BC.*

We indicate two general types of expansion.

(a) *Low intensity expansion.* In many applications the potential function has a natural scaling factor specifying intensity of interaction. In (3.3) this is the parameter v. Since Z is proportional to the expectation with respect to μ of $\exp(-U)$ the expansion of $\ln Z$ in powers of v is just the null cumulant expansion of U with argument $-v$. This approach has been used by Strauss (1975), (1977); it leads to "truncated" estimators such as $\tilde{\rho}$ in (5.2).

(b) *Low density expansions.* As an example, consider the square lattice model (3.8) with n sites occupied, $V-n$ empty, and N occupied nearest neighbor pairs:

$$P(S)=\exp(n\alpha+N\beta)/Z(\alpha,\beta). \tag{6.1}$$

If α is large and negative the density n/V will be small. Set $y=e^{\alpha}$ and $x=e^{-2\beta}$. It can be shown that, as $V\to\infty$, $(1/V)\ln Z$ tends to

$$-\frac{1}{2}\ln(xy)+\sum_{r=1}^{\infty} y^r g_r(x) \tag{6.2}$$

where the g_r are polynomials, the first 18 of which are tabulated by Domb (1974). Approximate maximum likelihood estimators for α,β may readily be obtained from (6.1) and (6.2) and could, for example, be compared with those from the coding methods of Besag (1974).

We note that when $\alpha=0$ the limit of $V^{-1}\ln Z$ is known in closed form: this is the classic Onsager solution (Baxter (1982, p. 110)), one of the few cases where Z is known explicitly.

6.2. A spatial model for colored points under "sparseness." We conclude by applying a low density expansion to obtain results for the colored spatial model with square-well potential (3.3). We begin with the one-color case. Let there be n points in a space of volume V. If n, $V\to\infty$ such that $\phi=n/V$ is constant, it is known that

$$\lim_{n,V\to\infty}\frac{1}{n}\ln Z=\sum_{k=1}^{\infty}\frac{\gamma_k}{k+1}\phi^k \tag{6.3}$$

where the γ_k are *irreducible cluster integrals* (Domb (1974)). Equation (6.3) is a *cluster* or *virial* expansion. Kubo (1962) gives a relatively short derivation. For a stable potential such as the square-well it is known that (6.3) has a positive radius of convergence.

The *sparseness* condition introduced by Saunders and Funk (1977) requires that $\lambda=n^2/V$ remains constant as $n,V\to\infty$ and that boundary effects are asymptotically negligible. This implies that we may neglect all powers beyond the first in (6.3). Thus we have

$$\ln Z=\frac{1}{2}\gamma_1\lambda.$$

The first cluster integral γ_1 is

$$\int_D\left[\exp\{-u(r_{12})\}-1\right]dr_{12}.$$

For the square-well potential in two dimensions this is

$$\gamma_1=\pi(e^{v}-1)(R^2-\varepsilon^2).$$

It can be shown that in the case of colored points there is a modified virial expansion in which γ_1 is replaced by its expectation over the coloring distribution prescribed by μ. We take μ such that for each n the numbers of points of color c are multinomially distributed with parameters n and $\{\theta_c\colon c\in C\}$. Let the pair-potential for colors c,d be

$$u_{cd}(r)=\begin{cases}\infty & \text{if } 0\leqq r<\varepsilon_{cd},\\ -v_{cd} & \text{if } \varepsilon_{cd}\leqq r\leqq R_{cd},\\ 0 & \text{otherwise}\end{cases} \tag{6.4}$$

with $\varepsilon_{cd}>0$. For any configuration S_n write

$$Y_{cd}^{(n)}=Y_{cd}^{(n)}(S_n)=\Sigma I\left[\varepsilon_{cd}\leqq r_{ij}\leqq R_{cd}\right],$$

the sum being over pairs (i,j) with colors c,d. Thus $Y_{cd}^{(n)}$ is the number of interacting pairs of color c,d. To indicate dependence on n and $v=\{v_{cd}\}$ in (6.4), we write the partition function as $Z^*(n,v)$. The joint cumulant generating function for the $Y_{cd}^{(n)}$ can be expressed as

$$K^{(n)}(t)=\ln Z^*(n,t+v)-\ln Z^*(n,v)$$

where t denotes $\{t_{cd}\}$. Hence $K^{(n)}(t)$ converges to

$$K_Y(t)=\frac{1}{2}\lambda\pi E\left[\{\exp(v_{cd}+t_{cd})-\exp(v_{cd})\}\left(R_{cd}^2-\varepsilon_{cd}^2\right)\right]$$

where E indicates expectation with respect to colorings. Thus

$$K_Y(t)=\frac{1}{2}\lambda\pi\sum_{c,d\in C}\theta_c\theta_d\exp v_{cd}(\exp t_{cd}-1)\left(R_{cd}^2-\varepsilon_{cd}^2\right). \tag{6.5}$$

Equation (6.5) asserts that as $n\to\infty$ the $Y_{cd}^{(n)}$ *converge to independent Poisson variables with means* $m_{cd}\exp(v_{cd})$ *where*

$$m_{cd}=\frac{1}{2}\lambda\pi\theta_c\theta_d\left(R_{cd}^2-\varepsilon_{cd}^2\right)$$

is the mean of Y_{cd} *in the null case. Thus the only effect of the interaction on the counts* Y_{cd} *is to inflate the means of their null (Poisson) distribution independently by factors* $\exp(v_{cd})$. It can also be seen that the sufficient statistics for the triples $(\varepsilon_{cd},R_{cd},v_{cd})$ are independent, so that inference for each color pair may be performed independently. Thus if the sparseness condition holds, estimation and hypothesis testing for the clustering parameters becomes straightforward.

This rather simple derivation illustrates the power of the cluster expansion. For example, in the one color case and in the absence of all interactions, the result reduces to Theorem 1 of Saunders and Funk (1977), and their Theorem 2 can be rapidly obtained if (6.4) is replaced by a suitable step function. The original proofs were lengthy.

REFERENCES

R. J. BAXTER, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.

J. E. BESAG, *Spatial interaction and the statistical analysis of lattice systems (with discussion)*, J. Roy. Stat. Soc. B, 36 (1974), pp. 192–236.

B. BERNE, ed., *Statistical Mechanics. Part* A: *Equilibrium Techniques*, Plenum Press, New York, 1977.

C. DOMB, *Ising model*, in Phase Transitions and Critical Pheonomena, Vol. 3, C. Domb and M. S. Green, eds., Academic Press, London, 1974, pp. 357–784.

C. DOMB AND M. S. GREEN, *Phase transitions and critical phenomena*, Vols. 1 and 3, Academic Press, London, 1974.

O. FRANK AND D. J. STRAUSS, *Markov graphs*, J. Amer. Statist. Assoc., 1986, to appear.

D. J. GATES AND M. WESTCOTT, *Clustering estimates in spatial point processes with unstable potentials*, unpublished, 1982.

R. B. GRIFFITHS, *Rigorous results and theorems*, in Phase Transitions and Critical Phenomena, Vol. 1, C. Domb and M. S. Green, eds., Academic Press, London, 1972, pp. 7–109.

J. M. HAMMERSLEY AND D. C. HANDSCOMB, *Monte Carlo Methods*, Methuen, London, 1964.

E. T. JAYNES, *Where do we stand on maximum entropy*? in The Maximum Entropy Formalism, E. Levine and M. Tribus, eds., MIT Press, Cambridge, MA, 1978, pp. 15–118.

F. P. KELLY AND B. D. RIPLEY, *A note on Strauss' model for clustering*, Biometrika, 63 (1976), pp. 357–360.

R. KUBO, *Generalized cumulant expansion method*, J. Phys. Soc. Japan, 17 (1962), pp. 1100–1120.

R. J. KRYSCIO, R. SAUNDERS AND G. M. FUNK, *Normal approximations for binary lattice systems*, J. Appl. Prob., 17 (1980), pp. 674–685.

Y. OGATA AND M. TANEMURA, *Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure*, Ann. Inst. Stat. Math., 33 (1981), pp. 315–338.

_____, *Likelihood analysis of spatial point patterns*, J. Roy. Stat. Soc. B, 46 (1984), pp. 496–518.

D. K. PICKARD, *Asymptotic inference for an Ising lattice*. II, Adv. Appl. Prob., 9, (1977), pp. 476–501.

B. D. RIPLEY, *Modelling spatial patterns* (*with discussion*), J. Roy. Stat. Soc. B, 39 (1977), pp. 172–212.

_____, *Spatial Statistics*, New York, John Wiley, 1981.

D. RUELLE, *Statisical Mechanics*: *Rigorous Results*, Benjamin, New York, 1969.

R. SAUNDERS AND G. M. FUNK, *Poisson limits for a clustering model of Strauss*, J. Appl. Prob., 14 (1977), pp. 776–784.

R. SAUNDERS, R. J. KRYSCIO AND G. M. FUNK, *Limiting results for arrays of binary random variables on rectangular lattices under sparseness conditions*, J. Appl. Prob., 16, (1979), pp. 554–566.

_____, *Poisson limits for a hard core clustering model*, Stoch. Proc. Appl., 12 (1982), pp. 97–106.

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

_____, *Clustering on colored lattices*, J. Appl. Prob., 14 (1977), pp. 135–143.

W. W. WOOD, *Monte Carlo studies of simple liquid models*, Physics of Simple Liquids, H. N. V. Temperley et al., eds., North-Holland, Amsterdam, 1968, pp. 115–230.