MARKOV FIELDS ON FINITE GRAPHS AND LATTICES*

by

J.M. Hammersley(1) and P. Clifford(2)

(1)University of California, Berkeley and Oxford University

(2)University of California, Berkeley and Bristol University

1. Introduction

A previous paper [1] by one of us gave an elementary description of the two-state Markov field on the simple cubic lattice, and referred the reader to Spitzer [2] for a proof that this Markov field was a Gibbsian ensemble. The main shortcoming of both papers [1] and [2] is that neither goes beyond this very special, and indeed rather trivial, case. Physicists are more interested in more general Markov fields on more general lattices; and the special case does not reveal what happens in general. For example, only pair interactions can arise on the simple cubic lattice, whereas quadruple interactions can occur on the (physically more interesting) face-centred cubic lattice . So, in the present paper, we attack the general situation. This requires more powerful mathematical tools than the trivial special case; and we develop an operational calculus (the blackening algebra) to prove our results. This algebra, once developed, yields quicker and clearer proofs than Spitzer's rather circuitous and opaque arguments.

Suitable notation is a source of trouble in this subject. It must not be so light and casual that ambiguity cannot be resolved by the context, and yet it must not be so complex and informative that it becomes

*This investigation was supported by PHS Research Grant No. GM-10525-08, National Institute of Health, Public Health Service.

(as with Dobrushin [3]) almost inpenetrable. Our notation here stems from [1]; but we have elided and lightened it in various ways, which we hope will improve its readability. Again, in [1] and [2], the three postulates of a Markov field (positivity, neighbourhood condition, and stationarity) were stated simultaneously; but it reveals things better, we believe, to state and investigate the three separately, and this we do here. In a sense, positivity and stationarity are minor irrelevances: the former has little physical import and is only a mathematical device for simplifying the proofs; while the latter restricts attention, perhaps unduly, to simple symmetric situations with simple conclusions. Indeed, we shall begin with arbitrary finite graphs, rather than lattices. For graphs, the stationarity condition does not even arise. From these, we deduce as a special case the situation for lattices under stationarity. The crucial postulate for Markov fields is the neighbourhood condition; but, after some reflection, we have decided to call it the Markov condition, thereby bringing it closer to cognate subjects elsewhere in the literature. Thus the present paper differs from [1] and [2] both in notation and approach; and to this end we have made it self-contained. Some of our readers may not be familiar with graph theory; so for their benefit (and also to allow other readers to accustom themselves to our notation) we begin with a fairly lengthy section on terminology and definitions before coming to grips with Markov fields themselves.

Spitzer [2] states that results "almost identical" to his are obtained by Averintzev [4]; but we have not been able to consult this to see how far it overlaps the present work.

2. Terminology for graphs

We shall be concerned with an arbitrary undirected finite graph Z. This consists of a finite number of sites (= vertices = nodes = points, to use equivalent terms in the literature on graphs) together with a given specification that certain pairs of sites are connected together by bonds (= arcs = loops = lines = edges). Two sites, which happen to be connected together by a bond, are called neighbours of each other. In all that follows we suppose that Z is fixed, and that we know the fixed specification of which sites are neighbours of which. We specifically exclude any bonds which connect a site to itself: thus no site is to be a neighbour of itself.

We write $z_1, z_2, \ldots, z_n$ for the sites of Z. We use capital letters (like $U, V, \ldots, Y$) for sets of sites contained in Z. We write $X+Y$ for the union of X and Y (= all sites contained in X or Y or both), XY for their intersection (= all sites common to both X and Y), and $X-Y$ for their difference (= all sites contained in X but not in Y). A set may be empty, in which case it is denoted by 0.

We write z for a typical site in Z, and ∂z for the set of its neighbours. We extend this notation to a set Y as follows: ∂Y denotes all sites, which do not belong to Y, but which are neighbours of one or more sites in Y. We call ∂Y the boundary of Y. A set Y is called a clique if either (i) it consists only of a single site, or (ii) each site in Y is a neighbour of every other site in Y. We write C for the family of all cliques in Z.

3. Colourings of the graph

We proceed to colour the sites of Z. We suppose that there are

c_i different colours available for colouring the site z_i. Thus there are $c_1 c_2 \ldots c_n$ different ways of colouring the whole graph Z. We write χ for a typical one of these $c_1 c_2 \ldots c_n$ available ways of colouring Z. Thus χ is a variable which can take $c_1 c_2 \ldots c_n$ different values. We shall need a reference colour, and we choose black for this purpose. We stipulate that black is an available colour for each and every site (this stipulation is a matter of notational convenience rather than necessity, and can be brought about by independently renaming the colours at each individual site); and, to avoid triviality, we stipulate that $c_i \geq 2$ for each $i = 1,2,\ldots,n$. Apart from these two stipulations, there are no other restrictions upon colourings: the available colours at different sites may be different or the same.

Now suppose that we are given a set of sites Y and some particular colouring χ of Z. This colouring χ assigns colours to each of the sites in Z, and in particular to the sites in Y. We define χY to be the colouring assigned to Y by χ. Thus χY is a partial colouring -- in fact, just that partial colouring of Z which χ assigns for the part of Z called Y. In particular, χz denotes the colour assigned by χ to the site z. Similarly, χZ is another way of writing χ. We labour this point deliberately, because it will be essential to distinguish the entity χY from the quite different entity χ_Y to be defined next. We define χ_Y to be the colouring of Z, which colours the whole of Y black, but which agrees with the original colouring χ at every site not in Y. Thus χ_Y is the colouring derived from χ by blackening Y; of course, some of the sites in Y may have been black originally under χ -- in which case they stay black. These two notations can, and will, be used in combination. Thus,

$\chi_U\partial V$ is the partial colouring assigned to the boundary of V by the derived colouring χ_U which blackens U but otherwise matches χ. If 0 is the empty subset of Z, then χ_0 will naturally be χ itself.

The set Y is said to be light relative to χ if no site in Y is black under the colouring χ. We define $L\chi$ to be the family of light cliques relative to χ.

We write Λ for the set of all colourings χ such that Z itself is light relative to χ. Thus a colouring of Λ never assigns black to any site.

4. The blackening algebra

Let $R(\chi)$ be a real-valued function of the colouring χ of the graph Z. We develop an algebra (= linear operational calculus) to study how R changes when χ is replaced by a derived colouring χ_Y. We define the pure operator B_Y by means of

$$B_Y R(\chi) = R(\chi_Y) . \tag{4.1}$$

If we apply first B_Y to R and then B_X to the result, we shall blacken first Y and then X; so finally the sites of the union $X + Y$ will be blackened. Thus

$$B_X B_Y R(\chi) = B_X R(\chi_Y) = R(\chi_{X+Y}) = B_{X+Y} R(\chi) . \tag{4.2}$$

Since the union $X + Y = Y + X$, we have in terms of operators alone

$$B_X B_Y = B_Y B_X = B_{X+Y} . \tag{4.3}$$

Thus the pure blackening operators commute.

We now extend the definition of operators to mixed operators, namely linear combinations of pure operators, by means of

$$(4.4)\qquad (uB_U + vB_V + \cdots + yB_Y)R(\chi) = uR(\chi_U) + vR(\chi_V) + \cdots + yR(\chi_Y) ,$$

where $u,v,\ldots,y$ are arbitrary real scalars. It is easy to verify that the mixed operators will multiply according to the rule

$$(4.5)\qquad (\sum_i u_i B_{U_i})(\sum_j v_j B_{V_j}) = \sum_{ij} u_i v_j B_{U_i+V_j} .$$

Thus the blackening operators form a commutative algebra. The identity operator in this algebra is $1 = B_0$, where the suffix 0 denotes the empty subset of Z.

An operator equal to its own square is called a <u>projector</u>. Putting $X = Y$ in (4.3) and noting that the union $Y + Y = Y$, we see that $B_Y^2 = B_Y$; so every pure operator is a projector. For any two sets X and Y, the operator

$$(4.6)\qquad B_X + B_Y - B_{X+Y} = B_X + B_Y(1-B_X)$$

is also a projector, because $X + (X+Y) = Y + (X+Y) = X + Y$ and

$$(4.7)\qquad \begin{aligned}(B_X + B_Y - B_{X+Y})^2 &= B_X^2 + B_Y^2 + B_{X+Y}^2 + 2B_{X+Y} - 2B_{X+Y} - 2B_{X+Y} \\ &= B_X + B_Y - B_{X+Y} .\end{aligned}$$

In particular, putting $Y = Z - (X+\partial X)$ in (4.6) and noting that $X + Y = Z - \partial X$ inasmuch as X and ∂X are disjoint by virtue of the definition of the boundary ∂X, we may define

$$(4.8)\qquad \beta_X = B_X + B_{Z-(X+\partial X)}(1-B_X) = B_X + B_{Z-(X+\partial X)} - B_{Z-\partial X} .$$

We shall also write, when X is the single site z_i, (4.9) $\beta_i = \beta_{z_i}', \quad b_i = B_{z_i}, \quad b_i^* = B_{Z-(z_i+\partial z_i)}, \quad (i = 1,2,\ldots,n)$.

Thus

(4.10) $$\beta_i = b_i + b_i^*(1-b_i)$$

is a projector because it is a particular case of the projector β_X, which in turn is a particular case of the projector (4.6). Finally we define

(4.11) $$\beta = \beta_1\beta_2\ldots\beta_n ,$$

and remark that β is a projector because $\beta^2 = \beta_1^2\beta_2^2\ldots\beta_n^2 = \beta_1\beta_2\ldots\beta_n = \beta$

Given any blackening operator B, we define its <u>invariant</u> <u>subset</u> $I(B)$ to be the set of all functions $R(\chi)$ which are left unchanged by B. In formal terms this definition means

(4.12) $$I(B) = \{R(\cdot)\colon BR(\chi) = R(\chi) \text{ for all } \chi\} .$$

We propose to study the invariant set $I(\beta)$ corresponding to (4.11). We claim that $I(\beta)$ consists of all functions of the form βR, where R is arbitrary:

(4.13) $$I(\beta) = \{\beta R\} .$$

In the first place, if R is arbitrary, then $\beta\beta R = \beta^2 R = \beta R$ because β is a projector; so βR is unchanged by β and belongs to $I(\beta)$. On the other hand, if Q belongs to $I(\beta)$ it is unchanged by β, so $Q = \beta Q$ and Q is of the form βR for some R, in fact $R = Q$. This proves (4.13).

Next we characterize $I(\beta)$ in terms of the light cliques of Z.

Let us write

(4.14) $$B_Y = \prod_{z_i \in Y} b_i, \quad B_Y^* = \prod_{z_i \in Y} b_i^*(1-b_i) .$$

In view of (4.3) and (4.9), the first equation in (4.14) is merely a reiteration of what we already know about pure operators; on the other hand, the second equation in (4.14) is a definition of B_Y^*, and B_Y^* is in general a mixed operator. We allow (4.14) to hold for all subsets Y of Z, including the empty subset 0, in which case we interpret $B_0 = B_0^* = 1$ as the identity operator. We write Ω for the collection of all subsets Y of Z, including 0 and Z. By virtue of (4.10) and (4.11), we have

(4.15) $$\beta = \prod_{z_i \in Z} [b_i + b_i^*(1-b_i)] = \sum_{Y \in \Omega} B_{Z-Y} B_Y^* ,$$

upon expanding the product in (4.15) and using (4.14).

Now let $R(\chi)$ be an arbitrary function of χ, let Y be any fixed subset of Z, and consider $B_{Z-Y} B_Y^* R(\chi)$. First we deal with the case where Y is not the empty subset of Z. From (4.9) and (4.14) we have

(4.16) $$B_Y^* R(\chi) = \prod_{z \in Y} B_{Z-(z+\partial z)}(1-B_z) R(\chi) .$$

We shall prove that this vanishes unless Y is a clique. For suppose that Y is not a clique. Then Y contains two distinct sites, say z and ζ, such that z is not a neighbour of ζ, and hence z does not belong to $\zeta + \partial\zeta$. The operator $B_{Z-(\zeta+\partial\zeta)}$ blackens all sites except those in $\zeta + \partial\zeta$, and so it blackens z. Then the operator B_z blackens z and leaves all other sites unaffected; but, since z

has already been blackened, B_z will leave every site unaffected, that is to say B_z will behave like the identity 1 when z is already black. Hence

(4.17) $$(1-B_z)B_{Z-(\zeta+\partial\zeta)}R(\chi) = 0 \quad .$$

However, the operators on the left-hand side of (4.17) occur in the product on the right-hand side of (4.16). Hence $B_Y^*R(\chi)$ vanishes when Y is not a clique. Next suppose that Y is a clique, but not a light clique relative to χ. Then Y will contain some site z which is already black under χ; and, as before, $(1-B_z)R(\chi) = 0$. Hence (4.16) vanishes unless Y is a light clique relative to χ. Finally, if Y is the empty subset of Z, we get

(4.18) $$B_{Z-Y}B_Y^*R(\chi) = B_ZR(\chi) = R(\chi_Z) \; .$$

Inserting these results into (4.15), and recalling that $L\chi$ denotes the family of light cliques relative to χ, we find

(4.19) $$\begin{aligned} \beta R(\chi) &= R(\chi_Z) + \sum_{Y \in L\chi} B_{Z-Y}B_Y^*R(\chi) \\ &= R(\chi_Z) + \sum_{Y \in L\chi} B_Y^*R(\chi_{Z-Y}) \; . \end{aligned}$$

Next suppose that $Y \in L\chi$ and consider any site $z \in Y$. The operator B_Y^* contains the operator $B_{Z-(z+\partial z)}$ as a factor, from (4.9) and (4.14). This latter operator blackens all sites not in $z+\partial z$, and leaves the remaining sites unaffected. But since Y is a clique, the latter operator can only blacken some of the sites in $Z-Y$, and these are already black under χ_{Z-Y}. Hence $B_{Z-(z+\partial z)}R(\chi_{Z-Y}) = R(\chi_{Z-Y})$ for all $z \in Y$. So, by (4.14) and (4.19), (4.20) $$\beta R(\chi) = R(\chi_Z) + \sum_{Y\varepsilon L\chi} \prod_{z\varepsilon Y} (1-B_z)R(\chi_{Z-Y}) .$$

Now $R(\chi)$ in (4.20) is arbitrary; and we may therefore define it in terms of another arbitrary function $S(\chi)$ as follows

(4.21) $$R(\chi) = S(\chi_Z) + \sum_{X\varepsilon L\chi} S(\chi_{Z-X}) = J_S(\chi) ,$$

where $J_S(\chi)$ is an abbreviation for the central term of (4.21). Here the sum is to be taken as zero if there are no members of $L\chi$; so

(4.22) $$R(\chi_Z) = S(\chi_Z) .$$

Now suppose χ given, and consider some $Y \varepsilon L\chi$. In the derived colouring χ_{Z-Y} all sites not in Y are blackened while all sites in Y are left unaltered and therefore remain light. So Y is also a light clique relative to χ_{Z-Y}. Every subset of a light clique is a light clique; so every subset of Y also belongs to $L\chi_{Z-Y}$. On the other hand, no clique with a site outside Y can belong to $L\chi_{Z-Y}$. Hence, if we substitute χ_{Z-Y} for χ in (4.21), we get

(4.23) $$R(\chi_{Z-Y}) = S(\chi_Z) + \sum_{X\subseteq Y} S(\chi_{Z-X}) .$$

We now operate on this equation with $1-B_z$, where $z \varepsilon Y$. Here we have to be careful, because the presence of Y in the range of summation makes $R(\chi_{Z-Y})$ behave (as it were) like a non-linear function of Y. However, replacing Y by $Y-z$ in (4.23), we have

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

On subtracting (4.24) from (4.23) we get

(4.25) $$(1-B_z)R(\chi_{Z-Y}) = \sum_{z\subseteq X\subseteq Y} S(\chi_{Z-X}) \ .$$

Similarly, we may operate on (4.25) with $1-B_\zeta$ where $\zeta \varepsilon Y$; and we get

(4.26) $$(1-B_\zeta)(1-B_z)R(\chi_{Z-Y}) = \sum_{z+\zeta\subseteq X\subseteq Y} S(\chi_{Z-X}) \ .$$

Proceeding in this fashion for all $z \varepsilon Y$, we finally obtain

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

Substitution of (4.22) and (4.27) into (4.20) yields

(4.28) $$\beta R(\chi) = S(\chi_Z) + \sum_{Y\varepsilon L_X} S(\chi_{Z-Y}) = R(\chi) \ ,$$

because X in (4.21) is only a dummy variable of summation.

Now let J denote the family of functions which can be written in the form $J_S(\chi)$ for some function S. Equation (4.28) shows that any member of J belongs to the invariant set of β. Thus

(4.29) $$J \subseteq I(\beta) \ .$$

On the other hand, for any arbitrary function R, we may remark that $\prod_{z\varepsilon Y}(1-B_z)R(\chi_{Z-Y})$ is some function of χ_{Z-Y}, say $S(\chi_{Z-Y})$; and likewise we may write $R(\chi_Z) = S(\chi_Z)$. So (4.20) shows that $\beta R(\chi)$ for arbitrary R can be written in the form $J_S(\chi)$ for some suitable S; and thus (4.13) establishes that

(4.30) $$I(\beta) \subseteq J \ .$$

Combination of (4.29) and (4.30) provides

(4.31) $$J = I(\beta) \ .$$

Finally in this section on the blackening algebra, we shall prove

(4.32) $$I(\beta) \subseteq I(\beta_X)$$

for any subset X of Z, where β_X is the projector defined in (4.8). Let S be an arbitrary function of χ; and let Y be an arbitrary clique. Clearly

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

Now β_X is a linear operator, and every member of $I(\beta) = J$ can be written in the form (4.21), where the sum is over light cliques; so it will suffice to show that

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

for an arbitrary clique Y. We notice first that X, ∂X, and $Z-(X+\partial X)$ are three disjoint sets whose union is Z. Further, the neighbours of sites in X lie in X or ∂X or both. So Y, being a clique, cannot lie partly in X and partly in $Z-(X+\partial X)$. Now consider the following table with columns indexed by operators and rows by sets.

	B_X	$B_{Z-(X+\partial X)}$	$B_{Z-\partial X}$
X	0	1	0
∂X	1	1	1
$Z-(X+\partial X)$	1	0	0

In the body of the table the entry in a particular column and row is 0 or 1 according as the indexing operator blackens all or none of the sites in the indexing set. If Y has no sites in $Z-(X+\partial X)$, the table shows that B_X and $B_{Z-\partial X}$ have identical effects upon any particular site of Y, while $B_{Z-(X+\partial X)}$ leaves the colouring of Y unchanged; thus

$$(4.35)\qquad B_X S(\chi_{Z-Y}) = B_{Z-\partial X} S(\chi_{Z-Y}), B_{Z-(X+\partial X)} S(\chi_{Z-Y}) = S(\chi_{Z-Y}) \quad ,$$

which implies (4.34). On the other hand, if Y has no sites in X, the table shows similarly that

$$(4.36)\qquad B_{Z-(X+\partial X)} S(\chi_{Z-Y}) = B_{Z-\partial X} S(\chi_{Z-Y}), B_X S(\chi_{Z-Y}) = S(\chi_{Z-Y}) \quad ,$$

which also implies (4.34). This exhausts the possibilities and proves (4.32).

As a special case of (4.32) when $X = z_i$, we get

$$(4.37)\qquad I(\beta) \subseteq I(\beta_i) \quad .$$

This is true for all i, we have

$$(4.38)\qquad I(\beta) \subseteq I(\beta_1)I(\beta_2)\ldots I(\beta_n) \quad .$$

On the other hand, if $R \in I(\beta_1)I(\beta_2)\ldots I(\beta_n)$, we have

$$(4.39)\qquad R = \beta_1 R = \beta_2 R = \cdots = \beta_n R \quad ,$$

and hence $R = \beta_1\beta_2\ldots\beta_n R = \beta R$, so $R \in I(\beta)$. Therefore

$$(4.40)\qquad I(\beta_1)I(\beta_2)\ldots I(\beta_n) \subseteq I(\beta) \quad ;$$

and we conclude from (4.38) and (4.40) that

$$(4.41)\qquad I(\beta_1)I(\beta_2)\ldots I(\beta_n) = I(\beta) \quad .$$

5. Probabilities of the colourings

We suppose that a colouring ω is selected at random from the $c_1c_2\ldots c_n$ possible colourings of Z. Thus ω is a random variable. We may specify its distribution by specifying the probability that the random colouring ω is equal to a prescribed colouring χ. We write this as

$$P(\omega=\chi) = P(\chi) \quad . \tag{5.1}$$

Thus $P(\chi)$ is regarded as a specified function of χ; and we shall later impose further conditions on this function. We adopt a notation, which avoids mention of ω as far as possible, although ω must be present at the back of our minds if we try to interpret the function $P(\chi)$ in physical terms.

Since P denotes a probability, we must have

$$\textstyle\sum_\chi P(\chi) = 1 \quad . \tag{5.2}$$

Here the summation is over all $c_1c_2\ldots c_n$ possible colourings.

We also define $P(\chi Y)$ to be the probability that the random colouring ω matches the specified colouring χ on the set Y. Thus

$$P(\chi Y) = \textstyle\sum P(\chi) \quad , \tag{5.3}$$

where the sum on the right is over all χ which yield a prescribed partial colouring χY on Y. We shall also utilize other fairly self-evident notations, employing the principle that $P(\cdots)$ denotes the probability that the random colouring fulfils all of the conditions inside the brackets $(\cdots)$. Thus $P(\chi X, \chi Y)$ is the probability that ω simultaneously has the partial colouring χX on X as well as the partial colouring χY on Y. We could have written this alternatively as $P(\chi X, \chi Y) = P[\chi(X+Y)]$, but we often prefer the former to the latter as being more easily printed and more easily read. To give a few more illustrations of the notation, $P(\chi_z)$ is the probability that ω equals the derived colouring χ_z; $P(\chi_z z)$ denotes the probability that ω is black at the site z; $P(\chi \partial z)$ denotes the probability that ω has the prescribed colouring χ on the boundary of z; $P(\chi_z z, \chi \partial z)$ denotes the probability that ω is black at z and simultaneously matches the prescribed colouring χ on the boundary of z; and so on. Any of these quantities $P(\cdots)$ can be obtained from the basic probabilities $P(\chi)$ by summing $P(\chi)$ over all χ which satisfy the conditions in $(\cdots)$.

6. The positivity condition

The positivity condition is a condition imposed by hypothesis on the function P; and it postulates that for all possible χ

$$P(\chi) > 0 \quad . \tag{6.1}$$

This postulate has two consequences. First, we may define the logarithmic likelihood

$$Q(\chi) = \log P(\chi) \quad . \tag{6.2}$$

Second, the probabilities of all events in the previous section will also be positive, since they are sums of the positive quantities in (6.1). For example, $P(\chi Y) > 0$; and so we may define conditional probabilities such as

$$P(\chi X | \chi Y) = P(\chi X, \chi Y)/P(\chi Y) \quad . \tag{6.3}$$

According to the usual laws of probability, $P(\chi X|\chi Y)$ is the conditional probability that ω has the prescribed partial colouring χX on X given that ω has the partial colouring χY on Y.

The positivity condition is mathematically convenient; but it hardly seems necessary. After all, the present work is merely an exercise in the manipulation of certain rational algebraic expressions involving the quantities $P(\chi)$. We shall postulate certain equations for these rational expressions, and deduce others. Any such equation can be reduced to a polynomial equation by multiplying out the denominators, after which there is no longer cause to worry about any denominator vanishing. It appears likely that we could do without the positivity condition by first working with it imposed to obtain the appropriate polynomial conclusions, after which we allow some probabilities to tend to zero in a limiting argument, thus relaxing the positivity condition. However in this approach there are certain technical difficulties, which we have not yet surmounted though we can scarcely imagine they are serious.

7. The Markov condition

The Markov condition (previously called the neighbourhood condition in [1] and [2]) is the main postulate. In essence, it postulates that a random partial colouring of a set X depends upon what happens on the boundary of X but not on anything beyond the boundary. Formally, the Markov condition for the set X is

(7.1) $$P[\chi X|\chi(Z-X)] = P(\chi X|\chi\partial X) \quad \text{for all } \chi .$$

We call this condition $M(X)$.

The Markov condition can occur in various forms, depending upon the sets X at which we decide to postulate $M(X)$. If we only postulate $M(z)$ for a single site z, then we say that P is Markovian at z. If we postulate $M(z_i)$ for all $i = 1,2,\ldots,n$ we say that P is locally Markovian. If we postulate $M(X)$ for all subsets X of Z, we say that P is globally Markovian. Obviously, a globally Markovian P is locally Markovian; but the converse (which is true) needs proof.

We begin by proving that $M(X)$ holds if and only if the logarithmic likelihood $Q(\chi)$ belongs to the invariant set $I(\beta_X)$ discussed in the blackening algebra.

Suppose first that $M(X)$ holds. Putting $\chi = \chi_X$ in (7.1) we get

$$P[\chi_X X | \chi_X(Z-X)] = P(\chi_X X | \chi_X \partial X) \quad . \tag{7.2}$$

However, X does not intersect either $Z-X$ or ∂X; so $\chi_X(Z-X) = \chi(Z-X)$ and $\chi_X \partial X = \chi \partial X$. Therefore

$$P[\chi_X X | \chi(Z-X)] = P(\chi_X X | \chi \partial X) \quad . \tag{7.3}$$

From (7.1) and (7.3)

$$\frac{P[\chi X | \chi(Z-X)]}{P[\chi_X X | \chi(Z-X)]} = \frac{P(\chi_X X | \chi \partial X)}{P(\chi_X X | \chi \partial X)} \tag{7.4}$$

The right-hand side of this equation depends only upon the partial colourings of X and ∂X, and is independent of the partial colouring of $Z-(X+\partial X)$. Thus we may substitute $\chi_{Z-(X+\partial X)}$ for χ in the left-hand side of (7.4) without altering its value. However $\chi_{Z-(X+\partial X)}X = \chi X$, and $(\chi_{Z-(X+\partial X)})_X X = \chi_X X$. So we get

(7.5) $$\frac{P[\chi X|\chi(Z-X)]}{P[\chi_X X|\chi(Z-X)]} = \frac{P[\chi X|\chi_{Z-(X+\partial X)}(Z-X)]}{P[\chi_X X|\chi_{Z-(X+\partial X)}(Z-X)]} .$$

Now multiply the numerator and denominator of the left-hand side of (7.5) by $P[\chi(Z-X)]$, and treat the right-hand side similarly with $P[\chi_{Z-(X+\partial X)}(Z-X)]$ as common multipliers. This gives

(7.6) $$\frac{P[\chi X,\chi(Z-X)]}{P[\chi_X X,\chi(Z-X)]} = \frac{P[\chi X,\chi_{Z-(X+\partial X)}(Z-X)]}{P[\chi_X X,\chi_{Z-(X+\partial X)}(Z-X)]} ,$$

which can be written as

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

Taking logarithms we obtain

(7.8) $$(1-B_X)Q(\chi) = Q(\chi) - Q(\chi_X) = Q(\chi_{Z-(X+\partial X)}) - Q(\chi_{Z-\partial X})$$
$$= (B_{Z-(X+\partial X)} - B_{Z-\partial X})Q(\chi) ,$$

which by (4.8) reduces to

(7.9) $$\beta_X Q(\chi) = Q(\chi) .$$

This proves that

(7.10) $$Q(\chi) \in I(\beta_X)$$

when $M(X)$ holds.

Next we deal with the converse. Suppose that (7.10) is true. Then we can retrace the steps of the above argument from (7.10) to (7.7). The right-hand side of (7.7) depends only upon the partial colouring of $X+\partial X$, since the remaining sites of Z are all black. Hence it is a function of $\chi(X+\partial X)$ only, say $\lambda[\chi(X+\partial X)]$. So (7.7) yields

$$P(\chi) = \lambda[\chi(X+\partial X)]P(\chi_X) \quad . \tag{7.11}$$

Summing (7.11) over all χ with $\chi(X+\partial X)$ fixed, we get

$$P[\chi(X+\partial X)] = \lambda[\chi(X+\partial X)]P[\chi_X(X+\partial X)] \quad . \tag{7.12}$$

Noting that $\chi_X\partial X = \chi\partial X$, we obtain from (7.11) and (7.12)

$$\frac{P(\chi X|\chi\partial X)}{P(\chi_X X|\chi\partial X)} = \frac{P(\chi X|\chi\partial X)P(\chi\partial X)}{P(\chi_X X|\chi_X\partial X)P(\chi_X\partial X)} = \frac{P[\chi(X+\partial X)]}{P[\chi_X(X+\partial X)]} \tag{7.13}$$

$$= \lambda[\chi(X+\partial X)] = \frac{P(\chi)}{P(\chi_X)} = \frac{P[\chi X,\chi(Z-X)]}{P[\chi_X X,\chi(Z-X)]}$$

$$= \frac{P[\chi x|\chi(Z-X)]}{P[\chi_X X|\chi(Z-X)]} \quad .$$

Hence

$$P(\chi X|\chi\partial X) = \mu[\chi(Z-X)]P[\chi X|\chi(Z-X)] \quad , \tag{7.14}$$

where

$$\mu[\chi(Z-X)] = \frac{P[\chi_X X|\chi\partial X]}{P[\chi_X X|\chi(Z-X)]} \tag{7.15}$$

depends only upon $\chi(Z-X)$. Now sum (7.14) over all χ which have the fixed partial colouring $\chi(Z-X)$, in which case the partial colouring $\chi\partial X$ is also fixed. We get

$$1 = \mu[\chi(Z-X)] \quad , \tag{7.16}$$

since the conditional probabilities must add up to unity. Substitution of (7.16) into (7.15) yields (7.1), as required.

Having established that $M(X)$ is equivalent to (7.10), the blackening algebra delivers the remaining goods automatically. Suppose that P is locally Markovian. Then $M(z_i)$ holds for each $i = 1,2,\ldots,n$.

Consequently $Q(\chi) \varepsilon I(\beta_i)$ for all i, and hence $Q(\chi)$ belongs to the intersection $I(\beta_1)I(\beta_2)\ldots I(\beta_n) = I(\beta) \subseteq I(\beta_X)$ for any X by (4.40) and (4.34). Hence P is globally Markovian. Thus the global and local Markovian properties are equivalent, and hereafter we need only say that P is <u>Markovian</u>. In view of (4.40), P is Markovian if and only if $Q(\chi) \varepsilon I(\beta)$; and hence, by (4.31), P is Markovian if and only if $Q(\chi) \varepsilon J$. This allows us to characterize P completely. P is Markovian if and only if it can be written in the form

$$P(\chi) = \exp[J_S(\chi)] \quad , \tag{7.17}$$

where

$$J_S(\chi) = S(\chi_Z) + \sum_{Y \varepsilon L_X} S(\chi_{Z-Y}) \quad , \tag{7.18}$$

S is an arbitrary function, and the sum is over all light cliques relative to χ. At first glance it may seem surprising that (7.17) will yield unity when summed over all χ. However, this is an automatic result of the form of $J_S(\chi)$. In fact $J_S(\chi_Z) = S(\chi_Z)$ on putting $\chi = \chi_Z$ in (7.18); so (7.17) provides

$$\frac{P(\chi)}{P(\chi_Z)} = \exp[\sum_{Y \varepsilon L_X} S(\chi_{Z-Y})] \quad . \tag{7.19}$$

Thus it is only the relative probabilities which are specified, and the constant of proportionality must still be determined from (5.2).

We can now interpret (7.19) in terms of Gibbsian ensembles, as follows. Given the graph Z, consider the family C of all cliques Y in Z, together with the set Λ of all light colourings of Z (i.e., colourings χ such that Z is light relative to χ). For each clique $Y \varepsilon C$ and each light colouring $\lambda \varepsilon \Lambda$, there will correspond a light colouring λY of Y. We take any arbitrary function of Y and its light colouring λY, and write $S(Y,\lambda Y)$ for this function. We call $S(Y,\lambda Y)$ the <u>light-coloured potential function</u>. Then (7.19) becomes

$$\frac{P(\chi)}{P(\chi_Z)} = \exp\Big[\sum_{Y\varepsilon L_\chi} S(Y,\chi Y)\Big] \quad . \tag{7.20}$$

We define a <u>Gibbsian ensemble</u> to be one whose probabilities can be expressed in the form (7.20) for some arbitrary light-coloured potential function defined on all cliques and all light colourings thereof. Then, more or less trivially by definition, a Markov field on a graph is a Gibbsian ensemble and vice versa. This is the generalization of Spitzer's result [2].

In [1] and [2] there was a good deal of unnecessary chatter about what happens when a Markov field in a region is conditional upon some fixed boundary values on the boundary of the region; and in [2] there was an unnecessary stipulation that the region should be connected. However, results of this kind follow immediately from (7.20) by calculating the appropriate conditional probability. For any subset X of Z, we get

$$\frac{P(\chi X|\chi\partial X)}{P(\chi_\chi X|\chi\partial X)} = \exp\Big[\sum_{Y\varepsilon L_\chi}^{X+\partial X} S(Y,\chi Y)\Big] \quad , \tag{7.21}$$

where $\sum_{Y\varepsilon L_\chi}^{X+\partial X}$ denotes summation over all light cliques in $X+\partial X$.

We have treated one particular colour, black, as the reference colour. But the choice of this colour was arbitrary; and it is interesting to note that we should have got the same result (with a suitably modified function S) by using any other colour as a reference.

Equation (7.20) expresses the Markovian probabilities in terms of the potentials. We can also solve (7.20) for the potentials in terms of the probabilities. For let Y be any clique of Z, and put $\chi = \lambda_{Z-Y}$ in (7.20). We get

(7.22)
$$\log \frac{P(\lambda_{Z-Y})}{P(\lambda_Z)} = \sum_{X \subseteq Y} S(X,\lambda X)$$

and from this we can extract the individual potentials $S(X,\lambda X)$ by the method of inclusion and exclusion. Thus

(7.23)
$$S(Y,\lambda Y) = [\prod_{z\varepsilon Y} (1-B_z)]\log P(\lambda_{Z-Y}) \quad .$$

8. Finite lattices

We obtain a finite lattice by defining the graph Z in terms of an arbitrary finite Abelian group G. For physical interpretation, we can regard the generators of G as vector displacements in a piece of Euclidean space wrapped on a torus. By taking the order of each generator large enough, and considering a suitable portion of the whole graph we can obtain the appropriate results for any bounded region of ordinary (non-toroidal) Euclidean space: in this case we shall need to fix the Markov field on the boundary of the region, and the appropriate version of (7.21) will give us the results we need.

Let $g_1, g_2, \ldots, g_n$ be the elements of G. We identify them with the corresponding sites $z_1, z_2, \ldots, z_n$ on Z. We take γ to be the identity element of G (it is, of course, one of the g_i). We write $\partial\gamma$ for a given subset of distinct elements from G with the properties (i) γ does not belong to $\partial\gamma$, (ii) $g \varepsilon \partial\gamma$ if and only if $g^{-1} \varepsilon \partial\gamma$, (iii) $\partial\gamma$ contains some set of generators of G. There are no other restrictions upon $\partial\gamma$: for instance, there is no need for the elements of $\partial\gamma$ to be independent -- indeed condition (ii) exhibits a pair of dependent elements. We write $\gamma_1,\gamma_2,\ldots,\gamma_m$ for the elements of $\partial\gamma$. Next, with any element $g \in G$ we associate a subset ∂g defined by

$$\partial g = \{g\gamma_i : i = 1,2,\ldots,m\} \quad . \tag{8.1}$$

We call any member of ∂g a <u>neighbour</u> of g. By condition (ii) above, an element g_i of G will be a neighbour of an element g_j if and only if g_j is a neighbour of g_i; and, by condition (i) no element of G is a neighbour of itself. We can now complete the structure of Z by specifying which sites are neighbours of which: naturally enough, we say that z_i and z_j are neighbours in Z if and only if their corresponding elements g_i and g_j are neighbours in G. Now let g and z denote corresponding elements of G and Z, and let h be any member of G. We write hz for the site in Z which corresponds to the element hg in G. Similarly, if $Y = \{z_a,z_b,\ldots,z_f\}$ is an arbitrary subset of Z, we write hY for the subset $\{hz_a,hz_b,\ldots,hz_f\}$. Thus multiplication in G carries over directly into Z in the natural fashion; and we can, for example, write (8.1) in the form $\partial g = \partial(g\gamma) = g\partial\gamma$. Indeed, it is easy to verify that elements of G distribute over all the set operations defined in section 2: thus $g(X+Y) = (gX)+(gY)$, $g(Z-X-\partial X) = (gZ)-(gX)-\partial(gX)$, and so on.

A <u>finite lattice</u> is, by definition, a graph Z defined in the foregoing manner.

9. The stationarity condition

The stationarity condition is a further condition which may be imposed on a Markov field when the graph is a lattice. It requires that

(9.1) $$P(\chi z|\chi\partial z) = P(\chi gz|\chi g\partial z)$$

for all sites $z \in Z$, all $g \in \partial\gamma$, and all possible colourings χ. Since $\partial\gamma$ generates G, we may replace $\partial\gamma$ by G in this postulate if we wish. We already know from section 7 that the quantities $P(\chi z|\chi\partial z)$ for all $z \in Z$ and all χ uniquely determine a Markov field. So, if they are invariant under the transformations provided by $g \in G$, all probabilities associated with the Markov field will similarly be invariant. In particular, the potential functions $S(Y,\lambda Y)$ defined for cliques and light colourings must be invariant, as (7.23) shows; that is to say

(9.2) $$S(Y,\lambda Y) = S(gY,\lambda gY)$$

for all $Y \in C$, and all $\lambda \in \Lambda$. Thus we need only specify the potential functions for cliques contained in $z+\partial z$ for any particular z; and all the remaining ones can be deduced from (9.2). Indeed not all these cliques in $z+\partial z$ will be needed, since some of them (certainly the one-site cliques) will be translated into others; and we need only one representative of each basically different clique. This will become apparent when we consider the examples in the next section.

10. Examples

The simplest case is the square lattice. Each site has four neighbours. There is just one basic type of 1-site clique, since all sites are equivalent under operations of G. There are two types of 2-site cliques, namely a vertical pair of neighbours and a horizontal pair of neighbours.

For the triangular lattice, there is just one 1-site clique, three 2-site cliques (corresponding to the three directions in the lattice), and two 3-site cliques (corresponding to triangles with orientations Δ and ∇ respectively).

The hexagonal lattice is the simplest kind of lattice in which more than one 1-site clique appears: there are two different kinds of sites according as the bonds to neighbours look like Y or λ.

For a comprehensive discussion of lattices which arise in physical applications, the reader may consult Tutton [5]. In general, the probability of a colouring will have the form

(10.1) $$P(\chi)/P(\chi_Z) = \exp(\sum_j \sigma_j s_j) \quad ,$$

where j runs over all types of distinct cliques (of all different sizes on the lattice) and over all types of light colouring of that clique, where s_j is an arbitrary function of j representing the potential function $S(Y,\lambda Y)$ for the $j^{\underline{th}}$ clique-colour combination, and where σ_j is the number of such clique-colour combinations found on Z under χ. When there are t_k different types of k-site cliques, the number of different values of j will be $\Sigma_k t_k \nu^k$ when $\nu+1$ colours (including black) are available at each site.

Values of t_k for some common lattices

Type of lattice	k = 1	k = 2	k = 3	k = 4
Square	1	2		
Triangular (= square with one diagonal)	1	3	2	
Hexagonal	2	3		
Square with both diagonals	1	4	4	1
Simple cubic	1	3		
Body-centred cubic	1	4		
Face-centred cubic	1	6	8	1
Tetrahedral (diamond)	2	4		
Simple cubic in D dimensions	1	D		

References

[1] J.M. Hammersley "Stochastic models for the distribution of particles in space" (this Symposium).

[2] F. Spitzer "Markov random fields and Gibbs ensembles" Amer. Math. Monthly 78 (1971), 142-154.

[3] R.L. Dobrushin "Gibbsian random fields for lattice systems with pairwise interactions." English translation of Funktional'nyi Analiz 2 (1968), 292-301.

[4] M.B. Averintzev "On a method of describing discrete parameter random fields." Problemy Peredači Informacii 6 (1970), 100-109.

[5] A.E.H. Tutton Crystallography and practical crystal measurement, Vols. 1 and 2 (1922) London: Macmillan.