Markov random fields and Markov chains on trees

Zachary, Stan

MARKOV RANDOM FIELDS AND MARKOV CHAINS ON TREES

Stan Zachary

(Heriot-Watt University)

A thesis submitted for the degree
of Ph.D., in the University of Durham

July, 1981

Markov random fields and Markov chains on trees

Stan Zachary

Abstract

We consider probability measures on a space S^A (where S and A are countable and the σ-field is the natural one) which are *Markov random fields* with respect to a given neighbour relation ~ on A. In particular, we study the set $G(\Pi)$ of Markov random fields corresponding to a given *Markov specification* Π, i.e. to a consistent family of "Markov" conditional probability distributions associated with the finite subsets of A.

First, we review the relation between Π and $G(\Pi)$. We consider also the representation of Π by a family of *interaction functions* associated with the simplices of the graph (A,~), together with some related problems.

The rest of the thesis is concerned with the case where (A,~) is a *tree*. We define *Markov chains* on S^A and consider their relation to the wider class of Markov random fields. We then derive analytical methods for the study of the set $M(\Pi)$ of Markov chains in $G(\Pi)$. These results are applied to homogeneous Markov specifications on regular infinite trees.

Finally, we consider Markov specifications which are either *attractive* or *repulsive* with respect to a total ordering on S. For these we obtain quite strong results, including an exact condition for $G(\Pi)$ to contain precisely one element. We thereby generalise results obtained by Preston and Spitzer for binary S.

Acknowledgements are due to:

my supervisor, Julian Besag, with whom I had many helpful discussions;

the Northern Ireland Department of Education, who provided generous financial assistance;

Mrs J Stewart, who typed the thesis, *and who showed infinite patience with the non-standard arrangement of letters on the italic golfball;*

my wife, who showed infinite patience generally.

CONTENTS

0. INTRODUCTION

This thesis is concerned with probability measures (p.m.s) on a measurable space (S^A, F), where S and A are countable sets, S^A is the corresponding product space, and F is the natural σ-field generated by the finite-dimensional cylinder sets. (Since A is countable, F simply consists of all subsets of S^A.) The p.m.s of interest are Markov random fields, defined below, with respect to a given *neighbour* relation ~ on the elements of A. These latter may then be regarded as the vertices of an (undirected) graph whose edges are defined by ~. We shall simply refer to A, taken together with ~, as the *graph* A. Our chief, though not sole, concern is with the special case where the graph A is a *tree*, that is, a connected graph which becomes disconnected when any one of its edges is removed.

Intuitively, and perhaps with regard to some physical applications in statistical mechanics and elsewhere, the elements of A may be thought of as *sites*. A state $x_i \in S$ is associated with each site $i \in A$. Thus the generic element $x_A = \{x_i\ , i \in A\}$ of S^A denotes the state of the entire system.

It is convenient to introduce the collection of coordinate random variables $\{X_i\ , i \in A\}$ where each random variable X_i maps $x_A \in S^A$ into its i^{th} coordinate x_i. Then a p.m. P on (S^A, F) defines a stochastic process which has P as its distribution. For each subset B of A we denote by $F(B)$ the σ-field generated by the vector of random variables $X_B = \{X_i\ , i \in B\}$. We also write x_B for the projection of $x_A \in S^A$ into S^B, i.e. $x_B = X_B(x_A)$. A p.m. on (S^A, F) is said to be a *Markov random field* (M.r.f.) (with respect to the neighbour relation ~), if for all finite subsets V of A

(1) $P(F/F(A-V)) = P(F/F(\partial V))$ P-a.s., for all $F \in F(V)$,

where ∂V, the *boundary* of V, is defined by

$$\partial V = \{i \in A-V : \text{there exists } j \in V \text{ with } i \sim j\}.$$

More intuitively, P is a M.r.f. if, for each finite subset V of A, the conditional distribution of the random variables X_V assciated with V, given the values of the remaining random variables X_{A-V} , depends in fact only on the values of the random variables $X_{\partial V}$ associated with the set ∂V of neighbours of V. The neighbour relation $\sim$ may be thought of as defining those pairs of sites between which some form of 'interaction' occurs.

It is well-known that when a p.m. P on (S^A,F) is *strictly positive*, i.e. when its finite-dimensional marginal or cylinder p.m.s have strictly positive densities, then for P to be a M.r.f. it is sufficient for the condition (1) to hold for all subsets V of A consisting of a single element. Thus many authors take this apparently weaker condition (which we will refer to as the *local Markov* property) as defining a M.r.f., but then usually also require strict positivity. We prefer to take the earlier, stronger, definition of a M.r.f., as on occasions it enables results to be carried through in the absence of the 'strict positivity' (or some similar) condition. (See, for example, section 2.1.)

The basic problem which we consider is a special case of one which was first studied by *Dobrushin (1968)*; it was subsequently taken up by many others, probably the most general (and most abstract) treatment being given by *Preston (1976)*. In the present context it may be stated as follows. For each member V of the collection V of all finite subsets of A, we are given a stochastic kernel $\pi_V : S^A \times F(V) \rightarrow R_+$ (where R_+ is the set of non-negative real numbers). The kernel π_V essentially defines a probability distribution on $(S^A, F(V))$, or if we prefer on $(S^V, F(V))$, conditional on each possible element x_{A-V} of S^{A-V} (so that we may alternatively regard π_V as a function $S^{A-V} \times F(V) \to R_+$). Thus π_V assigns probabilities to the possible states of the system 'inside' V, conditional on each possible state of the system 'outside' V. The family $\Pi = \{\pi_V, V \in V\}$ of such kernels is required to satisfy the obvious consistency condition; following *Föllmer (1975a,b)* such a family will be referred to as a *specification*. Let $G(\Pi)$ be the set of p.m.s on (S^A, F) corresponding to the specification Π, so that a p.m. P belongs to $G(\Pi)$ if and only if for each $V \in V$, $F \in F(V)$,

$$(2) \qquad P(F/F(A-V)) = \pi_V(\cdot, F) \qquad \text{P-a.s.} \quad .$$

(In fact it is convenient to extend the domain of each kernel π_V, so that it becomes a function $S^A \times F \to R_+$ and has the property that, if $P \in G(\Pi)$, the relation (2) holds for all $F \in F$.) We wish to say as much as possible about the set $G(\Pi)$.

We remark that it is obvious that $G(\Pi)$ is a convex set so that $|G(\Pi)|$, the number of elements in $G(\Pi)$, is equal to 0, 1 or ∞. Let $E(\Pi)$ be the set of extreme points of $G(\Pi)$. It is known that these extreme points 'exist', and indeed that every element of $G(\Pi)$ is a unique convex combination of them; that is, that $G(\Pi)$ may be put into a one-to-one correspondence with the set of p.m.s on $E(\Pi)$ (with a suitable σ-field); (see *Preston (1976)* or *Dynkin (1978)* for details.) Thus, in a certain sense, it is sufficient to study $E(\Pi)$.

Given the neighbour relation $\sim$ on A, we will say that a specification Π is *Markov* if for each $V \in V$, $F \in F(V)$, the random variable $\pi_V(\cdot, F)$ is $F(\partial V)$-measurable. Thus the p.m.s corresponding to a Markov specification are

M.r.f.s. In this thesis we consider only Markov specifications, and indeed (as previously remarked) most of our work is concerned with the case where the graph defined by $\sim$ is a tree. This latter problem may be regarded as a somewhat special case, and one unlikely to have direct physical applications - except perhaps in the important instance of the *one-dimensional integer lattice*, defined by taking A to be the set of integers and consecutive pairs of these to be neighbours. Nevertheless the 'tree' problem is of some interest; for here analysis is very much simpler than for more general graphs, so that, given Π, it is frequently possible to make quite detailed studies of the structure of $G(\Pi)$ and its various associated subsets; further we may often make explicit constructions of many of the elements of $G(\Pi)$. The methods of construction used are such as to provide some insight into the relation between a given Markov specification and the set of associated M.r.f.s, and there is perhaps a possibility that some of these methods might be extended to more general graphs. Lastly, there is some hope,as yet quite unfulfilled, that results for Markov specifications on trees may provide bounds for results about related specifications on more complex graphs. An example is the question of the number of elements of $G(\Pi)$. Because of its importance in statistical mechanics, the d-dimensional integer lattice ($A = Z^d$ with the obvious 'nearest' neighbour relation) is of particular interest in this respect.

We consider briefly the particular case where the tree A is the one-dimensional integer lattice defined above. Here, to show that a p.m. P on (S^A, F) is a M.r.f., it is sufficient to verify the defining relation (1) for each subset V of A which consists of a 'string' of consecutive integers; (see section 2.1). For any such V the 'conditioning' involved in (1) has a 'two-sided' nature; this is in contrast to the 'one-sided' conditioning involved in the usual definition of a Markov chain. It is easy to see that Markov chains are M.r.f.s, though the converse is not in general true. (For an example of a M.r.f. which is not a Markov chain, see *Cox (1977)*.) Now let Π be a given strictly positive, translation-invariant, specification. Many authors, especially *Spitzer (1975a)*, *Kesten (1976)* and *Cox (1979)*, have studied the general problem of characterising $G(\Pi)$. It is known (*Spitzer (1975a)*) that the elements of $E(\Pi)$ are Markov chains. Thus a complete description of the set $M(\Pi)$ of Markov chains in $G(\Pi)$, were it available, would provide an essentially complete description of $G(\Pi)$ itself. One complexity is that, even though Π is translation-invariant, the corresponding Markov chains need not be. Denote by $G_0(\Pi)$ the translation-invariant (i.e. stationary) M.r.f.s in $G(\Pi)$. *Kesten (1976)* showed that $G_0(\Pi)$ is either empty or else contains just one element which is then a (stationary) Markov chain. When $G_0(\Pi)$ is empty, $G(\Pi)$ may be empty or may contain infinitely many elements; (obviously it cannot contain just one). When $G_0(\Pi)$ contains one element, $G(\Pi)$ may be equal to $G_0(\Pi)$ or may contain infinitely many elements. *Spitzer (1975a)* showed that all these possibilities can actually occur. However, when the state space S is finite the situation is much simpler: it is known (*Dobrushin (1968)*, *Spitzer (1971)*) that $G(\Pi)$ always consists of a single stationary Markov chain. For countably infinite S a complete characterisation of $G(\Pi)$ is not in general known.

We now allow the graph A to be a general tree, and consider a general Markov specification Π. There is now a considerable increase in the variety of stochastic phenomena which may be exhibited by the M.r.f.s comprising $G(\Pi)$; some examples are given below. A natural generalisation of the concept of a Markov chain exists:

a p.m. P on (S^A, F) is a *Markov chain* (with respect to the neighbour relation ~ which defines the tree structure) if, for every finite *connected* subset V of A, the associated marginal p.m. P_V (induced by P on $(S^V, F(V))$) is a M.r.f. with respect to the restriction of ~ to V. (We will see in section 2.1 that on the one-dimensional integer lattice this definition agrees with the usual one for a Markov chain.) As previously, every Markov chain is a M.r.f., though the converse result is false; but for the given Markov specification Π the set $E(\Pi)$ of extreme points of $G(\Pi)$ is contained, sometimes strictly, in the set $M(\Pi)$ of Markov chains in $G(\Pi)$. Hence the problem of describing $G(\Pi)$ may again be reduced to that of describing $M(\Pi)$.

Much of this thesis is concerned with the development of analytical methods for studying $M(\Pi)$. It seems, however, that the usual approach to the study of Markov chains, which is via their transition matrices (so that the chains are thought of as 'evolving' sequentially), is not entirely natural in this situation; for example different chains in $M(\Pi)$ may have different (sets of) transition matrices. The approach taken here is based on the well-known representation of a Markov specification Π in terms of a family $\Phi = \{\phi_C , C \in C\}$ of *interaction functions* - one such function ϕ_C being associated with each member C of the set C of *cliques* or *simplices* of the graph. In this case, where the graph A is a tree, the cliques are simply its vertices and edges. Now if $P \in M(\Pi)$, the marginal p.m. P_V associated with each finite connected subset V of A corresponds to a family $\Phi^V = \{\phi_C^V , C \in C , C \subset V\}$ of interaction functions associated with the vertices and edges of V. The relations between the various families Φ^V, as V varies, form the basis of our study of $M(\Pi)$, and thus ultimately of $G(\Pi)$.

*Spitzer (1975b)** considered the regular infinite tree A with d + 1 edges meeting at each vertex, the binary state space S = {0,1}, and a strictly positive Markov specification Π which was *homogeneous* in the sense of being invariant under graph isomorphisms of the tree. Here, in spite of the simple nature of S, we find that even $M_0(\Pi)$ - defined to be the set of Markov chains in $M(\Pi)$ which are themselves homogeneous in the above sense - may contain more than one element; (it must always contain at least one). We further find that there may exist non-homogeneous Markov chains, and that a suitable mixture of these may result in an element of $G(\Pi)$ which is homogeneous but not a Markov chain. In addition to these and other results, Spitzer gave a precise and computable condition for $G(\Pi)$ to contain exactly one element (necessarily a homogeneous Markov chain).

In Chapters 3 and 4 of this thesis we also consider the regular infinite tree A defined above; we take a general countable state space S, and again seek to characterise $M(\Pi)$, and hence $G(\Pi)$, for a given homogeneous Markov specification Π. Spitzer's approach for the binary state space, which was essentially based on consideration of the transition matrices of the elements of $M(\Pi)$, seems difficult to extend to the general case; (see, however, section 3.3). We prefer to take the alternative approach outlined above. We also manage to make some, though not a total, relaxation of the requirement that Π be strictly positive. Many of the qualitative features of Spitzer's description of $G(\Pi)$ in the case where S is binary are reproduced in the more general case.

We summarise briefly the contents of four remaining chapters of the thesis. In Chapter 1 we review some basic results concerning, on the one hand, the relation between specifications and their corresponding probability measures

* See the note at the end of this chapter.

(or *random fields*), and on the other, the relation between families of interaction functions and their corresponding specifications. We discuss the Markov property, for a fairly general neighbour relation $\sim$ on A, in relation to all these objects. Finally, we derive some simple results concerning the representation by interaction functions of the marginal p.m.s P_V , $V \in \mathcal{V}$, of a given M.r.f. P on $(S^A, \mathcal{F})$.

Chapter 2 is concerned with the case where the neighbour relation $\sim$ is such that the graph A is a tree. We consider Markov chains and explore their properties, together with their relation to the wider class of Markov random fields. In particular we extend to our general tree Spitzer's result for the one-dimensional integer lattice, that every M.r.f. with (almost) trivial tail σ-field is a Markov chain. In the second half of the chapter we consider a Markov specification Π and present the key theorem which enables us to study the corresponding class $\mathcal{M}(\Pi)$ of Markov chains.

In Chapter 3 we specialise further to the consideration of a regular infinite tree and a homogeneous Markov specification Π. First, in section 3.1, we use the results of Chapter 2 to establish a one-to-one correspondence between the set $\mathcal{M}_0(\Pi)$ of corresponding homogeneous Markov chains, and the set of fixed points of a transformation on (roughly speaking) the space of functions $S \to R_+$. This correspondence effectively characterises $\mathcal{M}_0(\Pi)$ and in particular enables us - at least in principle, and often in practice - to determine the number of its elements. We also give a similar result for the wider set $\mathcal{M}_1(\Pi)$ of Markov chains in $\mathcal{M}(\Pi)$ which have the property of being invariant under those graph isomorphisms of the tree in which each of the vertices is translated an even number of 'steps'. 'Complementary' pairs of such Markov chains, belonging to $M_1(\Pi)$ but not $M_0(\Pi)$, may arise naturally when, as in Chapter 4, we consider *repulsive* specifications. In section 3.2 we take an increasing sequence $\{V_n\}$ of finite subsets of A such that $A = \bigcup_{n \geqslant 0} V_n$. For each n we take a p.m. $P^{(n)}$ on (S^A, F) such that the relation (2) is satisfied for all $V \subset V_n$, $F \in F(V)$. We consider briefly conditions for the weak convergence of the sequence $\{P^{(n)}\}$ to an element P of $M_0(\Pi)$. This leads, rather informally, to the idea of identifying the 'domain of attraction' of each of the elements of $M_0(\Pi)$. Section 3.3 is concerned with some connections between our approach to the study of $M_0(\Pi)$ and that adopted by Spitzer for the binary state space. Section 3.4 works out these connections for the binary state space itself. We thus obtain examples of all the possible phenomena considered in the preceding sections of the chapter. In section 3.5 we consider a further example with $|S| = 3$.

Chapter 4 is a continuation of Chapter 3 in which we require the homogeneous Markov specification Π to be either *attractive* or *repulsive* with respect to a given total ordering on the state space S. These concepts are natural generalisations of the familiar ones for binary S (where *every* Markov specification is either attractive or repulsive). (The concept of an attractive specification may be generalised further - see *Preston (1976)*.) If Π is attractive, and additionally S has *both* a minimal *and* a maximal element, then we may identify two, not necessarily distinct, Markov chains in $M_0(\Pi)$. If they are coincident then they represent the sole element, not only of $M_0(\Pi)$, but of $G(\Pi)$. Some similar results are at least implicit in the work of *Preston (1976)*, but the results here follow particularly simply from the ideas developed earlier, and avoid recourse to what is now known as the Holley-Preston inequality. If Π is repulsive, and additionally S has *either* a minimal *or* a maximal element, then we may identify a 'complementary' pair of Markov chains in $M_1(\Pi)$. Again, if these are coincident they represent the sole element of $G(\Pi)$.

We shall find it convenient to re-introduce all the above ideas and definitions at the various points where they naturally arise. However, most of the notation introduced above will be taken as standard throughout. If B is any subset of A, we shall on occasions wish to regard S^B as a space in its own right. Then x_B, defined above as the projection onto S^B of a generic point x_A of S^A, will simply denote a generic point of S^B. Similarly $F(B)$ will then denote the natural σ-field on S^B. We shall also make other such 'natural' identifications. The conditional probability of an event F with respect to the σ-field $G \subset F$ will be denoted in the usual way by $P(F/G)$, except that when we wish to regard $P(F/F(B))$ as a function on S^B and consider its value at the point x_B, we shall write $P(F/X_B = x_B)$. We shall often simply write i for the subset $\{i\}$ of A, e.g. x_{A-i} for $x_{A-\{i\}}$.

* Note added 'in proof'. Some of the results for $S = \{0,1\}$ and A a regular infinite tree, attributed to *Spitzer (1975b)*, were in fact first proved by *Preston (1974)*, as Spitzer's paper itself makes clear.

1 SPECIFICATIONS, INTERACTIONS, AND THE MARKOV PROPERTY

1.1 Specifications

Recall that $\mathcal{V}$ is the set of finite subsets of A. We now give a slightly more formal definition of a *specification* (*Föllmer (1975b), Preston (1976)*) associated with the space $(S^A, \mathcal{F})$, as a family $\Pi = \{\pi_V, V \in \mathcal{V}\}$ of functions $\pi_V : S^A \times \mathcal{F} \to R_+$, such that for each $V \in \mathcal{V}$,

$$
(1) \quad \begin{cases}
\text{(i) } \pi_V \text{ is a stochastic kernel, i.e. for each } x_A \in S^A,\ \pi_V(x_A, \cdot) \text{ is a p.m.} \\
\text{(ii) } \pi_V(\cdot, F) \text{ is } \mathcal{F}(A-V)\text{-measurable for all } F \in \mathcal{F} \\
\text{(iii) } \pi_V(\cdot, F) = \chi_F \quad \text{for all } F \in \mathcal{F}(A-V) \\
\text{(iv) } \pi_W \pi_V = \pi_W \quad \text{for all } V, W \in \mathcal{V} \text{ such that } V \subset W.
\end{cases}
$$

Here χ_F is the function $S^A \to R_+$ defined by $\chi_F(x_A) = 1$ if $x_A \in F$, and $\chi_F(x_A) = 0$ otherwise; $\pi_W \pi_V$ is the stochastic kernel defined by: for each $x_A \in S^A$, $F \in \mathcal{F}$,

$$
\pi_W \pi_V (x_A, F) = \int_{S^A} \pi_W(x_A, dy_A) \pi_V(y_A, F),
$$

i.e. $\pi_W \pi_V (x_A, F)$ is the expectation of the random variable $\pi_V(\cdot, F)$ with respect to the p.m. $\pi_W(x_A, \cdot)$,

It is only really necessary to define the p.m. $\pi_V(x_A, \cdot)$ on $(S^A, \mathcal{F}(V))$, but it is natural to extend it, by including the condition (iii) in (1) above, to a p.m. on $(S^A, \mathcal{F})$; this brings π_V into line with the general notion of conditional probability and makes more natural the expression of other relations, such as the consistency condition (iv). Further, by condition (ii) we may also regard each kernel π_V as a function $S^{A-V} \times F \to R_+$ (or indeed $S^{A-V} \times F(V) \to R_+$), so that we will frequently write $\pi_V(x_{A-V}, F)$ for $\pi_V(x_A, F)$. Thus it is seen that π_V is uniquely determined by its corresponding *conditional density function* $\tilde{\pi}_V : S^{A-V} \times S^V \to R_+$ given by

(2) $$\tilde{\pi}_V(x_{A-V}, x_V) = \pi_V(x_{A-V}, X_V = x_V)$$

We shall sometimes find it convenient to regard $\tilde{\pi}_V$ simply as a function $S^A \to R_+$. The specification Π will be called *strictly positive* if, for each $V \in V$, the conditional density function $\tilde{\pi}_V$ takes only strictly positive values.

In terms of the conditional density functions the consistency condition (1) (iv) becomes:

(3) for all V, W $\in V$ such that $V \subset W$,

$$\tilde{\pi}_W(x_{A-W}, x_W) = h_{W,V}(x_{A-V})\, \tilde{\pi}_V(x_{A-V}, x_V), \qquad x^A \in S^A$$

for some function $h_{W,V} : S^{A-V} \to R_+$.

The function $h_{W,V}$ may be determined, in terms of $\tilde{\pi}_W$, by noting that for each $x_{A-V} \in S^{A-V}$, $\sum_{y_V \in S^V} \tilde{\pi}_V(x_{A-V}, y_V) = 1$; thus $h_{W,V}(x_{A-V}) = \sum_{y_V \in S^V} \tilde{\pi}_W(x_{A-W}, x_{W-V} \times y_V)$. When $h_{W,V}(x_{A-V}) > 0$ (as is the case when Π is strictly positive), the density $\tilde{\pi}_V(x_{A-V}, \cdot)$ is uniquely determined by $\tilde{\pi}_W$.

A p.m. P on (S^A, F) is said to be a *random field* (or *stochastic field* or *Gibbs state*) corresponding to the specification Π if

(4) $$P(F/F(A-V)) = \pi_V(\cdot, F) \qquad \text{P-a.s.}, \qquad F \in F,\ V \in V$$

(Thus condition (1)(iv) simply corresponds to the relation

$$P(F/\mathcal{F}(A-W)) = E\{P(F/\mathcal{F}(A-V))/\mathcal{F}(A-W)\} \qquad P\text{-a.s.}, \; F \in \mathcal{F}$$

valid for all V, W $\in \mathcal{V}$ such that $V \subset W$.)

It is not in general known whether every p.m. P on $(S^A, \mathcal{F})$ corresponds to some specification. Certainly since $(S^A, \mathcal{F})$ is a Borel space, P has a regular conditional distribution which is almost a specification - only 'almost' because condition (1)(iv) is only guaranteed to be satisfied P-almost surely. This dependence on P however is a serious weakness. *Goldstein (1978)* showed that when S is *finite*, every p.m. on $(S^A, \mathcal{F})$ *does* correspond to some specification. In this thesis we take a specification as given, and study the set of corresponding p.m.s.

When the set A is finite, to each specification Π there corresponds precisely one p.m. P on $(S^A, \mathcal{F})$, which may be identified with the kernel π_A. If in addition Π is strictly positive, then it follows from our earlier remarks that Π is uniquely recoverable from P. Further *every* strictly positive p.m. on $(S^A, \mathcal{F})$ corresponds to (precisely one) specification. Thus for finite A we may effectively identify strictly positive specifications and strictly positive p.m.s on $(S^A, \mathcal{F})$.

Returning to the general case of countable A, let $\mathcal{G}(\Pi)$ denote the set of p.m.s on $(S^A, \mathcal{F})$ corresponding to a given specification Π. As remarked in the Introduction, $\mathcal{G}(\Pi)$ is obviously a convex set; thus $|\mathcal{G}(\Pi)|$, the number of elements in $\mathcal{G}(\Pi)$ is equal to 0,1 or ∞. It is well-known that all of these possibilities can occur. It is further known (see e.g. *Preston (1976), Theorem 2.1*) that the set $\mathcal{E}(\Pi)$ of extreme elements of $\mathcal{G}(\Pi)$ is precisely the set of p.m.s in $G(\Pi)$ with respect to which the tail σ-field $\hat{F} = \bigcap_{V \in V} F(A-V)$ is trivial, and that distinct elements of $E(\Pi)$ are mutually singular. Indeed since (S^A, F) is a Borel space we have the following integral representation (*Preston (1976), Theorem 2.2*), which will give us various results we shall need.

<u>Theorem 1.1.1</u> (Preston)

Suppose $G(\Pi) \neq \emptyset$. Then there exists a stochastic kernel $\pi: S^A \times F \to R_+$ such that

(i) $\pi(\cdot,F)$ is $\hat{F}$-measurable, for all $F \in F$

(ii) the p.m. $\pi(x_A,\cdot) \in E(\Pi)$, for all $x_A \in S^A$

(iii) $P(F/\hat{F}) = \pi(\cdot,F)$ P-a.s., for all $F \in F$, $P \in G(\Pi)$

(iv) for each $x_A \in S^A$, if $\Delta(x_A) = \{y_A \in S^A: \pi(y_A,\cdot) = \pi(x_A,\cdot)\}$, then $\pi(x_A,\Delta(x_A)) = 1$

(v) any p.m. P on (S^A,F) belongs to $G(\Pi)$ if and only if $P = P\pi$ (where $P\pi(\cdot,F)$ is the expectation of $\pi(\cdot,F)$ with respect to P).

The deduction which we require at present from this theorem is simply that (from (v)) the p.m.s comprising $G(\Pi)$ are the mixtures of those of the form $\pi(x_A,\cdot)$, $x_A \in S^A$, these latter being extreme points of $G(\Pi)$. Indeed, letting $\tilde{X}$ denote the the space of equivalence classes of S^A defined by calling x_A and y_A *equivalent* if $\pi(x_A,\cdot) = \pi(y_A,\cdot)$, *Preston (1976 , Proposition 2.4)* also showed how to make the more formal deduction that $G(\Pi)$ is in one-to-one correspondence with the set of all p.m.s on $(\tilde{X},\tilde{F})$, where $\tilde{F}$ is the σ-field in $\tilde{X}$ which corresponds naturally to the tail σ-field $\hat{F}$ in S^A. Here $E(\Pi)$ corresponds to the point masses on $(\tilde{X},\tilde{F})$.

Thus, for the specifications Π to be considered later, we seek to describe the p.m.s in $E(\Pi)$, these being most usefully identified by the alternative characterisation given above - as the set of p.m.s in $G(\Pi)$ with trivial tail σ-field. A complete description of $E(\Pi)$, at least when this set is countable, is in an obvious sense equivalent to a complete description of $G(\Pi)$. In particular we always have that $|G(\Pi)|$ is equal to 0,1, or ∞ according as $|E(\Pi)|$ is equal to 0,1, or is greater than 1.

1.2 Interactions

We define an *interaction* on (S,A) to be a family $\Phi = \{\phi_V, V \in \mathcal{V}\}$ of *interaction functions* $\phi_V : S^V \to R_+$ such that

(1)
- (i) for each $i \in A$, there are only finitely many $W \in \mathcal{V}$ such that $i \in W$ and ϕ_W is not identically equal to one,
- (ii) for each $V \in \mathcal{V}$, $x_{A-V} \in S^{A-V}$

$$0 < \sum_{x_V \in S^V} \theta_V(x_{A-V}, x_V) < \infty$$

where $\theta_V : S^{A-V} \times S^V \to R_+$ is defined by

$$\text{(2)} \qquad \theta_V(x_{A-V}, x_V) = \prod_{\substack{W \in \mathcal{V} \\ W \cap V \neq \emptyset}} \phi_W(x_W), \qquad x_A \in S^A$$

Here we are identifying $S^{A-V} \times S^V$ with S^A. Note that since V is finite, it follows from (1)(i) that the above product is taken over only a finite number of functions ϕ_W not identically equal to one, and so θ_V is well-defined.

It is easy to verify that the interaction Φ determines a corresponding specification Π_Φ, each kernel $\pi_{\Phi,V}$, $V \in \mathcal{V}$, being given by its equivalent conditional density function

$$\text{(3)} \qquad \tilde{\pi}_{\Phi,V}(x_{A-V}, x_V) = k_V(x_{A-V})\,\theta_V(x_{A-V}, x_V), \qquad x_A \in S^A$$

where the *normalising* function $k_V : S^{A-V} \to R_+$ is chosen so that

(4) $$\sum_{x_V \in S^V} \tilde{\pi}_{\Phi,V}(x_{A-V}, x_V) = 1, \quad x_{A-V} \in S^{A-V}$$

The interaction is not, of course, uniquely recoverable from the specification, unless additional conditions are imposed on the interaction functions. With suitable such conditions an interaction is just the exponential version of an *interaction potential* (see e.g. *Preston (1976)*). The former is more convenient here as we shall not always require that the interaction functions be strictly positive.

If $P \in G(\Pi_\Phi)$ we shall say simply that P corresponds to the interaction Φ. In particular if A is finite, P is uniquely determined by

(5) $$P(X_A = x_A) = k \prod_{W \in V} \phi_W(x_W) ,$$

where k is the appropriate normalising constant; further to show that P corresponds to Φ it is sufficient to verify this relation.

1.3 The Markov Property

Henceforth we shall assume that the set A is endowed with a symmetric binary relation $\sim$ on its elements. Elements i and j of A will be called *neighbours* if $i \sim j$. We shall also write $i \not\sim j$ if i and j are not neighbours. We shall further require that the set of neighbours of each element i of A is finite and does not contain i itself. The set A, taken together with $\sim$, may thus be regarded as a graph - we simply refer to it as the *graph* A. It will be convenient to define, for each subset B of A, all the following associated sets:

the *boundary* $\partial B = \{j \in A-B : \exists\ i \in B \text{ with } i \sim j\}$

the *environment* $\xi B = B \cup \partial B$

the *internal boundary* $\tilde{\partial} B = \partial(A-B)$

the *interior* $\eta B = B - \tilde{\partial} B$

Note that all of these sets are finite when B is finite, that $\xi(\eta B) \subset B$ for all B, and that $B' \subset \eta B$ if and only if $\xi B' \subset B$. We will in general write ∂i, ξi for $\partial\{i\}$, $\xi\{i\}$ for any $i \in A$.

A subset of C of A will be called a *clique* (or *simplex*) if $|C| = 1$, or if $|C| \geqslant 2$ and every pair of elements i and j of C are neighbours. We denote by $\mathcal{C}$ the set of all cliques, and by $\mathcal{C}(V)$ the set of all cliques which are subsets of the given set V in $\mathcal{V}$. Obviously we have $\mathcal{C} \subset \mathcal{V}$.

A specification $\Pi = \{\pi_V, V \in \mathcal{V}\}$ will be called *Markov* if for each $V \in \mathcal{V}$,

(1) $\qquad \pi_V(\cdot, F)$ is $\mathcal{F}(\partial V)$-measurable for all $F \in \mathcal{F}(V)$.

As usual this relation then holds automatically for all $F \in F(\xi V)$. The kernels π_V will be referred to as Markov kernels. It will frequently be convenient to regard their corresponding conditional density functions as functions $\tilde{\pi}_V : S^{\partial V} \times S^V \to R_+$ (or $S^{\xi V} \to R_+$).

The specification Π will be called *locally Markov* if we only require (1) to hold for V such that $|V| = 1$. It is known (see Theorem 1.3.1 below) that a specification which is both strictly positive and locally Markov is in fact Markov. Hence many authors take the former two properties as defining a Markov specification.

A p.m. P on (S^A, F) will be called a *Markov random field* (M.r.f.) if for each $V \in \mathcal{V}$,

(2) $$P(F/F(A-V)) = P(F/F(\partial V)) \quad \text{P-a.s.,} \quad \text{for all } F \in F(V)$$

In particular p.m.s which correspond to Markov specifications are M.r.f.s. Conversely, every strictly positive M.r.f. P corresponds to the Markov specification Π whose conditional density functions are given by

$$\tilde{\pi}_V(x_{\partial V}, x_V) = P(X_V = x_V / X_{\partial V} = x_{\partial V}), \qquad V \in \mathcal{V}$$

- for here it is easy to check that the consistency condition (1)(iv) of section 1.1 is satisfied.

To verify that a p.m. P on (S^A, F) corresponds to a given Markov specification Π it is (necessary and) sufficient to show:

(3) for some increasing sequence $\{V_n\}$ in $\mathcal{V}$ such that $V_n \uparrow A$,

$$P(F/F(V_n - V)) = \pi_V(\cdot, F) \quad \text{P-a.s.,}$$

for all $V \in \mathcal{V}$, $F \in F(V)$ and n such that $\xi V \subset V_n$.

For then, using (for instance) the martingale convergence theorem, we have for each $V \in \mathcal{V}$, $F \in \mathcal{F}(V)$,

$$P(F/\mathcal{F}(A-V)) = \pi_V(\cdot,F) \qquad P\text{-a.s.}$$

If P is known to be a M.r.f., then obviously it is sufficient to show that for each $V \in \mathcal{V}$, $F \in \mathcal{F}(V)$

(4) $$P(F/\mathcal{F}(\partial V)) = \pi_V(\cdot,F) \qquad P\text{-a.s.}$$

An interaction $\Phi = \{\phi_V, V \in \mathcal{V}\}$ on (S,A) will be called *Markov* (sometimes referred to as *nearest neighbour*) if each interaction function ϕ_V is identically equal to one whenever V does not belong to $\mathcal{C}$. It will thus be convenient to denote a Markov interaction by $\Phi = \{\phi_C, C \in \mathcal{C}\}$. Note that the Markov requirement makes condition (1)(i) of section 1.2 redundant.

Various versions of the following theorem, usually phrased in terms of interaction potentials, are well-known. (Part (i) of the theorem is trivial.)

Theorem 1.3.1

(i) If Φ is a Markov interaction on (S,A), Π_Φ is a Markov specification - the corresponding conditional density functions being given by

(5) $$\tilde{\pi}_{\Phi,V}(x_{\partial V}, x_V) = k_V(x_{\partial V}) \prod_{\substack{C \in \mathcal{C} \\ C \cap V \neq \emptyset}} \phi_C(x_C), \qquad V \in \mathcal{V}$$

where as usual each $k_V : S^{\partial V} \to R_+$ is the appropriate normalising function.

(ii) If Π is a strictly positive, locally Markov specification, then there exists a Markov interaction Φ on (S,A) such that $\Pi = \Pi_{\Phi}$.

A corollary is that a strictly positive, locally Markov specification is Markov.

The importance of part (ii) of the theorem lies in the result that at least all those Markov specifications which are strictly positive may be represented - and thus they and their corresponding M.r.f.s studied - in terms of Markov interactions.

Versions of part (ii) of the theorem have been proved for finite A by various authors (*Brook (1964)*, *Grimmett (1973)*, and others), and in this case it has also been shown that, given some arbitrary *basepoint* element $s_A \in S^A$, the interaction functions ϕ_C, $C \in \mathcal{C}$ may be chosen so that for each clique C, $\phi_C(x_C) = 1$ whenever $x_i = s_i$ for some $i \in C$; and further that under this condition the interaction functions are determined uniquely. When this condition is satisfied we shall say that both the interaction and the corresponding interaction functions are *adapted* with respect to the basepoint s_A. These additional results are well-known and may be found, for example, in *Grimmett (1973)*.

For the more general case of countable A, the theorem has been proved by *Spitzer (1971)* for the binary state space $S = \{0,1\}$, and by *Preston (1976)* for general countable S but under the assumption (in part (ii)) that the specification Π is Markov, rather than simply locally Markov. We show here how to prove part (ii) of the theorem as stated, assuming its truth, along with the additional results stated above, for finite A. The argument is related to one used by *Spitzer (1973)*, again for the binary state space.

Let $s_A \in S^A$ be a basepoint. Fix $C \in \mathcal{C}$, $i \in C$. For any $V \in \mathcal{V}$ such that $\xi i \subset V$ (and hence $C \subset V$), and for any $x_{A-V} \in S^{A-V}$, $\pi_V(x_{A-V},\cdot)$ is a p.m. on $(S^V, \mathcal{F}(V))$ which is easily seen to be strictly positive and locally Markov with respect to the restriction of the neighbour relation $\sim$ to V. Hence, since V is finite, $\pi_V(x_{A-V},\cdot)$ is Markov on V with a unique representation of the form

$$(6) \qquad \tilde{\pi}_V(x_{A-V}, x_V) = \hat{k} \prod_{C' \in \mathcal{C}(V)} \hat{\phi}_{C'}(x_{C'}), \qquad x_V \in S^V$$

where the functions $\hat{\phi}_{C'}$ (which, along with $\hat{k}$, may depend on V and x_{A-V}) are adapted with respect to the basepoint s_A. Now let x_V, $x'_V \in S^V$ be such that $x_{V-C} = s_{V-C}$, $x'_{V-i} = x_{V-i}$, $x'_i = s_i$. Then from (6) we have

$$(7) \qquad \prod_{\substack{C' \subset C \\ \text{s.t. } i \in C}} \hat{\phi}_{C'}(x_{C'}) = \frac{\tilde{\pi}_V(x_{A-V}, x_V)}{\tilde{\pi}_V(x_{A-V}, x'_V)}$$

$$= \frac{\tilde{\pi}_i(x_{A-i}, x_i)}{\tilde{\pi}_i(x_{A-i}, s_i)}$$

Since Π is locally Markov, this last expression depends only on $x_{\xi i}$, and so the left side of (7) is independent of all V such that $\xi i \subset V$, and all $x_{A-V} \in S^{A-V}$. The same argument may be applied to all $C' \subset C$ with $i \in C'$, and thus we deduce that $\hat{\phi}_C$ itself is independent of all V such that $\xi i \subset V$ and all $x_{A-V} \in S^{A-V}$. Further by considering those V such that $\xi C \subset V$ (so that for all $i \in C$ we have $\xi i \subset V$), we see that $\hat{\phi}_C$ is independent of all V, $x_{A-V} \in S^{A-V}$, such that for *some* $i \in C$ we have $\xi i \subset V$. Thus for each $C \in \mathcal{C}$, we define $\phi_C = \hat{\phi}_C$ where $\hat{\phi}_C$ is as constructed above.

It remains to show that the Markov interaction $\Phi = \{\phi_V, C \in \mathcal{C}\}$, thus defined, determines the specification Π. Given $W \in \mathcal{V}$, take $V = \xi W$, and for each $x_A \in S^A$ consider the representation of $\tilde{\pi}_V(x_{A-V}, \cdot)$ given by (6). It follows that for the appropriate normalising function $k' : S^{A-W} \to R_+$,

$$\tilde{\pi}_W(x_{A-W}, x_W) = k'(x_{A-W}) \prod_{\substack{C \in \mathcal{C} \\ C \cap W \neq \emptyset}} \hat{\phi}_C(x_C)$$

$$= k'(x_{A-W}) \prod_{\substack{C \in \mathcal{C} \\ C \cap W \neq \emptyset}} \phi_C(x_C)$$

(since $C \in \mathcal{C}$, $C \cap W \neq \emptyset$ and $V = \xi W$ together imply that there exists some $i \in C$ with $\xi i \subset V$). Thus $\pi_W = \pi_{\Phi,W}$ as required, and the locally Markov specification Π is seen to be Markov.

We now define a Markov interaction Φ to be *hereditary* with respect to a given basepoint $s_A \in S^A$, if each interaction function ϕ_C, $C \in \mathcal{C}$, satisfies $\phi_C(x_C) > 0$ whenever $x_i = s_i$ for some $i \in C$. We will also call a (necessarily Markov) specification Π *hereditary* if there exists a hereditary Markov interaction Φ such that $\Pi = \Pi_\Phi$. If this is the case, then we have

(8) $\quad \tilde{\pi}_i(x_{\partial i}, s_i) > 0 \quad$ for all $i \in A$ and $x_{\partial i} \in S^{\partial i}$,

and this condition in turn implies that *any* Markov interaction to which Π corresponds is hereditary.

We shall chiefly be concerned with hereditary Markov specifications, analysed in terms of the corresponding interactions. Theorem 1.3.1 ensures that at least every strictly positive, (locally) Markov specification is hereditary (the interaction functions here being strictly positive). The hereditary condition may therefore be seen as a slight - though on occasions desirable - relaxation of the strict positivity condition.

Finally we remark that it would be of interest to investigate the extent to which the condition (8) might be sufficient, or almost sufficient, to ensure that a Markov specification Π is hereditary. Consider the case where A is finite, and let Π satisfy (8). Then the corresponding M.r.f. P on (S^A, F) satisfies the condition that if $x_A \in S^A$ is such that $P(X_A = x_A) > 0$, and if x'_A is such that for all $i \in A$, x'_i is equal to either x_i or s_i, then $P(X_A = x'_A) > 0$. (This is what is *usually* referred to as the 'hereditary' condition.) It is well-known that this condition is sufficient to ensure that the M.r.f. P corresponds to some hereditary Markov interaction Φ. (See, for example, the proof of our Theorem 1.3.1 in *Grimmett (1973)*.) Then at least the specification Π_Φ, which is for all practical purposes equivalent to Π, is hereditary.

1.4 Marginal distributions

If $V \in \mathcal{V}$, and P is a p.m. on $(S^A, \mathcal{F})$ we shall denote by P_V the marginal (or cylinder) p.m. induced by P on $(S^V, \mathcal{F}(V))$.

Lemma 1.4.1

Let the p.m. P correspond to the Markov interaction $\Phi = \{\phi_C, C \in \mathcal{C}\}$ on (S,A) - so that P is a M.r.f.. Then P_V corresponds to an interaction $\Phi^V = \{\phi_W^V, W \subset V\}$ on (S,V) which satisfies $\phi_W^V = \phi_W$ for all W such that $W \cap \eta V \neq \emptyset$, (where if $W \notin \mathcal{C}(V)$, ϕ_W is taken to be identically equal to one).

Proof

We have

$$P(X_V = x_V) = P(X_{\tilde{\partial} V} = x_{\tilde{\partial} V})P(X_{\eta V} = x_{\eta V}/X_{\tilde{\partial} V} = x_{\tilde{\partial} V})$$

$$= \phi^V_{\tilde{\partial} V}(x_{\tilde{\partial} V}) \prod_{\substack{C \in \mathcal{C} \\ C \cap \eta V \neq \emptyset}} \phi_C(x_C), \qquad x_V \in S^V$$

for some function $\phi^V_{\tilde{\partial} V} : S^{\tilde{\partial} V} \to R_+$. The second equation follows since P corresponds to the Markov interaction Φ, and $\partial(\eta V) \subset \tilde{\partial} V$. The required result is now immediate from the remark following (5) of section 1.2.

Under the conditions of the lemma, it follows in particular that for each $V \in \mathcal{V}$, P_V is a M.r.f. with respect to the modified neighbour relation on V given by additionally defining as neighbours all pairs of distinct elements in $\tilde{\partial} V$. (P_V is not in general a M.r.f. with respect to the restriction of the original relation $\sim$ to V.)

This result may also be proved directly for *any* M.r.f. P on (S^A, F), as follows. Define two subsets of any graph to be *separate* if they are disjoint and if no element of one is a neighbour of any element of the other. It follows easily, from the defining relations for a M.r.f., that if P is a M.r.f. on (S^A, F), then for any separate $U, W \in V$, $F(U)$ and $F(W)$ are conditionally independent (under P) with respect to $F(A-(U \cup W))$. For finite A we also have the converse result that if this condition holds for all separate subsets U,W of A, then P is a M.r.f. on (S^A, F). (Obviously, corresponding remarks also hold for Markov specifications.) Now, for general countable A with neighbour relation $\sim$, and given $V \in V$, it is easy to see that if any two subsets of V are separate with respect to the graph V given by modifying $\sim$ as described above, then they are separate with respect to the original graph A. Thus, by the above remarks, if P is a M.r.f. on (S^A, F) (with respect to $\sim$), then P_V is a M.r.f. on $(S^V, F(V))$ with respect to the modified neighbour relation.

2 TREES

2.0 Introduction

Throughout this and subsequent chapters we assume that the graph defined by the neighbour relation ~ on A is a tree.

In section 2.1 we consider the special class of those M.r.f.s on (S^A, F) which may be called Markov chains (first introduced for trees by *Preston (1974)*). We develop the relation of Markov chains to more general M.r.f.s, and in particular we generalise a result of *Spitzer (1975a)* to show that if Π is a Markov specification, then the elements of the set $E(\Pi)$ of extreme points of $G(\Pi)$ are Markov chains.

In section 2.2 we consider the Markov chains which correspond to a given hereditary Markov specification Π (and thus also to a hereditary Markov interaction Φ such that $\Pi = \Pi_\Phi$). We present a basic theorem which provides the key to their analysis.

The class C of cliques in the tree A simply consists of those subsets of A which contain either a single element or else a pair of neighbouring elements. We will denote by N the subset of C consisting of the (unordered) pairs of neighbouring elements, and by $N(V)$ the set $N \cap C(V)$ - those elements of N which are subsets of any given $V \in V$. We will write a Markov interaction on (S,A) as $\Phi = \{p_i , q_{jk} ; i \in A , \{j,k\} \in N\}$, where for each $i \in A$ we have $p_i : S \to R_+$ and for each $\{j,k\} \in N$ we have $q_{jk} : S \times S \to R_+$. We will also write either $q_{jk}(x_j , x_k)$ or $q_{kj}(x_k , x_j)$ for $q_{jk}(x_{\{j,k\}})$.

We define two interactions to be *equivalent* if they generate the same specification. Any Markov interaction on (S,A) is equivalent to another which has the property that the *single-site* interaction functions p_i , $i \in A$, are identically equal to one. Therefore we could, without loss of generality, restrict consideration to Markov interactions consisting entirely of *pairwise* interaction functions q_{jk} , $\{j,k\} \in N$. However the advantages of retaining both types of interaction function (for example in the statement and proof of Theorem 2.2.1) outweigh the slightly more cumbersome expressions involved.

We also introduce the following additional notation. Because A is a tree, there is a unique *path* in A connecting any two distinct elements k, ℓ, i.e. a unique sequence of elements $k = k_0$, k_1 , k_2 , ... , $k_n = \ell$ of A such that $k_r \sim k_{r-1}$, $1 \leqslant r \leqslant n$. Now given distinct elements, i,j of A, let $\partial_j(i)$ denote the unique element k of ∂i which belongs to the path connecting i to j - in particular if $j \in \partial i$ we have $\partial_j(i) = j$. Further let $A_{i,j} = \{k \in A : \partial_k(i) = \partial_j(i)\}$. Note that for each $i \in A$, the sets $A_{i,j}$, $j \in \partial i$, taken together with $\{i\}$ form a partition of A. Finally let V^* be the subset of V consisting of the finite *connected* subsets of A. In those arguments which depend on consideration of the marginal p.m.s P_V of a given p.m. P on (S^A, F), it will be sufficient - and much easier - to consider those V belonging to V^* rather than to V.

2.1 Markov random fields and Markov chains

We define a p.m. P on (S^A, F) to be a *Markov chain*, if for each $V \in V^*$ the marginal p.m. P_V is a M.r.f. on $(S^V, F(V))$ with respect to (the tree obtained by) the restriction of the neighbour relation ~ to V. Note that any Markov chain P is itself a M.r.f. on (S^A, F): it is sufficient to observe that if $\{V_n\}$ is an increasing sequence in V^* such that $V_n \nearrow A$, then for each $V \in V$, $F \in F(V)$, we have

$$P(F/F(V_n - V)) = P(F/F(\partial V)) \qquad \text{P-a.s.}$$

for all n such that $\xi V \subset V_n$. Letting $n \to \infty$ we have

$$P(F/F(A-V)) = P(F/F(\partial V)) \qquad \text{P-a.s.}$$

Let M denote the class of all Markov chains on (S^A, F). We require some properties of these and some alternative characterisations of M. We start with the following easy result for finite A.

Lemma 2.1.1. Suppose A is finite.

(i) If P is a M.r.f. on (S^A, F), then for each $V \in V^*$, $i \in V$ we have

$$(1) \quad P(X_V = x_V) = P(X_i = x_i) \prod_{\substack{j \in V \\ j \neq i}} P(X_j = x_j / X_{\partial_i(j)} = x_{\partial_i(j)}),$$

(ii) Conversely, if for any p.m. P on (S^A, F), (1) holds with V = A and for some $i \in A$, then P is a M.r.f. on (S^A, F).

Proof. We prove (i) by induction on $|V|$. Given $V \in \mathcal{V}^*$, $i \in V$, choose $j \in V$ such that $j \neq i$ and also $\partial j \cap V = \{k\}$ for some k. Then $V' = V - \{j\}$ belongs to $\mathcal{V}^*$ and contains i. Now $A_{k,j}$ is finite, and so by the Markov property, $P(F/\mathcal{F}(V')) = P(F/\mathcal{F}(k))$ P-a.s. for all $F \in \mathcal{F}(A_{k,j})$. In particular we may take $F = \{X_j = x_j\}$. Thus, since $k = \partial_i(j)$, we have

$$P(X_V = x_V) = P(X_{V'} = x_{V'})P(X_j = x_j/X_{\partial_i(j)} = x_{\partial_i(j)})$$

and so if (1) holds for V', it also holds for V. Now it obviously holds for V such that $|V| = 1$, and so the general result follows. The converse result (ii) is trivial.

When A is finite, and P is a M.r.f. on $(S^A, \mathcal{F})$, the representation (1) of $P(X_A = x_A)$ effectively supplies a Markov interaction to which P corresponds. Note that (in contrast to the more general Theorem 1.3.1.), neither strict positivity nor the hereditary condition is required to obtain this result, but that in the absence of such a condition we do in general require P to be Markov, rather than simply locally Markov. The following example, adapted from one given by *Dobrushin (1968)* illustrates this : let $S = \{0,1\}$, $A = \{1,2,3,4,5\}$ and define $\sim$ on A by $i \sim j$ if and only if $|j-i| = 1$; let $P(X_A = x_A)$ equal ½ if either $x_A = (0,0,0,0,0)$ or $x_A = (1,1,0,1,1)$, and equal 0 otherwise. It is easy to check that P is a locally Markov, though not a Markov, random field - the defining relation (2) of section 1.3 for a M.r.f. fails to hold for the subsets {1,2} and {4,5} of A. Thus by part (i) of Theorem 1.3.1, P cannot correspond to any Markov interaction on (S,A).

Lemma 2.1.1 has the following immediate corollary.

<u>Corollary.</u> If A is finite, the classes of M.r.f.s and Markov chains on (S^A, F) coincide.

(This follows because, if P is a M.r.f. on (S^A, F), then for every finite connected subset V of A, the marginal p.m. P_V has a representation of the form (1) and is thus also a M.r.f.)

We now return to the general case where A is countable. Then neither the result (i) of Lemma 2.1.1, nor the above corollary, is in general true - as will be seen. We do have the following lemma.

<u>Lemma 2.1.2.</u> Let P be a p.m. on (S^A, F) Then the representation (1) of P_V holds for each $V \in \mathcal{V}^*$, $i \in V$ if and only if $P \in M$.

<u>Proof.</u> If $P \in M$, then for each $V \in \mathcal{V}^*$ the marginal p.m. P_V is a M.r.f. on $(S^V, F(V))$, (with respect to the restriction of ~ to V). Hence by Lemma 2.1.1(i) we obtain (1) for any $i \in V$. Conversely if (1) holds for each $V \in \mathcal{V}^*$, $i \in V$, then by Lemma 2.1.1(ii) each marginal p.m. P_V, $V \in \mathcal{V}^*$, is a M.r.f. and so $P \in M$.

<u>Remark.</u> The result which we really need is that a p.m. P on (S^A, F) belongs to M if and only if, for each $V \in \mathcal{V}^*$, the marginal p.m. P_V corresponds to a Markov interaction on (S,V). When P is strictly positive then this result may be deduced directly from the general Theorem 1.3.1 (arguing as in the proof of Lemma 2.1.2) - for when A is finite we may replace Markov specifications by M.r.f.s in the statement of that theorem.

The following corollary to Lemma 2.1.2 is of some interest, though it is not required at any later stage.

Corollary. If $P \in M$, then there is a Markov interaction $\Phi = \{p_i , q_{jk} ; i \in A , \{j,k\} \in N\}$ on (S,A) (and hence a Markov specification) to which P corresponds.

Proof. Choose any $i \in A$. Given $P \in M$ define p_i by $p_i(x_i) = P(X_i = x_i)$, $x_i \in S$; for each $j \in A-\{i\}$, let $k = \partial_i(j)$, define q_{jk} by $q_{jk}(x_j , x_k) = P(X_j = x_j/X_k = x_k)$, and also define p_j to be identically equal to one. This gives a complete Markov interaction $\Phi = \{p_h , q_{jk} ; h \in A , \{j,k\} \in N\}$ on (S,A). To show that P corresponds to Φ is straightforward: given $W \in V$, choose $V \in V^*$ such that both $\xi W \subset V$ and also $i \in V$; then since $P \in M$ we have by Lemma 2.1.2

$$P(X_V = x_V) = \prod_{h \in V} p_h(x_h) \prod_{\{j,k\} \in N(V)} q_{jk}(x_j , x_k)$$

and so for all $x_W \in S^W$

$$P(X_W = x_W/F(\partial W)) = \tilde{\pi}_{\Phi,W}(X_{\partial W} , x_W) \qquad P\text{-a.s.}$$

(recalling that $\tilde{\pi}_{\Phi,W}$ is the conditional density function associated with W by Φ - see section 1.2). Hence P corresponds to Φ as required.

Prompted by this result we conjecture that when A is a tree, every Markov specification, whether hereditary or not, corresponds to a Markov interaction. (The example following Lemma 2.1.1 can be modified to show that this result is not in general true for non-hereditary specifications which are simply locally Markov.)

We now provide a more conventional characterisation of the Markov chains on (S^A,F).

Lemma 2.1.3. If P is a p.m. on $(S^A, \mathcal{F})$, then the following statements are equivalent.

(i) $P \in \mathcal{M}$.

(ii) For each $i \in A$, the σ-fields $\mathcal{F}(A_{i,j})$, $j \in \partial i$, are conditionally independent (under P) relative to $\mathcal{F}(i)$.

(iii) For each $j \in A$, $i \in \partial j$, $F \in \mathcal{F}(j)$,

$$P(F/\mathcal{F}(A_{j,i})) = P(F/\mathcal{F}(i)) \qquad P\text{-a.s.}$$

Proof. We show (i) ⇒ (ii) ⇒ (iii) ⇒ (i).
(i) ⇒ (ii) : Given $i \in A$, for each $j \in \partial i$ let V_j be any finite subset of $A_{i,j}$; let $V \in \mathcal{V}^*$ be such that $i \in V$ and $V_j \subset V$ for all $j \in \partial i$. Then if $P \in \mathcal{M}$ it is immediate from the representation (1) of $P(X_V = x_V)$ that the σ-fields $\mathcal{F}(V_j)$ are conditionally independent under P relative to $\mathcal{F}(i)$. Since each V_j is an arbitrary subset of the corresponding $A_{i,j}$, the result (ii) follows in the usual manner.
(ii) ⇒ (iii) : Given $j \in A$, $i \in \partial j$, if P is such that (ii) holds then it is immediate that $\mathcal{F}(j)$ and $\mathcal{F}(A_{j,i} - \{i\})$ are conditionally independent under P relative to $\mathcal{F}(i)$, and this is equivalent to the result (iii).
(iii) ⇒ (i) : Consider any $V \in \mathcal{V}^*$. Let $i, i_1, i_2, \ldots, i_n$ be any ordering of the elements of V such that for each r $(1 \leqslant r \leqslant n)$, the set $V_r = \{i, i_1, i_2, \ldots, i_r\}$ also belongs to $\mathcal{V}^*$. It is easy to see that we then have $\partial_i(i_r) \subset V_{r-1} \subset A_{i_r,i}$ for each r. Now if P is such that (iii) holds we have,

$$P(X_V = x_V) = P(X_i = x_i) \prod_{1 \leqslant r \leqslant n} P(X_{i_r} = x_{i_r} / X_{V_{r-1}} = x_{V_{r-1}})$$

$$= P(X_i = x_i) \prod_{j \in V-i} P(X_j = x_j / X_{\partial_i(j)} = x_{\partial_i(j)})$$

Thus by Lemma 2.1.2, $P \in M$.

When A is equal to the set of integers, and consecutive pairs of these are defined to be neighbours - the one-dimensional integer lattice, the characterisation (iii) in the above Lemma is just the defining relation for a Markov chain in the usual sense; (ii) is just the well-known 'time-symmetric' alternative characterisation. Thus the concept of a Markov chain introduced here is simply a natural extension of the familiar one to more general trees.

*Spitzer (1975b)** introduced Markov chains on a tree, with the binary state space $S = \{0,1\}$. His motivation also was the study of the more general class of M.r.f.s on the tree. In a sense it was the characterisation (iii) of a Markov chain which was central to Spitzer's approach - thus any Markov chain whose distribution was strictly positive, and homogeneous in the sense of being invariant under graph isomorphisms of the tree, could be represented by its transition matrix. (See also section 3.3 of this work). In contrast the approach here is mostly based on our initial definition of a Markov chain. This will be developed in the next section.

Now let E be the class of all M.r.f.s on (S^A, F) with respect to which the tail σ-field $\hat{F}$ is (almost) trivial. Then we have the following result.

* and *Preston (1974)*.

<u>Theorem 2.1.4.</u> $\mathcal{E} \subset \mathcal{M}$.

<u>Proof.</u> (The argument is broadly analogous to one given by *Spitzer (1975a, Theorem 6)* for the one-dimensional integer lattice.) Given $P \in \mathcal{E}$, we must show that for each $V \in \mathcal{V}^*$, $W \subset V$, $F \in \mathcal{F}(W)$,

$$(2) \qquad P(F/\mathcal{F}(V-W)) = P(F/\mathcal{F}(V \cap \partial W)) \qquad P\text{-a.s.}$$

Let $\{V_n\}$ be any increasing sequence in $\mathcal{V}$ such that $V \subset V_n$ for all n, and $V_n \nearrow A$. Because P is a M.r.f., for each V_n in the given sequence, there is a version of $P(\cdot/\mathcal{F}(A-V_n))$ which at each point $x_A \in S^A$ is a p.m. whose marginal distribution on $(S^{V_n}, \mathcal{F}(V_n))$ is a M.r.f. with respect to the restriction of $\sim$ to V_n. Now V is a connected subset of V_n and thus, by the corollary to Lemma 2.1.1, the marginal distribution on $(S^V, \mathcal{F}(V))$ of each of the above p.m.s is also a M.r.f. (with respect to the corresponding restriction of $\sim$ to V). Since $F \in \mathcal{F}(W)$ we obtain

$$(3) \quad P(F/\mathcal{F}(V-W) \vee \mathcal{F}(A-V_n)) = P(F/\mathcal{F}(V \cap \partial W) \vee \mathcal{F}(A-V_n)) \qquad P\text{-a.s.}$$

As $n \to \infty$, we have $\mathcal{F}(V-W) \vee \mathcal{F}(A-V_n) \searrow \mathcal{F}(V-W) \vee \hat{\mathcal{F}}$ and $\mathcal{F}(V \cap \partial W) \vee \mathcal{F}(A-V_n) \searrow \mathcal{F}(V \cap \partial W) \vee \hat{\mathcal{F}}$. (These results follow easily since $\mathcal{F}(V-W)$ and $\mathcal{F}(V \cap \partial W)$ are countable.) Thus by the (reversed) martingale convergence theorem we obtain (3) with $\mathcal{F}(A-V_n)$ replaced by $\hat{\mathcal{F}}$. Since $P \in \mathcal{E}$, (2) follows immediately.

Now if Π is a Markov specification on $(S^A, \mathcal{F})$, we denote by $\mathcal{M}(\Pi)$ the class of corresponding Markov chains, i.e. $\mathcal{M}(\Pi) = \mathcal{M} \cap \mathcal{G}(\Pi)$. We have already defined $\mathcal{E}(\Pi)$ to be the set of extreme elements of $\mathcal{G}(\Pi)$, and noted that it is equivalent to the set of M.r.f.s in $\mathcal{G}(\Pi)$ with respect to which the tail σ-field $\hat{\mathcal{F}}$ is trivial, i.e. that $\mathcal{E}(\Pi) = \mathcal{E} \cap \mathcal{G}(\Pi)$.

By the above theorem $E(\Pi) \subset M(\Pi)$. Thus $G(\Pi)$, which we noted in section 1.1 to be (roughly speaking) the set of convex combinations of elements of $E(\Pi)$, is also the set of convex combinations of elements of $M(\Pi)$, so that its study reduces, in a sense, to that of $M(\Pi)$ - at least when the latter set turns out to be countable. The next section is devoted to the study of $M(\Pi)$ for hereditary Markov specifications Π.

We remark that in general the converse to Theorem 2.1.4 is false. *Spitzer (1975a)* gave a counter-example with S countably-infinite and A the one-dimensional integer lattice. Essentially this consists of two distinct (non-stationary) Markov chains P_1, P_2, say, which have the same stationary transition matrix, and so correspond to the same Markov specification Π. Then, for example, $\frac{1}{2}(P_1 + P_2)$ belongs to $M(\Pi)$ but not to $E(\Pi)$, and hence not to E.

2.2 Interactions for Markov chains

In this section we present a basic theorem concerning the Markov chains which correspond to a given Markov specification Π, hereditary with respect to some basepoint $s_A \in S^A$. By definition $\Pi = \Pi_\Phi$ for some hereditary Markov interaction Φ, i.e. for a Markov interaction $\Phi = \{p_i , q_{jk} ; i \in A , \{j,k\} \in N\}$ such that $p_i(s_i) > 0$ for each $i \in A$, and $q_{jk}(x_j , s_k) > 0$, $q_{jk}(s_j , x_k) > 0$ for each $\{j,k\} \in N$. (Recall that Theorem 1.3.1 ensures that every strictly positive (locally) Markov specification is hereditary with respect to any basepoint.) Without loss of generality we may take $p_i(s_i) = 1$ for all i. It is easiest to state the theorem in terms of a given interaction, rather than a specification.

<u>Theorem 2.2.1.</u> Let $\Phi = \{p_i , q_{jk} ; i \in A , \{j,k\} \in N\}$ be a Markov interaction on (S,A), hereditary with respect to the basepoint $s_A \in S^A$ and (for definiteness) satisfying $p_i(s_i) = 1$ for all $i \in A$.

(a) Let $P \in M(\Pi_\Phi)$. Then for each $V \in V^*$, we have the following representation of the marginal p.m. P_V on $(S^V, F(V))$:

$$(1) \qquad P(X_V = x_V) = a_V \prod_{i \in V} p_i^V(x_i) \prod_{\{j,k\} \in N(V)} q_{jk}(x_j , x_k)$$

where

$$(2) \qquad p_i^V(s_i) = 1 \qquad \text{for all } i \in V.$$

The functions p_i^V , $i \in V$ are uniquely determined by (1) and (2), and the normalising constant a_V is then determined by the requirement:

$$(3) \qquad \sum_{x_V \in S^V} P(X_V = x_V) = 1$$

For simplicity we will write p_i^j for $p_i^{\{i,j\}}$ for each $i \in V$, $j \in \partial i$. The collection of all the functions $\{p_i^V, V \in \mathcal{V}^*, i \in V\}$ then satisfy

(4) $$p_i^V(x) = p_i^{V \cap \xi i}(x) = k_i^V\, p_i(x) \prod_{j \in \partial i - V} (Q_{ij}p_j^i)(x), \quad x \in S,$$

(recalling that $\xi i = \{i\} \cup \partial i$), where the function $(Q_{ij}p_j^i) : S \to R_+$ is defined by

(5) $$(Q_{ij}p_j^i)(x) = \sum_{y \in S} q_{ij}(x,y)\, p_j^i(y), \quad x \in S,$$

and where k_i^V is a (strictly positive) constant determined by the requirement that (2) be satisfied. As special cases of (4) we have

(6) $$p_i^V = p_i \qquad \text{if } \xi i \subset V, \ V \in \mathcal{V}^*$$

and (writing k_i^j for $k_i^{\{i,j\}}$)

(7) $$p_i^j(x) = k_i^j p_i(x) \prod_{k \in \partial i - j} (Q_{ik}p_k^i)(x), \quad i \in A,\ j \in \partial i,\ x \in S.$$

Note also that for each $i \in A$, the following condition is satisfied:

(8) $$\sum_{x \in S} p_i(x) \prod_{j \in \partial i} (Q_{ij}p_j^i)(x) < \infty .$$

(b) Conversely, given a collection of functions $\{p_i^j : S \to R_+, i \in A, j \in \partial i\}$ such that

$$(9)\begin{cases} \text{(i) for each } i \in A,\ j \in \partial i,\ p_i^j(s_i) = 1, \\ \text{(ii) for each } i \in A,\ j \in \partial i, \text{ the relation (7) is satisfied for some (necessarily strictly positive) constant } k_i^j, \\ \text{(iii) for each } i \in A, \text{ the condition (8) is satisfied,} \end{cases}$$

then we may define a p.m. $P \in M(\Pi_\Phi)$, to which the functions p_i^j correspond as in part (a) of the theorem.

The theorem thus establishes a one-to-one correspondence between $M(\Pi_\Phi)$ *and the set of those collections of functions* $\{p_j^i\ ,\ i \in A,\ j \in \partial i\}$ *which satisfy* (9).

Remark In part (a) of the theorem it is important to note that when the parameter space A is finite, then the functions p_i^V are (under (2)) completely determined by the relations (4) and (5): for each $i \in A$ such that $\partial i = \{j\}$ for some $j \in A$ (i.e. for each *end-vertex* of the tree A), we have by (6) that $p_i^j = p_i$. We may therefore use (7) recursively to determine all the functions p_ℓ^k , $\{k,\ell\} \in N$, in each case simply by deleting end-vertices from the tree A until the set $\{k,\ell\}$ is obtained. We may then use (4) to determine the more general functions p_i^V. (The real application of this remark is to the consideration of Markov chains on $(S^V, F(V))$, where $V \in V^*$ for some countably-infinite A - an example occurs in the proof of part (b) of the theorem.)

Proof of Theorem 2.2.1. (a) We first establish, for each $V \in V^*$, the existence and uniqueness (under (2)) of the representation (1) of the marginal p.m. P_V. Since $V \in V^*$, no two elements of ∂V can be neighbours of each other and we also have $\tilde{\partial}(\xi V) = \partial V$. Thus from Lemma 1.4.1,

(10) $$P(X_{\xi V}=x_{\xi V}) = \phi_{\partial V}(x_{\partial V}) \prod_{i \in V} p_i(x_i) \prod_{\{j,k\}\in N(\xi V)} q_{jk}(x_j , x_k)$$

for some function $\phi_{\partial V} : S^{\partial V} \to R_+$.

Now $P(X_V = s_V) > 0$, and we have

(11) $$P(X_{\partial V} = x_{\partial V}/X_V = s_V) = c\phi_{\partial V}(x_{\partial V}) \prod_{i \in \partial V} q_{ij}(x_i , s_j)$$

where for each $i \in \partial V$, j is the unique element of $\partial i \cap V$, and where c is some strictly positive constant. Since P is a Markov chain and $\xi V \in V^*$, $P_{\xi V}$ is a M.r.f. on $(S^{\xi V}, F(\xi V))$, and so the random variables X_i , $i \in \partial V$, are conditionally independent (under P) relative to $F(V)$. Thus

(12) $$P(X_{\partial V} = x_{\partial V}/X_V = s_V) = \prod_{i \in \partial V} P(X_i = x_i/X_V = s_V)$$

It follows from (11) and (12) that we may write $\phi_{\partial V}(x_{\partial V}) = a_{\xi V} \prod_{i \in \partial V} p_i^{\xi V}(x_i)$, $x_{\partial V} \in S^{\partial V}$. Substituting this into (10) we have the representation (1) for $P_{\xi V}$ rather than P_V. By summing over $x_{\partial V} \in S^{\partial V}$, we obtain (1) for P_V itself. Note that the functions p_i^V , $i \in V$, may be chosen so as to satisfy (2), and that the special case of (4) given by the condition (6) is satisfied.

To see that the functions p_i^V are uniquely determined by (1) and (2), we observe that for each $i \in V$, (1) and (2) together give

(13) $$\frac{P(X_i=x_i/X_{V-i}=s_{V-i})}{P(X_i=s_i/X_{V-i}=s_{V-i})} = p_i^V(x_i) \prod_{j\in\partial i\cap V} \frac{q_{ij}(x_i , s_j)}{q_{ij}(s_i , s_j)}$$

for each $x_i \in S$.

We now establish the relation (4) (already established as (6) for V, i such that $\xi i \subset V$) in the general case. Given $V \in \mathcal{V}^*$, the representation (1) implies that P_V corresponds to the Markov interaction $\Phi^V = \{p_i^V, q_{jk} ; i \in V, \{j,k\} \in N(V)\}$. For each $i \in V$, we may therefore obtain the representation (1) for $P_{V \cap \xi i}$ by regarding the latter as the marginal p.m. induced by P_V on $(S^{V \cap \xi i}, F(V \cap \xi i))$: using the results obtained so far with V instead of A, $V \cap \xi i$ instead of V, we obtain immediately (from (6)) the first equation in (4). Now consider the representation (1) for $P_{\xi i}$. By (6) and the first part of (4) we may write this as

$$(14) \qquad P(X_{\xi i} = x_{\xi i}) = a_{\xi i} p_i(x_i) \prod_{j \in \partial i} q_{ij}(x_i, x_j) p_j^i(x_j)$$

Summing over all $x_{\partial i - V} \in S^{\partial i - V}$, we obtain an explicit expression of the form (1) for $P_{V \cap \xi i}$, from which it follows immediately that $p_i^{V \cap \xi i}$ is as given by the second equation in (4).

Lastly the relation (8) follows from the results so far obtained by noting that for each $i \in A$,

$$P(X_i = x) = a_i k_i^{\{i\}} p_i(x) \prod_{j \in \partial i} (Q_{ij} p_j^i)(x) \quad , \quad x \in S$$

(b) Now suppose we are given a collection of functions $\{p_i^j : S \to R_+ , i \in A, j \in \partial i\}$ satisfying the three conditions contained in (9). We seek to define a p.m. $P \in M(\Pi_\Phi)$ - to which the above collection of functions corresponds - via its marginal p.m.s P_V. We consider sets $V \in \mathcal{V}^*$ which satisfy the condition that, if $i \in \tilde{\partial} V$, then $|\partial i \cap V| = 1$. Let $\tilde{\mathcal{V}}$ be the set of all such (connected) sets V. (Note that it contains all sets of the form ξV, $V \in \mathcal{V}^*$.)

For each $V \in \tilde{\mathcal{V}}$ define the measure μ_V on $(S^V, \mathcal{F}(V))$ by

$$(15) \quad \mu_V(X_V = x_V) = \prod_{i \in \tilde{\partial} V} p_i^h(x_i) \prod_{i \in \eta V} p_i(x_i) \prod_{\{j,k\} \in N(V)} q_{jk}(x_j, x_k)$$

where for each $i \in \tilde{\partial}V$ in the first product above, h is the unique neighbour of i in V. We have to check (i) that each measure μ_V, $V \in \tilde{\mathcal{V}}$, may be normalised to a p.m. P_V on $(S^V, \mathcal{F}(V))$, (ii) that the family $(P_V, V \in \tilde{\mathcal{V}})$ of such p.m.s is consistent, so that they may be regarded as the marginals of a p.m. P on $(S^A, \mathcal{F})$, and (iii) that P thus defined belongs to $\mathcal{M}(\Pi_\Phi)$.

Given any $V \in \tilde{\mathcal{V}}$, then for any connected subset W of V, the marginal measure $\mu_{V,W}$ induced by μ_V on $(S^W, \mathcal{F}(W))$ may be calculated using the first part of the theorem - for this, with V instead of A, is obviously applicable to any positive measure on $(S^V, \mathcal{F}(V))$ whose density is given by a product of hereditary functions on the vertices and edges of V. (It is only necessary to allow that the functions p_i^W, $W \in \mathcal{V}^*$, $W \subset V$, may turn out to take infinite values.) Arguing exactly as in the remark which follows the statement of the theorem we have,

$$(16) \quad \mu_V(X_W = x_W) = c_{V,W} \prod_{i \in W} p_i^W(x_i) \prod_{\{j,k\} \in N(W)} q_{jk}(x_j, x_k),$$

where each function p_i^W (taking perhaps infinite values) is given by (4) with W instead of V and omitting the constant k_i^W, and where $c_{V,W}$ is a strictly positive (finite) constant, which may be computed from the constants k_i^j of (7).

In particular for any $i \in V$ we have,

$$\mu_V(X_i = x) = c_{V,i}\, p_i(x) \prod_{j \in \partial i} (Q_{i\,j} p_j^i)(x), \qquad x \in S$$

Thus by condition (iii) of (9), $\sum_{x_V \in S^V} \mu_V(X_V = x_V) < \infty$,

and so μ_V *may* be normalised to a p.m. $P_V = a_V \mu_V$. (It now follows that the functions p_i^W in (16) are finite.)

Now suppose that $V, W \in \tilde{V}$ and that $W \subset V$. Then from (16) we have that

$$P_V(X_W = x_W) = a'_W \prod_{i \in \tilde{\partial} W} p_i^h(x_i) \prod_{i \in \eta W} p_i(x_i) \prod_{\{j,k\} \in N(W)} q_{jk}(x_j, x_k)$$

for some constant a'_W, where for each $i \in \tilde{\partial}W$, h is the unique element of $\partial i \cap W$. Comparing this with (15) (with W instead of V) we see that $P_V(X_W = x_W) = P_W(X_W = x_W)$. Thus the family of p.m.s $(P_V, V \in \tilde{V})$ is consistent and so defines a p.m. P on (S^A, F).

To show $P \in M(\Pi_\Phi)$, we first note that, by construction, P_V is a M.r.f. for each $V \in \tilde{V}$ (and hence for each $V \in V^*$); this implies that P is a Markov chain. Now if $W \in V$, let $V \in \tilde{V}$ be such that $\xi W \subset V$. We then have, using (15)

$$P(X_W = x_W / X_{\partial W} = x_{\partial W}) = P_V(X_W = x_W / X_{\partial W} = x_{\partial W})$$

$$= k_W(x_{\partial W}) \prod_{i \in W} p_i(x_i) \prod_{\substack{\{j,k\} \in N \\ \{j,k\} \cap W \neq \emptyset}} q_{jk}(x_j, x_k)$$

$$= \tilde{\pi}_{\Phi, W}(x_{\partial W}, x_W)$$

(where $k_W : S^{\partial W} \to R_+$ is the appropriate normalising function). Hence $P \in M(\Pi_\Phi)$.

<u>Further remarks on Theorem 2.2.1</u>

1. In establishing the second part of the theorem it is sufficient that the condition (9)(iii) should hold for *some* $i \in A$. For then if $V \in \tilde{V}$ is such that $i \in V$, the measure μ_V of the proof is normalisable to a p.m., and so the above condition holds for all other $j \in V$; obviously every $j \in A$ belongs to some such set V.

2. When S is finite the condition 9(iii) (i.e. the normalisability condition (8)) is obviously unnecessary. When S is infinite we have the question of whether it might still be unnecessary, in the sense of being implied by 9(i) and 9(ii). An example will suffice to show that this is not the case. Let the graph A be the one-dimensional integer lattice (again), and consider the translation-invariant Markov interaction given by letting the single-site interaction functions p_i , $i \in A$, be identically equal to one, and the pairwise interaction functions $q_{i,i+1}$, $i \in A$, be equal to a hereditary function $q : S \times S \to R_+$ satisfying

$$\sum_{y \in S} q(x,y) = \sum_{y \in S} q(y,x) = 1$$

for all $x \in S$. It is obvious that this *does* define an interaction on (S,A), i.e. that the condition (1) of section 1.2 is satisfied. For each $i \in A$, $j \in \partial i$, define p_i^j to be identically equal to one. Then the collection of all such functions p_i^j satisfies the first two conditions of (9), but not the third for infinite S.

3. We postpone until section 3.4 a demonstration of the (perhaps surprising) necessity of the hereditary condition in the theorem. There we show by an example that in the absence of some such condition the conclusions of the theorem are false.

4. *Spitzer(1975a)* has proved a result similar to part (a) of the theorem for the special case where the graph A is the one-dimensional integer lattice, and the interaction Φ is strictly positive and translation invariant.

In the next chapter we show how the results of this one - in particular the above theorem - may be used to study in detail the classes of Markov chains corresponding to homogeneous hereditary Markov interactions on regular infinite trees.

3 HOMOGENEOUS MARKOV SPECIFICATIONS ON REGULAR TREES

3.0 Introduction

Throughout this and the subsequent chapter the graph A is the regular infinite tree with d + 1 edges meeting at each vertex. In particular when d = 1 we have the one-dimensional integer lattice.

A Markov specification, Markov interaction, or p.m. will be said to be *homogeneous* if it is invariant under graph isomorphisms of the tree onto itself. By the term *hereditary* we will mean hereditary with respect to a fixed basepoint $s_A \in S^A$, where for all $i \in A$, $s_i = s$ for some fixed $s \in S$. Our aim is to study various classes of M.r.f.s, in particular Markov chains, corresponding to homogeneous hereditary Markov (h.h.M.) specifications.

Any h.h.M.interaction on (S,A) may be represented by a pair of functions - $p : S \to R_+$ associated with the vertices of the tree A, and $q : S \times S \to R_+$ associated with its edges, i.e. with the elements of the set N introduced in the previous chapter. The function q is symmetric in its arguments, and the hereditary condition gives us that $p(s) > 0$, $q(s,x) > 0$ for all $x \in S$. We will additionally assume, without loss of generality, that p is strictly positive. (If this condition is not satisfied then the state space S may be reduced until it is.) We will write $\Phi(p,q)$ for such an interaction.

If $\Phi(p,q)$ is a h.h.M. interaction the corresponding specification $\Pi_{\Phi(p,q)}$ - we will write this simply as $\Pi(p,q)$ - is also h.h.M.. For any $V \in V$ the conditional density function associated with the kernel π_V is given by

$$\tilde{\pi}_V(x_{\partial V}, x_V) = k_V(x_{\partial V}) \prod_{i \in V} p(x_i) \prod_{\substack{\{j,k\} \in N \\ \{j,k\} \cap V \neq \emptyset}} q(x_j, x_k)$$

where as usual $k_V : S^{\partial V} \to R_+$ is the appropriate normalising function. Conversely, for any h.h.M. specification Π we can find a h.h.M. interaction $\Phi(p,q)$ such that $\Pi = \Pi(p,q)$. To see this, note that any hereditary Markov specification corresponds to a Markov interaction which is adapted (see section 1.3) with respect to the basepoint s_A : let $\Phi = \{p_i \, , \, q_{jk} \, ; \, i \in A \, , \, \{j,k\} \in N\}$ be *any* (necessarily hereditary) Markov interaction to which the specification corresponds; replace each function q_{jk} by q'_{jk} where

$$q'_{jk}(x,y) = \frac{q_{jk}(s,s)\,q_{jk}(x,y)}{q_{jk}(x,s)\,q_{jk}(s,y)}$$

and make appropriate adjustments to the functions p_i to obtain an equivalent adapted interaction. Now it is easy to see that the adapted (Markov) interaction is unique, and thus if the specification is additionally homogeneous, so too is this interaction. Our additional constraint on an h.h.M. interaction $\Phi(p,q)$, that the function p be strictly positive, is ensured by the corresponding additional constraint on $\Pi = \Pi(p,q)$, that the p.m. $\pi_i(s_{A-i}\, ,\cdot)$ be strictly positive for each $i \in A$.

We note also that it may be convenient for some applications to represent an h.h.M. specification by a hereditary Markov interaction which is *not* in general homogeneous. In section 3.3 we follow Spitzer and consider the representation of such a specification by an interaction $\Phi(1,q)$ where 1 is the function $S \to R_+$ which is identically one, and $q : S \times S \to R_+$ is a 'reversible', though not in general symmetric, stochastic matrix associated with a given 'direction' in the tree.

Given an h.h.M. specification Π, $G(\Pi)$, $E(\Pi)$ and $M(\Pi)$ will denote as usual the corresponding classes of, respectively, all M.r.f.s, extreme M.r.f.s (those with respect to which the tail σ-field $\hat{F}$ is (almost) trivial), and Markov chains.

We have $\mathcal{E}(\Pi) \subset \mathcal{M}(\Pi) \subset \mathcal{G}(\Pi)$. We will denote by $\mathcal{G}_0(\Pi)$, $\mathcal{E}_0(\Pi)$ and $\mathcal{M}_0(\Pi)$ the respective subclasses of $\mathcal{G}(\Pi)$, $\mathcal{E}(\Pi)$ and $\mathcal{M}(\Pi)$ whose elements have the additional property of being themselves homogeneous. In section 3.1 we will define a futher class $\mathcal{M}_1(\Pi)$ lying between $\mathcal{M}_0(\Pi)$ and $\mathcal{M}(\Pi)$. Note that $\mathcal{E}_0(\Pi)$ is a (perhaps proper) subset of the set of extreme elements of $\mathcal{G}_0(\Pi)$ - see the corollary to the corollary to Theorem 3.1.3.

Recall from Chapter 2 that $\mathcal{V}^*$ is the set of finite *connected* subsets of A, and that $\tilde{\mathcal{V}}$ is defined by

$$\tilde{\mathcal{V}} = \{V \in \mathcal{V}^* : \text{for all } i \in \tilde{\partial}V \text{ , } |\partial i \cap V| = 1\} \text{ .}$$

Most of this chapter is concerned with various applications of Theorem 2.2.1. The functions p_i^V , $V \in \mathcal{V}^*$, $i \in V$, introduced in that theorem were forced to be unique by the condition that they took the value one on the corresponding component of the basepoint. If we drop this condition - which it will now be convenient to do - then these functions remain uniquely determined up to an arbitrary, strictly positive, multiplicative constant. We therefore introduce the space Ψ of equivalence classes of functions $S \to R_+$ - omitting the function which is identically zero - where two such functions p_1 , p_2 are said to be equivalent if $p_1(x) = \lambda p_2(x)$ for all $x \in S$ and some strictly positive constant λ. We will denote a given element of Ψ by any function in the corresponding equivalence class, and where in consequence an equation involves an indeterminacy we will use the proportionality symbol $\propto$ (rather than =) to denote equality up to a strictly positive multiplicative constant. The reason for this change is that Ψ is the natural space in which the functions p_i^V are defined, and on occasions it will be convenient to represent a member p, say, of Ψ other than by the convention $p(s) = 1$. Indeed we now define for each $x \in S$, the particular element δ_x of Ψ given by: $\delta_x(x) > 0$, $\delta_x(y) = 0$ for all other $y \in S$. Where these functions are used, our former convention could only be accommodated by an otherwise unnecessary modification of the basepoint.

In section 3.1 we consider the class $M_0(\Pi)$ of homogeneous Markov chains corresponding to a given h.h.M. specification Π, as well as the (sometimes strictly) wider class $M_1(\Pi) \subset M(\Pi)$ to be introduced there. In section 3.2 we look (briefly) at the characterisation of any given Markov chain $\underset{\sim}{P}$ in $M_0(\Pi)$, as the limit of a sequence of p.m.s associated with a sequence $\{V_n\}$ in V^* such that $V_n \uparrow A$. In section 3.3 we consider the representation of h.h.M. specifications by 'reversible' stochastic matrices in the manner mentioned briefly above, and prove some analogues to the results of the first two sections. In particular this enables us to relate, in section 3.4, our results to those derived by *Spitzer (1975b)* for the binary state space $S = \{0,1\}$. Finally we consider in section 3.5, as a further example, the state space $S = \{0,1,2\}$ and those h.h.M. specifications which are invariant under interchange of any two of the states in S. Even under such restricted circumstances an interesting variety of behaviour may be observed.

3.1 The classes $M_0(\Pi)$ and $M_1(\Pi)$

Throughout this section Π is a given h.h.M. specification.

We consider first the problem of identifying the elements of $M_0(\Pi)$. For d = 1, so that our regular tree becomes the one-dimensional integer lattice, and under the weaker (and in this case more natural) condition that Π is simply translation invariant, the problem has been well-studied; some of the known results are summarised in the introductory chapter, but see also *Preston (1976, Chapter 5)*. In particular when d = 1 and S is finite, $M_0(\Pi)$ always containsprecisely one element, which is also the sole element of $G(\Pi)$. For d > 1, only the case where $|S| = 2$ has been studied in any detail - see section 3.4 for a discussion.

For general $d \geqslant 1$ and finite state space S, we have the result, essentially due to *Dobrushin (1968)*, that $M_0(\Pi)$ contains *at least* one element. Dobrushin actually considered the more important case of the d-dimensional integer lattice: he first used a simple compactness argument to construct a random field corresponding to any given specification satisfying certain conditions - automatically satisfied by any Markov specification; he then showed how, if the specification were homogeneous, the corresponding random field could be used to construct another which had the property of being itself homogeneous. All this is fairly readily adapted to specifications on trees, but we may also use the theory developed so far (explicitly for trees) to give a quick, essentially topological, proof of the result. This we do in the corollary to Theorem 3.1.1 below.

Now let $\Phi(p,q)$ be a fixed h.h.M. interaction to which Π corresponds. We wish to consider various applications of Theorem 2.2.1, and in order to simplify the expression of the relations involved in that theorem, we make the following definitions:

for any function $p^* : S \to R_+$ define $Qp^* : S \to R_+$ (where it exists) by

(1) $$(Qp^*)(x) = \sum_{y \in S} q(x,y)p^*(y) \quad , \qquad x \in S \quad ;$$

for any integer $t = 0, 1, \ldots, d+1$, and for any $p_1, \ldots, p_t \in \Psi$, define $T_t(p_1, \ldots, p_t) \in \Psi$ (where it exists) by

(2) $$[T_t(p_1, \ldots, p_t)](x) \propto p(x) \prod_{1 \leq r \leq t} (Qp_r)(x) \quad ;$$

(with $T_0 = p$)

and for any integer t as above, and any $p^* \in \Psi$, define $T'_t(p^*) \in \Psi$ (where it exists) by

(3) $$T'_t(p^*) = T_t(p^*, \ldots, p^*) \quad .$$

Note that when S is finite there are no problems of existence in these definitions. It also turns out that for countably-infinite S we have existence wherever it might reasonably be hoped for. Note also that whenever $T_t(p_1, \ldots, p_r)$ and $T'_t(p^*)$ exist, they are strictly positive at $s \in S$.

Theorem 2.2.1 establishes a one-to-one correspondence between $M(\Pi)$ and the set of those collections of functions $\{p_i^j, i \in A, j \in \partial i\}$ in Ψ which satisfy the conditions (9) of section 2.2. It is immediate from the theorem that a Markov chain belongs to $M_0(\Pi)$ if and only if the corresponding functions p_i^j are the same for all $i \in A$ and $j \in \partial i$. We thus have the following result.

Theorem 3.1.1. There is a one-to-one correspondence between $\mathcal{M}_0(\Pi)$ and the set of functions $\tilde{p} \in \Psi$ which satisfy

(i) $\tilde{p} = T'_d(\tilde{p})$

(ii) $$\sum_{x \in S} \sum_{y \in S} \tilde{p}(x)q(x,y)\tilde{p}(y) < \infty .$$

Given $\tilde{p} \in \Psi$ satisfying (i) and (ii), the corresponding homogeneous Markov chain P is given through its marginal densities associated with the sets $V \in \mathcal{V}^*$ by

$$P(X_V = x_V) \propto \prod_{i \in V} p_i^V(x) \prod_{\{j,k\} \in N(V)} q(x_j , x_k)$$

where $p_i^V = T'_{|\partial i - V|}(\tilde{p})$. In particular if $V \in \tilde{\mathcal{V}}$

$$p_i^V = \begin{cases} \tilde{p} , & i \in \tilde{\partial} V \\ p , & i \in \eta V . \end{cases}$$

(The marginal densities associated with the sets $V \in \tilde{\mathcal{V}}$ are sufficient of course to determine P.) Note that given the condition (i) above, the condition (ii) corresponds to (8) of section 2.2. It is redundant when S is finite. Note also that functions $\tilde{p} \in \Psi$ satisfying (i) are necessarily strictly positive.

A particular consequence of the theorem is that it enables us in principle, and in practice when $|S|$ is small, to determine $|\mathcal{M}_0(\Pi)|$. (This is as yet far from completely determining $|\mathcal{M}(\Pi)|$ and hence $|\mathcal{G}(\Pi)|$.) For some examples of the application of the theorem see sections 3.4 and 3.5.

<u>Corollary.</u> When S is finite, $|M_0(\Pi)| \geq 1$.

<u>Proof.</u> 'Normalise' Ψ by the requirement that for all $p^* \in \Psi$, $\sum_{x \in S} p^*(x) = 1$, and then regard it as a subset of the Banach space of all real-valued functions on S (effectively $R^{|S|}$) with, say, the norm $\| f \| = \sup_{x \in S} |f(x)|$ $(f : S \to R)$. The subset Ψ is then convex and compact. The mapping $T'_d : \Psi \to \Psi$ is given by

$$(4) \qquad [T'_d(p^*)](x) = \frac{p(x)[(Qp^*)(x)]^d}{\sum_{y \in S} p(y)[(Qp^*)(x)]^d} , \quad p^* \in \Psi , \ x \in S ;$$

for all $p^* \in \Psi$, the denominator of the right-hand side of (4) is strictly positive (consider y = s), and so the mapping T'_d is seen to be continuous. Thus by, for example, the Leray-Schauder-Tychonoff theorem (*Reed and Simon (1972), p 151*) T'_d has a fixed point.

We have already established that $E_0(\Pi) \subset M_0(\Pi)$. It would be of interest to know whether these two sets in fact coincide, i.e. whether the tail σ-field $\hat{F}$ is trivial with respect to every homogeneous Markov chain corresponding to Π. We conjecture, though cannot prove, that this is in general the case. For d = 1, the conjecture is of course true: for then any $P \in M_0(\Pi)$ is a stationary (and reversible) Markov chain in the usual one-dimensional sense, and since all states of S 'intercommunicate' with the state s, P is additionally irreducible, so that the result follows. For d > 1, *Spitzer (1975b)* claims to prove the result when $|S| = 2$. His proof, however, seems unclear - if correct it ought to be equally applicable to any countable state space S. What we *can* say is that the elements of $M_0(\Pi)$ are mutually singular (so that none is a mixture of any of the others): for let B be any subgraph of A which is isomorphic (as a graph) to the one-dimensional integer lattice; the marginal p.m.s on $(S^B, F(B))$ of the Markov chains comprising $M_0(\Pi)$ form a set of stationary ergodic Markov chains in the one-dimensional sense, so that these and hence the original Markov chains of $M_0(\Pi)$ are mutually singular. It follows from this result that if $|M_0(\Pi)| > 1$ (as is sometimes the case - see sections 3.4 and 3.5 for examples), then $G_0(\Pi)$ contains M.r.f.s which are not Markov chains, e.g. all those strictly convex combinations of any two distinct elements of $M_0(\Pi)$. The corollary to Theorem 3.1.3, taken with the examples of section 3.4 , shows that this can also happen when $|M_0(\Pi)| \leqslant 1$.

We now turn our attention to a (sometimes) wider class of Markov chains corresponding to the specification Π. Label the vertices of our regular tree A alternately 'even' and 'odd', denoting by E and O the respective sets of even and odd vertices - so that every pair of neighbours contains one member of each set. Define $M_1(\Pi)$ to be the set of those Markov chains in $M(\Pi)$ which are invariant under those graph isomorphisms of the tree mapping E onto E (and hence O onto O). We have $M_0(\Pi) \subset M_1(\Pi) \subset M(\Pi)$. Markov chains belonging to $M_1(\Pi)$ but not $M_0(\Pi)$ arise naturally when we consider the *repulsive* specifications (*Spitzer (1975b)*) discussed in section 3.4, and defined more generally in Chapter 4. If $P \in M_1(\Pi)$, define the *complement* of P to be its image under those graph isomorphisms of the tree mapping E onto O. Oviously this complement also belongs to $M_1(\Pi)$, so that we may think of the Markov chains in $M_1(\Pi)$ as coming in symmetrically related pairs; the two chains of such a pair are identical if and only if either belongs to $M_0(\Pi)$. Now recalling our h.h.M. interaction $\Phi(p,q)$ such that $\Pi = \Pi(p,q)$, we have the following obvious analogue to Theorem 3.1.1.

<u>Theorem 3.1.2.</u> There is a one-to-one correspondence between $M_1(\Pi)$ and the set of ordered pairs of functions (p^e, p^o) in Ψ which satisfy

$$\text{(i)}\quad p^o = T_d'(p^e) \quad ; \quad p^e = T_d'(p^o)$$

$$\text{(ii)}\quad \sum_{x \in S} \sum_{y \in S} p^e(x)q(x,y)p^o(y) < \infty$$

Given (p^e, p^o) satisfying (i) and (ii), the corresponding Markov chain P in $M_1(\Pi)$ is given through its marginal densities associated with the sets $V \in V^*$ by

$$P(X_V = x_V) \propto \prod_{i \in V} p_i^V(x_i) \prod_{\{j,k\} \in N(V)} q(x_j , x_k)$$

where

$$p_i^V = \begin{cases} T'_{|\partial i - V|}(p^o) , & i \in E \\ T'_{|\partial i - V|}(p^e) , & i \in O \end{cases}$$

In particular, if $V \in \tilde{V}$ we have

$$p_i^V = \begin{cases} p^e & \text{if} \quad i \in \tilde{\partial} V \cap E \\ p^o & \text{if} \quad i \in \tilde{\partial} V \cap O \\ p & \text{if} \quad i \in \eta V \end{cases}$$

<u>Notes.</u> (i) The complement of P above is given by interchanging p^e and p^o, and these two chains coincide if and only if $p^e = p^o$.

(ii) When S is finite, condition (ii) in the above theorem is redundant.

(iii) The functions p^e, p^o above are necessarily strictly positive.

(iv) We may of course re-express the above theorem by saying simply that there is a one-to-one correspondence between $M_1(\Pi)$ and functions $p^e \in \Psi$ satisfying $p^e = T'_d(T'_d(p^e))$ together with the amended 'normalisability' condition (ii), though this destroys the 'even-odd' symmetry of the statement of the result.

We also have the following result (the proof of which will follow its corollaries).

<u>Theorem 3.1.3.</u> Suppose P_1 , $P_2 \in M_1(\Pi)$ and $P_1 \neq P_2$. Then if $0 < \lambda < 1$, the p.m. $P_0 = \lambda P_1 + (1-\lambda)P_2$ belongs to $G(\Pi)$ but not to $M(\Pi)$.

In the next section we refer to some results of Spitzer which show that there sometimes do/exist complementary pairs of distinct Markov chains in $M_1(\Pi)$, so that the following corollary is of some interest.

<u>Corollary.</u> If P , P' form a complementary pair of distinct Markov chains in $M_1(\Pi)$, then $\frac{1}{2}(P+P')$ belongs to $G_0(\Pi)$ but not $M_0(\Pi)$.

A corollary to the corollary is that if the above Markov chains P , P' belong to $E(\Pi)$, then $E_0(\Pi)$ is a proper subset of the set of extreme points of $G_0(\Pi)$.

<u>Proof of the Theorem 3.1.3.</u> It is sufficient to prove that P_0 is not a Markov chain. Suppose the contrary reuslt that P_0 *is* a Markov chain. Then P_0 necessarily belongs to $M_1(\Pi)$. Thus, for $\ell = 0, 1, 2$, we have from Theorem 3.1.2 that there is a function $p_\ell : S \to R_+$ with (for definiteness) $p_\ell(s) = 1$, and a strictly positive constant a_ℓ such that for all $i \in O$,

$$P_\ell(X_{\partial i} = x_{\partial i}) = a_\ell \sum_{x_i \in S} p(x_i) \prod_{j \in \partial i} p_\ell(x_j) q(x_j , x_i)$$

We thus obtain

$$a_o \prod_{j \in \partial i} p_o(x_j) = a_1\lambda \prod_{j \in \partial i} p_1(x_j) + a_2(1-\lambda) \prod_{j \in \partial i} p_2(x_j)$$

valid for all $x_{\partial i} \in S^{\partial i}$. This can only be the case if $p_0 = p_1 = p_2$, implying, by Theorem 3.1.2, that $P_0 = P_1 = P_2$ in contradiction to our hypothesis.

3.2 Convergence to Markov chains

In this section we continue consideration of our h.h.M. specification Π, and consider briefly a topic capable of much more extensive investigation.

Let $\{V_n\}$ be an increasing sequence in V such that $V_n \nearrow A$. For each n let $P^{(n)}$ be a p.m. on $(S^{\xi V_n}, F(\xi V_n))$ with density of the form

$$(1) \qquad P^{(n)}(X_{\partial V_n} = x_{\partial V_n})\tilde{\pi}_{V_n}(x_{\partial V_n}, x_{V_n}) \quad ,$$

that is, the distribution of the random variables X_{V_n} conditional on the (σ-field generated by) the random variables $X_{\partial V_n}$ is in accordance with the specification Π.

We will say that the sequence of p.m.s $\{P^{(n)}\}$ *converges* **to** a p.m. P on (S^A, F), and will write $P^{(n)} \Rightarrow P$, if for each $V \in V$, the associated sequence of marginal p.m.s $P_V^{(n)}$ on $(S^V, F(V))$ (defined for all sufficiently large n) converges to the marginal p.m. P_V in the sense of pointwise convergence of the corresponding densities. (If we are prepared to extend each p.m. $P^{(n)}$, in any way we like, to a p.m. on (S^A, F), then this convergence is simply weak convergence in the usual sense.) It is easy to see that if $P^{(n)} \Rightarrow P$, then $P \in G(\Pi)$. Of particular interest are those sequences $\{P^{(n)}\}$ where each member $P^{(n)}$ assigns probability one to a given $x_{\partial V_n} \in S^{\partial V_n}$. It is a straightforward exercise to check that the limits of such sequences (where they exist) belong to $M(\Pi)$.

Trivially, if $P \in G(\Pi)$ then P is the limit of at least one sequence $\{P^{(n)}\}$ with densities of the form (1), namely that given by $P^{(n)} = P_{\xi V_n}$ for all n; of more interest would be the problem of identifying all such sequences which converge to P. Note that if S is finite and P is the *sole* element of $G(\Pi)$ (so that necessarily $P \in M_0(\Pi)$), then the usual form of argument based on compactness and the selection of subsequences shows that *every* such sequence $\{P^{(n)}\}$ converges to P.

We consider in detail only a special case: we take the state space S to be finite and consider the convergence of sequences of a particular kind to homogeneous Markov chains.

Let i_0 be a given element of A, and define the sequence $\{V_n\}$ by

(2) $$V_0 = \{i_0\} \ , \ V_n = \xi V_{n-1} \quad , \quad n \geqslant 1 \ .$$

As in section 3.1, let $\Phi(p,q)$ be a fixed h.h.M. interaction such that $\Pi = \Pi(p,q)$. Fix elements of Ψ by the convention

(3) $$\sum_{x \in S} p^*(x) = 1 \quad , \quad p^* \in \Psi \ .$$

For any $p^* \in \Psi$ and for any $n \geqslant 0$, define the (unnormalised) density $f_{p^*,n}$ on $S^{V_{n+1}}$ by

$$f_{p^*,n}(x_{V_{n+1}}) = \prod_{i \in \partial V_n} p^*(x_i) \prod_{i \in V_n} p(x_i) \prod_{\{j,k\} \in N(V_{n+1})} q(x_j \ , \ x_k)$$

Define the p.m. $P^{(p^*,n)}$ on $(S^{V_{n+1}}, F(V_{n+1}))$ by

$$P^{(p^*,n)}(X_{V_{n+1}} = x_{V_{n+1}}) = a_{p^*,n} \, f_{p^*,n}(x_{V_{n+1}})$$

where $a_{p^*,n}$ is the appropriate normalising constant. The sequence of p.m.s $\{P^{(p^*,n)}\}$ is of the form introduced at the beginning of this section. In particular if $x \in S$, and $\delta_x \in \Psi$ is as defined in section 3.0, then $P^{(\delta_x,n)}$ is just the restriction to $(S^{V_{n+1}}, F(V_{n+1}))$ of the p.m. $\pi_{V_n}(x_{\partial V_n}, \cdot)$, where $x_i = x$ for all $i \in \partial V_n$.

Now suppose $p_0 \in \Psi$. Define the sequence $\{p_n\}$ in Ψ by

(4) $$p_n = T_d'(p_{n-1}) \quad , \quad n \geq 1.$$

For any $m \geq 0$ consider the density of $P^{(p_0,m+1)}$ (on $S^{V_{m+2}}$). By summing over $x^{\partial V_{m+1}} \in S^{\partial V_{m+1}}$, and recalling the definition of T_d' we obtain the identity

$$P_V^{(p_0,m+1)} = P_V^{(p_1,m)}$$

valid for any $V \subset V_{m+1}$ (where as usual the suffix V denotes the marginal p.m. associated with the set V). Iterate this relation to obtain

(5) $$P_V^{(p_0,m+n)} = P_V^{(p_n,m)} \quad , \quad n,m \geq 0 \quad , \quad V \subset V_{m+1} \quad .$$

For any $\tilde{p} \in \Psi$, define the subset $D_{p,q}(\tilde{p})$ of Ψ to be the set of those $p_0 \in \Psi$ such that the sequence $\{p_n\}$ defined by (4) converges pointwise to $\tilde{p}$ (under our convention (3)). $D_{p,q}(\tilde{p})$ may be regarded as a kind of 'domain of attraction' in Ψ of $\tilde{p}$ with respect to the interaction $\Phi(p,q)$. The following theorem is now almost obvious.

<u>Theorem 3.2.1.</u> Given $\tilde{p}, p_0 \in \Psi$, the following two statements are equivalent.

(i) $p_0 \in D_{p,q}(\tilde{p})$

(ii) There exists $P \in M_0(\Pi)$ corresponding to $\tilde{p}$ in the sense of Theorem 3.1.1, and $P^{(p_0,n)} \Rightarrow P$.

<u>Proof.</u> Suppose first that $p_0 \in D_{p,q}(\tilde{p})$. Under our convention (3) the mapping $T'_d : \Psi \to \Psi$ is as given by (4) of section 3.1. Recall that the denominator of the right-hand side of this equation is strictly positive. It follows easily that because p_n converges to $\tilde{p}$, $T'_d(p_n)$ converges to $T'_d(\tilde{p})$; hence $\tilde{p} = T'_d(\tilde{p})$ and there exists $P \in M_0(\Pi)$ corresponding to $\tilde{p}$. For any $V \in V$, choose m such that $V \subset V_{m+1}$. As $n \to \infty$, $f_{p_n,m}$ converges pointwise to $f_{\tilde{p},m}$ on $S^{V_{m+1}}$. Summing over $x_{V_{m+1}} \in S^{V_{m+1}}$ we obtain also that $a_{p_n,m} \to a_{\tilde{p},m}$, and so deduce that $P_V^{(p_n,m)}$ converges to P_V (as $n \to \infty$). Using (5), $P_V^{(p_0,n)}$ converges to P_V, and since V is arbitrary it follows that $P^{(p_0,n)} \Rightarrow P$.

Now suppose that $P^{(p_0,n)} \Rightarrow P$ where $P \in M_0(\Pi)$ corresponds to $\tilde{p}$. In particular $P_{V_1}^{(p_0,n)}$ converges to P_{V_1}, and so by (5) $P_{V_1}^{(p_n,0)}$ converges to P_{V_1}. It follows that for any $x_{\partial V_0} \in S^{\partial V_0}$,

$$\prod_{i \in \partial V_0} \frac{p_n(x_i)}{p_n(s)} \to \prod_{i \in \partial V_0} \frac{\tilde{p}(x_i)}{\tilde{p}(s)}$$

and so $p_0 \in D_{p,q}(\tilde{p})$.

<u>Remarks.</u> 1. The above theorem may be regarded as identifying those sequences of the form $\{P^{(p_0,n)}\}$ which converge to a given $P \in \mathcal{M}_0(\Pi)$. If we regard Π as fundamental, then the statement of the theorem suffers from the same weakness as that of Theorem 3.1.1, that of being dependent on the choice of interaction $\Phi(p,q)$ such that $\Pi = \Pi(p,q)$.

2. It is not difficult to extend the theorem to the more general case of countable S. Given $p_0 \in \Psi$, it is necessary to re-express the definition of the convergence to $\tilde{p}$ of the sequence $\{p_n\}$ in Ψ given by (4), by saying that it occurs if

$$\frac{p_n(x)}{p_n(s)} \to \frac{\tilde{p}(x)}{\tilde{p}(s)} \quad ;$$

it is also necessary to consider only those p_0 such that the sequence $\{p_n\}$ actually exists, and satisfies the condition that for all n and all $x \in S$,

$$\frac{p_n(x)}{p_n(s)} \leqslant \rho(x)$$

for some function $\rho : S \to R_+$ such that

$$\sum_x \sum_y \rho(x) q(x,y) \rho(y) < \infty .$$

The theorem then goes through much as before, though there are tedious details concerned with verifying the existence of various quantities.

3.3 Markov specifications represented by stochastic matrices

In this section we consider the representation of an h.h.M. specification by a stochastic matrix associated with a given 'direction' in our regular tree A. This is natural when we start with a homogeneous Markov chain with known 'transition matrix', and wish to consider the associated specification. The results obtained will also enable us to tie up more easily the theory developed so far with that of Spitzer*for the binary state space.

Let $<$ be a partial ordering on the elements of A such that

(1) (i) for each $i,j \in A$ with $i \sim j$, either $i < j$ or $j < i$ but not both,

(ii) for each $i \in A$, there is precisely one neighbour j of i such that $j < i$ - denote this neighbour by α_i.

If α_i is thought of as the 'predecessor' of each i, then the tree A may be thought of as 'evolving' sequentially, with d 'branches' at each step. For each $i \in A$, define also A_i^* to be the set of those elements j of A such that α_i belongs to the path in the tree connecting j to i, i.e. $A_i^* = A_{i,\alpha_i}$ in the notation introduced in section 2.0.

The following lemma is little more than a restatement of part of Lemma 2.1.3.

Lemma 3.3.1. A p.m. P on $(S^A,\mathcal{F})$ is a Markov chain if and only if

(2) for each $i \in A$, $F \in \mathcal{F}(i)$, $P(F/\mathcal{F}(A_i^*)) = P(F/\mathcal{F}(\alpha_i))$.

* and Preston

<u>Proof.</u> If P is a Markov chain, then (2) is immediate from Lemma 2.1.3. Conversely, if P satisfies (2) then we proceed essentially as in the proof of the implication (iii) $\Rightarrow$ (i) of that lemma: for any $V \in \mathcal{V}^*$, let $i, i_1, i_2, \ldots, i_n$ be any ordering of the elements of V such that for each r, $1 \leqslant r \leqslant n$, α_{i_r} belongs to the set $\{i, i_1, \ldots, i_{r-1}\}$. The element i is obviously uniquely determined; for each remaining $j \in V$, $\alpha_j = \partial_i(j)$ (where $\partial_i(j)$ is as defined in section 2.0). Using (2) we deduce that

$$P(X_V = x_V) = P(X_i = x_i) \prod_{j \in V-i} P(X_j = x_j / X_{\alpha_j} = x_{\alpha_j}).$$

It follows from Lemma 2.1.2 that P is a Markov chain.

A Markov chain on $(S^A, \mathcal{F})$ will be called *stationary* (with respect to <) if it is invariant under those graph isomorphisms of the tree (onto itself) which preserve the partial ordering <. Homogeneous Markov chains are stationary, though the converse is not generally true.

Now let $q : S \times S \to R_+$ be a *stochastic matrix*, i.e. satisfy the condition that for all $x \in S$, $\sum_{y \in S} q(x,y) = 1$. A Markov chain P will be said to have q as its *transition matrix* (with respect to the partial ordering <) if for all $i \in A$, $x \in S$,

$$P(X_i = x / \mathcal{F}(\alpha_i)) = q(X_{\alpha_i}, x) \qquad \text{P-a.s.}$$

It follows, in an obvious generalisation of familiar one-dimensional Markov chain theory, that a Markov chain P with transition matrix q is stationary if and only if there exists $p' : S \to R_+$ (with $\sum_{x \in S} p'(x) = 1$) such that for all $i \in A$, $x \in S$, $P(X_i = x) = p'(x)$; that p' then satisfies $p'(x) = \sum_{y \in S} p'(y)q(y,x)$, $x \in S$; and if q is additionally irreducible, that such a stationary Markov chain P exists if and only if q is positive-recurrent, P then being unique.

Let $\Phi(1,q)$ denote the Markov interaction on (S,A) whose single-site interaction functions are identically one, and whose pairwise interaction functions are given by $q_{\alpha_i,i} = q$ for each $i \in A$. Let $\Pi(1,q)$ denote the corresponding Markov specification. If a Markov chain has transition matrix q, then it is easy to see that $P \in M(\Pi(1,q))$, though the converse need not be the case. (See the examples of sections 3.4 and 3.5.)

Henceforth we suppose that the stochastic matrix q additionally satisfies:

(3)
- (i) $q(x,s) > 0$, $q(s,x) > 0$ for all $x \in S$ (where $s \in S$ defines the basepoint); this implies that q is irreducible, and that $\Phi(1,q)$, $\Pi(1,q)$ are hereditary;
- (ii) q is positive-recurrent and *reversible*, in the sense that the corresponding one-dimensional stationary Markov chain is reversible.

The stationary Markov chain P on (S^A,F) with transition matrix q is then necessarily homogeneous. (Indeed a *homogeneous* Markov chain associated with the tree A is simply the natural generalisation of a one-dimensional *reversible* Markov chain.) We may show directly that the corresponding hereditary Markov specification $\Pi(1,q)$ is also homogeneous: since q is reversible there exists a *symmetric* function $q^* : S \times S \to R_+$, and a function $p : S \to R_+$ such that

(4) $$q(x,y) = q^*(x,y)p(y) \quad , \quad x,y \in S.$$

(The functions p and q^* are each uniquely determined up to a strictly positive multiplicative constant, so that p is unique when regarded as an element of Ψ.) The Markov interaction $\Phi(1,q)$ is therefore equivalent to the h.h.M. interaction $\Phi(p,q^*)$, and thus $\Pi(1,q)$ is h.h.M..

The homogeneous Markov chains corresponding to $\Pi(1,q)$ may be determined - at least in principle - via Theorem 3.1.1., though we must first find a h.h.M. interaction (e.g. $\Phi(p,q^*)$) equivalent to $\Phi(1,q)$. It is useful in applications, and perhaps provides some additional insight, to have an analogous statement directly in terms of $\Phi(1,q)$, i.e. in terms of our reversible stochastic matrix q. We give this below, first introducing some additional notation. Given any function $p^* : S \to R_+$ define (as in section 3.1) the function $Qp^* : S \to R_+$ (where it exists) by

$$(Qp^*)(x) = \sum_{y \in S} q(x,y)p^*(y),$$

and the function $p^*Q : S \to R_+$ (where it exists) by

$$(p^*Q)(x) = \sum_{y \in S} p^*(y)q(y,x) \quad .$$

<u>Theorem 3.3.2.</u> Suppose the stochastic matrix q satisfies the conditions (3). Let $p \in \Psi$ be as uniquely determined by the decomposition (4) of q. Then there is a one-to-one correspondence between $\mathcal{M}_0(\Pi(1,q))$ and the set of functions $\hat{p} \in \Psi$ satisfying

$$(5) \quad \begin{cases} \text{(i)} \ \hat{p}(x) \propto [(Q\hat{p})(x)]^d \\ \text{(ii)} \ \sum_{x \in S} p(x)\hat{p}(x)^{1+1/d} < \infty \end{cases}$$

Given $\hat{p}$ satisfying (5), the corresponding homogeneous Markov chain P is (uniquely) determined by its transition matrix $\hat{q}$ (with respect to the partial ordering $<$, or, since P is homogeneous, with respect to any other partial ordering satisfying the conditions (1)) given by

$$(6) \qquad \hat{q}(x,y) = \frac{q(x,y)\hat{p}(y)}{\sum_{z \in S} q(x,z)\hat{p}(z)}$$

and the corresponding 'stationary' distribution is given by

$$(7) \qquad P(X_i = x) \propto p(x)\hat{p}(x)^{1+1/d} \quad , \quad i \in A.$$

<u>Proof.</u> The correspondence asserted in the first part of the theorem may be deduced by a careful translation of the statement of Theorem 3.1.1 - simply note that the interaction $\Phi(1,q)$ is equivalent to the interaction $\Phi(p,q^*)$ where p and q^* are as given by (4). The function $\hat{p}$ of this theorem corresponds to the function $\tilde{p}$ of Theorem 3.1.1 via the relation

$$(8) \qquad \tilde{p}(x) \propto p(x)\hat{p}(x) \quad .$$

However, it is of some interest to deduce the correspondence directly, though in two distinct stages, from our basic Theorem 2.2.1.

1. Suppose first that $P \in \mathcal{M}(\Pi(1,q))$ is stationary. Consider the representation of the marginal p.m.s P_V, $V \in \mathcal{V}^*$, given by Theorem 2.2.1 with the interaction $\Phi(1,q)$. The functions p_i^j defined there satisfy

$$p_i^j = \begin{cases} \tilde{p} & , \quad i < j \\ \hat{p} & , \quad j < i \end{cases}$$

for some functions $\tilde{p},\hat{p} \in \Psi$, and we have

$$(9)\begin{cases} \text{(i)} & \hat{p}(x) \propto [\,(Q\hat{p})(x)\,]^{d} \\ \text{(ii)} & \tilde{p}(x) \propto (\tilde{p}Q)(x)\,[\,(Q\hat{p})(x)\,]^{d-1} \\ \text{(iii)} & \sum_{x \in S} \tilde{p}(x)(Q\hat{p})(x) < \infty \end{cases}$$

Conversely, given $\tilde{p},\hat{p} \in \Psi$ satisfying these conditions, there exists a unique stationary Markov chain $P \in M(\Pi(1,q))$ to which $\tilde{p},\hat{p}$ correspond.

2. Now let $q(x,y) = q^*(x,y)p(y)$ as in (4). By considering the sets $V \in \tilde{V}$, we see that a stationary Markov chain P in $M(\Pi(1,q))$ is homogeneous if and only if the corresponding functions $\tilde{p},\hat{p}$ satisfy (8), Therefore if P *is* homogeneous, (8) and (9) together give the result that $\hat{p}$ satisfies (5). Conversely if $\hat{p}$ satisfies (5), define $\tilde{p} \in \Psi$ via (8); again using the decomposition (4) of q, we see that $\tilde{p},\hat{p}$ together satisfy (9), and so there is a unique $P \in M_0(\Pi(1,q))$ to which $\hat{p}$ corresponds.

3. We complete the proof by considering $P \in M_0(\Pi(1,q))$ and $\tilde{p},\hat{p}$ corresponding to P as above. Let $i,j \in A$ be such that $i = \alpha_j$. Then by Theorem 2.2.1,

$$P(X_i = x,\ X_j = y) \propto \tilde{p}(x)q(x,y)\hat{p}(y)$$

Thus the transition matrix of P is as given by (6), and we also have

$$\begin{aligned} P(X_i = x) &\propto \tilde{p}(x)(Q\hat{p})(x) \\ &\propto p(x)\hat{p}(x)^{1+1/d} \end{aligned}$$

Remarks.* 1. It may be seen from the proof of the above theorem that $M(\Pi(1,q))$ contains a stationary, non-homogeneous Markov chain if and only if there exist $\tilde{p},\hat{p} \in \Psi$ satisfying (9) but not (8). It is not immediately obvious whether this can ever happen. If it can then $M(\Pi(1,q))$ must contain uncountably-infinitely many such Markov chains - one (or more) for each partial ordering $<$ satisfying the conditions (1).

2. It is now clear that for any h.h.M. specification Π, $M_0(\Pi)$ is in a one-to-one correspondence with the stochastic matrices q satisfying (3) and $\Pi = \Pi(1,q)$. In principle, this should lead to alternative derivations of Theorems 3.1.1 and 3.1.2, and indeed this *is* the basis of Spitzer's corresponding results for the binary case.

3. There is an obvious analogue of Theorem 3.2.2 for $M_1(\Pi(1,q))$; this may be similarly derived from Theorem 2.2.1, or by a translation exercise on the statement of Theorem 3.1.2. We do not state the result here, though we do use it in the next section.

Now suppose S is finite. Given our stochastic matrix q satisfying (3) and any $\hat{p} \in \Psi$, define $\bar{D}_{1,q}(\hat{p})$ to be the set of $p_0 \in \Psi$ such that the sequence $\{p_n\}$ given by $p_n(x) \propto [(Qp_{n-1})(x)]^d$, $n \geq 1$, converges to $\hat{p}$ in the sense introduced in the previous section. Then Theorem 3.2.1 may be translated into the result that $p_0 \in \bar{D}_{1,q}(\hat{p})$ if and only if there exists $P \in M_0(\Pi(1,q))$ corresponding to $\hat{p}$ in the sense of Theorem 3.3.2 and a suitably defined sequence of p.m.s $\{P^{(p_0,n)}\}$ converges to P.

* We should additionally remark that the function $p \in \Psi$ determined by (4) is just the 'stationary' distribution corresponding to the reversible stochastic matrix q. One solution of (5) is given by $\hat{p}(x) = 1$ for all x, yielding the stationary (homogeneous) Markov chain with transition matrix q.

3.4 The binary state space

We take the state space $S = \{0,1\}$, and consider h.h.M. specifications, where the basepoint $s_A \in S^A$ is given by $s_i = 0$ for all $i \in A$. The state space S is finite, and so we may take the approach of the preceding section and represent any such specification Π by the transition matrix

$$(1) \qquad q = q(\sigma,\tau) = \begin{pmatrix} \sigma & 1-\sigma \\ 1-\tau & \tau \end{pmatrix}$$

of any one of its corresponding homogeneous Markov chains. (By the corollary to Theorem 3.1.1 there is at least one such Markov chain.) The hereditary condition gives

$$(2) \qquad 0 < \sigma < 1 \quad , \quad 0 \leqslant \tau < 1 \quad .$$

Conversely, any such hereditary stochastic matrix q is necessarily reversible and so defines an h.h.M. specification $\Pi(1,q)$. (Note that we may also consider those stochastic matrices $q(\sigma,\tau)$ defined by $\sigma = 0$, $0 < \tau < 1$ - these are hereditary with respect to the basepoint s_A defined by $s_i = 1$ for all $i \in A$.) The stationary Markov chain P corresponding to q is homogeneous. Therefore, for a given h.h.M. specification Π we may characterise $M_0(\Pi)$ simply by identifying all hereditary stochastic matrices q such that $\Pi = \Pi(1,q)$.

Now for any such $q = q(\sigma,\tau)$ we have $\Pi = \Pi(1,q)$ if and only if, for all $i \in A$

$$(3) \qquad \frac{\tilde{\pi}_i(x_{\partial i},1)}{\tilde{\pi}_i(x_{\partial i},0)} = \frac{\tau^r(1-\tau)^{d-r}}{(1-\sigma)^{r-1}\,\sigma^{d+1-r}} \quad ,$$

where r, $0 \leqslant r \leqslant d+1$, is the number of neighbours j of i such that $x_j = 1$. (If $\Pi = \Pi(1,q)$ then (3) follows as an easy calculation, and the converse result is true because any hereditary Markov specification Π' is uniquely determined by the kernels π'_i , $i \in A$.) Therefore the Markov chains in $M_0(\Pi)$ are identified by those pairs (σ,τ) satisfying (2) and (3). This is the approach of *Spitzer (1975b)**. Our intention here is simply to show that Spitzer's results follow as an easy application of the theorems presented in the preceding sections of this chapter, and provide examples of all the various possible phenomena discussed there.

The right-hand side of (3) is an increasing function of r if $\sigma + \tau \geqslant 1$ - in which case the specification Π is *attractive*, and a decreasing function of r if $\sigma + \tau \leqslant 1$ - in which case Π is *repulsive*.

We now regard the hereditary transition matrix $q = q(\sigma,\tau)$ as fixed, and look at the Markov chains corresponding to $\Pi(1,q)$. Fix any element p of Ψ by our original convention $p(0) = 1$, and then write simply p for $p(1)$. (We allow $p = \infty$ to cover the case $p(0) = 0$, $p(1) > 0$.) Define the function $g : [0,\infty] \to (0,\infty)$ by

(4) $$g(p) = \left[\frac{1-\tau+\tau p}{\sigma+(1-\sigma)p}\right]^d \qquad \left(\text{with } g(\infty) = \left[\frac{\tau}{1-\sigma}\right]^d\right)$$

By Theorem 3.3.2 the elements of $M_0(\Pi(1,q))$ are in a one-to-one correspondence with the (necessarily strictly positive) solutions $\hat{p}$ of

(5) $$\hat{p} = g(\hat{p}) \ .$$

* and *Preston (1974)* : see the note at the end of Chapter 0.

(This is Spitzer's result.) For each such $\hat{p}$, the transition matrix of the corresponding homogeneous Markov chain P is given by $q(\hat{\sigma},\hat{\tau})$ where

$$\hat{\sigma} = \frac{\sigma}{\sigma+(1-\sigma)\hat{p}} \quad , \quad \hat{\tau} = \frac{\tau\hat{p}}{1-\tau+\tau p} \quad ;$$

and for all $i \in A$,

$$\frac{P(X_i = 1)}{P(X_i = 0)} = \frac{1-\sigma}{1-\tau}\,\hat{p}^{1+1/d}$$

One solution of (5) is of course $\hat{p} = 1$, giving the Markov chain with transition matrix q.

From the obvious analogue of Theorem 3.3.2 for $M_1(\Pi(1,q))$, (or by Theorem 3.1.2 plus a little work), the elements of this set are in a one-to-one correspondence with the ordered pairs (p^e,p^o) of solutions of

$$(6) \qquad p^o = g(p^e) \quad , \quad p^e = g(p^o) \quad .$$

For each such pair (p^e,p^o), the corresponding Markov chain P has a transition matrix from α_i to any $i \in E$ (recalling that E is the set of 'even' vertices in A, and that α_i is the 'predeccesor' of i), given by $q(\sigma^e,\tau^e)$ where

$$\sigma^e = \frac{\sigma}{\sigma+(1-\sigma)p^e} \quad , \quad \tau^e = \frac{\tau p^e}{1-\tau+\tau p^e} \quad ;$$

and we have

$$\frac{P(X_i = 1)}{P(X_i = 0)} = \frac{1-\sigma}{1-\tau}\,(p^e)^{1+1/d}$$

Corresponding results for any site $i \in O$ are given by replacing p^e, σ^e, τ^e by p^o, σ^o, τ^o. When $p^o = p^e$ we have a homogeneous Markov chain. The solutions of (6) with $p^o \neq p^e$ come in complementary pairs.

We may make the following deductions from these results. (See Spitzer's paper for details.)

(i) When the specification $\Pi(1,q)$ is attractive ($\sigma + \tau \geqslant 1$), the equation (5) has 1, 2, or 3 solutions, and so there are 1, 2 or 3 corresponding homogeneous Markov chains. For $d \geqslant 2$, all of these possibilities may occur. Further since g is increasing, the only solutions of (6) satisfy $p^o = p^e$, and so $M_1(\Pi(1,q)) = M_0(\Pi(1,q))$.

(ii) When the specification $\Pi(1,q)$ is repulsive ($\sigma + \tau \leqslant 1$), g is decreasing and so (5) has exactly one solution, implying that there is precisely one corresponding homogeneous Markov chain. For any $d \geqslant 2$, and suitable σ, τ, the equation (6) has at least one unordered solution pair (p^e, p^o) with $p^o \neq p^e$, so that $M_1(\Pi(1,q))$ is strictly larger than $M_0(\Pi(1,q))$.

Spitzer (1975b) gives a complete classification of results for the case $d = 2$ and all values of σ and τ, $0 < \sigma < 1$, $0 < \tau < 1$. He also shows that for general d and in both the attractive and repulsive cases, $|G(\Pi(1,q))| = 1$ if and only if $|M_1(\Pi(1,q))| = 1$, i.e. if and only if the equation $g\{g(p)\} = p$ has exactly one solution. His work relies on some results of *Preston (1974)* for attractive specifications with $S = \{0,1\}$ and a general neighbour relation on A. (The repulsive case is dealt with by the interchange of the labels 0 and 1 in S at the vertices in the subset O of A; this gives a transformed specification which is attractive, though not homogeneous.) In Chapter 4 we give a relatively quick derivation of these results, applicable to more general state spaces, but relying on our tree structure.

We now consider briefly, and without any formal proofs, the sets $\bar{D}_{1,q}(\hat{p})$ (see section 3.3) in Ψ associated with those $\hat{p} \in \Psi$ which correspond to the elements of $M_0(\Pi(1,q))$. Recall that under our current convention Ψ is simply the space $[0,\infty]$; hence, given $\hat{p}$, $\bar{D}_{1,q}(\hat{p})$ is simply the set of $p_0 \in [0,\infty]$ such that the sequence $\{p_n\}$, given by $p_n = g(p_{n-1})$, $n \geqslant 1$, converges to $\hat{p}$. We again consider separately the attractive and repulsive cases.

(i) When the specification $\Pi(1,q)$ is attractive and the equation (5) has exactly one solution $\hat{p}$, then it is not difficult to show that $\bar{D}_{1,q}(\hat{p}) = [0,\infty]$. Thus all the sequences of p.m.s of the general form considered in Theorem 3.2.1 converge to the homogeneous Markov chain P corresponding to $\hat{p}$. (From Spitzer's results, P is the sole element of $G(\Pi(1,q))$.) In an obvious sense the Markov chain P may be described as 'stable'. When the equation (5) has three solutions, $\hat{p}_1 < \hat{p}_2 < \hat{p}_3$, it is again easy to show that

$$\bar{D}_{1,q}(\hat{p}_1) = [0,\hat{p}_2) \quad , \quad \bar{D}_{1,q}(\hat{p}_2) = [p_2] \quad , \quad \bar{D}_{1,q}(\hat{p}_3) = (p_2,\infty] .$$

We thus have one 'unstable' and two 'stable' Markov chains. The case where (5) has two solutions corresponds to the coincidence of either $\hat{p}_1$ and $\hat{p}_2$ or $\hat{p}_2$ and $\hat{p}_3$ in the three-solution case.

(ii) When the specification $\Pi(1,q)$ is repulsive, so that (5) always has exactly one solution $\hat{p}$, then $\bar{D}_{1,q}(\hat{p}) = [0,\infty]$ if and only if the equation (6) has no additional 'solution pairs'. Then again the corresponding Markov chain is 'stable'. For the case d = 2 we may check that when such an additional, unordered, solution pair (p^e,p^o) exists $(p^o \neq p^e)$, then for any p_0 other than $\hat{p}$ the sequence $\{p_n\}$ defined above converges to the cycle $p^e,p^o,p^e,p^o,\ldots$. Thus the homogeneous Markov chain is 'unstable', and the remaining pair of chains in $M_1(\Pi(1,q))$ are in a sense jointly stable.

We conclude this section by considering the 'extremely repulsive' homogeneous non-hereditary specification $\Pi = \Pi(1,q)$ defined by the non-hereditary transition matrix

$$q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

This is not strictly a 'specification', as for each $V \in \mathcal{V}$ the corresponding p.m. $\pi_V(x_{\partial V}, \cdot)$ is not defined for all $x_{\partial V} \in S^{\partial V}$. The corresponding class of p.m.s $G(\Pi)$ is, however, well-defined, and consists of all convex combinations of the two elements P^e, P^o, say, of $E(\Pi)$, where P^e assigns probability one to the single element $x_A \in S^A$ defined by

$$x_i = \begin{cases} 1 , & i \in E \\ 0 , & i \in O , \end{cases}$$

and where P^o is the complement of P^e. All elements of $G(\Pi)$ are Markov chains, so that $G(\Pi) = M_1(\Pi)$. We attempt to apply the preceding results to Π. Note that the equation (5) correctly identifies the single element of $M_0(\Pi)$. However, if $d \geq 2$, equation (6) only identifies this and the extreme pair of chains P^e, P^o. Its failure to identify the remaining elements of $M_1(\Pi)$ seems to stem, not so much from the technical difficulties in defining an interaction or specification in this case, as from the failure of the key Theorem 2.2.1 in the absence of the hereditary condition.

3.5 An example with $|S| = 3$

We consider as a further example the state space $S = \{0,1,2\}$, and the h.h.M. interaction $\Phi(1,q)$ defined by the single-site interaction function which is identically one, and the pairwise interaction function $q : S \times S \to R_+$, where

$$q(x,y) = \begin{cases} \alpha & , \quad x = y \\ \beta & , \quad x \neq y \end{cases} ,$$

with $\alpha + 2\beta = 1$ and $0 < \alpha < 1$. We write Π for the corresponding h.h.M. specification $\Pi(1,q)$. Since q is also a stochastic matrix, we may use any of the results of sections 3.1, 3.2, and 3.3 to study the Markov chains corresponding to Π. (We will also contrive to use some of the results of section 3.4.)

Theorem 3.1.2 implies that $M_1(\Pi)$ is in a one-to-one correspondence with the set of ordered pairs of (equivalence classes of) functions (p^e, p^o) in Ψ satisfying

$$(1) \qquad p^o = T'_d(p^e) \quad , \quad p^e = T'_d(p^o)$$

where $T'_d : \Psi \to \Psi$, is the mapping introduced in section 3.1. For each such pair the corresponding Markov chain is as constructed in that theorem. The solutions of (1) with $p^o = p^e = \tilde{p}$, say, i.e. the solutions of

$$(2) \qquad \tilde{p} = T'_d(\tilde{p})$$

correspond to the homogeneous Markov chains in $M_1(\Pi)$ i.e. to the elements of $M_0(\Pi)$. For each such $\tilde{p}$ the corresponding Markov chain may be constructed either as in Theorem 3.1.1 or Theorem 3.3.2.

We will represent any element p of Ψ by (p_0, p_1, p_2), where for $x = 0,1,2$, $p_x = p(x)$. Here we are as usual identifying p with any member of the equivalence class of functions it represents, so that $(p_0,p_1,p_2) = (\lambda p_0, \lambda p_1, \lambda p_2)$ for any strictly positive constant λ. Since q is stochastic, one solution of (2) is always given by $\tilde{p} = (1,1,1)$. The corresponding homogeneous Markov chain, which we denote by P, is that with transition matrix q.

For each $x \in S$ define $M_1^x(\Pi)$ to be the set of Markov chains in $M_1(\Pi)$ which are invariant under interchange of the remaining two elements of S. Similarly define $M_0^x(\Pi) = M_1^x(\Pi) \cap M_0(\Pi)$.

The elements of $M_1^0(\Pi)$ correspond to the solutions of (1) with $p^e = (1,p_1^e,p_1^e)$ and $p^o = (1,p_1^o,p_1^o)$. These are given by the solutions of

$$(3) \qquad p_1^o = g(p_1^e) \quad , \quad p_1^e = g(p_1^o)$$

where the function $g : [0,\infty] \to (0,\infty)$ is defined by

$$(4) \qquad g(p) = \left[\frac{\frac{1}{2}(1-\alpha) + \frac{1}{2}(1+\alpha)p}{\alpha + (1-\alpha)p} \right]^d$$

and, of course, the elements of $M_0^0(\Pi)$ correspond to the functions $\tilde{p} = (1,\tilde{p}_1,\tilde{p}_1)$ in Ψ where

$$(5) \qquad \tilde{p}_1 = g(\tilde{p}_1) \quad .$$

Now note that (4) is just the equation (4) of the preceding section with $\sigma = \alpha$, $\tau = \frac{1}{2}(1+\alpha)$. We may therefore use Spitzer's results for the binary state space to make the following deductions.

(i) When $\alpha \leqslant \frac{1}{3}$ (corresponding to $\sigma + \tau \leqslant 1$), $M_0^0(\Pi)$ has P as its sole element. However $M_1^0(\Pi)$ may or may not contain additional elements, depending on d and the precise value of α. (For example, if d = 3 and $\alpha = 0.1$ then numerical investigation shows that $M_1^0(\Pi)$ consists of P, together with the non-homogeneous Markov chain defined by $p_1^e \simeq 2.4871$, $p_1^o \simeq 0.4698$, and the complement of this chain - given by interchanging p_1^e and p_1^o.) Note that when $\alpha = \frac{1}{3}$ the specification Π is trivial, $G(\Pi)$ consisting solely of P, which in this case is such that the random variables X_i, $i \in A$, are independent with $P(X_i = x) = \frac{1}{3}$ for all $i \in A$, $x \in S$.

(ii) When $\alpha \geqslant \frac{1}{3}$, $M_1^0(\Pi)$ is equal to $M_0^0(\Pi)$ and may contain 1, 2 or 3 elements according to d and the precise value of α.

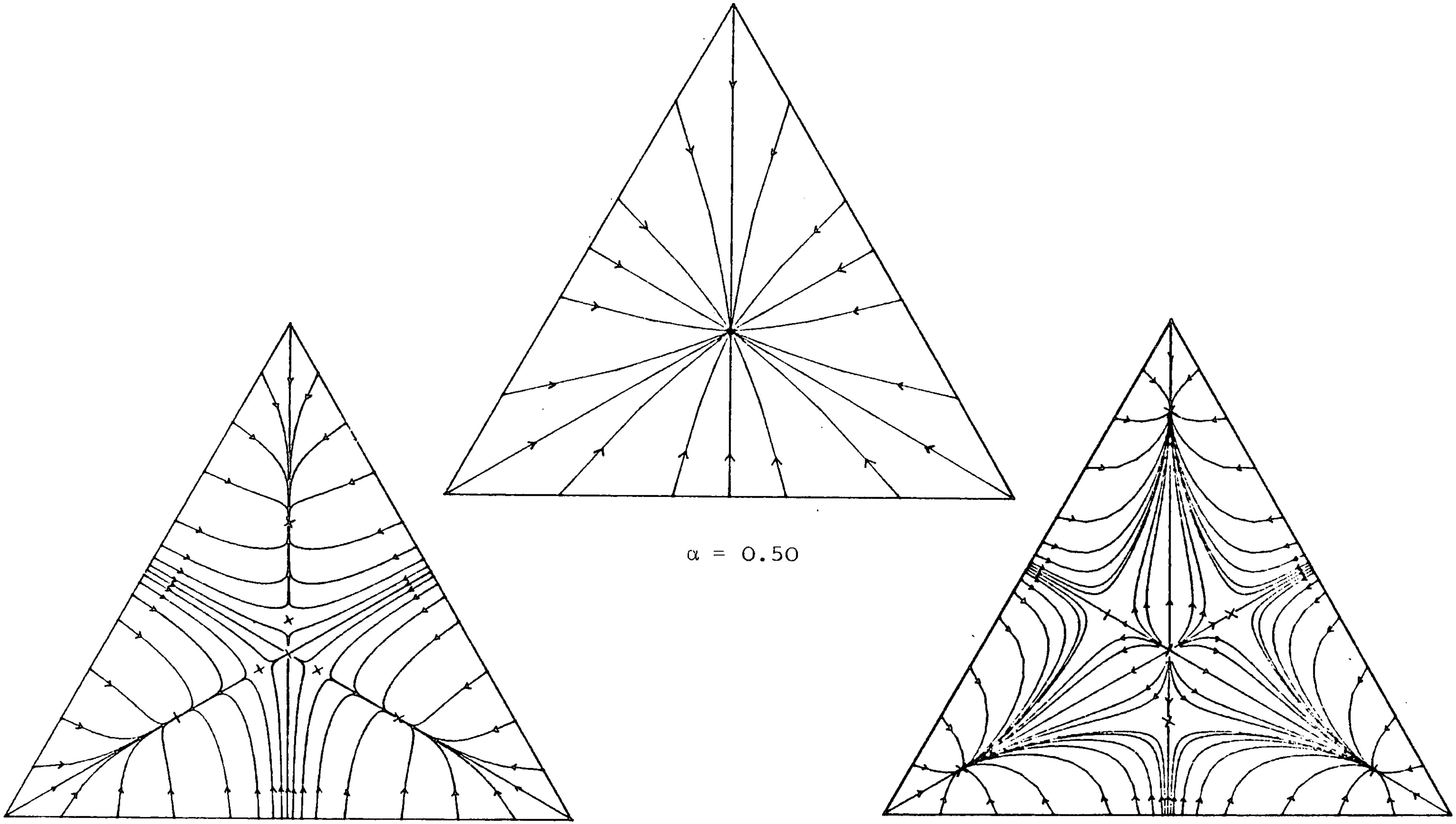
Note also that to each element of $M_1^0(\Pi)$ other than P, there correspond two further distinct elements of $M_1(\Pi)$ - one in $M_1^1(\Pi)$ and the other in $M_1^2(\Pi)$. Thus if $|M_1^0(\Pi)| = 2$ (respectively 3), then $|M_1(\Pi)| \geqslant 4$ (respectively 7).

We now consider in more detail the case where d = 2. Here we may compare (3), (4) and (5) with Spitzer's detailed results for d = 2 and binary state space, to obtain the following results for the present example. For all values of α the only solutions of (3) are given by $p_1^o = p_1^e = \tilde{p}_1$; thus $M_1^0(\Pi) = M_0^0(\Pi)$. An analysis of (5) shows that when $\alpha < 4\sqrt{2} - 5$ ($\simeq 0.6569$) this equation has only one solution $\tilde{p}_1 = 1$; thus P is the sole element of $M_0^0(\Pi)$; (for $M_0(\Pi)$ we can only say that $|M_0(\Pi)| \geqslant 1$). When $4\sqrt{2} - 5 < \alpha < 1$ and $\alpha \neq \frac{2}{3}$ the equation (5) has 3 solutions, $p_1 = 1$, p_1' and p_1'' ; thus $|M_0^0(\Pi)| = 3$ and $|M_0(\Pi)| \geqslant 7$. When $\alpha = 4\sqrt{2} - 5$ the two solutions p_1' , p_1'' above are coincident at the value $\frac{1}{2}$, and when $\alpha = \frac{2}{3}$ we have $p_1' = \frac{1}{4}$, $p_1'' = 1$: hence in both these cases $|M_0^0(\Pi)| = 2$ and $|M_0(\Pi)| \geqslant 4$.

Figure 1 shows for d = 2 and α = 0.50, 0.66 and 0.70, the computer-produced 'trajectories' in Ψ of the sequences $\{p^n\}$ defined by $p^n = T_d'(p^{n-1})$ and various starting values p^0. For symmetry between the 3 states of S we have chosen the 'normalisation'

$$(6) \qquad p_0 + p_1 + p_2 = 1 \quad , \quad p \in \Psi \ .$$

Thus Ψ is naturally represented by the equilateral triangle which forms the surface in R_+^3 defined by (6). We indicate by crosses the fixed points of the mapping T_d', defining the elements of $M_0(\Pi)$. For each of these values of α we may therefore identify (approximately for α = 0.66) the subsets $D_{1,q}(\tilde{p})$ of Ψ which form the 'domains of attraction' of the fixed points $\tilde{p}$ of T_d'. (It seems, in particular, that for α = 0.50, $D_{1,q}(\tilde{p}) = \Psi$ for the single-fixed point $\tilde{p} = (\frac{1}{3}\ ,\ \frac{1}{3}\ ,\ \frac{1}{3})$.) There are considerable qualitative differences between the 3 cases, in particular as regards the 'stability' of the homogeneous Markov chains. Note also that for both α = 0.66 and α = 0.70, it appears that $|M_0(\Pi)|$ is precisely 7. We suppose that for all values of α our earlier inequalities for $|M_0(\Pi)|$ (with d = 2) are in fact equalities.



α = 0.66

α = 0.50

α = 0.70

Figure 1

4 ATTRACTIVE AND REPULSIVE SPECIFICATIONS

4.0 Introduction

In this chapter we continue consideration of the situation where A is the regular infinite tree with d + 1 edges meeting at each vertex. The definitions, notation, and conventions given in section 3.0 continue to apply.

In section 4.1 we consider a given h.h.M. specification Π which has the property of being either *attractive* or *repulsive* with respect to a given total ordering $\leqslant$ on the state space S. We show that if $(S,\leqslant)$ satisfies the additional conditions mentioned in the introductory chapter, then it is frequently possible to draw much stronger conclusions - in particular about the sets $M(\Pi)$ and $G(\Pi)$ - than those of Chapter 3; and further that in the important special case d = 1, provided Π is strictly positive the set $G(\Pi)$ contains precisely one element, necessarily a homogeneous Markov chain.

In section 4.2 we give an example in which the state space S is the set of non-negative integers. The h.h.M. specification considered is such that the conditional distribution at each single site (or vertex) of A is Poisson; the specification is repulsive with respect to the natural ordering on S. We consider briefly the application of the results of section 4.1.

4.1 Definitions and Theorems

We will say that an h.h.M. interaction $\Phi(p,q)$ on (S,A) is *attractive* (with respect to the given total ordering $\leqslant$ on S), if for all $x_1,x_2,y_1,y_2 \in S$ with $x_1 \leqslant x_2$, $y_1 \leqslant y_2$

$$(1) \qquad q(x_1,y_1)q(x_2,y_2) \geqslant q(x_1,y_2)q(x_2,y_1) \quad .$$

We will say that $\Phi(p,q)$ is *repulsive* if we have '$\leqslant$' instead of '$\geqslant$' in (1) above.

Let Π be a given h.h.M. specification. We define Π to be *attractive* [*repulsive*] (with respect to $\leqslant$) if there exists an attractive [respectively repulsive] h.h.M. interaction $\Phi(p,q)$ such that $\Pi = \Pi(p,q)$.

Remarks. 1. For any $x,y \in S$ define $\pi^*(x,y)$ to be the common value of $\tilde{\pi}_i(x_{\partial i}, y)$ where $x_j = x$ for all $j \in \partial i$. Then Π is attractive [repulsive] if and only if for all $x_1 \leqslant x_2$, $y_1 \leqslant y_2$

$$\pi^*(x_1,y_1)\pi^*(x_2,y_2) \underset{[\leqslant]}{\geqslant} \pi^*(x_1,y_2)\pi^*(x_2,y_1)$$

Thus if Π is attractive [repulsive] *every* h.h.M. interaction $\Phi(p,q)$ such that $\Pi = \Pi(p,q)$ is attractive [repulsive].

2. Further, if q' is a reversible stochastic matrix such that $\Pi = \Pi(1,q')$ in the sense of section 3.3, by using the decomposition of q' given by (4) of that section, we see that Π is attractive or repulsive if and only if the appropriate version of (1) holds (with q' replacing q). We have at once that when the state space S is binary, the definitions of attractiveness and repulsiveness given here coincide with the well-known ones given (in terms of σ and τ) in section 3.4,

the definitions then being independent of the two possible orderings of the elements of S. In this case every h.h.M. specification is either attractive or repulsive, but for general S this is not true; (the h.h.M. specification of section 3.5 - where $|S| = 3$ - is neither attractive nor repulsive with respect to any ordering of the elements of S, except in the uninteresting case where $\alpha = \beta$).

3. In a considerably more general setting *Preston (1976, Chapter 9)* gives a somewhat different definition of an attractive specification; in the present situation specifications which are attractive in our sense also satisfy his definition: see his *Theorem 9.5*, which is essentially a generalisation of *Holley's inequality* (*Preston (1974)*). Of our two main theorems below, the first (for attractive specifications) is more or less implicit in the work of *Preston (1976)*; we give a shorter and considerably more direct proof, which is dependent, however, on the existence of our tree structure. The second theorem below (for repulsive specifications) then has a similar proof to the first.

If S has a *minimal* element x^-, say, (one such that $x^- \leq x$ for all $x \in S$), for each $V \in \mathcal{V}$ we define $\pi_V(-,\cdot)$ to be the p.m. $\pi_V(x_{\partial V},\cdot)$ where $x_i = x^-$ for all $i \in \partial V$; if S has a *maximal* element x^+, say, we similarly define the p.m. $\pi_V(+,\cdot)$.

Now let i_0 be a fixed reference element of A, and define the sequence $\{V_n\}$ in $\mathcal{V}^*$ as in section 3.2, i.e. by

$$V_0 = \{i_0\} \quad , \quad V_n = \xi V_{n-1} \quad , \quad n \geq 1 \quad .$$

Observe that $V_n \nearrow A$. We now state both the principal theorems of this chapter, before embarking on their (considerably overlapping) proofs.

Theorem 4.1.1. Suppose that S has both a minimal element x^- and a maximal element x^+, and that the given h.h.M. specification Π is attractive. Then there exist Markov chains P^-, P^+ (not necessarily distinct) in $M_0(\Pi)$ such that, as $n \to \infty$,

$$\pi_{V_n}(-,\cdot) \Rightarrow P^- \quad , \quad \pi_{V_n}(+,\cdot) \Rightarrow P^+$$

(where $\Rightarrow$ denotes weak convergence). If $P^- = P^+$ then $|G(\Pi)| = 1$, (and conversely). A sufficient condition for this to happen is that $P_i^- = P_i^+$ for some (and hence all) $i \in A$.

Remarks. 1. The p.m.s P^-, P^+ are obviously independent of the reference element i_0 of A.

2. The results of section 3.4 (for the binary state space) show that both $P^- = P^+$ and $P^- \neq P^+$ can occur.

3. A consequence of the theorem is that, under its hypotheses, $M_0(\Pi)$ always contains at least one element. In the most important special case of the theorem, that where S is finite, we know this already (from the corollary to Theorem 3.1.1).

Theorem 4.1.2. Suppose that S has a minimal element x^- (or equivalently a maximal element x^+), and that the given h.h.M. specification Π is repulsive. Then there exists a complementary pair of Markov chains P, P' (not necessarily distinct) in $M_1(\Pi)$ such that, as $n \to \infty$,

$$\pi_{V_{2n}}(-,\cdot) \Rightarrow P \quad , \quad \pi_{V_{2n+1}}(-,\cdot) \Rightarrow P' \quad .$$

If $P = P'$ then $|G(\Pi)| = 1$. A sufficient condition for this to happen is that $P_i = P_i'$ for some (and hence all) $i \in A$.

Remarks. 1. The p.m.s P, P' change places if the reference element i_0 is translated an odd number of steps in the tree, but are otherwise independent of i_0.

2. The results of section 3.4 show that both $P = P'$ and $P \neq P'$ can occur.

3. Again, a consequence of the theorem is that, under its hypotheses, $M_1(\Pi)$ always contains at least one element; when $P = P'$ this one element belongs to $M_0(\Pi)$, i.e. is a homogeneous Markov chain. It is of interest to speculate whether, when $P \neq P'$, $M_0(\Pi)$ still contains at least one (perhaps 'unstable') element; (we know, of course, that $M_0(\Pi)$ is non-empty when S is finite).

We now set up the apparatus common to the proofs of both theorems, before considering each separately. The most important item is a partial ordering $\prec$ on the space Ψ: given $p_1, p_2 \in \Psi$ we will say that $p_1 \prec p_2$ (or equivalently $p_2 \succ p_1$) if, for all $x_1, x_2 \in S$ with $x_1 \leqslant x_2$,

(2) $$p_1(x_1)p_2(x_2) \geqslant p_2(x_1)p_1(x_2) \quad ,$$

noting that this relation is independent of the particular member of the equivalence class of functions $S \to R_+$ chosen to represent each of p_1, p_2. The relation $\preccurlyeq$ is trivially reflexive. It is necessary to verify its transitivity: suppose $p_1, p_2, p_3 \in \Psi$ satisfy $p_1 \prec p_2$ and $p_2 \prec p_3$; for any $x_1, x_2 \in S$ such that $x_1 \leqslant x_2$

$$p_1(x_1)p_2(x_2) \geqslant p_2(x_1)p_1(x_2) \quad , \quad p_2(x_1)p_3(x_2) \geqslant p_3(x_1)p_2(x_2) \quad ;$$

it is straightforward that this implies $p_1(x_1)p_3(x_2) \geqslant p_3(x_1)p_1(x_2)$ except where we have

(3) $$p_2(x_1) = p_2(x_2) = 0 \quad , \quad p_3(x_1) \neq 0 \quad , \quad p_1(x_2) \neq 0 \quad ;$$

but (3) implies that $p_2(x) = 0$ for all $x \geqslant x_1$ (since $p_2 \prec p_3$), and that $p_2(x) = 0$ for all $x \leqslant x_2$ (since $p_1 \prec p_2$), so p_2 is identically zero in contradiction to the hypothesis $p_2 \in \Psi$; thus the condition (3) cannot occur and transitivity is established. Finally, we note that if $p_1 \prec p_2$ and $p_2 \prec p_1$ then $p_1 = p_2$.

Now let $\Phi(p,q)$ be a fixed h.h.M. interaction such that $\Pi = \Pi(p,q)$. Define Q, T_d, T'_d as in section 3.1. We will require the following lemma which says (roughly speaking) that when Π is attractive [repulsive] the partial ordering $\prec$ is preserved [reversed] by T_d (and thus in particular by T'_d).

Lemma 4.1.3. Suppose that for each t, $1 \leqslant t \leqslant d$, p_t, $p'_t \in \Psi$ are such that Qp_t, Qp'_t exist and that $p_t \prec p'_t$. Then if Π (and hence $\Phi(p,q)$) is attractive,

$$(4) \qquad T_d(p_1, \ldots, p_d) \prec T_d(p'_1, \ldots, p'_d) \quad ,$$

and if Π is repulsive, (4) holds with $\succ$ replacing $\prec$.

Proof. It is sufficient to show that if p_1, $p'_1 \in \Psi$ are such that Qp_1, Qp'_1 exist and $p_1 \prec p'_1$, then when $\Phi(p,q)$ is attractive $Qp_1 \prec Qp'_1$, and when $\Phi(p,q)$ is repulsive $Qp_1 \succ Qp'_1$. The results then follow easily from consideration of the definition of $\prec$. Thus if p_1, p'_1 are as above and $x_1 \leqslant x_2$,

$$(Qp_1)(x_1)(Qp_1')(x_2) - (Qp_1')(x_1)(Qp_1)(x_2)$$

$$= \sum_{y_1, y_2 \in S\times S} \{[q(x_1,y_1)q(x_2,y_2)-q(x_1,y_2)q(x_2,y_1)]p_1(y_1)p_1'(y_2)\}$$

$$= \sum_{\substack{y_1, y_2 \in S\times S \\ y_1 < y_2}} \{[q(x_1,y_1)q(x_2,y_2)-q(x_1,y_2)q(x_2,y_1)] [p_1(y_1)p_1'(y_2)-p_1'(y_1)p_1(y_2)]\}$$

$$\begin{cases} \geqslant 0 \; , & \text{if } \Phi(p,q) \text{ is attractive} \\ \leqslant 0 \; , & \text{if } \Phi(p,q) \text{ is repulsive.} \end{cases}$$

It follows from the lemma that a similar, but simpler, result holds for T_d'.

It is now convenient to define, for any $V \in \mathcal{V}$, the function $\mu_V : S^{\xi V} \to R_+$ by

$$\mu_V(x_{\xi V}) = \prod_{i \in V} p(x_i) \prod_{\substack{\{j,k\} \in N \\ \{j,k\} \cap V \neq \emptyset}} q(x_j \, , \, x_k)$$

When S has a minimal element x^-, define the sequence $\{\delta_n^-\}$ in Ψ by

$$\delta_0^- = \delta_{x^-} \quad (\text{i.e. } \delta_0^-(x^-) > 0 \; ; \; \delta_0^-(x) = 0 \text{ if } x \neq x^-) \; .$$

$$\delta_n^- = T_d'(\delta_{n-1}^-) \quad , \quad n \geqslant 1.$$

We may argue as in section 3.2 (or apply Theorem 2.2.1 with $A = V_{m+n+1}$) to deduce that for any $m,n \geqslant 0$

(5) $$\pi_{V_{m+n}}(-,X_{V_{m+1}} = x_{V_{m+1}}) \propto \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta_n^-(x_i) \quad ;$$

When S has a maximal element x^+, define similarly the sequence $\{\delta_n^+\}$ in Ψ. Again, for any $m,n \geq 0$

(6) $$\pi_{V_{m+n}}(+,X_{V_{m+1}} = x_{V_{m+1}}) \propto \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta_n^+(x_i) .$$

Note that (5) and (6) include the assertions that the functions δ_n^-, δ_n^+, $n \geq 0$, exist; note also that these functions are strictly positive for at least all $n \geq 2$. We must now consider separately the proofs of Theorems 4.1.1 and 4.1.2.

<u>Proof of Theorem 4.1.1.</u> We break the proof into three sections.

1. We first show how to construct P^-,P^+. For all $p_0 \in \Psi$ we have

$$\delta_0^- \prec p_0 \prec \delta_0^+ \quad ;$$

since Π is attractive, iterative application of Lemma 4.1.3 (with T_d' instead of T_d) to this result gives: for all $n \geq 0$

$$\delta_n^- \prec \delta_{n+1}^-$$

$$\delta_n^+ \succ \delta_{n+1}^+$$

$$\delta_n^- \prec p_n \prec \delta_n^+ \quad , \qquad \text{for all } p_0 \in \Psi,$$

where the sequence $\{p_n\}$ in Ψ is defined by $p_n = T_d'(p_{n-1})$,$n \geq 1$. Further the transitivity of $\prec$ implies that for all $m,n \geq 0$

$$\delta_n^- \prec \delta_m^+ \quad .$$

Thus under the convention

(7) $$p'(x^-) = 1 \qquad \text{for all } p' \in \Psi \text{ such that } p'(x^-) > 0 \quad ,$$

for each $x \in S$ the sequence $\{\delta_n^-(x)\}$ is increasing in n and is bounded above by $\delta_2^+(x)$, say. Hence there exists a strictly positive $\delta^- \in \Psi$ such that, under (7), $\delta_n^-(x) \to \delta^-(x)$, as $n \to \infty$. Further, under (7),

$$\delta_{n+1}^-(x) = \frac{p(x)[(Q\delta_n^-)(x)]^d}{p(x^-)[(Q\delta_n^-(x^-)]^d}$$

and so by dominated convergence $\delta_{n+1}^-(x) \to [T_d'(\delta^-)](x)$ as $n \to \infty$. We deduce that $\delta^- = T_d'(\delta^-)$. Further, under (7),

$$\sum_{x \in S} \sum_{y \in S} \delta^-(x) q(x,y) \delta^-(y) \leqslant \sum_{x \in S} \sum_{y \in S} \delta_2^+(x) q(x,y) \delta_2^+(y)$$

$$< \infty$$

(consider the density of $\pi_V(+,\cdot)$ where $V = \xi(\xi\{i,j\})$ for any $\{i,j\} \in N$). Thus, by Theorem 3.1.1, there exists a Markov chain $P^- \in M_0(\Pi)$ corresponding to δ^- as in that theorem.

Now, still under the convention (7), let $a_{m,n}^-$, $a_{m,n}^+$ be the normalising constants required in (5) and (6). We also have: for any fixed $m \geqslant 0$,

$$P^-(X_{V_{m+1}} = x_{V_{m+1}}) = a_m^- \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta^-(x_i) \;;$$

it follows from our results above that

$$\text{(8)} \quad \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta_n^-(x_i) \to \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta^-(x_i) \;;$$

the left-hand side of (8) is dominated by $\mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} \delta_2^+(x_i)$, whose sum over all $x_{V_{m+1}} \in S^{V_{m+1}}$ is $(a_{m,2}^+)^{-1}$; thus, again by dominated convergence, $a_{m,n}^- \to a_m^-$. Putting these results together, and recalling that m is arbitrary, we deduce that as $n \to \infty$,

$$\pi_V(-,\cdot) \Rightarrow P^- \quad .$$

Similarly there exists a strictly positive $\delta^+ \in \Psi$ such that the sequence $\{\delta_n^+\}$ converges pointwise to δ^+ (under the convention (7) with , or without , + replacing -); that δ^+ defines, via Theorem 3.1.1, a Markov chain $P^+ \in M_0(\Pi)$; and that as $n \to \infty$, $\pi_{V_m}(+,\cdot) \Rightarrow P^+$. Note also that if $p' \in \Psi$ satisfies $p' \succ \delta_n^-$ for all n, then $p' \succ \delta^-$; similarly if $p' \prec \delta_n^+$ for all n, then $p' \prec \delta^+$. We use this result in 2 below; (an immediate consequence is that $\delta^- \prec \delta^+$).

2. We now show that if $P^- = P^+$, then $|G(\Pi)| = 1$. Since $E(\Pi) \subset M(\Pi)$, it is sufficient to show that if $P^- = P^+$, then $|M(\Pi)| = 1$. Therefore consider $P \in M(\Pi)$; let $\{p_i^j , i \in A, j \in \partial i\}$ be the (unique) collection of functions in Ψ defined by Theorem 2.2.1 (relative to the interaction $\Phi(p,q)$). For all $n \geqslant 0$ we have the result:

$$(9) \qquad \delta_n^- \prec p_i^j \prec \delta_n^+ \quad , \quad \text{for all } i \in A,\ j \in \partial i.$$

(This follows by induction, noting first that it is trivially true for n = 0: by Theorem 2.2.1, $p_i^j = T_d(p_k^i : k \in \partial i - j)$ and so if (9) is true for a given n, by Lemma 4.1.3 it is also true for n + 1.) Thus,

$$\delta^- \prec p_i^j \prec \delta^+ \quad , \text{ for all } i \in A,\ j \in \partial i \quad .$$

Therefore, if $P^- = P^+$ then $\delta^- = \delta^+$ and so $p_i^j = \delta^-$ for all $i \in A$, $j \in \partial i$, implying (again by Theorem 2.2.1) that $P = P^-$.

3. Finally, we show that if $P_i^- = P_i^+$ for some $i \in A$, then $P^- = P^+$. From Theorem 3.1.1,

$$P^-(X_i = x) \propto [T'_{d+1}(\delta^-)](x)$$

$$\propto \frac{\delta^-(x)^{1+1/d}}{p(x)^{1/d}},$$

with a similar result for P^+. Thus $P_i^- = P_i^+$ implies that $\delta^- = \delta^+$, and so $P^- = P^+$.

<u>Remark.</u> In the above proof that if $P^- = P^+$ then $|G(\Pi)| = 1$, we actually show $|E(\Pi)| = 1$; that this implies the required result comes from (the quite deep) Theorem 1.1.1 (*Preston (1976)*), which, although conveniently available, is more than we really need here. We may, if we wish, proceed instead as follows. Fix $m \geq 0$; arguing as in the above proof, for any fixed $y_A \in S^A$, and for all $n \geq 0$,

$$\pi_{V_{m+n}}(y_A, X_{V_{m+1}} = x_{V_{m+1}}) \propto \mu_{V_m}(x_{V_{m+1}}) \prod_{i \in \partial V_m} p_n(x_i),$$

where, for each n, $p_n \in \Psi$ depends on y_A and satisfies

$$\delta_n^- \prec p_n \prec \delta_n^+ \quad .$$

Now if $P^- = P^+$, so that $\delta^- = \delta^+$, and if $P \in G(\Pi)$, we have from (4) of section 1.1 and the (reversed) martingale convergence theorem,

$$P(X_{V_{m+1}} = x_{V_{m+1}} / \hat{F}) = \lim_{n \to \infty} \pi_{V_{m+n}}(\cdot, X_{V_{m+1}} = x_{V_{m+1}}) \quad P\text{-a.s.},$$

$$= P^-(X_{V_{m+1}} = x_{V_{m+1}})$$

(recalling that $\hat{F}$ is the tail σ-field). Taking expectations with respect to P, and noting as usual that m is arbitrary, we obtain $P = P^-$.

Proof of Theorem 4.1.2. This is a variation on the proof of Theorem 4.1.1; we therefore concentrate on the differences. For any $p_0 \in \Psi$, $\delta_0^- \prec p_0$. Since Π is repulsive iterative application of Lemma 4.1.3 now gives that for all $n \geq 0$,

$$\delta_{2n}^- \prec \delta_{2n+2}^- \quad , \quad \delta_{2n+1}^- \succ \delta_{2n+3}^-$$

$$\delta_{2n}^- \prec p_{2n+1} \prec \delta_{2n+1}^- \quad , \quad \delta_{2n+1}^- \succ p_{2n+2} \succ \delta_{2n+2}^- \quad , \quad \text{for all } p_0 \in \Psi,$$

where the sequence $\{p_n\}$ in Ψ is defined by $p_n = T_d'(p_{n-1})$, $n \geq 1$. Further for all $m,n \geq 0$, we have $\delta_{2n}^- \prec \delta_{2m+1}^-$. Thus, the sequences $\{\delta_{2n}^-\}$ and $\{\delta_{2n+1}^-\}$ play essentially the same roles as the sequences $\{\delta_n^-\}$ and $\{\delta_n^+\}$ in the proof of Theorem 4.1.1. Arguing as in that proof, we deduce that there exist functions $\delta^e, \delta^o \in \Psi$ (with $\delta^e \prec \delta^o$) such that (under the convention (7)) the sequences $\{\delta_{2n}^-\}$, $\{\delta_{2n+1}^-\}$ converge pointwise to δ^e, δ^o respectively; that

$$\delta^o = T_d'(\delta^e) \quad , \quad \delta^e = T_d'(\delta^o) \quad ;$$

and that $\sum_{x \in S} \sum_{y \in S} \delta^e(x) q(x,y) \delta^o(y) < \infty$. Now let E be the set of 'even', and O the set of 'odd' vertices (or sites) of the tree A, taking O to include the reference element i_0. Define $P \in M_1(\Pi)$ via Theorem 3.1.2 by associating δ^e, δ^o with the vertices of E,O respectively (i.e. by letting the associated functions $p_i^j \in \Psi$ be equal to δ^e or δ^o according as i belongs to E or O). Define $P' \in M_1(\Pi)$ to be the complementary Markov chain, given by interchanging the roles of δ^e and δ^o. Thus, continuing to argue as in the proof of Theorem 4.1.1, we obtain that as $n \to \infty$,

$$\pi_{V_{2n}}(-,\cdot) \Rightarrow P \quad , \quad \pi_{V_{2n+1}}(-,\cdot) \Rightarrow P' \quad .$$

Now suppose P = P' (which is equivalent to $\delta^e = \delta^o$).
For any Markov chain P* in $M(\Pi)$ let $\{p_i^j , i \in A, j \in \partial i\}$ be the corresponding collection of functions in Ψ defined by Theorem 2.2.1; we may show by induction that for all $n \geqslant 0$,

$$\delta^-_{2n} \prec p_i^j \quad , \quad p_i^j \prec \delta^-_{2n+1} \quad , \quad \text{for all } i \in A \, , \, j \in \partial i \; ;$$

thus $p_i^j = \delta^e$ for all $i \in A$, $j \in \partial i$, and so P* = P. We deduce that $|M(\Pi)| = 1$ and so $|G(\Pi)| = 1$. (We may alternatively show this by mimicking the argument given in the remark which follows the proof of Theorem 4.1.1.)

Finally, if for any $i \in A$, $P_i = P_i'$, a variation of the argument for the corresponding result in Theorem 4.1.1 shows that P = P'.

Under the conditions of Theorem 4.1.1 we deduce that $|G(\Pi)| = 1$ if (and only if) $|M_0(\Pi)| = 1$; similarly under the conditions of either Theorem 4.1.1 or Theorem 4.1.2, $|G(\Pi)| = 1$ if and only if $|M_1(\Pi)| = 1$. Consider again the case where the state space S is binary; we have already remarked that here every h.h.M. specification is either attractive or repulsive (independently of the two possible orderings of S); we thus obtain Spitzer's result, mentioned in section 3.4, that $|M_1(\Pi)| = 1$ is always a sufficient condition for $|G(\Pi)| = 1$.

The following corollary to the above theorems is of some interest.

<u>Corollary.</u> For d = 1 (the one-dimensional integer lattice), under the conditions of either Theorem 4.1.1 or Theorem 4.1.2 and provided that Π is strictly positive, $G(\Pi)$ consists of a single element, necessarily a homogenenous Markov chain.

<u>Proof.</u> *Kesten (1976)* shows that for d = 1 and a given strictly positive translation-invariant Markov specification Π, $G_0(\Pi)$ consists of at most one element, necessarily a stationary Markov chain. Suppose Π satisfies the conditions of Theorem 4.1.1; we must then have equality of the Markov chains P^-, P^+ defined there (since they belong to $G_0(\Pi)$), and so the result follows by that theorem. Suppose instead Π satisfies the conditions of Theorem 4.1.2; if P,P' are as defined there, the p.m. $\frac{1}{2}(P + P')$ belongs to $G_0(\Pi)$ and hence, by Kesten's result, to $M_0(\Pi)$; thus, by the corollary to Theorem 3.1.3, P = P' and again the result follows.

<u>Remark.</u> The strict positivity condition of Kesten's result can presumably be relaxed to the hereditary condition used in this work; if this is so, then the proviso in the above corollary is redundant.

Now suppose again that the conditions of Theorem 4.1.1 are satisfied. Considering the method of construction of the p.m.s P^-, P^+ defined in that theorem, we would most certainly expect them to belong to the set $E(\Pi)$ of extreme points of $G(\Pi)$. This result would follow trivially if our conjecture of section 3.1 (that $M_0(\Pi) = E_0(\Pi)$) were true. We offer instead the proof below, based on the integral representation of Theorem 1.1.1. It does at least have the virtue that it requires only slight and obvious modifications to establish the corresponding result for repulsive specifications.

<u>Theorem 4.1.4.</u> (i) Under the conditions of Theorem 4.1.1, $P^-, P^+ \in E(\Pi)$.

(ii) Under the conditions of Theorem 4.1.2, $P, P' \in E(\Pi)$.

Proof of (i). Consider the stochastic kernel π defined by Theorem 1.1.1. For each $y_A \in S^A$, $\pi(y_A, \cdot)$ belongs to $\mathcal{E}(\Pi)$ and hence (Theorem 2.1.4) to $\mathcal{M}(\Pi)$; for each $i \in A$, $j \in \partial i$, let $p_i^j(y_A, \cdot) \in \Psi$ be the corresponding function defined by Theorem 2.2.1; as in part 2 of the proof of Theorem 4.1.1 we have

$$(10) \qquad \delta^- \prec p_i^j(y_A, \cdot) \prec \delta^+ \quad ;$$

and for each such pair of neighbours i,j in A,

$$(11) \qquad \pi(y_A ; X_i = x_i, X_j = x_j) \propto p_i^j(y_A, x_i) q(x_i, x_j) p_j^i(y_A, x_j) \quad .$$

Consider P^- : for each i,j as above,

$$(12) \qquad P^-(X_i = x_i, X_j = x_j) \propto \delta^-(x_i) q(x_i, x_j) \delta^-(x_j) \quad ;$$

it follows from (10), (11), (12) and the result $P^- = P^- \pi$ of Theorem 1.1.1 that

$$P^-(p_i^j(y_A, \cdot) = \delta^-) = 1 \quad ;$$

(this is easiest to see if we make use of the convention (7) to fix elements of Ψ). Thus,

$$P^-(p_i^j(y_A, \cdot) = \delta^- \text{ for all } i \in A,\ j \in \partial i) = 1 \quad ,$$

which (as in the proof of Theorem 4.1.1) is equivalent to

$$P^-(\pi(y_A, \cdot) = P^-) = 1 \quad .$$

We may now complete the proof in various ways, of which the simplest is to observe that this last result implies that $P^- = \pi(y_A, \cdot)$ for *some* $y_A \in S^A$, so that, by Theorem 1.1.1, $P^- \in \mathcal{E}(\Pi)$. Similarly, $P^+ \in \mathcal{E}(\Pi)$.

Examples of attractive and repulsive h.h.M. specifications are afforded by various of the *auto-models* of *Besag (1974)*. (Investigation of some of these was the starting point for the current work.) The specifications considered by Spitzer for $|S| = 2$ (and discussed further in section 3.4) are in fact examples of what Besag calls the *auto-logistic model*; on the d-dimensional integer lattice this corresponds to the Ising model of classical statistical mechanics. In the next section we consider as a further example the *auto-Poisson model*.

We conclude this section by remarking, *tentatively*, that both its results (with Markov chains replaced by M.r.f.s of a more general nature) and the methods used to obtain them, should not be too difficult to modify to cover the study of attractive and repulsive h.h.M. specifications (similarly defined) on the d-dimensional integer lattice Z^d; (for attractive specifications the *results* would be similar to those of *Preston (1976, Chapter 9)*). For each member V_n of a suitable increasing sequence of subsets of A it would be necessary to consider simultaneously all the sites comprising its boundary ∂V_n ; thus we would have to replace the space Ψ by spaces of positive real-valued functions (modulo a multiplicative constant) on finite products of the state space S, and either correspondingly modify the definition of the partial ordering $\prec$, or else replace it by a suitable metric on these spaces or on the space of p.m.s on (S^A, F). We have not attempted to check the details of all this.

4.2 The auto-Poisson specification

We consider a further example of a repulsive specification. Let the state space S be the set of non-negative integers, with the usual total ordering; take the minimal element 0 of S to define the basepoint in the usual way, and consider the h.h.M. interaction $\Phi(p,q)$ on (S,A) (where A remains our regular infinite tree) defined by

$$p(x) = \frac{\alpha^x}{x!} \qquad , \qquad 0 < \alpha < \infty$$

$$q(x,y) = \beta^{xy} \qquad , \qquad 0 \leqslant \beta \leqslant 1 \ .$$

The condition $0 \leqslant \beta \leqslant 1$ is necessary in order that $\Phi(p,q)$ should *be* an interaction, i.e. in order that the 'normalisability' condition (1)(ii) of section 1.2 should be satisfied; (see *Besag (1974), section 4.2.4*); the interaction $\Phi(p,q)$ is then easily seen to be repulsive.

Let $\Pi = \Pi(p,q)$ be the corresponding (repulsive) specification. For each $i \in A$, $x_{\xi i} \in S^{\xi i}$

$$(1) \qquad \tilde{\pi}_i(x_{\partial i}, x_i) = \exp(-\alpha\beta^{t_i}) \frac{(\alpha\beta^{t_i})^{x_i}}{x_i!}$$

where $t_i = \sum_{j \in \partial i} x_j$; the specification provides an example of an auto-model and, in view of (1), is called the *auto-Poisson* specification (*Besag (1974)*). (Naturally it may be similarly defined with respect to any other neighbour relation ~ on A.)

We may apply Theorem 4.1.2 to infer the existence of a complementary pair of (not necessarily distinct) Markov chains in $M_1(\Pi)$; if they are coincident they form a single homogeneous Markov chain and $|G(\Pi)| = 1$. Thus in particular $M(\Pi)$, and hence $G(\Pi)$, are non-empty.

By Theorem 3.1.2 the elements of $M_1(\Pi)$ correspond to the ordered pairs of functions p^e, p^o in Ψ satisfying

$$(2)\quad \begin{cases} \text{(i)}\ p^o(x) \propto \dfrac{\alpha^x}{x!}\left[\displaystyle\sum_{y \in S} \beta^{xy} p^e(y)\right]^d , \\[2ex] \qquad p^e(x) \propto \dfrac{\alpha^x}{x!}\left[\displaystyle\sum_{y \in S} \beta^{xy} p^o(y)\right]^d , \\[2ex] \text{(ii)}\ \displaystyle\sum_{x \in S}\ \sum_{y \in S} p^o(x)\beta^{xy}p^e(y) < \infty . \end{cases}$$

Full solutions of the equations (2) seems analytically quite intractable; although we conjecture that it might not be too difficult to obtain bounds on the set of (α,β) such that (2) had exactly one solution. However, for the important case $d = 1$, the corollary to the theorems of the previous section enables us to deduce directly that $G(\Pi)$ itself contains precisely one element, this being a homogeneous Markov chain.

References

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B, 36, 192-236.

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

Cox, J.T. (1977). An example of phase transition in countable one-dimensional Markov random fields. *J. Appl. Prob.*, 14, 205-211.

Cox, J.T. (1979). An alternative proof of a theorem of Kesten concerning Markov random fields. *Ann. Prob.*, 7, 377-378.

Dobrushin, R.L. (1968). Description of a random field by means of conditional probabilities and the conditions governing its regularity. *Theory of Prob. and its Appl.*, 13, 197-224.

Dynkin, E.B. (1978). Sufficient statistics and extreme points. *Ann. Prob.*, 6, 705-730.

Föllmer, H. (1975a). Phase transition and Martin boundary. *Sém. Prob. Strasbourg IX, Lecture Notes in Math. 465.* Springer.

Föllmer, H. (1975b). On the potential theory of stochastic fields. *Bull. Int. Statist. Inst.*, 46, 362-370.

Goldstein, J. (1978). A note on specifications. *Z. Wahrs. verw. Geb.*, 46, 45-51.

Grimmett, G.R. (1973). A theorem about random fields. *Bull. London. Math. Soc.*, 5, 81-84.

Kesten, H. (1976). Existence and uniqueness of countable one-dimensional Markov random fields. *Ann. Prob.*, 4, 557-569.

Preston, C.J. (1974). *Gibbs States on Countable Sets.* Cambridge University Press.

Preston, C.J. (1976). *Random Fields, Lecture Notes in Math. 534.* Springer.

Reed, M. and Simon, B. (1972). *Methods of Modern Mathematical Physics I : Functional Analysis.* Academic Press.

Spitzer, F. (1971). *Random Fields and Interacting Particle Systems.* Mathematical Association of America.

Spitzer, F. (1973). Introduction aux processus de Markov à paramètre dans Z_ν. *Lecture Notes in Math. 390.* Springer.

Spitzer, F. (1975a). Phase transition in one-dimensional nearest-neighbour systems. *J. Functional Analysis*, 20, 240-255.

Spitzer, F. (1975b). Markov random fields on an infinite tree. *Ann. Prob.*, 3, 387-398.