T. P. Speed

A Note on Nearest-Neighbour Gibbs and Markov Probabilities

Preprint
November
1978

12

Institute of Mathematical Statistics
University of Copenhagen

T.P. Speed*

A NOTE ON NEAREST-NEIGHBOUR GIBBS AND MARKOV PROBABILITIES

Preprint 1978 No. 12

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1978

* University of Western Australia

A NOTE ON NEAREST-NEIGHBOUR GIBBS AND MARKOV PROBABILITIES

by

T.P. Speed

INTRODUCTION

The basic result which asserts the coincidence, under a positivity assumption, of probabilities possessing a spatial Markov property with Gibbs states having the corresponding nearest-neighbour property is now very well known and has been proved many times, see references [2,3,6,7,10,16,19,20,21,23,25,26,27 and 28].

In this note we also give a proof of the result, and although it is not essentially different from those already published, we do give more emphasis to the underlying linear algebra. Our aim is to give the essential details of the proof, and then to comment upon various generalisations of the main result or aspects of the proof. It is not hard to pick out the proof itself: simply take the first main result of every section (together with the preceding discussion) and string these all together (i.e. 1.1, 2.1, 3.1 and 4.1).

Sincere thanks are offered to the many people who have discussed these matters with me, and particularly to Steffen Lauritzen.

1. FACTORISATIONS AND SUBSPACES

Let us take a system $\{I_\gamma : \gamma\varepsilon C\}$ of finite sets indexed by another finite set C. For any subset $a \subseteq C$ write $I_a = \Pi\{I_\gamma : \gamma\varepsilon a\}$, $I = I_C$, and denote the surjection $(i_\gamma : \gamma\varepsilon C) \mapsto (i_\gamma : \gamma\varepsilon a)$ from I onto I_a by π_a.

We will be discussing the Euclidean vector space $E = \mathbb{R}^I$ with inner product $\langle p,q\rangle = \sum_{\underline{i}\in I} p(\underline{i})q(\underline{i})$, $(p,q\in E)$, and certain of its subspaces. For $a\subseteq C$ the surjection $\pi_a: I\to I_a$ induces an injection $\pi_a^*: \mathbb{R}^{I_a}\to\mathbb{R}^I$ by writing $(\pi_a^* p)(\underline{i}) = p(\pi_a(\underline{i}))$, $(p\in\mathbb{R}^{I_a}, \underline{i}\in I)$ and the image $\pi_a^*(\mathbb{R}^{I_a})$ is a linear subspace of E which we will denote by E_a. The <u>a-marginal</u> $p_a\in E_a$ of an element $p\in E$ is defined on $\underline{i}\in I$ by

$$p_a(\underline{i}) = \sum_{\underline{j}_{a'}\in I_{a'}} p(\underline{i}_a\underline{j}_{a'}), \qquad \underline{i}_a = \pi_a(\underline{i}), \qquad \underline{j}_{a'} = \pi_{a'}(\underline{j}),\ a' = C\backslash a.$$

(We use the notation $a\cup b$, ab, $a\backslash b$, $a\Delta b$ for the set union, intersection, difference, and symmetric difference respectively of a and b.) The first lemma is well known but for completeness a proof will be given.

<u>1.1 Lemma</u> Let a and b be subsets of C with $a\cup b = C$, and suppose that $p\in E$ is strictly positive, i.e. $p(\underline{i})>0$, $\underline{i}\in I$. Then the following are equivalent:

(i) $\log p \in E_a+E_b$ (vector sum of subspaces);

(ii) $p = p_a p_b/p_{ab}$.

<u>Proof.</u> Suppose that $p = qr$ where $q\in E_a$ and $r\in E_b$, and further suppose that $q = \pi_a^*\tilde{q}$, $r = \pi_b^*\tilde{r}$ for $\tilde{q}\in R^{I_a}$, $\tilde{r}\in R^{I_b}$. Then

$$p_a p_b = (qr_a)(q_b r) = p(q_b r_a),$$

and the implication (i) $\Rightarrow$(ii) will be proved if we can show that $q_b r_a = (qr)_{ab}$. But this is an easy computation, for if $\underline{i}\in I$ we have

$$q_b(\underline{i})r_a(\underline{i}) = \sum_{\underline{j}_{a\backslash b}} q(\underline{i}_b\underline{j}_{a\backslash b}) \sum_{\underline{k}_{b\backslash a}} r(\underline{i}_a\underline{k}_{b\backslash a})$$

$$= \sum_{\underline{j}_{a\backslash b}} \sum_{\underline{k}_{b\backslash a}} \tilde{q}(\underline{i}_{ab}\underline{j}_{a\backslash b})\tilde{r}(\underline{i}_{ab}\underline{k}_{b\backslash a})$$

$$= \sum_{\underline{h}_{a\Delta b}} q(\underline{i}_{ab}\underline{h}_{a\Delta b})r(\underline{i}_{ab}\underline{h}_{a\Delta b})$$

$$= (qr)_{ab}(\underline{i}).$$

The converse is seen by simply taking logarithms of both sides of (ii). $\blacksquare$

1.2 Generalisations. Let us first consider a class $\{a,b,c\}$ consisting of three sets of union $a \cup b \cup c = C$. A computation similar to that just given shows that if

(i)′ $\log p \in E_a + E_b + E_c$,

then

(ii)′ $p = p_a p_b p_c p_{ab \cup bc \cup ca} / p_{a(b \cup c)} p_{b(c \cup a)} p_{c(a \cup b)}$.

The converse will fail in general, because the identity (ii)′ may be vacuous (simply take $a = \{1,2\}$, $b = \{2,3\}$, $c = \{1,3\}$, and observe that $ab \cup bc \cup ca = a \cup b \cup c = \{1,2,3\}$). However if certain inclusion relationships between ab, bc and ca hold, we can get a non-trivial result. For example, if $ac \subseteq bc$, we find that (i)′ holds if and only if $p = p_a p_b p_c / p_{ab} p_{bc}$.

Turning now to classes of four subsets $\{a,b,c,d\}$ with $a \cup b \cup c \cup d = C$, we are able to prove that

(i)″ $\log p \in E_a + E_b + E_c + E_d$,

implies

(ii)″ $p = p_a p_b p_c p_d p_{ab \cup ac \cup ad \cup bc \cup bd \cup cd} / p_{a(b \cup c \cup d)} p_{b(a \cup b \cup d)} p_{c(a \cup b \cup d)} p_{d(a \cup b \cup c)}$,

but again the problems raised by the union of the pairwise intersections being "too large" arise. As before we can make further progress by imposing conditions on these intersections, and many possibilities arise with four subsets. To illustrate, we remark that if $ac \subseteq bc$, $ad \subseteq cd$, and $bd \subseteq cd$, then (i)″ holds if and only if $p = p_a p_b p_c p_d / p_{ab} p_{bc} p_{cd}$, whilst if $ac \subseteq ab$, $ad \subseteq ab$, and $cd \subseteq ab$, (i)″ holds if and only if $p = p_a p_b p_c p_d / p_{ab}^3$.

A general discussion of this type of result involves the combinatorial structure of classes $\mathcal{C}$ of pairwise incomparable subsets of C whose union is C, see [14] and references therein, and Haberman [8, Chapter 5]. We will also mention this topic in 4.4 below.

1.3 Remark. An examination of the proof shows that 1.1 continues to hold even if the sets $I_\gamma, \gamma\varepsilon C$ are countably infinite (all terms are positive), but we certainly need to assume that C is finite to make sense of the operation of forming marginals.

2. INTERACTION SUBSPACES

Our next lemma involves certain subspaces of E which we term interaction subspaces because of their origins in the theory of factorial experiments. For a subset $b \subseteq C$ we write

$$F_b := E_b \cap \bigcap_{d \subsetneqq b} E_d^{\perp}.$$

The lemma is well known when C denotes a finite set of factors and I_γ the finite set of levels of the factor $\gamma\varepsilon C$, see Asmussen [1], Davidson [5], and Haberman [9]. A related result can be found in Kellerer [11, §1], and we refer to this below.

2.1 Lemma $E_c = \bigoplus_{b \subseteq c} F_b$ (direct sum of subspaces), $c \subseteq C$.

Proof We begin by showing that if a and b are distinct subsets of C, then $F_a \perp F_b$. This is most easily proved by making use of the orthogonal projections onto E_a and E_b, denoted by P_a and P_b respectively. The projection P_a is readily seen to be given by

$$P_a r = \frac{1}{|I_{a'}|} r_a, \qquad a' = C\backslash a, \; r\varepsilon E,$$

where $|\cdot|$ denotes cardinality, and the identity $P_a P_b = P_b P_a = P_{ab}$ is quickly checked. Under these circumstances, if $ab \subsetneqq a$ say, then $q\varepsilon F_a \subseteq E_a \cap E_{ab}^{\perp}$ and for any $r\varepsilon F_b \subseteq E_p$ we have

$$\langle q,r\rangle = \langle P_a q, P_b r\rangle = \langle P_b P_a q, r\rangle = \langle P_{ab} q, r\rangle = 0.$$

Now we prove the lemma by induction on $|c|$. It is clearly true for $|c| = 0$, for in this case $E_\phi = F_\phi = R$, and so we may suppose the lemma true for all proper subsets a of c where $|c| \geq 1$. This inductive hypothesis implies that $E_a \subseteq \sum_{b \subsetneqq c} F_b$ whenever $a \subsetneqq c$, and so we may sum over all such a to obtain

$$\sum_{a \subsetneqq c} E_a \subseteq \sum_{b \subsetneqq c} F_b .$$

Using this relationship and general facts concerning the calculus of subspaces of Euclidean spaces, we have

$$\begin{aligned} E_c &= \sum_{a \subsetneq c} E_a + E_c \cap \left(\sum_{a \subsetneq c} E_a \right) \\ &= \sum_{a \subsetneq c} E_a + F_c \\ &\subseteq \sum_{b \subsetneq c} F_b + F_c \\ &= \sum_{b \subseteq c} F_b \end{aligned}$$

proving the inductive step and thus the lemma. $\square$

2.2 Alternative proofs of 2.1 The foregoing proof was suggested to the author by Michael Meyer, and there are a number of other ways of obtaining the result. One makes use of certain simple facts concerning tensor products of finite-dimensional vector spaces (see, for example, Maclane and Birkhoff [15]), and runs as follows:

$$E_c \simeq \mathbb{R}^{I_c} \simeq \bigotimes_{\gamma \in c} \mathbb{R}^{I_\gamma} = \bigotimes_{\gamma \in c} (\mathbb{R} \oplus F_\gamma) \simeq \bigoplus_{b \subseteq c} \left(\bigotimes_{\gamma \in b} F_\gamma \right) \simeq \bigoplus_{b \subseteq c} F_b,$$

where $F_\gamma = \mathbb{R}^{I_\gamma} \ominus \mathbb{R}$, and $\simeq$ denotes (canonical) isomorphism of vector spaces.

A second method involves an explicit expression for the orthogonal projection Q_b of E onto F_b, namely

$$Q_b = \sum_{a \subseteq b} (-)^{|b \setminus a|} P_a.$$

This is easy to derive from the definition of F_b (since all the P_a commute, $a \subseteq C$), and one goes on to show that $P_c = \sum_{b \subseteq c} Q_b$, with $Q_a Q_b = 0$ if $a \neq b$.

Finally we mention Davidson's proof [5], which involves forming explicit bases for the subspaces F_b. This is perhaps the most elementary proof.

2.3 Corollary $E = \bigoplus_{b \subseteq c} F_b$.

<u>2.4 Corollary</u> $E_a + E_b = \oplus\{F_d : d \subseteq a \text{ or } d \subseteq b\}$, $a,b \subseteq C$.

<u>2.5 Variant on 1.1</u> A reformulation of 1.1 making use of 2.4 above might read (under the same assumptions): the following are equivalent (i) $\log p \in E_a + E_b$; (ii) $p = p_a p_b / p_{ab}$; (iii) $\log p \perp F_d$ for all $d \subseteq C$ with $d \cap (a \setminus b) \neq \emptyset$ and $d \cap (b \setminus a) \neq \emptyset$. The intuitive notion which (iii) makes precise is the idea that "a\b and b\a do not interact", equivalently, that "a and b interact only through ab". In a probabilistic setting (§4 below) this gives an equivalence between a precise condition of "no interaction" and the familiar (conditional) independence notion.

<u>2.6 Generalisations of 2.1</u> If we drop the assumption that each of the sets I_γ is finite and suppose instead that for each $\gamma \in C$ we have a measure space $(I_\gamma, \mathcal{I}_\gamma, m_\gamma)$, then there are at least two generalisations of 2.1 (which reduce to it when I_γ is finite, $\mathcal{I}_\gamma$ consists of all subsets of I_γ and m_γ is the uniform probability measure over $I_\gamma, \gamma \in C$). Let us write $m = \bigotimes_{\gamma \in C} m_\gamma$ for the product of the measures m_γ (when it exists), $\mathcal{I} = \bigotimes_{\gamma \in C} \mathcal{I}_\gamma$ for the product σ-algebra, and $I = \prod_{\gamma \in C} I_\gamma$ as usual.

In the first generalisation we suppose all the measures m_γ to be probability measures, and take for the space E the real Hilbert space $L^2(m)$ consisting of all (equivalence classes of) <u>m-square-integrable</u> $\mathcal{I}$-measurable real-valued functions defined on I. The definition of tensor product of Hilbert spaces and Fubini's Theorem gives the following equivalence

$$L^2\left(\bigotimes_{\gamma \in C} m_\gamma\right) \simeq \bigotimes_{\gamma \in C} L^2(m_\gamma).$$

Thus we may define $F_\gamma = \{f \in L^2(m_\gamma) : \int f\,dm_\gamma = 0\}$, and so $L^2(m_\gamma) \simeq R \oplus F_\gamma$, and argue as in 2.2 above to find as there that

$$E \simeq \bigoplus_{b \subseteq C} F_b, \qquad \text{with } F_b := \bigotimes_{\gamma \in b} F_\gamma,\ b \subseteq C.$$

This direct sum decomposition underlies some of the discussion of interaction in Lancaster [13], who (in our notation) expands the density dM/dm of a probability measure M on $(I,\mathcal{I})$ having marginals m_γ and being absolutely continuous with respect to their product m, at least in the (ϕ^2-bounded) case in which this density is m-square-integrable. Further discussion of this and its relation to other ideas in this note can be found in Darroch and Speed [4].

The second generalisation of 2.1 is given in Kellerer [11, §1] where, under certain assumptions regarding the measures m_γ, the space E (all in our notation) is taken to be the bounded $\mathcal{I}$-measurable real-valued functions defined on I. We refer to that paper for further details.

3. SUBSPACES DEFINED BY GRAPHS

We turn now to an examination of the preceding notions when C is the vertex set of a graph. More precisely, we consider a simple graph (i.e. an undirected graph having no loops or multiple edges) $\underset{\sim}{C} = (V(\underset{\sim}{C}), E(\underset{\sim}{C}))$ having vertices $V(\underset{\sim}{C}) = C$ and edges (unordered pairs of vertices) $E(\underset{\sim}{C})$. If $\{\alpha,\beta\}\varepsilon E(\underset{\sim}{C})$ we say that α and β are adjacent or neighbours, and write $\partial_{\underset{\sim}{C}}\gamma$ (or just $\partial\gamma$ when no confusion can result) for the set $\{\delta\varepsilon C: \{\gamma,\delta\}\varepsilon E(\underset{\sim}{C})\}$ of all neighbours of $\gamma\varepsilon C$. Similarly we write $\partial_{\underset{\sim}{C}} a = \partial a = \bigcup_{\alpha\varepsilon a}\partial\alpha$, and also put $a^- = a\cup\partial a$. Two disjoint subsets a and b of C are said to be separated by the third subset $d\subseteq C$ if every chain $\alpha = \gamma_o,\gamma_1,\ldots,\gamma_n = \beta$ $(\{\gamma_{m-1},\gamma_m\}\varepsilon E(\underset{\sim}{C}), 1\leq m\leq n)$ connecting some pair $\alpha\varepsilon a$ and $\beta\varepsilon b$ necessarily intersects d. A subset $b\subseteq C$ is called complete if every pair of distinct elements of b is an adjacent pair, and complete subsets which are maximal (under set inclusion) are known (by most authors) as cliques. The class of all cliques of the graph $\underset{\sim}{C}$ is denoted by $\mathcal{C}_{\underset{\sim}{C}}$ or just $\mathcal{C}$.

The following lemma is the key to the proof that nearest-neighbour Gibbs states and Markov states coincide, a result which we shall describe as the NNG = M proposition. Although the subspaces F_b are not mentioned in the statement of the lemma, vital use is made of them in the proof; see also §5 below. Recall that $\gamma- = \{\gamma\}\cup\partial\gamma$, and write $\gamma' = \{\gamma\}' = C\setminus\{\gamma\}$.

3.1 Lemma $$\sum_{c\in C} E_c = \bigcap_{\gamma\in C}(E_{\gamma-}+E_{\gamma'}).$$

Proof We begin with the easy observation that a subset $b\subseteq C$ is complete if and only if for all $\gamma\varepsilon C$ either $\gamma\notin b$ or $\gamma-\supseteq b$. Then by making repeated use of Lemma 2 we obtain

$$\begin{aligned}\sum_{c\in C} E_c &= \oplus\{F_b:\ b\subseteq c \text{ for some } c\in C\}\\ &= \oplus\{F_b:\ b \text{ complete}\}\\ &= \oplus\{F_b:\ \text{for all } \gamma\in C, \gamma\notin b \text{ or } \gamma-\supseteq b\}\\ &= \bigcap_{\gamma\in C}\oplus\{F_b:\ \gamma\notin b \text{ or } \gamma-\supseteq b\}\\ &= \bigcap_{\gamma\in C}(E_{\gamma'}+E_{\gamma-}).\quad \blacksquare\end{aligned}$$

3.2 Variants There are a number of variants on 3.1 above useful later when we consider different Markov properties. Two are the following:

$$\sum_{c\in C} E_c = \bigcap_{a\subseteq C}(E_{a-}+E_{a'})$$

and

$$\sum_{c\in C} E_c = \bigcap_{e,f}(E_e+E_f:\ e\backslash f,\ f\backslash e \text{ are separated by } ef,\ e\cup f = C).$$

Both are proved in exactly the same way as 3.1, and we leave this to the reader.

3.3 Orthogonal projections In certain problems, particularly those associated with multidimensional contingency tables [8] , explicit expressions for the orthogonal projections P_C onto the subspaces $E_C = \sum_{c\in C} E_c$ are needed. Because the projections $\{P_c:\ c\varepsilon C\}$ all commute, the following is one form:

$$P_C = \sum_{c\in C} P_c - \sum_{d\in D} P_d + \sum_{t\in T} P_t + \ldots \pm P_u$$

where $\mathcal{D}$ denotes the class of all intersections of distinct pairs of elements of $\mathcal{C}$, $\mathcal{T}$ of triples, ..., and $u = \bigcap\mathcal{C}$. It follows from 3.1 that we also have the formula

$$P_{\mathcal{C}} = \prod_{\gamma\in C} (P_{\gamma-}+P_{\gamma'}-P_{\partial\gamma})$$

and of course there are similar expressions based on 3.2 above.

The special case in which the graph $\underset{\sim}{C}$ is *decomposable* (see [14] for the definition and further details) is particularly interesting, because there is an *index* $\nu\colon \mathcal{D}\to\{0,1,2,\ldots\}$ defined on the intersections of pairs of distinct cliques of any decomposable graph $\underset{\sim}{C}$ such that

$$P_{\mathcal{C}} = \sum_{c\in\mathcal{C}} P_c - \sum_{d\in\mathcal{D}} \nu(d)P_d .$$

This formula was derived by Haberman [8, Chapter 5].

4. NNG = M

It is now possible to give a simple explanation of the coincidence of nearest-neighbour Gibbs with Markov states (or NNG = M, as we abbreviate it). The set C will be regarded as a finite set of *sites* γ, each having a finite *local state space* I_γ, and $\underset{\sim}{C}$ defines a graph over C (describing the spatial relationships between the sites) with maximal cliques $\mathcal{C}$. We are interested in (discrete) probabilities (states) p over $I = \Pi I_\gamma$, and only consider the *strictly positive* ones, i.e. those for which the probability $p(\underline{i})$ of every *configuration* $\underline{i}\in I$ is strictly positive.

Definition 1 The probability p is called a *nearest-neighbour Gibbs* (NNG) *state* relative to $\underset{\sim}{C}$ if $\log p \in \sum_{c\in\mathcal{C}} E_c$.

Corollary p is an NNG state if and only if $\log p \in \oplus(F_b\colon b \text{ complete})$.

This corollary gives a precise meaning to the notion that NNG states p are those whose "potential" log p has interaction only between (elements of) complete subsets, but this interpretation is not necessary for the main result.

For the next definition we remind the reader that p_a denotes the a-marginal of an element $p \in E$.

Definition 2   The probability p is said to be locally Markov relative to $\underset{\sim}{C}$ if for every $\gamma \in C$, $p = p_{\gamma^-} p_{\gamma'} / p_{\partial\gamma}$.

The purpose of this definition is best seen by considering an I-valued random variable $\underset{\sim}{X}: \Omega \to I$ defined on a probability space (Ω, F, P) and having distribution $p(\underline{i}) = P(\underset{\sim}{X} = \underline{i})$, $\underline{i} \in I$. For if we then write the a-marginal random variable $\underset{\sim}{X}_a = (X_\alpha : \alpha \in a)$, $a \subseteq C$, we see that for every $\gamma \in C$, X_γ and $\underset{\sim}{X}_{\gamma'}$ are conditionally P-independent given $\underset{\sim}{X}_{\partial\gamma}$ if and only if p is locally Markov. These conditional independence requirements (or their distributional equivalents) constitute a spatial Markov property - not by any means the most natural - which generalises the usual (discrete-time) Markov property. Other forms are mentioned below.

4.1 Proposition (NNG = M)   A strictly positive probability p is a NNG state if and only if it is locally Markov.

Proof   This is an immediate consequence of the two definitions, Lemma 3, and Lemma 1 (with $a = \gamma^-$, $b = \gamma'$, $ab = \partial\gamma$).

4.2 Other Markov properties   Two other properties which deserve the name Markov are now considered.

Definition 2'   The probability p is said to be regionally Markov if for every $a \subseteq C$, $p = p_{a^-} p_{a'} / p_{\partial a}$.

<u>Definition 2"</u> The probability p is said to be <u>globally Markov</u> if for every pair a and b of disjoint subsets of C separated in $\underset{\sim}{C}$ by $d \subseteq C$, we have $p = p_{a \cup d} p_{b \cup d} / p_d$.

In terms of the random variable $\underset{\sim}{X}$ with distribution p this last definition, for example, requires that $\underset{\sim}{X}_a$ and $\underset{\sim}{X}_b$ be conditionally P-independent given $\underset{\sim}{X}_d$ whenever d separates a and b in $\underset{\sim}{C}$. Every such pair a and b may be included in a pair of subsets e and f in such a way that $a \subseteq e \backslash f$, $b \subseteq f \backslash e$, $d = ef$, and $e \cup f = C$ (see Vorob'ev [30]; it is a simple maximality argument). Thus we can use the variants given in 3.2 to prove variants of 4.1 with locally replaced by regionally (resp. globally) in the Markov property.

The result of these remarks is the following: in the strictly positive case all three Markov properties and the NNG property are equivalent. We note that the equivalence of (say) the local and global Markov properties can be proved without the reference to NNG states, but that this is rather more tricky, and finally we mention that all of these equivalences fail if the probabilities in question are not strictly positive (see Moussouris [16] and Suomela [27]).

<u>4.3 Generalisations and variants of 4.1</u> (i) It would be quite straightforward to generalise the NNG = M result to cases in which the local state spaces I_γ are <u>countably infinite</u> equipped with a reference measure m_γ. The argument would be exactly the same as that already given but use one of the direct sum decompositions described in 2.2 above to establish an analogue of 3.1.

(ii) There is a version of NNG = M valid for <u>Gaussian distributions</u> which we state here without proof. It is the starting point for some work which will be reported elsewhere.

<u>Proposition'</u> Let $K = (K(\alpha,\beta))$ be a positive definite matrix defined on $C \times C$ where C is the vertex set of a graph $\underset{\sim}{C}$. The Gaussian density p having zero mean and covariance matrix K satisfies the local, regional and global Markov properties

(Definitions 2,2' and 2") if and only if we have $K^{-1}(\alpha,\beta) = 0$ whenever $\{\alpha,\beta\} \notin E(\underset{\sim}{C})$.

(iii) We now mention a result concerning conditional independence which, at the same time as allowing us to view NNG = M in a new light, also unifies the discrete version with the Gaussian one just noted. Suppose that $\mathcal{A}_1, \mathcal{A}_2$ and $\mathcal{B}$ are sub-σ-fields in a probability space $(\Omega, \mathcal{F}, \mu)$ and that $\mathcal{A}_1 \vee \mathcal{B} \vee \mathcal{A}_2 = \mathcal{F}$. Further suppose that $\mathcal{A}_1$ and $\mathcal{A}_2$ are conditionally μ-independent given $\mathcal{B}$, and that p is a μ-a.s. positive measurable function with $\int p\,d\mu = 1$. If we denote by $p.\mu$ the measure $F \mapsto \int_F p\,d\mu\ (F \in \mathcal{F})$, then it is not hard to prove that $\mathcal{A}_1$ and $\mathcal{A}_2$ are conditionally $p.\mu$-independent given $\mathcal{B}$ if and only if we can write $p = p_1 p_2$ a.s. μ with p_i $\mathcal{A}_i \vee \mathcal{B}$-measurable, $i = 1,2$. (For a proof of this and further discussion, see [24]).

Now let us consider an I-valued random variable $\underset{\sim}{X}$ defined on the probability space $(\Omega, \mathcal{F}, \mu)$, where $I = \prod_{\gamma \in C} I_\gamma$ and C is the vertex set of a graph $\underset{\sim}{C}$. Then $\underset{\sim}{X}$ is locally μ-Markov relative to $\underset{\sim}{C}$ if and only if for every $\gamma \in C$ the σ-fields $\mathcal{B}_{\gamma^-}$ and $\mathcal{B}_{\gamma'}$ generated by $\underset{\sim}{X}_{\gamma^-}$ and $\underset{\sim}{X}_{\gamma'}$ are conditionally μ-independent given $\mathcal{B}_{\partial\gamma}$, the σ-field generated by $\underset{\sim}{X}_{\partial\gamma}$. For an a.s. positive function p with $\int p\,d\mu = 1$ the preceding result tells us that $\underset{\sim}{X}$ is (locally) $p.\mu$-Markov relative to $\underset{\sim}{C}$ if and only if for every $\gamma \in C$ we can write $p = p_1 p_2$ a.s. μ with p_1 $\mathcal{B}_{\gamma^-}$-measurable and p_2 $\mathcal{B}_{\gamma'}$-measurable; Nelson [17]. Taking μ to be the uniform probability measure on $I = \prod I_\gamma$ (obviously Markov) this gives an alternative viewpoint on the NNG = M result, whilst we can also see that the Gaussian result above is covered, for there the density is $p(\underset{\sim}{x}) = \exp\left[-\tfrac{1}{2}\sum_{\alpha \neq \beta} K^{-1}(\alpha,\beta) x_\alpha x_\beta\right]$ relative to the product of the marginal measures on $I_\gamma = \mathbb{R}$, $\gamma \in C$.

(iv) There is an analogue of the NNG = M result valid for finite local state spaces but a countably infinite set C of sites. In this context the graph is usually locally finite, i.e. each site has only finitely many neighbours. The result in question has a number of forms, each depending in its proof on the one for finitely many sites, and further details can be found in Preston [20] or Suomela [28].

(v) Finally we mention Markov properties for uncountable sets C such as R, R^n, n>1, or more general index sets. Gaussian processes indexed by R^n are considered by Pitt [18] (following earlier work by other authors) and the result analogous to that noted in (ii) above is the equivalence of a locality condition on the reproducing kernel Hilbert space generated by the covariance function of the process to a (topological) local Markov property. Similarly Markov point processes have been discussed by Ripley and Kelly [22], where the Markov property is defined in terms of a symmetric reflexive relation (denoting identity or the neighbour relation).

4.4 Explicit formulae for Gibbs states The class of graphs termed decomposable in [14] have the interesting property that Markov measures relative to such graphs factorise into products of marginals in a very simple way. In terms of the index ν defined on the class $\mathcal{D}$ of all intersections of pairs of distinct cliques and the class $\mathcal{C}$ of all cliques of such a graph $\underset{\sim}{C}$, we can prove:

for a strictly positive probability p on $I = \Pi I_\gamma$, p is Markov relative to $\underset{\sim}{C}$ if and only if

$$p = \prod_{c\in\mathcal{C}} p_c \Big/ \prod_{d\in\mathcal{D}} p_d^{\nu(d)}.$$

This result (in a different form) is due to Vorob'ev [30] and Kellerer [12], and is intimately connected with the work of Haberman [8] mentioned in 1.2.

4.5 Relation of the foregoing to existing proofs of NNG = M It is not hard to see that any proof of NNG = M within the finite discrete framework will need to use, implicitly or explicitly, some form of 1.1. On the other hand it is less obvious, but appears to be the case, that all proofs need to use some sort of unique representation (cf. 2.1, 2.2) of elements of $\mathbb{R}^I$ as sums of functions involving subsets of the variables. The most frequently used form involves what we might term substitution operators defined relative to some fixed configuration $\underline{k}\in I$ by

$$(S_a p)(\underline{i}) = p(\underline{i}_a \underline{k}_{a'}),\quad a \subseteq C,\ \underline{i}\in I.$$

The subspace $\sum_{c \varepsilon C} E_c$ can be defined in terms of these operators, and in discussing this use is made of the inclusion-exclusion principle (also called Fourier analysis (mod 2), or Möbius inversion on the partially ordered set of subsets of C). Relations between these approaches and others to be found in the literature, as well as some general results, will be discussed in [4]. Finally we note that the property of complete subsets (b is complete if and only if for all $\gamma \varepsilon C$, $\gamma \not\subseteq b$ or $\gamma - \supseteq b$) is also used implicitly in most of the existing proofs, generally in the form of an include/exclude trick in an inclusion-exclusion expansion of the "potential" log p in terms of "interaction potentials".

REFERENCES

[1] ASMUSSEN, Søren — Statistik Forsøgsplanlægning. Institute of Mathematical Statistics, University of Copenhagen, (1975), 99p.

[2] AVERINTSEV, M.V. — A method of describing discrete parameter random fields. Problemy Peredaci Informatsii 6 (1970) 100-109.

[3] BESAG, J.E. — Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc., Series B 36 (1974) 192-236.

[4] DARROCH, J.N. and SPEED, T.P. — Multiplicative and additive models and interactions, Manuscript (1977).

[5] DAVIDSON, R. — Determination of confounding. In Stochastic Analysis edited by D.G. KENDALL and E.F. HARDING. John Wiley & Sons, London, New York, Sydney, Toronto, 1973.

[6] DOBRUSHIN, R.L. — The description of a random field by means of its conditional probabilities and conditions of its regularity. Theory Probability and Appl. 13 (1968) 197-224.

[7] GRIMMETT, G.R. — A theorem about random fields. Bull. Lond. Math. Soc. 5 (1973) 81-84.

[8] HABERMAN, Shelby J. — The Analysis of Frequency Data. The University of Chicago Press, Chicago and London 1974.

[9] HABERMAN, Shelby J. — Directe products and linear models for complete factorial tables. Ann. Statist. 3 (1975) 314-333.

[10] HAMMERSLEY, J.M. and CLIFFORD, P. — Markov fields on finite graphs and lattices. Unpublished manuscript (1971).

[11] KELLERER, Hans G. — Maßtheoretische marginalprobleme. Math. Ann. 153 (1964) 168-198.

[12] KELLERER, Hans G. — Verteilungsfunktionen mit gegeben Marginalverteilungen. Z. Warscheinlichkeitstheorie 3 (1964) 247-270.

[13] LANCASTER, H.O. — The Chi-squared Distribution. John Wiley & Sons, Inc. London, New York, Sydney 1969.

[14] LAURITZEN, Steffen L., SPEED, T.P. and VIJAYAN, K. — Decomposable graphs and hypergraphs.

[15] MACLANE, SAUNDERS and GARRETT BIRKHOFF — Algebra. The Macmillan Company, New York 1967.

[16] MOUSSOURIS, John — Gibbs and Markov random systems with constraints. J. Statistical Physics 10 (1974) 11-33.

[17] NELSON, E. Quantum fields and Markoff fields. Amer. Math. Soc. Summer Institute on Partial Differential Equations, Berkeley 1971.

[18] PITT, Loren D. A Markov property for Gaussian processes with multidimensional parameter. Archiv. for Rational Mech. and Anal. 43 (1971) 367-391.

[19] PRESTON, C.J. Generalised Gibbs states and Markov random fields. Advances in Applied Probability 5 (1973) 242-261.

[20] PRESTON, Christopher J. Gibbs States on Countable Sets. Cambridge University Press 1974.

[21] PRESTON, Christopher J. Random fields. Lectures notes in Mathematics No. 534, (1976). Springer-Verlag

[22] RIPLEY, B.D. and KELLY, F.P. Markov point processes. J. Lond. Math. Soc. (2) 15 (1977) 188-192.

[23] SHERMAN, S. Markov random fields and Gibbs random fields. Israel J. Math. 14 (1973) 92-103.

[24] SPEED, T.P. A factorisation theorem for adequate statistics. Austral. J. Statistics (to appear)

[25] SPITZER, F. Random fields and interacting particle systems. Notes on lectures given at the 1971 MAA Summer Seminar.

[26] SUOMELA, P. Factorrings of finite-dimensional distributions. Commentationes Physico-Mathematica (Helsinki) 42 (1972) 231-243.

[27] SUOMELA, P. Construction of Nearest-Neighbour Systems. Annales Acad. Scientiarum Fenn. Ser. A. Math. Diss. 10 (1976)

[28] SUOMELA, P. Dependence structures of discrete random fields. Report on Computer Science, Economics and Statistics, University of Jyväskylä 26/1978

[29] VOROB'EV, N.N. Consistent families of measures and their extensions. Theory Probability and Appl. 7 (1962) 147-163.

[30] VOROB'EV, N.N. Markov measures and Markov extensions. Theory Probability and Appl. 8 (1963) 420-429.

July 6, 1978. Sheffield, U.K.

Permanent address: Department of Mathematics,
University of Western Australia,
NEDLANDS, W.A. 6009,
Australia.

PREPRINTS 1977

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Asmussen, Søren & Keiding, Niels: Martingale Central Limit Theorems and Asymptotic Estimation Theory for Multitype Branching Processes.

No. 2 Jacobsen, Martin: Stochastic Processes with Stationary Increments in Time and Space.

No. 3 Johansen, Søren: Product Integrals and Markov Processes.

No. 4 Keiding, Niels & Lauritzen, Steffen L. : Maximum likelihood estimation of the offspring mean in a simple branching process.

No. 5 Hering, Heinrich: Multitype Branching Diffusions.

No. 6 Aalen, Odd & Johansen, Søren: An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations.

No. 7 Johansen, Søren: The Product Limit Estimator as Maximum Likelihood Estimator.

No. 8 Aalen, Odd & Keiding, Niels & Thormann, Jens: Interaction Between Life History Events.

No. 9 Asmussen, Søren & Kurtz, Thomas G.: Necessary and Sufficient Conditions for Complete Convergence in the Law of Large Numbers.

No. 10 Dion, Jean-Pierre & Keiding, Niels: Statistical Inference in Branching Processes.

PREPRINTS 1978

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Tjur, Tue: Statistical Inference under the Likelihood Principle.

No. 2 Hering, Heinrich: The Non-Degenerate Limit for Supercritical Branching Diffusions.

No. 3 Henningsen, Inge: Estimation in M/G/1-Queues.

No. 4 Braun, Henry: Stochastic Stable Population Theory in Continuous Time.

No. 5 Asmussen, Søren: On some two-sex population models.

No. 6 Andersen, Per Kragh: Filtered Renewal Processes with a Two-Sided Impact Function.

No. 7 Johansen, Søren & Ramsey, Fred L.: A Bang-Bang Representation for 3x3 Embeddable Stochastic Matrix.

No. 8 Braun, Henry: A Simple Method for Testing Goodness of Fit in the Presence of Nuisance Parameters.

No. 9 Lauritzen, S.L. & Speed, T.P. & Vijayan, K. : Decomposable Graphs and Hypergraphs.

No. 10 Hald, Anders: On the Statistical Theory of Sampling Inspection by Attributes.

No. 11 Darroch, J.N. & Lauritzen, S.L. & Speed, T.P.: Markov Fields and Log-linear Interaction Models for Contingency Tables.

No. 12 Speed, T.P.: A Note on Nearest-Neighbour Gribbs and Markov Probabilities.