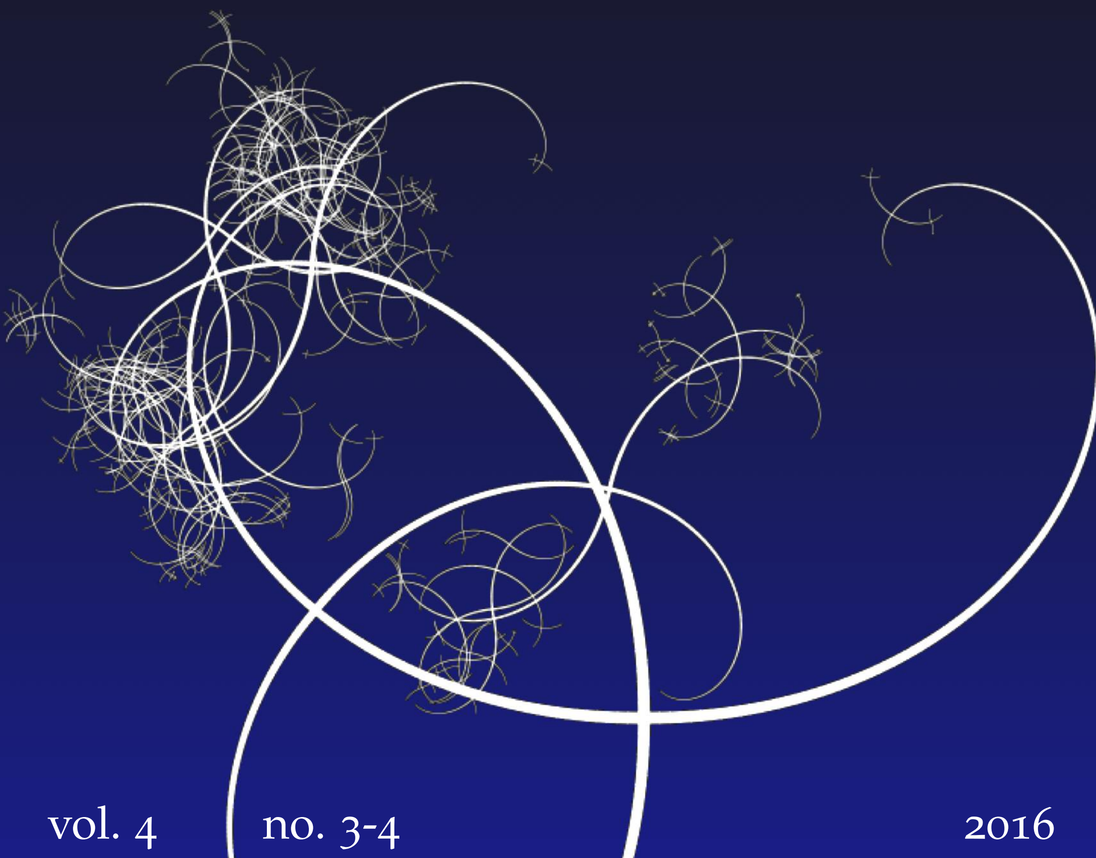
NISSUNA UMANA INVESTIGAZIONE SI PUO DIMANDARE VERA SCIENZIA
S'ESSA NON PASSA PER LE MATEMATICHE DIMOSTRAZIONI
LEONARDO DA VINCI



vol. 4 no. 3-4 2016

Mathematics and Mechanics of Complex Systems

Alberto Gandolfi and Pietro Lenarda

A NOTE ON GIBBS AND MARKOV RANDOM FIELDS WITH CONSTRAINTS AND THEIR MOMENTS

msp

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 4, No. 3-4, 2016
dx.doi.org/10.2140/memocs.2016.4.407

A NOTE ON GIBBS AND MARKOV RANDOM FIELDS WITH CONSTRAINTS AND THEIR MOMENTS

ALBERTO GANDOLFI AND PIETRO LENARDA

This paper focuses on the relation between Gibbs and Markov random fields, one instance of the close relation between abstract and applied mathematics so often stressed by Lucio Russo in his scientific work.

We start by proving a more explicit version, based on spin products, of the Hammersley–Clifford theorem, a classic result which identifies Gibbs and Markov fields under finite energy. Then we argue that the celebrated counterexample of Moussouris, intended to show that there is no complete coincidence between Markov and Gibbs random fields in the presence of hard-core constraints, is not really such. In fact, the notion of a constrained Gibbs random field used in the example and in the subsequent literature makes the unnatural assumption that the constraints are infinite energy Gibbs interactions on the same graph. Here we consider the more natural extended version of the equivalence problem, in which constraints are more generally based on a possibly larger graph, and solve it.

The bearing of the more natural approach is shown by considering identifiability of discrete random fields from support, conditional independencies and corresponding moments. In fact, by means of our previous results, we show identifiability for a large class of problems, and also examples with no identifiability. Various open questions surface along the way.

Personal acknowledgment. One of us (Gandolfi) learned about the theory of Gibbs and Markov random fields from Lucio Russo in a course based on [Ruelle 1978]. He is indebted to Lucio for his inspirational lectures and for many other things, such as an interest in percolation theory and statistical physics, a deep conviction of the close relation between abstract and applied mathematics, and an involvement in questions about the history of science.

This paper focuses on one instance of this close association between abstract and applied mathematics, namely the relation between Gibbs and Markov random fields; in spite of the great number of studies and applications of these models, this relationship has not been appropriately investigated in the literature.

Communicated by Raffaele Esposito.
MSC2010: primary 60J99, 82B20; secondary 44A60, 62B05, 62M40.
Keywords: Gibbs distributions, Markov random fields, hard-core constraints, moments, Hammersley–Clifford, Moussouris.

1. Introduction

Gibbs random fields are important models in equilibrium statistical mechanics, and Markov random fields are fundamental models in applications. They can both be defined in terms of a given graph, and they are almost two faces of the same phenomenon: Gibbs fields are defined from microscopic interactions and Markov fields by, in principle observable, conditional independencies.

Leaving the details for later, we can say in a nutshell that every Gibbs measure is Markov; the question is whether the opposite is also true. A celebrated result of Hammersley and Clifford [1971] states that with finite energy, i.e., the absence of zero probability configurations, every Markov random field is Gibbs. On the other hand, a famous example by Moussouris [1974] shows that in the presence of hard-core constraints, which is to say without the finite energy assumption, there are Markov fields which are not Gibbsian. Further studies have clarified that on a chordal graph Gibbsianity is equivalent to the global Markov property, regardless of finite energy [Lauritzen 1996], and that detailed algebraic conditions seem to be needed on nonchordal graphs [Geiger et al. 2006]. These results would seem to settle the issue.

There is, however, one weakness in this picture. When hard-core conditions, which can also be graph-based, appear in the results above, they are defined in terms of the same graph as the one used for the interactions. Such a choice seems to be justified by two simple remarks: first, one can always take the union of the two graphs, the one for interactions and the one for the hard-core conditions, as a common graph for both (as larger graphs accommodate more interactions or less conditional independence requirements); second, one can interpret the hard-core conditions as unbounded interactions, that are thus subject to the same geometrical dependence. In spite of these two remarks, however, the assumption of a unique graph is physically unwarranted; in general, the mechanisms which induce hard-core conditions are completely different from those generating interactions. For gravitational fields, for instance, the interaction is long-range while hard-core conditions can take care of the impenetrability of rigid bodies; on the other hand, in a canonical ensemble of short range interacting particles, the hard-core condition is long-range as opposed to the interaction. Even more importantly, the graph of Markov conditional independencies is naturally related to the one on which the interaction is based, which, as we just argued, has no relation to the one for hard-core conditions. Assuming a unique range for hard-core conditions and interactions hinders the more relevant relation between interactions and conditional independencies, and leads to confusing results.

The first consequence of the remarks above is the need of a more careful analysis, and of more explicit notation highlighting the importance of the graphs next to the notions of Markov or Gibbs; we present this in Section 2. Notice that a more careful distinction of the role of interactions and hard-core constraints already appears (albeit with less explicit notation) in the literature, chiefly in Ruelle's thermodynamic formalism [Ruelle 1978].

With the more explicit and natural identification of the graphs, it is still the case that a Gibbs random field is Markov, but the issue of whether a Markov random field on a given finite graph can always be obtained as a Gibbs field on the same graph with hard-core conditions based on a possibly larger graph falls outside the scope of past researches. In Section 5 we provide an answer to this question.

Prior to this, we give an alternative proof of the Hammersley–Clifford theorem, potentially more suitable for applications. One of the earlier proofs by Besag [1974] expresses the interaction in terms of products of spins, but works only for the binary case and has some problematic steps in the argument; the proof of Grimmett [1973], on the other hand, is valid for all finite and countable state spaces, but does not express the interaction as an explicit function of the spin values. The proof we present here expresses the interaction in terms of spin products and holds for all finite state spaces. In a sense, it exploits the fact that spin products are a basis of the interaction space. In so doing, we get an explicit calculation in terms of inverse Vandermonde matrices; we also get a more direct relation with the moments of the distribution.

Another noticeable consequence of our work is explored in Section 7, where we show that the statistical identifiability of a discrete random field by support, conditional independencies and moments can be analyzed by a combination of the moment related representation of Gibbs fields in Section 4, and of the clarification of the role of the graphs in the Markov–Gibbs relation in Section 5.

2. Definitions

Let Λ be a finite set of vertices, Ω_x be a finite set for each $x \in \Lambda$, and $\Omega = \prod_{x \in \Lambda} \Omega_x$. For a subset $A \subseteq \Lambda$, ω_A is a configuration in $\Omega_A = \prod_{x \in A} \Omega_x$; the same notation is used for the restriction of a configuration $\omega \in \Omega$ to A. Later on, we use the notation

$$[\bar{\omega}_1, *, \bar{\omega}_2, \dots, *] = \{\omega \in \Omega : \omega_i = \bar{\omega}_i, \text{ for all } i \text{ such that } \bar{\omega}_i \neq *\}$$

for cylinders.

In this paper we consider probabilities, generally denoted as P, on $(\Omega, \mathcal{P}(\Omega))$, where for every finite set S, $\mathcal{P}(S)$ indicates the set of all subsets of S.

To express the notions of interest here, we consider a graph $\mathcal{G} = (\Lambda, \mathcal{B})$ in which the set of undirected bonds is $\mathcal{B} \subseteq \{\{x, y\} \mid x \neq y, x, y \in \Lambda\}$. Given a probability P, we say that two sets $A, B \subseteq \Lambda$ are conditionally independent given a third set C,

$A \perp\!\!\!\perp B \mid C$, if ω_A and ω_B are conditionally independent given ω_C for all $\omega_S \in \Omega_S$, and $S = A, B, C$.

A probability P is pairwise Markov with respect to the bonds $\mathcal{B}$, or as we call it from now on, *$\mathcal{B}$-pair-Markov*, if

(1) for all pairs of vertices $x, y \in \Lambda$ which are not neighbors on $\mathcal{G}$, i.e., such that $\{x, y\} \notin \mathcal{B}$, $x \perp\!\!\!\perp y \mid (\Lambda \setminus \{x, y\})$,

and it is *$\mathcal{B}$-global-Markov* with respect to the bonds $\mathcal{B}$ if

(1′) for all pairs of disjoint sets $A, B \subseteq \Lambda$ which are not neighbors on $\mathcal{G}$, i.e., such that there is no bond in $\mathcal{B}$ connecting a vertex of A to a vertex of B, $A \perp\!\!\!\perp B \mid (\Lambda \setminus A \cup B)$.

P is *$\mathcal{B}$-pair-Markov$^>$*, or *$\mathcal{B}$-global-Markov$^>$*, if in addition

(2) $P(\omega) > 0$ for all $\omega \in \Omega$.

These are the notions of Markov probability or Markov random field generally used in the literature [Grimmett 1973; Lauritzen 1996; Geiger et al. 2006].

A $\mathcal{B}$-clique of the graph $\mathcal{G}$ is a maximal complete subgraph, possibly including single vertices, of $\mathcal{G}$. We denote by $\mathcal{C}\ell(\mathcal{B})$ the collection of subsets $A \subseteq \Lambda$ which are subsets of the vertex set of a clique of $\mathcal{G} = (\Lambda, \mathcal{B})$. A $\mathcal{B}$-interaction is a function $\phi : \cup_{A \in \mathcal{C}\ell(\mathcal{B})} \Omega_A \to \mathbb{R}$. Next, we consider a further collection $\mathcal{F} \subseteq \mathcal{P}(\Lambda)$ of subsets of Λ, and a (possibly empty) set $\overline{\Omega}_A$ of forbidden configurations for each $A \in \mathcal{F}$; notice that some authors focus on the set of allowed configurations (see [Ruelle 1978]), but our choice underlines the exceptionality of being forbidden and highlights the role of $\mathcal{F}$, as no restrictions can be imposed for sets not in $\mathcal{F}$. In greater generality, one can take as forbidden configurations those belonging to the set $\overline{\Omega}$ of zeros of a function ρ defined on Ω. It is convenient to deal with forbidden configurations by assigning them a probability anyway, which is then required to be zero.

A probability P is Gibbs with respect to the graph $\mathcal{G} = (\Lambda, \mathcal{B})$ and the allowed configurations on $\mathcal{F}$, or *$\mathcal{B}$-$\mathcal{F}$-Gibbs*, as we call it from now on, if for all $\omega \in \Omega$,

$$P(\omega) = \frac{1}{Z_\phi}\left(e^{\sum_{A \in \mathcal{C}\ell(\mathcal{B})} \phi(\omega_A)} \prod_{B \in \mathcal{F}} \mathbb{I}_{\Omega_B \setminus \overline{\Omega}_B}(\omega_B)\right), \tag{1}$$

where $\mathbb{I}_S$ the indicator function of the set S, ϕ is a $\mathcal{B}$-interaction, $\overline{\Omega}_B$ for $B \in \mathcal{F}$ is a collection of forbidden configurations, and Z_ϕ is a normalization factor. Notice that $\mathcal{B}$-$\varnothing$-Gibbs means that all configurations have positive probability; $\varnothing$-$\varnothing$-Gibbs is a Bernoulli distribution; and $\varnothing$-$\mathcal{F}$-Gibbs is a Bernoulli distribution constrained to have some zero probabilities. Moreover, $\mathcal{B}$-Λ-Gibbs means that the hard-core constraints can be imposed on the entire configuration, and $\mathcal{B}$-$\mathcal{B}$-Gibbs indicates the fact that both interaction and hard-core constraints are assigned on configurations defined on subsets of the cliques of the same graph $(\Lambda, \mathcal{B})$.

In abstract terms, ϕ and $\overline{\Omega}$ are measurable with respect to a σ-algebra of Ω, but as the σ-algebra can be expressed in terms of set $\mathcal{B}$ of bonds (in the sense that the σ-algebra is the one generated by $\cup_{A\in\mathcal{C}\ell(\mathcal{B})}\Omega_A$), we focus on $\mathcal{B}$ and $\mathcal{F}$ in the notation.

Finally, we sometimes use indices to distinguish the various collection of bonds $\mathcal{B}$. In general, we indicate it by $\mathcal{B}_m$ if m is the size of the largest clique. Notice that if $\mathcal{B} \subseteq \mathcal{B}'$ and $\mathcal{F} \subseteq \mathcal{F}'$, then $\mathcal{B}$-pair(global)-Markov implies $\mathcal{B}'$-pair(global)-Markov, and $\mathcal{B}$-$\mathcal{F}$-Gibbs implies $\mathcal{B}'$-$\mathcal{F}'$-Gibbs, so one is generally interested in the minimal such graphs and collections.

3. Previous results and one ensuing question

Global Markov implies pairwise Markov, but one can easily construct an example with enough configurations of zero probability showing that the opposite implication does not hold [Lauritzen 1996]. On the other hand, on some graphs there is no difference between pairwise and global Markov, as the only sets which can be separated are pairs. Such is, for instance, the graph with $\Lambda = \{1, 2, 3, 4\}$ and $\mathcal{B}_2 = \{(1, 2), (2, 3), (3, 4), (4, 1)\}$; this graph is used for several examples below.

$\mathcal{B}$-$\mathcal{F}$-Gibbs implies $\mathcal{B}$-global-Markov for any $\mathcal{F}$, which then implies $\mathcal{B}$-pair-Markov. The reversed implication is given for the case in which finite energy holds by the celebrated Hammersley–Clifford theorem [1971], which in our terminology can be phrased as follows:

Theorem 3.1 (Hammersley–Clifford). *Given a graph $\mathcal{G} = (\Lambda, \mathcal{B})$, a random field P is $\mathcal{B}$-pair-Markov$^{>}$ if and only if it is $\mathcal{B}$-$\varnothing$-Gibbs for some potential ϕ.*

There are various proofs of this result, probably starting from [Brook 1964] and the unpublished paper [Hammmersley and Clifford 1971] (see also [Grimmett 2010]). An explicit dependence of ϕ from the spin values appears for the binary case (i.e., $|\Omega_x| = 2$) in [Besag 1974] (with some unclear steps in the proofs); a general version including countable state spaces was proven by Grimmett [1973], but without the explicit dependence of ϕ from the spin values. A simpler statement, in which conditional probabilities are known instead of unconditional ones, is presented in [Onural 2016] and very likely elsewhere. For completeness, we prove Theorem 3.1 once again in Section 4 below; the proof we give is for all finite Ω_x but with ϕ explicitly expressed in terms of spin products. Our results can also be indirectly obtained from [Grimmett 1973] by decomposing ϕ on the basis of spin products.

The Hammersley–Clifford theorem has been generalized by Lauritzen [1996] by means of chordal graphs: an undirected graph is said to be chordal if every cycle of length 4 or more has a chord.

Theorem 3.2. *If the graph $\mathcal{G} = (\Lambda, \mathcal{B})$ is chordal, then a random field P is $\mathcal{B}$-global-Markov if and only if it is $\mathcal{B}$-$\mathcal{B}$-Gibbs for some potential ϕ.*

An algebraic interpretation and a slight generalization of these results is given in [Geiger et al. 2006].

In the opposite direction, Moussouris' counterexample [1974] shows that it is not true that every $\mathcal{B}$-global-Markov random field is $\mathcal{B}$-$\mathcal{B}$-Gibbs. This happens necessarily on a nonchordal graph.

Example 3.3 (Moussouris). Take $\Lambda = \{1,2,3,4\}$, $\mathcal{B}_2 = \{(1,2), (2,3), (3,4), (4,1)\}$ and let $P_{\mathcal{M}}$ be the uniform distribution on

$$\Omega' = \{(0,0,0,0),\ (1,0,0,0),\ (1,1,0,0),\ (1,1,1,0),\\ (1,1,1,1),\ (0,1,1,1),\ (0,0,1,1),\ (0,0,0,1)\}.$$

It is easily seen (and we will explicitly show a related statement in Lemma 5.2 below) that

$$\{1\} \perp\!\!\!\perp \{3\} \mid \{2, 4\} \quad \text{and} \quad \{2\} \perp\!\!\!\perp \{4\} \mid \{1, 3\}$$

so that $P_{\mathcal{M}}$ is $\mathcal{B}_2$-pair-Markov, and is (on this graph) also $\mathcal{B}_2$-global-Markov. On the other hand, $P_{\mathcal{M}}$ cannot be $\mathcal{B}_2$-$\mathcal{B}_2$-Gibbs; if it was so, then

$$\begin{aligned} P_{\mathcal{M}}(\omega) =& \frac{1}{Z} \prod_{A \in \mathcal{C}\ell(\mathcal{B}_2)} e^{\phi_B(\omega_B)} \prod_{B \in \mathcal{B}_2} \mathbb{I}_{\Lambda \setminus \overline{\Omega}_B}(\omega_B) \\ =& c(\omega)\psi_{(1,2)}(\omega_1, \omega_2)\psi_{(2,3)}(\omega_2, \omega_3)\psi_{(3,4)}(\omega_3, \omega_4)\psi_{(4,1)}(\omega_4, \omega_1) \end{aligned}$$

for suitable, not necessarily nonnegative, functions $\psi_{(i,j)}$ and $c(\omega) > 0$. But

$$\begin{aligned} P_{\mathcal{M}}(0, 1, 1, 0) &= \psi_{(1,2)}(0, 1)\psi_{(2,3)}(1, 1)\psi_{(3,4)}(1, 0)\psi_{(4,1)}(0, 0) = 0, \\ P_{\mathcal{M}}(0, 1, 1, 1) &= \psi_{(1,2)}(0, 1)\psi_{(2,3)}(1, 1)\psi_{(3,4)}(1, 1)\psi_{(4,1)}(1, 0) = \tfrac{1}{8}, \\ P_{\mathcal{M}}(0, 0, 0, 0) &= \psi_{(1,2)}(0, 0)\psi_{(2,3)}(0, 0)\psi_{(3,4)}(0, 0)\psi_{(4,1)}(0, 0) = \tfrac{1}{8}, \\ P_{\mathcal{M}}(1, 1, 1, 0) &= \psi_{(1,2)}(1, 1)\psi_{(2,3)}(1, 1)\psi_{(3,4)}(1, 0)\psi_{(4,1)}(0, 1) = \tfrac{1}{8}, \end{aligned}$$

are incompatible.

However, the probability in Moussouris example is $\mathcal{B}_2$-Λ-Gibbs, actually even $\varnothing$-Λ-Gibbs: to see this, it is enough to take the uniform Bernoulli distribution $\hat{P}(\omega) = 1/Ze^0 = 1/2^4$, with $\mathcal{F} = \Lambda$ and $\overline{\Omega}_\Lambda = \Omega \setminus \Omega'$; the graph of the hard-core constraints can actually be further reduced (see Section 6 below), although obviously not to $\mathcal{B}_2$. Moussouris' example is thus clearly not a counterexample to the following, more natural, question.

Question. *Is every Markov random field a constrained Gibbs random field on the same graph, in the more natural sense that every $\mathcal{B}$-global-Markov random field is $\mathcal{B}$-Λ-Gibbs?*

We take over this issue in Section 5 below.

4. Markov–Gibbs equivalence with no constraints: an explicit Hammersley–Clifford theorem

Before tackling the main question, we give a new proof of Theorem 3.1. We actually prove the following more explicit version, assuming that the possible states Ω_x are real numbers. It amounts to an explicit expansion of the interaction on the basis of spin products.

Lemma 4.1. *Let $\mathcal{G} = (\Lambda, \mathcal{B})$ be a given finite graph; $\Omega = \prod_{x\in\Lambda} \Omega_x$ for finite $\Omega_x \subseteq \mathbb{R}$; and P be a $\mathcal{B}$-pair-Markov$^>$ random field on $(\Omega, \mathcal{P}(\Omega))$.*

Next, let $\widetilde{\Omega}_x = \{0, 1, \dots, |\Omega_x| - 1\}$; $\widetilde{\Omega} = \prod_{x\in\Lambda} \widetilde{\Omega}_x$; and for $\sigma \in \widetilde{\Omega}$ and $\omega \in \Omega$, let $\omega^\sigma := \prod_{x\in\Lambda} \omega_x^{\sigma_x}$ (with the convention $0^0 = 1$, if needed). Moreover, let $\widetilde{\Omega}(\mathcal{B})$ be the set of σ such that $\{x : \sigma_x \neq 0\}$ is contained in a clique of $\mathcal{B}$. Then

$$P(\omega) = \frac{1}{Z} e^{\sum_{\sigma\in\widetilde{\Omega}(\mathcal{B})} J_\sigma \omega^\sigma} \tag{2}$$

with

$$J_\sigma = \sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega} \log P(\omega),$$

where $V^{-1}_{\sigma,\omega} = \prod_{x\in\Lambda} V^{-1}_{\sigma_x,\omega_x}(x)$ and $V^{-1}_{\sigma_x,\omega_x}(x)$ is the element in position (σ_x, ω_x) of the inverse $V^{-1}(x)$ of the Vandermonde matrix $V(x) = (r^s)_{r\in\Omega_x, s\in\widetilde{\Omega}_x}$.

Proof. The Vandermonde matrix is invertible as long as the elements of Ω_x are all different [Macon and Spitzbart 1958]. Next, for each $\omega \in \Omega$,

$$\begin{aligned}
e^{\sum_{\sigma\in\widetilde{\Omega}} J_\sigma \omega^\sigma} &= e^{\sum_{\sigma\in\widetilde{\Omega}} \omega^\sigma \sum_{\bar{\omega}\in\Omega} V^{-1}_{\sigma,\bar{\omega}} \log P(\bar{\omega})} \\
&= e^{\sum_{\bar{\omega}\in\Omega} \log P(\bar{\omega}) \sum_{\sigma\in\widetilde{\Omega}} \prod_{x\in\Lambda} \omega_x^{\sigma_x} V^{-1}_{\sigma_x,\bar{\omega}_x}(x)} \\
&= e^{\sum_{\bar{\omega}\in\Omega} \log P(\bar{\omega}) \prod_{x\in\Lambda} \left(\sum_{\sigma_x\in\widetilde{\Omega}_x} \omega_x^{\sigma_x} V^{-1}_{\sigma_x,\bar{\omega}_x}(x)\right)} \\
&= e^{\sum_{\bar{\omega}\in\Omega} \log P(\bar{\omega}) \prod_{x\in\Lambda} \delta_{\omega_x,\bar{\omega}_x}} \\
&= e^{\log P(\omega)} = P(\omega),
\end{aligned}$$

which gives (2) but with $\widetilde{\Omega}$ instead of $\widetilde{\Omega}(\mathcal{B})$. We now need to show that if $x \perp\!\!\!\perp y \mid (\Lambda \setminus \{x, y\})$, then $J_\sigma = 0$ for all σ such that $\sigma_x\sigma_y \neq 0$; it would then follow that for all $\sigma \in \widetilde{\Omega}$ such that $J_\sigma \neq 0$ we have that $\{x : \sigma_x \neq 0\}$ is contained in a clique of $\mathcal{G}$, as required. Indeed, if $x \perp\!\!\!\perp y \mid (\Lambda \setminus \{x, y\})$ then for all $\omega_x \in \Omega_x$, $\omega_y \in \Omega_y$, $\omega_{\setminus x,y} \in \Omega_{\setminus x,y} = \prod_{z\in\Lambda\setminus\{x,y\}} \Omega_z$ and $\omega = (\omega_x\omega_y\omega_{\setminus x,y})$ we have

$$P(\omega_x\omega_{\setminus x,y})P(\omega_y\omega_{\setminus x,y}) = P(\omega_{\setminus x,y})P(\omega_x\omega_y\omega_{\setminus x,y}),$$

i.e.,

$$\log P(\omega) = \log P(\omega_x\omega_y\omega_{\setminus x,y}) = \log P(\omega_x\omega_{\setminus x,y}) + \log P(\omega_y\omega_{\setminus x,y}) - \log P(\omega_{\setminus x,y}).$$

Therefore, if σ is such that $\sigma_x\sigma_y \neq 0$, we have

$$J_\sigma = \sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega} \log P(\omega)$$
$$= \sum_{\omega_{\setminus x,y}\in\Omega_{\setminus x,y}} \prod_{z\in\Lambda\setminus\{x,y\}} V^{-1}_{\sigma_z,\omega_z}(z) \sum_{\omega_x\in\Omega_x}\sum_{\omega_y\in\Omega_y} V^{-1}_{\sigma_x,\omega_x}(x)V^{-1}_{\sigma_y,\omega_y}(y)$$
$$\big(\log P(\omega_x\omega_{\setminus x,y}) + \log P(\omega_y\omega_{\setminus x,y}) - \log P(\omega_{\setminus x,y})\big), \quad (3)$$

which vanishes for the following reason: We have

$$\sum_{\omega_x\in\Omega_x} V^{-1}_{\sigma_x,\omega_x}(x) = \sum_{\omega_x\in\Omega_x} V^{-1}_{\sigma_x,\omega_x}(x)1 = [V^{-1}(x)V(x)]_{(\sigma_x,1)} = \delta_{\sigma_x,0}$$

since the first column of $V(x)$ is constantly equal to 1; here, $[A]_{i,j}$ denotes the element $\{i, j\}$ of the matrix A. This way, if $\sigma_x\sigma_y \neq 0$ then both $\sum_{\omega_x\in\Omega_x} V^{-1}_{\sigma_x,\omega_x}(x) = 0$ and $\sum_{\omega_y\in\Omega_y} V^{-1}_{\sigma_y,\omega_y}(y) = 0$, so we always get 0 for the right-hand side of (3) by taking the last two sums in the appropriate order.

It follows that if x and y are conditionally independent, then $J_\sigma = 0$ unless $\sigma_x\sigma_y = 0$. This implies that all bonds between vertices in $\{x : \sigma_x \neq 0\}$ belong to $\mathcal{G}$, and hence $\{x : \sigma_x \neq 0\}$ is contained in a clique of $\mathcal{G}$. Therefore, only $\sigma \in \widetilde{\Omega}(\mathcal{B})$ appear in (2), and the result is proven. □

Proof of Theorem 3.1. One direction is proven by Lemma 4.1. For the converse, if P is $\mathcal{B}$-$\varnothing$-Gibbs then $P(\omega) = 1/Z_\phi(e^{\sum_{A\in\mathcal{C}\ell(\mathcal{B})}\phi(\omega_A)})$ and if $A, B \subseteq \Lambda$ are two disjoint sets which are not neighbors on $\mathcal{G}$ then the probability factorizes, hence $A \perp\!\!\!\perp B|(\Lambda \setminus A \cup B)$. □

The interaction thus identified is unique, except for the value of $J_{\sigma(0)}$, where $\sigma(0)$ denotes the configuration such that $(\sigma(0))_x = 0$ for all $x \in \Lambda$.

Lemma 4.2. *If* $P(\omega) = \dfrac{1}{Z}e^{\sum_{\sigma\in\widetilde{\Omega}} J_\sigma\omega^\sigma}$, *then*

$$\sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\log P(\omega) = J_\sigma \textit{ for all } \sigma \neq \sigma(0).$$

Proof. For $\sigma \neq \sigma(0)$,

$$\sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\log P(\omega)$$
$$= \sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\Big(\log e^{\sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\omega^{\bar\sigma}} - \log Z\Big) = \sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\Big(\sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\omega^{\bar\sigma} - \log Z\Big)$$
$$= \sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\omega^{\bar\sigma} - \sum_{\omega\in\Omega} V^{-1}_{\sigma,\omega}\log Z$$
$$= \sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\sum_{\omega\in\Omega}\prod_{x\in\Lambda} V^{-1}_{\sigma_x,\omega_x}(x)\omega_x^{\bar\sigma_x} - (\log Z)\,\delta_{\sigma=\sigma(0)}$$
$$= \sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\prod_{x\in\Lambda}\sum_{\omega_x\in\Omega_x} V^{-1}_{\sigma_x,\omega_x}(x)\omega_x^{\bar\sigma_x} = \sum_{\bar\sigma\in\widetilde{\Omega}} J_{\bar\sigma}\prod_{x\in\Lambda}\delta_{\sigma_x=\bar\sigma_x} = \sum_{\bar\sigma\in\widetilde{\Omega}}\delta_{\sigma=\bar\sigma}J_{\bar\sigma} = J_\sigma. \quad □$$

5. Markov–Gibbs with hard-core constraints

We go back to our Question formulated on page 412 and answer it.

For a probability P on some $(\Lambda, \mathcal{P}(\Lambda))$, $\hat{P}$ is a strictly positive extension of P if $\hat{P}(\omega) > 0$ for all $\omega \in \Omega$, and

$$\hat{P}(\omega) = \frac{1}{\hat{Z}} P(\omega), \tag{4}$$

for some constant $\hat{Z}$, for all $\omega \in \Omega$ for which $P(\omega) > 0$.

Lemma 5.1. *On a graph $\mathcal{G} = (\Lambda, \mathcal{B})$, a probability P is $\mathcal{B}$-Λ-Gibbs if and only if it has a $\mathcal{B}$-global-Markov strictly positive extension $\hat{P}$.*

Proof. If P is $\mathcal{B}$-Λ-Gibbs, then define $\hat{P}(\omega) := 1/(Z(\hat{P}))e^{\sum_{A \in \mathcal{C}\ell(\mathcal{B})} \phi(\omega_A)}$, for a suitable constant $Z(\hat{P})$, which is $\mathcal{B}$-$\varnothing$-Gibbs (and hence $\mathcal{B}$-global-Markov). Moreover, $\hat{P}$ is strictly positive; if $P(\omega) > 0$ then $P(\omega) = 1/Z_\phi(e^{\sum_{A \in \mathcal{C}\ell(\mathcal{B})} \phi(\omega_A)}) = (Z(\hat{P})/Z_\phi)\hat{P}(\omega)$ so that (4) holds with $\hat{Z} = Z(\hat{P})/Z_\phi$.

Vice versa, if P has a $\mathcal{B}$-global-Markov strictly positive extension $\hat{P}$, then by the Hammersley–Clifford Theorem, $\hat{P}$ is a $\mathcal{B}$-$\varnothing$-Gibbs random field, i.e., $\hat{P}(\omega) = 1/(Z(\hat{P}))(e^{\sum_{A \in \mathcal{C}\ell(\mathcal{B})} \phi(\omega_A)})$ for some suitable ϕ. By (4),

$$P(\omega) = \hat{Z} \frac{1}{Z(\hat{P})} \left(e^{\sum_{A \in \mathcal{C}\ell(\mathcal{B})} \phi(\omega_A)} \prod_{B \in \mathcal{F}} \mathbb{1}_{\Omega_B \setminus \overline{\Omega}_B}(\omega_B) \right),$$

which is $\mathcal{B}$-Λ-Gibbs with $Z_\phi = Z(\hat{P})/\hat{Z}$. $\square$

Our Question has a negative answer.

Lemma 5.2. *There is a graph $\mathcal{G} = (\Lambda, \mathcal{B})$ and a $\mathcal{B}$-global-Markov random field* (*with hard-core constraints*) *which is not $\mathcal{B}$-Λ-Gibbs.*

Proof. In fact, we can take the same graph as Moussouris, with bond set $\mathcal{B}_2$. As support of the probability we take

$$\Omega'' = \{(1, 1, 1, 1), (0, 1, 1, 1), (1, 0, 1, 1), (0, 0, 1, 1), \\ (1, 1, 0, 0), (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 0, 0)\};$$

hence, we assume that all configurations in $\overline{\Omega} = \Omega \setminus \Omega''$ have zero probability: this is a $\mathcal{F} = \Lambda$ (or possibly a subgraph, see the next section below) constraint. This time, however, the probability P_* is taken as follows: $P_*(1, 1, 1, 1) = \frac{2}{9}$, and $P_*(\omega) = \frac{1}{9}$ for all other $\omega \in \Omega'' \setminus \{(1, 1, 1, 1)\}$.

We first show that

$$\{1\} \perp\!\!\!\perp \{3\} \mid \{2, 4\} \quad \text{and} \quad \{2\} \perp\!\!\!\perp \{4\} \mid \{1, 3\}.$$

For each $x, y \in \Lambda$, if $\bar{\omega} = \omega_{\Lambda \setminus \{x,y\}}$ has nonzero probability, then $x \perp\!\!\!\perp y \mid \Lambda \setminus \{x, y\}$ under a probability P is equivalent to

$$\begin{aligned} P(\omega_x = 1, \omega_y = 1, \omega_{\Lambda \setminus \{x,y\}} = \bar{\omega}) P(\omega_x = 0, \omega_y = 0, \omega_{\Lambda \setminus \{x,y\}} = \bar{\omega}) \\ = P(\omega_x = 1, \omega_y = 0, \omega_{\Lambda \setminus \{x,y\}} = \bar{\omega}) P(\omega_x = 0, \omega_y = 1, \omega_{\Lambda \setminus \{x,y\}} = \bar{\omega}), \end{aligned}$$

as easily seen by elementary calculations [Moussouris 1974; Lauritzen 1996]. To verify the claimed conditional independencies we the have to verify the following equalities [Lauritzen 1996; Geiger et al. 2006], in which $x = 1,\ y = 3$ in the first four equalities, and $x = 2,\ y = 4$ in the others:

$$\begin{aligned} (I) \quad & P(1,1,1,1)P(0,1,0,1) = P(0,1,1,1)P(1,1,0,1), \\ (II) \quad & P(1,0,1,1)P(0,0,0,1) = P(0,0,1,1)P(1,0,0,1), \\ (III) \quad & P(1,1,1,0)P(0,1,0,0) = P(0,1,1,0)P(1,1,0,0), \\ (IV) \quad & P(1,0,1,0)P(0,0,0,0) = P(0,0,1,0)P(1,0,0,0), \\ (V) \quad & P(1,1,1,1)P(1,0,1,0) = P(1,0,1,1)P(1,1,1,0), \\ (VI) \quad & P(0,1,1,1)P(0,0,1,0) = P(0,0,1,1)P(0,1,1,0), \\ (VII) \quad & P(1,1,0,1)P(1,0,0,0) = P(1,0,0,1)P(1,1,0,0), \\ (VIII) \quad & P(0,1,0,1)P(0,0,0,0) = P(0,0,0,1)P(0,1,0,0), \end{aligned} \tag{5}$$

with none having four zero values, so that the condition $P(\bar{\omega}) > 0$ is valid. These relations are easily seen to hold for P_*, as in each row there is exactly one configuration in $\overline{\Omega}$ on the right-hand side and one on the left-hand side of the equality. So, P_* is $\mathcal{B}_2$-pair-Markov; since on this graph the two notions coincide, P_* is also $\mathcal{B}_2$-global-Markov.

We now verify that P_* does not admit a $\mathcal{B}_2$-global-Markov strictly positive extension so that it cannot be $\mathcal{B}_2$-Λ-Gibbs. If such extension $\hat{P}$ existed, then all the above equalities would have to hold for $\hat{P}$ as well, as it would have to be $\mathcal{B}_2$-global-Markov; since $\hat{P}$ is an extension of P_*, it would have to be $\hat{P}(\omega)/\hat{P}(\omega') = P_*(\omega)/P_*(\omega')$ for all $\omega, \omega' \in \Omega''$. From the first equality above we would have

$$\frac{\hat{P}(0,1,0,1)}{\hat{P}(1,1,0,1)} = \frac{\hat{P}(0,1,1,1)}{\hat{P}(1,1,1,1)} = \frac{P_*(0,1,1,1)}{P_*(1,1,1,1)} = \frac{1}{2}. \tag{6}$$

On the other hand, from the seventh equality

$$\frac{\hat{P}(1,1,0,1)}{\hat{P}(1,0,0,1)} = \frac{\hat{P}(1,1,0,0)}{\hat{P}(1,0,0,0)} = \frac{P_*(1,1,0,0)}{P_*(1,0,0,0)} = 1, \tag{7}$$

or $\hat{P}(1,1,0,1) = \hat{P}(1,0,0,1)$; the second and eighth equality would give

$$\hat{P}(1,0,0,1) = \hat{P}(0,0,0,1) \quad \text{and} \quad \hat{P}(0,0,0,1) = \hat{P}(0,1,0,1),$$

respectively. Hence, it would be $\hat{P}(1, 1, 0, 1) = \hat{P}(0, 1, 0, 1)$, which would be a contradiction with (6) above. $\square$

6. Examples of minimal graphs

Given a graph $\mathcal{G} = (\Lambda, \mathcal{B})$, it would be interesting to classify $\mathcal{B}$-global-Markov or $\mathcal{B}$-pair-Markov random fields P in terms of the minimal graphs $\mathcal{B}' = \mathcal{B}'(\Omega, \mathcal{B}, P)$ and $\mathcal{B}'' = \mathcal{B}''(\Omega, \mathcal{B}, P)$ so that P is $\mathcal{B}'$-$\mathcal{B}''$-Gibbs; this problem can be given an explicit algebraic form, following the lines of [Geiger et al. 2006]. We have not been able to develop relevant results in this direction though, and therefore we limit ourselves to a review of the previous examples from this point of view.

Example 6.1 (minimal graphs in Moussouris' example). $\overline{\Omega} = \Omega \setminus \Omega'$ is measurable with respect to the σ-algebra generated by the cylinders

$$[10 * 1],\ [01 * 0],\ [*101],\ [*010],$$

where we use the previously introduced notation for cylinders. Therefore, a minimal collection $\mathcal{F}$ is $\{1, 2, 4\}$, $\{2, 3, 4\}$, which corresponds to the graph with bonds $\mathcal{B}_3 = \mathcal{B}_2 \cup \{\{2, 4\}\}$. As we already observed, to define the uniform probability we can take bonds $\mathcal{B}_0 = \varnothing$; hence, $P_{\mathcal{M}}$ is $\mathcal{B}_2$-global-Markov and $\mathcal{B}_0$-$\mathcal{B}_3$-Gibbs.

We see now that even changing the probability in Moussouris' example would not have yielded a counterexample to our Question.

Example 6.2 (minimal graphs in a modified Moussouris example). Consider the same graph with bonds $\mathcal{B}_2$ and Ω' as in Moussouris' example, but with any probability P strictly positive on Ω'. We can construct a $\mathcal{B}_2$-global-Markov strictly positive extension $\hat{P}$ as follows. Start from some configuration ([0101] for instance), let $\hat{P}(0, 1, 0, 1) = c$, and notice that the conditions for $\mathcal{B}$-global-Markov are those in (5). From relation (I), any $\mathcal{B}_2$-global-Markov strictly positive extension has

$$\hat{P}(1, 1, 0, 1) = \frac{P(1, 1, 1, 1)}{P(0, 1, 1, 1)} c;$$

next, from equality (VII), one gets

$$\hat{P}(1, 0, 0, 1) = \frac{P(1, 0, 0, 0)}{P(1, 1, 0, 0)} \hat{P}(1, 1, 0, 1) = \frac{P(1, 0, 0, 0)}{P(1, 1, 0, 0)} \frac{P(1, 1, 1, 1)}{P(0, 1, 1, 1)} c;$$

recursively, we get all probabilities as function of c, and finally c from normalization. This generates an extension $\hat{P}$ of P. We know $\hat{P}$ is strictly positive as P was strictly positive on Ω' and the above operations preserve positivity; furthermore, it is $\mathcal{B}_2$-global-Markov as the relations in (5) are all valid for $\hat{P}$ as well. As such, the Hammersley–Clifford theorem applies to $\hat{P}$, which is then $\mathcal{B}_2$-$\varnothing$-Gibbs. We have seen in Example 6.1 that the constraints are generated by the graphs with bonds $\mathcal{B}_3$;

therefore, any absolutely continuous modification P of Moussouris' example is $\mathcal{B}_2$-$\mathcal{B}_3$-Gibbs.

Example 6.3 (minimal graphs for P_*). The configurations in $\overline{\Omega} = \Omega \setminus \Omega''$ are generated by the cylinders

$$[**10],\ [**01];$$

hence, a minimal collection $\mathcal{F}$ is $\{3, 4\}$, which corresponds to the graph with bonds $\mathcal{B}_2' = \{\{3, 4\}\}$.

Next, observe that is not possible to express the probability in terms of a $\mathcal{B}_2$-Markov probability, as otherwise P_* would be $\mathcal{B}_2$-$\mathcal{B}_2$-Gibbs, and we know from Lemma 5.2 that it is not. On the other hand, as on $\mathcal{B}_3$ pairwise Markov is the same as global Markov, and $\mathcal{B}_3$ produces a separable graph, the result in [Lauritzen 1996] implies that P_* is $\mathcal{B}_3$-$\mathcal{B}_3$-Gibbs. Hence, P_* is certainly factorizable on $\mathcal{B}_3$. A simple calculation shows that all such factorizations have interactions

$$J_{\{3\}} = x_3, \qquad J_{\{4\}} = x_4, \qquad J_{\{3,4\}} = -(x_3 + x_4),$$
$$J_{\{4,1\}} = \log \tfrac{3}{2}, \quad J_{\{1,2,3\}} = \log 2, \quad J_{\{3,4,1\}} = \log \tfrac{2}{3},$$

and $J_A = 0$ for all other sets A. Therefore, taking

$$\mathcal{B}_3' = \{\{3\}, \{4\}, \{3, 4\}, \{4, 1\}, \{1, 2, 3\}, \{3, 4, 1\}\},$$

we have that P_* is $\mathcal{B}_3'$-$\mathcal{B}_3$-Gibbs. As we have already noticed that the hard-core constraints are generated by $\mathcal{B}_2'$, we have that P_* is $\mathcal{B}_3'$-$\mathcal{B}_2'$-Gibbs.

Back to the question and notation at the beginning of this section, it is not easy at this point to elaborate on the relationship between $\mathcal{B}$ and $\mathcal{B}'$ and $\mathcal{B}''$. For instance, in all the previous examples, one of the two graphs $\mathcal{B}'$ or $\mathcal{B}''$ was always contained in $\mathcal{B}$; but not even this holds in general.

Example 6.4. Consider Λ as in the examples above, and identical copies $\Lambda(1)$ and $\Lambda(2)$; in each copy consider a copy $\mathcal{B}_2(1)$ and $\mathcal{B}_2(2)$, respectively, of the edges in $\mathcal{B}_2$, in each copy between the appropriate vertices. The graph we consider is then $\mathcal{G} = (\Lambda(1) \cup \Lambda(2), \mathcal{B}_2(1) \cup \mathcal{B}_2(2))$. The configuration space is $\Omega = \Omega(1) \times \Omega(2)$, where $\Omega(i)$ is the copy over $\Lambda(i)$ of the configurations of $\{0, 1\}^\Lambda$. The probability P is taken to be the product $P = P_{\mathcal{M}} \times P_*$. It is easily seen that P is $(\mathcal{B}_2(1) \cup \mathcal{B}_2(2))$-global-Markov. On the other hand, the hard-core constraints are generated by cylinders in Examples 6.1 and 6.3 in the two copies, so that the graph of hard-core conditions is $\mathcal{B}_3'' = \mathcal{B}_3(1) \cup \mathcal{B}_2'(2)$. However, the interactions are generated by the graph $\mathcal{B}_3''' = \mathcal{B}_0(1) \cup \mathcal{B}_3'(2)$. So, altogether, P is $(\mathcal{B}_2(1) \cup \mathcal{B}_2(2))$-global-Markov and $\mathcal{B}_3''$- $\mathcal{B}_3'''$-Gibbs, but neither $\mathcal{B}_3''$ nor $\mathcal{B}_3'''$ are contained in $\mathcal{B}_2(1) \cup \mathcal{B}_2(2)$.

7. Identifiability of statistical models by support, moments and conditional independencies

As an application of our results, we turn to an identifiability problem in statistics. Suppose that of a discrete random field P on the configuration space Ω_Λ, for some finite set Λ, we have observations that determine the *support of the distribution*, the pairwise (or global) conditional independencies, leading to a *dependency graph* $\mathcal{G} = (\Lambda, \mathcal{B})$ with bonds $\mathcal{B}$, and finally the *collection of moments* determined by the cliques of $\mathcal{B}$, that is, all moments

$$m_\sigma = E_P(\omega^\sigma) \tag{8}$$

where $\sigma \in \tilde{\Omega}$ is such that $\{x : \sigma_x \neq 0\}$ is contained in a clique of $\mathcal{B}$. If (8) holds then we say that P *satisfies the* $\mathcal{B}$*-moments.*

We start by combining the results of Sections 4 and 5 to show that these are, in general, sufficient statistics to determine the distribution. Later we verify, however, that the counterexample of Lemma 5.2 leads to exceptional cases in which identifiability by the above statistics breaks down.

The result about identifiability is given in two steps. First, we assume that P is known to $\mathcal{B}$-$\mathcal{F}$-Gibbs for some $\mathcal{F}$.

Lemma 7.1. *Given* $\mathcal{B}$, $\mathcal{F}$, *the* $\overline{\Omega}_A$ *for* $A \in \mathcal{F}$, *and* $\mathcal{B}$*-moments* m_σ, *there is at most one* $\mathcal{B}$-$\mathcal{F}$*-Gibbs random field* P *satisfying the* $\mathcal{B}$*-moments.*

Proof. Suppose there are two $\mathcal{B}$-$\mathcal{F}$-Gibbs random fields P and P' satisfying the $\mathcal{B}$-moments. Each has a $\mathcal{B}$-global-Markov strictly positive extension by Lemma 5.1, which can be expressed as in Lemma 4.1 with interactions J and J', respectively. Now consider

$$f(t) = E_{P(t)}\bigg(\sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} \omega^\sigma (J'_\sigma - J_\sigma)\bigg),$$

where $P(t) = \frac{1}{Z} e^{\sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} \omega^\sigma (J_\sigma + t(J'_\sigma - J_\sigma))}$ for $t \in [0, 1]$. We have

$$f(0) = \sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} E_{P(0)}(\omega^\sigma)(J'_\sigma - J_\sigma) = \sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} E_{P(1)}(\omega^\sigma)(J'_\sigma - J_\sigma) = f(1)$$

by equality of $\mathcal{B}$-moments. Moreover, $f'(t) = \mathrm{Var}_{P_t}\big(\sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} \omega^\sigma (J'_\sigma - J_\sigma)\big) \geq 0$. Combined with $f(0) = f(1)$ this implies $f'(t) = 0$ for all $t \in [0, 1]$. Hence, $0 = f'(0) = \mathrm{Var}_{P_0}\big(\sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} \omega^\sigma (J'_\sigma - J_\sigma)\big) \geq 0$, which implies $\sum_{\sigma\in\tilde{\Omega}(\mathcal{B})} \omega^\sigma (J'_\sigma - J_\sigma) \equiv 0$ and that the two extensions of P and P' coincide. This implies that also the two random fields coincide. □

Theorem 7.2. *Let* $\mathcal{B}$ *and* $\mathcal{B}$*-moments* m_σ *be given.*

(1) *If finite energy holds then there is at most one $\mathcal{B}$-global-Markov random field P satisfying the $\mathcal{B}$-moments*; *in particular*, *P is $\mathcal{B}$-$\varnothing$-Gibbs.*

(2) *If $\mathcal{G} = (\Lambda, \mathcal{B})$ is chordal then there is at most one $\mathcal{B}$-global-Markov random field P satisfying the $\mathcal{B}$-moments*; *in particular*, *P is $\mathcal{B}$-$\mathcal{B}$-Gibbs.*

Proof. (1) If a random field is $\mathcal{B}$-global-Markov and completely positive then it is $\mathcal{B}$-$\varnothing$-Gibbs by Theorem 3.1, and uniqueness follows from Lemma 7.1.

(2) If $\mathcal{G} = (\Lambda, \mathcal{B})$ is chordal then P is $\mathcal{B}$-$\mathcal{B}$-Gibbs by Theorem 3.2, hence uniqueness follows from Lemma 7.1. □

Identifiability can break down. We first observe that in Moussouris' example identifiability still holds, and then show that it does not hold for P_* in Lemma 5.2.

Example 7.3. $\mathcal{B}_2$-moments in Moussouris' example are $E(\omega_i) = \frac{1}{2}, i = 1, \dots, 4$; $E(\omega_i\omega_{i+1}) = \frac{3}{8}, i = 1, \dots, 3$; $E(\omega_4\omega_1) = \frac{1}{8}$. With some algebra one can see that if a random field P satisfies these $\mathcal{B}_2$-moments then

$$\begin{aligned} P(1,0,1,1) = -P(0,0,1,0) - P(0,1,0,0) - P(0,1,0,1) \\ - P(0,1,1,0) - P(1,0,0,1) - P(1,0,1,0) - P(1,1,0,1), \end{aligned}$$

which implies that Ω' is the support of P; some further algebra shows that the other linear relations imply then that P is uniform on Ω', so that $P = P_{\mathcal{M}}$.

Example 7.4. The random field P_* in the proof of Lemma 5.2 has $\mathcal{B}_2$-moments $E(\omega_i) = \frac{5}{9}, i = 1, \dots, 4$; $E(\omega_i\omega_{i+1}) = \frac{3}{9}, i = 1, 2, 4$; $E(\omega_3\omega_4) = \frac{5}{9}$. We see that there is not a unique random field which has the same support Ω'' of P_*, is $\mathcal{B}_2$-global-Markov, and satisfies the above $\mathcal{B}_2$-moments. With a little algebra one can easily verify that every random field P with support in Ω'' and

$$\begin{aligned} &P(1,0,1,1) = \tfrac{1}{3} - \lambda, \quad P(1,1,0,0) = \tfrac{1}{3} - \lambda, \quad P(1,0,0,0) = -\tfrac{1}{9} + \lambda, \\ &P(0,1,1,1) = \tfrac{1}{3} - \lambda, \quad P(0,1,0,0) = -\tfrac{1}{9} + \lambda, \quad P(0,0,1,1) = -\tfrac{1}{9} + \lambda, \\ &\qquad\qquad P(0,0,0,0) = \tfrac{1}{3} - \lambda, \end{aligned}$$

with $\lambda = P(1, 1, 1, 1) \in \left[\frac{1}{9}, \frac{1}{3}\right]$, satisfies the above $\mathcal{B}_2$-moments and is also $\mathcal{B}_2$-global-Markov (as it is absolutely continuous with respect to P_*).

By Lemma 7.1, there is a unique $\mathcal{B}_2$-Λ-Gibbs satisfying the above $\mathcal{B}_2$-moments, namely the one with $\lambda \approx 0.2119$ equal to the real root of $\lambda\left(\lambda - \frac{1}{9}\right)^3 - \left(\frac{1}{3} - \lambda\right)^4 = -9 + 107x - 459x^2 + 729x^3 = 0$; one can get this equation directly by explicitly writing out the conditions for P to be $\mathcal{B}_2$-Λ-Gibbs, or indirectly by noticing that this must be the only value of λ which does not lead to a contradiction in (6) and (7). For all other values of $\lambda \in \left[\frac{1}{9}, \frac{1}{3}\right]$, including $\lambda = \frac{2}{9}$ as in Lemma 5.2, the random fields are not $\mathcal{B}_2$-Λ-Gibbs, so they constitute a family of counterexamples to our Question, all with the same $\mathcal{B}_2$-moments.

Remarks. (1) The last example raises the question (related to that of the minimal graph mentioned in Section 6) of the minimal set of moments which can identify a $\mathcal{B}$-Markov random field with hard-core constraints.

(2) The results above about identifiability of random fields from Markov properties and moments can be interpreted as follows. $\mathcal{B}$-moment conditions identify a simplex of probability measures, as described in detail in [Pitowsky 1989] for the violation of correlation conditions in quantum mechanics. When the $\mathcal{B}$-Markov conditional independencies are added, then the resulting algebraic variety reduces to a point in the interior of the simplex, or to a (possibly nontrivial) variety contained in the boundary of the original simplex. We do not know which additional observations could guarantee uniqueness.

8. Conclusions

We have reviewed the Hammersley–Clifford Theorem, which states the equivalence of Markov and Gibbs random fields when there are no hard-core conditions, giving a more explicit proof than usual with the interaction expressed in terms of spin products.

We then addressed the same problem when there are hard-core constraints. We argued that the hard-core constraints are more naturally represented in terms of a separate graph from that used to determine the interactions; in this respect, the counterexample of Moussouris as well as the subsequent literature on constrained Markov random fields do not address the appropriate issues.

We have shown that even allowing the largest possible graph for the hard-core constraints, there are cases in which it is not possible to restrict the graph of the interactions to the one for the Markov conditional independence.

This, in turn, has opened the question of finding minimal graphs for the hard-core conditions and the interactions given the graph of the Markov conditional independencies. We have not been able to address this issue, but we have provided examples of minimal graphs.

Finally, we discussed the statistical identifiability of a random field in terms of support, conditional independencies and moments, with the last two requirements based on the same graph. Our proof of the Hammersley–Clifford theorem allows us to easily show identifiability if finite energy is ascertained or the graph is chordal; while our counterexample allows us to exhibit a case in which support, conditional independencies and moments do not uniquely determine the random field.

References

[Besag 1974] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Roy. Statist. Soc. Ser. B* **36** (1974), 192–236.

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

[Geiger et al. 2006] D. Geiger, C. Meek, and B. Sturmfels, "On the toric algebra of graphical models", *Ann. Stat.* **34**:3 (2006), 1463–1492.

[Grimmett 1973] G. R. Grimmett, "A theorem about random fields", *Bull. Lond. Math. Soc.* **5** (1973), 81–84.

[Grimmett 2010] G. R. Grimmett, *Probability on graphs: random processes on graphs and lattices*, Institute of Mathematical Statistics Textbooks **1**, Cambridge University Press, 2010.

[Hammmersley and Clifford 1971] J. M. Hammmersley and P. Clifford, "Markov fields on finite graphs and lattices", 1971, www.statslab.cam.ac.uk/∼grg/books/hammfest/hamm-cliff.pdf.

[Lauritzen 1996] S. L. Lauritzen, *Graphical models*, Oxford Statistical Science Series **17**, Clarendon Press, New York, 1996.

[Macon and Spitzbart 1958] N. Macon and A. Spitzbart, "Inverses of Vandermonde matrices", *Amer. Math. Monthly* **65** (1958), 95–100.

[Moussouris 1974] J. Moussouris, "Gibbs and Markov random systems with constraints", *J. Stat. Phys.* **10** (1974), 11–33.

[Onural 2016] L. Onural, "Gibbs random fields and Markov random fields with constraints", preprint, 2016. arXiv 1603.01481

[Pitowsky 1989] I. Pitowsky, *Quantum probability—quantum logic*, Lecture Notes in Physics **321**, Springer, Berlin, 1989.

[Ruelle 1978] D. Ruelle, *Thermodynamic formalism: the mathematical structures of classical equilibrium statistical mechanics*, Encyclopedia of Mathematics and its Applications **5**, Addison-Wesley, Reading, MA, 1978.

Received 14 Sep 2016. Accepted 12 Jan 2017.

ALBERTO GANDOLFI: ag189@nyu.edu
NYU Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates

and

Dipartimento di Matematica e Informatica "Ulisse Dini", Università di Firenze, Viale Morgagni 67/a, 50134 Firenze, Italy

PIETRO LENARDA: pietro.lenarda@imtlucca.it
Multi-scale Analysis of Materials, IMT School for Advanced Studies Lucca, I-55100 Lucca, Italy

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

MEMOCS (ISSN 2325-3444 electronic, 2326-7186 printed) is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell'Aquila, Italy.

Cover image: "Tangle" by © John Horigan; produced using the *Context Free* program (contextfreeart.org).

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
vol. 4 no. 3-4 2016

Special issue in honor of
Lucio Russo

MEMOCS is a journal of the International Research Center for
the Mathematics and Mechanics of Complex Systems
at the Università dell'Aquila, Italy.

2326-7186(2016)4:3;1-A