2nd revision for *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* April 21, 2011

On conditional independence and log-convexity

František Matúš *

Abstract. If conditional independence constraints define a family of positive distributions that is log-convex then this family turns out to be a Markov model over an undirected graph. This is proved for the distributions on products of finite sets and for the regular Gaussian ones. As a consequence, the assertion known as Brook factorization theorem, Hammersley-Clifford theorem or Gibbs-Markov equivalence is obtained.

1. Main result

Let N be a finite set, X_i, $i \in N$, finite nonempty state spaces, X_I the product of X_i over $i \in I$, and π_I the coordinate projection of X_N to X_I, $I \subseteq N$. The marginal of a probability measure (pm) P on X_N to I is the pm $\pi_I P$ on X_I given by

$$\pi_I P(\pi_I x) = \textstyle\sum_{y \in X} P(y)\, \delta^I_{x,y}\,, \qquad x \in X_N\,,$$

where $\delta^I_{x,y}$ equals one if $\pi_I x = \pi_I y$ and zero otherwise.

A pm P on X_N satisfies a *conditional independence* (*CI*-) constraint if

$$(1) \qquad \pi_{ijK} P(\pi_{ijK} x) \cdot \pi_K P(\pi_K x) = \pi_{iK} P(\pi_{iK} x) \cdot \pi_{jK} P(\pi_{jK} x)\,, \qquad x \in X_N\,,$$

where $i, j \in N$ are different and $K \subseteq N \setminus ij$. By convention, an element $i \in N$ is not distinguished from the singleton $\{i\}$ and the sign $\cup$ for unions of subsets of N is omitted.

The family of all ordered couples $(ij|K)$ with i, j and K as above is denoted by $\mathcal{R}$. For $\mathcal{L} \subseteq \mathcal{R}$ let $\mathcal{P}_\mathcal{L}$ denote the family of pm's on X_N that satisfy the *CI*-constraint given by each $(ij|K) \in \mathcal{L}$ and $\mathcal{P}^+_\mathcal{L} = \mathcal{P}^+ \cap \mathcal{P}_\mathcal{L}$. Here, $\mathcal{P}^+$ is the family of pm's on X_N that are positive in the sense $P(x) > 0$ for all $x \in X_N$. Given a pm P, the projections π_i, $i \in N$, can be interpreted as the random variables jointly distributed according to P. Their distribution belongs to $\mathcal{P}_\mathcal{L}$ if and only if π_i and π_j are stochastically independent given $(\pi_k)_{k \in K}$ for all $(ij|K) \in \mathcal{L}$.

This work was partially supported by Grant Agency of Academy of Sciences of the Czech Republic, Grant IAA 100750603, and by Grant Agency of the Czech Republic, Grant 201/08/0539.

AMS 2000 Math. Subject Classification. Primary 62H05, 62M40; secondary 62H17, 62J10, 05C50, 11C20, 15A15.

Keywords: Conditional independence, Markov properties, factorizable distributions, graphical Markov models, log-convexity, Gibbs-Markov equivalence, Markov fields, Hammersley-Clifford theorem, contingency tables, Gibbs potentials, multivariate Gaussian distributions, positive definite matrices, covariance selection model.

*F. Matúš is with Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic; matus@utia.cas.cz

A nonempty subfamily of $\mathcal{P}^+$ is *log-convex* if it contains together with pm's P and Q also the pm proportional to $x \mapsto P^t(x)Q^{1-t}(x)$, $x \in X_N$, for all $0 < t < 1$. The general definition of log-convexity is recalled and discussed in Subsection 5.3.

A hypergraph $(N, \mathcal{A})$ consists of the vertex set N and a nonempty family $\mathcal{A}$ of subsets of N, called hyperedges. A pm P on X_N is *factorizable* w.r.t. the hypergraph if for each $I \in \mathcal{A}$ there exists a real-valued function ψ_I on X_I such that

$$P(x) = \prod_{I \in \mathcal{A}} \psi_I(\pi_I x)\,, \qquad x \in X_N\,. \tag{2}$$

The family of all pm's on X_N that factorize w.r.t. the hypergraph is denoted by $\mathcal{F}_\mathcal{A}$ and $\mathcal{F}_\mathcal{A}^+ = \mathcal{F}_\mathcal{A} \cap \mathcal{P}^+$. An undirected graph $G = (N, E)$ has a vertex set N and an edge set E, contained in the family $\binom{N}{2}$ of all two-element subsets of N. A set L of vertices is a clique of G if $\binom{L}{2} \subseteq E$. If $\mathcal{A}$ is the family of cliques of G, the notation $\mathcal{F}_G^+$ is preferred to $\mathcal{F}_\mathcal{A}^+$ and one speaks about the factorization w.r.t. G.

Theorem 1. *If $\mathcal{L} \subseteq \mathcal{R}$ and $\mathcal{P}_\mathcal{L}^+$ is log-convex then this family coincides with $\mathcal{F}_G^+$ where $G = (N, E)$ is the graph with $ij \in E$ if and only if $(ij|K) \notin \mathcal{L}$ for all $K \subseteq N \setminus ij$.*

In other words, given a family $\mathcal{L}$ the graph G is constructed on N by removing from $\binom{N}{2}$ all ij such that $(ij|K) \in \mathcal{L}$ for some $K \subseteq N \setminus ij$; Theorem 1 asserts that the log-convexity of $\mathcal{P}_\mathcal{L}^+$ implies $\mathcal{P}_\mathcal{L}^+ = \mathcal{F}_G^+$.

If $\mathcal{L}$ consists exclusively of the couples $(ij|K)$ having $K = N \setminus ij$ then it is log-convex. This follows easily from a well-known equivalent definition of *CI*-constraints, see (6) below. In this special case Theorem 1 implies what is called by statisticians Brook factorization theorem [5, 16] or Hammersley-Clifford theorem [4, 7] and by physicists Gibbs-Markov equivalence [26, 19].

Corollary 1. *If $G = (N, E)$ is an undirected graph and $\mathcal{L}$ consists of all couples $(ij|N \setminus ij)$ such that ij is not and edge of G then $\mathcal{P}_\mathcal{L}^+ = \mathcal{F}_G^+$.*

A proof of Theorem 1 is deferred to Section 4. It is based on properties of interaction spaces, summarized in Section 2, and new observations on the behaviour of the family $\mathcal{P}_\mathcal{L}$ around the uniform pm, presented in Section 3. Discussion and remarks are collected in Section 5 that contains also a short independent proof of Corollary 1 of geometric flavor. Section 6 is devoted to a version of Theorem 1 involving regular multidimensional Gaussian pm's, formulated as Theorem 2.

2. Interaction spaces and factorizability

The factorization (2) of positive pm's w.r.t. a hypergraph can be equivalently described on the exponential scale by a linear space. The space decomposes orthogonally to interaction spaces. The material of this section is standard, see [9] and [20, Appendix B.2]. The aim is to prepare arguments used in Sections 3 and 4, and supply self-contained proofs.

For $I \subseteq N$ let $\mathcal{V}_I$ denote the linear space of those functions v of $x \in X_N$ that depend on x only through $\pi_I x$, thus $v(x) = v(y)$ once $\pi_I x = \pi_I y$, $x, y \in X_N$. For $v \in \mathcal{V}_I$

$$\pi_I v(\pi_I x) = |X_{N \setminus I}| v(x)\,, \qquad x \in X_N\,, \tag{3}$$

where the marginalization of functions on X_N is defined analogously to that of pm's, thus $\pi_I v(\pi_I x)$ is the sum of $v(y)$ over $y \in X_N$ satisfying $\pi_I x = \pi_I y$.

Lemma 1. *A function u on X_N is orthogonal to $\mathcal{V}_J$, $J \subseteq N$, if and only if $\pi_J u = 0$.*

Proof. A function v belongs to $\mathcal{V}_J$ if and only if it can be written as the composition $v = v_J \pi_J$ where $v_J \colon X_J \to \mathbb{R}$. Since

$$\langle u, v\rangle = \textstyle\sum_{x \in X_N} u(x) \cdot v(x) = \sum_{x_J \in X_J} \pi_J u(x_J) \cdot v_J(x_J) = \langle \pi_J u, v_J\rangle$$

and v_J is arbitrary the assertion follows. ☐

For $I \subseteq N$ let $\mathcal{U}_I$ denote the orthogonal complement in $\mathcal{V}_I$ to the sum of $\mathcal{V}_J$ over $J \subsetneq I$, and be referred to as an *interaction space.*

Lemma 2. *If $J \subseteq N$ does not contain $I \subseteq N$ and $u \in \mathcal{U}_I$, then $\pi_J u = 0$.*

Proof. Since $\mathcal{U}_I \subseteq \mathcal{V}_I$ the marginal $\pi_J u$ of $u \in \mathcal{U}_I$ to J depends on $x_J \in X_J$ only through $\pi^J_{I\cap J} x_J$ where $\pi^J_{I\cap J}$ denotes the coordinate projection of X_J to $X_{I\cap J}$. It follows from (3), applied to $\pi_J u$ and $I \cap J$ in the roles of v and I, that $\pi^J_{I\cap J}\pi_J u(\pi^J_{I\cap J} x_J)$ equals $|X_{J\setminus I}|\pi_J u(x_J)$ for $x_J \in X_J$. Therefore, it suffices to prove that the marginal $\pi^J_{I\cap J}\pi_J u = \pi_{I\cap J} u$ of $u \in \mathcal{U}_I$ vanishes. By Lemma 1, this is equivalent to the orthogonality of u and $\mathcal{V}_{I\cap J}$ which holds by the definition of $\mathcal{U}_I$ and $I \cap J \neq I$. ☐

Lemma 3. *The spaces $\mathcal{U}_I$, $I \subseteq N$, are pairwise orthogonal.*

Proof. If $I, J \subseteq N$ are different then, up to symmetry, J does not contain I. By Lemma 2, if $u \in \mathcal{U}_I$ then $\pi_J u = 0$. Lemma 1 implies that u is orthogonal to $\mathcal{V}_J$. Thus, $\mathcal{U}_I$ and $\mathcal{U}_J$ are orthogonal. ☐

For a hypergraph $(N, \mathcal{A})$, let $\mathcal{W}_\mathcal{A}$ denote the sum of $\mathcal{V}_I$ over $I \in \mathcal{A}$.

Lemma 4. *For a hypergraph $(N, \mathcal{A})$ the space $\mathcal{W}_\mathcal{A}$ equals the orthogonal sum of $\mathcal{U}_I$ over all $I \subseteq N$ that are covered by some hyperedge from $\mathcal{A}$.*

Proof. By Lemma 3, the assertion follows from its special instance disregarding orthogonality. On account of the definition of $\mathcal{W}_\mathcal{A}$, it suffices to restrict to the hypergraphs with $\mathcal{A} = \{I\}$, $I \subseteq N$, in which case $\mathcal{W}_\mathcal{A} = \mathcal{V}_I$. Induction on the cardinality of I is employed to prove that $\mathcal{V}_I$ is the sum of $\mathcal{U}_J$ over $J \subseteq I$. If $I = \emptyset$ then $\mathcal{V}_\emptyset$ and $\mathcal{U}_\emptyset$ coincide with the space of constant functions on X_N and the assertion is trivial. Otherwise, $\mathcal{V}_I$ decomposes orthogonally into $\mathcal{U}_I$ and the sum of $\mathcal{V}_J$ over $J \subsetneq I$. By the induction assumption, this sum equals the sum of $\mathcal{U}_J$ over $J \subsetneq I$ whence the induction step is completed. ☐

Corollary 2. *The Euclidean space $\mathcal{V}_N = \mathbb{R}^{X_N}$ decomposes orthogonally to $\mathcal{U}_I$, $I \subseteq N$.*

Lemma 5. *For every hypergraph $(N, \mathcal{A})$ the family $\mathcal{F}^+_\mathcal{A}$ consists of the pm's that are proportional to e^w where w belongs to the orthogonal sum of $\mathcal{U}_I$ over all $I \subseteq N$ that are covered by some hyperedge from $\mathcal{A}$.*

Proof. If w belongs to the above sum of $\mathcal{U}_I$ then it is in the sum of $\mathcal{V}_I$ over $I \in \mathcal{A}$. Writing w as the sum of functions $v_I \pi_I$ where $v_I \colon X_I \to \mathbb{R}$, a pm proportional to e^w is given by $x \mapsto t \exp[\Sigma_{I \in \mathcal{A}} v_I(\pi_I x)]$, $x \in X_N$, with some constant $t > 0$. Such a pm belongs to $\mathcal{F}^+_{\mathcal{A}}$, choosing ψ_I in (2) proportional to e^{v_I}.

If $P \in \mathcal{F}^+_{\mathcal{A}}$ then, by (2), positive functions ψ_I exist such that $P(x) = e^{w(x)}$, $x \in X_N$, where w is the sum over $I \in \mathcal{A}$ of the functions $x \mapsto \ln \psi_I(\pi_I x)$. Thus, P is proportional to e^w, with the proportionality constant equal to 1. By definition, w belongs to $\mathcal{W}_{\mathcal{A}}$, equal to the above sum of $\mathcal{U}_I$ by Lemma 4. □

Corresponding to the interaction spaces $\mathcal{U}_I$, a special base α_y, $y \in X_N$, of the space $\mathbb{R}^{X_N}$ is constructed. To this end, it is assumed that each X_i has a distinguished element denoted by $\mathit{0}_i$ and $\mathit{0} = (\mathit{0}_i)_{i \in N}$. The function α_y is defined at $x \in X_N$ by

$$\alpha_y(x) = \begin{cases} (-1)^{|\mathsf{s}(y) \cap \mathsf{s}(x)|}, & x \sim y, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathsf{s}(y)$ denotes the support $\{i \in N \colon y_i \neq \mathit{0}_i\}$ of y and $x \sim y$ abbreviates the equality of projections of x and y onto $\mathsf{s}(x) \cap \mathsf{s}(y)$.

Lemma 6. *If $I \subseteq N$ then $\{\alpha_y \colon \mathsf{s}(y) \subseteq I\}$ is a base of $\mathcal{V}_I$ and $\{\alpha_y \colon \mathsf{s}(y) = I\}$ a base of $\mathcal{U}_I$.*

Proof. For $y, z \in X_N$ the scalar product of α_y and α_z equals the sum of $(-1)^{|[\mathsf{s}(y) \triangle \mathsf{s}(z)] \cap \mathsf{s}(x)|}$ over $x \in X_N$ such that $x \sim y$ and $x \sim z$. For $i \in \mathsf{s}(y) \setminus \mathsf{s}(z)$ the range of summation is partitioned into the pairs that differ only in the i-th coordinate, belonging to $\{\mathit{0}_i, y_i\}$. By the summations over the pairs the scalar product vanishes. Therefore, α_y and α_z are orthogonal when $\mathsf{s}(y) \neq \mathsf{s}(z)$. If y and z have the same support then $\alpha_y(z) = (-1)^{|\mathsf{s}(y)|} \delta^N_{y,z}$, and thus the functions α_y with $\mathsf{s}(y) = I$ are independent. It follows that $\{\alpha_y \colon \mathsf{s}(y) \subseteq I\}$ is an independent set. This set is a base of $\mathcal{V}_I$ because $\alpha_y \in \mathcal{V}_{\mathsf{s}(y)} \subseteq \mathcal{V}_I$ and $\dim \mathcal{V}_I = |X_I|$. Then, $\{\alpha_y \colon \mathsf{s}(y) = I\}$ is a base of $\mathcal{U}_I$ by Lemma 4. □

In particular, the space $\mathcal{U}_I$ has a positive dimension if and only if $|X_i| > 1$ for all $i \in I$. If $X_i = \{\mathit{0}_i, \mathit{1}_i\}$, $i \in N$, then each $\mathcal{U}_I$ is spanned by the single function α_y where $y = (y_i)_{i \in N}$ has y_i equal to $\mathit{1}_i$ for $i \in I$ and $\mathit{0}_i$ otherwise. These functions form an orthogonal base of $\mathbb{R}^{X_N}$ by Corollary 2.

The following technical lemma is prepared for the discussion in Subsection 5.2.

Lemma 7. *For any hypergraph $(N, \mathcal{A})$ the family $\mathcal{W}_{\mathcal{A}}$ is the direct sum of the spaces*

$$\mathcal{V}_{\mathit{0},I} = \{v \in \mathcal{V}_I \colon v(x) = 0 \text{ once } x_i = \mathit{0}_i \text{ for some } i \in I\}$$

over $I \subseteq N$ that can be covered by some $J \in \mathcal{A}$.

The space $\mathcal{V}_{\mathit{0},I}$ consists of the functions $v_I \pi_I$ where $v_I \colon X_I \to \mathbb{R}$ vanishes on each $x_I \in X_I$ having a coordinate $x_i = \mathit{0}_i$. Such a function v_I is said to be *adapted* to $\mathit{0}$.

Proof. For $I \subseteq N$ let ρ_I map $x \in X_N$ to $y \in X_N$ given by $\pi_I y = \pi_I x$ and $\pi_{N\setminus I} y = \pi_{N\setminus I} \mathit{0}$. Let $w \in \mathbb{R}^{X_N}$ be decomposed as the sum of $v_I \pi_I$ over $I \subseteq N$ where every v_I is adapted to $\mathit{0}$. If $x \in X_N$ and $I = \mathsf{s}(x)$ then

$$\begin{aligned} \textstyle\sum_{L\subseteq I} (-1)^{|I\setminus L|}\, w(\rho_L x) &= \textstyle\sum_{L\subseteq I} (-1)^{|I\setminus L|} \sum_{K\subseteq L} v_K(\pi_K \rho_L x) \\ &= \textstyle\sum_{K\subseteq I} v_K(\pi_K x) \sum_{K\subseteq L\subseteq I} (-1)^{|I\setminus L|} = v_I(\pi_I x)\,. \end{aligned} \tag{4}$$

Therefore, the functions v_I are unique and, in turn, the sum of all $\mathcal{V}_{\mathit{0},I}$, $I \subseteq N$, is direct.

If v is a function on X_N then

$$\textstyle\sum_{I\subseteq N} (-1)^{|N\setminus I|}\, v\rho_I \in \mathcal{V}_{\mathit{0},N}\,.$$

In fact, if $x \in X_N$ has a coordinate $x_i = \mathit{0}_i$ then $\rho_I x = \rho_{I\setminus i} x$ for $i \in I \subseteq N$, and grouping the summands into the pairs I, $I \setminus i$ the sum vanishes. Since $v = v\rho_N$ and $v\rho_I \in \mathcal{V}_I$ it follows that $\mathcal{V}_N = \mathbb{R}^{X_N}$ equals $\mathcal{V}_{\mathit{0},N}$ plus the sum of $\mathcal{V}_I$ over $I \subsetneq N$. By induction on the cardinality of N, $\mathcal{V}_N$ is the sum of $\mathcal{V}_{\mathit{0},I}$ over $I \subseteq N$, which implies the assertion. □

3. Conditional independence and interactions

In this section, the *CI*-constraints are related to interaction spaces. This is done via the relation $(ij|K) \diamond L$ between $(ij|K) \in \mathcal{R}$ and $L \subseteq N$ defined by the statement '$i \notin L$ or $j \notin L$ or $L \not\subseteq ijK$'. In other words, $(ij|K) \not\diamond L$ if and only if $ij \subseteq L \subseteq ijK$, see Subsection 5.3. From now on, m_I denotes $|X_I|^{-1}$, $I \subseteq N$.

Lemma 8. *If $(ij|K) \diamond L$ and $u \in \mathcal{U}_L$ then the measure given by $P(x) = m_N[1 + u(x)]$, $x \in X_N$, satisfies the* CI*-constraint given by $(ij|K)$.*

Proof. By Lemma 2, if ijK does not contain L then the marginal of P to ijK is uniform. Thus, the *CI*-constraint (1) reduces to $m_{ijK} m_K = m_{iK} m_{jK}$. Otherwise, $L \subseteq ijK$, and it follows from $(ij|K) \diamond L$ that $i \notin L$ or $j \notin L$. By the symmetry between i and j, it suffices to restrict to the case $i \notin L$. Then, $L \subseteq jK$ which implies that u belongs to $\mathcal{V}_{jK} \subseteq \mathcal{V}_{ijK}$. By the double use of (3) with ijK and jK in the role of I, the constraint (1) rewrites to

$$m_{ijK}\big[1 + u(x)\big] \cdot \pi_K P(\pi_K x) = \pi_{iK} P(\pi_{iK} x) \cdot m_{jK}\big[1 + u(x)\big]\,, \qquad x \in X_N\,. \tag{5}$$

If $j \in L$ then K and iK do not contain L. By Lemma 2, the marginals of P to K and iK are uniform whence (5) holds. If $j \notin L$ then $L \subseteq K$, and thus $u \in \mathcal{V}_K \subseteq \mathcal{V}_{iK}$. By the double use of (3) with K and iK in the role of I, $\pi_K P(\pi_K x) = m_K[1 + u(x)]$ and $\pi_{iK} P(\pi_{iK} x) = m_{iK}[1 + u(x)]$. Hence, (5) is always satisfied. □

For $\mathcal{L} \subseteq \mathcal{R}$ let $\mathcal{L}^\diamond$ denote the family of $L \subseteq N$ such that $(ij|K) \diamond L$ for all $(ij|K) \in \mathcal{L}$.

Corollary 3. *If $\mathcal{L} \subseteq \mathcal{R}$ and $L \in \mathcal{L}^\diamond$ then any pm on X_N that differs from the uniform one by a vector from $\mathcal{U}_L$, as in Lemma 8, belongs to $\mathcal{P}_{\mathcal{L}}$.*

An example of the construction $\mathcal{L} \mapsto \mathcal{L}^\diamond$ can be obtained from a graph $G = (N, E)$ and $\mathcal{K}_G = \{(ij|N \setminus ij) \colon ij \in \binom{N}{2} \setminus E\}$, arriving at the family $\mathcal{K}_G^\diamond$ of cliques of G.

For a later reference the following simple assertion is needed.

Lemma 9. *For any $\mathcal{L} \subseteq \mathcal{R}$, if $M \subseteq N$ and every different $i, j \in M$ can be covered by some $L_{ij} \in \mathcal{L}^\diamond$ contained in M then $M \in \mathcal{L}^\diamond$.*

Proof. If a couple $(ij|K) \in \mathcal{L}$ has $i, j \in M$ then, by the assumption, $ij \subseteq L_{ij} \subseteq M$ for some $L_{ij} \in \mathcal{L}^\diamond$. The definition of $\mathcal{L}^\diamond$ implies that $(ij|K) \diamond L_{ij}$, and thus $L_{ij} \not\subseteq ijK$ because $ij \subseteq L_{ij}$. Hence, $M \not\subseteq ijK$. It follows that $(ij|K) \diamond M$. In turn, $M \in \mathcal{L}^\diamond$. ☐

In the remaining part of this section, the *CI*-constraints (1) are analyzed via the mappings $\Psi_{ij|K}$ given by

$$\Psi_{ij|K}(w) = \big(\pi_{ijK} w(\pi_{ijK} x) \cdot \pi_K w(\pi_K x) - \pi_{iK} w(\pi_{iK} x) \cdot \pi_{jK} w(\pi_{jK} x)\big)_{x \in X_N} .$$

Here, $(ij|K) \in \mathcal{R}$ and w is a function on X_N. A pm can play the role of w as well.

Lemma 10. *The Jacobian of $\Psi_{ij|K}$ at the uniform pm on X_N is equal to*

$$\big(m_K \delta^{ijK}_{x,y} + m_{ijK} \delta^{K}_{x,y} - m_{jK} \delta^{iK}_{x,y} - m_{iK} \delta^{jK}_{x,y} \big)_{x,y \in X_N} .$$

Proof. The coordinate function of $\Psi_{ij|K}$ indexed by $x \in X_N$ equals

$$\textstyle\sum_{z \in X_N} \sum_{\underline{z} \in X_N} w(z) w(\underline{z}) \big[\delta^{ijK}_{x,z} \delta^{K}_{x,\underline{z}} - \delta^{iK}_{x,z} \delta^{jK}_{x,\underline{z}} \big] .$$

Differentiating w.r.t. $w(y)$, $y \in X_N$,

$$\textstyle\sum_{z \in X_N} \sum_{\underline{z} \in X_N} \big[\delta^{N}_{y,z} w(\underline{z}) + w(z) \delta^{N}_{\underline{z},y} \big] \big[\delta^{ijK}_{x,z} \delta^{K}_{x,\underline{z}} - \delta^{iK}_{x,z} \delta^{jK}_{x,\underline{z}} \big] .$$

When $w(z) = m_N$, $z \in X_N$, this equals

$$m_N \textstyle\sum_{\underline{z} \in X_N} \big[\delta^{ijK}_{x,y} \delta^{K}_{x,\underline{z}} - \delta^{iK}_{x,y} \delta^{jK}_{x,\underline{z}} \big] + m_N \sum_{z \in X_N} \big[\delta^{ijK}_{x,z} \delta^{K}_{x,y} - \delta^{iK}_{x,z} \delta^{jK}_{x,y} \big]$$

and the assertion follows. ☐

Let $\mathrm{Ker}_{(ij|K)}$ denote the kernel of the Jacobian of $\Psi_{ij|K}$ at the uniform pm.

Lemma 11. *For $\mathcal{L} \subseteq \mathcal{R}$ the intersection of $\mathrm{Ker}_{(ij|K)}$ over $(ij|K) \in \mathcal{L}$ is equal to the sum of $\mathcal{U}_L$ over $L \in \mathcal{L}^\diamond$.*

Proof. If $L \notin \mathcal{L}^\diamond$ then $ij \subseteq L \subseteq ijK$ for some $(ij|K) \in \mathcal{L}$. For $u \in \mathcal{U}_L$ and $v \in \mathrm{Ker}_{(ij|K)}$

$$\textstyle\sum_{x \in X_N} u(x) \sum_{y \in X_N} \big[m_K \delta^{ijK}_{x,y} + m_{ijK} \delta^{K}_{x,y} - m_{jK} \delta^{iK}_{x,y} - m_{iK} \delta^{jK}_{x,y} \big] v(y) = 0$$

because the inner sums equal zero due to Lemma 10. Since none of the sets K, iK and jK contains L the marginals of u to these sets vanish by Lemma 2 and the above equation rewrites to

$$\textstyle\sum_{x \in X_N} \sum_{y \in X_N} u(x) v(y) \delta^{ijK}_{x,y} = 0 .$$

Since $L \subseteq ijK$ the function u belongs to $\mathcal{V}_{ijK}$, and thus $u(x)\delta^{ijK}_{x,y} = u(y)\delta^{ijK}_{x,y}$. Therefore, u and v are orthogonal. In turn, $\mathrm{Ker}_{(ij|K)}$ is orthogonal to $\mathcal{U}_L$.

By Corollary 2, the intersection of $\mathrm{Ker}_{(ij|K)}$ over $(ij|K) \in \mathcal{L}$ is contained in the sum of $\mathcal{U}_L$ over $L \in \mathcal{L}^\diamond$. The opposite inclusion is a consequence of Corollary 3. ☐

4. Log-convexity and conditional independence

A nonempty family of positive pm's on X_N is *log-linear* if it contains together with pm's P and Q also the pm proportional to $x \mapsto P^t(x)Q^s(x)$, $x \in X_N$, for all real t and s. The family is *log-affine* if this is required only with $s = 1 - t$. The log-convexity assumes the additional restriction $0 < t < 1$. The log-affine families correspond to the full exponential families [20].

In this section, $\mathcal{L} \subseteq \mathcal{R}$.

Lemma 12. *If $\mathcal{P}^+_{\mathcal{L}}$ is log-convex then it is log-linear.*

Proof. Let $P, Q \in \mathcal{P}^+_{\mathcal{L}}$, $0 < t < 1$ and $R_t(x) = P^t(x)Q^{1-t}(x)$, $x \in X$. By the log-convexity, for $(ij|K) \in \mathcal{L}$ eqs. (1) hold with P replaced by R_t. Thus, the coordinate functions of $t \mapsto \Psi_{ij|K}(R_t)$ vanish when t ranges between 0 and 1. Since they are holomorphic they vanish identically. It follows that $\mathcal{P}^+_{\mathcal{L}}$ is closed to the log-affine combinations. The assertion follows because it is not difficult to see that a log-affine family that contains the uniform pm on X_N is log-linear. □

Lemma 13. *If $\mathcal{P}^+_{\mathcal{L}}$ is log-convex and a pm P on X_N is proportional to e^w for some $w \in \mathbb{R}^{X_N}$ then $P \in \mathcal{P}^+_{\mathcal{L}}$ if and only if w belongs to the sum of $\mathcal{U}_L$ over $L \in \mathcal{L}^\diamond$.*

Proof. The log-convex combinations of P with the uniform pm

$$P_t(x) = e^{tw(x)} \Big/ \textstyle\sum_{y \in X_N} e^{tw(y)}\,, \qquad x \in X_N\,,$$

are viewed as a curve parameterized by t. Its tangent vector at the uniform pm P_0 equals

$$m_N w - m_N^2 \textstyle\sum_{y \in X_N} w(y)\,.$$

If $P \in \mathcal{P}^+_{\mathcal{L}}$ then the log-convexity of $\mathcal{P}^+_{\mathcal{L}}$ implies that the curve ranges in $\mathcal{P}^+_{\mathcal{L}}$. Hence, the tangent belongs to $\mathrm{Ker}_{(ij|K)}$ for $(ij|K) \in \mathcal{L}$. By Lemma 11, the tangent is in the sum of $\mathcal{U}_L$ over $L \in \mathcal{L}^\diamond$. Since $\emptyset \in \mathcal{L}^\diamond$ and $\mathcal{U}_\emptyset$ consists of the constant functions the sum contains w.

Let w belong to the sum of $\mathcal{U}_L$ over $L \in \mathcal{L}^\diamond$. By Lemma 12, $\mathcal{P}^+_{\mathcal{L}}$ is log-linear. Therefore, to prove that $P \in \mathcal{P}^+_{\mathcal{L}}$ it suffices to confine to the case $w \in \mathcal{U}_L$ for some $L \in \mathcal{L}^\diamond$. The assertion is trivial if $L = \emptyset$, having w constant and P uniform. Otherwise, $L \neq \emptyset$ and w is orthogonal to $\mathcal{U}_\emptyset$ by Lemma 3. Then, $x \mapsto m_N[1 + \varepsilon w(x)]$ defines a pm for ε sufficiently close to 0. By Corollary 3, this pm belongs to $\mathcal{P}_{\mathcal{L}}$. It is proportional to $e^{\ln[1+\varepsilon w]}$. The log-affinity of $\mathcal{P}^+_{\mathcal{L}}$ implies that the pm P_ε proportional to e^{w_ε} with $w_\varepsilon = \frac{1}{\varepsilon}\ln[1 + \varepsilon w]$ belongs to $\mathcal{P}^+_{\mathcal{L}}$. Limiting with ε to 0 the functions w_ε converge to w and P_ε converges to $P \in \mathcal{P}^+$. Hence, $P \in \mathcal{P}^+_{\mathcal{L}}$. □

The *CI*-constraint (1) given by $(ij|K)$ can be equivalently expressed as

$$(6)\quad Q(x_i x_j x_K)\cdot Q(y_i y_j x_K) = Q(x_i y_j x_K)\cdot Q(y_i x_j x_K)\,, \quad x_i, y_i \in X_i, x_j, y_j \in X_j, x_K \in X_K\,,$$

where Q denotes the marginal of P to ijK, $x_i x_j x_K$ is the element of X_{ijK} that projects to x_i, x_j and x_K, and $y_i y_j x_K$, $x_i y_j x_K$ and $y_i x_j x_K$ have analogous meaning. If $K = N \setminus ij$ then $Q = P$ and it is easy to see that if two pm's satisfy the equations in (6) then the equations also hold for the log-linear combinations of the pm's.

Under the additional natural assumption $|X_i|>1$ for all $i\in N$, the following lemma and the proof of Theorem 1 can be slightly simplified but when not excluding $|X_i|=1$ only additional minor technicalities are needed.

Lemma 14. *If $\mathcal{P}^+_{\mathcal{L}}$ is log-convex, $L\in\mathcal{L}^\diamond$ and $|X_\ell|>1$ for each $\ell\in L$ then all subsets of L belong to $\mathcal{L}^\diamond$.*

Proof. It suffices to prove that $L\setminus\ell\in\mathcal{L}^\diamond$ for all $\ell\in L$. This is accomplished when the violation of $(ij|K)\diamond(L\setminus\ell)$ for some $(ij|K)$ implies that the couple is not in $\mathcal{L}$. Thus, assume $ij\subseteq L\setminus\ell\subseteq ijK$. By the assumption on cardinalities, there exists $y\in X_N$ with $\mathsf{s}(y)=L$. Let $z\in X_N$ have $\mathsf{s}(z)=\ell$ and the same ℓ-th coordinate as y, $y_\ell\neq \mathit{0}_\ell$. Since $L,\ell\in\mathcal{L}^\diamond$ Lemma 6 implies $\alpha_y\in\mathcal{U}_L$ and $\alpha_z\in\mathcal{U}_\ell$. By the log-convexity and Lemma 13, the pm P proportional to $e^{\alpha_y+\alpha_z}$ is in $\mathcal{P}^+_{\mathcal{L}}$.

There exists $t>0$ such that $t\cdot\pi_{N\setminus\ell}P(x_{N\setminus\ell})$ equals

$$e^{\alpha_y(\mathit{0}_\ell x_{N\setminus\ell})+\alpha_z(\mathit{0}_\ell x_{N\setminus\ell})}+e^{\alpha_y(y_\ell x_{N\setminus\ell})+\alpha_z(y_\ell x_{N\setminus\ell})}+\sum\nolimits_{x_\ell\in X_\ell\setminus\{\mathit{0},y_\ell\}}e^{\alpha_y(x_\ell x_{N\setminus\ell})+\alpha_z(x_\ell x_{N\setminus\ell})}$$
$$=e^{\alpha_y(\mathit{0}_\ell x_{N\setminus\ell})+1}+e^{\alpha_y(y_\ell x_{N\setminus\ell})-1}+|X_\ell|-2\,,\qquad x_{N\setminus\ell}\in X_{N\setminus\ell}\,,$$

where α_y and α_z vanish at $x_\ell x_{N\setminus\ell}$. Combining $ij\subseteq L\setminus\ell\subseteq ijK$ and $L\in\mathcal{L}^\diamond$, the set ijK cannot contain ℓ. Therefore, $\pi_{ijK}P=Q$ is a marginal of $\pi_{N\setminus\ell}P$. Since $\pi_{N\setminus\ell}P$ depends on $x_{N\setminus\ell}$ only through $\pi^{N\setminus\ell}_{ijK}x_{N\setminus\ell}$ it follows from (3) that $Q(\pi^{N\setminus\ell}_{ijK}x_{N\setminus\ell})$ is equal to $|X_{N\setminus\ell ijK}|\pi_{N\setminus\ell}P(x_{N\setminus\ell})$. Therefore,

$$Q(\mathit{0}_i\mathit{0}_j\mathit{0}_K)Q(y_iy_j\mathit{0}_K)-Q(\mathit{0}_iy_j\mathit{0}_K)Q(y_i\mathit{0}_j\mathit{0}_K)$$

is a positive multiple of $[e^2+e^{-2}+|X_\ell|-2]^2-|X_\ell|^2$. Using (6), this implies $(ij|K)\notin\mathcal{L}$. □

Proof of Theorem 1. Let $\mathcal{R}_{N,X}$ consist of the couples $(ij|N\setminus ij)$ with $|X_i|=1$ or $|X_j|=1$. By (1), $\mathcal{P}_{\mathcal{L}}=\mathcal{P}_{\mathcal{L}\cup\mathcal{R}_{N,X}}$ for all $\mathcal{L}\subseteq\mathcal{R}$. Let $G_{\mathcal{L}}=(N,E)$ be the graph with $ij\notin E$ if and only if $(ij|K)\in\mathcal{L}$ for some $K\subseteq N\setminus ij$. By (2), a pm factorizes w.r.t. $G_{\mathcal{L}}$ if and only if it does w.r.t. $G_{\mathcal{L}\cup\mathcal{R}_{N,X}}$. It follows that it suffices to prove the assertion of Theorem 1 under the additional assumption $\mathcal{R}_{N,X}\subseteq\mathcal{L}$. Hence, any $L\in\mathcal{L}^\diamond$ with at least two elements has $|X_\ell|>1$ for all $\ell\in L$. By Lemma 14, all subsets of any $L\in\mathcal{L}^\diamond$ belong to $\mathcal{L}^\diamond$; thus $\mathcal{L}^\diamond$ is hereditary. Using Lemma 9, a set $L\subseteq N$ with at least two elements belongs to $\mathcal{L}^\diamond$ if and only if $ij\in\mathcal{L}^\diamond$ for every $i,j\in L$; thus $\mathcal{L}^\diamond$ is conformal. It follows that $\mathcal{L}^\diamond$ is the family of cliques of $G_{\mathcal{L}}$. Hence, the assertion of Theorem 1 obtains from Lemmas 5 and 13. □

5. Discussion and remarks

5.1. **The main result and its corollary.** For an undirected graph $G=(N,E)$ let $\mathcal{K}_G$ consist of the couples $(ij|N\setminus ij)$ having $ij\notin E$. The pm's from $\mathcal{P}_{\mathcal{K}_G}$ are called *pairwise Markov w.r.t. G*. The families $\mathcal{P}^+_{\mathcal{K}_G}$ parameterized by undirected graphs G have been known in the statistical literature as the *Markov models over undirected graphs* [10, 20], over the contingency tables. They are log-convex because each family $\mathcal{P}^+{}_{\{(ij|N\setminus ij)\}}$ is log-linear by the argument following (6) and intersections of log-linear families are log-linear. Hence, Theorem 1 directly implies the assertion of Corollary 1, $\mathcal{P}^+_{\mathcal{K}_G}=\mathcal{F}^+_G$. Rephrased verbally, given an undirected graph, a positive pm is pairwise Markov if and only if it factorizes. The assumption of positivity matters, see [26, p. 22] and [20, Example 3.10].

Alternatively, $\mathcal{P}^+_{\mathcal{K}_G} = \mathcal{F}^+_G$ is a simple consequence of the lemmas on the interaction spaces from Section 2. In fact, it is rather straightforward that $P \in \mathcal{P}^+$ satisfies the *CI*-constraint $(ij|N \setminus ij)$ if and only if $\ln P \colon x \mapsto \ln P(x)$ belongs to the sum of $\mathcal{V}_{N\setminus i}$ and $\mathcal{V}_{N\setminus j}$, see e.g. [20, (3.6)]. This is the sum of $\mathcal{U}_I$ where I does not contain ij, on account of Lemma 4. Therefore, by Corollary 2, $P \in \mathcal{P}^+$ is pairwise Markov w.r.t. G if and only if $\ln P$ belongs to the sum of $\mathcal{U}_I$ where I contains no $ij \notin E$, thus I is a clique of G. To conclude $\mathcal{P}^+_{\mathcal{K}_G} = \mathcal{F}^+_G$ it suffices to evoke Lemma 5. It seems that this short geometric proof of Corollary 1 via the intersections of sums of the interaction spaces is new. The presented argumentation is coordinate-free; no projectors, Moebius transform, algebras, and a special choice of $\mathit{0} \in X_N$ are employed. The only comparable proof of Corollary 1 is the inductive one by Brook [5], see also [16, Theorem 7.1], which however has no geometric content.

For $\mathcal{L} \subseteq \mathcal{R}$, Theorem 1 and log-convexity of the Markov models imply that $\mathcal{P}^+_{\mathcal{L}}$ is log-convex if and only if it is Markov over an undirected graph. Thus, the class of these Markov models can be equivalently defined by means of the *CI*-constraints and log-convexity, without any reference to graphs. Here, both the conditional independence [11] and log-convexity [6], including the geometrical viewpoint of [1], have been recognized for decades as basic building stones in statistics. Theorem 1 seems to be a new bridge between them.

5.2. **Gibbs probability measures.** Given a distinguished element $\mathit{0} = (\mathit{0}_i)_{i\in N}$ of X_N and a hereditary hypergraph $(N, \mathcal{A})$, a pm $P \in \mathcal{P}^+$ is *Gibbsian* if there exist real-valued functions v_I on X_I such that

$$\ln P(x) = \textstyle\sum_{I\in\mathcal{A},\, I\subseteq s(x)} v_I(\pi_I x)\,, \qquad x \in X\,. \tag{7}$$

The family of such probability measures is denoted by $\mathcal{G}^+_{\mathit{0},\mathcal{A}}$. If $x_I \in X_I$ and for some $i \in I$ the i-th coordinate of x_I equals $\mathit{0}_i$ then the value $v_I(x_I)$ does not occur in (7). Therefore, it can be equivalently assumed in the above definition that all such values equal zero, thus v_I are adapted to $\mathit{0}$. In this situation, the summation in (7) can run equivalently over $I \in \mathcal{A}$. Thus, the Gibbs pm's factorize in a special way, $\mathcal{G}^+_{\mathit{0},\mathcal{A}} \subseteq \mathcal{F}^+_{\mathcal{A}}$. By Lemmas 5 and 7, if $P \in \mathcal{F}^+_{\mathcal{A}}$ then $P = e^w$ for $w \in \mathcal{W}_{\mathcal{A}}$ in the form $\sum_{I\in\mathcal{A}} v_I\pi_I$ with all v_I adapted to $\mathit{0}$. Therefore, $\mathcal{F}^+_{\mathcal{A}} \subseteq \mathcal{G}^+_{\mathit{0},\mathcal{A}}$. Thus, over a hypergraph, the notion of Gibbs pm's does not depend on the choice of $\mathit{0}$ and coincides with the factorizability of positive pm's. In literature, e.g. in [7, Theorem 2], [26, Theorem 2], typically (7) is stated to be equivalent to

$$v_K(\pi_K x) = \textstyle\sum_{I\subseteq K} (-1)^{|K\setminus I|} \ln P(\rho_I x)\,, \qquad K \in \mathcal{A}\,, x \in X_N\,, s(x) = K\,,$$

which is a reinterpretation of the computation in (4), based on Moebius transform.

The identity $\mathcal{P}^+_{\mathcal{K}_G} = \mathcal{F}^+_G$, disguised in Gibbs pm's and potentials, has been revealed and proved independently several times. Over point lattices it goes back to [2, 30]. Two early unpublished manuscripts [14, 17]† have been frequently cited. Other proofs are in [4, p. 198], [7, p. 22], [15, 27, 28], and three proofs in [26, Theorem 2]. For personal remarks see also Hammersley's discussion in [4, p. 230].

---

†The former work is not available to the author.

Under topological assumptions on the state spaces, Hammersley-Clifford theorem from [13, 20] asserts that over a graph the pairwise Markovness is equivalent to the factorization, for the pm's with continuous and positive densities w.r.t. a product measure.

5.3. **Miscellany.** When P and Q are pm's on a measurable space that are not mutually singular, p and q are their densities with respect to a dominating measure R and $0 < t < 1$, the log-convex combination of P and Q is defined as the pm with the R-density proportional to $p^t q^{1-t}$. The definition is not dependent on the choice of R. A family of pm's is called log-convex if it is closed to the log-convex combinations of pairs of not mutually singular pm's. Log-convex families of mutually absolutely continuous pm's are called 'geodesically convex' in [6]. Examples of log-convex sets comprise the exponential families with convex sets of canonical parameters and their extensions [8].

A crucial role in the proof of Theorem 1 is played by the binary relation $\diamond$ between $\mathcal{R}$ and the power set of N. It appeared previously prior to [23, Theorem 4]. In addition to the mapping $\mathcal{L} \mapsto \mathcal{L}^\diamond$, it gives rise to

$$\mathcal{A} \mapsto \mathcal{A}^\diamond = \{(ij|K) \in \mathcal{R} \colon (ij|K) \diamond L \text{ for all } L \in \mathcal{A}\}$$

where $\mathcal{A}$ is a family of subsets of N. The pair of mappings forms a Galois connection [3, V.7,8]. By the remark preceding [23, Theorem 4], the connection gives rise to an antiisomorphism between DCI-relations and C^w-families, similarly to [23, Theorem 1].

The implication of Lemma 9 is valid for the family of connected sets in any topological space in the role of $\mathcal{L}^\diamond$; it goes back at least to [18].

General conditional independence statements can be reduced to families of the CI-constraints (1) by [22, Lemma 3]. The families $\mathcal{P}_\mathcal{L}$ and $\mathcal{P}_\mathcal{L}^+$ have been, in spite of considerable effort, far from being understood in general [31, 13, 24].

6. Gaussian probability measures

In this section $N = \{1, 2, \ldots, n\}$. A pm on $\mathbb{R}^n$ is regular Gaussian (rG) if its density with respect to the Lebesgue measure has the form

$$x \mapsto (2\pi)^{-\frac{n}{2}} (\det A)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(x-\mu)^T A^{-1}(x-\mu)\right], \quad x \in \mathbb{R}^n,$$

where μ is a (column) vector from $\mathbb{R}^n$ and $A = (a_{i,j})_{i,j \in N}$ a real positive definite matrix. The standard notation $\det A$ is used for the determinant of A and ^{T} for the transposition.

An rG pm satisfies the CI-constraint given by $(ij|K) \in \mathcal{R}$ if and only if $\det A_{iK,jK}$ vanishes where $A_{iK,jK}$ is the submatrix of A whose rows are indexed by iK and columns by jK, see [21]. Hence, the pm is pairwise Markov w.r.t. an undirected graph $G = (N, E)$ if and only if $\det A_{N\setminus j, N\setminus i} = 0$ whenever $ij \notin E$; this equality is equivalent to vanishing of the element of A^{-1} indexed by i, j. It follows that the pm factorizes w.r.t. the graph [20, 5.1.4] which implies Markovness [20, Proposition 3.8]. Thus, as well-known, the pairwise Markovness of any rG pm is equivalent to a factorization, over an undirected graph. For the general CI-constraints on Gaussian variables see [29, 25, 21, 13, 12].

Log-convex combinations of two rG pm's parameterized by μ, A and ν, B are the rG pm's parameterized by some $\lambda_{\mu,\nu,t} \in \mathbb{R}^n$ and $[tA^{-1} + (1-t)B^{-1}]^{-1}$, $0 < t < 1$.

Theorem 2. *If the set of* rG *pm's on* $\mathbb{R}^n$ *that satisfy all* CI*-constraints from* $\mathcal{L} \subseteq \mathcal{R}$ *is log-convex then this set coincides with the set of all* rG *pm's that factorize w.r.t. the graph* (N, E) *that has* $ij \in E$ *if and only if* $(ij|K) \notin \mathcal{L}$ *for all* $K \subseteq N \setminus ij$.

Proof. Two sets of positive definite $n \times n$ matrices A are introduced: the first one $\boldsymbol{A}_{\mathcal{L}}$ is given by $\det A_{iK,jK} = 0$ for all $(ij|K) \in \mathcal{L}$, and the second one $\boldsymbol{B}_{\mathcal{L}}$ by $A_{i,j} = 0$ for $ij \notin E$. The set of *rG* pm's specified by the *CI*-constraints from $\mathcal{L}$ is parameterized by the pairs μ, A with $A \in \boldsymbol{A}_{\mathcal{L}}$. The set of *rG* pm's that factorize w.r.t. the graph is parameterized by the pairs μ, A with $A^{-1} \in \boldsymbol{B}_{\mathcal{L}}$.

By [21, Lemma 1], $\det A_{iK,jK} = 0$ if and only if $\det(A^{-1})_{N\setminus jK, N\setminus iK} = 0$, for all invertible matrices A. Thus, $A \in \boldsymbol{A}_{\mathcal{L}}$ is equivalent to $A^{-1} \in \boldsymbol{A}_{\mathcal{L}^\rceil}$ where $\mathcal{L}^\rceil$ is the set of $(ij|N \setminus ijK)$ having $(ij|K) \in \mathcal{L}$. Since $\boldsymbol{B}_{\mathcal{L}} = \boldsymbol{B}_{\mathcal{L}^\rceil}$ Theorem 2 asserts that $\boldsymbol{A}_{\mathcal{L}^\rceil} = \boldsymbol{B}_{\mathcal{L}^\rceil}$. The assumption of log-convexity is equivalent to the convexity of $\{A^{-1} \colon A \in \boldsymbol{A}_{\mathcal{L}}\} = \boldsymbol{A}_{\mathcal{L}^\rceil}$. Therefore, it suffices to prove that if $\boldsymbol{A}_{\mathcal{L}}$ is convex then $\boldsymbol{A}_{\mathcal{L}} = \boldsymbol{B}_{\mathcal{L}}$.

The unit matrix I always belongs to $\boldsymbol{A}_{\mathcal{L}}$. This set is closed to the positive multiples. These observations and the convexity of $\boldsymbol{A}_{\mathcal{L}}$ imply that if $A \in \boldsymbol{A}_{\mathcal{L}}$ then $A + tI$ belongs to $\boldsymbol{A}_{\mathcal{L}}$ for all $t > 0$. Therefore, $\det(A + tI)_{iK,jK} = 0$ for $(ij|K) \in \mathcal{L}$. The determinant is a polynomial in t with the leading coefficient $A_{i,j}$. Hence, $A_{i,j} = 0$ for $(ij|K) \in \mathcal{L}$ which implies $\boldsymbol{A}_{\mathcal{L}} \subseteq \boldsymbol{B}_{\mathcal{L}}$.

For $ij \in E$ let $I^{[ij]}$ be the symmetric matrix with the elements at the positions i, j and j, i equal to 1 and the remaining elements equal to zero. The matrix $I + \epsilon|E|I^{[ij]}$ belongs to $\boldsymbol{A}_{\mathcal{L}}$ for ϵ is sufficiently close to zero. By convexity, $I(\epsilon) = I + \epsilon \sum_{ij \in E} I^{[ij]}$ is in $\boldsymbol{A}_{\mathcal{L}}$. Let $\mathcal{K}$ denote the set of $(ij|K) \in \mathcal{R}$ such that $N \setminus ijK$ separates i from j. Then, [21, Theorem 1] implies that up to finitely many ϵ for all $(ij|K) \in \mathcal{R}$ the determinant of $I(\epsilon)_{iK,jK}$ vanishes if and only if $(ij|K) \in \mathcal{K}$. Thus, $\boldsymbol{A}_{\mathcal{K}} \subseteq \boldsymbol{A}_{\mathcal{L}}$. It suffices to show that $\boldsymbol{B}_{\mathcal{L}} \subseteq \boldsymbol{A}_{\mathcal{K}}$. Under the separation of i and j by $N \setminus ijK$, the set K partitions into K_i and K_j such that there is no edge between iK_i and jK_j. If $A \in \boldsymbol{B}_{\mathcal{L}}$ then $A_{k_i,k_j} = 0$ for $k_i \in iK_i$ and $k_j \in jK_j$, implying $\det A_{iK,jK} = 0$, and thus $A \in \boldsymbol{A}_{\mathcal{K}}$. □

The version of Theorem 2 that features the *rG* pm's with the means equal to the zero instead of the *rG* pm's also holds. This follows from $\lambda_{\mu,\nu,t} = 0$ whenever $\mu = 0$ and $\nu = 0$, and minor modifications in the argumentation of the above proof.

Acknowledgement

The author is indebted to G.R. Grimmett for calling his attention to the work of Brook [5] and to http://www.statslab.cam.ac.uk/∼grg/books/jmh.html, containing Volume in Honour of John M. Hammersley *Disorder in Physical Systems*, Edited by Geoffrey Grimmett and Dominic Welsh, with a link to the unpublished manuscript [17] by J.M. Hammersley and P.E. Clifford.

References

[1] S. Amari (1985) *Differential-Geometrical Methods in Statistics.* Lecture Notes in Statistics 28, Springer, Berlin.

[2] M.B. Averintsev (1970) On a method of describing random fields with discrete argument. *Problemy peredachi informacii* **6** 100–108.

[3] G. Birkhoff (1967) *Lattice Theory.* Providence, Rhode Island.

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

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

[6] N.N. Chentsov (1982) *Statistical Decision Rules and Optimal Inference.* Translations of Math. Monographs, AMS, Providence – Rhode Island, 1982 (Russian original: Nauka, Moscow, 1972).

[7] P. Clifford (1990) Markov random fields in statistics. In: G. Grimmett and D. Welsh (Eds.) Disorder in Physical Systems: A Volume in Honour of J.M. Hammersley, Oxford University Press, 19–32.

[8] I. Csiszár and F. Matúš (2008) Generalized maximum likelihood estimates for exponential families. *Probability Theory and Related Fields* **141** 213–246.

[9] J.N. Darroch and T.P. Speed (1983) Additive and multiplicative models and interactions. *The Annals of Statistics* **11** 724–738.

[10] J.N. Darroch, T.P. Speed and S.L. Lauritzen(1980) Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics* **8** 522–539.

[11] A.P. Dawid (1979) Conditional independence in statistical theory (with discussion). *J.R. Statist. Soc. B* **41** 1–31.

[12] M. Drton and H. Xiao (2010) Smoothness of Gaussian conditional independence models. *Contemporary Mathematics* **516** 155–177.

[13] M. Drton, B. Sturmfels, and S. Sullivant (2009) *Lectures on Algebraic Statistics.* Birkhäuser, Basel.

[14] D. Griffeath (1973) Markov and Gibbs fields with finite state space on graphs. (unpublished)

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

[16] G.R. Grimmett (2010) *Probability on Graphs.* Cambridge University Press, Cambridge.

[17] J.M. Hammersley and P.E. Clifford (1971) Markov fields on finite graphs and lattices. (unpublished)

[18] F. Hausdorff (1914) *Grundzüge der Mengenlehre.* Lepzig.

[19] J.G. Kemeny, J.L. Snell and A.W. Knapp (1976) *Denumerable Markov Chains.* Springer, New York.

[20] S.L. Lauritzen (1996) *Graphical Models.* Oxford University Press, Oxford.

[21] R. Lněnička and F. Matúš (2007) On Gaussian conditional independence structures. *Kybernetika* **43** 327–342.

[22] F. Matúš (1992) On equivalence of Markov properties over undirected graphs. *Journal of Applied Probability* **29** 745–749.

[23] F. Matúš (1992) Ascending and descending conditional independence relations. In: Transactions of the 11-th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Academia, Prague, Vol. B, 189–200.

[24] F. Matúš (1999) Conditional independences among four random variables III: final conclusion. *Combinatorics, Probability & Computing* **8** 269–276.

[25] F. Matúš (2005) Conditional independences in Gaussian vectors and rings of polynomials. In: G. Kern-Isberner, W. Rödder and F. Kulmann (Eds.) Proceedings of WCII 2002, LNAI 3301, Springer, Berlin, 152–161.

[26] J. Moussouris (1974) Gibbs and Markov random systems with constraints. *Journal of Statistical Physics* **10** 11–33.

[27] C.J. Preston (1973) Generalized Gibbs states and Markov random fields. *Adv. Appl. Probab.* **5** 242–261.

[28] S. Sherman (1973) Markov random fields and Gibbs random fields. *Israel J. Math.* **14** 92–103.

[29] T.P. Speed and H.T. Kiiveri (1986) Gaussian Markov distributions over finite graphs. *The Annals of Statistics* **14** 138–150.

[30] F. Spitzer (1971) Markov random fields and Gibbs ensembles. *The American Mathematical Monthly* **78** 142–154.

[31] M. Studený (2005) *Probabilistic Conditional Independence Structures.* Springer, New York.