*Proceedings of the Third Workshop on*
*Uncertainty Processing in Expert Systems*
Inst. Info. Th. Autom., Prague 1994, 181–198

# ON THE MAXIMUM-ENTROPY EXTENSIONS OF PROBABILITY MEASURES OVER UNDIRECTED GRAPHS[1]

František MATÚŠ

## ABSTRACT

Two problems: 1. whether a given family of discrete probability measures indexed by the maximal cliques of an undirected graph can be extended, and 2. how to compute explicitly the extension having maximal Shannon entropy, are examined from the point of view of graph decompositions. Close product formula for the extensions is presented and its algorithmical complexity is evaluated. Two conjectures on optimality of our decompositions in both problems are put forward.

## 1. INTRODUCTION

Let $N$ and $X_i$, $i \in N$, be finite nonempty sets. For $I \subset N$ we denote by $X_I$ the Cartesian product $\Pi_{i \in I} X_i$; $X_\emptyset = \{\emptyset\}$. Having a probability measure $P$ on $X_N$, in symbols $P \in \mathcal{P}(X_N)$, its image under the coordinate projection of $X_N$ onto $X_I$ (often called marginal of $P$) is denoted by $P^I \in \mathcal{P}(X_I)$.

Let $\mathcal{A}$ be a nonempty family of subsets of the basic set $N$. A system of probability measures $\mathbf{Q} = (Q_I \in \mathcal{P}(X_I);\ I \in \mathcal{A})$ is said to be *consistent* if $\mathbf{Q}$ is equal to the system $P^{\mathcal{A}} = (P^I;\ I \in \mathcal{A})$ for some $P \in \mathcal{P}(X_N)$. Every consistent system $\mathbf{Q}$ is *projective* in the sense $Q_I^{I\cap J} = Q_J^{I\cap J}$, $I, J \in \mathcal{A}$.

For every consistent $\mathbf{Q}$ the set $\{P \in \mathcal{P}(X_N);\ P^{\mathcal{A}} = \mathbf{Q}\}$ of all extensions of $\mathbf{Q}$ is nonempty, convex and compact. Denoting by $h_N(P)$ the Shannon entropy of $P \in \mathcal{P}(X_N)$, the function $h_N$ is continuous and strictly concave on $\mathcal{P}(X_N)$. Hence, the *maximum-entropy extension*

$$\pi_N(\mathbf{Q}) = \operatorname{argmax}\{h_N(P);\ P \in \mathcal{P}(X_N), P^{\mathcal{A}} = \mathbf{Q}\} \in \mathcal{P}(X_N)$$

[1]This research was supported by Internal grants of the Academy of Sciences of the Czech Republic (GA AV ČR) No. 275105 and No. 27564.

of $\mathbf{Q}$ is unique and well-defined.

Where $\mathbf{Q}$ is a projective system and $J$ a subset of the set $N$, a new projective system $\mathbf{Q}^J = (Q^K_{e(K)};\ K \in J \cap \mathcal{A})$ can be constructed by marginalization. Here, the indexing family $J \cap \mathcal{A}$ is equal to $\{J \cap I;\ I \in \mathcal{A}\}$ and $e$ is any mapping of $J \cap \mathcal{A}$ into $\mathcal{A}$ such that $K \subset e(K)$, $K \in J \cap \mathcal{A}$. The choice of $e$ does not influence the construction of $\mathbf{Q}^J$. If the system $\mathbf{Q}$ is even consistent then $\mathbf{Q}^J$ is consistent in context of $J$ and the maximum-entropy extension

$$\pi_J(\mathbf{Q}^J) = \operatorname{argmax}\{h_J(P_J);\ P_J \in \mathcal{P}(X_J), P_J^{J\cap\mathcal{A}} = \mathbf{Q}^J\} \in \mathcal{P}(X_J)$$

of $\mathbf{Q}^J$ in this context is obtained analogically as above. The symbol $h_J$ denotes the Shannon entropy function on $\mathcal{P}(X_J)$.

This paper aims at an insight into the following two notions.

**Definition 1.** *A family $\mathcal{B}$ of subsets of $N$ is called $\mathcal{A}$-admissible if consistence of every projective system $\mathbf{Q}$ indexed by $\mathcal{A}$ follows from the consistences of all $\mathbf{Q}^J$ for $J \in \mathcal{B}$.*

**Definition 2.** *A real-valued function $\kappa$ defined on subsets of $N$ is called $\mathcal{A}$-admissible if*

$$\pi_N(\mathbf{Q}) = \prod\nolimits_{J\subset N} \left[\pi_J(\mathbf{Q}^J)\right]^{\kappa(J)}$$

*is valid for all consistent systems $\mathbf{Q} = (Q_I;\ I \in \mathcal{A})$.*

Some comments are desirable. First, in both definitions there is no restriction to the sets $X_i$, $i \in N$; they are not fixed in advance to avoid the dependence of the notions on them. Second, the $J$-th term enters the product formula if and only if $\kappa(J) \neq 0$, $J \subset N$. Third, the above formula is to be understood as a symbolical substitute for this set of equations

$$\pi_N(\mathbf{Q})(x_N) = \prod\nolimits_{J\subset N} \left[\pi_J(\mathbf{Q}^J)(x_J)\right]^{\kappa(J)}$$

where $x_N = (x_i)_{i\in N} \in X_N$ and $x_J = (x_i)_{i\in J} \in X_J$, $J \subset N$, are the canonical projections of $x_N$. And finally, one supposes tacitly that $\pi_N(\mathbf{Q})(x_N) = 0$ implies $\pi_J(\mathbf{Q}^J)(x_J) = 0$ for at least one $J \subset N$ such that $\kappa(J) > 0$. These conventions apply to all formulas of this type throughout the paper.

We restrict ourselves only to the conformal families $\mathcal{A}$, i.e. $\mathcal{A}$ will always be the family of all maximal cliques of an undirected graph. In Section 2 basic notations and facts on graph decompositions are fixed and an index function is introduced. The next two sections contain fundamental examples of $\mathcal{A}$-admissible families and of $\mathcal{A}$-admissible functions, respectively, which are conjectured to be in some sense optimal. Then we discuss the presented material with regard to previous related works.

## 2. GRAPH-THEORETICAL PRELIMINARIES

In this section we collect necessary notations and facts on graph decompositions to be used later. Not to overload the content we postpone all proofs to an appendix.

We shall deal with simple undirected graphs $G = (N, E)$ where $N$ is the vertex set of $G$ and $E$, the edge set of $G$, is a subset of the set $N_{(2)}$ of all two-element subsets of $N$. A set of vertices $L$ is *connected* in $G$ if it cannot be partitioned into $I+J$, where $I, J \subset N$ are disjoint and nonempty, such that $E \cap L_{(2)} = (E \cap I_{(2)}) + (E \cap J_{(2)})$. The family of all sets $L$ that are connected in G and maximal (w.r.t. the set inclusion) with this property is denoted by $\vee_{\mathrm{co}}(G)$. For $K \subset N$ the vertex-induced subgraph $G_K$ of $G$ is the graph $(K, E \cap K_{(2)})$. The subgraphs $G_K$, $K \in \vee_{\mathrm{co}}(G)$, are the connected components of $G$; clearly $\vee_{\mathrm{co}}(G)$ partitions $N$.

A vertex set $L \subset N$ is a *clique* of $G$ if $L_{(2)} \subset E$. The family of all maximal cliques of $G$ is denoted by $\vee_{\mathrm{cl}}(G)$. A vertex set $L \subset N$ is *prime* in $G$ if for every clique $K$ of $G$ the set $L - K$ is connected in $G$. For the family of maximal prime sets of $G$ we reserve the symbol $\vee_{\mathrm{pr}}(G)$. A graph $G$ is prime if $N$ is prime in $G$ and it is straightforward that $L \subset N$ is prime in $G$ if and only if $G_L$ is prime. A set of vertices $K \subset N$ is *separating* in $G$ if $N - K$ is not connected in $G$. The notation $\wedge_{\mathrm{sc}}(G)$ concerns the family of minimal separating cliques of $G$. A graph is prime if it has no separating clique.

**Fact A.** If $L_1, L_2 \in \vee_{\mathrm{pr}}(G)$ and $L_1 \neq L_2$ then $L_1 \cap L_2$ is a clique of $G$.

A set $K \subset N$ is *attached* to $L \subset N - K$ in $G$ if to every vertex $k \in K$ there exists $l \in L$ such that $kl \in E$. Denoting by $\partial_G L$ the boundary $\{k \in N - L;\ \exists l \in L\ kl \in E\}$ of $L$ in $G$ we see that $K$ is attached to $L$ in $G$ if $K \subset \partial_G L$. If a separating set $K$ is attached to at least two different $J_1, J_2 \in \vee_{\mathrm{co}}(G_{N-K})$ then it is called *relatively minimal.* In this case no proper subset of $K$ is separating in $G_{K+J_1+J_2}$. The family of all relatively minimal separating cliques of $G$ is denoted by ${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G)$. Immediately, every $K \in \wedge_{\mathrm{sc}}(G) \subset {}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G)$ is attached to all $J \in \vee_{\mathrm{co}}(G_{N-K})$.

**Fact B.** If $L \in \vee_{\mathrm{pr}}(G)$ and $J \in \vee_{\mathrm{co}}(G_{N-L})$ then $L$ is not attached to $J$.

As a consequence we observe that $\vee_{\mathrm{pr}}(G) \cap {}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G) = \emptyset$ and validity of the following claim.

**Fact C.** If $K \in \wedge_{\mathrm{sc}}(G)$ then

$$\vee_{\mathrm{pr}}(G) = \sum\{\vee_{\mathrm{pr}}(G_{K+J});\ J \in \vee_{\mathrm{co}}(G_{N-K})\}.$$

Similarly, one can decompose the family ${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G)$ of all relatively minimal separating cliques.

**Fact D.** If $K \in \wedge_{\mathrm{sc}}(G)$ then

$${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G) = \{K\} + \sum\{{}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G_{K+J});\ J \in \vee_{\mathrm{co}}(G_{N-K})\}.$$

A prime graph is a *cycle* if it has the same number of vertices and edges. A graph $G$ is *triangulated* if no its subgraph $G_L$ for $L \subset N$ of the cardinality at least four is a cycle.

**Fact E.** The only triangulated prime graphs are $(N, N_{(2)})$.

**Fact F.** A graph is triangulated if and only if $\vee_{\mathrm{pr}}(G) = \vee_{\mathrm{cl}}(G)$.

Having a graph $G = (N, E)$ its *canonical triangularization* is the graph $G^{\triangle}$ with the same set of vertices and with the set of edges $E^{\triangle} = \bigcup\{L_{(2)};\ L \in \vee_{\mathrm{pr}}(G)\} \supset E$.

**Fact G.** $\vee_{\mathrm{pr}}(G) = \vee_{\mathrm{pr}}(G^{\triangle}) = \vee_{\mathrm{cl}}(G^{\triangle})$

**Fact H.** ${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G) = {}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G^{\triangle})$

Our key notion which will be used in a later construction of $\mathcal{A}$-admissible set functions is that of *index* of a clique $K \subset N$ of $G$ and it is defined as follows

$$\iota_G(K) = |\{J \in \vee_{\rm co}(G_{N-K});\ K \subset \partial_G J\}| - 1\ .$$

In words, $\iota_G(K)+1$ is equal to the number of connected components of $G_{N-K}$ to the vertex sets of which $K$ is attached. If $K \in \wedge_{\rm sc}(G)$ then $\iota_G(K) = |\vee_{\rm co}(G_{N-K})| - 1$. Evidently, $\iota_G(K) \geq 1$ if and only if $K \in {}_{\rm r}\wedge_{\rm sc}(G)$.

**Lemma 1.** *For $K \in \wedge_{\rm sc}(G)$ one has $\iota_G(L) = \iota_{G_{K+J}}(L)$ once $J \in \vee_{\rm co}(G_{N-K})$ and $L \subset K + J$ is a clique of $G$ different from $K$. Further, the equality $\iota_G(L) = \iota_{G^\triangle}(L)$ takes place for $L \in {}_{\rm r}\wedge_{\rm sc}(G)$.*

## 3. $\mathcal{A}$-ADMISSIBLE FAMILIES

A family $\mathcal{B}$ is $\mathcal{A}$-admissible if and only if the family of all maximal elements of $\mathcal{B}$ (w.r.t. the set inclusion) is $\mathcal{A}$-admissible. This makes possible to restrict in the investigation of this notion, and we prefer strictly to do it, to those families $\mathcal{B}$ which have incomparable elements ($I \subset J$ and $I, J \in \mathcal{B} \ \Rightarrow\ I = J$). It is also straightforward that $\mathcal{A}$-admissibility of some $\mathcal{B}_1$ implies $\mathcal{A}$-admissibility of any $\mathcal{B}_2$ such that every $I_1 \in \mathcal{B}_1$ is contained in some $I_2 \in \mathcal{B}_2$. Writting $\mathcal{B}_1 \preceq \mathcal{B}_2$ to denote this position of $\mathcal{B}_1$ and $\mathcal{B}_2$ we see that the most interesting will be $\preceq$-minimal $\mathcal{A}$-admissible families. In this section we introduce a candidate for such a family. Below, $\mathcal{A}$ is always taken as the family of all maximal cliques of a graph.

**Lemma 2.** *Where $G$ is a graph, the family $\vee_{\rm pr}(G) - \vee_{\rm cl}(G)$ is $\vee_{\rm cl}(G)$-admissible.*

**Proof.** We are going to apply induction on the cardinality of $\vee_{\rm pr}(G)$ to show that any projective $\mathbf{Q} = (Q_I;\ I \in \vee_{\rm cl}(G))$ is consistent if $\mathbf{Q}^L$ is consistent for every $L \subset N$ which is maximal prime and not a clique in $G$. If $G$ is prime, $\vee_{\rm pr}(G) = \{N\}$, then Lemma 2 is trivial.

Let $G$ be nonprime and $\mathbf{Q}$ a projective system with consistent $\mathbf{Q}^L$ for all $L \in \mathcal{B}$, $\mathcal{B} = \vee_{\rm pr}(G) - \vee_{\rm cl}(G)$. Let $K$ be a minimal separating clique of $G$. If $J \in \vee_{\rm co}(G_{N-K})$ then $\mathbf{Q}^{K+J}$ is projective and, due to Fact C, $(\mathbf{Q}^{K+J})^L = \mathbf{Q}^L$ is consistent provided $L \in \vee_{\rm pr}(G_{K+J}) - \vee_{\rm cl}(G_{K+J}) \subset \mathcal{B}$. By induction, the consistency of $\mathbf{Q}^{K+J}$ follows. Let us pick up an extension $P_{K+J} \in \mathcal{P}(X_{K+J})$ of $\mathbf{Q}^{K+J}$. We claim, that the system $(P_{K+J}; J \in \vee_{\rm co}(G_{N-K}))$ is projective. In fact, let $L$ be a maximal clique of $G$ containing $K$. Then for $J \in \vee_{\rm co}(G_{N-K})$ the set $(K + J) \cap L$ belongs to the index set of $\mathbf{Q}^{K+J}$ whence $P^K_{K+J} = (P^{(K+J)\cap L}_{K+J})^K = (Q^{(K+J)\cap L}_L)^K = Q^K_L$. At this moment one can construct the conditional product

$$Q^K_L \prod \{\, P_{K+J}\ /\ Q^K_L;\ J \in \vee_{\rm co}(G_{N-K})\,\} \ =\ P \in \mathcal{P}(X_N).$$

Obviously, if $I$ is a maximal clique of $G$ then $I \subset K + J$ for a uniquely determined set $J \in \vee_{\rm co}(G_{N-K})$ and $P^I = (P^{K+J})^I = P^I_{K+J} = Q_I$ yielding consistence of the system $\mathbf{Q}$. ∎

Let us remark that if the graph $G$ is triangulated then arbitrary family $\mathcal{B}$ is $\vee_{\rm cl}(G)$-admissible by Fact F and Lemma 2.

**Conjecture 1.** *For any prime graph $G$ which is not a clique there exists a nonconsistent system $\mathbf{Q} = (Q_I;\, I \in \vee_{\rm cl}(G))$ of probability measures such that $\mathbf{Q}^{N-i}$ is consistent for all $i \in N$.*

It is not difficult to see that such a system $\mathbf{Q}$ must be projective. The validity of this conjecture will imply, together with Lemma 2 and the discussion preceeding it, that a family $\mathcal{B}$ is $\vee_{\rm cl}(G)$-admissible if and only if $\mathcal{B} \succeq \vee_{\rm pr}(G) - \vee_{\rm cl}(G)$.

## 4. $\mathcal{A}$-ADMISSIBLE SET FUNCTIONS

The main result of this section on a product formula is preceded by a lemma concerning Markov probability measures. A measure $P \in \mathcal{P}(X_N)$ is called (globally) *Markov* w.r.t. a graph $G$ if $P = P^K \prod\{P^{K+J}/P^K;\, J \in \vee_{\rm co}(G_{N-K})\}$ for every separating set $K$ in $G$.

**Lemma 3.** *If $P$ is a probability measure on $X_N$ which is Markov w.r.t. a graph $G$ then it satisfies*

$$P = \frac{\prod\left\{P^L;\ L \in \vee_{\rm pr}(G)\right\}}{\prod\left\{[P^K]^{\iota_G(K)};\ K \in {}_{\rm r}\wedge_{\rm sc}(G)\right\}} \quad .$$

**Proof.** For prime graphs the formula is trivial, let $G$ be nonprime and let $I$ be a minimal separating set of $G$. Markovness of $P$ w.r.t. $G$ implies Markovness of $P^{I+J}$ w.r.t. $G_{I+J}$, $J \in \vee_{\rm co}(G_{N-I})$. Hence, by induction

$$P^{I+J} = \frac{\prod\left\{(P^{I+J})^L;\ L \in \vee_{\rm pr}(G_{I+J})\right\}}{\prod\left\{[(P^{I+J})^K]^{\iota_{G_{I+J}}(K)};\ K \in {}_{\rm r}\wedge_{\rm sc}(G_{I+J})\right\}} \quad .$$

After the multiplication of these formulae over $J \in \vee_{\rm co}(G_{N-I})$ it suffices to have a look at Fact C, Fact D and Lemma 1 and also to remember that $P$ is the conditional product of the measures $P^{I+J}$, $J \in \vee_{\rm co}(G_{N-I})$. ∎

Let us make the convention that if $\mathbf{Q} = (Q_I;\, I \in \vee_{\rm cl}(G))$ is a projective system of probability measures and $J$ a clique of $G$ then $Q_J$ will be the marginal $Q_I^J$ for some $I \in \vee_{\rm cl}(G)$ containing $J$.

**Theorem.** *Where $G$ is a graph and $\mathbf{Q} = (Q_I;\, I \in \vee_{\rm cl}(G))$ a consistent system of probability measures, the maximum-entropy extension of $\mathbf{Q}$ can be computed by the formula*

$$\pi_N(\mathbf{Q}) = \frac{\prod\left\{\pi_L(\mathbf{Q}^L);\ L \in \vee_{\rm pr}(G)\right\}}{\prod\left\{[Q_K]^{\iota_G(K)};\ K \in {}_{\rm r}\wedge_{\rm sc}(G)\right\}} \quad .$$

**Proof.** Under the assumptions it is not difficult to deduce that $\pi_N(\mathbf{Q})$ is Markov w.r.t. $G$. Since $Q_K$ for $K \in {}_{\rm r}\wedge_{\rm sc}(G)$ are marginals of $P = \pi_N(\mathbf{Q})$ we see that to apply Lemma 3 all we need is to verify that $\pi_N(\mathbf{Q})^L = \pi_L(\mathbf{Q}^L)$ for $L \in \vee_{\rm pr}(G)$. This is, however, trivial for $L \in \vee_{\rm pr}(G) \cap \vee_{\rm cl}(G)$ and thus if $G$ is triangulated then due to Fact F the proof is over.

Let us consider now the new system $\mathbf{R} = (R_L;\ L \in \vee_{\rm pr}(G))$ of probability measures where $R_L = \pi_L(\mathbf{Q}^L)$. It is indexed by the maximal cliques of the canonical triangularization $G^{\triangle}$ of $G$ (cf. Fact G). The system $\mathbf{R}$ is projective because for two different $L_1, L_2 \in \vee_{\rm pr}(G)$ the intersection $L_1 \cap L_2$ is a clique of $G$ by Fact A what causes that the $(L_1 \cap L_2)$-marginals of $\pi_{L_1}(\mathbf{Q}^{L_1})$ and $\pi_{L_2}(\mathbf{Q}^{L_2})$ are marginals of some measures from the projective system $\mathbf{Q}$. Projectivity of $\mathbf{R}$ and Lemma 2 imply that $\mathbf{R}$ is consistent. Knowing that our Theorem is valid for triangulated graphs we can apply it immediately to $G^{\triangle}$ and $\mathbf{R}$ to have

$$\pi_N(\mathbf{R}) \;=\; \frac{\prod\left\{\pi_L(\mathbf{R}^L);\ L \in \vee_{\rm pr}(G^{\triangle})\right\}}{\prod\left\{[R_K]^{\iota_{G^{\triangle}}(K)};\ K \in {}_{\rm r}\wedge_{\rm sc}(G^{\triangle})\right\}} \quad .$$

Using Fact G, Fact H and Lemma 1 we see that all triangles in this formula can be omitted and, moreover, instead of $\mathbf{R}$ the system $\mathbf{Q}$ can be substituted in the right-hand side

$$\pi_N(\mathbf{R}) \;=\; \frac{\prod\left\{\pi_L(\mathbf{Q}^L);\ L \in \vee_{\rm pr}(G)\right\}}{\prod\left\{[Q_K]^{\iota_G(K)};\ K \in {}_{\rm r}\wedge_{\rm sc}(G)\right\}} \quad .$$

Next, we observe that $\pi_N(\mathbf{R})$ is an extension of $\mathbf{Q}$. Namely, if $I \in \vee_{\rm cl}(G)$ then $I \subset L$ for some $L \in \vee_{\rm pr}(G)$ and consequently $\pi_N(\mathbf{R})^I = (\pi_N(\mathbf{R})^L)^I = \pi_L(\mathbf{Q}^L)^I = Q_I$. The entropy of $\pi_N(\mathbf{R})$

$$\sum\left\{h_L(\pi_L(\mathbf{Q}^L));\ L \in \vee_{\rm pr}(G)\right\} \\ - \sum\left\{\iota_G(K) h_K(Q_K);\ K \in {}_{\rm r}\wedge_{\rm sc}(G)\right\}$$

cannot exceed the entropy of $\pi_N(\mathbf{Q})$ which is equal (owing to Lemma 3) to

$$\sum\left\{h_L(\pi_N(\mathbf{Q})^L);\ L \in \vee_{\rm pr}(G)\right\} \\ - \sum\left\{\iota_G(K) h_K(Q_K);\ K \in {}_{\rm r}\wedge_{\rm sc}(G)\right\} \quad .$$

Hence

$$0 \le \sum\left\{h_L(\pi_N(\mathbf{Q})^L) \;-\; h_L(\pi_L(\mathbf{Q}^L))\,;\ L \in \vee_{\rm pr}(G)\right\}$$

and no of the terms in the summation can be positive. By the uniqueness of the maximum-entropy extensions we have $\pi_N(\mathbf{Q})^L = \pi_L(\mathbf{Q}^L)$ as was desired above. ∎

Let $\kappa_G$ be the set function equal identically to 1 on the family $\vee_{\rm pr}(G)$, equal to $-\iota_G$ on ${}_{\rm r}\wedge_{\rm sc}(G)$ and to zero otherwise. Then $\kappa_G$ is by Theorem $\vee_{\rm cl}(G)$-admissible.

**Conjecture 2.** *If $G$ is a prime and $\kappa$ any $\vee_{\rm cl}(G)$-admissible set function then necessarily $\kappa(N) = 1$.*

The validity of this conjecture will imply that over prime graphs there are no meaningful product formulae for maximum-entropy extension. Another consequence would claim that if $\kappa$ is $\vee_{\rm cl}(G)$-admissible and $L \in \vee_{\rm pr}(G)$ then $\kappa$ must be nonzero on some superset of $L$ and that means the formula from Theorem be in some sense optimal.

## 5. COMMENTS

This section gathers references to all relevant items of the material presented up to now. An example follows to clarify the used notions and to outline algorithmical aspects of the formula. A comparision is made with previous works.[2]

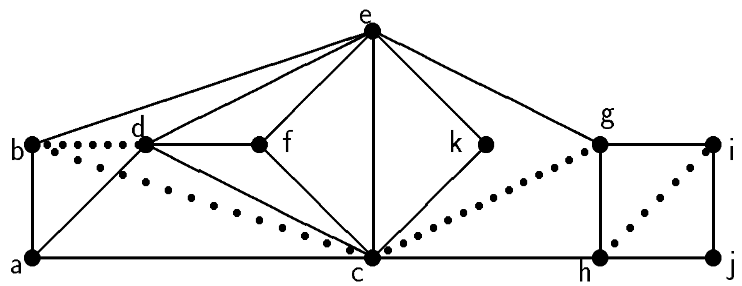
**5.1. Discussion of references.** First results on extensions of a system of probability measures on Cartesian products go back to Vorobjev (1962,1963) and Kellerer (1964ab). The problem of computing extensions and deciding whether they exist has been yet attractive for many researchers (Dall'Aglio et al., Eds.(1991), Malvestuto (1988), Matúš (1988), Matúš et al.(1993), etc.). It was found to be of some practical importance in probabilistic expert systems, see Hájek et al.(1992).

To recognize consistence of a given $\mathbf{Q}$ Kellerer (1964b) proposed to verify a huge set of inequalities. This being practically impossible the only way how to do the task is to run the iterative proportional fitting procedure, IPFP (cf. Fienberg(1983)), converging if and only if $\mathbf{Q}$ is consistent. Since one step of the procedure has the complexity $\mathcal{O}(|X_N|)$ this is, in general, not effective, too. Even when the procedure is implemented according to Jiroušek (1991), there is a strong motivation to decompose the problem in advance. First nontrivial steps in this direction were done by Malvestuto (1989) who proposed an algorithm decomposing the marginal problem; by his words "an 'informed' computational policy to reduce the use of IPFP to a minimum". We compare his algorithm with our approach below in great detail.

The facts on graph decompositions from Section 2 were taken over from Diestel (1990) and Leimer (1993) but they are mainly of an older origin. We present short proofs of them (with some fresh ideas) in Appendix to make the explanation selfconsistent.

For various Markov properties we refer the reader to Darroch et al.(1980) and Matúš (1992).

**5.2. Example.** Let us consider the graph $G = (N, E)$




having eleven vertices, $N = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \ldots, \mathsf{k}\}$, and eighteen edges, $E = \{\mathsf{ab}, \mathsf{ac}, \ldots, \mathsf{ek}\}$. To implement the results of Lemma 2 and Theorem we have to find all maximal

[2] The style of this section is in this WUPES working version of the paper much less formal.

prime sets of $G$, all relatively minimal separating cliques of $G$ and the index $\iota_G$. This can be performed as follows. 1. A minimal elimination ordering of the vertex set of $G$ is to be found (Rose et al.(1976)) and all edges of its fill-in to be added (Tarjan et al.(1984)). In our example, we obtain in this way e.g. $G' = (N, E')$ where $E' = E \cup \{\mathsf{bc}, \mathsf{bd}, \mathsf{cg}, \mathsf{hi}\}$ (the added edges as drawn by dots). The new graph $G'$ is a minimal triangularization of $G$. 2. The cliques of $G'$ are constructed and ordered to fulfil the running intersection property (RIP), e.g. by the maximum cardinality search (Tarjan et al.(1984)). Here, $\vee_{\mathrm{cl}}(G') = \{\mathsf{abcd}, \mathsf{bcde}, \mathsf{cdef}, \mathsf{ceg}, \mathsf{cgh}, \mathsf{ghi}, \mathsf{hij}, \mathsf{cek}\}$. 3. Leimer's (1993) modification of a decomposition algorithm by Tarjan (1985) is applied to the sequence of cliques; for our example

$$\vee_{\mathrm{pr}}(G) = \{\mathsf{abcde}, \mathsf{cdef}, \mathsf{cegh}, \mathsf{ghij}, \mathsf{cek}\}.$$

The elements of $\vee_{\mathrm{pr}}(G)$ are the cliques of $G^{\triangle}$ and in this order satisfy the RIP. Now, one proceeds as usually on triangulated graphs (Fact G, Fact H and Lemma 1)

$${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G) = \{\mathsf{ce}, \mathsf{gh}, \mathsf{cde}\}$$

and $\iota_G(\mathsf{ce}) = 2$ and $\iota_G(\mathsf{gh}) = \iota_G(\mathsf{cde}) = 1$, see also the next subsection.

Having a projective system $\mathbf{Q} = (Q_I;\, I \in \vee_{\mathrm{cl}}(G))$ of probability measures where

$$\vee_{\mathrm{cl}}(G) = \{\mathsf{ab}, \mathsf{be}, \mathsf{ch}, \mathsf{eg}, \mathsf{gh}, \mathsf{gi}, \mathsf{hj}, \mathsf{ij}, \mathsf{acd}, \mathsf{cek}, \mathsf{cdef}\}$$

it is consistent by Lemma 2 if and only if $\mathbf{Q}^{\mathsf{abcde}}, \mathbf{Q}^{\mathsf{cegh}}$ and $\mathbf{Q}^{\mathsf{ghij}}$ are consistent. If consistent, its maximum-entropy extension can be computed by the product formula of Theorem which looks graphically simplified like

$$\frac{\pi(\mathsf{abcde}) \cdot \pi(\mathsf{cegh}) \cdot \pi(\mathsf{ghij}) \cdot \mathsf{cek} \cdot \mathsf{cdef}}{\mathsf{ce}^2 \cdot \mathsf{cde} \cdot \mathsf{gh}} \quad .$$

The complexity of finding $\vee_{\mathrm{pr}}(G)$, ${}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G)$ and $\iota_G$ is that of the employed algorithms, i.e. $\mathcal{O}(|N||E|)$.

**5.3. Comparision of index functions.** Where $G$ is a graph, Darroch et al.(1980) and Lauritzen et al.(1984) defined their index $\nu_G$ of a clique $K \subset N$ of $G$ by

$$\nu_G(K) = 1 - |\{J \in \vee_{\mathrm{co}}(G_{N-K});\; K \notin \vee_{\mathrm{cl}}(G_{K+J})\}| \ .$$

Note that in the above example $\vee_{\mathrm{co}}(G_{N-\mathsf{cde}}) = \{\mathsf{f}, \mathsf{k}, \mathsf{ab}, \mathsf{ghij}\}$ whence $\nu_G(\mathsf{cde}) = 0$ whereas $\kappa_G(\mathsf{cde}) = -\iota_G(\mathsf{cde}) = -1$. If the graph $G$ is triangulated, then $\nu_G(K) = \kappa_G(K)$ for every clique $K$ of $G$. In fact, for maximal cliques it is straightforward by Fact G. The remainder follows from the observation that a clique $K$ of $G$ is attached to $J \in \vee_{\mathrm{co}}(G_{N-K})$ if and only if $K$ is not a maximal clique of $G_{K+J}$ (the "only if" implication is due to Fact B and Fact G).

Another kind of index comes from Haberman (1974) p. 174 where the strive was, similarly as with $\nu_G$, to find closed-form maximum-likelihood estimates (our approach is suitable for this purpose, too). We can describe it as $\lambda_G(K) = 1$ for $K \in \vee_{\mathrm{cl}}(G)$, $\lambda_G(K) = -\bar{\iota}_G(K)$ for $K \in \cap_2 \vee_{\mathrm{cl}}(G)$ and $\lambda_G(K) = 0$ otherwise.

Here $\overline{\iota}_G(K)$ is the "adjusted replication number" defined inductively on the poset $\cap_2 \vee_{\mathrm{cl}}(G) = \{L_1 \cap L_2;\ L_1 \neq L_2,\ L_1, L_2 \in \vee_{\mathrm{cl}}(G)\}$ by the formulae

$$1 + \sum\{\overline{\iota}_G(L);\ L \supset K,\ L \in \cap_2 \vee_{\mathrm{cl}}(G)\} = |\{J \in \vee_{\mathrm{cl}}(G);\ J \supset K\}|$$

starting from the maximal elements $K$ of $\cap_2 \vee_{\mathrm{cl}}(G)$ (w.r.t. the set inclusion). In the above example, one of the formulae is e.g. $1 + \overline{\iota}_G(\mathsf{h}) = 3$ whence $\lambda_G(\mathsf{h}) = -2$ whereas $\kappa_G(h) = -\iota_G(h) = 0$. It is not difficult to prove that for any triangulated graph $G$ the equality $\lambda_G = \kappa_G$ (on cliques of $G$) takes place again.

**5.4. Decompositions by Malvestuto.** To achieve similar product decompositions of maximum-entropy extensions Malvestuto (1989) proposed a two-stage decomposition algorithm working for arbitrary family $\mathcal{A}$ of subsets of the basic set $N$. We discuss its performance for conformal $\mathcal{A}$ below.

Let $\mathcal{A} = \vee_{\mathrm{cl}}(G)$ for a graph $G$. First, the elements of $\mathcal{A}$ must be linearly ordered. Then, for the first clique $K \in \mathcal{A}$ one has to verify whether the intersection graph of the family $\{L - K;\ L \in \mathcal{A},\ L \neq K\}$ has more connected components then the intersection graph of $\mathcal{A}$ (Step 1 of Decomposition algorithm from p. 305). We remark in passing that this is equivalent to the verification whether $K$ is a separating set of $G$ (in our terminology). If the answer is yes then the graph $G$ is decomposed into $G_{K+J}$, $J \in \vee_{\mathrm{co}}(G_{N-K})$ (and also $\mathcal{A}$ into $\mathcal{A}_J = \{L \in \vee_{\mathrm{cl}}(G);\ L \subset K + J\}$) and the same decomposition step will be performed with the second clique on its decomposition piece. If the answer is no, then $G$ is not decomposed and one will proceed with the second clique on the whole $G$. This decomposition phase goes on with all maximal cliques from $\mathcal{A}$ in the adopted order. It is clear that at the end $G$ is decomposed into a class of its subgraphs which have no maximal cliques that are separating in them.

In the second phase all obtained subgraphs are elaborated separately by the reduction algorithm. It works e.g. on $\mathcal{A}$ erasing repeatedly all elements of $N$ which are only in one set of $\mathcal{A}$ and all sets of $\mathcal{A}$ contained in some other sets of $\mathcal{A}$.

To illustrate this algorithm let us decompose the graph of our example. The order of its maximal cliques is as above. In the decomposition phase $G$ is not decomposed till the fifth step done with $K = \mathsf{gh}$. The pieces are $\mathsf{ghij}$ and $N - \mathsf{ij}$. Then the second piece is decomposed into $\mathsf{cdghk}$ and $N - \mathsf{ghij}$ in the tenth step and $N - \mathsf{ghij}$ into $\mathsf{abcdek}$ and $\mathsf{abcdef}$ in the last step. So, one obtains $\mathsf{ghij}$, $\mathsf{ceghk}$, $\mathsf{abcdek}$ and $\mathsf{abcdef}$. The last three subgraphs can be further reduced by the reduction algorithm, e.g. $\mathsf{ceghk}$ into $\mathsf{cek}$ and $\mathsf{cegh}$, etc.

It is clear that there is a lot of redundancy in the algorithm. Its output depends strongly on the adopted ordering and it is not unique often containing superfluous pieces. The complexity of the first step alone is $\mathcal{O}(|N||E|)$ (see Tarjan (1985) p. 233) so that our approach seems to be more effective.

Nevertheless, we have a proof that the algorithm finds all maximal primes of $G$ (even without repetitions in the reduction phase) those being always among the output subgraphs.

## 6. APPENDIX

**Proof of Fact A.** (cf. Leimer(1993), p. 102) Let us have $L_1, L_2, \in \vee_{\mathrm{pr}}(G)$ distinct. Then $G$ is not prime, let us chose any $K \in \wedge_{\mathrm{sc}}(G)$. Thus, $L_1 \in \vee_{\mathrm{pr}}(G_{K+J_1})$ and $L_2 \in \vee_{\mathrm{pr}}(G_{K+J_2})$ for unique $J_1, J_2 \in \vee_{\mathrm{co}}(G_{N-K})$. If $J_1 \neq J_2$ then $L_1 \cap L_2 \subset K$ is a clique of $G$, if $J_1 = J_2$ it suffices to apply an induction argument. ∎

**Proof of Fact B.** We proceed by contradiction, let us consider $L \in \vee_{\mathrm{pr}}(G)$, $J \in \vee_{\mathrm{co}}(G_{N-L})$, $L \subset \partial_G J$ and let the graph $G$ have the smallest possible number of vertices to admit this situation. As $J$ is nonempty the graph $G = G_{L+J}$ is not prime; let $K \in \wedge_{\mathrm{sc}}(G)$. Evidently, $K$ is different from $L$ and also $K \not\supset L$ otherwise $K$ would be a prime containing properly $L$. Hence, $L - K \subset I$ for some $I \in \vee_{\mathrm{co}}(G_{N-K})$ and $L \subset I + K \neq N$. Since $L - K$ is attached to $J$ we deduce that $L - K$ is contained in $\partial_{G_{I+K}}((I+K) - L)$. The separating clique $K$ cannot be a proper subset of $L$ as this would cause that $G_{N-K}$ be connected. It follows that $L \cap K \subset \partial_{G_{I+K}}(K - L)$. The set $J_0 = (I+K) - L$ is connected because $J$ is connected. We conclude $L \in \vee_{\mathrm{pr}}(G_{I+K})$ and $J_0 \in \vee_{\mathrm{co}}(G_{(I+K)-L})$ and $L \subset \partial_{G_{I+K}} J_0$ contradicting the minimality of $G$. ∎

Fact C is a direct consequence of Fact B and Fact D is implied by the first assertion of Lemma 1 proved below.

**Proof of Fact E.** Let $G$ be a triangulated prime and $N$ not a clique of $G$. Let $k, l \in N$ distinct, be chosen to have $kl \notin E$ with $K = \partial_G k \cap \partial_G l$ of the largest possible cardinality. Note that $K$ is a clique otherwise $G_{kl \cup K}$ contains a cycle of length four. Since $K$ does not separate $G$ there exist $j_0, \ldots, j_t \in N - K$, $t \geq 3$, satisfying $j_0 = k$, $j_t = l$ and $E \cap \{j_0, \ldots, j_t\} = \{j_0 j_1, \ldots, j_{t-1} j_t\}$. If $\partial_G j_s \supset K$ for $s = 1, 2$ we obtain for $k' = j_0$ and $l' = j_2$ a contradiction with the choice of $k$ and $l$. Thus, we can pick up $i \in K - \partial_G j_s$ for $s = 1$ or $s = 2$ and observe that the subgraph $G_{\{i, j_0, \ldots, j_t\}}$ contains a cycle of the length at least four. ∎

**Proof of Fact F.** If $\vee_{\mathrm{pr}}(G) = \vee_{\mathrm{cl}}(G)$ then every prime set $K$ of $G$ is a clique and thus $G$ is triangulated. The opposite implication can be proved by induction. For prime $G$ it follows from Fact E and the induction step comes from Fact C. ∎

**Proof of Fact G.** For prime graph $G$ the assertion is trivial (cf. Fact E). Let $G$ be not prime and $K \in \wedge_{\mathrm{sc}}(G)$. By Fact C we know that $K \in \wedge_{\mathrm{sc}}(G^{\triangle})$ and $(G_{K+J})^{\triangle} = (G^{\triangle})_{K+J}$ for $J \in \vee_{\mathrm{co}}(G)$. An induction argument based on double use of Fact C provides

$$\vee_{\mathrm{pr}}(G) = \sum \{\vee_{\mathrm{pr}}((G^{\triangle})_{K+J});\ J \in \vee_{\mathrm{co}}((G^{\triangle})_{N-K})\} = \vee_{\mathrm{pr}}(G^{\triangle}) \quad .$$

As $K$ is not a maximal clique in every $(G^{\triangle})_{K+J}$ we deduce $\vee_{\mathrm{pr}}(G^{\triangle}) = \vee_{\mathrm{cl}}(G^{\triangle})$ by induction, too. ∎

Fact H is a consequence of Fact D and induction arguments similarly as in the previous proof.

**Proof of Lemma 1.** Let $L$ be a clique of $G$ not contained in $K \in \wedge_{\mathrm{sc}}(G)$. Then $L \subset K + J$ for exactly one $J \in \vee_{\mathrm{co}}(G_{N-K})$. In the case $L \supset K$ we know that $L$ is attached to no set $I \in \vee_{\mathrm{co}}(G_{N-K}) - \{J\} \subset \vee_{\mathrm{co}}(G_{N-L})$ and thus $\iota_G(L) = \iota_{G_{K+J}}(L)$.

In the opposite case, $K - L \neq \emptyset$, the subgraph $G_{N-(J\cup L)}$ is connected and then the set $N - (J \cup L)$ is contained in some connected $I \in \vee_{\mathrm{co}}(G_{N-L})$. Further, $L$ is attached to $I$ if and only if it is attached to the set $I \cap (K + J) \in \vee_{\mathrm{co}}(G_{(K+J)-L})$. This yields again $\iota_G(L) = \iota_{G_{K+J}}(L)$. At this moment we know also that the Fact D and Fact H are proved.

The second assertion comparing $\iota_G$ and $\iota_{G^\triangle}$ is for prime $G$ trivial. For nonprime $G$ we go on as in the foregoing proof and see immediately that $\iota_G(K) = \iota_{G^\triangle}(K)$ for $K \in \wedge_{\mathrm{sc}}(G)$. Let $L \in {}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G)$, $L \neq K$ and thus (by Fact D) $L \in {}_{\mathrm{r}}\wedge_{\mathrm{sc}}(G_{K+J})$ for a unique $J \in \vee_{\mathrm{co}}(G_{N-K})$. By the first part of Lemma 1 we deduce $\iota_G(L) = \iota_{G_{K+J}}(L)$ and then by an induction argument $\iota_{G_{K+J}}(L) = \iota_{(G_{K+J})^\triangle}(L)$. Since, $(G_{K+J})^\triangle$ equals $(G^\triangle)_{K+J}$ we apply again the first assertion of this lemma to $G^\triangle$ and conclude that $\iota_G(L) = \iota_{G^\triangle}(L)$. ∎

**REFERENCES**


Dall'Aglio G., Kotz K. and Salinetti G., Eds. (1991) *Advances in Probability Distributions with Given Marginals*, Kluwer Academic Publishers, Dordrecht, Boston, London.

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

Diestel R. (1990) *Graph Decompositions: (A Study in Infinite Graph Theory)*, Clarendon Press, Oxford.

Fienberg E. and Meyer M.M. (1983) Iterative proportional fitting, in: *Encyclopedia of Statistical Sciences* (S. Kotz, N.L. Johnson, C.B. Read, Eds.), Vol. 4, John Wiley & Sons, New York, 275-279.

Haberman S.J. (1974) *The analysis of frequency data*, The University of Chicago Press, Chicago and London.

Hájek P., Havránek T. and Jiroušek R. (1992) *Uncertain Information Processing in Expert Systems*, CRC Press Inc., Boca Raton.

Jiroušek R. (1991) Solution of the marginal problem and decomposable distributions, *Kybernetika* **27**, 403-412.

Kellerer H.G. (1964a) Verteilungsfunktionen mit gegebenen Marginalverteilungen, *Zeitschrift fr Wahrscheinlichkeitstheorie und verw. Gebiete* **3**, 247-270.

Kellerer H.G. (1964b) Maßtheoretische Marginalprobleme, *Math. Annalen* **153**, 168-198.

Lauritzen S.L., Speed T.P. and Vijayan K. (1984) Decomposable graphs and hypergraphs, *J. Austral. Math. Soc. (Series A)* **36**, 12-29.

Leimer H.-G. (1993) Optimal decomposition by clique separators, *Discrete Math.* **113**, 99-123.

Malvestuto F.M. (1988) Existence of extensions and product extensions for discrete probability distributions, *Discrete Math.* **69**, 61-77.

Malvestuto F.M. (1989) Computing the maximum-entropy extension of given discrete probability distributions, *Comp. Statistics & Data Analysis* **8**, 299-311.

Matúš F. (1988) Discrete marginal problem for complex measures, *Kybernetika* **24**, 36-46.

Matúš F. (1992) On equivalence of Markov properties over undirected graphs, *J. Appl. Prob.* **29**, 745-749.

Matúš F. and Flusser J. (1993) Image representations via a finite Radon transform, *IEEE TPAMI* **15**, 996-1006.

Rose D.J., Tarjan R.E. and Lueker G.S. (1976) Algorithmic aspects of vertex elimination on graphs, *SIAM J. Comput.* **5**, 266-283.

Tarjan R.E. (1985) Decomposition by clique separators, *Discrete Math.* **55**, 221-232.

Tarjan R.E. and Yannakakis M. (1984) Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, *SIAM J. Comput.* **13**, 566-579.

Vorob'ev N.N. (1962) Consistent families of measures and their extensions, *Theory Prob. Applications* **7**, 153-169 (in Russian).

Vorob'ev N.N. (1963) Markov measures and Markov extensions, *SIAM Theory Prob. Applications* **8**, 420-429 (English translation).

František Matúš
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4
182 08 Prague 8, Czech Republic
E-mail: matus@utia.cas.cz