SUPPORT SETS OF DISTRIBUTIONS WITH GIVEN INTERACTION STRUCTURE

THOMAS KAHLE[1] AND NIHAT AY[1,2]

ABSTRACT. We study closures of hierarchical models which are exponential families associated with hypergraphs by decomposing the corresponding interaction spaces in a natural and transparent way. Here, we apply general results on closures of exponential families.

Index Terms – Closure of exponential family, graphical model, hierarchical model, interaction order.

1. INTRODUCTION

The set of probability measures on a Cartesian product of finite state sets of nodes allows for the analysis of interaction structures among the nodes [DS]. An important class of such structures, the so-called *graphical models*, is induced by graphical representations of interactions [Lau, St]. Given an undirected graph G, the set of strictly positive probability measures that satisfy corresponding Markov properties forms an exponential family [BN, Am]. Dealing with probability distributions associated with G that are not necessarily strictly positive requires the study of the closure of that exponential family. In this note, we apply general results from [BN, CMb] on closures of exponential families for the concise description of closures of *hierarchical models*. These models are associated with hypergraphs [Lau] and generalize the class of graphical models. By decomposing corresponding interaction spaces in terms of linear algebra we hope to approach a constructive method that specifies the closure of a hierarchical model.

2. PRELIMINARIES

Given a non-empty finite set $\mathcal{X}$, we denote the set of probability distributions on $\mathcal{X}$ by $\bar{\mathcal{P}}(\mathcal{X})$ which is a subset of $\mathbb{R}^{\mathcal{X}}$. The *support* of $P \in \bar{\mathcal{P}}(\mathcal{X})$ is defined as $\mathrm{supp}(P) := \{x \in \mathcal{X} : P(x) > 0\}$. For a subset $\mathcal{Y} \subseteq \mathcal{X}$ we consider the set $\mathcal{P}(\mathcal{Y})$ of probability vectors with support equal to $\mathcal{Y}$, and one obviously has

$$\bar{\mathcal{P}}(\mathcal{X}) = \bigcup_{\emptyset \neq \mathcal{Y} \subseteq \mathcal{X}} \mathcal{P}(\mathcal{Y}).$$

With the map

$$\exp : \mathbb{R}^{\mathcal{X}} \to \mathcal{P}(\mathcal{X}), \qquad f \mapsto \frac{\exp(f)}{\sum_{x \in \mathcal{X}} \exp(f(x))},$$

Date: July 29, 2006.

an *exponential family* (*in* $\mathcal{P}(\mathcal{X})$) is defined as the image $\exp(\mathcal{V})$ of a linear subspace $\mathcal{V}$ of $\mathbb{R}^{\mathcal{X}}$. The closure $\operatorname{cl}\mathcal{E}$ of an exponential family $\mathcal{E}$ is given with respect to the topology of $\mathbb{R}^{\mathcal{X}}$.

Now we assume a compositional structure of $\mathcal{X}$ induced by a set V of $1 \leq N < \infty$ nodes with state sets $\mathcal{X}_v$, $v \in V$. Here, we will only treat the binary case, i.e. $\mathcal{X}_v = \{0,1\}$ for all $v \in V$. Given a finite subset $A \subseteq V$, we write $\mathcal{X}_A$ instead of $\times_{v\in A}\mathcal{X}_v$, and we have the natural projections

$$X_A : \mathcal{X}_V \to \mathcal{X}_A, \qquad (x_v)_{v\in V} \mapsto (x_v)_{v\in A} \,.$$

With a probability vector P on $\mathcal{X}_V$, the X_A become random variables. We use the compositional structure of $\mathcal{X}_V$ in order to define exponential families in $\mathcal{P}(\mathcal{X}_V)$ given by interaction spaces. We decompose $x \in \mathcal{X}_V$ in the form $x = (x_A, x_{V\setminus A})$ with $x_A \in \mathcal{X}_A$, $x_{V\setminus A} \in \mathcal{X}_{V\setminus A}$, and define $\mathcal{I}_A$ to be the subspace of functions that do not depend on the configurations $x_{V\setminus A}$:

$$\mathcal{I}_A \;:=\; \Big\{ f \in \mathbb{R}^{\mathcal{X}} \;:\; f(x_A, x_{V\setminus A}) = f(x_A x'_{V\setminus A}) \quad \text{for all } x_A \in \mathcal{X}_A, \text{ and all } x_{V\setminus A}, x'_{V\setminus A} \in \mathcal{X}_{V\setminus A} \Big\} \,.$$

In the following, we apply these interaction spaces as building blocks for more general interaction spaces and associated exponential families [DS]. The most general construction is based on a set of subsets of V, a so-called *hypergraph* [Lau]. Given such a set $\mathscr{A} \subseteq 2^V$, we define the corresponding interaction space by

$$\mathcal{I}_{\mathscr{A}} \;:=\; \sum_{A\in\mathscr{A}} \mathcal{I}_A$$

and consider the corresponding exponential family $\mathcal{E}_{\mathscr{A}} := \exp(\mathcal{I}_{\mathscr{A}})$.

Example 1.
(1) Graphical models: Let $G = (V,E)$ be an undirected graph, and define

$$\mathscr{A}_G := \{C \subseteq V : C \text{ is a clique with respect to } G\}\,.$$

Here, a *clique* is a set C that satisfies the following property:

$$a, b \in C, \quad a \neq b \quad \Rightarrow \quad \text{there is an edge between } a \text{ and } b\,.$$

The exponential family $\mathcal{E}_{\mathscr{A}_G}$ is characterized by Markov properties with respect to G (see [Lau]).
(2) Interaction order: The hypergraph associated with a given interaction order $k \in \{0,1,2,\ldots,N\}$ is defined as

$$\mathscr{A}_k := \{A \subseteq V : |A| \leq k\} \,.$$

This gives us a corresponding hierarchy of exponential families studied in [Am, AK]:

$$\mathcal{E}_{\mathscr{A}_0} \;\subseteq\; \mathcal{E}_{\mathscr{A}_1} \;\subseteq\; \mathcal{E}_{\mathscr{A}_2} \subseteq\; \cdots \subseteq\; \mathcal{E}_{\mathscr{A}_N} = \mathcal{P}(\mathcal{X}_V)\,.$$

In Example (3), we will discuss $\mathscr{A}_i$ and $\mathcal{E}_{\mathscr{A}_i}$, $i = 1,2$, in the case of two units.

3. Problem Statement and the Main Result

Given a complete hypergraph $\mathscr{A}$ (i.e. $A \in \mathscr{A}, B \subseteq A \Rightarrow B \in \mathscr{A}$), we consider the closure $\operatorname{cl}\mathcal{E}_{\mathscr{A}}$ of the exponential family $\mathcal{E}_{\mathscr{A}}$, and the map

$$\operatorname{supp}: \operatorname{cl}\mathcal{E}_{\mathscr{A}} \to 2^{\mathcal{X}_V}, \qquad P \mapsto \operatorname{supp}(P),$$

that assigns to each $P \in \operatorname{cl}\mathcal{E}_{\mathscr{A}}$ the support $\operatorname{supp}(P)$. In our main result (Theorem 2) we characterize the image of this map. To this end, we define the following family of functions:

$$e_A : \mathcal{X}_V \to \mathbb{R}, \qquad x \mapsto (-1)^{E(A,x)}, \qquad (A \in \mathscr{A}) \tag{1}$$

where $E(A, x)$ denotes the number of entries of x in A that are equal to one. More formally,

$$E(A, x) := |\{v \in A : X_v(x) = 1\}| \,. \tag{2}$$

Obviously, the functions $e_A \in \mathbb{R}^{\mathcal{X}_V}$ can be represented by the canonical basis e_x, $x \in \mathcal{X}_V$, as follows:

$$e_A = \sum_{x \in \mathcal{X}_V} (-1)^{E(A,x)} e_x \,.$$

Now fix an arbitrary numbering of $\mathscr{A} \setminus \{\emptyset\}$, set $s := |\mathscr{A}| - 1$, and consider the following composed map:

$$e_{\mathscr{A}} : \mathcal{X}_V \to \mathbb{R}^s, \qquad x \mapsto (e_{A_1}(x), \dots, e_{A_s}(x)) \,.$$

The image of this map is a subset of the extreme points $\{-1, 1\}^s$ of the hypercube in $\mathbb{R}^s$. Note that for $\mathscr{A}_1$ (see Example 1), the image of $e_{\mathscr{A}_1}$ coincides with $\{-1, 1\}^s$. In general this is not the case, and Example 3 will illustrate this.

Let $\mathcal{F}_{\mathscr{A}}$ denote the set of (non-empty) faces of the polytope in $\mathbb{R}^s$ spanned by the image of $e_{\mathscr{A}}$. Our main result characterizes the support sets of the closure of $\mathcal{E}_{\mathscr{A}}$ in terms of $\mathcal{F}_{\mathscr{A}}$:

Theorem 2. *A subset $\mathcal{Y}$ of $\mathcal{X}_V$ is the support set of an element of $\operatorname{cl}\mathcal{E}_{\mathscr{A}}$ if and only if it is the preimage of a face $F \in \mathcal{F}_{\mathscr{A}}$ with respect to the map $e_{\mathscr{A}}$.*

The proof of the theorem will follow in Section 4.4. To illustrate the statement we consider the following instructive example:

Example 3. Consider the case of two binary units. We have $V = \{1, 2\}$, $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, and therefore $\mathcal{X}_V = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. The set of probability distributions is the three-dimensional simplex whose extreme points are the Dirac measures $\delta_{(x_1,x_2)}, x_1, x_2 \in \{0, 1\}$ (see Figure 1). As mentioned in Example 1 (2), we are going to discuss interactions of order one and two:

(1) For interactions of order one we have

$$\mathscr{A}_1 = \{\emptyset, \{1\}, \{2\}\} \,.$$

The exponential family $\mathcal{E}_1 := \mathcal{E}_{\mathscr{A}_1}$ coincides with the set of probability measures that factor over the two units. (It can be seen that $P \in \mathcal{E}_1 \Leftrightarrow P(x_1, x_2) = P_1(x_1)P_2(x_2)$ for all $x_1, x_2 \in \{0, 1\}$).

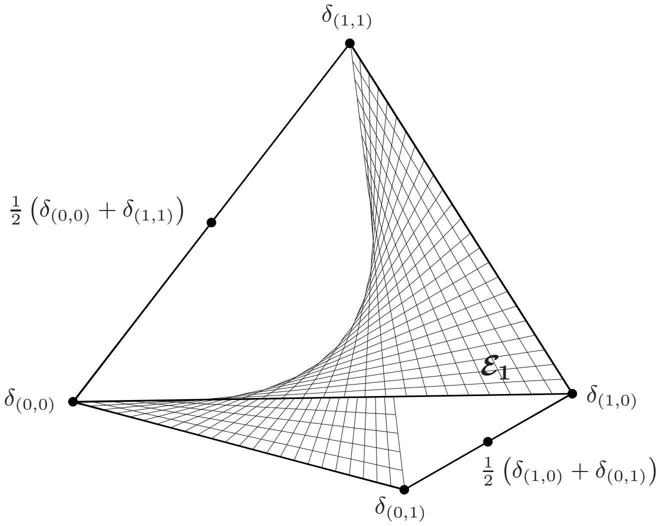

FIGURE 1. The exponential family $\mathcal{E}_1$ in the simplex of probability distributions.

The interaction space $\mathcal{I}_{\mathscr{A}_1}$ has dimension three, and one natural orthonormal basis (see Section 4.3) is the following:

$$\begin{aligned} e_{\emptyset} &= (1,1,1,1) \\ e_{\{1\}} &= (1,1,-1,-1) \\ e_{\{2\}} &= (1,-1,1,-1)\,. \end{aligned} \tag{3}$$

Here, the components are chosen with respect to the ordering $(e_{00}, e_{01}, e_{10}, e_{11})$ of the canonical basis of $\mathbb{R}^{(\{0,1\}^2)}$. The composed map is given as

$$e_{\mathscr{A}_1} : \mathcal{X}_V \to \mathbb{R}^2, \qquad x \mapsto (e_{\{1\}}(x), e_{\{2\}}(x))\,.$$

The image of that map consists of the four points $(-1,-1), (1,-1), (-1,1), (1,1)$ which have the square in $\mathbb{R}^2$ as their convex hull. Denoting the convex hull of points $p_1, \ldots, p_k$ by $[p_1, \ldots, p_k]$, we have the following (non-empty) faces in $\mathcal{F}_{\mathscr{A}_1}$:

$$F_1 = [(-1,-1), (-1,1), (1,-1), (1,1)]$$
$$F_2 = [(-1,-1), (-1,1)] \qquad F_3 = [(-1,-1), (1,-1)]$$
$$F_4 = [(-1,1), (1,1)] \qquad F_5 = [(1,-1), (1,1)]$$
$$F_6 = \{(-1,-1)\} \quad F_7 = \{(-1,1)\} \quad F_8 = \{(1,-1)\} \quad F_9 = \{(1,1)\}$$

The face F_1 is the square itself, F_2 to F_5 are the four edges, and F_6 to F_9 are the extreme points of the square. By Theorem 2, $\mathcal{Y}_i := e_{\mathscr{A}_1}^{-1}(F_i)$ are all support sets of probability measures in $\operatorname{cl}\mathcal{E}_1$ (compare with Figure 1):

$$\mathcal{Y}_1 = \{0,1\}^2$$
$$\mathcal{Y}_2 = \{(1,0),(1,1)\} \qquad \mathcal{Y}_3 = \{(0,1),(1,1)\}$$
$$\mathcal{Y}_4 = \{(0,0),(1,0)\} \qquad \mathcal{Y}_5 = \{(0,0),(0,1)\}$$
$$\mathcal{Y}_6 = \{(1,1)\} \quad \mathcal{Y}_7 = \{(1,0)\} \quad \mathcal{Y}_8 = \{(0,1)\} \quad \mathcal{Y}_9 = \{(0,0)\}\,.$$

(2) Now we consider the hypergraph of interactions of oder two, i.e.

$$\mathscr{A}_2 = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}.$$

The exponential family $\mathcal{E}_2 := \mathcal{E}_{\mathscr{A}_2}$ coincides with the whole simplex shown in Figure 1. The interaction space has dimension four, and the vector $e_{\{1,2\}} = (1,-1,-1,1)$ completes the basis (3) to an orthonormal basis of the space $\mathcal{I}_{\mathscr{A}_2}$. The image of $e_{\mathscr{A}_2}$ is given by $\{(-1,-1,1),(-1,1,-1),(1,-1,-1),(1,1,1)\}$ which is a subset of the extreme points of the cube in $\mathbb{R}^3$. It defines a simplex which is the image of the simplex in Figure 1 under the map $P \mapsto \mathbb{E}_P(e_{\mathscr{A}_2})$ (see Figure 2).

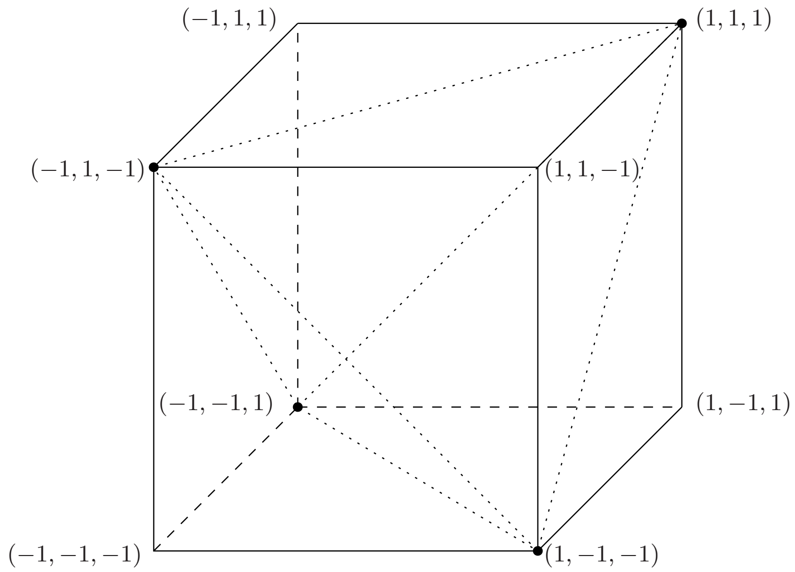

FIGURE 2. The convex hull of $\operatorname{im} e_{\mathscr{A}_2} = \operatorname{im}(e_{\{1\}}, e_{\{2\}}, e_{\{1,2\}})$ inside the cube in $\mathbb{R}^3$.

The faces in $\mathcal{F}_{\mathscr{A}_2}$ are given by

$$F_1 = [(-1,-1,1),(-1,1,-1),(1,-1,-1),(1,1,1)]$$
$$F_2 = [(-1,-1,1),(-1,1,-1),(1,1,1)] \quad F_3 = [(-1,-1,1),(-1,1,-1),(1,1,1)]$$
$$F_4 = [(-1,-1,1),(1,-1,-1),(1,1,1)] \quad F_5 = [(-1,1,-1),(1,-1,-1),(1,1,1)]$$
$$F_6 = [(-1,1,-1),(1,-1,-1)] \quad F_7 = [(1,1,1),(1,-1,-1)]$$
$$F_8 = [(1,1,1),(-1,1,-1)] \quad F_9 = [(1,1,1),(-1,-1,1)]$$
$$F_{10} = [(-1,1,=1),(1,-1,-1)] \quad F_{11} = [(-1,-1,1),(1,-1,-1)]$$
$$F_{12} = \{(-1,1,-1)\} \quad F_{13} = \{(-1,-1,1)\}$$
$$F_{14} = \{(1,-1,-1)\} \quad F_{15} = \{(1,1,1)\}\ .$$

The face F_1 is the nothing but the simplex in Figure 2, F_2 to F_5 are its four triangles, F_6 to F_{11} are the six edges, and the remaining faces are the extreme points. The preimages of these faces with respect to the map $e_{\mathscr{A}_2}$ are exactly the 15 non-empty subsets of $\{0,1\}^2$.

4. Proof of the Main Result

In this section, we are going to prove our main result in several steps. In the first step, we review a result of [BN, CMb] on closures of exponential families. The second step deals with the decomposition of the interaction spaces $\mathcal{I}_{\mathscr{A}}$ into orthogonal components, and a natural basis is constructed. Based on these two steps, finally, the proof of Theorem 2 is a straightforward implication.

4.1. **Closures of exponential families.** In a recent paper [CMb] several notions of closures of exponential families were studied. As a special case of these considerations, namely the case of finite configuration spaces, a classical result of [BN, pp. 154-155] appears. It is shown that $\operatorname{cl}\mathcal{E}$ can be written as a union of certain families with exponential structure. In order to review this result we have to introduce some further notation. Consider a Gibbs measure

$$P_{\theta,f}(x) := \frac{1}{Z}\exp\big(\langle\theta, f(x)\rangle\big) \qquad (x \in \mathcal{X})$$

on a non-empty finite set $\mathcal{X}$, where $Z = \sum_{x\in\mathcal{X}}\exp\big(\langle\theta, f(x)\rangle\big)$ is a normalization, $f : \mathcal{X} \to \mathbb{R}^d$ is a statistic, and $\theta \in \mathbb{R}^d$ is a vector of coefficients. As θ ranges over $\mathbb{R}^d$, the $P_{\theta,f}$ form an exponential family which we denote by

$$\mathcal{E}_f := \left\{P_{\theta,f} : \theta \in \mathbb{R}^d\right\}.$$

Since $\mathcal{X}$ is finite, the image of f is a finite subset of $\mathbb{R}^d$, and its convex hull $\mathcal{F}$ is a polytope. For every non-empty face F of $\mathcal{F}$ define

$$\mathcal{Y}^F := \{x \in \mathcal{X} : f(x) \in F\} = f^{-1}(F). \tag{4}$$

Finally, for every $\mathcal{Y}^F$ consider the restriction

$$\mathcal{E}_{\mathcal{Y}^F,f} := \begin{cases} \frac{1}{Z^F}\exp\left(\langle\theta, f(x)\rangle\right), & \text{if } x \in \mathcal{Y}^F \\ 0 & , \text{ otherwise}\end{cases}, \quad Z^F := \sum_{x'\in\mathcal{Y}^F}\exp\big(\langle\theta, f(x')\rangle\big).$$

The following statement is a special case of a more general result of [CMb]:

Theorem 4.

$$\operatorname{cl}(\mathcal{E}_f) = \bigcup_F \mathcal{E}_{\mathcal{Y}^F,f}$$

Remark. The statement in [CMa, CMb] refers to general measurable spaces and corresponding reference measures within the context of various notions of closure. In our case of finite $\mathcal{X}$ all notions coincide with the natural topological closure.

4.2. **Orthogonal decomposition of the interaction space.** In this section, we decompose the interaction space $\mathcal{I}_{\mathscr{A}}$ into orthogonal components by means of the construction of a basis. We then have an explicit description of the statistic that generates $\mathcal{E}_{\mathscr{A}}$ and can apply Theorem 4 to examine the closure $\operatorname{cl}\mathcal{E}_{\mathscr{A}}$. In what follows, all concepts of projections and orthogonality are meant with respect to the scalar product

$$\langle f, g\rangle := \frac{1}{2^N}\sum_{x\in\mathcal{X}_V} f(x)g(x).$$

In previous work [DS, Lau, AK], the spaces of *pure interactions among elements of $A \subseteq V$* were defined as follows:

$$\tilde{\mathcal{I}}_A := \mathcal{I}_A \cap \left(\bigcap_{B \subsetneq A} \mathcal{I}_B^{\perp} \right). \tag{5}$$

This implies an orthogonal decomposition

$$\mathcal{I}_A = \bigoplus_{B \subseteq A} \tilde{\mathcal{I}}_B\,, \tag{6}$$

where $\dim \tilde{\mathcal{I}}_A = 1$ (see [AK]). In particular, $\mathbb{R}^{\mathcal{X}_V} = \bigoplus_{A \subseteq V} \tilde{\mathcal{I}}_A$.

4.3. **A basis of the pure interaction spaces.** In Proposition 6, we prove that the functions e_A, $A \subseteq V$, which are defined according to (1) form an orthonormal basis of the interaction space $\mathcal{I}_{\mathscr{A}}$. To this end we need the following lemma:

Lemma 5. *Let $\emptyset \neq A \subseteq V$, then*

$$\sum_{x \in \mathcal{X}_V} (-1)^{E(A,x)} = 0.$$

Proof. Let i be an element of A, and define

$$\mathcal{X}_- := \{x \in \mathcal{X}_V \;:\; X_i(x) = 1\}, \qquad \mathcal{X}_+ := \{x \in \mathcal{X}_V \;:\; X_i(x) = 0\}\,.$$

Obviously, $E(A,x) = E(A \setminus \{i\}, x) + 1$ if $x \in \mathcal{X}_-$, and $E(A,x) = E(A \setminus \{i\}, x)$ if $x \in \mathcal{X}_+$. This implies

$$\sum_{x \in \mathcal{X}_V} (-1)^{E(A,x)} \;=\; \sum_{x \in \mathcal{X}_+} (-1)^{E(A \setminus \{i\},x)} - \sum_{x \in \mathcal{X}_-} (-1)^{E(A \setminus \{i\},x)} \;=\; 0$$

□

Proposition 6. *The vectors $(e_A)_{A \in \mathscr{A}}$ form an orthonormal basis of $\mathcal{I}_{\mathscr{A}}$.*

Proof. The e_A are normalized with respect to our scalar product. Since $\mathscr{A}$ is assumed to be complete, we have the decomposition

$$\mathcal{I}_{\mathscr{A}} = \bigoplus_{A \in \mathscr{A}} \tilde{\mathcal{I}}_A\,,$$

where $\dim \tilde{\mathcal{I}}_A = 1$, and it is sufficient to show that $e_A \in \tilde{\mathcal{I}}_A$. The case of $e_\emptyset$ is clear since $e_\emptyset = \sum_{x \in \mathcal{X}_V} e_x$ and $\tilde{\mathcal{I}}_\emptyset = \mathcal{I}_\emptyset$ is the space of constants. Now let A be non-empty and observe that, denoting by Π_B the projection onto $\mathcal{I}_B$, the definition (5) of the pure interaction spaces can be reformulated as

$$f \in \tilde{\mathcal{I}}_A \quad \Longleftrightarrow \quad f \in \mathcal{I}_A \text{ and } \Pi_B f = 0 \text{ for all } B \subsetneq A\,. \tag{7}$$

The projection onto the space $\mathcal{I}_A$ is given by

$$\Pi_A(f)(x_A, x_{V \setminus A}) = \frac{1}{2^{|V \setminus A|}} \sum_{x'_{V \setminus A} \in \mathcal{X}_{V \setminus A}} f(x_A, x'_{V \setminus A}).$$

We now check property (7):

$$\Pi_A(e_A)(x_A, x_{V\setminus A}) = \frac{1}{2^{|V\setminus A|}} \sum_{x'_{V\setminus A}\in\mathcal{X}_{V\setminus A}} e_A(x_A, x'_{V\setminus A}) = e_A\,.$$

This follows from the fact that changing x outside A does not alter the values of e_A. On the other hand, for a given subset $B \subsetneq A$ we have

$$\begin{aligned}\Pi_B(e_A)(x_B, x_{V\setminus B}) &= \frac{1}{2^{|V\setminus B|}} \sum_{x'_{V\setminus B}\in\mathcal{X}_{V\setminus B}} e_A(x_B, x'_{V\setminus B})\\ &= \frac{1}{2^{|V\setminus B|}} \sum_{x'_{V\setminus B}\in\mathcal{X}_{V\setminus B}} (-1)^{E(A,(x_B,x'_{V\setminus B}))}\\ &= \frac{1}{2^{|V\setminus B|}} \sum_{x'_{V\setminus B}\in\mathcal{X}_{V\setminus B}} (-1)^{(E(B,(x_B,x_{V\setminus B}))+E(A\setminus B,(x_B,x'_{V\setminus B})))}\\ &= 0\end{aligned}$$

The last equality holds since $(-1)^{E(B,(x_B,x'_{V\setminus B})}$ does not depend on $x'_{V\setminus B}$ and with $A\setminus B \neq \emptyset$ Lemma 5 implies

$$\sum_{x'\in\mathcal{X}_{V\setminus B}} (-1)^{E(A\setminus B,x')} = 0\,.$$

□

Remark (Orthonormal Basis). Since the e_A form an orthonormal basis, one can invert the transformation to find

$$e_x = \frac{1}{2^N} \sum_{A\in 2^V} (-1)^{E(A,x)} e_A\,,$$

and obviously none of the coefficients is zero.

Combining the results of the previous sections we can now proceed with proving Theorem 2.

4.4. **Proof.**

Proof of Theorem 2. From the above discussion it is clear that the exponential family under consideration can be written as

$$\mathcal{E}_{\mathscr{A}} = \left\{ \frac{1}{Z} \exp\left\{ \sum_{i=1}^{s} \theta^{A_i} e_{A_i}(x) \right\} : \theta = (\theta^{A_i})_{i=1,\ldots,s} \in \mathbb{R}^s \right\}$$

Thus, the exponential family has the form of Theorem 4

$$\mathcal{E}_{\mathscr{A}} = \bigcup_{F\in\mathcal{F}_{\mathscr{A}}} \mathcal{E}_{\mathcal{Y}^F, e_{\mathscr{A}}}$$

with the definition (4) of $\mathcal{Y}^F$ now becoming

$$\mathcal{Y}^F = e_{\mathscr{A}}^{-1}(F)\,.$$

□

5. Conclusions

Applying general results on closures of exponential families from [BN, CMb] we studied the closure of hierarchical models including graphical models [Lau]. Using a natural orthonormal basis of the corresponding interaction space allows for an explicit description of this closure. We hope that this description in terms of linear algebra will lead to a constructive method for specifying closures of hierarchical models.

Acknowledgments

Nihat Ay thanks the Santa Fe Institute for supporting him as an external professor.

References

[Am] S. Amari. *Information geometry on hierarchy of probability distributions*, IEEE Trans. IT **47**, 1701-1711 (2001)

[AK] N. Ay, A. Knauf. *Maximizing multi-information*, Kybernetika (2006), in press

[BN] O. Barndorff-Nielsen, *Information and exponential families statistical theory*, Wiley, 1978

[CMa] I. Csiszár, F. Matúš, *Information closure of exponential families and generalized maximum likelihood estimates* IEEE International Symposium on Information Theory 2002. Proceedings. (2002)

[CMb] I. Csiszár, F. Matúš, *Information Projections Revisited* IEEE Transactions on Information Theory, **49**(6): 1474-1490 (2003)

[DS] J.N. Darroch, T.P. Speed *Additive and multiplicative models and interactions* The Annals of Statistics, **11**(3): 724-738 (1983)

[Lau] S.L. Lauritzen *Graphical Models* Oxford Statistical Science Series. Oxford University Press (1996)

[St] M. Studeny. *Probabilistic Conditional Independence Structures*. Series: Information Science and Statistics, Springer 2005.

E-mail address: {kahle, nay}@mis.mpg.de

[1]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig,Germany

[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA