Caputo:

Graphical models with Koehler Symanowski distributions

Sonderforschungsbereich 386, Paper 104 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Graphical models with Koehler Symanowski distributions

By Angelika Caputo
University of Munich, Institute of Statistics,
Ludwigstr. 33, D-80539 Munich, Germany

Summary

In this paper, a multivariate distribution family introduced by Koehler and Symanowski (1995) is discussed as alternative assumption for graphical models which are typically connected with Conditional–Gaussian distributions. For that purpose, certain requirements which have to be fulfilled when formulating graphical models are checked. This leads to the introduction of graphical models with Koehler Symanowski distributions which are then investigated regarding some basic properties known for Gaussian graphical models.

1 Introduction

Graphical models are mainly defined on so–called concentration graphs and the family of Conditional–Gaussian (CG) distributions which allow to have simultaneously continuous and discrete variables under investigation (see for example Lauritzen, 1996, Lauritzen and Wermuth, 1989, Wermuth and Lauritzen, 1990). The family of CG distributions includes the multivariate normal and the multinomial distributions as special cases and satisfies the equivalence of the Markov properties for concentration graphs. The equivalence of pairwise, local, and global Markov properties is required when concluding from properties of the graph, like separation of two sets of vertices A and B by a third set C, to those of the joint distribution, like the conditional independence of X_A and X_B given X_C briefly written as $A \perp\!\!\!\perp B \mid C$ (Dawid, 1979, 1980). That means, missing edges in the underlying graph can be correctly interpreted as conditional independences. The conditions for marginalization, collapsibility, and decomposition of the ML estimation for graphical models with CG distribution can be found in Frydenberg (1990) and Frydenberg and Lauritzen (1989).

For so-called covariance graphs which represent marginal independences instead of conditional ones appropriate independence properties, also called Markov properties, have been proposed (Cox and Wermuth, 1993, Kauermann, 1996). For such graphs, the above mentioned equivalence holds under restrictive assumptions which are only fulfilled in special cases. Further properties, like marginalization and collapsibility, are merely shown for the multivariate normal distribution (Kauermann, 1996).

In this paper, graphical models on covariance graphs are defined using a joint distribution that represents an alternative to the usually applied distribution family, namely the family of multivariate normal distributions. Thus, the existing theory of graphical models is extended in two respects. On the one hand, another multivariate distribution is looked at. On the other hand, we focus on covariance graphs rather than on concentration graphs. The aim of this approach is to broaden the applicability of graphical models to practical problems where the multivariate normal or CG distributions are not adequate, as for instance in multivariate event history analysis. The class of multivariate distribution families introduced by Koehler and Symanowski (1995), here called KS distributions, allows to model complex associations between arbitrary subsets of the variable set as well as pairwise independences in the margins. These distributions can be constructed based on almost any given univariate marginal distributions by adding interaction terms and therefore promise a wide field of applications as, for instance, in the framework of graphical modeling.

The paper is organized as follows: In Section 2 some basic notations and properties concerning graphical models are given. The next section is devoted to the definition of the class of KS distributions. In addition, some important properties concerning conditional and marginal independence are established. Section 4 considers graphical models with KS distributions. The paper ends with a discussion giving among others an outlook on further research questions.

2 Graphical models

2.1 Basic terminology

A graph $\mathcal{G} = (V, E)$ is given by a set of vertices V representing the variables and a set of edges $E \subset V \times V$ with $(i, i) \notin E$ for all $i \in V$ reflecting associations among the variables. We identify the set of vertices with the index set of the vector X_V. The set E consists of ordered pairs (i, j), $i, j \in V$. We only consider so–called undirected graphs where $(i, j) \in E \Rightarrow (j, i) \in E$ holds, i.e. only symmetric associations are dealt with. For $A \subset V$, we define a subgraph $\mathcal{G}_A$ as $\mathcal{G}_A = (A, E \cap (A \times A))$. Vertices $i, j \in V$ with $(i, j) \in E$ are called neighbours. The boundary of a vertice $i \in V$ is given by $bd(i) = \{j \in V \mid (i, j) \in E\}$ and for any $A \subset V$ we define the boundary of A as $bd(A) = \{j | j \in bd(i), i \in A\} \setminus A$. A path from i to j is given by a sequence $i = i_0, i_1, \ldots, i_n = j$ with $(i_m, i_{m+1}) \in E$ for

$m = 0, \ldots, n-1$. For disjoint subsets A, B, C of V we say C separates A and B if every path from a vertice $i \in A$ to $j \in B$ includes at least one vertice $k \in C$. The subset $A \subset V$ is called complete if $(i, j) \in E$ for all $i, j \in A$ with $i \neq j$.

2.2 Markov properties

The basic idea of graphical models is to use a graph to represent particular properties of a family of multivariate distributions. A concentration graph reflects pairwise conditional independences: A missing edge between two vertices implies that the two variables are conditionally independent given all remainig variables, i.e. vertices in the graph. In a covariance graph missing edges are interpreted as marginal independences between pairs of variables. This property is called pairwise Markov property. Besides the pairwise Markov property the graphical representation suggests additional independence statements. As far as concentration graphs are concerned it seems to be sensible to say that a variable X_i is conditionally independent from all variables X_j which are not connected with X_i given the boundary of X_i what is also known as local Markov property. A detailed discussion of the Markov properties for concentration graphs can be found in Frydenberg (1990) and Lauritzen (1996, pp. 32). The corresponding properties for covariance graphs (Kauermann, 1996) are as follows:
A family $\mathcal{P}$ of distributions on a covariance graph $\mathcal{G} = (V, E)$ satisfies the

(i) pairwise Markov property according to $\mathcal{G}$ if $\{i\} \perp\!\!\!\perp \{j\}$ holds for all $(i, j) \notin E$ with $i \neq j$,

(ii) local Markov property according to $\mathcal{G}$ if $\{i\} \perp\!\!\!\perp V \setminus (\{i\} \cup bd(i))$ holds for all $i \in V$,

(iii) global Markov property according to $\mathcal{G}$ if $A \perp\!\!\!\perp B \mid C$ holds for all disjoint subsets A, B, C of V whenever A and B are separated by the set $D = V \setminus (A \cup B \cup C)$.

The global Markov property is the strongest of the three Markov properties. The local and the pairwise Markov property follow from the global. In addition, the local implies the pairwise Markov property, i.e. (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The implication (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) is not given in general but is a minimum requirement for the correct interpretation of a graph in the sense of a graphical model. This is fulfilled for concentration graphs under rather weak assumptions (Frydenberg, 1990). For covariance graphs, however, the conditions are more restrictive as can be seen from Proposition 2.1 (Kauermann, 1996).

Proposition 2.1 *Let A, B, C be disjoint subsets of V. The three Markov properties for covariance graphs are equivalent if for all $P \in \mathcal{P}$ the implication*

$$(A \perp\!\!\!\perp B \text{ and } A \perp\!\!\!\perp C) \Longrightarrow A \perp\!\!\!\perp (B \cup C), \tag{2.1}$$

is fulfilled.

Consider a covariance graph $\mathcal{G}$ and a family of distributions $\mathcal{P}$ for which (2.1) is fulfilled. $\mathcal{P}$ is called $\mathcal{G}$ Markov or a graphical ($\mathcal{G}$ Markov) model and is denoted as $\mathcal{M}(\mathcal{G})$ whenever $\mathcal{P}$ holds one of the above Markov properties. This provides the following interpretation. An independence statement which can be read off the graph is valid for each element of $\mathcal{M}(\mathcal{G})$. But $\mathcal{M}(\mathcal{G})$ usually contains distributions which fulfill more independences than those represented by the graph. An important question is whether the graph represents all independence statements which are fulfilled for each distribution in $\mathcal{M}(\mathcal{G})$.

Definition 2.2 *A graphical $\mathcal{G}$ Markov model $\mathcal{M}(\mathcal{G})$ is called $\mathcal{G}$ Markov perfect if $A \perp\!\!\!\perp B \mid C$ for all $P \in \mathcal{M}(\mathcal{G})$ implies that the sets A and B are separated by $D = V \setminus (A \cup B \cup C)$ in the covariance graph $\mathcal{G}$.*

3 Koehler Symanowski distributions

Let $V = \{1, \ldots, p\}$ be an index set and $\mathcal{V}$ the powerset of V. Let $X = X_V = (X_1, \ldots, X_p)^T$ denote a vector of random variables with marginal cumulative distribution functions (cdfs) $F_i(\cdot)$, $i \in V$. The joint distribution of X is assumed to be given by the following cdf

$$F(x_1, \ldots, x_p) = \prod_{i \in V} F_i(x_i) \prod_{I \in \mathcal{I}} c_I(x)^{-\alpha_I}. \tag{3.2}$$

For all sets $I \in \mathcal{I} = \{I \in \mathcal{V} \text{ with } |I| \geq 2\}$ let $\mathbb{R} \ni \alpha_I \geq 0$ and for all $i \in V$ let $\mathbb{R} \ni \alpha_i > 0$ with $\alpha_{i+} = \sum_{I \in \mathcal{V}, i \in I} \alpha_I < \infty$. For all $I \in \mathcal{I}$ the factors $c_I(x)$ in (3.2) are defined as

$$c_I(x) = \sum_{i \in I} \{\prod_{\substack{j \in I \\ j \neq i}} u_j(x_j)\} - (|I| - 1) \prod_{i \in I} u_i(x_i)$$

with $u_i(x_i) = F_i(x_i)^{\frac{1}{\alpha_{i+}}}$ for all $i \in V$. The distribution of X will be called KS distribution in the following. The definition given here slightly differs from that in Koehler and Symanowski (1995), since the parameters α_i are assumed to be strictly positive whereas in the original definition the case $\alpha_i = 0$ is included. The consequences of this restriction are described in Caputo (1998). The structure of the cdf is fairly easy. It factorizes into the product of the marginal cdfs and a product of association terms. If we assume that marginal density functions $f_i(\cdot)$ exist for all $i \in V$ it can easily be shown that the joint density function also exists (Koehler and Symanowski, 1995). However, in contrast to the cdf the functional representation of the density function is rather complicated. Besides the product of the marginal densities there are more complex factors with additive components which come into play due to the first derivation of F.

In the framework of graphical models, properties related to marginal and conditional independence are of interest as well as, for instance, the type of marginal and conditional distributions. The former are important whenever Markov properties of the distribution family are investigated. The family of KS distributions is closed under marginalization (Caputo, 1998) whereas conditional distributions are in general not again of KS type (Koehler and Symanowski, 1995). In the following the conditions for marginal and conditional independence are summarized. For more details we refer to Caputo (1998). Consider a vector $X = X_V$ with joint KS distribution and a partition of V into the sets A, B, and C. It can be shown that X_A and X_B are marginally and conditionally independent whenever for all $i \in A$ and $j \in B$ the condition $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i \in I$ and $j \in I$ is fulfilled. In addition, let X_A and X_B be marginally or conditionally independent and assume each X_i for $i \in V$ to be not degenerated, i.e. there exists at least one $\tilde{x}_i \in I\!R$ for which the corresponding cdf takes a value $F_i(\tilde{x}_i) = y_i$ with $0 < y_i < 1$. Then, for all $i \in A$ and $j \in B$ it holds that $\alpha_I = 0$ whenever $i \in I$ and $j \in I$. This proposition indicates that first independence between subvectors is derived by setting parameters to zero and second conditional and marginal independence cannot be distinguished.

4 Graphical models with KS distributions

4.1 Markov properties

In this section, it is discussed whether the family of KS distributions is a suitable assumption for graphical models and thus, a possible alternative to the family of CG distributions. As mentioned above the equivalence of the Markov properties is a basic requirement which is formulated in Theorem 4.1. In the following, let $\mathcal{P}$ denote the family of KS distributions with non degenerated marginal distributions.

Theorem 4.1 *Assuming $\mathcal{P}$, the pairwise, local, and global Markov property for covariance graphs are equivalent.*

Proof:
Let $\mathcal{G} = (W, E)$ be a covariance graph, A, B, C arbitrary subsets of W and $V = A \cup B \cup C$. Following Proposition 2.1 it has to be shown that for all $P \in \mathcal{P}$ the implication

$$(A \perp\!\!\!\perp B \text{ and } A \perp\!\!\!\perp C) \quad \Longrightarrow A \perp\!\!\!\perp (B \cup C)$$

is fulfilled. Since the family of KS distributions is closed under marginalization the distribution of X_V is again of KS type. For all $i \in V$, the distribution of X_i is assumed to be not degenerated. Thus, $A \perp\!\!\!\perp B$ yields that $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i, j \in I$ and $i \in A$,

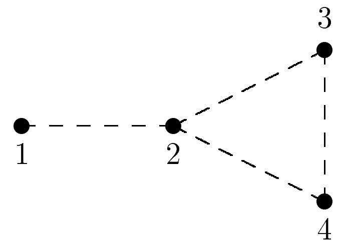
$j \in B$. Correspondingly, $A \perp\!\!\!\perp C$ implies $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i, j \in I$ and $i \in A$, $j \in C$. This can also be written as $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i, j \in I$ and $i \in A$, $j \in B \cup C$ which leads to $A \perp\!\!\!\perp (B \cup C)$. □

The equivalence of the Markov properties establishes a connection between a graph and a distribution family $\mathcal{P}$. If, for instance, $\mathcal{G}$ is a covariance graph the restrictions on the association parameters of the corresponding family of KS distributions are obtained as $\alpha_{ij} = 0$ for all $(i, j) \notin E$. In addition, for all pairs (i, j) with $\alpha_{ij} = 0$ it can be concluded that $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i \in I$ and $j \in I$. If vice versa a family $\mathcal{P}$ of KS distributions is given the corresponding covariance graph is built up connecting the vertices in I for all $I \in \mathcal{I}_2 = \{I \in \mathcal{V}, |I| = 2\}$ with $\alpha_I \neq 0$. The set of edges E reads as

$$E = \{(i, j) |\, \exists I \in \mathcal{I} \text{ with } \alpha_I \neq 0 \text{ and } i, j \in I\}.$$

A set of vertices $I \in \mathcal{I}$ with $\alpha_I \neq 0$ is complete in the corresponding graph $\mathcal{G}$.

Example 4.2 *Consider the covariance graph $\mathcal{G} = (V, E)$ with $V = \{1, 2, 3, 4\}$ and $E = \{(1, 2), (2, 1), (2, 3), (2, 4), (3, 2), (3, 4), (4, 2), (4, 3)\}$. Note that edges are represented as dashed lines following the conventions of covariance graphs (Cox and Wermuth, 1993).*



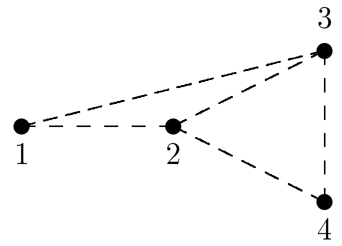
The family $\mathcal{P}$ of four–dimensional KS distributions with $\alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0$, association parameters α_{12}, α_{23}, α_{24}, α_{34}, $\alpha_{234} > 0$, and $\alpha_{13} = \alpha_{14} = \alpha_{123} = \alpha_{124} = \alpha_{134} = \alpha_{1234} = 0$ is $\mathcal{G}$ Markov. The Markov properties of $\mathcal{P}$ are given as

$$\{1\} \perp\!\!\!\perp \{3\},\ \{1\} \perp\!\!\!\perp \{4\},\ \{1\} \perp\!\!\!\perp \{3, 4\},\ \{1\} \perp\!\!\!\perp \{3\} \mid \{4\},\ \textit{and}\ \{1\} \perp\!\!\!\perp \{4\} \mid \{3\}.$$

The next question addresses again whether the graph represents all independence properties which hold for each element of $\mathcal{P}$. It is easy to see that a family of KS distributions implies in general more marginal and conditional independences as it can be concluded from the covariance graph by means of the three Markov properties.

Theorem 4.3 *The family $\mathcal{P}$ is not $\mathcal{G}$ Markov perfect.*

Proof:
Consider, for instance, the family $\mathcal{P}$ of KS distributions with $\alpha_i > 0$ for $i = 1, \ldots, 4$, association parameters α_{12}, α_{13}, α_{23}, α_{24}, α_{34}, α_{123}, $\alpha_{234} > 0$ and $\alpha_{14} = \alpha_{124} = \alpha_{134} = \alpha_{1234} = 0$. Define the sets $A = \{1\}$, $B = \{4\}$, $C = \{3\}$ and $D = \{2\}$. It holds that $A \perp\!\!\!\perp B \mid C$ because $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $1 \in I$ and $4 \in I$.
Suppose, $\mathcal{P}$ is $\mathcal{G}$ Markov perfect. Thus, the sets A and B have to be separated by the set $V \setminus (A \cup B \cup C) = \{2\} = D$ in the corresponding graph $\mathcal{G}$ which is, however, given as



and thus demonstrates that the sets $A = \{1\}$ and $B = \{4\}$ are not separated by the set $D = \{2\}$. $\square$

To cover all the independence properties which are fulfilled by the elements of a family $\mathcal{P}$ it seems to be necessary to introduce an additional Markov property.

Definition 4.4 *Let $\mathcal{P}$ be a family of distributions and $\mathcal{G} = (V, E)$ be a covariance graph. $\mathcal{P}$ is called total $\mathcal{G}$ Markov, i.e. fulfills the total Markov property for covariance graphs if*

$$A \perp\!\!\!\perp B \mid C \quad \textit{for all} \quad C \subseteq V \setminus (A \cup B)$$

whenever the disjoint subsets A, B of V are separated by $D = V \setminus (A \cup B)$.

Before investigating whether the total Markov property suffices to gather all independences of the underlying distribution family it has to be checked if the equivalence of the now four Markov properties is still given.

Theorem 4.5 *For the family $\mathcal{P}$ the global and the total Markov property for covariance graphs are equivalent.*

Proof:
Let $\mathcal{G} = (V, E)$ be a graph where the disjoint sets A, $B \subset V$ are separated by $D \subseteq V \setminus (A \cup B)$. In the first step it is shown that the global Markov property follows from the total Markov property.

Assume that $\mathcal{P}$ fulfills the total Markov property of $\mathcal{G}$, i.e. for all $C \subseteq D$ it holds that $A \perp\!\!\!\perp B \mid C$. This conditional independence holds particularly for all subsets C of D with $S = D \setminus C$ separates A and B and thus the global Markov property is fulfilled.
Now, it is shown that the total Markov property follows from the pairwise Markov property. Suppose $\mathcal{P}$ to be pairwise Markov with regard to $\mathcal{G}$. Theorem 4.1 then implies that $\mathcal{P}$ is $\mathcal{G}$ Markov, i.e. the local and global Markov properties of $\mathcal{G}$ are fulfilled, too. The assumption that the sets A and B are separated by $D \subseteq V \setminus (A \cup B)$ yields that there exists no path from $i \in A$ to $j \in B$ which only contains vertices of $A \cup B$. In particular, there exists no pair $(i, j) \in E$ with $i \in A$ and $j \in B$. Thus, the pairwise Markov property implies $\{i\} \perp\!\!\!\perp \{j\}$ for all $i \in A$, $j \in B$ and therefore $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i, j \in I$. As described in Section 3 this condition leads to $A \perp\!\!\!\perp B \mid C$ for all $C \subseteq D$. □

Since the above four Markov properties for covariance graphs are equivalent for the family of KS distributions $\mathcal{P}$ we call $\mathcal{P}$ $\mathcal{G}$ Markov whenever it satisfies one of the Markov properties.

Definition 4.6 *For a covariance graph $\mathcal{G} = (V, E)$ the set of all $P \in \mathcal{P}$ which are $\mathcal{G}$ Markov are called graphical $\mathcal{G}$ Markov model with KS distribution and is denoted with $\mathcal{KS}(\mathcal{G})$.*

We now come back to the discussion of the problem of Markov perfectness in the situation of a $\mathcal{G}$ Markov model with KS distribution starting with a more general definition of Markov perfectness.

Definition 4.7 *A graphical $\mathcal{G}$ Markov model with KS distribution $\mathcal{KS}(\mathcal{G})$ is said to be $\mathcal{G}$ Markov perfect in the wider sense if the sets A and B are separated by $D = V \setminus (A \cup B)$ in $\mathcal{G}$ whenever for disjoint subsets A, B, C of V the condition*

$$A \perp\!\!\!\perp B \mid C \quad \textit{for all } P \in \mathcal{KS}(\mathcal{G})$$

holds.

Theorem 4.8 *A graphical $\mathcal{G}$ Markov model with KS distribution $\mathcal{KS}(\mathcal{G})$ is $\mathcal{G}$ Markov perfect in the wider sense.*

<u>Proof:</u>
Let $A, B, C \subset V$ be disjoint sets. Assume for all $P \in \mathcal{P}$ that $A \perp\!\!\!\perp B \mid C$ and in addition that the sets A and B are not separated by $D = V \setminus (A \cup B)$ in the corresponding graph $\mathcal{G} = (V, E)$. Thus, there exists a pair $(i, j) \in E$ with $i \in A$ and $j \in B$ and it holds that $\{i\} \not\perp\!\!\!\perp \{j\}$. This leads to $\alpha_{ij} \neq 0$ which implies $A \not\perp\!\!\!\perp B \mid C$ but disagrees with the above assumption. □

A consequence of this property is that the separating set itself does not play an important role. The fact that there is no direct path between the sets A and B suffices to determine the Markov properties. Whenever there exists a set $C \subseteq V \setminus (A \cup B)$ which is not necessarily separating such that $A \perp\!\!\!\perp B \mid C$ it follows that there is no edge between a vertice i of A and a vertice j of B. Thus, the sets A and B are automatically separated by the set $V \setminus (A \cup B)$. This property will again be discussed later when the collapsibility of a graphical model with KS distribution is investigated. Here, it is of interest whether for a subset A of V the family of marginal distributions of X_A fulfill the Markov properties of the subgraph $\mathcal{G}_A$. In general, this is not the case since it is possible that a set C separates the sets A_1 and A_2 in the subgraph $\mathcal{G}_A$ but not in the original graph $\mathcal{G}$.

Up to now the focus has been on Markov properties of covariance graphs. If concentration graphs are considered in connection with KS distributions an interesting and striking observation can be made.

Theorem 4.9 *Let $\mathcal{G} = (V, E)$ be a graph. $\mathcal{P}$ fulfills the Markov properties of the covariance graph $\mathcal{G}$ if and only if $\mathcal{P}$ fulfills the Markov properties for the concentration graph $\mathcal{G}$.*

Proof:

First, it is shown that $\mathcal{P}$ fulfills the Markov properties of the concentration graph whenever $\mathcal{P}$ fulfills the Markov properties for the covariance graph. For that purpose, assume that the sets A and B are separated by C in the graph. Thus, the set $V \setminus (A \cup B)$ also separates A and B. Obviously, it holds that $C \subset V \setminus (A \cup B)$, and the total Markov property implies $A \perp\!\!\!\perp B \mid C$ which yields the global Markov property for concentration graphs.

Now, we assume that $\mathcal{P}$ holds the Markov properties for concentration graphs, i.e. $\mathcal{P}$ is in particular pairwise $\mathcal{G}$ Markov with respect to a concentration graph $\mathcal{G}$. This implies for all $(i, j) \notin E$ that $\{i\} \perp\!\!\!\perp \{j\} \mid V \setminus \{i, j\}$ and, because of the equivalence of marginal and conditional independence $\{i\} \perp\!\!\!\perp \{j\}$. Thus, $\mathcal{P}$ is pairwise $\mathcal{G}$ Markov with respect to the covariance graph $\mathcal{G}$ and following Theorem 4.1 $\mathcal{G}$ Markov. $\square$

The above theorem shows that the family of KS distributions cannot capture situations in which an observed marginal independence among subvectors does not imply a conditional independence given the remaining components or vice versa. These phenomena are well–known as Simpson's paradox which is described in detail in Simpson (1951), Dawid (1979), and Blyth (1972). The paradox which is in fact not really paradox can not occur in KS distributions. In other words, Simpson's paradox cannot be modeled assuming this distribution family.

Consider the subfamily of the KS distributions which arises if a priori all association parameters α_I with $|I| > 2$ equal zero. Strictly speaking, only second order associations are taken into account. This subfamily belongs to another distribution family described in Marshall and Olkin (1988). All members of this family are so–called associated distributions. A property of associated distributions is that all partial correlations are non–negative (Barlow and Proschan, 1975). Thus, the extremer version of Simpson's paradox cannot occur: a change of the sign of the partial correlation coefficient when the set of variables in the conditioning is changed. A similar situation is discussed in Cox and Wermuth (1994) for the family of quadratic exponential binary distributions. The authors compare this phenomenon with properties of the so–called multivariate MTP_2 normal distributions (Karlin and Rinott, 1980) which also share the occurence of non–negative partial correlation coefficients with the above described distributions. In addition, it should be noticed that the family of quadratic exponential binary distributions is one of the rare distribution families fulfilling the Markov properties of covariance graphs.

In the following, the above described phenomenon is embedded in a more general context.

Definition 4.10 *A distribution family $\tilde{\mathcal{P}}$ of $X = X_V$ is said to be resistant against Simpson's paradox if for disjoint subsets A, B, and C of V the following holds:*

(i) $A \perp\!\!\!\perp B$ implies $A \perp\!\!\!\perp B \mid C$ for all $C \subset V \setminus (A \cup B)$,

(ii) $A \perp\!\!\!\perp B \mid C$ for a set $C \subset V \setminus (A \cup B)$ implies $A \perp\!\!\!\perp B$

for all $P \in \tilde{\mathcal{P}}$.

Thus, the above Theorem 4.9 can also be generalized.

Theorem 4.11 *For a family of distributions $\tilde{\mathcal{P}}$ which is resistant against Simpson's paradox the global Markov properties for concentration and covariance graphs are equivalent.*

Proof:
Let A, B, C, and D be disjoint sets such that C separates the sets A and B and that $D = V \setminus (A \cup B \cup C)$.
Assume first $\tilde{\mathcal{P}}$ to be $\mathcal{G}$ Markov with respect to a concentration graph, i.e. it holds that $A \perp\!\!\!\perp B \mid C$. Part (ii) of Definition 4.10 yields $A \perp\!\!\!\perp B$. Thus, (i) leads to $A \perp\!\!\!\perp B \mid D$ because $D \subset V \setminus (A \cup B)$ which gives that $\tilde{\mathcal{P}}$ fulfills the Markov properties for covariance graphs.
Suppose now that $\tilde{\mathcal{P}}$ holds the Markov properties for covariance graphs, i.e. $A \perp\!\!\!\perp B \mid D$. Using again (ii) of Definition 4.10 yields $A \perp\!\!\!\perp B$ and thus with (i) $A \perp\!\!\!\perp B \mid C$ because $C \subset V \setminus (A \cup B)$. □

Theorem 4.12 *The family $\mathcal{P}$ is resistant against Simpson's paradox.*

Proof:
The statement is a direct consequence of the equivalence of marginal and conditional independence. □

4.2 Marginalization of graphical models with KS distribution

Another property discussed for graphical models is the so–called collapsibility onto a subset $A \subset V$. Here, two aspects have to be taken into account. First the distribution family under investigation has to be closed under marginalization. The second point is that the resulting family of marginal distributions can be identified with the graphical model corresponding to the subgraph $\mathcal{G}_A$, i.e. with the set of distributions fulfilling the Markov properties of $\mathcal{G}_A$. For $A \subset V$, we write $\mathcal{KS}(\mathcal{G})_A = \{f_A(\cdot) \mid f(\cdot) \in \mathcal{KS}(\mathcal{G})\}$ for the family of marginal distributions derived from $\mathcal{KS}(\mathcal{G})$. In general this holds only under certain restrictions on the set $B = V \setminus A$. A subgraph $\mathcal{G}_A$ is constructed from the original graph $\mathcal{G}$ by deleting all vertices in $B = V \setminus A$ and all edges between vertices in $B = V \setminus A$ and vertices in A. Thus, it is possible that separating sets arise in the subgraph which did not exist in $\mathcal{G}$ and lead to other Markov properties. As mentioned above, the separating set does not affect the Markov properties for families of distributions which are resistant against Simpson's paradox.

Theorem 4.13 *Let $\mathcal{G} = (V, E)$ be a graph, A a subset of V, and $\tilde{\mathcal{P}}$ a $\mathcal{G}$ Markov family of distributions which is resistant against Simpson's paradox. Then, it holds that $\tilde{\mathcal{P}}_A$ fulfills the Markov properties of the subgraph $\mathcal{G}_A$.*

Proof:

It has to be shown that $\mathcal{M}(\mathcal{G})_A$ holds the total $\mathcal{G}_A$ Markov property of $\mathcal{G}_A$. Let A_1, A_2 be disjoint subsets of A which are separated by the set $A \setminus (A_1 \cup A_2)$. This implies that there exists no path from vertices of A_1 to vertices of A_2 consisting only of vertices of $A_1 \cup A_2$. The subgraph $\mathcal{G}_A$ results from $\mathcal{G}$ by deleting the vertices $B = V \setminus A$ and all edges within B and between vertices of B and vertices of A. Edges within A and particularly edges between vertices of A_1 and vertices of A_2 are preserved. Thus, it is not possible that there exists a path between A_1 and A_2 which consists only of vertices of $A_1 \cup A_2$ in the original graph $\mathcal{G}$. As a consequence, A_1 and A_2 are separated by the set $V \setminus (A_1 \cup A_2)$ in $\mathcal{G}$. The fact that $\tilde{\mathcal{P}}$ is $\mathcal{G}$ Markov yields

$$A_1 \perp\!\!\!\perp A_2 \mid C \quad \text{for all} \quad C \subseteq V \setminus (A_1 \cup A_2).$$

Since A is a subset of V it holds that

$$A_1 \perp\!\!\!\perp A_2 \mid D \quad \text{for all} \quad D \subseteq A \setminus (A_1 \cup A_2).$$

Thus, $\tilde{\mathcal{P}}_A$ fulfills the total $\mathcal{G}_A$ Markov property. □

For a special choice of $\tilde{\mathcal{P}}$ it has to be checked whether the distribution family is closed under marginalization. For the family of KS distributions this is always the case. Thus, the following theorem can be formulated.

Theorem 4.14 *For a subvector X_A of X_V it holds that $\mathcal{KS}(\mathcal{G})_A = \mathcal{KS}(\mathcal{G}_A)$.*

Proof:
The inclusion $\mathcal{KS}(\mathcal{G})_A \subseteq \mathcal{KS}_\tau(\mathcal{G}_A)$ is a direct consequence of Theorem 4.13 and the fact that KS distributions are closed under marginalization. Let now $P_A \in \mathcal{KS}_\tau(\mathcal{G}_A)$ with cdf $G(x_A)$. The distribution P with cdf

$$F(x) = G(x_A) \prod_{i \in V \setminus A} F(x_i)$$

then belongs to $\mathcal{KS}_\tau(\mathcal{G})_A$. Thus, $\mathcal{KS}_\tau(\mathcal{G})_A \supseteq \mathcal{KS}_\tau(\mathcal{G}_A)$. □

5 Discussion

As most important result we have shown the equivalence of the Markov properties for covariance graphs for the family of KS distributions. Thus, a graph can be used to represent the association structure of the components of a KS distributed random vector. Missing edges in the graph can be interpreted as certain independences which can be formulated in terms of Markov properties. A second remarkable outcome is that marginal independence implies conditional independence and vice versa. A more detailed discussion of the consequences as well as the background of this property will be an interesting topic of further research. Another aspect of using the graphical representation concerns the possible simplification of ML estimation. In case of the multivariate normal distribution combined with a covariance graph (Kauermann, 1996) and the CG distribution combined with a concentration graph (Frydenberg and Lauritzen, 1989) it has been shown that a decomposition of the graph can be transferred to a decomposition of the estimation problem in some situations. The highdimensional estimation problem belonging to the original model, i.e. belonging to a graph $\mathcal{G}$, can be split into smaller ones each having a reduced number of parameters to be estimated and belonging to graphical models based on subgraphs of the original graph. It should be checked whether similar results can be obtained for KS distributions.

Acknowledgment
Financial support by the SFB 386, Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, New York.

Blyth, C.R. (1972). On Simpson's paradox and the sure–thing principle. *Journal of the American Statistical Association 67*, 364–366.

Caputo, A. (1998). *Some properties of the family of Koehler and Symanowski distributions.* SFB 386 – Discussion Paper 103, University of Munich.

Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statistical Science 8*, 204–283.

Cox, D.R. and Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika 81*, 403–408.

Dawid, A.P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society B 41*, 1–31.

Dawid, A.P. (1980). Conditional independence for statistical operations. *The Annals of Statistics 8*, 598–617.

Frydenberg, M. (1990). Marginalization and collapsibility in graphical interaction models. *The Annals of Statistics 18*, 790–805.

Frydenberg, M. and Lauritzen, S.L. (1989) Decomposition of maximum likelihood in mixed graphical interaction models. *Biometrika 76*, 539-555.

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis 10*, 467–498.

Kauermann, G. (1996). On a dualization of graphical gaussian models. *Scandinavian Journal of Statistics 23*, 105–116.

Koehler, K.J. and J.T. Symanowski (1995). Constructing multivariate distributions with specific marginal distributions. *Journal of Multivariate Analysis 55*, 261–282.

Marshall, A.W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association 83*, 834–841.

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

Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics 17*, 31–57.

Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B 13*, 238–241.

Wermuth, N. and Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence, graphs and graphical chain models. *Journal of the Royal Statistical Society B 52*, 21–50.