Caputo:

Decomposition of ML Estimation in Graphical Models with Koehler Symanowski distributions

Sonderforschungsbereich 386, Paper 105 (1998)

Decomposition of ML estimation in graphical models with Koehler Symanowski distributions

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

Summary

In the framework of graphical models the graphical representation of the association structure is used in manifold respects. One is the conclusion from a decomposition of the graph to a possible decomposition of the ML estimation. Results are well–known under the assumption of the Conditional Gaussian distribution. Here, graphical models with a family of distributions are considered which is introduced by Koehler and Symanowski (1995). This approach extends the existing theory of graphical models in two respects. The family of distributions we discuss forms an alternative to the usually applied multivariate normal distribution. Furthermore, the focus lies on covariance graphs rather than on concentration graphs. For these models the decomposability of ML estimation is examined.

1 Introduction

In the literature, 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 Markovian properties for concentration graphs. The equivalence of pairwise, local, and global Markovian 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 $A \perp B \mid C$. That means, missing edges in the underlying graph can be correctly interpreted as conditional independencies. Another aspect of using the graphical representation of the association structure among the variables concerns the possible simplification of ML estimation. It has been shown that under certain conditions the decomposition of the graph results in the decomposition of the estimation problem 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. 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 independencies instead of conditional ones appropriate independence properties, also called Markovian properties, have been proposed (Cox and Wermuth, 1993, Kauermann, 1996, Caputo, 1998b). For such graphs, the 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) and the family of Koehler Symanowski (KS) distributions (Caputo, 1998b) which represents an alternative to the usually applied distribution family, namely the family of multivariate normal distributions. The family of KS distributions allows to model on the one hand complex associations between arbitrary subsets of the variable set and on the other hand pairwise independencies in the margins. These distributions can be constructed for almost any given univariate marginal distributions by adding interaction terms and can be viewed as a generalization of the generalized Burr–Pareto–logistic distributions.

The following section gives a brief introduction to the family of KS distributions and its properties. Section 3 is devoted to the basic terminology of graph theory and the theory of graphical models including Markov properties for covariance graphs. In Section 4 the idea of decomposing estimation problems in graphical models is discussed in general. Section 5 deals with the estimation problem in case of graphical models with KS distributions. The paper ends with a discussion of the results.

2 The family of KS distributions

Let $V = \{1, \ldots, p\}$ be an index set and $\mathcal{V}$ the set of all subsets of V. Let $X = X_V = (X_1, \ldots, X_p)^T$ denote a vector of random variables with marginal cumulative density functions (cdf's) $F_i(\cdot)$, $i \in V$. Then, we define the joint distribution of X similarly to Koehler and Symanowski (1995) via the cdf given by

$$F(x_1, \ldots, x_p) = \prod_{i \in V} F_i(x_i) \prod_{I \in \mathcal{I}} c_I(x)^{-\alpha_I} \tag{2.1}$$

and call it KS distribution with parameters α_I, $I \in \mathcal{V}$. For all sets $I \in \mathcal{I} = \{I \in \mathcal{V}$ 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 $\alpha_{i+} = \sum_{I \in \mathcal{V}, i \in I} \alpha_I <$

∞. The factors $c_I(x)$ in Equation (2.1) are for all $I \in \mathcal{I}$ given by

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

with $u_i(x_i) = F_i(x_i)^{\frac{1}{\alpha_{i+}}}$ $(i \in V)$. The parameters α_I with $I \in \mathcal{I}$ are called association parameters. In the following, we assume that marginal density functions $f_i(\cdot)$ exist for all $i \in V$. Then, as shown in Koehler and Symanowski (1995), the joint density function, denoted by $f(\cdot)$, exists, too. 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 the cdf.

The family of KS distributions has a lot of interesting properties which are discussed in detail in Caputo (1998a). For instance, KS distributions are closed under marginalization 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 (Capuo, 1998a). 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 are equivalent.

3 Graph theory and Markov properties

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 $i \neq j$.
In the following, the Markov properties for covariance graphs are defined. 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 $\{i\} \perp \{j\}$ holds for all $(i, j) \notin E$ with $i \neq j$,

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

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

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

Caputo (1998b) shows that the above four Markovian properties for covariance graphs are equivalent for the family of KS distributions. Thus, a family $\mathcal{P}$ of KS distributions on X_V is called $\mathcal{G}$–Markov whenever it satisfies one of the Markov properties. A graphical model can now be defined on a covariance graph $\mathcal{G} = (V, E)$ with KS distribution, denoted by $\mathcal{KS}(\mathcal{G})$, as the family of all $\mathcal{G}$–Markov KS distributions. The model $\mathcal{KS}(\mathcal{G})$ is not Markov perfect with respect to the definition given for example in Kauermann (1996) but is Markov perfect referring to the additional total Markovian property: if $A \perp B \mid C$ for each $P \in \mathcal{KS}(\mathcal{G})$ and any disjoint sets $A, B, C \subset V$ it follows that A and B are separated by the set $D = V \setminus (A \cup B)$ (Caputo, 1998b).
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 $\mathcal{G}_A$. 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 $\mathcal{KS}(\mathcal{G})$. It can be found in Caputo (1998b) that $\mathcal{KS}(\mathcal{G})_A = \mathcal{KS}(\mathcal{G}_A)$ for any $A \subset V$, i.e. graphical models with KS distribution are collapsible onto arbitrary subsets $A \subset V$.

4 ML Estimation in graphical models

As mentioned above, a graphical model has the property that independence statements can be read off the graph. Another benefit of the graphical representation of the association structure concerns the possible simplification of the ML estimation. For special cases it has been shown that an appropriate decomposition of the graph results in a decomposition of the estimation problem 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. A detailed discussion of these problems in case of CG distribution can be found in Frydenberg and Lauritzen (1989), Frydenberg (1990) and Lauritzen (1996, pp. 175).

The decomposition of the ML estimation is possible whenever the likelihood function factorizes into functions that are distinct with respect to the unknown parameter vector, say ω. Thus, the maximization reduces to the separate maximization of each factor. This procedure is justified by a general concept proposed by Barndorff–Nielsen (1978). The idea is to show that the investigated family of distributions $\mathcal{P}$ can be written as a product space $\mathcal{P}_T \times \mathcal{P}^T$ for a statistic $T = T(X)$, where $\mathcal{P}_T$ denotes the set of distributions of T and $\mathcal{P}^T$ the set of conditional distributions of X given T. This implies on the one hand that any density $p \in \mathcal{P}$ factorizes into the product

$$p(x;\omega) = p_T(T(x);\omega_1)\, p^T(x;\omega_2|T(X) = T(x)) \tag{4.2}$$

for $p_T \in \mathcal{P}_T$ and $p^T \in \mathcal{P}^T$. On the other hand, it holds that any product (4.2) for arbitrary elements $p_T \in \mathcal{P}_T$ and $p^T \in \mathcal{P}^T$ is an element of $\mathcal{P}$. This defines the statistic T as a cut in $\mathcal{P}$ (Barndorff–Nielsen, 1978, p. 50). This argument is applied to a general graphical model $\mathcal{M}(\mathcal{G})$, i.e. $\mathcal{M}(\mathcal{G})$ is the set of all distributions P of a family of distributions $\tilde{\mathcal{P}}$ fulfilling the Markov property of a given graph $\mathcal{G} = (V, E)$. In the following, we identify elements $P \in \mathcal{M}(\mathcal{G})$ with their density and their cdf, respectively, i.e., $f(\cdot) \in \mathcal{M}(\mathcal{G})$ and $F(\cdot) \in \mathcal{M}(\mathcal{G})$ refer to the same element of $\mathcal{M}(\mathcal{G})$.

For a graph $\mathcal{G} = (V, E)$ and $A \subset V$ we define the set of marginal distributions as $\mathcal{M}(\mathcal{G})_A = \{f_A(\cdot) \mid f(\cdot) \in \mathcal{M}(\mathcal{G})\}$ and the set of conditional distributions of $\mathcal{M}(\mathcal{G})$ given X_A as $\mathcal{M}(\mathcal{G})^A = \{f_{V\setminus A|A}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{M}(\mathcal{G})\}$. These notations are also used to denote sets like $(\mathcal{M}(\mathcal{G})_{B\cup C})^C = \{f_{B|C}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{M}(\mathcal{G})_{B\cup C}\}$ for disjoint subsets B, C of V. Note that this differs from $\mathcal{M}(\mathcal{G}_{B\cup C})^C = \{f_{B|C}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{M}(\mathcal{G}_{B\cup C})\}$, insofar as $\mathcal{M}(\mathcal{G}_{B\cup C})$ denotes a graphical model corresponding to the graph $\mathcal{G}_{B\cup C}$ whereas $\mathcal{M}(\mathcal{G})_{B\cup C}$ is the family of marginal distributions of a graphical model corresponding to the graph $\mathcal{G}$. This difference is essential for the following argumentation. Within the scope of graphical models, the idea of a cut is used in a slightly modified way. For a partition A, B, and C of V the statistic $T(x) = x_{A\cup C}$ is said to be a cut in $\mathcal{M}(\mathcal{G})$ if the following three conditions hold:

(i) $\mathcal{M}(\mathcal{G})_{A\cup C} = \mathcal{M}(\mathcal{G}_{A\cup C})$,

(ii) $\mathcal{M}(\mathcal{G})^{A\cup C} = \mathcal{M}(\mathcal{G}_{B\cup C})^C$,

(iii) $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{G}_{A\cup C}) \times \mathcal{M}(\mathcal{G}_{B\cup C})^C$.

This definition of a cut is more restrictive than the one of Barndorff–Nielsen described above as two closure properties are called for besides the factorization criterion which shows up again in condition (iii). These additional properties (i) and (ii) guarantee that the factors of the product space are related to graphical models corresponding to subgraphs. For instance, condition (i) equals the so–called collapsibility of graphical models onto the set $A \cup C$ (cf. Frydenberg and Lauritzen, 1989) and describes that the family of marginal distributions $\mathcal{M}(\mathcal{G})_{A\cup C}$ of a graphical model $\mathcal{M}(\mathcal{G})$ coincides with the graphical model $\mathcal{M}(\mathcal{G}_{A\cup C})$ related with the subgraph $\mathcal{G}_{A\cup C}$ of $\mathcal{G}$.

Under the assumption $A \perp B \mid C$, the inclusion $\mathcal{M}(\mathcal{G}) \subseteq \mathcal{M}(\mathcal{G}_{A\cup C}) \times \mathcal{M}(\mathcal{G}_{B\cup C})^C$ which is part of condition (iii) follows directly from the well-known factorization of the joint density combined with (i) and (ii):

$$\mathcal{M}(\mathcal{G}) \subseteq \mathcal{M}(\mathcal{G})_{A\cup C} \times \mathcal{M}(\mathcal{G})^{A\cup C} = \mathcal{M}(\mathcal{G}_{A\cup C}) \times \mathcal{M}(\mathcal{G}_{B\cup C})^C. \tag{4.3}$$

The opposite direction, i.e.

$$\mathcal{M}(\mathcal{G}_{A\cup C}) \times \mathcal{M}(\mathcal{G}_{B\cup C})^C \subseteq \mathcal{M}(\mathcal{G}). \tag{4.4}$$

is in general more difficult to prove. For this purpose, it has to be shown that for arbitrary density functions $g_{A\cup C}(\cdot) \subseteq \mathcal{M}(\mathcal{G}_{A\cup C})$ and $h_{B|C}(\cdot) \subseteq \mathcal{M}(\mathcal{G}_{B\cup C})^C$ the product $g_{A\cup C}(x_A, x_C)h_{B|C}(x_B|x_C)$ is in $\mathcal{M}(\mathcal{G})$. In contrast to the above inclusion, this crucially depends on the properties of the underlying multivariate distribution especially on its factorization properties with respect to the parameter vector. That means, it must be checked whether the density of the assumed distribution fulfills the stated Condition (4.4) or not.

For a vector $X = X_V$ of random variables let $x_V^1, \ldots, x_V^n$ be an independent sample and $\mathcal{P}$ a family of distributions P on X_V with density functions $f(\cdot)$. Then, denote $L(f) = \prod_{j=1}^n f(x_V^j)$ for all $f(\cdot) \in \mathcal{P}$ the likelihood function of $f(\cdot)$ and $\hat{f}$ the ML estimation in the set $\mathcal{P}$ based on the sample $x_V^1, \ldots, x_V^n$ if $L(\hat{f}) \geq L(f)$ for all $f(\cdot) \in \mathcal{P}$.

Let $x_V^1, \ldots, x_V^n$ with $x_V^j = (x_i^j)_{i\in V} = (x_1^j, \ldots, x_p^j)$ for $j = 1, \ldots, n$ be an independent sample of X_V. For $A \subset V$ this yields the sample $x_A^1, \ldots, x_A^n$ of X_A with $x_A^j = (x_i^j)_{i\in A}$. Then

- $\hat{f}$ denotes the ML estimation in $\mathcal{M}(\mathcal{G})$ based on the sample $x_V^1, \ldots, x_V^n$,
- $\hat{f}_A$ denotes the ML estimation in $\mathcal{M}(\mathcal{G})_A$ based on $x_A^1, \ldots, x_A^n$,
- $\hat{f}_{[A]}$ denotes the ML estimation in $\mathcal{M}(\mathcal{G}_A)$ based on $x_A^1, \ldots, x_A^n$.

In addition, let B and C be disjoint subsets of V and $x_{B\cup C}^1, \ldots, x_{B\cup C}^n$ the corresponding sample of $X_{B\cup C}$, then

- $\hat{f}_{[B|C]}$ denotes the ML estimation in $\mathcal{M}(\mathcal{G}_{B\cup C})^C$ based on the sample $x^1_{B\cup C}, \dots, x^n_{B\cup C}$,
- $\hat{f}_{B|C}$ denotes the ML estimation in $(\mathcal{M}(\mathcal{G})_{B\cup C})^C$ based on the sample $x^1_{B\cup C}, \dots, x^n_{B\cup C}$.

Using the concept of a cut we derive the following general result for ML estimation in graphical models. The idea goes back to Frydenberg and Lauritzen (1989) for the case of graphical models with CG distribution. This can, however, be extended by a slight modification to more general situations.

Let $\mathcal{G} = (V, E)$ be a graph and A, B, and C a partition of V. Assume further that $x_{A\cup C}$ is a cut in $\mathcal{M}(\mathcal{G})$ and x_C is a cut in $\mathcal{M}(\mathcal{G}_{B\cup C})$. According to Barndorff–Nielsen (1978, p. 50) the ML estimation factorizes as $\hat{f} = \hat{f}_{A\cup C}\hat{f}_{B|A\cup C}$, whenever the corresponding distribution families satisfy the relation $\mathcal{M}(\mathcal{G}) = \mathcal{M}(\mathcal{G})_{A\cup C} \times \mathcal{M}(\mathcal{G})^{A\cup C}$. This is proven by inserting conditions (i) and (ii) of the above definition of a cut into (iii). Additionally, assumption (i) yields $\hat{f}_{A\cup C} = \hat{f}_{[A\cup C]}$ and from (ii) we get $\hat{f}_{B|A\cup C} = \hat{f}_{[B|C]}$. It follows that $\hat{f} = \hat{f}_{[A\cup C]}\hat{f}_{[B|C]}$ holds which implies that $\hat{f}$ exists whenever the factors $\hat{f}_{[A\cup C]}$ and $\hat{f}_{[B|C]}$ exist. Analogously, we get $\hat{f}_{[B\cup C]} = \hat{f}_{[C]}\hat{f}_{[B|C]}$ applying the same arguments to the statistic x_C. Thus, the final result reads as

$$\hat{f} = \frac{\hat{f}_{[A\cup C]}\hat{f}_{[B\cup C]}}{\hat{f}_{[C]}} \tag{4.5}$$

and we conclude that $\hat{f}$ exists whenever $\hat{f}_{[A\cup C]}$, $\hat{f}_{[B\cup C]}$, and $\hat{f}_{[C]}$ exist.

Summarizing, the above results allow the following interpretation. The maximization of the likelihood for the entire model $\mathcal{M}(\mathcal{G})$ can be reduced to the maximization of the likelihood functions of the models $\mathcal{M}(\mathcal{G}_{B\cup C})^C$ and $\mathcal{M}(\mathcal{G}_{A\cup C})$. The latter is, as far as $\mathcal{M}(\mathcal{G})$ is collapsible onto $A \cup C$ a graphical model corresponding to a subgraph of $\mathcal{G}$.

Now it still remains to find conditions for $x_{A\cup C}$ and x_C to be cuts in $\mathcal{M}(\mathcal{G})$ and in $\mathcal{M}(\mathcal{G}_{B\cup C})$, respectively. As mentioned above, these strongly depend on the type of the underlying distribution family. The solution to this problem can be found into Frydenberg and Lauritzen (1989) for the case of CG distribution. In the next section, this problem is discussed for the family of KS distributions.

The investigation of properties of graphical models with CG or multivariate normal distributions is often driven by exploiting the properties of the exponential family. The concept of a cut can be applied in this framework in a natural way because of the analytic structure of the density of these distributions. The family of KS distributions does not belong to the exponential family and is characterized by a multiplicative and clear structure of the cdf but not of the density. From this, two conclusions have to be drawn: A general solution to the above problem is not available because of the connection of the association parameters involved, i.e. a solution can only be found in special cases. The second point is that the proof of a possible decomposition will need other arguments than that which are applied in Frydenberg and Lauritzen (1989). It will be shown that in analogy to the well–known factorization of the joint density into a marginal and a conditional density which is used in the proof mentioned above a factorization of the cdf holds for KS distributions.

5 ML estimation in graphical models with KS distribution

In the original definition of the family of KS distributions the cdf is given by the parameters α_I $(I \in \mathcal{I})$ and α_i $(i \in V)$ with $\alpha_{i+} = \sum_{I \in \mathcal{V}, i \in I} \alpha_I < \infty$. Therefore, it is also possible to define the distribution via the parameters α_I and α_{i+}. We now restrict ourselves to a subfamily of the family of KS distributions ($\mathcal{KS}$), denoted by KS$^+$ distributions, given by the restriction that the values α_{i+} are fixed for all $i \in V$: $\mathcal{KS} \supset \mathcal{KS}^+ = \{P \in \mathcal{KS} \mid \alpha_{i+} = k_i \text{ for all } i \in V\}$. Correspondingly, we write $\mathcal{KS}^+(\mathcal{G})$ for the graphical model with KS$^+$ distribution. It is easily seen that the above properties, especially the collapsibility onto a subset $A \subset V$, hold for these submodels. Again $\mathcal{KS}^+(\mathcal{G})_A = \{f_A(\cdot) \mid f(\cdot) \in \mathcal{KS}^+(\mathcal{G})\}$ denotes the family of marginal distributions of a graphical model $\mathcal{KS}^+(\mathcal{G})$ and $\mathcal{KS}^+(\mathcal{G})^A = \{f_{V \setminus A|A}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{KS}^+(\mathcal{G})\}$ the family of conditional distributions of $\mathcal{KS}^+(\mathcal{G})$ given X_A.
As mentioned above, the density of a KS distribution is not easy to handle and therefore, the analysis of factorization properties fails in the simplest cases using only density functions. Here, a factorization formula for the cdf is shown which can be used to derive the conditions for a decomposition of the ML estimatiom indirectly via cdf's instead of densities.
In the following, some properties of the cdf of a KS distribution including the factorization property are introduced. Then, the conditions for the decomposition of the ML estimation is discussed for the special case of KS(2)$^+$ distributions. These distributions are given by the restriction that all parameters α_I equal zero for $|I| > 2$ (Caputo, 1998a). Subsequently, the results are generalized to the case of KS$^+$ distributions.

5.1 Factorization properties

For a partition A, B, and C of V with $A \perp B \mid C$ the general factorization of the density

$$f(x) = f_{A \cup B \cup C}(x_A, x_B, x_C) = f_{A \cup C}(x_A, x_C) f_{B|A \cup C}(x_B|x_A, x_C) \tag{5.6}$$

may be written as

$$f(x) = f_{A \cup C}(x_A, x_C) f_{B|C}(x_B|x_C) = \frac{f_{A \cup C}(x_A, x_C) f_{B \cup C}(x_B, x_C)}{f_C(x_C)} \tag{5.7}$$

for all $x \in I\!R^p$ with $f_C(x_C) \neq 0$. In the theory of graphical models, this formula is basically needed to show a possible decomposition of the ML estimation provided the assumption $A \perp B \mid C$ is fulfilled. From Theorem 2.16 (Caputo, 1998a) it can be directly concluded that

$$F(x) = F_{A \cup B \cup C}(x_A, x_B, x_C) = \frac{F_{A\cup C}(x_A, x_C) F_{B \cup C}(x_B, x_C)}{F_C(x_C)} \tag{5.8}$$

holds for the cdf of a KS distribution in the special case of conditional independence. In addition to the factorization of the joint cdf into marginal cdf's, it is possible to show that the joint cdf decomposes into a product of marginal and conditional cdf's analogously to the well–known result for density functions. To derive this result, the following proposition is needed.

Proposition 5.1 *Let $X = X_V$ be a vector of random variables and assume that $A \perp B \mid C$ holds for a partition A, B, and C of V. Then we have for all $x \in I\!R^p$ with $F_C(x_C) \neq 0$:*

$$\frac{\left\{\frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C)\right\} f_C(x_C)}{f_{A\cup C}(x_A, x_C) F_C(x_C)} = \frac{\left\{\frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C)\right\} f_C(x_C)}{f_{B\cup C}(x_B, x_C) F_C(x_C)} = 1. \tag{5.9}$$

Proof:
From $A \perp B \mid C$ it follows that (5.7) holds and with Equation (5.8) the joint density $f(x)$ can be written as

$$\begin{aligned}
f(x) &= f_{A\cup B\cup C}(x_A, x_B, x_C) = \frac{\partial^3}{\partial x_A \partial x_B \partial x_C} F_{A\cup B\cup C}(x_A, x_B, x_C) \\
&= \frac{\partial^3}{\partial x_A \partial x_B \partial x_C} \frac{F_{A\cup C}(x_A, x_C) F_{B\cup C}(x_B, x_C)}{F_C(x_C)} \\
&= \frac{1}{F_C^2(x_C)} \frac{\partial^2}{\partial x_A \partial x_B} \Bigg(\left\{ \frac{\partial}{\partial x_C} F_{A\cup C}(x_A, x_C) F_{B\cup C}(x_B, x_C) \right\} F_C(x_C) \\
&\qquad - F_{A\cup C}(x_A, x_C) F_{B\cup C}(x_B, x_C) f_C(x_C) \Bigg) \\
&= \frac{1}{F_C^2(x_C)} \frac{\partial^2}{\partial x_A \partial x_B} \Bigg(\left\{ \frac{\partial}{\partial x_C} F_{A\cup C}(x_A, x_C) \right\} F_{B\cup C}(x_B, x_C) F_C(x_C) \\
&\qquad + F_{A\cup C}(x_A, x_C) \left\{ \frac{\partial}{\partial x_C} F_{B\cup C}(x_B, x_C) \right\} F_C(x_C) \\
&\qquad - F_{A\cup C}(x_A, x_C) F_{B\cup C}(x_B, x_C) f_C(x_C) \Bigg) \\
&= \frac{1}{F_C^2(x_C)} \frac{\partial}{\partial x_A} \Bigg(\left\{ \frac{\partial}{\partial x_C} F_{A\cup C}(x_A, x_C) \right\} \left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} F_C(x_C) \\
&\qquad + F_{A\cup C}(x_A, x_C) f_{B\cup C}(x_B, x_C) F_C(x_C) \\
&\qquad - F_{A\cup C}(x_A, x_C) \left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} f_C(x_C) \Bigg) \\
&= \frac{1}{F_C^2(x_C)} \Bigg(f_{A\cup C}(x_A, x_C) \left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} F_C(x_C)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \left\{ \frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C) \right\} f_{B\cup C}(x_B, x_C) F_C(x_C) \\
&\qquad - \left\{ \frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C) \right\} \left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} f_C(x_C) \Big) \\
= \; & \frac{f_{A\cup C}(x_A, x_C) f_{B\cup C}(x_B, x_C)}{f_C(x_C)} \\
&\cdot \Big(\frac{\left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} f_C(x_C)}{f_{B\cup C}(x_B, x_C) F_C(x_C)} + \frac{\left\{ \frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C) \right\} f_C(x_C)}{f_{A\cup C}(x_A, x_C) F_C(x_C)} \\
&\quad - \frac{\left\{ \frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C) \right\} \left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} f_C^2(x_C)}{f_{A\cup C}(x_A, x_C) f_{B\cup C}(x_B, x_C) F_C^2(x_C)} \Big).
\end{aligned}
$$

Using the abbreviations

$$
a = \frac{\left\{ \frac{\partial}{\partial x_A} F_{A\cup C}(x_A, x_C) \right\} f_C(x_C)}{f_{A\cup C}(x_A, x_C) F_C(x_C)} \quad \text{and} \quad b = \frac{\left\{ \frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C) \right\} f_C(x_C)}{f_{B\cup C}(x_B, x_C) F_C(x_C)}
$$

it follows that $a + b - ab = 1$ or that $(1-a)(1-b) = 0$ which is fulfilled for $a = 1$ or $b = 1$. Since the two factors a and b have an identical structure it can be concluded that $a = b = 1$. □

From Equation (5.9) it is easily seen that

$$
f_{B|C}(x_B|x_C) = \frac{f_{B\cup C}(x_B, x_C)}{f_C(x_C)} = \frac{\frac{\partial}{\partial x_B} F_{B\cup C}(x_B, x_C)}{F_C(x_C)}.
$$

Therefore, we have

$$
F_{B|C}(x_B|x_C) = \int_{-\infty}^{x_B} f_{B|C}(u|x_C) du = \int_{-\infty}^{x_B} \frac{\frac{\partial}{\partial u} F_{B\cup C}(u, x_C)}{F_C(x_C)} du = \frac{F_{B\cup C}(x_B, x_C)}{F_C(x_C)}.
$$

Together with Equation (5.8), this result leads immediatly to a factorization formula which is dual to Equation (5.6) for the case that $A \perp B \mid C$ holds:

$$
F(x) = F_{A\cup C}(x_A, x_C) F_{B|C}(x_B|x_C). \tag{5.10}
$$

5.2 The case of KS(2) distributions

The following lemma formulates a proposition similar to Equation 5.8. But here, the cdf's involved need not fit together in the sense that they need not stem from the same joint cdf.

Lemma 5.2 *Let $\mathcal{G} = (V, E)$ be a graph and $A, B \subset V$ such that $C = V \setminus (A \cup B)$ separates A and B. Let further be $G_{A\cup C}(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G}_{A\cup C})$ with parameters α_{ij} for $i, j \in A \cup C$*

and $H_{B\cup C}(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})$ *with parameters* β_{kl} *for* $k,l \in B\cup C$ *arbitrarily chosen. Denote with* $H_C(\cdot)$ *the marginal cdf of the distribution of* X_C *derived from the joint cdf* $H_{B\cup C}(\cdot)$*. Then, it holds that*

$$\frac{G_{A\cup C}(\cdot)H_{B\cup C}(\cdot)}{H_C(\cdot)} = F(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G}),$$

i.e. $F(\cdot)$ *is the cdf of a KS(2)*$^+$ *distribution which fulfills the Markov properties of* $\mathcal{G}$*.*

<u>Proof:</u>
Let $V = \{1,\dots,p\}$ such that $A = \{1,\dots,m\}$, $C = \{m+1,\dots,n\}$, and $B = \{n+1,\dots,p\}$. From the definition of the KS distribution and the structure of the marginal cdf (Caputo, 1998a) it follows that

$$\begin{aligned}
F(x) &= \frac{\prod_{i=1}^n F_i(x_i)\prod_{i=1}^n\prod_{j=i+1}^n c_{ij}(x)^{-\alpha_{ij}}\prod_{k=m+1}^p F_k(x_k)\prod_{k=m+1}^p\prod_{l=k+1}^p c_{kl}(x)^{-\beta_{kl}}}{\prod_{k=m+1}^n F_k(x_k)\prod_{k=m+1}^n\prod_{l=k+1}^n c_{kl}(x)^{-\beta_{kl}}} \\
&= \prod_{i=1}^p F_i(x_i)\prod_{i=1}^n\prod_{j=i+1}^n c_{ij}(x)^{-\alpha_{ij}}\prod_{k=m+1}^n\prod_{l=n+1}^p c_{kl}(x)^{-\beta_{kl}}\prod_{r=n+1}^p\prod_{s=r+1}^p c_{rs}(x)^{-\beta_{rs}} \\
&= \prod_{i=1}^p F_i(x_i)\prod_{i=1}^p\prod_{j=i+1}^p c_{ij}(x)^{-\gamma_{ij}}
\end{aligned}$$

with $\gamma_{ij} = \alpha_{ij}$ $(i,j \in A\cup C, i<j)$, $\gamma_{ij} = 0$ $(i \in A, j \in B)$, $\gamma_{ij} = \beta_{ij}$ $(i \in C, j \in B)$, and $\gamma_{ij} = \beta_{ij}$ $(i,j \in B, i<j)$. Therefore, $F(\cdot)$ is the cdf of a KS(2)$^+$ distribution with $A \perp B \mid C$. □

With this lemma one of the main propositions of this paper can be proved which is formulated in the following theorem.

Theorem 5.3 *For a graph* $\mathcal{G} = (V,E)$ *and disjoint subsets* A *and* B *of* V *which are separated by the set* $C = V\setminus(A\cup B)$ *it holds that*

(i) $\mathcal{KS}(2)^+(\mathcal{G})_{A\cup C} = \mathcal{KS}(2)^+(\mathcal{G}_{A\cup C})$,

(ii) $\mathcal{KS}(2)^+(\mathcal{G})^{A\cup C} = \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C$,

(iii) $\mathcal{KS}(2)^+(\mathcal{G}) = \mathcal{KS}(2)^+(\mathcal{G}_{A\cup C}) \times \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C$,

i.e. $x_{A\cup C}$ *is a cut in* $\mathcal{KS}(2)^+(\mathcal{G})$*.*

<u>Proof:</u>
Statement (i) is a direct consequence from the collapsibility of graphical models with KS

distribution (Caputo, 1998b). Since the property (i) is also valid if A is replaced by B it holds that

$$\begin{aligned}\mathcal{KS}(2)^+(\mathcal{G})^{A\cup C} &= \{f_{B|A\cup C}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G})\} \\ &= \{f_{B|C}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G})_{B\cup C}\} \\ &= \{f_{B|C}(\cdot|\cdot) \mid f(\cdot) \in \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})\} = \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C\end{aligned}$$

which yields (ii).

As described in Section 4 the assumption $A \perp B \mid C$ leads directly to $\mathcal{KS}(2)^+(\mathcal{G}) \subseteq \mathcal{KS}(2)^+(\mathcal{G}_{A\cup C}) \times \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C$ whereas the opposite direction, i.e. $\mathcal{KS}(2)^+(\mathcal{G}_{A\cup C}) \times \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C \subseteq \mathcal{KS}(2)^+(\mathcal{G})$ is the more difficult part. It has to be shown that for arbitrary density functions $g_{A\cup C}(\cdot) \subseteq \mathcal{KS}(2)^+(\mathcal{G}_{A\cup C})$ and $h_{B|C}(\cdot) \subseteq \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C$ the product $g_{A\cup C}(x_A, x_C)h_{B|C}(x_B|x_C)$ is an element of $\mathcal{KS}(2)^+(\mathcal{G})$ which depends on the factorization properties of the underlying multivariate distribution. Let $G_{A\cup C}(\cdot)$ denote the cdf of $g_{A\cup C}(\cdot)$ and for

$$h_{B|C}(x_B|x_C) = \frac{h_{B\cup C}(x_B, x_C)}{h_C(x_C)}$$

with $h_{B\cup C}(\cdot) \in \mathcal{M}(\mathcal{G}_{B\cup C})$ and marginal density $h_C(\cdot) \in \mathcal{M}(\mathcal{G}_{B\cup C})_C$, let $H_{B\cup C}(\cdot)$ and $H_C(\cdot)$ denote the corresponding cdf's. The general idea is to show that

$$\frac{G_{A\cup C}(x_A, x_C)H_{B\cup C}(x_B, x_C)}{H_C(x_C)} = G_{A\cup C}(x_A, x_C)H_{B|C}(x_B|x_C) = F(x)$$

is the cdf of $g_{A\cup C}(x_A, x_C)h_{B|C}(x_B|x_C)$, i.e. that $f(\cdot) \in \mathcal{M}(\mathcal{G})$ is equivalent to $F(\cdot) \in \mathcal{M}(\mathcal{G})$. Therefore, define $H^*(x)$ as

$$H^*(x) = H_{B\cup C}(x_{B\cup C}) \prod_{i\in A} F_i(x_i)$$

with corresponding density function $h^*(\cdot)$. Then, it trivially holds that $H^*(\cdot) \in \mathcal{M}(\mathcal{G})$. Since $A \perp B \mid C$ is assumed to hold for all elements of $\mathcal{M}(\mathcal{G})$, Lemma 5.1 yields

$$1 = \frac{\left\{\frac{\partial}{\partial x_B} H^*_{B\cup C}(x_B, x_C)\right\} h^*_C(x_C)}{h^*_{B\cup C}(x_B, x_C)H^*_C(x_C)} = \frac{\left\{\frac{\partial}{\partial x_B} H_{B\cup C}(x_B, x_C)\right\} h_C(x_C)}{h_{B\cup C}(x_B, x_C)H_C(x_C)}.$$

In analogy to the proof of Lemma 5.1, the density $f(\cdot)$ of $F(\cdot)$ can be derived as

$$\begin{aligned}\frac{\partial}{\partial x}F(x) &= \frac{\partial^3}{\partial x_A \partial x_B \partial x_C} \frac{G_{A\cup C}(x_A, x_C)H_{B\cup C}(x_B, x_C)}{H_C(x_C)} \\ &= \frac{g_{A\cup C}(x_A, x_C)h_{B\cup C}(x_B, x_C)}{h_C(x_C)} \\ &\quad \cdot\Bigg(\frac{\left\{\frac{\partial}{\partial x_B} H_{B\cup C}(x_B, x_C)\right\} h_C(x_C)}{h_{B\cup C}(x_B, x_C)H_C(x_C)} + \frac{\left\{\frac{\partial}{\partial x_A} G_{A\cup C}(x_A, x_C)\right\} h_C(x_C)}{g_{A\cup C}(x_A, x_C)H_C(x_C)}\end{aligned}$$

$$- \frac{\left\{\frac{\partial}{\partial x_B} H_{B\cup C}(x_B, x_C)\right\} h_C(x_C)}{h_{B\cup C}(x_B, x_C) H_C(x_C)} \cdot \frac{\left\{\frac{\partial}{\partial x_A} G_{A\cup C}(x_A, x_C)\right\} h_C(x_C)}{g_{A\cup C}(x_A, x_C) H_C(x_C)} \Big)$$

$$= \frac{g_{A\cup C}(x_A, x_C) h_{B\cup C}(x_B, x_C)}{h_C(x_C)} = g_{A\cup C}(x_A, x_C) h_{B|C}(x_B|x_C) = f(x).$$

This implies that the condition $f(x) = g_{A\cup C}(\cdot)h_{B|C}(\cdot|\cdot) \in \mathcal{M}(\mathcal{G})$ is equivalent to $F(\cdot) = G_{A\cup C}(\cdot)H_{B|C}(\cdot|\cdot) \in \mathcal{M}(\mathcal{G})$ and it follows that

$$\mathcal{KS}(2)^+(\mathcal{G}_{A\cup C}) \times \mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})^C \subseteq \mathcal{KS}(2)^+(\mathcal{G}). \qquad \square$$

As described in Section 4 the decomposition of the ML estimation is carried out in two steps. This is done applying Theorem 5.3 twice.

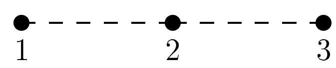
Theorem 5.4 *Let $\mathcal{G} = (V, E)$ be a graph and A, B subsets of V which are separated by the set $C = V \setminus (A \cup B)$. Then, it holds that x_C is a cut in $\mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})$.*

Proof:
In the subgraph $\mathcal{G}_{B\cup C}$ the set C separates $\emptyset$ and B. Thus, Theorem 5.3 can be applied and it follows that x_C is a cut in $\mathcal{KS}(2)^+(\mathcal{G}_{B\cup C})$. $\square$

Next, the above results are illustrated by a simple example before discussing the case of general KS^+ distributions.

5.3 A simple example

Consider a graphical model with KS(2) distribution given by the graph $\mathcal{G} = (V, E)$ with $V = \{1, 2, 3\}$ and $E = \{(1,2), (2,1), (2,3), (3,2)\}$:



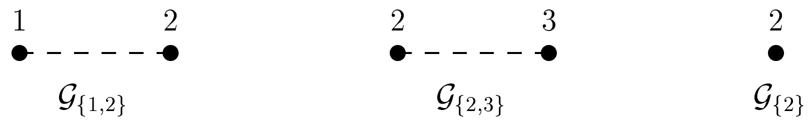
Here, the sets of vertices $A = \{1\}$ and $B = \{3\}$ are separated by $C = \{2\}$. Thus, for the corresponding vector $X = (X_1, X_2, X_3)$ of random variables with marginal cdf's $F_1(\cdot)$, $F_2(\cdot)$, and $F_3(\cdot)$ it holds that $X_1 \perp X_3$ and $X_1 \perp X_3 \mid X_2$, i.e. X_1 and X_3 are marginally independent and conditionally independent given X_2. The KS distribution of X is characterized by the fact that $\alpha_{13} = 0$. The graphical model $\mathcal{KS}(2)(\mathcal{G})$ is then given by the family of KS distributions with the five parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_{12}$, and α_{23} – or using the alternative parameterization by $\alpha_{1+}, \alpha_{2+}, \alpha_{3+}, \alpha_{12}$, and α_{23}.

For fixed values α_{1+}, α_{2+}, and α_{3+} we get the graphical model $\mathcal{KS}(2)^+(\mathcal{G})$ with parameters α_{12} and α_{23}. From Theorem 5.3 it can be concluded that $x_{A\cup C} = (x_1, x_2)^T$ is a cut in this model which yields that

$$\hat{f} = \frac{\hat{f}_{[A\cup C]} \cdot \hat{f}_{[B\cup C]}}{\hat{f}_{[C]}} = \frac{\hat{f}_{[\{1,2\}]} \cdot \hat{f}_{[\{2,3\}]}}{\hat{f}_{[\{2\}]}}. \tag{5.11}$$

The graphical models $\mathcal{KS}(2)^+(\mathcal{G}_{\{1,2\}})$ and $\mathcal{KS}(2)^+(\mathcal{G}_{\{2,3\}})$ are families of KS$(2)^+$ distributions with one parameter α_{12} and α_{23}, respectively. The family $\mathcal{KS}(2)^+(\mathcal{G}_{\{2\}})$ consists of one single element: the marginal distribution $F_2(\cdot)$ of X_2. That means that the estimation of the parameters α_{12} and α_{23} in the original model $\mathcal{KS}(2)^+(\mathcal{G})$ reduces to the estimation of α_{12} in the model $\mathcal{KS}(2)^+(\mathcal{G}_{\{1,2\}})$ and α_{23} in $\mathcal{KS}(2)^+(\mathcal{G}_{\{2,3\}})$.

The corresponding graphs of these models are given by the following subgraphs of $\mathcal{G}$:



5.4 Generalization to KS distributions

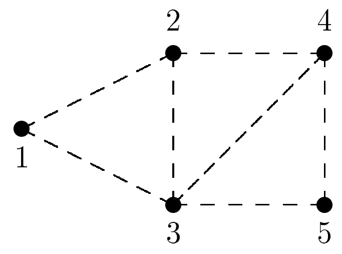
The inspection of the procedure which solves the decomposition problem for the case of KS$(2)^+$ distributions shows that only at one point the restriction to KS$(2)^+$ distributions is required: when Lemma 5.2 is called for. Thus, the search for the conditions for the decomposition of the ML estimation for the more general case of KS$^+$ distributions should focus on a proposition generalizing this lemma. In the proof it is taken advantage of the fact that all factors which appear in $H_C(\cdot)$ also appear in $H_{B\cup C}(\cdot)$ and thus, can be cancelled down. This property follows from the special structure of marginal KS distributions (Caputo, 1998a). For general KS$^+$ distributions factors can occur in the marginal cdf which have not been there in the joint cdf. This phenomenon strongly depends on the association structure of the components of the vector $X_{B\cup C}$, i.e. on the edges within vertices in $B\cup C$. In the following it will be discussed for which constellations of edges in $\mathcal{G}_{B\cup C}$ the spurious factors do not occur.

Definition 5.5 *Consider a graph $\mathcal{G} = (V, E)$ and disjoint subsets B, C of V. The set*

$$\mathcal{J}(B, C) = \{J \subseteq (bd(B) \cap C) \cup (bd(C) \cap B) \mid J \cap B \neq \emptyset, J \cap C \neq \emptyset\}$$

is called border of B and C.

Example 5.6 *For the graph $\mathcal{G} = (V, E)$:*



consider the disjoint sets $A = \{1\}$, $B = \{4, 5\}$, and $C = \{2, 3\}$. Thus, $bd(B) = \{2, 3\}$ and $bd(C) = \{1, 4, 5\}$. The border of B and C is then given as

$$\begin{aligned}
\mathcal{J}(B,C) &= \{J \subseteq (\{2,3\} \cap \{2,3\}) \cup (\{1,4,5\} \cap \{4,5\}) \mid J \cap \{4,5\} \neq \emptyset \wedge J \cap \{2,3\} \neq \emptyset\} \\
&= \{J \subseteq (\{2,3,4,5\} \mid J \cap \{4,5\} \neq \emptyset \wedge J \cap \{2,3\} \neq \emptyset\} \\
&= \{\{2,4\},\{2,5\},\{3,4\},\{3,5\},\{2,3,4\},\{2,3,5\},\{2,4,5\},\{3,4,5\},\{2,3,4,5\}\}.
\end{aligned}$$

By means of that term the decomposition of a graph can be defined.

Definition 5.7 *Let $\mathcal{G} = (V, E)$ be a graph and A, B, and C a partition of V. Whenever*

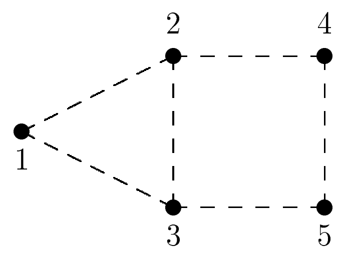
(i) C separates A and B and

(ii) all sets $J \in \mathcal{J}(B, C)$ with $|J \cap C| > 1$ and $|J| > 2$ are not complete,

the sets A, B, and C are called KS^+ decomposition of $\mathcal{G}$.

Note, that the sets A and B play a different role in the above definition. In the following, an example and a counter–example is given.

Example 5.8 *Recall Example 5.6. The sets A, B, and C do not build a KS^+ decomposition of $\mathcal{G}$. The set $\{2, 3, 4\} \in \mathcal{J}(B, C)$ is complete and it holds that $|\{2, 3, 4\} \cap C| = |\{2, 3\}| = 2 > 1$ and $|\{2, 3, 4\}| = 3 > 2$.*
Consider in contrast the graph



Here, $A = \{1\}$, $B = \{4, 5\}$, and $C = \{2, 3\}$ are a KS^+ decomposition of V.

Now, we are almost ready to formulate the following lemma in analogy to Lemma 5.2 and the final theorem which is proven by replacing Lemma 5.2 by Lemma 5.9 in the proofs of Theorem 5.3 and Theorem 5.4. But first, we have to introduce some more notations. For disjoint proper subsets A, B, and C of V let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ denote the corresponding powersets. We define $\mathcal{I}_{\mathcal{A}} = \mathcal{I} = \{I \in \mathcal{V} \text{ with } |I| \geq 2\}$ and the sets $\mathcal{I}_{\mathcal{B}}$, and $\mathcal{I}_{\mathcal{C}}$ analogously. In addition, let $\mathcal{AB} = \{I \subset (A \cup B) | I \cap A \neq \emptyset \text{ and } I \cap B \neq \emptyset\}$ and define in the same way $\mathcal{AC}$, $\mathcal{BC}$ and $\mathcal{ABC}$. To each of these sets $\mathcal{I}_{\mathcal{A}}, \ldots, \mathcal{I}_{\mathcal{ABC}}$ and $\mathcal{I}_{\mathcal{AB}_{AB}}, \ldots, \mathcal{I}_{\mathcal{ABC}_{\ldots}}$ are built as above.

Lemma 5.9 *Let $\mathcal{G} = (V, E)$ be a graph and $A, B,$ and C a KS$^+$ decomposition of V. Let further be*

$$\begin{array}{lllllll} & G_{A\cup C}(\cdot) & \in & \mathcal{KS}^+(\mathcal{G}_{A\cup C}) & \textit{with parameters} & \alpha_I & \textit{for} & I \in \mathcal{I}_{\mathcal{A}} \cup \mathcal{I}_{\mathcal{C}} \cup \mathcal{I}_{\mathcal{AC}} \\ \textit{and} & H_{B\cup C}(\cdot) & \in & \mathcal{KS}^+(\mathcal{G}_{B\cup C}) & \textit{with parameters} & \beta_J & \textit{for} & J \in \mathcal{I}_{\mathcal{B}} \cup \mathcal{I}_{\mathcal{C}} \cup \mathcal{I}_{\mathcal{BC}}. \end{array}$$

Denote $H_C(\cdot)$ the marginal cdf derived from $H_{B\cup C}(\cdot)$. Then, it holds that

$$\frac{G_{A\cup C}(\cdot) H_{B\cup C}(\cdot)}{H_C(\cdot)} = F(\cdot) \in \mathcal{KS}^+(\mathcal{G}).$$

Proof:
$A, B,$ and C are a KS$^+$ decomposition. Thus, $\beta_J = 0$ for all $J \in \mathcal{I}_{\mathcal{BC}}$ with $|J| > 2$. From Theorem 2.2 in Caputo (1998a) it can be concluded that

$$H_C(x_C) = \prod_{j \in C} F_i(x_i) \prod_{J \in \mathcal{I}_{\mathcal{C}}} c_J(x)^{-\beta_J}$$

which yields that $F(\cdot)$ is the cdf of a $\mathcal{KS}^+$ distribution with parameters $\gamma_I = \alpha_I$ for $I \in \mathcal{I}_{\mathcal{A}} \cup \mathcal{I}_{\mathcal{C}} \cup \mathcal{I}_{\mathcal{AC}}$, $\gamma_I = \beta_I$ for $I \in \mathcal{I}_{\mathcal{B}} \cup \mathcal{I}_{\mathcal{BC}}$, and $\gamma_I = 0$ for $I \in \mathcal{I}_{\mathcal{AB}} \cup \mathcal{I}_{\mathcal{ABC}}$. The condition $A \perp B \mid C$ is obviously fulfilled which leads to $F(\cdot) \in \mathcal{KS}^+(\mathcal{G})$. □

Theorem 5.10 *Let now $\mathcal{G} = (V, E)$ be a graph and A, B, C a KS$^+$ decomposition of $\mathcal{G}$. Then, $x_{A\cup C}$ is a cut in $\mathcal{KS}^+(\mathcal{G})$ and x_C is a cut in $\mathcal{KS}^+(\mathcal{G}_{B\cup C})$.*

5.5 Discussion

The investigations have shown that the connection between graph theory and the family of KS distributions can be useful in different ways. As discussed in Caputo (1998b) the introduction of graphical models with KS distribution allow a visualization and interpretation of the dependence structure of the variables involved and therefore, a better understanding of the association structure. In addition, the interplay can be exploited

to conclude from a separation in the graph to a possible simplification of the estimation problem which playes an important role whenever a larger number of variables is considered and a complex association structure between them is supposed.
One remaining question is how restrictive the assumptions are which have to be made to obtain the decomposition. The restriction to KS^+ distributions is of course heavy. On the other hand, the family of KS distributions is huge and perhaps a subfamily such as the KS^+ distributions may be sufficient for some practical problems.
Another aspect is the possible use of the decomposition within the scope of a two–stage estimation algorithm where in one step the parameters α_{i+} are plugged in as fixed values. This idea was implemented and tested with simulated data for some simple cases which showed that the decomposition reduces the computing time remarkably.
By comparison with the results derived for the case of CG distributions (Frydenberg and Lauritzen, 1989) it has to be stated that although the outcome for CG distributions appear to be more elegant and useful the restrictions are also rather restrictive. In conclusion it has been shown that the main properties known for the multivariate normal case or for the CG distributions similarly hold for graphical models with KS distribution. Thus, further research on this subject seems sensible and rewarding.

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

References

Barndorff–Nielsen, O.E. (1978). *Information and exponential families in statistical theory.* John Wiley and Sons, New York.

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

Caputo, A. (1998b). *Graphical models with Koehler and Symanowski distributions.* Discussion Paper 104, University of Munich.

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

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.

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.

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.

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.