Caputo:

Some properties of the family of Koehler Symanowski distributions

Sonderforschungsbereich 386, Paper 103 (1998)

Some properties of the family of Koehler Symanowski distributions

By Angelika Caputo

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

Summary

In this paper, a class of multivariate distributions introduced by Koehler and Symanowski (1995) is discussed with regard to whether it can be reasonably applied in the framework of graphical modeling. Therefore, the focus lies on properties like marginal and conditional independence, marginalization and the flexibility as far as the modeling of a dependence structure is concerned.

1 Introduction and Notations

Koehler and Symanowski (1995) introduce a class of multivariate distribution families which can be constructed for almost any given univariate marginal distributions and can be viewed as a generalization of the generalized Burr–Pareto–logistic distributions (Johnson, 1987; Cook and Johnson, 1981, 1986). The distribution can be defined in two different ways. One possibility is the definition via the cumulative distribution function (cdf) by adding interaction terms to the independence case, i.e. to the product of the marginal cdf's. Furthermore, it can be derived similarly to the generalized Burr–Pareto–logistic distributions by transforming independent exponential and gamma distributed random variables. The transformation rule can directly be translated into an algorithm to carry out simulation studies.

In contrast to the generalized Burr–Pareto–logistic distributions and a lot of other multivariate distributions like the different multivariate exponential distributions described in Johnson and Kotz (1972, pp. 268) the class of distributions discussed here allows to model on the one hand complex associations between arbitrary subsets of the variable set and on the other hand pairwise independences in the margins. This property is a minimum requirement in the framework of graphical models (Lauritzen and Wermuth, 1989, Wermuth and Lauritzen, 1990, Lauritzen, 1996) where the focus lies on conditional independencies between pairs or even subsets of variables. These independences are represented by missing edges in a graph which consists of vertices depicting the variables

and edges depicting associations between them. An association has to be understood as the absence of a conditional independence (Dawid, 1979, 1980). So–called Markov properties describe the conditions which have to be fulfilled by the distribution model to connect the graph with these independence statements (Frydenberg, 1990, Lauritzen, 1996). Distributions which are able to capture these demands are the multinomial, multivariate normal, and the conditional gaussian distribution (c.f. Lauritzen and Wermuth, 1989) which contains the former as special cases. In situations where these distributions are not adequate, for example when investigating multivariate survival times, alternative multivariate distributions are called for. Therefore, in this paper the class of distributions introduced by Koehler and Symanowski is inspected in the face of the mentioned requirements. Here, the importance is attached to basic properties like marginal and conditional independence and marginalization rather than the Markov properties which can be found in Caputo (1998).
The paper is organized as follows. The remaining part of this Section is dedicated to the definition of the class of KS distributions and a subclass which essentially serves for illustrating porposes. In Section 2, the form of marginal distributions and some special conditional probabilities are derived before the conditions for marginal and conditional independence are discussed. In the following section, the results of a simulation study are presented which considers the interpretation of the parameters of KS distributions. The paper ends with a brief discussion of the derived findings and an outlook to further research questions.

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

$$F(x_1, \ldots, x_p) = \prod_{i \in V} F_i(x_i) \prod_{I \in \mathcal{I}} c_I(x)^{-\alpha_I}. \tag{1.1}$$

For all sets $I \in \mathcal{I} = \{I \in \mathcal{V} \text{ with } |I| \geq 2\}$ let $I\!\!R \ni \alpha_I \geq 0$ and for all $i \in V$ let $I\!\!R \ni \alpha_i > 0$ 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 Equation (1.1) 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 $i \in V$. Here, the structure of the cdf is fairly easy. It factorizes into the product of the marginal cdf's and a product of association terms. If it is assumed 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 due to the derivation. The explicit formula for a subclass can be found in Koehler and Symanowski (1995).
In the following a special case of KS distributions is considered 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. A lot of properties which are valid in this special case denoted as KS(2) distribution as well as in general can be illustrated in a more transparent way.

Remark 1.1 Let $X = X_V = (X_1, \ldots, X_p)^T$ be a vector of random variables with marginal cdf's $F_i(\cdot)$ and for each set $I \in \mathcal{I}_2 = \{I \in \mathcal{V} \text{ with } |I| = 2\}$ let $\alpha_I \geq 0$ and for all $i \in V$ $\alpha_i > 0$ with

$$\alpha_{i+} = \alpha_i + \sum_{\substack{I \in \mathcal{I}_2 \\ i \in I}} \alpha_I = \alpha_i + \sum_{\substack{j \in V \\ j \neq i}} \alpha_{ij} < \infty.$$

The joint cdf given by

$$F(x_1, \ldots, x_p) = \prod_{i \in V} F_i(x_i) \prod_{I \in \mathcal{I}_2} c_I(x)^{-\alpha_I} = \prod_{i \in V} F_i(x_i) \prod_{\substack{j \in V \\ i < j}} c_{ij}(x)^{-\alpha_{ij}} \tag{1.2}$$

defines the KS(2) distribution. Note that the index in α_{ij} is to be interpreted as a set, i.e. as $\{i, j\}$. Therefore, $\alpha_{ji} = \alpha_{ij}$.

2 Properties

This section is devoted to the study of some important properties of KS distributions. First, the structure of marginal distributions and conditional probabilities is considered. Throughout the paper let A, B, and C be disjoint proper subsets of V and let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ denote the powersets of A, B, and C, respectively. Whenever A and B are a partition of V, the set $\mathcal{D}$ is defined as

$$\mathcal{D} = \{I \in \mathcal{V} | I \cap A \neq \emptyset \text{ and } I \cap B \neq \emptyset\}.$$

Recalling the above definition $\mathcal{I}_{\mathcal{V}}$ indicates the set $\mathcal{I}_{\mathcal{V}} := \mathcal{I} = \{I \in \mathcal{V} \text{ with } |I| \geq 2\}$. The sets $\mathcal{I}_{\mathcal{A}}$, $\mathcal{I}_{\mathcal{B}}$ and $\mathcal{I}_{\mathcal{D}}$ are defined correspondingly.
In addition, let

$$\begin{aligned} \mathcal{I}_{\mathcal{D}_{AB}} &= \{I \in \mathcal{I}_{\mathcal{D}} | I \cap A = J_A \in \mathcal{I}_{\mathcal{A}} \text{ and } I \cap B = J_B \in \mathcal{I}_{\mathcal{B}}\}, \\ \mathcal{I}_{\mathcal{D}_{A.}} &= \{I \in \mathcal{I}_{\mathcal{D}} | I \cap A = J_A \in \mathcal{I}_{\mathcal{A}} \text{ and } I \cap B = \{j\}, j \in B\}, \\ \mathcal{I}_{\mathcal{D}_{.B}} &= \{I \in \mathcal{I}_{\mathcal{D}} | I \cap A = \{j\}, j \in A, \text{ and } I \cap B = J_B \in \mathcal{I}_{\mathcal{B}}\} \\ \text{and} \quad \mathcal{I}_{\mathcal{D}_{..}} &= \{I \in \mathcal{I}_{\mathcal{D}} | I \cap A = \{j_A\}, j_A \in A, \text{ and } I \cap B = \{j_B\}, j_B \in B\}. \end{aligned}$$

Now, a partition A, B, and C of V is considered. In analogy to the above set $\mathcal{D}$ the set $\mathcal{AB} = \{I \subset (A \cup B) | I \cap A \neq \emptyset \text{ and } I \cap B \neq \emptyset\}$ and in the same way $\mathcal{AC}$, $\mathcal{BC}$ and $\mathcal{ABC}$ are defined. 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 formulated. Following the proceeding for the set $\mathcal{D}$ the sets $\mathcal{I}_{\mathcal{AB}_{AB}}, \ldots, \mathcal{I}_{\mathcal{ABC}_{ABC\ldots}}$ are defined.

Remark 2.1 (a) For a partition A and B of V, the following statements are valid:

(i) The sets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$ form a partition of $\mathcal{V}$.

(ii) $\mathcal{D} = \{I \in \mathcal{V} | \exists\, i \in A, j \in B \text{ with } i \in I \text{ and } j \in I\}$.

(iii) The sets $\mathcal{I}_{\mathcal{D}_{AB}}$, $\mathcal{I}_{\mathcal{D}_{A.}}$, $\mathcal{I}_{\mathcal{D}_{.B}}$, and $\mathcal{I}_{\mathcal{D}_{..}}$ form a partition of $\mathcal{I}_{\mathcal{D}}$.

(b) For a partition A, B, and C of V the sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{AB}$, $\mathcal{AC}$, $\mathcal{BC}$, and $\mathcal{ABC}$ form a partition of $\mathcal{V}$.

2.1 Marginal distributions and conditional probabilities

In the following the structure of marginal distributions and particular conditional probabilities are established. In the first theorem it is shown that KS distributions are closed under marginalization. Conditional distributions are in general not again of KS type. Some conditional probabilities, however, can be explicitly formulated.

Theorem 2.2 *Let $X = X_V = (X_1, \ldots, X_p)^T$ be a KS distributed random vector with parameters α_I, $I \in \mathcal{V}$, let further $A \neq \emptyset$ be a proper subset of V, and $\mathcal{J} = \{J \in \mathcal{A} \text{ with } |J| \geq 2\}$. The marginal distribution of $X_A = (X_j)_{j \in A}$ is then again a KS distribution with parameters β_J, $J \in \mathcal{A}$, where $\beta_J = \sum_{I \in \mathcal{I}, I \cap A = J} \alpha_I$ for all $J \in \mathcal{J}$ and $\beta_j = \alpha_j + \sum_{I \in \mathcal{I}, I \cap A = \{j\}} \alpha_I$ for all $j \in A$.*

Proof:
The distribution function of X_A is given by

$$F_A(x_A) = \lim_{x_{V \setminus A} \to \infty} F_V(x_V) = \prod_{j \in A} F_j(x_j) \prod_{I \in \mathcal{I}} \lim_{x_{V \setminus A} \to \infty} c_I(x)^{-\alpha_I}.$$

Inspecting the terms $\lim_{x_{V \setminus A} \to \infty} c_I(x)$ for $I \in \mathcal{I}$, the following four cases have to be taken into account:

(i) For $I \cap (V \setminus A) = \emptyset$, i.e. for $I \subseteq A$, it trivially holds that $\lim_{x_{V \setminus A} \to \infty} c_I(x) = c_I(x)$.

(ii) For $I \cap A = \emptyset$, i.e. for $I \subseteq (V \setminus A)$, note that

$$(*) \qquad \lim_{x_i \to \infty} u_i(x_i) = \lim_{x_i \to \infty} F_i(x_i)^{\frac{1}{\alpha_{i+}}} = 1.$$

Thus, $\lim_{x_{V \setminus A} \to \infty} c_I(x) = 1$.

(iii) For $I \cap A = J \neq \emptyset$ and $I \cap (V \setminus A) \neq \emptyset$ with $|J| > 1$, $(*)$ yields $\lim_{x_{V\setminus A}\to\infty} c_I(x) = c_J(x)$.

(iv) For $I \cap A = \{j\}$ with $j \in A$ and $I \cap (V \setminus A) \neq \emptyset$, $(*)$ implies $\lim_{x_{V\setminus A}\to\infty} c_I(x) = 1$.

Thus, $\beta_J = \sum_{I\in\mathcal{I}, I\cap A = J} \alpha_I$ for all $J \in \mathcal{J}$. In addition, for all $j \in A$ we obviously have that

$$\beta_{j+} = \alpha_{j+} = \sum_{\substack{I\in\mathcal{V}\\ j\in I}} \alpha_I = \alpha_j + \sum_{\substack{I\in\mathcal{I}\\ j\in I}} \alpha_I = \alpha_j + \sum_{\substack{I\in\mathcal{I}\\ j\in I\\ I\cap A=\{j\}}} \alpha_I + \sum_{\substack{I\in\mathcal{I}\\ j\in I\\ I\cap A\in\mathcal{J}}} \alpha_I$$

and $\beta_j = \alpha_j + \sum_{\substack{I\in\mathcal{I}\\ I\cap A=\{j\}}} \alpha_I$. □

Example 2.3 *Consider the case $p = 3$, i.e. $X = (X_1, X_2, X_3)^T$ is a KS distributed random vector. The two–dimensional marginal distribution of (X_1, X_2) is given by the following cdf*

$$\begin{aligned}
&\lim_{x_3\to\infty} F(x_1, x_2, x_3) \\
&= \lim_{x_3\to\infty} F_1(x_1)F_2(x_2)F_3(x_3)c_{12}(x)^{-\alpha_{12}}c_{13}(x)^{-\alpha_{13}}c_{23}(x)^{-\alpha_{23}}c_{123}(x)^{-\alpha_{123}} \\
&= F_1(x_1)F_2(x_2)c_{12}(x)^{-\alpha_{12}} \lim_{x_3\to\infty} c_{13}(x)^{-\alpha_{13}} \lim_{x_3\to\infty} c_{23}(x)^{-\alpha_{23}} \lim_{x_3\to\infty} c_{123}(x)^{-\alpha_{123}} \\
&= F_1(x_1)F_2(x_2)c_{12}(x)^{-\alpha_{12}}c_{12}(x)^{-\alpha_{123}} \\
&= F_1(x_1)F_2(x_2)c_{12}(x)^{-(\alpha_{12}+\alpha_{123})} \\
&= F_1(x_1)F_2(x_2)c_{12}(x)^{-\beta_{12}},
\end{aligned}$$

where $\beta_{12} = \alpha_{12} + \alpha_{123}$, $\beta_{1+} = \alpha_{1+}$, and $\beta_{2+} = \alpha_{2+}$. Thus, it can be concluded that $\beta_1 = \alpha_1 + \alpha_{13} > 0$ and $\beta_2 = \alpha_2 + \alpha_{23} > 0$.
This result can be obtained directly from the above theorem:
Here, $V = \{1, 2, 3\}$, $\mathcal{I} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$, $A = \{1, 2\}$, and $\mathcal{J} = \{\{1, 2\}\}$. That means, there is only one set $J \in \mathcal{J}$. In addition, only for the elements $I_1 = \{1, 2\}$ and $I_2 = \{1, 2, 3\}$ of $\mathcal{I}$ it holds that $I_1 \cap A = J$ and $I_2 \cap A = J$ which yields $\beta_{12} = \alpha_{12} + \alpha_{123}$. Correspondingly, $I = \{1, 3\} \in \mathcal{I}$ is the only set with $I \cap A = \{1\}$ and only for the set $I = \{2, 3\} \in \mathcal{I}$ it holds that $I \cap A = \{2\}$. Summarizing, it can be seen that $\beta_1 = \alpha_1 + \alpha_{13}$ and $\beta_2 = \alpha_2 + \alpha_{23}$.

Theorem 2.2 directly implies the following corollary. Note that (iii) of the proof does not occur.

Corollary 2.4 *For a KS(2) distributed vector $X = X_V = (X_1, \ldots, X_p)^T$ with parameters $\alpha_i > 0$ for all $i \in V$ and $\alpha_{ij} \geq 0$ for $i, j \in V$ with $i \neq j$ and a set $A \subset V$ the subvector $X_A = (X_j)_{j\in A}$ is again KS(2) distributed with parameters $\beta_{jk} = \alpha_{jk}$ for $j, k \in A$, $j \neq k$ and $\beta_j = \alpha_j + \sum_{k\in V\setminus A} \alpha_{jk}$ for all $j \in A$.*

Example 2.5 *Consider again the case $p = 3$. Let now $X = (X_1, X_2, X_3)^T$ be KS(2) distributed. The two–dimensional marginal distribution of (X_1, X_2) is then given as*

$$\begin{aligned}\lim_{x_3\to\infty} F(x_1,x_2,x_3) &= \lim_{x_3\to\infty} F_1(x_1)F_2(x_2)F_3(x_3)c_{12}(x)^{-\alpha_{12}}c_{13}(x)^{-\alpha_{13}}c_{23}(x)^{-\alpha_{23}} \\ &= F_1(x_1)F_2(x_2)c_{12}(x)^{-\alpha_{12}}.\end{aligned}$$

It is easy to see that $\beta_{12} = \alpha_{12}$, $\beta_{1+} = \alpha_{1+}$, and $\beta_{2+} = \alpha_{2+}$ which implies that $\beta_1 = \alpha_1+\alpha_{13}$ and $\beta_2 = \alpha_2 + \alpha_{23}$.

For a KS(2) distributed vector X_V and a set $A \subset V$ the association parameters β_J for $J \in \mathcal{J}_2 = \{J \in \mathcal{J} \text{ with } |J| = 2\}$ of the marginal distribution of the subvector X_A coincide with the association parameters α_I for $I \in \mathcal{J}_2$ of the distribution of X_V. Just the parameters β_j for $j \in A$ have to be modified in accordance with Corollary 2.4.
The explicit requirement that $\alpha_i > 0$ for all $i \in V$, i.e. the imperative inclusion of these parameters into the model guarantees that the distribution family is closed under marginalization. Koehler and Symanowski consider the simpler case $\alpha_i = 0$ for all $i \in V$. Hence, this subfamily is no longer closed under marginalization which can be seen by means of the above corollary. In this case, parameters $\beta_j > 0$ for $j \in A$ occur.
Formulae for certain conditional probabilities can be derived using the result of the functional representation of the marginal cdf. In the special case of KS(2) distributions these probabilities are given by simple expressions. The following corollary can be directly obtained from Theorem 2.2.

Corollary 2.6 *Let $X = X_V = (X_1, \ldots, X_p)^T$ be KS distributed and for $A \subset V$ define $B = V \setminus A$. Then for all $x \in I\!R^p$ with $F_B(x_B) \neq 0$ it holds*

$$\begin{aligned}&P(X_A \le x_A | X_B \le x_B) \\ &\quad = F_A(x_A) \prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} \left\{\frac{c_I(x)}{c_{I\cap A}(x)c_{I\cap B}(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} \left\{\frac{c_I(x)}{c_{I\cap A}(x)}\right\}^{-\alpha_I} \\ &\qquad\qquad \prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} \left\{\frac{c_I(x)}{c_{I\cap B}(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{..}}} c_I(x)^{-\alpha_I}.\end{aligned}$$

Proof:
The definition of conditional probabilities yields

$$P(X_A \le x_A | X_B \le x_B) = \frac{P(X_A \le x_A, X_B \le x_B)}{P(X_B \le x_B)} = \frac{F(x)}{F_B(x_B)}.$$

Using the above notations and the abbreviations $IA = I \cap A$ and $IB = I \cap B$ the last ratio can be rewritten as

$$
\begin{aligned}
\frac{F(x)}{F_B(x_B)} &= \frac{\prod_{i\in A} F_i(x_i)\prod_{i\in B} F_i(x_i)\prod_{I\in\mathcal{I}_A} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_B} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_\mathcal{D}} c_I(x)^{-\alpha_I}}{\prod_{i\in B} F_i(x_i)\prod_{I\in\mathcal{I}_B} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_{IB}(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} c_{IB}(x)^{-\alpha_I}}\\
&= F_A(x_A)\frac{\prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} c_I(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{..}}} c_I(x)^{-\alpha_I}}{\prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_{IA}(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} c_{IA}(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_{IB}(x)^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} c_{IB}(x)^{-\alpha_I}}\\
&= F_A(x_A)\prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}}\left\{\frac{c_I(x)}{c_{IA}(x)c_{IB}(x)}\right\}^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}}\left\{\frac{c_I(x)}{c_{IA}(x)}\right\}^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}}\left\{\frac{c_I(x)}{c_{IB}(x)}\right\}^{-\alpha_I}\prod_{I\in\mathcal{I}_{\mathcal{D}_{..}}} c_I(x)^{-\alpha_I}.
\end{aligned}
$$

□

Example 2.7 *Let $X = (X_1, X_2, X_3, X_4)^T$ be KS distributed with parameters α_I for $I \in \mathcal{V}$. Take $A = \{1,2\} \subset V$ and $B = V \setminus A = \{3,4\}$. Then, we get*

$$\mathcal{I}_\mathcal{D} = \{\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\},\{1,2,3,4\}\}$$

with the partition

$$
\begin{aligned}
\mathcal{I}_{\mathcal{D}_{AB}} &= \{\{1,2,3,4\}\},\\
\mathcal{I}_{\mathcal{D}_{A.}} &= \{\{1,2,3\},\{1,2,4\}\},\\
\mathcal{I}_{\mathcal{D}_{.B}} &= \{\{1,3,4\},\{2,3,4\}\}\\
\textit{and}\quad \mathcal{I}_{\mathcal{D}_{..}} &= \{\{1,3\},\{1,4\},\{2,3\},\{2,4\}\}.
\end{aligned}
$$

Thus, the two–dimensional conditional probability reads as

$$
\begin{aligned}
&P(X_1 \le x_1, X_2 \le x_2 | X_3 \le x_3, X_4 \le x_4)\\
&\quad = F_{12}(x_1,x_2)\left\{\frac{c_{1234}(x)}{c_{12}(x)c_{34}(x)}\right\}^{-\alpha_{1234}}\left\{\frac{c_{123}(x)}{c_{12}(x)}\right\}^{-\alpha_{123}}\left\{\frac{c_{124}(x)}{c_{12}(x)}\right\}^{-\alpha_{124}}\\
&\qquad \left\{\frac{c_{134}(x)}{c_{34}(x)}\right\}^{-\alpha_{134}}\left\{\frac{c_{234}(x)}{c_{34}(x)}\right\}^{-\alpha_{234}} c_{13}(x)^{-\alpha_{13}}\, c_{14}(x)^{-\alpha_{14}}\, c_{23}(x)^{-\alpha_{23}}\, c_{24}(x)^{-\alpha_{24}}.
\end{aligned}
$$

Corollary 2.8 *Let $X = X_V$ be KS(2) distributed and $A \subset V$, then it holds*

$$P(X_A \le x_A | X_{V\setminus A} \le x_{V\setminus A}) = F_A(x_A)\prod_{i\in A}\prod_{j\in V\setminus A} c_{ij}(x)^{-\alpha_{ij}}.$$

Proof:

The set $\mathcal{I}_\mathcal{D}$ from Corollary 2.6 is here seen to be

$$\mathcal{I}_\mathcal{D} = \mathcal{I}_{\mathcal{D}_{..}} = \{I \in \mathcal{I}_\mathcal{D} | I \cap A = \{i\},\ i \in A,\ \text{and } I \cap (V\setminus A) = \{j\},\ j \in V \setminus A\}$$

which implies the above assumption. □

Example 2.9 *For a KS(2) distributed vector $X = (X_1, X_2, X_3, X_4)^T$ and $A = \{1,2\}$ as in Example 2.7 we get*

$$
\begin{aligned}
&P(X_1 \le x_1, X_2 \le x_2 | X_3 \le x_3, X_4 \le x_4)\\
&\quad = F_{12}(x_1,x_2)c_{13}(x)^{-\alpha_{13}}\, c_{14}(x)^{-\alpha_{14}}\, c_{23}(x)^{-\alpha_{23}}\, c_{24}(x)^{-\alpha_{24}}.
\end{aligned}
$$

2.2 Marginal and conditional independence

Multivariate distributions do often not allow for association structures where pairs of components or subsets of the vector X are marginally or conditionally independent. In this subsection this problem is discussed for KS distributions. Combinations of the parameters and restrictions on them are considered which allow to model independences among components and subvectors of X.

Theorem 2.10 *Let $X = X_V = (X_1, \ldots, X_p)^T$ be a KS distributed vector with parameters α_I, $I \in \mathcal{V}$.*

(i) The variables $X_1, \ldots, X_p$ are marginally independent whenever $\alpha_I = 0$ holds for all $I \in \mathcal{I}$.

(ii) Let the variables $X_1, \ldots, X_p$ be marginally independent and assume each X_i ($i \in V$) to be not degenerated, i.e. there exists at least one $\tilde{x}_i \in I\!R$ for which the corresponding distribution function takes a value $F_i(\tilde{x}_i) = y_i$ with $0 < y_i < 1$, then it holds that $\alpha_I = 0$ for all $I \in \mathcal{I}$.

Proof:

(i) For $\alpha_I = 0$ $\forall I \in \mathcal{I}$, all association terms $c_I(x)^{-\alpha_I}$ regarded as functions of x equal the function $c(x) = 1$. Therefore, the distribution function is the product of the marginal distribution functions. As a result, the variables $X_1, \ldots, X_p$ are marginally independent.

(ii) Consider vice versa, $X_1, \ldots, X_p$ as marginally independent. Then the joint distribution function turns out to be the product of the marginal distribution functions. This is the case if all association terms for all values of $x \in I\!R^p$ with $F(x) \neq 0$ equal one.

Due to condition

$$0 < \alpha_{i+} = \sum_{\substack{I \in \mathcal{V} \\ i \in I}} \alpha_I < \infty$$

as part of the definition of the KS distribution in Section 1 the functions

$$c_I(\tilde{x}) = c_I((\tilde{x}_i)_{i \in I}^T) = \sum_{i \in I} \prod_{\substack{j \in I \\ j \neq i}} u_j(\tilde{x}_j) - (|I| - 1) \prod_{i \in I} u_i(\tilde{x}_i),$$

do not equal 1 for arbitrary sets $I \in \mathcal{I}$. Thus, $c_I(x)^{-\alpha_I} = 1$ for all $x \in I\!R^p$ if and only if $\alpha_I = 0$. $\square$

Here, the restriction $\alpha_{i+} < \infty$ is crucial. If it is dropped, for instance in the case of a KS(2) distribution, a pairwise association between X_i and X_j vanishes for $\alpha_{i+} = \alpha_{j+} = \infty$.

Lemma 2.11 *Let $X = X_V$ be KS distributed with parameters α_I for $I \in \mathcal{V}$ and A and B a partition of V.*

(i) The vectors X_A and X_B are marginally independent whenever for all $i \in A$ and all $j \in B$ the condition

$$\alpha_I = 0 \quad \textit{for all} \quad I \in \mathcal{I} \quad \textit{with} \quad i \in I \textit{ and } j \in I,$$

is fulfilled.

(ii) Let the vectors X_A and X_B be marginally independent and assume each X_i for $i \in V$ to be not degenerated. Then, for all $i \in A$ and $j \in B$ the condition $\alpha_I = 0$ is satisfied whenever $i \in I$ and $j \in I$.

The proof of this lemma is given in the Appendix.

Example 2.12 *For a vector $X = (X_1, X_2, X_3)^T$ with joint KS distribution and parameters $\alpha_1, \alpha_2, \alpha_3 > 0$ and $\alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{123} \geq 0$ the subvectors $(X_1, X_2)^T$ and X_3 are marginally independent if $\alpha_{13} = \alpha_{23} = \alpha_{123} = 0$. The joint cdf of X is then given as*

$$F(x_1, x_2, x_3) = F_1(x_1)F_2(x_2)F_3(x_3)c_{12}(x)^{-\alpha_{12}},$$

where the cdf of the marginal distribution of $(X_1, X_2)^T$ is

$$F_{12}(x_1, x_2) = F_1(x_1)F_2(x_2)c_{12}(x)^{-\alpha_{12}}.$$

Obviously, this leads to $F(x_1, x_2, x_3) = F_{12}(x_1, x_2)F_3(x_3)$.

Theorem 2.13 *Consider a vector $X = X_V$ with joint KS distribution and a partition of V into the sets A, B and C.*

(i) The vectors X_A and X_B are marginally independent whenever for all $i \in A$ and $j \in B$ the condition

$$\alpha_I = 0 \quad \textit{for all} \quad I \in \mathcal{I} \quad \textit{with} \quad i \in I \textit{ and } j \in I$$

is fulfilled.

(ii) Let X_A and X_B be marginally independent and assume each X_i for $i \in V$ to be not degenerated. Then, for all $i \in A$ and $j \in B$ it holds that $\alpha_I = 0$ whenever $i \in I$ and $j \in I$.

Proof:
From Theorem 2.2 it can be concluded that $(X_A^T, X_B^T)^T = (X_i)_{i\in A\cup B}$ follows a KS distribution. For $\mathcal{J} = \{J \subset (A\cup B) \text{ with } |J| \geq 2\}$ the corresponding association parameters β_J, $J \in \mathcal{J}$, are given by

$$(*) \quad \beta_J = \sum_{\substack{I\in\mathcal{I} \\ I\cap(A\cup B)=J}} \alpha_I \quad \text{for all } J \in \mathcal{J}.$$

Let for all $i \in A$, $j \in B$ $\alpha_I = 0$ whenever $i, j \in I$. Then, for each $J \in \mathcal{J}$ all terms of the sum in $(*)$ equal zero and thus, β_J itself equals zero. From Lemma 2.11 it can be seen that X_A and X_B are marginally independent which provides the proof of part (i) of the theorem.
Now, suppose that X_A and X_B are independent and consider the parameters of the corresponding marginal distribution of $(X_A^T, X_B^T)^T$. Then, for all $i \in A$ and $j \in B$ it holds according to Lemma 2.11 that

$$\beta_J = 0 \quad \text{for all} \quad J \in \mathcal{J} \text{ with } \quad i, j \in J.$$

Using $(*)$ it directly follows for all $I \in \mathcal{I}$, $I \cap (A\cup B) = J$ that $\alpha_I = 0$. $\square$

Remark 2.14 In particular, the components X_i and X_j, $i, j \in V$ and $i \neq j$, of a KS distributed vector X_V are marginally independent if $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i \in I$ and $j \in I$.

As a direct implication of Theorem 2.13 we have the following corollary.

Corollary 2.15 *Let $X = X_V$ be KS(2) distributed and let A, B be disjoint sets with $A \cup B \subseteq V$.*

(i) The subvectors X_A and X_B are marginally independent whenever $\alpha_{ij} = 0$ for all $i \in A$ and $j \in B$.

(ii) Let X_A and X_B be marginally independent and assume each X_i for $i \in V$ to be not degenerated. Then, for all $i \in A$ and $j \in B$ it holds that $\alpha_{ij} = 0$.

Theorem 2.16 *Let $X = X_V = (X_1, \ldots, X_p)^T$ be KS distributed and let the sets A, B, and C be a partition of V.*

(i) X_A and X_B are conditionally independent given X_C whenever for all $i \in A$ and $j \in B$ the condition

$$\alpha_I = 0 \quad \textit{for all} \quad I \in \mathcal{I} \quad \textit{with} \quad i \in I \textit{ and } j \in I$$

is fulfilled.

(ii) Let X_A and X_B be conditionally independent given X_C and assume each X_i for $i \in V$ to be not degenerated. Then, it holds that for all $i \in A$ and $j \in B$ $\alpha_I = 0$ whenever $i \in I$ and $j \in I$.

The proof is given in the Appendix.
The above theorem directly leads to Corollary 2.17.

Corollary 2.17 *Let $X = X_V$ be KS(2) distributed and let the sets $A, B,$ and C be a partition of V.*

(i) The subvectors X_A and X_B are conditionally independent given X_C whenever $\alpha_{ij} = 0$ for all $i \in A$ and $j \in B$.

(ii) Let X_A and X_B be conditionally independent given X_C and assume each X_i for $i \in V$ to be not degenerated. Then, it holds for all $i \in A$ and $j \in B$ that $\alpha_{ij} = 0$.

As summarizing result of Theorems 2.13 and 2.16 we get

Theorem 2.18 *For a KS distributed vector $X = X_V$ with not degenerated marginal distributions and a partition of V into the sets A, B, and C the vectors X_A and X_B are conditionally independent given X_C if and only if X_A and X_B are marginally independent.*

3 Simulation Study

The above results show that the absence of specific association parameters implies the marginal or conditional independence of components of the vector of random variables under consideration. In the case of a KS(2) distribution the incorporation of the parameter $\alpha_{ij} > 0$ $(i, j \in V, i \neq j)$ indicates that X_i and X_j are marginally dependent. But there is nothing known about a general functional relation between this parameter and common measures like Kendall's τ_{ij} which quantify the strength of the correlation between X_i and X_j. Merely, for the generalized Burr–Pareto–logistic distribution with $p = 2$ the relation between α and Kendall's τ is derived by Oakes (1982). Thus, a simulation study has been performed on a SUN Sparc–station 10 to interpret the association parameters. In the following, the results are presented for KS(2) distributions with $p = 3$. According to the instructions given in Koehler and Symanowski (1995) samples are taken from a three–dimensional KS(2) distribution for different values for $\alpha_1, \alpha_2, \alpha_3, \alpha_{12}, \alpha_{13}$, and α_{23}. For each sample the correlation coefficient as a measure for the linear and Kendall's τ as a measure for the monotone relation is calculated for all pairs of variables. As in a comparable study in Koehler and Symanowski (1995) the marginal distributions are Weibull with different parameter constellations.

Table 3.1 shows the results for a sample of size $n = 30000$ where the same parameter values are chosen as in Koehler and Symanowski (1995), i.e. the parameters of the marginal distributions are $\beta_1 = 5.75$, $\beta_2 = 4.24$, and $\beta_3 = 3.16$. However, the authors consider in their simulations only the special case that $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Table 3.1: Correlation coefficients and Kendall's τ for $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $\beta_1 = 5.75$, $\beta_2 = 4.24$, $\beta_3 = 3.16$

association parameter			Kendall's τ			correlation coefficient		
α_{12}	α_{13}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.01	0.326	0.327	0.330	0.393	0.400	0.413
0.10	0.10	0.01	0.388	0.400	0.048	0.513	0.535	0.068
0.50	0.01	0.01	0.479	0.020	0.022	0.684	0.040	0.042
0.10	0.02	0.001	0.690	0.156	0.010	0.822	0.250	0.016
0.50	0.50	0.50	0.147	0.146	0.146	0.223	0.220	0.217
1.00	0.10	0.10	0.276	0.061	0.062	0.434	0.095	0.094
1.60	0.50	0.50	0.140	0.073	0.076	0.230	0.115	0.117
5.00	0.10	0.01	0.098	0.018	0.001	0.164	0.032	0.001
5.00	5.00	5.00	0.027	0.023	0.027	0.041	0.035	0.041

Strikingly, the results differ slightly from those in Koehler and Symanowski. On the one hand this can be caused by the use of different random generators and on the other hand to the inversion of the cdf of the gamma distribution which can be carried out in various ways. A comparison of the first, fifth, and last row of Table 3.1 shows that the values of the correlation coefficients decrease for increasing parameter values. But if the first and third row are looked at where $\alpha_{13} = \alpha_{23} = 0.01$ the correlation between X_1 and X_2 is smaller for $\alpha_{12} = 0.01$ than for $\alpha_{12} = 0.50$. Here, the correlation increases for increasing association parameter. Thus, the results depend on the constellation of all parameters involved which can also be seen comparing the third and fifth row. In both cases α_{12} is set to be 0.50. The combination with $\alpha_{13} = \alpha_{23} = 0.01$ yields $\tau_{12} = 0.479$ whereas the numbers $\alpha_{13} = \alpha_{23} = 0.50$ result in $\tau_{12} = 0.147$. In addition, high values of the correlation measures only seldom arise.

The definition of the KS distribution induces that the choice of the parameters of the marginal distributions and even the type of the marginal distributions do not affect the strength of the association. This fact is illustrated by the results for $\beta_1 = \beta_2 = \beta_3 = 5$ and the same combinations for the parameters of the KS(2) distribution as above which are shown in Table 3.2. Koehler and Symanowski restrict their simulations to the case that

Table 3.2: Correlation coefficient and Kendall's τ for $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $\beta_1 = \beta_2 = \beta_3 = 5$

association parameter			Kendall's τ			correlation coefficient		
α_{12}	α_{13}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.01	0.326	0.327	0.327	0.393	0.390	0.391
0.10	0.10	0.01	0.398	0.399	0.049	0.519	0.525	0.067
0.50	0.01	0.01	0.481	0.019	0.019	0.689	0.028	0.029
0.10	0.02	0.001	0.688	0.158	0.008	0.818	0.229	0.009
0.50	0.50	0.50	0.146	0.146	0.147	0.222	0.222	0.225
1.00	0.10	0.10	0.273	0.067	0.060	0.433	0.099	0.090
1.60	0.50	0.50	0.138	0.082	0.077	0.228	0.125	0.120
5.00	0.10	0.01	0.081	0.014	0.003	0.140	0.025	0.005
5.00	5.00	5.00	0.022	0.021	0.022	0.041	0.036	0.041

$\alpha_i = 0$ for $i \in V$. Here, the focus lies on models where pairs of variables are marginally independent. Under the assumption $\alpha_i > 0$ for $i \in V$ this variables X_i and X_j with $i, j \in V$ and $i \neq j$ are marginally independent if and only if $\alpha_{ij} = 0$. Therefore in the following, situations are investigated in which α_1, α_2, and α_3 are strictly positive. To reveal the effect of the parameters α_1, α_2, and α_3 on the strength of the pairwise correlations the three parameters of the marginal distributions and the association parameters are chosen to be equal. In Table 3.3 some typical results for different situations with $\alpha_1 = \alpha_2 = \alpha_3$ are given. It can be seen that the correlation coefficients decrease for increasing values for $\alpha_1, \alpha_2, \alpha_3$. Thus, small values for $\alpha_1, \alpha_2, \alpha_3$ seem to be more sensible to discriminate pairwise dependence from pairwise independence. Table 3.4 shows the outcomes for the case that X_1 and X_3 are marginally independent, i.e. it holds that $\alpha_{13} = 0$. The results point out that the interpretation of the parameters with respect to the strength of the pairwise correlations is difficult. At first, the correlation coefficient and Kendall's τ get larger for increasing $\alpha_{12} = \alpha_{23}$ but then get again smaller. The same effect can be observed for different values for α_{12} and α_{23}: The constellation $\alpha_{12} = 0.01$ and $\alpha_{23} = 0.20$ in the eighth row leads to a pairwise correlation of $\rho_{23} = 0.778$ whereas in the last row $\rho_{23} = 0.683$ results for $\alpha_{12} = 0.01$ and $\alpha_{23} = 0.50$. Altogether, it can be determined that though KS distributions enable a wide range of pairwise correlations, the connection between the association parameters on the one and the correlation measures on the other hand is

Table 3.3: Correlation coefficient and Kendall's τ for $\beta_1 = \beta_2 = \beta_3 = 5$

$\alpha_1 = \alpha_2 = \alpha_3 = 0.01$								
association parameter			Kendall's τ			correlation coefficient		
α_{12}	α_{13}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.01	0.189	0.196	0.189	0.245	0.252	0.250
0.05	0.05	0.05	0.264	0.266	0.266	0.350	0.348	0.351
0.10	0.10	0.10	0.252	0.246	0.252	0.344	0.331	0.343
$\alpha_1 = \alpha_2 = \alpha_3 = 0.05$								
α_{12}	α_{13}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.01	0.077	0.077	0.078	0.108	0.105	0.109
0.05	0.05	0.05	0.170	0.173	0.175	0.231	0.233	0.236
0.10	0.10	0.10	0.201	0.200	0.198	0.277	0.273	0.271
$\alpha_1 = \alpha_2 = \alpha_3 = 0.1$								
α_{12}	α_{13}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.01	0.051	0.038	0.041	0.069	0.050	0.057
0.05	0.05	0.05	0.117	0.116	0.120	0.159	0.156	0.167
0.10	0.10	0.10	0.155	0.153	0.160	0.216	0.213	0.224

complex and not easy to interpret. The magnitude of the association parameters does not permit a direct conclusion on the strength of a possible linear or monotone relation among the variables. Thus, the association parameters describe other than linear and monotone associations. Up to know, an interpretation of the parameters is only possible via simulation studies.

4 Discussion

Our investigations have shown that the family of KS distributions offers possibilities which are not given by a wide range of other multivariate distributions. On the one hand KS distributions are able to model other than linear relations between the variables. Koehler and Symanowski (1995) discuss the shape of two–dimensional KS distributions. They compare, for instance, two–dimensional normal distributions with two–dimensional KS

Table 3.4: Correlation coefficient and Kendall's τ for $\alpha_1 = \alpha_2 = \alpha_3 = 0.01$ and $\alpha_{13} = 0$

association parameter		Kendall's τ			correlation coefficient		
α_{12}	α_{23}	τ_{12}	τ_{13}	τ_{23}	ρ_{12}	ρ_{13}	ρ_{23}
0.01	0.01	0.239	0.000	0.243	0.302	0.001	0.306
0.05	0.05	0.378	0.003	0.382	0.488	0.004	0.490
0.10	0.10	0.370	0.002	0.378	0.493	0.002	0.496
0.15	0.15	0.358	0.003	0.358	0.485	0.004	0.483
0.20	0.20	0.339	0.003	0.339	0.470	0.005	0.469
0.50	0.50	0.246	0.002	0.247	0.365	0.004	0.366
0.01	0.10	0.077	0.000	0.637	0.115	0.003	0.767
0.01	0.20	0.041	0.000	0.611	0.069	0.002	0.778
0.01	0.50	0.014	0.002	0.477	0.026	0.003	0.683

distributions having normally distributed margins likewise. By means of contour line plots of the density function they illustrate how manifold forms of appearence are obtained for different association parameters. The potentiality of modeling pairwise independencies is on the other hand a request which is not fulfilled for many multivariate distributions defined in the literature.

Nevertheless, the approach contains problems which reduce the scale of applications. It is questionable whether the flexibility concerning possible association structures is useful while there is no knowledge about type or strength of for example pairwise associations. In addition, it has to be checked what consequences have to be drawn from the equivalence of conditional and marginal independence. Another open question concerns the discussion of estimating and testing properties. The familiy of KS distributions is no exponential family which means that common results cannot be applied (Barndorff–Nielsen, 1978, Frydenberg and Lauritzen, 1987). Exponential families are characterized by a special structure of the density whereas the family of KS distributions is featured by the cdf. The corresponding density is given by a complicated formula which is not easy to treat analytically. Eventually, the discussion of a generalized form with uniform margins as it can be found for the generalized Burr–Pareto–logistic distributions in Johnson (1987) will be more easy and therefore sensible.

Summarizing, this distribution family seems to be worth being discussed in the framework of graphical models. Admittedly, attention has to be paid to the observed equivalence of marginal and conditional independence. Perhaps terms derived from the theory of graphical models will be helpful to interpret this property. In a next step it has to be examined whether the so–called Markov properties (Frydenberg, 1990) hold which make it possible to represent the association structure in a graph in such a way that missing edges can be interpreted as conditional or marginal independences which is investigated in a forthcoming paper.

Acknowledgment

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

Appendix

Proof of Lemma 2.11:

The joint distribution function of $(X_A^T, X_B^T)^T$ can be written as

$$\begin{aligned}
F(x_A, x_B) &= \prod_{i \in V} F_i(x_i) \prod_{I \in \mathcal{I}_\mathcal{V}} c_I(x)^{-\alpha_I} \\
&= \prod_{i \in A} F_i(x_i) \prod_{i \in B} F_i(x_i) \prod_{I \in \mathcal{I}_\mathcal{A}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_\mathcal{B}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_\mathcal{D}} c_I(x)^{-\alpha_I}.
\end{aligned}$$

The marginal distribution functions of X_A and X_B are of the form

$$\begin{aligned}
F_A(x_A) &= \prod_{j \in A} F_j(x_j) \prod_{J \in \mathcal{I}_\mathcal{A}} c_J(x)^{-\beta_J} \\
&= \prod_{j \in A} F_j(x_j) \prod_{J \in \mathcal{I}_\mathcal{A}} c_J(x)^{-\sum_{I \in \mathcal{I}_\mathcal{V}, I \cap A = J} \alpha_I} \\
&= \prod_{j \in A} F_j(x_j) \prod_{J \in \mathcal{I}_\mathcal{A}} c_J(x)^{-\alpha_J} \prod_{J \in \mathcal{I}_\mathcal{A}} c_J(x)^{-\sum_{I \in \mathcal{I}_\mathcal{V}, I \cap A = J, I \neq J} \alpha_I} \\
&= \prod_{j \in A} F_j(x_j) \prod_{J \in \mathcal{I}_\mathcal{A}} c_J(x)^{-\alpha_J} \prod_{\substack{I \in \mathcal{I}_\mathcal{D} \\ I \cap A = J \in \mathcal{I}_\mathcal{A}}} c_J(x)^{-\alpha_I}
\end{aligned}$$

$$\text{and} \qquad F_B(x_B) = \prod_{j \in B} F_j(x_j) \prod_{J \in \mathcal{I}_\mathcal{B}} c_J(x)^{-\alpha_J} \prod_{\substack{I \in \mathcal{I}_\mathcal{D} \\ I \cap B = J \in \mathcal{I}_\mathcal{B}}} c_J(x)^{-\alpha_I}.$$

The vectors X_A and X_B are marginally independent if the joint distribution function of $(X_A^T, X_B^T)^T$ is the product of the marginal distribution functions of X_A and X_B. This is the case if for all $x = (x_A^T, x_B^T)^T \in I\!R^p$ the following equation holds

$$\prod_{I \in \mathcal{I}_\mathcal{D}} c_I(x)^{-\alpha_I} = \prod_{\substack{I \in \mathcal{I}_\mathcal{D} \\ I \cap A = J \in \mathcal{I}_\mathcal{A}}} c_J(x)^{-\alpha_I} \prod_{\substack{I \in \mathcal{I}_\mathcal{D} \\ I \cap B = J \in \mathcal{I}_\mathcal{B}}} c_J(x)^{-\alpha_I}. \tag{4.3}$$

Let $\alpha_I = 0$ for all $\mathcal{I}_\mathcal{D}$ in (4.3). As a result, both sides of the equation equal 1, and part (i) of the lemma is proven.

To show (ii) the partition of the set $\mathcal{I}_\mathcal{D}$ given in Remark 2.1 is used. Thus, the terms of Equation (4.3) can be written as

$$\prod_{I\in\mathcal{I}_\mathcal{D}} c_I(x)^{-\alpha_I} = \prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_I(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} c_I(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} c_I(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{..}}} c_I(x)^{-\alpha_I},$$

$$\prod_{\substack{I\in\mathcal{I}_\mathcal{D}\\ I\cap A=J\in\mathcal{I}_\mathcal{A}}} c_J(x)^{-\alpha_I} = \prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_{I\cap A}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} c_{I\cap A}(x)^{-\alpha_I}$$

and
$$\prod_{\substack{I\in\mathcal{I}_\mathcal{D}\\ I\cap B=J\in\mathcal{I}_\mathcal{B}}} c_J(x)^{-\alpha_I} = \prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} c_{I\cap B}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} c_{I\cap B}(x)^{-\alpha_I}.$$

It can be seen that for all $x \in I\!R^p$ with $c_I(x) \neq 0$ for $I \in \mathcal{I}_\mathcal{D}$ (4.3) is equivalent to

$$1 = \prod_{I\in\mathcal{I}_{\mathcal{D}_{AB}}} \{\frac{c_I(x)}{c_{I\cap A}(x)c_{I\cap B}(x)}\}^{-\alpha_I} \tag{4.4}$$
$$\prod_{I\in\mathcal{I}_{\mathcal{D}_{A.}}} \{\frac{c_I(x)}{c_{I\cap A}(x)}\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{.B}}} \{\frac{c_I(x)}{c_{I\cap B}(x)}\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{D}_{..}}} c_I(x)^{-\alpha_I}.$$

Now, we will show that (4.4) is fulfilled if and only if $\alpha_I = 0$ for $I \in \mathcal{I}_\mathcal{D}$. In other words, the right hand side looked at as a function of x equals 1 only in this special case. For this purpose, the right hand side of the function is evaluated at suitably chosen numbers, as for instance, $\tilde{x}_j$ where $F_i(\tilde{x}_j) = y_j$ is assumed to properly lie between zero and one. If for two components j_A and j_B, such numbers $\tilde{x}_{j_A}$ and $\tilde{x}_{j_B}$ are taken, the function $c_{j_Aj_B}(\tilde{x}_{j_A}, \tilde{x}_{j_B})$ also lies inside the interval $(0; 1)$.
For $j_A \in A$ and $j_B \in B$ the limit of the above equation is analysed for $x_i \to \infty$ for all $i \in V \setminus \{j_A, j_B\}$ and $x_{j_A} = \tilde{x}_{j_A}$, $x_{j_B} = \tilde{x}_{j_B}$ which yields that the limits of all terms $c_{I\cap A}(x)$ and $c_{I\cap B}(x)$ of (4.4) equal 1. For the functions $c_I(x)$ the limit of all terms with $\{j_A, j_B\} \not\subset I$ equal 1 as well. For the remaining terms, i.e. for all $c_I(x)$ with $I \in \mathcal{I}_\mathcal{D}$ and $j_A, j_B \in I$ it holds that

$$\lim_{\substack{x_i\to\infty\\ i\in I\setminus\{j_A,j_B\}}} c_I(x) = c_{j_Aj_B}(\tilde{x}_{j_A}, \tilde{x}_{j_B}).$$

As a result, the limit of the right hand side of (4.4) turns out to be

$$\prod_{\substack{I\in\mathcal{I}_\mathcal{D}\\ j_A,j_B\in I}} c_{j_Aj_B}(\tilde{x}_{j_A}, \tilde{x}_{j_B})^{-\alpha_I}.$$

As assumed, $0 < F_i(\tilde{x}_i) = y_i < 1$ is satisfied which gives

$$0 < c_{j_Aj_B}(\tilde{x}_{j_A}, \tilde{x}_{j_B}) < 1.$$

Summarizing, (4.4) holds if and only if $\alpha_I = 0$ for all $I \in \mathcal{I}_\mathcal{D}$ with $j_A \in I$ and $j_B \in I$.

To complete the proof it remains to show that the above argumentation covers all sets I belonging to the index set in (4.4). Since the above arguments hold for any arbitrary pair j_A, j_B and since in addition, the set $\mathcal{D}$ is defined as $\mathcal{D} = \{I \in \mathcal{V} | I \cap A \neq \emptyset \text{ and } I \cap B \neq \emptyset\}$ the right hand side of (4.4) equals 1 only if $\alpha_I = 0$ for all $I \in \mathcal{I}_{\mathcal{D}}$. $\square$

Proof of Theorem 2.16

It can be shown that X_A and X_B are conditionally independent given X_C if for all $x \in I\!R^p$ with $F_C(x_C) \neq 0$ it holds that

$$P(X_A \leq x_A, X_B \leq x_B | X_C \leq x_C) = P(X_A \leq x_A | X_C \leq x_C) P(X_B \leq x_B | X_C \leq x_C).$$

According to the definition of conditional probabilities the above equation can be written as

$$\frac{P(X_A \leq x_A, X_B \leq x_B, X_C \leq x_C)}{P(X_C \leq x_C)} = \frac{P(X_A \leq x_A, X_C \leq x_C)}{P(X_C \leq x_C)} \frac{P(X_B \leq x_B, X_C \leq x_C)}{P(X_C \leq x_C)}$$

which is equivalent to

$$F_{A \cup B \cup C}(x_A, x_B, x_C) F_C(x_C) = F_{A \cup C}(x_A, x_C) F_{B \cup C}(x_B, x_C),$$

and to

$$\frac{F_{A \cup C}(x_A, x_C) F_{B \cup C}(x_B, x_C)}{F_{A \cup B \cup C}(x_A, x_B, x_C) F_C(x_C)} = 1. \tag{4.5}$$

The terms of this equation are given as

$$\begin{aligned} F(x) = & \prod_{i \in A} F_i(x_i) \prod_{i \in B} F_i(x_i) \prod_{i \in C} F_i(x_i) \\ & \prod_{I \in \mathcal{I}_{\mathcal{A}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{B}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{C}}} c_I(x)^{-\alpha_I} \\ & \prod_{I \in \mathcal{I}_{\mathcal{AB}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{AC}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{BC}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{ABC}}} c_I(x)^{-\alpha_I}, \end{aligned} \tag{4.6}$$

$$\begin{aligned} F_{A \cup C}(x_A, x_C) = & \prod_{i \in A} F_i(x_i) \prod_{i \in C} F_i(x_i) \prod_{I \in \mathcal{I}_{\mathcal{A}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{C}}} c_I(x)^{-\alpha_I} \\ & \prod_{I \in \mathcal{I}_{\mathcal{AC}}} c_I(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{AB}_{AB}}} c_{I \cap A}(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{AB}_{A.}}} c_{I \cap A}(x)^{-\alpha_I} \\ & \prod_{I \in \mathcal{I}_{\mathcal{BC}_{BC}}} c_{I \cap C}(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{BC}_{.C}}} c_{I \cap C}(x)^{-\alpha_I} \\ & \prod_{I \in \mathcal{I}_{\mathcal{ABC}_{ABC}}} c_{I \cap (A \cup C)}(x)^{-\alpha_I} \prod_{I \in \mathcal{I}_{\mathcal{ABC}_{A.C}}} c_{I \cap (A \cup C)}(x)^{-\alpha_I}, \end{aligned} \tag{4.7}$$

$$F_C(x_C) = \prod_{i \in C} F_i(x_i) \prod_{I \in \mathcal{I}_{\mathcal{C}}} c_I(x)^{-\alpha_I} \tag{4.8}$$

$$\prod_{I\in\mathcal{I}_{\mathcal{AC}_{AC}}} c_{I\cap C}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{AC}_{.C}}} c_{I\cap C}(x)^{-\alpha_I}$$
$$\prod_{I\in\mathcal{I}_{\mathcal{BC}_{BC}}} c_{I\cap C}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{BC}_{.C}}} c_{I\cap C}(x)^{-\alpha_I}$$
$$\prod_{I\in\mathcal{I}_{\mathcal{ABC}_{ABC}}} c_{I\cap C}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{A.C}}} c_{I\cap C}(x)^{-\alpha_I}$$
$$\prod_{I\in\mathcal{I}_{\mathcal{ABC}_{.BC}}} c_{I\cap C}(x)^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{..C}}} c_{I\cap C}(x)^{-\alpha_I}$$

and $F_{B\cup C}(x_B, x_C)$ according to $F_{A\cup C}(x_A, x_C)$. Equation (4.5) then reads as

$$\begin{aligned} 1 \quad = \quad & \prod_{I\in\mathcal{I}_{\mathcal{AB}_{AB}}} \left\{\frac{c_{I\cap A}(x)c_{I\cap B}(x)}{c_I(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{AB}_{A.}}} \left\{\frac{c_{I\cap A}(x)}{c_I(x)}\right\}^{-\alpha_I} \\ & \prod_{I\in\mathcal{I}_{\mathcal{AB}_{.B}}} \left\{\frac{c_{I\cap B}(x)}{c_I(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{AB}_{..}}} \left\{\frac{1}{c_I(x)}\right\}^{-\alpha_I} \\ & \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{ABC}}} \left\{\frac{c_{I\cap(A\cup C)}(x)c_{I\cap(B\cup C)}(x)}{c_I(x)c_{I\cap C}(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{A.C}}} \left\{\frac{c_{I\cap(A\cup C)}(x)}{c_I(x)c_{I\cap C}(x)}\right\}^{-\alpha_I} \\ & \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{.BC}}} \left\{\frac{c_{I\cap(B\cup C)}(x)}{c_I(x)c_{I\cap C}(x)}\right\}^{-\alpha_I} \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{..C}}} \left\{\frac{1}{c_I(x)c_{I\cap C}(x)}\right\}^{-\alpha_I} \\ & \prod_{I\in\mathcal{I}_{\mathcal{ABC}_{...}}} \left\{\frac{1}{c_I(x)}\right\}^{-\alpha_I}. \end{aligned} \tag{4.9}$$

It is known that

$$\begin{aligned} \mathcal{AB}\cup\mathcal{ABC} &= \{I\in\mathcal{I} | I\cap A\neq\emptyset, I\cap B\neq\emptyset\} \\ &= \{I\in\mathcal{I} | \,\exists\, i\in A, j\in B \text{ with } i\in I \text{ and } j\in I\}. \end{aligned}$$

As a consequence, let in Equation (4.9) be $\alpha_I = 0$ for all $I \in \mathcal{AB}\cup\mathcal{ABC}$. Then, the right hand side equals 1 and part (i) of the theorem is shown.

For the other part of the theorem it has to be shown that the conditional independence of X_A and X_B given X_C implies $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i \in I$ and $j \in I$ for all $i \in A$ and $j \in B$.

X_A and X_B are conditionally independent given X_C if for all $x_C \in I\!R^{|X_C|}$ with $f(x_C) > 0$ it holds that

$$F_{A\cup B|C}(x_A, x_B|x_C) = F_{A|C}(x_A|x_C)F_{B|C}(x_B|x_C). \tag{4.10}$$

The cdf of the conditional distribution of X_A given $X_C = x_C$ is defined as

$$\begin{aligned} F_{A|C}(x_A|x_C) &= \lim_{h\searrow 0} P(X_A \le x_A \mid x_C - h < X_C \le x_C) \\ &= \lim_{h\searrow 0} \frac{P(X_A \le x_A, x_C - h < X_C \le x_C)}{P(x_C - h < X_C \le x_C)} \end{aligned}$$

$$= \frac{\lim_{h\searrow 0}\frac{1}{h}P(X_A \le x_A, x_C - h < X_C \le x_C)}{\lim_{h\searrow 0}\frac{1}{h}P(x_C - h < Z \le x_C)}$$

$$= \frac{\int_{-\infty}^{x_A} f_{A\cup C}(t, x_C)dt}{f_C(x_C)}.$$

Note that $F_{B|C}$ and $F_{A\cup B|C}$ can be written analogously. Since all involved limits exist we get from Equation (4.10)

$$\lim_{h\searrow 0} P(X_A \le x_A \mid x_C - h < X_C \le x_C) \lim_{h\searrow 0} P(X_B \le x_B \mid x_C - h < X_C \le x_C)$$
$$= \lim_{h\searrow 0} P(X_A \le x_A, X_B \le x_B \mid x_C - h < X_C \le x_C)$$

which is equivalent to

$$\frac{\lim_{h\searrow 0}\frac{1}{h}P(X_A \le x_A, x_C - h < X_C \le x_C)}{\lim_{h\searrow 0}\frac{1}{h}P(x_C - h < X_C \le x_C)} \frac{\lim_{h\searrow 0}\frac{1}{h}P(X_B \le x_B, x_C - h < X_C \le x_C)}{\lim_{h\searrow 0}\frac{1}{h}P(x_C - h < X_C \le x_C)}$$
$$= \frac{\lim_{h\searrow 0}\frac{1}{h}P(X_A \le x_A, X_B \le x_B, x_C - h < X_C \le x_C)}{\lim_{h\searrow 0}\frac{1}{h}P(x_C - h < X_C \le x_C)}.$$

Simple algebraic transformations give

$$\lim_{h\searrow 0}\{P(X_A \le x_A, x_C - h < X_C \le x_C)P(X_B \le x_B, x_C - h < X_C \le x_C)$$
$$-P(X_A \le x_A, X_B \le x_B, x_C - h < X_C \le x_C)P(x_C - h < X_C \le x_C)\} = 0.$$

If all terms of the type $P(X_A \le x_A, x_C - h < X_C \le x_C)$ are replaced by $P(X_A \le x_A, X_C \le x_C) - P(X_A \le x_A, X_C \le x_C - h)$ the known expressions for the cdf's can be inserted. Consider now the limit of the above equation for $x_i \to \infty$ for all $i \in V \setminus \{j_A, j_B\}$ and $x_{j_A} = \tilde{x}_{j_A}$, $x_{j_B} = \tilde{x}_{j_B}$. The same argumentation as in the proof of Lemma 2.11 yields that Equation (4.10) is fulfilled for all $x \in I\!R^p$ if and only if for all $i \in A$ and $j \in B$ it holds that $\alpha_I = 0$ for all $I \in \mathcal{I}$ with $i \in I$ and $j \in I$. $\square$

References

Barndorff–Nielsen, O.E. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley and Sons, New York.

Caputo, A. (1998). Eine alternative Familie von Modellverteilungen für Kovarianz- and Konzentrationsgraphen. *Ph.D.Thesis*, Institute of Statistics, University of Munich.

Cook, R.D. and Johnson, M.E. (1981). A family of distributions for modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Association B*, 43, 210-218.

Cook, R.D. and Johnson, M.E. (1986). Generalized Burr-Pareto-logistic distributions with application to a uranium exploration data set. *Technometrics*, 28, 123-131.

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.

Johnson, M.E. (1987). *Multivariate Statistical Simulation.* John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley and Sons, New York.

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.