Записки научных
семинаров ПОМИ
Том 501, 2021 г.

N. Gauraha, D. von Rosen

TOTALLY ORDERED CONDITIONAL INDEPENDENCE MODELS

Abstract. A subclass of lattice conditional independence models is introduced. The new class of models is called totally ordered independence models. The class is based on an assumption that the index set which orders the random variables is a chain. It is shown that there is a jump in the chain if and only if there is a conditional independence relation. Some comparisons between the lattice conditional independence models and totally ordered independence models are presented.

§1. Introduction

In this article a class of multivariate normal models determined by so called totally ordered conditional independence (TOCI) restrictions on the covariance structure in normally distributed random vectors is introduced and studied. The class of models constitute a special class of the lattice conditional independence class of models (LCI) introduced by [3]. Considered is the set of subsets of a finite index set I which is totally ordered ($\equiv$ chain) by inclusion (precisely defined in Section 3). Given a non-decreasing chain, say $\mathcal{K}$, of indices $K_1 \subset \cdots \subset K_q$, where K_i denotes a subset of indices, the conditional independence relations in this article, i.e., the structure of the covariance matrix and factorization of the likelihood, are derived by considering adjacent pairs, K_i, K_{i+1}, of the elements of the chain $\mathcal{K}$.

The totally ordered conditional independence model $N(\mathcal{K}, \Sigma)$ is defined to be the class of all normal distributions on, say, $\mathbb{R}^p$ such that for the adjacent pair $\{K_i, K_{i+1}\} \in \mathcal{K}$, $i \in \{1, \dots, q-1\}$, the components of the multivariate normal vector $X \in \mathbb{R}^p$, indexed by the set difference $K_{i+1} \backslash K_i$, are mutually conditionally independent of the elements of X which are indexed by K_i. It will be assumed that the mean of the distribution, without

Key words and phrases: multivariate regression, lattice conditional independence models, totally ordered conditional independence, totally ordered multivariate linear models.

Dietrich von Rosen is supported by the Swedish Research Council (2017-03003).

loss of generality, equals 0 and the unknown non-singular covariance matrix is denoted Σ.

The reader is now further introduced to the TOCI model class through some basic examples:

Examples: Let $X = (X_1, X_2, X_3)^\top$ follow a multivariate normal distribution of dimension 3 with mean zero and non-singular covariance matrix Σ: 3×3, and $^\top$ denotes the transpose of a matrix. Let $I = \{1,\ 2,\ 3\}$ and let $\mathbb{P}(I)$ be the power set of I. Define a set $\mathcal{K} \subset \mathbb{P}(I)$ which is totally ordered by inclusion. We assume that $\mathcal{K}$ always contains the empty set $\{\varnothing\}$ and the index set I. For example, the trivial totally ordered chain is given by $\mathcal{K} = \{\{\varnothing\},\ I\}$. The difference $I \setminus \{\varnothing\}$ is the set I itself, therefore, since we are conditioning with respect to $\{\varnothing\}$, conditional independence means that for the trivially ordered set the three random variables corresponding to I are marginally independent which is denoted $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3$.

Another ordered set, say chain $\mathcal{K}$, defined on the same index set as in the previous paragraph, can be given by $\mathcal{K}_1 \subset K_2 \subset K_3$, e.g., $\{\varnothing\} \subset \{1\} \subset \{I\}$. The conditional independence structure implied by the first adjacent pair yields $K_2 \setminus K_1 = \{1\}$. Consider the second adjacent pair, then $K_3 \setminus K_2$ equals $\{1, 2, 3\} \setminus \{1\} = \{2, 3\}$ and therefore $(X_2 \perp\!\!\!\perp X_3) \mid X_1$, which denotes that X_2 and X_3 are independently distributed given X_1. Since we are working with multivariate normally distributed variables with a covariance matrix Σ, the conditional independence assumption among the random variables is equivalent to the well known covariance conditions $(\Sigma^{-1})_{23} = (\Sigma^{-1})_{32} = 0$ (using standard notation for block partitions of matrices). In this example the factorization of the parameter space and the likelihood function are given as follows:

$$\Sigma \equiv \{\Sigma_{11}, B_2, \Sigma_{22\bullet 1}, B_3, \Sigma_{33\bullet 1}\}, \quad B_2 = \Sigma_{21}\Sigma_{11}^{-1},\ B_3 = \Sigma_{31}\Sigma_{11}^{-1}, \tag{1}$$

$$f_x \propto f_{X_1} f_{X_2|X_1} f_{X_3|X_1}, \tag{2}$$

where $f_\bullet$s denote the multivariate normal density functions, $A \equiv B$ means that A can be expressed with the help of B (and some inserted 0), and the Schur complements $\Sigma_{22\bullet 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $\Sigma_{33\bullet 1} = \Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}$ are the covariance matrices in $f_{X_2|X_1}$ and $f_{X_3|X_1}$, respectively, meaning that the conditional independence model is determined by the totally ordered set $\mathcal{K} = \{\{\varnothing\},\ \{1\},\ I\}$. Moreover, (1) and (2) imply that the maximum likelihood estimators of the parameters in Σ given in (1) can be obtained from the conditional density functions in (2), using standard techniques from multivariate regression analysis (e.g., see [4]). From f_{X_1}

the variance Σ_{11} is estimated, from $f_{X_1|X_2}$ the conditional mean effect B_2 and the conditional variance $\Sigma_{22\bullet 1}$ are estimated and from $f_{X_3|X_1}$ the conditional mean effect B_3 and the conditional variance $\Sigma_{33\bullet 1}$ are estimated. These estimators build up the estimator of Σ, i.e.,

$$\Sigma_{11} = \Sigma_{11}, \quad \Sigma_{21} = B_2\Sigma_{11}, \quad \Sigma_{22} = \Sigma_{22\bullet 1} + \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$
$$\Sigma_{31} = B_3\Sigma_{11}, \quad \Sigma_{33} = \Sigma_{33\bullet 1} + \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}, \quad \Sigma_{23} = \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}.$$

If the normally distributed vector would have had a mean which differs from 0 it would be possible to express the mean structure with similar relations to those above.

The proposed normal TOCI models class can be viewed as a natural special case of the totally ordered multivariate linear models defined in [2]. It has been shown that general totally ordered multivariate linear models are amenable to explicit (non-iterative) likelihood analysis. We will investigate the relationship between the TOCI models and the lattice conditional independence (LCI) models, see [3], and it will be shown that the class of TOCI models is included in the class of LCI models.

It follows, as in the second example given above, that for all TOCI models the likelihood function and parameter space can be factored into the products of conditional likelihood functions and disjoint parameter spaces, respectively, and each conditional likelihood function, corresponds to an ordinary multivariate normal regression model from where explicit maximum likelihood estimators can be obtained. The main advantages with formulating TOCI models is that there is a very clear interpretation of the models which makes the model suitable for inference and straightforward model validations, a basic ingredient of the statistical paradigm.

It can be noted that there exists a lot of work on so called Graphical Markov models (ADG-models) and one article which is close to this work is [5] where also a literature review of ADG models is given. In [5] it is shown that normal linear ADG models among others include totally ordered normal linear models and lattice conditional independence models. The present work differs from the above mentioned models by supposing a chain structure in the model formulation with a focus on expressing conditional independence relations. In the considered models that explicit maximum likelihood estimators (non-iterative) can be obtained but looking very explicitely on the models one gets a better understanding for the model class and thereby model validation techniques can be developed which however will not take place in this article.

The organization of the paper is as follows. In Section 2, background concepts are introduced, and notation regarding ordered sets and lattices. In Section 3, the TOCI model class is defined, and in Section 4 the class of models is compared with LCI models, i.e., transitive directed acyclic graphical models (TDAG). Moreover, in Section 4 the notion of Markov equivalence among TOCI models is also discussed. A summary of the paper is provided in Section 5.

§2. Background

In this section some notations are introduced and a brief overview of ordered sets and lattices are presented. For more details on ordered set and lattices it is referred to [6].

2.1. Partially ordered and totally ordered sets. Throughout the article I is a finite index set. Let $\mathbb{P}(I)$ be the power set of the index set, I, which contains the set of all subsets of I, including the empty set and I itself. For example for a finite set $I = \{1, 2, 3\}$, its power set is given by $\mathbb{P}(I) = \{\{\varnothing\},\ \{1\},\ \{2\},\ \{3\},\ \{1,2\},\ \ \{1,3\},\ \{2,3\},\ \{1,2,3\}\}$. Order theoretic properties of a power set can be expressed in terms of the subset relation $\subseteq$. For example for all $A, B, C \subseteq \mathbb{P}(I)$ we have:

$$Reflexivity: \qquad A \subseteq A; \tag{3}$$

$$Antisymmetry: \qquad A \subseteq B \text{ and } B \subseteq A \implies A = B; \tag{4}$$

$$Transitivity: \qquad A \subseteq B \text{ and } B \subseteq C \implies A \subseteq C; \tag{5}$$

$$Comparability: \qquad A \subseteq B \text{ or } B \subseteq A. \tag{6}$$

In the following the partially ordered sets (poset) and totally ordered sets (chain) by inclusion ($\subset$) are defined. For illustration purpose we consider a finite set $I = \{1, 2, 3, 4\}$.

Definition 2.1. Relations satisfying (3) – (5) are called partial ordering relations, and the sets coupled with such relations are called partially ordered sets or posets.

For example, Let $\mathcal{L} = \{\{\varnothing\}, \{1\}, \{1, 2\}, \{1, 3\}, I\}$, then $(\mathcal{L}, \subset)$ is a poset.

Definition 2.2. A poset $(\mathcal{L}, \subset)$ that also satisfies (6) is called a totally ordered set or a chain.

For example, let $\mathcal{K} = \{\{\varnothing\}, \{1\}, \{1, 3\}, I\}$, then $(\mathcal{K}, \subset)$ is a chain.

Definition 2.3 (Lattice)**.** A poset $(\mathcal{L}, \subset)$ is a lattice if $\{A \cup B\}$ (least upper bound) and $\{A \cap B\}$ (greatest lower bound) is included in $\mathcal{L}$ for all $A, B \in \mathcal{L}$. The infimum $\cap$ and supremum $\cup$ can be characterized by the following set operations:

$$A \subset B \iff A \cap B = A,$$
$$A \subset B \iff A \cup B = B.$$

Definition 2.4 (Distributive lattice)**.** A lattice $(\mathcal{L}, \subset)$ is distributive if the following additional property holds for all $A, B, C \in \mathcal{L}$: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

For example, let $\mathcal{L} = \{\{\varnothing\}, \{1\}, \{1, 2\}, \{1, 3\}, \{1, 2, 3\}, I\}$, then $(\mathcal{L}, \subset)$ is a distributive lattice.

In linear models theory often linear spaces are involved and linear spaces satisfy the modular equality $A \cap (B \cup (A \cap C)) = (A \cap B) \cup (A \cap C)$, i.e., constitute a modular lattice, where A, B and C are linear subspaces. However, this algebraic property does not lead to unique decompositions of the likelihoods and therefore it is of advantage to consider more restricted structures. An example is a two-way unbalanced analysis of variance model which can be decomposed in several ways whereas a balanced two-way analysis of variance model has one unique decomposition. A distributive lattice is a modular lattice.

Proposition 2.1 $(\mathcal{K}(I) \subset \mathcal{L}(I))$**.** *Let $\mathcal{K}(I)$ be the set of all chains of the set $\mathbb{P}(I)$. Moreover, let $\mathcal{L}(I)$ be the set of all distributive lattices of $\mathbb{P}(I)$, then $\mathcal{K}(I) \subset \mathcal{L}(I)$.*

The proposition implies that every chain is a distributive lattice.

Definition 2.5 (Ascending chain)**.** A chain is called well numbered (ordered) chain or an ascending chain if its elements are ordered in an ascending way.

For example, let $\mathcal{K} = \{\{\varnothing\}, \{1\}, \{1, 2\}, I\}$ be a chain by inclusion which is well ordered. For the rest of the article, by a chain $\mathcal{K} = \{K_1, K_2, \ldots, K_{q+1}\}$, we mean a well ordered chain such that $K_1 \subset K_2 \subset \cdots \subset K_{q+1}$.

§3. Totally ordered conditional independence models

The class of multivariate normal models determined by totally ordered conditional independence (TOCI) restrictions on the covariance matrix will be introduced.

Let $I = \{1, ..., p\}$ be a finite index set, let $\mathbb{P}(I)$ denote the power set of I, and let $\mathcal{K} \subset \mathbb{P}(I)$ be a well-ordered chain. Throughout this article we shall consider a p-variate normally distributed random vector $X = (X_1, \ldots, X_p)^T$ with mean zero and a positive definite covariance matrix Σ. For convenience, we refer to a random variable, or a set of them, by their indexes, i.e., X_i as i and X_A as A. Let $M(I)$ denotes the set of all $p \times p$ positive definite matrices, and thus $\Sigma \in M(I)$. The TOCI model $N(\mathcal{K}; \Sigma)$ is defined to be the set of all normal distributions with respect to a totally ordered set $\mathcal{K}$, such that for every adjacent pair $K_i, K_{i+1} \in \mathcal{K}$, the elements of the set $K_{i+1} \setminus K_i$ are mutually conditionally independent given K_i. This is formally rewritten in the next definition.

Definition 3.1 (TOCI model and $N(\mathcal{K}; \Sigma)$)**.** Let $\mathcal{K} = \{K_1 \subset \cdots \subset K_{q+1}\}$ be a totally ordered set, where $|K| = q + 1$. The family of normal distributions $N(\mathcal{K}; \Sigma)$ is said to satisfy the totally ordered conditional independence property with respect to $\mathcal{K}$, i.e., being a TOCI model, if, for each adjacent pair of elements $K_i, K_{i+1} \in \mathcal{K}$ the following relations hold: for $i \in \{1, \ldots, q\}$ and vertices $v_1, \ldots, v_r$

$$\{v_1, \ldots, v_r\} = K_{i+1} \setminus K_i, \quad (v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp v_r) \mid K_i. \tag{7}$$

In the definition $(v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp v_r) \mid K_i$ means that the random variables which correspond to $(v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp v_r)$ are conditionally independent given the random variables which correspond to K_i.

Definition 3.2 (The difference set $D(\mathcal{K})$)**.** Let $\mathcal{K} = \{K_1 \subset \cdots \subset K_{q+1}\}$ be a chain, where the cardinality equals $|\mathcal{K}| = q + 1$. The difference set $D(\mathcal{K})$ of the chain $\mathcal{K}$ is defined as follows:

$$D(\mathcal{K}) = \{K_{i+1} \setminus K_i; \quad K_i, K_{i+1} \in \mathcal{K}, \forall i \in \{1, \ldots, q\}\}$$

and put $\{D_i\} = \{K_{i+1} \setminus K_i\}$.

The difference set $D(\mathcal{K}) = \{D_1, \ldots, D_q\}$, as defined in Definition 3.2, is an ordered partition of the index set I, and $\mathcal{K}$ is uniquely determined by $D(\mathcal{K})$. In fact for any $D(\mathcal{K})$ of the chain $\mathcal{K}$ the following holds: For $i \in \{1, \ldots, q\}$

$$K_{i+1} = D_1 \cup \cdots \cup D_i,$$

where $K_{q+1} = I = \bigcup(D_i \in D)$. This means that conditional independence of a chain $\mathcal{K}$ can be represented, more conveniently, in terms of the corresponding $D(\mathcal{K})$.

Proposition 3.1. *Let $N(\mathcal{K};\Sigma)$ be a TOCI model, and let the difference set be $D(\mathcal{K})$. The elements $D_i \in D$ are mutually conditionally independent given the corresponding subset $K_i \in \mathcal{K}$.*

A chain and its difference set, and conditional independence are illustrated in Table 1.

Table 1. An example of a chain $\mathcal{K} = \{\{\varnothing\}, \{1,2\}, I = \{1,2,3,4\}\}$, its difference set $D(\mathcal{K})$ and conditional independence relations (CIs)

	1	2	3
$\mathcal{K}$	$\varnothing$	$\{1,2\}$	I
$D(\mathcal{K})$	$\{1,2\}$	$\{3,4\}$	
CIs	$(1 \perp\!\!\!\perp 2)$	$(3 \perp\!\!\!\perp 4) \mid (1,2)$	

Theorem 3.1 (Factorization of the likelihood function)**.** *Let $N(\mathcal{K},\Sigma)$ be a TOCI model, with the difference set $D(\mathcal{K})$, $|\mathcal{K}| = q+1$. The likelihood function of the model can be factorized in conditional likelihood functions in terms of the elements of $D(\mathcal{K})$ as follows:*

$$f_I \propto \prod_{i=1}^{q}(f_{d\in D_i|K_i}).$$

Thus Theorem 3.1 shows how important the difference set is for performing statistical inference. However, in order to fully understand the conditioning we also have to see what happens with the parameter matrix Σ.

Theorem 3.2 (Factorization of parameter space)**.** *Let $N(\mathcal{K},\Sigma)$ be a TOCI model, with a difference set $D(\mathcal{K})$. The unknown parameter Σ can be reconstructed from its factors:*

$$\Sigma \equiv \{\Sigma_{\{d\in D_i\}\times K_i}\Sigma_{K_i}^{-1}, \Sigma_{\{d\in D_i\}\bullet K_i} \mid K_i \in \mathcal{K}, D_i \in D\},$$

where $\Sigma_A = \Sigma_{A\times A}$, $\Sigma_{A\times B}$ represents the $A\times B$ sub-matrix of Σ, and $\Sigma_{A\bullet B} = \Sigma_A - \Sigma_{A\times B}\Sigma_B^{-1}\Sigma_{B\times A}$.

For the proofs of Theorem 3.1 and Theorem 3.2 it is possible to take over a proof given in [3] because the class of TOCI models is a subclass of the lattice conditional independence class of models (LCI). The relation between TOCI models and LCI models is briefly discussed in Section 4.

The advantage with the TOCI models is that there is a clear order of the elements in the index set which then generates the difference set. To make use of the difference set it is of interest to define the concept of a jump.

Definition 3.3 (Jump)**.** Let $\mathcal{K} = \{K_1 \subset \cdots \subset K_{q+1}\}$ be a totally ordered set, where $|\mathcal{K}| = q+1$. Let $D(\mathcal{K})$ be the difference set of $\mathcal{K}$. The chain $\mathcal{K}$ is said to have a jump, if for $i \in \{1, \ldots, q-1\}$ at least one inequality $|D_i| \geq 2$ $(|K_{i+1}| \geq 2 + |K_i|)$ is satisfied.

A jump will always correspond to either an independence relation or a conditional independence relation. If the jump includes $\{\varnothing\}$ there is independence but otherwise the jump indicates conditional independence. The size of the jump tells how many variables are involved. If the jump size is t then there are $t-1$ independence relations (strict independecnce or conditional independence). The difference set immediately tells us how the corresponding variables are related which in turn defines the covariance structure in the TOCI model.

We will illustrate the results of Theorem 3.1 and Theorem 3.2 in detail using a series of examples with four variables. Let $X_1,\ X_2,\ X_3$ and X_4 follow a multivariate normal distribution with mean zero and non-singular covariance matrix Σ, that is $X = (X_1, X_2, X_3, X_4)^T \sim N_4(0, \Sigma)$, $\Sigma \in M(I)$. Here, $I = \{1,\ 2,\ 3,\ 4\}$ and let $\mathbb{P}(I)$ denote the power set of I. Now all possible non-isomorphic chains for the index set I are generated, and for each chain the difference set and the conditional independence relations implied by the difference set are derived. Consider a totally ordered set (chain) $\mathcal{K}_I \subset \mathbb{P}(I)$ where

$$\mathcal{K}_I = \{K_1, K_2, K_3, K_4\} = \{\{\varnothing\},\ \{1\},\ \{1,2\},\ \{1,2,3\},\ I\}. \qquad (8)$$

As there can be only one such non-isomorphic chain, the remaining non-isomorphic chains can be derived from this chain, which are subsets of $\mathcal{K}_I$. The set $\{\varnothing\}$ and I are included in all chains and for the remaining subsets $\{1\}$, $\{1,2\}$ and $\{1,2,3\}$, we have two choices, to include or not to include a set into the chain. Thus there are $2^{|I|-1} = 2^3 = 8$ total non-isomorphic chains connected to the index set I. The eight non-isomorphic chains for the index set I are now discussed one by one and focus is on conditional independence relations. Let $f_\bullet$ denote the normal density function and elements in the covariance matrix are indexed according to a mixture of the notation in Theorem 3.2 and standard notation which we hope will not cause any confusion.

(1) $\mathcal{K}_1 = \{\{\varnothing\},\ I\}$: The difference set is $D(\mathcal{K}_1) = I$. There is one jump of size four and since $\{\varnothing\}$ is included it means that there are three independence relations:

$$\begin{aligned} CI &: X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 \perp\!\!\!\perp X_4, \\ f_I &\propto f_1 f_2 f_3 f_4, \\ \Sigma &\equiv \{\Sigma_{11},\ \Sigma_{22},\ \Sigma_{33}, \Sigma_{44}\}. \end{aligned}$$

(2) $\mathcal{K}_2 = \{\{\varnothing\}, \{1\}, I\}$: The difference set equals $D(\mathcal{K}_2) = \{\{1\}, \{2,3,4\}\}$. There is one jump between $\{1\}$ and I which thus is of size three, generating two conditional independence relations. More precisely, the elements of the set $\{2,3,4\}$ are mutually conditionally independent conditioned on the corresponding $K \in \mathcal{K}$, which is $\{1\}$. The implied conditionally independence relations by the chain $\mathcal{K}_2$ are given as follows:

$$\begin{aligned} CI &: (X_2 \perp\!\!\!\perp X_3 \perp\!\!\!\perp X_4) \mid X_1, \\ f_I &\propto f_1 f_{2|1} f_{3|1} f_{4|1}, \\ \Sigma &\equiv \{\Sigma_{11}, \Sigma_{21}\Sigma_{11}^{-1}, \Sigma_{22\bullet 1}, \Sigma_{31}\Sigma_{11}^{-1}, \Sigma_{33\bullet 1}, \Sigma_{41}\Sigma_{11}^{-1}, \Sigma_{44\bullet 1}\}. \end{aligned}$$

(3) $\mathcal{K}_3 = \{\{\varnothing\},\ \{1,2\},\ \ I\}$: The difference set equals $D(\mathcal{K}_3) = \{\{1,2\},\ \{3,4\}\}$. There are two different jumps. The first one which involves $\{\varnothing\}$ is of size two and the jump between $\{1,2\}$ and I is also of size two. Therefore, the elements of the set $\{1,2\}$ are marginally independent as the "previous" set is the empty set $\{\varnothing\}$. The elements of the set $\{3,4\}$ are mutually conditionally independent conditioned on the union of all previous subsets which is $\{1,2\}$. The following are the implied conditional independency relations:

$$\begin{aligned} &CI : X_1 \perp\!\!\!\perp X_2,\ (X_3 \perp\!\!\!\perp X_4) \mid (X_1, X_2), \\ &f_I \propto f_1 f_2 f_{3|\{1,2\}} f_{4|\{1,2\}}, \\ &\Sigma \equiv \big\{\Sigma_{11}, \Sigma_{22}, \Sigma_{3\times\{1,2\})}\Sigma_{\{1,2\}}^{-1}, \Sigma_{33\bullet\{1,2\}}, \\ &\qquad\qquad \Sigma_{4\times\{1,2\}}\Sigma_{\{1,2\}}^{-1}, \Sigma_{44\bullet\{1,2\}}\big\}. \end{aligned}$$

(4) $\mathcal{K}_4 = \{\{\varnothing\},\ \{1\},\ \{1,2\},\ \ I\}$: The difference set is given by $D(\mathcal{K}_4) = \{\{1\},\ \{2\}, \{3,4\}\}$. There is one jump of size 2 between

$\{1,2\}$ and I. Thus,

$$CI : (X_3 \perp\!\!\!\perp X_4) \mid (X_1, X_2),$$
$$f_I \propto f_1 f_{2|1} f_{3|\{1,2\}} f_{4|\{1,2\}},$$
$$\Sigma \equiv \{\Sigma_{11}, \Sigma_{21}\Sigma_{11}^{-1}, \Sigma_{22\bullet 1}, \Sigma_{3\times\{1,2\}}\Sigma_{\{1,2\}}^{-1}, \Sigma_{33\bullet\{1,2\}}, \Sigma_{4\times\{1,2\}}\Sigma_{\{1,2\}}^{-1}, \Sigma_{44\bullet\{1,2\}}\}.$$

(5) $\mathcal{K}_5 = \{\{\varnothing\}, \{1,2,3\}, I\}$: The difference set $D(\mathcal{K}_5) = \{\{1,2,3\}, \{4\}\}$ appears. Since there is one jump of size three involving $\{\varnothing\}$ there are two independence relations. Thus, the implied conditional independence relations by the chain $\mathcal{K}_5$ are given by

$$CI : (X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3),$$
$$f_I \propto f_1 f_2 f_3 f_{4|\{1,2,3\}},$$
$$\Sigma \equiv \{\Sigma_{11}, \Sigma_{22}, \Sigma_{33}, \Sigma_{4\times\{1,2,3\}}\Sigma_{\{1,2,3\}}^{-1}, \Sigma_{44\bullet\{1,2,3\}}\}.$$

(6) $\mathcal{K}_6 = \{\{\varnothing\}, \{1\}, \{1,2,3\}, I\}$: The difference set is given by $D(\mathcal{K}_6) = \{\{1\}, \{2,3\}, \{4\}\}$. There is one jump between $\{1\}$ and $\{1,2,3\}$ which thus is of size two. From the chain $\mathcal{K}_6$ the following are the only implied conditional independence relations:

$$CI : (X_2 \perp\!\!\!\perp X_3) \mid X_1,$$
$$f_I \propto f_1 f_{2|1} f_{3|1} f_{4|\{1,2,3\}},$$
$$\Sigma \equiv \{\Sigma_{11}, \Sigma_{21}\Sigma_{11}^{-1}, \Sigma_{22\bullet 1}, \Sigma_{31}\Sigma_{11}^{-1}, \Sigma_{33\bullet 1}, \Sigma_{4\times\{1,2,3\}}\Sigma_{\{1,2,3\}}^{-1}, \Sigma_{44\bullet\{1,2,3\}}\}.$$

(7) $\mathcal{K}_7 = \{\{\varnothing\}, \{1,2\}, \{1,2,3\}, I\}$: The difference set is constructed as $D(\mathcal{K}_7) = \{\{1,2\}, \{3\}, \{4\}\}$. There is a jump between $\{\varnothing\}$ and $\{1,2\}$ describing one independence relation. Thus, according to $\mathcal{K}_7$ the only implied conditional independence (independence) relation is given by

$$CI : X_1 \perp\!\!\!\perp X_2,$$
$$f_I \propto f_1 f_2 f_{3|\{1,2\}} f_{4|\{1,2,3\}},$$
$$\Sigma \equiv \{\Sigma_{11}, \Sigma_{22}, \Sigma_{3\times\{1,2\}}\Sigma_{\{1,2\}}^{-1}, \Sigma_{33\bullet\{1,2\}}, \Sigma_{4\times\{1,2,3\}}\Sigma_{\{1,2,3\}}^{-1}, \Sigma_{44\bullet\{1,2,3\}}\}.$$

(8) $\mathcal{K}_8 = \{\{\varnothing\}, \{1\}, \{1,2\}, \{1,2,3\}, I\}$: The difference set is given by $D(\mathcal{K}_8) = \{\{1\}, \{2\}, \{3\}, \{4\}\}$. There are no jumps of size

larger than 1 and thus no conditional independence/independece relations implied by the chain $\mathcal{K}_8$, and since $\Sigma \in \mathbf{M}(I)$ is unrestricted and therefore no "natural" unique factorization can take place.

§4. Some properties of the TOCI class of models and its relation to LCI models

The following graph-theoretical notations and concepts will be used. A directed acyclic graph (DAG) D, is a pair $D = (V, E)$, where V is the set of vertices and E is the set of directed edges between certain pairs of distinct vertices such that no cycles are present. If we have a directed edge $(a \to b)$, we say that a is a parent of b, The set of all parents of a vertex b is denoted by $pa(b)$. For the edge from $(a \to b)$ and $(c \to b)$, but without an edge between a and c (i.e., $(a \to c)$ or $(c \to a)$) then a and c are immoral parents of b. A DAG is transitive (TDAG) if $(a \to b)$ and $(b \to c)$ imply that there exists a direct edge $(a \to c)$. A subset $A \in V$ is called ancestors of a vertex b when there is a direct path from each vertex $a \in A$ to b.

A graphical model is a set of distributions satisfying a set of conditional independence relations which usually are presented via a graph [8]. The vertices in the graph correspond to random variables. Absence of an edge between two vertices implies that the corresponding two random variables are interpreted to be conditionally independent. For example, the probability density function of the graphical model given by Figure 1 can be factorized as $f_x \propto f_{X_1} f_{X_2|X_1} f_{X_3|X_1}$, meaning that given X_1 the variables X_2 and X_3 are conditionally independent. For more details on graphical models it is referred to [7].

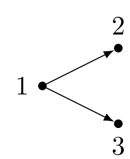

Figure 1. An example of a three-variable graphical model encoded by a graph.

In the following the lattice conditional independence (LCI) property is defined. For more details on LCI and its equality to transitive directed acyclic graphs (TDAGs) see [1]. The next definition should be compared with Definition 3.1. The normal distribution in the definition has once again a mean which equals 0 but this time the covariance structure is defined in such a way that the index set which generates the covariance structure follows a distributive lattice which not necessarily is a chain, i.e., instead of (3) – (6) only (3) – (5) hold.

Definition 4.1 (LCI models and $N(\mathcal{L}, \Sigma)$; see [1]). Let $\mathcal{L} = \{L_1, \ldots, L_{q+1}\}$ be a partially ordered set. The family of normal distributions $N(\mathcal{L}, \Sigma)$ is said to satisfy the LCI property with respect to $\mathcal{L}$ if, for each pair of elements $L_i, L_j \in \mathcal{L}$ the following holds:

$$L_i \perp\!\!\!\perp L_j \mid (L_i \cap L_j). \tag{9}$$

As discussed previously, the TOCI models can also be obtained via an ordered partition of the index set I. Since the ordered partition satisfies (9) the TOCI class is a subclass of the LCI class. To show that it is a proper subclass in the next proposition an interesting property which is valid for all TOCI models is given.

Proposition 4.1 (Totally ordered conditional independence property). *Let $D = (V, E)$ be a DAG. Given a well numbered (ascending order) of $v_1, v_2, \ldots, v_p$ of the elements of V. For the TOCI class any pair of nodes, v_i, v_j, satisfy*

$$pa(v_i) = pa(v_j) \implies (v_i \perp\!\!\!\perp v_j) \mid pa(v_i). \tag{10}$$

Proposition 4.2. *For each TOCI model there is a LCI model, equivalently a TDAG model, whereas for some LCI models it may not be possible to represent them as a TOCI model.*

Now via an example it will be shown that the TOCI model class is a proper subset of the LCI model class. Consider the graph in Figure 2. Then $pa(\{3\}) = pa(\{4\}) = \{1\}$ whereas according to the graph in Figure 2 $\{3\} \perp\!\!\!\perp \{4\} \mid \{1\}$ does not hold and therefore from (10) it follows that the LCI model described via Figure 2 is not a TOCI model.

To illustrate the proposition it is noticed that the LCI model encoded in Figure 2 has an inverse covariance structure which equals

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} & \sigma^{14} \\ \sigma^{21} & \sigma^{22} & 0 & 0 \\ \sigma^{31} & 0 & \sigma^{33} & \sigma^{34} \\ \sigma^{41} & 0 & \sigma^{43} & \sigma^{44} \end{pmatrix}. \tag{11}$$

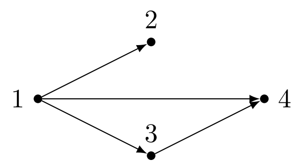

Figure 2. An example of a LCI (TDAG) model which cannot be expressed as a TOCI model.

A TOCI model with a similar inverse covariance structure was presented as Alternative 6 in Section 3 where the inverse covariance matrix equals

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} & \sigma^{14} \\ \sigma^{21} & \sigma^{22} & 0 & \sigma^{24} \\ \sigma^{31} & 0 & \sigma^{33} & \sigma^{34} \\ \sigma^{41} & \sigma^{42} & \sigma^{43} & \sigma^{44} \end{pmatrix}. \tag{12}$$

The difference between the two models can be further exploited via the likelihood. Corresponding to (11)

$$f_I \propto f_1 f_{2|1} f_{\{3,4\}|\{1\}} = f_1 f_{2|1} f_{3|1} f_{4|\{1,3\}}$$

whereas the likelihood corresponding to (12) satisfies

$$f_I \quad \propto \quad f_1 f_{2|1} f_{3|1} f_{4|\{1,2,3\}}.$$

Hence, the difference between the two factorizations lay in the terms $f_{4|\{1,3\}}$ and $f_{4|\{1,2,3\}}$. For the LCI model it holds that $f_{\{3,4\}|\{1\}} = f_{3|1} f_{4|\{1,3\}}$ which makes sense since there is a direction from $3 \to 4$ (see Figure 2). For TOCI models there is by definition always a direction and we only have to think about the jumps.

4.1. Jumps and immoralities. In Section 3 it was seen that the structure in the TOCI models was determined through the jumps. Below a result is presented where jumps and immoralities are connected which can be useful when evaluating a graph.

Theorem 4.1. *Let $\mathcal{K}$ be a chain and let $\mathcal{L}$ be the corresponding partially ordered index set. Then a jump in $\mathcal{K}$ (except the last one) coincides with an immorality in $\mathcal{L}$.*

Proof. First it is proven that a jump in $\mathcal{K}$ implies an immorality in $\mathcal{L}$. Consider a jump between K_i and $K_{i+1} \neq I$ in $\mathcal{K}$. By definition of TOCI, elements of the difference sets $D_i = K_{i+1} \setminus K_i$, given in Definition 3.2, are conditionally independent given K_i (see Proposition 3.1). Moreover, there are no direct edges among elements in the set D_i. Now because $K_{i+1} \neq I$, there is an edge from each element of the set D_i to its descendants $D_{i+1}, ..., I$. Thus, by definition of a jump, two (or more) elements in the set D_i are not directly linked but have a common child, that arises as an immorality.

Next it is established that an immorality in $\mathcal{L}$ implies a jump in $\mathcal{K}$. Consider an immorality in $\mathcal{L}$, such that there is no edge between $v, u \in E$, but $u \to w \in E$ and $v \to w \in E$. If there is no direct edge between u and v, then it follows that u and v belong to the same component of the difference set D_i, that is $u, v \in D_i$ but $u, v \notin K_i$. Hence $|D_i| \geq 2$ which implies a jump in $\mathcal{K}$. The "last jump" from, say, K_q to K_{q+1} or $|D_q| \geq 2$ does not imply a jump or immorality, because the last-jump corresponds to the last descendants and they can not have a common child. □

4.2. Markov equivalence of TOCI models. Two models are Markov equivalent if their corresponding graphs capture the same model of conditional independencies. It has been proven that two DAG models are Markov equivalent if and only if they have the same skeleton (same structure) and the same immoralities, see [9]. Therefore, the simplest way to determine Markov equivalence for TOCI models is by creating the TDAG and applying the DAG Markov equivalence concept. However, given a TOCI model, or the corresponding chain, it is trivial to say whether the TOCI model is unique in the sense that there does not exist any Markov equivalent TOCI model or otherwise it is easy to infer all Markov equivalent TOCI models. In the example in Section 3 only $\mathcal{K}_4$ and $\mathcal{K}_8$ have other Markov equivalent TOCI models, because they contain the (sub-chains) chains of a subset of I. Given a TOCI model determined by the chain $\mathcal{K}$, there can only be the following possibilities:

- $\mathcal{K}$ has a unique description of the implied conditional independencies. For example, the trivial totally ordered chain $\mathcal{K} = \{\{\varnothing\},\ I\}$, where the difference $I \setminus \{\varnothing\}$ is the set I itself, has no other equivalent TOCI models.
- $\mathcal{K}$ does not have a unique description of the implied conditional independencies. For example, $\mathcal{K} = \{\{\varnothing\},\ \{1\},\ \{1,2\},\ \{1,2,3\},\ \ I\}$: The difference set is given by $D(\mathcal{K}_8) = \{\{1\},\ \{2\},\ \{3\},\ \{4\}\}$. Then the chain can be represented in $|I| = p!$ ways. However, there is

no jump and thus no conditional independence constraints in this chain.

§5. Conclusion

The totally ordered conditional independence model class (TOCI) is proposed. It is shown that the TOCI class is a proper subclass of models of the lattice conditional independence models class (LCI models). As with the LCI class the TOCI models share the properties of the possibility to factorize the likelihood into products of conditional likelihood factors with disjoint sets of parameters. Moreover, each conditional likelihood function corresponds to an ordinary multivariate linear regression model and thus explicit estimators are available for all parameters. The benefit of studying a more restricted class of models than the LCI class is that the TOCI model class mainly considers jumps in the chain. The LCI models use besides jumps also the direction presented in a graph. It is easier to interpret results of a statistical analysis of a TOCI model in comparison with the analysis based on some LCI models which are not TOCI models. In future the idea is to link the causality concept (see [8]) to the TOCI class of models.

References

1. S. A. Andersson, D. Madigan, M. D. Perlman, Ch. M. Triggs, *On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs.* — J. Statist. Plann. Inference **48**, 25–46 (1995).
2. S. A. Andersson, J. I. Marden, M. D. Perlman, *Totally ordered multivariate linear models.* — Sankhyā Ser. A, 370–394 (1993).
3. S. A. Andersson, M. D. Perlman, *Lattice models for conditional independence in a multivariate normal distribution.* — The Annals of Statistics, 1318–1358 (1993).
4. S. A. Andersson, M. D. Perlman, *Normal linear models with lattice conditional independence restrictions*, *Multivariate analysis and its applications (Hong Kong, 1992)*, 97–110, IMS Lecture Notes Monogr. Ser., 24, Institute of Mathematical Statistics, Hayward, CA, 1994.
5. S. A. Andersson, M. D. Perlman, *Normal linear regression models with recursive graphical Markov structure.* — J. Multivar. Anal. **66**, 133–187 (1998).
6. G. Grätzer, *General Lattice Theory.* With appendices by B. A. Davey, R. Freese, B. Ganter, M. Greferath, P. Jipsen, H. A. Priestley, H. Rose, E. T. Schmidt, S. E. Schmidt, F. Wehrung and R. Wille. Reprint of the 1998 second edition. Birkhäuser Verlag, Basel, 2003.
7. S. L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series, 17. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1996.

8. S. L. Lauritzen, and Th. S. Richardson, *Chain graph models and their causal interpretations.* — J. Royal Statist. Soc. Ser. B Statistical Methodology **64**, 321–361 (2002).
9. T. S. Richardson, *A characterization of Markov equivalence for directed cyclic graphs. 1996 Uncertainty in AI (UAI '96) Conference.* — Inter. J. Approx. Reasoning **17**, 107–162 (1997).

Поступило 31 мая 2021 г.

Uppsala University,
Uppsala, Sweden
E-mail: Niharika.Gauraha@it.uu.se

Swedish University
of Agricultural Sciences, Sweden

Linköping University,
Linköping, Sweden
E-mail: dietrich.von.rosen@slu.se