JOURNAL OF ALGEBRAIC STATISTICS
Vol. 6, No. 2, 2015, 150-167
ISSN 1309-3452 – www.jalgstat.com

A linear-algebraic tool for conditional independence inference

Kentaro Tanaka[1,*], Milan Studený[2], Akimichi Takemura[3], Tomonari Sei[4]

[1] *Faculty of Economics, Seikei University, Tokyo, Japan*
[2] *Institute of Information Theory and Automation of the CAS, Prague, Czech Republic*
[3] *Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan*
[4] *Department of Mathematics, Faculty of Science and Technology, Keio University, Yokohama, Japan*

Abstract. In this note, we propose a new linear-algebraic method for the implication problem among conditional independence statements, which is inspired by the factorization characterization of conditional independence. First, we give a criterion in the case of a discrete strictly positive density and relate it to an earlier linear-algebraic approach. Then, we extend the method to the case of a discrete density that need not be strictly positive. Finally, we provide a computational result in the case of six variables.

2000 Mathematics Subject Classifications: 60A99, 68T15

Key Words and Phrases: Conditional independence inference, Automated theorem proving

---

1. Introduction

In this paper, we deal with the *conditional independence* (CI) implication problem, that is, testing whether a CI statement can be derived from a set of other CI statements.

It is well-known that there is no finite axiomatic characterization for the CI implication problem with general discrete probability distributions (Studený [14]). The situation is different if we restrict the class of CI statements. It is well-known that there exists a finite axiomatic characterization for each of the following restricted CI frames: unconditional independence statements (Geiger et al. [4], Matúš [10]); saturated CI statements (Geiger and Pearl [5], Malvestuto [8], Malvestuto and Studený [9]); CI statements represented by Markov networks (Pearl and Paz [12]), and so forth. See Niepert et al. [11] and Studený [15] for the comprehensive description.

Another way to approach the CI implication problem is based on algebra. The method of imsets by Studený [15] provides a powerful linear-algebraic method for testing the CI

*Corresponding author.

Email addresses: tanaken@econ.seikei.ac.jp (Kentaro Tanaka)

implications. By using the method of imsets, the CI implication problem is translated into relations among integer-valued vectors. In Bouckaert et al. [2], a method of linear programming for computer testing CI implications has been proposed. In this paper, we introduce another type of a linear-algebraic method for the CI implication problem which is particularly suitable in the case when the distribution is strictly positive.

The structure of the paper is as follows. In Section 2, we recall the method of imsets and formulate two lemmas. In Section 3, we give a criterion applicable in the case of a discrete strictly positive density. We also give some examples there to illustrate how to use it and discuss its relation to a former linear-algebraic sufficient condition for probabilistic CI implications. In Section 4, we deal with the case where discrete densities are not necessarily strictly positive. In Section 5, we present a computational example to demonstrate our method. Finally, in Conclusions, we summarize our results and discuss a possible relation of our approach to toric ideals.

2. Preliminaries

Throughout the paper, N is a finite indexing set for variables; to avoid the trivial case we assume $|N| \geq 2$. Given disjoint $A, B \subseteq N$ the symbol AB will be a shorthand for their union $A \cup B$.

2.1. Distributions and conditional independence

The *sample space* for our (discrete multivariate) probability distributions will be the direct product $\mathcal{X} := \prod_{i \in N} \mathcal{X}_i$, where $\mathcal{X}_i$, $i \in N$ are non-empty finite sets. Given a joint configuration of values $x \equiv [x_i]_{i \in N} \in \mathcal{X}$ and $A \subseteq N$, the symbol x_A will denote its *marginal configuration* $[x_i]_{i \in A}$. The *marginal sample space* for $A \subseteq N$ will be the collection $\mathcal{X}_A$ of marginal configurations for A. In particular, $\mathcal{X}_N \equiv \mathcal{X}$. Observe that for $A = \emptyset$ and $x \in \mathcal{X}$ the marginal configuration $x_\emptyset$ is the empty list $[x_i]_{i \in \emptyset}$. Thus, the marginal space for the empty set $\mathcal{X}_\emptyset$ is also introduced: it is a one-element set containing the empty configuration. Given $x \in \mathcal{X}_A$ and $y \in \mathcal{X}_B$ for disjoint $A, B \subseteq N$, the symbol $[x, y] \in \mathcal{X}_{AB}$ will denote their concatenation.

Any real-valued function on $\mathcal{X}_A$, for $A \subseteq N$, can formally be understood as a function on $\mathcal{X}$ which only depends on the components in A. In this case we say it is a *function of* A and denote the function as $q(A; *)$, where $* \in \mathcal{X}$ is the argument. Moreover, we will take advantage of the following flexible notation: given $x \in \mathcal{X}_D$, where $A \subseteq D \subseteq N$, we will write $q(A; x)$ to denote the value of the function q (of A) for $x_A \in \mathcal{X}_A$.

The *density* p of a probability distribution P on $\mathcal{X}$ is a function $p : \mathcal{X} \to [0, 1]$ such that $\sum_{x \in \mathcal{X}} p(x) = 1$. It is strictly *positive* if $p(x) > 0$ for every $x \in \mathcal{X}$. The *marginal density* of p for $A \subseteq N$ is a function on $\mathcal{X}_A$, usually denoted by $p(A; *)$:

$$p(A; x) := \sum_{y \in \mathcal{X}_{N \setminus A}} p([x_A, y]) \qquad \text{for } x \in \mathcal{X}.$$

In particular, $p(N;*) \equiv p(*)$. In our setting, conditional independence can be introduced as follows.

Definition 1. *For pairwise disjoint sets $A, B, C \subseteq N$ and the density p of a probability distribution P on $\mathcal{X}$, we say that A and B are conditionally independent given C with respect to P and write $A \perp\!\!\!\perp B \mid C\,[P]$ if the following equation holds:*

$$\forall\, x \in \mathcal{X} \qquad p(C;x) \cdot p(ABC;x) = p(AC;x) \cdot p(BC;x)\,. \tag{1}$$

2.2. Factorization characterization of conditional independence

Let $A, B, C \subseteq N$ be a triplet of pairwise disjoint sets. A well-known characterization of $A \perp\!\!\!\perp B \mid C\,[P]$ is in terms of factorization of the marginal density $p(ABC;*)$ to functions of AC and BC. We can formally extend this characterization to the case of functions of ACD and BCD, where D denotes $N \setminus ABC$. We give a straightforward proof.

Lemma 1. *For a probability distribution P on $\mathcal{X}$, $A \perp\!\!\!\perp B \mid C\,[P]$ is true if and only if there exist functions $q(ACD;\ *)$ and $r(BCD;*)$ such that the marginal density decomposes as follows:*

$$\forall\, x \in \mathcal{X} \qquad p(ABC;x) = q(ACD;x) \cdot r(BCD;x)\,. \tag{2}$$

As the left-hand side of (2) only depends on the components in ABC, the right-hand side of (2) does not depend on x_D, despite its factors $q(ACD;x)$ and $r(BCD;x)$ may depend on x_D.

Proof. Assume that (1) holds and put:

$$q(ACD;x) = \begin{cases} \frac{p(AC;x)}{p(C;x)} & \text{if } p(C;x) > 0, \\ 0 & \text{if } p(C;x) = 0, \end{cases} \qquad r(BCD;x) = p(BC;x) \qquad \text{for any } x \in \mathcal{X}.$$

Note that these particular functions $q(ACD;x)$ and $r(BCD;x)$ do not depend of x_D. If $p(C;x) = 0$, then $p(ABC;x) = 0$. Hence, (2) is valid. For the converse implication assume that (2) holds. Fix some $w \in \mathcal{X}_D$ and write for any $x \in \mathcal{X}$:

$$p(ABC;x) = p(ABC;[x_{ABC}, w]) = q(ACD;[x_{AC}, w]) \cdot r(BCD;[x_{BC}, w]),$$

$$p(AC;x) = \sum_{y \in \mathcal{X}_B} p(ABC;[y, x_{AC}, w]) = q(ACD;[x_{AC}, w]) \cdot \left\{ \sum_{y \in \mathcal{X}_B} r(BCD;[y, x_C, w]) \right\},$$

$$p(BC;x) = \sum_{z \in \mathcal{X}_A} p(ABC;[z, x_{BC}, w]) = r(BCD;[x_{BC}, w]) \cdot \left\{ \sum_{z \in \mathcal{X}_A} q(ACD;[z, x_C, w]) \right\},$$

$$p(C;x) = \left\{ \sum_{z \in \mathcal{X}_A} q(ACD;[z, x_C, w]) \right\} \cdot \left\{ \sum_{y \in \mathcal{X}_B} r(BCD;[y, x_C, w]) \right\}.$$

Hence, by substitution, we easily get (1), which was desired.

2.3. Imsets

For $S \subseteq N$, the symbol $\mathcal{P}(S)$ will denote its power set $\{T : T \subseteq S\}$. We will mainly deal with vectors in $\mathbb{R}^{\mathcal{P}(N)}$, respectively in $\mathbb{R}^{\mathcal{K}}$ for some $\mathcal{K} \subseteq \mathcal{P}(N)$. The symbol $\mathbf{0}$ will denote the zero vector. A well-known linear basis of $\mathbb{R}^{\mathcal{P}(N)}$ consists of the identifiers δ_T for sets $T \subseteq N$:

$$\delta_T(S) = \begin{cases} 1 & \text{if } S = T, \\ 0 & \text{if } S \neq T. \end{cases}$$

Given $\mathcal{L} \subseteq \mathcal{P}(N)$, we will denote the linear subspace of $\mathbb{R}^{\mathcal{P}(N)}$ spanned by $\{\delta_T : T \in \mathcal{L}\}$ by the symbol $L_{\mathcal{L}}$. Given $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$ and $\mathcal{K} \subseteq \mathcal{P}(N)$, the restriction of u to the components in $\mathcal{K}$, which is an element of $\mathbb{R}^{\mathcal{K}}$, will be denoted by the symbol $\mathsf{u}|_{\mathcal{K}}$. The following observation is evident.

$$\text{Given } \mathcal{L} \subseteq \mathcal{P}(N),\ \mathsf{u} \in L_{\mathcal{L}} \text{ if and only if } \mathsf{u}|_{\mathcal{K}} = \mathbf{0}\ , \text{ where } \mathcal{K} = \mathcal{P}(N) \setminus \mathcal{L}. \tag{3}$$

An *imset* is a vector in $\mathbb{R}^{\mathcal{P}(N)}$ whose components are integers, that is, an element of $\mathbb{Z}^{\mathcal{P}(N)}$. Any conditional independence statement over N corresponds to an ordered triplet $\langle A, B \,|\, C\rangle$ of pairwise disjoint sets $A, B, C \subseteq N$. Every such a triplet is assigned the respective *semi-elementary imset* $\mathsf{u}_{\langle A,B\,|\,C\rangle}$:

$$\mathsf{u}_{\langle A,B\,|\,C\rangle} := \delta_{ABC} - \delta_{AC} - \delta_{BC} + \delta_C\,. \tag{4}$$

The triplet is trivial if either $A = \emptyset$ or $B = \emptyset$, otherwise it is called *non-trivial.* Observe that $\mathsf{u}_{\langle A,B\,|\,C\rangle} = \mathbf{0}$ if and only if $\langle A, B \,|\, C\rangle$ is trivial.

Definition 2. *We say that* $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$ *is o-standardized if*

$$\sum_{S \subseteq N} \mathsf{u}(S) = 0 \quad \textit{and} \quad \forall i \in N \quad \sum_{S \subseteq N, i \in S} \mathsf{u}(S) = 0\,.$$

Apparently, every semi-elementary imset $\mathsf{u}_{\langle A,B\,|\,C\rangle}$ is o-standardized. As the set of o-standardized vectors is a linear subspace of $\mathbb{R}^{\mathcal{P}(N)}$, linear combinations of semi-elementary imsets are also o-standardized. Below, we employ the following auxiliary observation.

Lemma 2. *Let* $\mathcal{L} \subseteq \mathcal{P}(N)$ *be a class of sets closed under subsets, which means that* $S \in \mathcal{L},\ T \subseteq S \ \Rightarrow\ T \in \mathcal{L}$*; put* $\mathcal{K} = \mathcal{P}(N) \setminus \mathcal{L}$*. Then an o-standardized vector* $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$ *satisfies* $\mathsf{u}|_{\mathcal{K}} = \mathbf{0}$ *iff, for some non-negative integer* J*,*

$$\mathsf{u} = \sum_{j=1}^{J} \tau_j \cdot \mathsf{u}_{\langle E_j,F_j\,|\,G_j\rangle} \qquad \textit{with real coefficients } \{\tau_j\}_{j=1}^{J} \textit{ and } E_jF_jG_j \in \mathcal{L}.$$

Proof. It is obvious that if u is written as the sum above then $\mathsf{u}|_{\mathcal{K}} = 0$. We prove the converse by induction on $\ell = |\mathcal{L} \cap \{S \subseteq N : |S| \geq 2\}|$. If $\ell = 0$ then $\mathsf{u}|_{\mathcal{K}} = \mathbf{0}$ means $\mathsf{u} = \mathbf{0}$ due to the property of o-standardization and there is no non-trivial $\langle E_j, F_j \,|\, G_j\rangle$ with $E_jF_jG_j \in \mathcal{L}$. If $\ell > 0$ choose inclusion maximal $T \in \mathcal{L}$ with $|T| \geq 2$, find non-trivial

$\langle E_j, F_j \,|\, G_j \rangle$ with $E_jF_jG_j = T$, put $\tau_j \equiv \mathsf{u}(T)$ and apply the induction hypothesis to $\mathcal{L}\backslash\{T\}$ and the vector $\mathsf{u} - \tau_j \cdot \mathsf{u}_{\langle E_j,F_j \,|\, G_j\rangle}$. Note that $\mathsf{u} - \tau_j \cdot \mathsf{u}_{\langle E_j,F_j \,|\, G_j\rangle}$ is o-standardized because both u and $\mathsf{u}_{\langle E_j,F_j \,|\, G_j\rangle}$ are o-standardized. Moreover, the number of sets of cardinality at least 2 in $\mathcal{L} \setminus \{T\}$ is strictly less than ℓ, the number of such sets in $\mathcal{L}$. Thus, one can repeat this induction step until ℓ decreases to 0.

3. The case of a strictly positive distribution

Let $A, B, C \subseteq N$, respectively $A_i, B_i, C_i \subseteq N$ for $i = 1, \ldots, I$, are triplets of pairwise disjoint sets. We are dealing with the implication problem

$$\{A_i \perp\!\!\!\perp B_i \,|\, C_i \,[P]\}_{i=1}^{I} \quad \Rightarrow \quad A \perp\!\!\!\perp B \,|\, C \,[P], \tag{5}$$

for any distribution P with a strictly positive density. The main observation is as follows.

Theorem 1. *Let D denote $N \setminus ABC$. If there exist real numbers $\{\lambda_i\}_{i=1}^{I}$ such that*

$$\delta_{ABC} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i,B_i \,|\, C_i\rangle} \in L_{\mathcal{P}(ACD)\cup\mathcal{P}(BCD)} \tag{6}$$

then the implication (5) holds for any distribution P with a strictly positive density.

Proof. By (6), there exist real numbers α_S for $S \in \mathcal{P}(ACD) \cup \mathcal{P}(BCD)$ with

$$\sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i,B_i \,|\, C_i\rangle} = -\delta_{ABC} + \sum_{S\in\mathcal{P}(ACD)\cup\mathcal{P}(BCD)} \alpha_S \cdot \delta_S \,. \tag{7}$$

We show, for any distribution P with a strictly positive density satisfying the CI statements $\{A_i \perp\!\!\!\perp B_i \,|\, C_i \,[P]\}_{i=1}^{I}$, that

$$\begin{aligned} \forall\, x \in \mathcal{X} \qquad p(ABC; x) \;&=\; \prod_{S\in\mathcal{P}(ACD)\cup\mathcal{P}(BCD)} p(S;x)^{\alpha_S} \\ &=\; \underbrace{\prod_{S\in\mathcal{P}(ACD)} p(S;x)^{\alpha_S}}_{q(ACD;x)} \cdot \underbrace{\prod_{S\in\mathcal{P}(BCD)\setminus\mathcal{P}(ACD)} p(S;x)^{\alpha_S}}_{r(BCD;x)} \,, \end{aligned} \tag{8}$$

which clearly implies, by Lemma 1, that $A \perp\!\!\!\perp B \,|\, C \,[P]$. Since the density p of P is strictly positive, for any fixed $x \in \mathcal{X}$, one can write, by $A_i \perp\!\!\!\perp B_i \,|\, C_i \,[P]$ for $i \in \{1, \ldots, I\}$,

$$\left[\; \forall\, i \quad 1 = \frac{p(A_iB_iC_i;x)\cdot p(C_i;x)}{p(A_iC_i;x)\cdot p(B_iC_i;x)} \;\right] \quad \Longrightarrow \quad 1 = \prod_{i=1}^{I} \left(\frac{p(A_iB_iC_i;x)\cdot p(C_i;x)}{p(A_iC_i;x)\cdot p(B_iC_i;x)} \right)^{\lambda_i} .$$

The expression on the right-hand side there can be, by (7), re-written as

$$
\begin{aligned}
1 &= \prod_{i=1}^{I} \left(\frac{p(A_iB_iC_i;x)\cdot p(C_i;x)}{p(A_iC_i;x)\cdot p(B_iC_i;x)} \right)^{\lambda_i} = \prod_{i=1}^{I} \left(\prod_{T\subseteq N} p(T;x)^{\mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle}(T)} \right)^{\lambda_i} \\
&= \prod_{T\subseteq N} p(T;x)^{\sum_{i=1}^{I}\lambda_i\cdot \mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle}(T)} \stackrel{(7)}{=} \frac{1}{p(ABC;x)} \cdot \prod_{S\in\mathcal{P}(ACD)\cup\mathcal{P}(BCD)} p(S;x)^{\alpha_S}\,,
\end{aligned}
$$

and, since the term $p(ABC;x)$ here is strictly positive, one gets (8) by multiplying it by the factor $p(ABC;x)$.

3.1. Equivalent formulations of the condition

In this section, we give two equivalent formulations of the sufficient condition (6) for the implication (5) in the strictly positive case.

Lemma 3. *Let $A,B,C\subseteq N$, respectively $A_i,B_i,C_i\subseteq N$ for $i=1,\ldots,I$, be triplets of pairwise disjoint sets; denote $D\equiv N\setminus ABC$. Given a collection of real numbers $\{\lambda_i\}_{i=1}^{I}$ the following conditions are equivalent:*

(a) *the condition (6) holds, that is,*

$$\delta_{ABC} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle} \in L_{\mathcal{P}(ACD)\cup\mathcal{P}(BCD)}\,,$$

(b) *for $\mathcal{K}=\mathcal{P}(N)\setminus\{\mathcal{P}(ACD)\cup\mathcal{P}(BCD)\}$, one has*

$$(\mathsf{u}_{\langle A,B\,|\,C\rangle} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle})|_{\mathcal{K}} = \mathbf{0}\,, \tag{9}$$

(c) *there exist real numbers $\{\kappa_j\}_{j=1}^{J}$, $J\geq 0$ and pairwise disjoint triplets $\{\langle E_j,F_j\,|\,G_j\rangle\}_{j=1}^{J}$ such that, for any $j\in\{1,\ldots,J\}$, $E_jF_jG_j\in\mathcal{P}(ACD)\cup\mathcal{P}(BCD)$ and*

$$\mathsf{u}_{\langle A,B\,|\,C\rangle} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle} + \sum_{j=1}^{J} \kappa_j \cdot \mathsf{u}_{\langle E_j,F_j\,|\,G_j\rangle} = \mathbf{0}\,. \tag{10}$$

Proof. For the proof of (a)⇔(b) consider $\mathcal{L}\equiv\mathcal{P}(ACD)\cup\mathcal{P}(BCD)$. Since evidently $-\delta_{AC}-\delta_{BC}+\delta_C\in L_{\mathcal{L}}$ the condition in (a) means $\mathsf{u}\equiv\mathsf{u}_{\langle A,B\,|\,C\rangle}+\sum_{i=1}^{I}\lambda_i\cdot\mathsf{u}_{\langle A_i,B_i\,|\,C_i\rangle}\in L_{\mathcal{L}}$. Thus, (a)⇔(b) follows from (3) applied to u. The equivalence (b)⇔(c) then follows from Lemma 2, where one has $\kappa_j=-\tau_j$ for $j=1,\ldots,J$.

3.2. Some examples

Let us discuss which of the well-known CI implications can be derived by our method.

Example 1. Consider the case $|N| \geq 3$ and the CI implication, named *contraction* rule,

$$a \perp\!\!\!\perp b \,|\, c,\; a \perp\!\!\!\perp c \,|\, \emptyset \quad \Rightarrow \quad a \perp\!\!\!\perp bc \,|\, \emptyset\,, \tag{11}$$

where a, b, c are distinct elements of N. Since one has $\mathsf{u}_{\langle a,bc\,|\,\emptyset\rangle} - \mathsf{u}_{\langle a,b\,|\,c\rangle} - \mathsf{u}_{\langle a,c\,|\,\emptyset\rangle} = \mathbf{0}$ the condition (9) is fulfilled for $\langle A,B\,|\,C\rangle = \langle a,bc\,|\,\emptyset\rangle$, $\langle A_1,B_1\,|\,C_1\rangle = \langle a,b\,|\,c\rangle$, $\langle A_2,B_2\,|\,C_2\rangle = \langle a,c\,|\,\emptyset\rangle$ and $\lambda_1 = \lambda_2 = -1$. In particular, Lemma 3 and Theorem 1 imply that contraction is valid for P with a strictly positive density. Another positive example is the so-called *weak union* rule

$$a \perp\!\!\!\perp bc \,|\, \emptyset \quad \Rightarrow \quad a \perp\!\!\!\perp b \,|\, c\,, \tag{12}$$

in which case $\mathsf{u}_{\langle a,b\,|\,c\rangle} - \mathsf{u}_{\langle a,bc\,|\,\emptyset\rangle} + \mathsf{u}_{\langle a,c\,|\,\emptyset\rangle} = \mathbf{0}$ implies (10) is true for $\langle A,B\,|\,C\rangle = \langle a,b\,|\,c\rangle$, $\langle A_1,B_1\,|\,C_1\rangle = \langle a,bc\,|\,\emptyset\rangle$, $\lambda_1 = -1$, $\langle E_1,F_1\,|\,G_1\rangle = \langle a,c\,|\,\emptyset\rangle$ and $\kappa_1 = +1$. Thus, we have shown the weak union holds for the distributions with a strictly positive density.

The implications from Example 1 are valid for general discrete distributions. This is not the case in the following example.

Example 2. Consider the case $|N| \geq 3$ and the following CI implication case

$$a \perp\!\!\!\perp b \,|\, c,\; a \perp\!\!\!\perp c \,|\, b \quad \Rightarrow \quad a \perp\!\!\!\perp c \,|\, \emptyset\,. \tag{13}$$

Thus, we have $\langle A,B\,|\,C\rangle = \langle a,c\,|\,\emptyset\rangle$, $\langle A_1,B_1\,|\,C_1\rangle = \langle a,b\,|\,c\rangle$, $\langle A_2,B_2\,|\,C_2\rangle = \langle a,c\,|\,b\rangle$. To verify (9) realize that $\mathcal{K}$ consists of supersets of ac and observe the choice $\lambda_1 = +1$ and $\lambda_2 = -1$ reaches the goal. Thus, the implication (13) has been verified for any distribution with a strictly positive density. Note that there exists a discrete distribution for which (13) does not hold; see Example 2.3 of Studený [15]. In particular, (13) cannot be derived using the method discussed in Bouckaert et al. [2] applicable to general discrete distributions.

The next example shows that the repeated application of our new method makes sense.

Example 3. Consider the case $|N| \geq 3$. The *decomposition* rule

$$a \perp\!\!\!\perp bc \,|\, \emptyset \quad \Rightarrow \quad a \perp\!\!\!\perp c \,|\, \emptyset\,, \tag{14}$$

is an example of a valid CI implication, whose validity cannot be verified directly by means of our method. Here "directly" means that we only use our method *once* and do not use any other rules. In this case for $\langle A,B\,|\,C\rangle = \langle a,c\,|\,\emptyset\rangle$, $\langle A_1,B_1\,|\,C_1\rangle = \langle a,bc\,|\,\emptyset\rangle$ the condition (9) is not fulfilled for any $\lambda_1 \in \mathbb{R}$. Indeed, assume for a contradiction that such λ_1 exists. Since $abc, ac \in \mathcal{K}$, $\mathsf{u}_{\langle a,bc\,|\,\emptyset\rangle}(abc) = 1$ and $\mathsf{u}_{\langle a,c\,|\,\emptyset\rangle}(abc) = 0$ one has $0 \overset{(9)}{=} \mathsf{u}_{\langle a,c\,|\,\emptyset\rangle}(abc) + \lambda_1 \cdot \mathsf{u}_{\langle a,bc\,|\,\emptyset\rangle}(abc) = \lambda_1$. Then, however, $\mathsf{u}_{\langle a,c\,|\,\emptyset\rangle}(ac) + \lambda_1 \cdot \mathsf{u}_{\langle a,bc\,|\,\emptyset\rangle}(ac) = 1$ gives a contradiction with (9). Nevertheless, the decomposition rule (14) can be derived by repeated application of (6). That is, we first derive (12) and $a \perp\!\!\!\perp bc \,|\, \emptyset \Rightarrow a \perp\!\!\!\perp c \,|\, b$ in a similar way, and then obtain $a \perp\!\!\!\perp c \,|\, \emptyset$ using the result of (13).

Next, consider the well-known *intersection* rule

$$a \perp\!\!\!\perp b \,|\, c,\ a \perp\!\!\!\perp c \,|\, b \quad \Rightarrow \quad a \perp\!\!\!\perp bc \,|\, \emptyset\,. \tag{15}$$

It cannot be derived directly using our condition (6). Indeed, this time one has $\langle A, B \,|\, C\rangle = \langle a, bc \,|\, \emptyset\rangle$, $\langle A_1, B_1 \,|\, C_1\rangle = \langle a, b \,|\, c\rangle$, $\langle A_2, B_2 \,|\, C_2\rangle = \langle a, c \,|\, b\rangle$ and $ab, ac, abc \in \mathcal{K}$. Assume for a contradiction that $(\mathsf{u}_{\langle a,bc \,|\, \emptyset\rangle} + \lambda_1 \cdot \mathsf{u}_{\langle a,b \,|\, c\rangle} + \lambda_2 \cdot \mathsf{u}_{\langle a,c \,|\, b\rangle})|_{\mathcal{K}} = \mathbf{0}$ for $\lambda_1, \lambda_2 \in \mathbb{R}$. Then the facts $\mathsf{u}_{\langle a,bc \,|\, \emptyset\rangle}(ac) = \mathsf{u}_{\langle a,c \,|\, b\rangle}(ac) = 0$ and $\mathsf{u}_{\langle a,b \,|\, c\rangle}(ac) = -1$ imply $\lambda_1 = 0$. Analogously, $\mathsf{u}_{\langle a,bc \,|\, \emptyset\rangle}(ab) = \mathsf{u}_{\langle a,b \,|\, c\rangle}(ab) = 0$ and $\mathsf{u}_{\langle a,c \,|\, b\rangle}(ab) = -1$ implies $\lambda_2 = 0$ and $1 = \mathsf{u}_{\langle a,bc \,|\, \emptyset\rangle}(abc)$ gives a contradiction. Nevertheless, (15) can be derived by repeated application of (6); specifically, we first derive (13) and then obtain $a \perp\!\!\!\perp bc \,|\, \emptyset$ using the result of (11).

3.3. Relation to an earlier method

A natural question is whether there is a relation of our new condition (6) to a former imset-based sufficient condition for probabilistic CI implication proposed in § 6.2 of Studený [15]. That condition was a basis of linear-algebraic methods for computer testing CI implications applied by Bouckaert et al. [2] and can be re-phrased as follows.

Lemma 4. *If there exist pairwise disjoint triplets $\{\langle E_j, F_j \,|\, G_j\rangle\}_{j=1}^{J}$ and non-negative real numbers $\{\iota_i\}_{i=1}^{I}$ and $\{\kappa_j\}_{j=1}^{J}$, that is, $\iota_i, \kappa_j \geq 0$ for any i, j, such that*

$$\mathsf{u}_{\langle A,B \,|\, C\rangle} - \sum_{i=1}^{I} \iota_i \cdot \mathsf{u}_{\langle A_i,B_i \,|\, C_i\rangle} + \sum_{j=1}^{J} \kappa_j \cdot \mathsf{u}_{\langle E_j,F_j \,|\, G_j\rangle} = \mathbf{0}. \tag{16}$$

then the implication (5) holds for any (discrete) distribution P.

The first main difference is that (16) forces the implication (5) for any discrete probability distribution P, not just for the ones with a strictly positive density. On the other hand, the decomposition rule (14) from Example 3 shows that (16) need not imply (6). To characterize the case when (16) ⇒(6) we introduce the following terminology.

Definition 3. *Given disjoint $A, B \subseteq N$ we say a triplet $\langle E, F \,|\, G\rangle$* bridges *between A and B if both $(EFG) \cap A \neq \emptyset$ and $(EFG) \cap B \neq \emptyset$; otherwise, we say the triplet does not bridge between A and B.*

Equivalently, a triplet $\langle E, F \,|\, G\rangle$ does not bridge between sets A and B if and only if $EFG \in \mathcal{P}(N \setminus B) \cup \mathcal{P}(N \setminus A)$. A consequence of Theorem 1 and Lemma 3 is as follows.

Corollary 1. *If pairwise disjoint triplets $\{\langle E_j, F_j \,|\, G_j\rangle\}_{j=1}^{J}$ not bridging between A and B and real numbers $\{\iota_i\}_{i=1}^{I}$ and $\{\kappa_j\}_{j=1}^{J}$ exist such that* (16) *holds, that is,*

$$\mathsf{u}_{\langle A,B \,|\, C\rangle} - \sum_{i=1}^{I} \iota_i \cdot \mathsf{u}_{\langle A_i,B_i \,|\, C_i\rangle} + \sum_{j=1}^{J} \kappa_j \cdot \mathsf{u}_{\langle E_j,F_j \,|\, G_j\rangle} = \mathbf{0},$$

then the condition (6) holds, implying that (5) holds for any (discrete) distribution P with a strictly positive density.

Proof. This follows from the equivalence (a)⇔(c) in Lemma 3. We put $\lambda_i = -\iota_i$ for $i = 1, \ldots, I$ and observe that (16) turns into the condition (10). Then we apply Theorem 1.

Hence, the condition (16) implies our condition (6) under an additional technical assumption that none of the additionally considered triplets $\{\langle E_j, F_j \,|\, G_j\rangle\}_{j=1}^{J}$ bridges between A and B. Note that the condition (16) in Lemma 4 requires the non-negativity of the respective coefficients while the condition in Corollary 1 does not require the non-negativity constraints. Of course, there are some cases when (6) can be applied to derive (5) for any distribution P with a strictly positive density, despite (16) with non-negative coefficients is not applicable. From the point of view of computation, the condition (16) can be tested by linear programming tools as in Bouckaert et al. [2], while the conditions (9) and (10), which are equivalent to the condition (6), can be tested by solely checking the linear dependence among imsets.

3.4. Some interpretation

The observations in Lemma 4 and Corollary 1 allow one to interpret our new method as an approach motivated by the idea of "adding extra CI statements". Consider the implication problem (5) and "add" extra CI statements $\{E_j \perp\!\!\!\perp F_j \,|\, G_j\}_{j=1}^{J}$ to the antecedents in (5) such that none of the added triplets bridges between A and B and obtain

$$\{A_i \perp\!\!\!\perp B_i \,|\, C_i\}_{i=1}^{I} \cup \{E_j \perp\!\!\!\perp F_j \,|\, G_j\}_{j=1}^{J} \;\Rightarrow\; A \perp\!\!\!\perp B \,|\, C\,. \tag{17}$$

Provided we are able to verify the implication (17) by the method from Lemma 4 we can utilize the corresponding linear relation (16) in the context of Corollary 1. We may derive more than just (17) because we have no restriction to having only non-negative coefficients here. On the other hand, we only derive the validity of (5) for distributions with strictly positive density.

Example 4. Consider the case $|N| \geq 4$ and the following CI implication problem

$$a \perp\!\!\!\perp b \,|\, cd,\; c \perp\!\!\!\perp d \,|\, ab,\; c \perp\!\!\!\perp d \,|\, a,\; c \perp\!\!\!\perp d \,|\, b \quad\Rightarrow\quad c \perp\!\!\!\perp d \,|\, \emptyset\,, \tag{18}$$

where a, b, c, d are distinct elements in N. Let us "add" some CI statements not bridging between c and d, namely $a \perp\!\!\!\perp b \,|\, c$, $a \perp\!\!\!\perp b \,|\, d$ and $a \perp\!\!\!\perp b \,|\, \emptyset$ and get the implication problem

$$a \perp\!\!\!\perp b \,|\, cd,\; c \perp\!\!\!\perp d \,|\, ab,\; c \perp\!\!\!\perp d \,|\, a,\; c \perp\!\!\!\perp d \,|\, b,\; a \perp\!\!\!\perp b \,|\, c,\; a \perp\!\!\!\perp b \,|\, d,\; a \perp\!\!\!\perp b \,|\, \emptyset \quad\Rightarrow\quad c \perp\!\!\!\perp d \,|\, \emptyset\,.$$

The point is that even much stronger version of this implication can be derived by the method from Lemma 4, namely

$$a \perp\!\!\!\perp b \,|\, cd,\; c \perp\!\!\!\perp d \,|\, a,\; c \perp\!\!\!\perp d \,|\, b,\; a \perp\!\!\!\perp b \,|\, \emptyset \quad\Leftrightarrow\quad c \perp\!\!\!\perp d \,|\, ab,\; a \perp\!\!\!\perp b \,|\, c,\; a \perp\!\!\!\perp b \,|\, d,\; c \perp\!\!\!\perp d \,|\, \emptyset\,.$$

In fact, this can be derived from the following linear relation of respective imsets:

$$\mathsf{u}_{\langle c,d \,|\, \emptyset\rangle} + \mathsf{u}_{\langle a,b \,|\, c\rangle} + \mathsf{u}_{\langle a,b \,|\, d\rangle} + \mathsf{u}_{\langle c,d \,|\, ab\rangle} - \mathsf{u}_{\langle a,b \,|\, \emptyset\rangle} - \mathsf{u}_{\langle c,d \,|\, a\rangle} - \mathsf{u}_{\langle c,d \,|\, b\rangle} - \mathsf{u}_{\langle a,b \,|\, cd\rangle} = \mathbf{0}\,. \tag{19}$$

To verify (18) using Corollary 1 one can re-write (19) in the form

$$\mathsf{u}_{\langle c,d\,|\,\emptyset\rangle} + \left[-\mathsf{u}_{\langle a,b\,|\,cd\rangle} + \mathsf{u}_{\langle c,d\,|\,ab\rangle} - \mathsf{u}_{\langle c,d\,|\,a\rangle} - \mathsf{u}_{\langle c,d\,|\,b\rangle}\right] + \left\{\mathsf{u}_{\langle a,b\,|\,c\rangle} + \mathsf{u}_{\langle a,b\,|\,d\rangle} - \mathsf{u}_{\langle a,b\,|\,\emptyset\rangle}\right\} = \mathbf{0}\,,$$

where the four terms in the braces correspond to $\{A_i \perp\!\!\!\perp B_i \,|\, C_i\}_{i=1}^{I}$ and the three terms in curly brackets to $\{E_j \perp\!\!\!\perp F_j \,|\, G_j\}_{j=1}^{J}$.

Actually, by an analogous consideration, one can verify the validity of the implication

$$a \perp\!\!\!\perp b \,|\, cd,\; c \perp\!\!\!\perp d \,|\, ab,\; c \perp\!\!\!\perp d \,|\, a,\; c \perp\!\!\!\perp d \,|\, \emptyset \quad\Rightarrow\quad c \perp\!\!\!\perp d \,|\, b\,, \tag{20}$$

for distributions with a strictly positive density. The same arguments can be used to verify

$$a \perp\!\!\!\perp b \,|\, cd,\; c \perp\!\!\!\perp d \,|\, a,\; c \perp\!\!\!\perp d \,|\, b,\; c \perp\!\!\!\perp d \,|\, \emptyset \quad\Rightarrow\quad c \perp\!\!\!\perp d \,|\, ab. \tag{21}$$

Note that (18), (20) and (21) have been mentioned by Spohn [13] as special CI implications valid for distributions with a strictly positive density. Specifically, they are gathered in the property (S5) from [13]. The specialty of the implication (21) is that it holds even in the case of a general discrete distribution; see Corollary 2.1 and Example 4.1 of Studený [15].

Remark 1. The example that (18) does not hold for general discrete distributions is very simple: put $N = \{a,b,c,d\}$, $\mathcal{X}_i \equiv \{0,1\}$ for $i \in N$, and define the density by assigning the value 1/2 to the configurations of values $[\,0,0,0,0\,]$, $[\,1,1,1,1\,]$ and 0 to the remaining configurations. As concerns the example that (20) does not hold in general put $N = \{a,b,c,d\}$, $\mathcal{X}_a \equiv \{0,1,2,3\}$, $\mathcal{X}_i \equiv \{0,1\}$ for $i \in \{b,c,d\}$ and define the density by assigning 1/4 to the configurations of values $[\,0,0,0,0\,]$, $[\,1,1,0,1\,]$, $[\,2,1,1,0\,]$, $[\,3,0,1,1\,]$ and 0 to the remaining configurations.

Remark 2. This is to warn the reader not to misinterpret our motivational remark before Example 4. We say there that to verify (5) we "turn" it into an extended implication problem (17), where none of the "added" CI statements $\{E_j \perp\!\!\!\perp F_j \,|\, G_j\}_{j=1}^{J}$ bridges between A and B. However, this extended implication problem is *not equivalent* to the original one. A simple example is the following implication problem:

$$a \perp\!\!\!\perp b \,|\, c \quad \stackrel{?}{\Rightarrow} \quad a \perp\!\!\!\perp b \,|\, \emptyset\,.$$

This is not a valid CI implication even for distributions with a strictly positive density despite the extended implication problem

$$a \perp\!\!\!\perp b \,|\, c,\; a \perp\!\!\!\perp c \,|\, \emptyset \quad\Rightarrow\quad a \perp\!\!\!\perp b \,|\, \emptyset$$

is a valid CI implication. The crucial argument to verify (5) is the linear relation (16) and the fact that the only bridging triplets between A and B in (5) are of the form $\langle A_i, B_i \,|\, C_i\rangle$.

4. The case of not necessarily positive densities

To cover the case of general discrete distributions such as the implication (21) is valid, we use the following lemma. The difference from Lemma 3 is that $D = N \setminus ABC$ does not appear here.

Lemma 5. *Let $A, B, C \subseteq N$, respectively $A_i, B_i, C_i \subseteq N$ for $i = 1, \ldots, I$, are triplets of pairwise disjoint sets. Given a collection of real numbers $\{\lambda_i\}_{i=1}^I$ the following conditions are equivalent:*

(a) *one has*

$$\delta_{ABC} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle} \in L_{\mathcal{P}(AC) \cup \mathcal{P}(BC)}\,,$$

(b) *for $\mathcal{K} = \mathcal{P}(N) \setminus \{\mathcal{P}(AC) \cup \mathcal{P}(BC)\}$, one has*

$$(\mathsf{u}_{\langle A,B \,|\, C \rangle} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle})|_{\mathcal{K}} = \mathbf{0}\,,$$

(c) *real numbers $\{\kappa_j\}_{j=1}^J$ and pairwise disjoint triplets $\{\langle E_j, F_j \,|\, G_j \rangle\}_{j=1}^J$ exist such that $E_j F_j G_j \in \mathcal{P}(AC) \cup \mathcal{P}(BC)$ for $j = 1, \ldots, J$, and*

$$\mathsf{u}_{\langle A,B \,|\, C \rangle} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle} + \sum_{j=1}^{J} \kappa_j \cdot \mathsf{u}_{\langle E_j, F_j \,|\, G_j \rangle} = \mathbf{0}\,.$$

Proof. The proof is completely analogous to the proof of Lemma 3. The only difference is that one has $\mathcal{L} \equiv \mathcal{P}(AC) \cup \mathcal{P}(BC)$ instead.

Let $A, B, C \subseteq N$, respectively $A_i, B_i, C_i \subseteq N$ for $i = 1, \ldots, I$, are non-trivial triplets of pairwise disjoint sets. Recall that we are dealing with the implication problem (5) for any discrete distribution P. Given a triplet $\langle A, B \,|\, C \rangle$ of pairwise disjoint sets we introduce a special notation

$$\mathsf{u}^{\dagger}_{\langle A,B \,|\, C \rangle} := \ \delta_{ABC} - \delta_{AC} - \delta_{BC}. \tag{22}$$

The symbol $[\mathsf{u}]^+$ will denote the non-negative part of $\mathsf{u} \in \mathbb{R}^{\mathcal{P}(N)}$. The main result of this section is as follows.

Theorem 2. *If there exist real numbers $\{\lambda_i\}_{i=1}^I$ such that*

$$\delta_{ABC} + \sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle} \in L_{\mathcal{P}(AC) \cup \mathcal{P}(BC)} \tag{23}$$

and non-negative real numbers $\{\zeta_i^{\dagger}\}_{i=1}^I$, that is, $\zeta_i^{\dagger} \geq 0$ for any i, such that,

$$\left[\delta_{ABC} - \sum_{i=1}^{I} \zeta_i^{\dagger} \cdot \mathsf{u}^{\dagger}_{\langle A_i, B_i \,|\, C_i \rangle}\right]^+ \in L_{\mathcal{P}(AC) \cup \mathcal{P}(BC)} \tag{24}$$

holds, then the implication (5) is true for any discrete distribution P over N.

Proof. The formula (23) implies there exist real numbers α_S for $S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)$ such that one has

$$\sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle} = -\delta_{ABC} + \sum_{S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)} \alpha_S \cdot \delta_S \,. \tag{25}$$

We are going to show, for any distribution P satisfying $\{A_i \perp\!\!\!\perp B_i \,|\, C_i \, [P]\}_{i=1}^{I}$, that

$$\forall\, x \in \mathcal{X} \text{ with } p(ABC;x) > 0 \qquad p(ABC;x) = \prod_{S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)} p(S;x)^{\alpha_S} \,, \tag{26}$$

$$\forall\, x \in \mathcal{X} \text{ with } p(ABC;x) = 0 \qquad \text{either } \; p(AC;x) = 0 \; \text{ or } \; p(BC;x) = 0 \,. \tag{27}$$

This is enough to verify $A \perp\!\!\!\perp B \,|\, C \, [P]$ by Lemma 1: take there $N = ABC$ and put

$$q(AC;x) \;=\; \begin{cases} \prod_{S \in \mathcal{P}(AC)} p(S;x)^{\alpha_S} & \text{if } p(AC;x) > 0, \\ \quad 0 & \text{if } p(AC;x) = 0, \end{cases} \qquad \text{and}$$

$$r(BC;x) \;=\; \begin{cases} \prod_{S \in \mathcal{P}(BC) \setminus \mathcal{P}(AC)} p(S;x)^{\alpha_S} & \text{if } p(BC;x) > 0, \\ \quad 0 & \text{if } p(BC;x) = 0. \end{cases}$$

To verify (26) for a fixed $x \in \mathcal{X}$ with $p(ABC;x) > 0$ we basically repeat the consideration from the proof of Theorem 1. First, we choose and fix $\tilde{x} \in \mathcal{X}$ such that $\tilde{x}_{ABC} = x_{ABC}$ and $p(N;\tilde{x}) > 0$, which, of course, may differ from x. Now, we are sure that $p(A_iB_iC_i;\tilde{x}) > 0$ for any $i \in \{1,\ldots,I\}$ and can write by $A_i \perp\!\!\!\perp B_i \,|\, C_i \, [P]$:

$$\left[\, \forall\, i \quad 1 = \frac{p(A_iB_iC_i;\tilde{x}) \cdot p(C_i;\tilde{x})}{p(A_iC_i;\tilde{x}) \cdot p(B_iC_i;\tilde{x})} \,\right] \quad \Longrightarrow \quad 1 = \prod_{i=1}^{I} \left(\frac{p(A_iB_iC_i;\tilde{x}) \cdot p(C_i;\tilde{x})}{p(A_iC_i;\tilde{x}) \cdot p(B_iC_i;\tilde{x})} \right)^{\lambda_i} .$$

The expression on the right-hand side there can be, by (25), re-written as

$$\begin{aligned} 1 \;&=\; \prod_{i=1}^{I} \left(\frac{p(A_iB_iC_i;\tilde{x}) \cdot p(C_i;\tilde{x})}{p(A_iC_i;\tilde{x}) \cdot p(B_iC_i;\tilde{x})} \right)^{\lambda_i} = \prod_{i=1}^{I} \left(\prod_{T \subseteq N} p(T;\tilde{x})^{\mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle}(T)} \right)^{\lambda_i} \\ &=\; \prod_{i=1}^{I} \prod_{T \subseteq N} p(T;\tilde{x})^{\lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle}(T)} = \prod_{T \subseteq N} p(T;\tilde{x})^{\sum_{i=1}^{I} \lambda_i \cdot \mathsf{u}_{\langle A_i, B_i \,|\, C_i \rangle}(T)} \\ &\overset{(25)}{=}\; \frac{1}{p(ABC;\tilde{x})} \cdot \prod_{S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)} p(S;\tilde{x})^{\alpha_S} = \frac{1}{p(ABC;x)} \cdot \prod_{S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)} p(S;x)^{\alpha_S} \,. \end{aligned}$$

Since the term $p(ABC;x)$ above is strictly positive, multiplying by $p(ABC;x)$ gives (26).

To verify (27) for a fixed $x \in \mathcal{X}$ with $p(ABC;x) = 0$ it is enough to find some set $S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)$ with $p(S;x) = 0$. Indeed, note that $p(S;x) = 0$ implies $p(AC;x) = 0$ or $p(BC;x) = 0$. Realize that (24) means $\delta_{ABC}(T) - \sum_{i=1}^{I} \zeta_i^{\dagger} \cdot \mathsf{u}^{\dagger}_{\langle A_i, B_i \,|\, C_i \rangle}(T) \leq 0$ for any

$T \subseteq N$ such that $T \notin \mathcal{P}(AC) \cup \mathcal{P}(BC)$. Hence, the non-negativity assumption $\zeta_i^\dagger \geq 0$ for any i and the definition (22) implies the following observation:

$$\begin{array}{ll} \text{if} & T = ABC \\ \text{or} & [\, T \subseteq N,\, T \notin \mathcal{P}(AC) \cup \mathcal{P}(BC) \;\;\&\;\; \exists\, j \;\; \zeta_j^\dagger > 0 \text{ with } \mathsf{u}^\dagger_{\langle A_j, B_j \,|\, C_j \rangle}(T) = -1 \,] \\ \text{then} & [\, \exists\, i \;\; \zeta_i^\dagger > 0 \;\;\&\;\; \mathsf{u}^\dagger_{\langle A_i, B_i \,|\, C_i \rangle}(T) = +1 \,]. \end{array}$$

This observation then makes possible the following inductive consideration. Take $T \subseteq N$ such that $T \notin \mathcal{P}(AC) \cup \mathcal{P}(BC)$ and $p(T;x) = 0$. Of course, the starting set $T \subseteq N$ will be $T \equiv ABC$. If i is such that $\zeta_i^\dagger > 0$ and $\mathsf{u}^\dagger_{\langle A_i, B_i \,|\, C_i \rangle}(T) = +1$ one has $T = A_i B_i C_i$ by (22). Since $A_i \perp\!\!\!\perp B_i \,|\, C_i \,[P]$, by (1), either $p(A_i C_i; x) = 0$ or $p(B_i C_i; x) = 0$. Make a choice: put either $S \equiv A_i C_i$ or $S \equiv B_i C_i$ so that we have $p(S;x) = 0$. If $S \in \mathcal{P}(AC) \cup \mathcal{P}(BC)$, then the goal is reached. If $S \notin \mathcal{P}(AC) \cup \mathcal{P}(BC)$, then $\zeta_i^\dagger > 0$ and $\mathsf{u}^\dagger_{\langle A_i, B_i \,|\, C_i \rangle}(S) = -1$ allows one to apply the above observation again, this time to S. Since $|S| < |T|$, the process has to stop at some point, which means one has to reach a set in $\mathcal{P}(AC) \cup \mathcal{P}(BC)$ in this way. This completes the proof.

To deal with the general discrete case we modified in Theorem 2 the condition (6) from Theorem 1: the right-hand side of (23) is $L_{\mathcal{P}(AC) \cup \mathcal{P}(BC)}$, not $L_{\mathcal{P}(ACD) \cup \mathcal{P}(BCD)}$ as in (6). The second difference is that we need an additional condition (24) to cover the case of the events of probability zero. The following example shows that both assumptions in Theorem 2 are needed.

Example 5. To show that (24) cannot be omitted consider the CI implication problem

$$a \perp\!\!\!\perp b \,|\, c,\;\; a \perp\!\!\!\perp c \,|\, b,\;\; b \perp\!\!\!\perp c \,|\, a \;\;\stackrel{?}{\Rightarrow}\;\; ab \perp\!\!\!\perp c \,|\, \emptyset\,, \tag{28}$$

and observe that (23) holds in this case $N = \{a, b, c\}$ because

$$\delta_{abc} + \mathsf{u}_{\langle a,b \,|\, c \rangle} - \mathsf{u}_{\langle a,c \,|\, b \rangle} - \mathsf{u}_{\langle b,c \,|\, a \rangle} \in L_{\mathcal{P}(ab) \cup \mathcal{P}(c)}\,.$$

However, there exists a discrete distribution such that (28) is not valid for it. To this end put $\mathcal{X}_i = \{0, 1\}$ for any $i \in N$ and assign the value $1/2$ to the configurations $[\,0,0,0\,]$, $[\,1,1,1\,]$ and 0 to the remaining ones. One can check that the condition (24) does not hold in this case. To show that (23) cannot be omitted consider another implication problem

$$a \perp\!\!\!\perp b \,|\, cd,\;\; c \perp\!\!\!\perp d \,|\, a,\;\; c \perp\!\!\!\perp d \,|\, b \;\;\stackrel{?}{\Rightarrow}\;\; c \perp\!\!\!\perp d \,|\, ab\,, \tag{29}$$

and observe that (24) holds in this case for $N = \{a, b, c, d\}$ because

$$(\, \delta_{abcd} - \mathsf{u}^\dagger_{\langle a,b \,|\, cd \rangle} - \mathsf{u}^\dagger_{\langle c,d \,|\, a \rangle} - \mathsf{u}^\dagger_{\langle c,d \,|\, b \rangle} \,)|_{\mathcal{K}} = \mathbf{0} \qquad \text{for } \mathcal{K} = \{\, cd,\, acd,\, bcd,\, abcd \,\}. \tag{30}$$

On the other hand, (29) is not probabilistically valid: put $N = \{a, b, c, d\}$, $\mathcal{X}_i = \{0, 1\}$ for any $i \in N$ and assign non-zero density values to the following 6 configurations $[x_a, x_b, x_c, x_d]$:

$$[\,0,0,0,0\,] \mapsto 1/6\,, \quad [\,1,0,0,0\,] \mapsto 1/6\,, \quad [\,0,0,0,1\,] \mapsto 1/6\,,$$

$$[\,0,0,1,0\,] \mapsto 1/4\,, \quad [\,0,0,1,1\,] \mapsto 1/8\,, \quad [\,0,1,1,1\,] \mapsto 1/8\,.$$

In particular, (23) does not hold in this case.

To illustrate Theorem 2 we show that it can be applied to a formerly mentioned CI implication problem (21).

Example 6. We are going to show that (21), that is,

$$a \perp\!\!\!\perp b\,|\,cd,\ c \perp\!\!\!\perp d\,|\,a,\ c \perp\!\!\!\perp d\,|\,b,\ c \perp\!\!\!\perp d\,|\,\emptyset \quad \Rightarrow \quad c \perp\!\!\!\perp d\,|\,ab,$$

holds for any discrete distribution. Let us re-write the linear equality (19) in the form

$$\mathsf{u}_{\langle c,d\,|\,ab\rangle} + \left[\,-\mathsf{u}_{\langle a,b\,|\,cd\rangle} - \mathsf{u}_{\langle c,d\,|\,a\rangle} - \mathsf{u}_{\langle c,d\,|\,b\rangle} + \mathsf{u}_{\langle c,d\,|\,\emptyset\rangle}\,\right] + \left\{\,\mathsf{u}_{\langle a,b\,|\,c\rangle} + \mathsf{u}_{\langle a,b\,|\,d\rangle} - \mathsf{u}_{\langle a,b\,|\,\emptyset\rangle}\,\right\} = \mathbf{0},$$

which is nothing but the condition (c) in Lemma 5, an equivalent version of the condition (23) for (21), that is,

$$(\,\delta_{abcd} - \mathsf{u}_{\langle a,b\,|\,cd\rangle} - \mathsf{u}_{\langle c,d\,|\,a\rangle} - \mathsf{u}_{\langle c,d\,|\,b\rangle} + \mathsf{u}_{\langle c,d\,|\,\emptyset\rangle}\,)|_{\mathcal{K}} = \mathbf{0} \qquad \text{for } \mathcal{K} = \{\,cd,\,acd,\,bcd,\,abcd\,\}.$$

In this special case, condition (24) also holds. Actually, it can be obtained by minor modification of (23): it is enough to put $\zeta_i^\dagger = [\lambda_i]^-$, where $[\lambda]^-$ denotes the non-positive part of $\lambda \in \mathbb{R}$, and observe that the respective above relation (30) is valid, too.

Remark 3. Note that our new condition from Theorem 2 is neither stronger nor weaker than the one used by Bouckaert et al. [2]. For example, the decomposition rule (14), that is, $a \perp\!\!\!\perp bc\,|\,\emptyset \ \Rightarrow\ a \perp\!\!\!\perp c\,|\,\emptyset$, is a valid CI implication in the general discrete case and can be verified by the tools from Bouckaert et al. [2]. However, the validity of (14) for general discrete distributions cannot be derived using Theorem 2, even if its repeated application is allowed. Indeed, one can derive the weak union implications $a \perp\!\!\!\perp bc\,|\,\emptyset \ \Rightarrow\ a \perp\!\!\!\perp b\,|\,c$ and $a \perp\!\!\!\perp bc\,|\,\emptyset \ \Rightarrow\ a \perp\!\!\!\perp c\,|\,b$ in this way (cf. Example 1), but if one tries to verify

$$a \perp\!\!\!\perp bc\,|\,\emptyset,\ a \perp\!\!\!\perp b\,|\,c,\ a \perp\!\!\!\perp c\,|\,b \quad \Rightarrow \quad a \perp\!\!\!\perp c\,|\,\emptyset\,,$$

by means of Theorem 2, then the respective condition (24) is not fulfilled in this case. On the other hand, the implication (21) cannot be verified by the method in Bouckaert et al. [2] (see Example 4.1 of Studený [14]), while it can be derived by our new method as shown in Example 6. Hence, both methods have their strong and weak points.

5. A computational example

In this section, we present an example to demonstrate the methods described in the paper. The following CI implication has been found as the result of computational search experiments performed by the first author:

$$\begin{array}{c} c \perp\!\!\!\perp d\,|\,abef,\ a \perp\!\!\!\perp f\,|\,bde,\ b \perp\!\!\!\perp e\,|\,acf, \\ e \perp\!\!\!\perp f\,|\,ac,\ e \perp\!\!\!\perp f\,|\,bd,\ e \perp\!\!\!\perp f\,|\,ab \end{array} \Rightarrow e \perp\!\!\!\perp f\,|\,abcd\,. \tag{31}$$

We found this example by a random search combined with some heuristics. In this search, the set of CI statements in the left-hand side of (31) is given by adding CI statements which may factorize a density of a, b, c, d, e, f to functions of a, b, c, d, e and a, b, c, d, f. The above example involves six variables: $N = \{a, b, c, d, e, f\}$. While the CI structures in case $|N| \leq 5$ have been studied in detail in Studený et al. [16] and Hemmecke et al. [6], only a little bit is known about the case $|N| \geq 6$.

As we show below, the implication (31) is true. On the other hand, the method presented by Bouckaert et al. [2] does not allow one to verify (31). Furthermore, using the algorithm by Baioletti et al. [1], one can observe that (31) is not implied by the *graphoid properties*, which are well-known valid CI implications in the case of strictly positive distributions.

We consider the application of the method of Theorem 2. In case of (31), the class $\mathcal{K}$ from Lemma 5(b) consists of the supersets of ef and the condition (23) is fulfilled because

$$(\,\delta_{abcdef} - \mathsf{u}_{\langle c,d\,|\,abef\rangle} - \mathsf{u}_{\langle a,f\,|\,bde\rangle} - \mathsf{u}_{\langle b,e\,|\,acf\rangle} - \mathsf{u}_{\langle e,f\,|\,ac\rangle} - \mathsf{u}_{\langle e,f\,|\,bd\rangle} + \mathsf{u}_{\langle e,f\,|\,ab\rangle}\,)|_{\mathcal{K}} = \mathbf{0}. \quad (32)$$

In particular, the implication (31) holds for any discrete distribution with a strictly positive density. Nevertheless, one can also verify the condition (24) in this case, specifically:

$$(\,\delta_{abcdef} - \mathsf{u}^{\dagger}_{\langle c,d\,|\,abef\rangle} - \mathsf{u}^{\dagger}_{\langle a,f\,|\,bde\rangle} - \mathsf{u}^{\dagger}_{\langle b,e\,|\,acf\rangle} - \mathsf{u}^{\dagger}_{\langle e,f\,|\,ac\rangle} - \mathsf{u}^{\dagger}_{\langle e,f\,|\,bd\rangle}\,)|_{\mathcal{K}} = \mathbf{0}. \qquad (33)$$

Therefore, by Theorem 2, the implication (31) is valid for any discrete distribution.

Remark 4. We can also verify (32) by means of Lemma 5(c). Specifically, one can decompose a multiple by 10 of the respective imset as follows:

$$\begin{aligned}
&10 \cdot \mathsf{u}_{\langle e,f\,|\,abcd\rangle} \\
&\quad +[-10 \cdot \mathsf{u}_{\langle c,d\,|\,abef\rangle} - 10 \cdot \mathsf{u}_{\langle a,f\,|\,bde\rangle} - 10 \cdot \mathsf{u}_{\langle b,e\,|\,acf\rangle} \\
&\qquad - 10 \cdot \mathsf{u}_{\langle e,f\,|\,ac\rangle} - 10 \cdot \mathsf{u}_{\langle e,f\,|\,bd\rangle} + 10 \cdot \mathsf{u}_{\langle e,f\,|\,ab\rangle}] \\
&\quad + \{\, 2 \cdot \mathsf{u}_{\langle a,b\,|\,\emptyset\rangle} + 5 \cdot \mathsf{u}_{\langle c,d\,|\,\emptyset\rangle} + \mathsf{u}_{\langle c,f\,|\,\emptyset\rangle} + 3 \cdot \mathsf{u}_{\langle c,e\,|\,ab\rangle} \\
&\qquad + 7 \cdot \mathsf{u}_{\langle a,d\,|\,c\rangle} + 9 \cdot \mathsf{u}_{\langle b,d\,|\,ac\rangle} + 7 \cdot \mathsf{u}_{\langle b,e\,|\,ac\rangle} + \mathsf{u}_{\langle d,f\,|\,bc\rangle} + 10 \cdot \mathsf{u}_{\langle d,e\,|\,abc\rangle} + 9 \cdot \mathsf{u}_{\langle d,f\,|\,abc\rangle} \\
&\qquad + 6 \cdot \mathsf{u}_{\langle a,e\,|\,d\rangle} + 5 \cdot \mathsf{u}_{\langle b,e\,|\,ad\rangle} + 10 \cdot \mathsf{u}_{\langle a,f\,|\,bd\rangle} + \mathsf{u}_{\langle b,f\,|\,cd\rangle} + 5 \cdot \mathsf{u}_{\langle b,c\,|\,e\rangle} + 3 \cdot \mathsf{u}_{\langle c,d\,|\,ae\rangle} \\
&\qquad + 2 \cdot \mathsf{u}_{\langle c,d\,|\,be\rangle} + \mathsf{u}_{\langle a,d\,|\,bf\rangle} + \mathsf{u}_{\langle a,c\,|\,bdf\rangle} \\
&\qquad - 3 \cdot \mathsf{u}_{\langle a,e\,|\,\emptyset\rangle} - 4 \cdot \mathsf{u}_{\langle b,c\,|\,\emptyset\rangle} - \mathsf{u}_{\langle b,f\,|\,\emptyset\rangle} - \mathsf{u}_{\langle b,c\,|\,a\rangle} - 3 \cdot \mathsf{u}_{\langle c,e\,|\,b\rangle} - \mathsf{u}_{\langle c,f\,|\,ab\rangle} \\
&\qquad - 5 \cdot \mathsf{u}_{\langle d,e\,|\,ab\rangle} - 10 \cdot \mathsf{u}_{\langle d,f\,|\,ab\rangle} - 5 \cdot \mathsf{u}_{\langle d,e\,|\,c\rangle} - \mathsf{u}_{\langle a,c\,|\,d\rangle} - 3 \cdot \mathsf{u}_{\langle a,e\,|\,cd\rangle} \\
&\qquad - 2 \cdot \mathsf{u}_{\langle b,e\,|\,cd\rangle} - 8 \cdot \mathsf{u}_{\langle b,d\,|\,e\rangle} - 8 \cdot \mathsf{u}_{\langle a,d\,|\,be\rangle} - 2 \cdot \mathsf{u}_{\langle a,b\,|\,de\rangle} - \mathsf{u}_{\langle b,d\,|\,cf\rangle} \,\} = \mathbf{0}, \qquad (34)
\end{aligned}$$

where the imsets in the curly brackets in the above equation (34) correspond to additional conditional independence statements not bridging between e and f. Furthermore, from (34) we have

$$\begin{aligned}
&10 \cdot \mathsf{u}_{\langle c,d\,|\,abef\rangle} + 10 \cdot \mathsf{u}_{\langle a,f\,|\,bde\rangle} + 10 \cdot \mathsf{u}_{\langle b,e\,|\,acf\rangle} + 10 \cdot \mathsf{u}_{\langle e,f\,|\,ac\rangle} + 10 \cdot \mathsf{u}_{\langle e,f\,|\,bd\rangle} \\
&\qquad + 3 \cdot \mathsf{u}_{\langle a,e\,|\,\emptyset\rangle} + 4 \cdot \mathsf{u}_{\langle b,c\,|\,\emptyset\rangle} + \mathsf{u}_{\langle b,f\,|\,\emptyset\rangle} + \mathsf{u}_{\langle b,c\,|\,a\rangle} + 3 \cdot \mathsf{u}_{\langle c,e\,|\,b\rangle} + \mathsf{u}_{\langle c,f\,|\,ab\rangle}
\end{aligned}$$

$$
\begin{aligned}
&+ 5\cdot \mathsf{u}_{\langle d,e\,|\,ab\rangle} + 10\cdot \mathsf{u}_{\langle d,f\,|\,ab\rangle} + 5\cdot \mathsf{u}_{\langle d,e\,|\,c\rangle} + \mathsf{u}_{\langle a,c\,|\,d\rangle} + 3\cdot \mathsf{u}_{\langle a,e\,|\,cd\rangle}\\
&+ 2\cdot \mathsf{u}_{\langle b,e\,|\,cd\rangle} + 8\cdot \mathsf{u}_{\langle b,d\,|\,e\rangle} + 8\cdot \mathsf{u}_{\langle a,d\,|\,be\rangle} + 2\cdot \mathsf{u}_{\langle a,b\,|\,de\rangle} + \mathsf{u}_{\langle b,d\,|\,cf\rangle}\\
=\;& 10\cdot \mathsf{u}_{\langle e,f\,|\,abcd\rangle} + 10\cdot \mathsf{u}_{\langle e,f\,|\,ab\rangle} + 2\cdot \mathsf{u}_{\langle a,b\,|\,\emptyset\rangle} + 5\cdot \mathsf{u}_{\langle c,d\,|\,\emptyset\rangle} + \mathsf{u}_{\langle c,f\,|\,\emptyset\rangle} + 3\cdot \mathsf{u}_{\langle c,e\,|\,ab\rangle}\\
&+ 7\cdot \mathsf{u}_{\langle a,d\,|\,c\rangle} + 9\cdot \mathsf{u}_{\langle b,d\,|\,ac\rangle} + 7\cdot \mathsf{u}_{\langle b,e\,|\,ac\rangle} + \mathsf{u}_{\langle d,f\,|\,bc\rangle} + 10\cdot \mathsf{u}_{\langle d,e\,|\,abc\rangle} + 9\cdot \mathsf{u}_{\langle d,f\,|\,abc\rangle}\\
&+ 6\cdot \mathsf{u}_{\langle a,e\,|\,d\rangle} + 5\cdot \mathsf{u}_{\langle b,e\,|\,ad\rangle} + 10\cdot \mathsf{u}_{\langle a,f\,|\,bd\rangle} + \mathsf{u}_{\langle b,f\,|\,cd\rangle} + 5\cdot \mathsf{u}_{\langle b,c\,|\,e\rangle} + 3\cdot \mathsf{u}_{\langle c,d\,|\,ae\rangle}\\
&+ 2\cdot \mathsf{u}_{\langle c,d\,|\,be\rangle} + \mathsf{u}_{\langle a,d\,|\,bf\rangle} + \mathsf{u}_{\langle a,c\,|\,bdf\rangle}\,. \qquad (35)
\end{aligned}
$$

From (35) and the properties of structural imsets (Studený [15], Hemmecke et al. [6]), we obtain the following implication as a byproduct of the result.

$$
\begin{array}{c}
c \perp\!\!\!\perp d \,|\, abef,\\
a \perp\!\!\!\perp f \,|\, bde,\; b \perp\!\!\!\perp e \,|\, acf,\; e \perp\!\!\!\perp f \,|\, ac,\\
e \perp\!\!\!\perp f \,|\, bd,\; a \perp\!\!\!\perp e \,|\, \emptyset,\; b \perp\!\!\!\perp c \,|\, \emptyset,\; b \perp\!\!\!\perp f \,|\, \emptyset,\\
b \perp\!\!\!\perp c \,|\, a,\; c \perp\!\!\!\perp e \,|\, b,\\
c \perp\!\!\!\perp f \,|\, ab,\; d \perp\!\!\!\perp e \,|\, ab,\\
d \perp\!\!\!\perp f \,|\, ab,\; d \perp\!\!\!\perp e \,|\, c,\; a \perp\!\!\!\perp c \,|\, d,\; a \perp\!\!\!\perp e \,|\, cd,\\
b \perp\!\!\!\perp e \,|\, cd,\; b \perp\!\!\!\perp d \,|\, e,\; a \perp\!\!\!\perp d \,|\, be,\; a \perp\!\!\!\perp b \,|\, de,\\
b \perp\!\!\!\perp d \,|\, cf
\end{array}
\quad\Leftrightarrow\quad
\begin{array}{c}
e \perp\!\!\!\perp f \,|\, abcd,\\
e \perp\!\!\!\perp f \,|\, ab,\; a \perp\!\!\!\perp b \,|\, \emptyset,\; c \perp\!\!\!\perp d \,|\, \emptyset,\\
c \perp\!\!\!\perp f \,|\, \emptyset,\; c \perp\!\!\!\perp e \,|\, ab,\; a \perp\!\!\!\perp d \,|\, c,\; b \perp\!\!\!\perp d \,|\, ac,\\
b \perp\!\!\!\perp e \,|\, ac,\; d \perp\!\!\!\perp f \,|\, bc,\\
d \perp\!\!\!\perp e \,|\, abc,\; d \perp\!\!\!\perp f \,|\, abc,\\
a \perp\!\!\!\perp e \,|\, d,\; b \perp\!\!\!\perp e \,|\, ad,\; a \perp\!\!\!\perp f \,|\, bd,\; b \perp\!\!\!\perp f \,|\, cd,\\
b \perp\!\!\!\perp c \,|\, e,\; c \perp\!\!\!\perp d \,|\, ae,\; c \perp\!\!\!\perp d \,|\, be,\; a \perp\!\!\!\perp d \,|\, bf,\\
a \perp\!\!\!\perp c \,|\, bdf.
\end{array}
$$

Conclusions

Let us summarize the contributions of this note. We have proposed a new linear-algebraic method for derivation of (probabilistic) CI implications. The method mainly applies in the case of strictly positive discrete distributions, but it is also extended to the general case of discrete distributions. The method, which goes beyond the formerly known methods, has been illustrated by a few examples. The most complicated one has been obtained as the result of computational experiments.

The reader familiar with algebraic statistics knows that CI implication tasks can often be re-formulated in terms of (ideals of) polynomial rings. For example, the *conditional independence ideal* defined in § 3.1 of Drton et al. [3] consists of polynomials whose indeterminates correspond to configurations in the (fixed) joint sample space. Another approach is applied in § 5 of Hemmecke et al. [6], where the respective *toric ideal* consists of polynomials whose indeterminates correspond to elementary CI statements $i \perp\!\!\!\perp j \,|\, K$. The elements of the *Markov basis* for that toric ideal seem to correspond to CI implications that can be derived by the method of structural imsets described in Bouckaert et al. [2]. Hemmecke et al. [6] obtained 75,889 instances of CI implications (= elements of a minimal Markov basis) for $|N| = 5$, which decompose into 1,381 permutation equivalence classes (Kashimura et al. [7]).

The reader may wonder whether the condition (b), respectively (c), of Lemma 3 can also be modelled/represented in terms of polynomials, that is, by means of binomial relations. Perhaps there is a way to do that if one somehow considers Laurent polynomials whose indeterminates correspond to subsets of the set of variables N. To illustrate this rough idea, consider the CI problem (13) from Example 2 and assume that $t_{abc}, t_{bc}, t_{ac}, t_{ab}, t_c, t_b, t_a, t_\emptyset$ are the indeterminates. Then the inputs $a \perp\!\!\!\perp b \,|\, c$, $a \perp\!\!\!\perp c \,|\, b$ and the output $a \perp\!\!\!\perp c \,|\, \emptyset$ are represented as $z_{\langle a,b\,|\,c\rangle} \equiv t_{abc} t_c t_{ac}^{-1} t_{bc}^{-1}$, $z_{\langle a,c\,|\,b\rangle} \equiv t_{abc} t_b t_{ab}^{-1} t_{bc}^{-1}$ and $z_{\langle a,c\,|\,\emptyset\rangle} \equiv t_{ac} t_\emptyset t_a^{-1} t_c^{-1}$, respectively. One can model the restriction to $\mathcal{K} = \{ac, abc\}$ in the condition (b) of Lemma 3 by settings $t_{bc} = t_{ab} = t_c = t_b = t_a = t_\emptyset = 1$. Thus, one has $z_{\langle a,b\,|\,c\rangle}|_{\mathcal{K}} = t_{abc} t_{ac}^{-1}$, $z_{\langle a,c\,|\,b\rangle}|_{\mathcal{K}} = t_{abc}$ and $z_{\langle a,c\,|\,\emptyset\rangle}|_{\mathcal{K}} = t_{ac}$. Then one can introduce a binomial relation $(z_{\langle a,b\,|\,c\rangle}|_{\mathcal{K}})^{-1} \cdot z_{\langle a,c\,|\,b\rangle}|_{\mathcal{K}} - z_{\langle a,c\,|\,\emptyset\rangle}|_{\mathcal{K}} = 0$, which can, perhaps, be viewed as a kind of translation/interpretation of the condition (b) from Lemma 3 in the world of polynomials.

Acknowledgments

The research of Kentaro Tanaka has been supported by JSPS KAKENHI Grant Number 24700274. The research of Milan Studený has been supported from the grant project GAČR number 13-20012S. The research of Akimichi Takemura has been supported by JSPS KAKENHI Grant Number 25220001.

References

[1] Marco Baioletti, Giuseppe Busanello, and Barbara Vantaggi. Conditional independence structure and its closure: Inferential rules and algorithms. *International Journal of Approximate Reasoning*, 50:1097–1114, 2009.

[2] Remco Bouckaert, Raymond Hemmecke, Silvia Lindner, and Milan Studený. Efficient algorithms for conditional independence inference. *Journal of Machine Learning Research*, 11:3453–3479, 2010.

[3] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics. Oberwolfach Seminars 39.* Birkhäuser, Basel, 2009.

[4] Dan Geiger, Azaria Paz, and Judea Pearl. Axioms and algorithms for inferences involving probabilistic independence. *Information and Computation*, 91:128–141, 1991.

[5] Dan Geiger and Judea Pearl. Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics*, 21:2001–2021, 1993.

[6] Raymond Hemmecke, Jason Morton, Anne Shiu, Bernd Sturmfels, and Oliver Wienand. Three counter-examples on semi-graphoids. *Combinatorics, Probability and Computing*, 17:239–257, 2008.

[7] Takuya Kashimura, Tomonari Sei, Akimichi Takemura, and Kentaro Tanaka. Cones of elementary imsets and supermodular functions: a review and some new results. In *Proceedings of the Second CREST-SBM International Conference "Harmony of Gröbner Bases and the Modern Industrial Society"*, pages 117–152, 2012.

[8] Francesco M. Malvestuto. A unique formal system for binary decomposition of database relations, probability distributions and graphs. *Information Sciences*, 59:21–52, 1992.

[9] Francesco M. Malvestuto and Milan Studený. Comment on "a unique formal system for binary decomposition of database relations, probability distributions and graphs". *Information Sciences*, 63:1–2, 1992.

[10] František Matúš. Stochastic independence, algebraic independence and abstract connectedness. *Theoretical Computer Science*, 134:455–471, 1994.

[11] Mathias Niepert, Dirk Van Gucht, and Marc Gyssens. Logical and algorithmic properties of stable conditional independence. *International Journal of Approximate Reasoning*, 51:531–543, 2010.

[12] Judea Pearl and Azaria Paz. Graphoids, graph-based logic for reasoning about relevance relations. In L. Steels B. Du Boulay, D. Hogg, editor, *In Advances in Artificial Intelligence, Volume II*, pages 357–363, North-Holland, 1987.

[13] Wolfgang Spohn. On the properties of conditional independence. In Paul Humphreys, editor, *Patrick Suppes, Scientific Philosopher, Volume 1, Probability and Probabilistic Causality*, pages 173–196. Kluwer, 1994.

[14] Milan Studený. Conditional independence relations have no finite complete characterization. In J. Á. Víšek S. Kubík, editor, *Information Theory, Statistical Decision Functions and Random Processes, Transactions of the 11th Prague Conference, Volume B*, pages 377–396. Kluwer, 1992.

[15] Milan Studený. *Probabilistic Conditional Independence Structures.* Springer-Verlag, London, 2005.

[16] Milan Studený, Remco R. Bouckaert, and Tomáš Kočka. Extreme supermodular set functions over five variables. Research report 1977, Institute of Information Theory and Automation, Prague, 2000.