KYBERNETIKA — VOLUME *56* (2020), NUMBER 5, PAGES 850–874

CONTRIBUTION OF FRANTIŠEK MATÚŠ TO THE RESEARCH ON CONDITIONAL INDEPENDENCE

Milan Studený

An overview is given of results achieved by F. Matúš on probabilistic conditional independence (CI). First, his axiomatic characterizations of stochastic functional dependence and unconditional independence are recalled. Then his elegant proof of discrete probabilistic representability of a matroid based on its linear representability over a finite field is recalled. It is explained that this result was a basis of his methodology for constructing a probabilistic representation of a given abstract CI structure. His embedding of matroids into (augmented) abstract CI structures is recalled and his contribution to the theory of semigraphoids is mentioned as well. Finally, his results on the characterization of probabilistic CI structures induced by four discrete random variables and by four regular Gaussian random variables are recalled. Partial probabilistic representability by binary random variables is also mentioned.

Keywords: conditional independence, matroid, polymatroid, entropy function, semigraphoid, semimatroid

Classification: 62H05, 05B35, 68T30

1. INTRODUCTION

In the 1930s, *matroid theory* was introduced as an abstract theory of independence [41]. The classic matroid theory was inspired by the independence concepts appearing in linear algebra and graph theory; but analogous abstract independence structures were later recognized in other areas of mathematics, for example, in projective geometries, abstract algebra, and lattice theory; see [33] for details.

The concept of independence of random variables is inherent in probability theory, too. In fact, a more general concept of *conditional independence* (CI) has been studied there for many years; see [12, § 25.3]. The first attempts to formulate abstract properties of probabilistic CI occurred in the late 1970s [5, 36]. This abstract point of view was later emphasized in connection with probabilistic graphical models [34, 11]. The idea was to use graphs to depict mutual relationships among random variables and, in this context, some analogy was recognized between abstract properties of probabilistic CI and those of separation concept(s) in graphs. This led to introducing a concept of a *semigraphoid* [34], which can be viewed as an abstract model of CI structure; the same is true for an even more general concept of a *separoid*, introduced later [6].

DOI: 10.14736/kyb-2020-5-0850

Several open mathematical problems related to that concept became a topic of research for František ("Fero") Matúš and the author of this paper in the 1990s. One of those problems was whether probabilistic CI structures and certain substructures of these can be described in an axiomatic way. It was shown that discrete probabilistic CI structures cannot be described in this way [37] but some of their sensible substructures can be [16, 19].

This paper offers an overview of some results from Fero's papers on the topic of probabilistic CI. They were written in the period from the 1990s to time of his death. It is, of course, a limited choice from a variety of Fero's results on this topic.

The structure of this paper is as follows. In Section 2, basic concepts and facts are recalled. In Section 3, Fero's early results on axiomatic characterizations of certain substructures of probabilistic CI structures are reviewed. Section 4 deals with the relationship between matroids and CI structures. An elegant proof of Fero's basic result about probabilistic representability of a linear matroid is a starting point. His embedding of matroids into (augmented) abstract CI structures, as well as his methodology for constructing probabilistic representations of abstract CI structures are then recalled. Semigraphoids and Fero's results on this topic are discussed in Section 5. Fero's famous complete characterization of probabilistic CI structures over four discrete random variables, which was a result of an enormous effort of his, is reported in Section 6; the regular Gaussian case is also mentioned there. Finally, selected additional results concerning graphical models of CI structure and the binary case are mentioned in Section 7.

2. BASIC CONCEPTS

In this section, basic concepts, some notation and terminology are introduced. Throughout the paper, the abbreviation **CI** will stand for *conditional independence.*

2.1. Random vector over a basic set

By a (general) *random variable* will be meant any measurable function ξ from a probability space $(\Omega, \mathcal{A}, P)$ to a measurable space $(\mathsf{X}, \mathcal{B})$; the σ-algebra generated by ξ is $\{\xi^{-1}(\mathsf{Y}) \,:\, \mathsf{Y} \in \mathcal{B}\} \subseteq \mathcal{A}$. The random variable ξ is called *real* if $\mathsf{X} = \mathbb{R}$ is the set of all real numbers and $\mathcal{B}$ the σ-algebra of Borel subsets of $\mathbb{R}$. It is called *discrete* if $|\mathsf{X}| < \infty$ and $\mathcal{B} = \mathcal{P}(\mathsf{X}) := \{\mathsf{Y} \,:\, \mathsf{Y} \subseteq \mathsf{X}\}$ is the power set of X and *binary*, if, moreover, $|\mathsf{X}| = 2$.

Random variables are indexed by elements of a finite non-empty index set N, called the *basic set* (of variables). Shortened notation for some subsets of N will often be used in the paper: union of two subsets $I, J \subseteq N$ may alternatively be denoted by a juxtaposition of respective symbols: $IJ \equiv I \cup J$. Analogously, the symbol for an element $i \in N$ may also stand for the respective singleton subset of N: $i \equiv \{i\}$.

A *random vector* over N is an indexed collection $\boldsymbol{\xi} = [\xi_i]_{i \in N}$ of random variables, each ξ_i taking values in its individual measurable space X_i, being necessarily non-empty. A random vector $\boldsymbol{\xi}$ is named real/discrete/binary if each random variable ξ_i in it is real/discrete/binary. Its distribution $P_{\boldsymbol{\xi}}$ is a multidimensional probability measure on the *joint sample space* $\prod_{i \in N} \mathsf{X}_i$, equipped with the respective product σ-algebra; it is none other than the measure P transformed by $\boldsymbol{\xi}$: $P_{\boldsymbol{\xi}}(\mathsf{Y}) := P(\{\omega \in \Omega \,:\, \boldsymbol{\xi}(\omega) \in \mathsf{Y}\})$ for any product-measurable set $\mathsf{Y} \subseteq \prod_{i \in N} \mathsf{X}_i$.

A probabilistic CI structure induced by a random vector $\boldsymbol{\xi}$ is a discrete mathematical structure describing certain stochastic (in)dependence relations among subvectors of $\boldsymbol{\xi}$. Specifically, given a non-empty set $\emptyset \neq I \subseteq N$, the symbol $\xi_I := [\xi_i]_{i\in I}$ will denote the respective *random subvector* of $\boldsymbol{\xi}$; thus, $\xi_N \equiv \boldsymbol{\xi}$ and $\xi_{\{i\}} \equiv \xi_i$ for $i \in N$. All these subvectors are random variables in the general sense mentioned above, ξ_I taking values in $\mathsf{X}_I := \prod_{i\in I} \mathsf{X}_i$. Note that if one denotes by $\mathcal{A}_I$ the σ-algebra generated by ξ_I then $\emptyset \neq J \subseteq I \subseteq N$ implies $\mathcal{A}_J \subseteq \mathcal{A}_I \subseteq \mathcal{A}_N \subseteq \mathcal{A}$.

It is convenient to introduce a random variable representing the empty set: we accept a convention that $\xi_\emptyset$ is a constant function on Ω, representing a kind of trivial random variable in the general sense (in fact, a “non-random” variable). Such a convention is consistent with the above-described assignment $\xi_I \mapsto \mathcal{A}_I$ of generating σ-algebras to random subvectors because $\mathcal{A}_\emptyset = \{\emptyset, \Omega\}$ is a trivial algebra satisfying $\mathcal{A}_\emptyset \subseteq \mathcal{A}_I$ for any $I \subseteq N$.

From the point of view of our study, the particular values of $\boldsymbol{\xi}$, that is, the labels for elements in X_N, play an unimportant role. The substantial structural information is fully encoded in the mutual relationships among the σ-algebras $\mathcal{A}_I$, $I \subseteq N$, and, of course, in the values of probability measure P on $\mathcal{A}_N \subseteq \mathcal{A}$.

Another basic observation is that the joint sample space X_N can always be considered in place of Ω, the product σ-algebra $\mathcal{B}_N$ in place of $\mathcal{A}$ and the joint distribution $P_{\boldsymbol{\xi}}$ in place of P. Under this arrangement, every random variable ξ_i, $i \in N$, turns into a coordinate projection to X_i, and every σ-algebra $\mathcal{A}_I$, $I \subseteq N$, is none other than the respective coordinate sub-σ-algebra $\mathcal{B}_I$ of $\mathcal{B}_N$, with $\mathcal{B}_\emptyset := \{\emptyset, \mathsf{X}_N\}$. The CI structure induced by $\boldsymbol{\xi}$ is, therefore, defined solely in terms of the joint distribution $P_{\boldsymbol{\xi}}$.

2.2. Probabilistic conditional independence

In this paper, a general definition of probabilistic CI in terms of σ-algebras is omitted. Such a definition requires a technical auxiliary concept of conditional probability given a σ-algebra; the interested reader can find details in [40, Appendices A.6.4 and A.7]. For this paper, it is enough to recall the definition in the discrete case.

The distribution $P_{\boldsymbol{\xi}}$ of a discrete random vector $\boldsymbol{\xi}$ over N can be fully described by its joint density, which is a non-negative function on the joint sample space X_N whose values sum up to 1. Subvectors' distributions are then described by marginal densities obtainable from the joint density by summing. Formally, these are functions on the marginal sample spaces X_I, $\emptyset \neq I \subseteq N$, but it appears to be convenient to understand them as functions defined on the joint sample space X_N and depending only on components indexed by I.

Specifically, given a random vector $\boldsymbol{\xi}$ over N and $I \subseteq N$, the respective *marginal density* $p_I : \mathsf{X}_N \to [0,1]$ can be introduced by the formula

$$p_I(x) := P(\{\omega \in \Omega \,:\, \forall\, i \in I \quad \xi_i(\omega) = x_i\,\}) \qquad \text{for every } x = [x_i]_{i\in N} \in \mathsf{X}_N.$$

Then $p := p_N$ is the *joint density* of $\boldsymbol{\xi}$. Note that we have also introduced the marginal density $p_\emptyset$ for the empty set: it is a constant function on X_N taking the value 1.

Having three subsets $I, J, K \subseteq N$ of the basic set, we say that ξ_I is *conditionally independent* of ξ_J *given* ξ_K and write $\xi_I \perp\!\!\!\perp \xi_J | \xi_K$ if

$$\forall\, x \in \mathsf{X}_N \qquad p_{I\cup J\cup K}(x) \cdot p_K(x) = p_{I\cup K}(x) \cdot p_{J\cup K}(x)\,. \tag{1}$$

An alternative notation is $I \perp\!\!\!\perp J \,|\, K$ $[\boldsymbol{\xi}]$. Note that the definition (1) works for any triplet of sets I, J, K although these three sets are typically assumed to be pairwise disjoint. The case $K = \emptyset$ then corresponds to the classic (unconditional) stochastic independence, denoted by $\xi_I \perp\!\!\!\perp \xi_J$, alternatively by $I \perp\!\!\!\perp J$ $[\boldsymbol{\xi}]$.

A similar definition of CI works in the (marginally) continuous case, that is, in the case of a real random vector $\boldsymbol{\xi}$ whose distribution $P_{\boldsymbol{\xi}}$ on $\mathsf{X}_N \equiv \mathbb{R}^N$ is absolutely continuous with respect to a fixed product measure λ on $\mathbb{R}^N$, typically the $|N|$-dimensional Lebesgue measure. This is, for example, the case of a regular Gaussian distribution. In this continuous case, the density of $P_{\boldsymbol{\xi}}$ is defined as the Radon-Nikodym derivative of $P_{\boldsymbol{\xi}}$ with respect to the dominating σ-finite measure λ. One can analogously introduce its marginal densities f_I, $I \subseteq N$, as functions on $\mathsf{X}_N = \mathbb{R}^N$ depending on components indexed by I; however, these marginal densities are only determined uniquely in the λ-almost everywhere sense. The formula (1) with p_* replaced by f_* and with equality understood in the λ-almost everywhere sense then defines the respective CI statement.

Nonetheless, a straightforward algebraic characterization of the CI statement is at our disposal in the *regular Gaussian case*. This is the case of a real random vector $\boldsymbol{\xi}$ where $P_{\boldsymbol{\xi}} = \mathcal{N}(\mu, \Sigma)$ is the multivariate normal distribution with a mean vector $\mu \in \mathbb{R}^N$ and a positive definite *covariance matrix* $\Sigma \in \mathbb{R}^{N \times N}$; see [11, Appendix C] for a formal definition. Then, given pairwise disjoint $I, J, K \subseteq N$, the respective CI statement is characterized solely in terms of the covariance matrix Σ. Specifically, one has $\xi_I \perp\!\!\!\perp \xi_J | \xi_K$ iff the $I \times J$-submatrix of the inverse matrix to the $IJK \times IJK$-submatrix of Σ consists of zeros; see [42, Corolary 6.3.4] or [11, Proposition 5.2] in a special case.

Formally, the CI structure induced by a random vector $\boldsymbol{\xi}$ over N is a certain ternary relation on the power set $\mathcal{P}(N) := \{A \,:\, A \subseteq N\}$. Specifically, the *augmented CI structure* induced by a random vector $\boldsymbol{\xi}$ over N is

$$\{\, \langle I, J | K \rangle \in \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N) \,:\, \xi_I \perp\!\!\!\perp \xi_J | \xi_K \,\}\,.$$

We use a vertical line to separate the third component, which is interpreted as the conditioning variable set K. The *standard CI structure* induced by $\boldsymbol{\xi}$ is the above-mentioned ternary relation on $\mathcal{P}(N)$ confined to the triplets of pairwise disjoint subsets of N as traditionally required in [11, 34].

2.3. Matroids and polymatroids

A *matroid* over a basic set N is specified by a collection $\mathcal{I} \subseteq \mathcal{P}(N)$ of *independent sets* (in the matroid) satisfying the following axioms:

- $\emptyset \in \mathcal{I}$,
- $I \in \mathcal{I}$ and $J \subseteq I$ implies $J \in \mathcal{I}$,
- if $I, J \in \mathcal{I}$ with $|J| < |I|$ then there exists $i \in (I \setminus J)$ such that $J \cup \{i\} \in \mathcal{I}$.

Matroids are mathematical structures abstracting the concept of linear independence. Specifically, the following example is a classic instance of a matroid. Given a finite collection $\{x_i : i \in N\}$ of vectors in a linear space $\mathcal{E}$ over a field F the collection

$$\mathcal{I} \;=\; \{\, I \subseteq N \,:\, \{x_i\}_{i \in I} \;\text{ is a linearly independent set of vectors}\,\}$$

satisfies the axioms above and defines a matroid over N. Every matroid of this form is called *linearly representable* over the field F.

Several equivalent descriptions of a fixed matroid exists; see [33, Chapter 1]. One of them is the so-called *rank function* (of the matroid specified by $\mathcal{I}$) which is a function $r_{\mathcal{I}} : \mathcal{P}(N) \to \mathbb{Z}$ defined by the formula

$$r_{\mathcal{I}}(I) \;:=\; \max\{\,|J| \,:\, J \subseteq I \;\&\; J \in \mathcal{I}\,\} \qquad \text{for any } I \subseteq N.$$

Note that, in the case of a linearly representable matroid, the value $r_{\mathcal{I}}(I)$ for $I \subseteq N$ is the dimension of the linear subspace of $\mathcal{E}$ generated by $\{x_i : i \in I\}$ [33, § 1.3].

It is a well-known fact that an integer-valued set function $r : \mathcal{P}(N) \to \mathbb{Z}$ is a rank function of a matroid over N iff it satisfies the following conditions (see [33, Corollary 1.3.4]):

- if $I \subseteq N$ then $0 \le r(I) \le |I|$,
- if $J \subseteq I$ then $r(J) \le r(I)$,
- if $I, J \subseteq N$ then $r(I) + r(J) \ge r(I \cup J) + r(I \cap J)$.

The first condition means that r is non-negative and bounded from above by cardinality, the second one that r is *non-decreasing* and the third one that r is *submodular*. Note that the inverse relation is as follows: $\mathcal{I} = \{\, I \subseteq N \,:\, r_{\mathcal{I}}(I) = |I|\,\}$.

The description of a matroid in terms of a rank function leads to the following generalization of that concept; see [7, § 2.2]. By a rank function of a *polymatroid* over N we mean any *real* function $r : \mathcal{P}(N) \to \mathbb{R}$ which satisfies $r(\emptyset) = 0$, is non-decreasing and submodular. Polymatroids can also be assigned abstract CI structures (see Section 4.2).

2.4. Entropy function

Given a discrete random variable ξ with a sample space X, whose distribution is given by a density $p : \mathsf{X} \to [0,1]$, its *entropy* $H(\xi)$ is given by the formula

$$H(\xi) \;:=\; -\sum_{x \in \mathsf{X}: p(x) > 0} p(x) \cdot \ln(p(x))\,.$$

It is clearly a non-negative real number. Thus, every discrete random vector $\boldsymbol{\xi} = [\xi_i]_{i \in N}$ can be assigned its *entropy function* $h_{\boldsymbol{\xi}} : \mathcal{P}(N) \to [0, \infty)$ defined by

$$h_{\boldsymbol{\xi}}(I) \;:=\; H(\xi_I) \qquad \text{for every } I \subseteq N.$$

It follows from the basic information-theoretical inequalities [43, Appendix 14.A] that the entropy function $h_{\boldsymbol{\xi}}$ is a rank function of a *polymatroid*, that is, formally: $h_{\boldsymbol{\xi}}(\emptyset) = 0$, $h_{\boldsymbol{\xi}}(J) \le h_{\boldsymbol{\xi}}(I)$ whenever $J \subseteq I \subseteq N$, and

$$h_{\boldsymbol{\xi}}(I) + h_{\boldsymbol{\xi}}(J) \ge h_{\boldsymbol{\xi}}(I \cup J) + h_{\boldsymbol{\xi}}(I \cap J) \qquad \text{for any } I, J \subseteq N.$$

Note that the polymatroidal inequalities for $h_{\boldsymbol{\xi}}$ can equivalently be formulated as the condition that, for every $I, J, K \subseteq N$ (possibly intersecting), the expression

$$\Delta h_{\boldsymbol{\xi}}(I, J|K) \;:=\; h_{\boldsymbol{\xi}}(I \cup K) + h_{\boldsymbol{\xi}}(J \cup K) - h_{\boldsymbol{\xi}}(I \cup J \cup K) - h_{\boldsymbol{\xi}}(K)$$

is non-negative. This expression is none other than the so-called *conditional mutual information* between ξ_I and ξ_J given ξ_K, which is known to vanish just in the case of the validity of the respective CI statement; see [43, Theorem 2.34]. In particular, for (possibly intersecting) sets $I, J, K \subseteq N$, one has

$$\xi_I \perp\!\!\!\perp \xi_J | \xi_K \quad \Leftrightarrow \quad h_{\boldsymbol{\xi}}(I \cup K) + h_{\boldsymbol{\xi}}(J \cup K) = h_{\boldsymbol{\xi}}(I \cup J \cup K) + h_{\boldsymbol{\xi}}(K)\,,$$

see [43, § 13.5] for further details.

2.5. Lattices

A partially ordered set $(\mathcal{Z}, \preceq)$, abbreviated as a *poset* in this paper, is called a *lattice* [1, § I.4] if every two-element subset of $\mathcal{Z}$ has both the least upper bound, also called the supremum or the *join*, and the greatest lower bound, also called the infimum or the *meet*. A finite lattice is necessarily *complete*, which means that the requirement given above holds for any subset of $\mathcal{Z}$. A lattice $(\mathcal{Z}, \preceq)$ is *anti-isomorphic* to a lattice $(\mathcal{Z}', \preceq')$ if there is a one-to-one mapping ι from $\mathcal{Z}$ onto $\mathcal{Z}'$ which reverses the ordering: for $x, y \in \mathcal{Z}$, one has $x \preceq y$ iff $\iota(y) \preceq' \iota(x)$.

An element e of a lattice $\mathcal{Z}$ is called *meet-irreducible* if it cannot be written as the infimum of two elements from $\mathcal{Z}$ that would both be different from e. In a finite lattice, every element can be written as the infimum of a set of meet-irreducible elements. Analogously *join-irreducible* elements are defined, with supremum instead of infimum; of course, an anti-isomorphism maps meet-irreducible elements to join-irreducible ones and conversely.

Examples of join-irreducible elements are the least element $\mathsf{0}$ in the lattice and its *atoms*, which are elements $a \in \mathcal{Z}$ distinct from $\mathsf{0}$ such that the only elements $e \in \mathcal{Z}$ satisfying $e \preceq a$ are $e = a$ and $e = \mathsf{0}$. A (complete) lattice is called *atomic* [1, § VIII.9] if the only join-irreducible elements are the least element and the atoms. An equivalent definition is that every element $e \in \mathcal{Z}$ is the supremum of a set of atoms in $\mathcal{Z}$. An example of an atomic lattice is the *face-lattice* of a polyhedral cone [45, Theorem 2.7(v)].

3. AXIOMATIC CHARACTERIZATIONS IN SPECIAL CASES

In this section we discuss two elegant results of F. Matúš about axiomatic characterization of certain substructures of probabilistic CI structures.

3.1. Axiomatic characterization of functional dependence

The first journal paper by Fero touching on a topic related to probabilistic CI was his 1991 paper about abstract functional dependence [16]. In that paper, he gave, among other things, an axiomatic characterization of *stochastic functional dependence*; this concept can be viewed as a special case of the concept of probabilistic CI.

More specifically, given a random vector $\boldsymbol{\xi} = [\xi_i]_{i \in N}$ and subsets $I, J \subseteq N$, we say that ξ_J *functionally depends on* (or, equivalently, is a function of) ξ_I and write $I \to J\ [\boldsymbol{\xi}]$ if there exists a (correspondingly measurable) function $g : \mathsf{X}_I \to \mathsf{X}_J$ such that $\xi_J(\omega) = g(\xi_I(\omega))$ holds for any $\omega \in \Omega$. Verification of the following fact is left to the reader as an easy exercise.

Observation 3.1. Given a discrete random vector $\boldsymbol{\xi}$ over N,

$$I \to J\ [\boldsymbol{\xi}] \qquad \Leftrightarrow \qquad \xi_J \perp\!\!\!\perp \xi_J | \xi_I$$

holds for any $I, J \subseteq N$.

Recall that an alternative notation for $\xi_J \perp\!\!\!\perp \xi_J | \xi_I$ is $J \perp\!\!\!\perp J \,|\, I\ [\boldsymbol{\xi}]$. The functional dependence of J on I can thus be interpreted as self-independence of J conditioned by I. Every random vector $\boldsymbol{\xi}$ over N then induces a binary relation $* \to *\ [\boldsymbol{\xi}]$ on the power set $\mathcal{P}(N)$, which can be interpreted as a substructure of the augmented CI structure (defined in Section 2.2). It follows from the results of [16], specifically see [16, Remark 4], that a binary relation $\to$ on $\mathcal{P}(N)$ is induced by a discrete random vector $\boldsymbol{\xi}$ over N iff it satisfies the following implications, interpreted as "axioms" (for any $I, J, K \subseteq N$):

- $J \subseteq I$ implies $I \to J$,
- $I \to J$, $J \to K$ implies $I \to K$,
- $I \to J$, $I \to K$ implies $I \to (J \cup K)$.

The term an *FD-relation* on N was used in [16] to name any binary relation $\to$ on $\mathcal{P}(N)$ satisfying the three implications listed above (*FD* abbreviates *functional dependency*). Every FD-relation on N can be associated with a *closure system* on N, which is a collection $\mathcal{C} \subseteq \mathcal{P}(N)$ of subsets of N closed under intersection and containing N. Closure systems over N, sometimes called Moore families, are known to encode abstract *closure operators* on N; see [1, § V.1]. Recall that these are operators $\mathsf{c} : \mathcal{P}(N) \to \mathcal{P}(N)$ satisfying $I \subseteq \mathsf{c}(I)$, $\mathsf{c}(I) = \mathsf{c}(\mathsf{c}(I))$ and $I \subseteq J \ \Rightarrow\ \mathsf{c}(I) \subseteq \mathsf{c}(J)$ for any $I, J \subseteq N$.

Specifically, every FD-relation $* \to *$ on N defines a closure operator on N by

$$\mathsf{c}_{\to}(I) \ := \ \bigcup \{ J \subseteq N \ : \ I \to J \} \qquad \text{for any } I \subseteq N,$$

and a closure system by $\mathcal{C}_{\to} := \{ C \subseteq N \ : \ C = \mathsf{c}_{\to}(C) \}$. Conversely, any closure system $\mathcal{C} \subseteq \mathcal{P}(N)$ defines a closure operator on N by

$$\mathsf{c}_{\mathcal{C}}(I) \ := \ \bigcap \{ C \in \mathcal{C} \ : \ I \subseteq C \} \qquad \text{for any } I \subseteq N,$$

and an FD-relation on N by $I \to_{\mathcal{C}} J$ iff $J \subseteq \mathsf{c}_{\mathcal{C}}(I)$ for any $I, J \subseteq N$.

The above-mentioned (one-to-one) correspondence between FD-relations $\to$ and closure systems $\mathcal{C}$, defined through a shared closure operator $\mathsf{c}_{\to} = \mathsf{c}_{\mathcal{C}}$, was shown in [16] to be a special case of the *Galois connections* (see [1, Sections V.7,V.8]) between lattices $(\mathcal{P}(N) \times \mathcal{P}(N), \subseteq)$ and $(\mathcal{P}(N), \subseteq)$ induced by the following binary relation $\circ \subseteq [\mathcal{P}(N) \times \mathcal{P}(N)] \times \mathcal{P}(N)$:

$$\text{given } I, J, C \subseteq N, \qquad [I, J] \circ C \ \Leftrightarrow\ \{ I \setminus C \neq \emptyset \quad \text{or} \quad J \subseteq C \}\,.$$

The term FD-relation was justified in [16] by additional examples of such abstract relations arising in other contexts: lattice theory, theory of relational databases, polymatroids and logical dependency of Boolean variables. A generic example is as follows:

consider a poset $(\mathcal{Z}, \preceq)$ that has the greatest element and in which every pair of elements has the infimum $\wedge$ (= the so-called *meet semilattice*). Having a finite indexed collection $\{z_i : i \in N\}$ of elements in $\mathcal{Z}$ one can define, for any $I, J \subseteq N$,

$$I \to J \; [\mathcal{Z}] \quad := \quad \bigwedge_{i\in I} z_i \preceq \bigwedge_{j\in J} z_j \, .$$

The representation theorem was proved in [16] first for the above-specified generic case and then extended to other cases. In addition to the above stochastic functional dependence $* \to * \; [\boldsymbol{\xi}]$, we mention the case of an FD-relation induced by a rank function r of a *polymatroid* (see Section 2.3): we write $I \to J \; [r]$ for $I, J \subseteq N$ if $r(I) = r(I \cup J)$.

3.2. Axiomatic characterization of unconditional independence

Fero characterized stochastic independence structures in his thesis [15] and later published that result in a journal paper, [19]. Nevertheless, an analogous result was independently achieved in [8], where, however, possible self-independence of random variables has not been considered.

As explained in Section 2.2, every (discrete) random vector $\boldsymbol{\xi}$ over N induces a binary (unconditional independence) relation $* \perp\!\!\!\perp * \; [\boldsymbol{\xi}]$ on $\mathcal{P}(N)$, which is clearly a substructure of the augmented CI structure. It follows from [19, Theorem 2] that a binary relation $\perp\!\!\!\perp$ on $\mathcal{P}(N)$ is induced by a discrete random vector $\boldsymbol{\xi}$ over N in this way iff it is a non-empty relation and satisfies the following implications (for any $I, J, K \subseteq N$):

- $I \perp\!\!\!\perp I$ implies $I \perp\!\!\!\perp N$,
- $I \perp\!\!\!\perp J$, $K \subseteq J$ implies $I \perp\!\!\!\perp K$,
- $(I \cup J) \perp\!\!\!\perp K$, $I \perp\!\!\!\perp J$ implies $(J \cup K) \perp\!\!\!\perp I$.

Note that the triviality ($I \perp\!\!\!\perp \emptyset$ for any $I \subseteq N$) and the symmetry ($I \perp\!\!\!\perp J$ implies $J \perp\!\!\!\perp I$) properties follow from the above-mentioned conditions. The term an *I-relation* on N was used in [19] to name any non-empty binary relation $\perp\!\!\!\perp$ on $\mathcal{P}(N)$ satisfying those three implications (I stands for *independence*).

Every I-relation on N can be defined by means of its associated (abstract) class of *connected sets*, which is a non-empty collection $\mathcal{C} \subseteq \mathcal{P}(N)$ of subsets of N such that

- $C \in \mathcal{C}$, $C' \subseteq C$, $|C'| \leq 1$ implies $C' \in \mathcal{C}$,
- $C, C' \in \mathcal{C}$, $C \cap C' \neq \emptyset$ implies $(C \cup C') \in \mathcal{C}$.

The correspondence between I-relations $\perp\!\!\!\perp$ and such set systems $\mathcal{C}$ was introduced in [19] in terms of *Galois connections* [1, § V.7-8] induced by the following binary relation $\bullet \subseteq [\mathcal{P}(N) \times \mathcal{P}(N)] \times \mathcal{P}(N)$:

$$\text{for } I, J, C \subseteq N, \quad [I, J] \bullet C \;\; \Leftrightarrow \;\; \{\, I \cap J \cap C = \emptyset \text{ and } [C \setminus (I \cup J) \neq \emptyset \text{ or } C \subseteq I \text{ or } C \subseteq J\,]\,\}\,.$$

In [19], Fero gave two proofs of the representation theorem. The first one, following the idea from his thesis [15], used the tool of Fourier-Stieltjes transformation. The second proof, similar to the one from [8], was based on the Galois connections and utilized the observations discussed in the next Section 4. Moreover, it was shown in [19] that I-relations also arise in the context of lattice theory through the concept of *algebraic independence* and an analogous representation theorem was proved as [19, Theorem 3].

4. MATROIDS AND ABSTRACT CI STRUCTURES

This section is devoted to the relation of matroids and probabilistic CI structures, which was Fero's favorite research topic.

4.1. Probabilistically representable matroids

The following concept was already introduced in [18, § 3]; nonetheless, we use here a later terminology by Fero from [20, § 6]. We say that a matroid over N with a rank function $r : \mathcal{P}(N) \to \mathbb{Z}$ is *strongly probabilistically representable* if there exists a discrete random vector $\boldsymbol{\xi} = [\xi]_{i\in N}$ over N and a constant $k > 0$ such that

$$r(I) = k \cdot h_{\boldsymbol{\xi}}(I) \qquad \text{for every } I \subseteq N.$$

The following result by Fero, published as [18, Theorem 2], is highly important because it is a theoretical basis of his later methodology for constructing probabilistic representations of CI structures (see Sections 4.3 and 6.1). For this reason, its modified proof is given, which avoids an unexplained dual space interpretation from the original proof of [18, Theorem 2].

Theorem 4.1. Every matroid which is linearly representable over a finite field is also strongly probabilistically representable.

Proof. As explained in Section 2.3, the assumption means that there exist a linear space $\mathcal{E}$ over a finite field F and vectors $x_i \in \mathcal{E}$, $i \in N$, in it such that

$$r(I) = \dim\left(\mathrm{Lin}(\{x_i\}_{i\in I})\right) \qquad \text{for every } I \subseteq N,$$

where Lin(∗) denotes the linear hull. One can, without loss of generality, assume that $\mathcal{E} = \mathrm{Lin}(\{x_i\}_{i\in N})$ holds; otherwise, $\mathcal{E}$ would be replaced by its subspace. In such a case, one has $r(N) = \dim(\mathcal{E}) =: d$. Since $\mathcal{E}$ is finite-dimensional, one can interpret it as F^d and equip it with a non-degenerate symmetric bilinear form which maps $\mathcal{E} \times \mathcal{E}$ to F. Specifically, choose and fix a linear basis $\mathcal{L} \subseteq \mathcal{E}$, $|\mathcal{L}| = d$; then every element of $x \in \mathcal{E}$ can uniquely be written as $x = \sum_{l\in\mathcal{L}} \alpha^x_l \cdot l$ with $\alpha^x_l \in F$, $l \in \mathcal{L}$. Given $x, \omega \in \mathcal{E}$, we put $\langle x, \omega\rangle := \sum_{l\in\mathcal{L}} \alpha^x_l \cdot \alpha^\omega_l \in F$.

To verify the probabilistic representability of the matroid we define $\Omega := \mathcal{E}$, which is a finite set, and equip it with uniformly distributed probability measure P. Then we define random variables by $\xi_i(\omega) := \langle x_i, \omega\rangle$ for every $\omega \in \Omega = \mathcal{E}$ and $i \in N$. One can show (we skip the proof; see [18]) that $h_{\boldsymbol{\xi}}(I) = r(I) \cdot \ln(|F|)$ for every $I \subseteq N$. □

4.2. Embedding of matroids into abstract CI structures

Any ternary relation on $\mathcal{P}(N)$ can be interpreted as an *abstract CI structure* over N. Given $\mathcal{M} \subseteq \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N)$, the record $I \perp\!\!\!\perp J \mid K$ $[\mathcal{M}]$ for sets $I, J, K \subseteq N$ is read as "I is *conditionally independent of* J *given* (= under the condition) K *with respect to* (an abstract CI structure) $\mathcal{M}$". It means both $\langle I, J|K\rangle \in \mathcal{M}$ and that the triplets in $\mathcal{M}$ are interpreted as (conditional) *independence* statements. The record $I \not\perp\!\!\!\perp J \mid K$ $[\mathcal{M}]$ then means that $\langle I, J|K\rangle \notin \mathcal{M}$ holds, and that the triples outside $\mathcal{M}$ are interpreted as (conditional) *dependence* statements. If there is a mathematical object $\boldsymbol{o}$ over N inducing (= defining) the ternary relation $\mathcal{M}$ then we can write $I \perp\!\!\!\perp J \mid K$ $[\boldsymbol{o}]$ instead and say that the CI statement is *with respect to* $\boldsymbol{o}$, analogously with $I \not\perp\!\!\!\perp J \mid K$ $[\boldsymbol{o}]$.

Thus, we say that an abstract CI structure $\mathcal{M}$ over N is *probabilistically representable* if it is induced by a *discrete* random vector $\boldsymbol{\xi}$ over N in the sense of Section 2.2, that is,

$$\mathcal{M} = \{\, \langle I, J|K\rangle \in \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N) \,:\, \xi_I \perp\!\!\!\perp \xi_J | \xi_K \,\}\,.$$

An alternative record $I \perp\!\!\!\perp J \mid K$ $[\boldsymbol{\xi}]$ for $\xi_I \perp\!\!\!\perp \xi_J | \xi_K$ then follows our general notational convention. Our terminology here is, in fact, an unsubstantial modification of the one from [20, § 2]; see Section 5.1 for the description of the difference.

We will recognize two parts of any abstract CI structure $\mathcal{M}$. By the *functional dependence part* of $\mathcal{M}$ we will mean its intersection with the collection of triplets of the form $\langle J, J|K\rangle$, where $J, K \subseteq N$ are disjoint; these correspond to self-independence CI statements. Indeed, in the case of a probabilistic CI structure, this object is none other than the respective stochastic FD-relation; see Observation 3.1 in Section 3.1. The (pure) *conditional independence part* of $\mathcal{M}$ will be understood as the intersection of $\mathcal{M}$ with the collection of triplets of pairwise disjoint subsets of N. Indeed, in the case of a probabilistic CI structure, this is just the *standard* probabilistic CI structure.

The basic idea presented in [20, § 3] was that any matroid over N can be identified with an abstract CI structure over N, and this embedding is an injective mapping. Specifically, given a rank function r of a matroid over N and $I, J, K \subseteq N$, we put

$$\Delta r(I, J|K) \;:=\; r(I \cup K) + r(J \cup K) - r(I \cup J \cup K) - r(K).$$

Note that one always has $\Delta r(I, J|K) \geq 0$ (see Sections 2.3 and 2.4). Put

$$\mathcal{M}_r \;:=\; \{\, \langle I, J|K\rangle \in \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N) \,:\, \Delta r(I, J|K) = 0 \,\} \tag{2}$$

and interpret it as an abstract CI structure over N induced by r; thus, the symbol $I \perp\!\!\!\perp J \mid K$ $[r]$ for $I, J, K \subseteq N$ will mean $\Delta r(I, J|K) = 0$. It follows from the arguments in [20, § 3] that the mapping $r \mapsto \mathcal{M}_r$ is injective. More specifically, the collection $\mathcal{I}$ of independent sets in a matroid can be reconstructed from the functional dependence part of $\mathcal{M}_r$ with $r = r_{\mathcal{I}}$ as follows:

$$\mathcal{I} \;=\; \{\, I \subseteq N \,:\, \forall\, J, K \subseteq I, J \cap K = \emptyset \neq J \quad J \not\perp\!\!\!\perp J \mid K \; [r] \,\}\,.$$

This observation justifies Fero's terminology from [20], where he used the word "matroid" to name any abstract CI structure obtained from a rank function of a matroid in this way; more precisely, he used the word "matroid" to name the so-called local version of an abstract CI structure, see Section 5.1 for an explanation.

It is clear that the definition cited above also works in the case of a rank function of *polymatroid*. Thus, following Fero [20], one can say that an abstract CI structure is a *semimatroid* if it is induced by a rank function r of a polymatroid through (2).

Remark 4.2. A *loop* in a matroid over N is a singleton which is not an independent set in that matroid, that is, $\{i\} \notin \mathcal{I}$ for $i \in N$ [33, § 1.1.8]. In terms of the matroid rank function r it means $r(\{i\}) = 0$ and this definition can be extended to (the rank functions of) polymatroids. In terms of the induced abstract CI structure $\mathcal{M} = \mathcal{M}_r$ it hence means $i \perp\!\!\!\perp i \,|\, \emptyset \; [\mathcal{M}]$; this condition can be used as the definition of a loop within an abstract CI structure $\mathcal{M}$. Thus, any *loopless* matroid (= a matroid without loops) is assigned a loopless abstract CI structure. The point here is that a loopless matroid is uniquely determined by the pure conditional independence part of its induced CI structure; specifically, one can reconstruct the collection $\mathcal{I}$ of independent sets as follows:

$$\mathcal{I} \;=\; \{ L \subseteq N \,:\, \forall \text{ pairwise disjoint } I, J, K \subseteq L \quad I \perp\!\!\!\perp J \,|\, K \; [r_{\mathcal{I}}] \,\} \,.$$

Therefore, one can interpret the mapping $r \mapsto \mathcal{M}_r$ as an injective embedding of the class of loopless matroids into the class of *standard CI structures*.

Remark 4.3. The conditional independence parts of *semimatroids* over N appear to coincide with *structural semigraphoids* over N, which were introduced in [38] and later identified in [40, § 4.4.1] with standard abstract CI structures induced by the so-called *structural imsets*. Nonetheless, this observation is not immediate; it follows from a dual description of these structures in the terms of supermodular functions; see [40, Corollary 5.3].

The above embedding of matroids (and polymatroids) into abstract CI structures allowed Fero in [20] to extend the term of probabilistic representability to poly/matroids. Specifically, one can say that a polymatroid given by a rank function r is *probabilistically representable* if the respective semimatroid $\mathcal{M}_r$ is a probabilistic CI structure; that is, if there exists a discrete random vector $\boldsymbol{\xi} = [\xi_i]_{i \in N}$ over N such that

$$\forall\, I, J, K \subseteq N \qquad I \perp\!\!\!\perp J \,|\, K \; [r] \quad \Leftrightarrow \quad \xi_I \perp\!\!\!\perp \xi_J \,|\, \xi_K \,.$$

This terminology, applied to matroids, suggests that the following observation is true.

Observation 4.4. Any strongly probabilistically representable matroid over N is also probabilistically representable.

Proof. Recall that we assume $r = k \cdot h_{\boldsymbol{\xi}}$ with $k > 0$. As explained in Section 2.4, for $I, J, K \subseteq N$, one has $\xi_I \perp\!\!\!\perp \xi_J \,|\, \xi_K$ iff $0 = \Delta h_{\boldsymbol{\xi}}(I, J|K) = k^{-1} \cdot \Delta r(I, J|K)$. □

Let us recall two more results from [20]. Fero gave in [20, § 7] an example of a semimatroid over four variables that is not probabilistically representable. It leads to an example of a matroid over N, $|N| = 8$, which is not probabilistically representable; this is none other than the so-called *Vamos cube*, see [33, § 2.1.22]. The main result [20, Theorem in § 5] says that any probabilistic representation $\boldsymbol{\xi}$ of a *connected matroid* is special: it has to be a uniform probability measure on a particular subset of X_N.

4.3. Methodology for constructing probabilistic representations

Theorem 4.1, combined with Observation 4.4 was a basis of Fero's methodology for constructing probabilistic representations of abstract CI structures. He utilized his approach in [18] to extend the results reported in Section 3 and later in a series of papers [22, 23, 25] devoted to the case of four discrete variables (see Section 6.1).

4.3.1. Operations preserving probabilistic representability

Further ingredients of Fero's method were operations with abstract CI structures preserving probabilistic representability. The most important one among them is the (set) *intersection*: if structures $\mathcal{M}^1, \mathcal{M}^2 \subseteq \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N)$ are represented by discrete random vectors $\boldsymbol{\xi}^1 = [\xi_i^1]_{i\in N}, \boldsymbol{\xi}^2 = [\xi_i^2]_{i\in N}$ then $\mathcal{M}^1 \cap \mathcal{M}^2$ is represented by a "composed" random vector. Specifically, it is represented by $\boldsymbol{\xi} = [\xi_i]_{i\in N}$ where, for each $i \in N$, $\xi_i = [\xi_i^1, \xi_i^2]$ denotes a random vector composed of ξ_i^1 and ξ_i^2: it takes values in $\mathsf{X}_i^1 \times \mathsf{X}_i^2$ when ξ_i^1 takes values in X_i^1 and ξ_i^2 in X_i^2.

The second operation is a kind of *coarsening*. If $\phi : N' \to N$ is a mapping from a basic set N' onto N and $\mathcal{M}' \subseteq \mathcal{P}(N') \times \mathcal{P}(N') \times \mathcal{P}(N')$ is probabilistically representable then its "coarsened" structure $\mathcal{M}$ over N defined by

$$\mathcal{M} := \{ \langle I, J|K\rangle \in \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N) \ : \ \langle \phi^{-1}(I), \phi^{-1}(J)|\phi^{-1}(K)\rangle \in \mathcal{M}' \}$$

is probabilistically representable, too. Indeed, if a random vector $\boldsymbol{\xi}' = [\xi'_j]_{j\in N'}$ induces $\mathcal{M}'$ then $\boldsymbol{\xi} = [\xi_i]_{i\in N}$ with $\xi_i = [\xi'_j]_{j\in\phi^{-1}(i)}$ for each $i \in N$, induces $\mathcal{M}$. Note that Fero, however, used his special terminology in this context, motivated by [32], and said that $\mathcal{M}'$ is an *expansion* of $\mathcal{M}$. He also said (in [27, § 7]) that $\mathcal{M}$ is a *factor* of $\mathcal{M}'$, with the meaning that it is determined by an equivalence on N' defined by $j \sim j'$ iff $\phi(j) = \phi(j')$.

4.3.2. Ascending conditional independence structures

Using the above-described methodology, Fero succeeded at characterizing those *standard probabilistic* CI structures that are *monotone in the conditioning set*, [18]. His result concerning the superset-monotonicity can be re-phrased as follows: a ternary relation $* \perp\!\!\!\perp * \,|\, *$ on $\mathcal{P}(N)$ confined to pairwise disjoint triplets is induced by a discrete random vector $\boldsymbol{\xi}$ over N and closed under enlarging the conditioning set iff it satisfies the following conditions (for pairwise disjoint subsets $I, J, K, L, M \subseteq N$):

- $I \perp\!\!\!\perp \emptyset \,|\, K$,
- $\{ I \perp\!\!\!\perp J \,|\, KM$ and $I \perp\!\!\!\perp K \,|\, M \}$ is equivalent to $JK \perp\!\!\!\perp I \,|\, M$,
- $I \perp\!\!\!\perp J \,|\, K$ implies $I \perp\!\!\!\perp J \,|\, KL$,
- $I \perp\!\!\!\perp J \,|\, KM$, $I \perp\!\!\!\perp J \,|\, LM$, $K \perp\!\!\!\perp L \,|\, M$ implies $I \perp\!\!\!\perp J \,|\, M$.

Following the terminology used in [18, § 6], we call any ternary relation satisfying these four conditions an *ACI-relation* (where ACI stands for "ascending conditional independence"). One can modify Fero's method to construct a probabilistic representation of a given ACI-relation as described below. Introduce the following binary relation $\diamond$ between pairwise disjoint triplets from $\mathcal{P}(N)\times\mathcal{P}(N)\times\mathcal{P}(N)$ and elements of $\mathcal{P}(N)$:

$$\forall\, I,J,K,C\subseteq N \qquad \langle I,J|K\rangle\diamond C \;\Leftrightarrow\; \{\, C\cap K\neq\emptyset \text{ or } C\cap I=\emptyset \text{ or } C\cap J=\emptyset\,\}\,.$$

This relation allows us to introduce the Galois connections [1, § V.7-8] between the two respective posets (ordered by the set-theoretical inclusion $\subseteq$) and to observe that ACI-relations form a complete lattice. This procedure is analogous to those presented in Sections 3.1 and 3.2. Any ACI-relation can be defined in terms of its associated class $\mathcal{C}\subseteq\mathcal{P}(N)$ of sets where $\mathcal{C}$ contains all subsets of N of cardinality at most 1 and satisfies

- $\forall\, C\subseteq N \qquad \text{if } \;[\,\forall\, i,j\in C\;\; \exists\, C'_{ij}\in\mathcal{C} \text{ such that } i,j\in C'_{ij}\subseteq C\,] \quad \text{then } C\in\mathcal{C}.$

This allows one to observe that any ACI-relation is the intersection of special ACI-relations, namely of those induced by matroids with rank functions

$$r_C(I) := \min\{\,1, |I\cap C|\,\} \quad \text{for } I\subseteq N, \qquad \text{where } C\subseteq N.$$

These matroids are linearly representable over finite fields and, by Theorem 4.1 (and Observation 4.4), they are probabilistically representable. Finally, one realizes that probabilistic CI structures are closed under (set) intersection.

Note that Fero's technical presentation of his results about ACI-relations in [18] was slightly different from our presentation here. He introduced and treated ACI-relations in their local versions (see Section 5.1 for details), while our definition in this text is equivalent to his later global versions (of ACI-relations) from [18, § 6].

4.3.3. Descending conditional independence structures

Here comes a re-formulation of Fero's result concerning subset-monotone CI structures. A ternary relation $*\perp\!\!\!\perp *\,|\,*$ on $\mathcal{P}(N)$ confined to pairwise disjoint triplets is induced by a discrete random vector $\boldsymbol{\xi}$ over N and closed under reduction of the conditioning set iff it satisfies the following conditions (for pairwise disjoint subsets $I,J,K,L,M\subseteq N$):

- $I\perp\!\!\!\perp\emptyset\,|\,K$,
- $\{\, I\perp\!\!\!\perp J\,|\,KM$ and $I\perp\!\!\!\perp K\,|\,M\,\}$ is equivalent to $JK\perp\!\!\!\perp I\,|\,M$,
- $I\perp\!\!\!\perp J\,|\,KL$ implies $I\perp\!\!\!\perp J\,|\,K$,
- $I\perp\!\!\!\perp J\,|\,KM$, $I\perp\!\!\!\perp J\,|\,LM$, $K\perp\!\!\!\perp L\,|\,IJM$ implies $I\perp\!\!\!\perp J\,|\,KLM$.

Analogously, we call any ternary relation satisfying these conditions a *DCI-relation* (where DCI stands for "descending conditional independence"). The proof in [18] was based on duality considerations. Specifically, every matroid over N with a rank function r can be ascribed a *dual matroid* [33, § 2.1] whose rank function r^* is given by

$$r^*(I) = |I| + r(N\setminus I) - r(N) \quad \text{for } I\subseteq N.$$

A well-known fact is that the dual of a matroid that is linearly representable over a field F is linearly representable over the same field [33, Corollary 2.2.9]. Given pairwise disjoint subsets $I,J,K\subseteq N$, the formula

$$\Delta r^*(I,J|K) = \Delta r(I,J|L) \quad \text{with} \;\; L := N\setminus IJK$$

implies that the structure $\mathcal{M}_{r^*}$ is obtained from $\mathcal{M}_r$ by (self-inverse) *duality transformation* between (standard) abstract CI structures: $I \perp\!\!\!\perp J|K \leftrightarrow I \perp\!\!\!\perp J|L$. This transformation maps an ACI-relation to a DCI-relation and vice versa. This fact, together with what has already been observed about ACI-relations, allows us to conclude that any DCI-relation is an intersection of special DCI-relations, namely those induced by rank functions r^*_C, $C \subseteq N$. Since these relations are linearly representable over finite fields, an analogous consideration allows us to prove that any DCI-relation is probabilistically representable.

The aim of concluding remarks in [18, § 9] was to note that the derived observations about ACI/DCI-relations could be used to give alternative proofs of results from [8, 9, 14] about axiomatic characterizations of other substructures of probabilistic CI structures.

5. SEMIGRAPHOIDS

The augmented probabilistic CI structures, as defined in Section 2.2, are known to satisfy the following basic formal properties (for possibly intersecting subsets $I, J, K, M \subseteq N$):

- $J \subseteq K$ implies $I \perp\!\!\!\perp J \,|\, K$,
- $I \perp\!\!\!\perp J \,|\, K$ is equivalent to $J \perp\!\!\!\perp I \,|\, K$,
- $\{\, I \perp\!\!\!\perp J \,|\, KM$ and $I \perp\!\!\!\perp K \,|\, M \,\}$ is equivalent to $I \perp\!\!\!\perp JK \,|\, M$.

Indeed, the reader can easily verify those properties using (1) in the discrete case. Every augmented abstract CI structure satisfying those properties will be called a *general semigraphoid*. If we restrict our attention to pairwise disjoint triplets of subsets of N then the first property above can be replaced by a simpler one, namely, by

- $I \perp\!\!\!\perp \emptyset \,|\, K$,

and we get what is regarded as a *standard* definition of a *semigraphoid* [11, 34]. Semigraphoids became one of Fero's research topics.

Remark 5.1. Dawid introduced an even more general concept of a *separoid* in [6]; this concept can equivalently be defined as follows: it is a poset $(\mathcal{Z}, \preceq)$ that has the least element and in which every two-element set has the join $\vee$ (= the so-called *join semilattice*) equipped with a ternary relation $* \perp\!\!\!\perp * \,|\, *$ over elements of $\mathcal{Z}$ satisfying (for $I, J, K, M \in \mathcal{Z}$)

- if $J \preceq K$ then $I \perp\!\!\!\perp J \,|\, K$,
- $I \perp\!\!\!\perp J \,|\, K$ iff $J \perp\!\!\!\perp I \,|\, K$,
- $\{\, I \perp\!\!\!\perp J \,|\, (K \vee M)$ and $I \perp\!\!\!\perp K \,|\, M \,\}$ iff $I \perp\!\!\!\perp (J \vee K) \,|\, M$,

where $J \vee K$ denotes the join of elements J and K in $\mathcal{Z}$.

5.1. Global and local representations of a semigraphoid

To represent a semigraphoid over N in the memory of a computer, one need not record all involved triplets of subsets of N. One can limit one's attention to certain *elementary CI statements*. These elementary CI statements are encoded by triplets of the form $\langle i,j|K\rangle$, where $i,j \in N$, $i \neq j$, $K \subseteq N \setminus ij$, and by triplets $\langle i,i|N \setminus i\rangle$, where $i \in N$. The next observation was made by Fero in [17, Lemma 3] for standard semigraphoids but it can be extended to the case of general semigraphoids.

Observation 5.2. If $\mathcal{M}$ is a general semigraphoid over N and $I,J,K \subseteq N$ then

$$I \perp\!\!\!\perp J \,|\, K \; [\mathcal{M}] \quad \Leftrightarrow \quad \forall\, i \in I \setminus K,\, j \in J \setminus K,\, L : K \subseteq L \subseteq IJK \setminus ij \qquad i \perp\!\!\!\perp j \,|\, L \; [\mathcal{M}] \,.$$

Moreover, $i \perp\!\!\!\perp i \,|\, K \; [\mathcal{M}]$ iff $\{\, i \perp\!\!\!\perp i \,|\, Kj \; [\mathcal{M}]$ and $i \perp\!\!\!\perp j \,|\, K \; [\mathcal{M}] \,\}$ for disjoint i,j,K.

Proof. The first claim can easily be verified using the induction by $|IJ \setminus K|$ and is left to the reader. As concerns the additional claim, let us realize that

$$\begin{aligned} \{\, i \perp\!\!\!\perp i \,|\, Kj \; [\mathcal{M}] \;\&\; i \perp\!\!\!\perp j \,|\, K \; [\mathcal{M}] \,\} \;&\Leftrightarrow\; i \perp\!\!\!\perp ij \,|\, K \; [\mathcal{M}] \\ \Leftrightarrow \quad \{\, i \perp\!\!\!\perp j \,|\, Ki \; [\mathcal{M}] \;\&\; i \perp\!\!\!\perp i \,|\, K \; [\mathcal{M}] \,\} \;&\Leftrightarrow\; i \perp\!\!\!\perp i \,|\, K \; [\mathcal{M}] \end{aligned}$$

because $i \perp\!\!\!\perp j \,|\, Ki \; [\mathcal{M}]$ always holds in a general semigraphoid $\mathcal{M}$. □

In particular, every semigraphoid is uniquely determined by its *trace*, which is its intersection with the collection of all elementary CI statements. Fero characterized those sets of elementary CI statements which are traces of standard semigraphoids. The following is our re-phrasing of his result; we skip the proof and refer to [18, Proposition 1].

Observation 5.3. A set $\mathcal{E}$ of elementary CI statements over N is the intersection a *standard* semigraphoid over N with the collection of all elementary CI statements over N iff it satisfies (for distinct i,j and $L \subseteq N \setminus ij$, respectively for $k \in N \setminus ijL$):

- $i \perp\!\!\!\perp j \,|\, L \; [\mathcal{E}]$ is equivalent to $j \perp\!\!\!\perp i \,|\, L \; [\mathcal{E}]$,
- $\{\, i \perp\!\!\!\perp j \,|\, kL \; [\mathcal{E}]$ and $i \perp\!\!\!\perp k \,|\, L \; [\mathcal{E}] \,\}$ is equivalent to $\{\, i \perp\!\!\!\perp k \,|\, jL \; [\mathcal{E}]$ and $i \perp\!\!\!\perp j \,|\, L \; [\mathcal{E}] \,\}$.

One can distinguish between two ways of dealing with semigraphoids and CI structures. A traditional approach of [11, 34], described in Section 2.2, is to define and understand them as ternary relations on $\mathcal{P}(N)$. The lists of triplets of subsets of N can be viewed as *global versions* of semigraphoids. Fero, however, preferred to work with semigraphoids represented by their *local versions*. These are lists of symmetrized elementary CI statements: indeed, there is no reason to distinguish between $i \perp\!\!\!\perp j \,|\, L$ and $j \perp\!\!\!\perp i \,|\, L$ and they can both be represented by a pair $(\{i,j\},L)$ of disjoint sets. In the case of a general semigraphoid, one can additionally consider pairs $(\{i\},L)$ of disjoint sets representing $i \perp\!\!\!\perp i \,|\, L$.

During his career, Fero insisted on the local representation of CI structures. In his papers about semigraphoids and CI, he considered the local versions of CI structures to be the primary ones. He thus formally defined CI structures in their local modes and formulated abstract properties of CI in the form of implications between elementary CI statements. This approach allowed him to represent semigraphoids over a small variable set N by special diagrams; these were undirected graphs having N as the set of nodes whose edges were annotated by respective conditioning sets, see [24, § 2]. He and his student practically used such diagrams later in [13].

5.2. Semigraphoid inference

A related concept is that of the *semigraphoid closure* of a set $\mathcal{S}$ of CI statements (over N), which is the smallest semigraphoid (over N) containing $\mathcal{S}$. A natural question from the point of view of computer science is what the complexity is of obtaining the semigraphoid closure of $\mathcal{S}$ by consecutive applications of semigraphoid implications. The implementation of *semigraphoid inference* depends on the way of internal computer representation of semigraphoids.

The main topic of the paper [26] was the length of semigraphoidal inference sequences for the semigraphoids encoded by their *local versions*. The main result of [26] was that the length of such derivation sequence can be exponential in the cardinality of N. To prove that result, Fero used an auxiliary mathematical construction of an undirected graph with colored edges whose set of nodes is the set of elementary CI statements $i \perp\!\!\!\perp j \,|\, K$ where $i, j \in N$, $i \neq j$, $K \subseteq N \setminus ij$.

Remark 5.4. Note that, in this context, an alternative internal computer representation of semigraphoids comes from a *dominance* ordering between CI statements introduced in [39]: $I \perp\!\!\!\perp J \,|\, K$ dominates $I' \perp\!\!\!\perp J' \,|\, K'$ if

$$I' \subseteq I, \quad J' \subseteq J, \quad K \subseteq K', \quad \text{and} \quad I'J'K' \subseteq IJK\,.$$

One can encode any semigraphoid over N in the memory of a computer by the list of its dominant triplets and implement semigraphoid inference under this alternative mode of representation.

5.3. Matroidal approach to CI structures

Matroid theory has always been a source of inspiration for Fero in his research on CI. The idea of interpreting abstract CI structures as generalized matroids (see Section 4.2) led him to extending several traditional operations with matroids to the context of CI.

A basic operation of this type is that of a *minor* [33, Chapter 3]; it can be viewed as a combination of two separate shrinking operations. Given a rank function r' of a matroid over N' and a subset $N \subseteq N'$, let us assume that $N' \setminus N = D \cup E$ is partitioned into two disjoint sets. The respective minor is a matroid over N whose rank function r is given by $r(I) := r'(I \cup E)$ for any $I \subseteq N$. This can be interpreted as "extracting" information from E combined with "deletion" of D. The minor operation may naturally be extended to polymatroids, leading to the following operation with abstract CI structures. Given $\mathcal{M}' \subseteq \mathcal{P}(N') \times \mathcal{P}(N') \times \mathcal{P}(N')$, the respective minor is defined by

$$\mathcal{M} \;:=\; \{\, \langle I, J|K \rangle \in \mathcal{P}(N) \times \mathcal{P}(N) \times \mathcal{P}(N) \;: \quad \langle I, J|K \cup E \rangle \in \mathcal{M}' \,\}\,.$$

Note that, in the context of graphical models, this operation with abstract CI structures is interpreted as the *conditioning* on E combined with the *marginalization* to N. Fero already showed in [20, Lemma 2] that every minor of a probabilistically representable semimatroid is probabilistically representable.

Minors were used in [24] as tools to classify abstract CI structures. Fero characterized several classes of abstract CI structures in terms of finitely many *forbidden minors*; in particular, this concerns the class of semigraphoids and the class of structures induced by undirected graphs through a separation concept. On the other hand, it was shown in [24] that some classes cannot be characterized in terms of (finitely many) forbidden minors, for example, the class of semimatroids and probabilistically representable CI structures. The result that the class of discrete probabilistic CI structures has an infinite number of forbidden minors was interpreted by Fero in [24, § 6] as an analog of the result about the non-existence of a finite axiomatic characterization from [37].

The longest paper that was written by Fero as a sole author was his 2004 paper [27] devoted to the classification of semigraphoids. Being inspired by matroid theory, he introduced and discussed various operations with semigraphoids there. Besides minors and operations mentioned in Section 4.3 (intersection, coarsening, and duality transformation) he considered four types of extension operations. More specifically, he defined in a *direct sum* of semigraphoids, the operation called a *major* (being a special inverse to the minor mapping) and the so-called *parallel extension* [27, § 4]; moreover, he defined in [27, § 7] the operation of *expansion* (being a special inverse to the coarsening). The relationship of the local mode of semigraphoid representation to the concept of *dominance* from [39] (see Remark 5.4 in Section 5.2) was established in [27, § 8-9]. Fero also introduced a kind of *canonical representation* of a semigraphoid in [27, § 10]. In [27, § 11], the concept of *linear representability* for a semigraphoid was introduced, which is stronger than its probabilistic representability (see Section 4.2); the above-introduced operations with semigraphoids were shown to preserve linear representability. The main result ([27, Theorem 2]) says that any *semigraphoid with two generators*, that is, a semigraphoid which is the semigraphoid closure of a pair of (global) CI statements, is linearly representable; the result from [39] is thus strengthened.

6. THE CASE OF FOUR RANDOM VARIABLES

This section is devoted to Fero's results on CI structures over four random variables. Both the discrete case and the regular Gaussian case are discussed here.

6.1. Discrete CI structures

In a series of papers [22, 23, 25], Fero succeeded in characterizing all *augmented discrete* CI structures (defined in Section 2.2) over four variables. In this section, his methodological approach to this task is described in more detail.

First of all, let us recall that probabilistically representable CI structures are closed under permutations of variables. More specifically, let π be a one-to-one mapping from N onto N (= a permutation of N). It can be applied to components of a random vector as follows: $\boldsymbol{\xi} = [\xi_i]_{i\in N}$ is assigned a vector $\boldsymbol{\xi}' := [\xi_{\pi^{-1}(i)}]_{i\in N}$. The entropy function of $\boldsymbol{\xi}$ is then transformed analogously: $h_{\boldsymbol{\xi}'}(I) = h_{\boldsymbol{\xi}}(\pi^{-1}(I))$ for any $I \subseteq N$. The facts mentioned in Section 2.4 imply that $I \perp\!\!\!\perp J \,|\, K\ [\boldsymbol{\xi}]$ iff $\pi(I) \perp\!\!\!\perp \pi(J) \,|\, \pi(K)\ [\boldsymbol{\xi}']$ for all subsets $I, J, K \subseteq N$. In particular, if $\mathcal{M}$ is the CI structure induced by $\boldsymbol{\xi}$ then its permuted abstract CI structure $\mathcal{M}' := \{\, \langle \pi(I), \pi(J) | \pi(K) \rangle \,:\, \langle I, J | K \rangle \in \mathcal{M} \,\}$ is induced by $\boldsymbol{\xi}'$.

The way in which a permutation π of N is applied to an entropy function $h_{\boldsymbol{\xi}}$ can naturally be extended to an arbitrary set function: a function $r : \mathcal{P}(N) \to \mathbb{R}$ is assigned $r' : \mathcal{P}(N) \to \mathbb{R}$ given by $r'(I) = r(\pi^{-1}(I))$ for $I \subseteq N$. In particular, the space $\mathbb{R}^{\mathcal{P}(N)}$ of set functions disintegrates into equivalence classes of *permutational equivalence*, which are named *permutational types* below. Since a rank function r of a polymatroid is transformed by a permutation π to a rank function r' of polymatroid, one can confine the permutational equivalence to polymatroids. The permutational equivalence can also be introduced for induced abstract CI structures, called *semimatroids* in Section 4.2. In particular, the permutational types can be recognized both in the context of (rank functions of) polymatroids and in the context of their induced abstract CI structures.

6.1.1. The starting geometrical analysis

The first paper [22] in the series was a joint paper by Fero and myself. A starting point there was a conjecture that the class of augmented discrete CI structures over N with $|N| = 4$ coincides with the class of *semimatroids* over N (defined in Section 4.2). Our methodological approach was similar to that described in Section 4.3; specifically, we planned to find a small collection $\mathbb{C}$ of semimatroids over N such that

- every semimatroid over N would be the intersection of semimatroids from $\mathbb{C}$, and
- to verify that any semimatroid from $\mathbb{C}$ would be probabilistically representable.

The smallest possible collection $\mathbb{C}$ satisfying the first condition is the class of meet-irreducible elements (see Section 2.5) in the poset of semimatroids over N (ordered by $\subseteq$). We first found in [22, Lemma 3.1] that the lattice of semimatroids is anti-isomorphic to the face-lattice of a pointed convex cone of (the rank functions of) *polymatroids*. Since that cone is known to be a complete atomic lattice (cf. Section 2.5), the first step was to characterize the extreme rays of that cone. The collection $\mathbb{C}$ is then none other than the set of semimatroids induced by generators of these extreme rays and by the zero rank function. Fero found the extreme rays of the polymatroidal cone in the case $|N| = 4$ in a previous paper of his [21, § 5]. Specifically, one has 41 such extreme rays and they are divided into 11 permutational types. Thus, $\mathbb{C}$ has 42 elements (because one has to include the largest semimatroid induced by the zero rank function).

6.1.2. Probabilistic representability confirmation in the majority of cases

The next step was to realize that 27 extreme rays of the above polymatroidal cone are generated by rank functions of *matroids*. The zero function on $\mathcal{P}(N)$ is also a rank function of a matroid. Since matroids over N with $|N| \leq 7$ are linearly representable over finite fields [33, Proposition 6.4.10], one can apply Theorem 4.1 (and Observation 4.4) to show that the induced abstract CI structures are probabilistically representable.

The remaining 14 non-matroidal extreme rays of the cone fall into three permutational types and two of these three types are obtained by coarsening of the matroidal types over N' with $|N'| = 5$ (see Section 4.3.1). As the coarsening preserves the probabilistic representability, the respective semimatroids are probabilistically representable for the same reason.

6.1.3. Ingleton inequality and a modified conjecture

The last permutational type involves six extreme rays of the polymatroidal cone: any two-element subset $\{i, j\}$ of $N := \{i, j, k, l\}$ gives a ray generated by the function

$$r_{ij}(\emptyset) = 0, \quad r_{ij}(N) = r_{ij}(ij) = 4 \quad \text{and} \quad r_{ij}(S) = |S| + 1 \quad \text{for other } S \subseteq N.$$

The semimatroid induced by r_{ij} was found in [20, § 7] *not* to be probabilistically representable. In fact, there is a special *Ingleton inequality* [10] known to be valid for any rank function of a linearly representable matroid; it has the form $\Box\, r(ij) \geq 0$ where

$$\Box\, r(ij) := -r(ij) + r(ik) + r(il) + r(jk) + r(jl) + r(kl) - r(k) - r(l) - r(ikl) - r(jkl).$$

Clearly, r_{ij} does not satisfy this inequality. Further observation, made in [22, § 4], was that the subcone of the polymatroidal cone specified by the Ingleton inequalities has 35 extreme rays: these are just the above-mentioned probabilistically representable extreme rays of the polymatroidal cone. Nonetheless, a conjecture that this subcone determines all probabilistically representable semimatroids over four variables appeared not to be true.

6.1.4. Additional geometrical analysis and representability disproval techniques

A more detailed geometrical analysis of the polymatroidal cone was, therefore, done in [22, § 6]: it was shown there that the cone is a disjoint union of its subcone specified by the Ingleton inequalities and six pieces specified by $\Box\, r(ij) < 0$. For a fixed pair ij, the topological closure H_{ij} of the respective piece is the subcone of the polymatroidal cone given by $\Box\, r(ij) \leq 0$. The 15 extreme rays of H_{ij} were found in [22, Lemma 6.1], which helped us to characterize the *non-Ingleton semimatroids*, that is, the semimatroids which are induced by rank functions from H_{ij} but not by rank functions from the subcone given by the Ingleton inequalities.

The analysis of the non-Ingleton semimatroids was the topic of the second paper [23] in the series: the goal was to determine which of them were probabilistically representable and which were not. Several techniques to disprove their probabilistic representability were used there. The main such technique from [23, § 2] was based on finer properties of probabilistic CI applicable only within a wider framework of σ-algebras; see [31] for details. The other techniques, reported in [23, § 3-4], were more specific, based either on the interpretation of CI statements in terms of factorization or on the construction of special probability distributions and the application of information-theoretical inequalities to them.

6.1.5. Application of conditional information inequalities

A novel technique was applied in the third paper [25] in the series, published four years after the second one [23]. The disproval of probabilistic representability of some non-Ingleton semimatroids in [25, § 3] was based on *conditional information inequalities* for the entropy function; note that this new method was inspired by a breakthrough made by Yeung and Zhang in 1997 [44].

The probabilistic representability for a few remaining non-Ingleton semimatroids was confirmed by four direct constructions in [23, § 5] and [25, § 3]. The last section [25, § 4] of the third paper gathered all the observations, with a conclusion that every (augmented) discrete probabilistic CI structure over four variables is the intersection of some of 120 meet-irreducible elements in the lattice of probabilistically representable semimatroids, which fall into 16 permutational types.

6.2. Gaussian CI structures

Fero was also interested in characterizing *standard* CI structures induced by *regular Gaussian* distributions. In a conference contribution [28] he introduced methods of *algebraic statistics* in the research on CI structures. In [28, § 1], an elegant criterion was given for the validity of an *elementary* CI statement with respect to a regular Gaussian distribution. Specifically, if $i, j \in N$ are distinct, $K \subseteq N \setminus ij$ and $\boldsymbol{\xi}$ has a multivariate normal distribution with a covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$ then $i \perp\!\!\!\perp j \,|\, K \; [\boldsymbol{\xi}]$ holds iff the the determinant of the $iK \times jK$-submatrix of Σ vanishes. Note that the advantage of this criterion in comparison with the traditional criterion recalled in Section 2.2 is that one avoids computing the inverse of a sub-matrix of Σ.

6.2.1. Methods of algebraic statistics

The source of motivation for the paper [28] was the question of testing the validity of implications among elementary CI statements in a regular Gaussian case. An example of such specifically Gaussian implication is the property called *weak transitivity* in [34]:

$$\{\, i \perp\!\!\!\perp j \,|\, kL \;\text{ and }\; i \perp\!\!\!\perp j \,|\, L \,\} \quad \Rightarrow \quad \{\, i \perp\!\!\!\perp k \,|\, L \;\text{ or }\; j \perp\!\!\!\perp k \,|\, L \,\}\,.$$

One can consider many analogous *implication tasks*: given disjoint sets of (elementary) CI statements $\mathcal{A}$ and $\mathcal{C}$, decide whether, for any (regular) Gaussian distribution P, the simultaneous validity of the CI statements from $\mathcal{A}$ with respect to P implies that at least one of the CI statements from $\mathcal{C}$ is valid with respect to P.

Using Fero's simple Gaussian CI criterion, one can reformulate any such implication task as the problem of recognizing whether a certain semi-algebraic set (= a set in a Euclidean space specified by finitely many polynomial inequalities) is empty or not. More specifically, the space is $\mathbb{R}^{N \times N}$, the CI statements turn into polynomial equations on entries in Σ and the assumption that Σ is positive definite can be formulated in the form of polynomial inequalities. Fero suggested approximating this problem by relaxing the constraints so that the relaxed problem can be treated easily by tools of computational algebra [4], such as the Gröbner basis computation. For example, one can consider the space of symmetric matrices with complex entries whose main square sub-matrices are regular.

The main result, [28, Proposition 1], gives a necessary and sufficient condition for the emptiness of the relaxed semi-algebraic set, which is formulated in terms of rings of polynomials. This sufficient condition for the validity of the respective Gaussian

CI implication can then be tested by tools of computational algebra and allows one to confirm all valid Gaussian CI implications in the case of three variables. On the other hand, an example was given ([28, Example 3]) that not every valid Gaussian CI implication over four variables can be verified by this particular relaxation technique.

6.2.2. Gaussoids

In a later journal paper [13] Fero and his student Radim Lněnička introduced an abstract concept of a *gaussoid*: it is defined as a (standard) semigraphoid satisfying basic (regular) Gaussian CI implications. Every regular Gaussian CI structure is a gaussoid and the converse holds in the case of at most three variables. Both regular Gaussian CI structures and gaussoids were observed in [13, § 2] to be closed under the duality transformation mentioned in Section 4.3.3: the dual of a Gaussian CI structure corresponding to Σ corresponds to the inverse Σ^{-1}.

It was shown in [13, § 3] that every semigraphoid induced by a simple undirected graph over N through a natural separation test is probabilistically representable by a regular Gaussian distribution. This result implies that the same is true for the duals of these semigraphoids. In [13, § 4], all gaussoids over N with $|N| = 4$ were characterized: there are 679 such gaussoids falling into 58 permutational types, while the number of types of "separational" graphical semigraphoids is only 11. The characterization of those of them which are representable by regular Gaussian distributions was given in [13, § 5]: there are 629 regular Gaussian CI structures over four variables falling into 53 permutational types. All standard regular Gaussian CI structures over four variables had thus been characterized.

7. GRAPHICAL AND BINARY CI STRUCTURES

Three remaining papers on CI by Fero are mentioned in this last section.

7.1. Graphical models of CI structure

The first paper by Fero devoted specifically to graphical modeling of CI structures was his 1992 note [17]. It was about CI structures over N ascribed to *undirected graphs* over N which are simple, that is, without multiple edges and loops. Various ways to relate the (standard) CI structure of a random vector over N and an undirected graph G over N were proposed in [34]: one can specifically mention *pairwise* Markov property, *local* Markov property, and the *global* Markov property (defined through a separation test) with respect to G. Fero gave a graphical condition on G in [17, Proposition 1], characterizing situations when the local and global Markov properties with respect to G coincide; it is formulated in terms of forbidden induced subgraphs of G. Analogously, [17, Proposition 2] characterizes the coincidence of the local and pairwise Markov properties.

Statistical models assigned to undirected graphs were considered in Fero's 2012 paper [29]. These models are formally defined as classes of probability distributions on a fixed joint sample space X_N which satisfy the global Markov property with respect to graphs. One can distinguish the discrete case and the regular Gaussian case and, typically, restricts the attention to probability distributions with *strictly positive* densities. By the well-known Hammersley-Clifford theorem [11, § 3.2], globally Markovian probability distributions with respect to an undirected graph G coincide with those whose densities factorize according to G.

Any such a class of distributions $\mathcal{P}$ (= a statistical model ascribed to an undirected graph) is known to be *log-convex*, which means that, whenever $P, Q \in \mathcal{P}$ are determined by densities p and q on X_N and $\alpha \in (0,1)$, then the probability distribution whose density is proportional to the function $x \in \mathsf{X}_N \mapsto p(x)^\alpha \cdot q(x)^{1-\alpha}$ belongs to $\mathcal{P}$. The main result [29, Theorem 1] says that, whenever a class $\mathcal{P}$ of (positive) discrete distributions on X_N is log-convex and defined in terms of CI restrictions (over N), then it has to coincide with the statistical model assigned to an undirected graph G over N. An analogous result [29, Theorem 2] for a regular Gaussian case claims that the same implication is true for a class $\mathcal{P}$ of regular Gaussian distributions on $\mathbb{R}^N \equiv \mathsf{X}_N$.

7.2. Binary CI structures

The task of characterizing CI structures induced by binary random vectors is touched upon in Fero's last paper among those devoted to CI, [30]. The main result [30, Theorem 1] can be interpreted as an extension of an algorithmic characterization of partial binary representability. Specifically, the representability is *partial* in the sense that it concerns special CI statements $i \perp\!\!\!\perp j$ and $i \perp\!\!\!\perp j \,|\, k$, where the elements $i, j, k \in N$ are distinct. One has a prescribed set $\mathcal{L}$ of such special triplets over N and is interested in the existence of a *binary* random vector $\boldsymbol{\xi}$ over N such that a triplet belongs to $\mathcal{L}$ iff it represents a valid CI statement with respect of $\boldsymbol{\xi}$. The problem is extended by prescribing a pattern σ of *signs of covariances* between variables ξ_i and ξ_j for $i \neq j$, where an additional neutral sign is allowed to cover the case of independence $\xi_i \perp\!\!\!\perp \xi_j$.

A necessary and sufficient condition for the existence of a binary vector $\boldsymbol{\xi}$ over N having the prescribed $\mathcal{L}$ and σ is given, involving the solvability of a certain simple system of linear equalities and inequalities. The point here is that the existence of a solution to such a system of linear constraints can be tested by linear programming tools; from an algorithmic point of view, it is a problem of polynomial complexity. An additional observation says that the condition is also equivalent to (a seemingly stronger requirement of) the existence of a *positive binary* random vector $\boldsymbol{\xi}'$ representing $\mathcal{L}$ and σ. An interesting fact is that the main mathematical tool used in the proof was Fourier-Stieltjes transformation, which had already been used by Fero in his thesis [15].

The main result of [30] was formulated for standard CI structures that do not involve functional dependencies among random variables. It was then modified in [30, § 4] to cover the case of augmented CI structures that involve functional dependencies. A source of motivation for the results in [30, § 6], extending those from [3], was to offer methods to recognize abstract ternary relations of *causal betweenness* discussed in [2]; these relations have philosophical motivation dating back to Reichenbach [35].

ACKNOWLEDGEMENT

The research of the author is supported by GAČR Project No. 19-04579S.

(Received March 26, 2019)

REFERENCES

[1] G. Birkhoff: Lattice Theory. Third edition. American Mathematical Society, Colloquium Publications 25, Providence 1995.

[2] V. Chvátal and B. Wu: On Reichenbach's causal betweenness. Erkenntnis *76* (2012), 41–48. DOI:10.1007/s10670-011-9321-z

[3] V. Chvátal, F. Matúš, and Y. Zwólš: Patterns of conjuctive forks. A 2016 manuscript `arXiv/1608.03949`.

[4] D. Cox, J. Little, and D. O'Shea: Ideals, Varieties, and Algorithms. Springer, New York 1997.

[5] A. P. Dawid: Conditional independence in statistical theory. J. Royal Statist. Soc. B *41* (1979), 1, 1–31. DOI:10.1111/j.2517-6161.1979.tb01052.x

[6] A. P. Dawid: Separoids: a mathematical framework for conditional independence and irrelevance. Ann. Math. Artif. Intell. *32* (2001), 1/4, 335–372. DOI:10.1023/a:1016734104787

[7] S. Fujishige: Submodular Functions and Optimization. Second edition. Elsevier, Amsterdam 2005.

[8] D. Geiger, A. Paz, and J. Pearl: Axioms and algorithms for inferences involving probabilistic independences. Inform. Comput. *91* (1991), 1, 128–141. DOI:10.1016/0890-5401(91)90077-f

[9] D. Geiger and J. Pearl: Logical and algorithmic properties of conditional independence and graphical models. Ann. Statist. *21* (1993), 4, 2001–2021. DOI:10.1214/aos/1176349407

[10] A. W. Ingleton: Conditions for representability and transversality of matroids. In: Lecture Notes in Computer Science 211, Springer, 1971, pp. 62–67. DOI:10.1007/bfb0061075

[11] S. L. Lauritzen: Graphical Models. Clarendon Press, Oxford 1996.

[12] M. Loéve: Probability Theory, Foundations, Random Sequences. Van Nostrand, Toronto 1955.

[13] R. Lněnička and F. Matúš: On Gaussian conditional independence structures. Kybernetika *43* (2007), 3, 327–342. CORPUS:16748330

[14] F. M. Malvestuto: A unique formal system for binary decompositions of database relations, probability distributions and graphs. Inform. Sci. *59* (1992), 21–52. DOI:10.1016/0020-0255(92)90042-7

[15] F. Matúš: Independence and Radon Projections on Compact Groups (in Slovak). Thesis for CSc. Degree in Theoretical Computer Science, Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague 1988.

[16] F. Matúš: Abstract functional dependency structures. Theor. Computer Sci. *81* (1991), 117–126. DOI:10.1016/0304-3975(91)90319-w

[17] F. Matúš: On equivalence of Markov properties over undirected graphs. J. Appl. Probab. *29* (1992), 745–749. DOI:10.1017/s0021900200043552

[18] F. Matúš: Ascending and descending conditional independence relations. In: Trans. 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, volume B, Academia, Prague 1992, pp. 189–200.

[19] F. Matúš: Stochastic independence, algebraic independence and abstract connectedness. Theor. Computer Sci. *134* (1994), 455–471. DOI:10.1016/0304-3975(94)90248-8

[20] F. Matúš: Probabilistic conditional independence structures and matroid theory: background. Int. J. General Systems *22* (1994), 185–196. DOI:10.1080/03081079308935205

[21] F. Matúš: Extreme convex set functions with many non-negative differences. Discrete Math. *135* (1994), 177–191. DOI:10.1016/0012-365x(93)e0100-i

[22] F. Matúš and M. Studený: Conditional independences among four random variables I. Combinatorics, Probability and Computing *4* (1995), 269–278. DOI:10.1017/s0963548300001644

[23] F. Matúš: Conditional independences among four random variables II. Combinator. Probab. Comput. *4* (1995), 407–417. DOI:10.1017/s0963548300001747

[24] F. Matúš: Conditional independence structures examined via minors. Ann. Math. Artif. Intell. *21* (1997), 99–128. DOI:10.1023/a:1018957117081

[25] F. Matúš: Conditional independences among four random variables III: final conclusion. Combinator. Probab. Comput. *8* (1999), 269–276. DOI:10.1017/s0963548399003740

[26] F. Matúš: Lengths of semigraphoid inferences. Ann. Math. Artif. Intell. *35* (2002), 287–294. DOI:10.1023/a:1014525817725

[27] F. Matúš: Towards classification of semigraphoids. Discr. Math. *277* (2004), 115–145. DOI:10.1016/s0012-365x(03)00155-9

[28] F. Matúš: Conditional independences in Gaussian vectors and rings of polynomials. In: Proc. WCII 2002, Lecture Notes in Artificial Intelligence 3301, Springer, Berlin 2005, pp. 152–161. CORPUS:1893239

[29] F. Matúš: On conditional independence and log-convexity. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques *48* (2012), 4, 1137–1147. DOI:10.1214/11-aihp431

[30] F. Matúš: On patterns of conditional independences and covariance signs among binary variables. Acta Math. Hungar. *154* (2018), 2, 511–524. DOI:10.1007/s10474-018-0799-6

[31] M. Mouchart and J. M. Rolin: A note on conditional independences with statistical applications. Statistica *44* (1984), 557–584. ECON:630

[32] H. Q. Nguyen: Semimodular functions and combinatorial geometries. Trans. Amer. Math. Soc. *238* (1978), 355–383. DOI:10.1090/S0002-9947-1978-0491269-9

[33] J. G. Oxley: Matroid Theory. Oxford University Press, Oxford 1992.

[34] J. Pearl: Probabilistic Reasoning in Intelligent Systems – Networks of Plausible Inference. Morgan Kaufmann, San Francisco 1988.

[35] H. Reichenbach: The Direction of Time. University of California Press, Los Angeles 1956.

[36] W. Spohn: Stochastic independence, causal independence and shieldability. J. Philosoph. Logic *9* (1980), 1, 73–99. DOI:10.1007/bf00258078

[37] M. Studený: Conditional independence relations have no finite complete characterization. In: Trans. 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, volume B, Academia, Prague 1992, pp. 377–396.

[38] M. Studený: Structural semigraphoids. Int. J. General Syst. *22* (1994), 207–217. DOI:10.1080/03081079308935207

[39] M. Studený: Semigraphoids and structures of probabilistic conditional independence. Ann. Math. Artif. Intell. *21* (1997), 71–98. DOI:10.1023/a:1018905100242

[40] M. Studený: Probabilistic Conditional Independence Structures. Springer, London 2005.

[41] H. Whitney: On the abstract properties of linear dependence. Amer. J. Math. *57* (1935), 3, 509–533. DOI:10.2307/2371182

[42] J. Whittaker: Graphical Models in Applied Multivariate Statistics. John Wiley and Sons, Chichester 1990.

[43] R. W. Yeung: Information Theory and Network Coding. Springer, New York 2008.

[44] Z. Zhang and R. W. Yeung: A non-Shannon-type conditional inequality of information quantities. IEEE Trans. Inform. Theory *43* (1997), 1982–1986. DOI:10.1109/18.641561

[45] G. M. Ziegler: Lectures on Polytopes. Springer, New York 1995.

Milan Studený, Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: studeny@utia.cas.cz