Independence Models

Petr Šimeček
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Prague, Czech Republic
simecek@atrey.karlin.mff.cuni.cz

Abstract

In the graphical models field, a graph (nodes connected by edges) is used to represent the conditional independence structure induced by a set of random variables. The paper summarizes the known results on conditional independence structures that can be induced by four random variables and inspects a possibility of using these results to the structure learning.

1 Introduction

Theory of graphical models like Bayesian networks has become an essential part of probabilistic reasoning. The conditional independence relationships appear naturally in such highly structured stochastic systems. An interesting question coming originally from the work of J. Pearl (cf. [10]) is the problem of probabilistic representability, i.e. for which lists of conditional independence constraints (here called "independence models") there exists a random vector satisfying these and only these conditional independencies. The problem is usually examined in the given distributional framework.

It was proved by M. Studený in [15] that that there is no finite characterization (i.e. finite set of inference rules) of independence models representable by a collection of discrete variables. Analogical results (characterization by forbidden minors) also exist for binary and Gaussian distributional frameworks, cf. [12]. Therefore, the only hope to find the characterization of representable independence models is to restrict the number of random variables.

F. Matúš and M. Studený characterized models that are representable by a vector consisting of four (or less) discrete variables in a series of papers [5], [6] and [7]. Later on, F. Matúš and R. Lněnička found all models representable by a regular Gaussian distribution over three and four variables (cf. [8] and [2], respectively). These results were further generalized to (general) Gaussian distributional framework, cf. [13]. Partial results also exists for representability in binary and positive discrete frameworks, cf. [14].

The paper summarizes known answers to the representability question. Further, it is discussed using the knowledge of representable models to the structure learning. The most difficult obstacle is how to fit data to the given list of conditional independencies. Some simulations are presented in the regular Gaussian distributional framework.

2 Basics Concepts

For the reader's convenience, auxiliary results related to probability theory and independence models are recalled in this section.

The object of our interest will be a random vector $\boldsymbol{\xi} = (\xi_a)_{a\in N}$ indexed by a finite set $N = \{1,2,\dots,n\}$. Its distribution will be denoted by P. For $A \subseteq N$, a subvector $(\xi_a)_{a\in A}$ is denoted by $\boldsymbol{\xi}_A$; $\boldsymbol{\xi}_\emptyset$ is presumed to be a constant. Analogously, if $\boldsymbol{x} = (x_a)_{a\in N}$ is a constant vector then $\boldsymbol{x}_A$ is an appropriate coordinate projection.

The singleton $\{a\}$ will be shorten by a and the union of sets $A \cup B$ will be written simply as juxtaposition AB. For a square matrix $\mathbf{\Sigma}$, let $\mathbf{\Sigma}^{-1}$, $\mathbf{\Sigma}^{-}$, $|\mathbf{\Sigma}|$ and $\dim(\mathbf{\Sigma})$ denote its inverse, generalized inverse, determinant and rank, i.e. the dimension of the space spanned by its rows (or columns), respectively.

Provided A, B, C are pairwise disjoint subsets of N, "$\boldsymbol{\xi}_A \perp\!\!\!\perp \boldsymbol{\xi}_B | \boldsymbol{\xi}_C$" stands for a statement $\boldsymbol{\xi}_A$ and $\boldsymbol{\xi}_B$ are conditionally independent given $\boldsymbol{\xi}_C$. In particular, unconditional independence ($C = \emptyset$) is abbreviated as $\boldsymbol{\xi}_A \perp\!\!\!\perp \boldsymbol{\xi}_B$.

2.1 Gaussian distributional framework

A **Gaussian distribution** of a random vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ is a probability distribution specified by its characteristic function

$$\varphi_{\boldsymbol{\xi}}(\mathbf{t}) = \mathrm{E}\exp(i\mathbf{t}'\boldsymbol{\xi}) = \exp\left(i\mathbf{t}'\boldsymbol{\mu} - \frac{\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}{2}\right),$$

where the vector $\boldsymbol{\mu}$ and the symmetric positive semi–definite matrix $\mathbf{\Sigma}$ are mean and variance parameters, respectively. If $\mathbf{\Sigma}$ is regular, the distribution is called **regular Gaussian distribution** and has a density with respect to Lebesgue measure

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right).$$

Given a matrix $\mathbf{\Sigma} = (\sigma_{a\cdot b})_{a,b\in\{1,\dots,n\}}$ and A, B non-empty subsets of $\{1, 2, \dots, n\}$, the submatrix with A–rows and B–columns will be denoted by

$$\mathbf{\Sigma}_{A\cdot B} = (\sigma_{a\cdot b})_{a\in A,\ b\in B}.$$

The following lemma shows that both marginal and conditional distributions of a Gaussian distribution are also Gaussian.

Lemma 1. *Let A and B be disjoint subsets of $N = \{1, 2, \ldots, n\}$.*

i) The marginal distribution $\boldsymbol{\xi}_A$ is Gaussian distribution with the variance matrix $\boldsymbol{\Sigma}_{A \cdot A}$.

ii) The conditional distribution of $\boldsymbol{\xi}_A$ given $\boldsymbol{\xi}_B = \mathbf{x}_B$ is a Gaussian distribution with the variance matrix $\boldsymbol{\Sigma}_{A \cdot A|B} = \boldsymbol{\Sigma}_{A \cdot A} - \boldsymbol{\Sigma}_{A \cdot B} \boldsymbol{\Sigma}^{-}_{B \cdot B} \boldsymbol{\Sigma}_{B \cdot A}$, where $\boldsymbol{\Sigma}^{-}_{B \cdot B}$ is any generalized inverse of $\boldsymbol{\Sigma}_{B \cdot B}$.

iii) Moreover, if $\boldsymbol{\Sigma} > 0$ then also $\boldsymbol{\Sigma}_{A \cdot A} > 0$ and $\boldsymbol{\Sigma}_{A \cdot A|B} > 0$.

Proof. See [1], pp.256. □

Let us emphasize that the variance matrix of conditional distribution $\boldsymbol{\Sigma}_{A \cdot A|B}$ does not depend on the choice of $\mathbf{x}_B$.

Lemma 2. *Let a and b be distinct elements of $\{1, \ldots, n\}$ and $C \subseteq \{1, \ldots, n\} \setminus ab$:*

i) $\xi_a \perp\!\!\!\perp \xi_b \iff \sigma_{a \cdot b} = 0$.

ii) If $\boldsymbol{\Sigma}_{B \cdot B} > 0$, then $\xi_a \perp\!\!\!\perp \xi_c | \boldsymbol{\xi}_B \iff |\boldsymbol{\Sigma}_{aB \cdot Bc}| = 0$.

iii) If $D \subset C$ such that $\boldsymbol{\Sigma}_{D \cdot D} > 0$ and $\dim(\boldsymbol{\Sigma}_{D \cdot D}) = \dim(\boldsymbol{\Sigma}_{C \cdot C})$, then $\xi_a \perp\!\!\!\perp \xi_b | \boldsymbol{\xi}_C \iff \xi_a \perp\!\!\!\perp \xi_b | \boldsymbol{\xi}_D \iff |\boldsymbol{\Sigma}_{aD \cdot bD}| = 0$.

Proof. The first part is a well-known fact (cf. [1], pp.257). Let us expand the matrix $\Sigma_{aB \cdot cB}$ with using easily verifiable identity

$$\begin{pmatrix} R & S \\ T & U \end{pmatrix} = \begin{pmatrix} I & -SU^{-1} \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} R - SU^{-1}T & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} I & 0 \\ -U^{-1}T & I \end{pmatrix}^{-1},$$

to derive that under the assumption $|\boldsymbol{\Sigma}_{B \cdot B}| \neq 0$

$$|\boldsymbol{\Sigma}_{aB \cdot cB}| = 0 \Leftrightarrow \sigma_{a \cdot c} - \boldsymbol{\Sigma}_{a \cdot B} \boldsymbol{\Sigma}^{-1}_{B \cdot B} \boldsymbol{\Sigma}_{B \cdot c} = 0 \Leftrightarrow (\boldsymbol{\Sigma}_{ac \cdot ac|B})_{a \cdot c} = 0 \Leftrightarrow \xi_a \perp\!\!\!\perp \xi_c | \xi_B.$$

The third part follows from Lemma 1 ii) by constructing the generalized inverse (cf. [11], section 1b.5) as follows

$$\boldsymbol{\Sigma}^{-}_{C \cdot C} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{D \cdot D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

□

2.2 Discrete distributional framework

A random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ is called **discrete** if each ξ_a takes values in a state space X_a such that $1 < |X_a| < \infty$. In particular, $\boldsymbol{\xi}$ is called binary if $\forall a : \ |X_a| = 2$. Further, a discrete random vector $\boldsymbol{\xi}$ is called **positive** if

$$\forall \boldsymbol{x} \in \boldsymbol{X} = \prod_{a=1}^{n} X_a : \quad 0 < P(\boldsymbol{\xi} = \boldsymbol{x}) < 1.$$

In the case of discretely distributed random vector, variables ξ_a and ξ_b are independent given $\boldsymbol{\xi}_C$ if and only if for any $\boldsymbol{x}_{abC} \in \boldsymbol{X}_{abC}$

$$P(\boldsymbol{\xi}_{abC} = \boldsymbol{x}_{abC}) P(\boldsymbol{\xi}_C = \boldsymbol{x}_C) = P(\boldsymbol{\xi}_{aC} = \boldsymbol{x}_{aC}) P(\boldsymbol{\xi}_{bC} = \boldsymbol{x}_{bC}).$$

2.3 Independence models

Let $N = \{1, 2, \ldots, n\}$ be a finite set and $\mathcal{T}_N$ denotes the set of all pairs $\langle ab|C\rangle$ such that ab is an (unordered) couple of distinct elements of N and $C \subseteq N \setminus ab$.

Subsets of $\mathcal{T}_N$ will be referred here as formal **independence models** over N. The independence model $\mathcal{I}(\boldsymbol{\xi})$ induced by a random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ is the independence model over $\{1, 2, \ldots, n\}$ defined as follows

$$\mathcal{I}(\boldsymbol{\xi}) = \{\langle xy|Z\rangle;\ \xi_x \perp\!\!\!\perp \xi_y | \boldsymbol{\xi}_Z\}\,.$$

Let us emphasize that an independence model $\mathcal{I}(\boldsymbol{\xi})$ uniquely determines also all other conditional independencies among subvectors of $\boldsymbol{\xi}$ (cf. [3]).

Diagrams proposed by R. Lněnička will be used here for a visualisation of independence model I over N such that $|N| \leq 4$. Each element of N is plotted as a dot. If $\langle ab|\emptyset\rangle \in I$ then dots corresponding to a and b are joined by a line. If $\langle ab|c\rangle \in I$ then we put a line between dots corresponding to a and b and add small line in the middle pointing in c–direction. If both $\langle ab|c\rangle$ and $\langle ab|d\rangle$ are elements of I, then only one line with two small lines in the middle is plotted. Finally, if $\langle ab|cd\rangle \in I$ is visualised by a brace between a and b. See example in Figure 1.

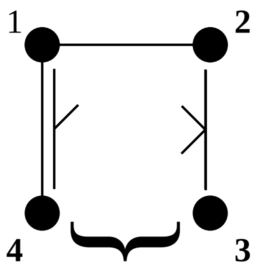

Figure 1: Diagram of the independence model $I = \big\{\langle 12|\emptyset\rangle,\ \langle 23|1\rangle,\ \langle 23|4\rangle,\ \langle 34|12\rangle,\ \langle 14|\emptyset\rangle,\ \langle 14|2\rangle\big\}$.

Two independence models I and J over N are **isomorphic** if there exists a permutation π on N such that

$$\langle xy|Z\rangle \in I \Leftrightarrow \langle \pi(x)\pi(y)|\pi(Z)\rangle \in J,$$

where $\pi(Z)$ stands for $\{\pi(z);\ z \in Z\}$. See Figure 2 for an example of three isomorphic models.

An equivalence class of independence models with respect to the isomorphic relation will be referred as **type**.

If I is an independence model over $N = \{1, \ldots, n\}$ and E, F are disjoint subsets of N then let us define the **minor** $I\,[^F_E$ as an independence model over $N \setminus EF$

$$I\,[^F_E = \{\langle ab|C\rangle;\ E \cap (abC) = \emptyset, \langle ab|CF\rangle \in I\}\,.$$

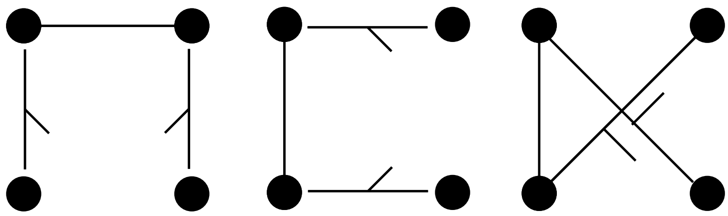

Figure 2: Example of three isomorphic models.

Note that $I\,[^{\emptyset}_{\emptyset} = I$ and $(I\,[^{F_1}_{E_1})\,[^{F_2}_{E_2} = I\,[^{F_1F_2}_{E_1E_2}$.

An independence model I is said to be Gaussian (g–), regular Gaussian (g^+–), discrete (d–) and/or positive discrete (d^+–) **representable** if there exists a random vector $\boldsymbol{\xi}$ such that it is Gaussian, regular Gaussian, discrete or positive discrete, respectively, and $I = \mathcal{I}(\boldsymbol{\xi})$.

It is easy to see that models of the same type are either all representable or none of them is representable. Consequently, we can classify the entire type as representable or non-representable.

Lemma 3. *Let a, b, c be distinct elements of $N = \{1, \ldots, n\}$ and $D \subseteq N \setminus abc$. If an independence models I over N is either discrete or Gaussian representable, then*

$$\big(\{\langle ab|cD\rangle, \langle ac|D\rangle\} \subseteq I\big) \iff \big(\{\langle ac|bD\rangle, \langle ab|D\rangle\} \subseteq I\big).$$

Moreover, if I is positively discrete or regular Gaussian representable, then

$$\big(\{\langle ab|cD\rangle, \langle ac|bD\rangle\} \subseteq I\big) \Longrightarrow \big(\{\langle ab|D\rangle, \langle ac|D\rangle\} \subseteq I\big).$$

Proof. These are so called "semigraphoid" and "graphoid" properties, cf. [1] or [16] for more details. □

In the discrete distributional framework, the intersection of representable models is also representable (cf. [18], pp. 5., for proof):

Lemma 4. *If $I = \mathcal{I}(\boldsymbol{\xi})$ and $I^* = \mathcal{I}(\boldsymbol{\xi}^*)$ are d–representable, then $I \cap I^*$ is also d–representable. In particular, if they are d^+–representable then $I \cap I^*$ is d^+–representable, too.*

Lemmas 1 and 4 follows that if I is an independence model representable in any of the above mentioned sense, then all its minors $I\,[^{F}_{E}$ are also representable in the same sense; i.e. the classes of g–/g^+–/d–/d^+–representable models are closed with respect to the operation of minorization[1].

Another useful result exists for g^+–representability (cf. [2] for proof):

Lemma 5. *If $I = \mathcal{I}(\boldsymbol{\xi})$ is g^+–representable model over N, then $I^* = \{\langle ab|N \setminus abC\rangle : \langle ab|C \in I\rangle\}$ is g^+–representable by Gaussian vector with the variance matrix that is inverse to the variance matrix of $\boldsymbol{\xi}$.*

[1] However, this property does not hold for the class of independence models representable by binary random vectors.

3 Representable models in the case $|N| = 4$

For N consisting of three or less elements, the problem of representability is easy to solve. In this case, all independence models not contradicting properties from Lemma 3 are discrete representable. The same hold for Gaussian representability with the exception of three types: $\{\langle ab|\emptyset\rangle, \langle bc|\emptyset\rangle\}$, $\{\langle ab|\emptyset\rangle, \langle bc|\emptyset\rangle, \langle ac|\emptyset\rangle\}$ and $\{\langle ab|\emptyset\rangle, \langle ab|c\rangle\}$ that are not g–representable (cf. [16] and [8]).

That is why we focus on $N = \{1, 2, 3, 4\}$ from now to the end of the section. The results presented here comes from [2] and [13] for the Gaussian distributional framework and from [17] and [4] for the discrete distributional framework. All results will be stated without proof. The reader should consult the cited sources.

The web page with data files and utilities related to these results is

`http://5r.matfyz.cz/skola/models`

3.1 Gaussian distributional framework

The g^+–representable models must necessarily have g^+–representable minors. The list of all types on $N = \{1, 2, 3, 4\}$ that have g^+–representable minors are plotted in Figure 3 on page 7, M1–M58. As shown in [2], the last 5 ones, M54–M58, are not g^+–representable. Therefore, there are 629 g^+–representable independence models corresponding to 53 types.

For the general Gaussian representability, in [13] the following two properties are shown:

Lemma 6. *Let I be a g–representable independence model and a, b, c and d distinct elements of N.*

i) if $\{\langle ab|c\rangle, \langle ac|b\rangle\} \subseteq I$ then either $\xi_b \simeq \xi_c$ or $\{\langle ab|\emptyset\rangle, \langle ac|\emptyset\rangle\} \subseteq I$,

ii) if $\{\langle ab|cd\rangle, \langle ac|bd\rangle\} \subseteq I$, then either $\xi_b \simeq \xi_c$ or $\{\langle ab|d\rangle, \langle ac|d\rangle\} \subseteq I$ or $\langle ad|bc\rangle \in I$,

where $\xi_b \simeq \xi_c$ stands for

$$\{\langle ab|c\rangle, \langle bd|c\rangle, \langle ac|b\rangle, \langle cd|b\rangle, \langle ab|cd\rangle, \langle bd|ac\rangle, \langle ac|bd\rangle, \langle cd|ab\rangle\} \subseteq I.$$

Thus, M1–M88 in Figure 3 are types having g–representable minors and not contradicting Lemma 6. However, it is proved in [12] that M54–M58 and M86–M88 are not g–representable. Therefore, there are 877 g–representable independence models corresponding to 80 types.

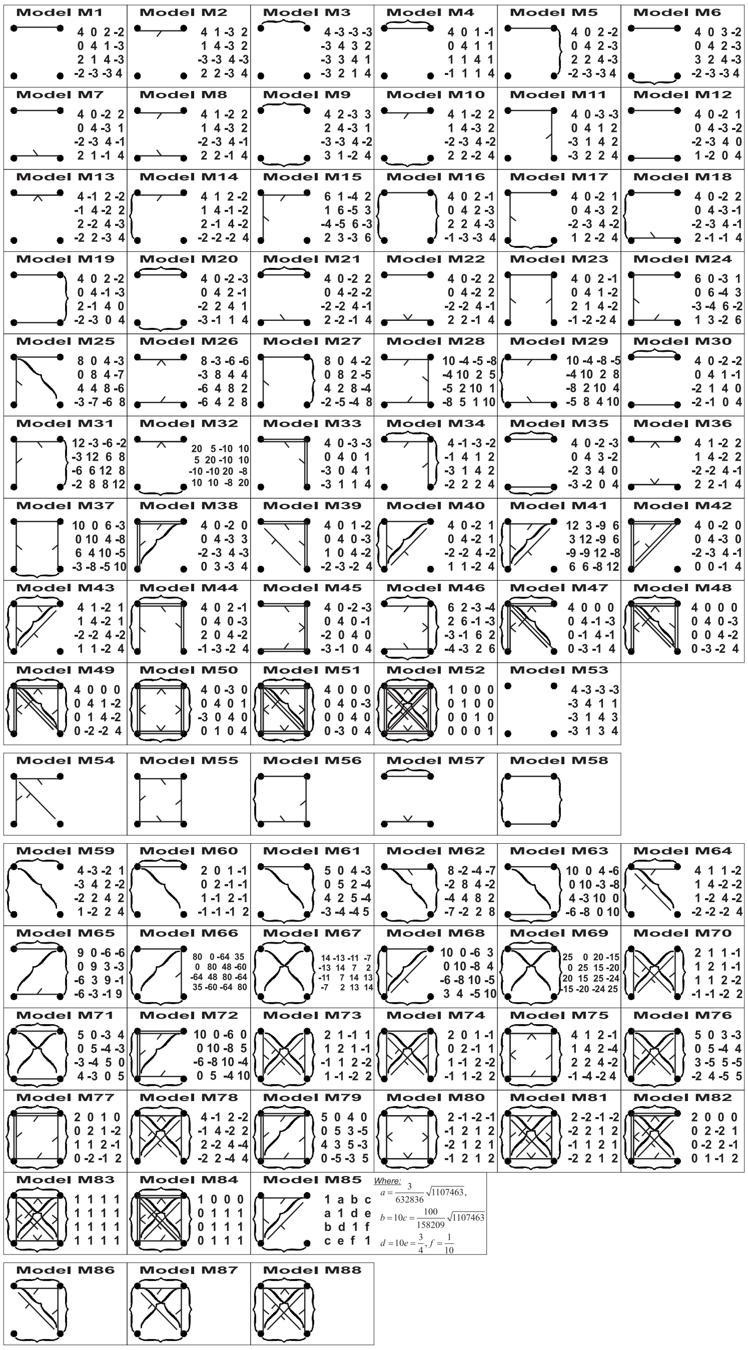

Figure 3: List of independence models M1–M88 and their g–representations.

3.2 Discrete distributional framework

In brief, due to Lemma 4 an intersection of two d–representable models is also d–representable. Therefore, the class of all d–representable models over N can be described by the set $\mathcal{C}$ of so called **irreducible** models, i.e. nontrivial d–representable models that cannot be written as an intersection of two other d–representable models. It is not difficult to evidence that a nontrivial independence model I is d–representable if and only if there exists $\mathcal{A} \subseteq \mathcal{C}$ such that

$$I = \bigcap_{C \in \mathcal{A}} C.$$

As shown in [5], [6] and [7] there are only only 13 types of irreducible models, see Figure 4, and 18478 d–representable independence models corresponding to 1098 types.

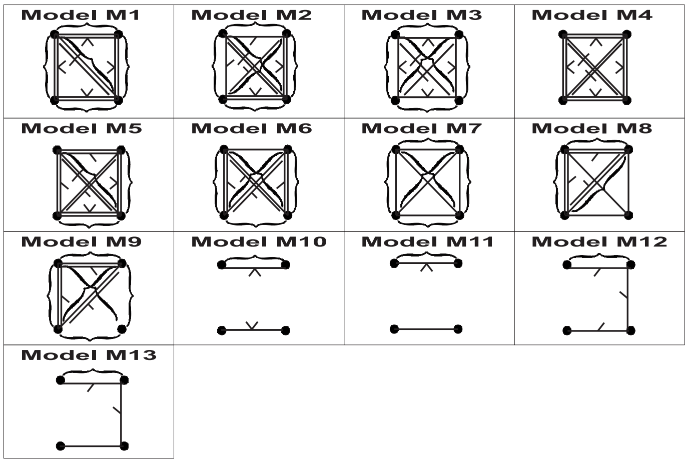

Figure 4: Irreducible types.

Surprisingly, only partial results exist for d^+–representability at this moment (cf. [14] for an overview).

4 Graphical Independence Models

In the applications, the independence model determining a class of probability distributions is not usually given as a list of prescribed conditional independencies but as an undirected or directed acyclic graph. That is not only for interpretation reasons but also because of nice properties of such models. The theory is described in details in [1] (or [9]). An independence model will be called **graphical** if it can be derived from some undirected graph by the global Markov property.

The main drawback of such approach is that there are much more representable models than just graphical ones. For instance, in the case $|N| = 4$ there are 11 graphical types only while there are 80 g–representable, 53 g^+–representable, 1098 d–representable and at least 299 d^+–representable types. On the other hand side, it is unfeasible to fit data to the most of non–graphical independence models.

To the end of the section we will restrict ourselves to the regular Gaussian distributional framework and take an advantage that all 53 g^+–representable models over $N = \{1, 2, 3, 4\}$ can be parametrized and fitted to data (by iterative likelihood maximization).

In the regular Gaussian distributional framework, the class of graphical models plays an important role: all graphical models are g^+–representable (cf. [2]) and they are regular exponential families (cf. [1]). More important, the latter property can also be reversed[2].

Lemma 7. *If the class of all regular Gaussian distribution corresponding to an independence model I is regular exponential family, then I is graphical.*

Proof. Without any loss of generality we may take into account just distributions with zero expected value, assume that the exponential family is minimal and the reference measure is Lebesgue.

Further, if $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^m)$ is a parameter then (to get the parametrization of Gaussian distributions) necessarily

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}'\left(\theta^1\boldsymbol{K}^1 + \cdots + \theta^m\boldsymbol{K}^m\right)\mathbf{x} - \psi_{\boldsymbol{\theta}}\right), \tag{1}$$

where $\boldsymbol{K}^1, \ldots, \boldsymbol{K}^m$ are some matrices and $\psi_{\boldsymbol{\theta}}$ is a normalization constant.

The independence model is graphical if and only if it contains only triplets of the type $\langle ab|N \setminus ab\rangle$ and triplets that can be derived from them by the lemma 3. In particular, for g^+–representation $\boldsymbol{\xi}$ with corresponding concentration matrix $\boldsymbol{K} = \boldsymbol{\Sigma}^{-1}$ it implies that $\mathcal{I}(\boldsymbol{\xi})$ is graphical if and only if

$$\forall abC \subseteq N: \ |\boldsymbol{K}_{aC\cdot bC}| = 0 \quad \Rightarrow \quad \text{There is a zero row/column in } \boldsymbol{K}_{\mathrm{aC\cdot bC}}. \tag{2}$$

On contrary, let us assume that for the parametrization (1) there exists a submatrix

$$\boldsymbol{K}_{aC\cdot bC} = \left(\theta^1\boldsymbol{K}^1 + \cdots + \theta^m\boldsymbol{K}^m\right)_{aC\cdot bC}$$

contradicting (2), i.e. having zero determinant for all $\boldsymbol{\theta}$ and no zero column at the same moment. However, that would mean that one column of $\boldsymbol{K}_{aC\cdot bC}$ is a linear combination of the others for any $\boldsymbol{\theta}$ following that for all $\boldsymbol{K}^1, \ldots, \boldsymbol{K}^m$ (and thus also $\boldsymbol{K}$) one column is the same linear combination of the others. But that is a contradiction that the exponential family really contains all distributions corresponding to given independence model. □

[2] I would like to thank F. Matúš for drawing my attention to that fact.

Now, lets start with the simulation. First, for each of 53 models the variance matrix and 1000 datalines have been generated. Then model search based on that dataset has been performed. The true model has been chosen in roughly half cases (45%). However, this number may be influenced by numerical instability of likelihood maximization for some models.

The future simulations should work out these problems, increase the sample size and inspect whether the "right" one of graphical independence models is chosen. The program for the statistical environment R may be downloaded from the above mentioned web page.

4.1 Acknowledgment

I would like to thank F. Matúš, M. Studený and R. Lněnička for inspiration and various help. This paper was supported by the grant GA ČR n. 201/05/H007.

References

[1] S. L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[2] R. Lněnička. On gaussian conditional independence structures. Technical report, ÚTIA AV ČR, 2005.

[3] F. Matúš. On equivalence of markov properties over undirected graphs. *Journal of Applied Probability*, 29:745–749, 1992.

[4] F. Matúš. Conditional independence structures examined via minors. *The Annals of Mathematics and Artificial Intelligence*, 21:99–128, 1997.

[5] F. Matúš and M. Studený. Conditional independences among four random variables I. *Combinatorics, Probability & Computing*, 4:269–278, 1995.

[6] F. Matúš. Conditional independences among four random variables II. *Combinatorics, Probability & Computing*, 4:407–417, 1995.

[7] F. Matúš. Conditional independences among four random variables III. *Combinatorics, Probability & Computing*, 8:269–276, 1999.

[8] F. Matúš. Conditional independences in gaussian vectors and rings of polynomials. In G. Kern-Isberner, W. Rödder, and F. Kulmann, editors, *Conditionals, Information, and Inference, International Workshop. WCII 2002*, pages 152–161. Springer, 2005.

[9] R.E. Neapolitan. *Learning Bayesian Networks.* Chapman & Hall/CRC, 2003.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, 1998.

[11] R. C. Rao. *Linear Statistical Inference and Its Applications.* John Wiley & Sons, 1973.

[12] P. Šimeček. Classes of gaussian, discrete and binary representable independence models have no finite characterization. Accepted to Prague Stochastics proceedings, 2006.

[13] P. Šimeček. Gaussian representation of independence models over four random variables. Accepted to COMPSTAT proceedings, 2006.

[14] P. Šimeček. A short note on discrete representability of independence models. Accepted to PGM proceedings, 2006.

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

[16] M. Studený. *Probabilistic Conditional Independence Structures.* Springer, 2005.

[17] M. Studený and P. Boček. CI-models arising among 4 random variables. In *Proceedings of the 3rd workshop WUPES*, pages 268–282, 1994.

[18] M. Studený and J. Vejnarová. The multiinformation function as a tool for measuring stochastic dependence. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 261–298. Kluwer, 1998.