Gaussian Representation of Independence Models over Four Random Variables

Petr Šimeček

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 18200 Prague, Czech Republic
simecek@atrey.karlin.mff.cuni.cz

Theory of graphical models like Bayesian networks has become an essential area in probabilistic reasoning. One way to introduce a model of this type is to give a list of conditional independence constraints. An interesting question coming originally from the work of J. Pearl (cf. [1]) is the problem of probabilistic representability, i.e. for which conditional independence constraints (here called *independence model*) there exists a random vector satisfying these and only these conditional independencies.

It was proved by M. Studený in [2] that there is no finite complete axiomatic characterization of representable independence models. Therefore, the only hope to find such characterization is to restrict the number of random variables. F. Matúš characterized models that are representable by a vector of four discrete variables in a series of papers [3], [4] and [5]. Later on, F. Matúš and R. Lněnička found all models representable by a regular Gaussian distribution over three and four variables (cf. [6] and [7], respectively).

In this paper, an alternative approach to the representability by a regular Gaussian vector over four variables is presented. In comparison with [7] it lacks mathematical beauty but it is much more straightforward. Moreover, the result is generalized to representations by general (=not necessarily regular) Gaussian distribution. For the reader's convenience definitions and lemmas related to Gaussian distribution and independence models are recalled in the first section. Several open problems can be found at the end of the paper.

The hope is that in the long run this approach could be utilized to make a model selection of Bayesian networks more effective. Classical model selection algorithms rely on the faithfulness assumption that data are generated from a distribution satisfying only the independencies given by some Bayesian network model. However, e.g. in the case of four variables there are only 25 different Bayesian networks while there are 18300 independence models representable by a discrete distribution (cf. [8], pp. 63). Estimating of independencies among quartets of variables instead of triplets as in PC algorithm (cf. [9], pp. 542) could better approximate the distribution by a Bayesian network.

1 Preliminaries

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

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

where the vector $\boldsymbol{\mu}$ and the symmetric positive semi–definite matrix $\boldsymbol{\Sigma}$ are mean and variance parameters, respectively. If $\boldsymbol{\Sigma}$ is regular, the distribution is called **regular Gaussian distribution**. Note all regular Gaussian distributions have finite multiinformation[1] and a density with respect to n–dimensional Lebesgue measure. For non-regular distributions (a part of $\boldsymbol{\xi}$ can be written as a linear combination of the rest) this generally does not hold.

Let us partition random vector $\boldsymbol{\xi}$ into two components

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{pmatrix}.$$

And let the variance matrix be partitioned accordingly into blocks as follows

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\varphi}} & \boldsymbol{\Sigma}_{\boldsymbol{\varphi},\boldsymbol{\psi}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\psi},\boldsymbol{\varphi}} & \boldsymbol{\Sigma}_{\boldsymbol{\psi}} \end{pmatrix}.$$

Lemma 1. *Marginal and conditional distributions of a Gaussian distribution are also Gaussian:*

i) The marginal distribution $\boldsymbol{\varphi}$ is Gaussian distribution with the variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}}$.
ii) The conditional distribution of $\boldsymbol{\varphi}$ given $\boldsymbol{\psi} = \mathbf{x}$ is Gaussian distribution with the variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}|\boldsymbol{\psi}} = \boldsymbol{\Sigma}_{\boldsymbol{\varphi}} - \boldsymbol{\Sigma}_{\boldsymbol{\varphi},\boldsymbol{\psi}}\boldsymbol{\Sigma}_{\boldsymbol{\psi}}^{-}\boldsymbol{\Sigma}_{\boldsymbol{\psi},\boldsymbol{\varphi}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}^{-}$ is any generalized inverse of $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}$.
iii) Moreover, if $\boldsymbol{\Sigma}$ is regular then the variance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}|\boldsymbol{\psi}}$ are also regular.

Proof. See [10], pp.256. □

Lemma 2. *Let x and y be distinct elements of $\{1, \ldots, n\}$ and Z equals $\{1, \ldots, n\} \setminus \{x, y\}$:*

i) The random variables ξ_x and ξ_y are independent (denoted by $\xi_x \perp\!\!\!\perp \xi_y$ or $\xi_x \perp\!\!\!\perp \xi_y|\emptyset$) if and only if the element of $\boldsymbol{\Sigma}$ in the row x and column y is zero.
ii) Provided the variance matrix of $\boldsymbol{\xi}_Z$ is regular: the random variables ξ_x and ξ_y are independent given $\boldsymbol{\xi}_Z$ (denoted by $\xi_x \perp\!\!\!\perp \xi_y|\boldsymbol{\xi}_Z$) if and only if a determinant of the matrix resulting from $\boldsymbol{\Sigma}$ by deleting the row x and column y is zero.

[1] that is Kullback–Leibler divergence, or relative entropy, between a joint distribution and a product of its marginals, cf. [8] pp. 24.

Proof. The first part is a well known fact (cf. [10], pp.257). It is possible to evidence the second part by an expansion of the determinant mentioned above and Lemma 1 ii). □

Note just the variance matrix $\boldsymbol{\Sigma}$ is used for determining conditional independencies. In addition a marginalization and Lemma 1 can be applied to judge the validity of independence statements such that Z is a proper subset of $\{1,\ldots,n\}\setminus\{x,y\}$. Provided $\boldsymbol{\xi}_Z$ is not regular, $\xi_x \perp\!\!\!\perp \xi_y|\boldsymbol{\xi}_Z$ if and only if $\xi_x \perp\!\!\!\perp \xi_y|\boldsymbol{\xi}_{Z^*}$ where Z^* is any subset of Z such that $\boldsymbol{\xi}_{Z^*}$ is regular and Z^* is maximal with respect to the relation "$\subseteq$".

An **independence model** I over a finite set N is a set of triples $\langle xy|Z\rangle$ where x, y are different elements of N and Z is a subset of $N\setminus\{x,y\}$. An independence model $\mathcal{I}(\boldsymbol{\xi})$ associated with a random vector $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_n)'$ is an independence model over $N=\{1,\ldots,n\}$ defined as follows

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

An independence model I is said to be generally/regularly **representable**[2] if there exists a general/regular Gaussian distribution $\boldsymbol{\xi}$ such that $I=\mathcal{I}(\boldsymbol{\xi})$. Let us emphasize that an independence model $\mathcal{I}(\boldsymbol{\xi})$ uniquely determines also all other conditional independencies among subvectors of $\boldsymbol{\xi}$ (cf. [11]).

Two independence models I and J over N are **permutably equivalent** if there exists a permutation π of 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 $\{\pi(z);\ z\in Z\}$.

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

Example 1. There are 5 regularly representable permutation classes of independence models over $N=\{1,2,3\}$:

$$\begin{aligned}
I_1 &= \emptyset\\
I_2 &= \{\langle 12|\emptyset\rangle\}\\
I_3 &= \{\langle 12|\{3\}\rangle\}\\
I_4 &= \{\langle 12|\emptyset\rangle,\langle 12|\{3\}\rangle,\langle 23|\emptyset\rangle,\langle 23|\{1\}\rangle\}\\
I_5 &= \{\langle 12|\emptyset\rangle,\langle 12|\{3\}\rangle,\langle 23|\emptyset\rangle,\langle 23|\{1\}\rangle,\langle 13|\emptyset\rangle,\langle 13|\{2\}\rangle\}
\end{aligned}$$

In addition there are two generally representable permutation classes that are not regularly representable:

$$\begin{aligned}
I_6 &= \{\langle 12|\{3\}\rangle,\langle 23|\{1\}\rangle\}\\
I_7 &= \{\langle 12|\{3\}\rangle,\langle 23|\{1\}\rangle,\langle 13|\{2\}\rangle\}
\end{aligned}$$

The proof is not extremely complicated and is left to the reader (or cf. [6]).

[2] In this paper,representable means Gaussian representable unless stated otherwise.

An independence model I over a set N after **marginalizing out** $k \in N$ is an independence model $I\lceil_k$ over a set $N \setminus k$ containing triples of I where k is not involved; that is

$$I\lceil_k = \{\langle xy|Z\rangle \in I;\ k \notin (\{x,y\} \cup Z)\}\ .$$

An independence model I over a set N after **conditioning on** $l \in N$ is an independence model $I\lceil^l$ over a set $N \setminus l$ containing triples of I where l is a part of the condition. More formally,

$$I\lceil^l = \{\langle xy|Z\rangle;\ \langle xy|Z \cup \{l\}\rangle \in I\}\ .$$

Independence models $I\lceil_k$ and $I\lceil^l$ are called **minors** of I.

Lemma 3. *If an independence model I is generally/regularly representable, then all its minors are generally/regularly representable.*

Proof. The representation of a minor is obtained by the corresponding marginalization or conditioning of the representation of I. □

2 Regular Gaussian Representations

In this section, all classes of permutation equivalence of regularly representable independence models over $N = \{1, 2, 3, 4\}$ will be enumerated. Let us recall that the problem was originally solved in [7] in a slightly different way.

Every regularly representable model must have regularly representable minors[3] (Lemma 3). Using a computer it was found out that there are 58 classes of permutation equivalence with regularly representable minors (see Appendix, M1 – M58).

The next step is a computer search for as many representations as possible. All integer symmetric positive definite matrices[4] with diagonal elements less or equal 24 were examined as a variance matrix and 53 regular representations were found (see Appendix, M1 – M53).

The remaining 5 classes cannot be representable because they do not fulfill rules derived from "independence implication". This is a complex inference tool based on so called structural imsets[5] and is beyond the scope of this short paper (cf. [8], pp. 114). It is possible to take advantage of "independence implication" by means of the java applet

`http://staff.utia.cas.cz/studeny/VerifyView.html` .

Moreover, in the next section it will be proved that these 5 classes are neither generally representable.

[3] See Example 1 for a list of possible minors.
[4] Some trivial symmetries were used to shorten the computation time.
[5] =integer vectors representing conditional independence relationships

3 General Gaussian Representation

In this section, all classes of permutation equivalence of (generally) representable independence models over $N = \{1, 2, 3, 4\}$ will be found. This is the main result of the paper. The following two lemmas seem to be crucial.

Lemma 4. *If an independence model I is representable $I = \mathcal{I}(\boldsymbol{\xi})$, then there exists its representation $\boldsymbol{\xi^*} = (\xi_1^*, \ldots, \xi_4^*)'$ such that*

$$\operatorname{Var} \xi_x^* = 1, \quad x = 1, \ldots, 4.$$

Proof. First, let us interchange each zero variance variable ξ_x and a new unit variance variable independent to a vector of remaining variables. After that just take

$$\boldsymbol{\xi^*} = \left((\operatorname{Var} \xi_1)^{-\frac{1}{2}} \cdot \xi_1, (\operatorname{Var} \xi_2)^{-\frac{1}{2}} \cdot \xi_2, (\operatorname{Var} \xi_3)^{-\frac{1}{2}} \cdot \xi_3, (\operatorname{Var} \xi_4)^{-\frac{1}{2}} \cdot \xi_4 \right)'. \square$$

Lemma 4 enables us to restrict our focus on distributions with a variance matrices of the following form

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{pmatrix}.$$

If two variables ξ_x and ξ_y are *functionally dependent* $\xi_x = \pm\xi_y$ (denoted by $\xi_x \simeq \xi_y$), then an element of $\boldsymbol{\Sigma}$ in the row x and column y is ± 1 and $\xi_x \perp\!\!\!\perp \xi_u | \xi_y$, $\xi_y \perp\!\!\!\perp \xi_u | \xi_x$, $\xi_x \perp\!\!\!\perp \xi_u | \boldsymbol{\xi_{\{y,v\}}}$ and $\xi_y \perp\!\!\!\perp \xi_u | \boldsymbol{\xi_{\{x,v\}}}$.

Lemma 5. *Let I be a representable independence model.*

i) If $\{\langle 12|\{3\}\rangle, \langle 13|\{2\}\rangle\} \subseteq I$ then either $\{\langle 12|\emptyset\rangle, \langle 13|\emptyset\rangle\} \subseteq I$ or $\xi_2 \simeq \xi_3$.
ii) If $\{\langle 12|\{3,4\}\rangle, \langle 13|\{2,4\}\rangle\} \subseteq I$, then either $\{\langle 12|\{4\}\rangle, \langle 13|\{4\}\rangle\} \subseteq I$ or $\langle 14|\{2,3\}\rangle \in I$ or $\xi_2 \simeq \xi_3$.

Proof. The required conditional independencies $\xi_1 \perp\!\!\!\perp \xi_2 | \xi_3$ and $\xi_1 \perp\!\!\!\perp \xi_3 | \xi_2$ are equivalent to $a = bd$ and $b = ad$ following either $a = b = 0$ ($\xi_1 \perp\!\!\!\perp \xi_2$ and $\xi_1 \perp\!\!\!\perp \xi_3$) or $|d| = 1$ ($\xi_2 \simeq \xi_3$). The proof of the second part is analogous. $\square$

There are 178 classes of permutation equivalence with representable minors. Lemma 5 allows us to show that 90 of them cannot be represented.

Exhaustive computer search for variance matrices found 79 representations and 1 more was found later without use of computer (see Appendix, M1–M53 and M59–M85).

The remaining 8 classes (M54–M58 and M86–M88) of permutation equivalence are not representable as proved in the following lemma.

Lemma 6. *If an independence model I is permutably equivalent to one of the following models $I_1, \ldots, I_8$*

i) $I_1 = \{\langle 34|\{1\}\rangle, \langle 14|\{2\}\rangle, \langle 24|\{3\}\rangle\}$
ii) $I_2 = \{\langle 12|\{3\}\rangle, \langle 34|\{2\}\rangle, \langle 24|\{1\}\rangle, \langle 13|\{4\}\rangle\}$
iii) $I_3 = \{\langle 34|\emptyset\rangle, \langle 12|\{3\}\rangle, \langle 24|\{1\}\rangle, \langle 13|\{2,4\}\rangle\}$
iv) $I_4 = \{\langle 34|\emptyset\rangle, \langle 12|\{3\}\rangle, \langle 12|\{4\}\rangle, \langle 34|\{1,2\}\rangle\}$
v) $I_5 = \{\langle 12|\emptyset\rangle, \langle 34|\emptyset\rangle, \langle 13|\{2,4\}\rangle, \langle 24|\{1,3\}\rangle\}$
vi) $I_6 = \{\langle 12|\{4\}\rangle, \langle 23|\{4\}\rangle, \langle 13|\{4\}\rangle, \langle 12|\{3,4\}\rangle, \langle 23|\{1,4\}\rangle, \langle 34|\{1,2\}\rangle, \langle 13|\{2,4\}\rangle\}$
vii) $I_7 = \{\langle 12|\{4\}\rangle, \langle 23|\{4\}\rangle, \langle 13|\{4\}\rangle, \langle 12|\{3,4\}\rangle, \langle 23|\{1,4\}\rangle, \langle 34|\{1,2\}\rangle, \langle 14|\{2,3\}\rangle, \langle 13|\{2,4\}\rangle, \langle 24|\{1,3\}\rangle\}$
viii) $I_8 = \{\langle 12|\{4\}\rangle, \langle 23|\{4\}\rangle, \langle 13|\{4\}\rangle, \langle 12|\{3\}\rangle, \langle 24|\{3\}\rangle, \langle 14|\{3\}\rangle, \langle 12|\{3,4\}\rangle, \langle 23|\{1,4\}\rangle, \langle 34|\{1,2\}\rangle, \langle 14|\{2,3\}\rangle, \langle 13|\{2,4\}\rangle, \langle 24|\{1,3\}\rangle\}$

then it is not representable.

Proof. **i)** The required conditional independencies follow $f = bc$, $c = ae$ and $e = df$. Therefore, either $c = e = f = 0$ or $\xi_1 \simeq \xi_2 \simeq \xi_3$. **ii)** In an analogical way it easily follows either $a = b = e = f = 0$ or $\xi_1 \simeq \xi_4$ and $\xi_2 \simeq \xi_3$. **iii)** Substituting $f = 0$, $a = bd$ and $e = ac$ into the equation for $\langle 13|\{2,4\}\rangle$ results in $b((1-d^2)+c^2d^2(1-b^2)) = 0$ and thus either $b = 0$ or $\xi_2 \simeq \xi_3$ or $\xi_2 \simeq \xi_4$. **iv)** Substituting $f = 0$, $a = bd$ and $e = \frac{a}{c}$ into the equation for $\langle 34|\{1,2\}\rangle$ results in $\frac{b}{c}(c^2(1-d^2)+d^2(1-b^2)) = 0$ and thus either $b = 0$ or $c = 0$ or $\xi_1 \simeq \xi_3$ or $\xi_1 \simeq \xi_2$. **v)** Unless $\xi_1 \simeq \xi_4$ or $\xi_2 \simeq \xi_3$ the substitution $a = f = 0$ into equations corresponding to $\langle 13|\{2,4\}\rangle$ and $\langle 24|\{1,3\}\rangle$ gives $b - be^2 + ecd = 0$ and $e - eb^2 + bcd = 0$. Substraction of the first equation multiplied by e and the second one multiplied by b results in $(b-e)(b+e) = 0$. Thus $b = 0$ or $e = 0$ or $b = \pm e$ and, similarly, $c = \pm d$ yielding $\xi_1 \simeq \xi_3$. **vi)-vii)** Unless $\xi_x \simeq \xi_y$ for distinct x and y, the substitution $a = ec$, $d = ef$ and $b = cf$ into equations for other independencies results in either $\xi_1 \simeq \xi_2$ or $\xi_1 \simeq \xi_4$ or $\xi_2 \simeq \xi_4$ yielding a contradiction. **viii)** If $\xi_3 \simeq \xi_4$ then the conditions $\xi_1 \perp\!\!\!\perp \xi_2 | \xi_3$ and $\xi_3 \perp\!\!\!\perp \xi_4 | \boldsymbol{\xi}_{\{1,2\}}$ results in $\xi_1 \simeq \xi_3$ or $\xi_2 \simeq \xi_3$. Otherwise, the argument from part vi) yields a contradiction. □

4 Conclusion and Open Problems

In this paper we have characterized all Gaussian representable independence models over four variables. A natural question arises whether it is possible to do the same for five variables. There are 366177 permutation equivalence classes with regularly representable minors. However, for most of them neither a representation nor the proof of its non-existence is known.

All numbers in variance matrices of representations in Appendix are rational, the only exception is a model M85. The question whether every Gaussian representable model is "rationally" representable remains open.

Last question is the relation between models representable by Gaussian and by categorical or binary distributions. It is easy to construct a model which is binary representable but not Gaussian representable. On the other hand, model M71 in Appendix is not categorically representable but it is generally (not regularly!) Gaussian representable. However, no regularly Gaussian independence model without a binary or categorical representation is known.

This research was supported by the grant GA ČR n. 201/05/H007.

References

1. Pearl J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann (1998)
2. Studený, M.: Conditional independence relations have no finite complete characterization. In: S. Kubík, J.A. Víšek (ed) Information Theory, Statistical Decision Functions and Random Processes. Transactions of the 11th Prague Conference. Kluwer (1992)
3. Matúš, F. and Studený, M.: Conditional independences among four random variables I. Combinatorics, Probability & Computing, **4**, 269–278 (1995)
4. Matúš, F.: Conditional independences among four random variables II. Combinatorics, Probability & Computing, **4**, 407–417 (1995)
5. Matúš, F.: Conditional independences among four random variables III. Combinatorics, Probability & Computing, **8**, 269–276 (1999)
6. Matúš, F.: Conditional independences in Gaussian vectors and rings of polynomials. In: Kern-Isberner G. (ed) Conditionals, Information, and Inference, International Workshop. WCII 2002. Springer, Berlin (2005)
7. Lněnička, R.: On conditional independences among four Gaussian variables. In: Kern-Isberner G. (ed) Proceedings of Conditionals, Information and Inference, International Workshop. WCII 2004. University Ulm (2004)
8. Studený, M.: Probabilistic Conditional Independence Structures. Springer(2005)
9. Neapolitan R.E.: Learning Bayesian Networks. Prentice Hall (2003)
10. Lauritzen, S.L.: Graphical Models. Oxford University Press (1996)
11. Matúš, F.: On equivalence of Markov properties over undirected graphs. Journal of Applied Probability, **29**, 745–749 (1992)

Appendix

Independence models mentioned in the text are plotted (accompanied with a variance matrix of the representation) in Figure 1 on the next page. Notation is adopted from [7]. Vertices of each square are numbered clockwise starting in the top left corner. A triple $\langle xy|\emptyset\rangle$ is visualized by a line between vertices x and y, a triple $\langle xy|\{u\}\rangle$ by a line between x and y with a small line in the middle pointing to u–direction. If both relations $\langle xy|\{u\}\rangle$ and $\langle xy|\{v\}\rangle$ take place then only one line with two small lines is plotted. Finally, a triple $\langle xy|\{u,v\}\rangle$ is coded by a brace between x and y. E.g., a model M57 is

$$\{\langle 12|\emptyset\rangle, \langle 34|\{1\}\rangle, \langle 34|\{2\}\rangle, \langle 12|\{3,4\}\rangle\}.$$

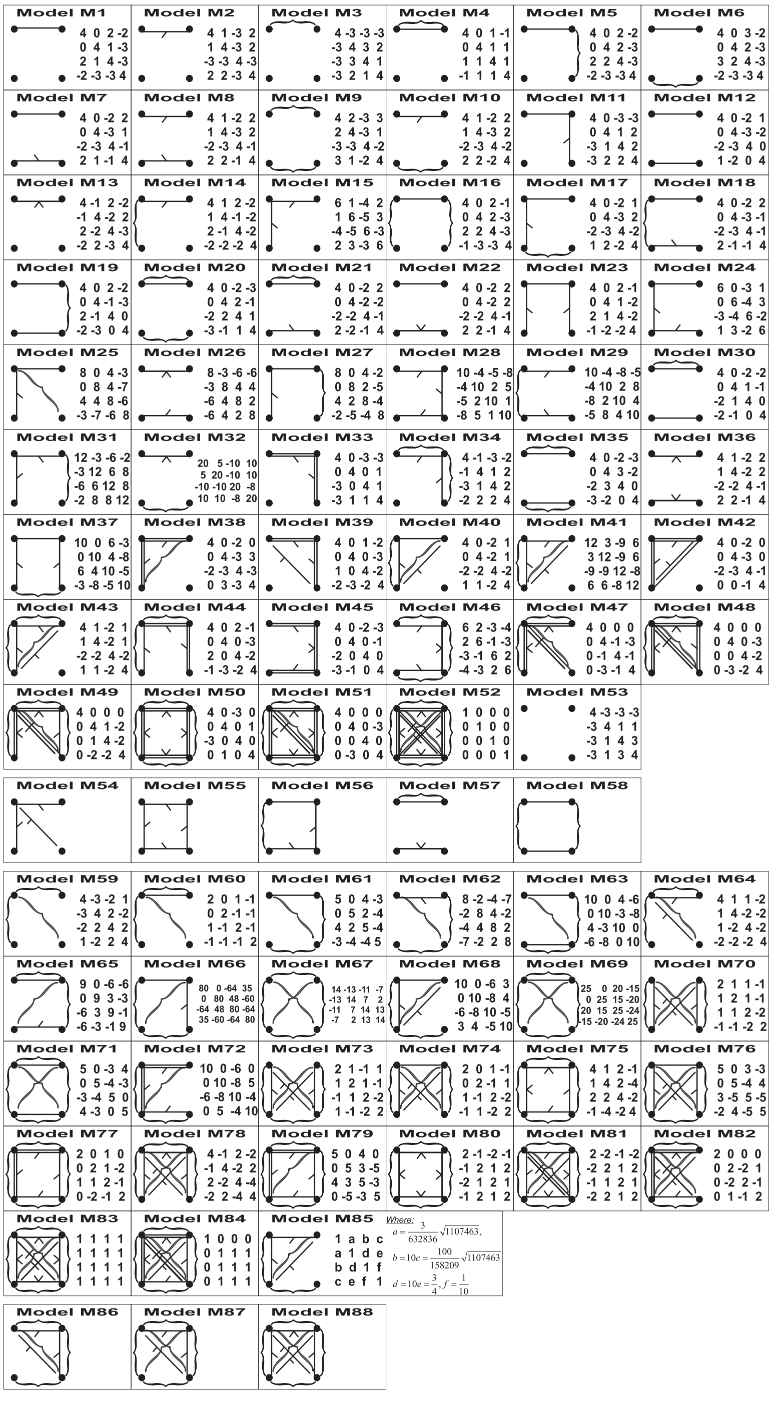

Fig. 1. List of independence models M1–M88 and their representations