Graphical Models and Independence Models

Yunshu Liu

ASPITRG Research Group

2014-03-04

References:

[1]. Steffen Lauritzen, *Graphical Models*, Oxford University Press, 1996
[2]. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Inc. 2006
[3]. Kevin P. Murphy, *Machine Learning - A Probabilistic Perspective*, The MIT Press, 2012
[4]. Petr Šimeček, *Independence Models*, Workshop on Uncertainty Processing(WUPES), 2006, Mikulov.
[5]. Radim Lněnička and František Matúš, *On Gaussian Conditional Independence Structures*, Kybernetika, Vol. 43(2007), No. 3, 327-342.

Outline

Preliminaries on Graphical Models
- Directed graphical model
- Undirected graphical model

Independence Models
- Gaussian Distributional framework
- Discrete Distributional framework

Outline

- Preliminaries on Graphical Models
- Independence Models

Preliminaries on Graphical Models

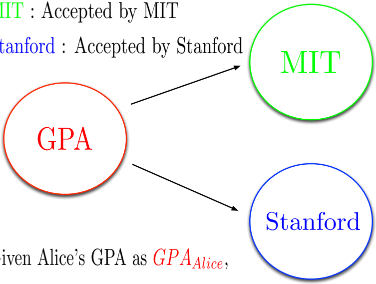
Definition of Graphical Models:

A graphical model is a probabilistic model for which a graph denotes the conditional dependence structure between random variables.

Example:
Suppose MIT and Stanford accepted undergraduate students only based on GPA

MIT : Accepted by MIT

Stanford : Accepted by Stanford



Given Alice's GPA as GPA_{Alice},

$\mathbb{P}(MIT|Stanford, GPA_{Alice}) = \mathbb{P}(MIT|GPA_{Alice})$

We say MIT is conditionally independent of Stanford given GPA_{Alice}

Sometimes use symbol $(MIT \perp Stanford | GPA_{Alice})$

Bayesian networks: directed graphical model

Bayesian networks

A Bayesian network consists of a collection of probability distributions P over $\mathbf{x} = \{x_1, \cdots, x_K\}$ that factorize over a directed acyclic graph(DAG) in the following way:

$$p(\mathbf{x}) = p(x_1, \cdots, x_K) = \prod_{k \in K} p(x_k | pa_k)$$

where pa_k is the direct parents nodes of x_k.

Alias of Bayesian networks:

- probabilistic directed graphical model: via directed acyclic graph(DAG)
- belief networks
- causal networks: directed arrows represent causal realtions

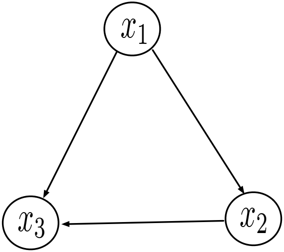
Bayesian networks: directed graphical model

Examples of Bayesian networks

Consider an arbitrary joint distribution $p(\mathbf{x}) = p(x_1, x_2, x_3)$ over three variables, we can write:

$$\begin{aligned} p(x_1, x_2, x_3) &= p(x_3|x_1, x_2)p(x_1, x_2) \\ &= p(x_3|x_1, x_2)p(x_2|x_1)p(x_1) \end{aligned}$$

which can be expressed in the following directed graph:



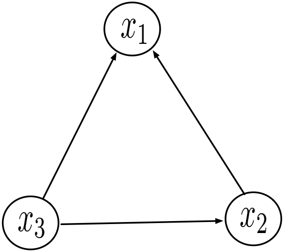
Bayesian networks: directed graphical model

Examples

Similarly, if we change the order of x_1, x_2 and x_3(same as consider all permutations of them), we can express $p(x_1, x_2, x_3)$ in five other different ways, for example:

$$
\begin{aligned}
p(x_1, x_2, x_3) &= p(x_1|x_2, x_3)p(x_2, x_3) \\
&= p(x_1|x_2, x_3)p(x_2|x_3)p(x_3)
\end{aligned}
$$

which corresponding to the following directed graph:



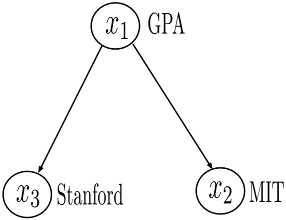
Bayesian networks: directed graphical model

Examples

Recall the previous example about how MIT and Stanford accept undergraduate students, if we assign x_1 to "GPA", x_2 to "accepted by MIT" and x_3 to "accepted by Stanford", then since $p(x_3|x_1,x_2) = p(x_3|x_1)$ we have

$$\begin{aligned} p(x_1,x_2,x_3) &= p(x_3|x_1,x_2)p(x_2|x_1)p(x_1) \\ &= p(x_3|x_1)p(x_2|x_1)p(x_1) \end{aligned}$$

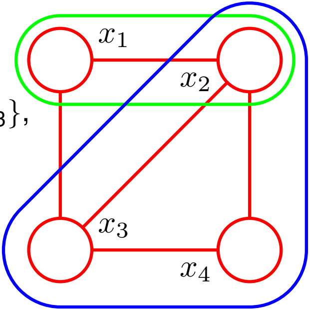
which corresponding to the following directed graph:



Markov random fields: undirected graphical model

In the undirected case, the probability distribution factorizes according to functions defined on the clique of the graph.
A clique is a subset of nodes in a graph such that there exist a link between all pairs of nodes in the subset.
A maximal clique is a clique such that it is not possible to include any other nodes from the graph in the set without it ceasing to be a clique.

Example of cliques:
$\{x_1,x_2\},\{x_2,x_3\},\{x_3,x_4\},\{x_2,x_4\},\{x_1,x_3\}$,
$\{x_1,x_2,x_3\}$, $\{x_2,x_3,x_4\}$
Maximal cliques:
$\{x_1,x_2,x_3\}$, $\{x_2,x_3,x_4\}$



Markov random fields: undirected graphical model

Markov random fields: Definition

Denote C as a clique, $\mathbf{x}_C$ the set of variables in clique C and $\psi_C(\mathbf{x}_C)$ a nonnegative potential function associated with clique C. Then a Markov random field is a collection of distributions that factorize as a product of potential functions $\psi_C(\mathbf{x}_C)$ over the **maximal cliques** of the graph:

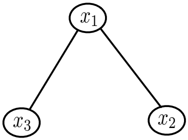
$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C)$$

where normalization constant $Z = \sum_{\mathbf{x}} \prod_C \psi_C(\mathbf{x}_C)$ sometimes called the partition function.

Markov random fields: undirected graphical model

Factorization of undirected graphs

Question: how to write the joint distribution for this undirected graph?



$(2 \perp 3|1)$ hold

Answer:

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{12}(x_1, x_2)\psi_{13}(x_1, x_3)$$

where $\psi_{12}(x_1, x_2)$ and $\psi_{13}(x_1, x_3)$ are the potential functions and Z is the partition function that make sure $p(\mathbf{x})$ satisfy the conditions to be a probability distribution.

Markov random fields: undirected graphical model

Markov property

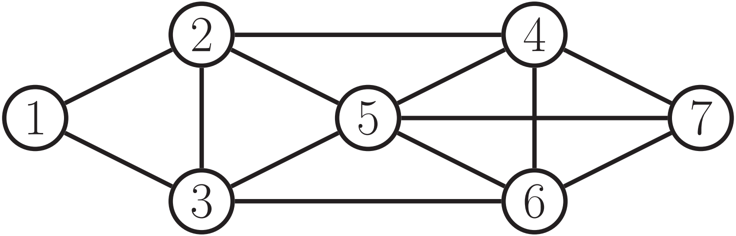
Given an undirected graph G = (V,E), a set of random variables $X = (X_a)_{a \in V}$ indexed by V , we have the following Markov properties:

- **Pairwise Markov property**: Any two non-adjacent variables are conditionally independent given all other variables: $X_a \perp X_b | X_{V \setminus \{a,b\}}$ if $\{a, b\} \notin E$
- **Local Markov property**: A variable is conditionally independent of all other variables given its neighbors:
  $X_a \perp X_{V \setminus \{nb(a) \cup a\}} | X_{nb(a)}$
  where nb(a) is the neighbors of node a.
- **Global Markov property**: Any two subsets of variables are conditionally independent given a separating subset:
  $X_A \perp X_B | X_S$, where every path from a node in A to a node in B passes through S(means when we remove all the nodes in S, there are no paths connecting any nodes in A to any nodes in B).

Markov random fields: undirected graphical model

Examples of Markov properties

Pairwise Markov property: $(1 \perp 7|23456)$, $(3 \perp 4|12567)$
Local Markov property: $(1 \perp 4567|23)$, $(4 \perp 13|2567)$
Global Markov property: $(1 \perp 67|345)$, $(12 \perp 67|345)$



Markov random fields: undirected graphical model

Relationship between different Markov properties and factorization property

(F): Factorization property; (G): Global Markov property;
(L): Local Markov property; (P):Pairwise Markov property

$$(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)$$

if assuming strictly positive $p(\cdot)$

$$(P) \Rightarrow (F)$$

which give us the Hammersley-Clifford theorem.

Markov random fields: undirected graphical model

The Hammersley-Clifford theorem(see Koller and Friedman 2009, p131 for proof)

Consider graph G, for strictly positive $p(\cdot)$, the following Markov property and Factorization property are equivalent:

Markov property: Any two subsets of variables are conditionally independent given a separating subset $(X_A, X_B | X_S)$ where every path from a node in A to a node in B passes through S.

Factorization property: The distribution p factorizes according to G if it can be expressed as a product of potential functions over maximal cliques.

Markov random fields: undirected graphical model

Example: Gaussian Markov random fields

A multivariate normal distribution forms a Markov random field w.r.t. a graph $G = (V, E)$ if the missing edges correspond to zeros on the concentration matrix(the inverse convariance matrix)

Consider $\mathbf{x} = \{x_1, \cdots, x_K\} \sim \mathcal{N}(0, \Sigma)$ with Σ regular and $K = \Sigma^{-1}$. The concentration matrix of the conditional distribution of (x_i, x_j) given $x_{\setminus\{i,j\}}$ is

$$K_{\{i,j\}} = \begin{pmatrix} k_{ii} & k_{ij} \\ k_{ji} & k_{jj} \end{pmatrix}$$

Hence

$$x_i \perp x_j | x_{\setminus ij} \Leftrightarrow k_{ij} = 0 \Leftrightarrow \{i, j\} \notin E$$

Markov random fields: undirected graphical model

Example: Gaussian Markov random fields

The joint distribution of gaussian markov random fields can be factorizes as:

$$\log p(x) = const - \frac{1}{2}\sum_i k_{ii}x_i^2 - \sum_{\{i,j\}\in E} k_{ij}x_i x_j$$

The zero entries in K are called structural zeros since they represent the absent edges in the undirected graph.

Outline

- Preliminaries on Graphical Models
- Independence Models

Gaussian Distributional framework

The multivariate Guassian

(Definition from Lauritzen)A d-dimensional random vector $X = (X_1, \cdots, X_d)$ has a multivariate Gaussian distribution on $\mathcal{R}^d$ if there is a vector $\xi \in \mathcal{R}^d$ and a $d \times d$ matrix Σ such that

$$\lambda^T X \sim \mathcal{N}(\lambda^T \xi, \lambda^T \Sigma \lambda) \quad \text{for all} \quad \lambda \in \mathcal{R}^d \tag{1}$$

We write $X \sim \mathcal{N}_d(\xi, \Sigma)$, ξ is the mean vector and Σ the covariance matrix of the distribution.
The validation of the definition requires that $\lambda^T \Sigma \lambda \geqslant 0$, i.e. if Σ is a positive semidefinite matrix.

Gaussian Distributional framework

Regular Guassian distribution

If furthermore, Σ is positive definite, i.e. if $\lambda^T \Sigma \lambda > 0$ for $\lambda \neq 0$, the distribution is called regular Guassian distribution and has a density

$$f(x|\xi, \Sigma) = (2\pi)^{-\frac{d}{2}} (\det K)^{\frac{1}{2}} exp(-\frac{(x-\xi)^T K (x-\xi)}{2}) \quad (2)$$

where $K = \Sigma^{-1}$ is the concentration matrix or precision matrix of the distribution.

Note: a positive semidefinite matrix is positive definite if and only if it is invertible, we also call Σ a regular matrix.

Gaussian Distributional framework

Properties of Gaussian distribution

Given a matrix $\Sigma = (\sigma_{a \cdot b})_{a,b \in \{1,\cdots,d\}}$ and A, B non-empty subsets of $\{1, \cdots, d\}$, we denote the submatrix with A-rows and B-columns be

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

Now let A be a subset of $\{1, \cdots, d\}$, then the marginal distribution ξ_A is also Gaussian distribution with variance matrix $\Sigma_{A \cdot A}$.

Gaussian Distributional framework

Properties of Gaussian distribution

Let us partition $\{1, \cdots, d\}$ into A and B such that $A \bigcup B = \{1, \cdots, d\}$ and $A \bigcap B = \emptyset$, the conditional distribution of ξ_A given $\xi_B = x_B$ is a Gaussian distribution with the variance matrix

$$\Sigma_{A|B} = \Sigma_{A \cdot A} - \Sigma_{A \cdot B} \Sigma_{B \cdot B}^{-} \Sigma_{B \cdot A}$$

where $\Sigma_{B \cdot B}^{-}$ is any generalized inverse of $\Sigma_{B \cdot B}$, which equals to $\Sigma_{B \cdot B}^{-1}$ if $\Sigma_{B \cdot B}$ is regular.

Gaussian Distributional framework

Properties of Gaussian distribution

Now assume $\Sigma_{B \cdot B}$ is regular, since

$$\begin{pmatrix} \Sigma_{A \cdot A} & \Sigma_{A \cdot B} \\ \Sigma_{B \cdot A} & \Sigma_{B \cdot B} \end{pmatrix}^{-1} = \begin{pmatrix} K_{A \cdot A} & K_{A \cdot B} \\ K_{B \cdot A} & K_{B \cdot B} \end{pmatrix}$$

we have

$$K_{A \cdot A}^{-1} = \Sigma_{A \cdot A} - \Sigma_{A \cdot B} \Sigma_{B \cdot B}^{-1} \Sigma_{B \cdot A} \tag{3}$$

$$K_{A \cdot A}^{-1} K_{A \cdot B} = -\Sigma_{A \cdot B} \Sigma_{B \cdot B}^{-1} \tag{4}$$

Thus

$$\Sigma_{A|B} = K_{A \cdot A}^{-1}$$

Gaussian Distributional framework

Properties of Gaussian distribution

Let A and B be a non-trivial partition of $\{1, \cdots, d\}$, we have

$$
\begin{aligned}
\xi_{A|B} &= \xi_A - K_{A \cdot A}^{-1} K_{A \cdot B}(x_B - \xi_B) \qquad (5) \\
&= \xi_A + \Sigma_{A \cdot B} \Sigma_{B \cdot B}^{-1}(x_B - \xi_B) \\
K_{A|B} &= K_{A \cdot A} \qquad (6)
\end{aligned}
$$

Thus X_A and X_B are independent if and only if $K_{A \cdot B} = 0$, giving $K_{A \cdot B} = 0$ if and only if $\Sigma_{A \cdot B} = 0$.

Gaussian Distributional framework

Properties of Gaussian distribution

Let a and b be distinct elements of $\{1, \cdots, d\}$ and $\mathbf{C} \subseteq \{1, \cdots, d\} \backslash ab$, if $\Sigma_{\mathbf{C}\cdot\mathbf{C}} > 0$, then the random variables ξ_a and ξ_b are independent given $\xi_{\mathbf{C}}$ if and only if $\det(\Sigma_{aC\cdot Cb}) = 0$ which is proved by

$$\det \Sigma_{a\mathbf{C}\cdot b\mathbf{C}} = 0 \Leftrightarrow \sigma_{a\cdot b} - \Sigma_{a\cdot\mathbf{C}}\Sigma^{-1}_{\mathbf{C}\cdot\mathbf{C}}\Sigma_{\mathbf{C}\cdot b}$$
$$\Leftrightarrow (\Sigma_{ab\cdot ab|\mathbf{C}})_{a\cdot b} = 0 \Leftrightarrow \xi_a \perp \xi_b | \xi_{\mathbf{C}}$$

Gaussian Distributional framework

Independence models

Let $D = \{1, 2, \cdots, d\}$ be a finite set and $\mathcal{T}_D$ denotes the set of all pairs $\langle ab|\mathbf{C}\rangle$ such that ab is an (unordered) couple of distinct emements of D and $\mathbf{C} \subseteq D \setminus ab$.
Subsets of $\mathcal{T}_D$ will be referred here as formal independence models over D. The independence model $\mathcal{I}(\xi)$ induced by a random vector $\xi = \xi_1, \cdots, \xi_d$ is the independence model over D defined as follows:

$$\mathcal{I}(\xi) = \{\langle ab|\mathbf{C}\rangle \,;\, \xi_a \perp \xi_b | \xi_{\mathcal{C}}\}$$

An independence model I is said to be generally/regularly Gaussian representable if there exists a general/regular Gaussian distribution ξ such that $I = \mathcal{I}(\xi)$.

Gaussian Distributional framework

Independence models: diagram

Visualisation of independence model I over D such that $|D| \leqslant 4$:
Each element of D is plotted as a dot.
If $\langle ab|\emptyset\rangle \in I$ then 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 $\langle ab|c\rangle \in I$ and $\langle ab|d\rangle \in I$ are elements of I, then only one line with two small lines in the middle is plotted;
If $\langle ab|cd\rangle \in I$ then a brace between a and b are added.

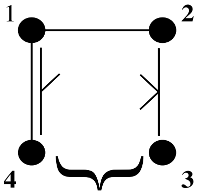


Diagram of independence model $I = \{\langle 12|\emptyset\rangle, \langle 23|1\rangle, \langle 23|4\rangle, \langle 34|12\rangle, \langle 14|\emptyset\rangle, \langle 14|2\rangle\}$

Gaussian Distributional framework

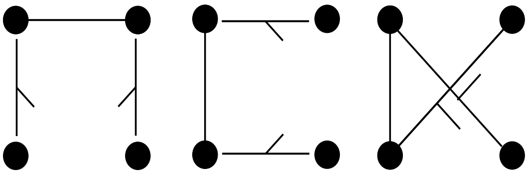
Independence models:permutation

Two independence models $\mathcal{I}$ and $\mathcal{J}$ over D are **isomorphic/permutably equivalent** if there exist a permutation π of D such that

$$\langle ab|C\rangle \in I \Leftrightarrow \langle \pi(a)\pi(b)|\pi(C)\rangle \in J$$

where $\pi(Z)$ stands for $\{\pi(z); z \in Z\}$.

Example of three isomorphic models



Gaussian Distributional framework

Independence models: example

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

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

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

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

Gaussian Distributional framework

Independence models: minor

If I is an independence model over $D = \{1, \cdots, d\}$ and E, F are disjoint subsets of D then let us define the minor $I[^F_E$ as an independence model over $D \backslash EF$

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

Notice that $I[^\emptyset_E = \{\langle ab|C\rangle \,; E \notin (\{a,b\} \cup C)\}$,
$I[^F_\emptyset = \{\langle ab|C\rangle \,; \langle ab|CF\rangle \in I\}$, $I[^\emptyset_\emptyset = I$, $(I[^{F_1}_{E_1})[^{F_2}_{E_2} = [^{F_1 F_2}_{E_1 E_2}$.

Properties involving minors

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

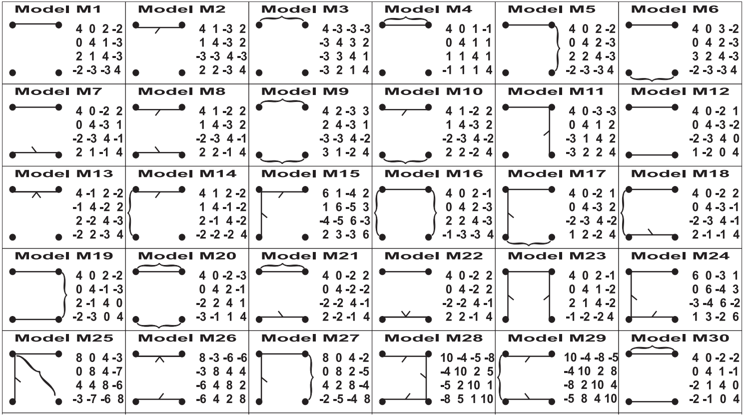
Gaussian Distributional framework

Regular Gaussian representable

Every regularly Gaussian representable model must have regularly representable minors. Using this property, among all the independence models for four variables, 58 classes of permutation equivalence with regularly representable minors are found for four variables.
Among the 58 types, 5 of them are not regularly representable, which left with 53 classes of permutation equivalent rugularly representable models(629 independence models after permutation).

Gaussian Distributional framework

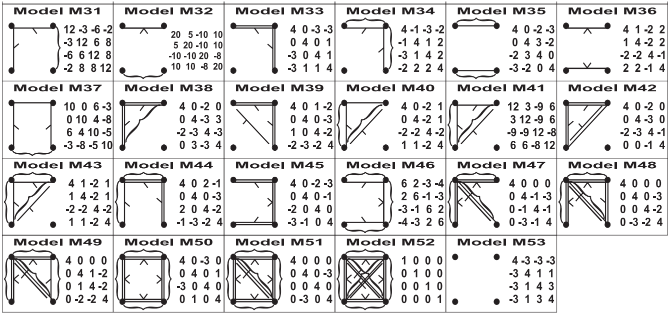
Regular Gaussian representable:M1-M30

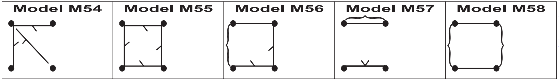


to be continue

Gaussian Distributional framework

Regular Gaussian representable:M31-M53



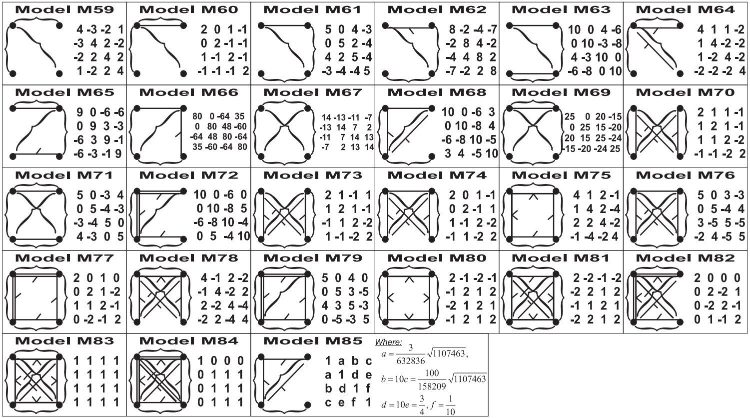


M54-M58 are
not regularly
Gaussian representable

Gaussian Distributional framework

General Gaussian representable:M1-M53, M59-M85

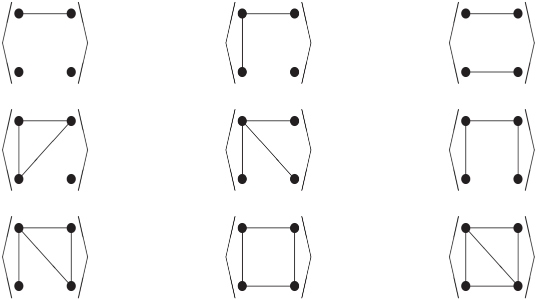
In addition, there are 27 Gaussian representable independence models(M59-M85) which are not regular representable.



Graphical Independence Models

Independence Models representable by undirected graph

For 4 variables, there are only 11 graphical(9 non-trivial) type of independence models, while there are 80 general gaussian representable, 53 regular gaussian representable independence models in total.



Discrete Distributional framework

Discrete distribution

A random vector $\xi = (\xi_1, \cdots, \xi_d)$ is called **discrete** if each ξ_a takes values in a state space X_a such that $0 < |X_a| < \infty$. In particular, ξ is called binary if $\forall a : |X_a| = 2$.
A discrete random vector ξ is called positive if

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

Conditional independence for discrete variables

For discrete distributioned random vector, variables ξ_a and ξ_b are independent given $\xi_{\mathbf{C}}$ if and only if for any $\mathbf{x_{abC}} \in X_{abC}$:

$$P(\xi_{\mathbf{abC}} = \mathbf{x_{abC}}) \cdot P(\xi_{\mathbf{C}} = \mathbf{x_{C}}) = P(\xi_{\mathbf{aC}} = \mathbf{x_{aC}}) \cdot P(\xi_{\mathbf{bC}} = \mathbf{x_{bC}})$$

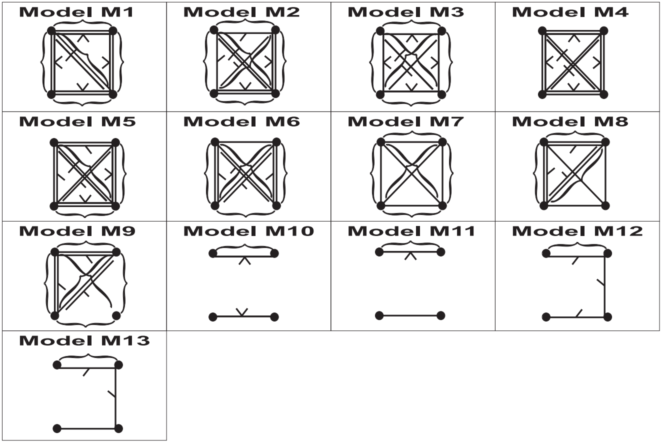
Discrete Distributional framework

Discrete representable models

The class of all discrete representable models can be described by the set $\mathcal{C}$ of irreducible models, i.e. nontrivial discrete representable models that cannot be written as an intersection of two other discrete representable models. There are 13 types of irreducible models for discrete distribution, and 18478 models in total corresponding to 1098 types.

Discrete Distributional framework

Discrete representable models



Notes: results on positive discrete representable models are not known yet.

Thanks!

Questions!