MODELE GRAFICZNE

Piotr GRACZYK

4. AXIOMS AND MARKOV PROPERTIES of GRAPHICAL MODELS.

This part is based on lectures of Prof. S. Lauritzen at CIMPA Summer School Hammamet 2011, with his kind permission.

For random variables X, Y, Z, and W it holds

(C1) If $X \perp\!\!\!\perp Y \mid Z$ then $Y \perp\!\!\!\perp X \mid Z$;

(C2) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp U \mid Z$;

(C3) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp Y \mid (Z, U)$;

(C4) If $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, then $X \perp\!\!\!\perp (Y, W) \mid Z$;

If density w.r.t. product measure $f(x, y, z, w) > 0$ also

(C5) If $X \perp\!\!\!\perp Y \mid (Z, W)$ and $X \perp\!\!\!\perp Z \mid (Y, W)$ then $X \perp\!\!\!\perp (Y, Z) \mid W$.

Proof of (C5): We have

$$X \perp\!\!\!\perp Y \mid (Z, W) \Rightarrow f(x, y, z, w) = a(x, z, w)b(y, z, w).$$

Similarly

$$X \perp\!\!\!\perp Z \mid (Y, W) \Rightarrow f(x, y, z, w) = g(x, y, w)h(y, z, w).$$

If $f(x, y, z, w) > 0$ for all (x, y, z, w) it thus follows that

$$g(x, y, w) = a(x, z, w)b(y, z, w)/h(y, z, w).$$

The left-hand side does not depend on z. So for fixed $z = z_0$:

$$g(x, y, w) = \tilde{a}(x, w)\tilde{b}(y, w).$$

Insert this into the second expression for f to get

$$f(x, y, z, w) = \tilde{a}(x, w)\tilde{b}(y, w)h(y, z, w) = a^*(x, w)b^*(y, z, w)$$

which shows $X \perp\!\!\!\perp (Y, Z) \mid W$.

Conditional independence can be seen as encoding abstract irrelevance. With the interpretation: *Knowing C, A is irrelevant for learning B,* (C1)–(C4) translate into:

(I1) If, knowing C, learning A is irrelevant for learning B, then B is irrelevant for learning A;

(I2) If, knowing C, learning A is irrelevant for learning B, then A is irrelevant for learning any part D of B;

(I3) If, knowing C, learning A is irrelevant for learning B, it remains irrelevant having learnt any part D of B;

(I4) If, knowing C, learning A is irrelevant for learning B and, having also learnt A, D remains irrelevant for learning B, then both of A and D are irrelevant for learning B.

The property analogous to (C5) is slightly more subtle and not generally obvious.

An *independence model* $\perp_\sigma$ is a ternary relation over subsets of a finite set V. It is *graphoid* if for all subsets A, B, C, D:

(S1) if $A \perp_\sigma B \mid C$ then $B \perp_\sigma A \mid C$ (symmetry);

(S2) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ (decomposition);

(S3) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid (C \cup D)$ (weak union);

(S4) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid (B \cup C)$, then $A \perp_\sigma (B \cup D) \mid C$ (contraction);

(S5) if $A \perp_\sigma B \mid (C \cup D)$ and $A \perp_\sigma C \mid (B \cup D)$ then $A \perp_\sigma (B \cup C) \mid D$ (intersection).

Semigraphoid if only (S1)–(S4) holds. It is *compositional* if also

(S6) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ then $A \perp_\sigma (B \cup D) \mid C$ (composition).

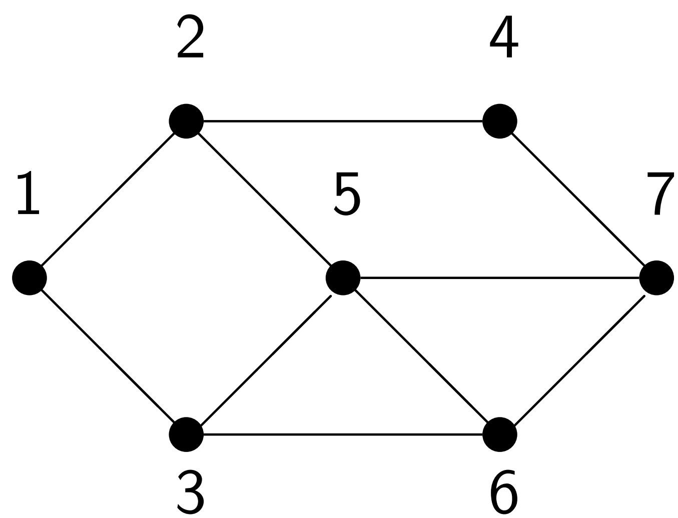
Separation in undirected graphs

Let $\mathcal{G} = (V, E)$ be finite and simple undirected graph (no self-loops, no multiple edges).

For subsets A, B, S of V, let $A \perp_{\mathcal{G}} B \mid S$ denote that *S separates A from B in $\mathcal{G}$*, i.e. that all paths from A to B intersect S.

Fact: *The relation $\perp_{\mathcal{G}}$ on subsets of V is a compositional graphoid.*

This fact is the reason for choosing the name 'graphoid' for such independence model.



For several variables, complex systems of conditional independence can for example be described by undirected graphs.
Then a set of variables A is conditionally independent of set B, given the values of a set of variables C if C separates A from B.

For example in picture above

$$1 \perp\!\!\!\perp \{4,7\} \mid \{2,3\}, \qquad \{1,2\} \perp\!\!\!\perp 7 \mid \{4,5,6\}.$$

Systems of random variables

For a system V of *labeled random variables* $X_v, v \in V$, we use the shorthand

$$A \perp\!\!\!\perp B \mid C \iff X_A \perp\!\!\!\perp X_B \mid X_C,$$

where $X_A = (X_v, v \in A)$ denotes the variables with labels in A.

The properties (C1)–(C4) imply that $\perp\!\!\!\perp$ *satisfies the semi-graphoid axioms* for such a system, and the *graphoid axioms if the joint density of the variables is strictly positive.*

A regular *multivariate Gaussian distribution*, defines a *compositional graphoid independence model.*

Definition. Let $\alpha \in V$ be a node of the graph $\mathcal{G}$.

If $v \sim w$ we say that v and w are **neighbours**.

The **boundary** $bd(\alpha) \stackrel{\text{df}}{=} \{v \in V |\ v \sim \alpha\}$ is the set of neighbours of α.

The **closure** $cl(\alpha) \stackrel{\text{df}}{=} \{\alpha\} \cup bd(\alpha)$.

$\mathcal{G} = (V, E)$ simple undirected graph; An independence model $\perp_\sigma$ satisfies

(P) *the pairwise Markov property* if

$$\alpha \not\sim \beta \Rightarrow \alpha \perp_\sigma \beta \mid V \setminus \{\alpha, \beta\};$$

(L) *the local Markov property* if

$$\forall \alpha \in V : \alpha \perp_\sigma V \setminus \mathrm{cl}(\alpha) \mid \mathrm{bd}(\alpha);$$

(G) *the global Markov property* if

$$A \perp_\mathcal{G} B \mid S \Rightarrow A \perp_\sigma B \mid S.$$

The terminology "Markov property" can be explained on the nearest neighbour "chain" model A_n

$$\mathcal{G} : \overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{k-1}{\bullet} - \overset{k}{\bullet} - \overset{k+1}{\bullet} - \cdots - \overset{n}{\bullet}$$

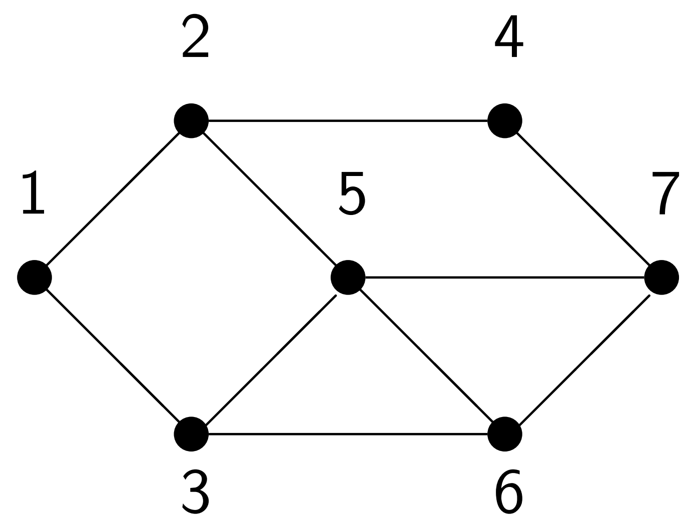
The global Markov property contains classical Markov properties of a Markov chain

$$k+1 \perp\!\!\!\perp \{1,\ldots,k-1\} |\ k$$

$$\{k+1,\ldots,n\} \perp\!\!\!\perp \{1,\ldots,k-1\} |\ k$$

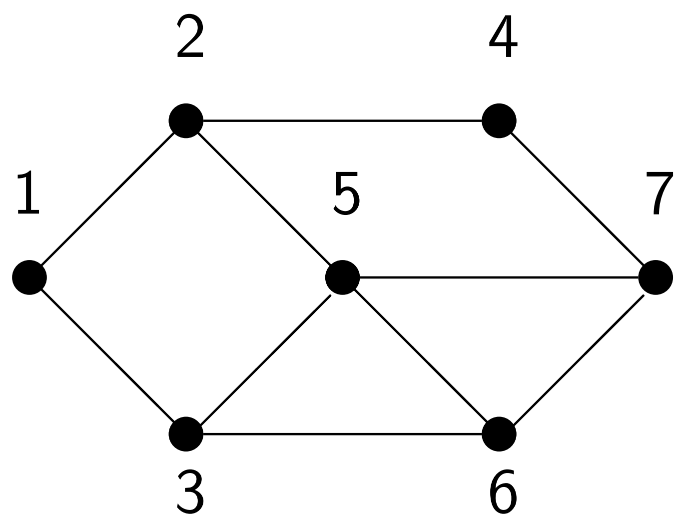
Graphical Markov properties (P), (L), (G) are a generalization of classical Markov properties from chains to graphs.

Pairwise Markov property



Any non-adjacent pair of random variables are conditionally independent given the remaning.
For example, $1 \perp_\sigma 5 \mid \{2, 3, 4, 6, 7\}$ and $4 \perp_\sigma 6 \mid \{1, 2, 3, 5, 7\}$.

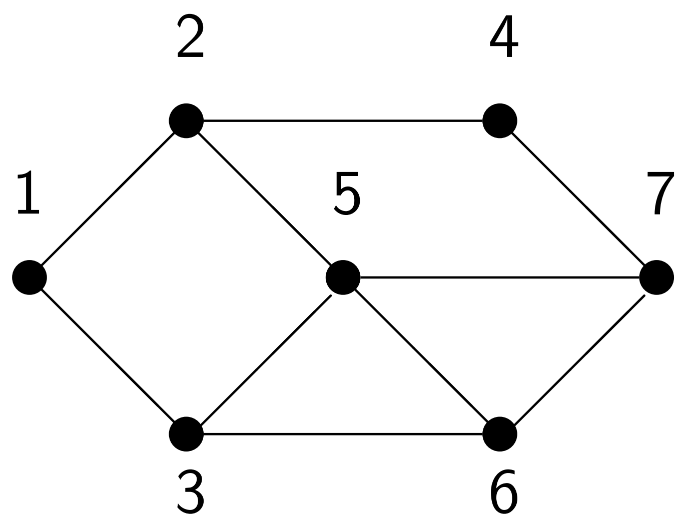
Local Markov property



Every variable is conditionally independent of the remaining, given its neighbours.

For example, $5 \perp_\sigma \{1, 4\} \mid \{2, 3, 6, 7\}$ and $7 \perp_\sigma \{1, 2, 3\} \mid \{4, 5, 6\}$.

Global Markov property



To find conditional independence relations, one should look for separating sets, such as $\{2,3\}$, $\{4,5,6\}$, or $\{2,5,6\}$
For example, it follows that $1 \perp_\sigma 7 \mid \{2,5,6\}$ and $2 \perp_\sigma 6 \mid \{3,4,5\}$.

For any semigraphoid it holds that

$$(\mathsf{G}) \Rightarrow (\mathsf{L}) \Rightarrow (\mathsf{P})$$

If $\perp_\sigma$ satisfies graphoid axioms it further holds that

$$(\mathsf{P}) \Rightarrow (\mathsf{G})$$

so that *in the graphoid case*

$$(\mathsf{G}) \iff (\mathsf{L}) \iff (\mathsf{P}).$$

The latter holds in particular for $\perp\!\!\!\perp$, when $f(x) > 0$.

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

(G) implies (L) because $\mathrm{bd}(\alpha)$ separates α from $V \setminus \mathrm{cl}(\alpha)$.

Assume (L). Then $\beta \in V \setminus \mathrm{cl}(\alpha)$ because $\alpha \not\sim \beta$. Thus

$$\mathrm{bd}(\alpha) \cup ((V \setminus \mathrm{cl}(\alpha)) \setminus \{\beta\}) = V \setminus \{\alpha, \beta\},$$

Hence by (L) and weak union (S3) we get that

$$\alpha \perp_\sigma (V \setminus \mathrm{cl}(\alpha)) \mid V \setminus \{\alpha, \beta\}.$$

Decomposition (S2) then gives $\alpha \perp_\sigma \beta \mid V \setminus \{\alpha, \beta\}$ which is (P).

(P) $\Rightarrow$ (G) for graphoids:
Assume (P) and $A \perp_{\mathcal{G}} B \mid S$. *We must show* $A \perp_{\sigma} B \mid S$.

Wlog assume A and B non-empty. Proof is reverse induction on $n = |S|$.

If $n = |V| - 2$ then A and B are singletons and (P) yields $A \perp_{\sigma} B \mid S$ directly.

Assume $|S| = n < |V| - 2$ and conclusion established for $|S| > n$:

First assume $V = A \cup B \cup S$. Then either A or B has at least two elements, say A. If $\alpha \in A$ then $B \perp_{\mathcal{G}} (A \setminus \{\alpha\}) \mid (S \cup \{\alpha\})$ and also $\alpha \perp_{\mathcal{G}} B \mid (S \cup A \setminus \{\alpha\})$ (as $\perp_{\mathcal{G}}$ is a semi-graphoid). Thus by the induction hypothesis

$$(A \setminus \{\alpha\}) \perp_{\sigma} B \mid (S \cup \{\alpha\}) \text{ and } \{\alpha\} \perp_{\sigma} B \mid (S \cup A \setminus \{\alpha\}).$$

Now intersection (S5) gives $A \perp_{\sigma} B \mid S$.

FACTORIZATION PROPERTY

with respect to a graph $\mathcal{G}$

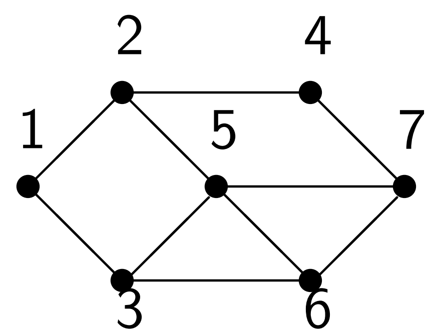
Assume density f w.r.t. product measure on $\mathcal{X}$.
For $a \subseteq V$, $\psi_a(x)$ denotes a function which depends on x_a only, i.e.

$$x_a = y_a \Rightarrow \psi_a(x) = \psi_a(y).$$

We can then write $\psi_a(x) = \psi_a(x_a)$ without ambiguity.
The distribution of X *factorizes w.r.t.* $\mathcal{G}$ or satisfies (F) if

$$f(x) = \prod_{a \in \mathcal{A}} \psi_a(x)$$

where $\mathcal{A}$ are *complete* subsets of $\mathcal{G}$.
Complete subsets of a graph are sets with all elements pairwise neighbours.



The *cliques* of this graph are the maximal complete subsets $\{1,2\}$, $\{1,3\}$, $\{2,4\}$, $\{2,5\}$, $\{3,5,6\}$, $\{4,7\}$, and $\{5,6,7\}$. A complete set is any subset of these sets.
The graph above corresponds to a factorization as

$$\begin{aligned} f(x) &= \psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)\psi_{24}(x_2,x_4)\psi_{25}(x_2,x_5) \\ &\times \psi_{356}(x_3,x_5,x_6)\psi_{47}(x_4,x_7)\psi_{567}(x_5,x_6,x_7). \end{aligned}$$

Let (F) denote the property that f factorizes w.r.t. $\mathcal{G}$ and let (G), (L) and (P) denote Markov properties w.r.t. $\perp\!\!\!\perp$. *It then holds that*

$$(\mathrm{F}) \Rightarrow (\mathrm{G})$$

and further: *If $f(x) > 0$ for all x*, $(\mathrm{P}) \Rightarrow (\mathrm{F})$.

The former of these is a simple direct consequence of the factorization whereas the second implication is more subtle and known as the *Hammersley–Clifford Theorem*.

Thus in the case of positive density (but typically only then), *all the properties coincide:*

$$(\mathrm{F}) \iff (\mathrm{G}) \iff (\mathrm{L}) \iff (\mathrm{P}).$$

Any joint probability distribution P of $X = (X_v, v \in V)$ has a *dependence graph* $G = G(P) = (V, E(P))$.

This is defined by letting $\alpha \not\sim \beta$ in $G(P)$ exactly when

$$\alpha \perp\!\!\!\perp_P \beta \mid V \setminus \{\alpha, \beta\}.$$

X will then satisfy the pairwise Markov w.r.t. $G(P)$ and $G(P)$ is smallest with this property, i.e. *P is pairwise Markov w.r.t. $\mathcal{G}$ iff*

$$G(P) \subseteq \mathcal{G}.$$

If $f(x) > 0$ for all x, *P is also globally Markov w.r.t. $G(P)$.*