Markov Properties for Graphical Models

Steffen Lauritzen, University of Oxford

Wald Lecture, World Meeting on Probability and Statistics

Istanbul 2012

Recall that a sequence of random variables $X_1, \ldots, X_n, \ldots$ is a *Markov chain* if it holds for all n that

$$P(X_{n+1} \in A \mid X_1, \ldots, X_n) = P(X_{n+1} \in A \mid X_n).$$

We express this by saying that X_{n+1} is *conditionally independent* of $X_1, \ldots, X_{n-1}$ given X_n and write symbolically

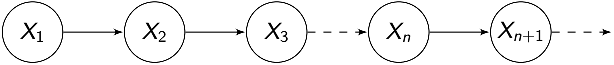
$$X_{n+1} \perp\!\!\!\perp (X_1, \ldots, X_{n-1}) \mid X_n$$

or

$$X_{n+1} \perp\!\!\!\perp_P (X_1, \ldots, X_{n-1}) \mid X_n$$

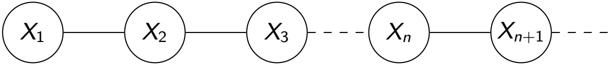
when we wish to emphasize that the statement is relative to a given probability distribution P.

A Markov chain is graphically represented as



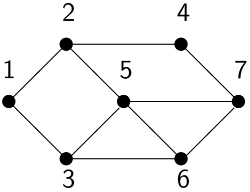
This is a so-called *directed acyclic graph* (DAG) representing one of many extensions of the Markov property.

Alternatively, we may consider an undirected representation



and derive a number of further conditional independence relations such as, for example, for $i < j < k < l < m$

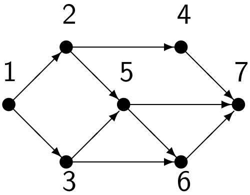
$$X_k \perp\!\!\!\perp (X_i, X_m) \mid X_j, X_l.$$



The graph above corresponds to a factorization as

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

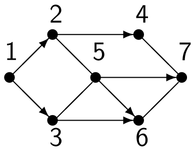
The *global Markov property* (Hammersley and Clifford, 1971) implies, for example, $1 \perp\!\!\!\perp 7 \mid \{3, 4, 5\}$.



The above graph corresponds to the factorization

$$\begin{aligned} f(x) &= f(x_1)f(x_2 \mid x_1)f(x_3 \mid x_1)f(x_4 \mid x_2) \\ &\times f(x_5 \mid x_2, x_3)f(x_6 \mid x_3, x_5)f(x_7 \mid x_4, x_5, x_6). \end{aligned}$$

The *global Markov property* (Pearl, 1986; Geiger et al., 1990; Lauritzen et al., 1990) implies, for example, $3 \perp\!\!\!\perp 4 \mid 1$.

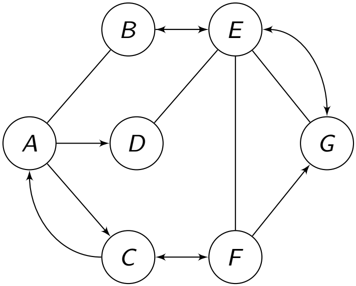


Chain components $\{1\}, \{2, 3, 5\}, \{4, 6, 7\}$; correspond to

$$\begin{aligned}
p(x) &= p(x_1)p(x_2, x_3, x_5 \mid x_1)p(x_4, x_6, x_7 \mid x_2, x_3, x_5) \\
p(x_2, x_3, x_5 \mid x_1) &= Z^{-1}(x_1)\psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_5)\psi(x_3, x_5) \\
p(x_4, x_6, x_7 \mid x_2, x_3, x_5) &= Z^{-1}(x_2, x_3, x_5) \\
&\quad \times\psi(x_2, x_4)\psi(x_4, x_7)\psi(x_5, x_7)\psi(x_6, x_7).
\end{aligned}$$

Global Markov property (Frydenberg, 1990) implies, for example, $1 \perp\!\!\!\perp 6 \mid \{2, 3\}$.

Graphical models have now developed a variety of ways of coding conditional independence relations using quite general graphs (Andersson et al., 1996; Cox and Wermuth, 1993; Koster, 2002; Richardson and Spirtes, 2002; Sadeghi, 2012) moving towards making sense of pictures as the following:



We shall in the following elaborate on this development.

Let V be a finite set. An *independence model* $\perp_\sigma$ over V is a ternary relation over subsets of a finite set V. The independence model is a *semi-graphoid* if it holds 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);

Probabilistic independence models

For a system V of *labeled random variables* $X_v, v \in V$ with distribution P we can define an independence model $\perp\!\!\!\perp_P$ by

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

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

General properties of conditional independence imply that *probabilistic independence models are semi-graphoids.*

If a semi-graphoid further satisfies

(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).

we say it is a *graphoid*.

In general, probabilistic independence models are neither graphoids nor compositional.

If P has strictly positive density wrt a product measure, $\perp\!\!\!\perp_P$ is a graphoid.

A *compositional graphoid* further satisfies.

(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).

If P is a regular multivariate Gaussian distribution, $\perp\!\!\!\perp_P$ is a compositional graphoid, but in general this is not the case.

The composition property ensures that pairwise conditional independence implies setwise conditional independence, i.e. that

$$A \perp_\sigma B \mid C \iff \alpha \perp_\sigma \beta \mid C, \forall \alpha \in A, \beta \in B.$$

Thus a *compositional graphoid* satisfies 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);

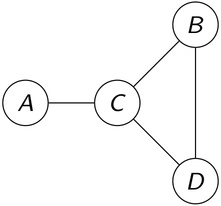
(S4) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid (B \cup C)$, then $A \perp_\sigma (B \cup D) \mid C$ (contraction);

(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);

(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).

Let $\mathcal{G} = (V, E)$ be finite and simple undirected graph. 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.

It is readily verified that *the relation $\perp_{\mathcal{G}}$ on subsets of V is a compositional graphoid.*



Graph basics

A *mixed graph* $\mathcal{G}$ over a finite set of vertices V has three types of edges: *arrows* (directed edges), *arcs* (bi-directed edges), and *lines* (undirected edges).

A *walk* is a list $\langle v_0, e_1, v_1, \cdots, e_k, v_k \rangle$ of nodes and edges such that for $1 \leq i \leq k$, the edge e_i has endpoints v_{i-1} and v_i.

A *path* is a walk with no repeated node or edge.

A *cycle* is a path with the modification that $v_0 = v_k$.

A path or cycle is *directed* if all edges are arrows and e_i points from v_{i-1} to v_i for all i.

More graph basics

If $u \to v$, u is a *parent* of v and v is a *child* of u. If $u \leftrightarrow v$, u and v are *spouses* and if $u \,—\, v$, u and v are *neighbours*. We write $u \sim v$ to denote that there is *some* edge between u and v and say that u and v are *adjacent*.

If there is a directed path from u to v, u is an *ancestor* of v and v is a *descendant* of u. The ancestors of u are $\mathrm{an}(u)$ and the descendants are $\mathrm{de}(u)$ and similarly for sets of nodes A we use $\mathrm{an}(A)$, and $\mathrm{de}(A)$.

A node v is a *collider* on a walk if two arrowheads of the walk meet head to head at v, i.e. if $\langle \cdots, \to v \leftarrow, \cdots \rangle$ or $\langle \cdots, \leftrightarrow v \leftarrow, \cdots \rangle$, or $\langle \cdots, \to, v, \leftrightarrow, \cdots \rangle$, or $\langle \cdots, \leftrightarrow, v, \leftrightarrow, \cdots \rangle$.

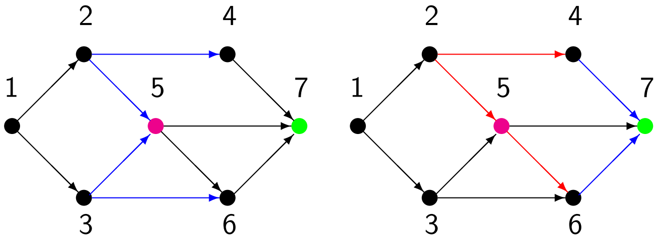
Consider a *directed acyclic graph* (DAG) $\mathcal{D}$, i.e. a graph where all edges are directed but no cycles are directed.

For $S \subseteq V$ we say that a path is rendered *active* by S if all its collider nodes are in $S \cup \mathrm{an}(S)$ and none of its other nodes are in S. A path that is not active is *blocked*.

If $A, B, S \subseteq V$ and all paths from A to B are blocked by S, we say that S *d-separates A from B*, and write $A \perp_d B \mid S$.

For any directed acyclic graph, the independence model $\perp_d$ is a compositional graphoid (Koster, 1999).

Separation by example



For $S = \{5\}$ or $S = \{7\}$, the path $(4, 2, 5, 3, 6)$ is *active*, whereas trails $(4, 2, 5, 6)$ and $(4, 7, 6)$ are *blocked* for $S = \{5\}$ and *active* for $S = \{7\}$.

For $S = \{3, 5\}$, they are all blocked.

Now consider a general *mixed graph*, with potentially three types of edges. The d-separation can be directly extended to a general, mixed graph (Richardson, 2003; Sadeghi, 2012).

For $S \subseteq V$ we say that a path is rendered *active* by S if all its collider nodes are in $S \cup \mathrm{an}(S)$ and none of its other nodes are in S. A path that is not active is *blocked*.

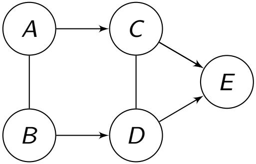
If $A, B, S \subseteq V$ and all paths from A to B are blocked by S, we say that S *m-separates A from B*, and write $A \perp_m B \mid S$.

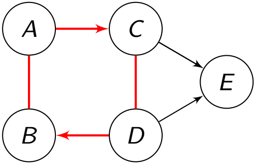
For any mixed graph, the independence model $\perp_m$ is a compositional graphoid (Sadeghi and Lauritzen, 2012).

Clearly, for a DAG $\mathcal{D}$, we have that $A \perp_d B \mid S$ if and only if $A \perp_m B \mid S$. But note also that for an undirected graph $\mathcal{G}$ it holds that $A \perp_g B \mid S$ if and only if $A \perp_m B \mid S$. *Thus m-separation extends and unifies standard independence models for DAGs and UGs.*

However, this is *not true for chain graphs,* not even with most alternative interpretations of such graphs discussed in the literature (Andersson et al., 2001; Drton, 2009).

A *standard chain graph* is a mixed graph with no multiple edges, no bi-directed edges, and *no directed or semi-directed cycles* i.e. no cycles with all arrows on the cycle pointing in the same direction.





The graph to the left is a chain graph, with *chain components* (connected components after removing arrows) $\{A, B\}, \{C, D\}, \{E\}$. The graph to the right is *not* a chain graph, due to the semi-directed cycle $\langle A \rightarrow C — D \rightarrow B — A \rangle$.

The separation criterion for standard chain graphs was developed by Studený and Bouckaert (1998) and further simplified by Studený (1998). It is similar to but different from m-separation.

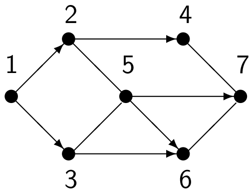
Firstly, as in Koster (2002), we are considering *walks* rather than paths, allowing repeated nodes.

Next, a *section* of a walk is a maximal cyclic subwalk with only directed edges, i.e. a subwalk of the form $\langle v - \cdots - v \rangle$.

It is a *collider section* on the walk if there are arrowheads meeting head-to-head at v.

For $S \subseteq V$ we say that a walk is rendered *active* by S if all its collider sections intersect with S and its other sections are disjoint from S. A walk that is not active is *blocked*.

If $A, B, S \subseteq V$ and all walks from A to B are blocked by S, we say that S *c-separates A from B*, and write $A \perp_c B \mid S$.



It is not difficult to verify that *for any chain graph, the independence model* $\perp_c$ *is a compositional graphoid.*

Studený (1998) shows that even though there are infinitely many walks, there is a local algorithm for checking c-separation.

The notion of *c-separation* $\perp_c$ *for chain graphs also coincides with standard separation* $\perp_g$ *in UGs and d-separation* $\perp_d$ *in DAGs* (Studený, 1998), but it is *distinct from m-separation* for general chain graphs or general mixed graphs.

Other independence models have been associated with chain graphs, see for example Drton (2009) who classifies them into four types, one of which has been described above.

The *multivariate regression Markov property* (Cox and Wermuth, 1996; Wermuth et al., 2009) would *correspond to m-separation in chain graphs with bi-directed chain components*, whereas the remaining two types (AMP and its dual) are different yet again.

Are the last two compositional graphoids?

The powerful features of graphical models are partly related to the fact that graphs and graph structures are *easy to communicate to computers,* but not least their *visual representation.*

The visual features are most immediate for undirected graphs, where separation is simple. But in general it is desirable that the graphs *represent* their independence model in the sense that

$$\alpha \not\sim \beta \iff \exists S \subseteq V \setminus \{\alpha, \beta\} : \alpha \perp_\sigma \beta \,|\, S$$

so missing edges represent conditional independence.

We say that a graph $\mathcal{G}$ with separation criterion $\perp_{\mathcal{G}}$ is *maximal* Richardson and Spirtes (2002) if adding an edge changes the independence model $\perp_{\mathcal{G}}$.

It then holds that *maximal graphs represent their independence models*.

UGs, DAGs and the various versions of chain graphs are always maximal, but other mixed graphs are not.

Let P be a joint distribution for random variable $X_v, v \in V$ and let $\mathcal{G}$ be a graph with independence model $\perp_{\mathcal{G}}$. A described earlier, P defines its own independence model $\perp\!\!\!\perp_P$.

We say that P is (globally) *Markov* w.r.t. $(\mathcal{G}, \perp_{\mathcal{G}})$ if it holds for all $A, B, S \subseteq V$ that

$$A \perp_{\mathcal{G}} B \mid S \Rightarrow A \perp\!\!\!\perp_P B \mid S.$$

We further say that P is *faithful* to $(\mathcal{G}, \perp_{\mathcal{G}})$ if also the converse holds

$$A \perp_{\mathcal{G}} B \mid S \iff A \perp\!\!\!\perp_P B \mid S.$$

Note that *if P is faithful to $(\mathcal{G}, \perp_{\mathcal{G}})$, $\perp\!\!\!\perp_P$ is a compositional graphoid*, whether or not it originates from a specific family of distributions.

If there is a P such that P is faithful to $(\mathcal{G}, \perp_{\mathcal{G}})$, we say that $\perp_{\mathcal{G}}$ is *probabilistically representable.*

Graphical independence models based on UGs, DAGs, chain graphs in all its variations, and *mixed graphs without ribbons* (Sadeghi, 2012) *are all probabilistically representable.*

In fact, for all these, a dimensional argument gives that *most P that are Markov w.r.t. $(\mathcal{G}, \perp_{\mathcal{G}})$ are indeed also faithful* (Meek, 1995, 1996).

Question: *Are all of the graphical independence models described here probabilistically representable? If not, which are?*

For certain purposes it can be helpful to consider weaker Markov properties than the global Markov property.

Pairwise Markov properties are of the type

$$\alpha \not\sim \beta \Rightarrow \alpha \perp\!\!\!\perp_P \beta \mid S(\alpha, \beta),$$

where, for example, $S(\alpha, \beta) = \mathrm{ant}(\alpha) \cup \mathrm{ant}(\beta) \setminus \{\alpha, \beta\}$, and *local Markov properties* of the type

$$\alpha \perp\!\!\!\perp_P V \setminus (\alpha \cup N(\alpha)) \mid N(\alpha)$$

for some kind of neighbourhood $N(\alpha)$, typically involving parents, spouses, and neighbours.

Pairwise and local properties are useful for establishing global Markov properties and hence conditions which ensure these to imply the global property are sought.

Typically the semi-graphoid properties of P are insufficient for these, with exceptions depending on the particular type of graph.

For example, *for ribbonless graphs with* $\perp_m$ *the pairwise Markov property implies the global Markov property if* $\perp\!\!\!\perp_P$ *is a compositional graphoid* (Sadeghi and Lauritzen, 2012).

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;
- Edges in E can have several types; at least three;

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;
- Edges in E can have several types; at least three;
- Markov condition defined by a relation $\perp_{\mathcal{G}}$ so that query $A \perp_{\mathcal{G}} B \mid C$ has clear resolution; preferably *path-based* criterion;

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;
- Edges in E can have several types; at least three;
- Markov condition defined by a relation $\perp_{\mathcal{G}}$ so that query $A \perp_{\mathcal{G}} B \mid C$ has clear resolution; preferably *path-based* criterion;
- Must *represent* their independence model: α and β are non-adjacent in $\mathcal{G}$ if and only if $\alpha \perp_{\mathcal{G}} \beta \mid S$ for some $S \subseteq V \setminus \{\alpha, \beta\}$; requires *maximality* in mixed graphs;

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;
- Edges in E can have several types; at least three;
- Markov condition defined by a relation $\perp_{\mathcal{G}}$ so that query $A \perp_{\mathcal{G}} B \mid C$ has clear resolution; preferably *path-based* criterion;
- Must *represent* their independence model: α and β are non-adjacent in $\mathcal{G}$ if and only if $\alpha \perp_{\mathcal{G}} \beta \mid S$ for some $S \subseteq V \setminus \{\alpha, \beta\}$; requires *maximality* in mixed graphs;
- $\perp_{\mathcal{G}}$ defines a compositional graphoid.

Graphical Independence Model

- Model determined by a graph $\mathcal{G} = (V, E)$;
- Edges in E can have several types; at least three;
- Markov condition defined by a relation $\perp_{\mathcal{G}}$ so that query $A \perp_{\mathcal{G}} B \mid C$ has clear resolution; preferably *path-based* criterion;
- Must *represent* their independence model: α and β are non-adjacent in $\mathcal{G}$ if and only if $\alpha \perp_{\mathcal{G}} \beta \mid S$ for some $S \subseteq V \setminus \{\alpha, \beta\}$; requires *maximality* in mixed graphs;
- $\perp_{\mathcal{G}}$ defines a compositional graphoid.
- Independence structure *probabilistically representable:* $\exists P$ so that $A \perp\!\!\!\perp_P B \mid S \iff A \perp_{\mathcal{G}} \mid S$, ie P is *faithful*.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1996). An alternative Markov property for chain graphs. In Horvitz, E. and Jensen, F. V., editors, *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, pages 40–48. Morgan Kaufmann Publishers, San Francisco, CA.

Andersson, S. A., Madigan, D., and Perlman, M. D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Stat.*, 28(1):33–85.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, 8:204–218; 247–277.

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies*. Chapman and Hall, London, United Kingdom.

Drton, M. (2009). Discrete chain graph models. *Bernoulli*, 15(3):736–753.

Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, 17:333–353.

Geiger, D., Verma, T. S., and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks*, 20(5):507–534.

Hammersley, J. M. and Clifford, P. E. (1971). Markov fields on finite graphs and lattices. Unpublished manuscript.

Koster, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26(3):413–431.

Koster, J. T. A. (2002). Marginalisation and conditioning in graphical models. *Bernoulli*, 8:817–840.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, 20:491–505.

Meek, C. (1995). Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, UAI'95, pages 411–418, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Meek, C. (1996). *Selecting Graphical Models: Causal and Statistical Reasoning*. PhD thesis, Carnegie–Mellon University.

Pearl, J. (1986). A constraint–propagation approach to probabilistic reasoning. In Kanal, L. N. and Lemmer, J. F., editors, *Uncertainty in Artificial Intelligence*, pages 357–370, Amsterdam, The Netherlands. North-Holland.

Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.*, 30(1):145–157.

Richardson, T. S. and Spirtes, P. (2002). Ancestral graph Markov models. *The Annals of Statistics*, 30:963–1030.

Sadeghi, K. (2012). Stable mixed graphs. *Bernoulli*. To appear.

Sadeghi, K. and Lauritzen, S. (2012). Markov properties for mixed graphs. arXiv:1109.5909.

Studený, M. (1998). Bayesian networks from the point of view of chain graphs. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 496–503, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Studený, M. and Bouckaert, R. R. (1998). On chain graph models for description of conditional independence structures. *The Annals of Statistics*, 26:1434–1495.

Wermuth, N., Cox, D., and Marchetti, G. (2009). Covariance chains. *Bernoulli*, 12:841–862.