Markov properties for directed acyclic graphs

Aarhus University, Fall 2003, Lecture 2 and 3

Steffen L. Lauritzen, Aalborg University

Directed acyclic graphs

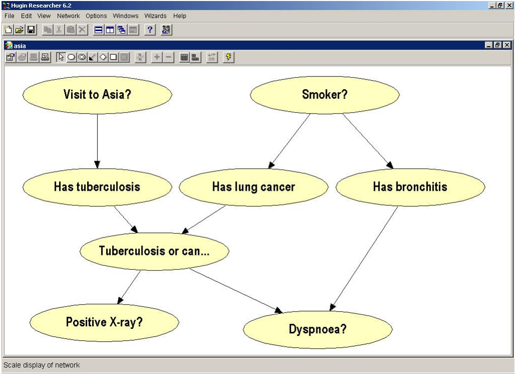
A *directed acyclic graph* $\mathcal{D}$ over a finite set V is a graph with all edges being directed, no multiple edges, and no directed cycles.

Absence of directed cycles means that, *following arrows in the graph, it is impossible to return to any point.*

Graphical models based on DAGs have proved fundamental and useful in a wealth of interesting applications, including expert systems, genetics, complex biomedical statistics, causal analysis, and machine learning.

The remainder of this course will give a glimpse into these applications.

Example of a directed graphical model



Factorisation with respect to a DAG

A probability distribution P over $\mathcal{X} = \mathcal{X}_V$ *factorizes* over a DAG $\mathcal{D}$ if it has density f w.r.t. a product measure $\mu = \otimes_{v \in V} \mu_v$, where f has the form

$$\text{(DF)}: \quad f(x) = \prod_{v \in V} k_v(x_v \mid x_{\mathrm{pa}(v)})$$

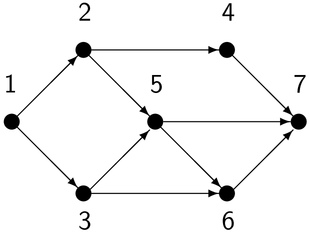
where $k_v \geq 0$ and $\int_{\mathcal{X}_v} k_v(x_v \mid x_{\mathrm{pa}(v)})\, d\mu_v(x_v) = 1$.

(DF) *is equivalent to* (DF*), where

$$(\text{DF}^*): \quad f(x) = \prod_{v \in V} f(x_v \mid x_{\mathrm{pa}(v)}),$$

i.e. it follows from (DF) that *k_v in fact are conditional densities.* Proof by induction!

Example of DAG factorization



The above graph corresponds to the factorization

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

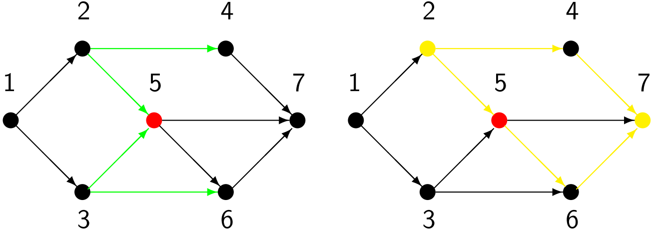
Separation in DAGs

A *trail* τ from node a to node b in a DAG $\mathcal{D}$ is *blocked* by S if it contains a node $n \in \tau$ such that

- either $n \in S$ and edges of τ do not meet head-to-head at n, or
- n and all its descendants are not in S, and edges of τ meet head-to-head at n.

A trail that is not blocked is *active.* Two subsets A and B of nodes are *d-separated* by S if all trails from A to B are blocked by S. We write $A \perp_{\mathcal{D}} B \mid S$.

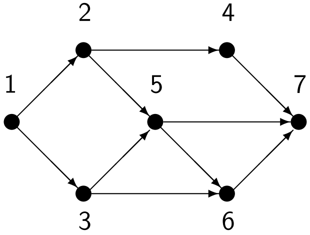
Separation by example



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

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

Returning to example



Hence $4 \perp_{\mathcal{D}} 6 \mid 3, 5$, but it is *not* true that $4 \perp_{\mathcal{D}} 6 \mid 5$ nor that $4 \perp_{\mathcal{D}} 6$.

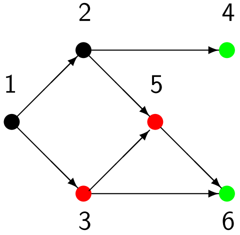
Alternative equivalent separation

1. Reduce to subgraph induced by ancestral set of $A \cup B \cup S$
2. Add undirected edges between unmarried parents in this subgraph
3. Drop directions on all edges. Process 2 and 3 is known as *moralization* and result is *moral graph*.
4. Say that *S m-separates A from B* and write $A \perp_m B \mid S$ if and only if S separates A from B in this undirected graph.

It then holds that *$A \perp_m B \mid S$ if and only if $A \perp_{\mathcal{D}} B \mid S$.*

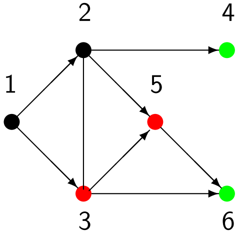
Sometimes (but not always) easier to use.

Forming ancestral set



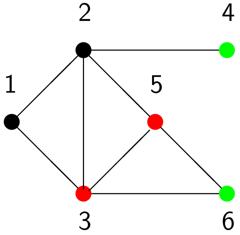
The subgraph induced by all ancestors of nodes involved in the query $4 \perp_m 6 \,|\, 3, 5$?

Adding links between unmarried parents



Adding an undirected edge between 2 and 3 with common child 5 in the subgraph induced by all ancestors of nodes involved in the query $4 \perp_m 6 \,|\, 3, 5$?

Dropping directions



Since $\{3, 5\}$ separates 4 from 6 in this graph, we can conclude that $4 \perp_m 6 \,|\, 3, 5$

Properties of d-separation

It holds for any DAG $\mathcal{D}$ that $\perp_{\mathcal{D}}$ *satisfies graphoid axioms*.

Clearly, this is then also true for $\perp_m$.

To show this is true, it is sometimes easy to use $\perp_m$, sometimes $\perp_{\mathcal{D}}$.

For example, (S2) is trivial for $\perp_{\mathcal{D}}$, whereas (S5) is trivial for $\perp_m$.

So, equivalence of $\perp_{\mathcal{D}}$ and $\perp_m$ is useful.

Local and global directed Markov properties

A semigraphoid relation $\perp_\sigma$ satisfies

(DL) *the local Markov property* if

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

(DG) *the global Markov property* if

$$A \perp_{\mathcal{D}} B \mid S \implies A \perp_\sigma B \mid S.$$

For a DAG it holds that (DG) $\iff$ (DL) also if $f(x) = 0$ is allowed.

There is also a pairwise property (DP). But it is less natural, it holds that (DL) $\implies$ (DP), but not conversely unless $f(x) > 0$.

Markov properties and factorization

In contrast to the undirected case, it holds for all f that (DF) holds if and only if $\perp\!\!\!\perp$ satisfies (DG), so global Markov property and factorisation property is equivalent:

Thus, in the directed case

$$\text{(DF)} \iff \text{(DG)} \iff \text{(DL)}$$

and this is true whether $f > 0$ or not.

Further properties

If P factorizes over $\mathcal{D}$, it factorizes over the moralised graph $\mathcal{D}^m$

This is seen directly from the factorization:

$$f(x) = \prod_{v \in V} f(x_v \mid x_{\mathrm{pa}(v)}) = \prod_{v \in V} \psi_{\{v\} \cup \mathrm{pa}(v)}(x),$$

since $\{v\} \cup \mathrm{pa}(v)$ are all complete in $\mathcal{D}^m$.

Hence if P satisfies any of the directed Markov properties w.r.t. $\mathcal{D}$, it satisfies all Markov properties for $\mathcal{D}^m$, so e.g. *any Markov chain is also a Markov field.*

Perfect DAGs

A DAG $\mathcal{D}$ is *perfect* if all parents are married. For a perfect DAG $\mathcal{D}$:

P satisfies (DG) w.r.t $\mathcal{D}$ if and and only if it satisfies (G) w.r.t. its skeleton $\sigma(\mathcal{D})$.

The *skeleton* is the undirected graph obtained from $\mathcal{D}$ by ignoring directions.

An undirected graph $\mathcal{G}$ can be oriented as a perfect DAG if and only if $\mathcal{G}$ is chordal, or *triangulated*, i.e. has *rigid circuits*,

An undirected graph is *chordal* if all cycles of length ≥ 4 have chords. Also known as *decomposable*.

Ancestral marginals

Consider a DAG $\mathcal{D}$ and an *ancestral subset* $A \subseteq V$.

If P factorizes w.r.t. $\mathcal{D}$, it factorizes w.r.t. $\mathcal{D}_A$.

Proof by induction, using that if A is ancestral and $A \neq V$, there is a terminal vertex v_0 with $v_0 \notin A$.

It now follows, that *if P factorizes w.r.t. $\mathcal{D}$:*

$$A \perp_m B \mid S \implies A \perp\!\!\!\perp B \mid S.$$

The equivalence of (DF), (DG) and (DL) now follows easily.

Only difficult proof of results on overheads is the equivalence of $\perp_{\mathcal{D}}$ and $\perp_m$. Proof in Lauritzen (1996) needs to allow self-intersecting paths to be correct.