Markov properties for directed graphs

Steffen Lauritzen, University of Oxford

Graphical Models, Lecture 10, Michaelmas Term 2010

November 8, 2010

Consider a *directed acyclic graph* $\mathcal{D}$ and associate for every vertex a random variable X_v. Consider now the equation system

$$X_v \leftarrow \alpha_v^\top X_{\mathrm{pa}(v)} + \beta_v + U_v, v \in V \tag{1}$$

where $U_v, v \in V$ are independent random disturbances with $U_v \sim \mathcal{N}(0, \sigma_v^2)$.

Such an equation system is known as a *recursive structural equation system*.

Structural equation systems are used heavily in social sciences and in economics. The term *structural* refers to the fact that the equations are assumed to be *stable under intervention* so that fixing a value of x_v^* would change the system only by removing the line in the equation system (1) defining x_v^*.

A recursive structural equation system defines a multivariate Gaussian distribution with joint density

$$\begin{aligned} f(x \mid \alpha, \sigma) &= \prod_v (2\pi)^{-1/2} \sigma_v^{-1} e^{-\frac{(x_v - \alpha_v^\top x_{\mathrm{pa}(v)} - \beta_v)^2}{2\sigma_v^2}} \\ &= (2\pi)^{-|V|/2} \left(\prod_v \sigma_v^{-1} \right) \\ &\quad \times e^{-\sum_v \frac{(x_v - \alpha_v^\top x_{\mathrm{pa}(v)} - \beta_v)^2}{2\sigma_v^2}}, \end{aligned}$$

from which the joint concentration matrix K can easily be derived.

Consider the system

$$\begin{aligned} X_1 &\leftarrow U_1 \\ X_2 &\leftarrow U_2 \\ X_3 &\leftarrow \alpha_{31}X_1 + U_3 \\ X_4 &\leftarrow \alpha_{42}X_2 + \alpha_{43}X_3 + U_4. \end{aligned}$$

The quadratic expression in the exponent becomes

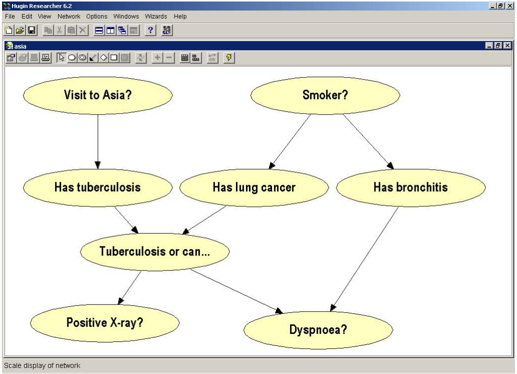
$$\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} + \frac{(x_3 - \alpha_{31}x_1)^2}{\sigma_3^2} + \frac{(x_4 - \alpha_{42}x_2 - \alpha_{43}x_3)^2}{\sigma_4^2}.$$

A *directed acyclic graph* $\mathcal{D}$ over a finite set V is a simple graph with all edges directed and *no directed cycles.* We use DAG for brevity.

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.

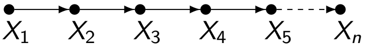
Example of a directed graphical model



A semigraphoid relation $\perp_\sigma$ satisfies *the local Markov property* (L) w.r.t. a directed acyclic graph $\mathcal{D}$ if

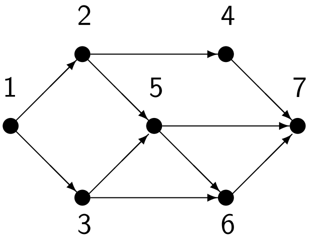
$$\forall \alpha \in V : \alpha \perp_\sigma \{\mathrm{nd}(\alpha) \setminus \mathrm{pa}(\alpha)\} \mid \mathrm{pa}(\alpha).$$

Here $\mathrm{nd}(\alpha)$ are the *non-descendants* of α.
A well-known example is a Markov chain:



with $X_{i+1} \perp\!\!\!\perp (X_1, \ldots, X_{i-1}) \mid X_i$ for $i = 3, \ldots, n$.

Local directed Markov property



For example, the local Markov property says
$4 \perp_\sigma \{1,3,5,6\} \mid 2$,
$5 \perp_\sigma \{1,4\} \mid \{2,3\}$
$3 \perp_\sigma \{2,4\} \mid 1$.

Suppose the vertices V of a DAG $\mathcal{D}$ are *well-ordered* in the sense that they are linearly ordered in a way which is compatible with $\mathcal{D}$, i.e. so that

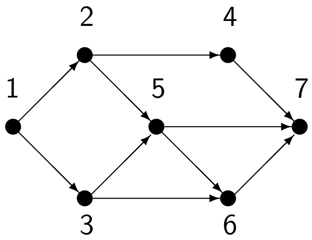
$$\alpha \in \mathrm{pa}(\beta) \Rightarrow \alpha < \beta.$$

We then say semigraphoid relation $\perp_\sigma$ satisfies the *ordered Markov property* (O) w.r.t. a well-ordered DAG $\mathcal{D}$ if

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

Here $\mathrm{pr}(\alpha)$ are the *predecessors* of α, i.e. those which are before α in the well-ordering..

Ordered Markov property



The numbering corresponds to a well-ordering. The ordered Markov property says for example

$4 \perp_\sigma \{1,3\} \mid 2,$
$5 \perp_\sigma \{1,4\} \mid \{2,3\}$
$3 \perp_\sigma \{2\} \mid 1.$

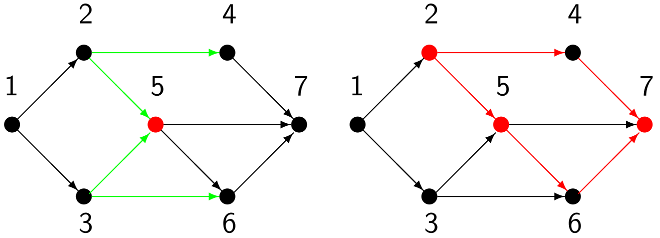
Separation in DAGs

A *trail* τ from vertex α to vertex β in a DAG $\mathcal{D}$ is *blocked* by S if it contains a vertex $\gamma \in \tau$ such that

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

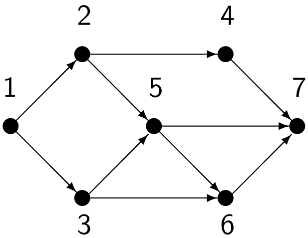
A trail that is not blocked is *active.* Two subsets A and B of vertices are *d-separated* by S if all trails from A to B are blocked by S. We write $A \perp_{\mathcal{D}} B \mid S$.

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 $4 \perp_{\mathcal{D}} 6$.

Equivalence of Markov properties

A semigraphoid relation $\perp_\sigma$ satisfies the *global Markov property* (G) w.r.t. $\mathcal{D}$ if

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

It holds for any DAG $\mathcal{D}$ and any semigraphoid relation $\perp_\sigma$ that all directed Markov properties are equivalent:

$$(\mathrm{G}) \iff (\mathrm{L}) \iff (\mathrm{O}).$$

There is also a pairwise property (P), but it is less natural than in the undirected case and it is weaker than the others.

A probability distribution P over $\mathcal{X} = \mathcal{X}_V$ *factorizes* over a DAG $\mathcal{D}$ if its density or probability mass function f has the form

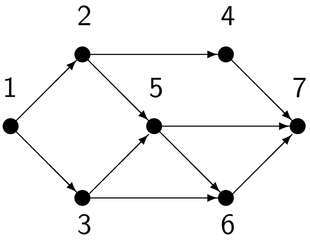
$$(\mathrm{F}): \quad f(x) = \prod_{v \in V} k_v(x_v \,|\, x_{\mathrm{pa}(v)})$$

where $k_v \geq 0$ and $\int_{\mathcal{X}_v} k_v(x_v \,|\, x_{\mathrm{pa}(v)}) \, \mu_v(dx_v) = 1$.
(F) *is equivalent to* (F*), where

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

i.e. it follows from (F) that *k_v in fact are conditional densities/pmf's.* Proof by induction!

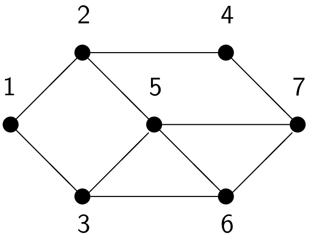
Example of DAG factorization



The above graph corresponds to the factorization

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

Contrast with undirected factorization



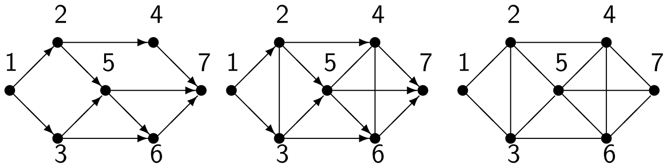
Factors ψ are typically not normalized as conditional probabilities:

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

In the directed case it is essentially *always true that (F) holds if and only if* $\perp\!\!\!\perp_P$ *satisfies (G)*,
so all directed Markov properties are equivalent to the factorization property!

$$(F) \iff (G) \iff (L) \iff (O).$$

The *moral graph* $\mathcal{D}^m$ of a DAG $\mathcal{D}$ is obtained by adding undirected edges between unmarried parents and subsequently dropping directions, as in the example below:



Undirected factorizations

If P factorizes w.r.t. $\mathcal{D}$, it factorizes w.r.t. 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 $\mathcal{D}^m$.*

Note the concentration matrix of the linear structural system considered:

$$K = \begin{pmatrix} \frac{1}{\sigma_1^2} + \frac{\alpha_{31}^2}{\sigma_3^2} & 0 & \frac{-\alpha_{31}}{\sigma_3^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} + \frac{\alpha_{42}^2}{\sigma_4^2} & \frac{\alpha_{42}\alpha_{43}}{\sigma_4^2} & \frac{-\alpha_{42}}{\sigma_4^2} \\ \frac{-\alpha_{31}}{\sigma_3^2} & \frac{\alpha_{42}\alpha_{43}}{\sigma_4^2} & \frac{1}{\sigma_3^2} + \frac{\alpha_{43}^2}{\sigma_4^2} & \frac{-\alpha_{43}}{\sigma_4^2} \\ 0 & \frac{-\alpha_{42}}{\sigma_4^2} & \frac{-\alpha_{43}}{\sigma_4^2} & \frac{1}{\sigma_4^2} \end{pmatrix}.$$

Hence this is Markov wrt the graph with cliques $\{1,3\},\{2,3,4\}$

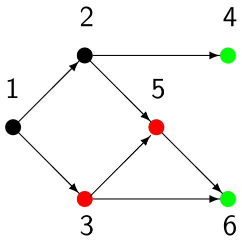
Alternative equivalent separation

To resolve query involving three sets A, B, S:

1. Reduce to subgraph induced by ancestral set $\mathcal{D}_{\mathrm{An}(A \cup B \cup S)}$ of $A \cup B \cup S$;
2. Moralize to form $(\mathcal{D}_{\mathrm{An}(A \cup B \cup S)})^m$;
3. 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.

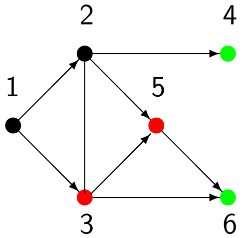
It then holds that *$A \perp_m B \mid S$ if and only if $A \perp_{\mathcal{D}} B \mid S$.*
Proof in Lauritzen (1996) needs to allow self-intersecting paths to be correct.

Forming ancestral set



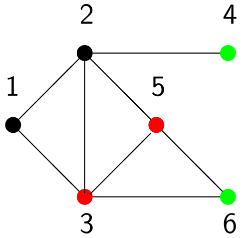
The subgraph induced by all ancestors of nodes involved in the query $4 \perp_m 6 \mid 3, 5$?

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 $4 \perp_m 6 \mid 3, 5$?

Dropping directions



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

Two DAGS $\mathcal{D}$ and $\mathcal{D}'$ are *Markov equivalent* if the separation relations $\perp_{\mathcal{D}}$ and $\perp_{\mathcal{D}'}$ are identical.

$\mathcal{D}$ and $\mathcal{D}'$ are equivalent if and only if:

1. $\mathcal{D}$ and $\mathcal{D}'$ have same *skeleton* (ignoring directions)
2. $\mathcal{D}$ and $\mathcal{D}'$ have same unmarried parents

so

Contrast with undirected case, where *two undirected graphs are Markov equivalent if and only if they are identical.*

Markov equivalence of directed and undirected graphs

A DAG $\mathcal{D}$ is *Markov equivalent* to an undirected $\mathcal{G}$ if the separation relations $\perp_{\mathcal{D}}$ and $\perp_{\mathcal{G}}$ are identical.
This happens if and only if $\mathcal{D}$ is perfect and $\mathcal{G} = \sigma(\mathcal{D})$. So, these are all equivalent

but not equivalent to