Expressive Power of Graphical Models

Draft; version from 10th January 2018; some things may change – don’t print

Michael Gutmann

Probabilistic Modelling and Reasoning (INFR11134)
School of Informatics, University of Edinburgh

Spring semester 2018
Recap

- Need for efficient representation of probabilistic models
- Restrict the number of interacting variables by making independence assumptions
- Restrict the form of interaction by making parametric family assumptions.
- Directed and undirected graphical models to represent independencies (I-map)
- Equivalences between independencies (Markov properties) and factorisation
- Rules for reading independencies from the graph that hold for all distributions that factorise over the graph.
1. Minimal independency maps

2. (Lossy) conversion between directed and undirected minimal I-maps
1. Minimal independency maps
   - Definition of I-maps, the goal of a perfect maps
   - Construction of undirected I-maps and their uniqueness
   - Construction of directed I-maps and their non-uniqueness
   - Equivalence of I-maps (I-equivalence)

2. (Lossy) conversion between directed and undirected minimal I-maps
Minimial I-maps

- A graph is an independency map for a set of independencies $\mathcal{I}$ if the independencies asserted by the graph are part of $\mathcal{I}$.
- Criterion is that the independency assertions are true.
- Is not concerned with the number of independency assertions.
- Full graph does not make any assertions. Empty set is trivially part of $\mathcal{I}$, so that the full graph is trivially an I-map.
- Minimal I-map: graph such that if you remove an edge (more independence assumptions), the graph is not an I-map any more.
- We want the graph to represent as many true independencies as possible: graph is sparser, and thus more informative, easier to understand, and facilitates learning and inference.
- If the graph represents all independencies in $\mathcal{I}$, the graph is said to be a perfect map (P-map). (May be hard to find and will not always exist!)
Example

- Let $p(x_1, x_2, x_3, x_4) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_4)$

- Minimal I-map:

- Non-minimal I-map ($x_1 \sim x_3$ edge could be removed)

- Not an I-map (wrongly claims $x_1 \perp \perp x_2, x_3$)
Example

Let $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)p(x_5|x_2)$

Minimal I-map:

(Non-minimal) I-map
($x_1 \rightarrow x_4$ could be removed)

Not an I-map
(wrongly claims $x_4 \perp \perp x_3$)
Constructing undirected minimal I-maps

Given a random variables $\mathbf{x} = (x_1, \ldots, x_d)$ with positive distribution $p > 0$

- Approaches based on pairwise and local Markov property
- Both yield same (unique) graph.
- For local Markov property approach: For each node:
  1. determine its Markov blanket $\text{MB}(x_i)$: minimal set of nodes $U$ such that
     
     $$x_i \perp \perp \text{all variables} \setminus (x_i \cup U) \mid U$$

     with respect to $p$.
  2. as we know that $x_i$ and $\text{MB}(x_i)$ must be neighbours in the graph: Connect $x_i$ to all nodes in $\text{MB}(x_i)$

- We need $p > 0$ because otherwise local independencies may not imply global ones.
Constructing directed minimal I-maps

Given a distribution $p$.

- We can use the ordered Markov property to derive a directed graph that is an I-map for $\mathcal{I}(p)$.

\[
x_{i} \perp \perp \text{pre}_i \setminus \text{pa}_i \mid \text{pa}_i
\]

- Procedure is exactly the same as the one used to simplify the factorisation obtained by the chain rule

1. Assume an ordering of the variables. Denote the ordered random variables by $x_1, \ldots, x_d$.
2. For each $i$, find a minimal subset of variables $\pi_i \subseteq \text{pre}_i$ such that

\[
x_{i} \perp \perp \text{pre}_i \setminus \pi_i \mid \pi_i
\]

holds in $\mathcal{I}(p)$.
3. Construct a graph with parents $\text{pa}_i = \pi_i$. 

Directed minimal I-maps are not unique

Consider $p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z)$

For ordering $(a, z, q, e, h)$

For ordering $(e, h, q, z, a)$

- Directed I-maps are not unique
- Different directed I-maps may make different independence assertions.
- Minimal I-maps of $\mathcal{I}(p)$ may not represent all independencies that hold for $p$, but generally only a subset of them.
I-equivalence for directed graphs

How do we determine whether two directed graphs make the same independence assertions (that they are “I-equivalent”)?

From d-separation: what matters is

- which node is connected to which irrespective of direction (skeleton)
- the set of collider (head-to-head) connections

| Connection | $p(x, y)$ | $p(x, y|z)$ |
|------------|-----------|-------------|
| $x \rightarrow z \rightarrow y$ | $x \not\rightarrow y$ | $x \perp y | z$ |
| $x \leftarrow z \rightarrow y$ | $x \not\rightarrow y$ | $x \perp y | z$ |
| $x \rightarrow z \leftarrow y$ | $x \perp y$ | $x \not\rightarrow y | z$ |
The situation $x \perp \perp y$ and $x \not\perp \perp y \mid z$ can only happen if there is no “covering edge” $x \rightarrow y$ or $x \leftarrow y$.

Colliders without covering edge are called “immoralities”.

Theorem: For two directed graphs $G_1$ and $G_2$:
$G_1$ and $G_2$ are I-equivalent $\iff G_1$ and $G_2$ have the same skeleton and the same set of immoralities.
Example

Not I-equivalent because of skeleton mismatch:

$G_1$: 

$G_2$: 

Example

Not I-equivalent because of immoralities mismatch:
I-equivalent (same skeleton, same immoralities):

$G_1$:  
\[ a \rightarrow q \rightarrow e \rightarrow h \rightarrow z \]

$G_2$:  
\[ a \rightarrow q \rightarrow e \rightarrow h \rightarrow z \]
For undirected graphs, I-map is unique.
Different graphs make different independence assertions.
Equivalence question does not come up.
1. Minimal independency maps

2. (Lossy) conversion between directed and undirected minimal I-maps
   - Moralisation for directed $\rightarrow$ undirected I-map
   - Example of non-existence of undirected perfect map
   - Triangulation for undirected $\rightarrow$ directed I-map
   - Example of non-existence of directed perfect map
   - Strengths and weaknesses of directed and undirected graphical models
Directed to undirected graphical model

Goal: undirected I-Map. Assume directed I-map $G$ given

- Probabilistic models factorises according to $G$ as

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} p(x_i | \text{pa}_i)$$

- Write each $p(x_i | \text{pa}_i)$ as factor $\phi_i(x_i, \text{pa}_i)$:

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} \phi_i(x_i, \text{pa}_i)$$

Gibbs distribution with normalisation constant equal to one

- Graph operation: Form cliques for $(x_i, \text{pa}_i)$
Directed to undirected graphical model

Goal: undirected I-Map. Assume directed I-map $G$ given

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} p(x_i | \text{pa}_i) = \prod_{i=1}^{d} \phi_i(x_i, \text{pa}_i)$$

- Graph operation: Form cliques for $(x_i, \text{pa}_i)$
- Remove arrows, and add edges between all the parents of $x_i$.
- Conversion from directed to undirected graphical model is called “moralisation”. Obtained undirected graph is the “moral graph” of $G$.
- Process above is equivalent to using the directed graph to determine the Markov blanket for each $x_i$. 
Example

Goal: Undirected I-map for
\[ p(a, z, q, e, h) = p(a)p(z)p(q|a, z)p(e|q)p(h|z) \]

Given: directed I-map

Moral graph:

Note: In undirected I-map, we do not have \( a \perp \perp z \). We lost that information.

Minimal I-maps of \( \mathcal{I}(p) \) may not represent all independencies that hold for \( p \), but generally only a subset of them.
Simpler example

Goal: Undirected I-map for \( p(x, y, z) = p(x)p(y)p(z|x, y) \)

Given: directed I-map

Only possible undirected I-map is full graph

No undirected I-map representing \( \mathcal{I} = \{ x \perp \perp y, x \not\perp y | z \} \)
Goal: directed I-Map. Assume undirected I-map $H$ given

- We can use the approach based on the local Markov property
- Read required independencies from the undirected graph
- Typically results in directed graphs that are larger than the undirected graph
- Directed graph will not have any immoralities for proof, see e.g. theorem 4.10 in Koller and Friedman’s book
- Results in chordal/triangulated graphs (longest loop without shortcuts is a triangle).
Example

Goal: Directed I-map for
\[ p(x, y, z, u) \propto \phi_1(x, y)\phi_2(y, z)\phi_3(z, u)\phi_4(u, x) \]

Given: undirected I-map
\[
\begin{align*}
x & \perp z \mid u, y \\
u & \perp y \mid x, z
\end{align*}
\]

Directed I-map (with ordering: \( x, y, u, z \))
\[
\begin{align*}
x & \perp z \mid u, y \\
u & \not\perp y \mid x, z
\end{align*}
\]

We lost information with the conversion. No directed I-map representing \( \mathcal{I} = \{x \perp z \mid u, y, u \perp y \mid x, z\} \)
Strengths and weaknesses

- Both directed and undirected graphical models have strengths and weaknesses.
- Some independencies are more easily represented with directed graphs, others with undirected graphs.
- Undirected graphs suitable when interactions are symmetrical and when there is no natural ordering of the variables, but cannot represent “explaining away” scenario (colliders).
- Directed graphs suitable when we have an idea of the data generating process (e.g. what is “causing” what, ancestral sampling), but they may force directionality where there is none, yielding unintuitive graphs (see triangulation).
- Combine individual strengths with mixed/partially directed graphs.
Program recap

1. Minimal independency maps
   - Definition of I-maps, the goal of a perfect maps
   - Construction of undirected I-maps and their uniqueness
   - Construction of directed I-maps and their non-uniqueness
   - Equivalence of I-maps (I-equivalence)

2. (Lossy) conversion between directed and undirected minimal I-maps
   - Moralisation for directed $\rightarrow$ undirected I-map
   - Example of non-existence of undirected perfect map
   - Triangulation for undirected $\rightarrow$ directed I-map
   - Example of non-existence of directed perfect map
   - Strengths and weaknesses of directed and undirected graphical models